



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE BIOTECNOLOGIA
BACHARELADO EM BIOTECNOLOGIA

GUSTAVO ANDREW MAHON MENDES PEREIRA

Identificação e validação *in silico* de proteínas essenciais de *Corynebacterium ulcerans* como potenciais alvos para fármacos: Estudo de caso com derivados de alcalóides tetraisoquinolínicos.

JOÃO PESSOA/PB

2020

GUSTAVO ANDREW MAHON MENDES PEREIRA

Identificação e validação *in silico* de proteínas essenciais de *Corynebacterium ulcerans* como potenciais alvos para fármacos: Estudo de caso com derivados de alcalóides tetraisoquinolínicos.

Trabalho de conclusão de curso apresentado ao Centro de Biotecnologia da Universidade Federal da Paraíba, como requisito para obtenção do título de Bacharel em Biotecnologia.

Orientador: Edson Luiz Folador

JOÃO PESSOA/PB

2020

Catálogo na publicação
Seção de Catalogação e Classificação

P436i Pereira, Gustavo Andrew Mahon Mendes.

Identificação e validação in silico de proteínas essenciais de *Corynebacterium ulcerans* como potenciais alvos para fármacos: Estudo de caso com derivados de alcalóides tetraisoquinolínicos. / Gustavo Andrew Mahon Mendes Pereira. - João Pessoa, 2020.

118 f. : il.

Orientação: Edson Luiz Folador.
Monografia (Graduação) - UFPB/Cbiotec.

1. *Corynebacterium ulcerans*. 2. Interolog mapping. 3. Rede de interação proteína-proteína. 4. Proteínas essenciais. 5. Docking molecular. 6. Alcalóides tetrahidroisoquinolínicos. 7. Candidatos a droga. I. Folador, Edson Luiz. II. Título.

UFPB/BC



UNIVERSIDADE FEDERAL DA PARAÍBA (UFPB)
CENTRO DE BIOTECNOLOGIA (CBiotec)
CAMPUS I – JOÃO PESSOA/PB
Coordenação do Curso de Bacharelado em
Biotecnologia



ATA DE DEFESA PÚBLICA DO TRABALHO DE CONCLUSÃO DE CURSO

Aos trinta e um dias do mês de julho de 2020, às 14:00h, em sessão pública realizada de forma remota, na Plataforma Google Meet, em atendimento à Portaria N°231/GR/REITORIA/UFPB, de 22 de julho de 2020 e a Resolução Consepe N° 13/2020, na presença da Banca Examinadora presidida pelo(a) Professor(a) Dr(a). Edson Luiz Folador (DB/CBIOTEC/UFPB) e composta pelo(a)s avaliadore(a)s: 1. Prof(a). Dr(a). Enéas Ricardo de Moraes Gomes (DB/CBIOTEC/UFPB); 2. Luis Cezar Rodrigues (DB/CBIOTEC/UFPB), o(a) discente Gustavo Andrew Mahon Mendes Pereira, matrícula 11512582, apresentou o Trabalho de Conclusão de Curso intitulado: **“Identificação e validação *in silico* de proteínas essenciais de *Corynebacterium ulcerans* como potenciais alvos para fármacos: Estudo de caso com derivados de alcaloides tetraisoquinolínicos”**, como requisito curricular indispensável para a integralização do Curso de Graduação em Biotecnologia. Após reunião em sessão reservada, a Banca Examinadora deliberou e decidiu pela APROVAÇÃO do referido trabalho, divulgando o resultado formalmente a(ao) discente e demais presentes e eu, Edson Luiz Folador, na qualidade de Presidente da Banca, lavrei a presente ata que será assinada por mim, pelo(a)s demais avaliadore(a)s e pelo(a) discente.

Presidente da Banca Examinadora

Avaliador(a) 1

Discente

Avaliador(a) 2

João Pessoa/PB, 31 de julho de 2020.

DEDICATÓRIA

Dedico primeiramente ao meu orientador, pois sem ele este trabalho não seria possível, sempre se fazendo presente na execução das atividades e no tirar das dúvidas, só tenho a agradecer pela oportunidade, pela paciência e por todo o seu tempo ao trabalho e a mim dedicado. Ao meu pai, que me inspirou como profissional e que trabalhou arduamente para me garantir uma educação de qualidade. À minha mãe, que sempre altruísta me inspirou como pessoa e me proporcionou as melhores condições de vida. Ao meu irmão, que por diversas vezes facilitou meu acesso à faculdade. Ao meu melhor amigo Igor Costa por todo dia ter com que compartilhar as melhores experiências e gostos pessoais. À Kisie Ainoã, pela duradoura amizade e por me aturar como dupla em atividades e avaliações durante todo o curso. Ao Lucas Matheus, pela amizade e por sempre estimular e ampliar meu conhecimento. Aos amigos de curso: Hércules de Almeida, Natália Santos e Ana Beatriz pelo companheirismo e preocupação mútua em momentos de dificuldade. Aos outros amigos e colegas de curso por aceitarem meu jeito de ser (pelo menos a maioria) e contribuírem com minha formação acadêmica e crescimento pessoal. E aos professores de curso, por todo o conhecimento e ajuda proporcionada, principalmente aos que foram flexíveis nas avaliações em momentos de dificuldade para falar em público.

AGRADECIMENTOS

Agradeço ao meu orientador, prof. e Dr. Edson Luiz Folador, por me permitir dar continuidade ao projeto em andamento e apresentá-lo na forma de Trabalho de Conclusão de Curso, assim como a oportunidade de trabalhar como bolsista durante a Iniciação Científica. Agradeço também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudos e recursos provenientes da chamada Universal 2016 sob processo 404275/2016-8. Aos colegas de trabalho Luiz Melo, Cayro Mendes e Annie Beltrão por toda a ajuda proporcionada diretamente com o trabalho. E aos professores pela dedicação em nos proporcionar ensino de qualidade mesmo sem os recursos disponíveis para isso.

RESUMO

Com o avanço e utilização das técnicas de alto rendimento uma enorme quantidade de interações proteína-proteína estão disponíveis em diversos bancos de dados. Interações proteína-proteína são importantes pois permitem determinar se uma proteína é essencial para a manutenção da vida de um organismo. Identificar dentre as proteínas de um organismo quais lhe são essenciais é importante para diversas aplicações, como o desenvolvimento de fármacos. Procuramos neste trabalho, pelo método de *interolog mapping*, prever a rede de interação proteína-proteína visando identificar e validar a essencialidade das proteínas *hub* de *Corynebacterium ulcerans*, sugerindo assim o seu uso como novos alvos para candidatos a drogas. O microrganismo em questão é uma das bactérias causadoras da difteria com maior potencial mutagênico, por ter como hospedeiros não só humano como também animais e por conta da difteria se manter endêmica em países como Índia e Venezuela. Identificamos 457 proteínas *hub* na rede de *C. ulcerans* e destas 421 (92,12%) tiveram essencialidade comprovada e 36 (7,88%) consideradas essenciais após análises funcionais e de enriquecimento. Com o potencial de serem utilizadas como alvo para drogas, tivemos 351 (76,8%) proteínas *hub* sem homologia com *H. sapiens*. Dentre estas 99 (21,66%) tiveram estrutura tridimensional predita, validada e com *pockets* drogáveis, viáveis para *docking*. No estudo de caso com as proteínas alvo dentre 42 derivados sintéticos de alcalóides tetrahydroisoquinolínicos foram identificados os seis mais promissores para serem usados como droga.

Palavras chaves: *Corynebacterium ulcerans*; *interolog mapping*; rede de interação proteína-proteína; proteínas essenciais; *docking* molecular; alcalóides tetrahydroisoquinolínicos; candidatos a droga;

ABSTRACT

With the advance and utilization of high throughput technologies a large number of protein–protein interactions were made available on a number of databases. Protein–protein interactions are important because they allow us to identify in a number of proteins which are essential for the maintenance of an organism's life. Proteins identified as essential for an organism can then be used in a number of applications, including drug design. In this paper, we aim to predict the protein-protein interaction network of *Corynebacterium ulcerans* by making use of the interolog mapping method, to identify and validate the essentiality of hub proteins and suggest its use as new targets for drug candidates. *C. ulcerans* is one of the most important bacteria that causes diphtheria, it has a considerable mutagenic potential for having not only human as hosts but animals too and because diphtheria is still endemic in countries like India and Venezuela. We managed to identify 457 hub proteins in the network of *C. ulcerans* and, of those, 421 (92,12%) had its essentiality verified and 36 (7,88%) considered essential after functional and enrichment analyses. With the potential of being used as targets for drugs, we had 351 (76.8%) hub proteins without homology with *H. sapiens*. Of those, 99 (21.66%) had a predicted and validated three-dimensional structure with druggable pockets, viable for docking. In the case study with the target proteins, of 42 synthetic tetrahydroisoquinoline alkaloids, the six most promising to be used as drugs were identified.

Keywords: *Corynebacterium ulcerans*; interolog mapping; protein-protein interaction network; essential proteins; molecular docking; tetrahydroisoquinoline alkaloids; drug candidates;

LISTA DE FIGURAS

Figura 1 - Representação das membranas de bactérias gram-negativas e gram-positivas	16
Figura 2 - Coloração da <i>Corynebacterium glutamicum</i> com vancomicina-FL mostrando a formação de dímeros celulares em forma de V	17
Figura 3 - Representação da estrutura do envelope celular corinebacteriano.....	18
Figura 4 - Exemplo de interação proteína-proteína	24
Figura 5 - Representação de uma interação na forma de grafos	24
Figura 6 - Rede direcionada e não direcionada.....	25
Figura 7 - Etapas para modelagem realizadas pelo Phyre2.....	36
Figura 8 - Visão geral dos ângulos de torção (ϕ , ψ , ω e χ_1) da cadeia central de proteínas	37
Figura 9 - Tratamentos e definições clássicas de ϕ e ψ	38
Figura 10 - Alinhamentos estruturais entre diferentes <i>softwares</i>	39
Figura 11 - Representação das forças de ligação	40
Figura 12 - Comparação estrutural entre isoquinolinas e tetrahydroisoquinolinas	42
Figura 13 - Fluxo representativo dos métodos CADD (<i>Computer-Aided Drug Design</i>)	44
Figura 14 - Fluxograma de predição das redes de interação proteína-proteína (RIP) das dez linhagens de <i>C. ulcerans</i>	50
Figura 15 - Fluxograma de seleção das proteínas alvo	52
Figura 16 - Fluxograma para execução do <i>virtual screening</i>	54
Figura 17 - <i>Docking</i> da molécula 23 competindo com o DNA na <i>UvrABC system protein B</i>	71
Figura 18 - Acoplamento e interações da molécula 23 com <i>UvrABC system protein B</i>	72
Figura 19 - Acoplamento e interações da molécula 27 com <i>UvrABC system protein B</i>	72
Figura 20 - Acoplamento e interações da molécula 24 com <i>UvrABC system protein B</i>	73

Figura 21 - Acoplamento e interações entre a molécula 26 e <i>UvrABC system protein B</i>	73
Figura 22 - Acoplamento e interações da molécula 41 com <i>Bifunctional RNase H/acid phosphatase</i>	74
Figura 23 - <i>Docking</i> da molécula 34 competindo pelo sítio de ligação com o ADP na <i>Bifunctional protein</i>	75
Figura 24 - Acoplamento e das interações da molécula 34 com <i>Bifunctional protein</i>	75
Figura 25 - Acoplamento e das interações da molécula 28 com <i>MviN-like protein</i> ...	76
Figura 26 - <i>Docking</i> da molécula 20 competindo pelo sítio de ligação com uridina-difosfato-n-acetilglicosamina na <i>UDP-N-acetylglucosamine 1-carboxyvinyltransferase</i>	77
Figura 27 - Acoplamento e interações da molécula 20 com <i>UDP-N-acetylglucosamine 1-carboxyvinyltransferase</i>	77
Figura 28 - Acoplamento e interações da molécula 27 com <i>UDP-N-acetylglucosamine 1-carboxyvinyltransferase</i>	78
Figura 29 - <i>Docking</i> da molécula 38 competindo com L-histidinol pelo mesmo sítio de ligação na <i>Histidinol dehydrogenase</i>	79
Figura 30 - Acoplamento e interações da molécula 38 com <i>Histidinol dehydrogenase</i>	79

LISTA DE GRÁFICOS

Gráfico 1 - Gráfico de distribuição de probabilidade $P(k)$ para o número de conexões (k) de cada proteína	32
Gráfico 2 - Distribuição do grau de interação das proteínas da RIP da linhagem FRC58	58
Gráfico 3 - Distribuição das interações nos organismos de origem.....	59
Gráfico 4 - Análise funcional das proteínas <i>hub</i> envolvidas com o metabolismo	61
Gráfico 5 - Análise funcional das proteínas <i>hub</i> envolvidas com o processamento e armazenamento de informações	62
Gráfico 6 - Análise funcional das proteínas <i>hub</i> envolvidas com processos e sinalização celular	63
Gráfico 7 - Análise funcional das proteínas <i>hub</i> fracamente categorizadas	64
Gráfico 8 - Análise de enriquecimento referente aos processos biológicos	66
Gráfico 9 - Análise de enriquecimento referente às funções celulares.....	67
Gráfico 10 - Análise de enriquecimento referente aos componentes celulares.....	68

LISTA DE TABELAS

Tabela 1 - Genomas de <i>Corynebacterium ulcerans</i> completamente sequenciados .	20
Tabela 2 - Lista das bases de dados públicas	49
Tabela 3 - Genoma completo de dez linhagens de <i>C. ulcerans</i>	49
Tabela 4 - Quantidade de alinhamentos e interações para cada linhagem de <i>C. ulcerans</i>	57
Tabela 5 - Ligantes com os melhores resultados de <i>docking</i>	70
Tabela 6 - Proteínas <i>hub</i> de <i>C. ulcerans</i> com os melhores resultados de <i>docking</i> ...	70
Tabela 7 - Dados e propriedades dos cinco ligantes com melhores médias de <i>score</i>	80

LISTA DE ABREVIATURAS E SIGLAS

3D	Tridimensional
ADME	Absorção, distribuição, metabolismo e excreção
ADP	Adenosina difosfato
AP-MS	<i>Affinity purification/mass spectrometry</i>
AR	Alinhamento recíproco
BHE	Barreira hematoencefálica
CADD	<i>Computer-aided drug design methods</i>
COG	<i>Cluster of Orthologous Genes</i>
CTP	Citidina trifosfato
DAT	Antitoxina diftérica (do inglês <i>diphtheria antitoxin</i>)
DEG	<i>Database of Essential Genes</i>
DEGREE	Grau de interação de um nodo (proteína)
DIP	<i>Database of Interacting Proteins</i>
DNA	Ácido desoxirribonucleico
DS	<i>Druggability score</i>
DSV	<i>Discovery Studio Visualizer</i>
DTP	<i>Diphtheria, Tetanus e Pertussis</i>
GO	<i>Gene Ontology</i>
HDH	<i>Histidinol dehydrogenase</i>
HMM	Modelo Oculto de Markov
HUB	Nodo (proteína) com alto grau de interação (<i>degree</i>)
HTS	<i>High-throughput screening</i>
HPRD	<i>The Human Protein Reference Database</i>
INTACT	<i>Molecular Interaction Database</i>
IPP	Interação proteína-proteína
LBVS	<i>Ligand Based Virtual Screening</i>
ME3D	Microscopia Eletrônica 3D
MS	Ministério da Saúde

NCBI	National Center for Biotechnology Information
OMS	Organização Mundial de Saúde
OPP	<i>Oligopeptide permeases</i>
QSAR	Relação Quantitativa Estrutura-Atividade
PDB	Protein Data Bank
RIP	Rede de interação proteína-proteína
RMN	Ressonância Magnética Nuclear
RMSD	<i>Root Mean Square Deviation</i>
RNA	Ácido ribonucleico
SMILES	<i>Simplified molecular-input line-entry system</i>
RO5	<i>Rule of five</i>
SBVS	<i>Structure Based Virtual Screening</i>
STRING	<i>Search Tool for the Retrieval of Interacting Genes/Proteins</i>
TGI	Trato Gastrointestinal
TRX	Tioredoxinas
TTP	Timidina trifosfato
TD	Toxina diftérica
UTP	Uridina trifosfato
WHO	World Health Organization
Y2H	<i>Yeast two hybrid</i>

SUMÁRIO

1 INTRODUÇÃO	15
2 FUNDAMENTAÇÃO TEÓRICA.....	16
2.1 <i>Corynebacterium</i>	16
2.1.1 Características	17
2.1.2 Espécies de <i>Corynebacterium</i>	19
2.2 <i>Corynebacterium ulcerans</i>	19
2.3 <i>Difteria</i>	21
2.4 Biologia de Sistemas	23
2.4.1 Interação Proteína-Proteína (IPP).....	23
2.4.2 Rede de interação proteína-proteína (RIP)	25
2.4.3 Propriedades topológicas de rede.....	26
2.4.4 Métodos para predição de IPP	27
2.4.4.1 Métodos experimentais.....	27
2.4.4.2 Métodos computacionais	28
2.4.5 Bases de dados de IPP	29
2.5 Proteínas essenciais.....	31
2.6 Estrutura tridimensional proteica	32
2.6.1 Métodos experimentais	33
2.6.2 Métodos computacionais de predição de estrutura tridimensional.....	34
2.6.3. Programas de modelagem	35
2.6.4 Validação de estruturas tridimensionais.....	36
2.7 Ligantes	40
2.7.1. Alcalóides tetrahydroisoquinolínicos	41
2.8 <i>Computer-Aided Drug Design Methods (CADD)</i>	43
2.8.1 <i>Ligand Based Virtual Screening (LBVS)</i>	44
2.8.2 <i>Structure Based Virtual Screening (SBVS)</i>	45
3. OBJETIVOS	48
3.1 Objetivo geral.....	48
3.2 Objetivos específicos	48
4 METODOLOGIA.....	49

4.1 Bancos de dados biológicos	49
4.2 Genomas completos de <i>C. ulcerans</i>	49
4.3 Predição das Redes de Interação Proteína-Proteína (RIP)	50
4.3.1 Alinhamento <i>C. ulcerans</i> contra STRING	50
4.3.2 Mapeamento das interações	51
4.4 Seleção das proteínas alvo	52
4.4.1 Identificação e validação das proteínas <i>hub</i>	53
4.4.2 Homologia com hospedeiro	53
4.5 Estudo de caso	54
4.5.1 Proteínas alvo	55
4.5.2 Derivados sintéticos de alcalóides tetrahydroisoquinolínicos	55
4.5.3 Execução do <i>virtual screening</i> (VS)	56
4.5.4 Análise de <i>docking</i>	56
5 RESULTADOS E DISCUSSÃO	57
5.1 Redes de Interação Proteína-Proteína	57
5.2 Validação da Rede de Interação	58
5.3 Proteínas <i>hub</i>	59
5.3.1 Análise funcional (COG)	60
5.3.2 Análise de Enriquecimento	64
5.4 Estudo de caso	68
5.4.1 Análise dos melhores complexos proteína-ligante	69
5.4.2 Análise dos candidatos a droga com melhor média de <i>score</i>	80
6 CONCLUSÃO	82
REFERÊNCIAS	83
Apêndice A - Estrutura química e propriedades dos 42 derivados sintéticos de alcalóides tetraisoquinolínicos	105
Apêndice B - Resultados de <i>docking</i> dos 42 derivados sintéticos de alcalóides tetraisoquinolínicos	108
Apêndice C – Informações gerais das 457 proteínas <i>hub</i>	110

1 INTRODUÇÃO

Corynebacterium ulcerans é um importante microrganismo cuja linhagens patogênicas causam doenças semelhantes à difteria em humanos (DE MATTOS GUARALDI; HIRATA; CARVALHO AZEVEDO, 2014) e mesmo linhagens não patogênicas causam infecções severas (SANTOS *et al.*, 2018).

Difteria é uma doença causada por bactérias produtoras de toxinas do gênero *Corynebacterium*, principalmente por *C. diphtheriae* e em menos casos por *C. ulcerans* e *C. pseudotuberculosis*. Comumente as toxinas produzidas por *Corynebacterium* toxigênicas causam a formação de pseudomembranas no trato respiratório superior e danifica outros órgãos, geralmente o miocárdio e nervos periféricos. As toxinas em questão podem levar o hospedeiro a morte por obstrução respiratória aguda, toxicidade sistêmica aguda, miocardite e complicações neurológicas (WHO, 2018). A taxa de mortalidade por difteria fica entre 5% e 10%, podendo atingir 20% para pessoas com mais de 40 anos e crianças com menos de 5 anos de idade (ATKINSON *et al.*, 2007).

Para prevenir a infecção por corinebactérias produtoras de toxinas, diferentes vacinas foram desenvolvidas, sendo a vacina tetravalente bacteriana (DTP+Hib) a mais recomendada pela Organização Mundial de Saúde (OMS), que além da difteria, também protege contra tétano, coqueluche e meningite (MONDIALE DE LA SANTÉ; WHO, 2017). Mesmo com programas de vacinação, 384.941 casos de difteria já foram reportados por todo o mundo de 1980 até 2019 de acordo com dados da OMS, com aproximadamente 8.395 casos reportados no Brasil. A Índia é o país onde a doença é mais endêmica, contribuindo com 43,66% do total de casos, ultrapassando a quantidade de casos do Brasil em 2018. Em anos recentes, por mais que os casos fossem isolados no Brasil, a doença teve certa incidência em países vizinhos, como a Venezuela, com cerca de 1.592 casos desde 2016 (WHO, 2020).

Considerando a possível reemergência de linhagens toxigênicas de *C. ulcerans*, procuramos por intermédio deste trabalho, identificar e validar proteínas essenciais de *C. ulcerans* com potencial para estudo e desenvolvimento de fármacos e, adicionalmente, testar a afinidade de derivados sintéticos de alcalóides contra estas proteínas. O presente trabalho foi escrito conforme as normas ABNT NBR 14724/2011.

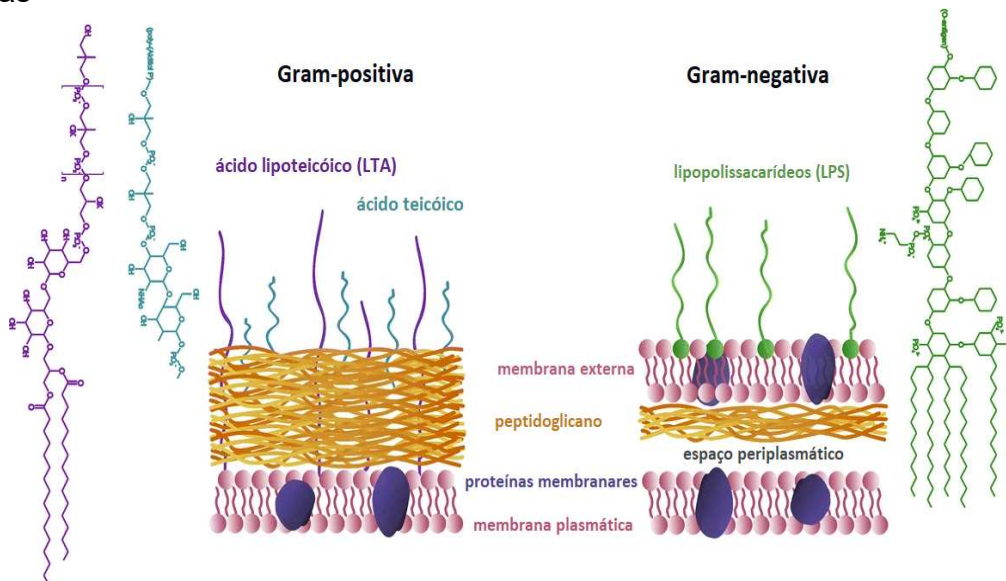
2 FUNDAMENTAÇÃO TEÓRICA

2.1 *Corynebacterium*

Bactérias são microrganismos procarióticos que apresentam uma grande diversidade morfológica. Dentre essa diversidade está a constituição da parede celular, que classificam as bactérias em gram-positivas e gram-negativas, nomenclaturas provindas da reação das células à coloração de Gram, um teste há muito utilizado para classificação bacteriana (GRAM, 1884). Somente bactérias pertencentes aos filos *Firmicutes* e *Actinobacteria* em particular possuem arranjo gram-positivo, enquanto a maioria se apresenta como gram-negativas (WALSH; AMYES, 2004). A razão para a classificação reside na parede celular mais espessa das bactérias gram-positivas, contendo camadas de peptidoglicanos e ácidos teicóicos. Já nas bactérias gram-negativas a camada de peptidoglicanos é menor, cercada por uma segunda membrana lipídica contendo lipopolissacarídeos e lipoproteínas (THOMSEN, 2016) (Figura 1).

Essas diferenças na estrutura da membrana celular são importantes pois podem produzir diferenças na suscetibilidade a antibióticos (WALSH; AMYES, 2004).

Figura 1 - Representação das membranas de bactérias gram-negativas e gram-positivas



Fonte: Adaptado de Pajerski *et al.* (2019).

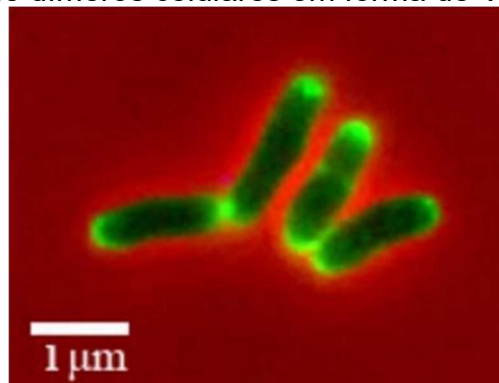
A membrana de bactérias gram-negativas (à direita) é coberta por uma camada de lipopolissacarídeos (LPS), uma fina camada de peptidoglicano, seguido pelo espaço periplasmático e membrana. Enquanto a membrana de gram-positivas (à esquerda) é coberta por múltiplas camadas de peptidoglicanos com ácidos teicóicos e lipoteicóicos as atravessando, e seguido pela membrana celular.

O gênero *Corynebacterium* foi descrito pela primeira vez no ano de 1896 por Lehmann e Neumann como um grupo taxonômico de bactérias semelhantes morfológicamente com o bacilo difteróide (BARKSDALE, 1970). Bacilos difteróides por sua vez, são gram-positivos, pleomórficos (que possuem a capacidade de alterar a morfologia, reprodução e função biológica em resposta às condições ambientais), aeróbicos e não esporulantes (CHANDRAN *et al.*, 2016). Como consequência, o gênero *Corynebacterium* por muito tempo incluiu diversos microrganismos Gram-positivos morfológicamente semelhantes, inclusive bactérias não patogênicas do solo (COLLINS, 1986). Em 1995, as análises da sequência gênica do 16S rRNA permitiram que as corinebactérias fossem definidas como um grupo monofilético (PASCUAL *et al.*, 1995), mesmo exibindo heterogeneidade considerável, como seu conteúdo de ácido micólico e composição de bases de DNA (FUNKE; LAWSON; COLLINS, 1995).

2.1.1 Características

As corinebactérias são microrganismos não-esporulantes com morfologia cúbica ou irregular que pertencem à classe das Actinobactérias, bactérias Gram-positivas, com alto índice de guanina e citosina, variando entre 46% a 74%, o que reflete em uma considerável diversidade genotípica (FUNKE; LAWSON; COLLINS, 1995). Durante a divisão celular desses microrganismos, é típica a formação de dímeros celulares em forma de “V”, conhecida como “*Snapping division*” ou “*V-snapping*” (BARKSDALE, 1970), representado na Figura 2.

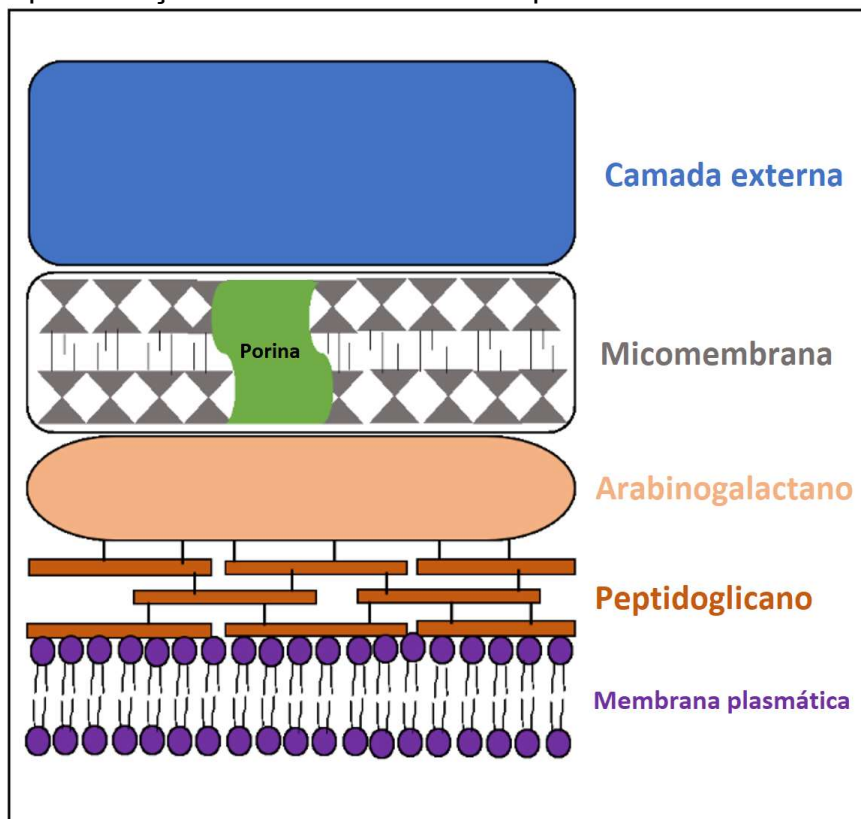
Figura 2 - Coloração da *Corynebacterium glutamicum* com vancomicina-FL mostrando a formação de dímeros celulares em forma de V



Fonte: Letek *et al.* (2008).

A membrana plasmática dessas bactérias é coberta por uma camada de peptidoglicanos ligada covalentemente ao arabinogalactano (Figura 3). Além disso, é encontrada uma camada externa de ácidos micólicos funcionalmente equivalente à membrana externa das bactérias gram-negativas. O material da superfície externa da camada superior é composto por polissacarídeos livres, glicolípídios e proteínas, que incluem proteínas da camada S, do pili (extensões da parede celular) e outras de superfície (DAFFÉ, 2005; EGGELING; BESRA; ALDERWICK, 2008).

Figura 3 - Representação da estrutura do envelope celular corinebacteriano



Fonte: Adaptado de Weerasekera *et al.* (2019)

As corinebactérias possuem uma membrana plasmática composta por uma bicamada de proteínas e fosfolipídios e coberta por uma camada de peptidoglicano covalentemente ligado ao arabinogalactano. A camada externa é composta por polissacarídeos, glicolípídios e proteínas, que incluem proteínas da camada S, pili, porina e outras proteínas da superfície. Entre a camada externa e o arabinogalactano está a micomembrana, composta por ácidos micólicos e porina.

A “Camada S” é composta por proteínas que se organizam de forma rígida, cobrindo a parte externa da célula (ENGELHARDT; PETERS, 1998) e atuando como uma barreira de difusão macromolecular, além de conferir proteção química e física. Essa camada também é conhecida por atuar como fatores de virulência em *Campylobacter* e por conter enzimas de superfície em *Bacillus stearothermophilus* (BEVERIDGE *et al.*, 1997).

2.1.2 Espécies de *Corynebacterium*

Atualmente existem 103 espécies atribuídas ao gênero *Corynebacterium* sequenciadas e depositadas no National Center For Biotechnology Information (NCBI, 2020a), sendo o gênero comumente representado pela espécie patogênica *C. diphtheriae*, principal causadora da difteria. Outras espécies patogênicas são importantes ou possuem importância médica, como: *C. amycolactum*, causadora de sepsis neonatal (BERNER *et al.*, 1997); *C. jeikeium* e *C. xerosis*, causadoras de endocardite infecciosa (MOOKADAM *et al.*, 2006; VELA *et al.*, 2006); *C. pseudotuberculosis*, causadora de linfadenite caseosa (ou linfangite ulcerativa) que atinge caprinos e ovinos (RUIZ *et al.*, 2011); *C. glutamicum*, *C. efficiens* e *C. ammoniagenes*, bem conhecidas por sua importância biotecnológica, como a produção industrial de nucleotídeos e aminoácidos (BURKOVSKI, 2013); e *C. ulcerans*, considerada muitas vezes comensal de animais, incluindo os seres humanos.

2.2 *Corynebacterium ulcerans*

Corynebacterium ulcerans é um microrganismo patogênico primeiramente isolado de lesões na garganta em humanos (GILBERT; STEWART, 1926) que infecta desde animais domésticos e selvagens até humanos (SCHUHEGGER *et al.*, 2009). *C. ulcerans* possui linhagens não-toxigênicas e toxigênicas, as últimas produzindo uma toxina diftérica (TD) semelhante ao da *C. diphtheriae* (SING; BIRSCHENK; HEESEMANN, 2003). Devido à capacidade de produzir a TD, os animais suscetíveis à infecção por linhagens toxigênicas de *C. ulcerans* são reservatórios para zoonoses (transmissíveis entre animais e seres humanos), podendo causar doenças semelhantes à difteria em humanos (DE MATTOS GUARALDI; HIRATA; DE CARVALHO AZEVEDO, 2014; SING *et al.*, 2003).

Independente da produção da TD, *C. ulcerans* causa síndromes clínicas do trato respiratório inferior, como pneumonia (HOMMEZ *et al.*, 1999), nódulos granulomatosos pulmonares (DESSAU *et al.*, 1995), que por vezes estão associados à resposta inflamatória sistêmica, e coagulação intravascular disseminada (NUREKI *et al.*, 2007). Infecções causadas por *C. ulcerans* podem ocorrer até em crianças, mesmo que previamente imunizadas contra difteria (LEEK *et al.*, 1990).

No caso de animais, a infecção por *C. ulcerans* pode causar: mastite em gado (seu principal reservatório), uma inflamação da glândula mamária devido a infecção por bactérias (KISELY; PRINCE; WARD, 1994); lesões ulcerativas em roedores aquáticos (EISENBERG *et al.*, 2015); lesões pulmonares em lontras (FOSTER *et al.*, 2002); abscesso em porcos selvagens (EISENBERG *et al.*, 2014); corças (RAU *et al.*, 2012); e pode ser assintomática, como é o caso de corujas (KATSUKAWA *et al.*, 2016).

Devido à semelhança entre espécies, até 1995 bactérias *C. ulcerans* faziam parte do grupo das *C. diphtheriae*, quando foi realizada sua análise genômica, determinando por fim a distinção de espécies (RIEGEL *et al.*, 1995). Atualmente estão publicamente disponíveis no NCBI (NCBI, 2020b) 29 genomas de *C. ulcerans*, sendo que somente 16 estão completamente sequenciados e montados (Tabela 1).

Tabela 1 - Genomas de *Corynebacterium ulcerans* completamente sequenciados

Linhagem	Tamanho (Mb)	GC%	Genes	Proteínas
NCTC7910	2,45	53,30	2.222	2.127
BR-AD22	2,60	53,40	2.392	2.288
0102	2,57	53,40	2.345	2.245
FH2016-1	2,57	53,40	2.356	2.255
0211	2,57	53,40	2.358	2.251
PO100/5	2,57	54,40	2.392	2.035
FRC58	2,54	53,30	2.313	2.220
210931	2,50	53,25	2.284	2.191
809	2,50	53,30	2.254	2.169
210932	2,48	53,30	2.234	2.142
131001	2,48	53,30	2.242	2.143
05146	2,46	53,30	2.214	2.130
NCTC8639	2,45	53,30	2.221	2.128
NCTC7908	2,45	53,30	2.220	2.121
FRC11	2,44	53,30	2.186	2.105
131002	2,43	53,40	2.182	2.099

Fonte: Adaptado do NCBI (2020b).

