UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

OTÁVIO HUGO AGUIAR GOMES

ANÁLISE ESTRUTURAL DE ORTÓLOGOS DA PROTEÍNA GLICOLATO OXIDASE (GOX) EM PLANTAS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

FORTALEZA
2022

OTÁVIO HUGO AGUIAR GOMES

ANÁLISE ESTRUTURAL DE ORTÓLOGOS DA PROTEÍNA GLICOLATO OXIDASE (GOX) EM PLANTAS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

Monografia apresentada ao Curso de Biotecnologia do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Biotecnologia.

Orientador: Murilo Alves Siqueira

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

G615a Gomes, Otávio Hugo Aguiar.
Análise estrutural de ortólogos da proteína Glicolato Oxidase (GOX) em plantas utilizando inteligência artificial / Otávio Hugo Aguiar Gomes. – 2022.
36 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Biotecnologia, Fortaleza, 2022.
Orientação: Prof. Dr. Murilo Siqueira Alves.

1. Modelagem computacional. 2. Glicolato Oxidase. 3. CPSMV. 4. RoseTTAFold. I. Título.

CDD 661

OTÁVIO HUGO AGUIAR GOMES

ANÁLISE ESTRUTURAL DE ORTÓLOGOS DA PROTEÍNA GLICOLATO OXIDASE (GOX) EM PLANTAS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

Monografia apresentada ao Curso de Biotecnologia do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Biotecnologia.

Aprovado em: ___/___/_____.

BANCA EXAMINADORA

_____

Prof. Dr. Murilo Siqueira Alves (Orientador)

Universidade Federal do Ceará (UFC)

_____

Prof. Dr. Bruno Anderson Matias da Rocha

Universidade Federal do Ceará (UFC)

_____

Ana Carolina Moreira da Costa

Universidade Federal do Ceará (UFC)

“Os erros são os portais da descoberta.”
(James Joyce)

“Toda a educação assenta nestes dois princípios: primeiro repelir o assalto fogoso das crianças ignorantes à verdade e depois iniciar as crianças humilhadas na mentira, de modo insensível e progressivo.”
(Franz Kafka)

“Mesmo pouco aprendizado é perigoso; beba profundamente se quiser provar a primavera de Pieira”
(Alexander Pope)

“And I run from the wolves, ooh
Tearing into me
Without teeth”
(Of Monsters and Men)

Resumo

O feijão tem inegável importância socioeconômica, cultural e nutritiva para o Brasil. Dentre as espécies de feijão mais consumidas no Brasil e no mundo encontra-se o feijão-caupi (*Vigna unguiculata [L.] Walp*), sendo uma das leguminosas de grãos mais amplamente adaptadas, versáteis e nutritivas. Mesmo apresentando resiliência a diversas condições de estresse ambiental, o feijão-caupi é suscetível a patógenos como o vírus do mosaico severo do caupi (CPSMV). Estudos recentes comprovaram que diversas proteínas são diferencialmente expressas durante a interação entre o CPSMV e o feijão caupi. Dentre essas proteínas está a enzima Glicolato Oxidase (GOX), cuja atividade é a oxidação do glicolato em glioxilato e H_2O_2. Para melhor entendimento sobre o papel biológico desta enzima em *V. unguiculata* e espécies filogeneticamente próximas, foi utilizada a ferramenta de modelagem de proteínas que utiliza algoritmos baseados em *deep learning*, RosettaFold, através do servidor Robetta, para predizer modelos estruturais de GOX em espécies de feijão (*Vigna unguiculata, Vigna radiata, Vigna angularis, Phaseolus vulgaris*), luzerna-cortada (*Medicago truncatula*), soja (*Glycine Max*), sorgo (*Sorghum bicolor*) e milho (*Zea mays*). Com a obtenção dos modelos estruturais foram utilizados 3 parâmetros fundamentais para qualificar os modelos obtidos, sendo eles o RMSD (raiz quadrada do desvio quadrático médio), TM-Score (Template Modeling Score) e o parâmetro de confiança obtido no servidor Robetta. As médias dos valores de RMSD de cada espécie variaram de 0,865 à 0,906, da confiança do Servidor Robetta foi de 0,85 a 0,88, já no TM-Score foi de 0,7059 a 0,9101 com exceção do *S. bicolor* que teve média de 0,1777. Os valores obtidos dos parâmetros comprovam a qualidade dos modelos produzidos e que também prova que o RosettaFold consegue prever estruturas protéicas a partir de sequências de aminoácidos com precisão. Além de fornecer modelos estruturais da proteína GOX de diferentes espécies de plantas para aplicação em estudos funcionais, espera-se que este trabalho sirva para incentivar mais estudos de modelos preditivos de estruturas protéicas relacionadas à resistência a estresse.

Palavras-chave: Modelagem computacional, RoseTTAFold, CPSMV, Vigna, Phaseolus, Zea, Glycine, Medicago, Sorghum

Abstract

Beans have undeniable socioeconomic, cultural and nutritional importance for Brazil. Cowpea (*Vigna unguiculata [L.] Walp*) is among the most consumed bean species in Brazil and in the world, being one of the most widely adapted, versatile and nutritious grain legumes. Even showing resilience to various environmental stress conditions, cowpea is susceptible to pathogens such as cowpea severe mosaic virus (CPSMV). Recent studies have shown that several proteins are differentially expressed during the interaction between CPSMV and cowpea. Among these proteins is the enzyme Glycolate Oxidase (GOX), whose activity is the oxidation of glycolate to glyoxylate and H_2O_2. For a better understanding of the biological role of this enzyme in V. unguiculata and phylogenetically close species, the protein modeling tool that uses algorithms based on deep learning, RosettaFold, through the Robetta server, was used to predict structural models of GOX in species of beans (*Vigna unguiculata, Vigna radiata, Vigna angularis, Phaseolus vulgaris*), barrelclover (*Medicago truncatula*), soybean (*Glycine Max*), sorghum (*Sorghum bicolor*) and maize (*Zea mays*). With the obtaining of the structural models, 3 fundamental parameters were used to qualify the models obtained, being the RMSD (square root of the mean square deviation), TM-Score (Template Modeling Score) and the confidence parameter obtained in the Robetta server. The mean values of RMSD for each species ranged from 0.865 to 0.906, the confidence of the Robetta Server ranged from 0.85 to 0.88, and the TM-Score ranged from 0.7059 to 0.9101 with the exception of *S. bicolor* which had a mean of 0.1777. The values obtained from the parameters prove the quality of the models produced and that also proves that RosettaFold can predict protein structures from amino acid sequences accurately. In addition to providing structural models of the GOX protein from different plant species for application in functional studies, it is expected that this work will serve to encourage further studies of predictive models of protein structures related to stress resistance.

Keywords: Computer modeling, RoseTTAFold, CPSMV, Vigna, Phaseolus, Zea, Glycine, Medicago, Sorghum

Sumário

1. Introdução

O feijão possui grande importância socioeconômica, cultural e nutritiva para a população brasileira, sendo um dos símbolos da alimentação nacional. Mesmo as espécies de feijão, como a *Vigna unguiculata [L.] Walp,* sendo conhecidas por possuir certa adaptação a diversos tipos de estresse, ainda sofrem com o ataque de patógenos, principalmente de vírus. O vírus da família do mosaico, dentre eles o vírus do mosaico severo do caupi (CPSMV), tem causado grandes perdas na produtividade das plantações ocasionando redução de até 85% da produção dependendo do cultivar de feijão e da época do ano (SILVA, 2016; BOOKER et al., 2005). Para uma melhor compreensão de como funcionam os mecanismos de resposta à infecção viral é necessário o estudo aprofundado de diversas moléculas diferentes, entre elas as proteínas. Diversas proteínas foram comprovadas serem diferencialmente expressas durante a infecção/interação com CPSMV, dentre essas proteínas se tem a enzima glicolato oxidase peroxissomal (EC: 1.1.3.15) também conhecida como GOX (AMORIM, 2018). Esta enzima tem atividade associada a catalase e media a oxidação do glicolato em glioxilato e H_2O_2 (PATRIOTA, 2022). Para melhor compreensão sobre GOX e como ela reage ao substrato é necessário estudar a estrutura tridimensional desta enzima. Com o melhoramento da computação e processamento de dados surgiu a tendência crescente em utilizar programas que fazem o uso de inteligência artificial para produzir modelos de predição destas estruturas. Dentre esses programas o RosettaFold tem ganhado destaque nos recentes anos devido sua precisão em obter modelos preditos de estruturas proteicas.

2. Fundamentação Teórica

2.1. Feijão

A importância do feijão para a alimentação da população brasileira é inquestionável sendo um clássico símbolo da alimentação no Brasil junto do arroz. De acordo com o IBGE (2022) em 2022 se terá uma produção estimada considerando as 3 safras de 3,2 milhões de toneladas de feijão. Dentre as espécies de feijão consumidas e produzidas no Brasil temos *Phaseolus vulgaris*, *Vigna unguiculata*, *Vigna radiata* e *Vigna angularis*.

