



Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Departamento de Química
Programa de Pós-Graduação em Química

Planejamento computacional e avaliação *in vitro* de novos potenciais inibidores da
fosfomanose isomerase (PMI) de interesse para o controle do cancro cítrico.

Mariana Pegrucci Barcelos

Dissertação apresentada à Faculdade de
Filosofia, Ciências e Letras de Ribeirão Preto da
Universidade de São Paulo, como parte das exigências para
a obtenção do título de Mestre em Ciências, Área: **Química**

RIBEIRÃO PRETO -SP

2020

MARIANA PEGRUCCI BARCELOS

Planejamento computacional e avaliação *in vitro* de novos potenciais inibidores da
fosfomanose isomerase (PMI) de interesse para o controle do cancro cítrico.

Versão Original

Dissertação apresentada à Faculdade de
Filosofia, Ciências e Letras de Ribeirão Preto da
Universidade de São Paulo, como parte das exigências para
a obtenção do título de Mestre em Ciências, Área: **Química**

Orientador: Prof. Dr. Carlos Henrique Tomich de Paula da Silva
Co-orientador: Dr. Leonardo Bruno Federico

RIBEIRÃO PRETO -SP

2020

Autorizo a reprodução total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Barcelos, Mariana Pegrucci

Planejamento computacional e avaliação *in vitro* de novos potenciais inibidores da fosfomanose isomerase (PMI) de interesse para o controle do cancro cítrico. Ribeirão Preto, 2020.

144 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Filosofia, Ciências e Letras Ribeirão Preto/USP – Área de concentração: Química.

Orientador: da Silva, Carlos Henrique Tomich de Paula.

1. Cancro cítrico. 2. Triagem virtual. 3. Modelagem molecular. 4. Planejamento de inibidores.

FOLHA DE APROVAÇÃO

Barcelos, M.P. **Planejamento computacional e avaliação *in vitro* de novos potenciais inibidores da fosfomanose isomerase (PMI) de interesse para o controle do cancro cítrico**, 2020. 86f. Dissertação (Mestrado). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2020.

Aprovado em:

Banca Examinadora

Prof.

Dr. _____

Instituição: _____

Julgamento: _____

Prof.

Dr. _____

Instituição: _____

Julgamento: _____

Prof.

Dr. _____

Instituição: _____

Julgamento: _____

*Dedico esse trabalho à minha
família, principalmente meus
pais, por sempre me apoiarem
em tudo.*

AGRADECIMENTOS

Primeiramente, gostaria de agradecer aos meus pais, José Anselmo e Elmara, meus maiores exemplos. Mãe, seu cuidado e dedicação me deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinha nessa caminhada.

Não poderia deixar de agradecer ao meu avô, Paulo, que todos os dias me levava para escola e jamais permitiu que eu faltasse.

Ao meu irmão, Lucas, por todo amor e carinho. Aos meus padrinhos, tios, primos e avó que sempre estiveram presentes. Ao meu namorado, Rafael, por todo amor, carinho, paciência e compreensão que tem me dedicado e também pelas risadas nos momentos de aflição.

Ao professor Carlos Henrique que, com muita paciência e atenção, dedicou seu valioso tempo para me orientar em cada passo desse trabalho e também pela contribuição em minha vida acadêmica e influência na minha futura vida profissional.

Aos meus colegas de laboratório do LCQF (Leonardo, João Gabriel, Suzane, Guilherme, Raí, Isaque e Lorane) e ao André do LBBMA. Obrigada pelos momentos de paciência, de brincadeira e, principalmente, pela mão que se estendia quando eu precisava.

Às minhas amigas e irmãs, Leticia, Julia, Marília e Patrícia por todo apoio e cumplicidade, porque mesmo distantes estavam presentes em minha vida.

Às coordenadoras Joze e Cristiane Colégio Objetivo de Batatais por todos conselhos, apoio, torcida, risadas, companheirismo e amizade. Aos alunos, professores e funcionários do Colégio, o meu muito obrigada.

Obrigada a todos que, mesmo não citados aqui, tanto contribuíram para a conclusão desta etapa e para quem eu sou hoje.

*“A ciência compõe-se de
erros que, por sua vez, são
os passos até a verdade”.*
Júlio Verne

RESUMO

Barcelos, M.P. **Planejamento computacional e avaliação *in vitro* de novos potenciais inibidores da fosfomanose isomerase (PMI) de interesse para o controle do cancro cítrico**, 2020. 86 f. Dissertação (Mestrado). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2020.

O cancro cítrico, causado pela bactéria *Xanthomonas citri* subsp. *citri* (XAC), é uma das doenças mais importantes na citricultura, e apresenta apresentando um significativo impacto na economia brasileira. Estudos de análise proteômica comparativa da superfície de células infectantes e não infectantes de XAC realizados no Laboratório de Bioquímica e Biologia Molecular Aplicada (LBBMA) do Departamento de Genética e Evolução (DGE) da Universidade Federal de São Carlos – UFSCar, apontaram para uma promissora enzima-alvo para o controle da doença, a fosfomanose isomerase (PMI). A PMI é uma enzima que catalisa a interconversão da frutose-6-fosfato (F6P) e da manose-6-fosfato (M6P) e é considerada um potencial alvo terapêutico devido a sua participação na sobrevivência e patogenicidade de diversos microrganismos como *Saccharomyces cerevisiae*, *Candida albicans*, *Mycobacterium smegmatis* e *Leishmania mexicana*. O presente projeto de pesquisa tem como meta o desenvolvimento de compostos químicos capazes de inibir a PMI de *Xanthomonas citri* subsp. *citri* encontrada na superfície bacteriana e, como consequência, investigar a contribuição para a diminuição da virulência da bactéria e/ou para o controle da infecção na planta. O estudo consistiu no desenvolvimento de um modelo tridimensional para o PMI de XAC utilizando o método de “modelagem molecular por homologia estrutural”. A metodologia utilizada para a seleção dos compostos foi a técnica de triagem virtual por forma e similaridade, além de simulações de *docking* e análises de propriedades farmacocinéticas e toxicológicas (ADME/Tox). Os modos de interações dos compostos selecionados pelos passos anteriores foram analisados através de uma inspeção visual, obtenção por fim, aqueles mais promissores. Essas moléculas foram adquiridas para serem testadas em ensaios de atividade *in vitro* utilizando o reagente de Seliwanoff.

Palavras-chaves: Cancro cítrico. Fosfomanose isomerase. Metodologias Computacionais

ABSTRACT

Barcelos, M.P. **Computational plannin and *in vitro* evaluation of new phophomannose isomerase (PMI) inhibitors of interest for citrus canker control**, 2020. 86 f. Dissertation (Master's Degree). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2020.

Citrus cancer, caused by the bacteria *Xanthomonas citri* subsp. *citri* (XAC), is one of the most important diseases in citrus, and has a significant impact on the Brazilian economy. Comparative protein analysis studies of XAC infecting and non-infecting cells performed at the Laboratory of Biochemistry and Applied Molecular Biology (LBBMA) of the Department of Genetics and Evolution (DGE) of the Federal University of São Carlos - UFSCar, pointing to a promising target enzyme for disease control, a phosphomannose isomerase (PMI). PMI is an enzyme that catalyzes the interconversion of fructose-6-phosphate (F6P) and mannose-6-phosphate (M6P), and is considered a potential therapeutic target due to its participation and pathogenicity of several microorganisms such as *Saccharomyces cerevisiae*, *Candida albicans*, *Mycobacterium smegmatis* and *Leishmania mexicana*. This research project aims to develop chemical compounds capable of inhibiting XAC PMI found on the bacterial surface and, as a consequence, to investigate a contribution to bacterial virulence and/or control. of infection in the plant. The study consisted of the development of a three-dimensional model for XAC PMI using the "structural homology molecular modeling" method. The methodology used for the selection of compounds was the virtual sorting technique by form and similarity, as well as *docking* simulations and pharmacokinetic and toxicological properties analysis (ADME/Tox). The interaction modes of the compounds selected by the previous steps were analyzed by visual inspection, finally obtaining the most promising ones. These molecules were purchased to be tested in *in vitro* activity assays using Seliwanoff reagent.

Key-words: Citrus canker. Phosphomannose isomerase. Computational Methodologies.

LISTA DE FIGURAS

Figura 1 - Sintomas do cancro cítrico presente nos frutos, folhas e caules de uma laranjeira.	20
Figura 2 - Ciclo do cancro cítrico mostrando como ocorre a infecção e os métodos de propagação da doença.	21
Figura 3 - Reação de isomerização entre manose-6-fosfato (M6P) e frutose-6-fosfato (F6P).	23
Figura 4 - Alterações de proteínas na superfície de <i>Xanthomonas citri</i> subsp. <i>citri</i> por análise proteômica em células cultivadas sob condições infecciosas e não-infecciosas e suas possíveis funções.	26
Figura 5 - Resultado obtido pelo BLASTp na procura de proteínas similares a fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i> (PDB ID: 2X5S, 2QH5, 2CU2 e 1H5R)	40
Figura 6 - Alinhamento múltiplo final das sequências de fosfomanose isomerase e as respectivas às estruturas indicadas como homólogas pelo BLASTp (PDB ID 1H5R, 2CU2, 2X5S e 2QH5) obtido pelo Clustal Omega	41
Figura 7 - Alinhamento múltiplo final das sequências de fosfomanose isomerase e as respectivas às estruturas indicadas como homólogas pelo BLASTp (PDB ID 1H5R, 2CU2, 2X5S e 2QH5) após o refinamento.	43
Figura 8 - Modelo gerado para a fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i> . utilizando o programa MODELLER	47
Figura 9 - Modelo final da fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i> , obtido por modelagem comparativa, com destaque para alguns aminoácidos presentes no sítio ativo.	48
Figura 10 - Ângulos diedros de uma proteína evidenciando os ângulos Phi e Psi.	49
Figura 11 - Diagrama de Ramachandram de <i>Xanthomonas citri</i> subsp. <i>citri</i> obtido pelo PROCHECK : 91,6% de resíduos em regiões favoráveis, 7,9% de resíduos em regiões permitidas, 0,4% de resíduos em regiões generosamente permitidas e 0% de resíduos em regiões proibidas.	50
Figura 12 - Resumo da avaliação estereoquímica do modelo selecionado construído por homologia estrutural para fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i> ., com um G-Factor de -0.1 e distorção tetraédrica do carbono alfa de -0.1.	51
Figura 13 - ERRAT obtido para o modelo de fosfomanose isomerase gerado de <i>Xanthomonas citri</i> subsp. <i>citri</i> com um fator de qualidade de 85.5019.	52
Figura 14 - Verify-3D obtido para o modelo de fosfomanose isomerase gerado de <i>Xanthomonas citri</i> subsp. <i>citri</i> : 91.70% dos aminoácidos obtiveram uma pontuação igual ou superior a 0,2.	53
Figura 15 - PROVE obtido para o modelo fosfomanose isomerase gerado de <i>Xanthomonas citri</i> subsp. <i>citri</i>	54
Figura 16 - Validação do modelo gerado para a fosfomanose isomerase <i>Xanthomonas citri</i> subsp. <i>citri</i> mostrando a posição do ligante glicose-1-fosfato e sua posição em relação do ligante cristalográfico quando utilizado os programas GOLD e Glide	56
Figura 17 - Resultado do Redocking da 1H5R para cálculo do RMSD, no qual foi obtido um valor de 2.5080 Å.	57
Figura 18 - Curva ROC obtida para o estudo dos 30 compostos relatados como mais potentes e menos no BindingDB pelo molde gerado para a proteína fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i>	60

Figura 19 - Curva de Previsão obtida para o estudo dos 30 compostos relatados como ativos e inativos no BindingDB pela proteína fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i>	61
Figura 20 - Curva de enriquecimento com um fator de enriquecimento de 10% obtida para o estudo dos 30 compostos relatados como ativos e inativos no BindingDB pela proteína fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i>	62
Figura 21 - Alinhamento entre as sequências da fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i> e <i>Homo sapiens</i> feita pelo programa Clustal Omega	64
Figura 22 - Alinhamento entre as sequências da fosfomanose isomerase de <i>Homo sapiens</i> e as sequências selecionadas feita pelo programa Clustal Omega	66
Figura 23 - Alinhamento múltiplo final das sequências de fosfomanose isomerase humana (PMI) e as respectivas às estruturas indicadas como homólogas pelo BLASTp . (PDB IDs 5NW7 e 3H1M).	67
Figura 24 - Diagrama de Ramachandram para a hipótese 3 obtido pelo PROCHECK : 89,2% de resíduos em regiões favoráveis, 8,9% de resíduos em regiões permitidas, 1,1% de resíduos em regiões generosamente permitidas e 0,8% de resíduos em regiões proibidas.	69
Figura 25 - Resumo da avaliação estereoquímica da hipótese 3, construída por homologia estrutural para PMI humana.	70
Figura 26 - ERRAT obtido para a hipótese 3 gerado de PMI humana.	71
Figura 27 - Verify-3D obtido para a hipótese 3 da PMI humana: 95,51% dos aminoácidos encontram-se dentro do perfil esperado.	71
Figura 28 - PROVE obtido para o a hipótese 3 da PMI humana.	72
Figura 29 - Curva ROC obtida para o estudo dos 30 compostos relatados como ativos e inativos no BindingDB para a proteína PMI de <i>Homo sapiens</i>	74
Figura 30 - Curva de Previsão obtida para o estudo dos 30 compostos relatados como ativos e inativos no BindingDB para a proteína PMI de <i>Homo sapiens</i>	75
Figura 31 - Curva de enriquecimento com um fator de enriquecimento de 10% obtida para o estudo dos 30 compostos relatados como ativos e inativos no BindingDB para PMI de <i>Homo sapiens</i>	76
Figura 32 - Estrutura da G1P, utilizada como referência para os estudos de triagem virtual por similaridade nos programas ROCS e EON	78
Figura 33 - Representação dos modos de interação do composto selecionado DL00392005 , DL03776644 , DL03874928 , DL03881790 , DL03999102 , DL04091967 , DL05234422 , DL07998085 , DL13518039 , DL13522617 , DL15021627 e DL35935889 na região do sítio ativo do modelo de XAC em duas dimensões obtidas como resultado do programa GOLD	84
Figura 34 – Reação química do método de Seliwanoff.	86
Figura 35 - Esquema do teste empregado na avaliação da atividade enzimática.	86
Figura 36 - Resultado para o ensaio de avaliação da atividade enzimática da PMI mostrando as colorações obtidas para os testes-controles e a reação entre a M6P e a enzima, após a adição do reagente de Seliwanoff e fervura por 5 minutos.	87
Figura 37 - Resultado para o ensaio de avaliação da atividade enzimática da PMI mostrando as colorações obtidas para os testes-controles e a reação entre a M6P e a enzima, após a adição do reagente de Seliwanoff e fervura por 5 minutos.	88

LISTA DE TABELAS

Tabela 1 - Relação do número de compostos presentes em cada base de dados utilizada. ...	33
Tabela 2 - Dados referentes aos ensaios in vitro realizados para a fosfomanose isomerase recombinante de <i>Xanthomonas citri</i> subsp. <i>citri</i>	37
Tabela 3 - Resultados de identidade e e-value de sequências de proteínas obtidas no servidor online BLAST usando o algoritmo para proteínas (BLASTp) na triagem de proteínas homologadas a fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i>	40
Tabela 4 - Apresentação dos resultados da Pontuação Geral (WHATIF), Regiões Proibidas (Ramachandran) e Fator G (PROCHECK) obtidos na etapa de validação dos modelos gerados.	46
Tabela 5 - Resultado da recuperação da seleção de inibidores (15 mais potentes e 15 menos potentes) do BindingDB pela fosfomanose isomerase de <i>Xanthomonas citri</i> subsp. <i>citri</i>	58
Tabela 6 - Resultados de identidade e e-value de sequências de proteínas obtidas no servidor online BLAST usando o algoritmo de explosão de proteínas (BLASTp) para a triagem de proteínas homologadas para PMI humana.	65
Tabela 7 - Apresentação dos resultados da Pontuação Geral (WHATIF), Regiões Proibidas (Ramachandran) e Fator G (PROCHECK) obtidos na etapa de validação dos modelos gerados.	68
Tabela 8 - Resultado da recuperação da seleção de inibidores do BindingDB	73
Tabela 9 - Relação entre os compostos sobreviventes das bases de dados após simulações de <i>docking</i> com esfera de 8 Å, predições ADME/Tox (DEREK e QikProp) e inspeção visual, utilizando a primeira estratégia.	79
Tabela 10 - Relação entre os compostos sobreviventes das bases de dados após estudos triagem virtual por similaridade (ROCS e EON), de <i>docking</i> , predições ADME/Tox (DEREK e QikProp) e inspeção visual, utilizando a segunda estratégia.	80
Tabela 11 - Relação entre os compostos sobreviventes das bases de dados após simulações de <i>docking</i> com esfera de 6 Å e 60 de fitness, triagem virtual por similaridade (ROCS e EON), predições ADME/Tox (DEREK e QikProp) e inspeção visual, utilizando a terceira estratégia.	81
Tabela 12 - Tipo e comprimeto de interações entre os compostos comprados e os aminoácidos presentes no sítio ativo.	83
Tabela 13 - Dados referentes aos ensaios in vitro realizados para a fosfomanose isomerase recombinante de <i>Xanthomonas citri</i> subsp. <i>citri</i>	88
Tabela 14 - Resultados obtidos para os ensaios de inibição da enzima PMI de <i>Xanthomonas citri</i> subsp. <i>citri</i> pelo inibidor DL03874928	89

LISTA DE ABREVIATURAS E SIGLAS

AUC	Área sob a curva
BEDROC	Discriminação reforçada por Boltzmann do ROC
BLAST	Basic Local Aligment Search Tool
DGE	Departamento de Genética
EF	Fator de enriquecimento
EMBL-	European Molecular Biology Laboratory - European
EBI	Bioinformatics Institute
F6P	frutose-6-fosfato
FDA	Food and Drug Administration
FFP	Frações Falso-Positivas
FVP	Frações Verdadeiras Positivas
G1P	Glicose-1-fosfato
G-Factor	General Factor
LBBMA	Laboratório de Bioquímica e Biologia Molecular Aplicada
M1P	Manose-1-fosfato
M6P	manose-6-fosfato
MDCK	Madin-Darby Canine Kidney
NCBI	National Center for Biotechnology Information
ONSA	Organization for Nucleotide Sequency and Analysis
PDB	Protein Data Bank
PMI	fosfomanose isomerase
RIE	Reforço inicial robusto
RMSD	Root-Mean-Square Deviation
ROC	Receiver Operating Characteristics
Sdf	Spatial Data Format
TG	Total padronizado
UFSCar	Universidade Federal de São Carlos
XAC	<i>Xanthomonas citri</i> subsp. <i>citri</i>
XAU-B	<i>Xanthomonas fuscans</i> subsp. <i>aurantifolli</i> tipo B
XAU-C	<i>Xanthomonas fuscans</i> subsp. <i>aurantifolli</i> tipo C

Sumário

1. INTRODUÇÃO	18
1.1 O cancro cítrico	18
1.2 O agente etiológico do cancro cítrico: a bactéria <i>Xanthomonas</i> sp.	21
1.3 Fosfomanose isomerase	23
1.4 Planejamento computacional de substâncias bioativas	26
2. OBJETIVOS	29
3. MATERIAIS E MÉTODOS	30
3.1. Construção do modelo de fosfomanose isomerase para <i>Xanthomonas citri</i> subsp. <i>citri</i>	30
3.1.1 Pesquisa e Alinhamento de sequências homólogas	30
3.1.2 Alinhamento das sequências	30
3.1.3 Construção e avaliação da qualidade do modelo	31
3.1.4. Validação do modelo de PMI gerado para simulações de <i>docking</i>	32
3.2 Triagem virtual de novos inibidores da fosfomanose isomerase	32
3.2.1. Geração de confôrmeros (OMEGA)	33
3.2.2. Seleção de inibidores baseados em ligantes (ROCS e EON)	34
3.2.3. Seleção de inibidores baseados em estrutura (GOLD).....	34
3.2.4 Predição das propriedades farmacocinéticas e toxicológicas (QikProp e DEREK)	34
3.3. Visualização Gráfica	35
3.4. Ensaio de atividade <i>in vitro</i> com a enzima PMI de <i>Xanthomonas citri</i> subsp. <i>citri</i>	35
3.4.1 Método de avaliação da atividade enzimática.....	36
3.4.2 Avaliação <i>in vitro</i> da inibição da fosfomanose isomerase pelos compostos selecionados	36
4. RESULTADOS E DISCUSSÃO	38
4.1 Construção do modelo da fosfomanose isomerase (PMI) para <i>Xanthomonas citri</i> subsp. <i>citri</i> (XAC)	38
4.1.1 Etapa de identificação da sequência de aprovação	38
4.1.2 Etapa de alinhamento das sequências	41
4.1.3 Etapa de construção e avaliação do modelo.....	44
4.1.4 Avaliação dos ângulos diédricos Ψ versus ϕ através do Diagrama de Ramachandran	48
4.1.5 Outros parâmetros importantes para a análise (Suite SAVE v5.0).....	50
4.1.6 Etapa de validação do modelo de PMI XAC.	54
4.2 Construção do modelo da fosfomanose isomerase (PMI) para <i>Homo sapiens</i>	63
4.1.1 Etapa de identificação da sequência de aprovação	65

4.1.2 Etapa de alinhamento	65
4.1.3 Etapa de construção e avaliação do modelo	67
4.1.4 Avaliação dos ângulos diédricos Ψ versus φ através do diagrama de Ramachandran	69
4.1.5 Outros parâmetros importantes para a análise dos modelos de PMI humana	70
4.1.6 Etapa de validação do modelo de PMI humana	72
4.3 Triagem virtual par seleção de inibidores da PMI de XAC.	77
4.4 Ensaios de atividade in vitro com a enzima PMI.	85
4.4.1. Método de avaliação da atividade enzimática	85
4.4.2. Avaliação in vitro da inibição da PMI pelo composto selecionado	87
5. CONCLUSÃO.....	90
6.PERSPECTIVAS FUTURAS	91
REFERÊNCIAS BIBLIOGRÁFICAS	92
APÊNDICES	99
APÊNDICE A - Dados relacionados a construção do modelo de PMI <i>Xanthomonas citri</i> subsp. <i>citri</i>	99
APÊNDICE B - Dados relacionados a construção do modelo de PMI <i>Homo sapiens</i>	115

1. INTRODUÇÃO

O Brasil é o maior produtor e exportador de laranjas do mundo, sendo a citricultura aquela com maior relevância dentro do agronegócio (CEPEA - CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA, [s.d.]). Dois fatores contribuem para que o país ocupe essa posição: o tamanho territorial e o clima favorável. De acordo com o Departamento de Agricultura dos Estados Unidos, a produção brasileira de laranja e suco de laranja apresentou uma média de 34% e 56% da produção mundial entre as safras de 2012 e 2015. Aliando essa participação com o avanço da tecnologia empregada nos cinturões, o Brasil apresenta uma participação de 76% do mercado mundial de citros (NEVES; TROMBIN, 2017). Apesar de todos os empenhos empregados para manter a alta produtividade nos pomares brasileiros, a citricultura tem enfrentado alguns problemas como algumas doenças que atacam os citros, sendo uma das principais o cancro cítrico (GOTTWALD et al., 2001).

1.1 O cancro cítrico

O cancro cítrico constitui uma das principais doenças ocasionadas em pomares de citros e é causado pela bactéria de gênero *Xanthomonas sp.*, sendo o tipo mais severo ocasionado por *Xanthomonas citri* subsp. *citri* (XAC). Os sintomas são lesões circulares, salientes, corticosas, de coloração castanha, bem como eruptivas, ocorrendo em folhas, ramos e frutos (BITTANCOURT, 1957; ROSSETTI, 2011). Em casos mais severos, ocorre a queda dos frutos imaturas impactando na perda direta da produção (BEHLAU; BELASQUE JR, 2014).

O controle da doença cítrica baseia-se em medidas de erradicação e exclusão, além de medidas regulamentárias determinantes da erradicação de plantas e pomares afetados pela doença (GOTTWALD et al., 2001). Em 1999, a Defesa Agropecuária do Estado de São Paulo determinou que todas as plantas dos talhões infestados deveriam ser eliminadas obrigatoriamente quando a incidência fosse superior a 0,5% e para aquelas em que a incidência fosse igual ou menor que 0,5%, as plantas doentes e as demais contidas em um raio de trinta metros também seriam eliminadas (BEHLAU; FONSECA; BELASQUE, 2016). Esta última medida foi tomada após o aparecimento da lagarta mineradora dos citros (*Phyllocnistis citrella*) nos laranjais, cuja presença contribuiu para o aumento da incidência do cancro cítrico (SEILLE; CHRISTIANO, 2006) e, conseqüentemente, tornou ineficiente a aplicação do raio de trinta metros para a retenção da doença. (BEHLAU; BARELLI; BELASQUE, 2014; FERENCO et al., 2018).

A incidência da doença voltou a crescer nos laranjais do estado de São Paulo desde que a legislação foi amenizada, em meados de 2009, excluindo a primeira exigência de erradicação de um talhão de plantas correspondente a contaminações iguais ou maiores a 0,5%. Entre os anos de 2001 e 2009, a doença manteve baixas incidências de talhões contaminados com valores variando entre 0,08% e 0,20%, mas, a partir de 2010, a contaminação avançou e atingiu valores de 0,44% de talhões contaminados. Por fim, em 2012, atingiu 1,39% registrando a maior incidência desde seu primeiro relato em 1957 (FUNDECITRUS, 2017a).

Em 2017, uma nova medida foi tomada: caso a doença apareça em uma área onde a praga ocorre com baixa incidência, o controle é feito principalmente pelo plantio de mudas sadias e pela eliminação das plantas cítricas contaminadas e daquelas suspeitas de contaminação. Por outro lado, caso haja uma área em que o cancro cítrico já esteja difundido entre as plantas, deverá ser realizado um conjunto de medidas de manejo, como a utilização de um composto de cobre, que visam reduzir o impacto da doença na produção (FUNDECITRUS, 2016, 2017b, 2017c)

A bactéria responsável pela doença cítrica possui a capacidade de infectar folhas, frutos e ramos (Figura 1), mesmo não possuindo um mecanismo ativo de penetração através da parede celular. Como sua propagação não necessita de um vetor específico, em condições naturais de infecção a disseminação ocorre a curtas distâncias pela ação de respingos de chuvas, principalmente quando associados com ventos, e a disseminação a longas distâncias ocorre pelo transporte de material vegetal infectado (DANOS, E; BERGER, R. D.; STALL, 1984; FUNDECITRUS, 2017a; GOTTWALD, T.R., MCGUIRE, R.G. & GARRAN, 1988). Por isso, é necessária a manutenção dos equipamentos de colheita através da desinfecção, além da pulverização dos veículos antes de entrar na propriedade para impedir a difusão da doença e assim manter a propriedade livre da mesma.

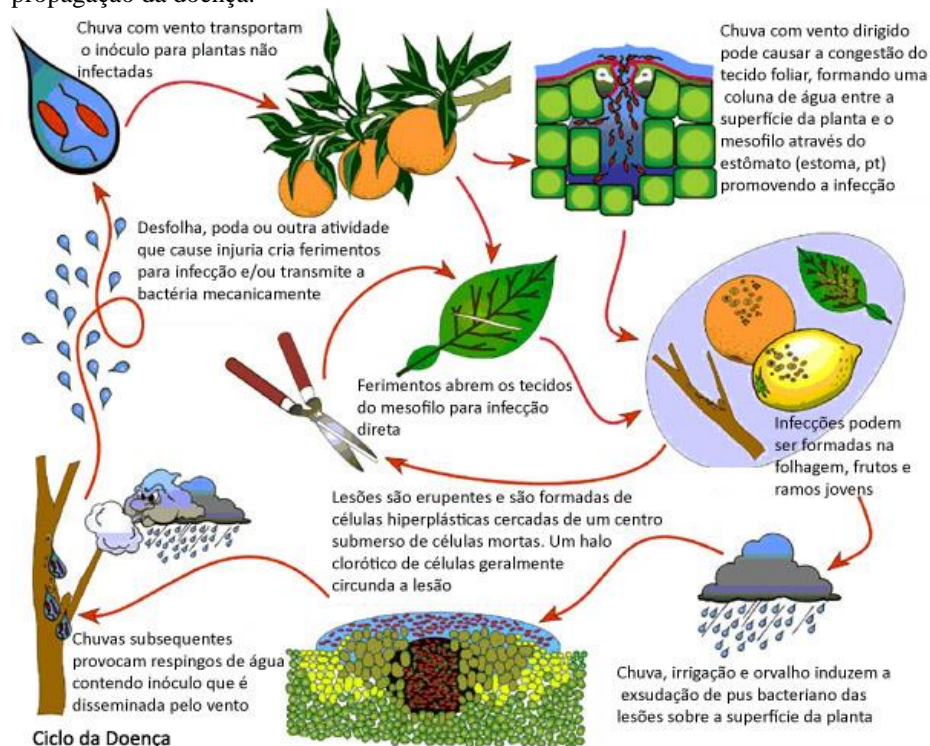
Figura 1 - Sintomas do cancro cítrico presente nos frutos, folhas e caules de uma laranjeira.



Fonte: Schubert, 2011.

O ciclo de infecção (Figura 2) começa pelo acúmulo anormal de água nas folhas, o que acarreta a formação de uma coluna de água entre a superfície foliar e o mesófilo dos estômatos e/ou hidatódios, permitindo a entrada do patógeno. Além disso, essa água causa a mobilização de bactérias para fora das lesões foliares, as quais são células hiperplásticas em torno de um centro de células mortas colapsadas, em que um anel ou halo clorótico frequentemente rodeiam a lesão. Após a saída da bactéria pela lesão foliar, o patógeno é carregado pelo vento, disseminando a doença e inoculando plantas não infectadas. Espinhos, a poda ou outras atividades mecânicas podem causar ferimentos nas folhas criando uma porta para a infecção e/ou transmissão bacteriana (GOTTWALD, TIM R.; GRAHAM, 2005) e o ciclo recomeça.

Figura 2 - Ciclo do cancro cítrico mostrando como ocorre a infecção e os métodos de propagação da doença.



Fonte: Cardoso, 2013.

1.2 O agente etiológico do cancro cítrico: a bactéria *Xanthomonas* sp..

O agente etiológico do cancro cítrico é uma bactéria pertencente ao gênero das *Xanthomonas*, com metabolismo oxidativo, Gram-negativa, aeróbica e não fixadora de nitrogênio (ARTIER, 2010). Há diferentes tipos de cancro cítrico causados pelas diferentes variantes da bactéria *Xanthomonas citri*. Como os sintomas são geralmente iguais, a sua distinção ocorre por meio do hospedeiro e pelas características fenotípicas e genotípicas, sendo assim, a doença é dividida em cancores do tipo “A”, “B”, “C”, “D” e “E” (GRAHAM et al., 2004), porém, neste estudo, será abordada apenas A, B e C, devido ao prejuízo econômico acometido pelas mesmas, pois são as com maior importância econômica.

A cancorese tipo **A** corresponde a forma mais severa da doença, sendo tipicamente da Ásia e causada por *Xanthomonas citri* subsp. *citri* (XAC), possuindo a maior gama de hospedeiros cítricos, como por exemplo, a laranja doce (*Citrus sinensis*), limão (*C. limon*) e lima ácida (*C. aurantifolia*) (STALL; SEYMOUR, 1983). É a estirpe que está mais difundida no mundo, atingindo o Japão, Argentina, Brasil, Paraguai, China, entre outros países.

A cancorese tipo **B** causada por *Xanthomonas fuscans* subsp. *aurantifolii* (Xau-B) é menos severa e de importância secundária quando comparada com a cancorese tipo **A** e atinge,

principalmente, os limoeiros verdadeiros da Argentina, Paraguai e Uruguai. A cancrose tipo **B** pode ser facilmente diferenciada do tipo **A**, mas não pela cancrose causada pelo tipo **C** (BEHLAU et al., 2007; GOTTWALD et al., 2001).

A cancrose tipo **C**, também conhecida como cancrose da limeira ácida “galego”, é causada por *Xanthomonas fuscans* subsp. *aurantifolii* tipo **C** (*Xau-C*) ocorrendo em único tipo de hospedeiro, a lima ácida. Ocorre somente no Brasil, afetando apenas alguns municípios do estado de São Paulo (BEHLAU et al., 2007; GOTTWALD et al., 2001). Como cada linhagem possui características próprias, marcadores específicos poderiam eventualmente auxiliar no diagnóstico de plantas infectadas no campo, especialmente em períodos que os sintomas não estejam presentes.

O sequenciamento completo dos genes da bactéria causadora da cancrose tipo **A** foi realizado pelos laboratórios da rede ONSA (*Organization for Nucleotide Sequencing and Analysis*) da FAPESP (DA SILVA et al., 2002a). Os resultados mostraram que o seu genoma é composto por 5.175.554 pares de bases, os quais codificam 4.313 proteínas. A comparação entre os diversos genomas das linhagens dessas bactérias poderia nos fornecer informações sobre o hospedeiro alvo e sua patogenicidade. Devido a esse fato, os projetos de sequenciamento das linhagens **B** (*XAU-B*) e **C** (*XAU-C*) também foram realizados por grupos de pesquisa da USP-SP e UNESP- Jaboticabal, com o suporte da FAPESP e do FUNDECITRUS (BODNAR et al., 2017; JALAN et al., 2011; MOREIRA et al., 2010; RICHARD et al., 2017) para que seja possível a comparação entre as sequências (ESCALON et al., 2013; GORDON et al., 2015; JALAN et al., 2013).

A *XAC* possui o sistema de secreção tipo III (SSTT) no qual a bactéria “percebe” a planta ou as condições do ambiente como favoráveis, o que ativa a produção proteica e o SSTT. Em sua grande maioria, as bactérias patogênicas gram-negativas de plantas têm dois grupos de genes modulares de suas interações com os hospedeiros, os genes *avr* (avirulência) responsáveis por determinar a especificidade do mesmo por uma interação genótipo-genótipo, e os genes *hrp* (resposta de hipersensibilidade e patogenicidade) relacionados à patogenicidade e indução de respostas de hipersensibilidade em plantas não-hospedeiras (BONAS, 1994; MOREIRA et al., 2004).

Os genes *hrp*, que se encontram ativados, produzem uma estrutura de adesão com a célula chamada de *pilus*, e através dele e da formação de um poro condutor, as proteínas são transferidas diretamente para o citosol da célula hospedeira. Uma dessas proteínas são as *avr* que, como dito anteriormente, causam uma reação de hipersensibilidade. Crê-se que algumas

dessas proteínas transportadas possam causar os sintomas da doença na planta hospedeira (BUTTNER, 2002).

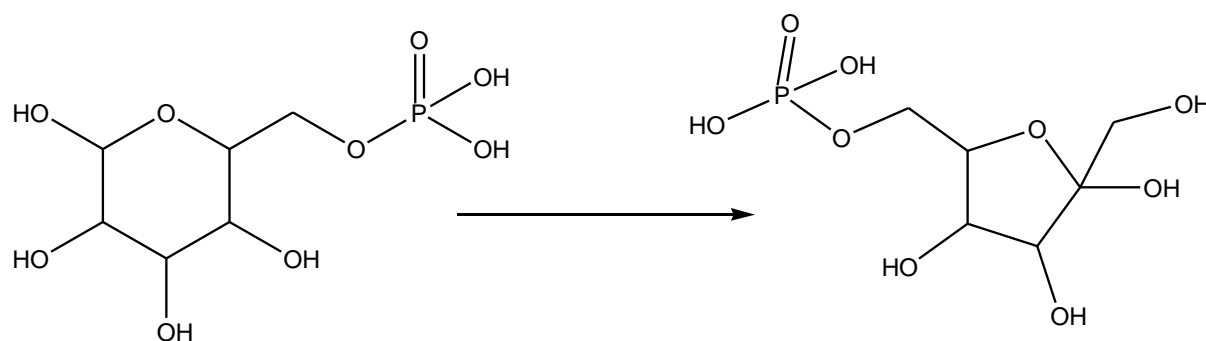
Além do Sistema de Secreção Tipo III, XAC., utilizam um mecanismo de conjugação para exportar moléculas efetoras (proteínas ou complexos DNA-proteínas) durante a infecção (YACOUBI, 2005). Estes sistemas são chamados de Sistemas de Secreção Tipo IV ou “Conjugação” Adaptada e secretam proteínas efetoras para o meio extracelular ou para o citosol das células hospedeiras, envolvendo o contato direto célula-célula (YACOUBI, 2005).

Diversos estudos sobre os mecanismos específicos das interações de XAC com seus hospedeiros têm sido realizados com a finalidade de entender a maneira pela qual a bactéria utiliza o Sistema de Secreção tipo III e o Sistema de Secreção tipo IV para interagir e modificar o metabolismo das plantas infectadas (FERENCE et al., 2018; ZHANG et al., 2015).

1.3 Fosfomanose isomerase

A fosfomanose isomerase (PMI), alternativamente chamada de manose-6-fosfato isomerase é uma enzima que catalisa a interconversão da frutose-6-fosfato (F6P) e da manose-6-fosfato (M6P) (Figura 3) além de permitir a síntese do difosfato de guanosina manose (GDP-manose) em organismos eucarióticos. A PMI é monomérica e uma metaloenzima, ou seja, é dependente de um íon metálico divalente para desempenhar as funções de atividade e catálise, sendo inibida pela eritose-4-fosfato, manitol-1-fosfato e, em menor grau, pelo anômero alfa da manose-6-fosfato (GAO; YU; LEARY, 2005).

Figura 3 - Reação de isomerização entre manose-6-fosfato (M6P) e frutose-6-fosfato (F6P).



Fonte: O autor

A fosfomanose isomerase converte uma aldose (manose) em uma cetose (frutose), além de abrir e fechar os anéis para esses açúcares (STRYER, L.; BERG, J.M., TYMOCZKO, 2007). Essa enzima mostra um alto grau de seletividade para o anômero beta de M6P enquanto que o

anômero alfa não possui atividade e pode, de fato, atuar como um inibidor (ROSE, IRWIN A.; O'CONNELL, EDWARD L.; SCHRAY, 1973).

Proudfoot e colaboradores (PROUDFOOT et al., 1994) definiram três tipos de PMIs com base no alinhamento da sequência e propriedades físico-químicas e cinéticas. As enzimas dos mamíferos e da maioria das bactérias pertencem às PMIs do Tipo I, que são proteínas monofuncionais homólogas que catalisam uma única reação de isomerização (PROUDFOOT et al., 1994; PROUDFOOT; PAYTON; WELLS, 1994). Aquelas pertencentes ao Tipo II são enzimas bifuncionais e possuem similaridade de sequência limitada às enzimas do Tipo I, exceto por um motivo muito curto (9 aminoácidos) (JENSEN; REEVES, 1998). Por fim, as enzimas do Tipo III catalisam reações de isomerização reversíveis, mas compartilham pouca identidade com sequências das enzimas do Tipo I e do Tipo II (SIGDEL et al., 2015).

A enzima presente em *Xanthomonas citri* subsp. *citri* é composta por 467 resíduos e pertence à família da manose-1-fosfato guanililtransferase e manose-6-fosfato isomerase, portanto, trata-se de uma enzima bifuncional (DA SILVA et al., 2002b).

A fosfomanose isomerase apresenta várias contribuições para as vias metabólicas necessárias. Ela permite que as células transformem o M6P em F6P, que poderá ser inserido na glicólise e também permitirá que as células convertam a F6P em M6P, um identificador celular glicolítico comum para o transporte celular e a identificação da membrana celular em organismos procarióticos e eucarióticos (STRYER, L.; BERG, J.M., TYMOCZKO, 2007). Esta enzima pode ser útil no desenvolvimento de novos tratamentos antifúngicos, uma vez que a falta de atividade da PMI em células de levedura pode levar à lise celular e a enzima pode ser alvo de inibição (CLEASBY et al., 1996). Isso pode ocorrer devido ao papel da PMI na formação das paredes celulares e na biossíntese de polissacarídeos capsulares, além do mais, a M6P é uma molécula de sinalização importante, especialmente para o transporte para os lisossomos. Distúrbios que afetam a atividade da PMI podem afetar a capacidade celular para produzir rapidamente a M6P a partir da F6P abundante e, portanto, o tráfego de vesículas para os lisossomos e os endossomos pode ser alterado, impactando negativamente a célula e acarretando uma lise celular (JAEKEN; MATTHIJS, 2001).