2.3 Difteria

A difteria é uma infecção do trato respiratório superior, caracterizada pela presença de uma pseudomembrana acinzentada no local da infecção, promovida pela produção de exotoxinas, multiplicação de células bacterianas e resposta imune do hospedeiro (LARTIGUE *et al.*, 2005). Geralmente é transmitida entre pessoas por contato direto, espirros ou tosse, além de poder ser disseminada por objetos contaminados (CDC, 2015).

A doença é causada principalmente por Corinebactérias produtoras de exotoxinas como a toxina diftérica (TD), presente no gene *Tox*. TD é produzida por linhagens patogênicas de *C. diphtheriae* (SING; BIERSCHEK; HEESEMANN, 2005; SEKIZUKA *et al.*, 2012), e em algumas linhagens de *C. ulcerans*, sugerindo que muitos outros fatores estão relacionados à capacidade de virulência de *C. ulcerans* (FUURSTED *et al.*, 2015). A forma clássica da doença tem a *C. diphtheriae* toxigênica como principal causadora (CERDEÑO-TÁRRAGA *et al.*, 2003).

A imunização passiva através da Antitoxina Diftérica (do inglês *Diphtheria Antitoxin*, DAT) é uma das maneiras de se tratar a difteria (HADFIELD *et al.*, 2000). DAT é um medicamento composto por anticorpos obtidos do sangue de cavalos que foram imunizados contra a TD (MONDIALE DE LA SANTÉ, 2017) e atua neutralizando as toxinas produzidas por *C. diphtheriae*. Os efeitos colaterais do DAT são comuns e incluem a doença do soro e reações alérgicas, como a anafilaxia (STUART; KOUIMTZI; HILL, 2009)

Em virtude da purificação da toxina diftérica inativa, o desenvolvimento da vacina foi possível, promovendo uma redução na incidência da difteria, porém, não sua erradicação. Diferente de infecções por *C. diphtheriae*, infecções por *C. ulcerans* em humanos estão aumentando, provavelmente devido à diversificação da toxina difteróide. Por conta da *C. diphtheriae* infectar apenas o humano, sua capacidade de adquirir mutações é reduzida quando comparado com *C. ulcerans*, que possui vários animais como reservatórios, possibilitando maior diversidade genética (OTSUJI *et al.*, 2019). Pela considerável diversidade entre linhagens *C. ulcerans*, outros fatores além da toxina diftérica podem estar relacionados à sua virulência (WEERASEKERA *et al.*, 2019). Ainda, linhagens de *C. diphtheriae* e de *C. ulcerans* não-toxigênicas podem adquirir a capacidade de produzir a toxina diftérica pelo processo de lisogenia com corinebacteriófagos carreadores do gene *Tox* (FREEMAN, 1951).

A ocorrência de cepas não-toxigênicas associadas a infecções invasivas tem sido cada vez mais relatadas. Recentemente, infecções do tipo difteria causadas pela *C. ulcerans* tornaram-se mais numerosas na Europa Ocidental, causando uma maior preocupação com transmissões zoonóticas. Apesar da importância médica e veterinária desses microrganismos, apenas alguns fatores de virulência foram caracterizados em detalhes, como interações hospedeiro-patógeno, estruturas como pili e outros fatores de adesão que desempenham um papel central na comunicação intracelular (WEERASEKERA *et al.*, 2019).

No ano de 2011, aproximadamente 75% dos casos de difteria zoonótica por *C. ulcerans* ocorreram em pacientes adultos, total ou parcialmente vacinados com toxóide diftérico (DIAS *et al.*, 2011). Houve relatos de sua presença em diferentes países, até mesmo onde houveram programas de imunização (BITRAGUNTA *et al.*, 2007). A frequência e gravidade de infecções humanas associadas a *C. ulcerans* parecem estar aumentando gradualmente em diferentes países, como no Japão (HATANAKA *et al.*, 2003; KOMIYA *et al.*, 2010; KIMURA *et al.*, 2011), Canadá (DEWINTER; BERNARD; ROMNEY, 2005) e França (BONMARIN *et al.*, 2009). Na Europa entre os anos de 2005 e 2010 *C. ulcerans* foi isolada com maior frequência em casos de difteria do que *C. diphtheriae* (DE ZOYSA *et al.*, 2005; WAGNER *et al.*, 2010; WAGNER *et al.*, 2012).

De acordo com o Ministério da Saúde (MS) a difteria está controlada no Brasil, se manifestando em casos isolados nos últimos anos. Porém, alguns países como Haiti e Venezuela surtos ativos de difteria ocorrem, iniciados respectivamente em 2014 e 2016. Até setembro de 2019 um total de 2.512 casos suspeitos de difteria foram relatados na Venezuela, provocando um total de 270 mortes. Esta situação é considerada alarmante, pois novos casos podem ser introduzidos no Brasil, principalmente pelo alto fluxo migratório venezuelano, conforme descrito na nota informativa Nº 9/2019-CGDT/DEVIT/SVS/MS (MS, 2019).

Mesmo com vários genomas sequenciados, montados e anotados, o mecanismo de patogenicidade de *C. ulcerans* ainda não é completamente elucidado, necessitando diferentes abordagens de pesquisa para entender sistematicamente a biologia deste microrganismo e identificar novos alvos farmacológicos. Neste sentido, as redes de interação proteína-proteína (RIP) se demonstraram eficientes ferramentas de pesquisa.

2.4 Biologia de Sistemas

A Biologia de Sistemas é o ramo da ciência que busca entender os organismos biológicos em todos os seus níveis, investigando o comportamento e relações dos elementos que os constituem (genes, RNAs, proteínas, metabólitos). Essas informações podem então ser integradas, exibidas graficamente e modeladas computacionalmente para melhor compreensão (IDEKER; GALITSKI; HOOD, 2001).

Um sistema refere-se a objetos interdependentes ou inter-relacionados que compreendem e funcionam como um todo. Uma rede complexa de entidades biologicamente relevantes está presente em um sistema biológico, que abrange várias escalas (BENSO *et al.*, 2020). Nessas diversas escalas estão presentes um grande número de componentes que interagem, como genes e proteínas, que desempenham suas funções dentro de uma complexa rede de interações, na qual um único componente pode afetar drasticamente todo o conjunto (CHAUTARD; THIERRY-MIEG; RICARD-BLUM, 2009).

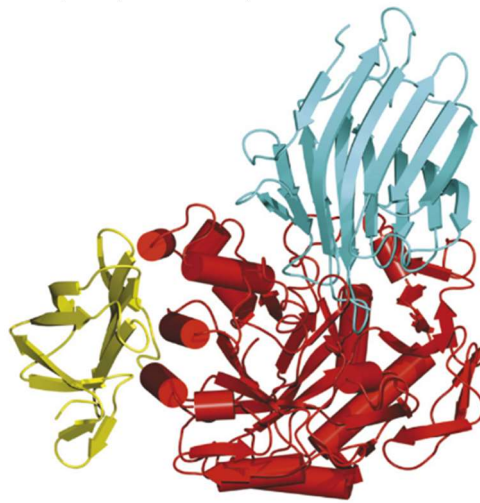
2.4.1 Interação Proteína-Proteína (IPP)

As proteínas são os principais catalisadores, elementos estruturais, mensageiros de sinalização e máquinas moleculares dos organismos (EISENBERG *et al.*, 2000). Muitos dos principais processos moleculares, como a replicação do DNA, são realizados por um grande número de proteínas que interagem entre si possibilitando que os processos sejam realizados (ANDREOPOULOS; LABUDDE, 2013). Logo, interações entre proteínas são importantes para orquestrar diversos eventos em uma célula, como a regulação da transcrição e a transdução de sinais (RAMAN, 2010). Estudos de genômica comparativa sugerem que interações que ocorrem entre as proteínas diferencia mais uma espécie de outra do que seus próprios genes individuais (VALENCIA; PAZOS, 2002). Processos e sistemas biológicos dependem dessas interações e a sua compreensão ajuda a elucidar como o sistema funciona, assim como aumentar o conhecimento sobre a evolução de organismos e suas funções biológicas (KELLY, 2009).

Existem diversos tipos de interações baseadas em proteínas: interações que ocorrem entre duas proteínas diferentes, conhecida como interação proteína-proteína (IPP) (TARASSOV *et al.*, 2008); interações entre proteínas e componentes da

sequência genética (BOONE; BUSSEY; ANDREWS, 2007); interações entre a proteína e ligantes; interações entre proteínas macromoléculas, como lipídeos e carboidratos; e interações entre proteínas e micromoléculas, como íons. No primeiro caso, as IPPs são realizadas mais especificamente entre domínios, componentes proteicos que existem independentes do restante da estrutura, com o mesmo domínio podendo ser encontrado em diferentes proteínas (KELLY, 2009). Para melhor representar, uma IPP típica entre duas proteínas pode ser visualizada na Figura 4.

Figura 4 - Exemplo de interação proteína-proteína

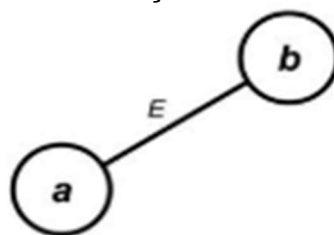


Fonte: BOMPARD-GILLES *et al.* (1996).

Figura mostra a estrutura da α -amilase pancreática suína (azul) ligada a um inibidor do tipo lectina de feijão (proteína com dois domínios, em vermelho e amarelo). A interação ocorre apenas entre os domínios azul e vermelho.

As IPPs podem ser de dois tipos: interação física ou direta, quando existe o contato entre duas proteínas; ou interação regulatória ou funcional, quando não existe o contato entre as duas proteínas. Em sistemas biológicos uma interação entre proteínas pode ser representada na forma de grafos através de uma coleção de nós ou vértices (V) e arestas ou conectores (E), representados na Figura 5.

Figura 5 - Representação de uma interação na forma de grafos

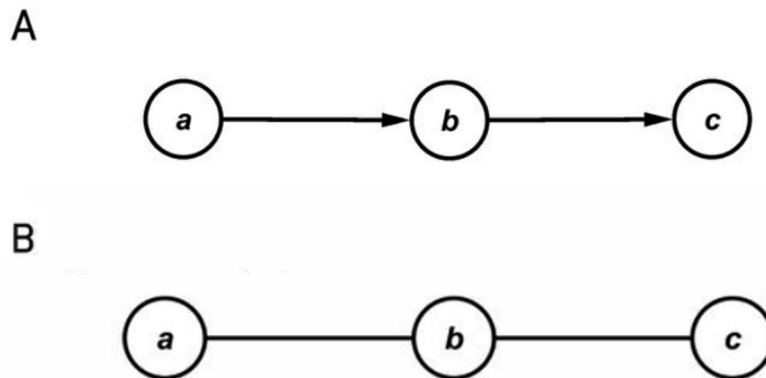


Fonte: Verli (2014).

Neste grafo está representada a interação de dois nós vizinhos (a e b), conectados pelo conector (aresta) E.

Os elementos do sistema biológico, como metabólitos, genes ou produtos gênicos (ex., proteínas), são representados pelos nós e as interações entre os elementos são representados pelas arestas. Um conjunto de IPP, par-a-par, podem formar uma complexa rede de interação proteína-proteína (RIP). Nas redes, uma interação (aresta) pode ser orientada em determinada direção, sendo assim chamadas de redes direcionadas ou diretas (Figura 6.A), existindo também as redes não-direcionadas como as RIPs, que apresentam conectores orientados em ambas as direções (Figura 6.B).

Figura 6 - Rede direcionada e não direcionada



Fonte: Verli (2014).

(A) Interação orientada, na qual “a” interage com “b” ($a \rightarrow b$), e “b” interage com “c” ($b \rightarrow c$). (B) Interação não orientada, com interação mútua entre “a” e “b” ($a \leftrightarrow b$) e “b” e “c” ($b \leftrightarrow c$).

2.4.2 Rede de interação proteína-proteína (RIP)

O conjunto completo de interações encontradas no proteoma de um organismo compreende então um interactoma ou uma rede de interação proteína-proteína (RIP), essenciais para entender as funções de genes e como são regulados à nível de organismo (SANCHEZ *et al.*, 1999). As interações proteicas são de grande importância por estarem envolvidas em processos imunes (antígeno-anticorpo), síntese de máquinas moleculares (ribossomos e proteassomos), metabolismo (interação enzima-substrato), sinalização (interação com moléculas mensageiras, hormônios e neurotransmissores) e expressão gênica (interação proteína-DNA) (SANCHEZ *et al.*, 1999; CHAUTARD; THIERRY-MIEG; RICARD-BLUM, 2009).

As redes podem ser usadas para descrever e representar interações biológicas, tais como interações em nível atômico, como entre domínios proteicos ou

entre metabólitos e proteínas, ou até entre organismos em uma população ou ecossistema (ALM; ARKIN, 2003).

De um modo geral, existem três importantes classes de redes biomoleculares: redes metabólicas, que representam o conjunto de processos metabólicos que determinam as propriedades bioquímicas de uma célula; redes reguladoras da transcrição, que descrevem as interações regulatórias entre diferentes genes; e redes de interação proteica, que consistem nas interações físicas ou regulatórias entre proteínas de um organismo. A partir dos dados fornecidos pela rede é possível investigar as propriedades estruturais das redes nas células vivas, identificar as principais propriedades e entender como elas podem ter evoluído biologicamente (MASON; VERWOERD, 2007).

Dada a natureza especial dos sistemas biológicos, houve uma necessidade de métodos de análise que pudessem extrair informações significativas de dados experimentais, considerando que as redes geradas pelas técnicas de alto rendimento são complexas demais para serem analisadas de maneira não sistemática, técnicas estas que serão explicadas mais a frente. Na matemática a teoria dos grafos é um dos objetos de estudo e análise dessas complexas redes biológicas (DIESTEL, 2005).

2.4.3 Propriedades topológicas de rede

Para que dados confiáveis e conclusivos sejam gerados, diversas propriedades são empregadas na análise de redes biológicas em um determinado sistema. Cada propriedade fornece uma diferente informação sobre os elementos constituintes e suas interações, podendo ser referente a nós individuais ou a rede como um todo.

Modularidade ou clusterização: conceito comum na área da biologia evolutiva, a modularidade pode ser calculada através do coeficiente de clusterização, que leva em conta a quantidade de arestas e nós na rede. Quanto maior seu valor, mais conectado de modo geral é o *cluster* (em port., aglomerado). Evolutivamente, as proteínas que compõem módulos altamente agrupados tendem a ser conservadas (VERLI, 2014).

Degree (Grau do nodo): se refere à quantidade de conexões diretas que um nodo faz com outros. Nodos que realizam uma grande quantidade de conexões são identificados e chamados de *hubs*. Numa visão biológica, proteínas *hub* se ligam a

várias outras proteínas e possuem uma função importante na rede, cuja deleção pode afetar diversas vias ou processos (VERLI, 2014).

Betweenness: parâmetro que calcula o número de caminhos mais curtos que passam por um nodo em questão, estimando a relação entre eles. Uma proteína com alto valor de *betweenness* apresentaria uma elevada capacidade de interação e/ou sinalização entre processos biológicos ou *clusters* (VERLI, 2014).

Closeness: é a tendência de aproximação ou isolamento de um nó, calculado pelo somatório dos caminhos mais curtos entre um nó e todos os outros da rede. Um alto valor de *closeness* para um nó indica que todos os outros estão próximos a ele, caso contrário se encontrariam distantes (VERLI, 2014).

Diâmetro: é um valor utilizado para identificar o grau de compactação da rede pelo cálculo da distância entre os dois nós mais afastados entre si de uma rede. Redes com diâmetro pequeno são consideradas mais agrupadas, uma vez que suas proteínas estão mais interligadas entre si (VERLI, 2014).

2.4.4 Métodos para predição de IPP

Com o intuito de identificar novas proteínas, melhor compreender os processos em que as proteínas estão envolvidas e identificar interações proteína-proteína ideais para intervenção farmacológica, mapear interações proteicas em um organismo é de extrema importância. Essas interações podem ser observadas usando uma variedade de métodos experimentais ou computacionais (*in silico*). Embora muitos métodos existam, apresentaremos alguns dos mais utilizados.

2.4.4.1 Métodos experimentais

Yeast two hybrid (Y2H): Técnica desenvolvida em 1989 (FIELDS; SONG, 1989) para detectar interações entre duas proteínas, bem como proteína e ácido nucleico, por meio de ensaios *in vivo* de leveduras. Método funciona de modo que duas proteínas de fusão são criadas: a isca (*bait*), uma proteína de interesse com um domínio de ligação (*binding domain*) ligado ao DNA; e a presa (*prey*), seu potencial parceiro de interação, ligado a um domínio de ativação (*activation domain*). Se a 'isca' e a 'presa' interagirem, sua ligação formará um ativador transcricional funcional (*transcriptional activator*), que por sua vez ativa genes repórteres ou marcadores selecionáveis (LIN; LAI, 2017). No entanto, devido a técnica Y2H ser realizada em

leveduras, boa parte das interações proteicas presentes em outros organismos podem não ser detectadas, por conta da baixa expressão, ausência de cofatores e/ou modificações pós-traducionais. Também é alta a quantidade de falsos-positivos, devido a necessidade de se auto expressar as proteínas em estudo, necessitando que interações biologicamente relevantes e verdadeiras sejam identificadas, ou seja, conhecimento prévio das proteínas em questão é indispensável. Além de que apenas interações entre proteínas que conseguem adentrar o núcleo são detectadas (HAMDI; COLAS, 2012).

Affinity purification/mass spectrometry (AP-MS): O princípio geral da técnica envolve a imobilização de uma proteína de interesse (isca) em um suporte sólido, como agarose ou esferas magnéticas. A proteína isca tem como função capturar as proteínas alvos de uma fase solúvel. Proteínas capturadas são então digeridas com proteases, como a tripsina, e submetidas a identificação por espectrometria de massa (DUNHAM; MULLIN; GINGRAS, 2012).

Protein microarray ou proteinchip: *Protein microarray* ajuda na detecção de interações *in vitro* de vários tipos, como interações proteína-proteína, proteína-lipídio, proteína-ácido-nucleico, antígeno-anticorpo, enzima-substrato e proteína-pequenas-moléculas. Funciona de modo que proteínas covalentemente ligadas a um suporte sólido são rastreadas com sondas marcadas com fluorescência (proteínas ou lipídios), para identificar interações (ZHU; SNYDER, 2003).

2.4.4.2 Métodos computacionais

Embora novas IPP sejam coletadas principalmente pelos métodos experimentais, métodos computacionais foram desenvolvidos para prever possíveis IPP em diferentes organismos se utilizando de dados previamente obtidos, sem o custo, mão-de-obra e tempo adicional que métodos experimentais exigem.

Fusão de domínio ou Rosetta Stone: método utilizado para prever proteínas que interagem funcionalmente e que podem estar envolvidas em IPPs diretas ou que participem da mesma via de sinalização ou metabólica. Se baseia na hipótese de que, se em um organismo os domínio A e B estão presentes em um polipeptídeo AB, então A e B estão funcionalmente ligados (TRUONG; IKURA, 2003).

Vizinhança conservada: dita que diferentes proteínas codificadas por diferentes genes sobre a ação do mesmo operon tendem a estar relacionadas. Este

método, portanto, associa a conservação da vizinhança entre genes com a funcionalidade das proteínas. É um importante indicador associativo para gerar RIPs no banco de dados STRING (ESCH; MERKL, 2020).

Perfis filogenéticos: identifica interações funcionais entre proteínas se baseando na ideia de que proteínas relacionadas funcionalmente muito provavelmente co-ocorrerem em diferentes genomas devido à pressão evolutiva (PAZOS *et al.*, 2005).

Co-evolução: métodos se baseiam na noção de que a pressão evolutiva pode levar à co-evolução de pares de proteínas, ou seja, se uma delas sofrer mutação é provável uma mutação subsequente ocorra na outra proteína para que a IPP e funções desempenhadas sejam mantidas (YIN; YAU, 2017). Como é o caso do método *in silico two-hybrid* (i2h), que detecta pares de proteínas que interagem fisicamente e identifica as regiões de sequência mais prováveis envolvidas nessas interações (PAZOS; VALENCIA, 2002).

Interolog mapping: O método se baseia na premissa de que, se duas proteínas interagem em uma espécie, é muito provável que as respectivas proteínas ortólogas presentes em outra espécie também interajam (WALHOUT *et al.*, 2000). Assim, se proteínas A e B interagem, seus ortólogos A' e B' provavelmente interagem. Desta forma, conhecendo um conjunto de interações, é possível mapear as interações ortólogas conservadas entre diferentes organismos. Para isso, quanto mais similaridade houver entre as sequências destas proteínas ortólogas, mais eficiente será o método (LEWIS *et al.*, 2012). Mapeados os *interologs* com base na transferência por homologia, eles podem ser utilizados para anotações entre genomas (YU *et al.*, 2004) e para predição da RIP (LI *et al.*, 2004; HUANG; LIN; KAO, 2007; GEISLER-LEE *et al.*, 2007). Por existirem disponíveis nas bases de dados públicas uma grande quantidade de organismos, proteínas e interações, o mapeamento de interações ortólogas é uma metodologia viável.

2.4.5 Bases de dados de IPP

As interações proteína-proteína determinadas por métodos experimentais ou computacionais são depositadas em bancos de dados públicos, possibilitando o acesso pela comunidade científica. Alguns dos vários mais importantes e utilizados banco de dados de IPP disponíveis publicamente são:

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins):

Banco de dados de redes biológicas que contém interações físicas (diretas) e funcionais (indiretas) determinadas experimentalmente ou preditas *in silico*. Em fevereiro de 2020 o STRING possuía 5.090 organismos, 24,6 milhões de proteínas e mais de 3 bilhões de interações depositadas, disponível em <https://string-db.org/> (SZKLARCZYK *et al.*, 2019).

HPRD (The Human Protein Reference Database): Banco de dados de organismos modelo que contém informações sobre funções de proteínas humanas, incluindo interações proteína-proteína, modificações pós-traducionais, relações enzima-substrato e associações com doenças. Anotações são feitas por curadoria manual, utilizando literatura publicada por especialistas na área da biologia e por análises das sequências de proteínas por bioinformatas. Em fevereiro de 2020 o HPRD possuía 30 mil proteínas e mais de 40 mil interações, disponível em <http://www.hprd.org/> (KESHAVA PRASAD *et al.*, 2009).

DIP (Database of Interacting Proteins): Banco de dados que permite aos usuários pesquisar e visualizar, estaticamente ou dinamicamente, proteínas que interagem, manualmente curadas por especialistas ou automaticamente por meios computacionais. Ferramenta permite aos usuários enviar novas interações proteína-proteína e atualizar entradas do banco de dados. Em fevereiro de 2020 o DIP possuía 834 organismos, 28.850 proteínas e 81.923 interações, disponível em <https://dip.doe-mbi.ucla.edu> (SALWINSKI *et al.*, 2004).

IntAct (Molecular Interaction Database): Banco de dados público de interações derivadas de literatura curada ou submetidas diretamente por usuários. Em fevereiro de 2020 o IntAct possuía, 115.255 proteínas e 1.043.844 interações, disponível em <https://www.ebi.ac.uk/intact/> (KERRIEN *et al.*, 2012).

Constantemente bancos de dados são criados e oferecem diferentes recursos e dados disponibilizados. Os bancos citados anteriormente não são os únicos, existem vários outros que disponibilizam IPP para a comunidade científica analisar, possibilitando a identificação de proteínas essenciais.

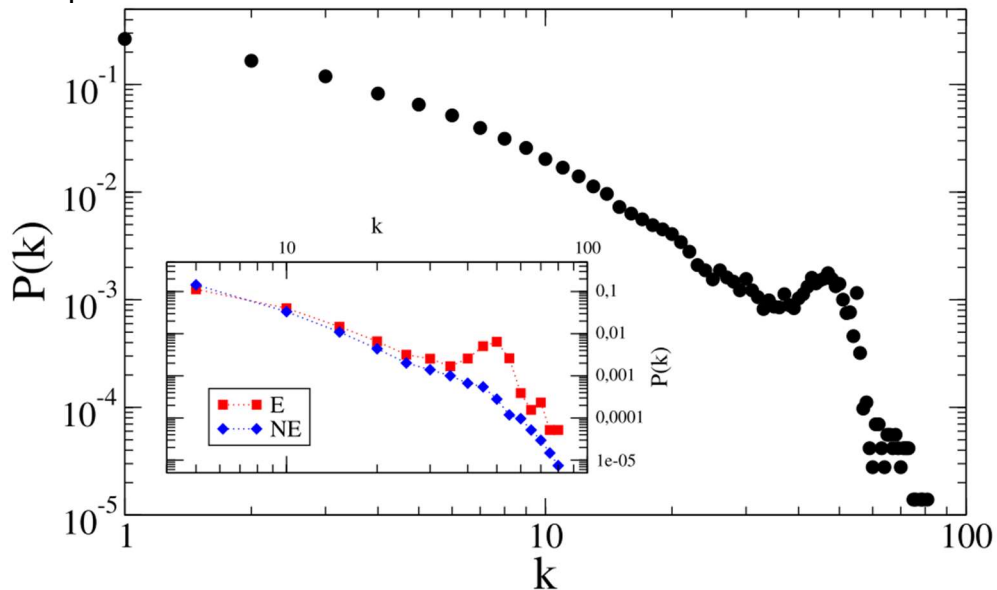
2.5 Proteínas essenciais

Quando se trata de essencialidade, aminoácidos essenciais são aqueles que não podem ser sintetizados *de novo* (do zero) por um organismo (ex., o humano) e que formam as proteínas. No caso de proteínas, a essencialidade é devido à indispensabilidade para a viabilidade de um organismo (ACENCIO; LEMKE, 2009), pois a exclusão de apenas uma dessas proteínas pode ser letal ao organismo (WINZELER *et al.*, 1999) ou tornar uma célula inviável (YU *et al.*, 2004). Deste modo, a identificação de proteínas essenciais para um organismo é importante não apenas para a compreender seus requisitos mínimos para viabilidade celular, mas também para fins práticos, como o desenvolvimento de medicamentos (WANG *et al.*, 2011) com as proteínas essenciais como potenciais alvos (PARK *et al.*, 2008).

Existem técnicas experimentais que são aplicadas para identificação de proteínas essenciais, como nocautes de um único gene (GIAEVER *et al.*, 2002), interferência de RNA (CULLEN; ARNDT, 2005) e nocautes condicionais (ROEMER *et al.*, 2003), porém geralmente elas são trabalhosas e demoradas, o que restringe seus usos. Considerando essas restrições experimentais, uma abordagem computacional para identificar proteínas essenciais seria de grande importância. As redes de interação proteína-proteína servem também para este propósito, pois inúmeros experimentos mostraram que, em uma rede de interação, as proteínas com alto grau de interação (*hubs*) estão correlacionadas com sua essencialidade (JEONG *et al.*, 2001; UETZ *et al.*, 2000; RAIN *et al.*, 2001).

A distribuição do grau de interação das proteínas em uma RIP não é aleatória, sendo justificado pela pressão seletiva que as interações estão submetidas nos organismos. Assim, uma das características de redes biológicas é apresentar uma distribuição livre de escala (*scale free*) com tendência a lei de potência para o grau de interação, justificada pela existência das proteínas *hub* (Gráfico 1) (JEONG *et al.*, 2000). Devido à pressão seletiva, as proteínas *hub* tendem a ser mais conservadas a nível de sequência, entre os organismos e estão correlacionadas com a essencialidade, como foi demonstrado por Dilucca, Cimini e Giansanti (2017).

Gráfico 1 - Gráfico de distribuição de probabilidade $P(k)$ para o número de conexões (k) de cada proteína



Fonte: Dilucca, Cimini e Giansanti (2017).

Inserto: Gráfico $P(k)$ para genes essenciais (E) e não essenciais (NE) com valor médio dos genomas anotados no DEG. É interessante notar que o grau de interação na média é maior para E do que para NE, comprovando que proteínas essenciais fazem um maior número de conexões.

Assim, as proteínas *hub*, são importantes alvos farmacológicos, entretanto, para que sejam realizados testes por métodos computacionais as suas estruturas tridimensionais devem ser elucidadas.

2.6 Estrutura tridimensional proteica

Para desempenhar sua função biológica as proteínas se dobras em conformações específicas guiadas por interações não covalentes, como ligações de hidrogênio, interações iônicas, forças de Van der Waals e interações hidrofóbicas. Cada conformação possui um arranjo tridimensional de átomos diferente, que se dá o nome de estrutura tridimensional proteica.

Para se trabalhar com proteínas pela abordagem computacional é necessário que as estruturas estejam em um formato que o computador consiga interpretar. Na década de 70 este formato foi desenvolvido e conhecido até hoje como PDB, nome este derivado do banco de dados onde é depositado, o Protein Data Bank (BERMAN *et al.*, 2000). Em um arquivo PDB estão contidas as localizações tridimensionais de cada átomo de uma proteína (coordenadas atômicas), além de informações relacionadas à estrutura, sequência gênica e experimento realizado para obtê-las. Com o arquivo PDB da estrutura, é possível visualizar a proteína em interface gráfica

através de programas como PyMOL (DELANO, 2002), ADT do pacote MGLTools (MORRIS *et al.*, 2009) e UCSF Chimera (PETTERSEN *et al.*, 2004).

Porém, obter coordenadas atômicas de estruturas proteicas não é algo tão simples. São diversos os métodos utilizados para este fim, sejam eles experimentais ou com auxílio de computadores.

2.6.1 Métodos experimentais

Os métodos experimentais são mais precisos para se obter uma estrutura em alta resolução, mas são custosos e requerem equipamentos específicos (WU; ZHANG, 2009). Dentre os diversos métodos experimentais, alguns dos que mais contribuem com o depósito de estruturas tridimensionais de proteínas são:

Cristalografia de raios-X: Neste método, a proteína de interesse é purificada, cristalizada e submetida a um intenso feixe de raios-X. A proteína cristalizada difrata o feixe em várias direções, que são analisados para determinar a distribuição de elétrons em uma proteína e conseqüentemente a localização espacial de cada átomo. Através dessa técnica é possível obter informações precisas sobre ligantes, inibidores, íons e outras moléculas que também estão incorporadas ao cristal. Embora seja o método mais preciso em determinar a estrutura 3D proteica, sua utilização é menor do que se espera devido ao tempo, custo e inabilidade de algumas proteínas formarem cristais ideais, muitas vezes não refletindo na estrutura real da proteína. Por conta da cristalografia depender das moléculas estarem alinhadas, o processo de cristalização é difícil e limitado às proteínas rígidas (BERMAN *et al.*, 2000; YADA; JACKMAN; NAKAI, 1988). Em dezembro de 2019 mais de 132.200 proteínas estavam depositadas no PDB com estrutura determinada pelo método de cristalografia de raios-X.

Espectroscopia de Ressonância Magnética Nuclear (RMN): Na RMN a proteína é purificada, colocada em um forte campo magnético e depois sondada com ondas de rádio, gerando um conjunto de ressonâncias distintas (picos), que serão observadas e analisadas. Uma vantagem do método é fornecer informações sobre proteínas ainda em solução, sem a necessidade de ser cristalizada ou preparada para ser visualizada em microscópio. Portanto, a espectroscopia de RMN é o principal método experimental para estudar as estruturas atômicas de proteínas flexíveis, porém, o método é limitado a proteínas pequenas ou médias (BERMAN *et al.*, 2000).

Estavam depositadas no PDB em dezembro de 2019 mais de 11.200 proteínas com estrutura determinada pelo método de Espectroscopia de RMN.

Microscopia Eletrônica 3D (ME3D): Para se obter uma estrutura tridimensional proteica por este método é realizada a captura de várias imagens 2D em diferentes ângulos, por microscópios eletrônicos de transmissão que utilizam de feixes de elétrons e um sistema de lentes de elétrons. Em termos de detalhes moleculares e atômicos, o método produz resultados similares à cristalografia, permitindo a visualização de domínios proteicos, elementos estruturais secundários, cadeias laterais de aminoácidos, moléculas de água de superfície e ligantes não covalentemente ligados (BERMAN *et al.*, 2000). Estavam depositadas no PDB em dezembro de 2019 mais de 3.000 proteínas com estrutura determinada pelo método de Microscopia Eletrônica 3D.

2.6.2 Métodos computacionais de predição de estrutura tridimensional

Embora o número de estruturas de proteínas determinadas experimentalmente esteja aumentando, a quantidade de sequências de proteínas sem estrutura que estão sendo constantemente depositadas aumenta ainda mais. Em maio de 2020 o TrEMBL continha ao menos 180.690.000 proteínas automaticamente anotadas e sem estrutura, enquanto no Swiss-prot apenas 562.000 proteínas manualmente anotadas e com estrutura. Devido a enorme diferença entre a quantidade de sequências e estruturas tridimensionais, se faz necessário a utilização de métodos computacionais, que predizem a estrutura de proteínas com uma certa rapidez e confiabilidade, são eles:

Predição por enovelamento (*Threading*): *Threading* identifica e alinha a estrutura da proteína alvo com outras do PDB com base em seus enovelamentos ou conhecidos *motifs* estruturais (conexões entre estruturas secundárias) (ROY; KUCUKURAL; ZHANG, 2010). Utilizado quando a modelagem comparativa, baseada na sequência proteica, não identifica proteínas homólogas no PDB (XU; JIAO; YU, 2008).

Método *ab initio*: Também conhecido como modelagem *de novo* e denominado modelagem sem modelo, é baseado nos princípios da física, química e na hipótese termodinâmica de Anfinsen (ANFINSEN, 1973). Método alternativo

quando não existe estrutura homóloga no PDB nem pode ser detectada pelo *threading*, gera a estrutura da proteína a partir do zero (WU; ZHANG, 2009).

Modelagem comparativa: Apesar do considerável progresso na predição da estrutura *de novo* (BRADLEY; MISURA; BAKER, 2005), os métodos de modelagem comparativa, quando aplicáveis (proteínas homólogas presentes), fornecem modelos mais confiáveis e precisos de estrutura proteica (BAKER; SALI, 2001). A modelagem comparativa se baseia na premissa de que sequências relacionadas evolutivamente possuem estruturas tridimensionais semelhantes (CHOTHIA; LESK, 1986), permitindo construir a estrutura de uma proteína de interesse a partir de proteína com estrutura conhecida, que compartilham similaridade à nível de sequência. O procedimento tradicional de modelagem comparativa consiste em várias etapas consecutivas, geralmente repetidas até que uma estrutura satisfatória seja obtida: encontrar proteínas molde adequadas e relacionadas ao alvo; alinhar sequências do alvo e do modelo; identificar regiões estruturalmente conservadas; prever regiões estruturalmente variáveis, incluindo inserções e terminais N e C ausentes; modelagem das cadeias laterais; e por último refinamento e avaliação do modelo resultante (GINALSKI, 2006).

Da área de bioinformática estrutural, estes métodos estão embutidos, não exclusivamente, nos programas utilizados para modelagem tridimensional de estruturas proteicas.

