UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

GUSTAVO MACHADO ÁLVARES DE LIMA

Estudos estruturais de enzimas da via de biossíntese de folatos de
Xanthomonas albilineans para o desenvolvimento de novos candidatos a
agroquímicos para a cultura de cana-de-açúcar

São Carlos

2017

GUSTAVO MACHADO ÁLVARES DE LIMA

Estudos estruturais de enzimas da via de biossíntese de folatos de
Xanthomonas albilineans para o desenvolvimento de novos candidatos a
agroquímicos para a cultura de cana-de-açúcar

Tese apresentada ao Programa de Pós-
Graduação em Física no Instituto de
Física de São Carlos da Universidade de
São Paulo, para obtenção do título de
Doutor em Ciências.

Área de Concentração: Física Aplicada
Opção: Física Biomolecular
Orientador: Prof. Dr. Rafael V. C. Guido

Versão Original

São Carlos

2017

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

Machado Alvares de Lima, Gustavo
Estudos estruturais de enzimas da via de
biossíntese de folatos de *Xanthomonas albilineans*
para o desenvolvimento de novos candidatos a
agroquímicos para a cultura de cana-de-açúcar /
Gustavo Machado Alvares de Lima; orientador Rafael
Victório Carvalho Guido -- São Carlos, 2017.
153 p.

Tese (Doutorado - Programa de Pós-Graduação em
Física Básica) -- Instituto de Física de São Carlos,
Universidade de São Paulo, 2017.

1. Cristalografia. 2. Escaldadura das folhas. 3.
Cana-de-açúcar. 4. *Xanthomonas albilineans*. 5.
Biologia estrutural. I. Victório Carvalho Guido,
Rafael, orient. II. Título.

FOLHA DE APROVAÇÃO

Gustavo Machado Alvares de Lima

Tese apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Doutor em Ciências. Área de Concentração: Física Aplicada - Opção: Física Biomolecular.

Aprovado(a) em: 22/09/2017

Comissão Julgadora

Dr(a). Rafael Victório Carvalho Guido

Instituição: (IFSC/USP)

Dr(a). João Renato Carvalho Muniz

Instituição: (IFSC/USP)

Dr(a). Marcio Vinícius Bertacine Dias

Instituição: (ICB/USP)

Dr(a). Marie Anne van Sluys

Instituição: (IB/USP)

Dr(a). Artur Torres Cordeiro

Instituição: (CNPEM/Campinas)

Aos meus pais, João Carlos e Eliana.
Que eles sintam por mim o orgulho que eu sinto por eles.

AGRADECIMENTOS

Aos meus pais, por 30 anos de uma fé inabalável em mim. Que as minhas conquistas sejam sempre as deles.

Aos meus irmãos Guilherme e Maria Clara, pelo apoio e amor.

Ao Professor Rafael V. C. Guido, pela orientação e apoio em tempos conturbados.

À Fernanda C. Costa, quem sempre foi um exemplo e uma amiga, por ter sido sempre a voz da razão e nem sempre a voz da amizade. Pelos longos anos de uma amizade sem igual. Foi por muitas vezes a minha força.

À Anna Carolina C. Aguiar pela amizade, pela forma incrível de sua empatia. Pelo jeito simples e doce que sempre teve comigo, fundamentais principalmente no último ano de doutorado. Não seria possível sem.

Aos Professores Antônio José da Costa Filho, Richard Charles Garratt e Glaucius Oliva, pelos exemplos de professor e cientista. Fazerem parte da minha formação desde o início.

Aos Professores Paul Michels e Malcolm Walkinshaw, pela oportunidade de viver e aprender em uma realidade que não a minha.

Aos amigos Fernando V. Maluf, Maria Amélia e Fábio Dotta, pela amizade dentro e fora da vida científica, pela parceria e conselhos, por serem minha família.

Aos amigos que fiz durante a pós-graduação. À Renata Bueno, por me entender e me criticar. Ao Andrew Oliveira, a pessoa mais altruísta que convivi durante o doutorado. Partilho esta conquista com eles.

Aos que enfrentaram a improvável tarefa de ir até o final da primeira turma de Ciências Físicas e Biomoleculares, Hilde Buzzá, Paula Amaral, Victor Caldas, Ivan Silva e Rodrigo Berté.

Aos técnicos e funcionários do IFSC por todo o suporte à realização dos experimentos, em especial ao Dr. Humberto D'Muniz Pereira.

Ao serviço de pós-graduação do IFSC, em especial Patrícia G Ferreira, Ricardo Vital e Sílvio C Athayde, pelo empenho incomum e muito valioso.

Aos funcionários da biblioteca do IFSC, especialmente Neusa Aguiar, que a calma e prestatividade foram fundamentais para a conclusão do doutorado.

Aos laboratórios de dentro e fora do IFSC que foram suporte à minha pesquisa.

Às agências de fomento FAPESP e CAPES, pelo apoio financeiro.

“Já sei que, se para ser
o Homem, escolher pudera,
Ninguém o papel quisera
do sofrer e padecer.
Todos quiseram fazer
o de mandar e reger
Sem advertir e sem ver
que, em ato tão singular,
Aquilo é representar
mesmo ao pensar que é viver”

Pedro Calderón de la Barca

RESUMO

LIMA, G. M. A. Estudos estruturais de enzimas da via de biossíntese de folatos de *Xanthomonas albilineans* para o desenvolvimento de novos candidatos a agroquímicos para a cultura de cana-de-açúcar. 2017. 153 p. Tese (Doutorado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

O destaque da cana-de-açúcar como fonte de energia renovável tem motivado a busca por melhorias no processo de cultivo e processamento da planta, buscando o aumento da produtividade dos canaviais. Dentre os fatores limitantes para o aumento da produção de cana-de-açúcar, destaca-se a ocorrência de fitodoenças, como por exemplo a escaudadura das folhas, causada pela bactéria *Xanthomonas albilineans*. Essa doença causa a diminuição do valor energético do caldo extraído da planta, e a ausência de tratamento químico ou biológico acarreta na necessidade de reforma precoce dos canaviais, causando perdas significativas para o agronegócio brasileiro. Portanto, existe uma grande necessidade no desenvolvimento de novas moléculas capazes de atuar como defensivos agrícolas com eficiência e especificidade. Neste trabalho, duas estratégias foram empregadas para a identificação de moléculas como potenciais candidatos a agroquímicos. Na primeira, avaliou-se o potencial inibidor de moléculas de origem natural ou sintética em ensaios fenotípicos contra a cultura de *X. albilineans*, identificando dois derivados, um de nitrotiofeno e um de nitrofurano, capazes de inibir o crescimento bacteriano. Na segunda estratégia, duas enzimas da via de biossíntese de folatos, 2-amino-4-hidroxi-6-hidroximetil di-hidropterina pirofosfoquinase (*XaHPPK*) e 7,8-di-hidroneopterina aldolase (*XaDHNA*), foram selecionadas como alvos moleculares. Métodos em biologia molecular e biologia estrutural foram empregados para a obtenção de proteínas puras e solúveis, visando-se a elucidação das estruturas tridimensionais e caracterização bioquímica dos alvos. Estudos de desnaturação térmica foram conduzidos para determinação da afinidade da *XaHPPK* pelo substrato natural 7,8-di-hidropterina ($K_d = 97 \pm 3 \mu\text{M}$). Os experimentos de biologia estrutural resultaram na determinação da estrutura da *XaHPPK* a 2,3 Å de resolução (PDB ID, 5VSP) com o grupo espacial P2₁2₁2₁. A obtenção da estrutura a alta resolução permitiu a aplicação de estratégias computacionais para a descoberta de potenciais inibidores. Os estudos da *XaDHNA* possibilitaram a coleta de um conjunto de dados à 3,5 Å de resolução, no grupo espacial I422. Os experimentos e resultados obtidos neste doutorado ampliam o conhecimento sobre a via de biossíntese de folatos em *X. albilineans* e abrem o caminho para a descoberta de novos inibidores antibacterianos utilizando técnicas baseadas na estrutura do alvo.

Palavras-chave: Cana-de-açúcar. Escaudadura das folhas. *Xanthomonas albilineans*

ABSTRACT

LIMA, G. M. A. Estudos estruturais de enzimas da via de biossíntese de folatos de *Xanthomonas albilineans* para o desenvolvimento de novos candidatos a agroquímicos para a cultura de cana-de-açúcar. 2017. 153 p. Tese (Doutorado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

The importance of sugarcane as a source of renewable energy has pushed researchers to search for improvements in the process of cultivation and processing of the plant, looking after increase the productivity of sugarcane crops. Among known facts that limits sugarcane production, we highlight the occurrence of plant diseases, such as leaf scald, a bacterial disease caused by *Xanthomonas albilineans*. This disease decreases the energetic value of the broth extracted from the cane, and the absence of chemical or biological treatment entails the need for early refresh sugarcane plantations, causing significant losses for Brazilian agribusiness. Therefore, the economic interest in the development of new molecules capable of acting as agricultural defenses with efficiency and specificity is motivating. Acting on the essential biochemical pathways of the pathogen is a recurring strategy in the development of bioactive molecules. Thus, several inhibitor identification strategies were used. Firstly, the inhibitory potential of molecules from natural or synthetic origins were evaluated in phenotypic assays against the culture of *X. albilineans*, identifying two nitrothiophene and nitrofuran derivatives capable of inhibiting bacterial growth at 2 mM. In another strategy, the essential pathway of *X. albilineans* for folate biosynthesis was selected and two proteins of this pathway investigated: 2-amino-4-hydroxy-6-hydroxymethyl dihydropteridine pyrophosphokinase (*XaHPPK*) and 7,8-dihydroneopterin aldolase (*XaDHNA*). Both molecular biology and structural biology methods were used to obtain pure and soluble proteins, aiming the obtaining of three-dimensional structures and biochemical characterization of targets. Studies with the *XaHPPK* protein for kinetic characterization made possible to calculate the K_d apparent ($K_d = 97 \pm 3 \mu\text{M}$) for the substrate 7,8-dihydropterin. Structural biology experiments resulted in a 2.3 Å resolution structure in the space group $P2_12_12_1$, available in the PDB under the 5VSP identifier. Structure based drug design for *XaHPPK* structure, SBDD, were performed using ZINC15 database with hierarchical filters and the Glide docking software. The *XaDHNA* studies enabled the collection of a set of data at 3.5 Å resolution in the space group $I422$. The experiments and results obtained in this doctorate increase the knowledge about the folate biosynthesis pathway in *X. albilineans* and open the way for the discovery of new antibacterial inhibitors using structure based drug design.

Keywords: Sugarcane. Leaf Scald. *Xanthomonas albilineans*

LISTA DE FIGURAS

Figura 1 –	Expansão histórica do cultivo da cana-de-açúcar até sua chegada ao Brasil.	37
Figura 2 –	Distribuição geográfica da produção de cana-de-açúcar safra 2015/2016, ⁴ colorida com base na porcentagem da área cultivada.	38
Figura 3 –	Distribuição geográfica do total de cana-de-açúcar cultivado por Unidade da Federação, safra 2015/2016, ⁶ colorida com base na porcentagem da área cultivada.	39
Figura 4 –	Distribuições geográficas das culturas de cana-de-açúcar e escaldadura das folhas. ¹⁶	41
Figura 5 –	Sintomas da escaldadura das folhas em cana-de-açúcar. (A) Linhas esbranquiçadas na direção da nervura da folha. (B) Estágio avançado da doença no qual as folhas já estão necrosadas. (C) Clareamento parcial das folhas no estágio crônico da doença. (D) Linhas esbranquiçadas com algumas manchas avermelhadas. (E) Secção longitudinal de uma muda de cana-de-açúcar infectada, em estágio avançado da fitodoença. (F) Clorose das folhas com o aspecto avermelhado, forma mais característica da doença.....	42
Figura 6 –	Classificação taxonômica da <i>Xanthomonas albilineans</i>	44
Figura 7 –	(A) (1) Incorporação de uma molécula de pABA a uma molécula de MCA por uma ação em conjunto das proteínas PKS e NRPS do módulo Alb01 do cluster alb. (2) Incorporação de uma molécula de Cya pela NRPS do módulo Alb04, por um mecanismo trans-complementar de duas etapas: i. a adenilação da extremidade ácido α -carboxílico para a formação de um intermediário tioéster e ii. a fosforilação do oxigênio da amida na cadeia lateral. (3) O aminoácido ciano-L-alanina é incorporado ao esqueleto precursor da albicidina. (4) Uma molécula de pABA é ativada por uma NRPS, proveniente do módulo Alb01, e ligada ao produto da etapa anterior. (5) e (6) As duas últimas etapas são realizadas por duas NRPSs, NRPS-5 e NRPS-6, traduzidas a partir de sequências contidas no módulo Alb09, que incorporam os dois últimos resíduos de pMBA, pMBA-5 e pMBA-6,	

respectivamente, liberando a molécula de albicidina. (B) Representação tridimensional da estrutura da albicidina.	47
Figura 8 – Estrutura química do tetra-hidrofolato (THF).....	49
Figura 9 – Modificações na porção pterina precursora do THF pelas enzimas <i>XaDHNA</i> e <i>XaHPPK</i>	50
Figura 10 – Representação gráfica do cálculo da área fracional.....	63
Figura 11 – Modelo das interações entre a cauda de hexa-histidina e o metal divalente imobilizado, em uma matriz de agarose.	67
Figura 12 – Condições tamponantes utilizadas para os experimentos de DSF baseadas na proposta de NIESEN.	68
Figura 13 – Organograma do processo de obtenção dos dados do experimento de difração de raios X, processamento dos dados, refinamento do modelo tridimensional e deposição no banco de dados de estruturas PDB.....	72
Figura 14 – Distribuição das moléculas utilizadas para a triagem virtual frente à enzima <i>XaHPPK</i> , organizadas por massa molecular e LogP. As características utilizadas para a obtenção deste subconjunto foram: Similaridade tridimensional, sem grupos reativos (Clean), disponíveis para compra ainda que com pequeno tempo de espera (Wait OK), pH referência, qualquer carga e filtro de Tice.	73
Figura 15 – Curvas de crescimento de <i>X. albilineans</i> na presença de DMSO.	79
Figura 16 – Curvas de crescimento da <i>X. albilineans</i> na presença de antibióticos. (A) Curvas de crescimento na presença de tobramicina (B) Ajuste de Gompertz para os dados de tobramicina (C) Curvas de crescimento na presença de tetraciclina (D) Ajuste de Gompertz para os dados de tetraciclina.	80
Figura 17 – Estrutura molecular dos compostos com atividade inibitória frente a <i>Xanthomonas albilineans</i> . (A) Composto FG011 (B) Composto FG031.....	81
Figura 18 – Curvas de crescimento de <i>X. albilineans</i> . (A), (B) representam as curvas de controle positivo, negativo e negativo + DMSO. (C) Curvas de crescimento de <i>X. albilineans</i> na presença do composto FG011. (D) Curvas de crescimento de <i>X. albilineans</i> na presença do composto FG031. As	

concentrações dos compostos representam as concentrações utilizadas para a triagem inicial de atividade.	82
Figura 19 – Porcentagem de inibição dos compostos FG011 e FG031 em diferentes concentrações, calculadas a partir da área fracional sob a curva de crescimento.	82
Figura 20 – Mapa do vetor pET28a(+) com as regiões promotoras, repressoras, resistência a antibiótico e etiquetas de auxílio à purificação. As setas em verde e laranja representam quadros abertos de leitura, coloridos baseado no sentido de leitura, senso ou anti-senso em relação origem de replicação.	84
Figura 21 – Ilustração da estratégia de clonagem de ligação independente. (A1) Segmento gênico amplificado pela reação de PCR. A região em vermelho denota o segmento a ser digerido pela T4 DNA polimerase. (A2) Segmento gênico com as bordas coesivas geradas pela atividade exonuclease 3'-5' da T4 DNA polimerase na presença de dATP. (B1) Vetor de expressão. A região em vermelho denota o segmento a ser digerido pela T4 DNA polimerase. (B2) Vetor de expressão com as bordas coesivas geradas pela atividade exonuclease 3'-5' da T4 DNA polimerase na presença de dTTP. (C) Produto da reação de hibridização. As bases em vermelho representam as regiões com ausência da ligação fosfodiéster entre as bases, a ser realizada pela bactéria em vez da T4 DNA ligase. (D) Passos realizados para a obtenção de um clone com o plasmídeo contendo o segmento gênico de interesse.	86
Figura 22 – Eletroforese em gel de agarosa 1%. Segmentos gênicos nos diferentes vetores: pETM11, pETTrx-1a pETNus-a; amplificados por PCR.	87
Figura 23 – Cromatograma da Gel Filtração da XaHPPK e eletroforese em gel de poliacrilamida 12% com as frações coletadas durante a purificação da XaHPPK. (A) Fração eluída na primeira IMAC. (B) Fração após a clivagem da proteína fusionada com a TEV protease (C) Fração que não interage com a resina cromatográfica durante a segunda IMAC, contendo a proteína HPPK. (D) Fração eluída durante a segunda IMAC, contendo a TEV protease, a XaHPPK e a construção 6-His-Trx-XaHPPK remanescente	

da reação de clivagem. (E) Fração correspondente ao pico da gel filtração, contendo uma única forma oligomérica da <i>XaHPPK</i>	89
Figura 24 – Curvas de fluorescência do Sypro Orange no experimento de DSF para a proteína <i>XaHPPK</i>	90
Figura 25 – Diagrama de fases de cristalização. Em vermelho estão as mudanças sofridas pela gota de proteína durante um experimento de cristalização por difusão de vapor.....	91
Figura 26 – Esquema experimental da otimização da condição de cristalização para a <i>XaHPPK</i>	93
Figura 27 – Cristais obtidos durante o experimento de otimização da condição de cristalização e suas respectivas condições.	94
Figura 28 – Estatística das estruturas depositadas no banco de dados PDB. Os gráficos foram traçados com os fatores R e R_{free} em função da resolução da estrutura depositada, analisados em abril 2017.	98
Figura 29 – Alinhamento da sequência completa da <i>XaHPPK</i> com a estrutura depositada 5VSP. As regiões não modeladas estão representadas pelos espaços no alinhamento.....	103
Figura 30 – (A) Topologia da estrutura da <i>XaHPPK</i> gerada utilizando o PDBsum. ¹⁵⁸ (B) Representação da estrutura primária da proteína <i>XaHPPK</i> e suas respectivas estruturas secundárias.	103
Figura 31 – Representação da estrutura da <i>XaHPPK</i> com a disposição biológica, depositada no PDB sob o código 5VSP. (A) Diagrama de Ramachandran para a <i>XaHPPK</i> , gerada na ferramenta RAMPAGE. ¹⁵⁹ (B) Monômero da <i>XaHPPK</i> com detalhe da coordenação do Mg^{2+} pelos resíduos Aps92 e Asp94 e duas moléculas de água. (C) Visão superior do homo-tetrâmero da <i>XaHPPK</i> , sob um dos eixos de simetria de ordem 2 e visão lateral com detalhe da interface entre as cadeias A e B.	104
Figura 32 – Mapa dos contatos dos monômeros que compõem o homo-tetrâmero. As letras A – Z com fundo amarelo representam os contatos entre resíduos em interfaces não cristalográficas. As letras A – Z em itálico representam os contatos em interfaces cristalográficas. O símbolo # identifica contatos que estão em um eixo de simetria de ordem 2, no caso da <i>XaHPPK</i> . As	

letras em vermelho identificam os contatos a uma distância menor que 3.2 Å, enquanto letras em preto para distâncias entre 3.2 Å e 5.0 Å.	105
Figura 33 – Comparação entre estruturas homólogas da <i>XaHPPK</i> disponíveis no PDB. (A) Representação da conservação de resíduos (vermelho – conservado, branco – não conservado) entre as homólogas analisadas e RMSD médio entre as posições dos Cα 5VSP-homóloga (maior espessura na representação – maior RMSD médio). (B) Conservação dos resíduos do sítio ativo que interagem diretamente com o substrato 6HMPH2. (C) Alinhamento das sequências primárias das proteínas homólogas analisadas. (D) Comparação filogenética das todas as sequências com estrutura disponível no PDB, geradas pela ferramenta PhyloDendron.	106
Figura 34 – Curvas de desnaturação térmica da enzima <i>XaHPPK</i> na presença do substrato 6HMPH2. A concentração do substrato foi variada entre 20 mM e 100 nM.	108
Figura 35 – Ajuste dos dados de desnaturação térmica da enzima <i>XaHPPK</i> na presença do substrato 6HMPH2 utilizando o modelo de cooperação simples do substrato. ¹⁰⁸ O ponto de inflexão da curva é numericamente igual à constante de dissociação aparente $K_d = 97 \pm 3 \mu\text{M}$	109
Figura 36 – Representação da região definida para busca de interações no sítio ativo e os resíduos que a compõe.	113
Figura 37 – Resultado da aplicação da metodologia de acomodação do substrato 6HMPH2 na cavidade definida para busca de interações na proteína <i>XaHPPK</i> . (A) Sobreposição das estruturas cristalográfica e predita pelo Glide do substrato 6HMPH2. Em verde estão representados os resíduos mais importantes para o modo de ligação segundo o avaliador <i>GlideScore</i> . (B) Representações ortogonais do ligante em sua pose cristalográfica e predita.	114
Figura 38 – Modos de ligação preditos pela ferramenta Glide. (A1-A2) Detalhes das interações entre o composto melhor classificado pelo <i>GlideScore</i> e os resíduos Phe50, Asp92, Asp94 e Phe120. (B-C) Comparação das interações de duas poses diferentes de um mesmo composto. A interação entre uma	

das hidroxilas do grupo catecol com o resíduo Asn52 influencia os termos de interação de hidrogênio e Van der Waals no cálculo do <i>GlideScore</i>	117
Figura 39 – Cromatograma da Gel Filtração da <i>XaDHNA</i> e eletroforese em gel de poliacrilamida 15% com as frações coletadas durante a purificação. (A) Fração insolúvel da expressão. (B) Fração solúvel da expressão. (C) Fração eluída na IMAC. (D) Fração que não interagiu com a resina durante a IMAC. (E) Fração precipitada após 12 h à temperatura ambiente. (F) Fração solúvel após a gel filtração em coluna XK 26/1000 com resina Superdex 200, contendo uma única forma oligomérica da <i>XaDHNA</i>	119
Figura 40 – Curvas de fluorescência do Sypro Orange no experimento de DSF para a proteína <i>XaDHNA</i>	120
Figura 41 – Cristais de <i>XaDHNA</i> obtidos durante o experimento de otimização da condição de cristalização utilizando o <i>kit</i> Detergent Screen HT e suas respectivas condições.	122
Figura 42 – Alinhamento da sequência completa da <i>XaDHNA</i> com o modelo em refinamento.	124
Figura 43 – (A) Topologia da estrutura da <i>XaDHNA</i> gerada utilizando o PDBsum. ¹⁵⁸ (B) Representação da estrutura primária da proteína <i>XaDHNA</i> e suas respectivas estruturas secundárias.	125
Figura 44 – Representação da posição do substrato da reação da DHNA na estrutura homóloga de <i>E. coli</i> PDB 2O90.	125
Figura 45 – Representação do modelo da <i>XaDHNA</i> com a disposição biológica, ainda em estágio de refinamento. (A) Diagrama de Ramachandran da <i>XaDHNA</i> . (B) Representação do contato feito entre dois monômeros de metades diferentes do octâmero da <i>XaDHNA</i> (C) Visão superior do homotetrâmero da <i>XaDHNA</i> , sobre o eixo de simetria de ordem 4 e sua visão ortogonal, sobre um dos eixos de simetria de ordem 2.....	126
Figura 46 – Comparação entre estruturas homólogas da <i>XaDHNA</i> disponíveis no PDB. (A) Representação da conservação de resíduos (vermelho – conservado, branco – não conservado) entre as homologas analisadas e RMSD médio entre as posições dos α 5VSP-homóloga (maior espessura na representação – maior RMSD médio). (B) Conservação dos resíduos do	

sítio ativo que interagem diretamente com o substrato 6HMPH2. (C) Alinhamento das sequências primárias das proteínas homólogas analisadas. (D) Comparação filogenética das todas as sequências com estrutura disponível no PDB, geradas pela ferramenta PhyloDendron.127

Figura 47 – Diagrama do algoritmo de desenho de oligonucleotídeos iniciadores por tipo de objetivo. O detalhamento dos nucleotídeos entre parênteses indica a atual sequência (antes do ponto) e a sequência que será construída (após o ponto).131

Figura 48 – (A) Página principal do HTP Oligo Designer. (B) Calculadora de Tm. A ferramenta está disponível no endereço <http://www.ifsc.usp.br/htpoligo>.132

Figura 49 – Análise da utilização do HTP Oligo Designer no período entre 16 de maio de 2016 e 17 de maio de 2017. O mapa mostra a escala de utilização entre os países: Reino Unido, Rússia e Brasil são os três países com o maior número de acessos, nesta ordem.132

LISTA DE TABELAS

Tabela 1 – Oligonucleotídeos iniciadores utilizados para identificar culturas de <i>X. albilineans</i>	62
Tabela 2 – Características dos vetores de expressão utilizados.	64
Tabela 3 – Sequências de oligonucleotídeos iniciadores utilizados para amplificação dos segmentos gênicos de interesse.....	65
Tabela 4 – Parâmetros utilizados para a minimização de energia do sistema, equilíbrio do sistema e dinâmica molecular.	75
Tabela 5 – Valores de MIC e NIC para os antibióticos tetraciclina e tobramicina. Os valores foram calculados para cultura de <i>X. albilineans</i> IBSBF 740 utilizando o método descrito no item 3.2.	81
Tabela 6 – Condições tamponantes identificadas no experimento de DSF para a proteína <i>XaHPPK</i>	89
Tabela 7 – Classificação dos kits comerciais utilizados neste projeto quanto ao planejamento de sua formulação.	92
Tabela 8 – Condições de cristalização nas quais foram coletados conjuntos de dados para os cristais da proteína <i>XaHPPK</i> , com suas respectivas formulações.	93
Tabela 9 – Parâmetros de cela unitária, simetria do cristal, qualidade do conjunto de dados e estatística do processamento <i>XaHPPK</i> . Entre parênteses estão os valores para a camada de mais alta resolução.	101
Tabela 10 – Comparação entre HPPKs homólogas com mais de 40% de identidade disponíveis no PDB com a <i>XaHPPK</i> . Os valores de RMSD são calculados pela superposição das estruturas da <i>XaHPPK</i> com a estrutura homóloga em destaque.....	107
Tabela 11 – Definição das participações dos termos da equação do <i>GlideScore</i> . ^{178,179}	112
Tabela 12 – Resultado da triagem virtual de ligantes utilizando o ORCA para cálculo quântico das cargas das moléculas e o Glide para predição do modo de ligação.....	115
Tabela 13 – Condições tamponantes identificadas no experimento de DSF para a proteína <i>XaDHNA</i>	120

Tabela 14 – Classificação dos <i>kits</i> comerciais utilizados neste projeto quanto ao planejamento de sua formulação.....	121
Tabela 15 – Condição de cristalização na qual foi coletado o conjunto de dados para o cristal da proteína <i>XaDHNA</i> , com sua respectiva formulação.....	122
Tabela 16 – Parâmetros de cela unitária, simetria do cristal, qualidade do conjunto de dados e estatística do processamento <i>XaDHNA</i> . Entre parênteses estão os valores para a camada de mais alta resolução.....	123
Tabela 17 – Comparação entre DHNAs homólogas disponíveis no PDB com a <i>XaHPPK</i> . Os valores de RMSD são calculados pela superposição das estruturas da <i>XaDHNA</i> com a estrutura homóloga em destaque.	128
Tabela 18 – Guia de entrada de informações para planejamento de mutagênese no HTP Oligo Designer.	129

LISTA DE ABREVIATURAS E SIGLAS

6-His	Cauda de hexa-histidina
A.C.	Antes de Cristo
ATP	Adenosina Trifosfato
CC _{1/2}	Coeficiente de correlação
CCD	Dispositivo de carga acoplada
CCP4	Consórcio colaborativo – Projeto #4
CMA	Ácido carboximetil aspártico
CoA	Coenzima A
crio-ME	Crio microscopia eletrônica
CTAB	Brometo de cetiltrimetilamônio
Cya	β-ciano-L-alanina
D.C.	Depois de Cristo
D.O. _{600 nm}	Densidade óptica a 600 nm
dATP	Desoxi Adenosina Trifosfato
DHFR	Di-hidro folato redutase
DHNA	Di-hidro neopterina aldolase
DHPS	Di-hidro pteroato sintase
DMSO	Dimetil Sulfóxido
DNA	Ácido desoxirribonucleico
DSF	Fluorimetria diferencial de varredura
DTT	Ditiotreitol
dTTP	Desoxi Timina Trifosfato
FBDD	Planejamento de inibidores baseado em fragmentos
GST	Glutathione S-transferase
HPPK	2-amino-4-hidroxi-6-hidroximetil di-hidropteridina pirofosfoquinase
HtpG	Proteína de choque térmico G
IMAC	Cromatografia de afinidade a metal imobilizado
IPTG	Isopropil 2-tiogalacto piranosídeo