A PMI é considerada como potencial alvo terapêutico por causa da participação na sobrevivência e patogenicidade de diversos microrganismos como *Saccharomyces cerevisiae* (PAYTON et al., 1991), *Candida albicans* (CLEASBY et al., 1996), *Mycobacterium smegmatis* (PATTERSON et al., 2003) e *Leishmania mexicana* (GARAMI; ILG, 2001). Dessa forma, a informação estrutural sobre essa enzima presente em diferentes organismos ajudará na

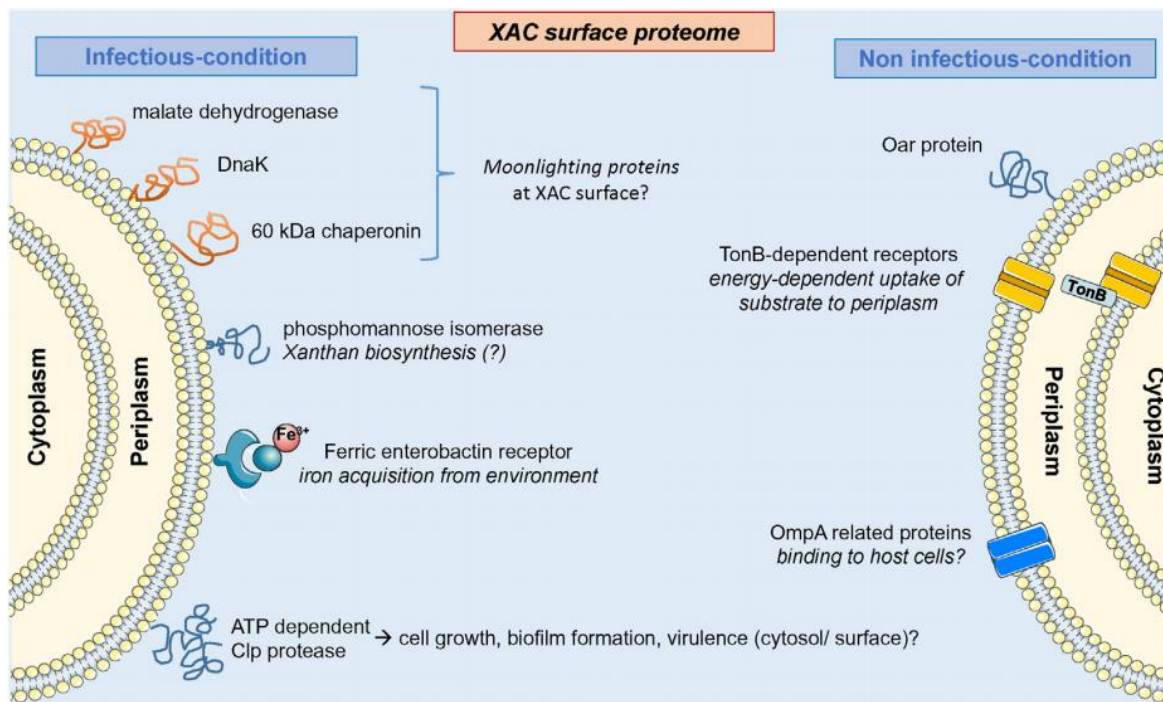
concepção e otimização de compostos que inibam seletivamente as enzimas de organismos patogênicos, de tal forma que não afetem as células do hospedeiro, mantendo-as saudáveis.

A primeira estrutura tridimensional dessa enzima foi depositada em 2004 no Protein Data Bank (PDB) (KIRCHMAIR et al., 2008) e pertence à espécie *Bacillus subtilis* (código PDB 1QWR, a 1.8 Å de resolução). Essa estrutura revelou a existência de um local ativo na superfície do domínio central com um único átomo de zinco ligado. Estudos cinéticos demonstraram a necessidade deste cofator de metal para a atividade da PMI (NOLTMANN, 1968). A *Ca*-PMI mostrou ser inibida por fosfatos de açúcares como o 4-fosfato de eritrose em concentrações milimolares (PROUDFOOT; PAYTON; WELLS, 1994). Além do zinco no local catalítico, demonstrou-se que PMI liga Zn^{2+} em dois outros locais com menor afinidade (PROUDFOOT et al., 1994). Após a resolução dessa estrutura acima descrita, mais quatro estruturas de MPI contendo Zn^{2+} foram depositadas, sendo duas com o substrato F6P (PMI de *Salmonella thyphimurium*, depositada em 2009 sob o código PDB 3HLI, a 2.08 Å; PMI de *Pyrococcus furiosus*, depositada em 2006 – código PDB 2GC3, a 2.1 Å) e duas sem (códigos PDB 3H1M de 2.5 Å e 3H1W de 1.94 Å depositadas em 2009, de *S. thyphimurium*).

O papel do cofator de zinco na catálise ainda não está claro. Uma vez que, a elucidação dos mecanismos catalíticos e regulatórios da PMI é essencial para o planejamento racional de fármacos, várias outras técnicas, como estudos de inibição, (ROUX et al., 2004; WELLS et al., 1995; WELLS; PAYTON; PROUDFOOT, 1994) rótulos específicos dos sítios (COULIN et al., 1993; PAPOUTSOPOULOU; KYRIAKIDIS, 1997; WELLS; SCULLY; MAGNENAT, 1994) e mutações direcionadas ao sítio (WELLS et al., 1995; WELLS; SCULLY; MAGNENAT, 1994) têm sido empregadas para gerar informações complementares.

Em *Xanthomonas*, a fosfomanose isomerase está relacionada com a produção de goma xantana, a qual é um mecanismo de defesa da bactéria para evitar sua desidratação em ambientes nocivos (PAPOUTSOPOULOU; KYRIAKIDIS, 1997). Em estudos realizados anteriormente no LBBMA-UFSCar por Carnielli e colaboradores (CARNIELLI et al., 2017), a fosfomanose isomerase foi encontrada na superfície da bactéria *Xanthomonas citri subsp. citri* em condições infecciosas (Figura 4), o que demonstra o caráter inédito e inovador desta pesquisa em identificar novos inibidores para essa enzima.

Figura 4 - Alterações de proteínas na superfície de *Xanthomonas citri* subsp. *citri* por análise proteômica em células cultivadas sob condições infecciosas e não-infecciosas e suas possíveis funções.



Fonte: Carnielli et al., 2017.

1.4 Planejamento computacional de substâncias bioativas

A descoberta de novas tecnologias de informação e novas ferramentas computacionais, durante os últimos anos, aumentaram a quantidade de dados em relação a moléculas, estruturas, proteínas, alvos e mecanismos de ação de fármacos (WERMUTH, 2008).

Devido ao avanço de técnicas como cristalografia e ressonância magnética nuclear (RMN), diversas estruturas estão disponíveis em bancos de dados e podem ser aplicadas em estratégias de planejamento computacional de fármacos (JORGENSEN, 2004, 2009).

Atualmente, o planejamento racional *in silico* é a metodologia mais eficiente e menos dispendiosa para o desenvolvimento de novos compostos, capaz de contribuir em todos os estágios do processo, desde a descoberta de protótipos (também conhecidos como “compostos de partida” ou “lead compounds”), sua otimização (com respeito à afinidade, especificidade, eficácia e efeitos colaterais), até a elaboração de compostos candidatos a testes clínicos (OPREA, 2005). Esta metodologia é baseada na inibição ou estimulação da atividade biológica de macromoléculas, tais como proteínas ou ácidos nucleicos (DNA ou RNA) associados a diferentes doenças. A informação estrutural do bioreceptor (e ligantes) permite a descoberta e direciona a síntese de compostos com complementaridade estérica, hidrofóbica e eletrostática ao seu sítio de ligação, os quais podem vir a se tornar potentes fármacos. Essa abordagem, em

sua essência, caracteriza o planejamento racional de fármacos por vias computacionais. O que torna ainda mais atrativo tal *approach*, quando utilizado com proteínas, é o conhecimento de que 78% dos fármacos atuais têm como alvo receptor esse tipo de biomolécula (OPREA, 2005).

Desde a concepção do alvo biológico até a descoberta de um novo fármaco, um processo que pode levar em média 11 anos ou mais, a química computacional vem se tornando, a cada dia, indispensável no planejamento racional de novos fármacos, já com inúmeros casos de sucesso envolvendo o emprego de suas metodologias (OPREA, 2005). Para Christopher Watson, da Novartis Institutes for BioMedical Research, o processo de descoberta e desenvolvimento de novos fármacos é hoje totalmente dependente das metodologias computacionais envolvidas (PEITSCH, 2004).

Em química computacional (ou “químioinformática”), genericamente citada como modelagem molecular, destaca-se o *docking* molecular como um dos métodos mais empregados. Com essa metodologia, são investigadas as possíveis orientações que determinada molécula assume no interior do sítio ligante de um bioreceptor, ou simplesmente entre duas moléculas, tal como é o caso da interação entre proteína-proteína ou proteína-DNA, caracterizando o “*docking* macromolecular”.

Os métodos de *docking*, em geral, envolvem uma função de energia contendo parâmetros eletrostáticos, de van der Waals, de ligações de hidrogênio e, algumas vezes, hidrofóbicos, baseados em modelos matemáticos que predizem as melhores orientações do ligante, segundo uma lista de escores de energia. O “Problema do *docking*” é assim centrado na geração e avaliação (quase sempre por intervenção do usuário) de plausíveis estruturas de complexos intermoleculares (CODDING, 1998). As mais recentes versões dos *programas* de maior sucesso em *docking*, **GOLD** (NISSINK et al., 2002) e **GLIDE** (FRIESNER et al., 2004), consideram a flexibilidade do ligante e, também, de algumas cadeias laterais do sítio receptor.

O método “*screening* (ou triagem) virtual” é utilizado para a seleção de novos protótipos, e a literatura vem reportando, há alguns anos, diversos casos de sucesso com o uso dessa sistemática, tal como a descoberta de isoflavonóides como inibidores não-esteroidais de 5 α -redutase, utilizando *constraint* farmacofórica (BRENK et al., 2003; CHEN; LIU; GILSON, 2001). Nessa era pós-genômica, o *screening* virtual complementa os conhecidos métodos experimentais de *screening* em larga escala no processo de descoberta de novos protótipos (KUNZ, 1992). O sucesso do *screening* virtual, e em geral de *docking*, depende do conhecimento de detalhes estruturais finos do sítio receptor (CARLSON; MASUKAWA; MCCAMMON, 1999).

Para o desenvolvimento de um novo fármaco também é possível hoje estimar propriedades farmacocinéticas e farmacodinâmicas, bem como propriedades “*drug-like*” ou “*lead-like*” de diferentes compostos, selecionando, durante as diversas etapas da modelagem, somente aqueles com potencial de serem fármacos. Como exemplo, a ‘Regra dos 5’ (RO5), de Lipinski, que é uma regra que os fármacos oralmente ativos e em geral seguem: peso molecular menor ou igual a 500, $\log P$ menor ou igual a 5, número de grupos doadores de ligações de hidrogênio menor ou igual a 5 e número de grupos aceptores de ligações de hidrogênio menor ou igual a 10 (LIPINSKI; HOPKINS, 2004). Citações em *CAS SciFinder* do artigo original da RO5, de 1997, excederam 1000 somente no ano de 2004, e continuam crescendo (LIPINSKI, 2004). Uma variação dessa regra é aplicável a protótipos e, além disso, novas regras empíricas vêm sendo reportadas, tais como o número de ligações rotacionáveis em fármacos ser menor do que 8 ou, ainda, a área superficial ser menor ou igual a 140 \AA^2 (OPREA, 2005).

Sendo assim, propriedades tais como absorção, distribuição, metabolismo, excreção e toxicidade (ADME/Tox) podem ser preditas a partir do *screening* em bancos de dados contendo essas informações, as quais são computadas para uma grande variedade de compostos (BAJORATH, 2004; CARLSON; MASUKAWA; MCCAMMON, 1999). Entre os *programas* mais utilizados para esses fins, destacam-se o **ADMET Predictor** (ADMET), **Meteor** (metabolismo) e **Derek** (toxicidade).

Dessa forma, utilizamos tais técnicas computacionais para o estudo de novos inibidores capazes de atuar na enzima PMI de XAC, com o objetivo de controlar o cancro cítrico.

2. OBJETIVOS

O presente estudo tem como objetivo o planejamento de novos potenciais inibidores de PMI utilizando metodologias computacionais, adquirir comercialmente alguns desses compostos selecionados e testá-los *in vitro*.

As metodologias computacionais incluem, especificamente, experimentos de *screening* (ou triagem) virtual em diferentes bases de dados de compostos comerciais contendo propriedades de fármacos, utilizando variadas estratégias de Química Medicinal Computacional.

Os objetivos específicos do trabalho incluem ainda:

- Desenvolvimento, validação e avaliação de um modelo proposto para a PMI de XAC através de técnicas de modelagem de proteínas por homologia estrutural;
- Desenvolvimento, validação e avaliação de um modelo proposto para a PMI humana através de técnicas de modelagem de proteínas por homologia estrutural visando sua diferença em relação ao sítio ativo com a PMI de XAC;
- Tratamento de 10 bases de dados e geração da biblioteca de confôrmeros utilizados nos estudos de triagem virtual;
- Estudos de triagem virtual baseado no ligante (similaridade por forma e eletrostática) para a investigação de compostos com propriedades inibidoras à PMI;
- Estudos do perfil farmacocinético e toxicológico dos compostos identificados, buscando selecionar os que apresentam um perfil mais adequado a novos potenciais defensivos agrícolas e que não cause interferência em humanos;
- Realizar ensaios biológicos junto aos compostos selecionados e otimizados, buscando verificar os níveis de atividade dos mesmos, tentando racionalizar os resultados por meio das relações estrutura-atividade assim observados.

3. MATERIAIS E MÉTODOS

Para a pesquisa de novos inibidores de PMI, foi usada a técnica de triagem virtual por forma e similaridade, além de simulações de *docking* e análises de propriedades farmacocinéticas e toxicológicas (ADME/Tox) e seus modos de interações foram analisados através de uma inspeção visual, obtendo por fim, os compostos mais promissores. Essas moléculas selecionadas foram adquiridas para serem testadas em ensaios de atividade *in vitro*.

Como não existem estruturas depositadas no PDB (*Protein Data Bank*) para a PMI de XAC, um modelo tridimensional foi desenvolvido utilizando o método de “modelagem molecular por homologia estrutural”.

3.1. Construção do modelo de fosfomanose isomerase para *Xanthomonas citri* subsp. *citri*

O primeiro passo para esse estudo foi a construção de um modelo da proteína fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*, utilizando a técnica de modelagem por homologia estrutural, já que tal estrutura ainda não existe nos bancos de dados de estruturas de proteínas como o PDB. Os estudos de homologia estrutural são baseados em três passos: (1) alinhamento (a etapa crucial do processo), (2) a construção do modelo e (3) a validação do modelo.

3.1.1 Pesquisa e Alinhamento de sequências homólogas

Uma busca por estruturas com sequências homólogas à de interesse (PMI de XAC), contidas no PDB, foi realizada previamente utilizando o algoritmo **BLASTp**, um algoritmo pertencente à família **BLAST** disponibilizado pelo *National Center for Biotechnology* e que realiza a comparação entre sequências proteicas (ALTSCHUL et al., 1990).

Quatro proteínas foram selecionadas para a geração do modelo, sendo uma com identidade sequencial de 29% e outras três com identidade sequencial de 38%. A estrutura de menor identidade (PDB ID: 1H5R) foi utilizada devido ao fato de possuir um ligante em seu sítio ativo e, dessa forma, ser possível observar os dobramentos ocasionados pela formação do complexo ligante receptor.

3.1.2 Alinhamento das sequências

Após a obtenção das sequências do PDB, foi realizado o alinhamento dessas sequências, para o qual foi utilizado o servidor *online* **Clustal Omega** (SIEVERS et al., 2011).

Esse servidor é disponibilizado pela European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBO).

O **Clustal Omega** tem cinco etapas principais para gerar o alinhamento de múltiplas sequências. O primeiro passo consiste em produzir um alinhamento aos pares (*pairwise*) para, em seguida, serem agrupados (BLACKSHIELDS et al., 2010). A terceira etapa é a construção de uma árvore guia, em que os dois grupos mais próximos são combinados e repetidos até que a árvore final possa ser avaliada. Então, na etapa final, o alinhamento de múltiplas sequências é, de fato, produzido (EDDY, 2011).

Um dado importante analisado durante o alinhamento das sequências diz respeito a quais alterações de resíduos de aminoácidos essas proteínas sofreram ao longo da Evolução, estimando que a função deveria ser preservada nos trechos onde a identidade (ou mesmo a similaridade) sequencial deveria ser maior.

3.1.3 Construção e avaliação da qualidade do modelo

O programa **MODELLER** foi utilizado para a construção do modelo da PMI e trata-se de um algoritmo automatizado utilizado para a modelagem de estrutura proteica através de comparação com estruturas tridimensionais resolvidas. Esse programa utiliza pelo menos uma estrutura de entrada como referência (ou molde) e, a partir de outros alinhamentos fornecidos, gera um modelo tridimensional homólogo (FISER; ŠALI, 2003).

Após a construção do modelo tridimensional, o próximo passo consiste na avaliação da qualidade do modelo gerado. Para a etapa de avaliação foram utilizados os servidores *online* **WHATIF** (VRIEND, 1990) para a validação dos contatos atômicos e **SAVE v5.0**, que contém os programas **PROCHECK**, **VERIFY3D**, **ERRAT** e **PROVE**.

O servidor *online* **WHATIF** analisa contatos atômicos, entre maus contatos e contatos favoráveis (como interações intermoleculares), gerando pontuações para cada resíduo. O **PROCHECK** (LASKOWSKI et al., 1993) permite uma comparação entre os resíduos do modelo e estruturas de proteínas de alta-resolução cristalográfica. Através desse programa foram gerados: o Diagrama de Ramachandran, os gráficos de planaridade das ligações peptídicas, de distorção tetraédrica dos carbonos alfa, de energias das ligações de hidrogênio, de medidas das interações não-ligantes (van der Waals, ligações de hidrogênio, interações eletrostáticas) e General Factor (G-Factor), o qual, este último, consiste em uma média da estereoquímica geral do modelo.

O **VERIFY 3D** (LUTHY; BOWIE; EISENBERG, 1992) analisa a compatibilidade entre o modelo 3D e a sua sequência de aminoácido, baseado na localização (alfa, beta, loop) e uma sequência de ambientes químicos (polar, apolar), utilizando uma matriz proposta por Bowie e colaboradores (BOWIE; LTCY; EISENBERG, 1990). O **ERRAT** (COLOVOS; YEATES, 1993) analisa as interações não covalentes entre átomos diferentes no modelo em comparação com estruturas altamente refinadas. Por fim, o **PROVE** (PONTIUS; RICHELLE; WODAK, 1996) utiliza um algoritmo que calcula o volume dos átomos presentes nas macromoléculas, assumindo-os como esferas rígidas; e gera um desvio estatístico denominado Z-score. Esse dado estatístico provém da comparação do modelo com estruturas refinadas de alta resolução depositadas no PDB.

3.1.4. Validação do modelo de PMI gerado para simulações de *docking*

Após a criação do modelo, foi realizado um experimento para a validação do mesmo, utilizando os programas **GOLD** (SUITE, 2015) e **GLIDE** (FRIESNER et al., 2004). Definiu-se, a partir do substrato da estrutura cristalográfica original (1H5R), a melhor região do sítio ativo na qual seriam realizados os cálculos. Em seguida foram feitas simulações de *docking* com o substrato da 1H5R para definir qual seria o melhor programa para os próximos estudos. Escolhido o programa, experimentos de *docking* foram realizados com a estrutura de PMI, utilizando inibidores reportados no *web servidor* **BindingDB** (CHEN; LIU; GILSON, 2001), um repositório de estruturas e dados de atividade focado na interação entre proteínas e moléculas com potenciais atividades terapêuticas.

3.2 Triagem virtual de novos inibidores da fosfomanose isomerase

Acessando o *web servidor* **BindingDB**, um total de 959 hits são descritos como inibidores de PMI, sendo que o mais potente possui $IC_{50} = 4 \mu M$. Sendo assim, torna-se desejável a descoberta e o desenvolvimento de novas classes de compostos e que possam inibir mais efetivamente a enzima aqui investigada. O objetivo é realizar experimentos de triagem virtual (por *docking* e similaridade) com a PMI em diferentes bases de dados de compostos disponíveis comercialmente. As bases de dados de compostos incluem: *ZINC* (subcoleções *Natural Stock*, *Drug-database* e *Drug-like*), *BindingDB* subdivisão *Drugs FDA*, ChemBridge (subcoleções *DIVERSetTM-EXPEXPRESS-PickTM Collection(DIVERSetTM-EXP)* e *DIVERSet Core Library (DIVERSetTM-CL)*), *Maybridge* subdivisão *Screening Collection*,

Princeton e *IBS* (subdivisão *Natural* e *Synthetic*) e na tabela abaixo (Tabela 1) é mostrado o número de compostos presentes em cada base de dados utilizada.

Tabela 1 - Relação do número de compostos presentes em cada base de dados utilizada.

Base de dados	Número de compostos	Referências
<i>ZINC</i> subdivisão <i>Drug-database</i>	2924	(IRWIN; SHOICHET, 2005)
<i>BindingDB</i> subdivisão <i>Drugs</i> <i>FDA</i>	6144	(CHEN; LIU; GILSON, 2001)
<i>Chembridge</i> subdivisão <i>DIVERSet-EXP</i>	50000	(CHEMBRIDGE, [s.d.])
<i>Maybridge</i> subdivisão <i>Screening Collection</i>	53000	(MAYBRIDGE, 2004)
<i>Chembridge</i> subdivisão <i>DIVERSet CL</i>	60000	(CHEMBRIDGE, [s.d.])
<i>IBS</i> subdivisão <i>Natural</i>	68497	
<i>IBS</i> subdivisão <i>Synthetic</i>	483510	
<i>ZINC</i> subdivisão <i>Natural Stock</i> <i>Princeton</i>	450000 1300000	(IRWIN; SHOICHET, 2005)
<i>ZINC</i> subdivisão <i>Drug-Like</i>	17900742	
TOTAL	20374817	

Fonte: o autor

Algumas dessas bases contêm os compostos em formato *spatial data format* (sdf), os quais já foram reescritos em formato mais apropriado, mol2, contendo cargas atômicas, tipos de átomos, ordens de ligação e estruturas tridimensionais bem definidas e mais adequadas para as simulações de triagem virtual e propósitos gerais de modelagem molecular. Também foram geradas novas bases/bibliotecas de confôrmeros, a partir dessas, utilizando o *programa OMEGA* (BOSTRÖM; GREENWOOD; GOTTFRIES, 2003).

3.2.1. Geração de confôrmeros (OMEGA)

Para a geração de confôrmeros, o programa **OMEGA** da empresa **OpenEye** foi utilizado. Esse programa realiza a geração de confôrmeros em três dimensões com baixo

desprendimento computacional e um alto efeito no alcance da conformação bioativa entre os confôrmeros gerados.

No presente estudo, os parâmetros utilizados no **OMEGA** para a geração de confôrmeros foram: 300/molécula, com uma energia de tensão de até 9 kcal/mol e um desvio médio quadrático de 0.6 Å, segundo protocolo otimizado por Carlos H.T.P. Silva (DA SILVA; TAFT, 2017).

3.2.2. Seleção de inibidores baseados em ligantes (**ROCS e EON**)

Após a preparação das bases de dados, um segundo tipo de triagem virtual foi ainda realizado, o qual é baseado em diferentes tipos de similaridade: de forma e eletrostática. Ambos os programas utilizados nessa etapa, **ROCS** e **EON**, da **OpenEye**, tem revelado compostos promissores para o tratamento de várias doenças (NAYLOR et al., 2009).

O programa **ROCS** realiza uma triagem virtual por forma, através da comparação entre os volumes e conformação de dois compostos, a molécula de referência e as moléculas presentes em bancos de dados; e o programa **EON** realiza a comparação e classificação dos compostos bases em seus mapas de superfície eletrostática. Os parâmetros adotados para ambos programas não foram alterados.

3.2.3. Seleção de inibidores baseados em estrutura (**GOLD**)

Os programas de *docking* analisam a afinidade ligante-proteína e para essa etapa do estudo foi utilizado o programa **GOLD**.

A localização do centróide utilizado para a PMI de *XAC* foi $x = 23.977$, $y = 15.659$ e $z = 43.886$ com uma esfera de 8 Å. O tamanho dessa esfera foi modificado para algumas bases de dados devido à problemas encontrados para alguns ligantes e, nesse caso, foi utilizado uma esfera de 6 Å e 60 de fitness. Para os estudos executados na PMI de *Homo sapiens*, para a validação do modelo gerado, o centróide utilizado foi $x = 20.21$, $y = 16.63$ e $z = 10.00$ e com uma esfera de 7 Å.

3.2.4 Predição das propriedades farmacocinéticas e toxicológicas (**QikProp e DEREK**)

Após os experimentos de triagem virtual realizados, somente as moléculas que passaram pelo crivo de propriedades adequadas a um potencial defensivo agrícola avançaram para a próxima etapa, avaliando-se, principalmente, a potencial toxicidade dos compostos frente ao

organismo humano. As moléculas foram avaliadas com respeito às suas propriedades ADME/Tox, utilizando-se os programas **QikProp** (SCHRODINGER, [s.d.]) (da empresa Schroedinger) e **DEREK** (MARCHANT; BRIGGS; LONG, 2008) (da empresa Lhasa).

Os compostos sobreviventes à predição de propriedades toxicológicas, foram submetidos à análises farmacocinéticas realizadas pelo **QikProp**, os descritores analisados foram: log CACO2, log MDCK, porcentagem de absorção oral e qualidade do modelo de absorção oral humana. Foram rejeitados todos os compostos que tiveram valores de log CACO2 e log MDCK maiores ou iguais a 500, potencial de absorção oral maior ou igual a 80% e qualidade do modelo de absorção humana igual a alta.

Na análise toxicológica realizada pelo **DEREK**, os compostos rejeitados foram aqueles que apresentaram algum alerta de toxicidade segundo *Custom Prediction* e *Lhasa Prediction*.

3.3. Visualização Gráfica

Os programas utilizados para a visualização gráfica foram o **Discovery Studio** (DASSAULT SYSTÈMES BIOVIA, 2015) da empresa *Biovia* bem como o **Maestro** (SCHRODINGER, 2018), da empresa *Schroedinger*.

O **Discovery Studio** foi empregado na visualização da sobreposição dos modelos obtidos pela modelagem de proteínas por homologia estrutural. Já o **Maestro** foi utilizado na análise visual dos resultados obtidos a partir dos estudos de simulações de *docking* e, conseqüentemente, na seleção das moléculas finais.

3.4. Ensaio de atividade *in vitro* com a enzima PMI de *Xanthomonas citri* subsp. *citri*.

A enzima recombinante de XAC foi produzida por expressão heteróloga em sistema bacteriano e foi cedida por Alexandrino *et al* (dados não publicados) (LBBMA/ UFSCar- São Carlos).

Durante o ensaio foi utilizado um sal dissódico hidratado de D-manose-6-fosfato (M6P) da Sigma (cod. M6876-10) com pureza maior ou igual a 97% e um sal dissódico hidratado de D-frutose-6-fosfato (F6P) da Sigma (cod. F3627-100) com pureza maior ou igual a 98%. Ambos sais apresentam uma massa molar (MM) de 304,10 g/mol e foram diluídos em água, de modo a obter uma concentração final de 0,3 M.

Uma solução estoque (em água) de 0,1 M do inibidor **DL03874928** (MM = 286,28 g/mol) e diluições subseqüentes a partir dessa solução-estoque foram feitas de modo a obter concentrações de 1 mM, 100 µM e 10 µM.

O reagente de Seliwanoff utilizado para a detecção colorimétrica foi feito no mesmo dia dos ensaios e consistiu em 0,025 g de resorcinol, e 50 mL de ácido clorídrico concentrado para um volume final de 50 mL.

3.4.1 Método de avaliação da atividade enzimática

Como dito anteriormente, a fosfomanose isomerase é uma enzima que atua na interconversão da frutose-6-fosfato (F6P) e da manose-6-fosfato (M6P). Com isso, a atividade enzimática foi realizada por meio de um teste de detecção de frutose baseado em um método colorimétrico (SHAHIDULLAH, M; KHORASANI, 1972) com o reagente de Seliwanoff.

O procedimento consistiu na adição de 5 µL de uma solução contendo 0,3 M de M6P (0,456 mg) em 95 µL de uma solução da enzima recombinante PMI de XAC em tampão (50 mM Tris-HCl pH 8,0, 10 mM NaCl) para um volume final de 200 µL. Após incubação da mistura por 1h a 30°C foram adicionados 3 mL do reagente de Seliwanoff e a mistura foi mantida a 100°C por 5 minutos.

Os ensaios foram feitos como um teste exploratório a 25°C e a leitura de absorbância foi realizada em comprimento de onda de 415 nm no leitor de microplacas *iMark Microplate Absorbance Reader* (Bio-Rad). O experimento definitivo já está em andamento.

3.4.2 Avaliação *in vitro* da inibição da fosfomanose isomerase pelos compostos selecionados

A avaliação da inibição da PMI pelo composto **DL03874928**, um dos inibidores mais promissores pois apresenta o maior número de interações com os aminoácidos do modelo gerado, foi realizada com a mesma metodologia do passo anterior.

Em cada um dos tubos foram adicionados 95 µL de enzima e 100 µL do inibidor nas concentrações de 10 µM, 100 µM e 0,1 M, de modo a obter as concentrações finais de 5 µM, 50 µM e 50 mM (Tabela 2). Essa solução foi incubada à temperatura ambiente por 10 minutos. Em seguida, foram adicionados aos tubos 5 µL da solução M6P 0,3 M; sendo que essa solução formada por inibidor, enzima e substrato foi incubada por 1h à 30°C.

Passado o tempo de incubação, cada ensaio recebeu 3 mL de Reagente de Seliwanoff e foi levado à fervura por 5 minutos. As absorbâncias das soluções foram medidas com a ajuda do leitor de microplacas *iMark Microplate Absorbance Reader* (Bio-Rad) sob comprimento de onda de 415 nm.

Tabela 2 - Dados referentes aos ensaios *in vitro* realizados para a fosfomanose isomerase recombinante de *Xanthomonas citri* subsp. *citri*.

REAGENTES	Volume (μL)					
	Controle	Controle	Reação com M6P e Concentrações finais do inibidor			
	com F6P	com M6P	sem inibidor	5 μM	50 μM	50 mM
M6P 0,3M	0	5	5	5	5	5
F6P 0,3M	5	0	0	0	0	0
Enzima	0	0	95	95	95	95
Inibidor 10 μM	0	0	0	100	0	0
Inibidor 100 μM	0	0	0	0	100	0
Inibidor 0,1 M	0	0	0	0	0	100
Tampão (50 mM Tris-HCl pH 8,0, 10 mM NaCl)	195	195	100	0	0	0

4. RESULTADOS E DISCUSSÃO

A seguir serão apresentadas e discutidas as etapas para construção e validação dos modelos de fosfomanose isomerase para *Xanthomonas citri* subsp. *citri*, e *Homo sapiens*. Também serão apresentados os resultados da triagem virtual para os inibidores de XAC e avaliação *in vitro* para um dos compostos selecionados.

4.1 Construção do modelo da fosfomanose isomerase (PMI) para *Xanthomonas citri* subsp. *citri* (XAC)

A criação de um modelo tridimensional para uma proteína através da homologia estrutural consiste na comparação entre duas ou mais sequências de proteínas similares. Para que esse método seja aplicado é necessário que pelo menos três situações ocorram: (a) uma estrutura proteica seja experimentalmente resolvida; (b) compartilhar um valor de identidade sequencial acima de 25% com a proteína desejada; e (c) correto alinhamento entre as sequências envolvidas (FOMBY; CHERLIN, 2011; MARTÍ-RENOM et al., 2000; SCHWEDE et al., 2003).

A criação de um modelo é possível devido à relativa conservação que ocorre no enovelamento de uma determinada família de proteínas ao longo da evolução. Assim, caso a estrutura primária de duas proteínas sejam similares, o modelo tridimensional terá grandes chances de ser similar também. A criação de um modelo por homologia estrutural é constituída por: pesquisa por sequências homólogas, alinhamento das sequências, construção e avaliação do modelo (MARTÍ-RENOM et al., 2000; SANCHEZ; SAIL, [s.d.]).

Nos próximos tópicos são apresentadas as 4 etapas de predição de modelagem de proteínas por homologia estrutural aplicadas para construção do modelo de PMI de XAC.

4.1.1 Etapa de identificação da sequência de aprovação

A etapa inicial do processo de predição de estruturas tridimensionais de proteínas é a identificação de estruturas de proteínas homologadas por sequências em busca que possuam dados cristalográficos, ou seja, sua estrutura resolvida e relatada em um banco de dados de estruturas cristalográficas, como o PDB. Para maior confiabilidade do modelo gerado, é desejável que as identidades sequenciais estejam acima de 25% em relação à sequência de proteínas em estudo (RODRIGUEZ; VRIEND, 1997).

Para a construção do modelo da PMI inicialmente obtivemos o sequenciamento de aminoácidos da mesma através do banco de dados GenBank (AAM38423.1) (“NCBI-National

Center for Biotechnology Information”, 2018). Em seguida, utilizando o algoritmo do servidor online **BLASTp** foi possível identificarmos proteínas com estruturas cristalinas disponíveis no PDB e que detêm identidades sequenciais acima de 25% compartilhadas com a PMI de interesse.

O algoritmo **BLAST** (*Basic Local Alignment Search Tool*) disponibilizado pelo NCBI (*National Center for Biotechnology*) detecta similaridades entre sequências nucleotídicas ou de proteínas. Os programas da família **BLAST** foram feitos para comparar: a) sequências de consultas de DNA; b) proteínas com bancos de dados de DNA; e c) proteínas em qualquer combinação (ALTSCHUL et al., 1990).

Uma das variantes dos algoritmos da família **BLAST** é o **BLASTp**, utilizado para comparações entre sequências de proteínas, através de uma busca por 3 aminoácidos semelhantes (*hit*) da sequência-modelo em outras sequências depositadas em diversos bancos de dados. Os alinhamentos desses *hits* recebem uma pontuação chamada de *score* e, depois, a análise é expandida nos dois lados para o restante da sequência-modelo com o intuito de se obter um maior *score*.

Uma outra pontuação analisada pelo **BLAST** é o *e-value*, com o intuito de determinar a aleatoriedade com que uma sequência homóloga foi obtida. Essa pontuação indica a probabilidade de se encontrar um alinhamento em outra sequência, aleatoriamente. Por isso, quanto menor o valor de *e-value*, mais confiável é o resultado obtido.

Por fim, um outro resultado importante fornecido pelo **BLAST** é a porcentagem de identidade sequencial, a qual é responsável por mostrar o grau de correspondência entre duas subsequências. Nesse caso, valores entre 18-25% indicam uma potencial semelhança estrutural, enquanto valores iguais ou maiores a 25% indicam adicionalmente uma potencial semelhança de função

Portanto, neste trabalho, o algoritmo **BLASTp** ajudou a detectar 3 estruturas homólogas à PMI (PDB ID 2CU2, 2X5S e 2QH5), com percentual de identidade sequencial de 36% e 38%, além de altas resoluções cristalográficas (de 1,9 a 2,35 Å), tidas como adequadas para a modelagem por homologia (ou comparativa). Foi detectada também uma quarta estrutura, sendo essa a única contendo ligante (o substrato) no sítio ativo (PDB ID 1H5R) e que apresenta 29% de identidade (Figura 5) (Tabela 3).

Figura 5 - Resultado obtido pelo **BLASTp** na procura de proteínas similares a fosfomanose isomerase de *Xanthomonas citri* subsp. *citri* (PDB ID: 2X5S, 2QH5, 2CU2 e 1H5R)

Description	Max score	Total score	Query cover	E value	Ident	Accession
Chain A, Crystal Structure Of Mannose-6-phosphate Isomerase From Helicobacter Pylori	207	207	97%	3e-65	40%	2QH5_A
Chain A, Crystal Structure Of T. Maritima Gdp-Mannose Pyrophosphorylase In Apo State	177	177	97%	1e-53	40%	2X5S_A
Chain A, Crystal Structure Of Mannose-1-Phosphate Geranyltransferase From Thermus Thermophilus Hb8	171	171	97%	6e-51	38%	2CU2_A
Chain A, Crystal structure of glucose 1-phosphate thymidyltransferase from Aneurimibacillus thermoaerophilus	41.2	41.2	67%	8e-04	29%	4H00_A
Chain A, Thymidyltransferase Complexed With Tmp	38.1	38.1	69%	0.007	29%	1H5S_A
Chain D, Thymidyltransferase Complexed With Tmp	38.1	38.1	67%	0.007	29%	1H5S_D
Chain A, Crystal Structure Of Glucose-1-phosphate Thymidyltransferase RmlA Complex With Dtdp	37.7	37.7	67%	0.009	25%	1LVW_A
Chain A, Thymidyltransferase Complexed With Thymidylidiphosphate- Glucose	37.0	37.0	67%	0.018	29%	1H5T_A
Chain B, Thymidyltransferase Complexed With Thimidine And Glucose- 1-Phosphate	36.6	36.6	67%	0.019	29%	1H5R_B
Chain A, Crystal Structure Of Glucose-1-phosphate Thymidyltransferase From Burkholderia Vietnamensis In Complex With 2'-deoxyuridine-5'- Monophosphate And 2'-deoxy-thy	35.8	35.8	11%	0.045	58%	5IFY_A
Chain A, Mycobacterium tuberculosis RmlA in complex with dTDP-glucose	35.4	35.4	94%	0.055	26%	6B5E_A
Chain A, L89T VARIANT OF S. ENTERICA RmlA	34.7	34.7	67%	0.099	28%	1MP3_A

Fonte: (“NCBI-National Center for Biotechnology Information”, 2018)

As estruturas selecionadas (PDB ID 2QH5, 2X5S, 2CU2 e 1H5R) são as seguintes enzimas: manose-6-fosfato isomerase, GDP-manose pirofosforilase, glicose-1-fosfato timidiltransferase e uma timidililtransferase.

A manose-6-fosfato isomerase e a GDP-manose pirofosforilase realizam ambas atividades desempenhadas pela PMI de *XAC*, isto é, a isomerização entre manose-6-fosfato e frutose-6-fosfato e a transferência de manose-1-fosfato (M1P) para a guanosina trifosfato (GTP). Por fim, a glicose-1-fosfato timidiltransferase e a timidililtransferase realizam a transferência de timidila produzindo, por exemplo, a DTDP-glicose.

Devido ao fato de o modelo a ser gerado ser de interesse para os estudos de descoberta de novas moléculas baseados em estruturas, a escolha da estrutura cristalográfica para sua construção levou em consideração a presença de um ligante complexado no sítio da proteína. Assim, a escolha da proteína homóloga, usada para prever uma estrutura PMI *XAC*, foi a cadeia B da proteína com código PDB 1H5R (*Escherichia coli*, cepa K12 timidililtransferase) (Tabela 3). Como todas as estruturas apresentam uma identidade acima de 25%, elas compartilham com a estrutura de interesse uma potencial semelhança estrutural e de função.

Tabela 3 - Resultados de identidade e *e-value* de sequências de proteínas obtidas no servidor online **BLAST** usando o algoritmo para proteínas (**BLASTp**) na triagem de proteínas homologadas a fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*.

PDB ID	Identidade	E-value
2QH5_A	38%	7×10^{-63}
2X5S_A	38%	5×10^{-64}
2CU2_A	38%	2×10^{-59}
1H5R_B	29%	0,033

2X5S	EKPDLETAKKFVESGRFLWNSGMFLWKAREFIEEVKVCEPSIYENLKDV-----D	219
2QH5	EKPSLDKAIEFQKSGGFYFNDSGMFVFQAGVFLDELKKHAPTILKGCERAFESLENAYFFE	227
XANTHO	EKPDLATAQSYLASGEYYWNSGMFLFRASRYLEELRKFHPAIDACQKAWENGKRDADF-	231
	*: .: . : : : : . : * : :	
1H5R	-----TDINRIYLEQGRLSVAMMGRGYAWLDTGTHQSLIEASNFIATIE	244
2CU2	----ASLEEVYAGLPKISIDYG-VMEKAERVRVVLGRFPWDDVGNWRALERVFSQDPHEN	273
2X5S	PRNFEELKKAYEKVPSISVDYA-VMEKSKKVRVVKADFEWSDLGNWSSVREIEGYTEES-	277
2QH5	KKIARLSEKSMQDLEDMSIDIA-LMQQSHKIKMVELNAKWSDLGNFNALFEEAANEPKEN	286
XANTHO	---TRLDKDAFAASPSDSIDYA-VMEKTADAVVPLDAGWNDVGSWSSLLDVSNDQAQGN	287
	.: : : : : * * * . : :	
1H5R	ER--QGL--KVSCPEEIAF-RKGFIDVEQVRKLA-----VPLIKNNYGQYLYKMTKDS	292
2CU2	VVLGEGRHVALDTFGCVVYADRGVVATLGVSGLVVAKVGDEVLLVVPKDWAREVREVVKRL	333
2X5S	---DEVIL--VSDRVFVKTHNKPIAVVGLSDVIVIDTPNGILICKEEYAKVREVVKKL	332
2QH5	VSLNQTTPV-----FAK-ESEEGHHHHHH-----	308
XANTHO	AHHGDVIQ--LDCQNTYAY-GSRLIAMVGLEDVVVETPDVAVLVGHRDRIQEVKDVVSQI	344
	: .	
1H5R	N-----	293
2CU2	EAQE-----	337
2X5S	FRTS-----	336
2QH5	-----	308
XANTHO	KTAGRSEATWHRKVYRPGWYDSDIDMGQRHQVKRITVKPGAVLSLQMHHHRAEHWIVVSG	404
	-----	293
	-----	337
	-----	336
	-----	308
	TAEVTRGEEVLLLLENQSTYIPLGVTHRLRNPGKLPLELIEVQSGSYLGEDDIVRFEDTY	464
1H5R	---	293
2CU2	---	337
2X5S	---	336
2QH5	---	308
XANTHO	GRA	467

A simbologia de asterisco (*) representa a presença de um único resíduo totalmente conservado, dois pontos (:) indica a conservação entre grupos de propriedades fortemente semelhantes e um ponto (.) indica a conservação entre grupos de propriedades fracamente semelhantes.