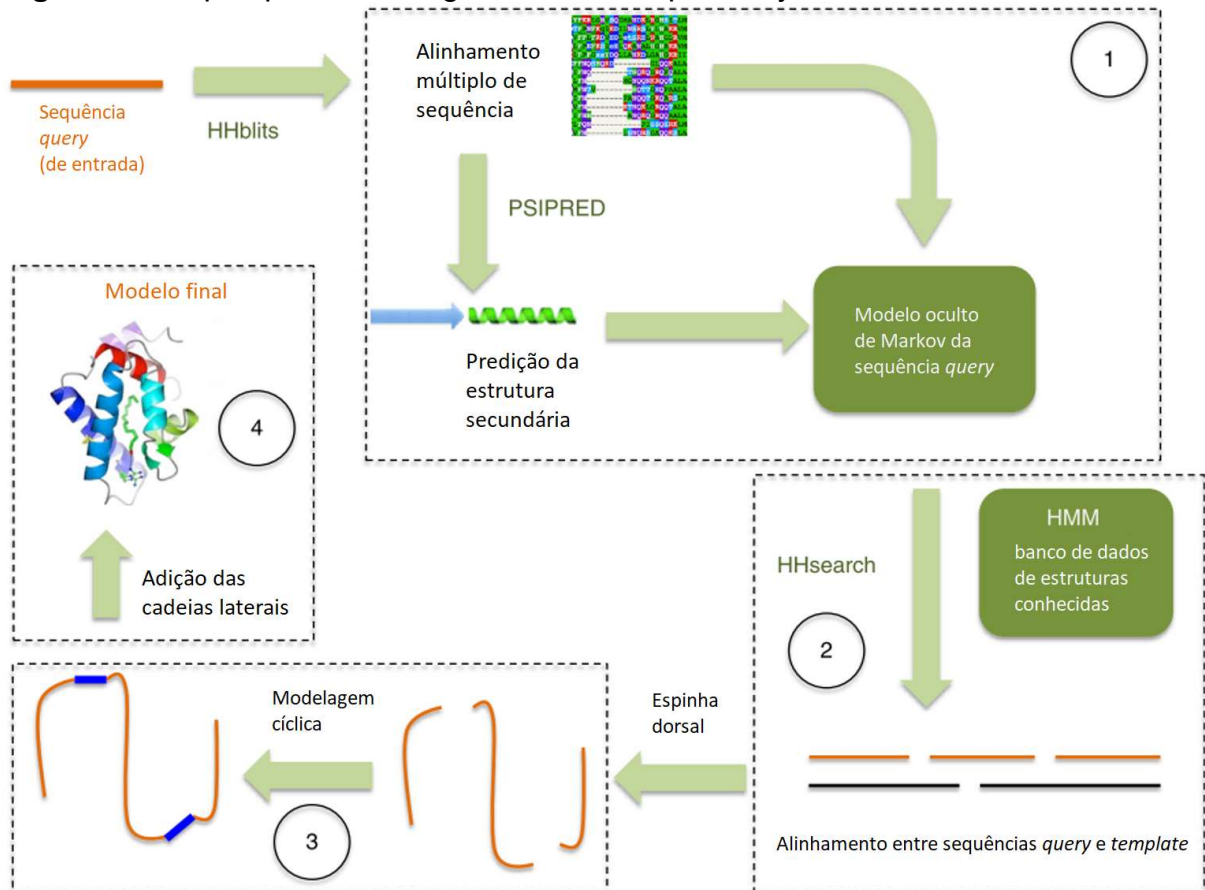
2.6.3. Programas de modelagem

Existem diversos programas utilizados para a predição de estruturas tridimensionais, diferindo pouco na precisão com que o fazem (MOULT *et al.*, 2011). Alguns dos servidores Web mais usados para modelagem de proteínas são Phyre2 (KELLEY *et al.*, 2015), i-TASSER (ROY; KUCUKURAL; ZHANG, 2010), Swiss-Model (ARNOLD *et al.*, 2006), HHpred (SÖDING, 2005), predição de estrutura secundária baseada em PSI-BLAST (PSIPRED) (LOBLEY; SADOWSKI; JONES, 2009), Rosetta (RAMAN *et al.*, 2009) e Raptor (KÄLLBERG *et al.*, 2012; KELLEY *et al.*, 2015).

O sistema Phyre2, desenvolvido em 2011 e sucessor do Phyre (KELLEY; STERNBERG, 2009), é uma combinação de diversos algoritmos próprios para, entre outras funções, prever estruturas 3D proteicas através da modelagem comparativa.

Seu método para gerar um modelo 3D é composto por quatro estágios descritos e ilustrados na Figura 7.

Figura 7 - Etapas para modelagem realizadas pelo Phyre2



Fonte: Adaptado de Kelley (2015).

Etapa 1 (coleta de sequências homólogas): a sequência de uma proteína é alinhada contra diversas outras sequências de proteínas curadas, com o objetivo de capturar a posição preferível para cada resíduo. O resultante alinhamento múltiplo de sequências é usado para prever a estrutura secundária com o auxílio do PSIPRED. Tanto o alinhamento múltiplo quanto a estrutura secundária predita são usados para gerar um modelo oculto de Markov (HMM). Etapa 2 (alinhamento contra biblioteca de estruturas secundárias): O modelo HMM gerado é alinhado contra estruturas secundárias de proteínas conhecidas do banco de dados com perfil HMM conhecidos. Os alinhamentos com pontuação mais alta são usados para construir modelos com apenas a espinha dorsal (cadeia polipeptídica sem o grupo R). Etapa 3 (modelagem cíclica): os *indels* (inserções e deleções) nesses modelos são corrigidos pela modelagem cíclica. Etapa 4 (adição de cadeias laterais): cadeias laterais (grupo R) de cada aminoácido são adicionadas para gerar o modelo Phyre2 final.

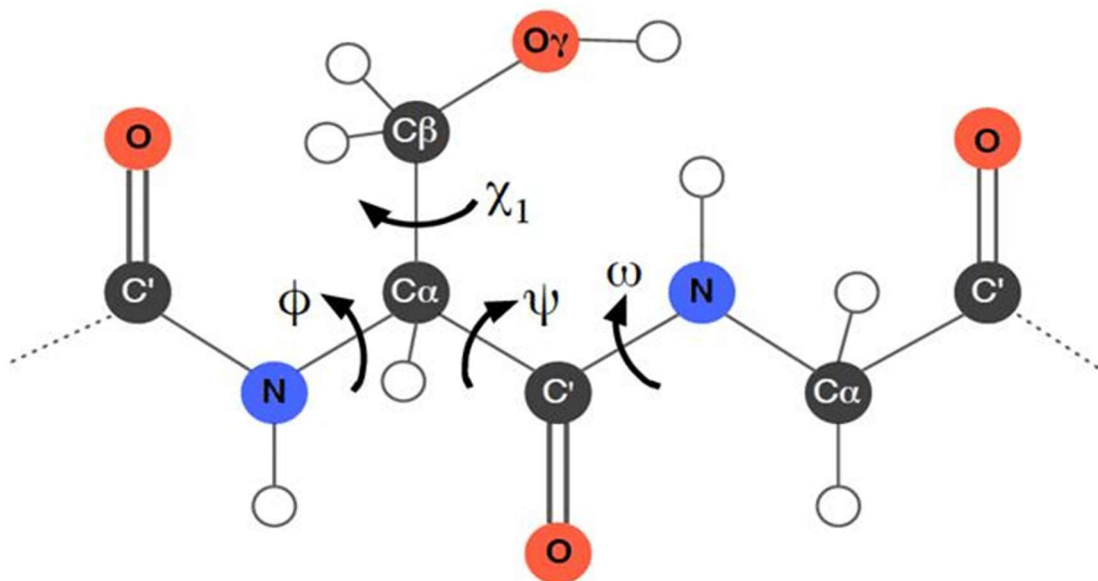
2.6.4 Validação de estruturas tridimensionais

As estruturas 3D geradas pelos programas de modelagem devem ser avaliadas, verificando a qualidade estereoquímica e possíveis erros estruturais. Para a validação são usados principalmente o gráfico de Ramachandran (RAMACHANDRAN, 1963) e o cálculo do RMSD (*Root Mean Square Deviation*),

baseados nos ângulos de torção da cadeia principal da proteína e na distância média entre os átomos da proteína molde para validação, respectivamente.

No gráfico de Ramachandran, cada resíduo da cadeia principal de uma proteína possui quatro ângulos de torção ϕ (phi), ψ (psi), ω (omega) e χ_1 , determinantes para o enovelamento de uma proteína (Figura 8). A faixa permitida de ângulos é restritiva, portanto, variações nos ângulos de torção não dão muita variedade conformacional (HOOFT; SANDER; VRIEND, 1997). O gráfico de Ramachandran é um gráfico 2D dos ângulos de torção ϕ - ψ que, comparados com uma distribuição determinada, permite visualizar a conformação proteica, de uma forma simplificada (RAMACHANDRAN, 1963). Esses gráficos ϕ - ψ foram aceitos como um aspecto central da validação da estrutura proteica, sendo rotineiramente aplicados durante a deposição de estruturas no PDB (BERMAN *et al.*, 2000).

Figura 8 - Visão geral dos ângulos de torção (ϕ , ψ , ω e χ_1) da cadeia central de proteínas

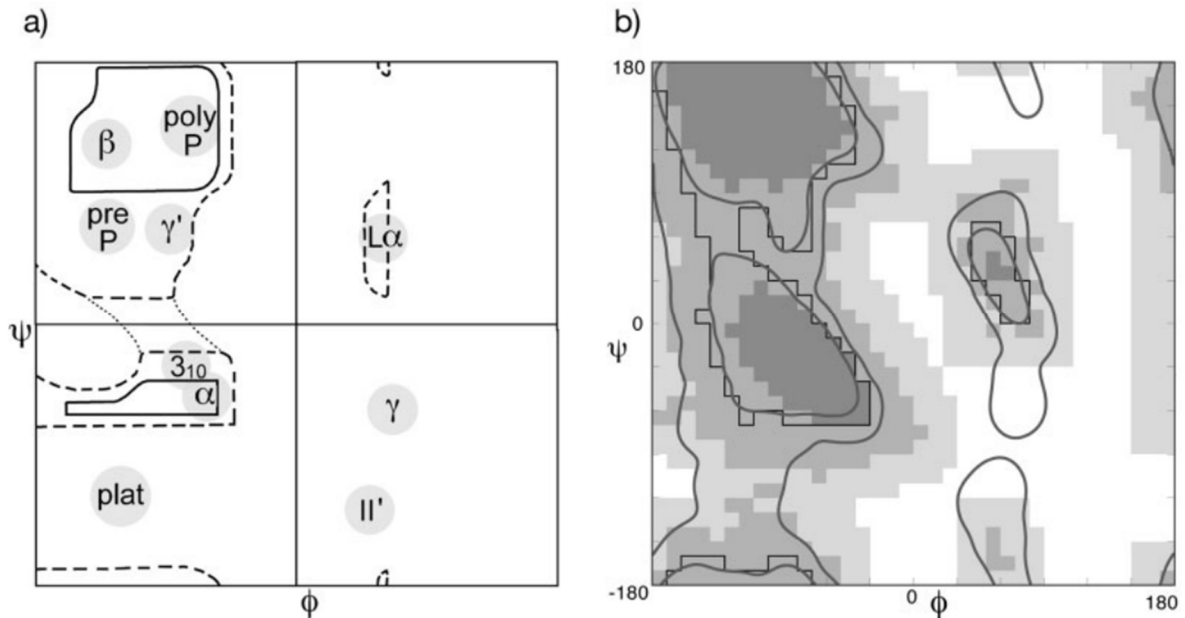


Fonte: Gibbs (2001).

Os ângulos ϕ e ψ se agrupam em regiões distintas no gráfico de Ramachandran, correspondendo a uma estrutura secundária específica (HO; BRASSEUR, 2005). Os gráficos de ϕ e ψ para proteínas individuais se tornaram principais para a validação da estrutura, porque os valores de ϕ e ψ não são otimizados no processo de refinamento e, portanto, fornecem um bom indicativo de áreas locais problemáticas (MORRIS *et al.*, 1992). A localização aproximada e a forma

tomada pelas regiões mais favoráveis de baixa energia podem ser visualizadas na Figura 9a e Figura 9b.

Figura 9 - Tratamentos e definições clássicas de ϕ e ψ



Fonte: Kleywegt e Jones (1996); Lovell *et al.* (2003).

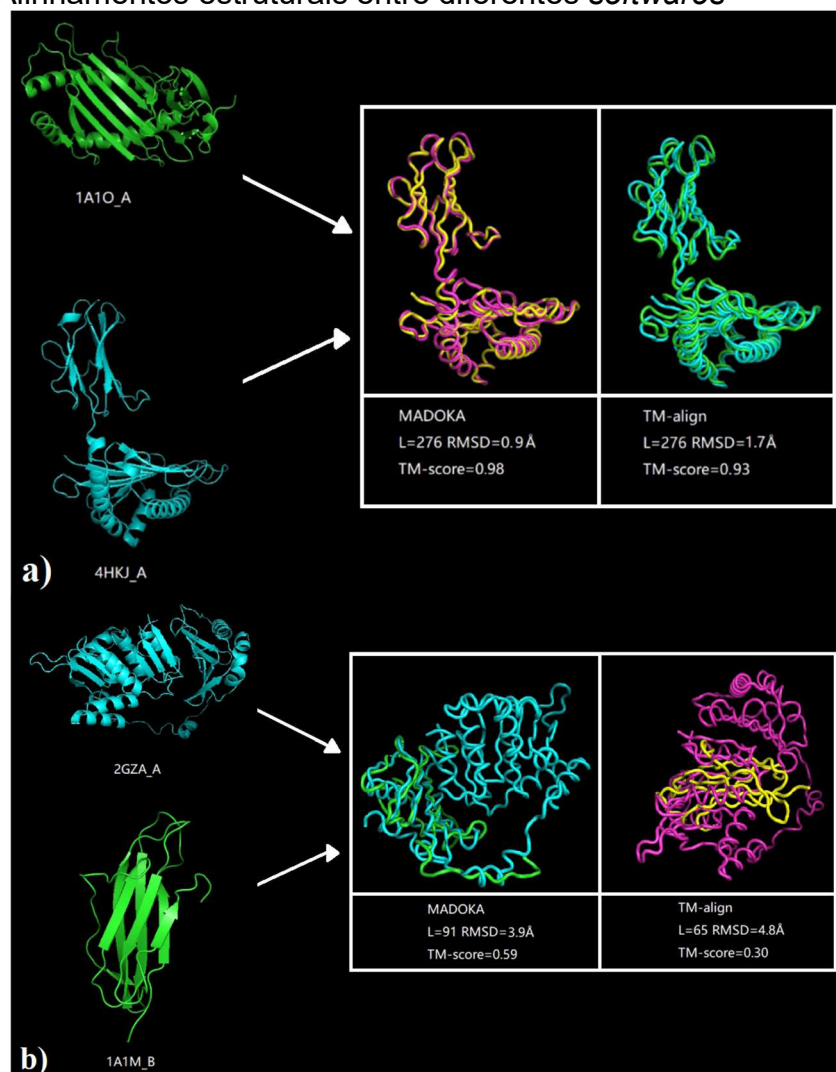
a) Limites definidos pelas sobreposições atômicas de esfera dura (RAMACHANDRAN; SASISEKHARAN, 1968); Imagem mostra a localização aproximada das regiões $\alpha+3_{10}$, β +polyPro e $L\alpha$; linhas tracejadas e pontilhadas incluem regiões permitidas. **b)** Regiões ϕ e ψ usadas para validação de estruturas protéicas experimentais; as áreas escuras, cinza escuro e cinza claro são as regiões "mais favoráveis", "adicionalmente permitidas" e "generosamente permitidas", respectivamente, do ProCheck (LASKOWSKI *et al.*, 1993); o contorno preto inclui a região única "estritamente permitida" do WhatCheck.

O cálculo do Desvio Médio Quadrático (do inglês Root Mean Square Deviation, abreviado RMSD), é a medida em angstroms da distância média entre os átomos correspondentes da proteína alvo (predita) com a proteína molde (experimental), de modo que quanto menor for a distância, mais semelhantes são as duas estruturas (REVA; FINKELSTEIN; SKOLNICK, 1998). São considerados bons modelos preditos aqueles com RMSD menor ou igual a 3Å, no caso de proteínas homólogas, ou seja, com similaridade maior ou igual a 30% (CHOTHIA; LESK, 1986). Também é importante a quantidade de átomos considerados na comparação, pois alguns algoritmos tendem a desconsiderar pares de átomos em posições muito divergentes para se obter melhor valor de RMSD.

Ferramentas computacionais são comumente utilizadas para calcular o valor de RMSD, através do alinhamento entre as estruturas proteicas. São inúmeras as ferramentas que propõem realizar o alinhamento ideal sem desconsiderar muitos

resíduos, como DALI (HOLM; SANDER, 1993), CE (SHINDYALOV; BOURNE, 1998), SAL (KIHARA; SKOLNICK, 2003), FATCAT (YE; GODZIK, 2003), TM-align (ZHANG; SKOLNICK, 2005), Fr-TM-align (PANDIT; SKOLNICK, 2008), FAST (ZHU; WENG, 2005), CASSERT (MROZEK; MAŁYSIAK-MROZEK, 2013), DeepAlign (WANG *et al.*, 2013), MICAN-SQ (MINAMI *et al.*, 2018), ou até pelo PyMOL (DELANO, 2002). Na Figura 10 estão exemplos que podem ser considerados modelos bons ou ruins, obtidos por ferramentas diferentes.

Figura 10 - Alinhamentos estruturais entre diferentes *softwares*



Fonte: Adaptado de DENG *et al.* (2019).

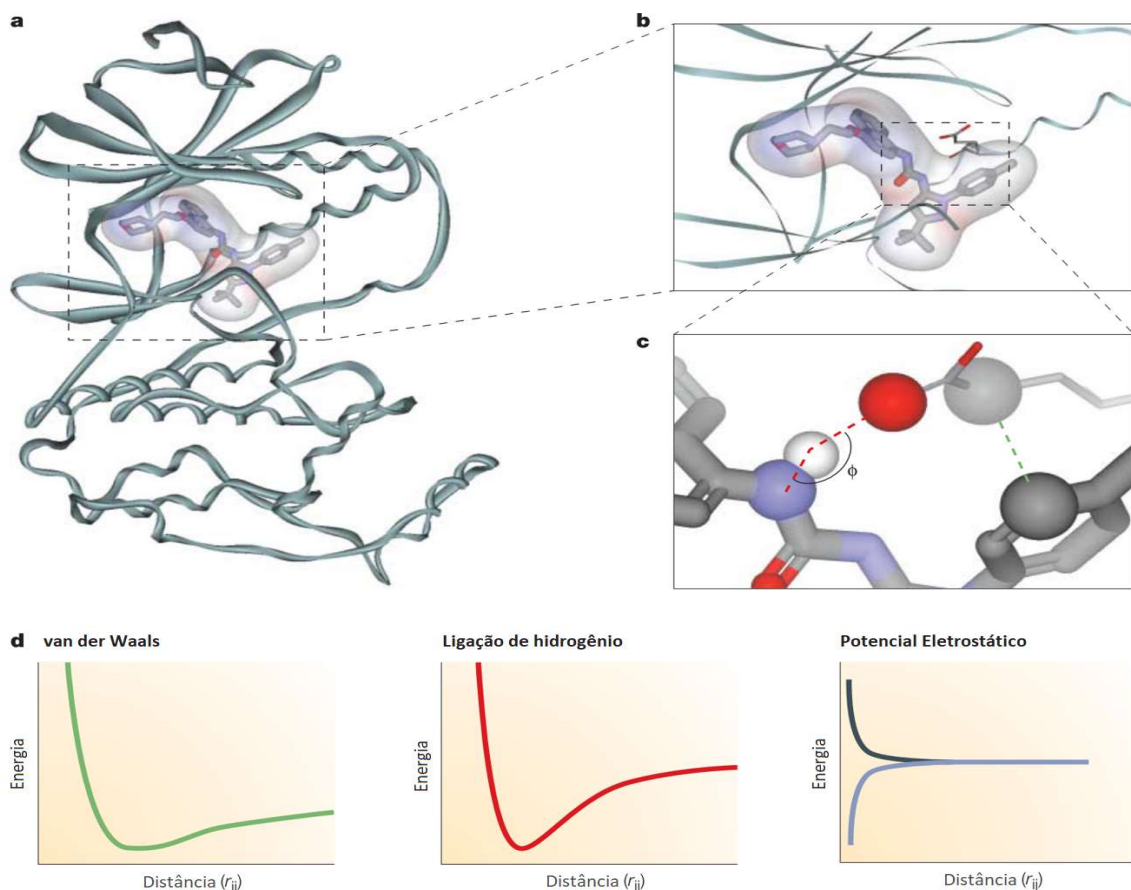
Para o alinhamento entre as duas proteínas, as ferramentas MADOKA (DENG *et al.*, 2019) e TM-align (ZHANG; SKOLNICK, 2005) foram utilizadas. **a)** alinhamento entre 1A1O_A (276 resíduos) e 4HKJ_A (277 resíduos), com RMSD variando de 0,9Å até 1,7Å. **b)** alinhamento entre 2GZA_A (334 resíduos) e 1A1M_B (99 resíduos), com RMSD variando de 3,9Å até 4,8Å. O cálculo do TM-score também avalia a qualidade do alinhamento entre os resíduos alinhados, mas independe do tamanho da proteína, quanto maior melhor, enquanto L dita o tamanho original da estrutura original.

Independentemente do método utilizado para se obter a estrutura tridimensional da proteína, a afinidade dos ligantes por seu sítio ativo podem ser testados por métodos computacionais.

2.7 Ligantes

As funções desempenhadas por muitas proteínas envolvem interações intermoleculares com uma variedade de moléculas, como hormônios, metabólitos, neurotransmissores e peptídeos, que damos o nome de ligante. O potencial de cada força intermolecular depende da distância entre os dois átomos envolvidos na interação (Figura 11) (GOVINDARAJ; BRYLINSKI, 2018; KITCHEN *et al.*, 2004).

Figura 11 - Representação das forças de ligação



Fonte: Adaptado de Kitchen *et al.* (2004).

a. Estrutura de uma proteína quinase interagindo com o seu inibidor. **b.** Visão ampla do sítio de ligação. **c.** Visão ampla da interação entre a proteína e o ligante, com a ligação de hidrogênio e van der Waals representadas em vermelho e verde, respectivamente. **d.** Representações do potencial energético das diferentes forças intermoleculares em função da distância entre átomos (i e j). À esquerda está representado o potencial da interação van der Waals e no meio da ligação de hidrogênio, ambos com energia em e distância inversamente proporcionais. À direita está o potencial eletrostático, representado em azul dois átomos com carga igual e em preto com cargas opostas.

Essa interação pode ser específica e reversível, podendo ativar ou desativar um receptor, que ao induzir ou impedir que mudanças na conformação ocorram, permite ou impossibilita de desempenhar suas funções, respectivamente.

Forças intermoleculares contribuem aumentando a afinidade entre a proteína e o ligante, como ligações iônicas, ligações de hidrogênio e forças de Van der Waals, com exceção de ligações covalentes irreversíveis, pois são infrequentes em sistemas biológicos.

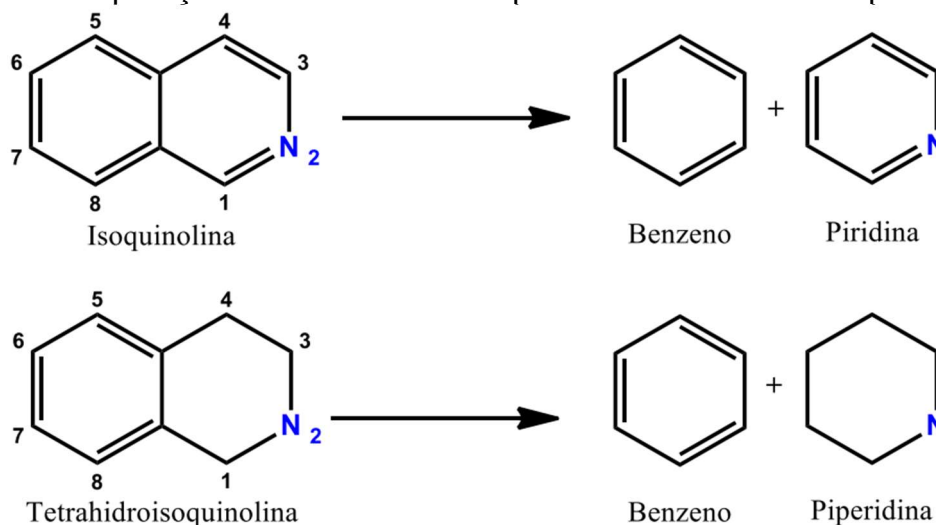
A depender da força de ligação e do sítio onde o ligante interage com uma proteína alvo, seu potencial para ser um candidato a droga pode ser avaliado. Para o ligante não ser tóxico ao corpo humano ele deve atender à regra dos cinco de Lipinski e possuir propriedades farmacocinéticas favoráveis de absorção, distribuição, metabolismo e excreção (ADME) (KALYAANAMOORTHY; CHEN, 2011). A regra dos cinco de Lipinski (RO5, do inglês *Rule of five*) dita que o ligante não deve ter mais do que cinco doadores de hidrogênio, dez aceptores de hidrogênio, massa menor que 500 daltons, coeficiente de partição que não exceda a cinco, e não mais do que uma dessas regras violadas (LIPINSKI, 2004).

Ligantes que atendam às RO5 e propriedades de ADME podem ser buscados em bancos de dados de compostos comercialmente disponíveis, como é o caso do ZINC (STERLING; IRWIN, 2015) ou desenhados para testes *in silico*. Para o estudo de caso, os ligantes desenhados são derivados sintéticos de alcalóides tetrahydroisoquinolínicos, considerados candidatos a droga.

2.7.1. Alcalóides tetrahydroisoquinolínicos

Alcalóides são moléculas encontradas na natureza possuindo baixo peso molecular e contendo em sua estrutura o nitrogênio, o que lhes garante uma característica básica, sendo a maioria derivados de aminoácidos, podendo ser classificados em isoquinolínicos e tetrahydroisoquinolínicos (Figura 12).

Figura 12 - Comparação estrutural entre isoquinolinas e tetrahydroisoquinolinas



Fonte: Adaptado de Silva (2016).

Os alcalóides isoquinolínicos são compostos orgânicos heterocíclicos aromáticos produzidos por plantas que possuem propriedades farmacológicas comprovadas, como analgésico, relaxante muscular e antimicrobiana (HAGEL; FACCHINI, 2013). São formados pela junção de um anel benzênico e piridina, e têm como origem o aminoácido aromático tirosina, podendo formar aproximadamente 2500 diferentes estruturas, como por exemplo, a (S)-norcoclaurina, um precursor importante para formação de ainda outros diversos compostos (STADLER; ZENK, 1990).

Já os alcalóides tetrahydroisoquinolínicos são compostos orgânicos derivados das isoquinolinas presentes em plantas como *Cactaceae*, *Papaveraceae* e *Fumariaceae*. O que difere as tetrahydroisoquinolinas das isoquinolinas é que no lugar da piridina se encontra a piperidina, que em conjunto com o anel benzênico forma sua estrutura base (Figura 12).

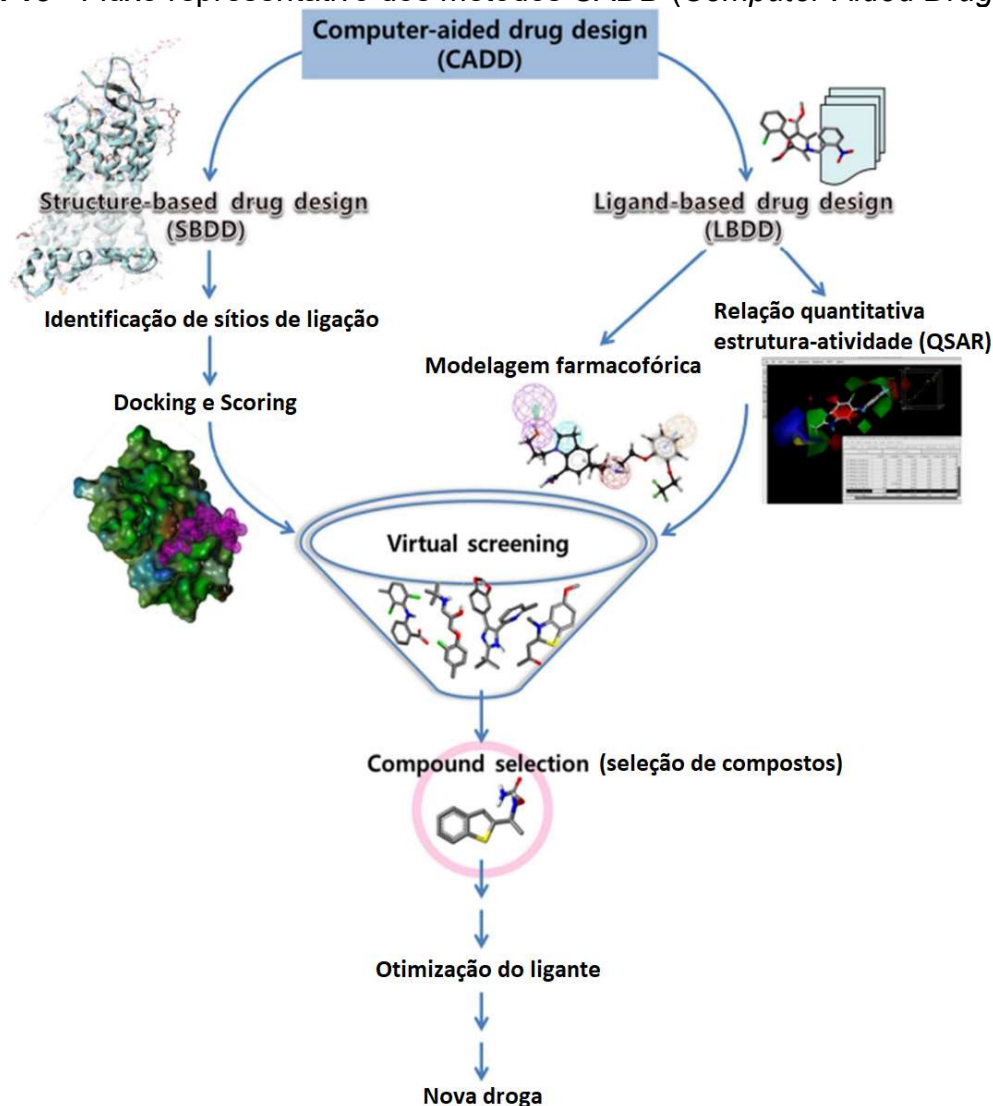
Um dos principais objetivos na descoberta de fármacos é identificar novos ligantes que possuam alta afinidade por proteínas alvo a fim de provocar uma resposta biológica desejada. Considerando as propriedades conhecidas de alcalóides, derivados de tetrahydroisoquinolínicos são potenciais ligantes para avaliarmos a interação com nossas proteínas *hub*. Para testar o potencial de interação e identificar novos candidatos a droga (ligantes), métodos de desenho de drogas auxiliados por computadores (CADD) estão sendo cada vez mais utilizados.

2.8 Computer-Aided Drug Design Methods (CADD)

Existem diversos métodos para identificar novos “compostos líderes” (*lead compounds*) que interajam com proteínas de interesse, o mais relevante sendo a triagem em larga escala ou *high-throughput screening* (HTS), que utiliza da robótica para testar uma grande quantidade de compostos químicos contra seus alvos. Porém o HTS é muito custoso e requer todo um espaço laboratorial ideal para a realização da técnica, além do alto custo de síntese dos compostos, necessários para a triagem. Um método alternativo é a triagem virtual ou *virtual screening* (VS) pois utiliza apenas poder computacional (*in silico*) para o mesmo propósito sem a necessidade de síntese e de espaço laboratorial, tentando identificar e otimizar compostos líderes, ao predizer quais diferentes conformações dos mesmos interagem bem com seus respectivos alvos (SHOICHET, 2004).

O *virtual screening* faz parte dos métodos para se desenhar drogas com o auxílio de computadores, conhecido como *Computer-Aided Drug Design* (CADD), que utilizam de ferramentas matemáticas na forma de programas para manipular e quantificar as propriedades de uma droga em potencial (YU; MACKERELL, 2017). É possível então, com a utilização de computadores, descobrir, desenvolver e analisar drogas e moléculas com atividade biológica desejada (KUMAR; JHA, 2017). Em geral, são dois os métodos CADD utilizados em *virtual screening*, o baseado no ligante (LBVS) e na estrutura proteica (SBVS), representados na Figura 13.

Figura 13 - Fluxo representativo dos métodos CADD (*Computer-Aided Drug Design*)



Fonte: Adaptado de Macalino *et al.*, 2015.

A abordagem *structure-based* (baseado na estrutura e à esquerda na figura) consiste em usar a estrutura 3D do alvo (receptor ou enzima) para identificar possíveis ligantes. Pelo método SBDD são identificados os sítios de ligação (*binding site identification*) do alvo, testados (pelo *docking*) os ligantes de interesse e obtidos os valores respectivos de *score*. Porém, a abordagem *ligand-based* (baseado no ligante e à direita na figura) são obtidas informações de um conjunto de ligantes ativos contra um alvo relevante, sendo duas as técnicas, a modelagem farmacofórica ou relação quantitativa estrutura-atividade (QSAR). Seja pelo SBDD ou LBDD, o *virtual screening* é realizado, selecionados e otimizados os melhores ligantes, e possivelmente desses processos obtido uma nova droga.

2.8.1 Ligand Based Virtual Screening (LBVS)

Nos casos em que a estrutura 3D da proteína alvo está ausente e seus sítios de ligação são desconhecidos, as informações obtidas de um conjunto de ligantes ativos contra um alvo relevante (receptor ou enzima) podem ser usadas para identificar propriedades estruturais ou físico-químicas significativas responsáveis pela atividade biológica observada e otimizar drogas *lead-like* (MACALINO *et al.*, 2015;

HAMZA; WEI; ZHAN, 2012). Nos métodos baseados no ligante se supõe que compostos estruturalmente ou físico-quimicamente similares exibam resposta biológica e interação similares com o mesmo alvo (PRATHIPATI; DIXIT; SAXENA, 2007). Os métodos LBVS podem ser divididos em Relação Quantitativa Estrutura-Atividade (QSAR) e em farmacóforos.

Relação Quantitativa Estrutura-Atividade (QSAR): São baseados na premissa de que alterações na atividade biológica (ou farmacológica) de compostos estão associadas a variações estruturais e moleculares (LAVECCHIA, 2013). Um modelo estatístico é gerado a partir dessa correlação para desenvolver e prever matematicamente a propriedade biológica de novos compostos (MELO-FILHO; BRAGA; ANDRADE, 2014). Existem inúmeras estratégias QSAR, que vão desde 3D até 6D, com o intuito de resolver as deficiências da estratégia QSAR multidimensional menos avançada. Estratégias 3D-QSAR incluem Topomer CoMFA (CRAMER, 2003), Relação Estrutura-Atividade Espectral (S-SAR) (PUTZ; LACRĂMĂ, 2007), Adaptação de Campos para Comparação Molecular (AFMoC) (GOHLKE; KLEBE, 2002), Análise de Interação de Resíduo Comparativo (CoRIA) (DATAR *et al.*, 2006), e o mais utilizado, o Análise de Campo Molecular Comparativo (CoMFA) (CRAMER; PATTERSON; BUNCE, 1988).