Vigna unguiculata [L.] Walp. é uma leguminosa granífera de grande importância para alimentação humana sendo uma das principais fontes de proteína na dieta de populações pobres especialmente na América Latina e África (MELO et al., 2021; SILVA et al., 2016). É popularmente conhecida como feijão-caupi, feijão-de-corda, feijão-frade, feijão-fradinho, feijão-miúdo e feijão-massacar. Presente em todo o país, é predominantemente cultivada nas regiões Norte e Nordeste, do qual o Ceará é o segundo maior produtor deste feijão. (IBGE, 2015). O feijão-de-corda é de grande importância para os continentes Africano, Asiático e Sul-Americano (GONÇALVES, 2016). Essa leguminosa pertence ao grupo das dicotiledôneas, da família das Fabaceae, subfamília das Faboideae tribo Phaseoleae, subtribo Phaseolinae (SMARTT, 1990). As características de resistência à estresses bióticos e abióticos presentes no feijão-caupi são amplamente conhecidas, principalmente sua adaptabilidade a temperaturas elevadas (20-35 °C) e a solos com pouca água (MARTINS et al., 2020). Contudo, diferentes genótipos podem apresentar resistências a estresses diferentes. De acordo com Pereira (2013), às médias (% base seca) dos valores de proteínas totais, lipídios, cinzas, fibra alimentar total e carboidratos digeríveis do feijão-caupi foram 22,95; 1,60; 3,84; 17,80 e 53,87, respectivamente. Já os teores de ferro e zinco foram em média 4,62 e 3,21 mg/100g de semente, respectivamente.

Phaseolus vulgaris L. mais conhecido como feijão comum é um dos principais alimentos da população brasileira, sendo uma das principais fontes de proteína na dieta alimentar. Seu teor proteico pode chegar a 33% com valor energético de 341 cal/100g (OLIVEIRA et al., 2006; SILVA & WANDER, 2013). Seu teor nutricional varia de 20 a 35% de proteína, 60 a 65% de carboidrato (17 a 23 % de carboidratos na forma de fibra alimentar), de potássio (cerca de 1 %), de fósforo (cerca de 0,04

%), de ferro (cerca de 0,007 %) e quantidades bastante baixas de cálcio, cobre, zinco, magnésio e sódio (PINTO, 2016).

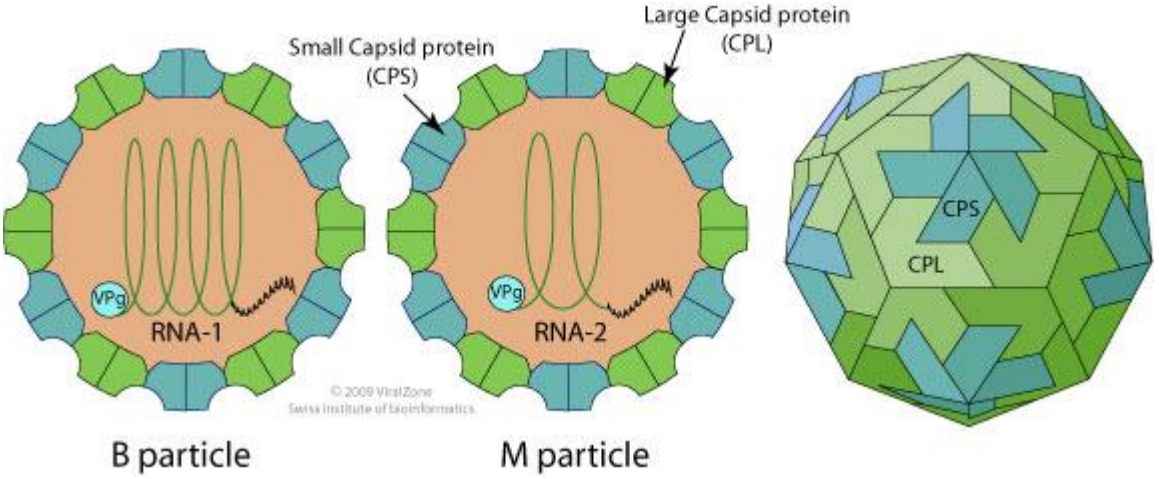
Vigna radiata (L.) R. Wilczek é conhecida no Brasil como feijão moyashi, feijão mungo, feijão mungo verde ou feijão-da-china. É muito utilizada para a obtenção de brotos de feijão para consumo humano (ARAÚJO et al., 2011). Também muito utilizado para a alimentação de bebês devido a sua quase ausência de fatores causadores de flatulências (ADSULE et al., 1986). O grão tem cerca de 19,05-23,86% de proteína (ULLAH et al., 2014).

Vigna angularis também conhecida como feijão-azuqui, uma leguminosa muito popular no Japão e muito consumida. É utilizada na fabricação de diversos doces. O conteúdo protéico varia 18,34 g a 23,81 g para cada 100 g de grão.

2.2. Cowpea severe mosaic virus (CPSMV)

O vírus mosaico severo do caupi (Cowpea severe mosaic virus, CPSMV) é um vírus isométrico (diâmetro entre 28 e 30 nm) pertencente à família Comoviridae e gênero Comovirus (Figura 01), com capsídeo constituído por 60 cópias de uma proteína maior de 43 kDa (CPL) e 60 cópias de uma proteína menor de 23 kDa (CPS). O CPSMV possui RNA (positivo) bipartido, denominados RNA-1 e RNA-2, como material genético, sendo desprovido de envelope viral. Este vírus pode causar perdas de até 85% da produção dependendo do cultivar de feijão caupi e da época do ano (BOOKER et al., 2005). Também pode afetar outras espécie como *Glycine max*, *Phaseolus vulgaris, Macroptilium lathyroides, Canavalia brasiliensis, Canavalia ensiformis, Psophocarpus tetragonolobus* e *Crotalaria juncea* e *Crotalaria paulinea* (BRIOSO et al., 1994; BERTACINI et al., 1998; LIMA et al., 2005).

Figura 01 - Modelo esquemático da estrutura de um Comovírus



Fonte: ViralZone, 2016.

2.3. Mecanismos de resposta a estresses bióticos em plantas

As plantas passam constantemente por diversas situações de estresse tanto biótico quanto abiótico e conseguem modular respostas de defesa de forma a superar tais estresses e retornar ao metabolismo normal (SOARES et al., 2007). O estresse biótico causado por um patógeno possui duas classificações para a interação planta-patógeno, sendo a interação compatível e a interação incompatível. Na interação compatível todo o ciclo de vida do patógeno acontece no hospedeiro. Já na interação incompatível, parte do ciclo de vida do patógeno acontece fora do hospedeiro, normalmente uma ou mais fases do ciclo reprodutivo (AMORIM, 2018).

Os mecanismos de defesa contra patógenos são divididos em duas categorias. O primeiro são as barreiras constitutivas formadas por barreiras químicas ou físicas inatas da planta como cutícula, parede celular e fitoanticipinas que limitam o acesso à célula vegetal. O segundo mecanismo é a resposta de defesa induzida que envolve a percepção do patógeno e a sinalização para induzir a uma resposta bioquímica para combater a infecção (DODDS & RATHJEN, 2010; SILVA, 2016).

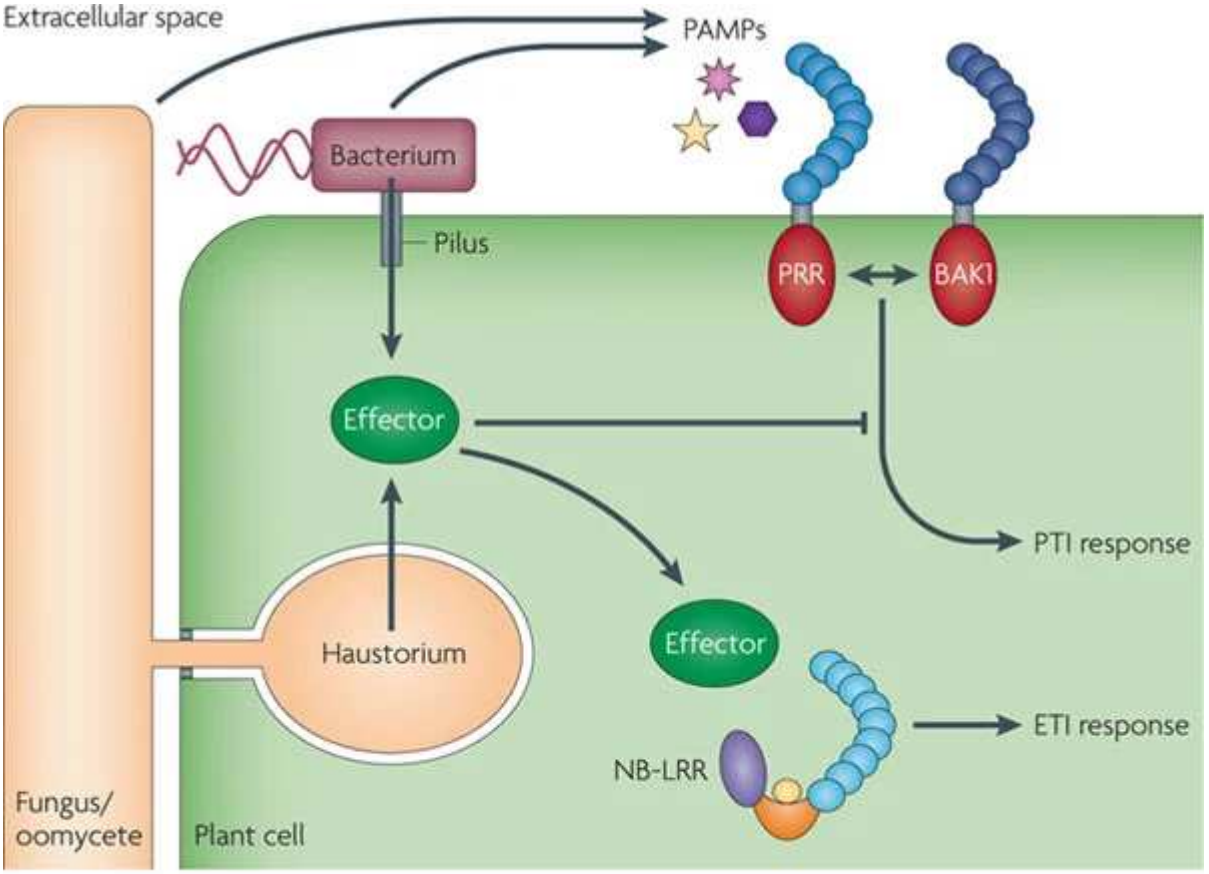
Por sua vez, os patógenos desenvolvem estratégias para burlar o maquinário de defesa das plantas. Não só a planta consegue reconhecer e se defender do patógeno, como o patógeno deve ter a capacidade de manipular a processos biológicos da planta para criar um ambiente favorável ao seu ciclo de vida. Em

consequência disso, tanto o patógeno quanto a planta têm evoluído um conjunto de genes que permitem essa interação.