ITC	Calorimetria de titulação isotérmica
K_d	Constante de dissociação
LB	Caldo lisogênico
LIC	Clonagem de ligação independente
LLG	Logarítmo do ganho de verossimilhança
LNLS	Laboratório Nacional de Luz Síncrotron
LogP	Coeficiente de partição entre água e hexano
MAD	Dispersão anômala de múltiplos comprimentos de onda
MCA	ácido <i>p</i> -cumárico
MCS	Sítio de clonagem múltipla
ME	Microscopia eletrônica
MIC	Concentração inibitória mínima
MM	Massa molecular
MST	Termoforese capilar
NIC	Concentração não inibitória
NRPS	Peptídeo sintase não ribossomal
NTA	Ácido nitriloacético
OPLS-AA	Potencial para simulação de líquidos otimizado para todos os átomos
ORF	Quadro aberto de leitura
pABA	Ácido <i>p</i> -amino benzoico
PAINS	Compostos interferentes em ensaios gerais
PCR	Reação em cadeia de DNA polimerase
PDB	Banco de dados de proteínas
PEG	Poli etileno glicol
pET	Plasmídeo de expressão com promotor T7
PKS	Policetídeo sintase
pMBA	Ácido 4-amino-2-hidróxi-3-metoxibenzóico
PMSF	Fluoreto de fenilmetano sulfonil
PPT	Fosfopanteteinil transferase
qPCR	Reação em cadeia quantitativa de DNA polimerase

RMN	Ressonância magnética nuclear
RNA	Ácido ribonucleico
SAD	Dispersão anômala com único comprimento de onda
SAR	Relações estrutura-atividade
SBDD	Planejamento de inibidores base na estrutura do alvo
SDS-PAGE	Eletroforese em gel de poliacrilamida desnaturante
SPR	Ressonância plasmônica de superfície
SUMO	Modificadores pequenos similares à ubiquitina
TEV	Virus da gravura de tabaco
TFZ	Z-Score da função de translação
THF	Tetra hidrofolato
T _m	Temperatura de desnaturção
TRX	Tiorredoxina
VLS	Triagem virtual de ligantes
XAS	Meio seletivo de <i>Xanthomonas albilineans</i>

LISTA DE SÍMBOLOS

K	Kelvin
°C	Graus Celsius
rpm	Rotações por minuto
s	Segundo
min	Minuto
h	Hora
Å	Angstrom
nm	Nano metro
nL	Nano litro
μL	Micro litro
mL	Mili litro
L	Litro
Kb	Quilo base
ΔT_m	Variação da temperatura de desnaturação
m/v	Relação massa volume
pH	Potencial de hidrogênio
mg	Mili grama
mM	Mili molar
M	Molar
μ_0	Constante de permeabilidade magnética do vácuo
KDa	Quilo Dalton

SUMÁRIO

1	INTRODUÇÃO	37
1.1	CANA-DE-AÇÚCAR - INDÚSTRIA SUCROALCOOLEIRA E O DESENVOLVIMENTO SUSTENTÁVEL	37
1.2	DOENÇAS DA CANA-DE-AÇÚCAR	40
1.3	A ESCALDADURA DAS FOLHAS	40
1.4	XANTHOMONAS ALBILINEANS.....	43
1.5	A ALBICIDINA	45
1.6	SELEÇÃO DE ALVOS MOLECULARES PARA A ESCALDADURA DAS FOLHAS	48
1.7	VIA DE BIOSÍNTESE DE FOLATOS E AS ENZIMAS DI-HIDROPTERINA PIROFOSFOQUINASE E DI-HIDRONEOPTERINA ALDOLASE	48
1.8	CRISTALOGRAFIA DE RAIOS X.....	50
2	OBJETIVOS	57
3	MATERIAL E MÉTODOS.....	61
3.1	CULTURAS DE XANTHOMONAS ALBILINEANS – OBTENÇÃO, CULTIVO, CARACTERIZAÇÃO E EXTRAÇÃO DE DNA GENÔMICO	61
3.2	PADRONIZAÇÃO DO ENSAIO E TRIAGEM FENOTÍPICA DE X. ALBILINEANS	62
3.3	CLONAGEM DOS SEGMENTOS GÊNICOS FOLB E FOLK.....	64
3.3.1	<i>Método da clonagem de ligação independente</i>	<i>64</i>
3.4	EXPRESSÃO E PURIFICAÇÃO DAS PROTEÍNAS DE X. ALBILINEANS	65
3.4.1	<i>Testes de expressão das proteínas XaDHNA e XaHPPK</i>	<i>65</i>
3.4.2	<i>Expressão e purificação das proteínas XaDHNA e XaHPPK</i>	<i>66</i>
3.5	CARACTERIZAÇÃO DA PROTEÍNA XAHPPK	68
3.5.1	<i>Avaliação da estabilidade da proteína XaHPPK.....</i>	<i>68</i>
3.5.2	<i>Determinação do parâmetro cinético K_D para a proteína XaHPPK.....</i>	<i>69</i>
3.6	ENSAIOS DE CRISTALIZAÇÃO DA XAHPPK E XADHNA	69
3.6.1	<i>Triagem de condições de cristalização utilizando kits comerciais</i>	<i>69</i>
3.6.2	<i>Montagem do cristal e coleta de dados.....</i>	<i>70</i>
3.6.3	<i>Processamento do conjunto de dados e obtenção do modelo tridimensional</i>	<i>71</i>
3.7	DESCOBERTA DE INIBIDORES	72
3.7.1	<i>Triagem virtual para a enzima XaHPPK</i>	<i>72</i>

3.7.2	<i>Cálculo de cargas de pequenas moléculas por métodos quânticos</i>	74
3.7.3	<i>Dinâmica molecular para a enzima XaHPPK</i>	74
4	RESULTADOS	79
4.1	CULTURAS DE XANTHOMONAS ALBILINEANS – OBTENÇÃO, CULTIVO, CARACTERIZAÇÃO E EXTRAÇÃO DE DNA GENÔMICO	79
4.2	PADRONIZAÇÃO DO ENSAIO FENOTÍPICO DE X. ALBILINEANS	79
4.3	TRIAGEM FENOTÍPICA DE X. ALBILINEANS	81
4.4	CLONAGEM DOS SEGMENTOS GÊNICOS FOLB E FOLK	83
4.4.1	<i>Sistema de expressão heteróloga utilizando vetores pET</i>	83
4.4.2	<i>Método da clonagem de ligação independente</i>	85
4.5	2-AMINO-4-HIDROXI-6-HIDROXIMETIL DI-HIDROPTERIDINA PIROFOSFOQUINASE (HPPK)	87
4.5.1	<i>Expressão e purificação da proteína HPPK de X. albilineans</i>	87
4.5.2	<i>Avaliação da purificação e estabilidade da proteína XaHPPK</i>	88
4.5.3	<i>Técnica da difusão de vapor em gota assentada</i>	90
4.5.4	<i>Triagem e otimização de condições de cristalização da XaHPPK</i>	92
4.5.5	<i>Avaliadores de qualidade do conjunto de dados</i>	94
4.5.6	<i>O problema das fases</i>	95
4.5.7	<i>Avaliação da qualidade do modelo tridimensional</i>	97
4.5.8	<i>Definição da resolução da estrutura final dos modelos</i>	99
4.5.9	<i>Coleta e processamento de dados de difração de raios X para a XaHPPK ...</i>	100
4.5.10	<i>Estrutura da XaHPPK de Xanthomonas albilineans</i>	102
4.5.11	<i>Comparação da XaHPPK com suas proteínas homólogas</i>	105
4.5.12	<i>Determinação do parâmetro cinético K_d aparente para a proteína XaHPPK</i>	107
4.5.13	<i>Triagem virtual de ligantes</i>	110
4.5.13.1	<i>Descoberta de inibidores baseado na estrutura do alvo</i>	110
4.5.13.2	<i>Acoplamento molecular</i>	110
4.5.13.3	<i>Triagem virtual</i>	113
4.6	7,8-DI-HIDRONEOPTERINA ALDOLASE	118
4.6.1	<i>Expressão e purificação da proteína DHNA de X. albilineans</i>	118
4.6.2	<i>Avaliação da purificação e estabilidade da proteína XaDHNA</i>	119
4.6.3	<i>Triagem e otimização de condições de cristalização da proteína XaDHNA .</i>	120
4.6.4	<i>Coleta e processamento de dados de difração de raios X para a XaDHNA .</i>	122

4.6.5	<i>Modelo DHNA de Xanthomonas albilineans</i>	124
4.6.6	<i>Comparação da XaDHNA com suas proteínas homólogas</i>	127
4.7	DESENVOLVIMENTO DA FERRAMENTA HTP OLIGO DESIGNER.....	128
4.7.1	<i>Motivação científica</i>	128
4.7.2	<i>Algoritmo de determinação da sequência para o oligonucleotídeo iniciador</i>	129
5	CONCLUSÃO	135
6	PERSPECTIVAS	137
	REFERÊNCIAS	139
	APÊNDICE A – Produção científica	153

Introdução

1 INTRODUÇÃO

1.1 Cana-de-açúcar - Indústria sucroalcooleira e o desenvolvimento sustentável

A cana-de-açúcar é originária do leste asiático, provavelmente na atual Papua Nova Guiné.¹ No início do século V A.C. aproximadamente, as primeiras mudas foram levadas à região onde hoje é a Índia, onde as tecnologias de extração e processamento do caldo da cana-de-açúcar foram desenvolvidas.¹ Fatores como guerras e comércio levaram o cultivar para o oriente médio, onde o aperfeiçoamento e difusão destas tecnologias adicionaram o caráter econômico à cana-de-açúcar, de modo que entre os séculos X e XII D.C. a expansão da cultura rapidamente atingiu o norte do continente africano e sul do continente europeu, como as regiões onde hoje se situam Egito, Itália, Grécia e Espanha.¹ A difusão dentro da península Ibérica levou a cana-de-açúcar a ser o principal produto agrícola nas colônias tropicais portuguesas e espanholas, como Madeira, Açores e Ilhas Canário e posteriormente, em 1532, Brasil (Figura 1).¹

Figura 1 – Expansão histórica do cultivo da cana-de-açúcar até sua chegada ao Brasil.
Fonte: Elaborada pelo autor.

A planta foi introduzida no Brasil por Martim Afonso de Souza, em 1532, e hoje constitui uma das principais culturas para o agronegócio brasileiro. O setor sucroalcooleiro brasileiro se destaca no cenário internacional, movimentando aproximadamente US\$ 100 bilhões em exportações anuais.²⁻³ Recentemente, além do açúcar e do bioetanol, as usinas

de processamento de cana-de-açúcar passaram a produzir bioeletricidade, álcoolquímicos, bem como a comercialização de créditos de carbono.

A cana-de-açúcar é uma das principais fontes de energia renovável, constituindo a matéria-prima mais importante na busca de energia limpa e sustentável. A matriz energética brasileira é marcada pela forte utilização de energias renováveis, que representam aproximadamente 40% de toda energia utilizada no País. Quando comparado a outros países em desenvolvimento ou industrializados, nos quais apenas 6% e 13% da energia, respectivamente, provém de fontes renováveis, a proeminência da utilização de bioenergia no Brasil torna-se mais impactante.²⁻³

Figura 2 – Distribuição geográfica da produção de cana-de-açúcar safra 2015/2016, ⁴ colorida com base na porcentagem da área cultivada.

Fonte: Elaborada pelo autor.

O principal biocombustível derivado de cana-de-açúcar, o etanol, atingiu na safra de 2015/2016 a produção de 30,5 bilhões de litros.⁵⁻⁶ A produção de etanol a partir de cana-de-açúcar é social e economicamente relevante para o País, uma vez que estimula a geração direta e indireta de empregos pelo modo descentralizado de produção e processamento do

cultivar, amplamente distribuído entre os estados brasileiros. Ainda, a utilização do bioetanol combustível é capaz de reduzir em até 90% a produção de gases de efeito estufa, quando comparado à gasolina.⁷ Além disso, o uso da biomassa gerada pelo processamento da planta pode ser utilizado para a produção de bioeletricidade, biopolímeros, ração animal, aglomerados, fertilizantes, entre outros.

O Brasil é o líder mundial na produção de cana-de-açúcar, responsável por cerca de 40% do total cultivado em 2016, seguido por Índia e China, responsáveis por aproximadamente 19% e 7% da produção, respectivamente (Figura 2).⁴ Analisando a produção interna na safra 2015/2016, 95% desta se concentra em oito estados, ordenados pelo percentual de área total de cana-de-açúcar cultivado: São Paulo, Goiás, Minas Gerais, Mato Grosso do Sul, Paraná, Alagoas, Pernambuco e Mato Grosso (Figura 3).⁵

Figura 3 – Distribuição geográfica do total de cana-de-açúcar cultivado por Unidade da Federação, safra 2015/2016,⁶ colorida com base na porcentagem da área cultivada.

Fonte: Elaborada pelo autor.

Diante do cenário econômico e social do setor sucroalcooleiro no Brasil, é de extrema importância o controle de fatores que possam comprometer o cultivo e a produtividade dos canaviais, com o objetivo de garantir a expansão da produção para o futuro.

1.2 Doenças da cana-de-açúcar

Entre os diversos fatores que afetam a produtividade de um canavial, destaca-se a ocorrência de doenças nas culturas de cana-de-açúcar, causadas por bactérias, fungos e vírus, além do ataque de diversas pragas.⁸⁻⁹ Dentre as mais de duzentas doenças de cana-de-açúcar identificadas em todo o mundo, destacam-se cinco como as mais importantes por causarem danos substanciais às lavouras:¹⁰ i. carvão – causado pelo fungo *Ustilago scitaminea*; ii. ferrugem – causada pelos fungos *Puccinia melanocephala* e *Puccinia kuehnii*; iii. raquitismo das soqueiras – causado pela bactéria *Leifsonia xyli*; iv. mosaico – causado pelo vírus do mosaico; v. escaldadura das folhas – causada pela bactéria *Xanthomonas albilineans*.

1.3 A escaldadura das folhas

A doença está presente em praticamente todos os países onde a cana-de-açúcar é cultivada (**Figura 4**). Entre os prejuízos causados estão a diminuição da produtividade, reforma precoce dos canaviais e a diminuição do valor energético do caldo extraído¹¹⁻¹², este último causado pela redução no teor de sacarose na planta. O impacto real causado pela escaldadura das folhas é subestimado devido às dificuldades no diagnóstico e diferenciação dos prejuízos causados por outras doenças, como por exemplo o raquitismo das soqueiras. Os dados existentes para as variedades de cana-de-açúcar cultivadas em Guadalupe (B 69379), Ilhas Maurício (M 695/69) e México (Mex 64-1487) apontam para um decréscimo de produtividade por hectare entre 15 e 30%.¹³⁻¹⁵

Figura 4 – Distribuições geográficas das culturas de cana-de-açúcar e escaldadura das folhas.¹⁶
Fonte: Elaborada pelo autor

A escaldadura das folhas acomete lavouras de plantas monocotiledôneas da Família *Poaceae*, incluindo a *Saccharum* spp. Em cana-de-açúcar, a doença pode ser assintomática (estado latente), e se mantém desta forma utilizando mecanismos ainda não conhecidos.¹⁷⁻
¹⁸ Em alguns casos, a doença se manifesta apenas após a colheita, a partir da soqueira infectada.¹⁸ O estágio crônico da doença é caracterizado por diversos sintomas externos, sendo o mais comum a presença de linhas esbranquiçadas com larguras entre 1 e 2 mm na direção da bainha ao limbo foliar (Figura 5 **A**) com uma borda difusa de tamanho variável e podendo conter regiões avermelhadas (Figura 5 **F**).¹⁸⁻²⁰ A progressão natural da escaldadura das folhas leva à necrose das células, iniciando pelas extremidades da folha (lateral ou apical) e espalhando-se por toda a folha, caracterizando a fase aguda da doença. Nesse momento, os aspectos esbranquiçados e contorcidos das folhas doentes se assemelham ao de folhas escaldadas pelo sol, o que dá nome à doença. O último estágio da doença é a morte da planta.¹⁷⁻²⁰

Figura 5 – Sintomas da escaldadura das folhas em cana-de-açúcar. (A) Linhas esbranquiçadas na direção da nervura da folha. (B) Estágio avançado da doença no qual as folhas já estão necrosadas. (C) Clareamento parcial das folhas no estágio crônico da doença. (D) Linhas esbranquiçadas com algumas manchas avermelhadas. (E) Secção longitudinal de uma muda de cana-de-açúcar infectada, em estágio avançado da fitodoença. (F) Clorose das folhas com o aspecto avermelhado, forma mais característica da doença.

Fonte: Adaptada de ROTT.²¹

O principal meio de disseminação da doença é a utilização de mudas infectadas.²² Outras formas de disseminação incluem a ação dos ventos, chuvas ou durante a colheita, quando os ferimentos nos colmos causados pela colheita, mecanizada ou não, transportam bactérias por toda a lavoura.²¹ O fitopatógeno é capaz de sobreviver no restolho da cana-de-açúcar, mas não de resistir a longos períodos no bagaço de cana.^{15,30} Alternativamente, a

bactéria *X. albilineans* é capaz de utilizar hospedeiros intermediários, em sua maioria gramíneas, como o capim-napiê (*Pennisetum purpureum*).²¹

Atualmente, existe uma carência em estratégias de controle e combate à doença, resumindo-se à substituição de variedades suscetíveis por variedades resistentes. Assim, variedades resistentes são escolhidas durante o processo de seleção de variedades comerciais em detrimento de características desejadas, como alta produtividade e alto teor de sacarose no caldo extraído. Esse fato é ainda agravado pelo surgimento de variações genéticas nos patógenos que acarretam a quebra de resistência, consequentemente, transformando cultivares resistentes em altamente suscetíveis.²³⁻²⁴ Nesse cenário, que urge a necessidade pelo desenvolvimento de novas alternativas para o controle da escaldadura das folhas, cria-se a oportunidade para a descoberta de alvos moleculares atrativos para o desenvolvimento de novas moléculas bioativas, potentes e seletivas, capazes de contribuir significativamente para o controle seguro e eficiente da doença.

1.4 *Xanthomonas albilineans*

A *Xanthomonas albilineans*, agente causador da escaldadura das folhas, é uma bactéria gram-negativa da família *Xanthomonadaceae*, pertencente à ordem *Xanthomonadales* da subdivisão gama do filo *Proteobacteria* (**Figura 6**).²⁵⁻²⁶ Morfologicamente, a *X. albilineans* apresenta o formato de bastão alongado, uniflagelada, não esporulante, aeróbica, crescimento lento, aspecto amarelado e colônias não mucoidais. Sua adaptação à cana-de-açúcar se caracteriza pelo crescimento lento e limitado às células do xilema da planta, especialmente nas células cilíndricas e porosas, os traqueídeos.

Em comparação com outros membros da família *Xanthomonadaceae*, como a bactéria *Xylella fastidiosa*, ambas possuem um genoma reduzido e ausente do sistema de secreção do tipo III, o qual é composto pelos genes de controle *hrp* (do inglês, *hypersensitive reaction and pathogenicity*).²⁶ O cluster de genes *hrp-T3SS* (do inglês, *hypersensitive reaction and pathogenicity - type three secretion system*) é comum à maioria das bactérias fitopatogênicas, como *Pseudomonas syringae*, *Erwinia amylovora* e *Ralstonia solanacearum*, desempenhando papel essencial na interação planta-patógeno.²⁷

Figura 6 – Classificação taxonômica da *Xanthomonas albilineans*.

Fonte: Adaptada de BIRCH.¹⁷

A análise filogenética de representantes dos gêneros *Xanthomonas* spp, *Stenotrophomonas* spp e *Xylella* spp confirmou que, embora exista alta relação entre estes, ainda é possível diferenciar *X. campestris*, *X. axonopodis* e *X. albilineans* em um subgrupo mais específico.²⁸ Esta observação, aliada à comparação dos genes constitutivos *ihfA* e *efp*, sugerem que a *X. albilineans* foi um intermediário evolutivo entre diversas espécies de *Xanthomonas* spp e a *Xylella fastidiosa*, e que estas são geneticamente mais próximas entre si do que entre membros de seus respectivos gêneros.²⁶

Outro mecanismo comum às espécies fitopatogênicas e ausente na *X. albilineans* é a biossíntese de goma xantana. Alternativamente, a *X. albilineans* incorporou o *cluster* de genes SPI-1-T3SS (do inglês, *Salmonella pathogenicity island-1 type three secretion system*) comumente presente em bactérias patogênicas de animais.²⁹⁻³⁰

1.5 A albicidina

Em *X. albilineans*, a resistência à ação da albicidina se dá por diversos mecanismos distintos: i. mimetização da molécula de DNA; ii. biossíntese de albicidinas altamente regulada; iii. efluxo eficiente de albicidina. O *cluster* de genes XALB1 possui uma sequência que codifica uma proteína similar às que conferem resistência à quinolonas, denominada *albG*, que possui uma região pentapeptídica responsável por mimetizar a molécula de DNA e impedir a interação da albicidina com a subunidade A da DNA girase.^{17-18,25} Em contraste com o observado em outros organismos que produzem moléculas inibidoras de DNA girase, não existe evidências de subunidades A ou B resistentes a albicidinas em *X. albilineans*. A regulação de biossíntese de albicidinas, embora não relacionada ao crescimento bacteriano, é feita baseada na abundância de nutrientes no meio, principalmente aminoácidos e fontes de fósforo e nitrogênio.³³ O aumento da quantidade do aminoácido histidina no meio leva a uma drástica redução na produção de albicidina devido à sua interação com a proteína AlbA.³³⁻³⁴ Condições adversas, como a baixa concentração de fosfatos no meio, estimula a produção da albicidina, em concordância com as observações do aparecimento dos sintomas da escaldadura das folhas nas lavouras após período estresse, como seca prolongada. Adicionalmente, a resistência também é conferida pelo efluxo de albicidina para o meio extracelular mediado pela proteína DHA14. A análise da atividade antimicrobiana da albicidina frente a *E. coli* mostrou um aumento de até três mil vezes no MIC em cepas que expressavam o gene *albF*, codificante para DHA14.³⁵

A *X. albilineans* ainda possui um *cluster* de genes exclusivo desta espécie responsável pela biossíntese da albicidina, o *alb*.¹⁷⁻³⁶⁻³⁸ Este *cluster* possui aproximadamente 49 Kb, denominado XALB1, e duas regiões adicionais de aproximadamente 3 Kb cada, denominadas XALB2 e XALB3, necessárias para a biossíntese da albicidina. Dentro da região denominada XALB1, que possui um total de 20 quadros abertos de leitura (ORF, do inglês, *Open Reading Frame*), estão genes para três grandes poliketideo sintases, PKS (do inglês, *polyketide synthase*), e uma peptídeo sintase não-ribossomal, NRPS (do inglês, *nonribosomal peptide synthase*), assim como outras regiões reguladoras e relacionadas a resistência à albicidina.³¹⁻³⁷⁻³⁹⁻⁴² A região denominada XALB2 contém um único gene que leva a expressão da proteína fosfopanteteinil transferase, PPT (do inglês, *phosphopantetheinyl transferase*), que regula a

ativação pós traducional das PKSs e NRPSs utilizando a transferência do grupo fosfopanteteína da Coenzima A (CoA) para um resíduo de serina nas enzimas inativas.²⁰⁻³⁷⁻⁴⁰⁻

⁴¹⁻⁴³ A região XALB3 codifica a proteína HtpG (do inglês, *heat-shock protein G*), uma proteína de membrana que auxilia as PKSs e NRPSs na biossíntese de albicidinas.³⁷⁻⁴⁴

A molécula de albicidina é formada por um bloco de pentapeptídeo unidos por seu N-terminal a um derivado do ácido p-cumárico (MCA-1), o ácido 3-(4-hidroxifenil)-2-metilacrílico.³⁶ O pentapeptídeo, por sua vez, é formado pelo α -aminoácido não proteinogênico β -ciano-L-alanina (Cya-3), os δ -aminoácidos aromáticos ácido p-aminobenzóico (pABA-2 e pABA-4) e pelo ácido 4-amino-2-hidróxi-3-metoxibenzóico (pMBA-5 e pMBA-6).³⁶ Os números que procedem as siglas das moléculas relacionam estas à **Figura 7**. As seis etapas da síntese, proposta por Cociancich e colaboradores baseada na estrutura molecular da albicidina e no arranjo dos genes no cluster *alb*,³⁶ é iniciada pela incorporação de uma molécula de pABA (pABA-2) a uma molécula de MCA (MCA-1) por uma ação em conjunto das proteínas PKS e NRPS-1 do módulo Alb01 do *cluster alb*, na qual a molécula de pABA é carregada ou por genes intrínsecos ao cluster, *alb17* e *alb18*, quanto por seus parálogos provenientes do metabolismo primário *pabAB* e *pabC*.³⁶⁻⁴⁵ Em seguida, a síntese procede com a incorporação de uma molécula de Cya (Cya-3) pela NRPS-2* do módulo Alb04, por um mecanismo trans-complementar similar ao observado na síntese de bleomicina de duas etapas: i. a adenilação da extremidade ácido α -carboxílico para a formação de um intermediário tioéster e ii. a fosforilação do oxigênio da amida na cadeia lateral. Seguida por desfosforilação, é liberado o aminoácido ciano-L-alanina, que ligada ao esqueleto de albicidina em formação da continuidade à biossíntese.³⁶⁻⁴⁰ Na quarta etapa da rota biossintética, mais uma molécula de pABA (pABA-4) é ativada por uma NRPS-4, proveniente do módulo Alb01, e ligada ao produto da etapa anterior.³⁶ As duas últimas etapas são realizadas por duas NRPSs, NRPS-5 e NRPS-6, traduzidas a partir de sequências contidas no módulo Alb09, que incorporam os dois últimos resíduos de pMBA, pMBA-5 e pMBA-6, respectivamente, liberando a molécula de albicidina.³⁶

O modo de ação da albicidina é a inibição seletiva da subunidade A da DNA girase do cloroplasto, principalmente nos tecidos parenquimáticos como os parênquimas paliçádico e lacunoso, esclerênquima e colênquima,^{13-14,16-18} inibindo a introdução de super espirais negativos no DNA bacteriano necessários para contrapor a criação das super espirais

positivos durante a replicação do DNA circular ou processos de transcrição.³¹⁻³²⁻⁴⁶ A DNA girase de procariotos é composta por duas subunidades, A e B, formando um complexo tetramérico A_2B_2 .⁴⁷⁻⁴⁸ A atividade da DNA girase em organismos procariotos inicia-se pela interação da fita super espiralada de DNA com a subunidade A, na qual está o sítio ativo responsável pela clivagem transiente do DNA.³²⁻⁴⁸ A reação é mediada pela quebra de moléculas de ATP, feitas na subunidade B do complexo. O produto desta reação é a adição de duas super espirais negativas ao DNA super espiralado.

A

B

Figura 7 – (A) (1) Incorporação de uma molécula de pABA a uma molécula de MCA por uma ação em conjunto das proteínas PKS e NRPS do módulo Alb01 do cluster alb. (2) Incorporação de uma molécula de Cya pela NRPS do módulo Alb04, por um mecanismo trans-complementar de duas etapas: i. a adenilação da extremidade ácido α -carboxílico para a formação de um intermediário tioéster e ii. a fosforilação do oxigênio da amida na cadeia lateral. (3) O aminoácido ciano-L-alanina é incorporado ao esqueleto precursor da albicidina. (4) Uma molécula de pABA é ativada por uma NRPS, proveniente do módulo Alb01, e ligada ao produto da etapa anterior. (5) e (6) As duas últimas etapas são realizadas por duas NRPSs, NRPS-5 e NRPS-6, traduzidas a partir de sequências contidas no módulo Alb09, que incorporam os dois últimos resíduos de pMBA, pMBA-5 e pMBA-6, respectivamente, liberando a molécula de albicidina. (B) Representação tridimensional da estrutura da albicidina.

Fonte: Adaptada de COCIANCICH.³⁶

Dada a sua natureza essencial para a célula e suas diferenças para as equivalentes em eucariotos, esta enzima tem sido alvo constante do desenvolvimento de novos agentes antibacterianos, ⁴⁹⁻⁵⁰ como quinolonas (*e.g.*, ácido nalidíxico e ácido oxolínico) que interferem na religação das fitas de DNA separados pela subunidade A da DNA girase, cumarinas (*e.g.*, coumermicina A₁ e novobiocina) que competem pelo sítio de ligação de ATP na subunidade B e simociclononas que inibem a interação inicial do DNA com a subunidade A. ⁵⁰⁻⁵¹

1.6 Seleção de alvos moleculares para a escaldadura das folhas

No planejamento de substâncias bioativas como candidatos a inibidores para doenças infecciosas, avalia-se as vias bioquímicas essenciais do patógeno, explorando as divergências moleculares ente o patógeno e o hospedeiro, a fim de modular seletivamente os alvos moleculares. A elucidação do genoma de *X. albilineans* forneceu valiosas informações sobre rotas biossintéticas e sobre as proteínas envolvidas nesses processos, ²⁶ oferecendo oportunidades científicas para o desenvolvimento de novas formas de controle da doença.