Fonte: o autor

Essa sobreposição estrutural nos permitiu visualizar os consensos de todas as estruturas secundárias das proteínas de referência, entre hélices-alfa e fitas de folha-beta. Devido a isso, pudemos analisar se as inserções ou deleções sugeridas pelo **Clustal Omega** estavam de acordo com o esperado, estruturalmente falando. Após a análise dos alinhamentos, juntamente com as sobreposições das estruturas de referência, obteve-se o alinhamento final.

Os aminoácidos com coloração azul escuro pertencem a folha β enquanto que aqueles com coloração verde pertencem às α -hélices. Os aminoácidos que tiveram seus rotâmeros modificados estão realçados em azul. A simbologia de asterisco () representa a presença de um único resíduo totalmente conservado, dois pontos (:) indica a conservação entre grupos de propriedades fortemente semelhantes e um ponto (.) indica a conservação entre grupos de propriedades fracamente semelhantes.*

Fonte: o autor.

A parte final da sequência da XAC (a partir do aminoácido 345) foi descartada devido ao não alinhamento com as outras sequências selecionadas e tal passo é possível pois essa região não está associada ao sítio ativo do modelo a ser gerado. Além disso, o alinhamento apresenta valores de 58% de identidade e 33% de similaridade no sítio ativo.

4.1.3 Etapa de construção e avaliação do modelo

Após a conclusão da etapa de alinhamento, a próxima etapa a ser conduzida é a construção do próprio modelo. Tal passo foi possível através do programa **MODELLER**, que gerou 5 hipóteses de modelo.

Além da construção do modelo, **MODELLER** pode realizar tarefas auxiliares adicionais, incluindo atribuição de “voltas”, alinhamento de duas sequências de proteínas ou seus perfis (MARTÍ-RENOM et al., 2000), alinhamento múltiplo de sequências de proteínas e/ou estruturas (ESWAR et al., 2007), cálculo de árvores filogenéticas e modelagem de alças ou *loops* em estruturas de proteínas (FISER; ŠALI, 2003).

Para a construção de um modelo tridimensional, o **MODELLER** utiliza algumas restrições espaciais definidas entre as sequências escolhidas e o molde. Algumas dessas restrições são: distâncias e ângulos diedros, comprimento e ângulos de ligação e restrições criadas manualmente como, por exemplo, enovelamento da estrutura secundária, ligações cruzadas, entre outros.

Primeiramente, são analisadas as distâncias existentes entre os resíduos das sequências alinhadas, enquanto, que os comprimentos de ligações, ângulos de ligação e de diedro são provenientes do alinhamento com o modelo. Em seguida, utiliza-se uma função objetiva para determinar os comprimentos e ângulos de ligação entre átomos não-ligados. Após esses passos, surge um modelo por satisfação dessas restrições espaciais. Subsequente ao surgimento do modelo, pode ocorrer uma avaliação e modelagem das regiões de alças/*loops* através de um algoritmo, seguido pela otimização estrutural do modelo gerado (ESWAR et al., 2007; SANTOS FILHO; BICCA DE ALENCASTRO, 2003).

Em posse dos modelos gerados, o próximo passo é a etapa de avaliação para determinar o modelo mais confiável e adequado para novos estudos. Na etapa de avaliação, é aconselhável avaliar a presença de contatos atômicos bons ou ruins e também comparar os resíduos do modelo com base nos parâmetros observados nas estruturas proteicas cristalográficas de alta resolução. Esses estudos foram desenvolvidos com o servidor on-line **WHATIF** e o programa **PROCHECK**, respectivamente, e os principais resultados são exibidos na Tabela 4.

O servidor *online* **WHATIF** analisa contatos atômicos gerando pontuações para cada resíduo e, adicionalmente, proporciona um ambiente flexível para exibir, manipular e analisar pequenas moléculas, proteínas, ácidos nucleicos e suas interações, para validação e correção de estruturas protéicas.

O **PROCHECK** (LASKOWSKI et al., 1993) realiza uma validação estereoquímica através da comparação entre os resíduos do modelo com base em parâmetros observados em estruturas de proteínas de alta-resolução cristalográfica. Uma das medidas geradas por esse programa é o General Factor (G-Factor), o qual, este último, consiste em uma média da estereoquímica geral do modelo.

Dados os resultados obtidos pelo **WHATIF**, pode-se observar que a melhor hipótese avaliada foi a Hipótese 4, que obteve uma pontuação total - 0,758, sendo que o valor de referência para este estudo é - 0,813, que corresponde ao escore geral da estrutura cristalográfica (PDB ID 1H5R) usada como referência (Tabela 4).

Outro ponto de atenção são as regiões identificadas do Gráfico de Ramachandran. O Diagrama Ramachandran é dividido em quatro regiões: altamente favorável, favoravelmente permitido, generosamente permitido e proibido. Quando um aminoácido está na região proibida, isso significa que qualquer distância interatômica entre átomos não ligados é menor que a distância de Van der Waals correspondente a esses átomos. No caso da Hipótese 4, nenhum resíduo de aminoácido é encontrado na região proibida, favorecendo sua seleção como modelo (Tabela 4).

Um ponto importante a considerar é o fator G, uma pontuação baseada nos parâmetros estereoquímicos observados de uma determinada estrutura. No caso do modelo PMI, as hipóteses apresentaram um fator G de - 0,1, um valor que está dentro da região para uma proteína com uma resolução de 1,9 Å (referente ao 1H5R) (Tabela 4).

Tabela 4 - Apresentação dos resultados da Pontuação Geral (**WHATIF**), Regiões Proibidas (Ramachandran) e Fator G (**PROCHECK**) obtidos na etapa de validação dos modelos gerados.

Modelos	Pontuação Geral (WHATIF)	Regiões Proibidas (Ramachandran Plot)	G-Factor
<i>1H5R</i>	- 0.813	1	0
<i>Hipótese 1</i>	- 0.697	0	- 0.1
<i>Hipótese 2</i>	- 0.745	1	- 0.1
<i>Hipótese 3</i>	- 0.734	0	- 0.1
<i>Hipótese 4</i>	- 0.758	0	- 0.1
<i>Hipótese 5</i>	- 0.667	1	- 0.1

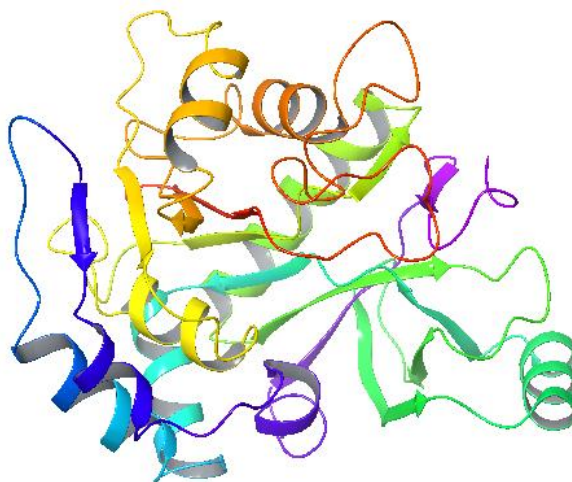
Fonte: o autor

Devem ser analisadas nesta fase do processo, as conformações dos resíduos de aminoácidos que fazem parte do local de interação em estudo. Essa inspeção visual tem como objetivo analisar e, se necessário, modificar a conformação de determinados resíduos através do uso de uma biblioteca de rotâmeros. A escolha do melhor rotâmero deve levar em consideração a direção dos resíduos da estrutura utilizada como referência para o desenvolvimento do modelo e a ausência de contato com os resíduos vizinhos.

Destacamos que, para o modelo PMI gerado (Figura 8) foi necessário modificar alguns resíduos, o que pode ser visto na Figura 7. Com essa mudança de rotâmeros, o valor de **WHATIF** para a hipótese 4 foi alterado para - 0.782, aproximando-se do valor da estrutura de referência (1H5R), além de estar entre - 1 e - 0,5 indicando que se trata de um bom modelo.

Na figura 8 é apresentado o modelo gerado para a PMI de XAC com o programa Modeller, em que é possível observar a presença das estruturas secundárias (alfa-hélices, folhas beta e alças) além de mostrar o seu enovelamento tridimensional.

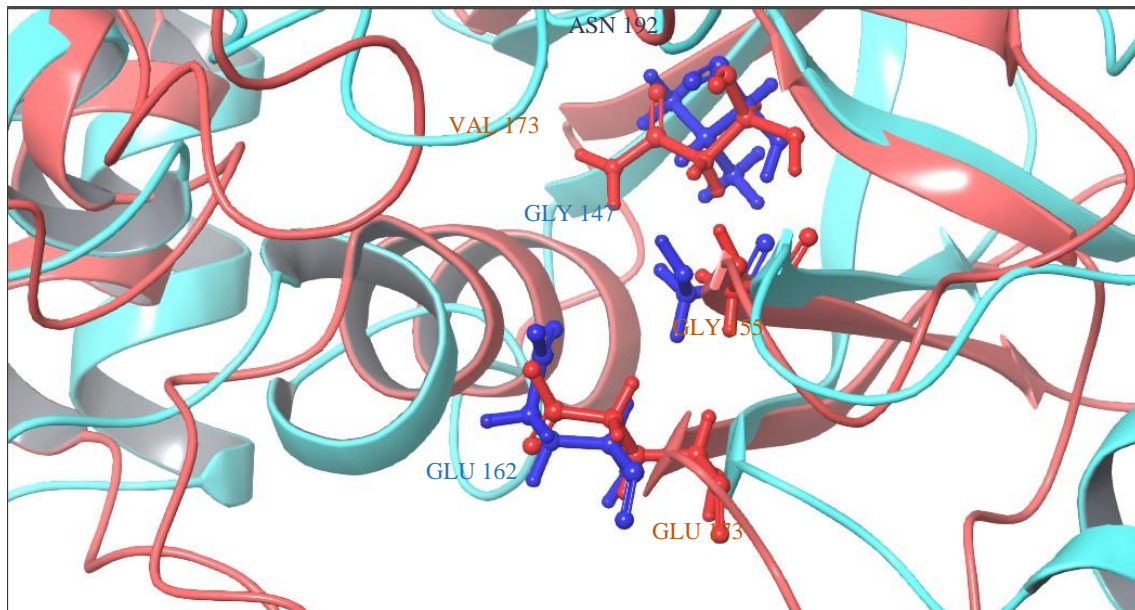
Figura 8 - Modelo gerado para a fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*. utilizando o programa **MODELLER**.



Fonte: o autor.

A sobreposição do modelo gerado (hipótese 4) com a estrutura referência (PDB ID 1H5R) revela ainda dados importantes que aumentam a confiabilidade do modelo obtido (Figura 9). Pode ser observado, na região do sítio ativo, que os resíduos de aminoácidos descritos por Zuccotti *et al.* (2001) como responsáveis pela formação do complexo ligante-PMI, Gln147, Glu162 e Val173, permanecem alocados na mesma região, quando comparado o modelo com a estrutura de referência de PDB ID 1H5R (Figura 9). Porém, no modelo criado baseado na sequência da XAC, o aminoácido valina é substituído por uma asparagina, mas ambos (asparagina do modelo e valina da estrutura referência) apresentam-se na mesma região e sobre a mesma conformação de cadeia principal (Figura 9).

Figura 9 - Modelo final da fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*, obtido por modelagem comparativa, com destaque para alguns aminoácidos presentes no sítio ativo.



Em cor azul, os resíduos pertencentes ao modelo que ficaram alocados na mesma região observada para a estrutura aqui usada como referência (PDB ID 1H5R), que se encontra na cor vermelha.

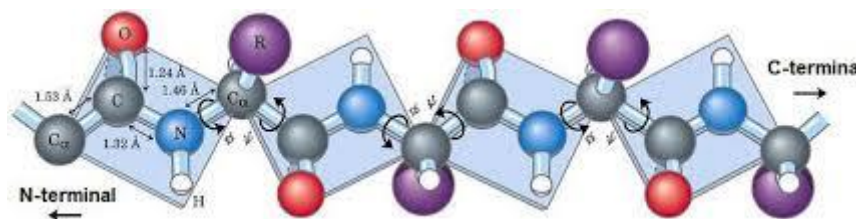
Fonte: o autor.

4.1.4 Avaliação dos ângulos diédricos Ψ versus ϕ através do Diagrama de Ramachandran

O Diagrama de Ramachandran, fornecido pelo **PROCHECK**, é uma ferramenta de análise importante pois indica erros na estrutura terciária dessas macromoléculas. O diagrama é dividido em regiões energeticamente favoráveis e desfavoráveis e mostra quais resíduos se encontram em tal regiões, permitindo uma avaliação de qualidade dos modelos teóricos ou experimentais (RAMACHANDRAN et al., 1963).

No gráfico é possível observar dois eixos contendo os ângulos da cadeia principal, Phi e Psi (Figura 10), com valores entre -180° e $+180^\circ$, compreendendo todas as combinações energéticas/conformações possíveis que os resíduos de aminoácidos podem adquirir de acordo com esses ângulos diedros (RAMACHANDRAN et al., 1963). Entretanto há dois aminoácidos que podem ser encontrados em regiões desfavoráveis, mas não causariam dano; são eles: glicina e prolina.

Figura 10 - Ângulos diedros de uma proteína evidenciando os ângulos Phi e Psi.

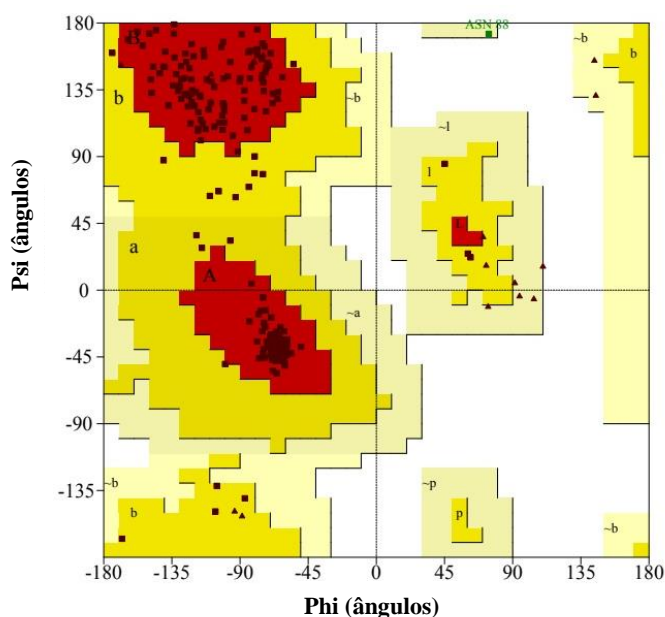


Fonte: Sancho, n.d.

A glicina possui apenas um hidrogênio como cadeia lateral, sendo menos estericamente impedido em comparação com outros aminoácidos. Este fato permite com que a cadeia polipeptídica dos resíduos de glicina assumam conformações que são ‘proibidas’ para outros resíduos. Assim, a glicina ocorre frequentemente em regiões de proteínas onde qualquer outro resíduo seria estericamente impedido. Por outro lado, a prolina é um iminoácido e possui uma cadeia lateral cíclica, tornando sua rotação restrita em torno da ligação. No caso da prolina, os valores de phi são restritos a valores em torno de -60° e, portanto, trata-se do resíduo de aminoácido mais conformacionalmente restrito.

O gráfico obtido para o modelo PMI gerado (Figura 11) não indicou ângulos phi e psi desfavoráveis para os resíduos presentes no modelo, especialmente na região do local de interação. O modelo apresentou 91,6% de seus resíduos em regiões altamente favoráveis e, embora o resíduo ASN88 apareça em uma região “generosamente permitida”, seus ângulos de torção foram “herdados” das estruturas cristalográficas utilizadas como referência.

Figura 11 - Diagrama de Ramachandram de *Xanthomonas citri* subsp. *citri* obtido pelo **PROCHECK**: 91,6% de resíduos em regiões favoráveis, 7,9% de resíduos em regiões permitidas, 0,4% de resíduos em regiões generosamente permitidas e 0% de resíduos em regiões proibidas.



Estatísticas de plotagem		
Resíduos em regiões mais favoráveis [A,B,L]	219	91.6%
Resíduos em regiões adicionais permitidas [a,b,l,p]	19	7.9%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	1	0.4%
Resíduos em regiões proibidas	0	0.0%
Número de resíduos não glicinas e não prolina	239	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	19	
Número de resíduos de prolina	17	
Número total de resíduos	277	

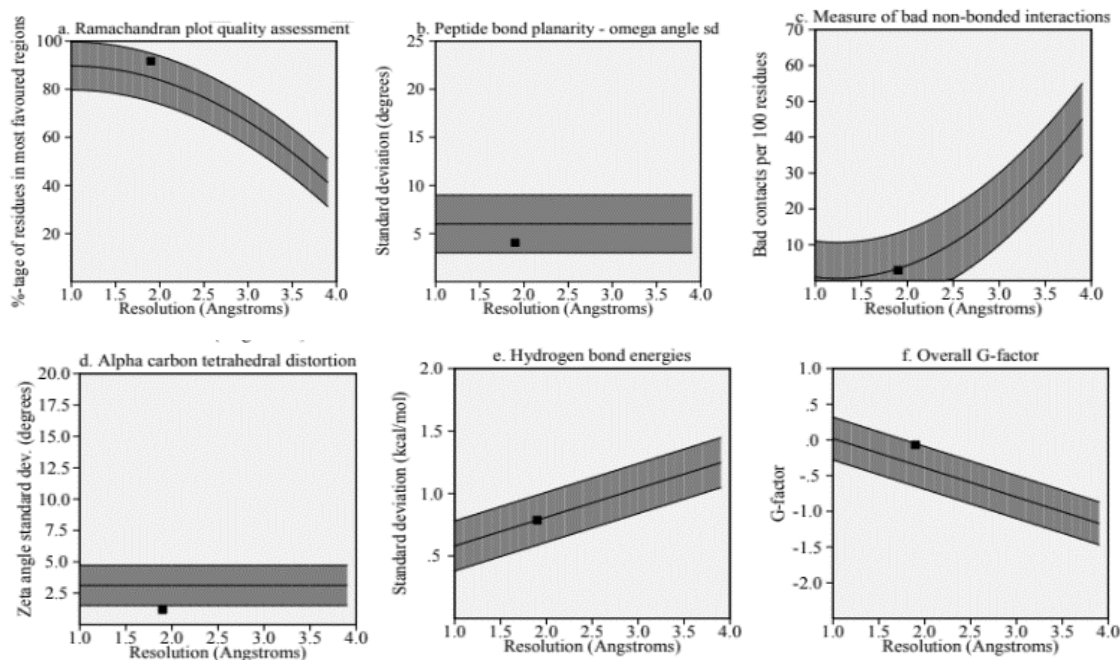
Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

Fonte: o autor.

4.1.5 Outros parâmetros importantes para a análise (Suite **SAVE v5.0**)

Outras ferramentas importantes para analisar um modelo de proteína predito podem ser encontradas no pacote **SAVE v5**. O uso do **PROCHECK** permite a análise da planaridade das ligações peptídicas, contatos atômicos fracos, distorção tetraédrica dos carbonos alfa, energia das ligações hidrogênio e fator G. Para o modelo PMI gerado, todos os valores estão dentro dos valores esperados para estruturas de proteína de resolução de 1,9 Å (referindo-se à estrutura PDB ID 1H5R, usada aqui como referência para modelagem) (Figura 12) e a distorção tetraédrica dos carbonos alfa apresentou um melhor valor que o cristalográfico.

Figura 12 - Resumo da avaliação estereoquímica do modelo selecionado construído por homologia estrutural para fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*., com um G-Factor de -0.1 e distorção tetraédrica do carbono alfa de -0.1.



Fonte: o autor.

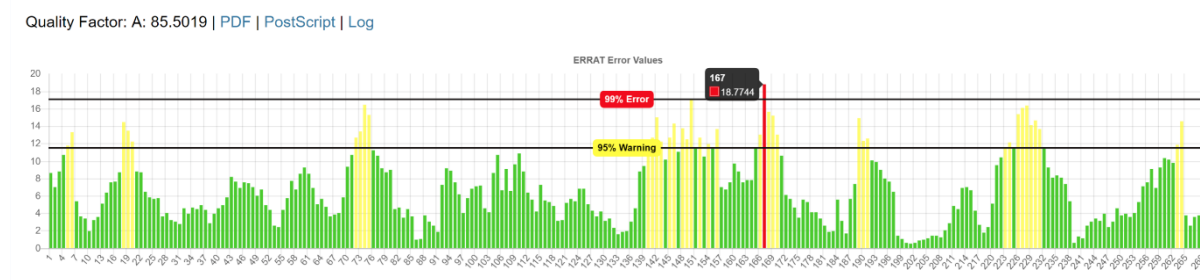
Os parâmetros apresentados juntos nas Figuras 11 e 12 ajudam a avaliar a qualidade estereoquímica geral do modelo selecionado. Os gráficos mostram que o modelo obteve valores dentro do intervalo de referência **PROCHECK** e, ao analisar os valores apresentados pelo **WHATIF** para os aminoácidos pertencentes ao local de interação do PMI, nenhum apresentou valores abaixo de -5, que, se existirem, devem ser revisados e reformados.

Além das ligações atômicas que ocorrem na cadeia peptídica, há também interações entre átomos não-ligados. A estrutura terciária de uma proteína é altamente dependente dessas interações e existem seis tipos de interações prováveis entre átomos não ligados e elas não ocorrem aleatoriamente. Assim, é possível através do programa **ERRAT** identificar regiões e fazer uma análise estatística dessas interações.

O **ERRAT** (COLOVOS; YEATES, 1993) analisa as interações intermoleculares entre átomos diferentes no modelo em comparação com estruturas altamente refinadas, gerando um gráfico de uma função de erro por uma janela formada por nove resíduos. De um modo geral, é um método mais eficaz que a **Perfil 3-D** proposto por Bowie e colaboradores (BOWIE; LTCY; EISENBERG, 1990), porém não é melhor que o **PROCHECK**.

Observando o gráfico desenvolvido pela **ERRAT** para o modelo PMI (Figura 13), é possível identificar três regiões distintas: uma verde, uma amarela e uma vermelha. A região verde mostra que todos os aminoácidos estão dentro dos valores esperados para estruturas altamente resolvidas e refinadas. A região amarela corresponde aos aminoácidos que excedem o resultado esperado e podem ser rejeitados com 95% de confiança. Espera-se que uma boa estrutura proteica tenha 5% de sua estrutura nessa região. Finalmente, as regiões vermelhas são aquelas que podem ser rejeitadas com 99% de confiança. O modelo gerado para o PMI possui poucas regiões amarelas e apenas um único aminoácido na região vermelha (Ala167). Nenhum aminoácido presente nessas regiões faz parte do site ativo; portanto, não haverá problemas para testes futuros. Além desses dados, o fator de qualidade, valor relacionado às interações atômicas não ligadas de 85.5019, determina uma estrutura de alta qualidade (resultado para valores acima de 50).

Figura 13 - ERRAT obtido para o modelo de fosfomanose isomerase gerado de *Xanthomonas citri* subsp. *citri* com um fator de qualidade de 85.5019.



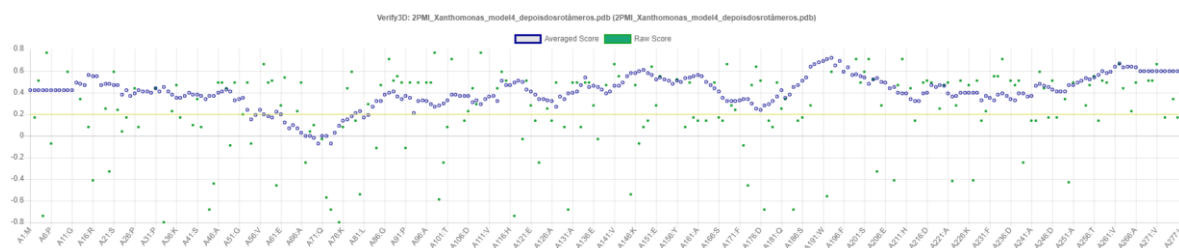
A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas. Fonte: o autor.

Outra análise interessante é verificar se um resíduo de aminoácido corresponde ao ambiente em que está, ou seja, se há compatibilidade do modelo com sua sequência de aminoácidos. Essa análise pode ser realizada usando o **Verify-3D**.

O **VERIFY 3D** (LUTHY; BOWIE; EISENBERG, 1992) analisa a compatibilidade entre o modelo 3D e a sua sequência de aminoácido, baseado na localização (alfa, beta, loop) e uma sequência de ambientes químicos (polar, apolar), utilizando uma matriz proposta por Bowie e colaboradores (BOWIE; LTCY; EISENBERG, 1990). Essa matriz permite uma análise de probabilidade de se encontrar os 20 aminoácidos em diferentes ambientes químicos, com base no tipo de aminoácido e natureza do ambiente químico. Em seguida, é realizado o alinhamento entre o modelo 3D e a sequência de aminoácidos; permitindo detectar as regiões mal resolvidas.

No estudo do PMI, pode-se observar que pelo menos 80% dos aminoácidos tiveram uma pontuação igual ou superior a 0,2 no perfil 3D-1D (Figura 14). Além disso, há uma região (Ala66 a Leu82) que caiu abaixo do limite, indicando uma provável dobra nessa região, no entanto, não é motivo de preocupação porque não está na região do site ativo do modelo.

Figura 14 - Verify-3D obtido para o modelo de fosfomanose isomerase gerado de *Xanthomonas citri* subsp. *citri*: 91.70% dos aminoácidos obtiveram uma pontuação igual ou superior a 0,2.

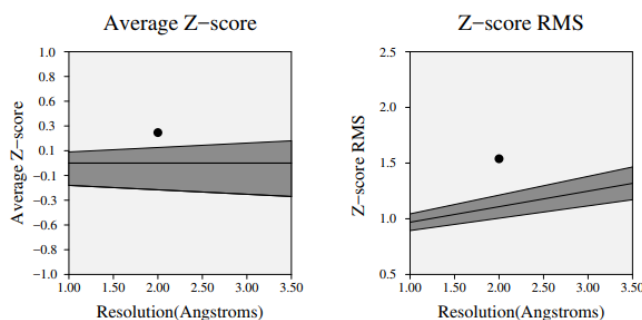


Fonte: o autor.

Finalmente, como última análise, é possível verificar o volume dos átomos. O programa **PROVE** correlaciona o volume obtido para cada aminoácido da proteína com os volumes de proteínas de alta resolução.

Como resultado, o **PROVE** fornece um valor de pontuação chamado Z-score que corresponde ao desvio estatístico programado de estruturas altamente resolvidas e refinadas. No caso do modelo PMI em estudo, tanto o Z-score quanto o Z-score RMS estavam acima da região de interesse; sendo que o Z-score possui uma média de 0,274 e o Z-score RMS com valor de 1,539 (Figura 15). Além disso, 4,6% dos resíduos analisados foram considerados outliers, indicando que mais atenção deve ser dada ao modelo. No entanto, juntamente com todas as outras análises realizadas, pode-se determinar que este é um modelo confiável para o PMI de *XAC*, e permite dar continuidade aos estudos.

Figura 15 - PROVE obtido para o modelo fosfomanose isomerase gerado de *Xanthomonas citri* subsp. *citri*.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: o autor

4.1.6 Etapa de validação do modelo de PMI XAC.

Após a criação do modelo, foi realizado uma simulação para a validação do mesmo, utilizando os programas **GOLD** e **GLIDE**. Definiu-se, a partir do substrato da estrutura cristalográfica original (1H5R), a melhor região do sítio ativo na qual seriam realizados os cálculos.

Os simulações de *docking* são cálculos que analisam a potencial afinidade entre o ligante e o sítio de ligação do receptor, propondo a pose (conformação + orientação), ou predição do modo de ligação com a maior afinidade entre ligante e receptor, estimada por uma pontuação calculada com base em uma função empírica associada ao método de mecânica molecular associado. A seguir serão descritos os métodos utilizados pelos programas **GOLD** (SUITE, 2015) e **GLIDE**.

O posicionamento dos ligantes no sítio do receptor resulta de um algoritmo adotado pelo programa escolhido para os simulações de *docking*. Os algoritmos mais usados são os genéticos, baseados em Monte Carlo ou busca sistemática. O **GOLD** utiliza o algoritmo genético enquanto o **GLIDE** utiliza um algoritmo de busca sistemática associada ao método de Monte Carlo.

No programa **GOLD**, o processo para gerar possibilidades conformacionais para a molécula ligante (e também para seletas cadeias laterais de resíduos de aminoácidos), se inicia gerando populações contendo conformações aleatórias de um dado ligante, em que cada

indivíduo dessa população é codificado e chamado de “cromossomo”. Os cromossomos possuem, então, as informações dos átomos dos ligantes como possibilidades de fazer interações. Além disso, eles também possuem informações sobre liberdades translacionais, rotacionais e conformacionais, chamadas de “genes”.

Com a ideia de gerar conformações mais “evoluídas”, os cromossomos utilizam mecanismos chamados operadores genéticos, no qual haverá uma análise de mutações e recombinações. Assim, diversas etapas como essa ocorrem até que seja obtida a população final (SUITE, 2015), evoluída de seus “ancestrais”, em processo tendenciado à maior afinidade com o sítio receptor.

Após a criação da população final, cada cromossomo/ligante é ancorado diversas vezes no sítio com diferentes orientações e cada pose é pontuada através de uma função chamada *fitness score*.

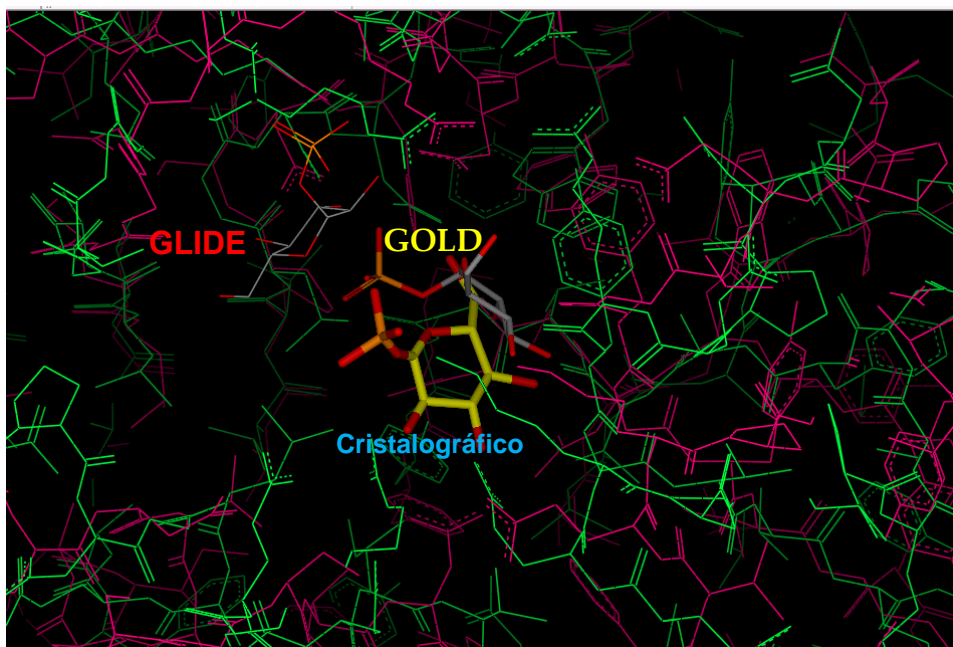
O primeiro passo do método utilizado pelo **GLIDE** consiste na construção de um *grid* no sítio de interação do receptor, determinando sua forma e propriedades. Nesse *grid* são feitas análises da presença de átomos existentes, fornecendo informações sobre os tipos de interações presentes e a pontuação para a pose do ligante. Em seguida, ocorre a geração de confôrmeros, os quais serão analisados nessa região.

A geração desses confôrmeros pelo programa **GLIDE** é baseado nas conformações de cada ligante, a qual varia conforme do número de ligações livres. A busca pelos melhores ligantes para um determinado receptor começa com a procura por regiões em que podem ocorrer interações, tanto no ligante quanto no receptor. Se ocorrer uma correspondência entre a conformação e o sítio de ancoragem, essa conformação segue para a próxima etapa.

A segunda etapa analisa os átomos do ligante, em relação à cavidade do receptor, de forma a evitar choques estéricos. Além disso, são feitas avaliação em relação ao diâmetro do ligante e suas rotações. O ligante recebe uma pontuação e caso ela seja ‘satisfatória’ de acordo com os critérios estabelecidos, ele segue em frente. Em seguida, é calculada uma nova pontuação, chamada de “gananciosa”, que irá avaliar a posição do ligante ao receptor quando esse se move em torno de 1Å nas coordenadas x, y e z. Esse movimento gera uma nova pontuação chamada de ChemScore, na qual penaliza contatos estéricos além de reconhecer diversos tipos de interações como hidrofóbica, metálicas e ligações de hidrogênio. Após isso, o ligante é novamente movimentado nas coordenadas e sua pontuação é recalculada. Essa segunda etapa ocorre a redução dos números de confôrmeros. Por fim, no terceiro passo, o algoritmo Monte Carlos explora a minimização de energia dos ligantes, pontuando as poses novamente através da função *GlideScore*.

O primeiro passo para a etapa de validação do modelo consistiu na escolha do programa para as simulações de *docking*, o **GOLD** ou o **GLIDE**. Para tanto foi feito um *redocking*, para ver qual programa conseguia posicionar melhor o ligante cristalográfico, glicose-1-fosfato (G1P), em relação à sua posição original na estrutura cristalográfica PDB ID 1H5R. Foram rodadas 10 simulações para cada programa. Como pode ser observado na Figura 15, o **GOLD** foi o programa que conseguiu um melhor posicionamento, sendo ele selecionado para futuros simulações de *docking* (Figura 16).

Figura 16 - Validação do modelo gerado para a fosfomanose isomerase *Xanthomonas citri* subsp. *citri* mostrando a posição do ligante glicose-1-fosfato e sua posição em relação do ligante cristalográfico quando utilizado os programas **GOLD** e **Glide**.



A estrutura em verde corresponde ao modelo gerado (hipótese 4) e a em rosa corresponde a estrutura cristalográfica 1H5R.

Fonte: o autor.

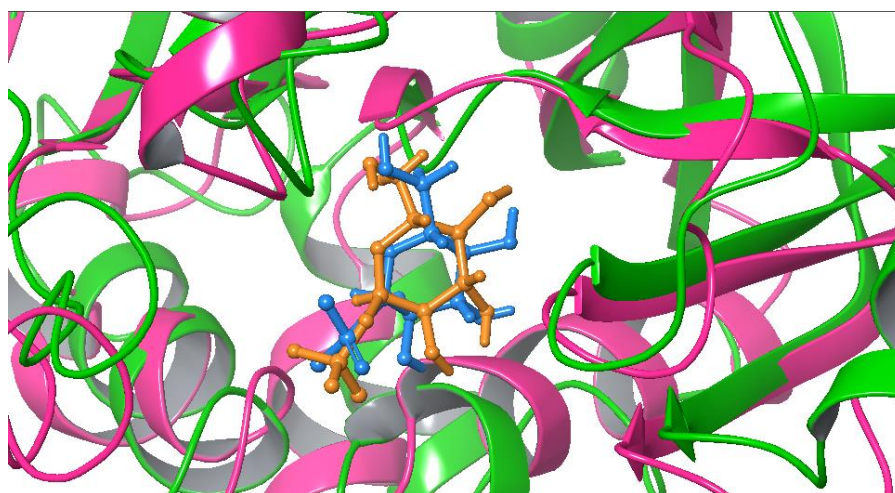
O sucesso de um programa na previsão de uma posição de ligante é geralmente medido pelo desvio médio quadrático, ou RMSD (Root Mean Square Deviation), no qual é possível determinar o grau de similaridade entre duas estruturas (COLE et al., 2005). Esse dado é calculado entre a posição do ligante observada experimentalmente e a posição do ligante prevista pelo programa, com base nas distâncias interatômicas de seus átomos, sendo o valor obtido através da seguinte expressão matemática (1):

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|r_i^A - r_i^B\|^2} \quad (1)$$

onde N é o número total de átomos pareados entre as duas estruturas, r_i é o vetor posição do átomo i , A é o modelo gerado e B é a estrutura molde.

A estatística mais usada é a que os modelos gerados deveriam obter um $RMSD < 2\text{Å}$ e, embora seja comumente utilizada, é um limite arbitrário (GOHLKE; HENDLICH; KLEBE, 2000; KONTOYIANNI; MCCLELLAN; SOKOL, 2004; KRAMER; RAREY; LENGAUER, 1999). Por outro lado, alguns autores como Vieth e Bursulaya (BURSULAYA et al., 2003; VIETH et al., 1998), consideram valores de $RMSD$ entre 2Å e 3Å como sucessos parciais. Para o modelo gerado da PMI de XAC foi encontrado um $RMSD = 2.5080\text{Å}$ (Figura 17) que, de acordo com os autores citados acima, seria considerado como um sucesso parcial.

Figura 17 - Resultado do *Redocking* da 1H5R para cálculo do $RMSD$, no qual foi obtido um valor de 2.5080Å .



A estrutura cristalográfica PDB ID 1H5R está em cor rosa, o modelo de PMI XAC em cor verde, o ligante cristalográfico em cor azul e o ligante do redocking em cor laranja.
Fonte: o autor.

Uma vez selecionado e refinado o modelo final, avaliamos sua eficiência em estudos de triagem virtual. Para essa avaliação, foram escolhidos 30 inibidores conhecidos da PMI, depositados e provenientes do servidor online **BindingDB**, contendo valores descritos de potência (IC_{50}), reportados na literatura. O objetivo dessa etapa era avaliar a capacidade que o modelo possui de recuperar os ligantes sabidamente ativos e ordená-los de forma coerente em relação ao seus IC_{50} , com base em função de pontuação (**GOLD Fitness**) proveniente do programa **GOLD**.

Para que o objetivo fosse alcançado, primeiramente era necessária a definição do sítio de *docking*, a qual foi obtida através do posicionamento do ligante proveniente da PDB ID 1H5R em relação ao modelo. Com essa posição, foi possível obter-se uma esfera com as

seguintes coordenadas espaciais: $x = 23,977$, $y = 15,659$, $z = 43,886$, com uma esfera de raio de 8 Å (região onde seriam realizadas as simulações de *docking*). Os resultados dessa etapa são mostrados na Tabela 5.

Tabela 5 - Resultado da recuperação da seleção de inibidores (15 mais potentes e 15 menos potentes) do **BindingDB** pela fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*.

Número do Inibidor	Nome PCid	Gold.PLP.Fitness	IC50
1	2998795	62.216	1.30E+03
2	1256813	60.097	1.00E+03
3	5771379	59.096	1.00E+03
4	25199560	57.191	>1,00E+5
5	4089709	56.362	707
6	25199559	55.438	>1,00E+5
7	2311082	54.774	270
8	2248499	48.79	829
9	1489253	48.566	1.23E+03
10	3093671	45.896	1.27E+03
11	4120895	44.867	>1,00E+5
12	22416235	44.529	1.25E+03
13	4917291	44.379	709
14	387196	44.267	>1,00E+5
15	25241622	43.919	>1,00E+5
16	1510389	43.654	985
17	3311395	41.868	>1,00E+5
18	25181243	41.807	1.03E+03
19	1519965	41.795	1.00E+03
20	4879298	41.788	>1,00E+5
21	3194	41.062	190
22	25181239	36.225	>1,00E+5
23	25181212	35.122	>1,00E+5
24	22765927	35.095	>1,00E+5
25	25241626	33.333	>1,00E+5
26	6112173	32.755	>1,00E+5
27	25241624	32.246	>1,00E+5
28	6033619	27.892	>1,00E+5
29	5951502	26.357	>1,00E+5
30	25199533	22.889	1.07E+03

Os inibidores mais potentes são mostrados na cor verde e os inibidores menos potentes são mostrados na cor vermelha.