Modelagem Farmacofórica: Esta abordagem visa identificar compostos contendo diferentes estruturas, mas com um arranjo 3D semelhante dos principais grupos funcionais que interagem com o alvo (VUORINEN; SCHUSTER, 2015). Ao explorar a conformação bioativa dos compostos candidatos, as informações dos sítios de ligação (*binding site*) podem ser incorporadas neste modelo. A modelagem de farmacóforo também pode ser realizada no estágio de alinhamento molecular dos estudos de modelagem QSAR (MELO-FILHO; BRAGA; ANDRADE, 2014). Vários programas são comumente utilizados para geração automática de farmacóforo, incluindo Discovery Studio, PHASE (DIXON *et al.*, 2006), LigandScout (WOLBER; LANGER, 2005) e MOE (MOE, 2005).

2.8.2 Structure Based Virtual Screening (SBVS)

Nos casos em que a estrutura 3D de uma proteína alvo está disponível, a triagem virtual baseada em estrutura (SBVS) é o método de escolha nos estágios iniciais do processo de descoberta de medicamentos (CAVASOTTO; ORRY;

ANDREW, 2007). SBVS visa explorar a afinidade de um ligante por uma proteína para identificar compostos que se ligam fortemente aos sítios de ligação de alvos biologicamente relevantes (GHOSH *et al.* 2006).

Após a predição ou obtenção da estrutura 3D proteica, diversas etapas são necessárias para realização da triagem virtual: 1) identificação de *pockets* (bolsões) drogáveis; 2) preparo das proteínas alvo e ligantes, ao adicionar cargas e converter os arquivos no formato do PDB (Protein Data Bank) em PDBQT (adição da carga parcial, Q, e o tipo do átomo, T); 3) processo de *docking*, que tenta prever precisamente as melhores conformações de um composto dentro do *pocket* da proteína; 4) análise de *docking* com base na pontuação (*score*) de cada conformação, medida que tenta prever a probabilidade do ligante realmente se ligar ao alvo; 5) priorização de compostos para teste (CAVASOTTO; ORRY; ANDREW, 2007; LYNE, 2002).

Como primeira etapa, a identificação e delimitação de *pockets* é de suma importância para predição eficiente de conformações e interações do ligante com a proteína por programas de *docking*. *Pockets* ou bolsões são cavidades na superfície ou no interior de proteínas que possuem propriedades adequadas para interagirem com ligantes. Bolsões podem ser bons ou ruins, a depender de sua capacidade de fornecer sítios de ligação para drogas. Com base na pontuação da propriedade *druggability score*, os *pockets* podem ser classificados em quatro categorias: não drogável ($0,0 \leq DS < 0,2$), pouco drogável ($0,2 \leq DS < 0,5$), drogável ($0,5 \leq DS < 0,7$) e altamente drogável ($0,7 \leq DS \leq 1,0$) (RAMOS *et al.*, 2018).

Diversos *softwares* desempenham o papel de predizer *pockets*, alguns deles são: I-TASSER (ROY; KUCUKURAL; ZHANG, 2010), SCREEN (NAYAL; HONIG, 2006), POCASA (YU *et al.*, 2010), CASTp (BINKOWSKI; NAGHIBZADEH; LIANG, 2003), MSPocket (ZHU; PISABARRO, 2011), fPocket (LE GUILLOUX; SCHMIDTKE; TUFFERY, 2009) e PocketDepth (KALIDAS; CHANDRA, 2008).

Depois de preparadas as proteínas alvo e os ligantes, diversos programas de *docking* podem ser utilizados, diferindo no algoritmo, no manuseio da flexibilidade de ligantes e proteínas, na utilização ou não de moléculas de água, no emprego das funções de pontuação e no tempo necessário para o computador encaixar uma molécula em um determinado *pocket* da proteína. Alguns programas comumente utilizados são: Dock (EWING *et al.*, 2001), FlexX (RAREY *et al.*, 1996), Slide

(SCHNECKE; KUHN, 2000), Gold (JONES; WILLETT; GLEN, 1995), ICM (ABAGYAN; TOTROV; KUZNETSOV, 1994), bem como AutoDock Vina (TROTT; OLSON, 2010).

Inicialmente, a maioria dos algoritmos de *docking* foram desenvolvidos com base em proteínas rígidas e ligantes flexíveis, ignorando qualquer alteração conformacional proteica que possa ocorrer antes ou após a ligação. No entanto, ao se perceber a importância da flexibilidade no *docking*, novos algoritmos foram desenvolvidos. A flexibilidade de uma proteína pode ser introduzida por diferentes métodos: implicitamente, ao executar múltiplas etapas de *docking* a partir de diversas conformações geradas (*ensemble docking*); ou explicitamente, permitindo flexibilidade na cadeia lateral durante o encaixe ou em uma etapa de aprimoramento. Porém, para realização de *docking* flexível, é necessário o conhecimento das regiões flexíveis e mudanças conformacionais que possam ocorrer na proteína, muitas vezes derivado de simulações de dinâmica molecular, um processo computacionalmente custoso (BONVIN, 2006).

Independente do *docking* ser rígido ou flexível, uma vez que as inúmeras conformações do ligante foram geradas pelo *software*, elas devem ser ranqueadas de acordo com a qualidade da mesma em relação à outras conformações nas bases de dados ou do mesmo ligante. Programas pontuam essas conformações de pelo menos três formas diferentes: *physical-based (force-field)*, método pontua com base nos campos de força dos átomos, como Amber ou CHARMM, ao limitar uma região (*grid*) para *docking*; *empirical*, baseado em propriedades físico-químicas como a quantidade de ligações de hidrogênio para se obter um valor estimado de energia livre; ou *knowledge-based*, método estatístico baseado na estrutura do complexo proteína-ligante que calcula a soma de energias livres de contatos interatômicos. Programas citados como AutoDock, Dock e ICM utilizam de *force-field* para pontuação, enquanto FlexX, Slide e Gold utilizam de *empirical*.

3. OBJETIVOS

3.1 Objetivo geral

Identificar e validar proteínas essenciais de *C. ulcerans* com potencial uso para alvo no desenvolvimento de fármaco.

3.2 Objetivos específicos

- Gerar e validar rede de interação proteína-proteína para *C. ulcerans*.
- Identificar, validar e caracterizar as proteínas essenciais.
- Testar 42 derivados sintéticos de alcalóides tetrahydroisoquinolínicos contra as proteínas essenciais de *C. ulcerans*.

4 METODOLOGIA

4.1 Bancos de dados biológicos

Para a execução do trabalho, em suas diversas etapas, foram utilizadas diversas bases de dados publicamente disponíveis (Tabela 2).

Tabela 2 - Lista das bases de dados públicas

Nome	Descrição	Proteínas	Versão	Local	Referência
COG	Cluster de grupos ortólogos de proteínas		2014	http://clovr.org/	(TATUSOV <i>et al.</i> , 2001)
DEG	Base de dados de proteínas essenciais	53.885	15.2	http://tubic.tju.edu.cn/deg/	(LUO <i>et al.</i> , 2014)
NCBI	Banco de dados de genomas	na	na	https://www.ncbi.nlm.nih.gov/	(NCBI, 1988)
PDB	Banco de dados de proteínas com estrutura tridimensional definida	161.273	na	https://www.rcsb.org/	(ROSE <i>et al.</i> , 2015)
STRING	Bases de dados de interação proteína-proteína	24.584.628	11.0	http://string-db.org/	(SZKLARCZYK <i>et al.</i> , 2019)
UNIPROT	Banco de dados de anotação de sequências proteicas	178.316.438	2019-10	https://www.uniprot.org/	(CONSORTIUM, U., 2019)
GO	Base de dados com informações sobre funções de produtos genéticos	1.359.256	2020-02	http://geneontology.org/	(CONSORTIUM, G. O., 2019)

Fonte: Autor. Legenda: na, sem versão.

4.2 Genomas completos de *C. ulcerans*

Dez genomas de *C. ulcerans* com sequenciamento completo, disponíveis no NCBI (NCBI, 2020b), foram utilizados para construir as RIPs (Tabela 3).

Tabela 3 - Genoma completo de dez linhagens de *C. ulcerans*

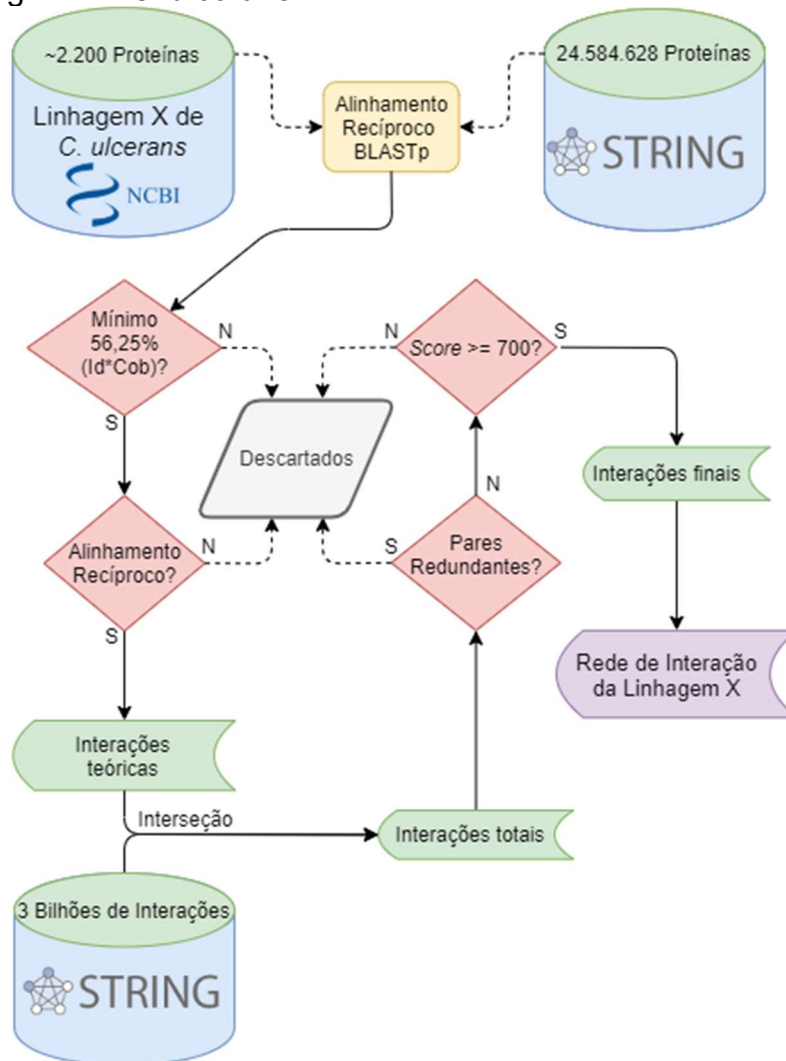
Linhagem	Tamanho (Mb)	GC%	Genes	Proteínas	Referência
102	2.579	53.40	2339	2250	(SEKIZUKA <i>et al.</i> , 2012)
809	2.502	53.30	2250	2176	(TROST <i>et al.</i> , 2011)
5146	2.466	53.30	2212	2131	(TATUSOVA <i>et al.</i> , 2015)
131001	2.483	53.30	2242	2143	na
131002	2.434	53.40	2179	2099	(TATUSOVA <i>et al.</i> , 2015)
210931	2.509	53.25	2280	2196	(TATUSOVA <i>et al.</i> , 2015)
210932	2.484	53.30	2230	2145	(VIANA <i>et al.</i> , 2014)
BR-AD22	2.606	53.40	2373	2276	(TROST <i>et al.</i> , 2011)
FRC11	2.442	53.30	2182	2105	(DE JESUS BENEVIDES <i>et al.</i> , 2015)
FRC58	2.542	53.30	2298	2218	(DE SOUSA SILVA <i>et al.</i> , 2014)

Fonte: adaptado do NCBI (2020b). Legenda: na, sem referência.

4.3 Predição das Redes de Interação Proteína-Proteína (RIP)

Para a predição da RIP o método utilizado foi o *interolog mapping*, seguindo o fluxograma da Figura 14 e detalhado subsequentemente.

Figura 14 - Fluxograma de predição das redes de interação proteína-proteína (RIP) das dez linhagens de *C. ulcerans*



Fonte: Autor.

4.3.1 Alinhamento *C. ulcerans* contra STRING

As sequências proteicas de cada uma das dez linhagens de *C. ulcerans* foram alinhadas contra todas as 24.584.628 sequências proteicas presentes no banco de dados STRING, visando identificar proteínas de *C. ulcerans* homólogas a algum organismo de STRING. Após, as proteínas de STRING foram alinhadas contra as proteínas de cada linhagem de *C. ulcerans*, visando construirmos alinhamento

recíproco. O programa BLASTp com *e-value* configurado para e^{-10} foi utilizado para o alinhamento (CAMACHO *et al.*, 2009).

Objetivando mapear das bases de dados somente as interações realmente ortólogas, foram desconsiderados então os alinhamentos que apresentaram identidade*cobertura menor que 56,25% (valor que representa uma média de 75% de identidade e 75% de cobertura). Adicionalmente, foram identificadas e somente consideradas as proteínas de *C. ulcerans* com alinhamento recíproco (AR) contra o STRING.

4.3.2 Mapeamento das interações

A partir das proteínas com alinhamento recíproco (AR) foram geradas as interações teóricas, um conjunto de IPP contendo o produto cartesiano entre todas as proteínas com AR de *C. ulcerans*:

$$\text{Interações teóricas} = \text{AR} \times \text{AR}$$

Assim, as interações teóricas foram mapeadas nas interações do STRING, obtendo-se as interações reais totais, ou seja, a intersecção entre ambos conjuntos:

$$\text{Interações reais} = \text{Interações teóricas} \cap \text{Interações de STRING}$$

O conjunto de interação totais contém interações redundantes mapeada de organismos diferentes de STRING. As interações redundantes foram retiradas, permanecendo somente a interação com maior identidade*cobertura, a mais similar à *C. ulcerans*.

Adicionalmente, visando mapear da base de dados STRING as interações altamente confiáveis ou experimentais, foram consideradas somente as interações distintas com *score* de confiança ≥ 700 .

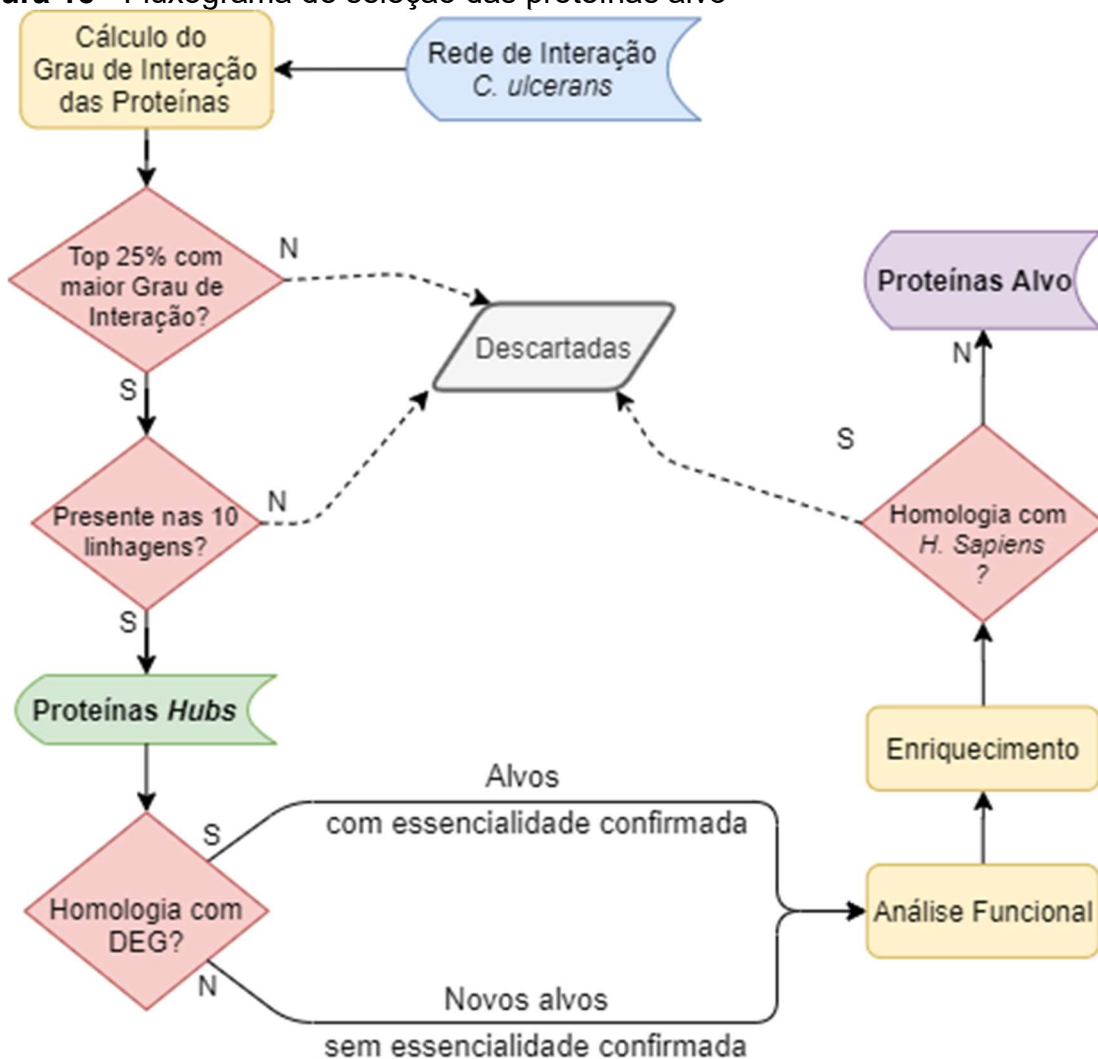
Assim, as interações distintas, não redundantes e com *score* ≥ 700 foram utilizadas para gerar as RIPs de cada linhagem.

Adicionalmente, objetivando validar as RIP geradas, foi plotado o gráfico de distribuição do grau de interação pelo programa Cytoscape (SHANNON *et al.*, 2003).

4.4 Seleção das proteínas alvo

Para a seleção de proteínas alvo se levou em consideração o grau de interação de cada proteína na RIP, sua essencialidade e a homologia com o hospedeiro, processo resumido na Figura 15.

Figura 15 - Fluxograma de seleção das proteínas alvo



Fonte: Autor.

4.4.1 Identificação e validação das proteínas *hub*

Visando identificar as proteínas *hub* de *C. ulcerans* e validar a essencialidade, as RIPs das dez linhagens foram submetidas à análise topológica. O grau de interação foi calculado pelo *plugin* NetworkAnalyzer (ASSENOV *et al.*, 2008) do Cytoscape (SHANNON *et al.*, 2003). Do total de proteínas das RIPs, foram selecionadas somente aquelas com um grau de interação maior que zero.

Como o grau de interação tende a ser maior em proteínas essenciais (DILUCCA; CIMINI; GIANANTI, 2017), foi selecionado os primeiros 25% de proteínas com maior grau de interação e presentes nas dez linhagens para compor as proteínas *hub*. A linhagem FRC58 foi utilizada como referência para representar as proteínas *hub* de *C. ulcerans* conservadas nas 10 linhagens.

Para validar a essencialidade, as proteínas *hub* foram alinhadas contra a base de dados de genes essenciais DEG (LUO *et al.*, 2014) utilizando o programa BLASTp (ALTSCHUL *et al.*, 1990) com *evaluate* $1e^{-10}$, filtro de baixa complexidade desativado e matriz de substituição BLOSUM62. As proteínas *hub* com homólogos no DEG tiveram a essencialidade confirmada, já as proteínas sem homólogos no DEG tiveram essencialidade verificada por anotação, sendo consideradas novos alvos.

Adicionalmente, as proteínas *hub*, com ou sem homólogos no DEG, passaram pela análise funcional e de enriquecimento. As proteínas foram classificadas em categorias do COG (*Cluster of Orthologous Genes*) pela ferramenta eggNog-mapper (HUERTA-CEPAS *et al.*, 2017), tendo suas funções anotadas e discutidas. A ferramenta BLAST2GO (GÖTZ *et al.*, 2008) foi utilizada para o enriquecimento e classificação das proteínas *hub* nos termos do *Gene Ontology* (GO). Realizadas todas estas etapas, as proteínas *hub* passaram a ser denominadas proteínas alvo.

4.4.2 Homologia com hospedeiro

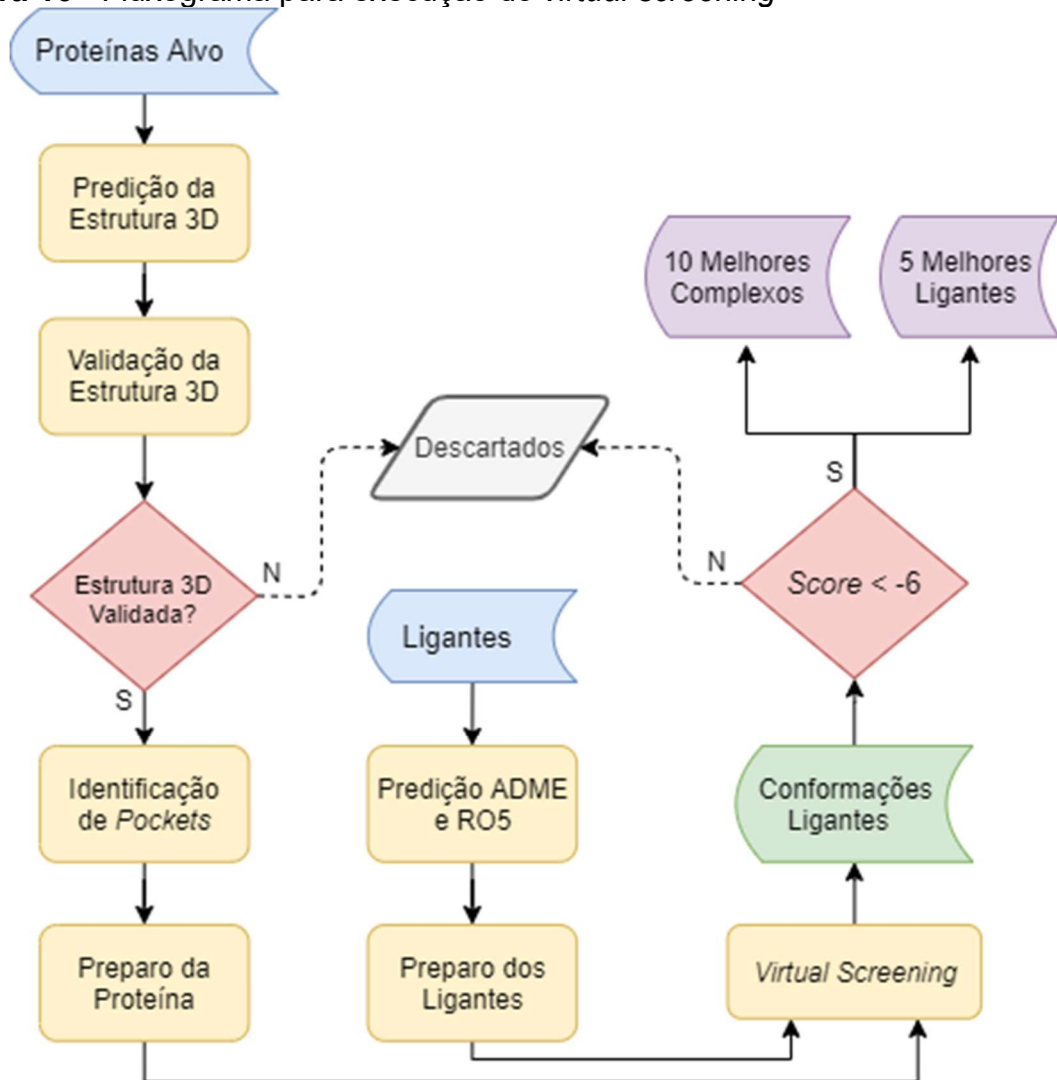
Um bom alvo para drogas deve ser específico de forma a não ter homólogo no organismo hospedeiro. Assim, foi verificada a existência de proteínas ortólogas de *C. ulcerans* no hospedeiro (*H. sapiens*) com uso do programa BLASTp (ALTSCHUL *et al.*, 1990) para alinhar as proteínas *hub*. Foi utilizado o valor padrão para os parâmetros do BLASTp, com *e-value* configurado para $1e^{-10}$. As proteínas essenciais

de *C. ulcerans* com identidade*cobertura maior ou igual a 30% às proteínas de *H. sapiens* foram descartadas.

4.5 Estudo de caso

Para testar a possível interação entre as proteínas alvo (*hub*) e moléculas candidatas a fármaco, diversas etapas de preparo e predição foram realizadas, possibilitando em seguida executar o *virtual screening*, processo representado na Figura 16.

Figura 16 - Fluxograma para execução do *virtual screening*



Fonte: Autor.

4.5.1 Proteínas alvo

Para serem utilizadas no estudo de caso as proteínas alvo tiveram suas estruturas tridimensionais preditas, validadas, preparadas e seus *pockets* identificados.

A predição das estruturas das proteínas alvo foi realizada por métodos de homologia e *threading* ao submeter as sequências das proteínas à ferramenta Phyre2 com configurações padrão (KELLEY *et al.*, 2015). Para validar as estruturas tridimensionais preditas foi usado o gráfico de Ramachandran, sendo excluindo os modelos que tiveram menos de 90% de seus aminoácidos nas regiões permitidas.

Para preparo das proteínas, os arquivos em formato PDB de cada proteína foram convertidos para PDBQT pelo programa "prepare_receptor4.py" do pacote do MGLTools versão 4.1 (MORRIS *et al.*, 2009), adicionando cargas parciais gasteiger, combinando hidrogênios não-polares e detectando carbonos aromáticos.

Os *pockets* de cada proteína foram identificados com o programa fpocket project (LE GUILLOUX; SCHIMIDTKE; TUFFERY, 2009), sendo removidos aqueles com a propriedade *druggability score* menor que 0,5.

4.5.2 Derivados sintéticos de alcalóides tetrahydroisoquinolínicos

Para o estudo de caso foram selecionados 42 derivados sintéticos de alcalóides tetrahydroisoquinolínicos (ligantes e candidatos a droga) desenhados no laboratório de Síntese Orgânica pelo grupo do prof. Luis Cezar Rodrigues da Universidade Federal da Paraíba.

A fim de prever os parâmetros ADME e se atendem à RO5, os ligantes foram submetidos ao SwissADME (DAINA; MICHIELIN; ZOETE, 2017) no formato SMILES (*Simplified molecular-input line-entry system*).

Para preparo dos ligantes, os arquivos em formato PDB de cada ligante foram convertidos para PDBQT pelo programa "prepare_ligand4.py" do pacote do MGLTools versão 4.1 (MORRIS *et al.*, 2009), adicionando cargas parciais gasteiger, hidrogênios não-polares, detectando carbonos aromáticos e flexibilizando o ligante ao adicionar torções. Foi também utilizado o parâmetro opcional de reparo das ligações de hidrogênio.

Após considerar as propriedades farmacológicas e potencial de atuarem como drogas, os 42 ligantes tiveram sua afinidade testada com as proteínas alvo por processo de *docking*.

4.5.3 Execução do *virtual screening* (VS)

Para executar o *docking*, além de ter proteínas e ligantes preparados, é necessário informar ao programa uma série de informações através de um arquivo de configuração, como: a proteína e o ligante que serão testados, o espaço de busca baseado nas coordenadas do *pocket*, o número máximo de conformações (*num_modes*), o tempo máximo levado para procurar as conformações (*exhaustiveness*) e a diferença máxima entre o melhor e pior *score* (*energy_range*).

Algoritmos *in house* foram desenvolvidos para gerar os arquivos de configuração (.conf) necessários para automação do VS no programa AutoDock Vina (TROTT; OLSON, 2010). Os arquivos de configuração foram distribuídos entre diversos computadores disponíveis para uma execução mais rápida.

Ao executar o *docking*, inúmeras conformações para cada ligante são geradas para cada *pocket* de cada proteína, cada conformação com seu respectivo valor de *score* em kcal/mol. Quanto mais negativo o *score* melhor, representando uma boa afinidade do ligante com a proteína. Assim, foram descartados *dockings* com valores de *score* maiores que -6 (seis negativos), considerados irrelevantes (SHITYAKOV; FÖRSTER, 2014) e executadas triplicatas com os valores de *score* iguais ou menores que -6. Em seguida, foram selecionados para análise os dez complexos proteína-ligante com melhores *scores* e, os cinco candidatos a droga (ligantes) com melhor média de *score*.

4.5.4 Análise de *docking*

Para análise dos complexos com melhor *score*, foi utilizado o *PyMOL Molecular Graphics System* versão 2.3.4 (DELANO, 2002), possibilitando gerar imagens. O programa *Discovery Studio Visualizer* (DSV) versão 20.1.0.19295 (BIOVIA, 2017) foi utilizado para visualizar os tipos de interações existentes entre a proteína alvo e o ligante.

5 RESULTADOS E DISCUSSÃO

5.1 Redes de Interação Proteína-Proteína

As proteínas de todas as linhagens de *C. ulcerans* alinharam reciprocamente contra 24.584.628 proteínas do STRING, resultando em 3.044.329 alinhamentos em média por linhagem. Como os procedimentos foram realizados para cada linhagem os valores serão representados em média. Após retirar os alinhamentos com identidade*cobertura menor que 0,5625, a quantidade de alinhamentos reduziu para 167.584 proteínas ortólogas entre *C. ulcerans* e STRING.

Após eliminar as proteínas sem alinhamento recíproco (AR) entre *C. ulcerans* e STRING, restaram 71.444 proteínas com AR, utilizadas para gerar o conjunto de interações teóricas, cuja intersecção com as interações do STRING resultaram na predição de 673.450 interações totais. Após remoção das interações redundantes (identificadas em múltiplos organismos) e das interações com score abaixo de 700 no STRING, resultaram em 22.347 interações preditas por linhagem de *C. ulcerans*, então usadas para gerar as redes de interação proteína-proteína (RIP). Os valores de alinhamentos e interações em cada etapa podem ser visualizados na Tabela 4.

Tabela 4 - Quantidade de alinhamentos e interações para cada linhagem de *C. ulcerans*

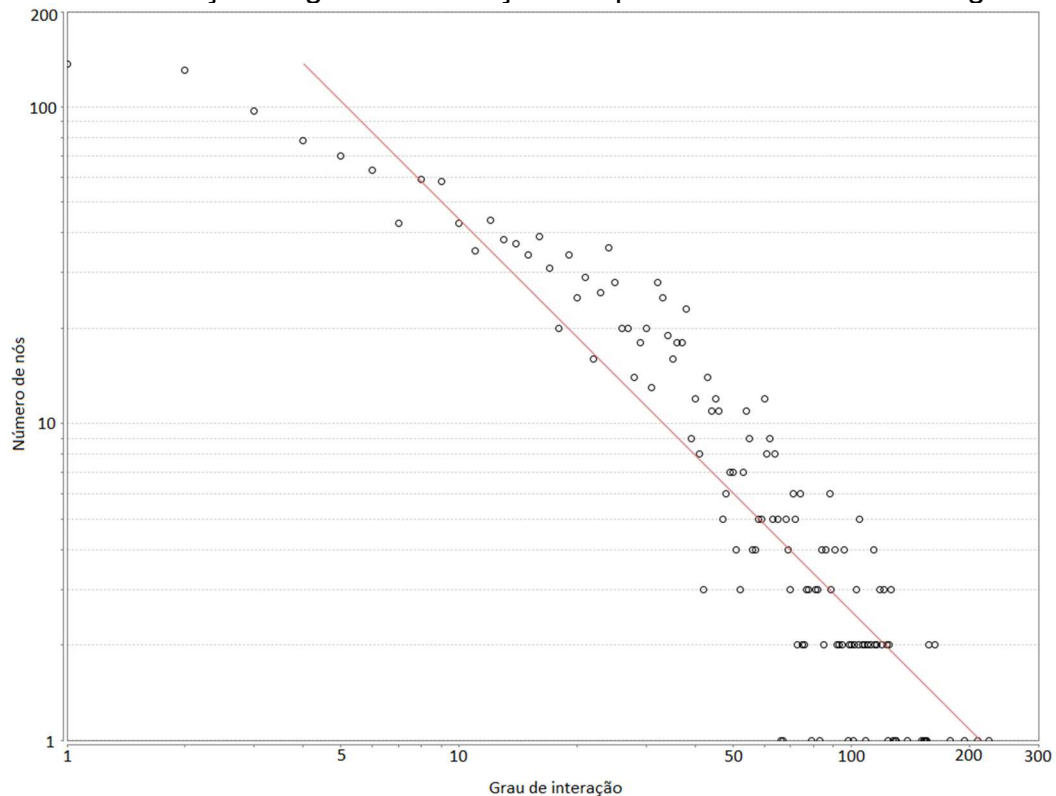
Linhagem	Alinhamentos BLASTp	Alinhamento com Identidade*cobertura ≥ 0.5625	Alinhamentos Recíprocos	Interações totais	Interações distintas	Interações com score ≥ 700
102	2.976.035	158.73	72.066	623.762	29.028	22.483
5146	2.965.008	157.739	71.615	620.932	28.847	22.340
131001	3.223.286	208.004	71.671	896.727	32.866	22.199
131002	3.245.802	205.054	69.426	885.625	36.634	22.117
210931	2.973.931	156.869	71.295	612.754	28.694	22.209
210932	2.882.944	157.422	71.450	617.988	28.591	22.165
809	3.039.516	157.757	71.778	620.220	29.011	22.517
BR-AD22	2.988.283	158.196	71.842	617.031	29.697	22.867
FRC11	2.978.140	157.306	71.340	618.022	28.538	22.116
FRC58	3.170.345	158.755	71.960	621.445	28.940	22.457
Média de alinhamentos ou interações	3.044.329	167.584	71.444	673.450	30.084	22.347

Fonte: Autor.

5.2 Validação da Rede de Interação

As RIPs previstas para cada linhagem de *C. ulcerans* foram submetidas a análise a fim de verificar se a distribuição do grau de interação possui características de redes biológicas. Todos os gráficos gerados foram similares, assim, utilizamos a linhagem FRC58 para representar todas as outras linhagens (Gráfico 2).

Gráfico 2 - Distribuição do grau de interação das proteínas da RIP da linhagem FRC58



Fonte: Autor.