1.7 Via de biossíntese de folatos e as enzimas di-hidropterina pirofosfoquinase e di-hidroneopterina aldolase

Os folatos são moléculas formadas por três componentes, caracterizados pelo i. anel pterina, ii. ácido para-aminobenzóico (pABA) e iii. cauda glutamilada, cuja composição varia de 1–8 resíduos glutamato (**Figura 8**). ⁵² Os folatos diferenciam-se pelo estado de oxidação do anel pterina, pela natureza do substituinte nas posições N⁵ e N¹⁰ e pelo número de unidades de glutamato conjugadas durante a formação da cauda oligo-γ-glutâmica. A estrutura geral do tetra-hidrofolato (THF), está representada na **Figura 8**.

Os folatos desempenham função central nas vias metabólicas responsáveis pela sobrevivência celular de eucariotos e procariotos, sendo o THF doador de unidades de carbono em diversos processos metabólicos, incluindo a síntese de purinas, de timidilato e de metionina. ⁵³⁻⁵⁵ Enquanto os animais adquirem folatos a partir da dieta pelo transporte

ativo através de proteínas transportadoras associadas à membrana celular, plantas, fungos, alguns protistas e a maioria das bactérias são dependentes da síntese *de novo* de folatos.⁵³

Uma vez que a via de biossíntese de folatos é essencial para a manutenção de níveis celulares adequados de THF, as enzimas envolvidas na síntese e modificação deste cofator são alvos moleculares para o desenvolvimento de novos fármacos.⁵⁴⁻⁵⁶⁻⁵⁷ As sulfonamidas, introduzidas na terapêutica na década de 1930 e as sulfonas, na década seguinte, são exemplos de inibidores da via dos folatos que atuam como análogos estruturais do pABA, inibindo a enzima di-hidropteroato sintase (DHPS).⁵³ Outra enzima bastante estudada e situada na etapa final da via biossintética dos folatos é a di-hidrofolato redutase (DHFR), na qual atuam fármacos da classe de diaminopirimidinas.⁵³ As diaminopirimidinas são análogos estruturais das pteridinas, bloqueando a síntese bacteriana de THF. O efeito sinérgico observado na associação de sulfonamidas e diaminopirimidinas fez com que essas duas classes fossem amplamente utilizadas no tratamento de infecções.⁵⁴

Figura 8 – Estrutura química do tetra-hidrofolato (THF).

Fonte: Adaptada de de CRÉCY-LAGARD et al.⁵⁸

Além das enzimas DHPS e DHFR, para as quais diversos fármacos foram desenvolvidos, outras enzimas da via dos folatos podem constituir alvos moleculares para o desenvolvimento de novos inibidores. Uma vez que análogos estruturais do substrato podem atuar em múltiplas enzimas da mesma cascata metabólica,⁵⁴ levando à inibição enzimática simultânea, maior eficácia do tratamento e diminuição da resistência bacteriana,

a caracterização estrutural e funcional de todas as enzimas de uma via biossintética já validada é uma estratégia interessante na busca por novas substâncias inibidoras.

As enzimas 7,8-di-hidroneopterina aldolase (*XaDHNA*, E.C. 4.1.2.25) e 2-amino-4-hidroxi-6-hidroximetil-di-hidropteridina pirofosfoquinase (*XaHPPK*, E.C. 2.7.6.3), codificadas pelos genes *folB* e *folK*, respectivamente, estão envolvidas nas etapas iniciais da síntese de folatos (**Figura 9**). A *XaDHNA*, uma liase homo-octamérica de 107,5 KDa formada por monômeros de 119 resíduos de aminoácidos, cliva a ligação entre carbonos do substituinte do anel pirazina da molécula 7,8-di-hidroneopterina, liberando um glicolaldeído e 6-hidroximetil-7,8-di-hidropterina. A reação subsequente é catalisada pela *XaHPPK*, uma quinase homo-tetramérica de 71,7 KDa formada por monômeros de 167 resíduos de aminoácidos, que na presença de Mg^{2+} , realiza a transferência do pirofosfato de um ATP ao produto da *XaDHNA*, em um mecanismo bi-bi, formando 6-hidroximetil-7,8-di-hidropterina pirofosfato e liberando AMP.⁵⁹

Figura 9 – Modificações na porção pterina precursora do THF pelas enzimas *XaDHNA* e *XaHPPK*.

Fonte: Elaborada pelo autor.

1.8 Cristalografia de raios X

Os estudos científicos de sólidos cristalinos tiveram início no final do século XVIII com o padre francês Rene-Just Haüy. Seu esforço em compreender o arranjo interno do cristal levou-o a concluir que estes deveriam possuir planos cristalinos e serem estruturas periódicas formadas por pequenos poliedros, aos quais denominou moléculas integrais (do francês, *molécules integrantes*), formulando a Lei dos Índices Racionais.⁶⁰⁻⁶¹ No começo do século XIX, o físico alemão Moritz Frankenheim atribuiu elementos de simetria aos sistemas cristalinos definidos por Friedrich Mohs e Christian Weiss, identificando trinta e duas classes de cristais, hoje conhecidas como grupos pontuais, e possibilitou o francês Auguste Bravais

derivar os 14 tipos de redes de Bravais.^{60,62} Bravais também notou não ser possível explicar todos os 32 grupos pontuais com apenas as 14 redes propostas, mas seria possível se operações de rotação e reflexão fossem adicionadas.^{60,63} Apenas no final do século XIX, Arthur Schoenflies e Evgraf Fedorov catalogaram todos os 230 possíveis grupos espaciais, embora em cristalografia de proteínas, devido a exclusão enantiomérica, apenas 65 grupos espaciais sejam efetivamente observados.^{60,63}

A descoberta dos raios X em 1895 após observações da fluorescência do platinocianeto de bário pelo físico e engenheiro mecânico alemão/holandês Wilhelm Röntgen foi seguida por constantes aplicações desta radiação em diversas áreas da ciência.⁶⁴⁻⁶⁶ O início da década de 1910 foi marcada pela demonstração do padrão de difração em estruturas cristalinas por Max von Laue⁶⁷ e pela derivação de uma equação mais geral que determina a condição única para interferência construtiva de ondas difratadas em diferentes planos cristalinos dado um comprimento de onda e um ângulo de incidência por William Bragg, a Lei de Bragg.⁶⁸⁻⁶⁹

A possibilidade de se produzir cristais a partir de polímeros biológicos foi demonstrada pela primeira vez em meados da década de 1920 pelo químico norte-americano James Sumner, motivando cientistas como Max Perutz e John Kendrew na busca por mais cristais proteicos, a exemplo de hemoglobina e mioglobina.⁶⁶ No final da década de 1950, o trabalho com a mioglobina da baleia cachalote gerou a primeira estrutura cristalina com resolução atômica de uma macromolécula, resolvida utilizando a técnica de substituição isomórfica múltipla, MIR (do inglês, *Multiple Isomorphous Replacement*),⁷⁰ para solucionar o problema das fases, por J. Kendrew,⁷¹⁻⁷² e veio a confirmar as predições feitas sobre possíveis estruturas secundárias feitas quase 10 anos antes por Linus Pauling e Robert Corvey.⁷³⁻⁷⁵ Estes trabalhos resultaram no aparecimento de uma nova área, a biologia molecular, posteriormente denominada biologia estrutural ou biologia molecular estrutural.

O final da década de 1970 foi marcado por um avanço significativo na capacidade de produção heteróloga de proteínas após a descoberta de enzimas de restrição capazes de digerir DNA em regiões específicas, facilitando a clonagem de segmentos gênicos em plasmídeos para expressão em bactéria *Escherichia coli*.^{66,76-77} Antes, apenas proteínas abundantes em tecidos ou órgãos eram utilizadas para estudos, como a lisozima da clara do ovo de galinha, mioglobinas e hemoglobinas.^{66,78} A utilização desta técnica tornou-se

extremamente popular com o desenvolvimento de novos vetores de clonagem e expressão, novas cepas de *E. coli* com diferentes características e sistemas eucariontes de expressão.

A fim de acompanhar o rápido desenvolvimento da biologia molecular, desenvolveu-se novas técnicas de cristalização, como a utilização de sulfato de amônio como principal precipitante.^{79–81} Esta técnica foi amplamente utilizada até meados da década de 1990, quando novos *kits* comerciais e reagentes ganharam espaço. Avanços importantes também foram feitos na engenharia dos dispositivos de geração de raios X, como a utilização de ânodos rotatórios no final da década de 1970.^{66,82} Posteriormente, a *Bremsstrahlung* gerada em aceleradores síncrotron possibilitou o ajuste de comprimento de onda do feixe de raios X pela modulação da energia dos fótons,⁸³ permitindo resolver o problema das fases por métodos diretos utilizando de técnicas como dispersões e espalhamentos anômalos monocromáticos ou com múltiplos comprimentos de onda, SAD (do inglês, *Single-wavelength Anomalous Dispersion*) e MAD (do inglês, *Multiwavelength Anomalous Diffraction*), respectivamente.^{66–72–84–85} Avanços no sistema de detecção, como a utilização de detectores CCD no início da década de 1990,⁸⁶ evoluindo para sistemas de livres de obturador no final da década de 2000, o PILATUS,^{87–89} abriram possibilidades para a cristalografia em larga escala.

Os sistemas de montagem do cristal proteico e técnicas de coleta de imagem também foram importantes para a cristalografia. O desenvolvimento do protocolo de oscilação e rotação do cristal em relação ao feixe de raios X no final da década de 1970 se tornou rapidamente a principal metodologia para coleta de dados.⁹⁰ Em meados da década de 1990, o dano por radiação sofrido pelos cristais foi drasticamente reduzido resfriando-se o cristal em nitrogênio líquido, a aproximadamente 100 K, o que permitiu a utilização de apenas um cristal para a obtenção de um conjunto de dados, eliminando a necessidade de unir conjuntos provenientes de diferentes cristais.^{63,91}

A década de 1990 também foi marcada pelo desenvolvimento de ferramentas computacionais fundamentais para a resolução de estruturas e métodos de avaliação de qualidade de modelo tridimensional.^{92–94} Este desenvolvimento possibilitou uma disseminação da biologia estrutural uma vez que eliminou a necessidade do cristalógrafo dominar conceitos avançados de matemática e física. Iniciativas como o projeto colaborativo computacional número 4, CCP4 (do inglês, *Collaborative Computational Project, number 4*),

⁹² agrupou diferentes soluções para determinação de estruturas cristalográficas, como automatização do método de substituição molecular para determinação das fases. ⁹⁵⁻⁹⁶ O começo da década de 2000 trouxe novos pacotes computacionais, como o PHENIX e Coot, alternativas aprimoradas do CNS e O, respectivamente.

A necessidade de um banco de dados para armazenar as estruturas tridimensionais de proteínas levou o Laboratório Nacional de Brookhaven, EUA, a criar o *Protein Data Bank* (PDB) no começo da década de 1970. Inicialmente com apenas sete estruturas determinadas utilizando difração de raios X, dentre elas a quimiotripsina, carboxipeptidase A, papaína e subtilisina, o banco possui, em abril de 2017, um total de 129.541 estruturas depositadas, sendo 84% destes depósitos referentes a proteínas determinadas por difração de raios X. O banco conta hoje com estruturas de proteínas e ácidos nucleicos determinadas utilizando, além da difração de raios X, a ressonância magnética nuclear (RMN), microscopia eletrônica (ME) e crio-microscopia eletrônica de transmissão (crio-ME).

Objetivos

2 OBJETIVOS

O objetivo deste projeto de Doutorado é realizar estudos estruturais de enzimas da via de biossíntese de folatos de *Xanthomonas albilineans* para o desenvolvimento de novos candidatos a agroquímicos para a cultura de cana-de-açúcar. Esse processo envolveu a elucidação da estrutura tridimensional dos alvos selecionados, padronização de bioensaio, triagem fenotípica e virtual para a descoberta de candidatos a inibidores. Para tanto, foram estabelecidos os objetivos específicos:

- Clonagem das sequências codificadoras das proteínas HPPK e DHNA de *X. albilineans*
- Expressão e purificação das proteínas HPPK e DHNA de *X. albilineans*;
- Elucidação das estruturas tridimensionais das proteínas HPPK e DHNA de *X. albilineans*;
- Padronização e caracterização cinética da proteína HPPK de *X. albilineans*;
- Padronização de um ensaio fenotípico para identificação de inibidores de crescimento da cultura de *X. albilineans*;
- Triagem biológica de compostos, naturais e sintéticos, dos candidatos a inibidores da cultura de *X. albilineans*;
- Triagem virtual de compostos candidatos a inibidores da enzima HPPK de *X. albilineans*

Material e Métodos

3 MATERIAL E MÉTODOS

3.1 Culturas de *Xanthomonas albilineans* – Obtenção, cultivo, caracterização e extração de DNA genômico

As cepas IBSBF1374, IBSBF326, IBSBF654 e IBSBF740 de *X. albilineans* foram adquiridas do banco de fitopatógenos do Instituto Biológico de Campinas. As culturas, liofilizadas, foram recuperadas utilizando o meio YSG sólido⁹⁷ por 5 dias a 28 °C (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) conforme sugerido pelo curador do banco. Após recuperadas, colônias isoladas foram inoculadas em meio seletivo para *X. albilineans* XAS⁹⁸ (1 g.L⁻¹ sacarose, 5 g.L⁻¹ bactopectona, 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) por 5 dias a 28 °C sob agitação constante de 175 rpm.

Para a obtenção do DNA genômico purificado, a cepa IBSBF740 foi escolhida por suas características morfológicas mais próximas das descritas na literatura, como crescimento lento, colônias amareladas e aspecto não mucoidal.¹⁷ O protocolo de extração utilizando o tampão CTAB⁹⁹ foi adaptado e otimizado para o trabalho com cepas *Xanthomonas sp.*. Brevemente, a cepa de *X. albilineans* IBSBF740 cultivada foram centrifugadas a 4.000 g durante 20 min. Em seguida, o meio contendo a bactéria foi ressuspenso em tampão de lise (2% brometo de cetiltrimetil amônio - CTAB, 1,4 M NaCl, 100 mM Tris-HCl pH 8, 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 (16.000 g 10 min). Sobre a fase superior isolada, adicionou-se 1 volume de clorofórmio/álcool isoamílico (24:1) e separou-se novamente por centrifugação (16.000 g 10 min). Os ácidos nucleicos presentes na fase superior foram recuperados pela adição de 1 volume de isopropanol (-20 °C) seguido de incubação a -20 °C 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%.

A caracterização molecular de *X. albilineans* foi realizada baseada na amplificação de uma região espaçadora das subunidades ribossomais 16S-23S com 288 pares de bases.¹⁰⁰ A amplificação dessa região utilizou os oligonucleotídeos iniciadores contidos na **Tabela 1**.

Tabela 1 – Oligonucleotídeos iniciadores utilizados para identificar culturas de *X. albilineans*.

Nome	Sequência	Tm
PGBL1	5'- CTTTGGGTCTGTAGCTCAGG -3'	61
PGBL2	5'- GCCTCAAGGTCATATTCAGC -3'	61

Fonte: Elaborada pelo autor.

3.2 Padronização do ensaio e triagem fenotípica de *X. albilineans*

A padronização dos ensaios e a triagem fenotípica foram conduzidas utilizando-se a incubadora Bioscreen Microbiological Analyser (LabSystem, Helsinki, Finland) em placas de 100 poços (10x10). Para a padronização, a *X. albilineans* foi cultivada utilizando-se o meio seletivo XAS a temperatura de 28 °C e 175 rpm. O crescimento bacteriano no meio de cultura foi monitorado a cada 10 min durante 72 h ($\lambda = 600$ nm).

As curvas de crescimento foram tratadas com a ferramenta GraphPad Prism 6 (GraphPad Software Inc., Califórnia). Os dados foram ajustados ao modelo matemático descrito pela equação de Gompertz.¹⁰¹ Os valores da concentração inibitória mínima (MIC) e concentração não inibitória (NIC) foram determinados calculando-se a área fracional sob a curva em função do logaritmo das concentrações dos controles positivos. A área fracional ($A_{\text{fracional}}$) sob a curva é definida como

$$A_{\text{fracional}} = \frac{A_{\text{composto}}}{A_{\text{controle}}} ; 0 \leq A_{\text{fracional}} \leq 1 \quad (1)$$

onde A_{composto} representa a área sob a curva de uma cultura bacteriana tratada com um determinado composto, e A_{controle} a área sob a curva de uma cultura contendo apenas o

meio XAS.¹⁰² O valor de $A_{\text{fracional}}$ está entre 0, situação na qual o composto inibe completamente o crescimento bacteriano, e 1, situação na qual o composto não apresenta nenhuma atividade inibitória.

Figura 10 – Representação gráfica do cálculo da área fracional.

Fonte: Elaborada pelo autor.

Para validação do ensaio e identificação de controles positivos foram avaliados diferentes os antibióticos: tobramicina, gentamicina, ampicilina, tetraciclina e ampicilina. As concentrações dos antibióticos variaram entre 5 g.L^{-1} e 5 mg.L^{-1} . Foram avaliadas ainda a interferência do DMSO no crescimento bacteriano em concentrações entre 10% e 0,25% de DMSO. Todas as medidas foram realizadas em triplicata e as médias de densidade ótica para cada concentração foram utilizadas para os cálculos dos valores de MIC e NIC.

As triagens fenotípicas utilizaram as mesmas condições experimentais de crescimento empregadas durante a padronização do ensaio. Os compostos triados, recebidos de grupos colaboradores, foram solubilizados em DMSO 100% de modo que a concentração final de DMSO nos ensaios fosse de 1% (1:100). A concentração da solução padrão para todos os compostos foi de 200 mM, exceto nos casos nos quais os compostos apresentavam problemas de solubilidade nessa concentração, quando então adicionava-se um volume de DMSO, reduzindo a concentração à metade até a solubilização completa do composto.

Inicialmente, a triagem dos compostos foi feita utilizando três concentrações de compostos: 2 mM (ou a máxima possível na diluição 1:100) e mais duas diluições seriadas.

Os compostos que apresentam atividade inibitória máxima nesta triagem são submetidos ao experimento de determinação de MIC e NIC. Neste experimento o composto é diluído serialmente a fim de se obter uma variação maior do que quatro ordens de grandeza entre a maior e a menor concentração testada. Os dados são tratados utilizando o ajuste à equação de Gompertz.

3.3 Clonagem dos segmentos gênicos folB e folK

3.3.1 Método da clonagem de ligação independente

Nesta técnica, o segmento gênico (~150 ng) é tratado em uma reação de digestão pela T4 DNA Polimerase (Fermentas) por 30 min a 22 °C contendo apenas o nucleosídeo dATP para a formação das extremidades coesivas, fazendo uso da atividade exonuclease no sentido 3' – 5' da enzima. Os vetores (**Tabela 2**) são tratados em uma reação análoga contendo apenas o nucleosídeo dTTP.

Tabela 2 – Características dos vetores de expressão utilizados.

Vetor de expressão	Proteína de fusão	Promotor	M.M. da proteína fusionada	Protease para clivagem da fusão	Resistência a antibiótico	Origem
pETM11	Cauda 6-His	T7/Lac	3 KDa	TEV Protease	Canamicina	pBR322
pETTrx-1a	Tiorredoxina	T7/Lac	14 KDa	TEV Protease	Canamicina	pBR322
pETNUS-1a	NusA	T7/Lac	57 KDa	TEV Protease	Canamicina	pBR322

Fonte: Elaborada pelo autor.

As sequências dos segmentos gênicos utilizados para a clonagem foram obtidas do banco de dados de sequências de nucleotídeos GenBank sob os códigos XALC_0418 e XALC_0832, referentes à folB e folK respectivamente. Os oligonucleotídeos iniciadores utilizados para a amplificação (**Tabela 3**) foram elaborados utilizando a ferramenta HTP-Oligo Designer.¹⁰³

Tabela 3 – Sequências de oligonucleotídeos iniciadores utilizados para amplificação dos segmentos gênicos de interesse.

Segmento gênico	Sequência	T _m (°C)	Bases
Senso			
folB	CAGGGCGCCATGGATAAAGTCTTATCGAAGGTCTG	59	24
folK	CAGGGCGCCATGACCACCGTGCTACTGAGTC	60	19
Anti-senso			
folB	GACCCGACGCGGTTATCCACACGGGCG	58	12
folK	GACCCGACGCGGTTACCCTGGAGTCCCGG	61	14

Fonte: Elaborada pelo autor.

3.4 Expressão e purificação das proteínas de *X. albilineans*

3.4.1 Testes de expressão das proteínas *XaDHNA* e *XaHPPK*

A avaliação da expressão proteica para as diferentes combinações de genes e vetores foi realizada com cultivos aeróbicos a 37 °C em meio líquido LB (do inglês, *Lysogeny Broth*). As culturas bacterianas foram crescidas até que as densidades ópticas a 600 nm (D.O._{600 nm}) estivessem entre 0,6 e 0,8. Na sequência, a expressão foi induzida pela ativação irreversível do operon lac por 1 mM Isopropil β-D-1-tiogalactopiranosídeo (IPTG). Após a indução, as colônias foram mantidas a 18 °C por 22 h. Ao final do período de expressão, as células foram separadas por centrifugação (4000 g, 45 min, 4 °C) e ressuspensas em tampão de lise (50 mM Tris-HCl pH 8,0, 500 mM NaCl). Em seguida, as células foram incubadas por 30 min com 250 µg.mL⁻¹ de lisozima, 4 mM de DTT e 1 mM PMSF. Após a incubação, as células foram submetidas ao desmembramento ultrassônico e as frações insolúveis e solúveis foram obtidas por centrifugação a 16.000 g por 20 min. O perfil de expressão foi analisado por eletroforese em gel de poliácridamida 12,5% (SDS-PAGE), corado com 0,01% de *Comassie Brilliant Blue G-250*.

3.4.2 Expressão e purificação das proteínas *XaDHNA* e *XaHPPK*

A expressão em larga escala das proteínas utilizou o meio auto indutor ZYM-5052.¹⁰⁴ O pré-inóculo (10 mL para cada 1 L de cultura) provém de uma cultura previamente crescida em meio LB por 16 h a 37 °C. Em seguida, a cultura inoculada em meio ZYM-5052 foi colocada sob agitação constante de 175 rpm a 37 °C e a D.O._{600 nm} foi acompanhada até que seu valor estivesse entre 0,6 e 0,8. Nesse momento, diminui-se a temperatura para 18 °C e manteve-se a agitação por 22 h. O meio auto indutor utiliza fontes de carbono provenientes da glicose, glicerol e lactose. A indução ocorre pela metabolização natural da lactose após esgotar-se a glicose do meio, gerando o ativador do *operon lac*, a alolactose (molécula à qual o IPTG é um análogo). A recuperação da massa celular, lise e obtenção das frações solúveis seguiram o mesmo protocolo utilizado seção 3.4.1.

A construção dos vetores de expressão, todos com cauda de hexa-histidina (6-His-tag), possibilita a utilização do mesmo protocolo de expressão para as três proteínas, independente da proteína de fusão utilizada. Todas as etapas cromatográficas foram realizadas em sistema semiautomatizado ÄKTA Pure 25 (GE Healthcare) com monitoramento da absorbância em $\lambda = 280$ nm (UV). As etapas cromatográficas utilizadas foram: i. cromatografia por afinidade a metal imobilizado, IMAC (do inglês, *Immobilized Metal Affinity Chromatography*), utilizando a resina Ni Sepharose High Performance (GE Healthcare) e TALON Crude (GE Healthcare) para a *XaHPPK* e *XaDHNA*, respectivamente, nas quais se explora a afinidade da HisTag por metais divalentes. Nesta técnica, proteínas com HisTag interagem com o metal divalente imobilizado em uma matriz de agarose, utilizando uma extensão de derivado de ácido carboximetil aspártico (CMA), separando-se de outras proteínas presentes no clarificado (**Figura 11**).¹⁰⁵

A recuperação da proteína de interesse é feita utilizando um gradiente de imidazol (20 – 500 mM), que compete com a HisTag pelo metal liberando a proteína; ii. cromatografia de gel filtração utilizando uma resina Sephadex G-25 fine (GE Healthcare), na qual se explora a diferença de tamanhos entre as proteínas e sais em solução e suas diferentes difusões entre a resina da coluna para a retirada do imidazol; iii. cromatografia por afinidade a metal imobilizado (IMAC). Após a retirada do imidazol, a proteína foi incubada por 20 h a 8 °C com TEV protease (na proporção 1 mg de TEV protease para 20 mg

de proteína de interesse) e 4 mM de DTT, quando a purificação foi realizada em coluna HisTrap HP, e 1 mM de β -mercaptoetanol quando realizada em coluna HiTrap HP TALON Crude. Nessas condições, a TEV protease reconhece a sequência ENLYFQ\S, na qual “\” denota a ligação peptídica clivada,¹⁰⁶ contida na região espaçadora entre a proteína de interesse e a proteína fusionada (que possui a HisTag), com 7 resíduos de aminoácido, clivando a construção, de modo a tornar possível a separação da proteína de fusão, que interage especificamente com o metal imobilizado, e a proteína de interesse, que não interage. iv. cromatografia de gel filtração utilizando a resina Superdex 200, na qual se explora os diferentes tamanhos globulares entre as proteínas em solução e sua difusão pela resina. A pureza da amostra e seu progresso durante as etapas de purificação foram acompanhados por eletroforese em gel de poliacrilamida 12,5%.

Figura 11 – Modelo das interações entre a cauda de hexa-histidina e o metal divalente imobilizado, em uma matriz de agarose.

Fonte: Elaborada pelo autor.

3.5 Caracterização da proteína XaHPPK

3.5.1 Avaliação da estabilidade da proteína XaHPPK

Para avaliar a estabilidade da proteína XaHPPK em diferentes soluções tamponantes, foi utilizada a técnica de fluorimetria de varredura diferencial – DSF (do inglês, *Differential Scanning Fluorimetry*) em um termociclador Mx3005p qPCR System (Agilent) ¹⁰⁷ com os kits comerciais JBScreen FUNDAMENT e JBScreen SPECIFIC (Jena Bioscience) e as condições propostas por NIESEN e colaboradores (**Figura 12**).

	Final	mM	Final	mM	Final	%	Final	pH
A1	Sodium acetate-AcOH	100	NaCl	150	Glicerol	0	5.0	
A2	Sodium cacodylate-HCl	100	NaCl	150	Glicerol	0	6.0	
A3	MES-NaOH	100	NaCl	150	Glicerol	0	6.0	
A4	Imidazole-HCl	100	NaCl	150	Glicerol	0	7.0	
A5	HEPES-NaOH (REFERENCE)	100	NaCl	150	Glicerol	0	7.5	
A6	MOPS-NaOH	100	NaCl	150	Glicerol	0	7.2	
A7	Tricine-NaOH	100	NaCl	150	Glicerol	0	8.0	
A8	Tris-HCl	100	NaCl	150	Glicerol	0	8.1	
A9	Bicine-NaOH	100	NaCl	150	Glicerol	0	8.3	
A10	HEPES-NaOH (REFERENCE)	100	NaCl	150	Glicerol	0	7.5	
A11	CHES-NaOH	100	NaCl	150	Glicerol	0	9.5	
A12	Ethanolamine-NaOH	100	NaCl	150	Glicerol	0	9.5	
B1	CAPS-NaOH	100	NaCl	150	Glicerol	0	10.0	
B2	HEPES-NaOH	100	NaCl	0	Glicerol	0	7.5	
B3	HEPES-NaOH	100	NaCl	50	Glicerol	0	7.5	
B4	HEPES-NaOH	100	NaCl	500	Glicerol	0	7.5	
B5	HEPES-NaOH	100	NaCl	150	Glicerol	5	7.5	
B6	HEPES-NaOH (REFERENCE)	100	NaCl	150	Glicerol	0	7.5	
B7	PIPES-NaOH	100	NaCl	150	Glicerol	0	7.0	
B8	Glycine-NaOH	100	NaCl	150	Glicerol	0	9.0	
B9	HEPES-NaOH	100	NaCl	150	Glicerol	10	7.5	
B10	HEPES-NaOH	100	NaCl	500	Glicerol	5	7.5	

Figura 12 – Condições tamponantes utilizadas para os experimentos de DSF baseadas na proposta de NIESEN.