Fonte: o autor.

Observando os dados obtidos pelo estudo de triagem virtual, o próximo passo consistiu na obtenção de curvas ROC (Receiver Operating Characteristic), de predição e de enriquecimento para melhor entendimento dos resultados. Para a criação desses gráficos, foi

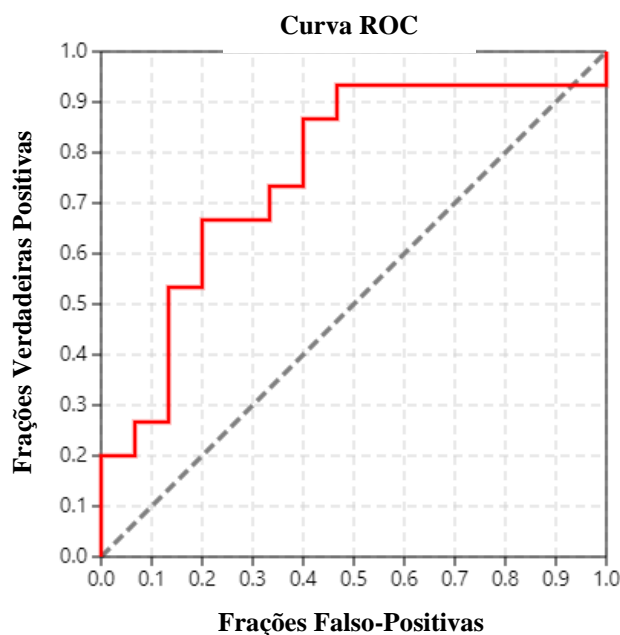
utilizado o *web servidor* Screening Explorer (EMPEREUR-MOT; ZAGURY; MONTES, 2016) com os parâmetros selecionados como padrões.

Em estudos de triagem virtual, a análise dos dados é feita com base em uma curva ROC. Trata-se de um gráfico dividido em frações positivas verdadeiras (FVP) localizadas no eixo y e frações falso-positivas (FFP) localizadas no eixo x. Um par de FVP/FFP é representado por cada ponto dentro da curva ROC e corresponde a uma fração específica de um conjunto de moléculas. As curvas ROC resumem a capacidade geral de um método para distinguir entre compostos ativos e inativos (EMPEREUR-MOT et al., 2015; EMPEREUR-MOT; ZAGURY; MONTES, 2016).

Uma função de pontuação com perfeita discriminação é aquela em que no canto superior esquerdo do gráfico, temos uma sensibilidade perfeita ($FVP = 1$) e uma especificidade perfeita ($FFP = 0$). A sensibilidade corresponde a capacidade do sistema em prever corretamente quais moléculas possuem certas características (proporção de positivos verdadeiras) ; enquanto especificidade corresponde em prever a ausência de características em algumas moléculas (proporção de falsos-positivos) (BERTRAND; TRIBALLEAU, 2005).

A linha diagonal presente no gráfico corresponde a experimentos em que a função de pontuação não seria discriminada. Uma curva entre a linha diagonal e a parte superior esquerda do gráfico é considerada boa, sendo que quanto maior a distância da linha diagonal, melhor será o sistema. De forma qualitativa, quanto mais próximo à esquerda do gráfico estiver a curva, maior será a precisão geral do estudo (BERTRAND; TRIBALLEAU, 2005; EMPEREUR-MOT et al., 2015; EMPEREUR-MOT; ZAGURY; MONTES, 2016). Abaixo, na Figura 18, é mostra o gráfico de curva ROC obtido para o estudo realizado com os 30 inibidores do BindingDB e o modelo de PMI XAC gerado. Nele é possível observar que grande parte do gráfico encontra-se distante da linha diagonal, mostrando que o estudo apresenta uma boa precisão, conseguindo diferenciar os compostos mais ativos dos menos ativos.

Figura 18 - Curva ROC obtida para o estudo dos 30 compostos relacionados como mais potentes e menos no **BindingDB** pelo molde gerado para a proteína fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*.

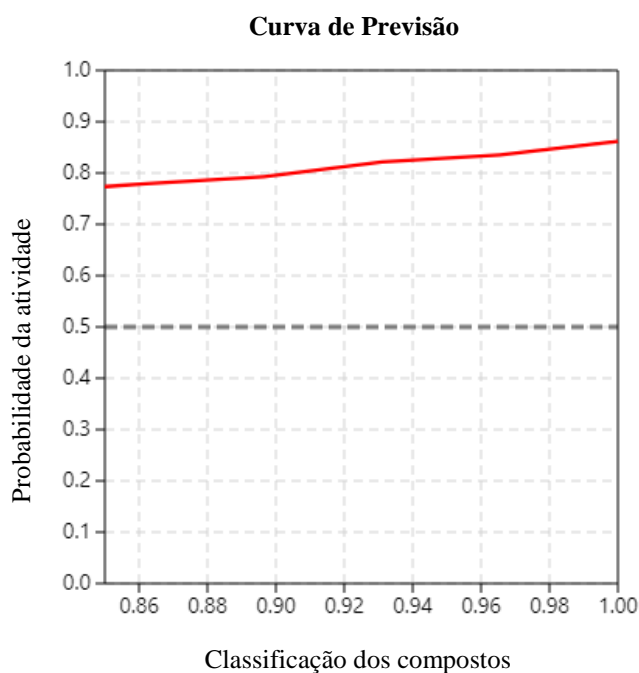


Fonte: o autor.

Caso fique difícil analisar uma curva ROC, é possível utilizar uma medida padrão chamada de área sob a curva (AUC). O AUC é um número que varia de 0 a 1 obtido por métodos de integração numérica e mostra o desempenho geral do estudo, isto é, a probabilidade de o modelo classificar um dado positivo como negativo e vice-versa. Teoricamente, quanto maior o valor obtido para AUC, melhor é o sistema utilizado (EMPEREUR-MOT; ZAGURY; MONTES, 2016). O valor de AUC ROC obtido para esse estudo foi de 0.760, um bom valor para o estudo realizado já que se encontra próximo a 1 (valor ideal).

Um outro dado obtido pelo *web server* é a curva de previsão, que quantifica a associação entre as variações de pontuação e detecção dos compostos ativos. No eixo y do gráfico é mostrado a probabilidade de atividade para cada composto presente no conjunto de dados e no eixo x, esses compostos são mostrados (EMPEREUR-MOT et al., 2015). Consequentemente, a curva de previsão permite detectar a variação emitidas por uma função de pontuação na detecção dos compostos ativos, isto é, se as variações de probabilidade de atividade são importantes o suficiente para induzir a seleção de um limiar para triagens virtuais (EMPEREUR-MOT et al., 2015). Na Figura 19 é mostrado a curva de previsão para esse estudo e ela se encontra acima da linha limite, tornando-se um bom resultado e mostrando que o modelo foi capaz de realizar uma discriminação entre os compostos utilizados.

Figura 19 - Curva de Previsão obtida para o estudo dos 30 compostos relatados como ativos e inativos no **BindingDB** pela proteína fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*.



Fonte: o autor.

Um dado calculado a partir da curva de previsão é o ganho total padronizado (TG) (BURA; GASTWIRTH, 2001), que resume a discriminação dos compostos ativos em relação à variação das pontuações ao longo do conjunto de dados das moléculas (EMPEREUR-MOT et al., 2015).

A TG possui seu valor entre 0 e 1, sendo que sua análise deve estar associado ao valor obtido para a AUC. Valores de TG acima de 0,25 combinados com uma AUC ROC acima de 0,5, significa que as variações de pontuação são relevantes na discriminação dos compostos ativos; já valores de TG acima de 0,4 associados com uma AUC ROC acima de 0,5 indica, que o método de triagem utilizado teve um bom desempenho e seria reproduzível em condições experimentais semelhantes (EMPEREUR-MOT; ZAGURY; MONTES, 2016).

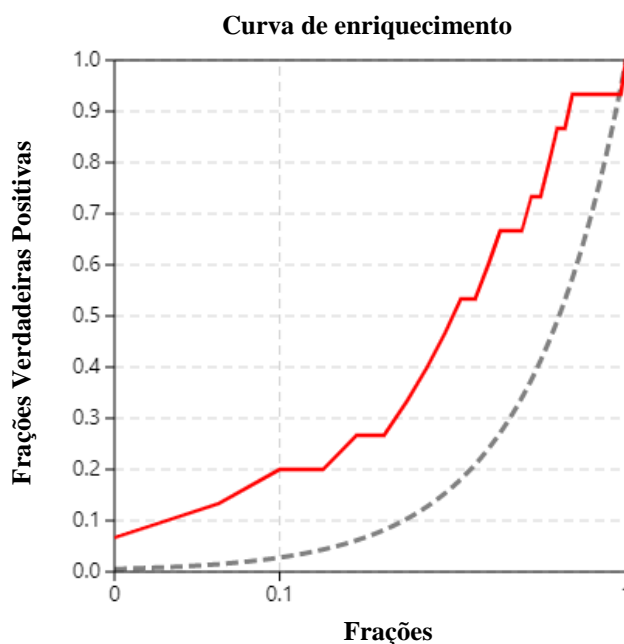
No nosso estudo, o valor de TG obtido foi de 0,312 que, associado a um AUC ROC de 0,760, mostra que o sistema foi capaz de discriminar os compostos ativos para tal modelo proteico.

Por fim, um ultimo dado obtido é a curva de enriquecimento, a qual permite avaliar a recuperação precoce de compostos ativos presentes em estudos de triagem virtual. O gráfico em escala logaritmica mostra no eixo y as FVP e no eixo x o conjunto de dados (TRUCHON; BAYLY, 2007).

Assim, como as curvas ROC, as de enriquecimento possuem uma linha diagonal que inicia no canto inferior esquerdo, subindo até o canto superior direito e que representa uma distribuição aleatória. Quanto mais distante a curva gerada estiver da linha diagonal, melhor é o estudo realizado (BERTRAND; TRIBALLEAU, 2005).

Embora as curvas de enriquecimento sejam mais fáceis de traçar, elas possuem uma grande desvantagens. Essas curvas dependem diretamente da proporção de moléculas ativas presentes no conjunto; portanto, quando a proporção desses ativos aumenta, a curva de enriquecimento ideal se aproxima da linha de classificação aleatória (BERTRAND; TRIBALLEAU, 2005). Além disso, o espaço é limitado pelas curvas ideal e aleatória, prejudicando comparações com outros dados. Na Figura 20 é mostrada a curva de enriquecimento para o estudo em questão.

Figura 20 - Curva de enriquecimento com um fator de enriquecimento de 10% obtida para o estudo dos 30 compostos relatados como ativos e inativos no **BindingDB** pela proteína fosfomanose isomerase de *Xanthomonas citri* subsp. *citri*.



Fonte: o autor.

Um dado relacionado à curva de enriquecimento é o fator de enriquecimento (EF), que corresponde à fração de compostos ativos obtidos quando uma determinada porcentagem de compostos inativos é encontrada (TRUCHON; BAYLY, 2007) Esse dado possui uma alta variabilidade e seu valor máximo depende diretamente da proporção de moléculas ativas no conjunto (TRUCHON; BAYLY, 2007) e do valor de compostos inativos considerados

(PICCIRILLO; DO AMARAL, 2018). Nesse estudo, foi considerado a porcentagem de 10% dos compostos inativos e o valor obtido para EF foi de 2.00, mostrando que para 10 compostos inativos recuperados, 2 deles são compostos ativos.

Além dessas curvas, temos alguns outros dados para levar em consideração: discriminação reforçada por Boltzmann do ROC (BEDROC) (MUEGGE; ENYEDY, 2004) e o reforço inicial robusto (RIE) (CORNELL, 2006). Truchon & Bayly (2007) discutiram vários métodos para abordar o reconhecimento precoce, como o BEDROC e o RIE. Ambas as métricas são utilizadas para quantificar o reconhecimento precoce dos compostos ativos utilizados em um estudo de triagem virtual (EMPEREUR-MOT; ZAGURY; MONTES, 2016). O BEDROC é limitado por um valor de 0 a 1 e trata da probabilidade de um ativo ser classificado antes de um composto selecionado aleatoriamente. Por outro lado, o RIE, desenvolvido por Sheridan (SHERIDAN et al., 2001), utiliza a ponderação exponencial, atribuindo um maior peso para os ativos reconhecidos precocemente (ZHAO et al., 2009).

Uma das desvantagens do RIE é que o seu valor mínimo e máximo depende diretamente da quantidade de compostos ativos e a quantidade de compostos totais presente nos conjuntos. O valor obtido para essa métrica foi 1,832. Por outro lado, o valor obtido pra BEDROC foi 0,916 encontra-se muito próximo a 1, mostrando que o modelo é capaz de separar os compostos ativos e inativos.

Com isso, através desse resultado de validação e daqueles de avaliação mostrados anteriormente, podemos afirmar que o modelo para a PMI XAC está apto para ser utilizado nos posteriores estudos de triagem virtual.

4.2 Construção do modelo da fosfomanose isomerase (PMI) para *Homo sapiens*.

Após a construção do modelo para a PMI de XAC foi decidido realizar um estudo em relação a proteína PMI de *Homo sapiens*. O primeiro passo foi realizar o alinhamento entre as sequências de aminoácidos de ambas proteínas, com o intuito de analisar as porcentagens de identidade e similaridade. Elas compartilham uma identidade de 7% e uma similaridade de 11% (Figura 21).

Figura 21 - Alinhamento entre as sequências da fosfomanose isomerase de *Xanthomonas citri* subs. *citri* e *Homo sapiens* feita pelo programa **Clustal Omega**.

<i>Xanthomonas</i>	1	-----	0
<i>Homo sapiens</i>	1	MAAPRVFPLSCAVQYAWGKMGNSSEVARLLASSDPLAQIAEDKPYAELW	50
<i>Xanthomonas</i>	1	-----	0
<i>Homo sapiens</i>	51	MGTHPRGDAKILDNRISQKTLQWIAENQDSLGSVKVDTFNGNLPFLFKV	100
<i>Xanthomonas</i>	1	-----	0
<i>Homo sapiens</i>	101	LSVETPLSIQAHPNKELAEKHLHQAPQHYPDANHKPEMAIALTPFQGLCG	150
<i>Xanthomonas</i>	1	-----MSDVLPIILSGSGTRLWPLSRESYPKQFLPL	32
<i>Homo sapiens</i>	151	FRPVEEIVTFLKTAAGNNMEDIF-----GELLLQL-HQQYP-----	185
<i>Xanthomonas</i>	33	VGDKSMLQSTWLRAPV-AGHAPIVVANEEHRFMAAEQLQ-----Q	72
<i>Homo sapiens</i>	186	-GDIGCFAIYFLNLLTLKPGEAMFLEANVPHAYLKGDCVECMACSDNTVR	234
<i>Xanthomonas</i>	73	LGVKPSAI----LLE-----PKGRNTAPAIAVAALAEATRDGADPLLL	110
<i>Homo sapiens</i>	235	AGLTPKFIDVPTLCEMLSYTPSSSKDRLFLP-----TRSQEDPYLS	275
<i>Xanthomonas</i>	111	V----LPSDHVIQNEAAFQAAVT----LAATAAEQGKLVTFGIKPTAPET	152
<i>Homo sapiens</i>	276	IYDPPVPDFTIMKTEV--PGSVTEYKVLALDSASILLMVQGTVIASPTTT	323
<i>Xanthomonas</i>	153	-----GYGIKASAGAGASAVERFVEKPDLATAQSYLASGEYYWNS	193
<i>Homo sapiens</i>	324	QTPIPLQGGVLFI----GANESVSLKLTEPKDL-----	353
<i>Xanthomonas</i>	194	GMFLFRASRYLEELRKFPAPAIADACQKAWENGKRDADFTRLDKDAFAASP	243
<i>Homo sapiens</i>	354	--LIFRACCLL-----	362
<i>Xanthomonas</i>	244	SDSIDYAVMEKTADAVVVPLDAGWNDVGSWSLLDVSNDQAQGNAAHGDV	293
<i>Homo sapiens</i>	363	-----	362
<i>Xanthomonas</i>	294	IQLDCQNTYAYGSRLIAMVGLEDVVVVETPDAVLVGHDRDRIQEVKDVVSQ	343
<i>Homo sapiens</i>	363	-----	362
<i>Xanthomonas</i>	344	IKTAGRSEATWHRKVYRPWGAYDSIDMGQRHQVKRITVKPGAVLSLQMH	393
<i>Homo sapiens</i>	363	-----	362
<i>Xanthomonas</i>	394	HRAEHWIVVSGTAEVTRGEEVLLLTENQSTYIPLGVTHRLRNPGLPLEL	443
<i>Homo sapiens</i>	363	-----	362
<i>Xanthomonas</i>	444	IEVQSGSYLGEDDIVRFEDTYGRA	467
<i>Homo sapiens</i>	363	-----	362

Fonte: o autor

Após essa etapa, foi criado o modelo para a PMI humana para comprovar que se trata de estruturas totalmente diferentes. O modelo também foi construído por modelagem de homologia estrutural, utilizando todos os mesmos passos que foram utilizados anteriormente, isto é, alinhamento, construção do modelo e avaliação do modelo.

4.1.1 Etapa de identificação da sequência de aprovação

Para a construção do modelo da PMI, inicialmente obtivemos o sequenciamento de aminoácidos da mesma através do banco de dados GenBank (CAA53657.1) (NCBI, 2019). Em seguida, o algoritmo do servidor online **BLASTp** identificou duas estruturas (PDB ID 5NW7 e 3H1M) com identidades sequenciais acima de 25% (Tabela 6), sendo que uma possui identidade de aproximadamente 35% e a outra de 40%. As identidades dessas estruturas acima de 25% mostram que elas compartilham com a estrutura de interesse uma potencial semelhança estrutural e de função

Tabela 6 - Resultados de identidade e *e-value* de sequências de proteínas obtidas no servidor online **BLAST** usando o algoritmo de explosão de proteínas (**BLASTp**) para a triagem de proteínas homologadas para PMI humana.

<i>PDB ID</i>	<i>Identidade</i>	<i>E-value</i>
5NW7-A	40.32%	2×10^{-100}
3H1M_A	34.48%	5×10^{-64}

Fonte: o autor

De posse das sequências homólogas da sequência em estudo, o próximo passo foi determinado pelo alinhamento das sequências e sua avaliação.

4.1.2 Etapa de alinhamento

A partir dos resultados obtidos com o **BLASTp** foi feito o alinhamento das sequências através do servidor online **Clustal Omega** (Figura 22).

Figura 23 - Alinhamento múltiplo final das seqüências de fosfomanose isomerase humana (PMI) e as respectivas às estruturas indicadas como homólogos pelo **BLASTp**. (PDB IDs 5NW7 e 3H1M).

PMI H.	-----MAAPRVFPLSCAVQYAWGKMGNSSEVARLLASSDPLAQIAEDKPY	46
3H1M:A	-----ASMQKLINSVQNYAWGSK-TALTELYG-----IANPQQQPM	35
5NW7:A	MRGSHHHHHHGMASMSSEKLFRIQCQGYQNYDWGKIGSSSAVAQFVHNSDPSITIDETKPY	60
	: : . *:* ** . : : : : : : *	
PMI H.	AELWMGTHPRGDAKILDNRISQKTLISQ-WIAENQDSLGSVKVDTF--NGNLPFLFKVLSV	103
3H1M:A	AELWMAHPKSSSRITTANGETVSLRDAIEKNKTAMLGEAVANRFG---ELPFLFKVLCA	92
5NW7:A	AELWMGTHPSVPSKAID--LNNQTLRDLVTAKPQEYLGESIIITKFGSSKELPFLFKVLSI	118
	*****:** : : . :* : : : ** . : * :*****.	
PMI H.	ETPLSIQAHPNKELAEKHLQAPQ-----HYPDANHKPEMAIALTPFQGLCGFRP	153
3H1M:A	AQPLSIQVHPNKRNSEIGFAKENAAGIPMDAAERNYKDPNHKPELVFALTPFLAMNAFRE	152
5NW7:A	EKVLSIQAHDPKKLGAQLHAADPK-----NYPDDNHKPEMAIAVTDTEGFCGFKP	168
	****.**:* . . : : * * *****:.*:* * . : .*	
PMI H.	VEEIVTFLLKQVPEFQFLIGDEAATHLKQTM-----SHDSQAVASSLQSCFSHLMKSEK	206
3H1M:A	FSDIVSLLQP---VAGA--HSAIAHFLOVP-----NAERLSQLFASLLNMQG	194
5NW7:A	LDQLAKTLATVPELNEIIGQELVDEFISGIKLPAEVGSQDDVNNRKLQKVFGLMNTDD	228
	...:.. * * : * . : : . . * . . * . * : : :	
PMI H.	KVVVEQLNLLVKRISQQAAGNMMEDIFGELLQLHQYYPGDIGCFAI-YFLNLLTLKPG	265
3H1M:A	EEKSRALAVLKAALNSQQ-----GEPWQTIRVISEYYPDDSGLFSP-LLLNVLKLNPG	246
5NW7:A	DVIKQQTAKLLERTDREPQVFKDIDSRLPELIQRLNKQFPNDIGLFCGCLLLNHVGLNKG	288
	. . * : : . : : : : : * . * * . : ** : * : *	
PMI H.	EAMFLEANVPHAYLKGDVCECMACSDNTVRAGLTPKFIDVPTLCEMLSYPSSSKDRFLF	324
3H1M:A	EAMFLFAETPHAYLQGVALEVMANSDNVLRAGLTPKYIDIPELVANVKFEPKPAGELLTA	305
5NW7:A	EAMFLQAKDPHAYISGDIIECMAASDNVVRAGFTPKFKDVKNLVEMLTYSYESVEKQKMP	348
	***** * : ***** . * : * ** *** . : *** : *** : * : * : : . .	
PMI H.	PTRSQ----EDPYLSIYDPPVDFTIMKTEVPGSVTEYKVL---ALDSASILLMVQGTVI-A	379
3H1M:A	PVKSG-----AELDFPIPVDDFA--FSLHDLALQETSIGQH----SAAILFCVEGEAVLR	355
5NW7:A	LQEFPRSKGDAVKSVLYDPPIAEFVSLQTIQFQKS--KGGKQVIEGLNGPSIVVIATNGKGTIQ	408
	* . : * : : * . . : . . . : * : : . : *	
PMI H.	STPTTQTPPIPLQRGGVLFIGANE-SVSLKLTEP---KDLLIFRACLL	423
3H1M:A	-----KDEQRLVLKPGESAFIGADESPVNASGT-----GRLARVYNKL	393
5NW7:A	ITGDDSTKQKIDTGYVFFVAPG--SSIELTADSANQDQDFTTYRAFVEA	455
	. : . * * : . . * : : .	

Os aminoácidos com coloração azul escuro pertencem a folha β enquanto que aqueles com coloração verde pertencem às α -hélices. A simbologia de asterisco (*) representa a presença de um único resíduo totalmente conservado, dois pontos (:) indica a conservação entre grupos de propriedades fortemente semelhantes e um ponto (.) indica a conservação entre grupos de propriedades fracamente semelhantes.

Fonte: o autor

4.1.3 Etapa de construção e avaliação do modelo

Após a etapa de alinhamento ser concluída, a próxima etapa foi conduzida através da construção do próprio modelo. Tal passo foi possível através do programa **MODELLER**, que foi utilizado e, assim, gerou 5 hipóteses de modelo. De posse dos modelos gerados, surge a

etapa de avaliação para determinar o modelo mais confiável e adequado para novos estudos. As análises de avaliação foram desenvolvidas com o servidor *online* **WHATIF** e o programa **PROCHECK**.

Dados os resultados obtidos pelo **WHATIF**, pode-se que temos três boas hipóteses: a 3, 4 e 5. Elas obtiveram um escore total do WHAT IF de -0.675, -0.675 e -0.690, respectivamente. Esses valores são próximos entre si, porém distantes quando comparamos com os valores obtidos para as estruturas cristalográficas (Tabela 7). Outro ponto de atenção são as regiões identificadas do gráfico Ramachandran. No caso das Hipóteses 3, 4 e 5, foram encontrados 3 e 4 resíduos em regiões proibidas (Tabela 7).

Outro ponto importante a considerar é o fator G. As hipóteses apresentaram um fator G de -0,2, um valor que está dentro da região para uma proteína com uma resolução de 1,85 Å (referente ao PDB ID5NW7) (Tabela 7).

Tabela 7 - Apresentação dos resultados da Pontuação Geral (**WHATIF**), Regiões Proibidas (Ramachandran) e Fator G (**PROCHECK**) obtidos na etapa de validação dos modelos gerados.

Modelos	Pontuação Geral (WHATIF)	Regiões Proibidas (Ramachandran Plot)	G-Factor
<i>3HIM</i>	- 0.408	2	0.0
<i>5NW7</i>	-0.231	0	-0.1
<i>Hipótese 1</i>	-0.611	7	-0.2
<i>Hipótese 2</i>	-0.707	3	-0.2
<i>Hipótese 3</i>	-0.675	3	-0.2
<i>Hipótese 4</i>	-0.675	3	-0.2
<i>Hipótese 5</i>	0.690	4	-0.1

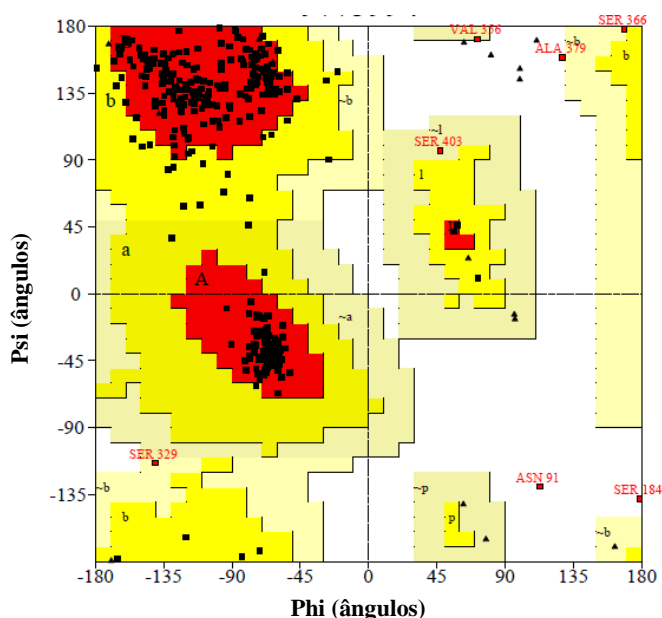
Fonte: o autor

Para que podemos realizar uma melhor seleção da hipótese escolhida, devemos analisar outros parâmetros. Em seguida foi feita a análise dos ângulos diédricos e do diagrama de Ramachandran, para poder seguir para outras análises.

4.1.4 Avaliação dos ângulos diédricos Ψ versus ϕ através do diagrama de Ramachandran

Os diagramas de **Ramachandran** obtidos para as hipóteses apresentaram alguns resíduos em regiões proibidas. As hipóteses 1 e 2 foram descartadas logo no início devido ao grande número de aminoácidos presentes em regiões proibidas, restando as hipóteses 3, 4 e 5 para análise. A hipótese escolhida foi 3, na qual apresenta 89,2% de seus resíduos em regiões altamente favoráveis e apenas 0,8% em regiões proibidas sendo os resíduos Ala (379), Asn (91) e Ser (184) (Figura 24).

Figura 24 - Diagrama de Ramachandran para a hipótese 3 obtido pelo **PROCHECK**: 89,2% de resíduos em regiões favoráveis, 8,9% de resíduos em regiões permitidas, 1,1% de resíduos em regiões generosamente permitidas e 0,8% de resíduos em regiões proibidas.



Estatísticas do gráfico

Resíduos em regiões mais favoráveis [A,B,L]	330	89.2%
Resíduos em regiões adicionais permitidas [a,b,l,p]	33	8.9%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	4	1.1%
Resíduos em regiões proibidas	3	0.8%
	----	----
Número de resíduos não glicinas e não prolina	370	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	21	
Número de resíduos de prolina	30	

Número total de resíduos	423	

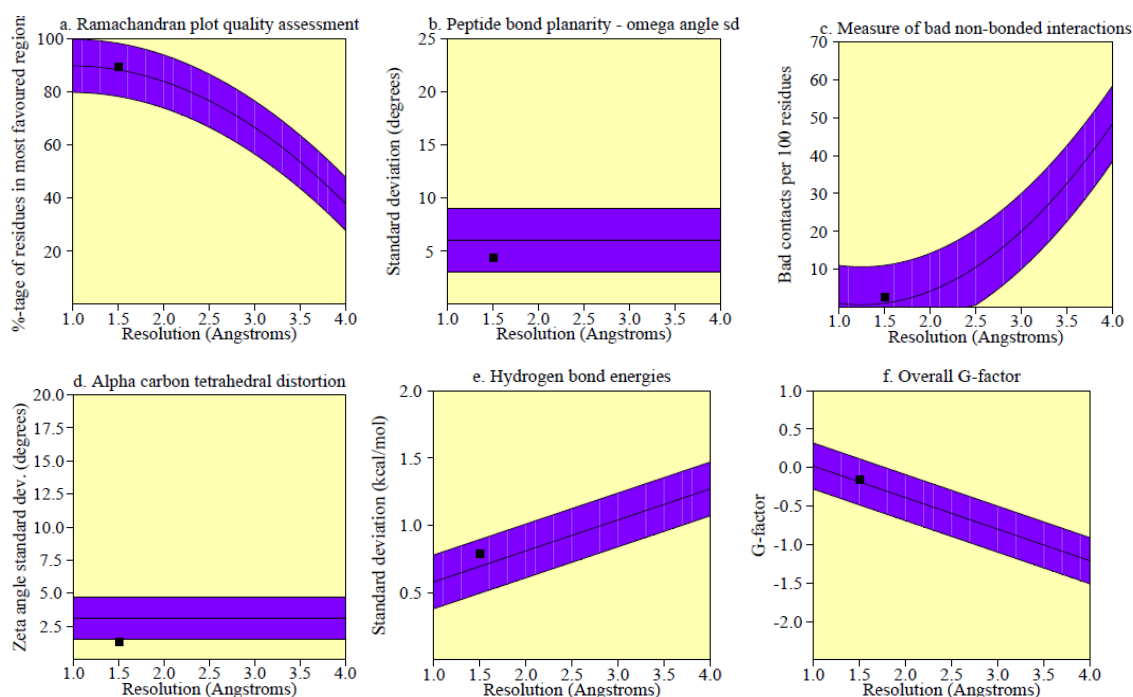
Fonte: o autor.

Embora a hipótese apresente resíduos em regiões proibidas no Ramachandran, nenhum deles se encontra na região do sítio ativo. Apenas utilizando os dados apresentados até agora, não foi possível escolher qual o melhor modelo gerado. Para que a escolha seja feita corretamente, foram utilizados outros parâmetros para a análise.

4.1.5 Outros parâmetros importantes para a análise dos modelos de PMI humana

Outras ferramentas importantes para analisar um modelo de proteína predito podem ser encontradas no pacote **SAVE v5.0**. Para a hipótese selecionada, todos os valores estão dentro dos valores esperados para estruturas de proteína de resolução de 1,85 Å (referindo-se à estrutura PDB ID 5NW7) e a distorção tetraédrica dos carbonos alfa apresentou um melhor valor que o do cristalográfico (Figura 25).

Figura 25 - Resumo da avaliação estereoquímica da hipótese 3, construída por homologia estrutural para PMI humana.

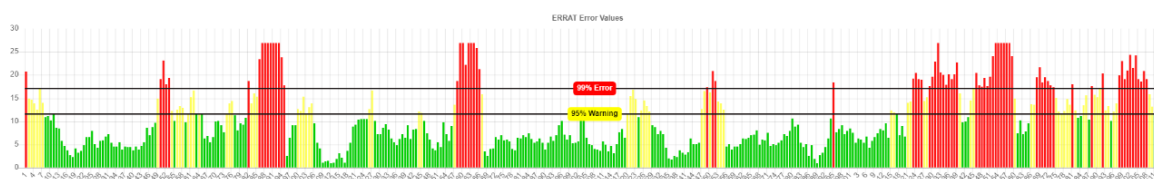


Fonte: o autor.

Além das ligações atômicas que ocorrem na cadeia peptídica, há também interações entre átomos não-ligados. A estrutura terciária de uma proteína é altamente dependente dessas interações e existem seis tipos de interações prováveis entre átomos não ligados e elas não ocorrem aleatoriamente. Assim, é possível através do programa **ERRAT** identificar regiões e fazer uma análise estatística dessas interações.

Observando o gráfico desenvolvido pela **ERRAT**, é possível identificar três regiões distintas: uma verde, uma amarela e uma vermelha. A hipótese gerada apresenta grande quantidade de aminoácidos listados como vermelho, porém esses aminoácidos não estão na região do sítio ativo. O fator de qualidade, valor relacionado às interações atômicas não ligadas, apresentados pela hipótese (61.6867) estão acima de 50, mostrando que é uma estrutura de alta qualidade (Figura 26).

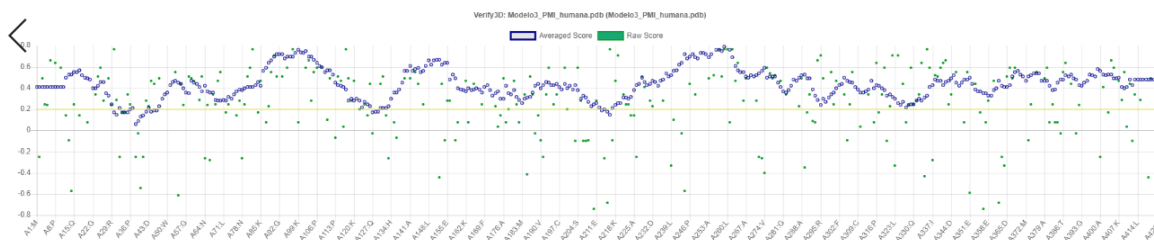
Figura 26 - ERRAT obtido para a hipótese 3 gerado de PMI humana.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas. Fonte: o autor.

Outra análise interessante é verificar se um resíduo de aminoácido corresponde ao ambiente em que está, ou seja, se há compatibilidade do modelo com sua sequência de aminoácidos. Essa análise pode ser realizada usando o **Verify-3D**. No estudo da hipótese 3, pode-se observar que 95,51% dos aminoácidos teve uma pontuação igual ou superior a 0,2 no perfil 3D-1D (Figura 27). Além disso, há regiões que caiu abaixo do limite, indicando uma provável dobragem nessa região; no entanto, não é motivo de preocupação porque não está na região do site ativo das hipóteses.

Figura 27 - Verify-3D obtido para a hipótese 3 da PMI humana: 95,51% dos aminoácidos encontram-se dentro do perfil esperado.

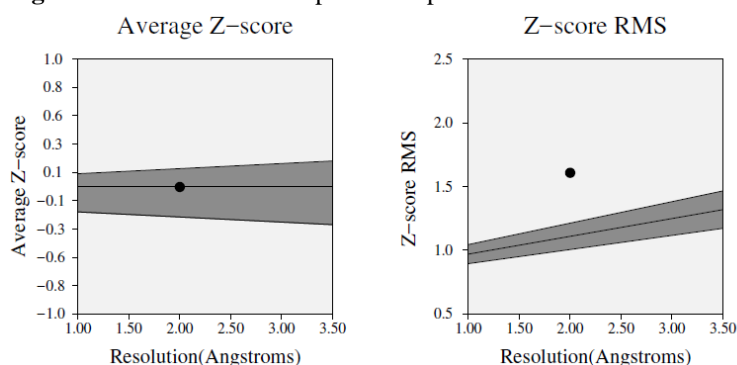


Fonte: o autor.

Finalmente, como última análise, é possível verificar o volume de átomos. O programa **PROVE** correlaciona o volume obtido para cada aminoácido da proteína com os volumes de

proteína de alta resolução. Como resultado, o **PROVE** fornece um valor de pontuação chamado escore Z que corresponde ao desvio estatístico programado de estruturas altamente resolvidas e refinadas. No caso da hipótese escolhida do modelo PMI em estudo, o Z-score encontra-se dentro da faixa esperada e o Z-score RMS acima da região de interesse (Figura 28). No entanto, juntamente com todas as outras análises realizadas, pode-se determinar que se trata da melhor hipótese de modelo para a PMI humana.

Figura 28 - PROVE obtido para o a hipótese 3 da PMI humana.



O escore Z corresponde aos desvios dos valores do volume atômico em relação aos valores padrão enquanto o Z-score rms mede a média de volumes irregulares na estrutura da proteína.

Fonte: o autor.

4.1.6 Etapa de validação do modelo de PMI humana

Uma vez selecionado e refinado o modelo final, avaliamos sua eficiência em estudos de triagem virtual. Para essa avaliação, foram escolhidos 30 inibidores conhecidos da PMI, depositados e provenientes do servidor online **BindingDB**, contendo valores descritos de potência (IC_{50}), reportados na literatura. O objetivo dessa etapa era avaliar a capacidade que o modelo possui de recuperar os ligantes sabidamente ativos (ditos “contaminantes”) e ordená-los de forma coerente em relação ao seus IC_{50} , com base em função de pontuação (**GOLD Fitness**) proveniente do programa **GOLD**.

Para que o objetivo fosse alcançado, primeiramente era necessária a definição do sítio de *docking*, uma esfera com as seguintes coordenadas espaciais: $x = 20.21$, $y = 16.63$, $z = 10.00$, com uma esfera de raio de 7 Å. Os resultados dessa etapa são mostrados na Tabela 8.

Tabela 8 - Resultado da recuperação da seleção de inibidores do *BindingDB*.

Número de Inibidores	Inibidores	GOLD PLP Fitness	IC50 (nM)
1	2998795	76,169	1.30E+03
2	5771379	76,061	1.00E+03
3	4089709	72,416	707
4	25199560	67,963	>1,00E+6
5	25199559	67,128	>1,00E+5
6	2248499	66,702	829
7	4917291	61,567	709
8	1256813	60,232	1.00E+03
9	3311395	58,643	>1,00E+5
10	1489253	58,532	>1,00E+5
11	1519965	58,281	1.00E+03
12	4120895	58,255	>1,00E+5
13	2311082	57,179	270
14	25181243	56,910	1.03E+03
15	1510389	54,689	985
16	25241622	54,330	>1,00E+5
17	387196	53,533	>1,00E+5
18	3093671	53,365	1.27E+03
19	22416235	52,628	1.25E+03
20	3194	51,576	190
21	4879298	46,232	>1,00E+5
22	25181212	45,228	>1,00E+5
23	25181239	44,972	>1,00E+5
24	6112173	44,039	>1,00E+5
25	22765927	43,708	>1,00E+5
26	25241626	43,102	>1,00E+5
27	25241624	42,484	>1,00E+5
28	25199533	41,919	>1,00E+5
29	6033619	40,274	>1,00E+5
30	5951502	40,144	>1,00E+5

Os inibidores mais potentes estão em verde enquanto os menos potentes estão em vermelho.

Fonte: o autor.

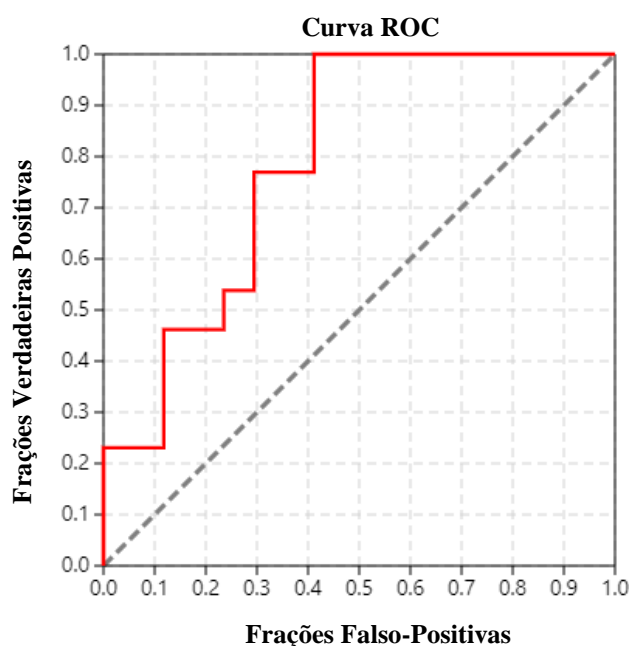
Observando os dados obtidos pelo estudo de triagem virtual, o próximo passo consistiu na obtenção de curvas ROC, de predição e de enriquecimento para melhor entendimento dos resultados. Para a criação desses gráficos, foi utilizado o *web servidor* Screening Explorer (EMPEREUR-MOT; ZAGURY; MONTES, 2016) com os parâmetros selecionados como padrões.