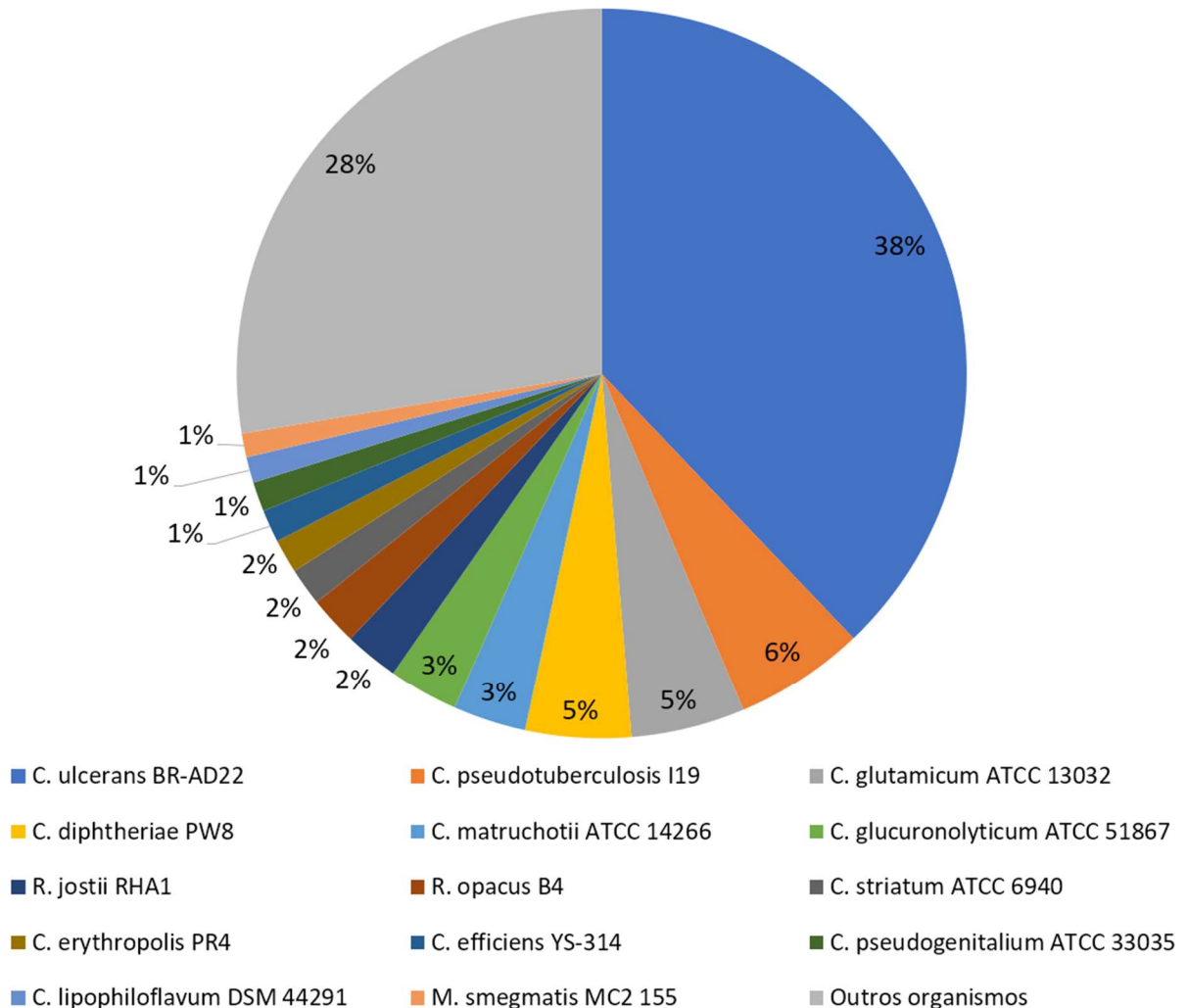
Relação entre o grau de interação (*degree*) e número de nós (*node* ou proteína). Lei de potência perfeita representada na linha em vermelho.

No gráfico plotado observamos que a distribuição do grau de interação das proteínas da RIP teve uma distribuição livre de escala (não-aleatória) com tendência a lei de potência, ou seja, apresentando muitas proteínas com poucas conexões e poucas proteínas com muitas conexões, tendo a RIP característica de rede biológica.

Considerando que as interações mapeadas em *C. ulcerans* foram provindas de diferentes organismos, é muito provável que IPP de organismos filogeneticamente próximos também ocorram em *C. ulcerans*, assim, para validar a RIP foi plotado o gráfico de distribuição das interações nos organismos de origem. Como previsto, dentro os 5090 organismos de STRING, foram mapeadas as IPP em maior parte de

organismos filogeneticamente próximos, principalmente do gênero *Corynebacterium* (71,51%), aumentando a confiabilidade no método de predição de RIP (Gráfico 3).

Gráfico 3 - Distribuição das interações nos organismos de origem



Fonte: Autor.

Legenda: C. = *Corynebacterium*; R. = *Rhodococcus*; M. = *Mycolicibacterium*.

5.3 Proteínas *hub*

De um total de 2.218 proteínas da linhagem FRC58, 1.939 proteínas possuem um grau de interação maior que zero na rede, ou seja, interagem com pelo menos uma outra proteína. Quando organizadas em ordem decrescente pelo grau de interação, 457 fazem parte dos 25% de proteínas com maior grau de interação e estão presentes nas dez linhagens de *C. ulcerans*, sendo então classificadas como proteínas *hub*. As proteínas *hub*, devido à grande quantidade de interações que fazem

em um organismo, tendem a ser essenciais, sendo consideradas bons alvos para fármacos.

Visando identificar dentre as proteínas *hub* aquelas com essencialidade determinada experimentalmente, as 457 proteínas *hub* foram alinhadas contra todas as proteínas da base de dados DEG. Como resultado, 421 proteínas *hub* (92,12%) alinharam e se apresentaram ortólogas às proteínas dos organismos do DEG, demonstrando que o grau de interação das proteínas tem relação direta com sua essencialidade, sendo as RIPs um método válido para identificar proteínas essenciais e conseqüentemente bons alvos para drogas.

Pelo fato de participarem de muitas interações, as 36 proteínas *hub* (7,88%) não homólogas no DEG tem o potencial de serem essenciais para *C. ulcerans*, sendo necessário passarem pela análise funcional antes de serem consideradas novos alvos para drogas. É importante notar também que a metodologia utilizada no trabalho para seleção de proteínas *hub* não foi conservadora ao selecionar 25% das proteínas com maior grau de interação, certamente com um percentual menor a quantidade de proteínas *hub* homólogas na base de dados DEG seria maior.

5.3.1 Análise funcional (COG)

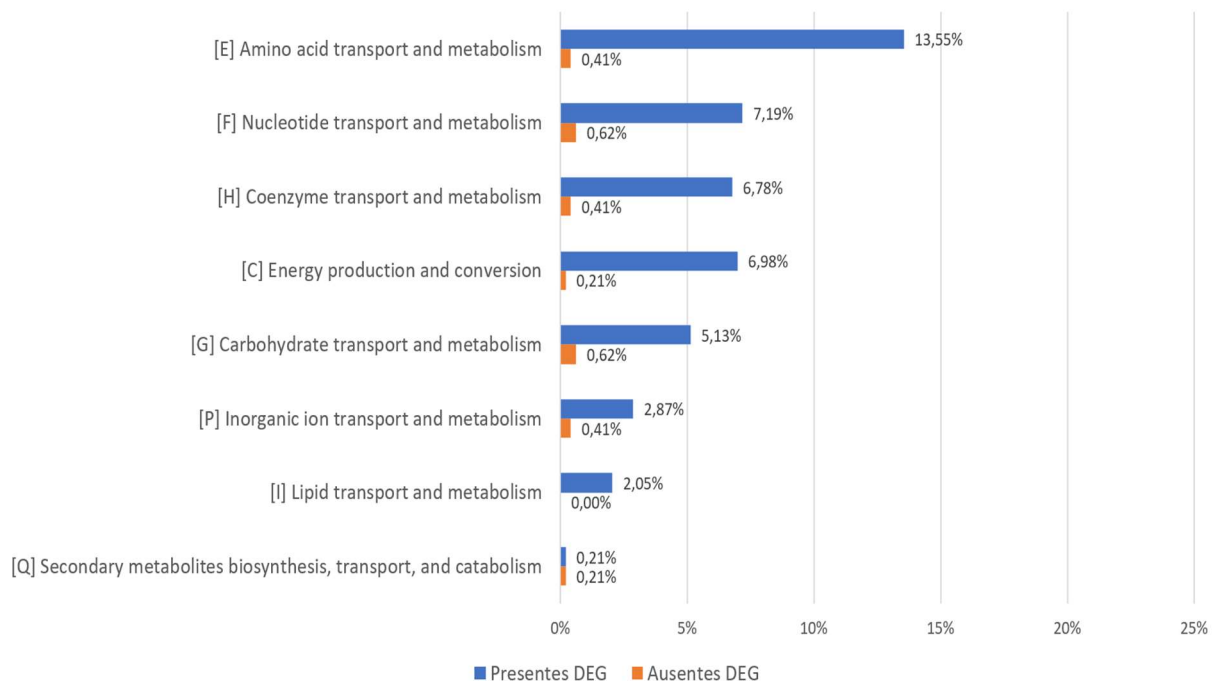
As 457 proteínas *hub* foram submetidas a análise funcional na ferramenta EggNog-Mapper, classificando-as em quatro grupos do COG: metabolismo; processamento e armazenamento de informações; processos e sinalização celular; e proteínas fracamente categorizadas. Cada grupo também possui outras categorias que melhor especificam as funções das proteínas.

Metabolismo: este grupo foi composto por 47,64% de todas as proteínas *hub*, com 93,93% delas presentes no DEG (Gráfico 4). A categoria mais representativa dentro deste grupo foi o metabolismo e transporte de aminoácidos (E) composta majoritariamente por proteínas transferase, sintase e sintetase e com 97% das proteínas presentes no DEG. Os 3% de proteínas que não estavam presentes no DEG dentro da categoria E foram representados por: *chorismate mutase*, enzima que catalisa a conversão do corismato em pterfenato, etapa crucial na via de produção de fenilalanina e tirosina (SASSO *et al.*, 2009), aminoácidos essenciais na síntese de proteínas e precursores de metabólitos secundários (PÉREZ *et al.*, 2015); e

oligopeptide-binding protein (OppA), proteína extracelular essencial para o transporte de oligopeptídeos de 2 a 18 aminoácidos, substratos para a célula e na percepção do meio externo (MONNET, 2003). Importante notar que o sistema de transporte de oligopeptídeos possui outras subunidades além da OppA, sendo as proteínas OppC e OppD *hubs* e presentes no DEG, indicando a OppA como um novo alvo.

A segunda categoria mais representativa foi o metabolismo e transporte de nucleotídeos (F) com 92% das proteínas presente no DEG, sendo composta majoritariamente por proteínas desidrogenases, transferases e quinases. Ficaram ausentes do DEG dentro da categoria F: *orotidine 5'-phosphate decarboxylase*, enzima essencial para biosíntese *de novo* das pirimidinas CTP, TTP e UTP (HARRIS *et al.*, 2000) e *bifunctional protein PyrR*, que regula a transcrição das pirimidinas, nucleotídeos essenciais para todas as formas de vida; e *7,8-dihydro-8-oxoguanine-triphosphatase*, enzima que previne a disfunção e morte celular induzida pelo estresse oxidativo (CONSORTIUM, U., 2019), essencial para a sobrevivência do organismo em casos de estresse oxidativo.

Gráfico 4 - Análise funcional das proteínas *hub* envolvidas com o metabolismo

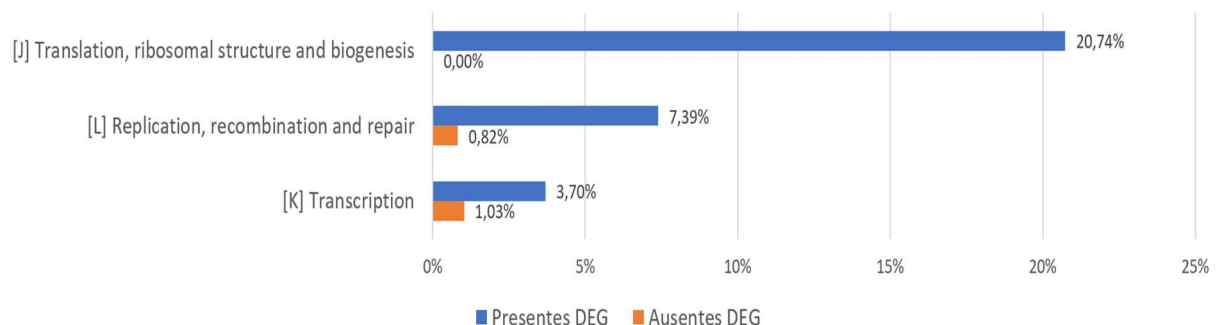


Fonte: Autor.

Processamento e armazenamento de informações: grupo que compõe a segunda maior parte das proteínas *hub* (33,68%) estando 94,5% delas presentes no DEG. O

processamento e armazenamento de informações é representado pelo processo de tradução, biogênese e estrutura ribossomal (J) com 100% das proteínas presentes no DEG, pela replicação, recombinação e reparo (L) com 90,01% presente no DEG e pela transcrição (K) com 78,22% presente no DEG, todos processos considerados essenciais à vida de qualquer organismo. Proteínas ribossomais, ligases e transferases foram as que mais se fizeram presentes na categoria J, todas com ortólogas no DEG. A segunda maior categoria (L) é representada majoritariamente por proteínas helicase, polimerase e de reparo, estando ausentes no DEG: (i) duas *DNA helicases*, que ao separar fitas duplas de DNA permitem que processos celulares como reparo, transcrição e replicação do DNA sejam realizados, necessários para manutenção da vida em qualquer organismo; (ii) uma *endonuclease*, enzima que cliva ligações fosfodiéster nas extremidades de fitas simples de DNA (CONSORTIUM, U., 2019); e (iii) uma *DNA repair protein*, proteína responsável pelo reparo de DNA, essencial em casos de danos ao DNA, principalmente quando este dano impede o acesso à informações essenciais para o bom funcionamento da célula. Enquanto na categoria K, representada por proteínas polimerase, quinase e de transcrição, estiveram ausentes no DEG: (i) duas proteínas responsáveis por regular a transcrição; (ii) duas *phage shock proteins A*, proteínas com papel importante de competir pela sobrevivência do organismo/da célula em condições de nutrientes ou energia limitada (BRISSETTE *et al.*, 1991); e (iii) *hypothetical protein*, proteína predita cuja função desempenhada não foi totalmente elucidada (Gráfico 5).

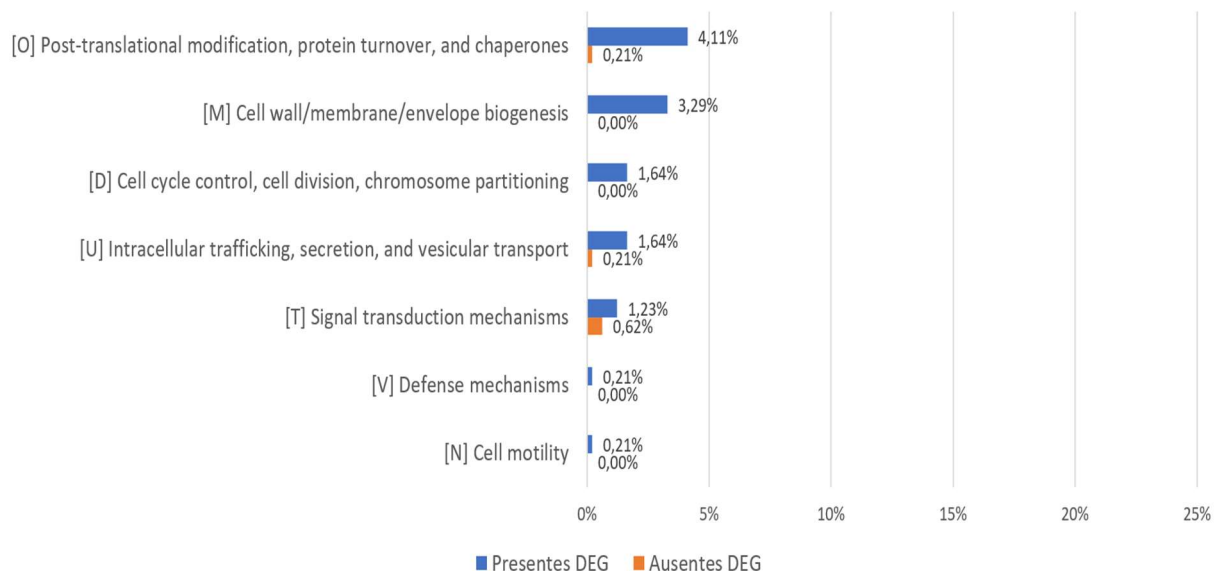
Gráfico 5 - Análise funcional das proteínas *hub* envolvidas com o processamento e armazenamento de informações



Fonte: Autor.

Processos e sinalização celular: representa um dos menores grupos, contendo 13,37% das proteínas *hub* com função como chaperonas, turnover proteico e processos de modificação pós-traducional (O), estando 95,35% presente DEG, composta majoritariamente por chaperonas e tioredoxinas. Chaperonas, por terem um papel importante no enovelamento e desenovelamento de proteínas e na desmontagem de macromoléculas foram todas validadas como essenciais. Presentes no DEG também estavam todas as tioredoxinas (TRX), pequenas proteínas que doam elétrons para espécies reativas de oxigênio (mecanismos de defesa por estresse oxidativo) e participam de reações redox, conservadas em quase todos os organismos, comprovando sua essencialidade (MICHELET; ZAFFAGNINI; LEMAIRE, 2009). A única proteína não presente no DEG na categoria O foi a *Glutaredoxin domain protein*, também envolvida com reações redox, cuja função ainda não foi bem elucidada (KALITA *et al.*, 2018). A segunda (M) e terceira (D) categoria mais representativas tiveram todas suas proteínas validadas pelo DEG. As únicas proteínas dentro das categorias de mecanismos de defesa (V) e motilidade celular (N) foram proteínas hipotéticas, ambas validadas pelo DEG (Gráfico 6).

Gráfico 6 - Análise funcional das proteínas *hub* envolvidas com processos e sinalização celular

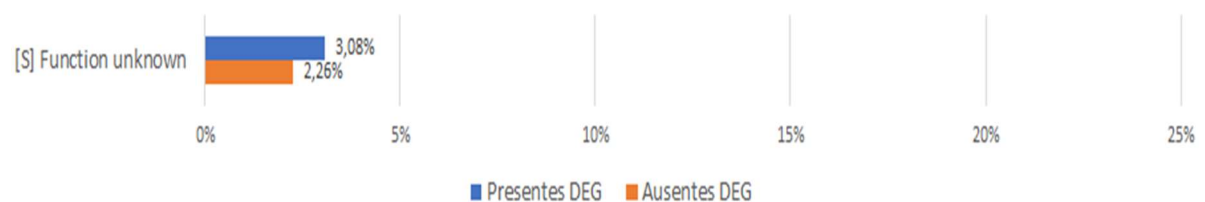


Fonte: Autor.

Fracamente categorizadas: apenas 5,34% das proteínas *hub* possuíam funções desconhecidas (S), destas 57,68% estiveram presentes no DEG. Esta categoria foi composta em sua maioria por proteínas hipotéticas e transferases. Ausentes no DEG

estiveram: *phosphotransferase enzyme family protein*, proteína predita que apenas se sabe que catalisa reações de fosforilação (CONSORTIUM, U., 2019); *RecB family nuclease*, proteína nuclease que faz parte do complexo enzimático RecBCD, responsável pelo reparo e recombinação do DNA (SMITH, 2012), essencial para sobrevivência de diversos organismos em situações específicas, comprovado ser essencial para a bactéria *Pseudomonas syringae* (PAVANKUMAR, SINHA, RAY, 2010); *Septum formation*, proteína envolvida com a síntese da parede celular que se forma entre duas células em divisão, importante para a reprodução bacteriana; e oito proteínas hipotéticas (Gráfico 7).

Gráfico 7 - Análise funcional das proteínas *hub* fracamente categorizadas



Fonte: Autor.

Com exceção das proteínas sem função conhecida (Gráfico 7), a maioria das proteínas *hub* foram categorizadas em grupos e categorias consideradas essenciais, majoritariamente composta por proteínas homólogas no DEG. Nestas categorias, estão presentes muitas proteínas *hub* com funções aparentemente essenciais à *C. ulcerans*. Adicionalmente, revisão na literatura referente às proteínas ausentes no DEG, indicou que possuem funções importantes. Assim, todas as proteínas *hub* analisadas, funcionalmente foram consideradas alvos promissores para fármacos, fato que também será demonstrado na análise de enriquecimento.

5.3.2 Análise de Enriquecimento

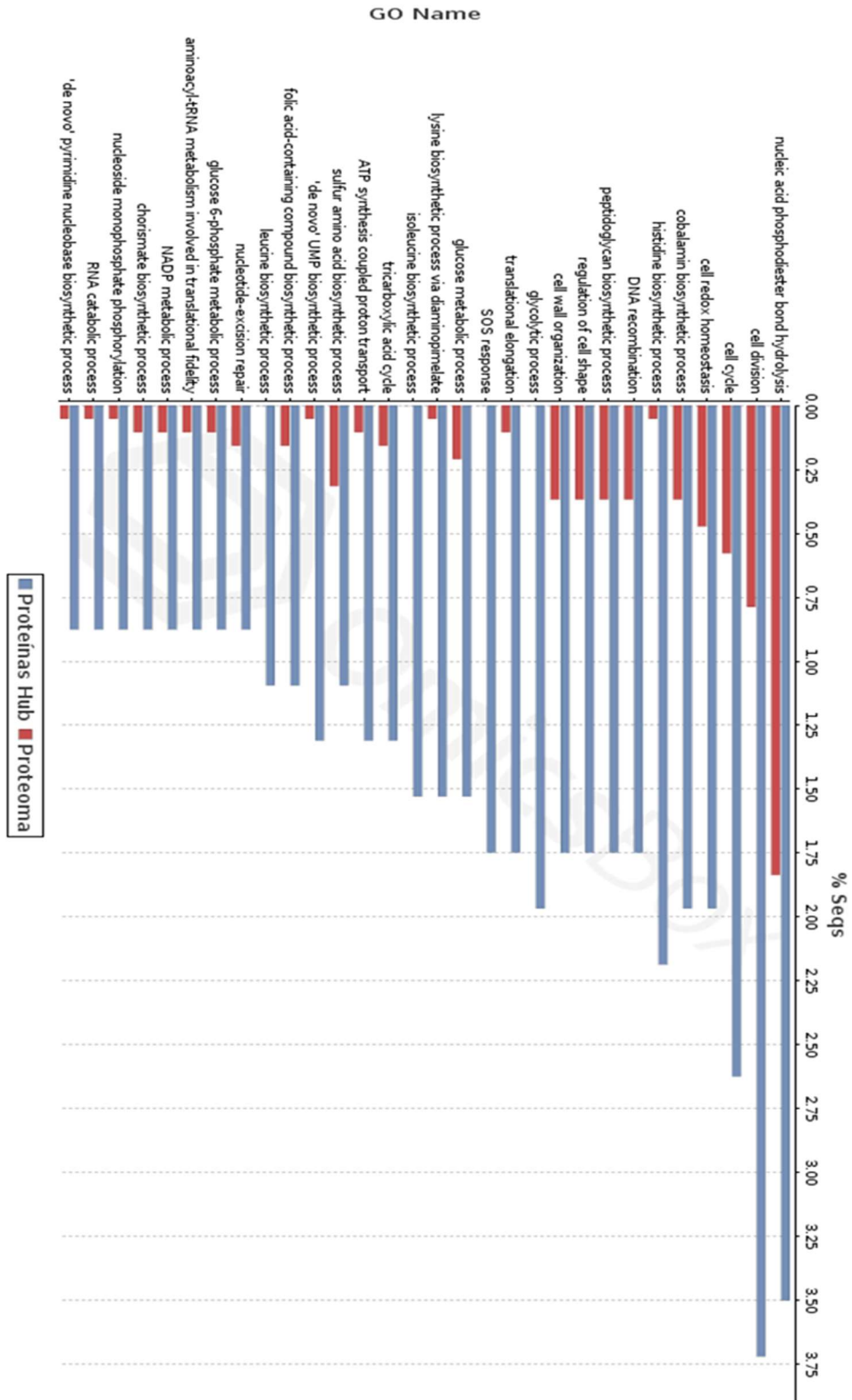
Com o objetivo de melhor caracterizar as proteínas *hub* identificadas no trabalho, também foi realizada a análise de enriquecimento conforme os termos do Gene Ontology (CONSORTIUM, G. O., 2019) para as categorias: processo biológico, função molecular e componente celular. Com o auxílio da ferramenta Blast2GO foram identificados os termos GO estatisticamente mais significativos ($p < 0,05$) nas 457 proteínas *hub* quando comparado com todo o proteoma de *Corynebacterium ulcerans* (2.218 proteínas).

Em se tratando de processos biológicos, alguns são mais importantes que outros para a sobrevivência de um organismo, devendo refletir em maior quantidade de proteínas *hub* para os termos GO importantes. Processos biológicos essenciais (Gráfico 8) para *C. ulcerans* estão mais representados em proteínas *hub* que nas demais proteínas do organismo, como: (i) ciclo e divisão celular, necessários para reprodução bacteriana; (ii) organização da parede celular; (iii) recombinação celular, para manter a estrutura e forma da parede; (iv) e processos glicolíticos, essencial para o metabolismo de carboidratos.

Por conta das proteínas *hub* interagirem direta ou indiretamente com muitas moléculas além de proteínas, função celular (Gráfico 9) que incluem a ligação (*binding*) foram super representados nas proteínas *hub* quando comparados ao proteoma. Os termos de ligação incluem interações com: (i) nucleotídeos (*DNA*, *tRNA*, *ATP*, *GTP*, *NAD* e *NADP*), para a execução de diversas atividades essenciais para a célula; (ii) *pyridoxal phosphate* e *thiamine pyrophosphate* (cofator da vitamina B6 e derivado da vitamina B1, respectivamente), coenzimas essenciais envolvidas em diversos processos celulares; (iii) íons magnésio e zinco, essenciais para o crescimento bacteriano; (iv) proteínas não enoveladas, para enovelamento ou quebra da proteína; (v) ácidos monocarboxílicos; (vi) e ribossomos. Também bem representadas estão outras funções, como de contribuir com a integridade estrutural de ribossomos, possuírem atividade helicase de RNA, GTPase, oxidoreductase, de alongamento de cadeias polipeptídicas, entre outros.

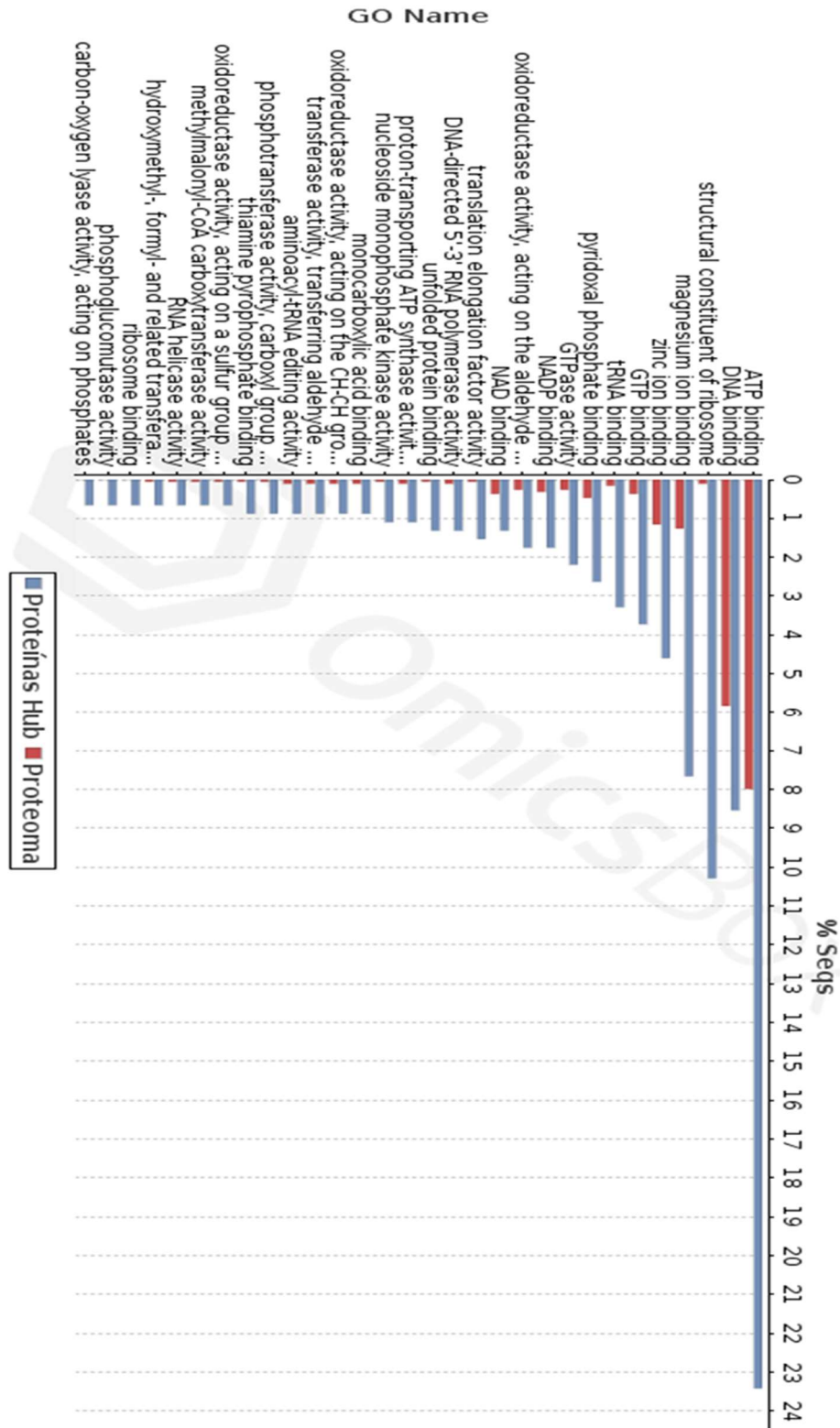
Termos GO relacionados com ribossomos também são bem representados quando se trata de componentes celulares (Gráfico 10) das proteínas *hub*, como é o caso de *large* e *small ribosomal subunit*. Adicionalmente, estão bem representados em componentes celulares proteínas relacionadas com: o transporte de prótons entre o meio externo e interno da célula; o reparo de nucleotídeos do sistema UvrABC; a biosíntese de ácidos graxos, de termo *Acetyl-CoA carboxylase (ACC)*; a formação do DNA a partir de RNA, de termo *ribonucleotide diphosphate reductase (rNDP)*; e proteínas que se associam com cromossomos.

Gráfico 8 - Análise de enriquecimento referente aos processos biológicos



Fonte: Autor
 Processos biológicos mais representativos para as proteínas *hub* ($p < 0,05$). Gráfico gerado pelo BLAST2GO versão acadêmica (GÖTZ *et al.*, 2008).

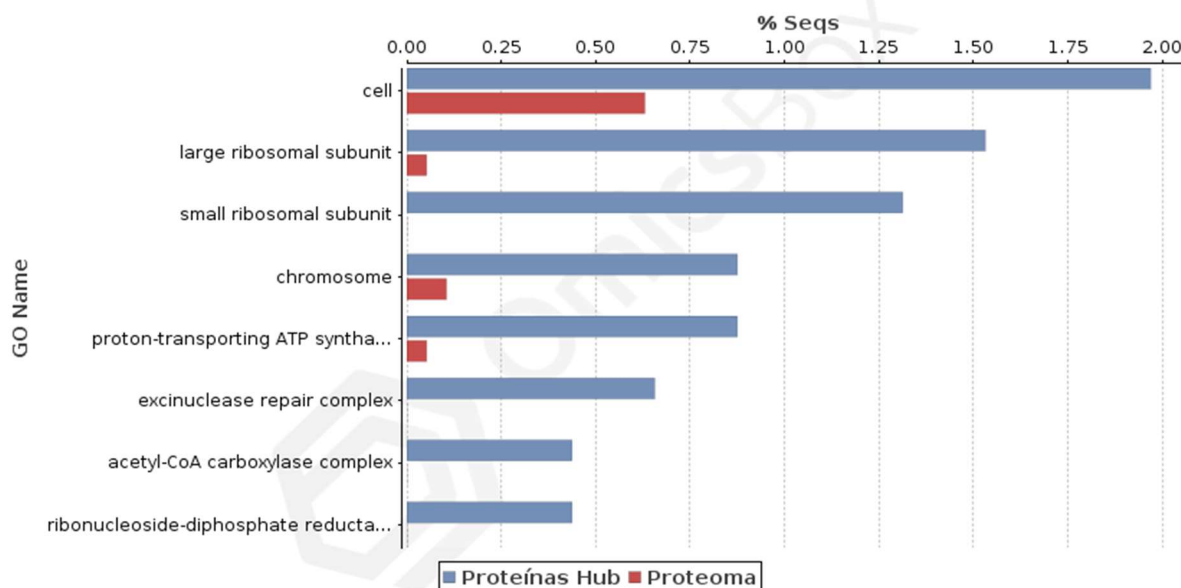
Gráfico 9 - Análise de enriquecimento referente às funções celulares



Fonte: Autor

Funções celulares mais representativas para as proteínas *hub* ($p < 0,05$). Gráfico gerado pelo BLAST2GO versão acadêmica (GÖTZ *et al.*, 2008).

Gráfico 10 - Análise de enriquecimento referente aos componentes celulares
Enriched Bar Chart



Fonte: Autor

Componentes celulares mais representativos para as proteínas *hub* ($p < 0,05$). Gráfico gerado pelo BLAST2GO versão acadêmica (GÖTZ *et al.*, 2008).

De forma geral, os termos GO mais representativos, atribuídos às proteínas dentro de cada categoria, correspondem à essencialidade das proteínas *hub*.

Para as proteínas serem consideradas bons alvos para drogas, além de selecionar, validar e caracterizar as proteínas *hub* como essenciais para *C. ulcerans*, elas não devem possuir homologia com o hospedeiro, neste caso, o humano. Assim, após o alinhamento realizado das proteínas *hub* contra as proteínas de *H. sapiens* do NCBI, 351 não obtiveram homologia com o hospedeiro, podendo todas serem consideradas possíveis alvos farmacológicos e testadas no estudo de caso.

5.4 Estudo de caso

A fim de estudar a afinidade dos 42 derivados sintéticos de alcalóides com as proteínas alvo, visando em trabalhos futuros selecionar as moléculas mais viáveis de serem sintetizadas para testes *in vitro*, o teste de *virtual screening* foi realizado. Antes, porém, as estruturas dos 42 alcalóides foram submetidas ao SwissADME para verificar os parâmetros ADME e se atendem à regra dos cinco, resultando que todos os alcalóides possuem alta absorção pelo trato gastrointestinal e estão de acordo com a RO5, possuindo características de fármacos (Apêndice A).

Para serem obtidos resultados de *docking* mais confiáveis, as estruturas tridimensionais das 351 proteínas não homólogas ao hospedeiro, preditas por homologia na ferramenta Phyre2 (KELLEY *et al.*, 2015), passaram pela validação de Ramachandran. Dentre as 351 proteínas, 119 apresentaram mais de 90% de seus resíduos em regiões favoráveis no gráfico de Ramachandran, apresentando estrutura modelada com qualidade estereoquímica. Dentre estas, 99 proteínas *hub* apresentaram 145 *pockets* com *druggability score* superior a 0,5. Assim, o *docking* das 42 moléculas derivadas de alcalóides foi realizado com cada uma dessas 99 proteínas. É importante destacar que as proteínas *hub* removidas por não apresentarem bons *pockets* continuam sendo potencialmente essenciais à *C. ulcerans*, só não foram exploradas neste estudo de caso.

Após os ligantes e proteínas alvo passarem pela etapa de preparação, o *virtual screening* foi executado, gerando inúmeras conformações para o ligante, cada uma com seu *score* individual. No total foram realizados 16.762 *dockings*, ou 6.090 (145*42), se desconsiderar as triplicatas.

Como resultado do VS, foram analisados os dez complexos proteína-ligante mais promissores, aqueles com melhor *score* e, adicionalmente, os cinco ligantes com melhor média de *score* em todas as proteínas. Valores de melhor *score*, melhor média de *score* de todas as moléculas, assim como os *pockets* com os quais possuíram afinidade podem ser visualizadas no Apêndice B. Informações gerais sobre as 457 proteínas *hub* podem ser visualizadas no Apêndice C, como seu grau de interação, categoria do COG, se tiveram homólogas no DEG, quais delas foram para o *docking*, entre outros dados.