Fonte: Elaborada pelo autor

Nesta técnica, utiliza-se uma sonda hidrofóbica, o SYPRO *Orange Protein Gel Stain* (ThermoFisher Scientific), para avaliar a desnaturação da proteína em estudo pelo aumento da temperatura. A condição mais estável é aquela na qual a proteína precisa de mais energia para se desenovelar, ou seja, em uma temperatura maior. Os dados foram tratados seguindo a o protocolo descrito por Niesen.¹⁰⁷

Nos experimentos, à 15 µL de XaHPPK a 10 µM, inicialmente em tampão Tris-HCl 20 mM pH 8,0 e 200 mM NaCl, adiciona-se 5 µL de uma solução tamponante, com Sypro Orange 5x, à qual deseja-se testar a mudança de estabilidade. O qPCR Mx3005p foi então programado para estabilizar todas as amostras a 25 °C por 5 min e, por 50 ciclos de 1 min, variar a temperatura à taxa de 1 °C/ciclo.

3.5.2 Determinação do parâmetro cinético K_d para a proteína XaHPPK

Utilizando o protocolo de determinação de K_d aparente desenvolvido por Harmer e colaboradores,¹⁰⁸ variou-se a concentração do substrato 6HMPH2 entre 20 mM e 100 nM utilizando diluição seriada, mantendo-se a concentração de proteína constante e igual a 10 µM. A solução final continha, além de proteína, condição tamponante (Tris-HCl 20 mM pH 8,0 e 200 mM NaCl) e substrato, Sypro Orange 5x, similar ao descrito no item 3.5.1 desta tese. As medidas foram feitas em duplicatas utilizando os mesmos estoques de proteína de substrato.

3.6 Ensaios de cristalização da XaHPPK e XaDHNA

3.6.1 Triagem de condições de cristalização utilizando *kits* comerciais

Após quatro etapas cromatográficas, a proteína XaHPPK que se encontrava pura e monodispersa, foi concentrada utilizando o concentrador Amicon Ultra-15 30 KDa (Millipore). Foram montadas placas Intelliplate 96-2 (Art Robbins Instruments) com duas concentrações distintas, 5 e 10 mg.mL⁻¹, e mantidas a 18 °C utilizando a técnica de difusão de vapor em gota assentada. A triagem foi feita utilizando os *kits* comerciais Morpheus (Molecular Dimensions), *Index*, *Crystal Screen* e *Crystal Screen Cryo* (Hampton Research) e

PEGs I, PEGs II, MPD *Suite*, AmSO4 *Suite*, JCSG *Core Suite* I, II, III e IV (Qiagen). As gotas continham uma mistura 1:1 de proteína e solução dos *kits* comerciais, totalizando 2 μL . O experimento foi acompanhado em dias intercalados por 2 semana, quando se passou a acompanhá-los a cada 7 dias.

As condições do *kit* JCSG *Core Suite* I F3 (200 mM Formato de magnésio, 20% (m/v) PEG 3350) e F5 (200 mM Cloreto de magnésio, 20% (m/v) PEG 3350) produziram cristais na concentração de 10 mg.mL^{-1} de proteína após 10 dias e, a condição A12 (200 mM Acetato de magnésio, 20% (m/v) PEG 3350), após 21 dias. Essas condições foram expandidas variando-se o tampão (Tris-HCl 20mM e HEPES 50 mM), o pH (entre 7,0 e 8,5) e a concentração do agente precipitante PEG 3350 (entre 15 e 25%).

A proteína XaDHNA seguiu os mesmos procedimentos de purificação e triagem de condições de cristalização realizados com a XaHPPK, adicionando-se os *kits* de cristalização SaltRX HT (Hampton Research) e Midas (Molecular Dimensions). A triagem de condições da XaDHNA, utilizando a técnica de difusão de vapor em gota assentada, foi feita com o auxílio do robô de cristalização Gryphon LCP (Art Robbins Instruments), em placas Intelliplate 96-3 (Art Robbins Instruments), com gotas na proporção 1:1, 1:2 e 1:3 de proteína para solução dos *kits*, respectivamente, com volume de 200 nL por parte, totalizando 400 nL, 600 nL e 800 nL. As condições do *kit* SaltRx D2 (2,5 M Nitrato de amônio, 100 mM BIS-Tris propano pH 7,0), D3 (2,5 M Nitrato de amônio, 100 mM Tris-HCl pH 8,5) e D6 (6 M Nitrato de amônio, 100 mM Tris-HCl pH 8,5) produziram cristais na concentração de 15 mg.mL^{-1} de proteína após 28 dias.

3.6.2 Montagem do cristal e coleta de dados

Utilizando uma alça de poli-imida ou nylon (MiTeGen e Hampton Research, respectivamente), os cristais obtidos durante os experimentos de triagem e expansão de condições foram retirados de suas gotas e transferidos para uma solução crioprotetora contendo o tampão de cristalização e 10% de um crioprotetor (e.g., etilenoglicol, glicerol, PEG 200, etc.).

Os cristais de *XaHPPK* provenientes da triagem de condições foram levados ao Laboratório Nacional de Luz Sincrotron (LNLS) – Campinas/SP e submetidos à difração de raios X na linha MX-2, utilizando um detector PILATUS 2M. A coleta foi planejada pelo programa iMosflm 7.1.0 utilizando duas difrações ortogonais,¹⁰⁹ resultando em uma rotação do cristal de 270° coletados em 540 imagens (*fine – ϕ – slicing*; $\phi = 0,2^\circ$), com exposição de 3 segundos por imagem. Os cristais de *XaHPPK* obtidos nos experimentos de expansões das condições identificadas na triagem foram coletados na linha MX-2 do LNLS, e devido às limitações da mesa de coleta, sobre a qual está o detector PILATUS 2M, a coleta foi feita em 360° em 1800 imagens com 2 segundos de exposição por imagem, utilizando a mesma velocidade angular para o goniômetro de 0,2°/2 s.

De maneira similar, os cristais de *XaDHNA* obtidos durante os experimentos de triagem foram retirados de suas gotas e transferidos para uma solução crioprotetora contendo o tampão de cristalização e 25% de glicerol e resfriados a 100 K em nitrogênio líquido e mantidos nessa condição. Estes cristais foram levados ao Laboratório Nacional de Luz Sincrotron (LNLS) – Campinas/SP e submetidos à difração de raios X na linha MX-2, utilizando um detector PILATUS 2M. A coleta foi planejada pelo programa iMosflm 7.1.0 utilizando duas difrações ortogonais, resultando em uma rotação do cristal de 360° coletados em 720 imagens (*fine – ϕ – slicing*; $\phi = 0,5^\circ$), com exposição de 5 segundos por imagem.

3.6.3 Processamento do conjunto de dados e obtenção do modelo tridimensional

Os padrões de difração foram processados e integrados utilizando o pacote XDS,¹¹⁰ que converte as imagens bidimensionais obtidas em um arquivo de intensidades (XDS_ASCII.HKL) e reduzidos utilizando a ferramenta Aimless (CCP4).^{93,146}

Figura 13 – Organograma do processo de obtenção dos dados do experimento de difração de raios X, processamento dos dados, refinamento do modelo tridimensional e deposição no banco de dados de estruturas PDB.

Fonte: Elaborada pelo autor.

A determinação das fases iniciais do modelo da *XaHPPK* foi feita pelo método de substituição molecular utilizando o programa Phaser MR (CCP4), baseando-se na estrutura da proteína homóloga de *Staphylococcus aureus* (PDBID 4CYU) com 40% de identidade sequencial, utilizando os avaliadores LLG e TFZ foram para validar o sucesso da substituição molecular. A determinação das fases iniciais do modelo da *XaDHNA* foi feita pelo método de substituição molecular utilizando o programa Phaser MR (CCP4), baseando-se na estrutura da proteína homóloga de *Vibrio cholerae* (PDBID 3O1K) com 53% de identidade sequencial, utilizando os avaliadores LLG e TFZ foram para validar o sucesso da substituição molecular.

3.7 Descoberta de inibidores

3.7.1 Triagem virtual para a enzima *XaHPPK*

A base de dados ZINC15¹¹² foi selecionada para a triagem virtual. Esta base de dados contém mais de 100 milhões de compostos disponíveis para compra, cujas estruturas 3D estão disponíveis gratuitamente.

A base de dados foi filtrada utilizando as seguintes propriedades: i. formato 3D, na qual os compostos baixados da base de dados vem prontos para experimentos de *docking*; ii. Reatividade *Clean*, na qual são removidos funcionais como aldeídos, tióis, aceptores de Michael e moléculas com padrões PAINS (do inglês, *Pan Assay Interference Compounds*), que

são compostos com alto potencial de interferir no ensaio de atividade (e.g., rodaminas, barbituratos de alquilideno, bases fenólicas de Mannich, hidroxifenilhidrazonas, alquilideno heterocíclicos, etc.).¹¹³ iii. Disponibilidade para compra *Wait OK*, contendo compostos disponíveis imediatamente para compra e disponíveis sob demanda; iv. pH *Reference 7,4*, relevante pois a base foi filtrada pelo subconjunto 3D; v. Carga líquida entre -2 e +2. Outros filtros aplicados à seleção do subconjunto, como propriedades físico-químicas, foram o coeficiente de partição $\text{LogP} \leq 5$ e massa molecular $\text{MM} \leq 500$ Da, e eliminação de moléculas similares que não excluíssem o esqueleto molecular necessário para eventuais estudos de relação estrutura-atividade, SAR (do inglês, *Structure-Activity Relationship*), utilizando o coeficiente de Tanimoto para similaridades maior que 75% em uma rotina própria escrita utilizando a biblioteca *vegan* disponível para R e Python.¹¹⁴⁻¹¹⁵ Desse conjunto de dados filtrados, com aproximadamente 445 mil compostos, uma etapa adicional de preparação molecular e criação de conformações alternativas foi aplicada para adicionar flexibilidade aos compostos da base de dados e estados de protonação alternativos utilizando o programa *Ligand Preparation* (Glide Software, Schrödinger).

Massa molecular (menor que, Daltons)													
LogP (menor que)		200	250	300	325	350	375	400	425	450	500	>500	Totais, por LogP
	-1	46	66	164	61	55	19	11	22	4	11	0	413
	0	220	312	511	385	142	141	104	20	22	43	0	1,680
	1	1,080	2,437	5,456	3,860	3,881	4,176	483	103	93	149	0	20,638
	2	1,174	7,691	24,519	20,174	11,053	2,472	2,383	453	659	1,053	18	70,457
	2.5	443	5,954	24,323	20,518	6,445	3,195	2,858	617	1,044	1,315	16	66,269
	3	213	4,829	27,255	27,035	23,413	4,866	4,487	953	1,283	2,541	50	96,662
	3.5	84	2,591	21,995	22,382	5,562	5,520	6,064	1,310	1,776	3,957	32	71,157
	4	38	787	9,212	15,047	4,505	5,900	7,074	1,537	2,314	5,604	84	51,980
	4.5	5	294	3,034	6,480	3,792	5,443	6,503	1,352	2,558	6,802	109	36,258
	5	0	94	1,145	2,233	3,200	4,545	6,945	991	2,236	8,426	90	29,815
	>5	0	17	561	938	3,095	3,889	5,712	1,651	3,478	19,947	682	0
Totais, por Massa		0	25,055	117,614	118,175	62,048	36,277	36,912	7,358	11,989	29,901	0	445K Protómeros 1.3K Subgrupos

Figura 14 – Distribuição das moléculas utilizadas para a triagem virtual frente à enzima XaHPPK, organizadas por massa molecular e LogP. As características utilizadas para a obtenção deste subconjunto foram: Similaridade tridimensional, sem grupos reativos (*Clean*), disponíveis para compra ainda que com pequeno tempo de espera (*Wait OK*), pH referência, qualquer carga e filtro de Tice.

Fonte: Elaborada pelo autor.

A estrutura da proteína XaHPPK foi utilizada para os experimentos de acoplamento (*docking*) molecular. A proteína foi preparada para este experimento utilizando o programa *Protein Preparation Workflow* (Glide Software, Schrödinger), no qual seus resíduos carregados têm suas cargas determinadas utilizando um campo de forças para aminoácidos (e.g., *Optimized Potential for Liquid Simulations* – OPLS), os hidrogênios foram adicionados e seus possíveis rotâmeros identificados, seguida pela determinação da grade de busca por interações (*grid*) e avaliação de sua representatividade biológica. As interações entre os candidatos a inibidores e os resíduos do sítio ativo foram analisadas, bem como o valor do parâmetro de avaliação *Glide Score*, e estabeleceu-se um limite de corte deste valor para a identificação de potenciais inibidores.

3.7.2 Cálculo de cargas de pequenas moléculas por métodos quânticos

A capacidade preditiva do modo de ligação proteína-ligante está diretamente relacionada com a precisão do cálculo das cargas elétricas da pequena molécula.¹¹⁶ Deste modo, as cargas das moléculas mais bem classificadas na etapa de triagem virtual foram recalculadas utilizando o programa ORCA em uma rotina escrita em Python, em colaboração com o Dr. José Fernando Ruggiero Bachega. Este cálculo de cargas utilizou o método baseado em rede aplicada a cargas derivadas de potenciais eletrostáticos (CHELPG) com a aproximação de Hartree-Fock e a base 3-21G. Com as cargas recalculadas, uma nova rotina de *docking* molecular foi feita utilizando o algoritmo de alta precisão *XP-dock* com cargas recalculadas por métodos híbridos QM/MM e os resultados foram avaliados de acordo com seu *GlideScore*.¹¹⁶⁻¹¹⁸ As rotinas utilizadas estão disponíveis no repositório GitHub, podendo ser acessado pelo endereço <https://github.com/gustalima/orca/>.

3.7.3 Dinâmica molecular para a enzima XaHPPK

Os melhores compostos tiveram as interações com suas respectivas proteínas avaliadas utilizando o método de dinâmica molecular utilizando o pacote GROMACS,¹¹⁹ com a intenção de observar se a interação estática predita pelo algoritmo do programa Glide reproduz uma estabilidade dinâmica com a proteína, o que de fato levaria a inibição de sua atividade. Os parâmetros utilizados para minimização, equilíbrio e evolução dinâmica do sistema estão na **Tabela 4**.

Neste método, acompanha-se a evolução dinâmica de um sistema composto pela biomolécula, pelo ligante e pelo solvente, comumente água e íons, quando submetido a campo de força (e.g., CHARMM, AMBER, OPLS, GROMOS, etc.). Brevemente, as etapas

necessárias para a dinâmica molecular envolvem: i. gerar as topologias para a proteína e para o ligante em estudo, que contém tanto informações estruturais (ligações, ângulos e diedros) quanto não estruturais (tipos de átomos e cargas); ii. definir uma região de solvatação com condições de contorno periódicas, respeitando a convenção de imagem mínima e continuidade; iii. solvatar a região definida e adicionar íons para neutralização das cargas do sistema; iv. minimizar a energia do sistema, garantido que a estrutura que será usada para a dinâmica molecular tenha parâmetros estruturais e não estruturais razoáveis. O número de passos necessários para uma minimização eficiente depende do tamanho e complexidade do sistema analisado; v. equilibrar o sistema para uma temperatura e pressão desejadas, a fim de evitar que as moléculas do solvente mais próximas da proteína ou do ligante tenham temperaturas muito mais altas que o restante dos átomos, o que pode levar o sistema a um colapso; vi. dinâmica molecular. A análise da evolução do sistema proteína-ligante revela uma informação mais precisa sobre suas interações intermoleculares.

Tabela 4 – Parâmetros utilizados para a minimização de energia do sistema, equilíbrio do sistema e dinâmica molecular.

Parâmetro	Valor
Minimização	
Passos de minimização	50.000
Algoritmo	<i>Steepest Descent Minimization</i>
Equilíbrio	
Tempo de simulação	5 ns
Integrador	Clássico (<i>Leap-frog integrator</i>)
Dinâmica	
Modelo de água	<i>Simple Point Charge/Effective</i>
Tipo de caixa de simulação	Cúbica. 2 nm > proteína
Íons	Na ⁺ ; Cl ⁻
Campo de força	OPLS-AA/L <i>all-atom force field</i>
Tempo de simulação	50 ns

Fonte: Elaborada pelo autor.

A fim de maximizar o tempo de simulação pelo custo computacional, os experimentos de dinâmica molecular foram realizados utilizando a biblioteca de programação paralelizável Nvidia CUDA® e openMPI.^{120–122}

Resultados

4 RESULTADOS

4.1 Culturas de *Xanthomonas albilineans* – Obtenção, cultivo, caracterização e extração de DNA genômico

A cepa IBSBF740 de *X. albilineans* adquiridas da Coleção de Culturas de Fitobactérias foi cultivada em meio seletivo XAS por um período de 5 a 7 dias. As características descritas na literatura como crescimento lento, colônias amareladas e aspecto não mucoidal foram identificadas nas colônias em placa.¹⁷ As avaliações de crescimento, caracterização e extração de DNA genômico estão descritas em estudos anteriores, parte da minha dissertação de mestrado.¹²³

4.2 Padronização do ensaio fenotípico de *X. albilineans*

Durante a padronização dos ensaios fenotípicos, avaliou-se a influência do DMSO e de diversos antibióticos no crescimento bacteriano, buscando uma condição ótima para solubilização de compostos a serem testados e um controle positivo, respectivamente. Das concentrações testadas (entre 10% e 0,25% de DMSO), identificou-se que valores menores que 1% de DMSO poderiam ser utilizados sem alteração no crescimento da *X. albilineans* (Figura 15).

Figura 15 – Curvas de crescimento de *X. albilineans* na presença de DMSO.

Fonte: Elaborada pelo autor.

Figura 16 – Curvas de crescimento da *X. albilineans* na presença de antibióticos. (A) Curvas de crescimento na presença de tobramicina (B) Ajuste de Gompertz para os dados de tobramicina (C) Curvas de crescimento na presença de tetraciclina (D) Ajuste de Gompertz para os dados de tetraciclina.

Fonte: Elaborada pelo autor.

A inibição completa do crescimento de *X. albilineans* pela tetraciclina e tobramicina possibilitaram calcular o MIC e NIC para cada um dos compostos, utilizando o ajuste à equação de Gompertz. Variando a concentração dos antibióticos entre 4 mg.mL⁻¹ e 100 ng.mL⁻¹, avaliou-se o crescimento da cultura bacteriana e construiu-se um gráfico do log da concentração de antibiótico pela área fracional **Figura 16**, calculada segundo o descrito no item 3.2. Conforme apresentado na **Tabela 5**, foi possível observar um valor de MIC de $22 \pm 4 \mu\text{g.mL}^{-1}$ e $9,2 \pm 1,4 \mu\text{g.mL}^{-1}$ para os antibióticos tobramicina e tetraciclina, respectivamente. Os valores de NIC calculados para a tobramicina e tetraciclina foram $1,3 \pm 0,2 \mu\text{g.mL}^{-1}$ e $1,2 \pm 0,2 \mu\text{g.mL}^{-1}$, respectivamente.

Tabela 5 – Valores de MIC e NIC para os antibióticos tetraciclina e tobramicina. Os valores foram calculados para cultura de *X. albilineans* IBSF 740 utilizando o método descrito no item 3.2.

Antibiótico	Estrutura	MIC ($\mu\text{g.mL}^{-1}$)	NIC ($\mu\text{g.mL}^{-1}$)
Tobramicina		22 ± 4	$1,3 \pm 0,2$
Tetraciclina		$9,2 \pm 1,4$	$1,2 \pm 0,2$

Fonte: Elaborada pelo autor.

4.3 Triagem fenotípica de *X. albilineans*

A triagem fenotípica representa uma importante fonte de identificação de inibidores pela simplicidade experimental e capacidade de triagem de compostos, seja de original natural ou sintética. Os ensaios em microplacas Honeycomb de 100 poços permitem o teste de uma grande quantidade de compostos isoladamente, medindo-se a turbidez do meio em períodos de tempo predeterminados.

Figura 17 – Estrutura molecular dos compostos com atividade inibitória frente a *Xanthomonas albilineans*. (A) Composto FG011 (B) Composto FG031.

Fonte: Elaborada pelo autor.

Os testes de atividade inibitória foram realizados com uma série de 20 derivados de nitrotiofeno e nitrofurano, gentilmente cedida pelo Prof. Gustavo Trossini da Faculdade de Ciências Farmacêuticas da USP. A triagem identificou dois compostos com atividades promissoras: FG011 e FG031 (**Figura 17**). Os gráficos do crescimento bacteriano mostram que estes dois compostos, na concentração de 2 mM, foram capazes de inibir

completamente o crescimento da *X. albilineans*, condição fundamental para a construção do experimento para determinação do MIC e NIC (Figura 18).

Figura 18 – Curvas de crescimento de *X. albilineans*. (A), (B) representam as curvas de controle positivo, negativo e negativo + DMSO. (C) Curvas de crescimento de *X. albilineans* na presença do composto FG011. (D) Curvas de crescimento de *X. albilineans* na presença do composto FG031. As concentrações dos compostos representam as concentrações utilizadas para a triagem inicial de atividade.

Fonte: Elaborada pelo autor.

Figura 19 – Porcentagem de inibição dos compostos FG011 e FG031 em diferentes concentrações, calculadas a partir da área fracional sob a curva de crescimento.

Fonte: Elaborada pelo autor.

As porcentagens de inibição do crescimento das culturas de *X. albilineans* em diferentes concentrações de cada um dos compostos está representada na **Figura 19**, calculadas a partir da área fracional sob a curva de crescimento após 48 h.

Nesse sentido, novos experimentos serão realizados para a determinação dos valores de MIC e NIC bem como novos derivados serão avaliados para o estabelecimento da relação estrutura-atividade (SAR) para essa série de nitrotiofeno e nitrofurano.

4.4 Clonagem dos segmentos gênicos folB e folK

4.4.1 Sistema de expressão heteróloga utilizando vetores pET

A expressão heteróloga das proteínas *XaHPPK* e *XaDHNA* foi conduzida utilizando o sistema pET (do inglês, *Plasmid; Expression vector; T7 promoter*), um dos sistemas de expressão de proteínas recombinantes mais utilizados. Inicialmente desenvolvido por Studier e colaboradores no final da década de 80,¹²⁴⁻¹²⁵ a versão moderna destes vetores foi otimizada pela Novagen com a adição de regiões de reconhecimento por enzimas de restrição, etiquetas de identificação em ensaios imunoquímicos, genes de resistência a antibióticos para seleção positiva de clones e regiões para facilitação de purificação. Neste sistema, a super expressão ocorre devido à alta seletividade da T7 RNA polimerase de bacteriófago pela sequência de seu promotor e a alta eficiência de tradução mediada pelos sinalizadores de início de tradução do T7 gene 10. A inserção da sequência codificante para a proteína no sistema pET é feita em sentido invertido do promotor T7 e do T7 gene 10. O controle da expressão é feito pela ativação, reversível ou irreversível, do promotor *lacUV5* contido no *operon lac*. Dadas a especificidade do promotor T7, expressões basais do alvo clonado são mínimas e, a alta eficiência de tradução após ativado, levam a proteína clonada a ser o produto proteico nas células em algumas horas.

Figura 20 – Mapa do vetor pET28a(+) com as regiões promotoras, repressoras, resistência a antibiótico e etiquetas de auxílio à purificação. As setas em verde e laranja representam quadros abertos de leitura, coloridos baseado no sentido de leitura, senso ou anti-senso em relação origem de replicação.

Fonte: Elaborada pelo autor.

Mesmo na ausência de IPTG, existe uma baixa expressão da T7 RNA polimerase pelo promotor *lacUV5* em lisógenos de λ DE3. Nos casos de expressão de proteínas tóxicas, essa expressão basal pode ser suficiente para prevenir a manutenção do plasmídeo nestas células, de modo que os vetores atuais do sistema pET, por exemplo o vetor pET-28a(+) (**Figura 20**), contêm também um promotor denominado T7 *lac*,¹²⁶ localizado antes do promotor T7. Associado ao promotor T7 *lac* está localizado o repressor *lac* (*lacI*) orientado em sentido contrário ao promotor T7, que na ausência do IPTG bloqueia tanto o sítio o promotor *lacUV5*, inibindo a transcrição da T7 RNA polimerase pela maquinaria da célula hospedeira, quanto o promotor T7 *lac*, bloqueando qualquer transcrição que a T7 RNA polimerase possa fazer.¹²⁶ A existência de um sítio de clonagem múltipla, MCS (do inglês, *Multiple Cloning Site*), permite a utilização de uma variedade de enzimas de restrição para a

linearização do plasmídeo e subsequente inserção da sequência de interesse. Essa associação de repressão e promoção da transição do fragmento gênico de interesse dentro do plasmídeo reduz significativamente a expressão basal no sistema pET.

4.4.2 Método da clonagem de ligação independente

Para maximizar as chances de sucesso na expressão de proteína solúvel, utiliza-se a clonagem dos segmentos gênicos de interesse em diferentes vetores de expressão, comumente com a adição de uma proteína de fusão (e.g., Nus-A, tiorredoxina, GST, SUMO). A otimização da purificação foi feita pela escolha de vetores que tenham afinidade por algum íon ou molécula específica, como o caso da afinidade por metais divalentes (e.g., Ni^{2+} , Co^{2+} , Cu^{2+} , Fe^{2+}) exploradas pela presença de uma cauda hexa ou deca histidina, a forte interação entre a Glutathione S-Transferase (GST) e a forma reduzida da glutathione exploradas pela adição da etiqueta GST no C-terminal a proteína de interesse. Os genes *folB* e *folK* foram amplificados a partir do DNA genômico por uma reação em cadeia da polimerase – PCR (do inglês, *Polimerase Chain Reaction*). A clonagem foi feita utilizando o sistema de clonagem de ligação independente – LIC (do inglês, *Ligation Independent Cloning*)¹²⁷ com os vetores pETM-11, pETTrx-1a e pETNus-a adaptados originalmente a partir do vetor pET28a (Figura 20)

A clonagem das proteínas *XaDHNA* e *XaHPPK* iniciou-se com a amplificação dos segmentos gênicos *folB* (327 pb, codificando uma proteína com 13,4 KDa) e *folK* (501 pb, codificando uma proteína com 17,9 KDa), respectivamente. Após a confirmação do tamanho dos fragmentos por PCR analisados em gel de agarose 1%, os produtos das reações foram purificados utilizando o kit comercial *Wizard® SV Gel and PCR Clean-Up System* (PROMEGA). Utilizando o sistema de clonagem de ligação independente, os fragmentos foram inseridos nos vetores de expressão pETM11, pETTrx-1a e pETNus-a, confirmados pela análise do sequenciamento dos fragmentos ligados aos plasmídeos (Figura 21).

Figura 21 – Ilustração da estratégia de clonagem de ligação independente. (A1) Segmento gênico amplificado pela reação de PCR. A região em vermelho denota o segmento a ser digerido pela T4 DNA polimerase. (A2) Segmento gênico com as bordas coesivas geradas pela atividade exonuclease 3'-5' da T4 DNA polimerase na presença de dATP. (B1) Vetor de expressão. A região em vermelho denota o segmento a ser digerido pela T4 DNA polimerase. (B2) Vetor de expressão com as bordas coesivas geradas pela atividade exonuclease 3'-5' da T4 DNA polimerase na presença de dTTP. (C) Produto da reação de hibridização. As bases em vermelho representam as regiões com ausência da ligação fosfodiéster entre as bases, a ser realizada pela bactéria em vez da T4 DNA ligase. (D) Passos realizados para a obtenção de um clone com o plasmídeo contendo o segmento gênico de interesse.

Fonte: Elaborada pelo autor.

Após o tratamento dos vetores e dos segmentos gênicos, iniciou-se a reação de hibridização, realizada a temperatura ambiente por 30 min. O produto da reação de hibridização foi transformado em cepas bacterianas *E. coli* DH5α e *E. coli* Rosetta 2 (DE3) para clonagem e expressão, respectivamente. A cepa *E. coli* DH5α é útil para a clonagem de fragmentos gênicos recém ligados a um plasmídeo por possuir uma mutação no gene *hdsR17* que acarreta na eliminação da endonuclease EcoKI, responsável por degradar fitas de DNA não metiladas, embora ainda possua o gene *hdsM* para metilação de fitas de DNA sintetizadas pela célula.¹²⁸ A ausência do gene *recA*, responsável pela recombinação homóloga, reduz consideravelmente eventos de deleção e multimerização do plasmídeo contendo o segmento gênico de interesse,¹²⁹ e a mutação do gene *endA1* inativa

endonucleases celulares que podem degradar o plasmídeo durante a execução de métodos de extração deste.¹³⁰ A cepa *E. coli* Rosetta 2 (DE3) foi desenvolvida para melhorar a expressão de proteína recombinante, principalmente de organismos eucariotos, pela incorporação de tRNA para os códons AUA, AGG, AGA, CUA, CCC, GGA e CGG em um plasmídeo que também confere resistência à cloranfenicol, raramente utilizados naturalmente pela *E. coli* e por isso denominados códons raros. A referência DE3 indica que esta cepa é um lisógeno de λ DE3, isto é, possui uma cópia do segmento gênico da T7 RNA polimerase sob o controle do promotor *lacUV5*.

Os plasmídeos contendo os fragmentos de interesse foram inseridos em células competentes *E. coli* Rosetta 2 (DE3) em uma reação de transformação por choque térmico. Novamente, uma reação de PCR foi utilizada para confirmar o sucesso na clonagem dos segmentos gênicos (**Figura 22**), utilizando como molde para amplificação o produto de uma reação de *miniprep*.