As plantas são capazes de reconhecer patógenos no meio intracelular e extracelular, sendo essas as duas estŕatégias de resposta de defesa induzida. No meio extracelular, a célula faz uso de proteínas presentes nas membranas plasmáticas denominadas receptores de reconhecimento de padrão (pattern recognition receptors, PRR), que reconhecem elicitores microbianos conservados conhecidos como padrões moleculares associados a patógenos (pathogen-associated molecular patterns, PAMPs), desencadeando respostas de defesa contra os patógenos conhecidas como PTI (PAMP-triggered immunity) (DODDS & RATHJEN, 2010). Contudo, o patógeno, com a evolução, foi capaz de suprimir os diferentes componentes da PTI por meio de proteínas efetoras, que são proteínas secretadas pelos patógenos para acentuar a infecção e que ultrapassam as barreiras constitutivas e os PRRs. Nesse caso, uma segunda linha de defesa das plantas evoluiu, na qual ocorre a expressão de genes R, esses genes codificam proteínas R que interagem, direta ou indiretamente, com esses efetores dos patógenos. Esse reconhecimento desencadeia uma reação de resistência forte conhecida como ETI (Effector-Triggered Immunity) (DANGL & JONES, 2001). A ETI leva a uma dinâmica de evolução totalmente diferente da PTI, sistema que envolve moléculas conservadas indispensáveis à sobrevivência dos patógenos. Os efetores, por sua vez, apresentam alta taxa de variação (DODDS & RATHJEN, 2010; SILVA, 2016). Por isso, uma ampla diversidade de receptores e de efetores de patógenos no sistema ETI pode ser observada.

Apesar das respostas das similaridades entre os dois tipos de imunidade, as respostas causadas pela ETI tendem a ser mais fortes, envolvendo frequentemente apoptose de células infectadas, que resulta na denominada resposta hipersensível (DODDS & RATHJEN, 2010; SILVA, 2016). A PTI costuma ser eficiente contra patógenos não adaptados, devido às moléculas envolvidas serem conservadas ao longo das gerações. Por outro lado, ETI é eficaz contra patógenos adaptados, uma vez que há notável diversificação dos receptores e efetores envolvidos nesse sistema (TSUDA & KATAGIRI, 2010).

Figura 02 - Modelo esquemático de estratégias de respostas de defesa induzida de plantas.



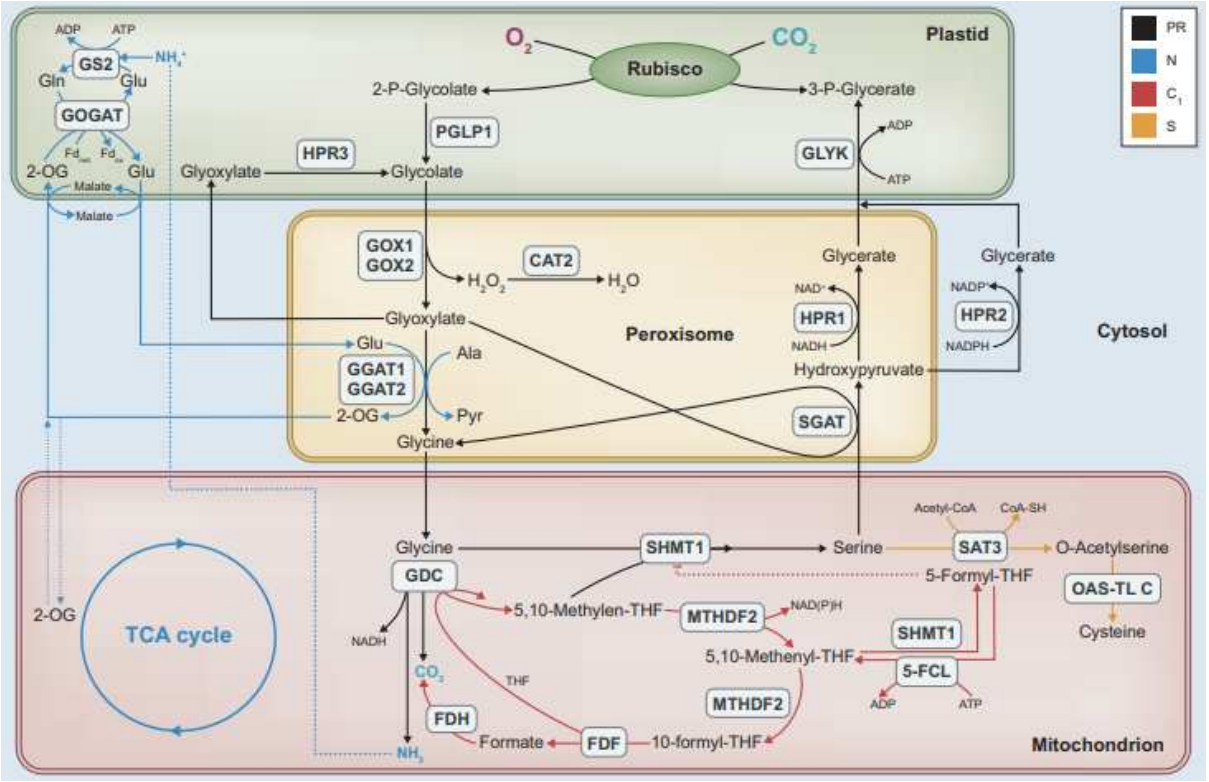
Fonte: DODDS & RATHJEN, 2010
As figuras em formatos de estrelas e hexágonos representam os PAMPs, que são liberados pelos patógenos. Os PAMPs são reconhecidos na região extracelular pelo domínio LRR (estruturas azuis) dos PRRs, localizados nas barreiras constitutivas da célula. Os PRRs apresentam o domínio quinase intracelular (em vermelho). A detecção das PAMPs pelos PRRs induz a respostas via PTI. Efetores liberados por patógenos no meio intracelular (em verde) geralmente atuam para suprimir a PTI. Porém, muitos são reconhecidos por receptores NB-LRRs na região intracelular, induzindo respostas via ETI. Receptores NB-LRRs são constituídos por um domínio LRR (azul claro), um domínio central NB (laranja) ligado a ATP ou ADP (amarelo) e um domínio TIR ou CC (roxo).

2.4. Glicolato Oxidase (GOX)

A proteína (S)-2-hydroxiácido oxidase peroxissomal (EC: 1.1.3.1 no PDB Primary Data e EC: 1.1.3.15 no UniProt) ou glicolato oxidase (GOX) tem 369 aminoácidos, 40,286 Daltons e é classificada como uma oxirredutase (http://www.uniprot.org/uniprot/P05414). Atua na oxidação do glicolato em glioxilato e H_2O_2 nos peroxissomos (Figura 03). Tem atividade catalítica sobre agrupamentos de CH-OH com oxigênio como aceptor final de elétrons. Utiliza um mononucleotídeo de flavina (FMN) como cofator e participa do metabolismo do glioxilato e do

dicarboxilato (STERNBERG & LINGQVIST, 1997; AMORIM, 2018). Tem sua atividade nos peroxissomos associada a catalase (E.C. 1.11.1.6) (PATRIOTA, 2022).

Figura 03 - Metabolismo fotorrespiratório em *Arabidopsis thaliana.*



Fonte: Eisenhut et al. New Phytologist (2019, p. 1764)
Metabolismo fotorrespiratório em *A. thaliana.* Interdependência da fotorrespiração com o metabolismo do nitrogênio, C_1 e enxofre. Abreviações: PGLP1, phosphoglycolate phosphatase 1; GOX1, glycolate oxidase 1; GOX2, glycolate oxidase 2; CAT2, catalase 2; GGAT1, glutamate:glyoxylate aminotransferase 1; GGAT2, glutamate:glyoxylate aminotransferase 2; GDC, glycine decarboxylase complex; SHMT1, serine hydroxymethyltransferase 1; SGAT, serine:glyoxylate aminotransferase; HPR1, hydroxypyruvate reductase 1; HPR2, hydroxypyruvate reductase 2; HPR3, hydroxypyruvate reductase 3, GLYK, glycerate kinase, GS2, glutamine synthetase 2; GOGAT, glutamine:oxoglutarate aminotransferase; THF, tetrahydrofolate; MTHDF2, bifunctional 5,10-methylene-THF dehydrogenase/5,10-methenyl-THF cyclohydrolase; 5-FCL, 5-formyl-THF cycloligase; FDF, 10-formyl-THF deformylase; FDH, formate dehydrogenase; SAT3, serine o-acetyltransferase; OAS-TL C, o-acetylserine lyase isoform C.