5.4.1 Análise dos melhores complexos proteína-ligante

Com o objetivo de analisar somente os complexos proteína-ligante com melhor resultado de *docking*, selecionamos dez ligantes que obtiveram os melhores *scores* (Tabela 5) ancorados nas respectivas proteínas (Tabela 6). Para visualizar as cavidades das proteínas complexadas com as moléculas foi usado o programa PyMOL e, para as interações entre os ligantes e as proteínas o programa Discovery Studio Visualizer.

Tabela 5 - Ligantes com os melhores resultados de *docking*

Ligante	Melhores Scores	Fórmula química	Núm. de Aceptores de H	Núm. de Doadores de H	Violações RO5	Absorção TGI	BHE permeável
Molécula 23	-9.9	C26H29NO6	7	2	0	Alta	Não
Molécula 27	-9.9 -9.3	C25H26N2O6	7	1	0	Alta	Não
Molécula 24	-9.4	C27H31NO7	8	2	0	Alta	Não
Molécula 26	-9.3	C25H27NO5	6	2	0	Alta	Sim
Molécula 41	-9.6	C23H20N2O4	5	0	0	Alta	Sim
Molécula 34	-9.5	C18H18N2O3	4	0	0	Alta	Sim
Molécula 28	-9.4	C25H27NO5	6	2	0	Alta	Sim
Molécula 20	-9.4	C20H21N3O3S	4	0	0	Alta	Não
Molécula 38	-9.3	C24H23NO4	5	1	0	Alta	Sim

Fonte: Adaptado do SWISSADME.

Importante notar que o número de aceptores e doadores considerados pelo SWISSADME é diferente dos considerados por Lipinski, pois a contagem realizada pelo SWISSADME é baseada em regras mais elaboradas, como o flúor alifático que é considerado acceptor e nitrogênios de anilina que não são aceptores tampouco doadores.

Legenda: TGI (Trato Gastrointestinal); BHE (Barreira hematoencefálica).

Tabela 6 - Proteínas *hub* de *C. ulcerans* com os melhores resultados de *docking*

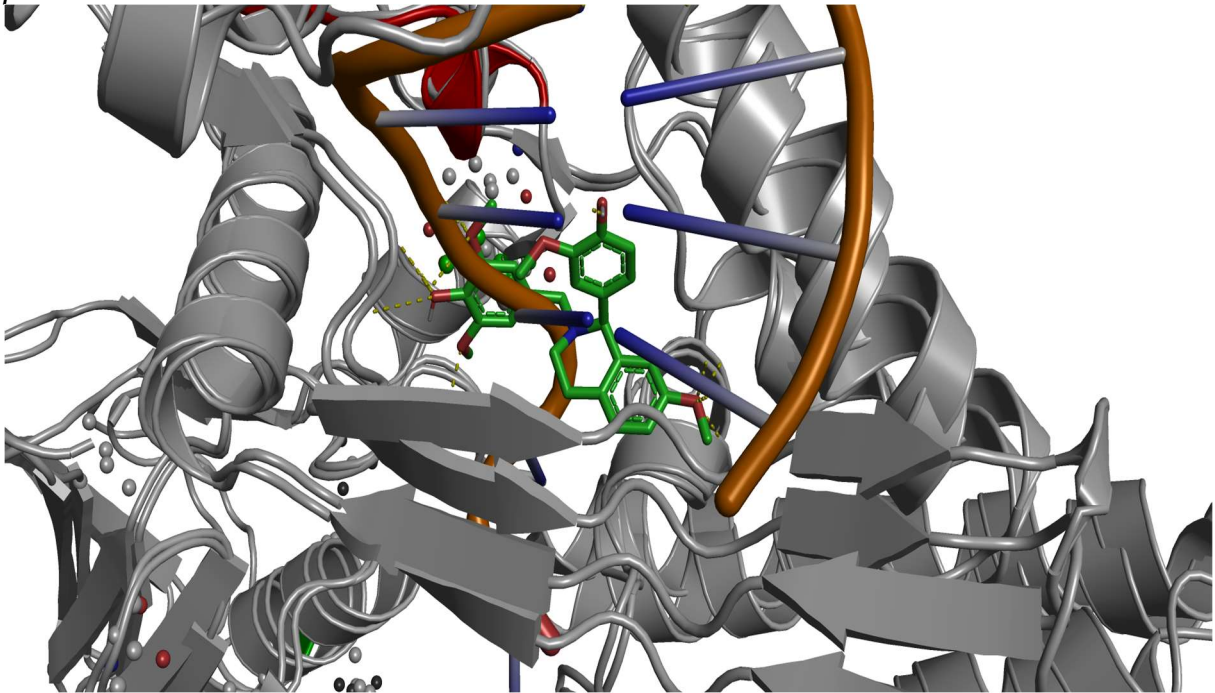
Código PDB	Proteína	Identidade *cobertura	Ramachandran	Categoria COG	DEG	Degree	Ligantes
AKN76931	<i>UvrABC system protein B</i>	3,5%	93,50%	L	4	54	Molécula 23 Molécula 27 Molécula 24 Molécula 26
AKN77488	<i>Bifunctional RNase H/acid phosphatase</i>	0,0%	90,40%	GL	2	46	Molécula 41
AKN77659	<i>Bifunctional protein</i>	18,8%	91,00%	H	25	46	Molécula 34
AKN78203	<i>MviN-like protein</i>	2,8%	91,70%	S	4	44	Molécula 28
AKN77763	<i>UDP-N-acetylglucosamine 1-carboxyvinyltransferase</i>	0,0%	91,50%	M	31	54	Molécula 20 Molécula 27
AKN77374	<i>Histidinol dehydrogenase (HDH)</i>	2,3%	92,40%	E	7	57	Molécula 38

Fonte: Autor.

Legenda: Identidade*cobertura, % do alinhamento da proteína de *C. ulcerans* contra *H. sapiens*; Ramachandran, quantidade de resíduos nas regiões favoráveis; DEG, Quantidade de organismos com proteínas homólogas no DEG; Degree, grau de interação da proteína.

Proteína *UvrABC system protein B*: UvrB faz parte do sistema de reparo UvrABC, que em conjunto com UvrA procura e reconhece danos no DNA. Por ser uma proteína de reparo, foi lhe dada a categoria L do COG. Quanto à homologia com hospedeiro, apenas 3,5% da sequência proteica possui identidade*cobertura com o proteoma de *H. sapiens*. A essencialidade da proteína foi validada no DEG, tendo homólogas em quatro diferentes organismos. Esse sistema de reconhecimento pode ser interrompido por ligantes que venham a interagir fortemente com a proteína em *pockets* envolvidos no processo de reconhecimento. Exemplo desta interrupção pode ser visualizado na Figura 17 com a molécula 23 ancorada. Além da molécula 23 (Figura 18), as moléculas 27 (Figura 19), 24 (Figura 20) e 26 (Figura 21) também obtiveram bons scores, todas ancorando no mesmo *pocket*, com diferentes poses e interações com a proteína.

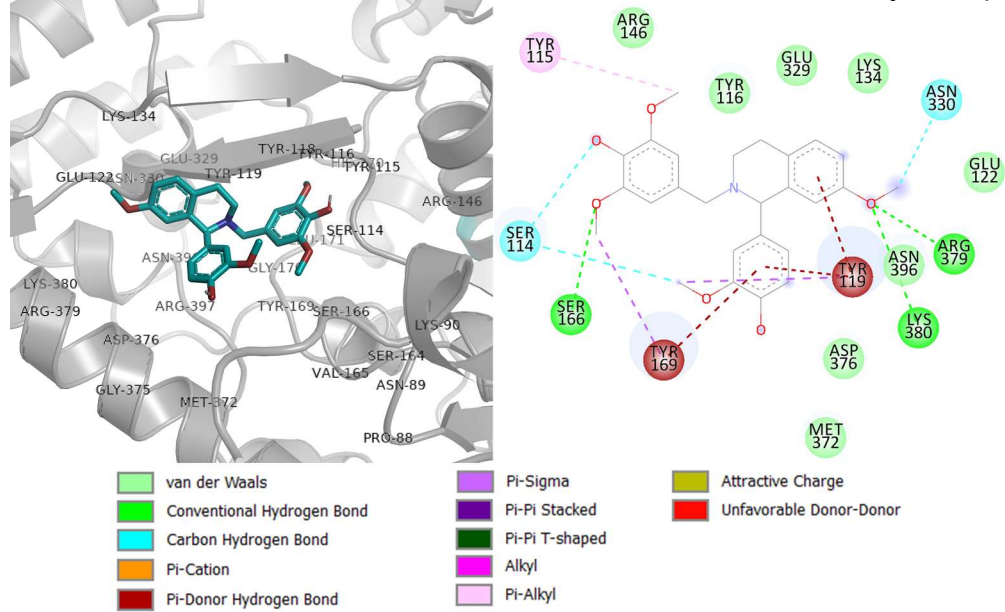
Figura 17 - *Docking* da molécula 23 competindo com o DNA na *UvrABC system protein B*



Fonte: Autor.

Em laranja e azul está o DNA que viria a ser reconhecido pela *UvrABC system protein B* e em verde a molécula 23.

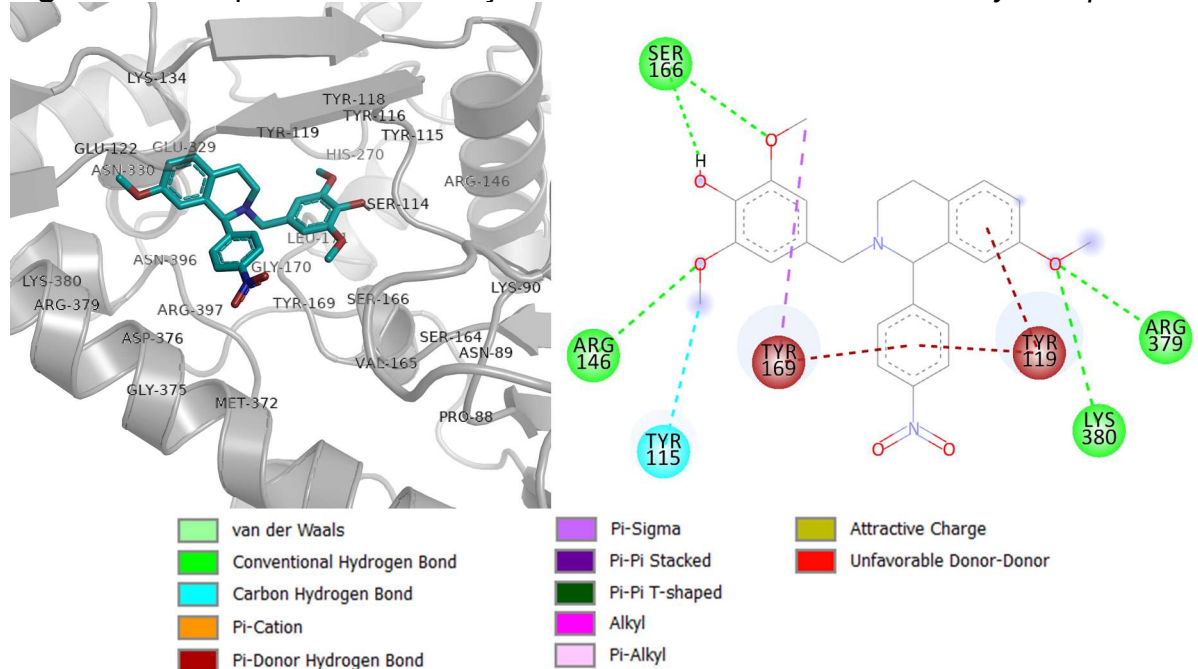
Figura 18 - Acoplamento e interações da molécula 23 com UvrABC system protein B



Fonte: Autor.

a. Visualização da molécula 23 (em ciano) acoplada na proteína *UvrABC system protein B*; **b.** Interações do DSV: Van der Waals (ARG146, ASP376, ASP396, GLU122, GLU32, LYS134, MET372 e TYR116); Hidrogênio (ARG379, LYS380 e SER 166); Hidrogênio Carbono (ASN330, SER114 e TYR 169); Hidrogênio Pi-doador (2xTYR119 e TYR169); Pi-Sigma (TYR119, TYR169); e Pi-Alquila (TYR115).

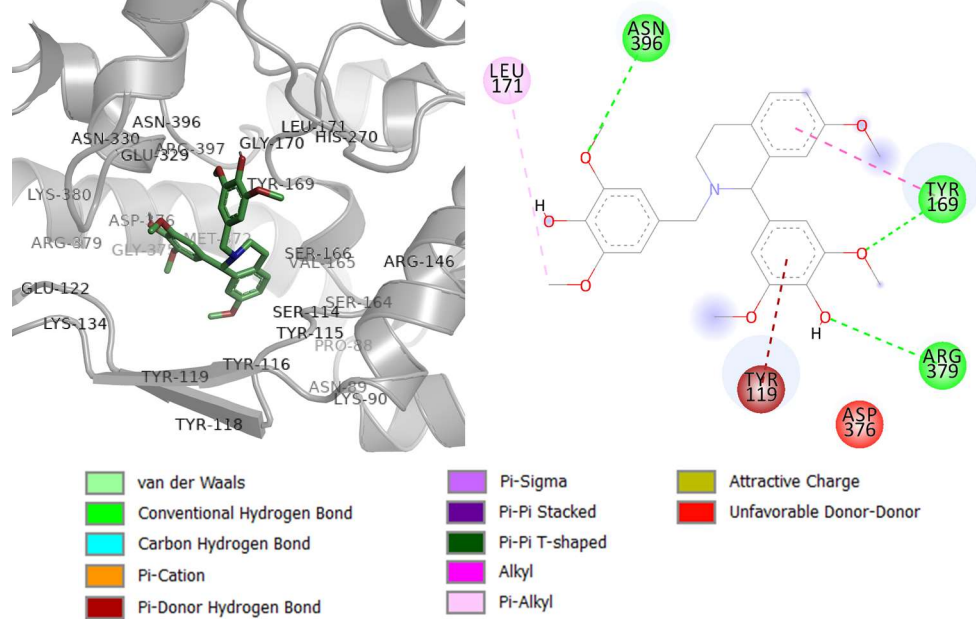
Figura 19 - Acoplamento e interações da molécula 27 com UvrABC system protein B



Fonte: Autor.

a. Molécula 27 em ciano acoplada na proteína; **b.** Interações do DSV: Van der Waals (SER114); Hidrogênio (ARG146, ARG379, LYS380 e SER166); Hidrogênio Pi-doador (2xTYR119 e TYR169); Hidrogênio Carbono (TYR115); Pi-Sigma (TYR169); e Pi-Alquila (TYR115).

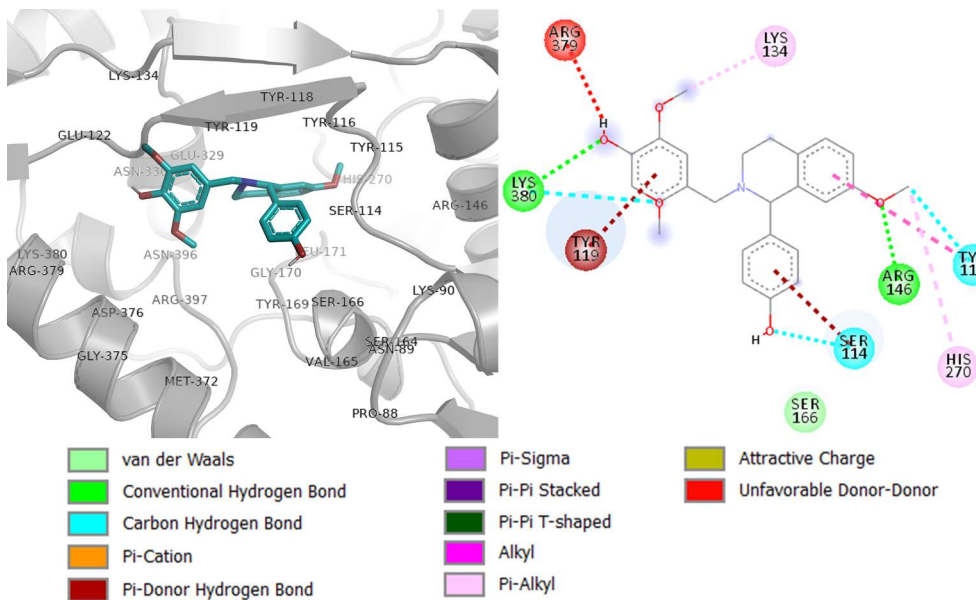
Figura 20 - Acoplamento e interações da molécula 24 com *UvrABC system protein B*



Fonte: Autor.

a. Molécula 24 em verde acoplada na proteína pelo PyMOL; **b.** Interações do DSV: Van der Waals (ARG146, GLY170, SER114, TYR114 e TYR115); Hidrogênio (2xASN396, ARG379 e TYR169); Hidrogênio Pi-doador (TYR119); Pi-Pi Stacked (TYR169); Pi-Alquila (LEU171); e não favorável Aceptor-Aceptor (ASP376).

Figura 21 - Acoplamento e interações entre a molécula 26 e *UvrABC system protein B*

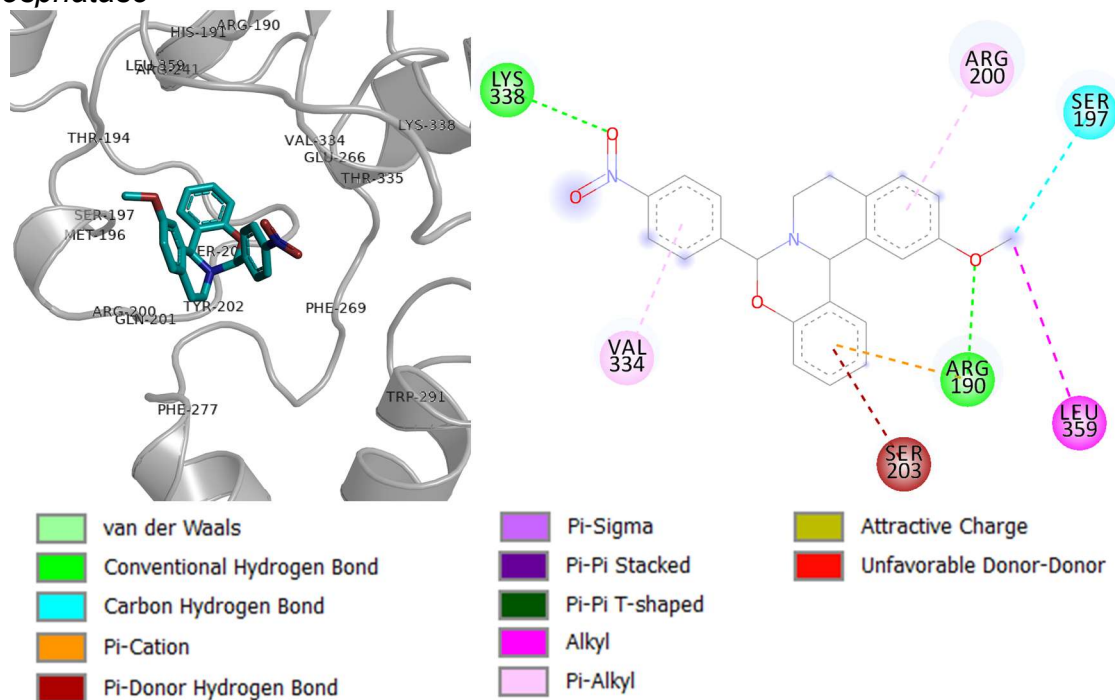


Fonte: Autor.

a. Molécula 26 em ciano acoplada na proteína pelo PyMOL; **b.** Interações do DSV: Van der Waals (SER166); Hidrogênio (ARG146 e LYS380); Hidrogênio Carbono (LYS380 e SER114); Hidrogênio Pi-Doador (SER114 e TYR119); Pi-Pi Stacked (TYR115); Alquilas (LYS134); Pi-Alquila (HIS270 e TYR115); Não favorável Aceptor-Aceptor (ARG379).

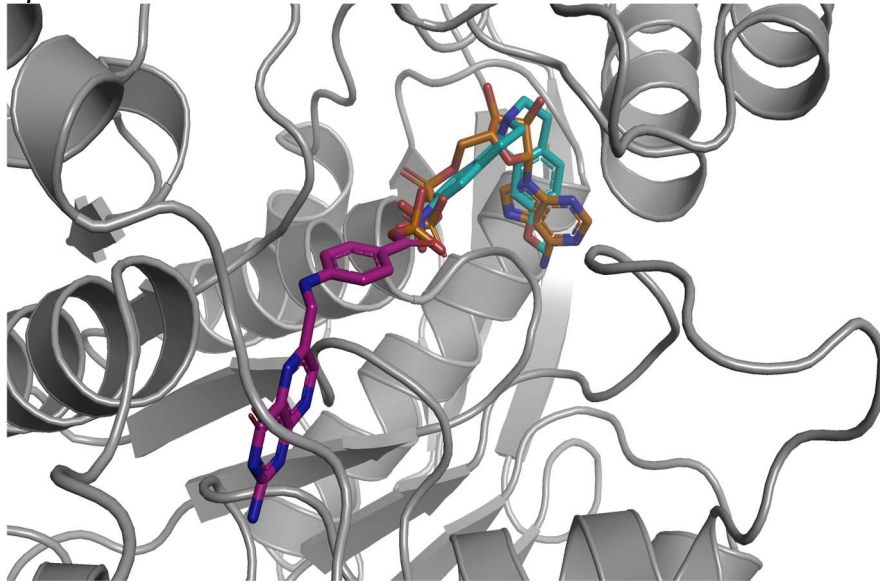
Bifunctional RNase H/acid phosphatase: proteína com atividade ribonuclease, catalisando a quebra do RNA em moléculas menores. A molécula 41 apresentou melhor *score* de *docking* com a RNase (Figura 22), porém, não estão disponíveis no PDB estruturas terciárias similares da proteína desempenhando seu papel de RNase, logo, não foi possível identificar, representar e comparar o sítio de interação do RNA com o sítio de interação da molécula 41.

Figura 22 - Acoplamento e interações da molécula 41 com *Bifunctional RNase H/acid phosphatase*



Interações com *Bifunctional protein*: proteína anotada no InterPro como *tetrahydrofolylpolyglutamate synthase activity*, considerada essencial por atuar na síntese de dihidrofolato e folilpoliglutamato, desempenhando ambas as atividades sintases no mesmo sítio de ligação (SHENG *et al.*, 2008). A molécula 34 obteve melhor *score* de *docking* (-9,5) com a *Bifunctional protein*, competindo pelo mesmo sítio que o ADP (Adenosina difosfato), molécula nativa cristalizada junto com a proteína (Figura 23). O acoplamento e interações podem ser visualizadas na Figura 24.

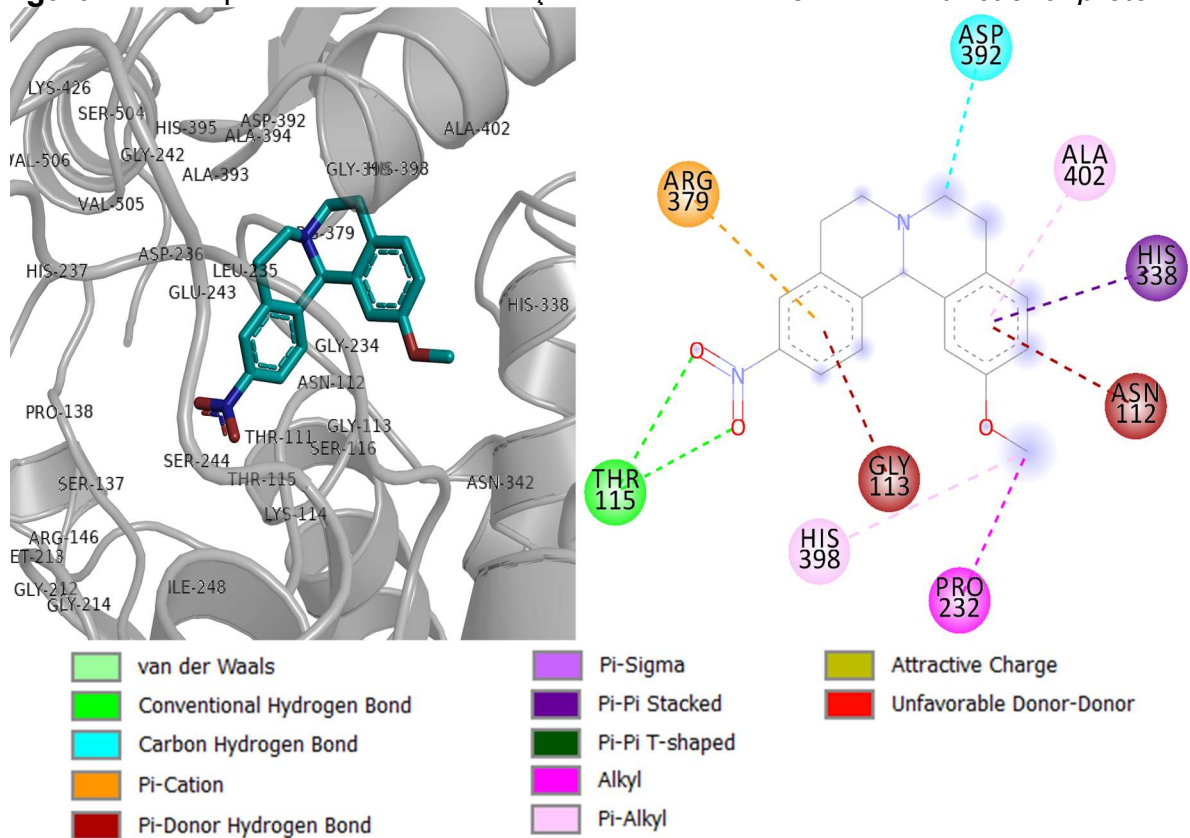
Figura 23 - Docking da molécula 34 competindo pelo sítio de ligação com o ADP na *Bifunctional protein*



Fonte: Autor.

Em azul está representada a molécula 34, em laranja o ADP e em roxo o dihidropteroato.

Figura 24 - Acoplamento e das interações da molécula 34 com *Bifunctional protein*

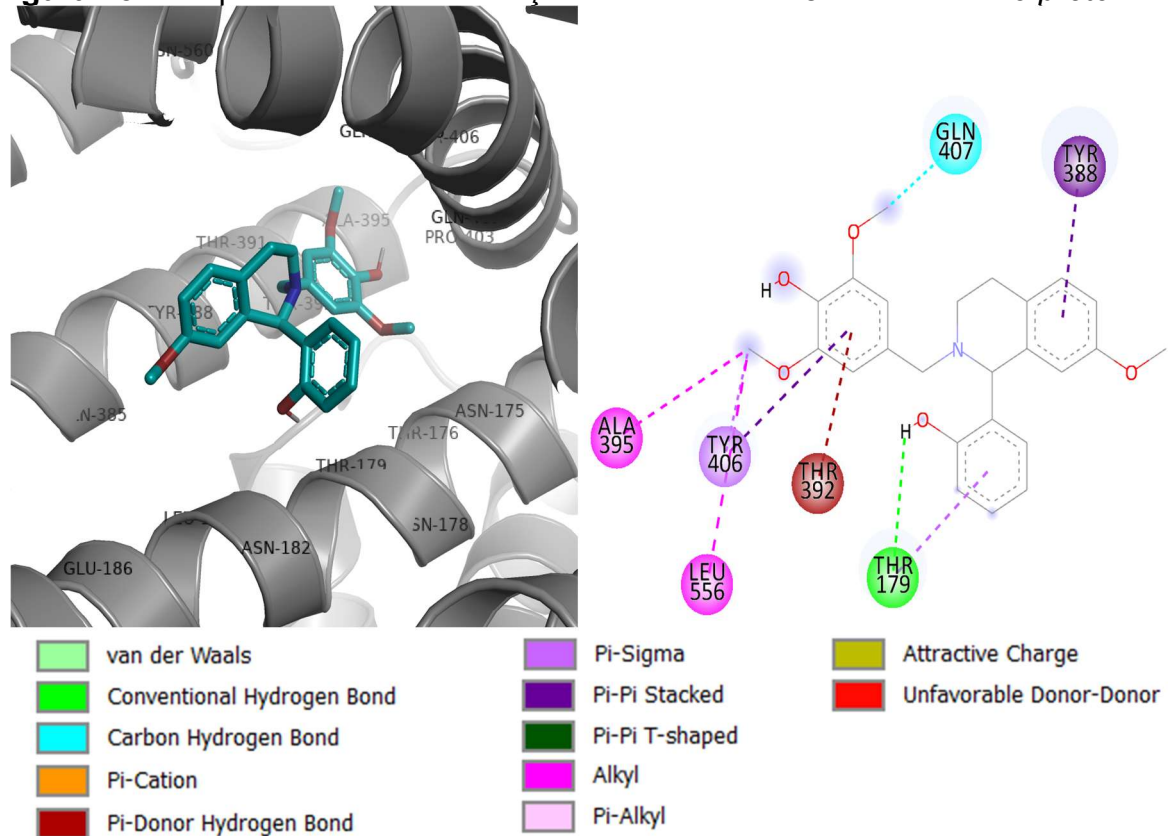


Fonte: Autor.

a. Molécula 34 em ciano acoplada na proteína. b. Interações: Hidrogênio (HIS266); Pi-Sigma (THR241); e Pi-Alquila (HIS373, PRO216, TYR367 e VAL374).

Interações com *MviN-like protein*: Proteína sem função elucidada, sendo componente integral da membrana celular (CONSORTIUM, G. O., 2019). A molécula 28 apresentou o melhor *score* de *docking* (-9,4) quando acoplada com a proteína (Figura 25).

Figura 25 - Acoplamento e das interações da molécula 28 com *MviN-like protein*

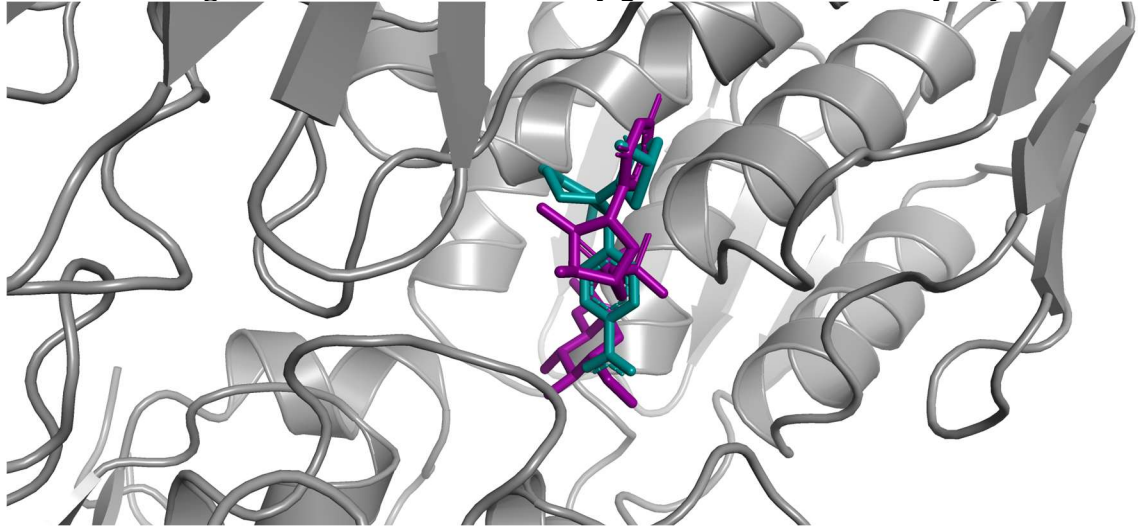


Fonte: Autor.

a. Molécula 28 em ciano acoplada na proteína; **b.** Interações: Van der Waals (LEU183 e TRP410); Hidrogênio (THR179); Hidrogênio Carbono (GL407); Hidrogênio Pi-doador (THR392); Pi-Sigma (TYR406 e TYR179); Pi-Pi Stacked (TYR388 e TYR406); e Alquilas (ALA395 e LEU556).

UDP-N-acetylglucosamine 1-carboxyvinyltransferase: Proteína envolvida com a síntese do peptidoglicano ao adicionar enolpiruvil ao UDP-N-acetylglucosamine. As moléculas 20 e 27 obtiveram os melhores *scores* de *docking* (-9,4 e -9,3 respectivamente), ambas ancorando no mesmo sítio de ligação que a uridina-difosfato-n-acetilglicosanima, uma molécula nativa cristalizada junto com a proteína (Figura 26). O acoplamento e interações das moléculas 20 e 27 podem ser visualizadas na Figura 27 e Figura 28, respectivamente.

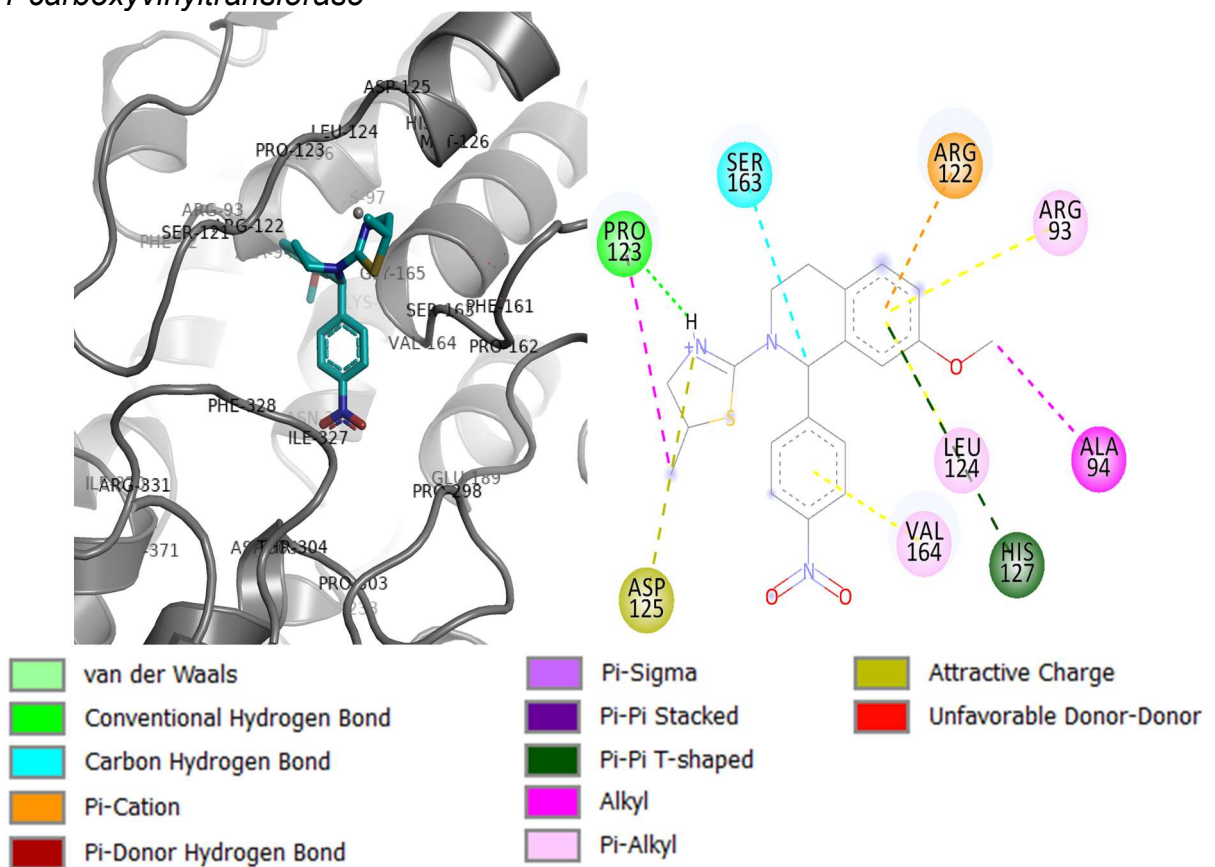
Figura 26 - Docking da molécula 20 competindo pelo sítio de ligação com uridina-difosfato-n-acetilglicosamina na *UDP-N-acetylglucosamine 1-carboxyvinyltransferase*



Fonte: Autor.