Figura 22 – Eletroforese em gel de agarose 1%. Segmentos gênicos nos diferentes vetores: pETM11, pETTrx-1a pETNus-a; amplificados por PCR.

Fonte: Elaborada pelo autor.

4.5 2-amino-4-hidroxi-6-hidroximetil di-hidropteridina pirofosfoquinase (HPPK)

4.5.1 Expressão e purificação da proteína HPPK de *X. albilineans*

A eficiência da expressão de proteínas em sistemas heterólogos, como por exemplo *E. coli*, é influenciada por fatores biológicos, como a escolha da cepa bacteriana, sistemas de

co-expressão, fatores bioquímicos como sistema de indução de expressão, meio de cultura e, por fim, fatores físico-químicos, como temperatura, agitação e tempo de expressão. A expressão da proteína e *XaHPPK* utilizou o seguinte protocolo: expressão com proteína de fusão Tiorredoxina em meio autoindutor ZYM5052 por 18 h a 20 °C sob agitação constante de 175 rpm.

A purificação da proteína foi realizada em 4 etapas cromatográficas (2 etapas de IMAC e 2 etapas de gel filtração). A primeira etapa cromatográfica consiste na separação de proteínas com afinidade por metal imobilizado, entre elas a proteína de interesse, das proteínas do lisado celular, eluindo a proteína ligada à coluna por meio de um gradiente de imidazol 20-500 mM. A segunda etapa, uma etapa de gel filtração utilizando resina Sephadex G-25 fine em coluna XK 26/400 teve por objetivo a remoção do imidazol da amostra, de modo a adequar o tampão às condições reacionais requeridas pela TEV protease. Após a reação de clivagem da construção 6-His-Trx-*XaHPPK* com TEV protease, uma segunda etapa de IMAC foi utilizada para separar a proteína de interesse, *XaHPPK*, neste momento sem a cauda de hexa-histidina, da proteína 6-His-Trx, TEV protease e a construção remanescente da clivagem, 6-His-Trx-*XaHPPK*. Finalmente, outra etapa de exclusão por tamanho molecular utilizando uma resina Superdex 200 em uma coluna XK 26/1000 foi utilizada para a obtenção de uma amostra monodispersa, fundamental para os experimentos de cristalização.

4.5.2 Avaliação da purificação e estabilidade da proteína *XaHPPK*

Com as etapas de purificação otimizadas foi possível obter um bom rendimento e um alto grau de pureza da proteína *XaHPPK* (**Figura 23**), o que possibilitou a avaliação da estabilidade da *XaHPPK* em diferentes tampões, utilizando o sistema de qPCR Mx3005p (Agilent), proposto por Vedadi e colaboradores, para a realização da técnica de DSF.¹⁰⁷ Nessa técnica, utiliza-se a sonda Sypro Orange que quando em água, tem sua fluorescência extinguida. Ao se aumentar a temperatura da amostra, a desnaturação da proteína expõe as regiões hidrofóbicas, às quais a sonda se liga inespecificamente, aumentando a fluorescência. Em um ensaio no qual as condições tamponantes são variadas, das curvas de fluorescência versus temperatura pode-se identificar condições de maior estabilidade da proteína. Esta técnica, que pode ser usada para se otimizar o tampão de uma dada proteína,

foi aplicada à *XaHPPK* a fim de se identificar uma condição ótima para prosseguir com os experimentos de cristalização (**Figura 24**).

Figura 23 – Cromatograma da Gel Filtração da *XaHPPK* e eletroforese em gel de poliácridamida 12% com as frações coletadas durante a purificação da *XaHPPK*. (A) Fração eluída na primeira IMAC. (B) Fração após a clivagem da proteína fusionada com a TEV protease (C) Fração que não interage com a resina cromatográfica durante a segunda IMAC, contendo a proteína *XaHPPK*. (D) Fração eluída durante a segunda IMAC, contendo a TEV protease, a *XaHPPK* e a construção 6-His-Trx-*XaHPPK* remanescente da reação de clivagem. (E) Fração correspondente ao pico da gel filtração, contendo uma única forma oligomérica da *XaHPPK*.

Fonte: Elaborada pelo autor.

O experimento de triagem foi realizado em duplicatas e permitiu a identificação de duas condições que confere estabilidade para a proteína *XaHPPK* (**Tabela 6**).

Tabela 6 – Condições tamponantes identificadas no experimento de DSF para a proteína *XaHPPK*

	Tampão	Sal	pH	ΔT_m
Condição 1	HEPES 100 mM	NaCl 150 mM	8	7,3
Condição 2	Tris-HCl 50 mM	NaCl 50 mM	7,5	6,8

Fonte: Elaborada pelo autor.

O gráfico da **Figura 24** mostra a leitura de fluorescência versus a temperatura, acrescida da temperatura inicial de 25 °C. As curvas com o ponto de inflexão no maior ciclo correspondem às condições mais estáveis.

Figura 24 – Curvas de fluorescência do Sypro Orange no experimento de DSF para a proteína XaHPPK.
Fonte: Elaborada pelo autor.

4.5.3 Técnica da difusão de vapor em gota assentada

A técnica de difusão de vapor é um dos métodos mais populares para cristalização de proteínas. Nesta técnica, a gota contendo proteína em uma condição tamponante e uma solução mãe (ou solução do poço) gradativamente perde água por evaporação e difusão em um sistema hermeticamente isolado (*i.e.*, não existe trocas de soluto ou solvente entre os meios interno e externo), aumentando a concentração da proteína.^{63,117-118} A difusão ocorre sempre da gota para o poço devido às características experimentais, uma vez que a gota contém uma proporção variável de solução com proteína e solução mãe (e.g., 1:1, 1:2, 1:3, 2:1) de modo que a concentração final dos componentes da solução do poço na gota é reduzida, promovendo a difusão do solvente da gota, menos concentrada, para o poço, mais concentrado.^{63,117-120}

Figura 25 – Diagrama de fases de cristalização. Em vermelho estão as mudanças sofridas pela gota de proteína durante um experimento de cristalização por difusão de vapor.

Fonte: Elaborada pelo autor.

O aumento gradativo da concentração da proteína pode levar uma quantidade razoável de moléculas se associarem no espaço, formando agregados termodinamicamente estáveis, chamado núcleos críticos, propiciando o crescimento de um cristal. Esse evento inicial e fundamental para obtenção de um cristal de proteína é denominado nucleação.^{131–135} Seguido da nucleação, a difusão das proteínas em solução para a superfície do cristal leva ao crescimento ordenado dos núcleos formados, em um evento denominado fase de crescimento, como mostrado no diagrama de fases de cristalização (**Figura 25**). A formação da rede cristalina depende fortemente de interações específicas e espacialmente orientadas entre as proteínas ainda em solução e os núcleos críticos, fazendo deste um evento dependente da homogeneidade oligomérica da amostra de proteína. O crescimento do cristal pode cessar por diferentes motivos, como a depleção de proteína em solução, deformações na estrutura cristalina ou acúmulo de impurezas na superfície do cristal de modo a impedir o arranjo de novas moléculas neste.^{117–120,122}

4.5.4 Triagem e otimização de condições de cristalização da XaHPPK

Uma das etapas que mais geram dificuldades nos projetos de biologia estrutural é a cristalização da macromolécula, dada a quantidade de variáveis que atuam sobre o processo de nucleação e crescimento do cristal. Uma abordagem bem-sucedida é a amostragem multidimensional de condições de cristalização, como arranjos ortogonais, triagem reversa, fatoriais incompletos e matrizes esparsas.^{137–140} Embora os três primeiros métodos sejam interessantes por fornecer uma distribuição sistemática e com parâmetros estatisticamente analisáveis, o sucesso obtido pelas matrizes esparsas as tornaram notavelmente populares.

Tabela 7 – Classificação dos *kits* comerciais utilizados neste projeto quanto ao planejamento de sua formulação.

Fabricante	Nome do <i>kit</i>
Matriz esparsa	
Qiagen	JCSG Core Suite I, II, III, IV;
Hampton	Crystal Screen Cryo; Crystal Screen
Fatorial incompleto	
Molecular Dimensions	Morpheus
Qiagen	MPD Suite; PEGs I, II; AmSO ₄ Suite
Hampton	Index HT

Fonte: Elaborada pelo autor.

As distribuições não sistemáticas dos agentes precipitantes nas matrizes esparsas podem levar a algumas condições no entorno das que estão presentes no *kit* e que poderiam gerar nucleação a serem perdidas, diferente do que acontece no fatorial incompleto por exemplo, embora um levantamento feito com os reagentes mais comuns presentes na base de dados PDB mostra uma tendência, ainda que enviesada pela facilidade na utilização ou custo, de alguns reagentes serem predominantes nas composições das condições de cristalização.^{141–142}

	1	2	3	4	5	6	7	8	9	10	11	12
A	Tris-HCl pH 7,5 200 mM MgCl ₂						Tris-HCl pH 8 200 mM MgCl ₂					
B	Tris-HCl pH 8,5 200 mM MgCl ₂						HEPES pH 6,5 200 mM MgCl ₂					
C	HEPES pH 7,0 200 mM MgCl ₂						Tris-HCl pH 7,5 200 mM Mg(HCO ₂) ₂					
D	Tris-HCl pH 8 200 mM Mg(HCO ₂) ₂						Tris-HCl pH 8,5 200 mM Mg(HCO ₂) ₂					
E	HEPES pH 6,5 200 mM Mg(HCO ₂) ₂						HEPES pH 7,0 200 mM Mg(HCO ₂) ₂					

PEG 3350 - 15 % a 25 %

PEG 3350 - 15 % a 25 %

Figura 26 – Esquema experimental da otimização da condição de cristalização para a XaHPPK.

Fonte: Elaborada pelo autor.

Tabela 8 – Condições de cristalização nas quais foram coletados conjuntos de dados para os cristais da proteína XaHPPK, com suas respectivas formulações.

Condição de cristalização	Tampão	Sal	pH	Precipitante
XaHPPK				
JCSG I - F5	-	0,2 M Cloreto de Magnésio	-	20% PEG 3350
Otimização F5	50 mM Tris-HCl	0,2 M Cloreto de Magnésio	8	25% PEG 3350

Fonte: Elaborada pelo autor.

Para os ensaios de triagem de condições de cristalização, a XaHPPK foi concentrada a 5 e 10 mg.mL⁻¹ e avaliada em doze kits comerciais, totalizando 2.304 condições distintas. As gotas foram acompanhadas sistematicamente e as condições do kit JCSG Core Suite I F3 (200 mM Formato de magnésio, 20% (m/v) PEG 3350) e F5 (200 mM Cloreto de magnésio, 20% (m/v) PEG 3350) produziram cristais na concentração de 10 mg.mL⁻¹ de proteína após 10 dias e, a condição A12 (200 mM Acetato de magnésio, 20% (m/v) PEG 3350), após 21 dias. Estes cristais foram testados na fonte de luz síncrotron MX-2 do LNLS, confirmando serem cristais de proteína. Os padrões de difração porém mostraram-se de baixa qualidade (e.g.,

baixa resolução, anisotrópicos, alta mosaicidade) e incapazes de serem processados ou refinados. Sendo assim, experimentos de otimização foram realizados visando-se a obtenção de cristais maiores, não geminados e com maior poder de difração (**Figura 26**).

Deste experimento, sessenta novas condições foram testadas explorando-se as concentrações de PEG 3350, pH e agente tamponante, mantendo a concentração de proteína em 10 mg.mL⁻¹ e a de sal de magnésio em 200 mM. Cristais com as características desejáveis foram obtidos em quatro novas condições foram obtidas com cristais (**Figura 27**) com as características desejáveis: 1) Tris 100 mM pH 8,5, MgCl₂ 200 mM, PEG 3350 25%; 2) HEPES 100 mM pH 7,0, Mg(CHO₂)₂ 200 mM, PEG 3350 25%; 3) Tris 100 mM pH 7,5, Mg(CHO₂)₂ 200 mM, PEG 3350 23%; 4) Tris-HCl 50 mM pH 8,0, MgCl₂ 200 mM, PEG 3350 25%, que foram submetidos à difração de raios X. Os conjuntos coletados e as formulações das suas condições de cristalização estão na Tabela 8.

Figura 27 – Cristais obtidos durante o experimento de otimização da condição de cristalização e suas respectivas condições.

Fonte: Elaborada pelo autor.

4.5.5 Avaliadores de qualidade do conjunto de dados

A qualidade do conjunto de dados foi avaliada utilizando os parâmetros R_{meas} (**equação 3**) em vez do comumente utilizado R_{merge} (**equação**), uma vez que o valor de R_{merge} para conjuntos com alta redundância não é capaz de representar a real qualidade destes. Isso ocorre porque o somatório dos termos positivos da **equação 2** é divergente, de modo

que conjuntos redundantes, nos quais se espera que os valores médios calculados para cada elemento I_{hkl} seja mais próximo do valor esperado, tem uma avaliação pior do que conjuntos com baixa redundância.¹⁴³⁻¹⁴⁴ A solução para essa limitação do avaliador R_{merge} foi proposta do Diederichs e Karplus,¹⁴³ com um fator de correção baseado na multiplicidade, $\sqrt{n/(n-1)}$, que ajusta o valor de R_{meas} para valores similares em conjuntos de baixa e alta redundâncias (**equação 3**). Este fator de correção deve estar dentro do somatório em (h,k,l) , de modo que cada reflexão tenha sua correção baseada na própria redundância em vez de uma correção para a redundância média do conjunto de dados. Isso também implica em um valor de R_{meas} de um conjunto de dados formado pela união de outros conjuntos próximo à média dos valores de R_{meas} de cada conjunto individualmente, que embora os valores de $R_{meas} > R_{merge}$ para um mesmo conjunto de dados, a utilização do R_{meas} fornece indicativos não tendenciados por redundância e simetria do conjunto de dados.¹⁴³

$$R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,(j)} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,(j)}} \quad (2)$$

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_j |I_{hkl,(j)} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,(j)}} \quad (3)$$

O valor da intensidade I é obtido pela seguinte relação:

$$I = \frac{\sum \frac{I_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} ; \sigma^2 = \frac{1}{\sum \frac{1}{\sigma_i^2}} \quad (4)$$

4.5.6 O problema das fases

O problema das fases, solucionado pelo método de substituição molecular, decorre de uma incapacidade instrumental de se detectar toda a informação contida nos fótons difratados por um cristal, uma vez que os tanto métodos antigos como a utilização de chapas fotográficas e placas de imagem quanto métodos mais modernos como CCD são apenas

capazes de fornecer informação sobre a intensidade dos fótons. Em um detector CCD, a detecção é feita pela medida da carga acumulada em cada pixel do detector, gerada pela energia transferida pelo fluxo dos fótons que colidem com uma determinada superfície dos fotopontos, que pode ser calculada pelo vetor de Poynting $\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B}$, onde μ_0 é constante magnética, $\vec{E}$ é o vetor campo elétrico e $\vec{B}$ o vetor campo magnético.¹⁴⁵⁻¹⁴⁶ Este sinal é processado por um conversor analógico-digital integrado ao CCD resultando em um valor de intensidade por pixel.¹⁴⁷

$$F_{h,k,l} = \sum_{j=1}^{\text{átomos}} f_j e^{[2\pi \cdot i(hx_j + ky_j + lz_j)]} \quad (5)$$

A intensidade de uma onda de raios X difratada por um cristal de proteína é proporcional ao quadrado do fator de estrutura deste cristal, $|F_{hkl}|$, que é um escalar.⁶³ O fator de estrutura (**equação 5**) é um somatório dos fatores de espalhamento atômicos f_j em todos os átomos j da rede cristalina. Os fatores de espalhamento atômicos descrevem como cada tipo de átomo espalha radiação em função do comprimento de onda e ângulo de difração, conseqüentemente a amplitude da onda difratada é proporcional ao número atômico Z , o que explica a dificuldade de se observar espalhamentos dos átomos de hidrogênio. Essa característica do fator de espalhamento atômico também é explorada em conjunto com os efeitos de dispersão anômala para outra técnica de determinação de fases, a MAD.¹⁴⁸ A **equação 6** também possibilita calcular o fator de estrutura para uma dada distribuição atômica conhecida, como por exemplo uma estrutura proteica.¹⁴⁸⁻¹⁵⁰

Em cristalografia, o interesse é pelo inverso, isto é, conhecendo-se os fatores de estrutura é possível calcular a distribuição eletrônica da macromolécula. Matematicamente, a densidade eletrônica $\rho_{(x,y,z)}$ pode ser obtida pela transformada de Fourier da **equação 6**, resultado na **equação 6**. A informação fornecida pelo detector de raios X, a intensidade da onda difratada, é o produto escalar do fator de estrutura por seu complexo conjugado, $I = F \cdot F^*$, o que implica na impossibilidade física de se obter o segundo termo da exponencial da **equação 6**, $i\phi(hkl)$, resultando no chamado problema das fases.¹⁴⁸

$$\rho_{x,y,z} = \frac{1}{V} \sum_h \sum_k \sum_l |F_{h,k,l}| e^{-2\pi i(hx+ky+lz)+i\phi(hkl)} \quad (6)$$

4.5.7 Avaliação da qualidade do modelo tridimensional

A avaliação da qualidade do modelo final refinado foi feita utilizando os parâmetros R (também denominado R_{work}) e R_{free} , ambos calculados utilizando a **equação 7**. Este parâmetro compara os fatores de estrutura obtidos a partir do conjunto de dados com o fator de estrutura calculado a partir do modelo (arquivo de coordenadas) da estrutura de interesse,¹⁴⁹⁻¹⁵⁰ definindo um valor de similaridade normalizado entre estas duas grandezas, de modo que para modelos que reproduzem satisfatoriamente a estrutura da proteína cristalina, o fator de estrutura observado F^{obs} será próximo ao valor do fator de estrutura calculado F^{calc} , diminuindo o valor do parâmetro R e R_{free} .¹⁵¹⁻¹⁵² Os valores de R_{free} são calculados baseados num conjunto fixo escolhido aleatoriamente a partir das reflexões originais do cristal, e é usado para avaliar o ajuste excessivo do modelo ao conjunto de dados. Os valores de R_{free} serão sempre maiores do que os valores de R pois o modelo gerado não utilizou estas reflexões em sua construção, mas não arbitrariamente maiores, uma vez que um modelo preditivo deve se ajustar a todo o conjunto de dados e não apenas àqueles que o geraram.¹⁵¹ Os valores típicos para estes parâmetros estão associados à resolução do modelo gerado, mas comumente entre 0,15 e 0,3 para estruturas entre 0,5 e 4 Å (**Figura 28**).¹⁵³

$$R, R_{free} = \frac{\sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}|}{\sum_{hkl} F_{hkl}^{obs}} \quad (7)$$

Figura 28 – Estatística das estruturas depositadas no banco de dados PDB. Os gráficos foram traçados com os fatores R e R_{free} em função da resolução da estrutura depositada, analisados em abril 2017.

Fonte: Elaborada pelo autor

4.5.8 Definição da resolução da estrutura final dos modelos

A definição da resolução do modelo foi feita baseando-se nas discussões propostas por Karplus e Diederichs em seus trabalhos recentes.¹⁵⁴⁻¹⁵⁵ Nestes trabalhos discute-se a influência do corte de resolução nos diferentes estágios da resolução da estrutura de uma macromolécula (e.g., processamento dos dados, refinamento do modelo), bem como quais parâmetros observar para definir um modelo como satisfatório e quais não devem ser utilizados para definir qualidade do modelo ou do conjunto de dados. Neste trabalho, a resolução da estrutura final foi definida como a resolução máxima a partir da qual não se adiciona informações estatisticamente relevantes e maximiza a relação $R_{\text{work}}/R_{\text{free}}$ como proposto por Evans e Diederichs, respectivamente.^{93,141,143}

O conjunto de dados não sofreu cortes de resolução durante seu processamento, uma vez que estes cortes podem melhorar a precisão do conjunto de dados, consequentemente melhorando os parâmetros de qualidade do conjunto de dados, mas diminuem a exatidão das medidas, aumentando o efeito negativo causado por erros sistemáticos os quais frequentemente não são mensuráveis (e.g., vibração do cristal devido ao fluxo de nitrogênio líquido, instabilidade no fluxo e direção do feixe, tempo morto no detector).^{141-142,144} Outro parâmetro observado foi o coeficiente de correlação interna do conjunto de dados, denominado $CC_{1/2}$ (**equações 8 e 9**). Este parâmetro analisa dois subconjuntos aleatórios do conjunto de dados e compara os valores das intensidades obtidas. Por ser fundamentado em um tratamento estatístico, é possível utilizar outros conceitos estatísticos em conjuntos cristalográficos bem como comparar os parâmetros obtidos em diferentes momentos da resolução da estrutura, como por exemplo a qualidade do modelo final com a qualidade do conjunto de dados.¹⁵⁴ A relevância da correlação dos subconjuntos está estritamente relacionada com o número de observações dentro deste subconjunto de modo que, num caso real no qual se possui cerca de 1000 reflexões únicas por camada de resolução, um $CC_{1/2} > 0,1$ é significativo com $p > 0,01$. Deste modo, uma definição arbitrária da relação sinal ruído $I/\sigma(I)$ para o conjunto de dados não foi o fator limitante para o corte de resolução, embora a aproximação proposta por Karplus e Diederichs correlacione as duas grandezas (**equação 9**).¹⁵⁴

$$CC_{1/2} \sim \frac{1}{1 + \frac{4}{\left\langle \frac{I}{\sigma} \right\rangle_{mrgd}^2}} \quad (8)$$

O modelo refinado foi definido como a estrutura para a proteína de estudo quando ciclos sucessivos de refinamento não mudaram significativamente os parâmetros de avaliação de qualidade. Para confirmar os valores de R_{free} e R_{work} como próximos ao melhor possível para o conjunto de dados, comparou-se os valores de CC^* (**equação 11**) com os valores CC_{work} e CC_{free} .^{141-142,144}

$$CC_{1/2} = \frac{\langle I^2 \rangle - \langle I \rangle^2}{\langle I^2 \rangle - \langle I \rangle^2 + \sigma_\epsilon^2} \quad (9)$$

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}} \quad (10)$$

Onde σ_ϵ erro médio dentro de um dos subconjuntos e CC^* é o coeficiente de correlação calculado para intensidades de um conjunto de dados agrupado (*merged*).¹⁵⁴

4.5.9 Coleta e processamento de dados de difração de raios X para a XaHPPK

Devido às características da mesa de coleta da linha MX-2 do LNLS, foram coletados 360° do cristal de XaHPPK na condição A12 da otimização em 720 imagens, com exposição de 3 s por imagem, a uma distância de 135 mm entre o cristal e o detector PILATUS 2M, com resolução de borda de 1.9 Å. O conjunto de dados foi processado utilizando o programa XDS, o qual identificou a rede de Bravais como ortorrômbica primitiva, com o grupo pontual P222, embora a identificação exata do grupo espacial requiera a determinação estrutural.

A redução dos dados foi feita utilizando a ferramenta Aimless (CCP4), que reúne as funções encontradas nos programas SCALA e Pointless. Este último utiliza as restrições geométricas impostas pelas dimensões da célula unitária identificada pelo XDS para confirmar a escolha do grupo espacial. O Aimless determina um conjunto de reflexões únicas, obtidas pela união das reflexões parciais (aquelas reflexões contidas em diferentes

imagens, mas proveniente do mesmo átomo) e da média das reflexões totais, incluindo as relacionadas por simetria interna. A **Tabela 9** mostra os parâmetros cristalinos e avaliadores estatísticos da estrutura final. A determinação do conteúdo da cela unitária foi feita pelo cálculo do coeficiente de Matthews (CPP4), resultando numa probabilidade de 89% para 4 moléculas na cela unitária e composta por 42% de solvente.

Tabela 9 – Parâmetros de cela unitária, simetria do cristal, qualidade do conjunto de dados e estatística do processamento XaHPPK. Entre parênteses estão os valores para a camada de mais alta resolução.

	XaHPPK
Comprimento de onda	1,459
Resolução	32,0 - 2,3 (2,382 - 2,3)
Grupo espacial	P 2 ₁ 2 ₁ 2 ₁
Célula unitária	(a,b,c) = 54,8 73,7 158,0 (α,β,γ) = 90 90 90
Reflexões totais	57727 (5371)
Reflexões únicas	28902 (2693)
Multiplicidade	2,0 (2,0)
Completeza (%)	98,84 (94,05)
I/σ(I) médio	29,24 (1,66)
B-fator Wilson	45,55
R_{meas}	0,023 (0,20)
CC_{1/2}	1 (0,981)
R_{work}	0,18 (0,22)
R_{free}	0,23 (0,30)
Resíduos da proteína	611
RMS (ligações)	0,012
RMS (ângulos)	1,12
Ramachandran favoráveis (%)	97,63
Ramachandran permitidos (%)	2,20
Ramachandran proibidos (%)	0,17
Rotâmeros outliers (%)	2,43
B-fator médio	53,40
B-fator Macromoléculas	53,46
B-fator Ligantes	85,00
B-fator Solvente	43,97
Número de grupos TLS	27

Fonte: Elaborada pelo autor.

A sequência da XaHPPK foi obtida do bando de estruturas primárias UniProt (D2UCY7_XANAP) e utilizada como sequência de entrada para a busca por estruturas depositadas no PDB na ferramenta BLASTp. Dentre as estruturas identificadas pelo BLASTp, a proteína tetramérica XaHPPK homóloga de *S. aureus* 4CYU com 40% de identidade foi utilizada como molde para a substituição molecular por meio do programa Phaser MR. O algoritmo de máxima verossimilhança utilizado pelo Phaser MR avalia a concordância entre os dados e o modelo gerado por meio de operações de rotação, translação, empacotamento e refinamento. A avaliação da qualidade do modelo proferido pelo logaritmo do ganho de verossimilhança – LLG (do inglês, *log of likelihood gain*) indica o quão melhor a solução encontrada pelo Phaser é melhor que uma solução randômica para os dados apresentados. Valores de LLG maiores que 40 indicam soluções corretas para o conjunto de dados. O valor de Z-score para a função de translação – TFZ (do inglês, *translation function Z-score*) para que uma solução seja definitiva são TFZ > 8. A busca automatizada do Phaser considerou a composição da célula unitária identificada pelo coeficiente de Matthews e identificou uma solução com valor de LLG igual a 259,13 e TFZ igual a 8,9.

4.5.10 Estrutura da XaHPPK de *Xanthomonas albilineans*

A estrutura da XaHPPK é formada por 4 monômeros idênticos de 167 resíduos cada, dos quais 154 estão modelados (**Figura 29**). Cada um dos monômeros, que possui 7 fitas- β e 5 hélices- α , pertence à classe das alfa-beta proteínas (a+b), formada principalmente por fitas- β antiparalelas separando as hélices- α . O enovelamento destes monômeros é do tipo ferredoxina, com o padrão de estruturas secundárias $\beta\alpha\beta\beta\alpha\beta$ (**Figura 30**).

Estas estruturas secundárias são formadas por dois grampos $\beta\alpha\beta$ se conectam por meio de interações de hidrogênio entre as fitas β_1 e β_3 , formando uma folha- β de 4 fitas antiparalelas (**Figura 31**). O sítio de ligação aos substratos se situa logo acima do átomo de magnésio (em verde, **Figura 31 B**). Em outras estruturas depositadas no PDB com proteínas homólogas, é possível identificar dois átomos de Mg^{2+} quando o substrato está presente na estrutura, bem como a alça que conecta os resíduos Asp78 e Asp88. O resíduo Phe45 e Ala43 estão dispostos na alça entre as β_3 e β_4 , que possui um alto B-fator devido a sua participação na acomodação do substrato. Estes resíduos possuem densidade eletrônica

razoavelmente bem definida na subunidade A do tetrâmero (não modelados nas outras subunidades).

Figura 29 – Alinhamento da sequência completa da *XaHPPK* com a estrutura depositada 5VSP. As regiões não modeladas estão representadas pelos espaços no alinhamento.

Fonte: Elaborada pelo autor.

Figura 30 – (A) Topologia da estrutura da *XaHPPK* gerada utilizando o PDBsum.¹⁵⁸ (B) Representação da estrutura primária da proteína *XaHPPK* e suas respectivas estruturas secundárias.

Fonte: Elaborada pelo autor.

A topologia da estrutura mostra que associado ao enovelamento do tipo ferredoxina, que se estende até o resíduo Cys101 aproximadamente, existe um grampo- β formado pelas fitas $\beta 5$ e $\beta 6$. A estrutura ainda possui três hélices- α ($\alpha 3$, $\alpha 4$ e $\alpha 5$) que fazem parte da interface entre os monômeros, que se estendem por uma região de alça até o último resíduo

modelado Ala163. O mapa detalhado dos contatos realizados em cada uma das interfaces, cristalográficas ou não, foi elaborado utilizando a ferramenta ENDScript (Figura 32). 160

Figura 31 – Representação da estrutura da *XaHPPK* com a disposição biológica, depositada no PDB sob o código 5VSP. (A) Diagrama de Ramachandran para a *XaHPPK*, gerada na ferramenta RAMPAGE.¹⁵⁹ (B) Monômero da *XaHPPK* com detalhe da coordenação do Mg^{2+} pelos resíduos Aps92 e Asp94 e duas moléculas de água. (C) Visão superior do homo-tetrâmero da *XaHPPK*, sob um dos eixos de simetria de ordem 2 e visão lateral com detalhe da interface entre as cadeias A e B.