Através da pesquisa de Han et al. (2022) é relatado que o inibidor H_2S atenua o estresse oxidativo em célula de *Arabidopsis thaliana* ao reprimir a atividade da GOX. O aumento de H_2S e sua enzima produtora L-cisteína desulfurase apresentaram aumento transitório em resposta ao estresse oxidativo intracelular.

Figura 04 - Estrutura da enzima 1GOX de *Spinacia oleracea*

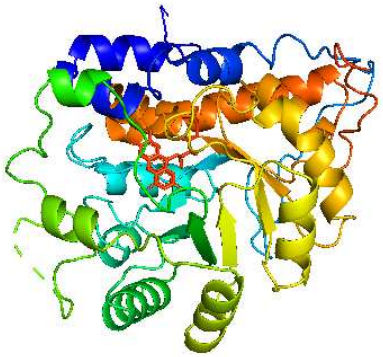


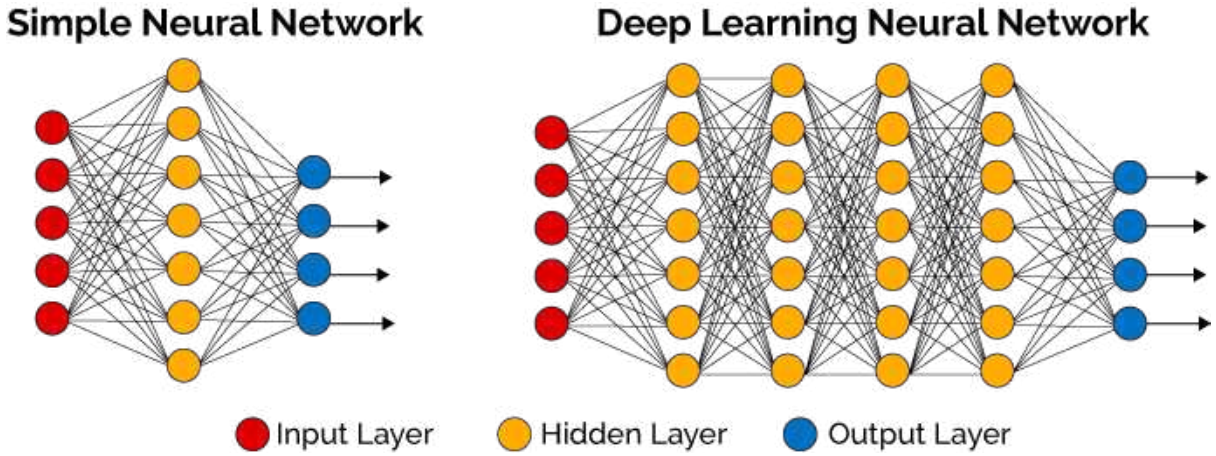
Imagem produzida pelo PyMOL Molecular Graphic System da estrutura de 1 GOX de *Spinacia oleracea* a partir da estrutura disponibilizada do Protein Data Bank.

2.5. Pacote Rosetta e Servidor Robetta

Deep Learning (aprendizado profundo, em portugues) trata-se de um conjunto de técnicas de Machine Learning (aprendizado de máquinas, em portugues) que utilizam redes neurais artificiais profundas, com muitas camadas intermediárias entre a camada de entrada e a de saída (Figura 05) (LECUN et al., 2015). Na literatura sobre *Deep Learning*, destacam-se alguns conceitos como sendo modelos que consistem em múltiplas camadas ou estágios de processamento de informação não linear; e também são métodos para a aprendizagem supervisionada ou não supervisionada da representação de características em camadas consecutivamente mais altas e mais abstratas. O *Deep Learning* abrange as áreas de pesquisa de redes neurais, inteligência artificial, modelagem gráfica, otimização, reconhecimento de padrões e processamento de sinais (HOSAKI & RIBEIRO, 2021).

Métodos de aprendizado profundo possibilitaram a evolução de diversas áreas de aprendizado de máquina especialmente por duas razões: disponibilidade de bases de dados com milhares de dados e computadores eficazes no processamento e em tempo consideravelmente reduzido (DENG et al., 2009; RUSSAKOVSKY et al., 2015).

Figura 05 - Modelo gráfico do funcionamento de Rede Neural simples e Aprendizado Profundo



Fonte: https://www.deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/

Rosetta é um pacote de softwares que possui algoritmos para a modelagem computacional e análise de estruturas tridimensionais de proteínas baseado em algoritmos de *deep learning*. Permitiu avanços científicos notáveis em biologia computacional, incluindo design de proteína *de novo*, design de enzima, docking molecular e previsão de estrutura de macromoléculas biológicas e complexos macromoleculares. O desenvolvimento do Rosetta começou no laboratório do Dr. David Baker na Universidade de Washington como uma ferramenta de previsão de estrutura, mas desde então foi adaptado para resolver problemas macromoleculares computacionais comuns (https://www.rosettacommons.org/software/).

Robetta é um servidor de previsão de estrutura de proteínas desenvolvido pelo Baker Lab da Universidade de Washington. Em seu núcleo está o conjunto de modelagem macromolecular Rosetta desenvolvido pelo Rosetta Commons, um grupo multi-institucional colaborativo de pesquisa e desenvolvimento de software. O principal serviço de Robetta é prever a estrutura tridimensional de uma proteína dada a sequência de aminoácidos. Cinco opções são fornecidas para a previsão de estrutura: (1) Um método baseado em deep learning, RoseTTAFold, (2) Outro método baseado em deep learning, TrRosetta, (3) Modelagem Comparativa Rosetta (RosettaCM), (4) Rosetta Ab Initio (RosettaAB) e (5) um pipeline totalmente automatizado que primeiro prevê domínios como unidades de dobragem independentes, modela cada unidade com (3) ou (4) e depois os monta em modelos

de cadeia completa. Os usuários têm a opção de renunciar a esses métodos fornecendo seus próprios modelos e alinhamentos e, opcionalmente, modificar os alinhamentos, adicionar restrições personalizadas e muito mais por meio de uma interface interativa na página de envio (https://robetta.bakerlab.org/faqs.php/).

3. Materiais e Métodos

3.1. Obtenção das sequências de aminoácidos

Foi obtida a sequência da proteína (S)-2-hydroxiácido oxidase peroxissomal GOX de *Spinacia oleracea* através do UniProt (http://www.uniprot.org/). Sua escolha foi devido a curadoria do banco de dados e por ser uma sequência já validada de GOX. As espécies alvo foram *Vigna unguiculata*, *Vigna radiata*, *Vigna angularis*, *Phaseolus vulgaris*, *Medicago truncatula*, sorgo (*Sorghum bicolor)*, soja (*Glycine Max)* e milho (*Zea mays)*.

Após a obtenção da sequência de aminoácidos pelo Uniprot, a sequência foi utilizada como referência para realizar buscas utilizando a ferramenta BLASTp (https://www.ncbi.nlm.nih.gov/BLAST/), que recebe uma sequência de aminoácidos como input (entrada), contra as espécies alvo nos bancos de dados disponíveis na ferramenta. Foi utilizado o banco de dados Non-redundant protein sequence(nr). Os critérios para seleção dos genes resultantes da busca foram o menor e-value possível e o maior Max Score para cada espécie.

3.2. Obtenção das estruturas tridimensionais

Para obtenção dos modelos de estruturas tridimensionais foi utilizada a ferramenta RoseTTAFold através do servidor Robetta (https://robetta.bakerlab.org/). Foram utilizadas as sequências obtidas resultante do BLASTp como dado de entrada. A ferramenta utiliza como input as sequências de aminoácidos no modelo FASTA e gera 5 modelos de estruturas no formato PDB para cada sequência de aminoácidos e um gráficos de erro estimado de cada aminoácido em Ångström para cada modelo gerado.

3.3. Análise de qualidade dos modelos de estruturas

Para a análise de qualidade dos modelos foram utilizados os parâmetros de root-mean-square deviation (RMSD) e Template modelling score (TM-Score).

3.3.1. O RMSD

O RMSD é uma medida comumente usada de semelhança entre duas estruturas de proteínas. Quanto menor o rmsd entre duas estruturas, mais semelhantes são essas duas estruturas. Na predição da estrutura da proteína, é

necessário o rmsd entre as estruturas preditas e experimentais para as quais uma predição pode ser considerada bem sucedida (REVA et al., 1998). Para a obtenção do RMSD foi utilizada a ferramenta PyMOL Molecular Graphic System (GL version 3.1), através do alinhamento entre a estrutura da 1GOX de *S. oleracea* obtida no UniProt e os modelos obtidos das estruturas alvos, em formato PDB.

3.3.2. TM-Score

O TM-score é uma métrica para avaliar a similaridade topológica de estruturas de proteínas. Ele foi projetado para resolver dois grandes problemas em métricas tradicionais, como o RMSD. O TM-score pondera erros de distância menores mais fortes do que erros de distância maiores e torna o valor da pontuação mais sensível à similaridade de dobra global do que às variações estruturais locais. O TM-score introduz uma escala dependente do comprimento para normalizar os erros de distância e torna a magnitude do TM-score independente do comprimento para pares de estruturas aleatórias. TM-score tem o valor entre 0 e 1, onde 1 indica uma combinação perfeita entre duas estruturas. Seguindo estatísticas estritas de estruturas no PDB, pontuações abaixo de 0,17 correspondem a proteínas não relacionadas escolhidas aleatoriamente (ZHANG & SKOLNICK, 2004; XU & ZHANG, 2010).