Em estudos de triagem virtual, a análise dos dados é feita com base em uma curva ROC (Receiver Operating Characteristic), resumem a capacidade geral de um método para

distintuar entre compostos ativos e inativos (EMPEREUR-MOT et al., 2015; EMPEREUR-MOT; ZAGURY; MONTES, 2016). Utilizando os mesmos análises feitas no item anterior, quanto maior a distância entre a linha do gráfico e a linha diagonal, melhor será a precisão geral do estudo (BERTRAND; TRIBALLEAU, 2005; EMPEREUR-MOT et al., 2015; EMPEREUR-MOT; ZAGURY; MONTES, 2016).

Abaixo, na Figura 29, é mostra o gráfico de curva ROC obtido para o estudo realizado com os 30 inibidores do BindingDB e o modelo de PMI *Homo sapiens* gerado. Nele é possível observar que grande parte do gráfico encontra-se distante da linha diagonal, mostrando que o estudo apresnta uma boa precisão, conseguindo diferenciar os compostos mais ativos dos menos ativos. Inclusive uma parte da linha pertencente ao gráfico encontra-se na parte superior, mostrando que esses inibidores encontram-se com a sensibilidade e especificidade perfeitas.

Figura 29 - Curva ROC obtida para o estudo dos 30 compostos relatados como ativos e inativos no **BindingDB** para a proteína PMI de *Homo sapiens*.



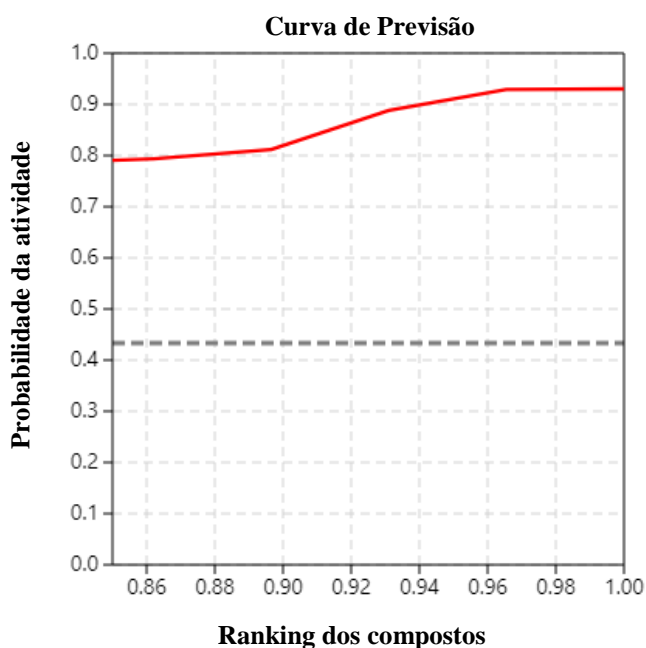
Fonte: o autor.

Outro dado que pode ser utilizado é a AUC, que mostra o desempenho geral do estudo. O valor de AUC ROC obtido para esse estudo foi de 0.792, um bom valor para o estudo realizado já que se encontra próximo a 1 (valor ideal).

Um outro dado obtido pelo *web server* é a curva de previsão, que quantifica se as variações de probabilidade de atividade são importantes o suficiente para induzir a seleção de um limiar para triagens virtuais (EMPEREUR-MOT et al., 2015). Na Figura 30 é mostrado a

curva de previsão para esse estudo e ela se encontra acima da linha limite, tornando-se um bom resultado e mostrando que o modelo foi capaz de realizar uma discriminação entre os compostos utilizados.

Figura 30 - Curva de Previsão obtida para o estudo dos 30 compostos relatados como ativos e inativos no **BindingDB** para a proteína PMI de *Homo sapiens*.



Fonte: o autor.

Um dado calculado a partir da curva de previsão é o ganho total padronizado (TG) (BURA; GASTWIRTH, 2001), que resume a discriminação dos compostos ativos em relação à variação das pontuações ao longo do conjunto de dados das moléculas (EMPEREUR-MOT et al., 2015).

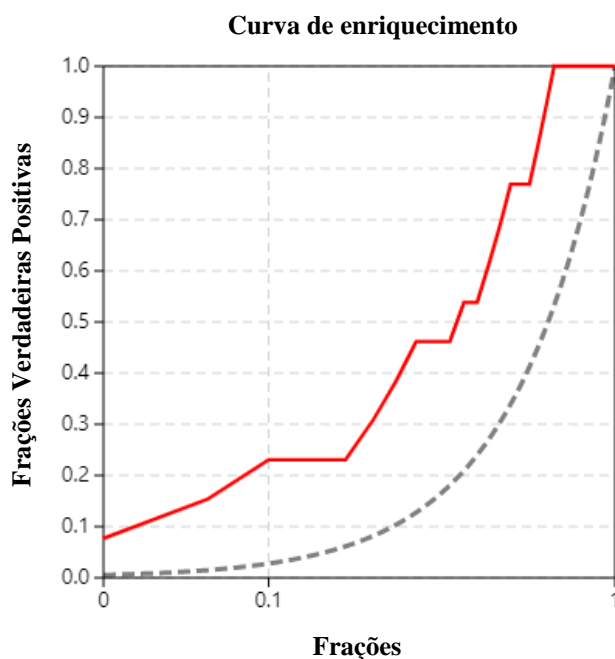
A TG possui seu valor entre 0 e 1, sendo que sua análise deve estar associado ao valor obtido para a AUC. Valores de TG acima de 0,4 associados com uma AUC ROC acima de 0,5 indica, que o método de triagem utilizado teve um bom desempenho e seria reproduzível em condições experimentais semelhantes (EMPEREUR-MOT; ZAGURY; MONTES, 2016).

No nosso estudo, o valor de TG obtido foi de 0,444 que, associado a um AUC ROC de 0,792, mostra que o método utilizado teve um bom desempenho, podendo ser reprodutível em condições experimentais semelhantes.

Por fim, um ultimo dado obtido é a curva de enriquecimento, a qual permite avaliar a recuperação precoce de compostos ativos presentes em estudos de triagem virtual. Assim,

quanto mais distante a curva gerada estiver da linha diagonal, melhor é o estudo realizado (BERTRAND; TRIBALLEAU, 2005). Na Figura 31 é mostrado a curva de enriquecimento para o estudo em questão.

Figura 31 - Curva de enriquecimento com um fator de enriquecimento de 10% obtida para o estudo dos 30 compostos relatados como ativos e inativos no **BindingDB** para PMI de *Homo sapiens*.



Fonte: o autor.

Um dado relacionado à curva de enriquecimento é o fator de enriquecimento (EF), que corresponde à fração de compostos ativos obtidos quando uma determinada porcentagem de compostos inativos é encontrada (TRUCHON; BAYLY, 2007) Nesse estudo, foi considerado a porcentagem de 10% dos compostos inativos e o valor obtido para EF foi de 2.31.

Além dessas curvas, temos alguns outros dados para levar em consideração: discriminação reforçada por Boltzmann do ROC (BEDROC) (MUEGGE; ENYEDY, 2004) e o reforço inicial robusto (RIE) (CORNELL, 2006). Ambas as métricas são utilizadas para quantificar o reconhecimento precoce dos compostos ativos utilizados em um estudo de triagem virtual (EMPEREUR-MOT; ZAGURY; MONTES, 2016). O valor obtido para o RIE foi 2,069; enquanto que o valor obtido pra BEDROC foi 0,897 encontra-se muito próximo a 1, mostrando que o modelo é capaz de separar os compostos ativos e inativos.

Com isso, através desse resultado de validação e daqueles de avaliação mostrados anteriormente, podemos afirmar que o modelo para a PMI *Homo sapiens* provou ser capaz de diferenciar os inibidores mais potentes dos menos potentes escolhidos no **BindingDB**.

4.3 Triagem virtual par seleção de inibidores da PMI de XAC.

Um dos primeiros e mais importantes passos nos estudos de planejamento e desenvolvimento de moléculas *in silico* são as conformações que os ligantes podem assumir. A posição e distância que os grupos funcionais assumem em relação aos resíduos de aminoácidos dentro do alvo molecular são fundamentais para que ocorra uma interação entre eles e, conseqüentemente, a formação de um complexo (LAPLANTE et al., 2014). Assim, avaliar e explorar o máximo de conformações que um ligante pode assumir torna-se uma atividade essencial em diversas metodologias computacionais (BAI et al., 2010).

O primeiro passo foi preparar as 6 bases de dados, descritas anteriormente, com o programa **OMEGA**. Esse passo tem o intuito de corrigir e criar diversas conformações das moléculas presentes nas bases de dados. Esse programa realiza a geração de confôrmeros em três dimensões com baixo desprendimento computacional e um alto efeito no alcance da conformação bioativa entre os confôrmeros gerados. O **OMEGA** obtém os confôrmeros em três passos. O primeiro passo, para a obtenção dos confôrmeros, consiste ao acesso à biblioteca de fragmentos, gerada através do fragmento de moléculas presente em um extenso banco de dados, para a construção da estrutura em três dimensões (3D). Os fragmentos são ligados com base em valores de distância de ligações, sendo ajustados posteriormente e minimizados por um campo de força ajustado.

Em seguida, todas as torções rotativas são numeradas de acordo com valores de um banco dados presente no programa. As rotações são ajustadas com base em ângulos e energias torcionais derivadas de um conjunto de estruturas cristalinas. Novas torções são geradas no ligante, prestando atenção a choques e duplicações, com a finalidade de gerar uma grande quantidade de confôrmeros.

Enfim, na última fase, ocorre a análise da estrutura em relação aos seus valores energéticos e geometria. Nessa etapa, a molécula pertencente ao mínimo global tem o seu RMSD comparado à outras, eliminando aquelas não pertencentes a parametrização fornecida pelo usuário (HAWKINS et al., 2010; HAWKINS; NICHOLLS, 2012).

Após a preparação das bases de dados, um tipo de triagem virtual foi realizado, o qual é baseado em diferentes tipos de similaridade: de forma e eletrostática. Essa metodologia explora o conceito de similaridade, ou seja, a triagem virtual em bases de dados de compostos

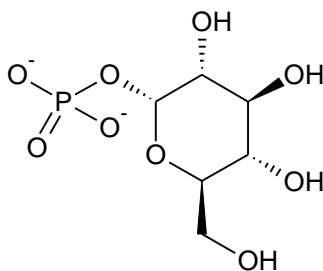
será realizada com o intuito de explorar diversidade química e selecionar novos compostos (e novas classes químicas) que sejam similares em forma e eletrostática ao substrato presente a estrutura cristalográfica PDB ID 1H5R.

O posicionamento de um ligante em um sítio ativo de uma biomolécula é um fato determinante para a atividade biológica do mesmo, já que esse sítio ativo apresenta um determinado formato para que a interação ocorra. Dessa forma, a estrutura tridimensional de um ligante é importante para que ocorra a formação do complexo ligante-proteína (HAIGH et al., 2005). Ambos os programas utilizados na etapa de triagem por similaridade e forma, **ROCS** e **EON**, da **OpenEye**, tem revelado compostos promissores para o tratamento de várias doenças (NAYLOR et al., 2009).

O programa **ROCS** realiza uma triagem virtual por forma, através da comparação entre os volumes e conformação de dois compostos, a molécula de referência (G1P) (Figura 32) e as moléculas presentes em bancos de dados. Os cálculos dos volumes são feitos através da superfície exposta de esferas de inserção colocadas na molécula referência.

Após o **ROCS**, uma outra triagem de similaridade foi feita comparando os mapas de potencial eletrostático das moléculas. O programa utilizado nessa etapa foi o **EON**, sendo que seus arquivos de entrada foram os provenientes do **ROCS**. Com esse programa se realiza a comparação e classificação dos compostos bases em seus mapas de superfície eletrostática. Os parâmetros adotados para ambos programas não foram alterados.

Figura 32 - Estrutura da G1P, utilizada como referência para os estudos de triagem virtual por similaridade nos programas **ROCS** e **EON**.



Fonte: O autor.

Para a obtenção dos melhores inibidores possíveis para a PMI de XAC, foram utilizadas três estratégias. A primeira estratégia corresponde a apenas a etapa de *docking* com o programa **GOLD** (8Å) utilizando o modelo por homologia como receptor, seguido por previsões ADME/TOX com os programas **QikProp** e **DEREK** respectivamente. A segunda estratégia

inicia-se com uma triagem virtual por similaridade utilizando os programas **ROCS** e **EON**, seguido pelo *docking* e, novamente, predições ADME/Tox. Por fim, na terceira estratégia, a triagem virtual por similaridade é realizada após o *docking*; seguindo para análises de predições ADME/Tox. Em todas as estratégias realizadas, o último passo corresponde à análise visual da interação dos compostos com o receptor.

Na primeira estratégia, as bases de dados foram submetidas ao *docking* utilizando o centroide de $x = 23.977$, $y = 15.659$ e $z = 43.886$ e um raio de esfera de 8 \AA para as bases de dados Chembridge CL, Chembridge EXP, DrugDatabase, DrugLike, *FDA Binding*, *Maybridge* e *Natural Stock*. Os arquivos finais de cada base de dados saíram com 500 compostos e foram utilizados em estudos de predições ADME/TOX. Em seguida, esses compostos passaram por uma análise visual para verificar quais possuíam uma maior interação com o modelo da PMI XAC. Abaixo (Tabela 9) são mostrados quantos compostos sobreviveram a todas as etapas.

Tabela 9 - Relação entre os compostos sobreviventes das bases de dados após simulações de *docking* com esfera de 8 \AA , predições ADME/Tox (**DEREK** e **QikProp**) e inspeção visual, utilizando a primeira estratégia.

Bases de dados	<i>Docking</i>	DEREK	QikProp	Inspeção Visual
<i>Chembridge CL</i>	500	155	5	0
<i>Chembridge EXP</i>	500	226	4	0
<i>DrugDatabase</i>	500	83	41	0
<i>FDA Binding</i>	500	92	28	5
<i>Maybridge</i>	500	5	5	0
<i>Natural Stock</i>	500	93	91	0

Fonte: o autor.

Na segunda estratégia, a primeira triagem feita foi em relação à similaridade por forma entre a molécula referência e as bases de dados. Nessa triagem foi realizada com o programa **ROCS** e foram obtidos os 2000 melhores resultados para as bases de dados *Chembridge CL*, *Chembridge EXP*, *DrugDatabase*, *DrugLike*, *FDA Binding*, *Maybridge* e *Natural Stock*. Os compostos selecionados nessa etapa seguiram para os estudos de similaridade por mapa eletrostático com o programa **EON**, no qual foram selecionados 1000 melhores resultados.

Em seguida, os arquivos obtidos na etapa anterior foram submetidos ao *docking* utilizando o centroide de $x = 23.977$, $y = 15.659$ e $z = 43.886$ e um raio de esfera de 6 \AA e 60 de fitness. O tamanho da esfera teve que ser modificado devido a problemas encontrados para alguns ligantes em relação a sua torção e aumento do tamanho das ligações. Os arquivos finais de cada base de dados saíram com 500 compostos e foram utilizados em estudos de predições ADME/Tox. Em seguida, esses compostos passaram por uma análise visual para verificar quais

possuíam uma maior interação com o modelo da PMI de XAC. Abaixo (Tabela 10) são mostrados quantos compostos sobreviveram a todas as etapas.

Tabela 10 - Relação entre os compostos sobreviventes das bases de dados após estudos triagem virtual por similaridade (**ROCS** e **EON**), de *docking*, predições ADME/Tox (**DEREK** e **QikProp**) e inspeção visual, utilizando a segunda estratégia.

Bases de dados	ROCS	EON	<i>Docking</i>	DEREK	QikProp	Inspeção Visual
<i>Chembridge CL</i>	2000	1000	500	8	0	0
<i>Chembridge EXP</i>	2000	1000	500	1	0	0
<i>DrugDatabase</i>	2000	1000	500	13	13	0
<i>DrugLike</i>	2000	1000	500	265	83	38
<i>FDA Binding</i>	2000	1000	500	31	18	1
<i>Maybridge</i>	2000	1000	0	0	0	0
<i>Natural Stock</i>	2000	1000	500	3	3	1

Fonte: o autor.

Na terceira e final estratégia, as bases de dados *Princeton*, *IBS Natural* e *IBS Synthetic* foram submetidas ao *docking* utilizando o centroide de $x = 23.977$, $y = 15.659$ e $z = 43.886$ e um raio de esfera de 6 Å e 60 de fitness. Nessa etapa foram obtidos 5000 compostos de cada base de dados. O próximo passo corresponde a busca de compostos com similariedade, por forma e eletrostática, utilizando o arquivo proveniente do **GOLD**. Nessa triagem foram utilizados os programas **ROCS** e **EON**, do qual foram obtidos os 3000 e 500 melhores resultados respectivamente.

Os arquivos finais de cada base de dados obtiveram como *hits* 500 compostos (provenientes do **EON**) e foram utilizados em estudos de predições ADME/Tox. Em seguida, esses compostos passaram por uma análise visual para verificar quais possuíam uma maior interação com o modelo da PMI de XAC. Abaixo (Tabela 11) são mostrados quantos compostos sobreviveram a todas as etapas.

Tabela 11 - Relação entre os compostos sobreviventes das bases de dados após simulações de *docking* com esfera de 6 Å e 60 de fitness, triagem virtual por similaridade (**ROCS** e **EON**), previsões ADME/Tox (**DEREK** e **QikProp**) e inspeção visual, utilizando a terceira estratégia.

Bases	<i>Docking</i>	ROCS	EON	DEREK	QikProp	Inspeção Visual
<i>Princeton</i>	5000	3000	500	105	9	0
<i>IBS Natural</i>	5000	3000	500	0	0	0
<i>IBS Synthetic</i>	5000	3000	500	135	4	0

Fonte: o autor.

As análises de propriedades farmacocinéticas são extremamente importantes para o desenvolvimento de novos compostos, pois o fato de algumas moléculas não apresentarem essas propriedades adequadas são um dos principais insucessos em pesquisas voltadas para essa área. A descoberta dessas propriedades antes dos ensaios biológicos, permite um menor gasto de tempo e dinheiro para substâncias biologicamente ativas (BOOBIS et al., 2002; PRENTIS; LIS; WALKER, 1988; VENKATESH; LIPPER, 2000).

O cálculo de descritores é um método de predição de atividades farmacocinéticas e um importante filtro em estudos de triagem virtual. Esses cálculos podem ser realizados através do programa **QikProp** (SCHRODINGER, [s.d.]). O programa **Qikprop**, da empresa **Schrodinger** realiza a predição de propriedades farmacocinéticas (ADME) utilizando descritores e com baixo desprendimento computacional. A estrutura da molécula é analisada como um todo e a predição é baseada em sua estrutura 3D.

Diversas importantes propriedades farmacocinéticas podem ser preditas e comparadas com um extenso banco de dados. Entre essas propriedades, destacamos aquelas utilizadas: *log CACO2*, *log MDCK*, porcentagem de absorção oral e qualidade do modelo de absorção oral humana.

O *log CACO2* é uma medida de absorção pela mucosa intestinal por difusão passiva, realizado através de estudos em células do epitélio intestinal diferenciadas, Caco-2 (QPPCaco). A linhagem celular Caco-2 é derivada de um carcinoma colo-retal humano e é fonte de para a absorção de fármacos *in-vitro* (ARTURSSON; KARLSSON, 1991). Junto com o modelo de células Caco-2, a permeabilidade das células MDCK (*Madin-Darby canine kidney*) também desempenham um importante fator a ser analisado no estudo de permeabilidade de fármacos *in-vitro* (HORSTER; STOPP, 1986; IRVINE et al., 1999), sendo um importante mimetizador da barreira sangue-cérebro.

Os valores utilizados como critério para a eliminação dos compostos nos estudos de triagem virtual foram: *log CACO2* e *log MDCK* acima ou igual a 500, porcentagem de absorção oral maior ou igual a 80% e qualidade do modelo de absorção oral humana igual a “alta”.

As predições toxicológicas foram calculadas através do programa **DEREK**, um programa em que a toxicidade em moléculas é baseado no conhecimento. A estrutura total da molécula é analisada por similaridade 2D e indica quais possíveis riscos toxicológicos que as estruturas possam ter. Além disso, também é feita uma comparação entre grupos toxicóforos, isto é, fragmentos da molécula e fragmentos presentes em bancos de dados. Dentre as predições toxicológicas mostradas por esse programa temos a mutagenicidade, carcinogenicidade, sensibilização da pele, irritação, entre outras. Essa etapa permite excluir moléculas potencialmente nocivas para futuros ensaios (COLE; BURSAVICH, 2010). Esse programa permite escolher a espécie para qual as predições serão feitas e, nesse estudo, as predições foram feitas para humanos enquanto o critério de exclusão foi baseado em qualquer alerta de toxicidade segundo Custom Prediction e Lhasa Prediction.

Após as análises farmacocinéticas e toxicológicas, todas as bases foram submetidas à uma análise visual, que possui o intuito de verificar uma maior interação com o modelo da PMI XAC através da quantidade de interações químicas. As análises dos resultados gerados serviram como critério final para a seleção de compostos, sendo um total de 45 inibidores. Ao final dos estudos, a relação dos compostos sobreviventes em cada base de dado são mostrados nas tabelas acima.

Os programas de *docking* analisam a afinidade ligante-proteína em estudos de avaliação. Essa etapa do estudo teve sua eficiência aumentada graças à evolução do processamento computacional e automação de buscas sistemáticas (JAIN, 2006; KITCHEN et al., 2004; YURIEV; AGOSTINO; RAMSLAND, 2011; YURIEV; RAMSLAND, 2013). O posicionamento dos ligantes em cavidades dos receptores é realizada por um algoritmo, sendo no caso do programa **GOLD**, o algoritmo genético.

Classificar as poses mais efetivas baseadas em uma pontuação também trata-se de uma grande dificuldade encontrada nos programas de *docking* (JAIN, 2006; KITCHEN et al., 2004; YURIEV; AGOSTINO; RAMSLAND, 2011; YURIEV; RAMSLAND, 2013). Há diferentes funções de pontuações adotadas por cada programa e, algumas delas, são baseadas no conhecimento, consenso, dados experimentais além da combinação de mecânica quântica e mecânica molecular (JAIN, 2006; KITCHEN et al., 2004; YURIEV; AGOSTINO; RAMSLAND, 2011; YURIEV; RAMSLAND, 2013).

Os resultados dos simulações de *docking* nos permitiram visualizar como os compostos se comportam no sítio ativo do modelo. Sendo assim, a seleção final se baseou prioritariamente na possibilidade dos compostos interagirem com os resíduos Arg173 Asn192, Asp249, Asp270

e Gly155; sendo as interações com os aminoácidos arginina (173) e ácido aspártico (249) as mais importantes.

O programa **GOLD** indicou que alguns compostos podem interagir de forma satisfatória no sítio ativo do modelo e, dessa forma, elegemos estes compostos como os mais promissores. Foram selecionados 1 composto da base de dados *Natural Stock*, 6 compostos da *FDA Binding* e 38 compostos da *DrugLike*.

De todos os compostos selecionados, alguns foram comprados e em seguida serão submetidos a ensaios *in vitro* e *in vivo*, para a comprovação da ação inibidora da enzima PMI de *Xanthomonas citri* subsp. *citri*. Doze inibidores foram selecionados para compra: **DL00392005**, **DL03776644**, **DL03874928**, **DL03881790**, **DL03999102**, **DL04091967**, **DL05234422**, **DL07998085**, **DL13518039**, **DL13522617**, **DL15021627** e **DL35935889**; além disso também foi comprado um inibidor de PMI humana para ser testado. Esses inibidores que partiram para os ensaios biológicos foram selecionados com base nas interações com aminoácidos interessantes que compõem o sítio ativo e quantidade dessas interações; além de serem moléculas com perfis polares e não muito rígidas (Tabela 12).

Tabela 12 - Tipo e comprimento de interações entre os compostos comprados e os aminoácidos presentes no sítio ativo.

Composto	Aminoácidos	Tipo de interação	Comprimento da interação (Å)
DL00392005	Gly155 / Asn192	Ligações de hidrogênio	2,08 e 1,68
	Glu 173 / Asp249	Iônica	2,97 e 2,94
DL03776644	Gly 153 / Gly155	Ligações de hidrogênio	2,00 e 2,99 / 2,10
	Asn192		2,77 e 2,91
DL03874928	Gly155 / Glu 173	Ligações de hidrogênio	2,04 e 1,83
	Asn192 / Asp249		1,73 e 2,10 / 2,04 e 2,10
DL03881790	His116 / Gly155	Ligações de hidrogênio	2,19 / 2,08
	Glu173 / Asp249		2,05 / 1,69
DL03999102	Gly155 / Glu173	Ligações de hidrogênio	2,07 / 2,06
	Asn192		2,01, 2,03 e 2,10
DL04091967	Gly155 / Asp249	Ligações de hidrogênio	1,85 / 2,60
	Asn192		1,79 e 1,96
DL05234422	Gly155 / Glu173	Ligações de hidrogênio	1,82 / 1,72
	Asn192 / Ser193		1,69 e 2,17 / 2,18
DL07998085	His116 / Gly155	Ligações de hidrogênio	2,10 / 2,08
	Asn192		1,73 e 2,10
DL13518039	Gly155 / Glu173	Ligações de hidrogênio	2,07

	Asn192		1,70 e 2,00
DL13522617	Gly153 / Gly155	Ligações de hidrogênio	2,10 e 2,11 / 2,06
	Asn192 / Asp249		1,99 e 2,03 / 2,06
DL15021627	Gly155 / Glu173	Ligações de hidrogênio	2,02 / 2,10
	Asn192 / Asp270		1,66 e 2,07 / 2,06
DL35935889	Gly155 / Glu173	Ligações de hidrogênio	1,88 / 1,99
	Asn192 / Asp249		1,86 e 2,10 / 2,57

Fonte: o autor

A seguir estão mostradas todas as estruturas dos compostos selecionados (Figura 33) para compras e, conseqüentemente, os testes biológicos.

Figura 33 - Representação dos modos de interação do composto selecionado **DL00392005**, **DL03776644**, **DL03874928**, **DL03881790**, **DL03999102**, **DL04091967**, **DL05234422**, **DL07998085**, **DL13518039**, **DL13522617**, **DL15021627** e **DL35935889** na região do sítio ativo do modelo de XAC em duas dimensões obtidas como resultado do programa **GOLD**.

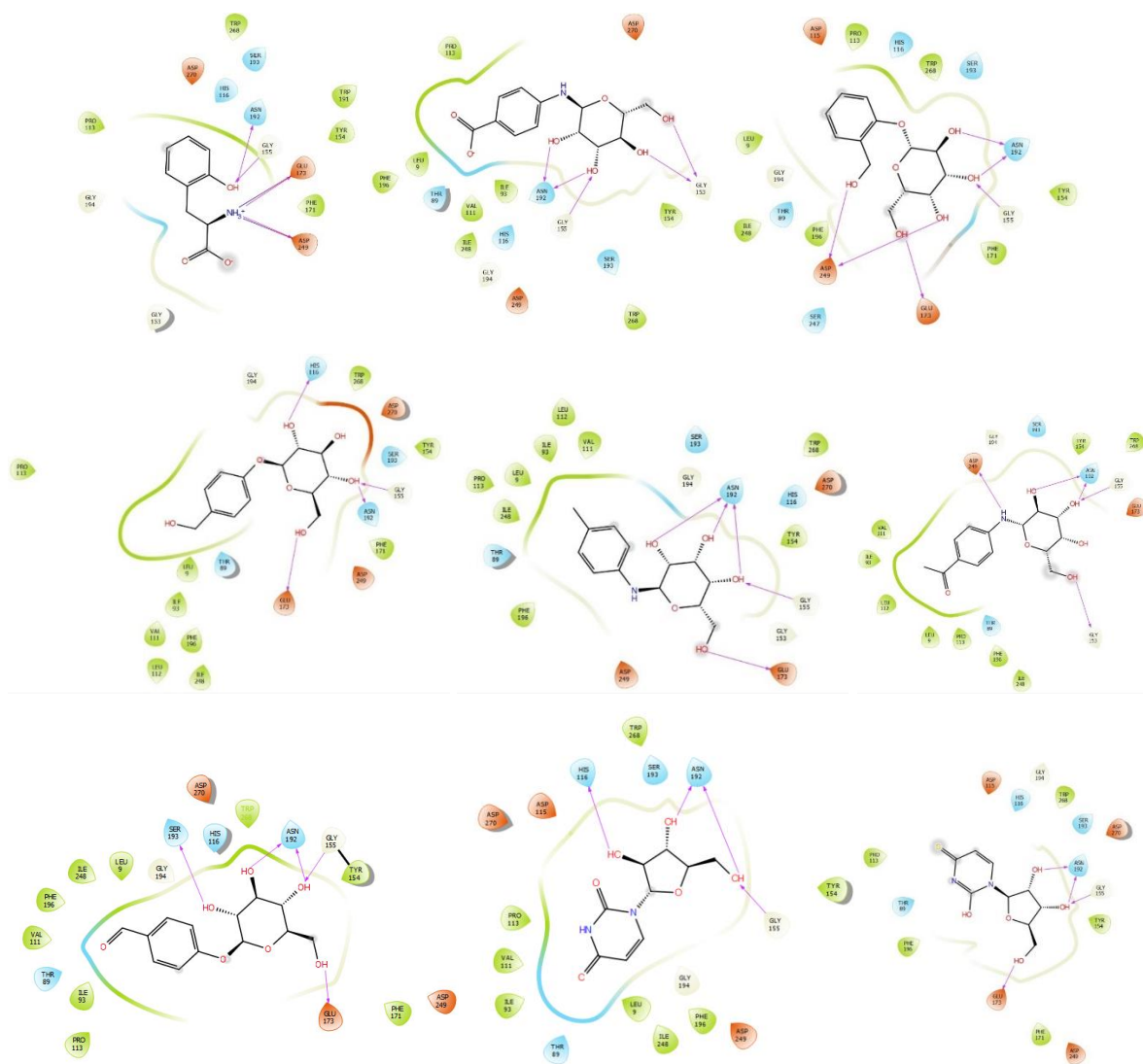
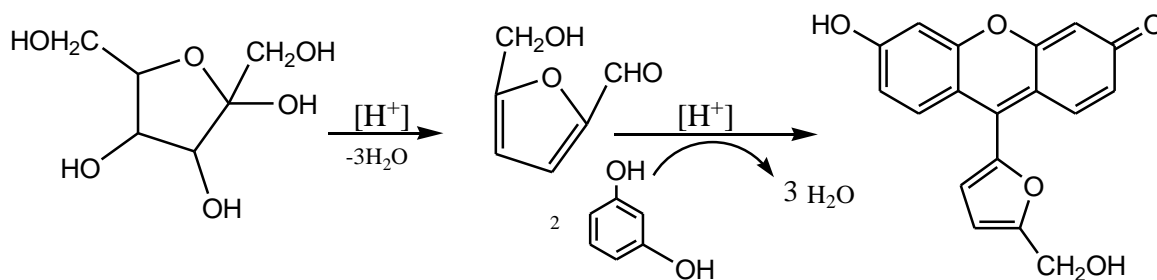
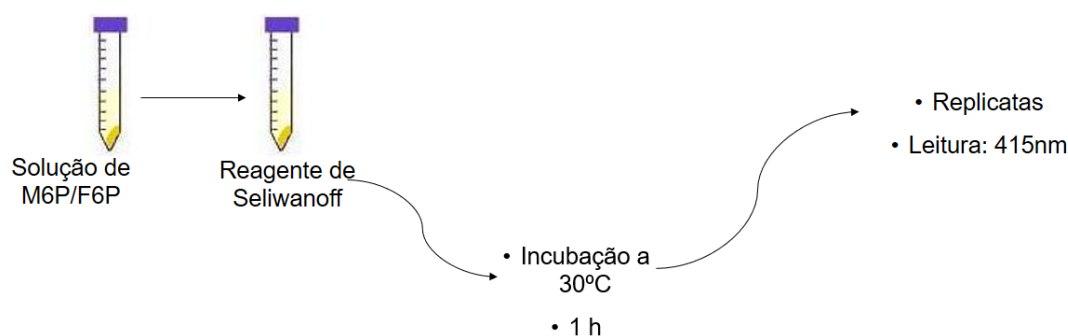


Figura 34 – Reação química do método de Seliwanoff.

Fonte: o autor

Considerando a disponibilidade do reagente, os testes realizados para a avaliação e inibição de PMI são de caráter exploratório. O primeiro teste realizado foi para verificar se a enzima PMI estava viável para os ensaios de inibição. Nos testes-controle foram realizados com a mesma quantidade (0,456 μ g) de frutose-6-fosfato (F6P) e manose-6-fosfato (Figura 80).

Figura 35 - Esquema do teste empregado na avaliação da atividade enzimática.

Fonte: o autor

Notou-se uma mudança rápida de coloração no controle com F6P, enquanto o controle com M6P e a reação (PMI + M6P) levaram em torno de 3 minutos para apresentarem alguma mudança em sua coloração. O primeiro tubo (aquele com a F6P) atingiu uma coloração alaranjada enquanto os outros dois apresentaram uma coloração rósea; porém, o tubo que continha a M6P e a enzima, apresentou uma coloração com intensidade maior do que a solução controle com M6P (Figura 36).

Figura 36 - Resultado para o ensaio de avaliação da atividade enzimática da PMI mostrando as colorações obtidas para os testes-controles e a reação entre a M6P e a enzima, após a adição do reagente de Seliwanoff e fervura por 5 minutos.



Fonte: o autor

A presença da coloração mais clara no tubo de controle da reação de PMI com M6P comparada ao controle com F6P é que nem toda M6P foi convertida em F6P pois é uma reação de equilíbrio (GRACY; NOTTMANN, 1968). Com esses resultados, é possível concluir que a enzima apresentava a sua atividade de interconversão da M6P em F6P, permitindo que os testes com os inibidores pudessem seguir.

4.4.2. Avaliação in vitro da inibição da PMI pelo composto selecionado

O primeiro composto selecionado para os ensaios de atividade com a enzima PMI, fornecida por Alexandrino, foi o **DL03874928** pois ele realiza ligações entre os seus grupos hidroxilas e os aminoácidos Gly155, Glu173, Asn192 e Asp249. Esse composto foi o primeiro selecionado pois realiza ligações com os aminoácidos Asn192 e Gly155 igual o ligante cristalográfico (G1P); além de apresentar interações com os aminoácidos Glu173 e Asp249 presente no modelo PMI *Xanthomonas citri* subsp *citri* gerado.

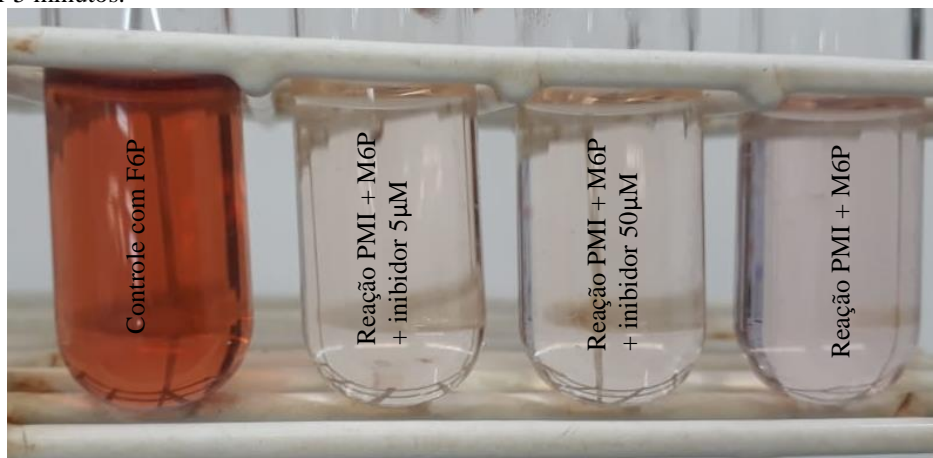
Foram testadas três concentrações finais para o inibidor **DL03874928**: 5 μ M, 50 μ M e 50mM (Tabela 13). As concentrações de 5 μ M e 50 μ M apresentaram uma coloração rósea, porém com menor intensidade do que a solução com F6P e a solução de enzima com M6P sem inibidor (Figura 37). Para essas duas concentrações do inibidor foram medidas as absorvâncias de 0,035 e 0,028, indicando que quanto maior a concentração do composto, maior é a inibição da enzima.

Tabela 13 - Dados referentes aos ensaios *in vitro* realizados para a fosfomanose isomerase recombinante de *Xanthomonas citri* subsp. *citri*.

REAGENTES	Volume (μL)					
	Controle com F6P	Controle com M6P	Reação com M6P e Concentrações finais do inibidor			
			sem inibidor	5 μM	50 μM	50 mM
M6P 0,3M	0	5	5	5	5	5
F6P 0,3M	5	0	0	0	0	0
Enzima	0	0	95	95	95	95
Inibidor 10 μM	0	0	0	100	0	0
Inibidor 100 μM	0	0	0	0	100	0
Inibidor 0,1 M	0	0	0	0	0	100
Tampão	195	195	100	0	0	0

(50 mM Tris-HCl pH 8,0, 10 mM NaCl)

Figura 37 - Resultado para o ensaio de avaliação da atividade enzimática da PMI mostrando as colorações obtidas para os testes-controles e a reação entre a M6P e a enzima, após a adição do reagente de Seliwanoff e fervura por 5 minutos.



Fonte: o autor.

Através de cálculos simples foi possível obter a absorbância relativa e a capacidade de inibição de cada concentração utilizada (Tabela 14). A absorbância relativa foi calculada pela divisão entre absorbância obtida para o respectivo inibidor e absorbância da solução M6P +

enzima. Os resultados obtidos para 5 μ M e 50 μ M foram, respectivamente, 0,54 e 0,43 (Tabela 12). Por sua vez, considerando que a absorbância na ausência de inibidor (M6P + enzima, valor de 0,065, Tabela 12) corresponde a 100% de atividade, a porcentagem de inibição relativamente ao 100% corresponde à diferença das absorbâncias apresentadas na ausência de inibidor (0,065) e na sua presença (0,035, por exemplo). Os resultados obtidos foram 46% e 57% para as concentrações de 5 μ M e 50 μ M respectivamente (Tabela 12).

Tabela 14 - Resultados obtidos para os ensaios de inibição da enzima PMI de *Xanthomonas citri* subsp. *citri* pelo inibidor *DL03874928*.

Concentração	Absorbância (415nm)	Absorbância relativa	Inibição (%)
F6P	0.902		
M6P + Enzima	0.065	1	0
5 μM do inibidor	0.035	0,54	46
50μM do inibidor	0.028	0,43	57

Fonte: o autor.

Por outro lado, o ensaio com o inibidor *DL03874928* a uma concentração de 50mM, apresentou problemas já que formou um aglomerado e não foi possível a sua detecção. Após esse fato, foi realizado um teste do inibidor com a enzima em presença do reagente de Seliwanoff e foi verificado, que em altas concentrações, o reagente utilizado no ensaio reage com o composto inibidor.

É possível notar que, embora a concentração do inibidor tenha sido aumentada em 10 vezes das concentrações de 5 μ M para 50 μ M, a inibição não ocorre de forma linear. Esse fato, aliado ao ensaio desse inibidor em altas concentrações, poderia indicar que o limite máximo de inibição foi atingido.

Em seguida a esse trabalho, todos os compostos comprados serão testados em relação a inibição da proteína PMI. Os resultados serão apresentados em triplicatas.

5. CONCLUSÃO

A estrutura tridimensional de um receptor é um dado de suma importância para químicos medicinais e computacionais que atuam na área de modelagem molecular. Explorar a região de interação do complexo ligante-receptor garante uma abordagem mais completa de estudos da área e amplia a gama de ferramentas e possibilidades a serem utilizadas.

O modelo de PMI de XAC construído se demonstrou confiável através das validações e da avaliação dos resíduos do sítio de interação. Dessa forma, tal modelo pode ser utilizado em triagens virtuais em bases de dados de compostos, na busca por novos quimiotipos ou classes químicas que poderiam interagir com o mesmo alvo, visando ao tratamento e eliminação do cancro cístico.

As estratégias de triagem virtual utilizadas foram validadas *in silico* e mostraram-se eficazes para a obtenção de diversas moléculas com potencial de inibição para a PMI. Um total de 45 compostos foram obtidos na etapa de triagem virtual, porém 12 foram comprados e passarão por ensaios biológicos para validar e quantificar suas atividades biológicas assim estimadas. Os resultados serão racionalizados do ponto de vista estrutura-atividade, com vistas a um posterior refinamento ou otimização bem como ao desenvolvimento de um futuro bactericida útil ao tratamento do cancro cístico.