Fonte: Elaborada pelo autor.

No mapa dos contatos dos monômeros que compõem o homo-tetrâmero (**Figura 32**), as letras A – Z com fundo amarelo representam os contatos entre resíduos em interfaces não cristalográficas. As letras A – Z em itálico representam os contatos que ocorrem em interfaces cristalográficas. O símbolo # identifica contatos que estão em um eixo de simetria de ordem 2, no caso da *XaHPPK*. As letras em vermelho identificam os contatos a uma distância menor que 3.2 Å, enquanto letras em preto para distâncias entre 3.2 Å e 5.0 Å. As legendas *hyd* e *acc* representam a hidrofobicidade e a acessibilidade ao solvente de cada um dos resíduos, respectivamente.

Figura 32 – Mapa dos contatos dos monômeros que compõem o homo-tetrâmero. As letras A – Z com fundo amarelo representam os contatos entre resíduos em interfaces não cristalográficas. As letras A – Z em itálico representam os contatos em interfaces cristalográficas. O símbolo # identifica contatos que estão em um eixo de simetria de ordem 2, no caso da *XaHPPK*. As letras em vermelho identificam os contatos a uma distância menor que 3.2 Å, enquanto letras em preto para distâncias entre 3.2 Å e 5.0 Å.

Fonte: Elaborada pelo autor.

4.5.11 Comparação da *XaHPPK* com suas proteínas homólogas

A comparação de seis proteínas homólogas à *XaHPPK* disponíveis no PDB (**Tabela 10**) mostra uma baixa conservação nas alças ℓ_1 (A41-A49) e ℓ_2 (G78-D88), responsáveis pelo ajuste induzido aos substratos.¹⁶⁰ O sítio de ligação ao ATP e 6HMPH2 apresenta alta

os monômeros sensivelmente diferentes dos observados na XaHPPK. Na XaHPPK os monômeros interagem por contatos polares entre α_1 de um dos monômeros com a α_6 de um monômero adjacente, enquanto na estrutura 5ETL os contatos ocorrem entre ambas as alças ℓ_1 e ℓ_2 de monômeros adjacentes, formando uma superfície de contatos mantida pelas interações ℓ_2 - ℓ_1 e ℓ_1 - ℓ_2 que mantém a forma biológica dimérica. Na estrutura 4CYU, apenas contatos cristalográficos fracos entre α_6 - ℓ_1 e β_2 - ℓ_2 de monômeros justapostos interagem. Todos os contatos foram preditos pela ferramenta PISA¹⁶¹ e inspecionados no PyMOL.

Tabela 10 – Comparação entre HPPKs homólogas com mais de 40% de identidade disponíveis no PDB com a XaHPPK. Os valores de RMSD são calculados pela superposição das estruturas da XaHPPK com a estrutura homóloga em destaque.

PDB ID	Organismo	Identidade	RMSD	Disposição biológica
5VSP	<i>Xanthomonas albilineans</i>	-	-	Tetrâmero
4CYU	<i>Staphylococcus aureus</i>	41%	0,8 Å	Monômero
3QBC	<i>Staphylococcus aureus</i>	41%	1,0 Å	Monômero
5ETQ	<i>Staphylococcus aureus</i>	41%	0,7 Å	Monômero
2QX0	<i>Yersinia pestis</i>	40%	0,9 Å	Dímero
1CBK	<i>Haemophilus influenzae</i>	40%	1,0 Å	Dímero
5ETL	<i>Escherichia coli</i>	40%	1,2 Å	Monômero

Fonte: Elaborada pelo autor.

4.5.12 Determinação do parâmetro cinético K_d aparente para a proteína XaHPPK

A estabilidade conferida por um ligante ao interagir com uma proteína pode ser medida indiretamente pelo incremento na energia necessária para desenovelar o complexo proteína-ligante. A técnica de DSF, por meio da medida de temperatura de desnaturação, pode ser utilizada para a determinação do K_D aparente.¹⁰⁸⁻¹⁶² Brevemente, analisa-se a variação da temperatura de desnaturação (ΔT_m °C) em um gráfico de ΔT_m vs log([ligante]) para diferentes concentrações do ligante. Em seguida, faz-se o ajuste da curva utilizando o modelo de cooperação simples descrito por Vivoli e colaboradores,¹⁰⁸ na qual o parâmetro determinado pelo ajuste é o valor de K_D de acordo com a **equação 11**:

$$y = A + (B - A) * \left(\left(\frac{x}{K_d} \right)^n \cdot \left(\frac{1}{1 + \left(\frac{x}{K_d} \right)^n} \right) \right) \quad (11)$$

onde **A** é T_m sem ligante, **B** é o T_m na maior concentração de ligante, **n** é o coeficiente de cooperação de Hill, **K_d** é a constante de dissociação e **x** é a variável concentração de ligante.

A otimização dos parâmetros e determinação do K_d aparente foram feitos utilizando o programa GraphPad Prism 6 e sua ferramenta de personalização de curvas.

A constante de dissociação (K_d) mede a afinidade de uma proteína por uma pequena molécula, por exemplo, um ligante. Dentre os fatores que influenciam o valor dessa constante, destaca-se as interações não-covalentes, como ligações de hidrogênio, forças de van der Waals e interações hidrofóbicas. A medida do K_d aparente utilizando DSF é vantajosa quando comparado a técnicas como ITC (do inglês, *Isothermal Titration Calorimetry*) e SPR (do inglês, *Surface Plasmon Resonance*) por ser um experimento menos complexo e pouco dispendioso.

Figura 34 – Curvas de desnaturação térmica da enzima XaHPPK na presença do substrato 6HMPH2. A concentração do substrato foi variada entre 20 mM e 100 nM.

Fonte: Elaborada pelo autor.

Os dados foram ajustados ao modelo de interação cooperativa simples de ligante da **equação 11** com os parâmetros de assíntota superior $B = 42,64\text{ }^{\circ}\text{C}$, assíntota inferior $A = 33,94\text{ }^{\circ}\text{C}$ e concentração de proteína $P = 0,10\text{ mM}$. O ponto de inflexão da curva ajustada representa o equilíbrio bioquímico " $X\alpha\text{HPPK} + 6\text{HMPH2} \rightleftharpoons X\alpha\text{HPPK-6HMPH2}$ ", que é por definição a concentração de substrato igual ao valor de K_d . Para a enzima $X\alpha\text{HPPK}$ em relação ao substrato 6HMPH2, o valor desta constante cinética é $K_d = 97 \pm 3\text{ }\mu\text{M}$ (**Figura 35**). Quando comparado aos valores valor superior aos identificados para HPPKs homólogas. Nas proteínas homólogas existe uma predominância de resíduos carregados no sítio ativo, enquanto na $X\alpha\text{HPPK}$ existe apenas o resíduo Asp46 presente no $\ell 1$, o que pode explicar o arranjo tetramérico cooperativo da $X\alpha\text{HPPK}$ a fim de superar a menor afinidade pelo substrato e aumentar a atividade da proteína. A análise da curva de T_m pela concentração de substrato em diversos modelos de interação mostrou o melhor ajuste ao de cooperação simples do substrato com coeficiente de Hill calculado igual a 1,6.

Figura 35 – Ajuste dos dados de desnaturação térmica da enzima $X\alpha\text{HPPK}$ na presença do substrato 6HMPH2 utilizando o modelo de cooperação simples do substrato.¹⁰⁸ O ponto de inflexão da curva é numericamente igual à constante de dissociação aparente $K_d = 97 \pm 3\text{ }\mu\text{M}$.

Fonte: Elaborada pelo autor.

4.5.13 Triagem virtual de ligantes

4.5.13.1 Descoberta de inibidores baseado na estrutura do alvo

A descoberta de inibidores baseado na estrutura do alvo, SBDD (do inglês, *Structure-based Drug Design*) é a técnica de desenvolvimento de moléculas bioativas guiada por estruturas de alvos biológicos para a identificação de pequenas moléculas que possam modular a atividade catalítica deste alvo e que, em algum grau, possa ser otimizada para a obtenção das propriedades farmacocinéticas e farmacodinâmicas desejadas.¹⁶³⁻¹⁶⁴ A capacidade de diminuir custos e tempo de pesquisa para a identificação de candidatos a novos inibidores colocou as metodologias computacionais em destaque, sendo parte integrante tanto em indústrias farmacêuticas como laboratórios acadêmicos.

A associação da técnica SBDD à triagem virtual de ligantes, VLS (do inglês, *Virtual Ligand Screening*), é uma das abordagens mais promissoras para a identificação de ligantes para alvos biológicos.^{149,151} Nesta metodologia, utiliza-se programas capazes de prever interações entre proteína e pequenas moléculas (e.g., Autodock, Glide, GOLD) e suas diferentes funções de pontuação para prever e avaliar a energia de interação de um conjunto de moléculas, que buscam varrer um espaço químico abrangente. Apesar da dificuldade em se avaliar resultado de triagens virtuais em bibliotecas de compostos muito heterogêneas e dos desafios intrínsecos dos campos de força moleculares em prever com exatidão as diversas funções químicas presentes nos compostos triados, a técnica SBDD é superior a métodos baseados na estrutura do ligando como a análise comparativa de campo molecular, CoMFA (do inglês, *Comparative Molecular Field Analysis*).^{149,152-155}

4.5.13.2 Acoplamento molecular

A ferramenta utilizada para os experimentos de acoplamento molecular foi o Glide (do inglês, *Grid-based Ligand Docking with Energetics*), parte do Schrodinger Maestro, escolhido baseado na máxima capacidade preditiva e capacidade de personalização das rotinas, não em custo computacional.¹⁷⁰⁻¹⁷⁶ O algoritmo do Glide foi desenvolvido para ter resultados similares às buscas exaustivas de posição, conformação e orientação possíveis

para um dado ligante, porém com custo computacional suficientemente baixo de modo que seja possível utilizá-lo para triagem de compostos em larga escala. Para isso, uma série de filtros hierárquicos são aplicados.

O primeiro deles faz uma busca exaustiva no sítio de ligação definido por possíveis orientações e regiões de interação, construindo uma rede de pontos separados por 2 Å e determinando pontos dentro desta rede, denominados pontos do sítio, para os que satisfizeram a condição de proximidade entre um ponto na superfície da proteína ao centro de massa de um determinado ligante.¹⁷⁷⁻¹⁷⁸ O segundo filtro analisa as posições atômicas do ligantes em relação à superfície da proteína para uma variedade pré-definida de orientações do ligantes em relação à proteína, eliminando as orientações com muitos choques estéricos. Em seguida, o filtro analisa as rotações das orientações escolhidas pelo filtro anterior e identifica átomos capazes de fazer interações de hidrogênio com a proteína ou que possam interagir com átomos metálicos, gerando um subgrupo e calculando a função de pontuação para este grupo baseando-se apenas nestas interações. Caso estas interações tenham um valor suficientemente bom, todas as interações do complexo proteína-ligante são avaliadas e o *GlideScore* é definido para cada pose. A função de pontuação derivada da função gananciosa utilizada pelo Glide avalia não apenas sua posição em relação ao receptor, mas também pela melhor pontuação possível movendo este átomo ± 1 Å em x, y, ou z, de maneira a eliminar as limitações impostas pela rede espaçada de 2 Å criada pelo primeiro filtro. A última etapa do segundo filtro é reclassificar as melhores poses utilizando um refinamento rígido do ligando em torno do seu centro de massa.¹⁷⁷⁻¹⁷⁸ O terceiro filtro minimiza a energia das poses definidas pelo segundo filtro utilizando um campo de força molecular, como o OPLS-AA, possibilitando rotações, translações e torções no ligante, permitindo que as melhores poses sejam submetidas a uma nova amostragem de conformações melhorando sua energia de interação com a proteína. Finalmente, as poses minimizadas são avaliadas novamente para uma determinação final do *GlideScore*. O *GlideScore* é uma função derivada da função empírica *ChemScore*, incluindo a esta função termos de choque estérico, termos de regiões polares abaixo da superfície da proteína, a fim de eliminar falsos positivos nas predições de interação, e ajustes aos outros termos descritos na função *ChemScore*, resultando na **equação 12**:

$$\begin{aligned}
\text{GlideScore} = & C_{lipo} \sum f(r_{lr}) + C_{hbond} \sum_{i=nn,nc,cc} g_i(\Delta r) h_i(\Delta \alpha) \\
& + C_{max-meta-ion} \sum f(r_{lm}) + C_{roth} H_{roth} \\
& + C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} \\
& + \text{termos de solavatação} + \text{termos do sítio de ligação}
\end{aligned} \tag{12}$$

onde cada um dos termos está descritos na **Tabela 11**.¹⁷⁷⁻¹⁷⁸

Tabela 11 – Definição das participações dos termos da equação do GlideScore.¹⁷⁷⁻¹⁷⁸

Termo	Definição
$C_{lipo} \sum f(r_{lr})$	Termo derivado das interações hidrofóbicas
$\sum_{i=nn,nc,cc} g_i(\Delta r) h_i(\Delta \alpha)$	Termo das interações de hidrogênio. Diferentes pesos são utilizados para interações entre doadores e aceptores neutros (nn), para quando um é carregado (nc) e quando ambos são carregados (cc)
$C_{max-meta-ion} \sum f(r_{lm})$	Termo de interação do ligante com metais com carga líquida diferente de zero.
$C_{roth} H_{roth}$	Termo que penaliza interações que travam ligações rotacionáveis
$C_{polar-phob} V_{polar-phob}$	Termo que penaliza interações com grupos polares sob a superfície da proteína
$C_{coul} E_{coul}$	Termo da energia de Coulomb. Este termo utiliza as cargas iônicas reduzidas para grupos com carga formal não nula, como metais, carboxilatos e guanidinas
$C_{vdW} E_{vdW}$	Termo da energia de Van der Waals. Este termo utiliza as cargas iônicas reduzidas para grupos com carga formal não nula, como metais, carboxilatos e guanidinas
Termos de solvatação	Termo que contabiliza as restrições impostas por grupos polares ou carregados dos ligantes que necessitam de solvatação
Termos do sítio de ligação	Termo para interações polares dentro do sítio de ligação que não correspondem a interações de hidrogênio

Fonte: Elaborada pelo autor

4.5.13.3 Triagem virtual

A primeira etapa do processo de triagem virtual foi a definição (**Figura 36**) e validação da região para busca de interações, uma vez que sua representatividade impacta diretamente a taxa de sucesso da triagem virtual de ligantes. Deste modo, buscou-se no PDB complexos ternários de proteínas homólogas à *Xa*HPPK com o substrato 6-hidroximetil-7,8-di-hidropterina com a maior conservação do sítio ativo e menor valor da raiz quadrada do desvio médio, RMSD (do inglês, *Root-mean-square deviation*), dos resíduos que compõe a cavidade catalítica entre relação à *Xa*HPPK, resultando na escolha da estrutura da HPPK de *E. coli*, com 1,5 Å de resolução e 40% de identidade sequencial, sob o código PDB 1F9H. Esta seleção deve ser feita cuidadosamente pelo fato de não possuir complexo ternário da *Xa*HPPK com o substrato 6HMPH2. A representatividade foi medida utilizando uma metodologia de acomodação do substrato utilizando os mesmos parâmetros utilizados para a triagem virtual e comparando a posição relativa entre a estrutura cristalográfica e a predita pelo algoritmo, como mostrado na **Figura 37**.

Figura 36 – Representação da região definida para busca de interações no sítio ativo e os resíduos que a compõe.

Fonte: Elaborada pelo autor.

Figura 37 – Resultado da aplicação da metodologia de acomodação do substrato 6HMPH2 na cavidade definida para busca de interações na proteína XaHPPK. (A) Sobreposição das estruturas cristalográfica e predita pelo Glide do substrato 6HMPH2. Em verde estão representados os resíduos mais importantes para o modo de ligação segundo o avaliador GlideScore. (B) Representações ortogonais do ligante em sua pose cristalográfica e predita.

Fonte: Elaborada pelo autor.

A avaliação desta sobreposição utilizou o valor de RMSD calculado utilizando a adição do PyMOL *Pair Fit*. O valor de RMSD para a predição do Glide em relação à estrutura cristalográfica é $\text{RMSD}=1.1 \text{ \AA}$, que indica um bom poder preditivo. A partir deste resultado, o subconjunto do ZINC15 filtrado de acordo com o item 3.8.4 desta tese foi submetido ao processo de triagem virtual de ligantes.

O ORCA foi utilizado para cálculo das cargas das 1.000 moléculas mais bem avaliadas pela função de pontuação do Glide. Subsequentemente, uma nova avaliação de interação com a proteína XaHPPK levou a um arranjo final das poses, de acordo com a **Tabela 12**. Finalmente, as 20 melhores poses foram submetidas a uma dinâmica molecular utilizando o pacote GROMACS 5.1.4, com tempo de simulação de 50 ns, para avaliação da estabilidade do modo de ligação predito pelo Glide, e o GlideScore foi recalculado para a pose final. A **Tabela 12** mostra a relação do GlideScore das 20 melhores poses antes e depois do cálculo quântico das cargas e dinâmica molecular, bem como a classificação final de cada uma delas.

Tabela 12 – Resultado da triagem virtual de ligantes utilizando o ORCA para cálculo quântico das cargas das moléculas e o Glide para predição do modo de ligação.

Classificação		ZINC ID	GlideScore	Estrutura
pós ORCA FINAL	pré ORCA			
1	3	4159251	-8.50
2	17	33760400	-8.18
3	7	4721730	-8.12
4	101	13990750	-8.07
5	4	33760400	-8.00
6	12	39321433	-7.99
7	79	225597661	-7.98
8	155	39098400	-7.98
9	91	72195962	-7.92

continua

conclusão

10	2	6491833	-7.88
11	102	72190816	-7.86
12	1	5030605	-7.83
13	167	72194819	-7.81
14	16	72233477	-7.81
15	204	72198983	-7.80
16	354	9118413	-7.80
17	606	13538921	-7.80
18	79	27420751	-7.78
19	10	72233477	-7.78
20	8	14725220	-7.78

Fonte: Elaborada pelo autor.

A característica polar com carga negativa dos resíduos de ácido aspártico (nominalmente Asp92 e Asp94), responsáveis por coordenar a ligação do substrato no sítio ativo mediada por íons de magnésio, privilegiam interações de grupos positivamente carregados, como aminas primárias e secundárias e azidas. Um dos mais potentes inibidores reportados para a reação de hidrólise de ATP,¹⁷⁹ o segundo substrato da reação bi-bi catalisada pela HPPK, é uma molécula que possui a função azida que interage com os resíduos carregados Asp92 e Asp94 (**Figura 38 A1-A2**).

Figura 38 – Modos de ligação preditos pela ferramenta Glide. (A1-A2) Detalhes das interações entre o composto melhor classificado pelo GlideScore e os resíduos Phe50, Asp92, Asp94 e Phe120. (B-C) Comparação das interações de duas poses diferentes de um mesmo composto. A interação entre uma das hidroxilas do grupo catecol com o resíduo Asn52 influencia os termos de interação de hidrogênio e Van der Waals no cálculo do GlideScore.

Fonte: Elaborada pelo autor.

Os resíduos hidrofóbicos Phe50 e Phe120, diametralmente dispostos no sítio de ligação, possibilita a formação interações do tipo π , em especial o empilhamento π - π e interação cátion- π . As interações promovidas por este par de fenilalaninas é capaz de orientar o ligante dentro do sítio e, caso exista um grupo aceptor de interação de hidrogênio em posição adequada como nos compostos de classificação final 1-5, 14, 18 e 19, a energia de interação entre proteína e ligante são largamente favorecidas. Outra interação identificada pelo Glide como relevante foi a interação do grupo catecol e seus derivados com os resíduos de Asn52 e Asp92 (**Figura 38 B-C**), pois este permitiu a orientação das moléculas para interagir com a carbonila da cadeia principal do resíduo Ala42, bloqueando a cavidade que acomoda o substrato 6HMPH2, como nos compostos de classificação final 2, 5, 8 e 20. As moléculas selecionadas nesse experimento serão futuramente testadas quanto à sua atividade inibitória frente a *XaHPPK*.

4.6 7,8-di-hidroneopterina aldolase

4.6.1 Expressão e purificação da proteína DHNA de *X. albilineans*

A expressão da proteína e *XaDHNA* utilizou um protocolo similar ao descrito no item 4.6.1, porém sem uma proteína de fusão, apenas a cauda hexa-histidina. A purificação da proteína foi realizada em 2 etapas cromatográficas (1 etapa de IMAC e 1 etapa de gel filtração). A primeira etapa cromatográfica consiste na separação de proteínas com afinidade por metal imobilizado, entre elas a proteína de interesse, das proteínas do lisado celular, eluindo a proteína ligada à coluna por meio de um gradiente de imidazol 20-500 mM. Dada a alta estabilidade da proteína *XaDHNA* à temperatura, utilizou-se a precipitação térmica à temperatura ambiente de proteínas contaminantes, removidas da amostra utilizando uma centrifugação a 12.900 rpm por 20 minutos. A segunda etapa cromatográfica, uma etapa de gel filtração, utilizou uma resina Superdex 200 em uma coluna XK 26/1000 foi utilizada para a obtenção de uma amostra monodispersa, fundamental para os experimentos de cristalização (**Figura 39**).

Figura 39 – Cromatograma da Gel Filtração da XaDHNA e eletroforese em gel de poliacrilamida 15% com as frações coletadas durante a purificação. (A) Fração insolúvel da expressão. (B) Fração solúvel da expressão. (C) Fração eluída na IMAC. (D) Fração que não interagiu com a resina durante a IMAC. (E) Fração precipitada após 12 h à temperatura ambiente. (F) Fração solúvel após a gel filtração em coluna XK 26/1000 com resina Superdex 200, contendo uma única forma oligomérica da XaDHNA.

Fonte: Elaborada pelo autor.

4.6.2 Avaliação da purificação e estabilidade da proteína XaDHNA

A avaliação da purificação e de estabilidade da proteína XaDHNA foi feita com a mesma metodologia descrita no item 4.5.2 desta tese. A condição inicial da proteína era Tris-HCl 20 mM, 250 mM NaCl e 5% de glicerol, pH 8,0. O qPCR Mx3005p foi então programado para estabilizar todas as amostras a 25° C por 5 min e, por 50 ciclos de 1 min, variar a temperatura à taxa de 1° C/ciclo. O gráfico da **Figura 40** mostra as curvas de desnaturação térmica da XaDHNA, acrescida da temperatura inicial de 25 °C. As duas condições correspondentes às maiores estabilidades estão descritas na **Tabela 13**, assim como suas variações de temperatura de desenovelamento em relação ao referencial (condição inicial). A força iônica aparenta diminuir a estabilidade do octâmero de XaDHNA pela diminuição das interações intermoleculares.

Experimento de DSF para a proteína XaDHNA

Figura 40 – Curvas de fluorescência do Sypro Orange no experimento de DSF para a proteína XaDHNA.

Fonte: Elaborada pelo autor.

Tabela 13 – Condições tamponantes identificadas no experimento de DSF para a proteína XaDHNA.

	Tampão	Sal	Glicerol	pH	ΔT_m
Condição 1	HEPES 100 mM	NaCl 50 mM	0	7,5	6,7
Condição 2	HEPES 100 mM	NaCl 500 mM	5%	7,5	5,5

Fonte: Elaborada pelo autor.

4.6.3 Triagem e otimização de condições de cristalização da proteína XaDHNA

Uma das etapas que mais geram dificuldades nos projetos de biologia estrutural é a cristalização da macromolécula, dada a quantidade de variáveis que atuam sobre o processo de nucleação e crescimento do cristal. Uma abordagem bem-sucedida é a amostragem multidimensional de condições de cristalização, como arranjos ortogonais, triagem reversa, fatoriais incompletos e matrizes esparsas.^{137–140} Embora os três primeiros métodos sejam interessantes por fornecer uma distribuição sistemática e com parâmetros estatisticamente analisáveis, o sucesso obtido pelas matrizes esparsas as tornaram notavelmente populares.

Tabela 14 – Classificação dos *kits* comerciais utilizados neste projeto quanto ao planejamento de sua formulação.

Fabricante	Nome do <i>kit</i>
Matriz esparsa	
Qiagen	JCSG Core Suite I, II, III, IV;
Hampton	Crystal Screen Cryo; Crystal Screen
Fatorial incompleto	
Molecular Dimensions	Morpheus; Midas
Qiagen	MPD Suite; PEGs I, II; AmSO ₄ Suite
Hampton	SaltRx HT; Index HT

Fonte: Elaborada pelo autor.

As distribuições não sistemáticas dos agentes precipitantes nas matrizes esparsas podem levar a algumas condições no entorno das que estão presentes no *kit* e que poderiam gerar nucleação a serem perdidas, diferente do que acontece no fatorial incompleto por exemplo, embora um levantamento feito com os reagentes mais comuns presentes na base de dados PDB mostra uma tendência, ainda que enviesada pela facilidade na utilização ou custo, de alguns reagentes serem predominantes nas composições das condições de cristalização.¹⁴¹⁻¹⁴²

Para os ensaios de triagem de condições de cristalização, a XaDHNA foi concentrada a 15 mg.mL⁻¹ e avaliada em quatorze kits comerciais com três proporções distintas, totalizando 4.032 condições distintas. As gotas foram acompanhadas sistematicamente e a condição do *kit* SaltRx HT D3 (100 mM Tris-HCl; 2,5 M Nitrato de amônio; pH 8,5) produziram cristais em todas as concentrações de proteína após 21 dias. Estes cristais foram testados na fonte de luz sincrotron MX-2 do LNLS, confirmando serem cristais de proteína. Os padrões de difração porém mostraram-se de baixa qualidade (e.g., baixa resolução e alta mosaicidade) e incapazes de serem processados ou refinados.

O experimento de otimização da condição de cristalização teve como objetivo a obtenção de cristais maiores e de melhor qualidade. Assim, a condição identificada na triagem foi utilizada como base para a triagem do *kit* Detergent Screen HT (Hampton Research). Quatro novas condições foram obtidas com cristais com as características

desejáveis (descrito apenas os aditivos): A5) Ácido 3a,12a-Di-hidroxi-5b-colano-24-óico, sal monossódico; B4) n-Tridecil-b-D-maltopiranosídeo; B7) BRIJ®-35, C12E23, a-Dodecil-w-hidroxi-poli-(oxi-1,2-etanodiil), Polietileno glicol (23), monododecil éter; C8) 1-s-Decil-b-D-tiomaltosídeo, n-Decil-b-D-tiomaltopiranosídeo, que foram submetidos à difração de raios X (Figura 41).

Figura 41 – Cristais de *XaDHNA* obtidos durante o experimento de otimização da condição de cristalização utilizando o *kit* Detergent Screen HT e suas respectivas condições.

Fonte: Elaborada pelo autor.

4.6.4 Coleta e processamento de dados de difração de raios X para a *XaDHNA*

Devido às características da mesa de coleta da linha MX-2 do LNLS, foram coletados 360° do cristal de *XaDHNA* na condição C8 do *kit* Detergent Screen HT (Tabela 15) em 900 imagens, com exposição de 5 s por imagem, a uma distância de 150 mm entre o cristal e o detector PILATUS 2M, com resolução de borda de 3.3 Å.

Tabela 15 – Condição de cristalização na qual foi coletado o conjunto de dados para o cristal da proteína *XaDHNA*, com sua respectiva formulação.

Condição de cristalização	Tampão	Sal	pH	Detergente
<i>XaDHNA</i>				
Detergent Screen HT A5	100 mM Tris-HCl	2,5 M Nitrato de amônio	8,5	Ácido 3a,12a-Di-hidroxi-5b-colano-24-óico, sal monossódico
Detergent Screen HT C8	100 mM Tris-HCl	2,5 M Nitrato de amônio	8,5	1-s-Decil-b-D-tiomaltosídeo, n-Decil-b-D-tiomaltopiranosídeo

Fonte: Elaborada pelo autor.

Tabela 16 –Parâmetros de cela unitária, simetria do cristal, qualidade do conjunto de dados e estatística do processamento *XaDHNA*. Entre parênteses estão os valores para a camada de mais alta resolução.