Em ciano está representada a molécula 20 e em rosa a uridina-difosfato-n-acetilglicosamina.

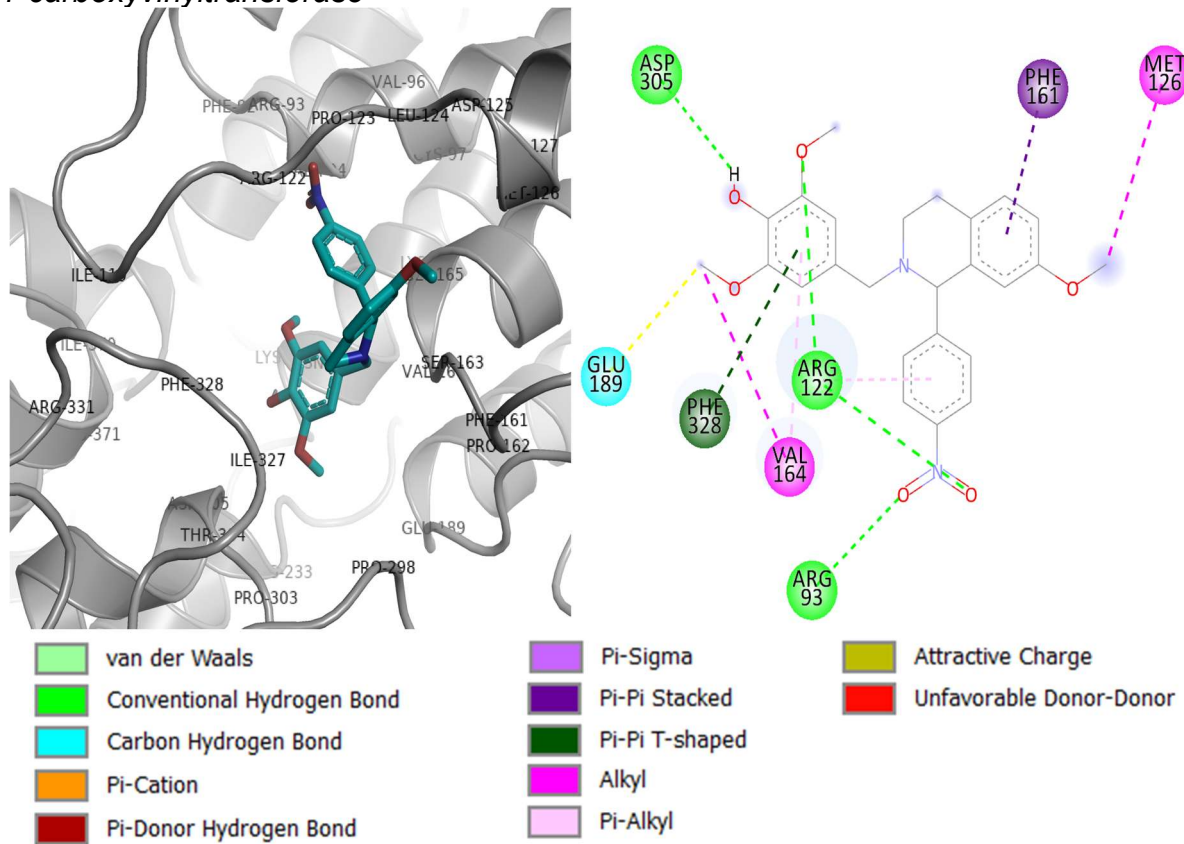
Figura 27 - Acoplamento e interações da molécula 20 com *UDP-N-acetylglucosamine 1-carboxyvinyltransferase*



Fonte: Autor.

a. Molécula 20 em ciano acoplada na proteína; **b.** Interações: Carga Atrativa (ASP125); Hidrogênio (PRO123); Hidrogênio Carbono (SER163); Pi-Cátion (ARG122); Pi-Pi T-shaped (HIS127); Alquilas (ALA94 e PRO123); e Pi-Alquila (ARG93, LEU124 e VAL164).

Figura 28 - Acoplamento e interações da molécula 27 com *UDP-N-acetylglucosamine 1-carboxyvinyltransferase*

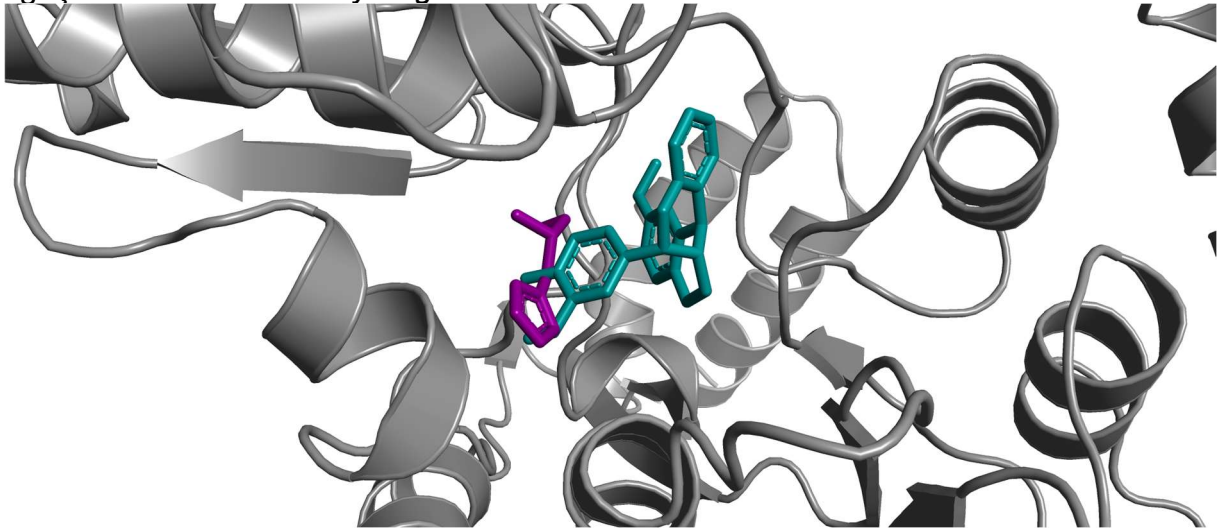


Fonte: Autor.

a. Molécula 27 em ciano acoplada na proteína; **b.** Interações: Hidrogênio (ARG93, 2xARG122 e ASP305); Hidrogênio-Carbono (GLU189); Pi-Pi Stacked (PHE161); Pi-Pi T-shaped (2xPHE328); Alquilas (MET126 e VAL164); e Pi-Alquila (ARG122 e VAL164).

Histidinol dehydrogenase (HDH): Proteína envolvida com a síntese de L-histidina ao catalizar a oxidação de L-histidinol para L-histidinaldeído e depois para L-histidina. Esta via está presente apenas em bactérias, arqueobactérias, alguns eucariotos e plantas, mas ausente em mamíferos, o que explica não haver proteínas homólogas em humanos. Estudos realizados em linhagens mutantes de *Mycobacterium tuberculosis* com *knockout* do gene *hisD* comprovaram a essencialidade da HDH, o que a torna bom alvo para drogas (PARISH, 2003; PARISH *et al.*, 1999; NUNES *et al.*, 2011). Dos alcalóides, a molécula 38 foi a que obteve melhor *score* de *docking* (-9,3) com a HDH (Figura 30) e acabou por competir com o L-histidinol pelo mesmo sítio de ligação na proteína (Figura 29).

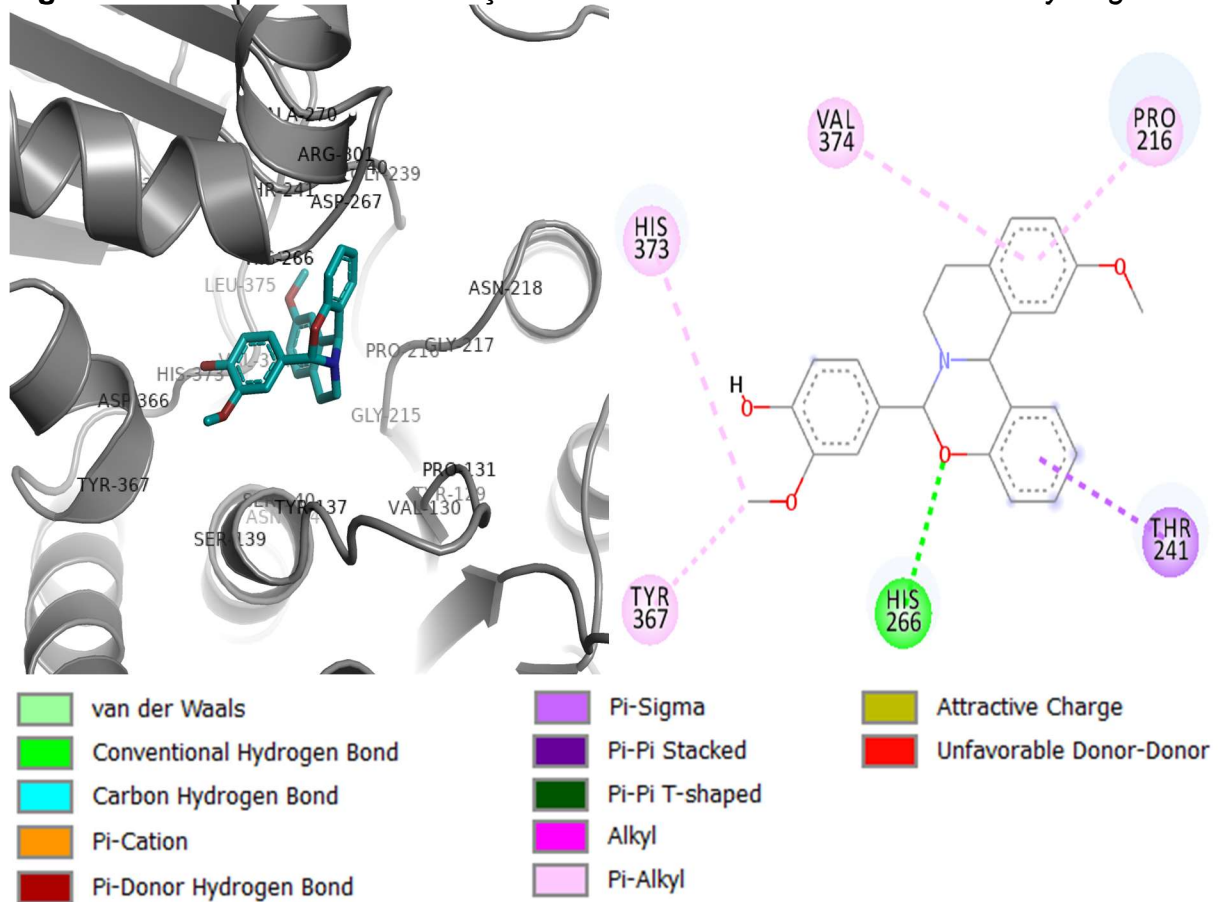
Figura 29 - Docking da molécula 38 competindo com L-histidinol pelo mesmo sítio de ligação na *Histidinol dehydrogenase*



Fonte: Autor.

Em ciano está representada a molécula 38 e em rosa o L-histidinol.

Figura 30 - Acoplamento e interações da molécula 38 com *Histidinol dehydrogenase*



Fonte: Autor.

a. Molécula 38 em ciano acoplada na proteína pelo PyMOL; **b.** Interações do DSV: Hidrogênio (HIS266); Pi-Sigma (THR241); e Pi-Alquila (HIS373, PRO216, VAL374 e TYR367).

Como observado, candidatos a droga não competem com outros ligantes que interagem naturalmente com a proteína, porém, a presença do candidato em sítios ativos ou alostéricos pode conferir alteração estrutural da proteína e impedir que desempenhe sua atividade, ou ainda impedir a interação com outras proteínas ou ligantes, podendo comprometer a sobrevivência do organismo.

5.4.2 Análise dos candidatos a droga com melhor média de *score*

Diferente dos melhores complexos proteína-ligante, os ligantes com melhor média de *score* interagem com uma grande quantidade de proteínas, mantendo ainda bons valores de *score* individual (Tabela 7), e por isso detém maior chance de interagir com pelo menos uma das proteína com que possui afinidade, impedir sua atividade e comprometer a sobrevivência de *C. ulcerans*. Alguns destes ligantes, além de terem obtido melhor média de *score*, também apresentaram melhor *score* individual, como as moléculas 27 (-9,9), 41 (-9,6) e 34 (-9,5).

Dentre os candidatos a droga com melhor média de *score* consideramos como mais promissores: (i) a molécula 27, por ter o melhor *score* de *docking* (-9,9), interagir com uma grande quantidade de *pockets* (71), possuir o maior número de aceptores de hidrogênio (7) e por ser a única não-permeável pela barreira hematoencefálica; (ii) a molécula 34, por interagir com a maior quantidade de *pockets* (82); (iii) ou ainda a molécula 42 por ter a melhor média de *score*.

Tabela 7 - Dados e propriedades dos cinco ligantes com melhores médias de *score*

Ligante	Média de Scores	Melhor Score	<i>Pockets</i>	Núm. de Aceptores de H	Núm. de Doadores de H	Violações RO5	Absorção TGI	BHE permeável
Molécula 42	-7,22	-9,3	68	3	0	0	Alta	Alta
Molécula 34	-7,21	-9,5	82	4	0	0	Alta	Sim
Molécula 27	-7,2	-9,9	71	7	1	0	Alta	Não
Molécula 22	-7,19	-9,1	70	5	1	0	Alta	Sim
Molécula 41	-7,19	-9,6	61	5	0	0	Alta	Sim

Fonte: Autor.

Dados sobre a absorção pelo TGI (trato gastrointestinal), BHE (barreira hematoencefálica) permeável, violações da RO5 e o número de aceptores e doadores de hidrogênio foram fornecidos pelo SwissADME. Para o cálculo da média somente foram consideradas interações com *score* menor que -6.

Por fim, dentre as 42 moléculas testadas no estudo de caso, tanto as moléculas participantes dos melhores complexos quanto as com melhor média de *score* são boas candidatas para otimização por processo de desenho racional de fármaco ou de serem sintetizadas para testes *in vitro* em trabalhos futuros. A otimização das moléculas antes da síntese poderá garantir que o ligante tenha maior afinidade pela proteína alvo e iniba sua atividade, evitando também que o ligante interaja com outras proteínas ou moléculas do hospedeiro.

6 CONCLUSÃO

Geramos pelo método *interolog mapping* a RIP de *Corynebacterium ulcerans* para identificar as proteínas *hubs*, consideradas essenciais. Dentre as proteínas com alto grau de interação (*hub*) 36 não possuíam homólogas no DEG e 421 foram validadas por homologia contra o DEG. As 36 proteínas não-validadas tiveram a essencialidade caracterizada pela análise funcional e análise de enriquecimento, concluindo que mesmo não presentes no DEG, por diversos motivos identificados na literatura, podem ser essenciais para a manutenção da vida de *C. ulcerans*.

Adicionalmente, obtivemos 99 proteínas *hub* de *C. ulcerans* com estruturas tridimensionais preditas por modelagem, validadas pelo gráfico de Ramachandran, não homólogas ao hospedeiro e com *pockets* viáveis, nas quais testamos 42 derivados sintéticos de alcalóides tetrahydroisoquinolínicos, identificando os derivados mais promissores para futuro processo de otimização, síntese e testes *in vitro*.

Tivemos como melhor complexo a molécula 23 ancorada à proteína *UvrABC system protein B*, justificado pela essencialidade e mecanismo de ação da proteína, pela quantidade de interações preditas entre o ligante e a proteína e pela possível competição pelo sítio ativo. Também consideramos como promissores as cinco moléculas candidatas a droga analisadas: 27, 34, 42, 22 e 41.

A síntese de uma única molécula é um processo caro e demorado, assim, o estudo de caso foi importante, evitando gastos com a síntese indiscriminada dos 42 derivados de alcalóides, selecionando os mais promissores a serem deletérios a *C. ulcerans* para testes *in vitro*.

Mesmo analisando apenas os derivados sintéticos de alcalóides tetrahydroisoquinolínicos, conhecendo-se as proteínas *hub* (essenciais) de *C. ulcerans*, qualquer outra droga poderá ser testada *in silico* ou ainda poderão ser direcionados testes *in vitro* com as proteínas como alvos.

O presente trabalho foi inédito ao usar RIP para identificar proteínas *hub* de *C. ulcerans* e as caracterizar como essenciais, assim como testar a afinidade de 42 derivados de alcalóides tetrahydroisoquinolínicos contra as proteínas essenciais de *C. ulcerans*. Adicionalmente, a realização deste trabalho foi extremamente enriquecedora para a minha formação acadêmica e por possivelmente ajudar em um futuro combate à difteria causada pela *Corynebacterium ulcerans*.

REFERÊNCIAS

ABAGYAN, R.; TOTROV, M.; KUZNETSOV, D. ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. **Journal of computational chemistry**, 15, n. 5, p. 488-506, 1994.

ACENCIO, M. L.; LEMKE, N. Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information. **BMC bioinformatics**, 10, n. 1, p. 290, 2009.

ALDERWICK, L. J. *et al.* The mycobacterial cell wall—peptidoglycan and arabinogalactan. **Cold Spring Harbor perspectives in medicine**, 5, n. 8, p. a021113, 2015.

ALM, E.; ARKIN, A. P. Biological networks. **Current opinion in structural biology**, 13, n. 2, p. 193-202, 2003.

ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of molecular biology**, 215, n. 3, p. 403-410, 1990.

ANDREOPOULOS, W.; LABUDDE, D. Protein-protein interaction networks. *In: Protein purification and analysis I: Methods and applications*: iConcept Press, 2013.

ANFINSEN, C. B. Principles that govern the folding of protein chains. **Science**, 181, n. 4096, p. 223-230, 1973.

ARNOLD, K. *et al.* The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. **Bioinformatics**, 22, n. 2, p. 195-201, 2006.

ASSENOV, Y. *et al.* Computing topological parameters of biological networks. **Bioinformatics**, 24, n. 2, p. 282-284, 2008.

ATKINSON, W. *et al.* Mumps. **Epidemiology and prevention of vaccine-preventable diseases. 10th ed.** Washington: National Immunization Program, CDC (US), p. 149-158, 2007.

- BAKER, D.; SALI, A. Protein structure prediction and structural genomics. **Science**, 294, n. 5540, p. 93-96, 2001.
- BARKSDALE, L. *Corynebacterium diphtheriae* and its relatives. **Bacteriological Reviews**, 34, n. 4, p. 378, 1970.
- BENSO, A. *et al.* Modeling biological complexity using Biology System Description Language (BiSDL). 2020.
- BERMAN, H. M. *et al.* The protein data bank. **Nucleic acids research**, 28, n. 1, p. 235-242, 2000.
- BERNER, R. *et al.* Fatal sepsis caused by *Corynebacterium amycolatum* in a premature infant. **Journal of clinical microbiology**, 35, n. 4, p. 1011-1012, 1997.
- BEVERIDGE, T. J. *et al.* V. Functions of S-layers. **FEMS microbiology reviews**, 20, n. 1-2, p. 99-149, 1997.
- BINKOWSKI, T. A.; NAGHIBZADEH, S.; LIANG, J. CASTp: computed atlas of surface topography of proteins. **Nucleic acids research**, 31, n. 13, p. 3352-3355, 2003.
- BIOVIA, D. S. Discovery studio visualizer. **San Diego, CA, USA**, 936, 2017.
- BITRAGUNTA, S. *et al.* Safety and immunogenicity of single dose of tetanus–diphtheria (Td) vaccine among non/partially immune children against diphtheria and/or tetanus, Hyderabad, India, 2007. **Vaccine**, 28, n. 37, p. 5934-5938, 2010.
- BOMPARD-GILLES, C. *et al.* Substrate mimicry in the active center of a mammalian α amylase: structural analysis of an enzyme–inhibitor complex. **Structure**, 4, n. 12, p. 1441-1452, 1996.
- BONMARIN, I. *et al.* Diphtheria: a zoonotic disease in France? **Vaccine**, 27, n. 31, p. 4196-4200, 2009.
- BONVIN, A. M. Flexible protein–protein docking. **Current opinion in structural biology**, 16, n. 2, p. 194-200, 2006.
- BOONE, C.; BUSSEY, H.; ANDREWS, B. J. Exploring genetic interactions and networks with yeast. **Nature Reviews Genetics**, 8, n. 6, p. 437-449, 2007.

BRADLEY, P.; MISURA, K. M.; BAKER, D. Toward high-resolution de novo structure prediction for small proteins. **Science**, 309, n. 5742, p. 1868-1871, 2005.

BRISSETTE, J. L. *et al.* Characterization and sequence of the Escherichia coli stress-induced psp operon. **Journal of molecular biology**, 220, n. 1, p. 35-48, 1991.

BURKOVSKI, A. Cell envelope of corynebacteria: structure and influence on pathogenicity. **ISRN microbiology**, 2013, 2013.

CAMACHO, C. *et al.* BLAST+: architecture and applications. **BMC bioinformatics**, 10, n. 1, p. 421, 2009.

CAVASOTTO, C. N.; ORRY, W.; ANDREW, J. Ligand docking and structure-based virtual screening in drug discovery. **Current topics in medicinal chemistry**, 7, n. 10, p. 1006-1014, 2007.

CERDEÑO-TÁRRAGA, A. *et al.* The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. **Nucleic acids research**, 31, n. 22, p. 6516-6523, 2003.

CDC. **About Diphtheria.** 2020. Disponível em: <<https://www.cdc.gov/diphtheria/about/index.html>> Acesso em: 01 março de 2020.

CHANDRAN, R. *et al.* Diphtheroids-important nosocomial pathogens. **Journal of clinical and diagnostic research: JCDR**, 10, n. 12, p. DC28, 2016.

CHAUTARD, E.; THIERRY-MIEG, N.; RICARD-BLUM, S. Interaction networks: from protein functions to drug discovery. A review. **Pathologie Biologie**, 57, n. 4, p. 324-333, 2009.

CHOTHIA, C.; LESK, A. M. The relation between the divergence of sequence and structure in proteins. **The EMBO journal**, 5, n. 4, p. 823-826, 1986.

COLLINS, M. Genus Corynebacterium. **Bergey's manual of systematic bacteriology**, 2, p. 1266-1276, 1986.

CONSORTIUM, G. O. The gene ontology resource: 20 years and still GOing strong. **Nucleic acids research**, 47, n. D1, p. D330-D338, 2019.

CONSORTIUM, U. UniProt: a worldwide hub of protein knowledge. **Nucleic acids research**, 47, n. D1, p. D506-D515, 2019.

CRAMER, R. D. Topomer CoMFA: a design methodology for rapid lead optimization. **Journal of medicinal chemistry**, 46, n. 3, p. 374-388, 2003.

CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, 110, n. 18, p. 5959-5967, 1988.

CULLEN, L. M.; ARNDT, G. M. Genome-wide screening for gene function using RNAi in mammalian cells. **Immunology and cell biology**, 83, n. 3, p. 217-223, 2005.

DAFFÉ, M. The cell envelope of corynebacteria. **Handbook of Corynebacterium glutamicum**, p. 121-148, 2005.

DAINA, A.; MICHIELIN, O.; ZOETE, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. **Scientific reports**, 7, p. 42717, 2017.

DATAR, P. A. *et al.* Comparative residue interaction analysis (CoRIA): a 3D-QSAR approach to explore the binding contributions of active site residues with ligands. **Journal of computer-aided molecular design**, 20, n. 6, p. 343-360, 2006.

DE JESUS BENEVIDES, L. *et al.* Genome sequence of *Corynebacterium ulcerans* strain FRC11. **Genome announcements**, 3, n. 2, 2015.

DE MATTOS GUARALDI, A. L.; HIRATA, R.; DE CARVALHO AZEVEDO, V. A. *Corynebacterium diphtheriae*, *Corynebacterium ulcerans* and *Corynebacterium pseudotuberculosis*—general aspects. *In: Corynebacterium diphtheriae and Related Toxigenic Species*: Springer, 2014. p. 15-37.

DE SOUSA SILVA, A. d. S. *et al.* Draft genome sequence of *Corynebacterium ulcerans* FRC58, isolated from the bronchitic aspiration of a patient in France. **Genome announcements**, 2, n. 1, 2014.

DE ZOYSA, A. *et al.* Characterization of toxigenic *Corynebacterium ulcerans* strains isolated from humans and domestic cats in the United Kingdom. **Journal of Clinical Microbiology**, 43, n. 9, p. 4377-4381, 2005.

DELANO, W. L. Pymol: An open-source molecular graphics tool. **CCP4 Newsletter on protein crystallography**, 40, n. 1, p. 82-92, 2002.

DENG, L. *et al.* MADOKA: an ultra-fast approach for large-scale protein structure similarity searching. **BMC bioinformatics**, 20, n. 19, p. 1-10, 2019.

DESSAU, R. *et al.* Pulmonary nodules due to *Corynebacterium ulcerans*. **European Respiratory Journal**, 8, n. 4, p. 651-653, 1995.

DEWINTER, L. M.; BERNARD, K. A.; ROMNEY, M. G. Human clinical isolates of *Corynebacterium diphtheriae* and *Corynebacterium ulcerans* collected in Canada from 1999 to 2003 but not fitting reporting criteria for cases of diphtheria. **Journal of clinical microbiology**, 43, n. 7, p. 3447-3449, 2005.

DIAS, A. A. d. S. d. *et al.* *Corynebacterium ulcerans* diphtheria: an emerging zoonosis in Brazil and worldwide. **Revista de saude publica**, 45, p. 1176-1191, 2011.

DIESTEL, R. Graph theory. 2005. **Grad. Texts in Math**, 101, 2005.

DILUCCA, M.; CIMINI, G.; GIANSAITI, A. Topological transition in bacterial protein-protein interaction networks ruled by gene conservation, essentiality and function. **arXiv preprint arXiv:1708.02299**, 2017.

DIXON, S. L. *et al.* PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. **Journal of computer-aided molecular design**, 20, n. 10-11, p. 647-671, 2006.

DUNHAM, W. H.; MULLIN, M.; GINGRAS, A. C. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. **Proteomics**, 12, n. 10, p. 1576-1590, 2012.

EGGELING, L.; BESRA, G. S.; ALDERWICK, L. Structure and synthesis of the cell wall. **Corynebacteria: genomics and molecular biology**. Caister Academic Press, Norfolk, UK, p. 267-294, 2008.

EISENBERG, D. *et al.* Protein function in the post-genomic era. **Nature**, 405, n. 6788, p. 823-826, 2000.

EISENBERG, T. *et al.* Nontoxigenic tox-bearing *Corynebacterium ulcerans* infection among game animals, Germany. **Emerging infectious diseases**, 20, n. 3, p. 448, 2014.

EISENBERG, T. *et al.* Outbreak with clonally related isolates of *Corynebacterium ulcerans* in a group of water rats. **BMC microbiology**, 15, n. 1, p. 42, 2015.

ENGELHARDT, H.; PETERS, J. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer–cell wall interactions. **Journal of structural biology**, 124, n. 2-3, p. 276-302, 1998.

ESCH, R.; MERKL, R. Conserved genomic neighborhood is a strong but no perfect indicator for a direct interaction of microbial gene products. **BMC bioinformatics**, 21, n. 1, p. 1-8, 2020.

EWING, T. J. *et al.* DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. **Journal of computer-aided molecular design**, 15, n. 5, p. 411-428, 2001.

FIELDS, S.; SONG, O.-k. A novel genetic system to detect protein–protein interactions. **Nature**, 340, n. 6230, p. 245-246, 1989.

FOSTER, G. *et al.* *Corynebacterium ulcerans* in free-ranging otters. **The Veterinary record**, 150, n. 16, p. 524-524, 2002.

FREEMAN, V. J. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. **Journal of bacteriology**, 61, n. 6, p. 675, 1951.

FUNKE, G.; LAWSON, P. A.; COLLINS, M. D. Heterogeneity within human-derived centers for disease control and prevention (CDC) coryneform group ANF-1-like

bacteria and description of *Corynebacterium auris* sp. nov. **International Journal of Systematic and Evolutionary Microbiology**, 45, n. 4, p. 735-739, 1995.

FUURSTED, K. *et al.* Non-toxigenic tox gene-bearing *Corynebacterium ulcerans* in a traumatic ulcer from a human case and his asymptomatic dog. **Microbes and infection**, 17, n. 10, p. 717-719, 2015.

GEISLER-LEE, J. *et al.* A predicted interactome for *Arabidopsis*. **Plant physiology**, 145, n. 2, p. 317-329, 2007.

GHOSH, S. *et al.* Structure-based virtual screening of chemical libraries for drug discovery. **Current opinion in chemical biology**, 10, n. 3, p. 194-202, 2006.

GIAEVER, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. **nature**, 418, n. 6896, p. 387-391, 2002.

GIBBS, N. **Ab initio protein tertiary structure prediction using restricted ramachandran geometries and physio-chemical potentials**. 2001. -, University of Bristol.

GILBERT, R.; STEWART, F. *Corynebacterium ulcerans*; a pathogenic microorganism. **J. Lab. Clin. Med.**, 12, p. 756-761, 1926.

GINALSKI, K. Comparative modeling for protein structure prediction. **Current opinion in structural biology**, 16, n. 2, p. 172-177, 2006.

GOHLKE, H.; KLEBE, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. **Angewandte Chemie International Edition**, 41, n. 15, p. 2644-2676, 2002.

GÖTZ, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic acids research**, 36, n. 10, p. 3420-3435, 2008.

GOVINDARAJ, R. G.; BRYLINSKI, M. Comparative assessment of strategies to identify similar ligand-binding pockets in proteins. **BMC bioinformatics**, 19, n. 1, p. 91, 2018.

GRAM, C. Ueber die isolirte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten. **Fortschritte der Medicin**, 2, p. 185-189, 1884.

HADFIELD, T. L. *et al.* The pathology of diphtheria. **Journal of Infectious Diseases**, 181, n. Supplement_1, p. S116-S120, 2000.

HAGEL, J. M.; FACCHINI, P. J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. **Plant and Cell Physiology**, 54, n. 5, p. 647-672, 2013.

HAMDI, A.; COLAS, P. Yeast two-hybrid methods and their applications in drug discovery. **Trends in pharmacological sciences**, 33, n. 2, p. 109-118, 2012.

HAMZA, A.; WEI, N.-N.; ZHAN, C.-G. Ligand-based virtual screening approach using a new scoring function. **Journal of chemical information and modeling**, 52, n. 4, p. 963-974, 2012.

HARRIS, P. *et al.* Structural basis for the catalytic mechanism of a proficient enzyme: orotidine 5'-monophosphate decarboxylase. **Biochemistry**, 39, n. 15, p. 4217-4224, 2000.

HATANAKA, A. *et al.* *Corynebacterium ulcerans* diphtheria in Japan. **Emerging infectious diseases**, 9, n. 6, p. 752, 2003.

HO, B. K.; BRASSEUR, R. The Ramachandran plots of glycine and pre-proline. **BMC structural biology**, 5, n. 1, p. 14, 2005.

HOLM, L.; SANDER, C. Protein structure comparison by alignment of distance matrices. **Journal of molecular biology**, 233, n. 1, p. 123-138, 1993.

HOMMEZ, J. *et al.* Identification of nonlipophilic corynebacteria isolated from dairy cows with mastitis. **Journal of clinical microbiology**, 37, n. 4, p. 954-957, 1999.

HOOFT, R. W.; SANDER, C.; VRIEND, G. Objectively judging the quality of a protein structure from a Ramachandran plot. **Bioinformatics**, 13, n. 4, p. 425-430, 1997.

HUANG, T.-W.; LIN, C.-Y.; KAO, C.-Y. Reconstruction of human protein interolog network using evolutionary conserved network. **BMC bioinformatics**, 8, n. 1, p. 152, 2007.

HUERTA-CEPAS, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. **Molecular biology and evolution**, 34, n. 8, p. 2115-2122, 2017.

IDEKER, T.; GALITSKI, T.; HOOD, L. A new approach to decoding life: systems biology. **Annual review of genomics and human genetics**, 2, n. 1, p. 343-372, 2001.

JEONG, H. *et al.* Lethality and centrality in protein networks. **Nature**, 411, n. 6833, p. 41-42, 2001.

JEONG, H. *et al.* The large-scale organization of metabolic networks. **Nature**, 407, n. 6804, p. 651-654, 2000.

JONES, G.; WILLETT, P.; GLEN, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. **Journal of molecular biology**, 245, n. 1, p. 43-53, 1995.

KALIDAS, Y.; CHANDRA, N. PocketDepth: a new depth based algorithm for identification of ligand binding sites in proteins. **Journal of structural biology**, 161, n. 1, p. 31-42, 2008.

KALITA, P. *et al.* Role of the glutaredoxin domain and FAD in the stabilization of thioredoxin glutathione reductase. **Archives of biochemistry and biophysics**, 656, p. 38-45, 2018.

KÄLLBERG, M. *et al.* Template-based protein structure modeling using the RaptorX web server. **Nature protocols**, 7, n. 8, p. 1511, 2012.

KALYAANAMOORTHY, S.; CHEN, Y.-P. P. Structure-based drug design to augment hit discovery. **Drug discovery today**, 16, n. 17-18, p. 831-839, 2011.

KATSUKAWA, C. *et al.* Toxigenic *Corynebacterium ulcerans* isolated from a wild bird (ural owl) and its feed (shrew-moles): comparison of molecular types with human isolates. **BMC research notes**, 9, n. 1, p. 181, 2016.

KELLEY, L. A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. **Nature protocols**, 10, n. 6, p. 845, 2015.

KELLEY, L. A.; STERNBERG, M. J. Protein structure prediction on the Web: a case study using the Phyre server. **Nature protocols**, 4, n. 3, p. 363, 2009.

KELLY, W. P. **On the analysis of protein interaction networks**. 2009. -, Imperial College London.

KERRIEN, S. *et al.* The IntAct molecular interaction database in 2012. **Nucleic acids research**, 40, n. D1, p. D841-D846, 2012.

KESHAVA PRASAD, T. *et al.* Human protein reference database—2009 update. **Nucleic acids research**, 37, n. suppl_1, p. D767-D772, 2009.

KIHARA, D.; SKOLNICK, J. The PDB is a covering set of small protein structures. **Journal of molecular biology**, 334, n. 4, p. 793-802, 2003.

KIMURA, Y. *et al.* Acute peritonitis due to *Corynebacterium ulcerans* in a patient receiving continuous ambulatory peritoneal dialysis: a case report and literature review. **Clinical and experimental nephrology**, 15, n. 1, p. 171-174, 2011.

KISELY, S.; PRICE, S.; WARD, T. '*Corynebacterium ulcerans*': a potential cause of diphtheria. **Communicable disease report. CDR review**, 4, n. 5, p. R63-64, 1994.