6.PERSPECTIVAS FUTURAS

As estratégias de triagem virtual executadas nesse estudo foram efetivas do ponto de vista teórico. Como esse Mestrado faz parte de um projeto mais amplo, novos estudos serão realizados em relação ao potencial de inibição das outras moléculas selecionadas.

As moléculas que apresentarem maior potencial de inibição, seguirão para ensaios biológicos com a bactéria *XAC* em *Citrus aurantifolia*. Após a inoculação e passados os dias necessários, as folhas serão destacadas e fotografadas para constatação e comparação do processo infeccioso da *XAC*. Após a conclusão dessa etapa, os compostos que apresentarem propriedades desfavoráveis a novos defensivos agrícolas serão otimizados de modo a apresentarem um perfil mais adequado. O objetivo é que a otimização desses compostos, após extenso estudo de relação estrutura-atividade, leve desenvolvimento de um futuro bactericida útil ao tratamento do cancro cítrico.

REFERÊNCIAS BIBLIOGRÁFICAS

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

ARTIER, J. Análise proteômica diferencial da fração periplasmática de *Xanthomonas citri* subsp. *citri*: proteínas relacionadas com a indução da patogenicidade in vitro Juliana Artier Análise proteômica diferencial da fração periplasmática de *Xanthomonas citri* subsp. 2010.

ARTURSSON, P.; KARLSSON, J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (CACo2) cells. **Biochemical and Biophysical Research Communications**, v. 175, n. 3, p. 880–885, 1991.

BAI, F. et al. Bioactive conformational generation of small molecules: A comparative analysis between force-field and multiple empirical criteria based methods. **BMC Bioinformatics**, v. 11, n. November, 2010.

BAJORATH, J. Understanding chemoinformatics: A unifying approach. **Drug Discovery Today**, v. 9, n. 1, p. 13–14, 2004.

BEHLAU, F. et al. Incidência e severidade de cancro cítrico em laranja “Pêra Rio” sob condições de controle químico e proteção com quebra-vento. **Fitopatologia Brasileira**, v. 32, n. 4, p. 311–317, 2007.

BEHLAU, F. ; BELASQUE JR, J. **Cancro cítrico: A Doença e seu Controle**. 1 Ed. ed. Araraquara: Fundecitrus, 2014.

BEHLAU, F.; BARELLI, N. L.; BELASQUE, J. Lessons from a case of successful eradication of citrus canker in a citrus-producing farm in s?o paulo state, brazil. **Journal of Plant Pathology**, v. 96, n. 3, p. 561–568, 2014.

BEHLAU, F.; FONSECA, A. E.; BELASQUE, J. A comprehensive analysis of the Asiatic citrus canker eradication programme in São Paulo state, Brazil, from 1999 to 2009. **Plant Pathology**, 2016.

BERTRAND, H.; TRIBALLEAU, N. Virtual Screening Workflow Development Guided by the ROC Curve Approach . Application to High-Throughput *Docking* on Metabotropic Glutamate Receptor Subtype 4 Accepting the Uncertainty ? **Receptor**, p. 2534–2547, 2005.

BITTANCOURT, A. . O cancro cítrico. **O Biológico**, v. 23, p. 101–111, 1957.

BLACKSHIELDS, G. et al. Sequence embedding for fast construction of guide trees for multiple sequence alignment. **Algorithms for Molecular Biology**, v. 5, n. 1, p. 1–11, 2010.

BODNAR, A. M. et al. Complete genome sequences of three *Xanthomonas citri* strains from Texas. **Genome Announcements**, 2017.

BONAS, U. Hrp genes of phytopathogenic bacteria. In: **Current Topics in Microbiology and Immunology**. [s.l: s.n.]. v. 192p. 79–84.

BOOBIS, A. et al. In silico prediction of ADME and pharmacokinetics: Report of an expert meeting organised by COST B15. **European Journal of Pharmaceutical Sciences**, v. 17, n. 4–5, p. 183–193, 2002.

BOSTRÖM, J.; GREENWOOD, J. R.; GOTTFRIES, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. **Journal of Molecular Graphics and Modelling**, v. 21, n. 5, p. 449–462, 2003.

BOWIE, J. U.; LTCY, R.; EISENBERG, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. **Science**, v. 253, n. 58, p. 164–170, 1990.

BRENK, R. et al. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. **Journal of Medicinal Chemistry**, v. 46, n. 7, p. 1133–1143, 2003.

- BURA, E.; GASTWIRTH, J. L. The Binary Regression Quantile Plot: Assessing the Importance of Predictors in Binary Regression Visually. **Biometrical Journal**, v. 43, n. 1, p. 5–21, 2001.
- BURSULAYA, B. D. et al. Comparative study of several algorithms for flexible ligand *docking*. **Journal of Computer-Aided Molecular Design**, v. 17, n. 11, p. 755–763, 2003.
- BUTTNER, D. NEW EMBO MEMBER'S REVIEW: Getting across--bacterial type III effector proteins on their way to the plant cell. **The EMBO Journal**, v. 21, n. 20, p. 5313–5322, 2002.
- CARDOSO, J. E. Cancro Cítrico. **The Plant Health Instructor**, 2013.
- CARLSON, H. A.; MASUKAWA, K. M.; MCCAMMON, J. A. Method for including the dynamic fluctuations of a protein in computer-aided drug design. **Journal of Physical Chemistry A**, v. 103, n. 49, p. 10213–10219, 1999.
- CARNIELLI, C. M. et al. Xanthomonas citri subsp. citri surface proteome by 2D-DIGE: Ferric enterobactin receptor and other outer membrane proteins potentially involved in citric host interaction. **Journal of Proteomics**, v. 151, p. 251–263, 2017.
- CEPEA - CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA. **PIB-AGRO/CEPEA: Agronegócio Favorece Crescimento Do Pib Nacional E Ajuda No Controle Da Inflação.**
- CHEMBRIDGE. **The Glod standard in small molecule screening libraries and building blocks.**
- CHEN, X.; LIU, M.; GILSON, M. K. BindingDB: A web-accessible molecular recognition database. **J. Combi. Chem. High-Throughput Screen**, v. 4, p. 719–725, 2001.
- CLEASBY, A. et al. The x-ray crystal structure of phosphomannose isomerase from Candida albicans at 1.7 angstrom resolution. **Nature structural biology**, v. 3, n. 5, p. 470–9, 1996.
- CODDING, P. W. **Introduction to Structure-Based Drug Design.** Dordrecht: [s.n.].
- COLE, D. C.; BURSAVICH, M. G. **Aspartic Acid Proteases as Therapeutic Targets.** [s.l.] John Wiley & Sons, 2010.
- COLE, J. C. et al. Comparing protein-ligand *docking* programs is difficult. **Proteins: Structure, Function and Genetics**, v. 60, n. 3, p. 325–332, 2005.
- COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Science**, v. 2, p. 1511–1519, 1993.
- CORNELL, W. D. Recent evaluations of high throughput *docking* methods for pharmaceutical lead finding. **Annual Reports in Computational Chemistry**, v. 2, p. 297–323, 2006.
- COULIN, F. et al. Identification of Cys-150 in the Active Site of Phosphomannose Isomerase from Candida albicans. **Biochemistry**, v. 32, n. 51, p. 14139–14144, 1993.
- DA SILVA, A. C. R. et al. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. **Nature**, v. 417, n. 6887, p. 459–463, 2002a.
- DA SILVA, A. C. R. et al. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. **Nature**, v. 417, n. 6887, p. 459–463, maio 2002b.
- DA SILVA, C. H. T. DE P.; TAFT, C. A. 3D descriptors calculation and conformational search to investigate potential bioactive conformations, with application in 3D-QSAR and virtual screening in drug design. **Journal of Biomolecular Structure and Dynamics**, v. 35, n. 13, p. 2966–2974, 3 out. 2017.
- DANOS, E; BERGER, R. D.; STALL, R. E. **Temporal and Spatial Spread of Citrus Canker Within Groves**Phytopathology, 1984.

DASSAULT SYSTÈMES BIOVIA. **Discovery Studio Modeling Environment** San Diego, CA, USA Dassault Systèmes, , 2015.

EDDY, S. R. Accelerated Profile HMM Searches. **PLoS Computational Biology**, v. 7, n. 10, p. e1002195, 20 out. 2011.

EMPEREUR-MOT, C. et al. Predictiveness curves in virtual screening. **Journal of Cheminformatics**, v. 7, n. 1, p. 1–17, 2015.

EMPEREUR-MOT, C.; ZAGURY, J. F.; MONTES, M. Screening Explorer-An Interactive Tool for the Analysis of Screening Results. **Journal of Chemical Information and Modeling**, v. 56, n. 12, p. 2281–2286, 2016.

ESCALON, A. et al. Variations in type III effector repertoires, pathological phenotypes and host range of *Xanthomonas citri* pv. *citri* pathotypes. **Molecular Plant Pathology**, 2013.

ESWAR, N. et al. Comparative protein structure modeling using MODELLER. (Supplementary). **Current protocols in protein science / editorial board, John E. Coligan ... [et al.]**, v. Chapter 2, n. November, p. Unit 2.9, 2007.

ERENCE, C. M. et al. Recent advances in the understanding of *Xanthomonas citri* ssp. *citri* pathogenesis and citrus canker disease management. **Molecular Plant Pathology**, 2018.

FISER, A.; ŠALI, A. MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. **Methods in Enzymology**, v. 374, p. 461–491, 2003.

FOMBY, P.; CHERLIN, A. J. NIH Public Access. v. 72, n. 2, p. 181–204, 2011.

FRIESNER, R. A. . et al. Glide: A New Approach for Rapid, Accurate *Docking* and Scoring. Method and Assessment of *Docking* Accuracy. **Journal of Medicinal Chemistry**, v. 47, p. 1739–1749, 2004.

FUNDECITRUS. Novos Caminhos para o Cancro Citríco. **Revista Citricultor**, p. 16, 2016.

FUNDECITRUS. **Levantamento de Doenças dos Citros: HLB, CVC e Cancro Cítrico no Cinturão Citrícola de São Paulo e Triângulo/Sudoeste Mineiro**. [s.l: s.n.].

FUNDECITRUS. **Cancro Cítrico - Medidas Essenciais de Controle**. [s.l: s.n.].

FUNDECITRUS. Mitigação de Risco. **Revista Citricultor**2, p. 16, 2017c.

GAO, H.; YU, Y.; LEARY, J. A. Mechanism and kinetics of metalloenzyme phosphomannose isomerase: Measurement of dissociation constants and effect of zinc binding using ESI-FTICR mass spectrometry. **Analytical Chemistry**, v. 77, n. 17, p. 5596–5603, 2005.

GARAMI, A.; ILG, T. Disruption of mannose activation in *Leishmania mexicana*: GDP-mannose pyrophosphorylase is required for virulence, but not for viability. **EMBO Journal**, v. 20, n. 14, p. 3657–3666, 2001.

GOHLKE, H.; HENDLICH, M.; KLEBE, G. Knowledge-based scoring function to predict protein-ligand interactions. **Journal of Molecular Biology**, v. 295, n. 2, p. 337–356, 2000.

GORDON, J. L. et al. Comparative genomics of 43 strains of *Xanthomonas citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host ranges. **BMC Genomics**, 2015.

GOTTWALD, T.R., MCGUIRE, R.G. & GARRAN, S. Asiatic citrus canker: spatial and temporal spread in simulated new planting situations in Argentina. **Phytopathology**, v. 78, p. 739–745, 1988.

GOTTWALD, TIM R.; GRAHAM, J. . **Cancro Cítrico**. Disponível em: <<https://www.apsnet.org/edcenter/disandpath/prokaryote/pdlessons/Pages/CitrusCankerPort.aspx>>.

GOTTWALD, T. R. et al. The Citrus Canker Epidemic in Florida: The Scientific Basis of Regulatory Eradication Policy for an Invasive Species. **Phytopathology**, v. 91, n. 1, p. 30–34, 2001.

GRACY, R. W. .; NOTTMANN, E. A. Studies on Phosphomannose Isomerase - A mechanism for catalysis and for the role of zinc in the enzymatic and the nonenzymatic isomerization. **The Journal of Biological Chemistry**, v. 243, n. 20, p. 5410–5419, 1968.

GRAHAM, J. H. et al. *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradication of citrus canker. **Molecular and Plant Pathology**, v. 5, n. 1, p. 1–15, 2004.

HAIGH, J. A. et al. Small Molecule Shape-Fingerprints. n. 1, p. 673–684, 2005.

HAWKINS, P. C. D. et al. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. **Journal of Chemical Information and Modeling**, v. 50, n. 4, p. 572–584, 2010.

HAWKINS, P. C. D.; NICHOLLS, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. **Journal of Chemical Information and Modeling**, v. 52, n. 11, p. 2919–2936, 2012.

HORSTER, M. F.; STOPP, M. Transport and metabolic functions in cultured renal tubule cells. **Kidney International**, v. 29, n. 1, p. 46–53, 1986.

IRVINE, J. D. . et al. MDCK (Madin-Darby Canine Kidney) Cells: A Tool for Membrane Permeability Screening. v. 88, n. 1, p. 28–33, 1999.

IRWIN, J. J. .; SHOICHET, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. **Journal of Chemical Information and Modeling**, v. 45, p. 177–182, 2005.

JAEKEN, J.; MATTHIJS, G. Congenital disorders of glycosylation. **Annu Rev Genomics Hum Genet**, n. 2, p. 129–151, 2001.

JAIN, A. N. Scoring Functions for Protein-Ligand *Docking*. **Current Protein & Peptide Science**, v. 7, n. 5, p. 407–420, 2006.

JALAN, N. et al. Comparative genomic analysis of *Xanthomonas axonopodis* pv. *citrumelo* F1, which causes citrus bacterial spot disease, and related strains provides insights into virulence and host specificity. **Journal of Bacteriology**, 2011.

JALAN, N. et al. Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. **BMC Genomics**, 2013.

JENSEN, S. O.; REEVES, P. R. Domain organisation in phosphomannose isomerases (types I and II). **Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology**, v. 1382, n. 1, p. 5–7, 1998.

JORGENSEN, W. L. **The Many Roles of Computation in Drug Discovery** Science, 2004.

JORGENSEN, W. L. Efficient drug lead discovery and optimization. **Accounts of Chemical Research**, 2009.

KIRCHMAIR, J. et al. The Protein Data Bank (PDB), its related services and software tools as key components for in silico guided drug discovery. **Journal of medicinal chemistry**, v. 51, n. 22, p. 7021–40, nov. 2008.

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. 935–949, 2004.

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, 2004.

KRAMER, B.; RAREY, M.; LENGAUER, T. Evaluation of the FlexX incremental construction algorithm for protein- ligand *docking*. **Proteins: Structure, Function and Genetics**, v. 37, n. 2, p. 228–241, 1999.

KUNZ, I. D. Structure-based strategies for drug design and discovery. **Science**, v. 257, n. 5470, p. 1078–1082,

1992.

LAPLANTE, S. R. et al. Ligand bioactive conformation plays a critical role in the design of drugs that target the hepatitis C virus NS3 protease. **Journal of Medicinal Chemistry**, v. 57, n. 5, p. 1777–1789, 2014.

LASKOWSKI, R. A. et al. Procheck Programs to check the Stereochemical Quality of Protein Structures. **Journal of Applied Crystallography**, v. 26, p. 283–291, 1993.

LIPINSKI, C. A. Lead- and drug-like compounds: The rule-of-five revolution. **Drug Discovery Today: Technologies**, v. 1, n. 4, p. 337–341, 2004.

LIPINSKI, C.; HOPKINS, A. Navigating chemical space for biology and medicine. **Nature**, v. 432, n. 7019, p. 855–861, 2004.

LUTHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v. 83–85, 1992.

MARCHANT, C. A. ; BRIGGS, K. A. ; LONG, A. In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic. **Toxicology Mechanism and Methods**, v. 18, n. 2–3, p. 177–187, 2008.

MARTÍ-RENOM, M. A. et al. Comparative protein structure modeling of genes and genomes. v. 29, p. 291–325, 2000.

MAYBRIDGE. **The Maybridge Screening Collection**. Disponível em: <<http://www.maybridge.com>>.

MOREIRA, L. M. et al. Comparative Genomics Analyses of Citrus-Associated Bacteria. **Annual Review of Phytopathology**, v. 42, n. 1, p. 163–184, 2004.

MOREIRA, L. M. et al. Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. *aurantifolii*. **BMC Genomics**, v. 11, n. 1, 2010.

MUEGGE, I. ; ENYEDY, I. J. Virtual screening for kinase targets. **Current Medicinal Chemistry**, v. 11, n. 6, p. 693–707, 2004.

NAYLOR, E. et al. Identification of a chemical probe for NAADP by virtual screening. **Nature Chemical Biology**, v. 5, n. 4, p. 220–226, 2009.

NCBI-National Center for Biotechnology Information. Disponível em: <<https://www.ncbi.nlm.nih.gov/protein/AAM38423.1>>.

NEVES, M. F. ; TROMBIN, V. G. **Anuário da Citricultura 2017**Citrus Br. São Paulo: [s.n.].

NISSINK, J. W. M. et al. Proteins: Structure, Function and Genetics. **Proteins**, v. 9, p. 457–471, 2002.

NOLTMANN, A. on Phosphomannose. n. 20, 1968.

OPREA, T. I. **Introduction to Chemoinformatics in Drug Discovery**. Weinheim: Wiley-VCH, 2005.

PAPOUTSOPOULOU, S. V; KYRIAKIDIS, D. A. Phosphomannose isomerase of *Xanthomonas campestris*: a zinc activated enzyme. **Molecular and cellular biochemistry**, v. 177, n. 1, p. 183–191, 1997.

PATTERSON, J. H. et al. Mannose metabolism is required for mycobacterial growth. **Biochemical Journal**, v. 372, n. 1, p. 77–86, 2003.

PAYTON, M. A. et al. A novel *Saccharomyces cerevisiae* secretory mutant possesses a thermolabile phosphomannose isomerase. **Journal of Bacteriology**, v. 173, n. 6, p. 2006–2010, 1991.

PEITSCH, M. C. Manuel Peitsch discusses knowledge management and informatics in drug discovery. **Drug Discovery Today: BIOSILICO**, v. 2, n. 3, p. 94–96, 2004.

- PICCIRILLO, E.; DO AMARAL, A. T. Virtual screening of bioactive compounds: Concepts and applications. **Química Nova**, v. 41, n. 6, p. 662–677, 2018.
- PONTIUS, J.; RICHELLE, J.; WODAK, S. J. Deviations from Standard Atomic Volumes as a Quality Measure for Protein Crystal Structures. **Journal of Molecular Biology**, v. 264, n. 121–136, 1996.
- PRENTIS, R.; LIS, Y.; WALKER, S. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985). **British Journal of Clinical Pharmacology**, v. 25, n. 3, p. 387–396, 1988.
- PROUDFOOT, A. E. I. et al. Purification cDNA cloning and heterologous expression of human phosphomannose isomerase. **European Journal of Biochemistry**, v. 219, n. 1–2, p. 415–423, 1994.
- PROUDFOOT, A. E. I.; PAYTON, M. A.; WELLS, T. N. C. Purification and characterization of fungal and mammalian phosphomannose isomerases. **Journal of Protein Chemistry**, v. 13, n. 7, p. 619–627, 1994.
- RICHARD, D. et al. Complete genome sequences of six copperresistant *Xanthomonas citri* pv. *citri* strains causing Asiatic citrus canker, obtained using long-read technology. **Genome Announcements**, 2017.
- RODRIGUEZ, R.; VRIEND, G. Professional Gambling. In: **Biomolecular Structure and Dynamics**. Dordrecht: Springer Netherlands, 1997. p. 79–119.
- ROSE, IRWIN A.; O'CONNELL, EDWARD L.; SCHRAY, K. J. Mannose 6-Phosphate : Anomeric Isomerase and Its LEpimerization Isomerase * Form Used by Phosphomannose by w Phosphoglucose. **The Journal of Biological Chemistry**, v. 248, n. 6, p. 2232–2234, 1973.
- ROSSETTI, V. . **Manual Ilustrado de Doenças dos Citros**. [s.l: s.n.].
- ROUX, C. et al. Inhibition of Type I and Type II Phosphomannose Isomerases by the Reaction Intermediate Analogue 5-Phospho-D-Arabinonohydroxamic Acid Supports a Catalytic Role for the Metal Cofactor. **Biochemistry**, v. 43, n. 10, p. 2926–2934, 2004.
- SANCHEZ, R.; SAIL, A. Advances in comparative protein-structure modelling. p. 206–214, [s.d.].
- SANCHO, J. **Estructura de Macromoléculas**. Disponível em: <<http://www.bifi.es/~jsancho/estructuramacromoleculas/3peptidicoyES/3peptidicoyES.htm>>. Acesso em: 31 dez. 2019.
- SANTOS FILHO, O. A.; BICCA DE ALENCASTRO, R. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253–259, 2003.
- SCHRODINGER. **Maestro**New York, 2018.
- SCHRODINGER, L. **QikProp: Rapid ADME predictions of drug candidates**. Disponível em: <<https://www.schrodinger.com/qikprop>>.
- SCHUBERT, T. **No Title**Florida Department of Agriculture and Consumer Services, 2011.
- SCHWEDE, T. et al. SWISS-MODEL: An automated protein homology-modeling server. **Nucleic Acids Research**, v. 31, n. 13, p. 3381–3385, 2003.
- SEILLE, R.; CHRISTIANO, C. Universidade de São Paulo Escola superior de Agricultura “Luiz de Queiroz” Cancro cítrico (. 2006.
- SHAHIDULLAH, M; KHORASANI, S. S. M. A. The sensitivity and selectivity of the Seliwanoff test for frutose. **Analytica Chimica Acta**, v. 61, p. 317–319, 1972.
- SHERIDAN, R. P. . et al. Protocols for bridging the peptide to nonpeptide gap in topological similarity searches. **Journal of Chemical Information and Computer Sciences**, v. 41, n. 5, p. 1395–1406, 2001.
- SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal

Omega. **Molecular Systems Biology**, v. 7, n. 539, p. 1–6, 2011.

SIGDEL, S. et al. Characterization of a Mannose-6-Phosphate Isomerase from *Bacillus amyloliquefaciens* and Its Application in Fructose-6-Phosphate Production. **PLoS ONE**, v. 10, n. 7, p. 1–15, 2015.

STALL, R. E.; SEYMOUR, C. P. Canker, a Threat to Citrus in the Gulf-Coast States. **American Phytopathological Society**, v. 67, n. 5, p. 581–585, 1983.

STRYER, L.; BERG, J.M., TYMOCZKO, J. **Biochemistry**. 6^o ed ed. New York: [s.n.].

SUITE, G. **GOLD User Guide A Component of the GOLD Suite**, 2015. Disponível em: <<https://www.ccdc.cam.ac.uk>>

TRUCHON, J. F.; BAYLY, C. I. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. **Journal of Chemical Information and Modeling**, v. 47, n. 2, p. 488–508, 2007.

VENKATESH, S.; LIPPER, R. A. Role of the development scientist in compound lead selection and optimization. **Journal of Pharmaceutical Sciences**, v. 89, n. 2, p. 145–154, 2000.

VIETH, M. et al. Assessing energy functions for flexible *docking*. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1612–1622, 1998.

VRIEND, G. WHAT IF: a molecular modeling and drug design program. **Journal of Molecular Graphics**, v. 8, n. 1, p. 52–6, 1990.

WELLS, T. N. C. et al. Mechanism of Irreversible Inactivation of Phosphomannose Isomerases by Silver Ions and Flamazine. **Biochemistry**, v. 34, n. 24, p. 7896–7903, 1995.

WELLS, T. N. C.; PAYTON, M. A.; PROUDFOOT, A. E. I. Inhibition of Phosphomannose Isomerase by Mercury Ions. **Biochemistry**, v. 33, n. 24, p. 7641–7646, 1994.

WELLS, T. N. C.; SCULLY, P.; MAGNENAT, E. Arginine 304 Is an Active Site Residue in Phosphomannose Isomerase from *Candida albicans*. **Biochemistry**, v. 33, n. 19, p. 5777–5782, 1994.

WERMUTH, C. G. **The Practice of Medicinal Chemistry**. [s.l.] Elsevier, 2008.

YACOUBI, B. EL. **BACTERIAL CITRUS CANKER: MOLECULAR ASPECTS OF A COMPATIBLE PLANT-MICROBE INTERACTION**. [s.l.: s.n.].

YURIEV, E.; AGOSTINO, M.; RAMSLAND, P. A. Challenges and advances in computational *docking*: 2009 in review. **Journal of Molecular Recognition**, v. 24, n. 2, p. 149–164, 2011.

YURIEV, E.; RAMSLAND, P. A. Latest developments in molecular *docking*: 2010–2011 in review. **Journal of Molecular Recognition**, v. 26, n. 5, p. 215–239, 2013.

ZHANG, Y. et al. Positive selection is the main driving force for evolution of citrus canker-causing *Xanthomonas*. **ISME Journal**, 2015.

ZHAO, W. et al. A statistical framework to evaluate virtual screening. **BMC Bioinformatics**, v. 10, p. 1–13, 2009.

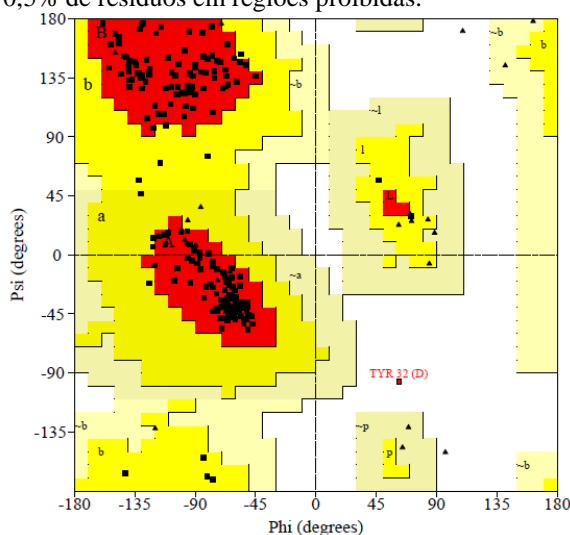
ZUCCOTTI, S. et al. Kinetic and Crystallographic Analyses Support a Sequential-Ordered Bi Bi Catalytic Mechanism for *Escherichia Coli* Glucose-1-Phosphate Thymidyltransferase. **J.Mol.Biol.**, v. 313, p. 831, 2001.

APÊNDICES

APÊNDICE A - Dados relacionados a construção do modelo de PMI *Xanthomonas citri* subsp. *citri*.

Nesse apêndice são apresentados os dados obtidos em relação as estruturas cristalográficas 1H5R, 2CU2, 2X5S e 2QH5; além dos outros 4 modelos preditos por modelagem por homologia estrutural. Os dados obtidos são os gráficos de Ramachandran, **PROCHECK**, **VERIFY3D**, **ERRAT** e **PROVE** com os respectivos valores encontrados para cada estrutura proteica.

Figura A1 - Diagrama de Ramachandran da estrutura cristalográfica 1H5R obtido pelo **PROCHECK**: 93% de resíduos em regiões favoráveis, 6,5% de resíduos em regiões permitidas, 0% de resíduos em regiões generosamente permitidas e 0,5% de resíduos em regiões proibidas.



Estatísticas de plotagem

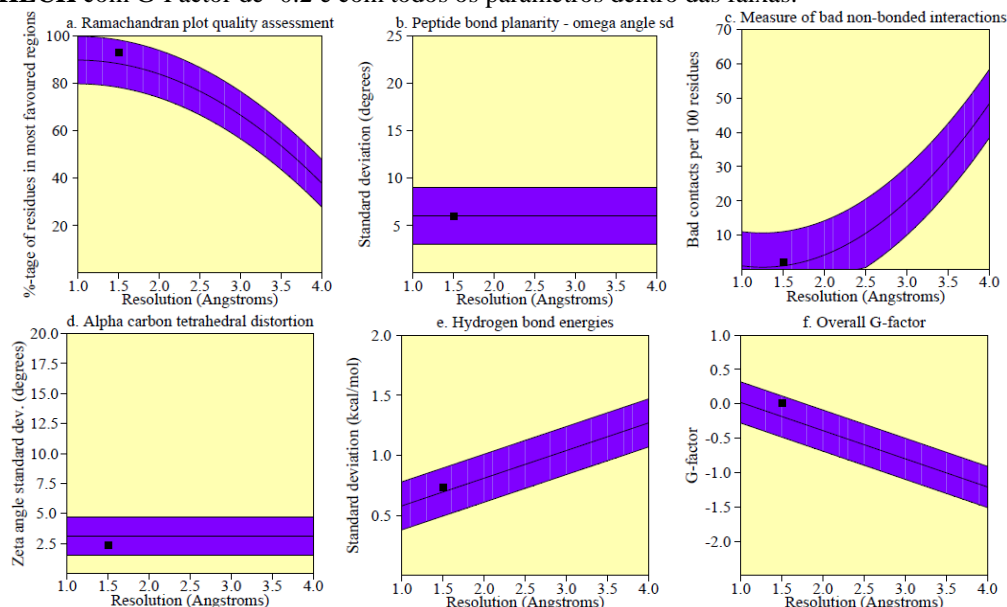
Resíduos em regiões mais favoráveis [A,B,L]	185	93.0%
Resíduos em regiões adicionais permitidas [a,b,l,p]	13	6.5%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	0	0.0%
Resíduos em regiões proibidas	1	0.5%
	----	----
Número de resíduos não glicinas e não prolinas	199	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	23	
Número de resíduos de prolina	13	

Número total de resíduos	237	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

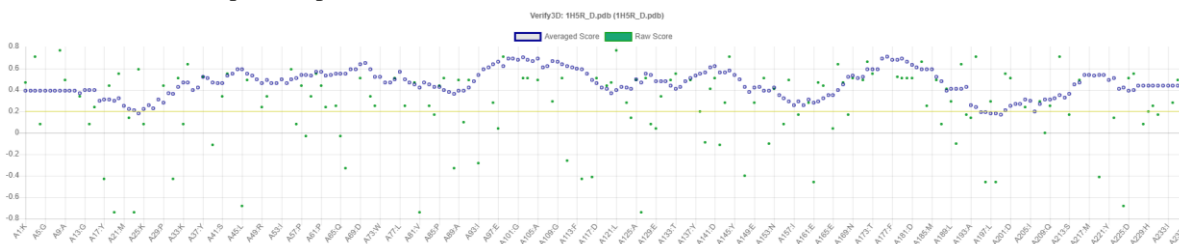
Fonte: O autor

Figura A2 - Resumo da avaliação estereoquímica da estrutura cristalográfica 1H5R realizado pelo PROCHECK com G-Factor de -0.2 e com todos os parâmetros dentro das faixas.



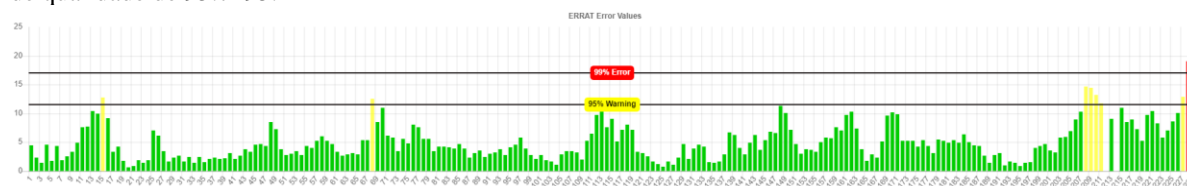
Fonte: O autor

Figura A3 - Gráfico para a estrutura cristalográfica 1H5R obtido pelo VERIFY3D: 97,47% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

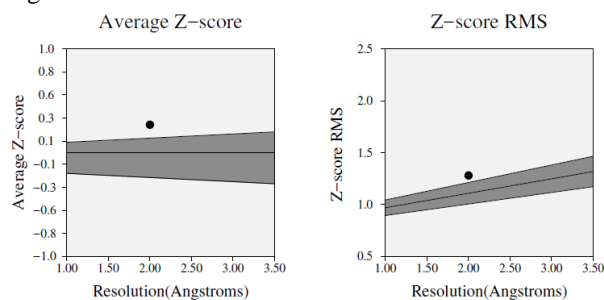
Figura A4 - Gráfico para a estrutura cristalográfica 1H5R obtido pelo ERRAT. Essa estrutura possui um fator de qualidade de 95.7295.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

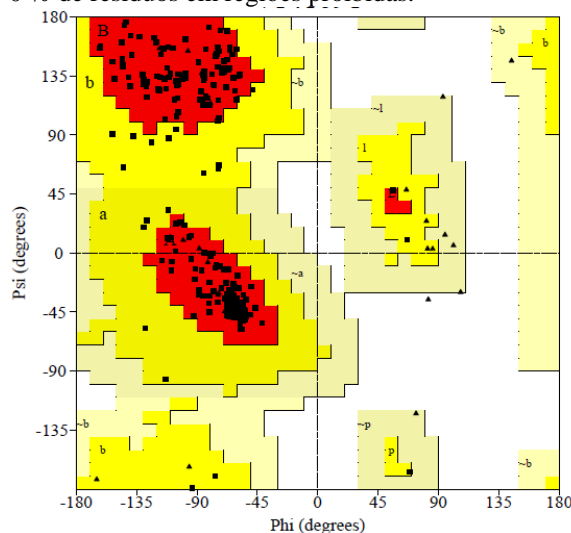
Figura A5 - Gráfico para a estrutura cristalográfica 1H5R obtido pelo **PROVE**. Essa estrutura apresenta 2,6% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A6 - Diagrama de Ramachandram da estrutura cristalográfica 2CU2 obtido pelo **PROCHECK**: 91,9% de resíduos em regiões favoráveis, 8,1% de resíduos em regiões permitidas, 0% de resíduos em regiões generosamente permitidas e 0 % de resíduos em regiões proibidas.



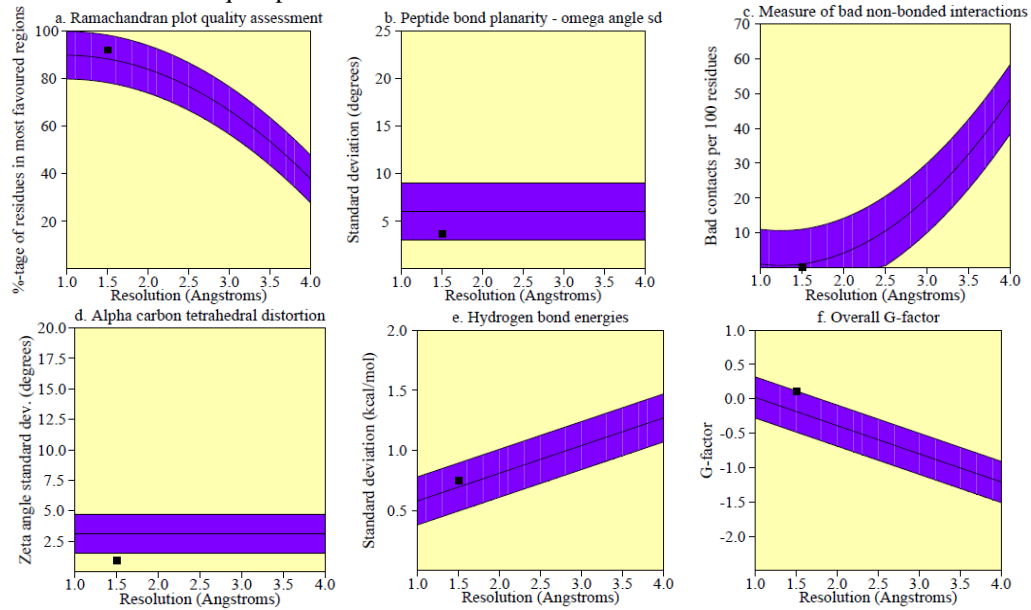
Estadísticas de plotagem

Resíduos em regiões mais favoráveis [A,B,L]	204	91.9%
Resíduos em regiões adicionais permitidas [a,b,l,p]	18	8.1%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	0	0.0%
Resíduos em regiões proibidas	0	0.0%
Número de resíduos não glicinas e não prolina	222	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	25	
Número de resíduos de prolina	18	
Número total de resíduos	267	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

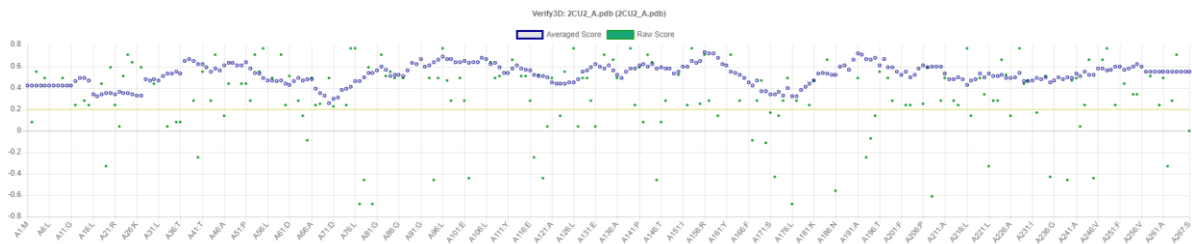
Fonte: O autor

Figura A7 - Resumo da avaliação estereoquímica da estrutura cristalográfica 2CU2 realizado pelo **PROCHECK** com G-Factor de 0.1 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta resultados melhores com um valor de 0.9.



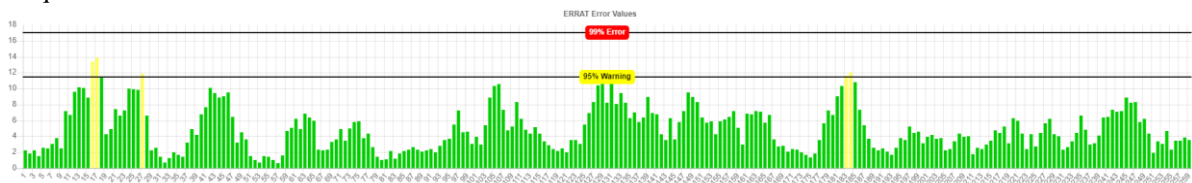
Fonte: O autor

Figura A8 - Gráfico para a estrutura cristalográfica 2CU2 obtido pelo **VERIFY3D**: 100% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

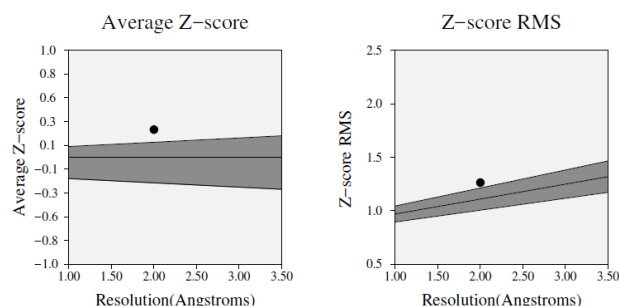
Figura A9 - Gráfico para a estrutura cristalográfica 2CU2 obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 98.0695.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

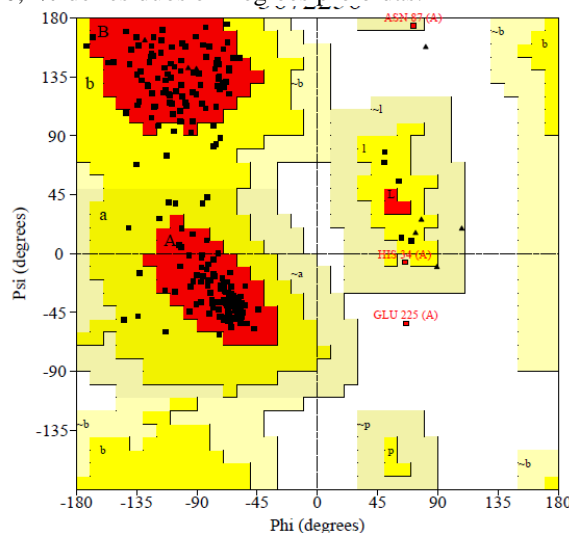
Figura A10 - Gráfico para a estrutura cristalográfica 2CU2 obtido pelo **PROVE**. Essa estrutura apresenta 1,9% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A11 - Diagrama de Ramachandram da estrutura cristalográfica 2QH5 obtido pelo **PROCHECK**: 87,6% de resíduos em regiões favoráveis, 11,2% de resíduos em regiões permitidas, 0,9% de resíduos em regiões generosamente permitidas e 0,4% de resíduos em regiões proibidas.