Para a obtenção do parâmetro do TM-Score foi utilizado a plataforma do Zhang Lab (https://zhanggroup.org/TM-score/). Para o cálculo foram utilizados os arquivos do modelo da estrutura da 1GOX de *S. oleracea* e os arquivos de modelo de estruturas obtidos pelo servidor Robetta em formato PDB. A ferramenta realiza um alinhamento/sobreposição entre as duas estruturas de entrada para realizar o cálculo.

3.3.3. Confiança pelo Servidor Robetta

Através do processo de se obter uma predição de estrutura pelo Robetta se obtém um valor métrico de Confidence (confiança) baseado no trabalho de Hiranuma et al. (2020). Para domínios de modelagem comparativa, a confiança corresponde à concordância na estrutura entre os modelos de rosca parcial (partial threaded models) a partir do alinhamento superior de cada método de alinhamento independente. A confiança para os domínios *ab initio* corresponde à TM-Score médio dos pares dos 10 principais modelos de pontuação Rosetta. Essas métricas

foram encontradas para se relacionar com o Teste Global de Distância (global distance test, GDT) real para nativo. Para os domínios RoseTTAFold, a confiança corresponde ao (Local Distance Difference Test, l-DDT ou lDDT) previsto usando DeepAccNet. lDDT é uma pontuação livre de superposição local para comparar estruturas e modelos de proteínas usando testes de diferença de distância (MARIANI et al, 2013).

4. Resultados e Discussão

4.1. Sequência de Aminoácidos

Através da busca por sequências similares pelo BLASTp fora obtido seguinte resultados:

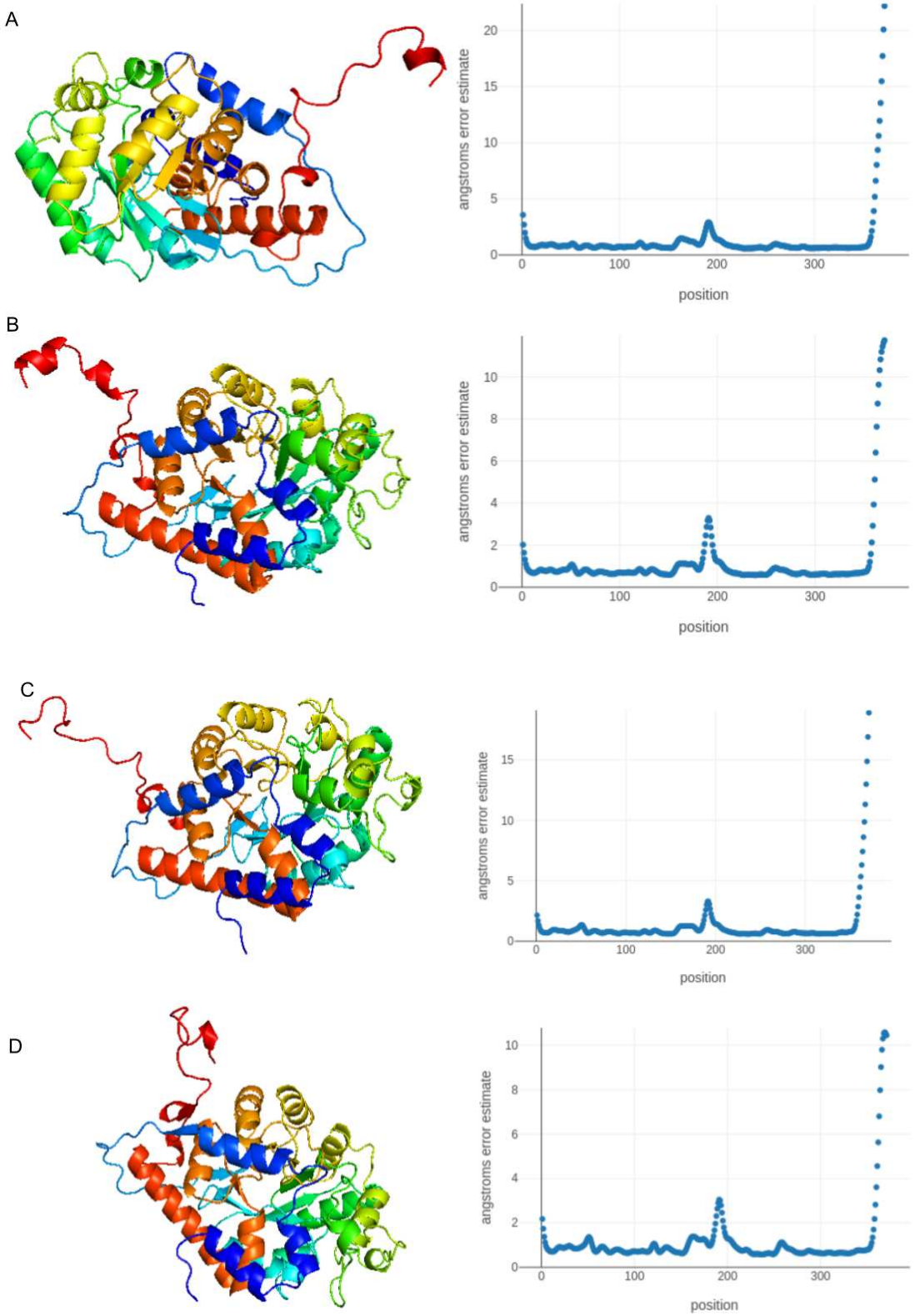
Tabela 01 - Resultado do BLASTp para GOX de *S. oleracea*

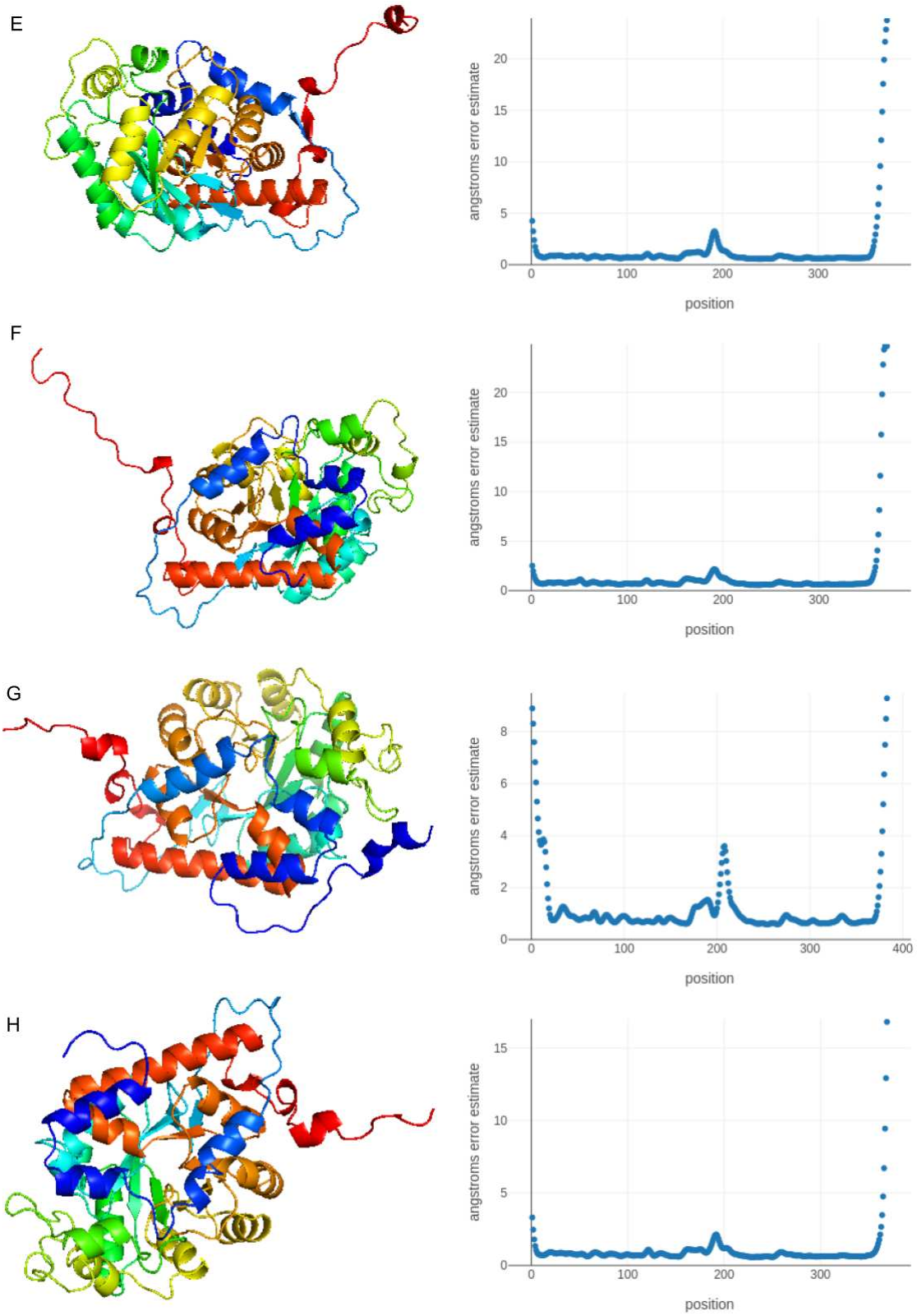
Descrição	Max Score	Total Score	E value	Accession
(S)-2-hydroxy-acid oxidase GLO1 isoform X1 [*Vigna unguiculata*]	669	669	0.0	XP_027936846.1
peroxisomal (S)-2-hydroxy-acid oxidase isoform X1 [*Medicago truncatula*]	667	667	0.0	XP_024633287.1
PREDICTED: peroxisomal (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [*Vigna angularis*]	665	665	0.0	XP_017425390.1
peroxisomal glycolate oxidase [*Phaseolus vulgaris*]	659	659	0.0	AGV54360.1
peroxisomal (S)-2-hydroxy-acid oxidase GLO1-like [*Glycine max*]	663	663	0.0	NP_001241302.1
peroxisomal (S)-2-hydroxy-acid oxidase GLO1 [*Vigna radiata var. radiata*]	659	659	0.0	XP_014501443.1
peroxisomal (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [*Sorghum bicolor*]	636	636	0.0	XP_021302176.1
RecName: Full=Glycolate oxidase 1; Short=GOX [*Zea mays*]	631	631	0.0	A0A3L6E0R4.1