	<i>XaDHNA</i>
Comprimento de onda	1,459
Resolução	40,23 -3,5 (3,625 - 3,5)
Grupo espacial	I 4 2 2
Célula unitária	(a,b,c) = 69,04 69,04 142,061 (α,β,γ) = 90 90 90
Reflexões totais	4760 (467)
Reflexões únicas	2383 (234)
Multiplicidade	2,0 (2,0)
Completeza (%)	99,71 (100,00)
I/σ(I) médio	6,39 (2,18)
B-fator Wilson	74,09
R_{meas}	0,1282 (0,5232)
CC_{1/2}	0,992 (0,709)
R_{work}	0,2969 (0,3452)
R_{free}	0,3342 (0,3501)
Resíduos da proteína	114
RMS (ligações)	0,021
RMS (ângulos)	3,05
Ramachandran favoráveis (%)	84,46
Ramachandran permitidos (%)	11,79
Ramachandran proibidos (%)	3,75
Rotâmeros outliers (%)	9,73
B-fator médio	86,54
Número de grupos TLS	6

Fonte: Elaborada pelo autor.

O conjunto de dados foi processado utilizando o programa XDS, o qual identificou a rede de Bravais como tetragonal cubica de corpo centrado, com o grupo pontual I422. A determinação do conteúdo da cela unitária foi feita pelo cálculo do coeficiente de Matthews (CPP4), resultando numa probabilidade de 99% para 1 molécula na cela unitária e composta

por 62% de solvente. A redução dos dados e determinação do conteúdo da cela unitária foi feita pela metodologia descrita no item 4.7.4 desta monografia.

A sequência da *Xa*DHNA foi obtida do bando de estruturas primárias UniProt (D2UBF2_XANAP) e utilizada como sequência de entrada para a busca por estruturas homólogas depositadas no PDB na ferramenta BLASTp. Dentre as estruturas identificadas a proteína tetramérica *Xa*DHNA homóloga de *V. cholerae* 3O1K com 53% de identidade foi utilizada como molde para a substituição molecular por meio do programa Phaser MR. A busca automatizada do Phaser considerou a composição da célula unitária identificada pelo coeficiente de Matthews e identificou uma solução com valor de LLG igual a 265 e TFZ igual a 8,9.

4.6.5 Modelo DHNA de *Xanthomonas albilineans*

O modelo da *Xa*DHNA é formado por 8 monômeros idênticos de 119 resíduos cada, dos quais 113 estão modelados (**Figura 42**). Cada um dos monômeros, que possui 4 fitas- β e 4 hélices- α , pertence à classe das alfa-beta proteínas (a+b), formada principalmente por fitas- β antiparalelas. O enovelamento destes monômeros é do tipo T, com o padrão de estruturas secundárias $\beta\beta\alpha\alpha\alpha\beta\beta$ (**Figura 43**).

	1	10	20	30	40	50	60
DHNA_SaltRx_D3	MDKVFIEGLEIDALIGIYAWERRIRQTLRFDLEMGFNRIIPAASDDIADTLNYKAISKRM						
tr D2UBF2 D2UBF2_XANAP	MDKVFIEGLEIDALIGIYAWERRIRQTLRFDLEMGFNRIIPAASDDIADTLNYKAISKRM						

	70	80	90	100	110
DHNA_SaltRx_D3	VELVEQSDHGLVETLAERCAAMVLDEFVRWLRRLKLSKPGAVRGAQAVGVMIER.....				
tr D2UBF2 D2UBF2_XANAP	VELVEQSDHGLVETLAERCAAMVLDEFVRWLRRLKLSKPGAVRGAQAVGVMIERSRPG				

Figura 42 – Alinhamento da sequência completa da *Xa*DHNA com o modelo em refinamento.

Fonte: Elaborada pelo autor.

O sítio ativo desta proteína não está localizado no centro do barril, como ocorre no enovelamento do tipo barril TIM, mas na interface de dois monômeros, mais precisamente na região formada entre o C-terminal da α_4 de um monômero com o C-terminal da α_3 e a alça que conecta β_1 em β_2 , de modo que cada octâmero biologicamente ativo possui 8 sítios

enovelamento do tipo T formam barris oligoméricos de diferentes tamanhos, como o observado no modelo da *XaDHNA* (**Figura 45**).

Cada uma das metades do octâmero é formada pelas interações de hidrogênio das fitas β_1 de uma subunidade com a β_4 de uma subunidade imediatamente ao lado, antiparalelas entre si, formando um tetrâmero toroidal como observado nas **Figura 45 C**. Estas duas metades são mantidas pelos resíduos carregados presentes no C-terminal de cada subunidade com a cadeia principal da alça entre α_4 e β_3 de uma subunidade presente na outra metade do octâmero (**Figura 45 C**).

Figura 45 – Representação do modelo da *XaDHNA* com a disposição biológica, ainda em estágio de refinamento. (A) Diagrama de Ramachandran da *XaDHNA*. (B) Representação do contato feito entre dois monômeros de metades diferentes do octâmero da *XaDHNA* (C) Visão superior do homotetrâmero da *XaDHNA*, sobre o eixo de simetria de ordem 4 e sua visão ortogonal, sobre um dos eixos de simetria de ordem 2.

Fonte: Elaborada pelo autor.

4.6.6 Comparação da XaDHNA com suas proteínas homólogas

A comparação de quatro proteínas homólogas à XaDHNA disponíveis no PDB (**Tabela 17**) mostra uma alta conservação da sua estrutura primária em toda a estrutura. O sítio de ligação à neopterina apresenta alta conservação de seus resíduos (**Figura 46**), com mutações conservativas apenas em um de seus resíduos de valina para isoleucina na estrutura de *Y. pestis*.

Figura 46 – Comparação entre estruturas homólogas da XaDHNA disponíveis no PDB. (A) Representação da conservação de resíduos (vermelho – conservado, branco – não conservado) entre as homologas analisadas e RMSD médio entre as posições dos Ca 5VSP-homóloga (maior espessura na representação – maior RMSD médio). (B) Conservação dos resíduos do sítio ativo que interagem diretamente com o substrato 6HMPH2. (C) Alinhamento das sequências primárias das proteínas homólogas analisadas. (D) Comparação filogenética das todas as sequências com estrutura disponível no PDB, geradas pela ferramenta PhyloDendron.

Fonte: Elaborada pelo autor.

O estado oligomérico biológico octamérico da XaDHNA, baseado no cromatograma da Gel Filtração feita em uma resina Superdex 200 empacotada em uma coluna XK 26/1000, é comum entre as estruturas disponíveis no PDB (análise que inclui 15 entradas do PDB).

Tabela 17 – Comparação entre DHNAs homólogas disponíveis no PDB com a *XaHPPK*. Os valores de RMSD são calculados pela superposição das estruturas da *XaDHNA* com a estrutura homóloga em destaque.

PDB ID	Organismo	Identidade	RMSD	Disposição biológica
-	<i>Xanthomonas albilineans</i>	-	-	Octâmero
3O1K	<i>Vibrio cholerae</i>	53%	1,0 Å	Octâmero
2O90	<i>Escherichia coli</i>	52%	1,1 Å	Octâmero
3R2E	<i>Yersinia pestis</i>	49%	1,3 Å	Octâmero
1NBU	<i>Mycobacterium tuberculosis</i>	32%	1,6 Å	Octâmero

Fonte: Elaborada pelo autor.

Os resultados obtidos com a *XaDHNA*, embora não tenham resolução para serem usados em planejamentos de inibidores baseados na estrutura do alvo, abrem espaço para otimizações experimentais futuras. Entre elas, a influência das pontes de dissulfeto que mantem coeso os monômeros de cada um dos tetrâmeros deve ser mantida evitando-se agentes redutores no processo de lise e purificação. Outro fator identificado pelo experimento de DSF como importante para a estabilidade proteica foi a baixa concentração de sal nos tampões. É provável que o uso destas informações para novos experimentos gere cristais com melhor poder de difração, dando continuidade ao projeto de descoberta de novos agentes antibacterianos para a cultura de cana-de-açúcar.

4.7 Desenvolvimento da ferramenta HTP Oligo Designer

4.7.1 Motivação científica

O processo de clonagem de um novo alvo inicia-se pela amplificação da sequência de DNA de um quadro aberto de leitura contendo a proteína de interesse. Essa etapa requer a utilização de oligonucleotídeos iniciadores específicos em uma PCR. A ausência de uma ferramenta capaz de gerar estes oligonucleotídeos em média ou larga escala levou à criação do HTP-Oligo Designer, uma ferramenta *web* gratuita e de livre acesso.

A grande quantidade de dados de genomas e transcriptoma são uma fonte quase inesgotável de informação de genes não caracterizados, os quais podem ser estudados por técnicas bioquímicas e biofísicas.¹⁸⁰⁻¹⁸¹ Grupos de pesquisa têm desenvolvido e implementado plataforma de clonagem e expressão em larga escala, na qual a seleção de alvos viáveis utiliza dados como triagem de expressão, rendimento e atividade da proteína

de interesse.^{182–187} Deste modo, uma vez que o planejamento de oligonucleotídeos iniciador não automatizada demanda uma grande quantidade de tempo e é comumente fonte de erros não sistemáticos, uma ferramenta automatizada, que não requer conhecimento de bioinformática para o usuário final, é fundamental para o sucesso dos projetos nos quais uma grande quantidade de amplificações de sequências de DNA são utilizadas.

Alternativas disponíveis no momento da idealização do HTP Oligo Designer ofereciam apenas opção de entrada para uma única sequência de DNA, e alternativas com a opção de múltiplas sequências de entrada como o BatchPrimer3¹⁸⁸ não oferecem opção de planejamento de iniciadores para clonagem e mutagênese. Com a finalidade de prover a funcionalidade não disponível no momento, o HTP Oligo Designer foi desenvolvido.

4.7.2 Algoritmo de determinação da sequência para o oligonucleotídeo iniciador

A ferramenta HTP Oligo Designer foi desenvolvida em JavaScript 6 e jQuery 1.11.3, compatíveis com HTML 5. A interface do usuário foi elaborada utilizando as diretrizes do Google Material Design. Com o objetivo de maximizar a compatibilidade para o usuário final, o HTP Oligo Designer foi desenvolvido como uma aplicação *web*, utilizando JavaScript, CSS e HTML5, capaz de ser executado em qualquer navegador com suporte a JavaScript independente do sistema operacional, incluindo plataformas portáteis. As funções de alinhamento, otimização de oligonucleotídeos iniciadores e cálculo da temperatura de desnaturação – T_m (do inglês, *Melting Temperature*) foram criadas a partir das definições teóricas,¹⁸⁹ enquanto as energias relacionadas à energia necessária para desfazer ligações de hidrogênio entre as fitas de DNA foram baseadas em trabalhos prévios.^{190–191}

Tabela 18 – Guia de entrada de informações para planejamento de mutagênese no HTP Oligo Designer.

Tipo de mutação	Exemplo	Descrição
Mutação Pontual	TG(C.A)TTA	Substitui C por A na sequência TGCTTA, resultando em TGATTA
Deleção	TG(C.)TTA	Remove C na sequência TGCTTA, resultando em TGTTA
Inserção	TGC(.GAT)TTA	Insere GAT na sequência TGCTTA, resultando em TGCGATTTA

Fonte: Adaptada de LIMA et al.¹⁰³

Para cada entrada no formato FASTA, o algoritmo que constrói sequências amplificação de um segmento gênico separa as sequências de seus identificadores e cria uma lista de possíveis oligonucleotídeos senso iniciadores. Baseando-se nas sequências da lista, é calculado o valor de T_m o elemento de tamanho médio desta. A partir desse resultado, o algoritmo calcula valores de T_m para os elementos de tamanhos maiores ou menores do previamente calculado. Após a obtenção de um dado número de possíveis respostas, ou seja, oligonucleotídeos com valor de T_m com até 3 °C de diferença com o valor desejado, o algoritmo procura por sequências que terminem com os nucleotídeos C ou G, a fim de aumentar a estabilidade da hibridização com a fita de DNA molde. A sequência com o valor de T_m mais próximo do valor desejado é retornado para o usuário. Os oligonucleotídeos iniciadores anti-senso são construídos de maneira idêntica, utilizando o valor de T_m do seu par senso como meta.¹⁰³

A construção de oligonucleotídeos iniciadores para mutagênese sitio dirigida é feita após a escolha do tipo de mutação desejada (i.e., mutação pontual, deleção ou inserção) utilizando o sistema identificador próprio e a metodologia de criação de sequências (i.e., por tamanho ou por temperatura de desnaturação) (**Figura 47**). As sequências são geradas a partir do resíduo escolhido para a edição, de modo que o oligonucleotídeo senso final conterà a mutação no meio de sua sequência. A sequência anti-senso é construída a partir do primeiro nucleotídeo na fita complementar na direção 3'-5', utilizando o T_m de seu par senso como meta.¹⁰³ O método adotado pelo HTP Oligo Designer para definir as regiões de mutação, inserção ou deleção de bases está descrito na **Tabela 18**.

A ferramenta para desenho de oligonucleotídeos para amplificação gênica e para mutagênese sitio dirigida HTP Oligo Designer foi desenvolvida utilizando as linguagens JavaScript, CSS e HTML5, a fim de maximizar a compatibilidade do aplicativo em diferentes navegadores e sistemas operativos. O tempo médio de criação de um par de oligonucleotídeos iniciadores, em um computador de mesa simples, é de 5 ms, e não existe limitação técnica sobre a quantidade de sequências de entrada.

Os oligonucleotídeos iniciadores são desenhados baseados em uma lista de sequências no formato FASTA e nos parâmetros de entrada do usuário: T_m para amplificação gênica e T_m ou tamanho para mutagênese sítio dirigida, concentração de sal e concentração de oligonucleotídeos iniciadores (**Figura 48**). O algoritmo desenvolvido busca

sequências que melhor se adequem aos parâmetros de entrada e à possibilidade de conter as bases citosina ou guanina nas extremidades, assegurando uma hibridização ótima nos experimentos de PCR. Além disso, uma ferramenta para cálculo do T_m para sequências fornecidas pelo usuário está disponível.¹⁰³

Amplificação

ATGCTAGCTGACTAGCTATCTATCCTGACTAGCTATCTATCCTGACTAGCTATCTATCCTGACTAGCTAGCTAGCTATCTATCCTGACTAGCT
TACGATCGACTGATCGATAGATAGGACTGATCGATAGATAGGACTGATCGATAGATAGGACTGATCGATAGATAGGACTGATCG

Mutagênese sítio dirigida

Mutação pontual

ATCGTAAGCTGGGGAGGAGGAACCCCTATC(CG.TA)GACATTCTGAGGGGAGGAGACTACGACATGC

Senso

ATCGTAAGCTGGGGAGGAGGAACCCCTATC(CG.TA)GACATTCTGAGGGGAGGAGACTACGACATGC

Anti-senso

ATCGTAAGCTGGGGAGGAGGAACCCCTATC(CG.TA)GACATTCTGAGGGGAGGAGACTACGACATGC

Deleção

ATCGTAAGCTGGGGAGGAGGAACCCCTATC(CCC.)GAACATTCTGAGGGGAGGAGACTACGACATGC

Inserção

ATCGTAAGCTGGGGAGGAGGAACCCCTATC(.ATACTA)CATTCTGAGGGGAGGAGACTACGACATGC

Figura 47 – Diagrama do algoritmo de desenho de oligonucleotídeos iniciadores por tipo de objetivo. O detalhamento dos nucleotídeos entre parênteses indica a atual sequência (antes do ponto) e a sequência que será construída (após o ponto).

Fonte: Elaborada pelo autor.

A ferramenta está disponível no endereço <http://www.ifsc.usp.br/htpoligo> de forma gratuita e tem sido amplamente utilizada em diversos países de acordo com a análise feita pelo Google Analytics no período entre 16 de maio de 2016 e 17 de maio de 2017 (**Figura 49**).

Os países que lideram a utilização da ferramenta são: Reino Unido (25%), Rússia (16,5%) e Brasil (16%), sendo Londres a cidade que mais utiliza o serviço, correspondendo a 64% dos acessos do Reino Unido. A funcionalidade mais utilizada é a amplificação de genes para clonagem, seguida pela mutagênese sitio dirigida utilizando o tamanho do oligonucleotídeo iniciador como delimitador e, por último, o cálculo da temperatura de desnaturação para uma determinada sequência de DNA.

Figura 48 – (A) Página principal do HTP Oligo Designer. (B) Calculadora de Tm. A ferramenta está disponível no endereço <http://www.ifsc.usp.br/httpoligo>.

Fonte: Elaborada pelo autor.

Figura 49 – Análise da utilização do HTP Oligo Designer no período entre 16 de maio de 2016 e 17 de maio de 2017. O mapa mostra a escala de utilização entre os países: Reino Unido, Rússia e Brasil são os três países com o maior número de acessos, nesta ordem.

Fonte: Elaborada pelo autor.

Conclusão e Perspectivas

5 CONCLUSÃO

A escaldadura das folhas, a doença de cana-de-açúcar objeto de estudo deste trabalho, causa danos significativos em canaviais ao redor do mundo, e o real impacto econômico é ainda mais preocupante, visto que sua característica latente ou assintomática é fonte de sub valorização do prejuízo. A ausência de tratamento químico ou biológico para a doença é outro fator motivante para a pesquisa de descoberto de novos agentes antimicrobianos para cultura de *X. albilineans*.

Durante os 48 meses de doutoramento foram desenvolvidos trabalhos nas áreas de biologia molecular, biologia estrutural, biologia celular, bioensaios e bioinformática. Entre eles destacam-se:

A padronização de um ensaio celular para triagem de compostos é uma importante fonte de informação de moléculas bioativas provenientes de colaboradores e atualmente sendo utilizada por todos os integrantes do grupo. A identificação de duas moléculas com potencial inibitório possibilita o estudo de séries homólogas, capazes de gerar relações de estrutura e atividade, ainda que sem o conhecimento do alvo molecular sobre o qual atuam. Estes resultados, somados a resultados de outros pesquisadores do nosso grupo, demonstram a capacidade deste ensaio de funcionar como plataforma de descoberta de novos agentes antimicrobianos para culturas de *X. albilineans*.

O sucesso nas etapas de padronização da cultura de *X. albilineans* e extração do DNA genômico possibilitaram a clonagem dos segmentos gênicos folB e folK nos diferentes vetores de expressão. Com a metodologia de clonagem utilizando sistema LIC, avaliou-se o perfil de expressão solúvel em diferentes construções de proteína de interesse com uma proteína de fusão. A prospecção de informação bioquímica da via biossintética de folatos constitui uma abordagem conhecida para desenvolvimento de antibióticos, pois é essencial para a manutenção do ciclo de vida da bactéria e é substancialmente diferente em eucariotos superiores, quando não ausente. As enzimas *Xa*DHNA e *Xa*HPPK constituem dois passos consecutivos de uma reação que leva a formação da molécula precursora do folato reduzido, THF, e ainda não se conhece antibióticos que atuam sobre estas duas enzimas. A evolução dos estudos com cada uma das proteínas foi guiado pelos desafios propostos por cada biomolécula, abrindo uma gama de situações que foram exploradas em cada caso.

Ambas as proteínas foram expressadas em sistemas heterólogos (Rosetta 2 (DE3)) e empregaram protocolos de purificação eficientes em sistema semi automatizado ÄKTA.

A obtenção de um cristal de *XaHPPK*, sua otimização e a difração deste cristal por raios X gerou um conjunto de dados capazes de construir um modelo em alta resolução (2,3 Å). Além disso, com a utilização de uma metodologia rápida e barata para determinação do parâmetro cinético K_d foi possível determinar o valor da constante afinidade da *XaHPPK* pelo substrato, corroborando o enovelamento funcional da proteína. As particularidades do sítio ativo da *XaHPPK*, em especial das alças que se ajustam para acomodação de ambos os substratos ATP e 6HMPH2, são interessantes do ponto de vista de planejamento, pois oferecem oportunidades de otimização de propriedades físico-químicas de inibidores para aumentar a especificidade e seletividade. A estrutura resolvida em alta resolução permitiu a aplicação do método de triagem virtual de candidatos a ligantes da enzima alvo. Para tanto, foi utilizada a base de dados ZINC15 e uma abordagem complexa que incluiu a definição de sitio de ligação flexível, cálculo quântico das cargas dos ligantes e dinâmica molecular para a classificação final das moléculas. Essa estratégia forneceu informações relevantes para o planejamento e condução de ensaios de atividade com a *XaHPPK*.

Em relação a *XaDHNA*, a otimização da condição de cristalização proporcionou um conjunto de dados ligeiramente melhor do que o padrão gerado pela condição inicial, porém o cristal de *XaDHNA* não possuiu um poder de difração suficientemente alto para que uma estrutura de alta resolução fosse determinada, mas as peculiaridades identificadas na estrutura a 3,5 Å, como ligações de dissulfeto intramoleculares (reduzidas na estrutura obtida), são capazes de guiar passos seguintes de expressão e purificação.

Dada a característica desse projeto, houve a possibilidade de contribuir para o desenvolvimento de uma nova ferramenta útil no planejamento em larga escala de oligonucleotídeos. O desenvolvimento da ferramenta computacional HTP Oligo Designer se provou útil não apenas para os projetos internos do grupo, mas também para a comunidade científica, que a tem utilizado extensamente.

Por fim, O período de pesquisas no exterior, no Centro de Biologia Química Translacional da Universidade de Edimburgo, sob supervisão do Prof. Malcolm D.

Walkinshaw, possibilitou ampliar a vivência laboratorial e a interação com grupos multidisciplinares.

6 PERSPECTIVAS

Com os resultados referentes ao *XaHPPK*, será possível iniciar a etapa avaliação de inibição de atividade da atividade enzimática utilizando a informação da triagem virtual de inibidores e da triagem de fragmentos. Estes dados serão pontos de partida para otimização de compostos na busca por novos possíveis agentes antimicrobianos. Os resultados obtidos com a *XaDHNA* serão importantes para se delinear os novos experimentos de purificação e cristalização a fim de se obter cristais com alto poder de difração, úteis para as etapas seguintes de SBDD e FBDD.

REFERÊNCIAS

- 1 MIRANDA, J. R. **História da cana-de-açúcar**: history of sugarcane. Campinas: Komedi, 2008 167 p.
- 2 BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Relações Internacionais do Agronegócio. **Balança comercial do agronegócio** - Dez/2012. 2013. Disponível em: <<http://www.agricultura.gov.br/assuntos/relacoes-internacionais/estatisticas-de-comercio-exterior>>. Acesso em: 14 maio 2017.
- 3 BRASIL. Ministério de Minas e Energia. Empresa de Pesquisa Energética. **Balanco energético nacional**. Disponível em: <https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2013.pdf>. Acesso em 14 maio 2017.
- 4 FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Sugarcane production for all countries**. 2016. Disponível em: <<http://www.fao.org/faostat/en/#data>>. Acesso em: 16 maio 2017.
- 5 BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar**: safra 2016/2017. Brasília: Conab, 2016. v. 2, n. 1. 33 p.
- 6 SÃO PAULO. União da Indústria de Cana-de-Açúcar. **Relatório final da safra 2015/2016**: região centro-sul do Brasil. 2016. Disponível em: <<http://www.unicadata.com.br/listagem.php?idMn=92>>. Acesso em: 23 jan. 2017.
- 7 SOUSA, E. L. Etanol e bioeletricidade : a cana-de-açúcar no futuro da matriz energética. 2009. Disponível em: <http://www.senado.gov.br/comissoes/ci/ap/AP20091111_Eduardo_Leao_de_Souza.pdf>. Acesso em: 23 jan. 2016.
- 8 TOKESHI, H. Doenças da cana-de-açúcar (híbridos de *Saccharum* spp.). In: KIMATI, H. et al. (Ed.). **Manual de fitopatologia**: doenças das plantas cultivadas. São Paulo: Agronômica Ceres, 1997. v. 2. p. 207–225.
- 9 BLANCH, M.; LEGAZ, M.-E.; VICENTE, C. Xanthan production by *Xanthomonas albilineans* infecting sugarcane stalks. **Journal of Plant Physiology**, v. 165, n. 4, p. 366-374, 2008.
- 10 JACKSON, R. W. **Plant pathogenic bacteria**: genomics and molecular biology. New York: Horizon Scientific Press, 2009.
- 11 DAVIS, M. J. et al. Intraspecific genomic variation within *xanthomonas albilineans*, the sugarcane leaf scald pathogen. **Phytopathology**, v. 87, n. 3, p. 316–324, 1997.
- 12 JAUFEEERALLY-FAKIM, Y. 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. 287–293, 2000.

13 COCHEREAU, P.; JEAN-BART, A. Etat phytosanitaire des cultures de canne à sucre à Marie-Galante (Guadeloupe). **Bulletin Agronomique des Antilles et de la Guyane**, n. 9, p. 13–19, 1989.

14 ROTT, P. et al. Leaf scald (*Xanthomonas albilineans*) incidence and its effect on yield in seven sugarcane cultivars in Guadeloupe. **Plant Pathology**, v. 44, n. 6, p. 1075–1084, 1995.

15 BORKAR, S. G.; YUMLEMBAM, R. A. **Bacterial diseases of crop plants**. New York: CRC Press, 2016.

16 CROP protection compendium. Disponível em: <<http://www.cabi.org/cpc/restricted/?target=%2Fcpc%2Fdatasheet%2F56901>>. Acesso em: 2 maio 2017.

17 BIRCH, R. G. *Xanthomonas albilineans* and the antipathogenesis approach to disease control. **Molecular Plant Pathology**, v. 2, n. 1, p. 1-11, 2001.

18 MENSI, I. et al. Breaking dogmas: the plant vascular pathogen *Xanthomonas albilineans* is able to invade non-vascular tissues despite its reduced genome. **Open Biology**, v. 4, n. 2, p. 130116, 2014.

19 BIRCH, R. G.; PATIL, S. S. Evidence that an albicidin-like phytotoxin induces chlorosis in sugarcane leaf scald disease by blocking plastid DNA replication. **Physiological and Molecular Plant Pathology**, v. 30, n. 2, p. 207-214, 1987.

20 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.

21 ROTT, P. et al. **Sugarcane leaf scald**. 2017. Disponível em: <<http://edis.ifas.ufl.edu/pdf/SC/SC00100.pdf>>. Acesso em: 23 jun. 2017.

22 RICAUD, C.; RYAN, C. C. Leaf scald. In: RICAUD, C. et al. (Ed.). **Diseases of Sugarcane: major diseases**. New York: Elsevier, 1989. cap. 3, p. 39–58.

23 SILVA, M. S.; BEDENDO, I. P.; CASAGRANDE, M. V. 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.

24 DABBAS, K. M. 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.

25 VERSLYPPE, B. et al. StrainInfo introduces electronic passports for microorganisms. **Systematic and Applied Microbiology**, v. 37, n. 1, p. 42–50, 2014.

- 26 PIERETTI, I. et al. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. **BMC Genomics**, v. 10, n. 1, p. 616, 2009.
- 27 LINDGREN, P. B. The role of hrp genes during plant-bacterial interactions. **Annual Review of Phytopathology**, v. 35, n. 1, p. 129–152, 1997.
- 28 NESME, X. et al. Diversity and genetic relatedness within genera *xanthomonas* and *stenotrophomonas* using restriction endonuclease site differences of pcr-amplified 16s rRNA gene. **Systematic and Applied Microbiology**, v. 18, n. 1, p. 127–135, 1995.
- 29 PIERETTI, I. et al. What makes *Xanthomonas albilineans* unique amongst xanthomonads? **Frontiers in Plant Science**, v. 6, p. 289, 2015.
- 30 LU, H. et al. Acquisition and evolution of plant pathogenesis–associated gene clusters and candidate determinants of tissue-specificity in *xanthomonas*. **PLoS One**, v. 3, n. 11, p. e3828, 2008.
- 31 HASHIMI, S. M. et al. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. **Antimicrobial Agents Chemotherapy**, v. 51, n. 1, p. 181–7, 2007.
- 32 HASHIMI, S. M. et al. DNA gyrase from the albicidin producer *Xanthomonas albilineans* has multiple-antibiotic-resistance and unusual enzymatic properties. **Antimicrobial Agents and Chemotherapy**, v. 52, n. 4, p. 1382–90, 2008.
- 33 ZHANG, L.; XU, J.; BIRCH, R. G. Factors affecting biosynthesis by *Xanthomonas albilineans* of albicidin antibiotics and phytotoxins. **Journal of Applied Microbiology**, v. 85, n. 6, p. 1023–1028, 1998.
- 34 ZHANG, L.; XU, J.; BIRCH, R. G. High affinity binding of albicidin phytotoxins by the AlbA protein from *Klebsiella oxytoca*. **Microbiology**, v. 144, n. 2, p. 555–559, 1998.
- 35 BOSTOCK, J. M. et al. A DHA14 drug efflux gene from *Xanthomonas albilineans* confers high-level albicidin antibiotic resistance in *Escherichia coli*. **Journal of Applied Microbiology**, v. 101, n. 1, p. 151–160, 2006.
- 36 COCIANCICH, S. et al. The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine. **Nature Chemical Biology**, v. 11, n. 3, p. 195-197, 2015.
- 37 CHAMPOISEAU, P.; DAUGROIS, J.-H. Variation in albicidin biosynthesis genes and in pathogenicity of *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. **Phytopathology**, v. 96, n. 1, p. 33–45, 2006.
- 38 ELI LILLY AND COMPANY. Karl H. Michel; Ralph E. Kastner. **Antibiotics and process for production thereof**. US4492650 A, 27 Sept. 1982, 8 Jan. 1985.
- 39 COCIANCICH, S. et al. The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine. **Nature Chemical Biology**, v. 11, n. 3, p. 195-197, 2015.