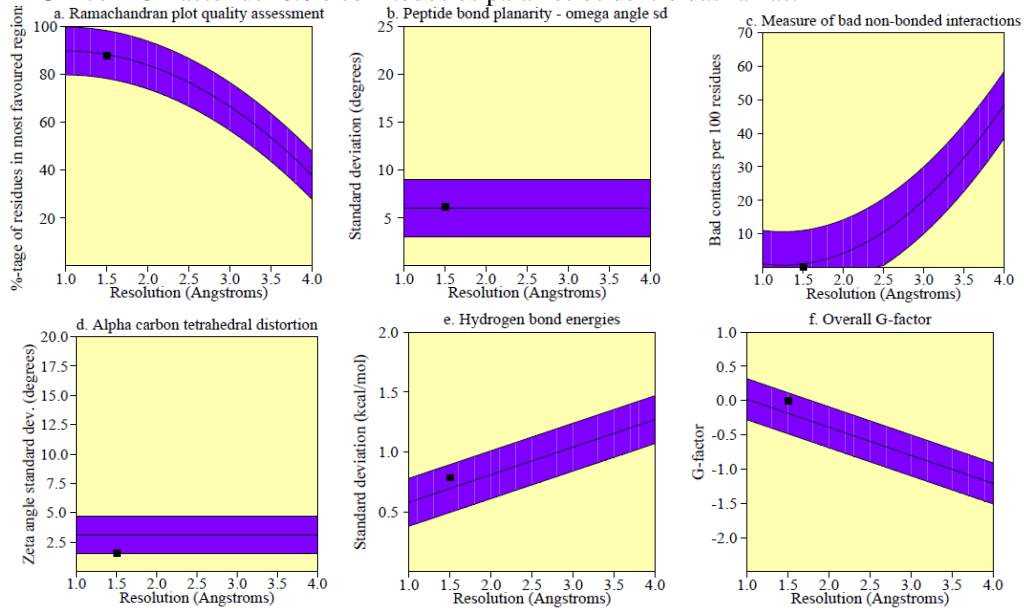
Estadísticas de plotagem

Resíduos em regiões mais favoráveis [A,B,L]	204	87.6%
Resíduos em regiões adicionais permitidas [a,b,l,p]	26	11.2%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	2	0.9%
Resíduos em regiões proibidas	1	0.4%
Número de resíduos não glicinas e não prolina	233	100.0%
Número de resíduos finais (exc. Glicina e prolina)	3	
Número de resíduos de glicina (mostrados como triângulos)	11	
Número de resíduos de prolina	6	
Número total de resíduos	253	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

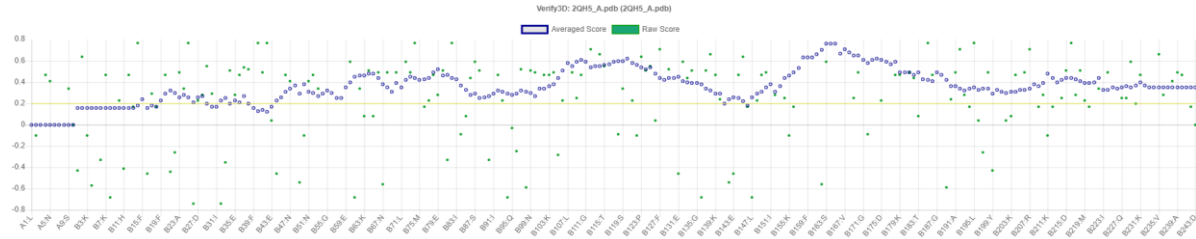
Fonte: O autor

Figura A12 - Resumo da avaliação estereoquímica da estrutura cristalográfica 2QH5 realizado pelo PROCHECK com G-Factor de -0.0 e com todos os parâmetros dentro das faixas.



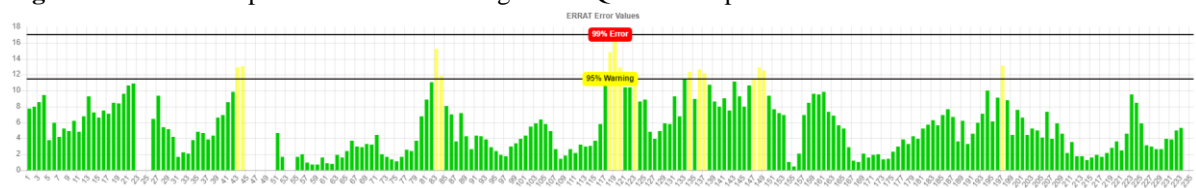
Fonte: O autor

Figura A13 - Gráfico para a estrutura cristalográfica 2QH5 obtido pelo VERIFY3D: 86,17% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

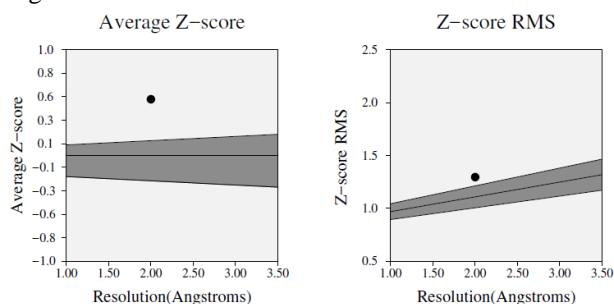
Figura A14 - Gráfico para a estrutura cristalográfica 2QH5 obtido pelo ERRAT.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

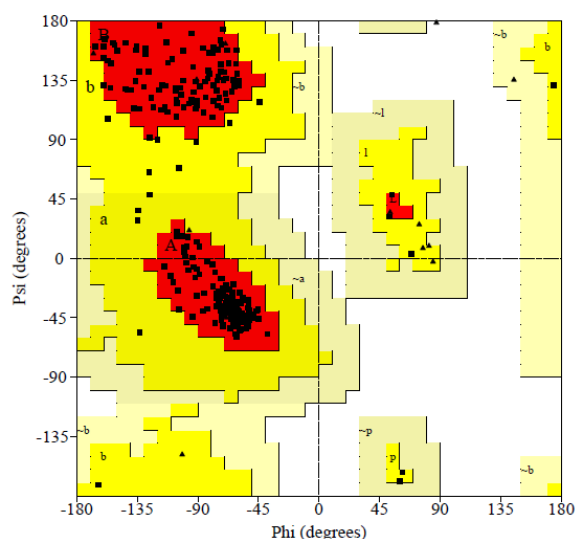
Figura A15 - Gráfico para a estrutura cristalográfica 2QH5 obtido pelo **PROVE**. Essa estrutura apresenta 3,2% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A16 - Diagrama de Ramachandram da estrutura cristalográfica 2X5S obtido pelo **PROCHECK**: 93,7% de resíduos em regiões favoráveis, 6,3% de resíduos em regiões permitidas, 0,0% de resíduos em regiões generosamente permitidas e 0,0% de resíduos em regiões proibidas.



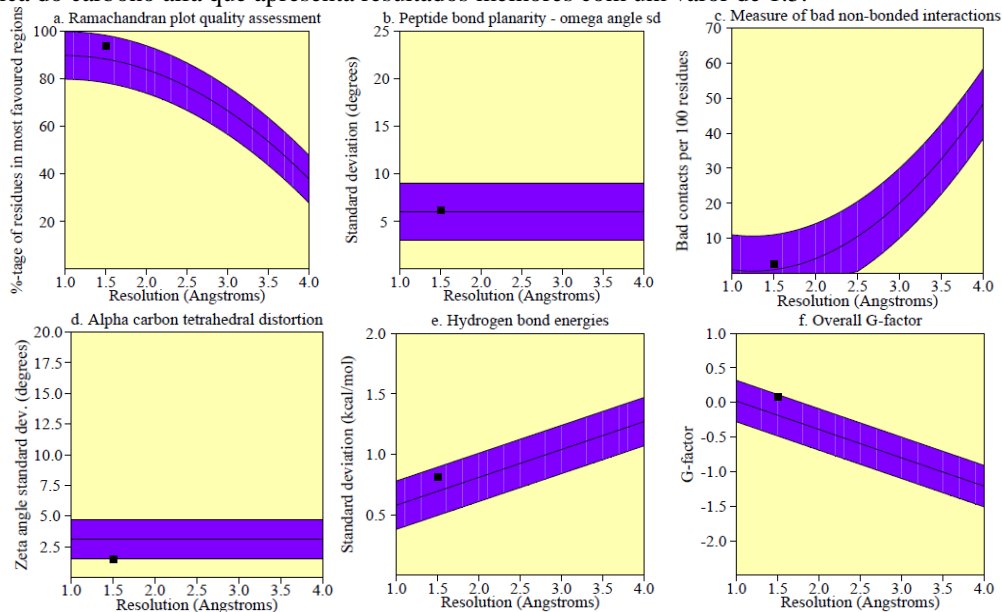
Estadísticas do gráfico

Resíduos em regiões mais favoráveis [A,B,L]	224	93.7%
Resíduos em regiões adicionais permitidas [a,b,l,p]	15	6.3%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	0	0.0%
Resíduos em regiões proibidas	0	0.0%
Número de resíduos não glicinas e não prolina	239	100.0%
Número de resíduos finais (exc. Glicina e prolina)	1	
Número de resíduos de glicina (mostrados como triângulos)	15	
Número de resíduos de prolina	17	
Número total de resíduos	272	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

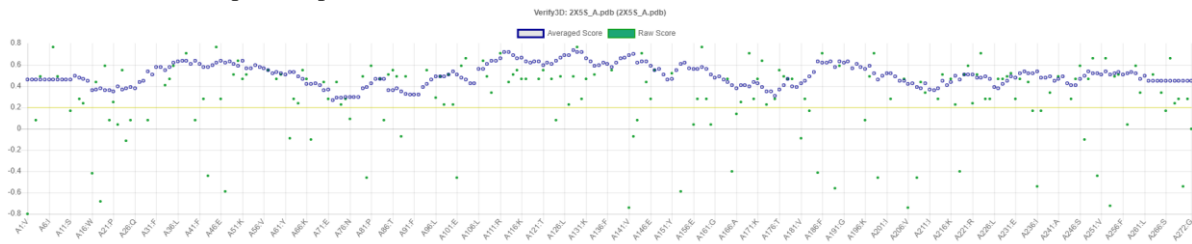
Fonte: O autor

Figura A17 - Resumo da avaliação estereoquímica da estrutura cristalográfica 2X5S realizado pelo **PROCHECK** com G-Factor de 0.1 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta resultados melhores com um valor de 1.5.



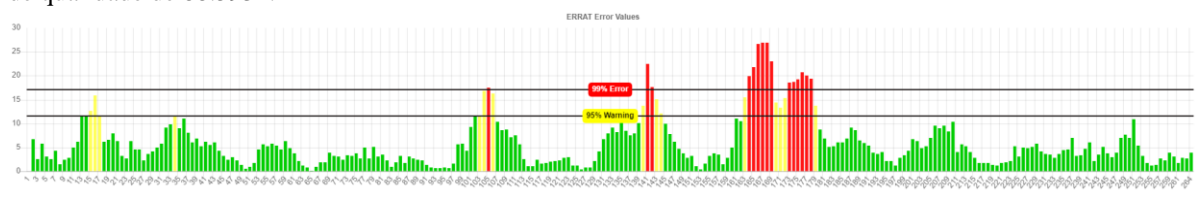
Fonte: O autor

Figura A18 - Gráfico para a estrutura cristalográfica 2X5S obtido pelo **VERIFY3D**: 100% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

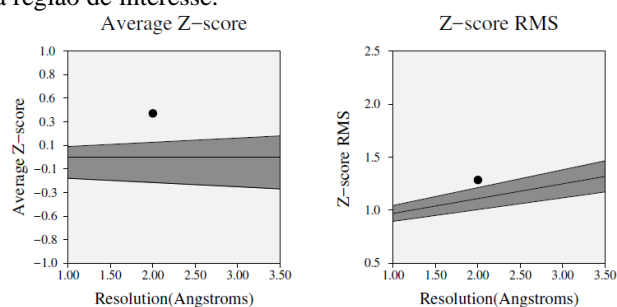
Figura A19 - Gráfico para a estrutura cristalográfica 2X5S obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 88.5932.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

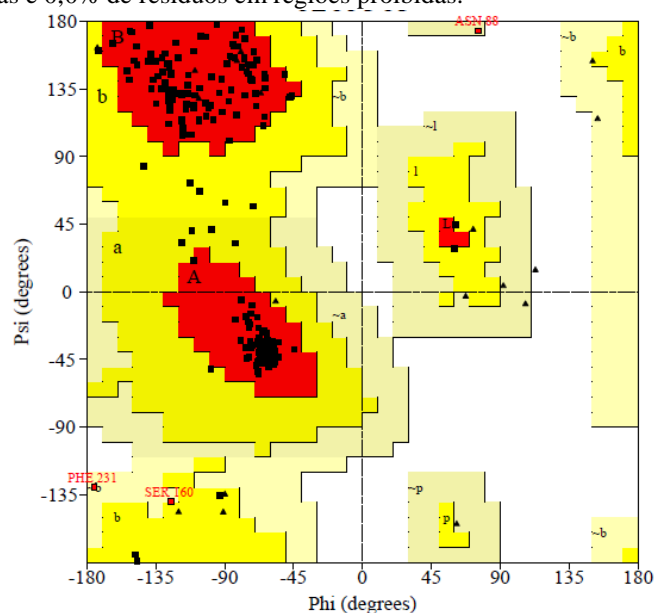
Figura A20 - Gráfico para a estrutura cristalográfica 2X5S obtido pelo **PROVE**. Essa estrutura apresenta 2,8% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A21 - Diagrama de Ramachandram da hipótese 1 gerada para PMI XAC obtido pelo **PROCHECK**: 91,6% de resíduos em regiões favoráveis, 7,1% de resíduos em regiões permitidas, 1,3% de resíduos em regiões generosamente permitidas e 0,0% de resíduos em regiões proibidas.



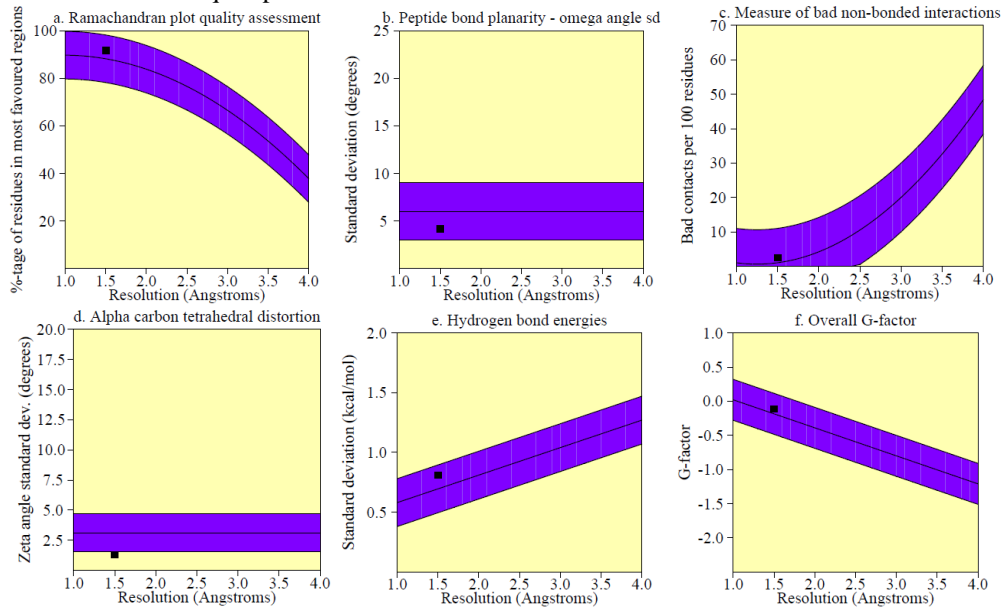
Estadísticas do gráfico

Resíduos em regiões mais favoráveis [A,B,L]	219	91.6%
Resíduos em regiões adicionais permitidas [a,b,l,p]	17	7.1%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	3	1.3%
Resíduos em regiões proibidas	0	0.0%
Número de resíduos não glicinas e não prolina	239	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	19	
Número de resíduos de prolina	17	
Número total de resíduos	267	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

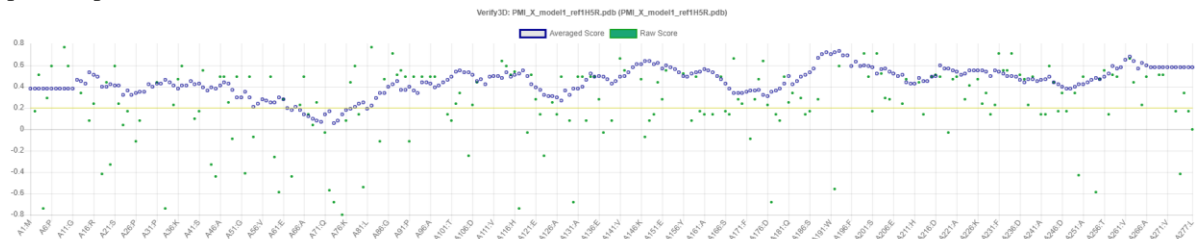
Fonte: O autor

Figura A22 - Resumo da avaliação estereoquímica da hipótese 1 gerada para PMI XAC realizado pelo **PROCHECK** com G-Factor de -0.1 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta resultados melhores com um valor de 1.3.



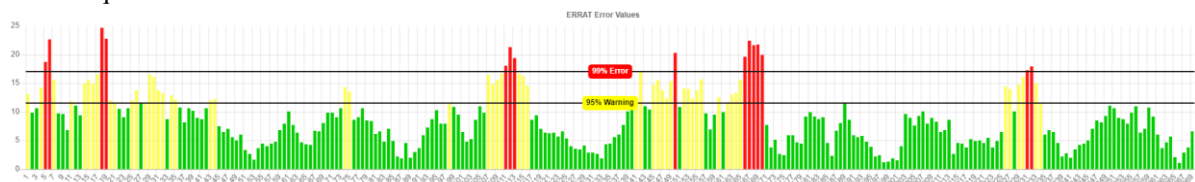
Fonte: O autor

Figura A23 - Gráfico para a hipótese 1 obtido pelo **VERIFY3D**: 94,58% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

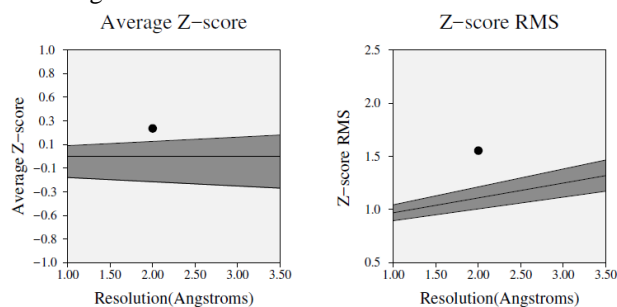
Figura A24 - Gráfico para a hipótese 1 gerada para PMI XAC obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 74.3494.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

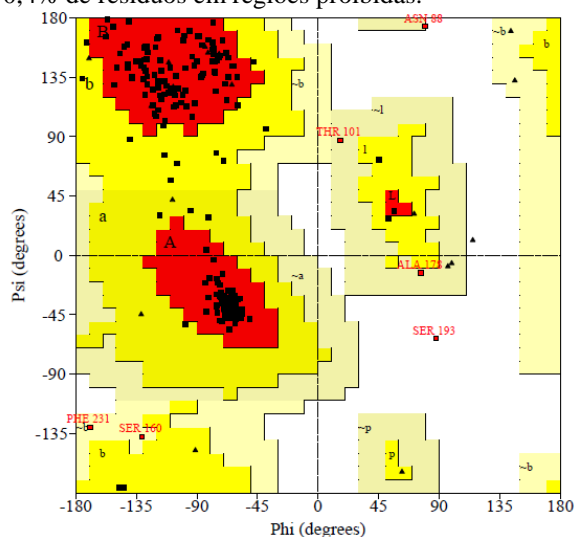
Figura A25 - Gráfico para hipótese 1 gerada para PMI XAC obtido pelo **PROVE**. Essa estrutura apresenta 4,9% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A26 - Diagrama de Ramachandram da hipótese 2 gerada para PMI XAC obtido pelo **PROCHECK**: 89,5% de resíduos em regiões favoráveis, 7,9% de resíduos em regiões permitidas, 2,1% de resíduos em regiões generosamente permitidas e 0,4% de resíduos em regiões proibidas.



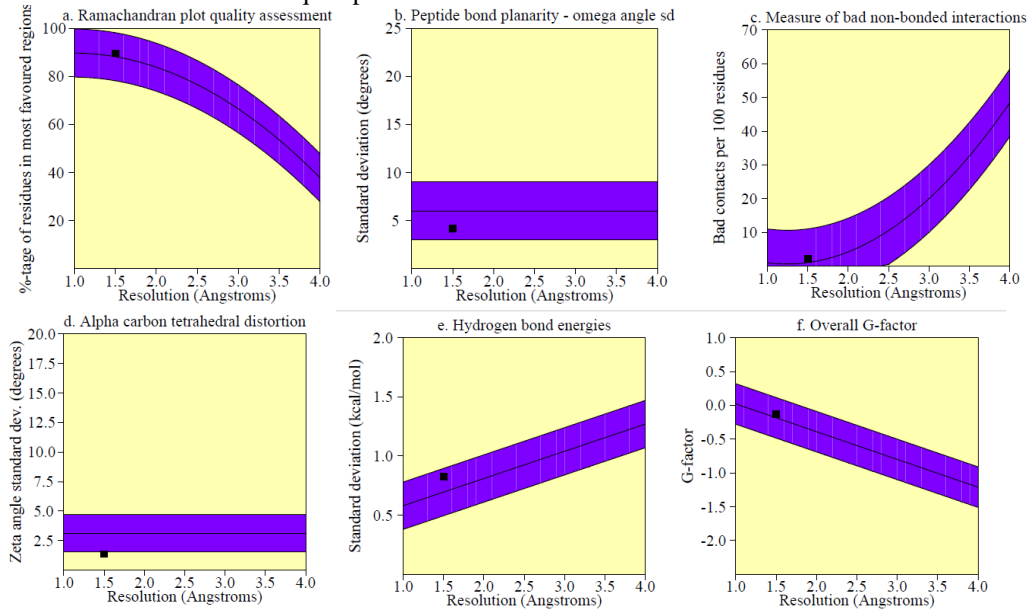
Estadísticas de gráfico

Resíduos em regiões mais favoráveis [A,B,L]	214	89.5%
Resíduos em regiões adicionais permitidas [a,b,l,p]	19	7.9%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	5	2.1%
Resíduos em regiões proibidas	1	0.4%
Número de resíduos não glicinas e não prolina	239	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	19	
Número de resíduos de prolina	17	
Número total de resíduos	277	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

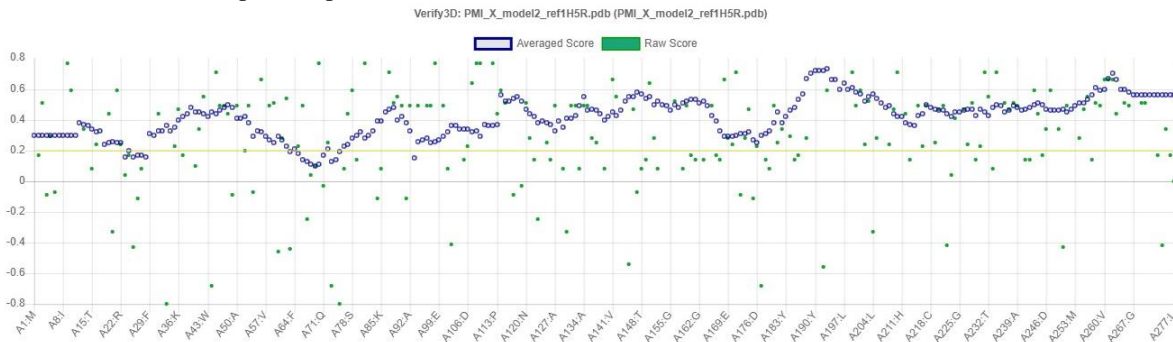
Fonte: O autor

Figura A27 - Resumo da avaliação estereoquímica da hipótese 2 gerada para PMI XAC realizado pelo **PROCHECK** com G-Factor de -0.1 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta resultados melhores com um valor de 1.4.



Fonte: O autor

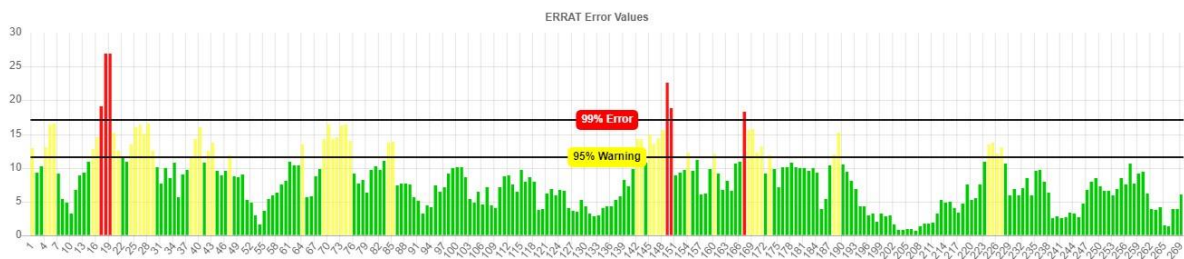
Figura A28 - Gráfico para a hipótese 2 gerada para PMI XAC obtido pelo **VERIFY3D**: 93,88% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

Figura A29 - Gráfico para a hipótese 2 gerada para PMI XAC obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 79.5539.

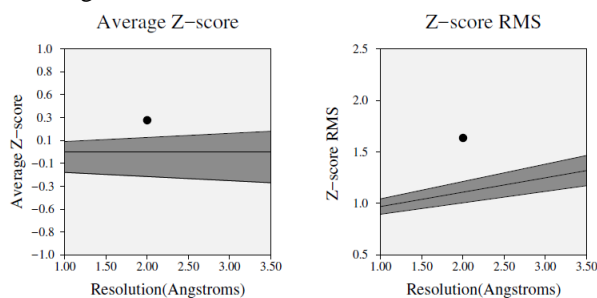
Quality Factor: A: 79.5539 | PDF | PostScript | Log



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

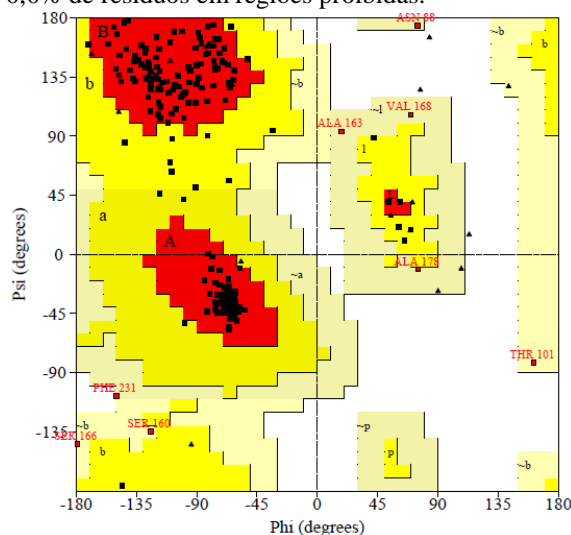
Figura A30 - Gráfico para a hipótese 2 gerada para PMI XAC obtido pelo **PROVE**. Essa estrutura apresenta 5,6% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A31 - Diagrama de Ramachandram da hipótese 3 gerada para PMI XAC obtido pelo **PROCHECK**: 89,5% de resíduos em regiões favoráveis, 7,1% de resíduos em regiões permitidas, 3,3% de resíduos em regiões generosamente permitidas e 0,0% de resíduos em regiões proibidas.



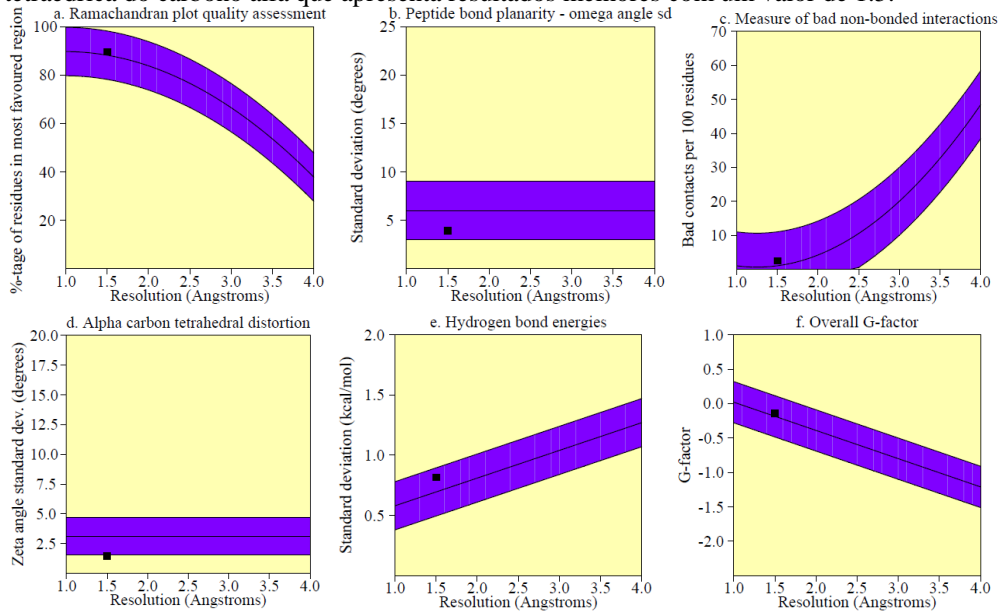
Estadísticas de plotagem

Resíduos em regiões mais favoráveis [A,B,L]	214	89.5%
Resíduos em regiões adicionais permitidas [a,b,l,p]	17	7.1%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	8	0.0%
Resíduos em regiões proibidas	0	0.0%
Número de resíduos não glicinas e não prolina	222	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	19	
Número de resíduos de prolina	17	
Número total de resíduos	277	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

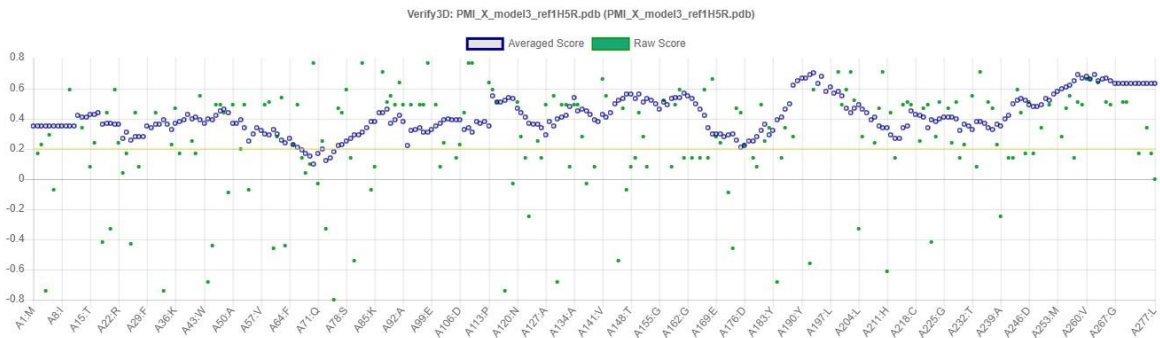
Fonte: O autor

Figura A32 - Resumo da avaliação estereoquímica da hipótese 3 gerada para PMI XAC realizado pelo **PROCHECK** com G-Factor de -0.1 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta resultados melhores com um valor de 1.5.



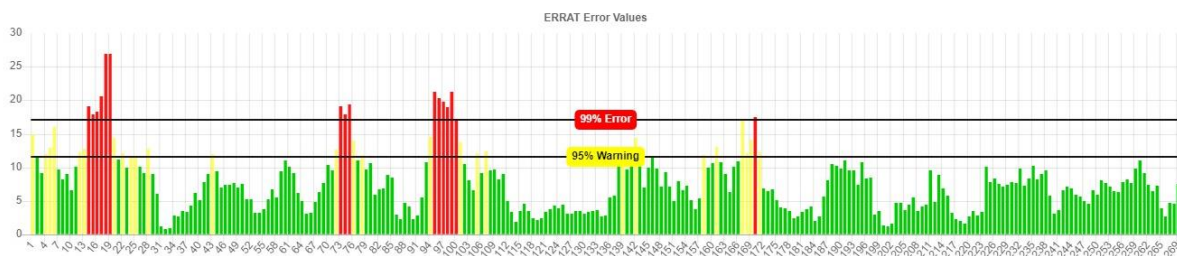
Fonte: O autor

Figura A33 - Gráfico para a hipótese 3 obtido pelo **VERIFY3D**: 97,11% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

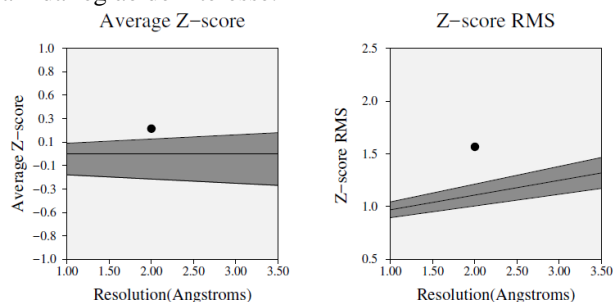
Figura A34 - Gráfico para a hipótese 3 gerada para PMI XAC obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 84.0149.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

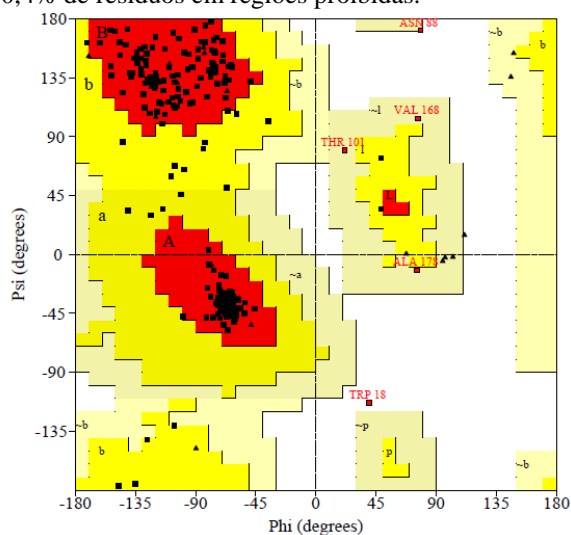
Figura A35 - Gráfico para a hipótese 3 gerada para PMI XAC obtido pelo **PROVE**. Essa estrutura apresenta 5,7% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura A36 - Diagrama de Ramachandram da hipótese 5 gerada para PMI XAC obtido pelo **PROCHECK**: 89,1% de resíduos em regiões favoráveis, 8,8% de resíduos em regiões permitidas, 1,7% de resíduos em regiões generosamente permitidas e 0,4% de resíduos em regiões proibidas.



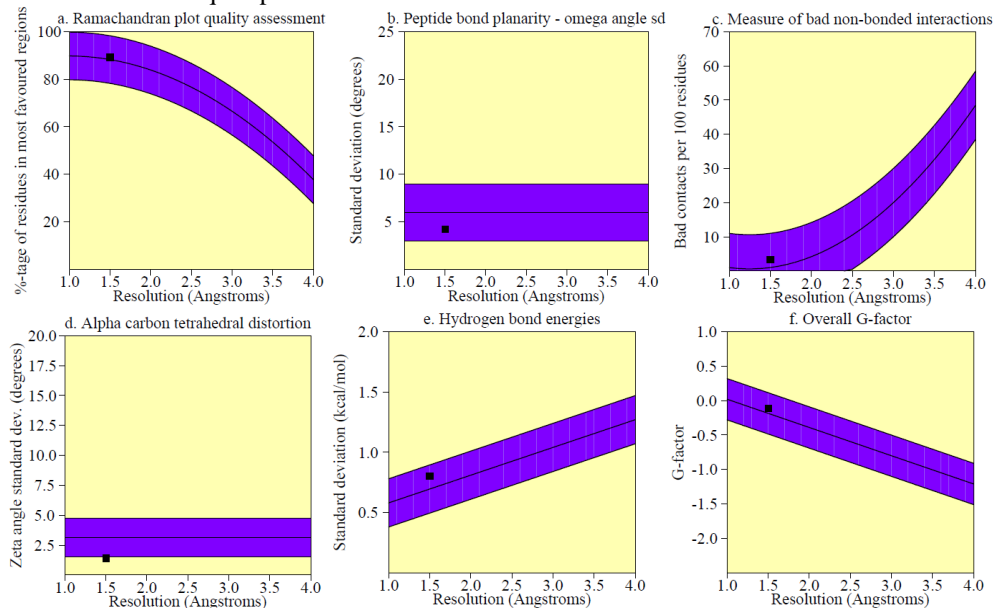
Estadísticas de plotagem

Resíduos em regiões mais favoráveis [A,B,L]	213	89.1%
Resíduos em regiões adicionais permitidas [a,b,l,p]	21	8.8%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	4	1.7%
Resíduos em regiões proibidas	1	0.4%
Número de resíduos não glicinas e não prolina	239	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	19	
Número de resíduos de prolina	17	
Número total de resíduos	277	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

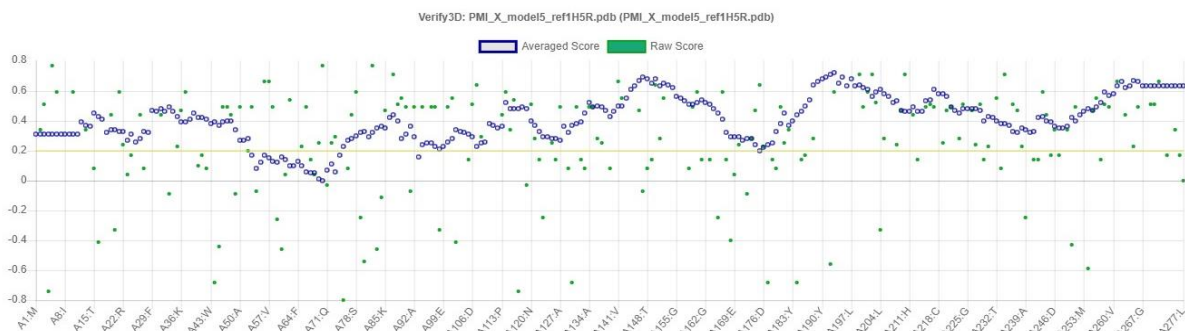
Fonte: O autor

Figura A37 - Resumo da avaliação estereoquímica da hipótese 5 gerada para PMI XAC realizado pelo **PROCHECK** com G-Factor de -0.1 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta resultados melhores com um valor de 1.4.



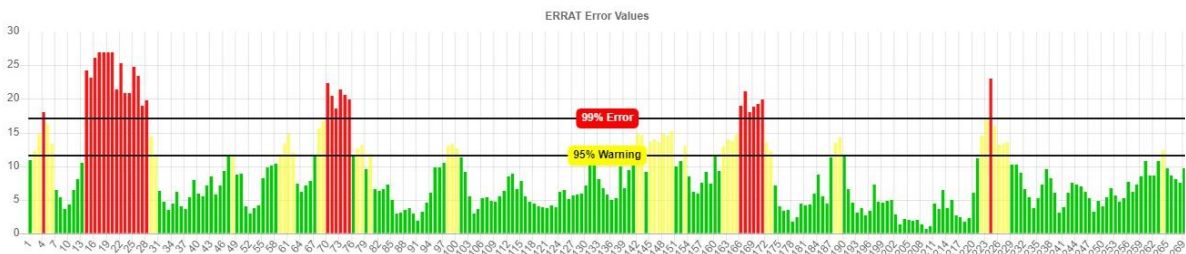
Fonte: O autor

Figura A38 - Gráfico para a hipótese 5 gerada para PMI XAC obtido pelo **VERIFY3D**: 91,70% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

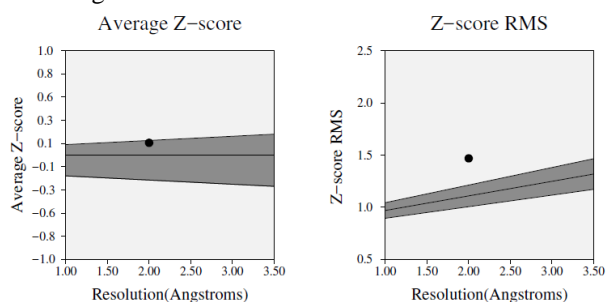
Figura A39 - Gráfico para a hipótese 5 gerada para PMI XAC obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 73.2342.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

Figura A40 - Gráfico para a hipótese 5 gerada para PMI XAC obtido pelo **PROVE**. Essa estrutura apresenta 4,1% dos aminoácidos foram da região de interesse.