4.2. Modelos preditos de estruturas tridimensionais de GOX

Através da ferramenta RosettaFold pelo servidor Roberta foram obtidos cinco modelos de estrutura para cada uma das sequências de GOX, ou seja, cinco estruturas por espécie, totalizando 40 estruturas. Em conjunto com os modelos obtidos um gráfico de erro estimado em Ångström (Å) para cada posição aminoácido para cada modelo obtido. Segue imagens das estruturas e gráficos abaixo.

Figura 06 - Modelos de estrutura tridimensional de GOX e seus respectivos gráficos de erro estimado por Ångström para cada aminoácido





Modelos de estrutura tridimensional de cada sequência obtida de GOX modelado através do RosettaFold pelo servidor Robetta e seus respectivos gráficos de erro estimado de cada aminoácido em Ångström. Região N-terminal em azul e região C-terminal em vermelho. A- *Vigna unguiculata*; B- *Vigna radiata*; C- *Vigna angularis*; D- *Phaseolus vulgaris*; E- *Medicago truncatula*; F- *Glycine max*; G- *Sorghum bicolor* e H- *Zea mays*.

As regiões das proteínas que apresentaram maior erro (Figura 06) foram as regiões C-Terminal (em vermelho), N-Terminal (em azul) e a região entre os aminoácidos 150 a 220 (região em verde e amarelo). A região final, os últimos 20 aminoácidos, de cada modelo obtido apresentou maior erro por distância devido a mau enovelamento da região reduzindo o grau de confiança. As principais hipóteses para o mal enovelamento das regiões terminais são:

1. Softwares que fazem modelagem computacional de proteínas tendem a ter um mal enovelamento das regiões terminais;
2. A dificuldade em medir a influência das outras regiões da molécula nas regiões terminais devido a distância.

4.3. Qualidade do modelos das estruturas

Através dos cálculos de RMSD, TM-Score e Confiança do Servidor Robetta foi obtido os seguintes valores:

Tabela 02 - Valores do RMSD para GOX de cada espécie

Espécie	RMSD					
	1	2	3	4	5	Média
Vigna unguiculata	0,840	0,871	0,877	0,882	0,880	0,870
Medicago truncatula	0,894	0,893	0,855	0,916	0,900	0,892
Vigna angularis	0,885	0,881	0,906	0,903	0,944	0,904
Phaseolus vulgaris	0,857	0,905	0,877	0,930	0,959	0,906
Glycine max	0,888	0,889	0,882	0,901	0,853	0,883
Vigna radiata var. radiata	0,913	0,883	0,852	0,852	0,859	0,872
Sorghum bicolor	0,909	0,898	0,814	0,848	0,854	0,865
Zea mays	0,837	0,867	0,871	0,890	0,900	0,873

Os valores obtidos de RMSD (Tabela 02) através do alinhamento entre as estruturas obtidas do RosettaFold contra a estrutura GOX de S. oleracea comprovam uma boa modelagem das estruturas. Todos obtiveram valores acima de 0,85, exceto uma estrutura de Z. mays que obteve 0,837 e um de V. unguiculata que recebeu 0,840. Contudo outros parâmetros foram necessários para reiterar essa afirmação.

Tabela 03 - Valores do TM-Score de GOX de cada espécie

Espécie	TM-Score					
	1	2	3	4	5	Média
Vigna unguiculata	0,7058	0,7063	0,7060	0,7059	0,7054	0,7059
Medicago truncatula	0,7067	0,7068	0,7074	0,7085	0,7071	0,7073
Vigna angularis	0,9116	0,9094	0,9101	0,9097	0,9091	0,9100
Phaseolus vulgaris	0,9097	0,9083	0,9094	0,9073	0,9085	0,9086
Glycine max	0,9099	0,9098	0,9092	0,9086	0,9105	0,9096
Vigna radiata var. radiata	0,9099	0,9090	0,9104	0,9109	0,9103	0,9101
Sorghum bicolor	0,1778	0,1788	0,1778	0,1766	0,1773	0,1777
Zea mays	0,7118	0,7119	0,7114	0,7121	0,7120	0,7118

Os valores do TM-Score de cada estrutura (Tabela 03) foram maiores que 0,7, com exceção do TM-Score das estruturas de *S. bicolor* que apresentaram média de 0,1777 e com nenhum acima de 0,2, sendo próximo ao valor dado para proteínas aleatórias não relacionadas (0,17) de acordo com os criadores do parâmetro Zhang e Skolnick (2004). Essa discrepância em relação às outras espécies pode ser dada pelo fato do TM-Score ponderar erros de distâncias menores mais fortes do que os erros de distâncias maiores. A principal hipótese para essa diferença filogenética de *S. bicolor* das outras espécies comparadas (Figura 07).

Figura 07 - Dendograma das sequências de Glicolato Oxidase

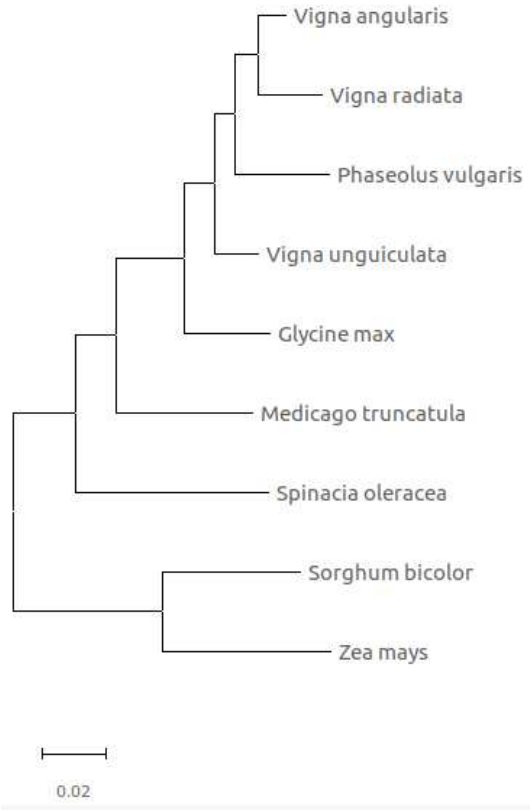


Tabela 04 - Valores de Confiança obtidos do Robetta para GOX de cada espécie

Espécie	Confiança
Vigna unguiculata	0.86
Medicago truncatula	0.87
Vigna angularis	0.86
Phaseolus vulgaris	0.85
Glycine max	0.88
Vigna radiata var. radiata	0.87
Sorghum bicolor	0.85
Zea mays	0.88

Com o parâmetro de confiança gerado pelo servidor Robetta (Tabela 04) ao ser utilizado para modelagem, todos obtiveram valores acima ou igual a 0,85. Mesmo com essa redução do grau confiança causada pelo mau enovelamento da região C-Terminal se obteve um alto grau de confiança (confiança ≥ 0.8). Com ressalvas ao *S. bicolor* que apresentou alta taxa de erro na região do C-Terminal e N-Terminal.

A ferramenta RosettaFold através do servidor Robetta é capaz de obter modelos de estruturas confiáveis devido a utilizar método *ab initio* junto com método *de novo* com maior precisão devido a sua constante melhoria por ser um método baseado em inteligência artificial/aprendizado profundo. A utilização do RosettaFold de necessita alto poder computacional para que sua atividade seja realizada de forma hábil e eficiente. Com a utilização do servidor Robetta é possível produzir modelos estruturais sem a necessidade de alto poder computacional com apenas o acesso a internet. Porém o servidor disponibiliza só o grau de confiança para verificar a qualidade da modelagem da estrutura. E o RobettaFold, fora do servidor Robetta, disponibiliza diversos dados adicionais para a utilização de outros parâmetros para verificação da qualidade do modelo.

Vale ressaltar que TM-Score e RMSD são parâmetros comparativos entre duas estruturas proteicas enquanto o parâmetro de confiança é baseado no numa estrutura de aprendizado profundo (DeepAccNet) que estima a precisão por resíduo e o erro assinado da distância resíduo-resíduo em modelos de proteínas e usa essas previsões para orientar o refinamento da estrutura da proteína Rosetta (HIRANUMA et al. 2020).

A variação dos resultados obtidos para cada uma das estruturas de uma mesma espécie pode ser explicada pela flexibilidade das estruturas proteicas, já que são estruturas dinâmicas que variam sua conformidade durante sua atividade.