- 40 ROYER, M. et al. Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. **Molecular Plant-Microbe Interactions**, v. 17, n. 4, p. 414–27, 2004.
- 41 HUANG, G.; ZHANG, L.; BIRCH, R. G. A multifunctional polyketide-peptide synthetase essential for albidin biosynthesis in *Xanthomonas albilineans*. **Microbiology**, v. 147, pt 3, p. 631–642, 2001.
- 42 ZHANG, L.; XU, J.; BIRCH, R. G. Engineered detoxification confers resistance against a pathogenic bacterium. **Nature Biotechnology**, v. 17, n. 10, p. 1021–1024, 1999.
- 43 WALL, M. K.; BIRCH, R. G. Genes for albidin 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.
- 44 VIVIEN, E. et al. *Xanthomonas albilineans* HtpG is required for biosynthesis of the antibiotic and phytotoxin albidin. **FEMS Microbiology Letters**, v. 251, n. 1, p. 81–89, 2005.
- 45 DOSSELAERE, F.; VANDERLEYDEN, J. A metabolic node in action: chorismate-utilizing enzymes in microorganisms. **Critical Reviews in Microbiology**, v. 27, n. 2, p. 75–131, 2001.
- 46 COLLIN, F.; KARKARE, S.; MAXWELL, A. Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. **Applied Microbiology and Biotechnology**, v. 92, n. 3, p. 479–497, 2011.
- 47 CHAMPOUX, J. J. DNA topoisomerases: structure, function, and mechanism. **Annual Review of Biochemistry**, v. 70, n. 1, p. 369–413, 2001.
- 48 REECE, R. J.; MAXWELL, A. DNA gyrase: structure and function. **Critical Reviews in Biochemistry and Molecular Biology**, v. 26, n. 3-4, p. 335-375, 1991.
- 49 TSE-DINH, Y.-C. Exploring DNA topoisomerases as targets of novel therapeutic agents in the treatment of infectious diseases. **Infectious Disorders Drug Targets**, v. 7, n. 1, p. 3–9, 2007.
- 50 MAXWELL, A. DNA gyrase as a drug target. **Trends in Microbiology**, v. 5, n. 3, p. 102–109, 1997.
- 51 FLATMAN, R. H. et al. Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. **Antimicrobial Agents and Chemotherapy**, v. 49, n. 3, p. 1093–1100, 2005.
- 53 BERMINGHAM, A.; DERRICK, J. P. The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. **Bioessays**, v. 24, n. 7, p. 637–48, 2002.
- 54 BOURNE, C. Utility of the biosynthetic folate pathway for targets in antimicrobial discovery. **Antibiotics**, v. 3, n. 1, p. 1–28, 2014.

- 55 STOVER, P. J. One-carbon metabolism-genome interactions in folate-associated pathologies. **Journal of Nutrition**, v. 139, n. 12, p. 2402–2405, 2009.
- 56 LAMBIE, D. G.; JOHNSON, R. H. Drugs and folate metabolism. **Drugs**, v. 30, n. 2, p. 145–155, 1985.
- 57 RENGARAJAN, J. et al. The folate pathway is a target for resistance to the drug para-aminosalicylic acid (PAS) in mycobacteria. **Molecular Microbiology**, v. 53, n. 1, p. 275–282, 2004.
- 58 DE CRÉCY-LAGARD, V. et al. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. **BMC Genomics**, v. 8, n. 1, p. 245, 2007.
- 59 BLASZCZYK, J. et al. Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. **Structure**, v. 12, n. 3, p. 467–75, 2004.
- 60 MUECK, L. **Milestone 1:** imagine a crystal's inner life. Disponível em: <<https://www.nature.com/milestones/milecrystal/pdf/milecrystal01.pdf>>. Acesso em: 23 jan. 2016.
- 61 MEYER, G. Review of basic concepts of crystallography. an outcome from crystal symmetry basic concepts of crystallography. an outcome from crystal symmetry , by EmilZolotoyabko, Wiley-VCH : Weinheim, Germany , 2011. 266 p. US \$65. ISBN 978-3-527-33009-6 . **Crystal Growth & Design**, v. 12, n. 11, p. 5833–5833, 2012.
- 62 ROUYET, J. P.; ZAK, J. Bravais-lattice operator in crystals. **Physical Review B**, v. 23, n. 2, p. 561–564, 1981.
- 63 RUPP, B. **Biomolecular crystallography:** principles, practice, and application to structural biology. New York: Garland Science, 2010.
- 64 RÖNTGEN, W. C. Ueber eine neue Art von Strahlen. **Annals of Physics**, v. 300, n. 1, p. 12–17, 1898.
- 65 FRANKEL, R. I. Centennial of Röntgen's discovery of x-rays. **Western Journal of Medicine**, v. 164, n. 6, p. 497–501, 1996.
- 66 SHI, Y. et al. A Glimpse of Structural Biology through X-ray crystallography. **Cell**, v. 159, n. 5, p. 995–1014, 2014.
- 67 MARTIRADONNA, L. **Milestone 2:** a photograph of crystal order. 2014. Disponível em: <<https://www.nature.com/milestones/milecrystal/pdf/milecrystal02.pdf>>. Acesso em: 23 jan. 2016.
- 68 BRAGG, W. L. The diffraction of short electromagnetic waves by a crystal. **Proceedings of the Cambridge Philosophical Society**, v. 17, p. 43–57, 1913.

69 GEVAUX, D. **Milestone 3**: the equation to bridge worlds. 2014. Disponível em: <<https://www.nature.com/milestones/milecrystal/pdf/milecrystal03.pdf>>. Acesso em 23jan. 2016.

70 GREEN, D. W.; INGRAM, V. M.; PERUTZ, M. F. The structure of haemoglobin. IV. sign determination by the isomorphous replacement method. **Proceedings of the Royal Society A: mathematical, physical & engineering sciences**, v. 225, n. 1162, 1954.

71 KENDREW, J. C. et al. Structure of myoglobin: a three-dimensional fourier synthesis at 2 Å. resolution. **Nature**, v. 185, n. 4711, p. 422–427, 1960.

72 KIRK, R. **Milestone 12**: the first of its kind. 2014. Disponível em: <<https://www.nature.com/milestones/milecrystal/pdf/milecrystal12.pdf>>. Acesso em: 30 out. 2016.

73 PAULING, L.; COREY, R. B. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. **Proceedings of the National Academy of Sciences of the United States of America**, v. 37, n. 5, p. 235–40, 1951.

74 PAULING, L.; COREY, R. B. Configuration of Polypeptide Chains. **Nature**, v. 168, n. 4274, p. 550–551, 1951.

75 PAULING, L.; COREY, R. B.; BRANSON, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. **Proceedings of the National Academy of Sciences of the United States of America**, v. 37, n. 4, p. 205–11, 1951.

76 ROGGENKAMP, R.; KUSTERMAN-KUHN, B.; HOLLENBERG, C. P. Expression and processing of bacterial beta-lactamase in the yeast *Saccharomyces cerevisiae*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 78, n. 7, p. 4466–70, 1981.

77 GREENE, P. J. et al. A general method for the purification of restriction enzymes. **Nucleic Acids Research**, v. 5, n. 7, p. 2373–80, 1978.

78 PERUTZ, M. F. et al. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. **Nature**, v. 185, n. 4711, p. 416–22, 1960.

79 MULLIN, J. W.; CHAKRABORTY, M.; MEHTA, K. Nucleation and growth of ammonium sulphate crystals from aqueous solution. **Journal of Applied Chemistry**, v. 20, n. 12, p. 367–371, 2007.

80 LARSON, M. A.; MULLIN, J. W. Crystallization kinetics of ammonium sulphate. **Journal of Crystal Growth**, v. 20, n. 3, p. 183–191, 1973.

81 MCPHERSON, A. A comparison of salts for the crystallization of macromolecules. **Protein Science**, v. 10, n. 2, p. 418–22, 2001.

82 SMYTH, M. S.; MARTIN, J. H. X-ray crystallography. **Molecular Pathology**, v. 53, n. 1, p. 8–14, 2000.

83 BLUMENTHAL, G. R. & GOULD, R. J. Bremsstrahlung, synchrotron radiation, and compton scattering of high-energy electrons traversing dilute gases. **Reviews of Modern Physics**, v. 42, n. 2, p. 237–270, 1970.

84 LEGATE, K. R. **Milestone 19**: anomalous diffraction tackles phasing. 2014. Disponível em: <<https://www.nature.com/milestones/milecrystal/pdf/milecrystal19.pdf>>. Acesso em: 23 jan. 2016.

85 WILLIAMS, G. P. A general review of synchrotron radiation, its uses and special technologies. **Vacuum**, v. 32, n. 6, p. 333–345, 1982.

86 PHILLIPS, W. C. et al. A CCD-based area detector for X-ray crystallography using synchrotron and laboratory sources. **Nuclear Instruments and Methods in Physics Research A**, v. 334, n. 2–3, p. 621–630, 1993.

87 BROENNIMANN, C. et al. The PILATUS 1M detector. **Journal of Synchrotron Radiation**, v. 13, n. 2, p. 120–130, 2006.

88 THOMPSON, A. C. et al. A large area CMOS detector for shutterless collection of x-ray diffraction data. **Journal of Physics: conference series**, v. 493, p. 12019, 2014.

89 KRAFT, P. et al. Performance of single-photon-counting PILATUS detector modules. **Journal of Synchrotron Radiation**, v. 16, n. pt 3, p. 368–75, 2009.

90 KABSCH, W. The rotation method in crystallography Edited by U. W. Arndt and A. J. Wonacott. **Acta Crystallographica B**, v. 34, n. 3, p. 1049–1050, 1978.

91 RODGERS, D. W. et al. Cryocrystallography. **Structure**, v. 2, n. 12, p. 1135–40, 1994.

92 WINN, M. D. et al. Overview of the CCP4 suite and current developments. **Acta Crystallographica D**, v. 67, n. pt 4, p. 235–42, 2011.

93 EVANS, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. **Acta Crystallographica D**, v. 67, n. pt 4, p. 282–292, 2011.

94 MURSHUDOV, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. **Acta Crystallographica D**, v. 67, n. pt 4, p. 355–67, 2011.

95 DOERR, A. **Milestone 13**: a method ahead of its time. 2014. Disponível em: <<https://www.nature.com/milestones/milecrystal/pdf/milecrystal13.pdf>>. Acesso em: 23 jan. 2016.

96 ROSSMANN, M. G. & BLOW, D. M. The detection of sub-units within the crystallographic asymmetric unit. **Acta Crystallographica**, v. 15, n. 1, p. 24–31, 1962.

97 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.

- 98 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.
- 99 JAUFEERALLY-FAKIM, Y.; DOOKUN, A. Extraction of high quality DNA from polysaccharides-secreting xanthomonads. **Science and Technology**, v. 6, p. 33–40, 2000.
- 100 PAN, Y.-B. 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.
- 101 LAMBERT, R. J. W.; PEARSON, J. Susceptibility testing: accurate and reproducible minimum inhibitory concentration (MIC) and non-inhibitory concentration (NIC) values. **Journal of Applied Microbiology**, v. 88, n. 5, p. 784–790, 2000.
- 102 MURAKAMI, C. J. et al. A method for high-throughput quantitative analysis of yeast chronological life span. **Journals of Gerontology A**, v. 63, n. 2, p. 113–21, 2008.
- 103 LIMA, G. M. A. et al. HTP-OligoDesigner: an online primer design tool for high-throughput gene cloning and site-directed mutagenesis. **Journal of Computational Biology**, v. 23, n. 1, p. 27–29, 2016.
- 104 STUDIER, F. W. Protein production by auto-induction in high density shaking cultures. **Protein Expression and Purification**, v. 41, n. 1, p. 207–34, 2005.
- 105 BORNHORST, J. A.; FALKE, J. J. Purification of proteins using polyhistidine affinity tags. **Methods in Enzymology**, v. 326, p. 245–54, 2000.
- 106 WAUGH, D. S. An overview of enzymatic reagents for the removal of affinity tags. **Protein Expression and Purification**, v. 80, n. 2, p. 283–93, 2011.
- 107 NIESEN, F. H., BERGLUND, H. & VEDADI, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. **Nature Protocols**, v. 2, n. 9, p. 2212–21, 2007.
- 108 VIVOLI, M. et al. Determination of protein-ligand interactions using differential scanning fluorimetry. **Journal of Visualized Experiments**, n. 91, p. 51809, 2014. doi:10.3791/51809
- 109 BATTYE, T. G. G. et al. iMOSFLM : a new graphical interface for diffraction-image processing with MOSFLM. **Acta Crystallographica D**, v. 67, n. 4, p. 271–281, 2011.
- 110 KABSCH, W. Integration, scaling, space-group assignment and post-refinement. **Acta Crystallographica D**, v. 66, n. pt 2, p. 133–44, 2010.
- 111 EVANS, P. Scaling and assessment of data quality. **Acta Crystallographica D**, v. 62, n. pt 1, p. 72–82, 2006.
- 112 IRWIN, J. J. et al. ZINC: a free tool to discover chemistry for biology. **Journal of Chemical Information and Modeling**, 2012. doi: 10.1021/ci3001277.

- 113 BAELE, J. B.; HOLLOWAY, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. **Journal of Medicinal Chemistry**, v. 53, n. 7, p. 2719–2740, 2010.
- 114 GODDEN, J. W.; XUE, L.; BAJORATH, J. Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and tanimoto coefficients. **Journal of Chemical Information and Modeling**, v. 40, n. 1, p. 163-166, 1999.
- 115 BENDER, A.; GLEN, R. C. Molecular similarity: a key technique in molecular informatics. **Organic & Biomolecular Chemistry**, v. 2, n. 22, p. 3204–18, 2004.
- 116 CHO, A. E. et al. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. **Journal of Computational Chemistry**, v. 26, n. 9, p. 915–31, 2005.
- 117 PARK, K.; SUNG, N. K.; CHO, A. E. Importance of accurate charges in binding affinity calculations: a case of neuraminidase series. **Bulletin of the Korean Chemical Society**, v. 34, n. 2, p. 545–548, 2013.
- 118 NEESE, F. The ORCA program system. **Wiley Interdisciplinary Reviews: computational molecular science**, v. 2, n. 1, p. 73–78, 2012.
- 119 ABRAHAM, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. **Software X**, v. 1–2, p. 19–25, 2015.
- 120 NICKOLLS, J. et al. Scalable parallel programming with CUDA. **Queue**, v. 6, n. 2, p. 40, 2008.
- 121 COOK, S. **CUDA programming : a developer's guide to parallel computing with GPUs**. Amsterdam: Elsevier; Morgan Kaufmann, 2013.
- 122 GABRIEL, E. et al. **OpenMPI: goals, concept, and design of a next generation MPI implementation**. 2004. Disponível em: <<https://www.open-mpi.org/papers/euro-pvmmpi-2004-overview/euro-pvmmpi-2004-overview.pdf>>. Acesso em: 23 jan. 2016.
- 123 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. 121p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.
- 124 ROSENBERG, A. H. et al. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. **Gene**, v. 56, n. 1, p. 125–35, 1987.
- 125 STUDIER, F. W. & MOFFATT, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. **Journal of Molecular Biology**, v. 189, n. 1, p. 113–30, 1986.
- 126 DUBENDORFF, J. W.; STUDIER, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. **Journal of Molecular Biology**, v. 219, n. 1, p. 45–59, 1991.

- 127 ASLANIDIS, C.; DE JONG, P. J. Ligation-independent cloning of PCR products (LIC-PCR). **Nucleic Acids Research**, v. 18, n. 20, p. 6069–74, 1990.
- 128 MESELSON, M.; YUAN, R. DNA Restriction Enzyme from *E. coli*. **Nature**, v. 217, n. 5134, p. 1110–1114, 1968.
- 129 WITKIN, E. M. Ultraviolet mutagenesis and inducible DNA repair in *Escherichia coli*. **Bacteriological Reviews**, v. 40, n. 4, p. 869–907, 1976.
- 130 TAYLOR, R. G.; WALKER, D. C.; MCINNES, R. R. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. **Nucleic Acids Research**, v. 21, n. 7, p. 1677–8, 1993.
- 131 BENVENUTI, M. & MANGANI, S. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. **Nature Protocols**, v. 2, n. 7, p. 1633–1651, 2007.
- 132 RUSSO KRAUSS, I. et al. An overview of biological macromolecule crystallization. **International Journal of Molecular Sciences**, v. 14, n. 6, p. 11643–11691, 2013.
- 133 MCPHERSON, A. A brief history of protein crystal growth. **Journal of Crystal Growth**, v. 110, n. 1–2, p. 1–10, 1991.
- 134 DESSAU, M. A.; MODIS, Y. Protein crystallization for X-ray Crystallography. **Journal of Visualized Experiments**, n. 47, 2011. doi: 10.3791/2285.
- 135 RIÈS-KAUTT, M. M.; DUCRUIX, A. F. Crystallization of basic proteins by ion pairing. **Journal of Crystal Growth**, v. 110, n. 1–2, p. 20–25, 1991.
- 136 ABERGEL, C.; NESA, M. P.; FONTECILLA-CAMPS, J. C. The effect of protein contaminants on the crystallization of turkey egg white lysozyme. **Journal of Crystal Growth**, v. 110, n. 1–2, p. 11–19, 1991.
- 137 KINGSTON, R. L.; BAKER, H. M.; BAKER, E. N. Search designs for protein crystallization based on orthogonal arrays. **Acta Crystallographica D**, v. 50, n. 4, p. 429–440, 1994.
- 138 JANCARIK, J.; KIM, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. **Journal of Applied Crystallography**, v. 24, n. 4, p. 409–411, 1991.
- 139 STURA, E. A. et al. Reverse screening. **Acta Crystallographica D**, v. 50, n. 4, p. 448–455, 1994.
- 140 CARTER, C. W.; CARTER, C. W. Protein crystallization using incomplete factorial experiments. **Journal of Biological Chemistry**, v. 254, n. 23, p. 12219–23, 1979.
- 141 GORREC, F. Protein crystallization screens developed at the MRC laboratory of molecular biology. **Drug Discovery Today**, v. 21, n. 5, p. 819–825, 2016.
- 142 KIRKWOOD, J. et al. Analysis of crystallization data in the Protein Data Bank. **Acta Crystallographica F**, v. 71, n. 10, p. 1228–1234, 2015.

- 143 DIEDERICHS, K.; KARPLUS, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. **Nature Structural Biology**, v. 4, n. 4, p. 269–275, 1997.
- 144 WEISS, M. S.; HILGENFELD, R. On the use of the merging R factor as a quality indicator for X-ray data. **Journal of Applied Crystallography**, v. 30, n. 2, p. 203–205, 1997.
- 145 POYNTING, J. H. On the transfer of energy in the electromagnetic field. **Philosophical Transactions of the Royal Society**, v. 175, p. 343–361, 1884. doi: 10.1098/rstl.1884.0016.
- 146 GRIFFITHS, D. J. **Introduction to electrodynamics**. New York: Pearson Education, 2014.
- 147 STRAUSS, M. G. et al. CCD-based detector for protein crystallography with synchrotron X-rays. **Nuclear Instruments and Methods in Physics Research A**, v. 297, n. 1–2, p. 275–295, 1990.
- 148 RUPP, B. High-Throughput Crystallography at an Affordable Cost: The TB Structural Genomics Consortium Crystallization Facility. **Accounts of Chemical Research**, v. 36, n. 3, p. 173–181, 2003.
- 149 WAASMAIER, D.; KIRFEL, A. New analytical scattering-factor functions for free atoms and ions. **Acta Crystallographica A**, v. 51, n. 3, p. 416–431, 1995.
- 150 **INTERNATIONAL TABLES FOR CRYSTALLOGRAPHY**. Dordrecht: International Union of Crystallography, 2016.
- 151 BRÜNGER, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. **Nature**, v. 355, n. 6359, p. 472–475, 1992.
- 152 MORRIS, A. L. et al. Stereochemical quality of protein structure coordinates. **Proteins**, v. 12, n. 4, p. 345–364, 1992.
- 153 KLEYWEGT, G. J.; JONES, T. A. Homo crystallographicus--quo vadis? **Structure**, v. 10, n. 4, p. 465–72, 2002.
- 154 KARPLUS, P. A. & DIEDERICHS, K. Assessing and maximizing data quality in macromolecular crystallography. **Current Opinion in Structural Biology**, v. 34, p. 60–68, 2015.
- 155 DIEDERICHS, K. Crystallographic data and model quality. **Methods in molecular Biology**, v. 1320, p. 147–173, 2016. doi: 10.1007/978-1-4939-2763-0_10.
- 156 EVANS, P. R.; MURSHUDOV, G. N. How good are my data and what is the resolution? **Acta Crystallographica D**, v. 69, n. 7, p. 1204–1214, 2013.
- 157 KARPLUS, P. A.; DIEDERICHS, K. Linking crystallographic model and data quality. **Science**, v. 336, n. 6084, p. 1030–1033, 2012.
- 158 DE BEER, T. A. P. et al. PDBsum additions. **Nucleic Acids Research**, v. 42, n. d1, p. d292–D296, 2014.

- 159 LOVELL, S. C. et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. **Proteins**, v. 50, n. 3, p. 437–50, 2003.
- 160 YANG, R. et al. Loop conformation and dynamics of the Escherichia coli HPPK apo-enzyme and its binary complex with MgATP. **Biophysical Journal**, v. 89, n. 1, p. 95–106, 2005.
- 161 KRISSINEL, E.; HENRICK, K. Inference of Macromolecular Assemblies from Crystalline State. **Journal of Moleculer Biology**, v. 372, n. 3, p. 774–797, 2007.
- 162 PANTOLIANO, M. W. et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. **Journal of Biomolecular Screening**, v. 6, n. 6, p. 429–40, 2001.
- 163 LOUNNAS, V. et al. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. **Computational and Structural Biotechnology Journal**, v. 5, p. e201302011, 2013.
- 164 ANDRICOPULO, A. D.; SALUM, L. B.; ABRAHAM, D. J. Structure-based drug design strategies in medicinal chemistry. **Current Topics in Medicinal Chemistry** v. 9, n. 9, p. 771–90, 2009.
- 165 GUIDO, R. V. C.; OLIVA, G.; ANDRICOPULO, A. D. Virtual screening and its integration with modern drug design technologies. **Current Medicinal Chemistry**, v. 15, n. 1, p. 37–46, 2008.
- 166 SEDDON, G. et al. Drug design for ever, from hype to hope. **Journal of Computer-Aided Molecular Design**, v. 26, n. 1, p. 137–150, 2012.
- 167 KITCHEN, D. B. 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.
- 168 HUNFIELD, F.L. Astex, structural genomix, and syrrx. **Chemistry & Biology**, v. 10, n. 2, p. 95–98, 2003.
- 169 KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations **Drug Discovery Today**, v. 11, n. 13–14, p. 580–94, 2006.
- 170 KIM, R.; SKOLNICK, J. Assessment of programs for ligand binding affinity prediction. **Journal of Computational Chemistry**, v. 29, n. 8, p. 1316–1331, 2008.
- 171 IRWIN, J. J. et al. Automated docking screens: a feasibility study. **Journal of Medicinal Chemistry**, v. 52, n. 18, p. 5712–5720, 2009.
- 172 OKIMOTO, N. et al. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. **PLOS Computational Biology**, v. 5, n. 10, p. e1000528, 2009.
- 173 SLIWOSKI, G. et al. Computational Methods in Drug Discovery. **Pharmacological Reviews**, v. 66, n. 1, p. 334–395, 2013.

- 174 KONTOYIANNI, M; MCCLELLAN, L. M.; SOKOL, G. S. Evaluation of docking performance: comparative data on docking algorithms. **Journal of Medicinal Chemistry**, v. 47, n. 3, p. 558-565, 2003.
- 175 PLEWCZYNSKI, D. et al. Can we trust docking results? evaluation of seven commonly used programs on PDBbind database. **Journal of Computational Chemistry**, v. 32, n. 4, p. 742–755, 2011.
- 176 CROSS, J. B. et al. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. **Journal of Chemical Information and Modeling**, v. 49, n. 6, p. 1455–1474, 2009.
- 177 FRIESNER, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. **Journal of Medicinal Chemistry**, v. 47, n. 7, p. 1739–1749, 2004.
- 178 SCHRÖDINGER, I. **Glide 5.5: user manual**. 2009. Disponível em: <https://isp.ncifcrf.gov/files/isp/uploads/2010/07/gli55_user_manual.pdf>. Acesso em: 23 jan. 2016.
- 179 BOWLER, M. W. et al. How azide inhibits ATP hydrolysis by the F-ATPases. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 23, p. 8646–8649, 2006.
- 180 METZKER, M. L. Sequencing technologies - the next generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31–46, 2010.
- 181 MARDIS, E. R. A decade's perspective on DNA sequencing technology. **Nature**, v. 470, n. 7333, p. 198–203, 2011.
- 182 STEVENS, R. C.; YOKOYAMA, S.; WILSON, I. A. Global efforts in structural genomics. **Science**, v. 294, n. 5540, p. 89–92, 2001.
- 183 LESLEY, S. A. et al. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 18, p. 11664–9, 2002.
- 184 BUNKOCZI, G. et al. Mechanism and substrate recognition of human holo ACP synthase. **Chemistry & Biology**, v. 14, n. 11, p. 1243–53, 2007.
- 185 SAUDER, M. J. et al. High throughput protein production and crystallization at NYSGXRC. **Methods in Molecular Biology**, v. 426, p. 561–75, 2008.
- 186 XIAO, R. et al. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. **Journal of Structural Biology**, v. 172, n. 1, p. 21–33, 2010.
- 187 VINCENTELLI, R. et al. High-throughput protein expression screening and purification in *Escherichia coli*. **Methods**, v. 55, n. 1, p. 65–72, 2011.

188 YOU, F. M. et al. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. - google search. **BMC Bioinformatics**, v. 9, n. 1, p. 253, 2008.

189 RABHI, I. et al. A novel simple and rapid PCR-based site-directed mutagenesis method. **Molecular Biotechnology**, v. 26, n. 1, p. 27–34, 2004.

190 BRESLAUER, K. J. et al. Predicting DNA duplex stability from the base sequence. **Proceedings of the National Academy of Sciences of the United States of America.**, v. 83, n. 11, p. 3746–3750, 1986.

191 SUGIMOTO, N. et al. Improved Thermodynamic Parameters and Helix Initiation Factor to Predict Stability of DNA Duplexes. **Nucleic Acids Research**, v. 24, n. 22, p. 4501–4505, 1996.

APÊNDICE A – Produção científica

GODOY, A. S.; LIMA, G. M. A.; OLIVEIRA, K. I. Z.; TORRES, N. U.; MALUF, F. V.; GUIDO, R. V. C.; OLIVA, G. Crystal structure of Zika virus NS5 RNA-dependent RNA polymerase. ***Nature Communication***, v. 8, p. 14764, 2017.

SOUZA-MOREIRA, T. M.; ALVES, T. B.; PINHEIRO, K. A.; FELIPPE, L. G.; LIMA, G. M. A.; WATANABE, T. F.; BARBOSA, C. C.; SANTOS, V. A. F. F. M.; LOPES, N. P.; VALENTINI, S. R.; GUIDO, R. V. C.; FURLAN, M.; ZANELLI, C. F. Friedelin Synthase from *Maytenus ilicifolia*: Leucine 482 Plays an Essential Role in the Production of the Most Rearranged Pentacyclic Triterpene. ***Scientific Reports***, v. 6, p. 36858, 2016.

REMLINGER, M.; LIMA, G. M. A.; GUIDO, RAFAEL V. C.; SOUZA-MOREIRA, T. M. ; VALENTINI, S. R. ; FURLAN, M.; ZANELLI, C. F. Análise in silico dos resíduos de aminoácidos da enzima friedelina sintase de *Maytenus ilicifolia*. ***Revista de Ciências Farmacêuticas Básica e Aplicada***, v. 37, p. 235, 2016.

DA SILVA, M. T. ALVES; SILVA-JARDIM, I.; PORTAPILLA, G. B.; LIMA, G. M. A.; COSTA, F. C.; ANIBAL, F. F.; THIEMANN, O. H. In vivo and in vitro auranofin activity against *Trypanosoma cruzi*: Possible new uses for an old drug. ***Experimental Parasitology***, p. 189-193, 2015.

LIMA, G. M. A.; CAMILO, C. M.; MALUF, F. V.; GUIDO, R. V. C.; POLIKARPOV, I. HTP-OligoDesigner: an online primer design tool for high-throughput gene cloning and site-directed mutagenesis. ***Journal of Computational Biology***, v. 23, p. 27-29, 2015.