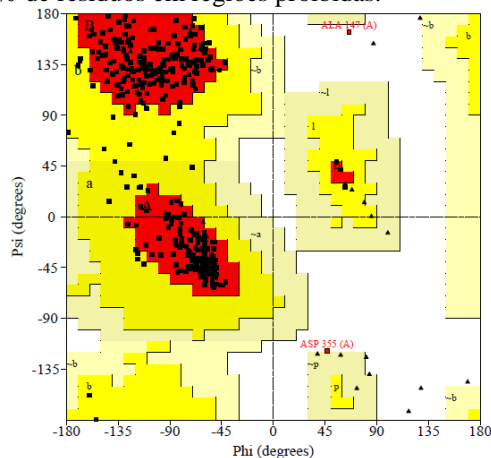
O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

APÊNDICE B - Dados relacionados a construção do modelo de PMI *Homo sapiens*

Nesse apêndice são apresentados os dados obtidos em relação as estruturas cristalográficas 3H1M e 5NW7; além dos outros 4 modelos preditos por modelagem por homologia estrutural. Os dados obtidos são os gráficos de Ramachandran, **PROCHECK**, **VERIFY3D**, **ERRAT** e **PROVE** com os respectivos valores encontrados para cada estrutura proteica.

Figura B1 - Diagrama de Ramachandram da estrutura cristalográfica 3H1M obtido pelo **PROCHECK**: 90,6% de resíduos em regiões favoráveis, 8,8% de resíduos em regiões permitidas, 0,0% de resíduos em regiões generosamente permitidas e 0,6% de resíduos em regiões proibidas.



Estatísticas de plotagem

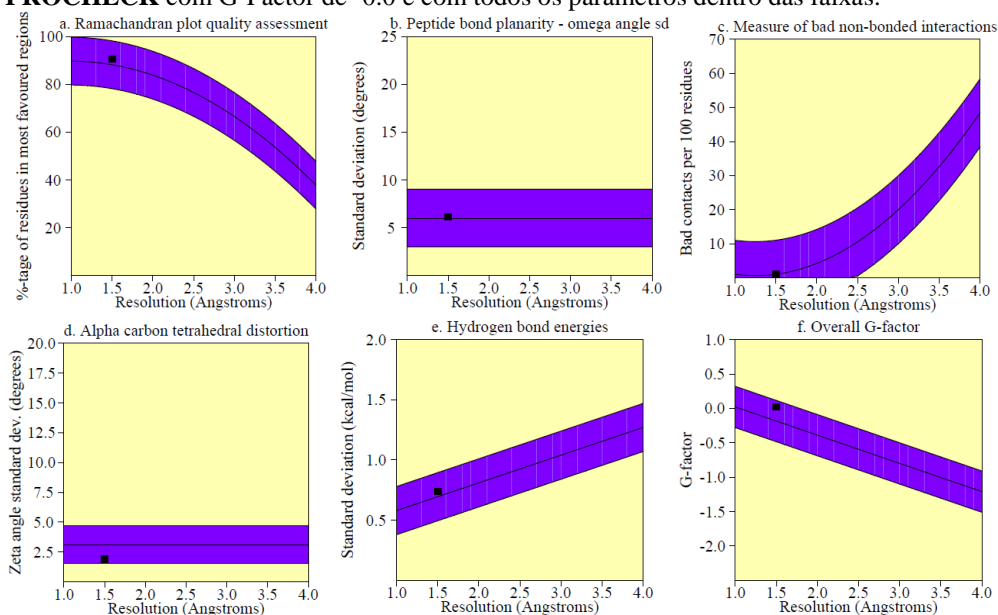
Resíduos em regiões mais favoráveis [A,B,L]	308	90.6%
Resíduos em regiões adicionais permitidas [a,b,l,p]	30	8.8%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	0	0.0%
Resíduos em regiões proibidas	2	0.6%
	----	----
Número de resíduos não glicinas e não prolina	340	100.0%
Número de resíduos finais (exc. Glicina e prolina)	3	
Número de resíduos de glicina (mostrados como triângulos)	23	
Número de resíduos de prolina	26	

Número total de resíduos	392	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

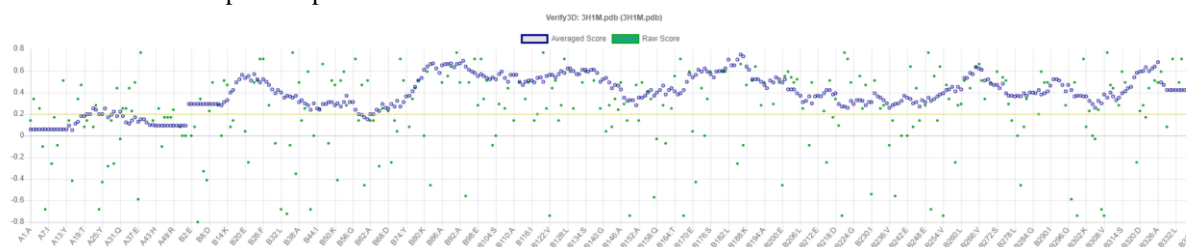
Fonte: O autor

Figura B2 - Resumo da avaliação estereoquímica da estrutura cristalográfica 3H1M realizado pelo **PROCHECK** com G-Factor de -0.0 e com todos os parâmetros dentro das faixas.



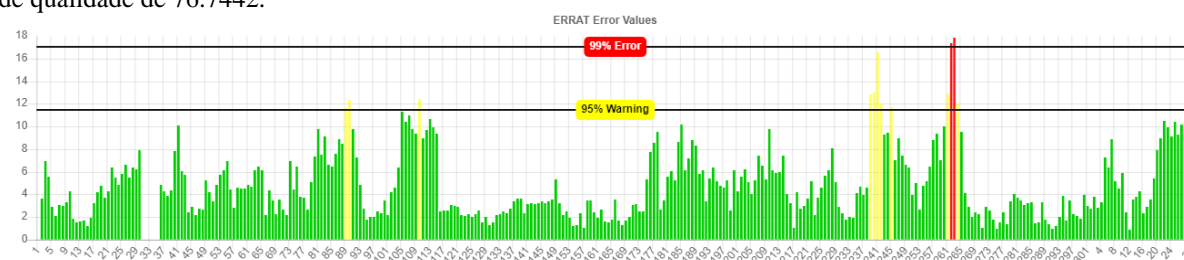
Fonte: O autor

Figura B3 - Gráfico para a estrutura cristalográfica 3H1M obtido pelo **VERIFY3D**: 88,01% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

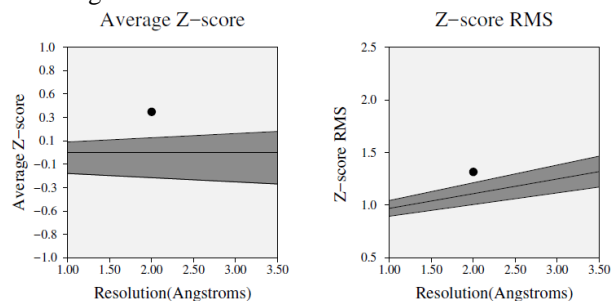
Figura B4 - Gráfico para a estrutura cristalográfica 3H1M obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 76.7442.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

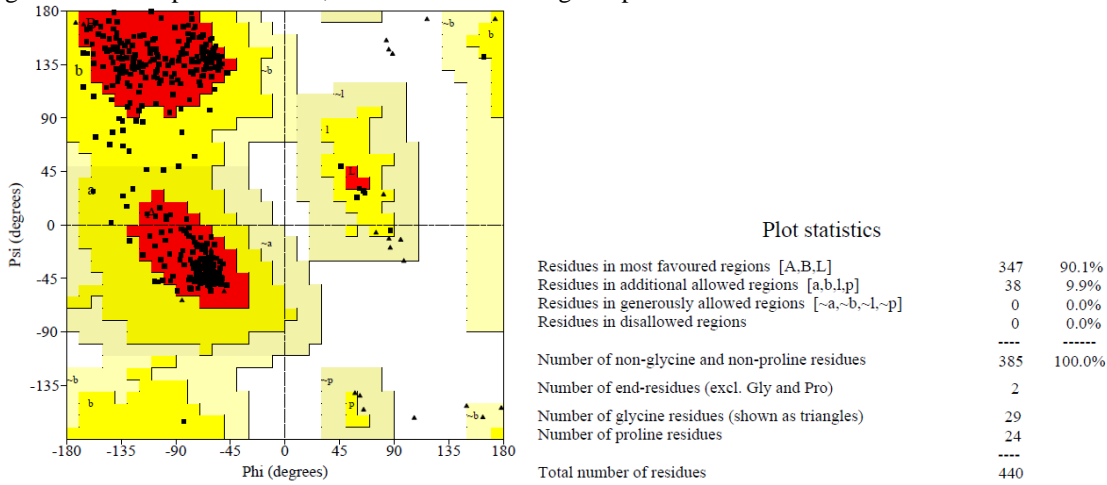
Figura B5 - Gráfico para a estrutura cristalográfica 3H1M obtido pelo **PROVE**. Essa estrutura apresenta 3,6% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura B6 - Diagrama de Ramachandram da estrutura cristalográfica 5NW7 obtido pelo **PROCHECK**: 90,1% de resíduos em regiões favoráveis, 9,9% de resíduos em regiões permitidas, 0,0% de resíduos em regiões generosamente permitidas e 0,0% de resíduos em regiões proibidas.



Estadísticas de plotagem

Resíduos em regiões mais favoráveis [A,B,L]	347	90.1%
Resíduos em regiões adicionais permitidas [a,b,l,p]	38	9.9%
Regiões em generosamente permitidas [-a, ~b, ~l, ~p]	0	0.0%
Resíduos em regiões proibidas	0	0.0%

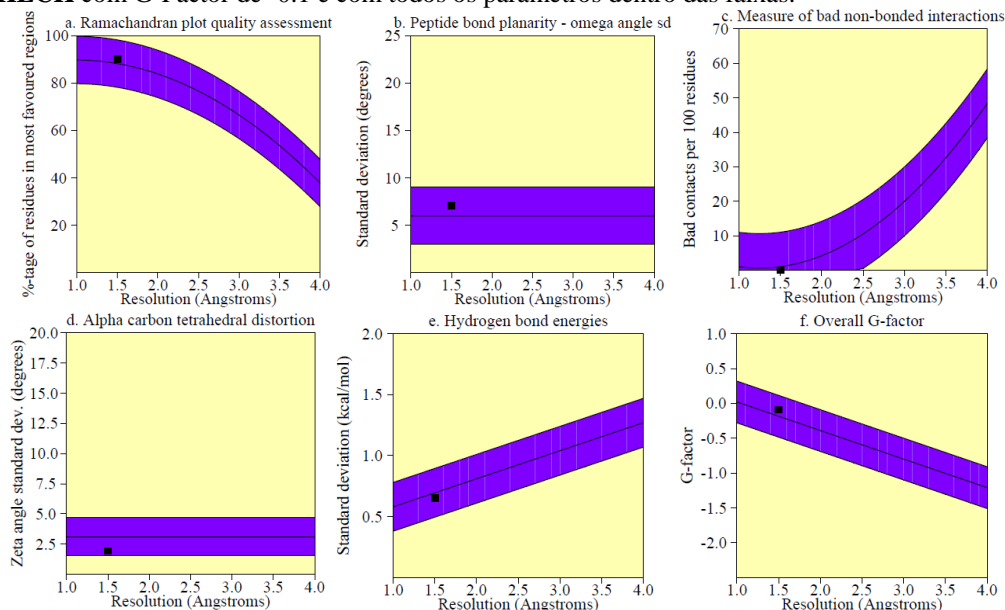
Número de resíduos não glicinas e não prolina	385	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	29	
Número de resíduos de prolina	24	

Número total de resíduos	440	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

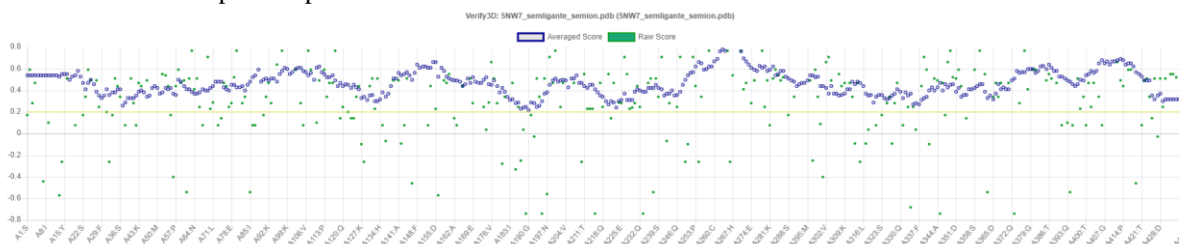
Fonte: O autor

Figura B7 - Resumo da avaliação estereoquímica da estrutura cristalográfica 3H1M realizado pelo **PROCHECK** com G-Factor de -0.1 e com todos os parâmetros dentro das faixas.



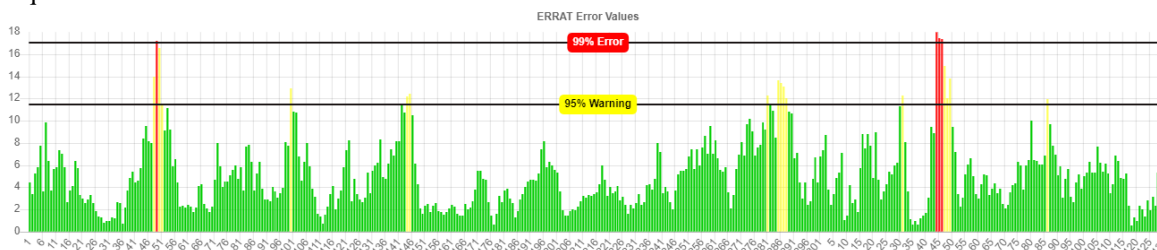
Fonte: O autor

Figura B8 - Gráfico para a estrutura cristalográfica 3H1M obtido pelo **VERIFY3D**: 100% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

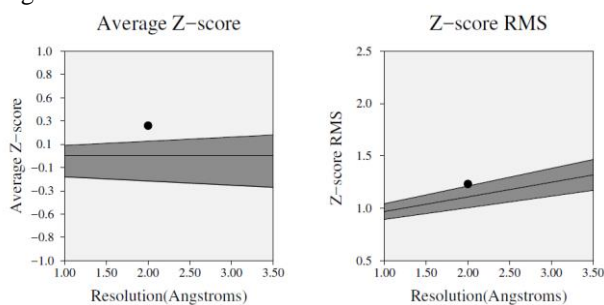
Figura B9 - Gráfico para a estrutura cristalográfica 5NW7 obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 95.3704.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

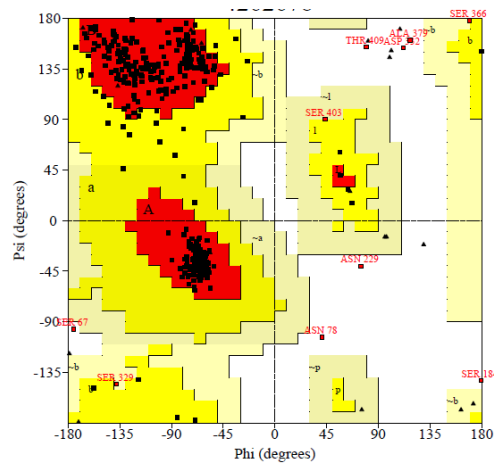
Figura B10 - Gráfico para a estrutura cristalográfica 3H1M obtido pelo **PROVE**. Essa estrutura apresenta 2,9% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura B11 - Diagrama de Ramachandram da hipótese 1 gerada para PMI *Homo sapiens* obtido pelo **PROCHECK**: 88,4% de resíduos em regiões favoráveis, 8,9% de resíduos em regiões permitidas, 0,8% de resíduos em regiões generosamente permitidas e 1,9% de resíduos em regiões proibidas.



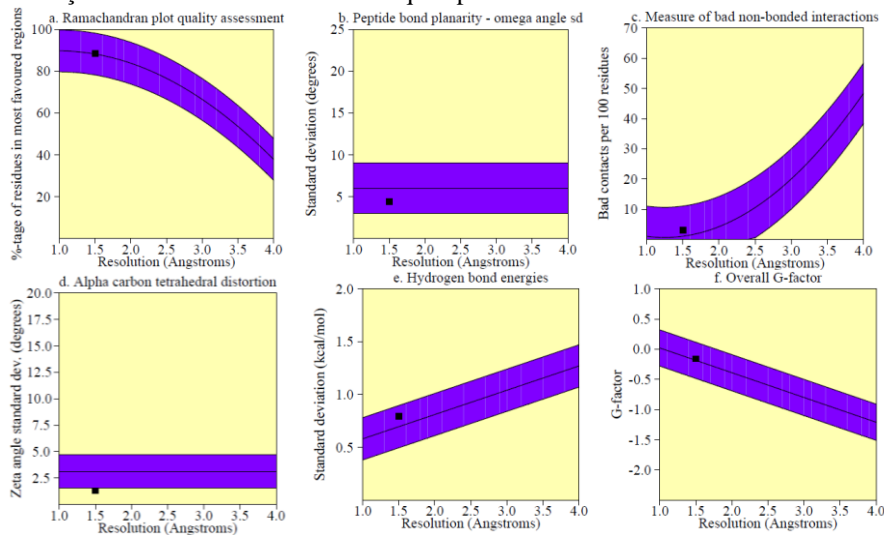
Estadísticas de plotagem

Resíduos em regiões mais favoráveis [A,B,L]	327	88.4%
Resíduos em regiões adicionais permitidas [a,b,l,p]	33	8.9%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	3	0.8%
Resíduos em regiões proibidas	7	1.9%
	----	----
Número de resíduos não glicinas e não prolina	370	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	21	
Número de resíduos de prolina	30	
	----	----
Número total de resíduos	425	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

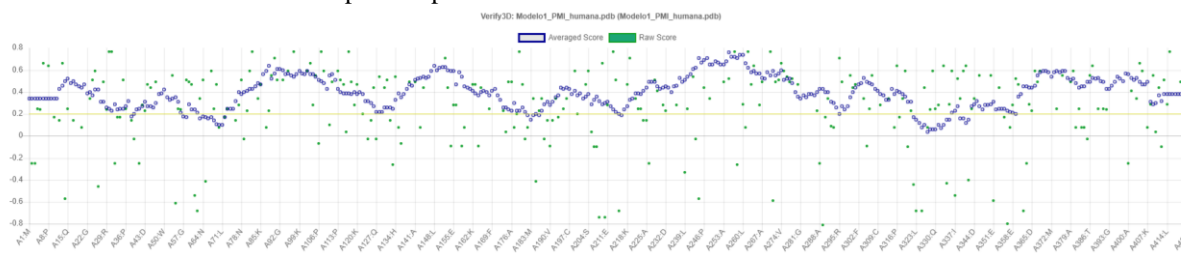
Fonte: O autor

Figura B12 - Resumo da avaliação estereoquímica da hipótese 1 gerada para PMI *Homo sapiens* realizado pelo **PROCHECK** com G-Factor de -0.2 e com todos os parâmetros dentro das faixas, com exceção da distorção tetraédrica do carbono alfa que apresenta um melhor resultado com valor de 1.3.



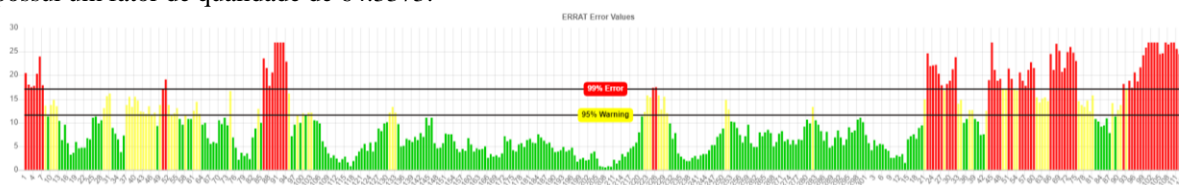
Fonte: O autor

Figura B13 - Gráfico para a hipótese 1 gerada para PMI *Homo sapiens* obtido pelo **VERIFY3D**: 91,49% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

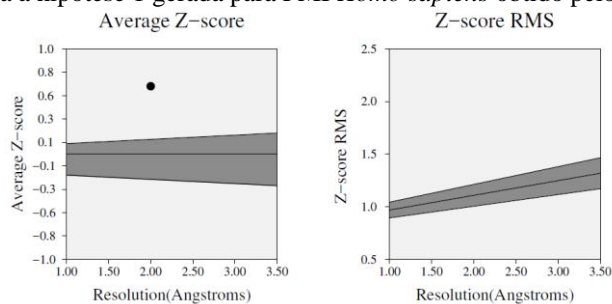
Figura B14 - Gráfico para a hipótese 1 gerada para PMI *Homo sapiens* obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 64.3373.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

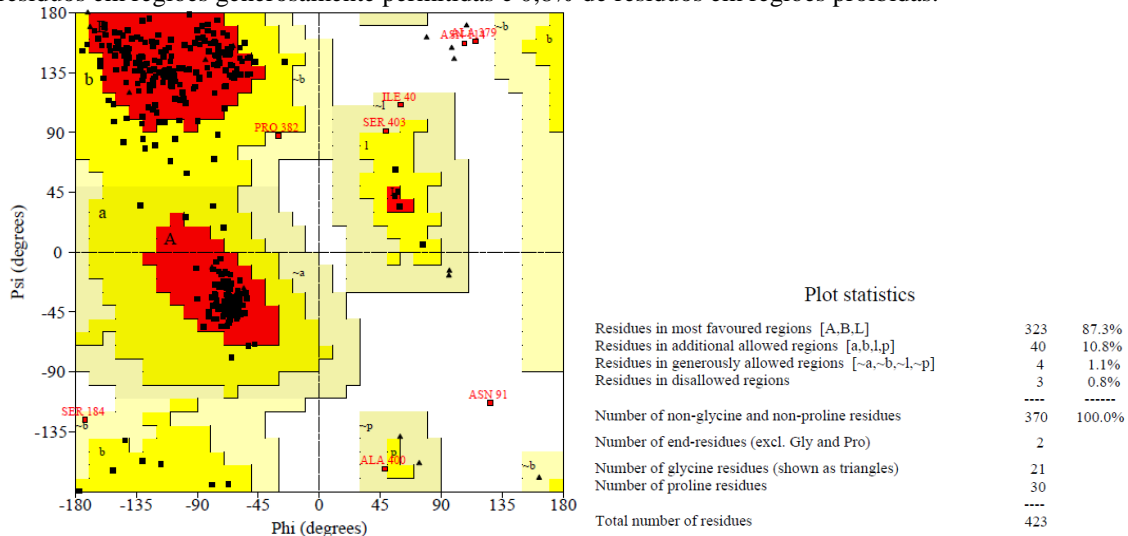
Figura B15 - Gráfico para a hipótese 1 gerada para PMI *Homo sapiens* obtido pelo **PROVE**.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura B16 - Diagrama de Ramachandram da hipótese 2 gerada para PMI *Homo sapiens* obtido pelo PROCHECK: 87,3% de resíduos em regiões favoráveis, 10,8% de resíduos em regiões permitidas, 1,1% de resíduos em regiões generosamente permitidas e 0,8% de resíduos em regiões proibidas.



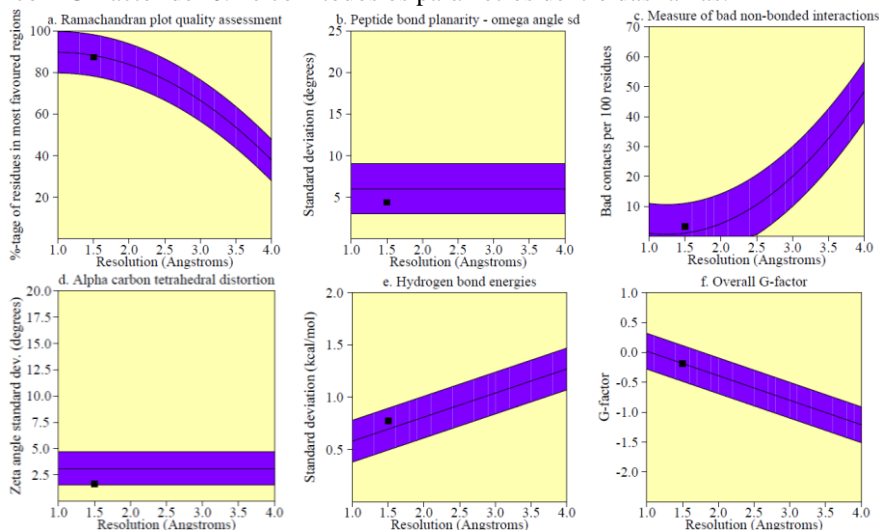
Estadísticas de gráfico

Resíduos em regiões mais favoráveis [A,B,L]	323	87.3%
Resíduos em regiões adicionais permitidas [a,b,l,p]	40	10.8%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	4	1.1%
Resíduos em regiões proibidas	3	0.8%
Número de resíduos não glicinas e não prolina	370	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	21	
Número de resíduos de prolina	30	
Número total de resíduos	423	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

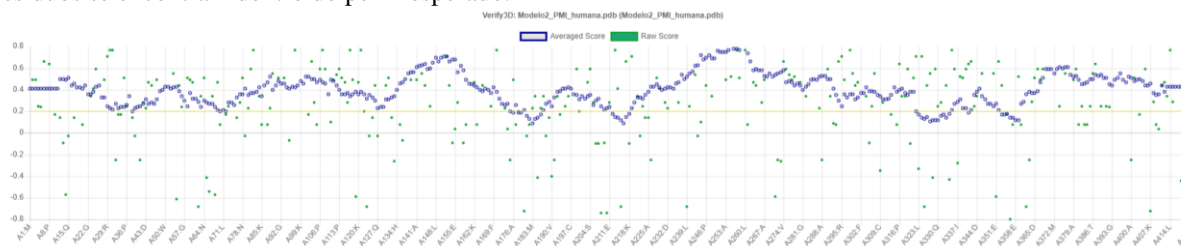
Fonte: O autor

Figura B17 - Resumo da avaliação estereoquímica da hipótese 2 gerada para PMI *Homo sapiens* realizado pelo PROCHECK com G-Factor de -0.2 e com todos os parâmetros dentro das faixas.



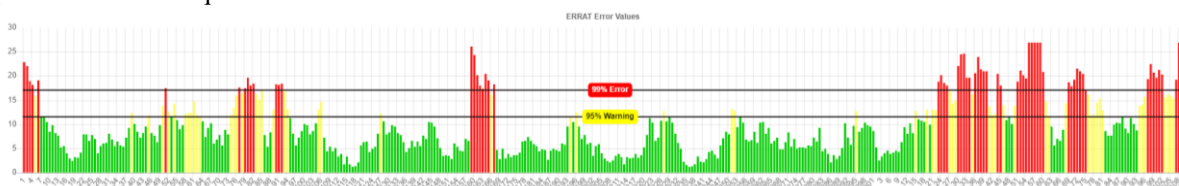
Fonte: O autor

Figura B18 - Gráfico para a hipótese 2 gerada para PMI *Homo sapiens* obtido pelo **VERIFY3D**: 91,49% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

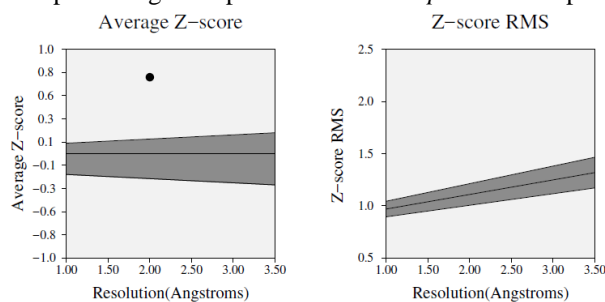
Figura B19 - Gráfico para a hipótese 2 gerada para PMI *Homo sapiens* obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 68.6747.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

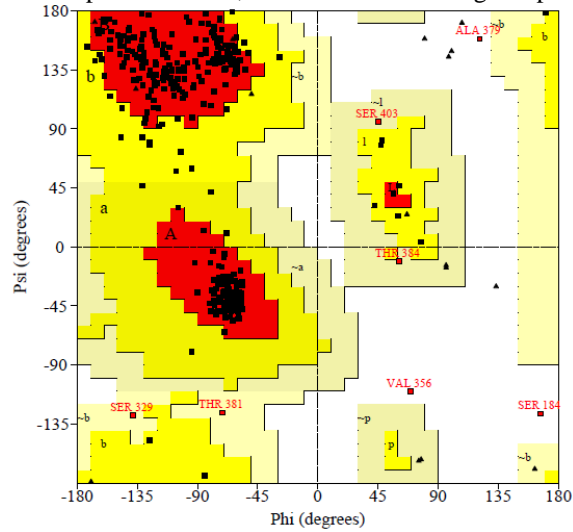
Figura B20 - Gráfico para a hipótese 2 gerada para PMI *Homo sapiens* obtido pelo **PROVE**.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura B21 - Diagrama de Ramachandram da hipótese 4 gerada para PMI *Homo sapiens* obtido pelo PROCHECK: 88,9% de resíduos em regiões favoráveis, 9,2% de resíduos em regiões permitidas, 1,1% de resíduos em regiões generosamente permitidas e 0,8% de resíduos em regiões proibidas.



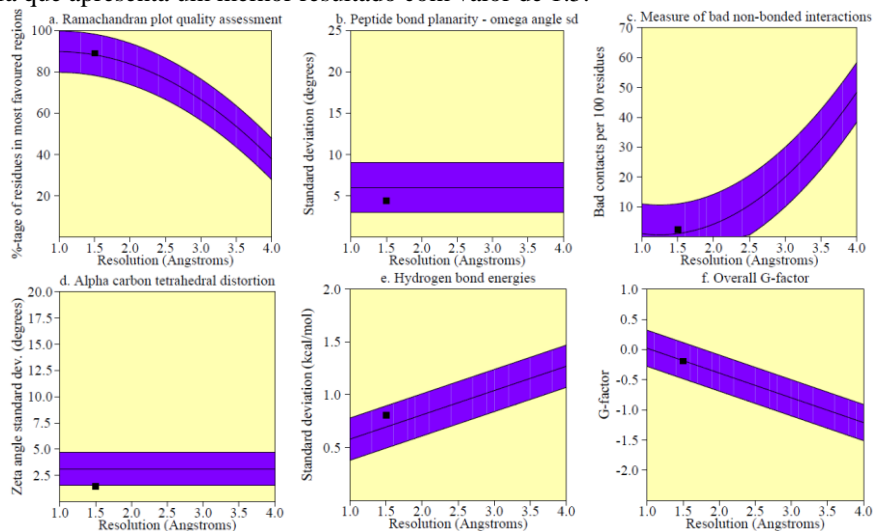
Estadísticas de gráfico

Resíduos em regiões mais favoráveis [A,B,L]	329	88.9%
Resíduos em regiões adicionais permitidas [a,b,l,p]	34	9.2%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	4	1.1%
Resíduos em regiões proibidas	3	0.8%
	----	----
Número de resíduos não glicinas e não prolina	370	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	21	
Número de resíduos de prolina	30	
	----	----
Número total de resíduos	423	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

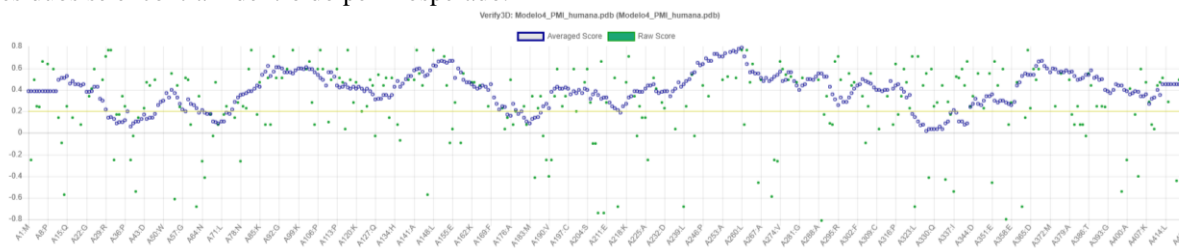
Fonte: O autor

Figura B22 - Resumo da avaliação estereoquímica da hipótese 4 gerada para PMI *Homo sapiens* realizado pelo PROCHECK com G-Factor de -0.2 e com todos os parâmetros dentro das faixas, exceto a distorção tetraédrica do carbono alfa que apresenta um melhor resultado com valor de 1.5.



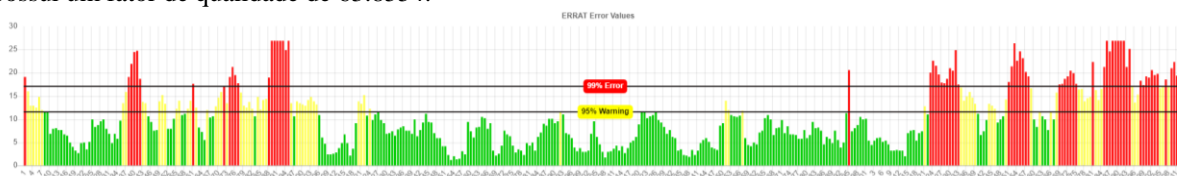
Fonte: O autor

Figura B23 - Gráfico para a hipótese 4 gerada para PMI *Homo sapiens* obtido pelo **VERIFY3D**: 86,05% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

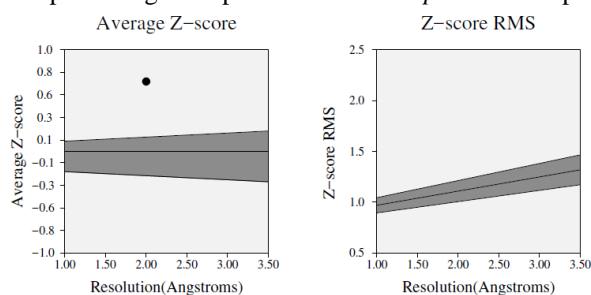
Figura B24 - Gráfico para a hipótese 4 gerada para PMI *Homo sapiens* obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 63.8554.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

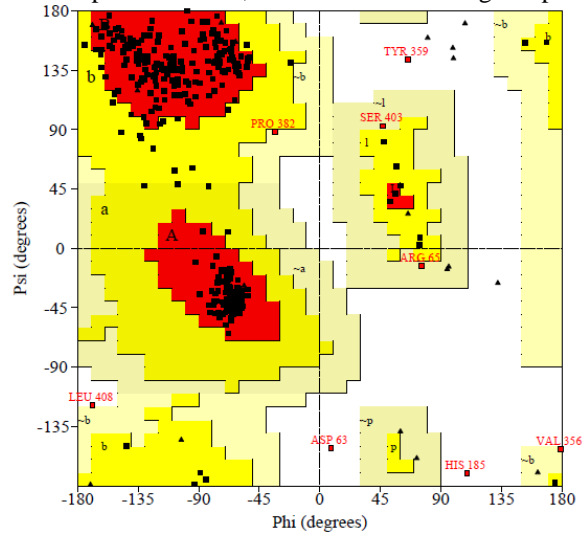
Figura B25 - Gráfico para a hipótese 4 gerada para PMI *Homo sapiens* obtido pelo **PROVE**.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor

Figura B26 - Diagrama de Ramachandram da hipótese 5 gerada para PMI *Homo sapiens* obtido pelo PROCHECK: 89,2% de resíduos em regiões favoráveis, 8,9% de resíduos em regiões permitidas, 0,8% de resíduos em regiões generosamente permitidas e 1,1% de resíduos em regiões proibidas.



Estadísticas de plotagem

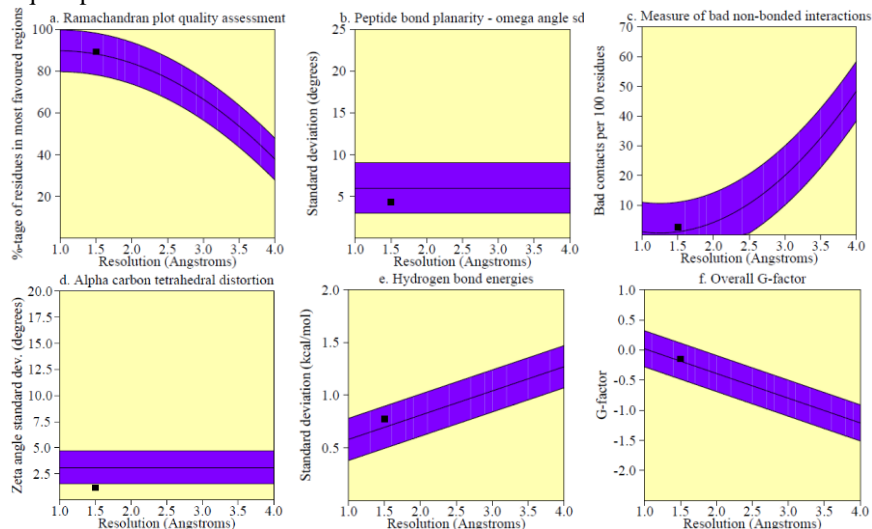
Resíduos em regiões mais favoráveis [A,B,L]	330	89.2%
Resíduos em regiões adicionais permitidas [a,b,l,p]	33	8.9%
Regiões em generosamente permitidas [~a, ~b, ~l, ~p]	3	0.8%
Resíduos em regiões proibidas	4	1.1%
	----	----
Número de resíduos não glicinas e não prolina	370	100.0%
Número de resíduos finais (exc. Glicina e prolina)	2	
Número de resíduos de glicina (mostrados como triângulos)	21	
Número de resíduos de prolina	30	

Número total de resíduos	423	

Com base na análise de 118 estruturas de resolução de pelo menos 2.0 Å e fator R não superior a 20%, espera-se que um modelo de boa qualidade tenha mais de 90% nas regiões mais favorecidas.

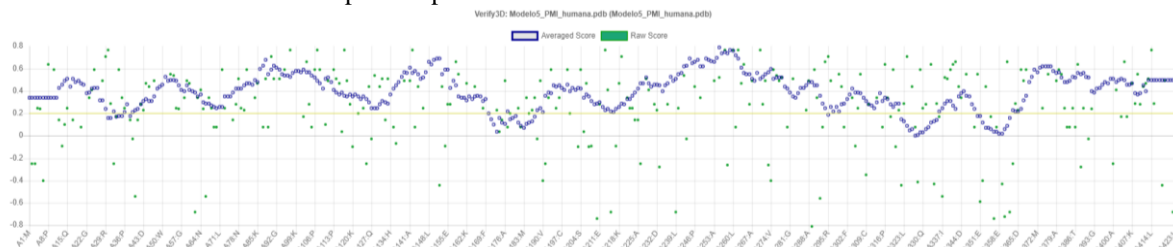
Fonte: O autor

Figura B27 - Resumo da avaliação estereoquímica da hipótese 5 gerada para PMI *Homo sapiens* realizado pelo PROCHECK com G-Factor de -0.1 e com todos os parâmetros dentro das faixas, exceto a distorção tetraédrica do carbono alfa que apresenta um melhor resultado com valor de 1.2.



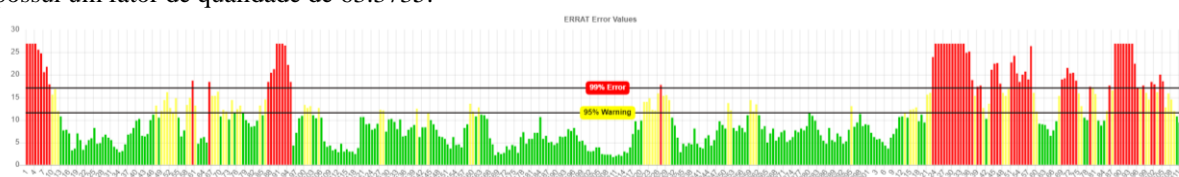
Fonte: O autor

Figura B28 - Gráfico para a hipótese 5 gerada para PMI *Homo sapiens* obtido pelo **VERIFY3D**: 88,18% dos resíduos se encontram dentro do perfil esperado.



Fonte: O autor

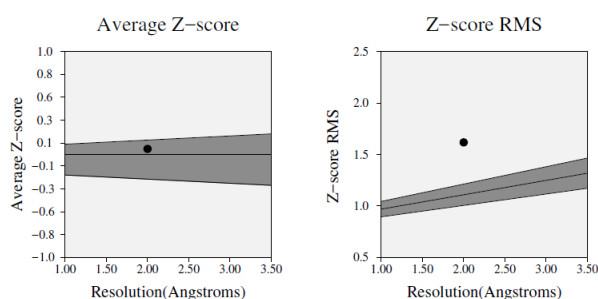
Figura B29 - Gráfico para a hipótese 5 gerada para PMI *Homo sapiens* obtido pelo **ERRAT**. Essa estrutura possui um fator de qualidade de 63.3735.



A barra vermelha indica o aminoácido com erro acima de 99%, as barras amarelas indicam regiões com erros entre 95% e 99% e as barras verdes indicam as regiões com menor porcentagem de erro para dobrar proteínas.

Fonte: O autor

Figura B30 - Gráfico para a hipótese 5 gerada para PMI *Homo sapiens* obtido pelo **PROVE**. Essa estrutura apresenta 6,4% dos aminoácidos foram da região de interesse.



O Z-score corresponde aos desvios dos valores do volume atômico em relação aos valores padrão e o Z-score RMS mede a média de volumes irregulares na estrutura da proteína.

Fonte: O autor