A realização de docking molecular para verificar a interação proteína-ligante não foi possível devido à 5 erros que aconteceram ao tentar abrir o arquivo da estrutura da proteína no formato PDB na ferramenta de docking AutoDockTools na versão 1.5.6 e na versão 1.5.7. Outros arquivos de proteínas não modeladas através de RoseTTAFold não apresentaram o problema.

5. Conclusão

A Pacote Rosetta para modelagem de proteínas conseguiu modelar com qualidade bastante satisfatória dos modelos preditos obtidos para GOX em *Vigna unguiculata [L.] Walp, Vigna radiata, Vigna angularis, Phaseolus vulgaris, Medicago truncatula*, *Glycine Max* e *Zea mays*. Os resultados dos parâmetros avaliativos RMSD, TM-Score e Confiança reiteram esta afirmação. O resultado da modelagem de *Sorghum bicolor* não foi considerado satisfatório na avaliação do TM-Score e necessita de outros parâmetros avaliativos para verificar a qualidade dos modelos gerados.

Para estudos mais aprofundados sobre a estrutura de GOX é necessário resultados da cristalografia desta proteína nas espécies citadas nesta pesquisa e docking entre os modelos obtidos e o ligante para verificar a interação proteína-ligante.

Referências Bibliográficas

AMORIM, Deborah Douglas Damasceno. **Análises in silico de genes de vigna unguiculata (l.) walp envolvidos na interação compatível com o Cowpea Severe Mosaic Virus (CPSMV).** 2018. 68 f. Trabalho de Conclusão de Curso (Graduação em Biotecnologia) – Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2018.

BERTACINI, P. V.; ALMEIDA, A. M. R.; LIMA, J. A. A.; CHAGAS, C. M. Biological and physicochemical properties of cowpea severe mosaic Comovirus isolated from soybean in the State of Paraná. **Brazilian Archives of Biology and Technology**, Curitiba, v. 41, p. 409-416, 1998.

BOOKER, H. M., UMAHARAN, P., MCDAVID, C. R. Effect of Cowpea severe mosaic virus on crop growth characteristics and yield of cowpea. **Plant Disease**, v. 89, p. 515-520, 2005.

BRIOSO, P. S. T., SANTIAGO, L. J. M., ANJOS, J. R. N., OLIVEIRA D. E. Identificação de espécies do gênero Comovirus através de “polynerase chain reaction”. **Fitopatologia Brasileira,** v. 21, p. 219-225, 1996.

DANGL, J. L.; JONES, J. D. G. Plant pathogens and integrated defence responses to infection. **Nature**, v. 411, p. 826–833, 2001.

DENG, J. et al. ImageNet: A Large Scale Hierarchical Image Database. In **CVPR09**, 2009.

DODD, A. N., KUDLA, J., SANDERS, D. The Language of Calcium Signaling. **Annual Review of Plant Biology.** v. 61, p. 593–620, 2010.

DODDS, P. N.; RATHJEN, J. P. Plant immunity: towards an integrated view of plant–-pathogen interactions. **Nature Publishing Group**, v. 11, n. 8, p. 539–548, 2010.

EISENHUT, M., ROELL M.S., WEBER, A.P.M. Mechanistic understanding of photorespiration paves the way to a new green revolution. **New Phytologist**, v. 223: p. 1762–1769, 2019.

GONÇALVES, Alexandre et al. Cowpea (Vigna unguiculata L. Walp.), a renewed multipurpose crop for a more sustainable agri-food system: nutritional advantages and constraints. **Journal of the Science Food Agriculture**, v. 96, n. 9, p. 2941–2951, 2016.

HIRANUMA , N., PARK, H., BAEK, M., ANISHCHANKA, I., DAUPARAS, J. BAKER, D. Improved protein structure refinement guided by deep learning based accuracy estimation. **bioRxiv.** Novembro**,** 2020.

HOSAKI, G.Y., RIBEIRO, D.F. DEEP LEARNING: Ensinando a aprender. **Revista de Gestão e Estratégia.** v. 3, n. 1, 2021.

JONES, J. D. G.; DANGL, J. L. The plant immune system. **Nature,** v. 444, p. 323–329, 2006.

LECUN, Y. et al. Deep learning. **Nature**, 521(7553):436-444. 2015.

LIMA, J. A. A., NASCIMENTO, A. K., SILVA, G. S., CAMARÇO, R. F., GONÇALVES, F. B. Crotalaria paulinea, novo hospedeiro natural do vírus do mosaico severo do Caupi. **Fitopatologia Brasileira**, Brasília, v. 30, n. 4, p. 429-433, 2005.

MARIANI V, BIASINI M, BARBATO A, SCHWEDE T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. **Bioinformatics.** v. 29, n. 21, p. 2722-2728, 2013.

MELO, A. S., MELO, Y. L.,LACERDA, C. F., VIÉGAS, P. R. A., FERRAZ, R. L. S., GHEYI H. R. Water restriction in cowpea plants [Vigna unguiculata (L.) Walp.]: Metabolic changes and tolerance induction. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.26, n.3, p.190-197, 2022.

PATRIOTA, M. A. **Relationship between catalase and glycolate oxidase activities in response to contrasting photorespiratory conditions induced by different light intensities.** 2022. 52 f. Dissertação (Mestrado em Bioquímica) - Universidade Federal do Ceará, Fortaleza, 2022.

PEREIRA, R. F. **Caracterização bioquímica, nutricional e funcional de genótipos elite de feijão-caupi [Vigna unguiculata (L.) Walp].** 2013. 79 f. Dissertação (Mestrado em Bioquímica) - Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2013.

PINTO, J. V. **Propriedades físicas, químicas, nutricionais e tecnológicas de feijões (Phaseolus vulgaris L.) de diferentes grupos de cor.** 2016. 167 f. Dissertação (mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Goiás, Goiânia, 2016.

REVA, A.B., FINKELSTEIN, A.V., SKOLNICK, J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? **Folding & Design**, v. 3, p. 141–147, Março, 1998.

RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, 115(3):211-252, 2015

SILVA, R. G. G. **Respostas fisiológicas e bioquímicas do feijão-de-corda [Vigna unguiculata [L]. Walp] submetido ao estresse hídrico, infectado com o Vírus do Mosaico Severo do Caupi, e sob pressão dos dois estresses combinados.** 2016. Tese (Doutorado em Bioquímica) – Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2016.

SMARTT, J. Grain legumes: evolution and genetic resources. New York: **Cambridge University Press**, 1990.

SOARES, A.M.S., MACHADO, O.L.T. Defesa de plantas: Sinalização química e espécies reativas de oxigênio. **Revista Trópica – Ciências Agrárias e Biológicas,** v.1, n. 1, p. 9-19, 2007

STERNBERG, K., LINDQVIST, Y. Three-dimensional structures pf glycolate oxidase with bound active-site inhibitors. **Protein Science**, v. 6, p. 1009–1015. 1997.

TSUDA, K., KATAGIRI, F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. **Current Opinion in Plant Biology**, v. 13, p. 459-465, 2010.

ULLAH, R., ULLAH, Z., AL-DEYAB, S.S., ADNAN, M., TARIQ, A. Nutritional Assessment and Antioxidant Activities of Different Varieties of Vigna radiata. **The Scientific World Journal**, 2014.

WANG, L., MU, X., CHEN, X., HAN, Y. Hydrogen sulfide attenuates intracellular oxidative stress via repressing glycolate oxidase activities in Arabidopsis thaliana. **BMC Plant Biology.** 2022.

WANG, Y., YAO, X., SHEN, H., ZHAO, R., LI, Z., SHEN, X., WANG, F., CHEN, K., ZHOU, Y., LI, B., ZHENG, X., LU, S.Nutritional Composition, Efficacy, and Processing of Vigna angularis (Adzuki Bean) for the Human Diet: An Overview. **Molecules**, v. 27, 2022

XU, J. ZHANG, Y. How significant is a protein structure similarity with TM-score=0.5? **Bioinformatics**, v. 26, p. 889-895, 2010.

ZHANG, Y., SKOLNICK, J. Scoring function for automated assessment of protein structure template quality, **Proteins**, v. 57,p. 702-710, 2004.

Apêndices

Apêndice A - Tabela com as informações das sequências-alvo utilizadas resultantes da pesquisa por BLASTp

Descrição	Espécie	Max Score	Total Score	Query Cover (%)	E value	Percent Identity	Accession Length	Accession
(S)-2-hydroxy-acid oxidase GLO1 isoform X1 [*Vigna unguiculata*]	*Vigna unguiculata*	669	669	97%	0.0	91.39%	372	XP_027936846.1
peroxisomal (S)-2-hydroxy-acid oxidase isoform X1 [*Medicago truncatula*]	*Medicago truncatula*	667	667	99%	0.0	89.19%	372	XP_024633287.1
PREDICTED: peroxisomal (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [*Vigna angularis*]	*Vigna angularis*	665	665	99%	0.0	89.40%	371	XP_017425390.1
peroxisomal glycolate oxidase [*Phaseolus vulgaris*]	*Phaseolus vulgaris*	659	659	99%	0.0	88.32%	371	AGV54360.1
peroxisomal (S)-2-hydroxy-acid oxidase GLO1-like [*Glycine max*]	*Glycine max*	663	663	100%	0.0	88.68%	371	NP_001241302.1
peroxisomal (S)-2-hydroxy-acid oxidase GLO1 [*Vigna radiata var. radiata*]	*Vigna radiata var. radiata*	659	659	99%	0.0	88.32%	371	XP_014501443.1
peroxisomal (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [*Sorghum bicolor]*	*Sorghum bicolor*	636	636	97%	0.0	87.43%	383	XP_021302176.1
RecName: Full=Glycolate oxidase 1; Short=GOX [*Zea mays*]	*Zea mays*	631	631	98%	0.0	86.46%	369	A0A3L6E0R4.1

Apêndice B

Sequências-alvo no modelo FASTA

>XP_027936846.1 (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [Vigna unguiculata]
MGEITNVSEYEAIAKQKLPKMVFDYYASGAEDQWTLQENRNAFSRILFRPRILIDVS
KIDMTTTVLGFKISMPIMIAPTAMQKMAHPEGEYATARAASAAGTIMTLSSWATSSVE
EVASTGPGIRFFQLYVYKDRNVVAQLVRRAERAGFKAIALTVDTPRLGRREADIKNR
FTLPPFLTLKNFEGLDLGKMDKADDSGLASYVAGQIDRTLSWQDVKWLQTITTLPILV
KGVLTAEDTRIAVQSGAAGIIVSNHGARQLDYVPATISALEEVVKAAEGRLPVFLDGG
VRRGTDVFKALALGASGIFIGRPVVFSLAAEGEAGVRKVLQMLREEFELTMALSGC
RSLKEITRDHIVTDWDQPRVQPRVRAL

>XP_024633287.1 peroxisomal (S)-2-hydroxy-acid oxidase isoform X1 [Medicago truncatula]
MGEITNISEYEEIARQKLPKMAFDYYASGAEDQWTLQENRNAFSRILFRPRILIDVSKI
DLSTTVLGFKISMPIMIAPTAFQKMAHPEGEYATARAASAAGTIMTLSSWATSSVEEV
ASTGPGIRFFQLYVYKDRNVVAQLVRRAEKAGFKAIALTVDTPRLGRREADIKNRFV
LPPFLTLKNFEGLNLGKMDEANDSGLASYVAGQIDRTLSWKDVKWLQTITSLPILVK
GVLTAEDARLAVQSGAAGIIVSNHGARQLDYVPATISALEEVVKAAQGRVPVFLDGG
VRRGTDVFKALALGASGIFIGRPVVYSLAAEGEVGVRKVLQMLRDEFELTMALSGC
RSLKEITRDHIVADWDTPRVNPRAIPRL

>XP_017425390.1 PREDICTED: peroxisomal (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [Vigna angularis]
MEVTNVSEYEAIAKQKLPKMAFDYYASGAEDQWTLQENRNAFSRILFRPRILIDVSKI
DMTTTVLGFKISMPIMIAPTAFQKMAHPEGEYATARAASAAGTIMTLSSWATSSVEE
VASTGPGIRFFQLYVYKDRNVVAQLVRRAERAGFKAIALTVDTPRLGRREADIKNRF
TLPPFLTLKNFEGLNLGTMDKADDSGLASYVAGQIDRTLSWKDVKWLQTITSLPILVK
GVLTAEDTRIAVQSGAAGIIVSNHGARQLDYVPATISALEEVVKAAEGRIPVFLDGGV
RRGTDVFKALALGASGIFIGRPVVYSLAAEGEAGVRKVLQMLREEFELTMALSGCR
SLKEITRDHIVTDWDHPRIQPRARALL

>AGV54360.1 peroxisomal glycolate oxidase [Phaseolus vulgaris]
MEVTNVSEYEAIAKQKLPKMAFDYYASGAEDQWTLQENRNAFSRILFRPRILIDVSKI
DITTTVLGFKISMPIMIAPTAFQKMAHPEGEYATARAASAAGTIMTLSSWATSSVEEV
ASTGPGIRFFQLYVYKDRNGVAQLGRRAERAGFKAIALTVDTPILGRREADIKNRFTL
PPFLTLKNFEGLDLGKMDKANDSGLASYVSGQIDRTLSWKDVKWLQTITSLPILVKG
VLTAEDTRIAIQSGAAGIIVSNHGARQLDYVPATISALEEVVKAAEGRLPVFLDGGVR
RGTDVFKALALGASGIFIGRPVVFSLAAEGEAGVRKVLQMLREEFELTMALSGCRSL
KEITRDHIVTDWDQPRTHPRTRALL

>NP_001241302.1 peroxisomal (S)-2-hydroxy-acid oxidase GLO1-like [Glycine max]
MEITNVSEYEAIAKQKLPKMVFDYYASGAEDQWTLQENRNAFSRILFRPRILIDVSKI
DITTTVLGFKISMPIMLAPTAMQKMAHPEGEYATARAASAAGTIMTLSSWATSSVEEV
ASTGPGIRFFQLYVYKDRNVVAQLVRRAERAGFKAIALTVDTPRLGRREADIKNRFT
LPPFLTLKNFEGLDLGKMDKADDSGLASYVAGQIDRTLSWKDVKWLQTITKLPILVK
GVLTAEDTRIAVQSGAAGIIVSNHGARQLDYVPATISALEEVVKAAEGRVPVFLDGGV
RRGTDVFKALALGASGIFIGRPVVFSLAAEGEAGVRNVLRMLREEFELTMALSGCTS
LKDITRDHIVTDWDQPRTIPRALPRL

>XP_014501443.1 peroxisomal (S)-2-hydroxy-acid oxidase GLO1 [Vigna radiata var. radiata]
MEITNVSEYEAIAKQKLPKNAFDYYASGAEDQWTLQENRNAFSRILFRPRILIDVSKI
DLTTTVLGFKISMPIMIAPTAFQKMAHPEGEYATARAASAAGTIMTLSSWATSSVEEV
ASTGPGIRFFQLYVYKDRNVVAQLVRRAERAGFKAIALTVDTPRLGRREADIKNRFT
LPPHLTLKNFEGLNLGKMDKADDSGLASYVAGQIDRTLSWKDVKWLQTITSLPILVK
GVLTAEDTRIALQNGAAGIIVSNHGARQLDYVPATISALEEVVKAAEGRVPVFLDGGV
RRGTDVFKALALGASGIFIGRPVVFSLAAEGEAGVRKVLQLLREEFELTMALSGCRS
LKEITRDHVVTDWDHPRIQPRARALL

>XP_021302176.1 peroxisomal (S)-2-hydroxy-acid oxidase GLO1 isoform X1 [Sorghum bicolor]
MTAEYSVEGAPVADTMGEITNVMEYQAIAKQKLPKMAYDYYASGAEDEWTLKENR
EAFSRILFRPRILIDVSKIDMTTSVLGFKISMPIMVAPTAMQKMAHPDGEYATARAASA
AGTIMTLSSWATSSVEEVASTGPGIRFFQLYVHKDRKVVEQLVRRAERAGFKAIALT
VDTPRLGRREADIKNRFVLPPHLTLKNFEGLDLGKMDQANDSGLASYVAGQIDRTL

SWKDVKWLQSITSMPILVKGVVTAEDARLAVHSGAAGIIVSNHGARQLDYVPATISAL
EEVVKAAQGRIPVYLDGGVRRGTDVFKALALGAAGIFVGRPVVFALAAEGEAGVRN
VLRMLRDEFELTMALSGCTTLADINRSHVLTEGDRLRPTPRL

>sp|A0A3L6E0R4.1|GOX1_MAIZE RecName: Full=Glycolate oxidase 1; Short=GOX
MGEITNVMEYQAIAKQKLPKMAYDYYASGAEDEWTLQENREAFSRILFRPRILIDVS
KIDMTTTVLGFKISMPIMVAPTAMQKMAHPDGEYATARAAAAAGTIMTLSSWATSSV
EEVASTGPGIRFFQLYVYKDRKVVEQLVRRAERAGFKAIALTVDTPRLGRREADIKN
RFVLPPHLTLKNFEGLDLGKMDQAADSGLASYVAGQVDRTLSWKDVKWLQTITTLP
ILVKGVLTAEDTRLAVANGAAGIIVSNHGARQLDYVPATISALEEVVKAARGQLPVFV
DGGVRRGTDVFKALALGAAGVFVGRPVVFSLAAAGEAGVSNVLRMLRDEFELTMA
LSGCTSLAEITRKHIITESDKLSAIPSRL

Apêndice C

Tutorial do RosettaFold através do servidor Robetta

1. Acessar a plataforma do servidor através do link (https://robetta.bakerlab.org/)
2. Para conseguir utilizar o servidor é necessário realizar um cadastro.
   - **2.1.** Vá em **Register** (Canto superior direito da tela) -> Preencher com os dados necessários e finalize o cadastro.
3. Fazendo a modelagem da proteína
   - **3.1.** **Structure Prediction** -> **Submit** (Canto superior esquerdo);
   - **3.2.** Preencha o formulário com um nome para identificação no **Target Name** e em **Protein Sequence** coloque a sequência da proteína em formato FASTA ou faça o upload do arquivo FASTA;
   - **3.3.** Tem as 4 opções de modelagem:
     - RoseTTAFold
     - CM (Modelagem comparativa)
     - AB (Ab Initio)
     - Predict domains
   - **3.4.** Selecione RoseTTAFold
4. Responda a operação de 3+2=
5. Se tem a opção de manter o resultado privado para isso é só marcar a caixa **Keep Private**
6. Submeta a modelagem clicando em **Submit**
7. Para ver o resultado da modelagem é só acessar no canto superior direito a parte do usuário e clicar em **My Queue** onde mostra os resultados ou pode acessar pelo email que será recebido no endereço de email que foi registrado na conta.

OBS: Seus resultados têm um prazo para serem acessados, depois que o prazo acaba os dados da modelagem são automaticamente deletados.