KITCHEN, D. B. *et al.* Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature reviews Drug discovery**, 3, n. 11, p. 935-949, 2004.

KLEYWEGT, G. J.; JONES, T. A. Phi/psi-chology: Ramachandran revisited. **Structure**, 4, n. 12, p. 1395-1400, 1996.

KOMIYA, T. *et al.* Two Japanese *Corynebacterium ulcerans* isolates from the same hospital: ribotype, toxigenicity and serum antitoxin titre. **Journal of medical microbiology**, 59, n. 12, p. 1497-1504, 2010.

KUMAR, A.; JHA, A. Drug Development Strategies. **Anticandidal Agents**, p. 63-71, 2017.

LARTIGUE, M.-F. *et al.* Corynebacterium ulcerans in an immunocompromised patient with diphtheria and her dog. **Journal of clinical microbiology**, 43, n. 2, p. 999-1001, 2005.

LASKOWSKI, R. A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of applied crystallography**, 26, n. 2, p. 283-291, 1993.

LAVECCHIA, A.; DI GIOVANNI, C. Virtual screening strategies in drug discovery: a critical review. **Current medicinal chemistry**, 20, n. 23, p. 2839-2860, 2013.

LE GUILLOUX, V.; SCHMIDTKE, P.; TUFFERY, P. Fpocket: an open source platform for ligand pocket detection. **BMC bioinformatics**, 10, n. 1, p. 168, 2009.

LEEK, M. *et al.* Diphtheria with a difference—a rare Corynebacterium fatality with associated apoptotic cell death. **Histopathology**, 16, n. 2, p. 187-189, 1990.

LETEK, M. *et al.* Cell growth and cell division in the rod-shaped actinomycete Corynebacterium glutamicum. **Antonie Van Leeuwenhoek**, 94, n. 1, p. 99-109, 2008.

LEWIS, A. C. *et al.* What evidence is there for the homology of protein-protein interactions? **PLoS Comput Biol**, 8, n. 9, p. e1002645, 2012.

LI, S. *et al.* A map of the interactome network of the metazoan C. elegans. **Science**, 303, n. 5657, p. 540-543, 2004.

LIN, J.-S.; LAI, E.-M. Protein–protein interactions: yeast two-hybrid system. *In*: **Bacterial Protein Secretion Systems**: Springer, 2017. p. 177-187.

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

LOBLEY, A.; SADOWSKI, M. I.; JONES, D. T. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. **Bioinformatics**, 25, n. 14, p. 1761-1767, 2009.

LOVELL, S. *et al.* Structure validation by C α geometry: phi, psi and C β deviation. **Proteins** 50, 437e450. 2003.

LUO, H. *et al.* DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. **Nucleic acids research**, 42, n. D1, p. D574-D580, 2014.

LYNE, P. D. Structure-based virtual screening: an overview. **Drug discovery today**, 7, n. 20, p. 1047-1055, 2002.

MACALINO, S. J. Y. *et al.* Role of computer-aided drug design in modern drug discovery. **Archives of pharmacal research**, 38, n. 9, p. 1686-1701, 2015.

MASON, O.; VERWOERD, M. Graph theory and networks in biology. **IET systems biology**, 1, n. 2, p. 89-119, 2007.

MELO-FILHO, C. C.; BRAGA, R. C.; ANDRADE, C. H. 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. **Curr Comput Aided Drug Des**, 10, n. 2, p. 148-159, 2014.

MICHELET, L.; ZAFFAGNINI, M.; LEMAIRE, D. Thioredoxins and related proteins. *In: The Chlamydomonas Sourcebook*: Elsevier, 2009. p. 401-443.

MINAMI, S. *et al.* MICAN-SQ: a sequential protein structure alignment program that is applicable to monomers and all types of oligomers. **Bioinformatics**, 34, n. 19, p. 3324-3331, 2018.

MOE, C. Montreal. **Quebec, Canada, Chemical Computing Group Inc**, 2005.

MONDIALE DE LA SANTÉ, O.; ORGANIZATION, W. H. Diphtheria vaccine: WHO position paper–August 2017–Vaccin antidiphthérique: Note de synthèse de l’OMS–août 2017. **Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire**, 92, n. 31, p. 417-435, 2017.

MONNET, V. Bacterial oligopeptide-binding proteins. **Cellular and Molecular Life Sciences CMLS**, 60, n. 10, p. 2100-2114, 2003.

MOOKADAM, F. *et al.* *Corynebacterium jeikeium* endocarditis: a systematic overview spanning four decades. **European Journal of Clinical Microbiology and Infectious Diseases**, 25, n. 6, p. 349-353, 2006.

MORRIS, A. L. *et al.* Stereochemical quality of protein structure coordinates. **Proteins: Structure, Function, and Bioinformatics**, 12, n. 4, p. 345-364, 1992.

MORRIS, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. **Journal of computational chemistry**, 30, n. 16, p. 2785-2791, 2009.

MOULT, J. *et al.* Critical assessment of methods of protein structure prediction (CASP)—round IX. **Proteins: Structure, Function, and Bioinformatics**, 79, n. S10, p. 1-5, 2011.

MROZEK, D.; MAŁYSIAK-MROZEK, B., 2013, **CASSERT: a two-phase alignment algorithm for matching 3D structures of proteins**. Springer. 334-343.

MS. **Nota Informativa N° 9/2019-CGDT/DEVIT/SVS/MS**. Brasília-DF, 12 de fev de 2019. Disponível em: <<https://portalarquivos2.saude.gov.br/images/pdf/2019/abril/08/Nota-informativa-N09-2019-CGDT-DEVIT-SVS-MS.pdf>> Acesso em: 27 fev. de 2020.

NAYAL, M.; HONIG, B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. **Proteins: Structure, Function, and Bioinformatics**, 63, n. 4, p. 892-906, 2006.

NCBI. National Center for Biotechnology Information. 1988. **Bethesda (MD): National Library of Medicine (US)**. Disponível em: <<https://www.ncbi.nlm.nih.gov>>.

NCBI. Taxonomy Browser. 2020a. **Bethesda (MD): National Library of Medicine (US)**. Disponível em: <<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>>. Acesso em: 10 de março de 2020.

NCBI. Genome Information by Organism. 2020b. **Bethesda (MD): National Library of Medicine (US)**. Disponível em: <<https://www.ncbi.nlm.nih.gov/genome/genomes/3461>>. Acesso em: 10 de março de 2020.

NUNES, J. E. *et al.* Molecular, kinetic, thermodynamic, and structural analyses of *Mycobacterium tuberculosis* hisD-encoded metal-dependent dimeric histidinol

dehydrogenase (EC 1.1. 1.23). **Archives of biochemistry and biophysics**, 512, n. 2, p. 143-153, 2011.

NUREKI, S.-i. *et al.* Corynebacterium ulcerans infection of the lung mimicking the histology of Churg-Strauss syndrome. **Chest**, 131, n. 4, p. 1237-1239, 2007.

OTSUJI, K. *et al.* Mutation and Diversity of Diphtheria Toxin in Corynebacterium ulcerans. 2019.

PAJERSKI, W. *et al.* Attachment efficiency of gold nanoparticles by Gram-positive and Gram-negative bacterial strains governed by surface charges. **Journal of Nanoparticle Research**, 21, n. 8, p. 186, 2019.

PANDIT, S. B.; SKOLNICK, J. Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. **BMC bioinformatics**, 9, n. 1, p. 531, 2008.

PARISH, T. Starvation survival response of Mycobacterium tuberculosis. **Journal of bacteriology**, 185, n. 22, p. 6702-6706, 2003.

PARISH, T. *et al.* Production of mutants in amino acid biosynthesis genes of Mycobacterium tuberculosis by homologous recombination. **Microbiology**, 145, n. 12, p. 3497-3503, 1999.

PARK, D. *et al.* Analysis of human disease genes in the context of gene essentiality. **Genomics**, 92, n. 6, p. 414-418, 2008.

PASCUAL, C. *et al.* Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences. **International journal of systematic and evolutionary microbiology**, 45, n. 4, p. 724-728, 1995.

PAVANKUMAR, T. L.; SINHA, A. K.; RAY, M. K. All three subunits of RecBCD enzyme are essential for DNA repair and low-temperature growth in the Antarctic Pseudomonas syringae Lz4W. **PLoS One**, 5, n. 2, p. e9412, 2010.

PAZOS, F. *et al.* Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. **Journal of molecular biology**, 352, n. 4, p. 1002-1015, 2005.

PAZOS, F.; VALENCIA, A. In silico two-hybrid system for the selection of physically interacting protein pairs. **Proteins: Structure, Function, and Bioinformatics**, 47, n. 2, p. 219-227, 2002.

PÉREZ, E. *et al.* The importance of chorismate mutase in the biocontrol potential of *Trichoderma parareesei*. **Frontiers in microbiology**, 6, p. 1181, 2015.

PRATHIPATI, P.; DIXIT, A.; SAXENA, A. K. Computer-aided drug design: Integration of structure-based and ligand-based approaches in drug design. **Current Computer-Aided Drug Design**, 3, n. 2, p. 133-148, 2007.

PUTZ, M. V.; LACRĂMĂ, A.-M. Introducing spectral structure activity relationship (S-SAR) analysis. Application to ecotoxicology. **International journal of molecular sciences**, 8, n. 5, p. 363-391, 2007.

RAIN, J.-C. *et al.* The protein–protein interaction map of *Helicobacter pylori*. **Nature**, 409, n. 6817, p. 211-215, 2001.

RAMACHANDRAN, G. N. Stereochemistry of polypeptide chain configurations. **J. Mol. Biol.**, 7, p. 95-99, 1963.

RAMACHANDRAN, G. t.; SASISEKHARAN, V. Conformation of polypeptides and proteins. *In: Advances in protein chemistry*: Elsevier, 1968. v. 23, p. 283-437.

RAMAN, K. Construction and analysis of protein–protein interaction networks. **Automated experimentation**, 2, n. 1, p. 2, 2010.

RAMAN, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. **Proteins: Structure, Function, and Bioinformatics**, 77, n. S9, p. 89-99, 2009.

RAMOS, P. I. P. *et al.* An integrative, multi-omics approach towards the prioritization of *Klebsiella pneumoniae* drug targets. **Scientific reports**, 8, n. 1, p. 1-19, 2018.

RAREY, M. *et al.* A fast flexible docking method using an incremental construction algorithm. **Journal of molecular biology**, 261, n. 3, p. 470-489, 1996.

RAU, J. *et al.* Corynebacterium ulcerans infection in roe deer (*Capreolus capreolus*). **Berliner und Munchener tierarztliche Wochenschrift**, 125, n. 3-4, p. 159-162, 2012.

REVA, B. A.; FINKELSTEIN, A. V.; SKOLNICK, J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? **Folding and Design**, 3, n. 2, p. 141-147, 1998.

RIEGEL, P. *et al.* Taxonomy of Corynebacterium diphtheriae and related taxa, with recognition of Corynebacterium ulcerans sp. nov. nom. rev. **FEMS microbiology letters**, 126, n. 3, p. 271-276, 1995.

ROEMER, T. *et al.* Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery. **Molecular microbiology**, 50, n. 1, p. 167-181, 2003.

ROSE, P. W. *et al.* The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. **Nucleic acids research**, 43, n. D1, p. D345-D356, 2015.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature protocols**, 5, n. 4, p. 725, 2010.

RUIZ, J. C. *et al.* Evidence for reductive genome evolution and lateral acquisition of virulence functions in two Corynebacterium pseudotuberculosis strains. **PloS one**, 6, n. 4, 2011.

SALWINSKI, L. *et al.* The database of interacting proteins: 2004 update. **Nucleic acids research**, 32, n. suppl_1, p. D449-D451, 2004.

SANCHEZ, C. *et al.* Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database. **Nucleic acids research**, 27, n. 1, p. 89-94, 1999.

SANTOS, A. S. *et al.* Searching whole genome sequences for biochemical identification features of emerging and reemerging pathogenic Corynebacterium species. **Functional & integrative genomics**, 18, n. 5, p. 593-610, 2018.

SASSO, S. *et al.* Structure and function of a complex between chorismate mutase and DAHP synthase: efficiency boost for the junior partner. **The EMBO journal**, 28, n. 14, p. 2128-2142, 2009.

SCHNECKE, V.; KUHN, L. A. Virtual screening with solvation and ligand-induced complementarity. *In: Virtual Screening: An Alternative or Complement to High Throughput Screening?*: Springer, 2000. p. 171-190.

SCHUHEGGER, R. *et al.* Pigs as source for toxigenic *Corynebacterium ulcerans*. **Emerging infectious diseases**, 15, n. 8, p. 1314, 2009.

SEKIZUKA, T. *et al.* *Corynebacterium ulcerans* 0102 carries the gene encoding diphtheria toxin on a prophage different from the *C. diphtheriae* NCTC 13129 prophage. **BMC microbiology**, 12, n. 1, p. 72, 2012.

SHANNON, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, 13, n. 11, p. 2498-2504, 2003.

SHENG, Y. *et al.* Mutagenesis of folylpolyglutamate synthetase indicates that dihydropteroate and tetrahydrofolate bind to the same site. **Biochemistry**, 47, n. 8, p. 2388-2396, 2008.

SHINDYALOV, I. N.; BOURNE, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. **Protein engineering**, 11, n. 9, p. 739-747, 1998.

SHITYAKOV, S.; FÖRSTER, C. In silico predictive model to determine vector-mediated transport properties for the blood–brain barrier choline transporter. **Advances and applications in bioinformatics and chemistry: AABC**, 7, p. 23, 2014.

SHOICHET, B. K. Virtual screening of chemical libraries. **Nature**, 432, n. 7019, p. 862-865, 2004.

SILVA, L. A. d. A. **Novos derivados sintéticos de alcalóides tetrahidroisoquinolínicos**. Dissertação (Mestrado) - UFPB/CCS. Paraíba, p. 83.

2016. Disponível em: <<https://repositorio.ufpb.br/jspui/handle/tede/9080>> Acesso em: 01 março de 2020.

SILVER, S.; LUSK, J. E. Bacterial magnesium, manganese, and zinc transport. *In: Ion Transport in Prokaryotes*: Elsevier, 1987. p. 165-180.

SING, A.; BIERSCHEK, S.; HEESEMANN, J. Classical diphtheria caused by *Corynebacterium ulcerans* in Germany: amino acid sequence differences between diphtheria toxins from *Corynebacterium diphtheriae* and *C. ulcerans*. **Clinical infectious diseases**, 40, n. 2, p. 325-326, 2005.

SING, A. *et al.* Detection of differences in the nucleotide and amino acid sequences of diphtheria toxin from *Corynebacterium diphtheriae* and *Corynebacterium ulcerans* causing extrapharyngeal infections. **Journal of clinical microbiology**, 41, n. 10, p. 4848-4851, 2003.

SMITH, G. R. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. **Microbiology and Molecular Biology Reviews**, 76, n. 2, p. 217-228, 2012.

SÖDING, J. Protein homology detection by HMM–HMM comparison. **Bioinformatics**, 21, n. 7, p. 951-960, 2005.

STADLER, R.; ZENK, M. H. A revision of the generally accepted pathway for the biosynthesis of the benzyltetrahydroisoquinoline alkaloid reticuline. **Liebigs Annalen der Chemie**, 1990, n. 6, p. 555-562, 1990.

STERLING, T.; IRWIN, J. J. ZINC 15–ligand discovery for everyone. **Journal of chemical information and modeling**, 55, n. 11, p. 2324-2337, 2015.

STUART, M. C.; KOUIMTZI, M.; HILL, S. R. **WHO model formulary 2008**. World Health Organization, 2009. 9241547650.

SZKLARCZYK, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. **Nucleic acids research**, 47, n. D1, p. D607-D613, 2019.

TARASSOV, K. *et al.* An in vivo map of the yeast protein interactome. **Science**, 320, n. 5882, p. 1465-1470, 2008.

TATUSOV, R. L. *et al.* The COG database: new developments in phylogenetic classification of proteins from complete genomes. **Nucleic acids research**, 29, n. 1, p. 22-28, 2001.

TATUSOVA, T. *et al.* RefSeq microbial genomes database: new representation and annotation strategy. **Nucleic acids research**, 43, n. 7, p. 3872, 2015.

THOMSEN, T. T. **Peptide antibiotics for ESKAPE pathogens: Past, present and future perspectives of antimicrobial peptides for the treatment of serious Gram-negative and Gram-positive infections**. 2016. -, Department of Biology, Faculty of Science, University of Copenhagen.

TROST, E. *et al.* Comparative analysis of two complete *Corynebacterium ulcerans* genomes and detection of candidate virulence factors. **BMC genomics**, 12, n. 1, p. 383, 2011.

TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of computational chemistry**, 31, n. 2, p. 455-461, 2010.

TRUONG, K.; IKURA, M. Domain fusion analysis by applying relational algebra to protein sequence and domain databases. **BMC bioinformatics**, 4, n. 1, p. 16, 2003.

UETZ, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. **Nature**, 403, n. 6770, p. 623-627, 2000.

VALENCIA, A.; PAZOS, F. Computational methods for the prediction of protein interactions. **Current opinion in structural biology**, 12, n. 3, p. 368-373, 2002.

VELA, A. *et al.* Isolation of *Corynebacterium xerosis* from animal clinical specimens. **Journal of clinical microbiology**, 44, n. 6, p. 2242, 2006.

VERLI, H. *Bioinformática: da biologia à flexibilidade molecular*. 2014.

VIANA, M. V. C. *et al.* Genome sequence of *Corynebacterium ulcerans* strain 210932. **Genome announcements**, 2, n. 6, 2014.

VUORINEN, A.; SCHUSTER, D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. **Methods**, 71, p. 113-134, 2015.

WAGNER, K. *et al.* Diphtheria in the United Kingdom, 1986–2008: the increasing role of *Corynebacterium ulcerans*. **Epidemiology & Infection**, 138, n. 11, p. 1519-1530, 2010.

WAGNER, K. S. *et al.* Diphtheria in the postepidemic period, Europe, 2000–2009. **Emerging infectious diseases**, 18, n. 2, p. 217, 2012.

WALHOUT, A. J. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. **Science**, 287, n. 5450, p. 116-122, 2000.

WALSH, F. M.; AMYES, S. G. Microbiology and drug resistance mechanisms of fully resistant pathogens. **Current opinion in microbiology**, 7, n. 5, p. 439-444, 2004.

WANG, J. *et al.* Identification of essential proteins based on edge clustering coefficient. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, 9, n. 4, p. 1070-1080, 2011.

WANG, S. *et al.* Protein structure alignment beyond spatial proximity. **Scientific reports**, 3, p. 1448, 2013.

WEERASEKERA, D. *et al.* Beyond diphtheria toxin: cytotoxic proteins of *Corynebacterium ulcerans* and *Corynebacterium diphtheriae*. **Microbiology**, 165, n. 8, p. 876-890, 2019.

WHO. **Diphtheria reported cases**. 2020. Disponível em: <http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tsincidence/diphtheria.html> Acesso em: 01 março de 2020.

WHO. **Diphtheria Vaccine-Preventable Diseases Surveillance Standards Diphtheria**. 2018. Disponível em: <https://www.who.int/immunization/monitoring_surveillance/burden/vpd/WHO_Surveil

lanceVaccinePreventable_04_Diphtheria_R2.pdf?ua=1> Acesso em: 01 março de 2020.

WINZELER, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. **science**, 285, n. 5429, p. 901-906, 1999.

WOLBER, G.; LANGER, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. **Journal of chemical information and modeling**, 45, n. 1, p. 160-169, 2005.

WU, S.; ZHANG, Y. Protein structure prediction. *In*: **Bioinformatics**: Springer, 2009. p. 225-242.

XU, J.; JIAO, F.; YU, L. Protein structure prediction using threading. *In*: **Protein Structure Prediction**: Springer, 2008. p. 91-121.

YADA, R.; JACKMAN, R.; NAKAI, S. Secondary structure prediction and determination of proteins—a review. **International journal of peptide and protein research**, 31, n. 1, p. 98-108, 1988.

YE, Y.; GODZIK, A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. **Bioinformatics**, 19, n. suppl_2, p. ii246-ii255, 2003.

YIN, C.; YAU, S. S.-T. A coevolution analysis for identifying protein-protein interactions by Fourier transform. **PLoS One**, 12, n. 4, p. e0174862, 2017.

YU, H. *et al.* Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. **Genome research**, 14, n. 6, p. 1107-1118, 2004.

YU, J. *et al.* Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. **Bioinformatics**, 26, n. 1, p. 46-52, 2010.

YU, W.; MACKERELL, A. D. Computer-aided drug design methods. *In*: **Antibiotics**: Springer, 2017. p. 85-106.

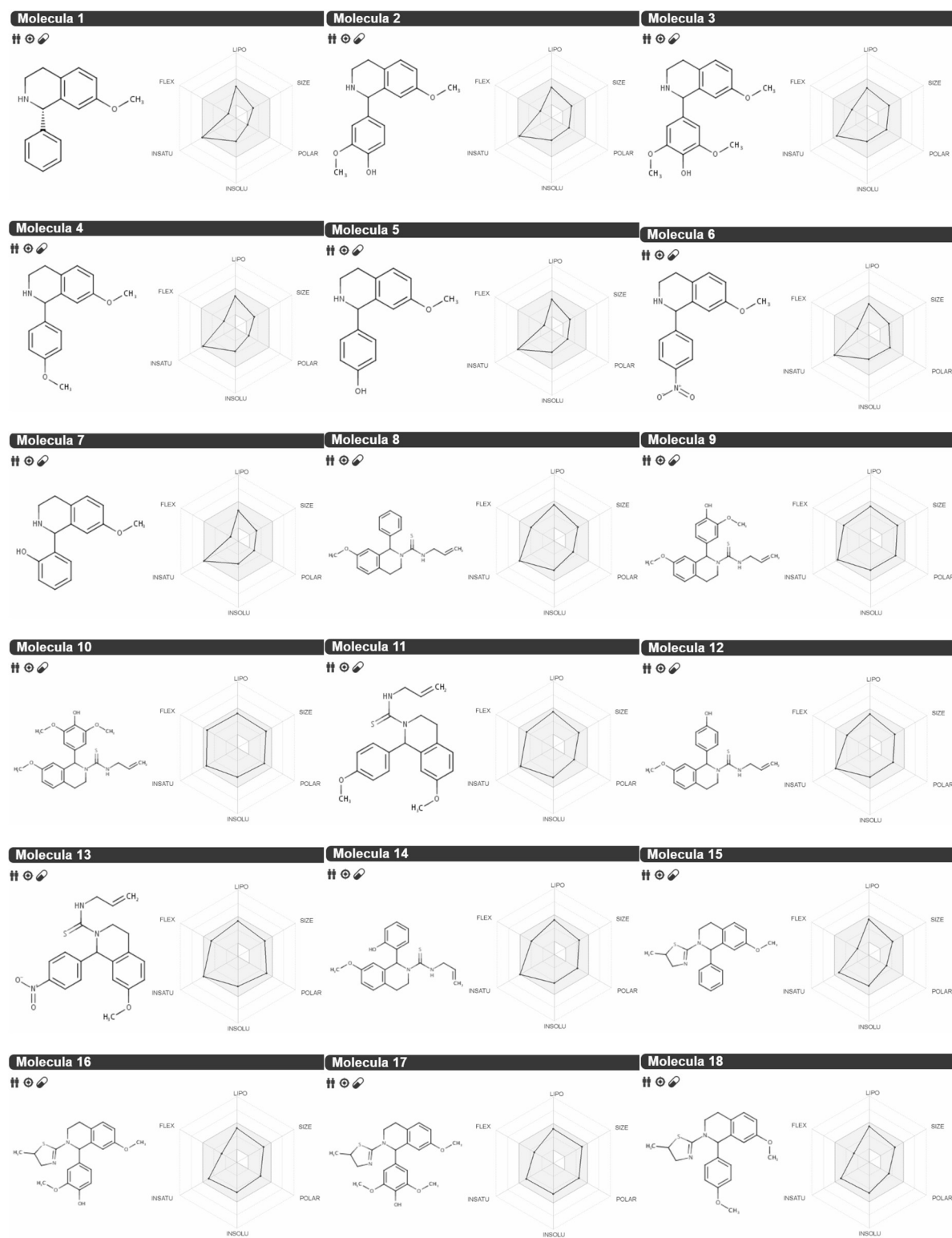
ZHANG, Y.; SKOLNICK, J. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic acids research**, 33, n. 7, p. 2302-2309, 2005.

ZHU, H.; PISABARRO, M. T. MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets. **Bioinformatics**, 27, n. 3, p. 351-358, 2011.

ZHU, H.; SNYDER, M. Protein chip technology. **Current opinion in chemical biology**, 7, n. 1, p. 55-63, 2003.

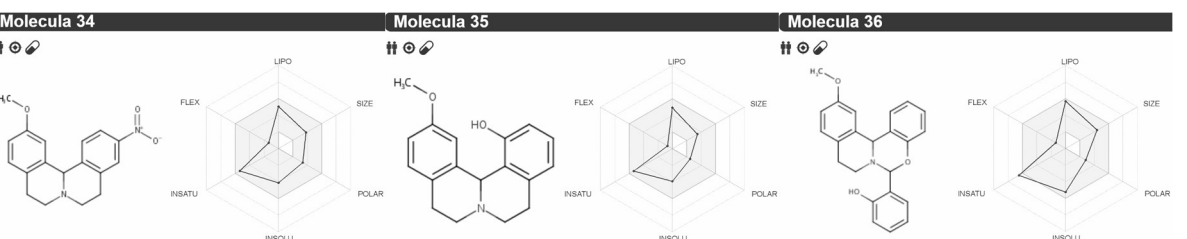
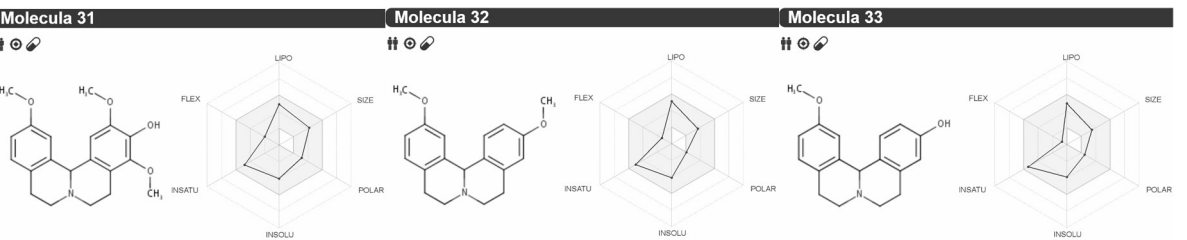
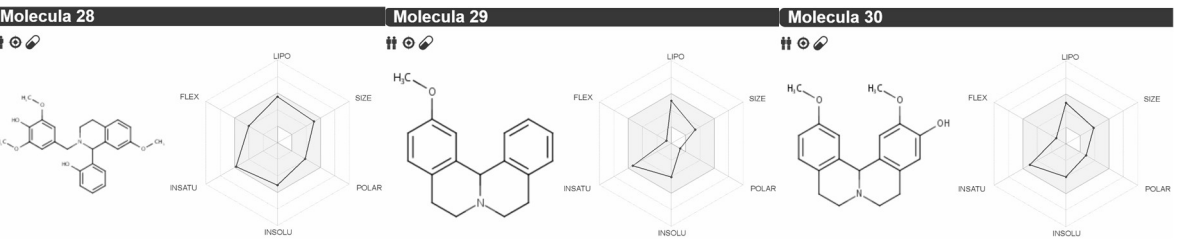
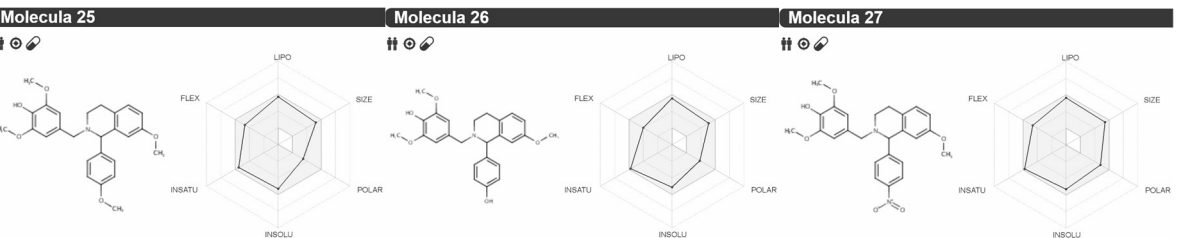
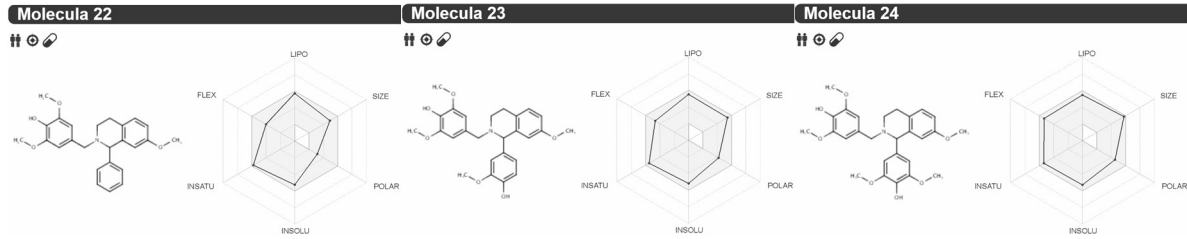
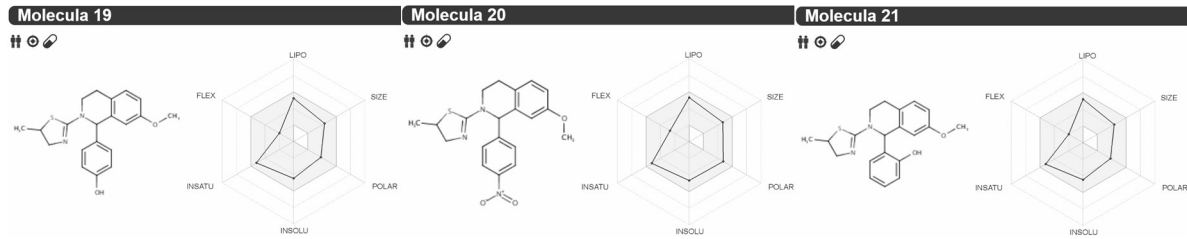
ZHU, J.; WENG, Z. FAST: a novel protein structure alignment algorithm. **PROTEINS: Structure, Function, and Bioinformatics**, 58, n. 3, p. 618-627, 2005.

Apêndice A - Estrutura química e propriedades dos 42 derivados sintéticos de alcalóides tetraisoquinolínicos



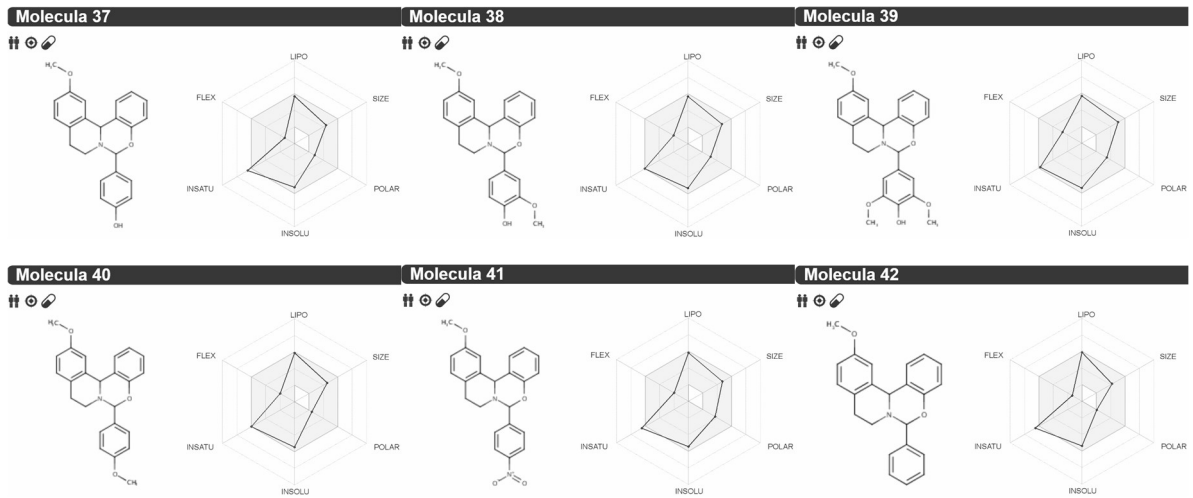
(Continua)

(Continuação)



(Continua)

(Continuação)



Fonte: SwissADME

À esquerda de cada figura está representada a estrutura química de cada um dos derivados sintéticos de alcalóides e à direita as seis propriedades que, quando dentro do campo ideal cinza, determinam a biodisponibilidade oral: FLEX (flexibilidade), LIPO (lipofilicidade), SIZE (tamanho), INSATU (insaturação), INSOLU (insolubilidade), POLAR (polaridade).

Apêndice B - Resultados de *docking* dos 42 derivados sintéticos de alcalóides tetraisoquinolínicos

Ligantes	Melhor score	Score médio	Pockets
Molécula 42	-9,3	-7,22	68
Molécula 34	-9,5	-7,21	82
Molécula 27	-9,9	-7,2	71
Molécula 22	-9,1	-7,19	70
Molécula 41	-9,6	-7,19	61
Molécula 26	-9,3	-7,17	69
Molécula 29	-9,2	-7,16	79
Molécula 33	-9	-7,15	76
Molécula 32	-9,2	-7,15	75
Molécula 23	-9,9	-7,14	72
Molécula 15	-9	-7,12	75
Molécula 28	-9,4	-7,11	76
Molécula 19	-8,9	-7,1	72
Molécula 6	-9	-7,1	84
Molécula 38	-9,3	-7,08	61
Molécula 20	-9,4	-7,07	79
Molécula 37	-9,2	-7,06	64
Molécula 35	-9	-7,06	75
Molécula 36	-8,9	-7,04	65
Molécula 39	-9,2	-7,04	55
Molécula 30	-8,8	-7,03	72
Molécula 2	-8,7	-7,01	75
Molécula 16	-8,7	-7,01	65
Molécula 5	-8,6	-7,01	77
Molécula 40	-9,2	-7,01	60
Molécula 25	-9,2	-7	70
Molécula 24	-9,4	-7	54
Molécula 7	-9	-7	72
Molécula 1	-8,9	-7	84
Molecular 4	-8,7	-6,98	77
Molécula 21	-8,9	-6,97	69
Molécula 18	-9,1	-6,95	71
Molécula 17	-9,1	-6,92	61
Molécula 3	-8,3	-6,9	62
Molécula 31	-8,5	-6,86	59
Molécula 8	-8,5	-6,86	64
Molécula 13	-8,8	-6,82	64

(Continua)

(Continuação)

Molécula 12	-8,2	-6,82	64
Molécula 11	-8,4	-6,8	58
Molécula 9	-8,2	-6,76	61
Molécula 10	-8,5	-6,76	45
Molécula 14	-7,8	-6,75	51

Fonte: Autor.

Apêndice C – Informações gerais das 457 proteínas *hub*

Código	Gene	Proteína	Ramachandran	Degree	COG	DEG	Docking
AKN76394	<i>guaA</i>	GMP synthase	89,0%	224	F	Sim	Não
AKN76391	<i>guaB</i>	Inosine-5'-monophosphate dehydrogenase	90,5%	209	F	Sim	Não
AKN77468	<i>CulFRC58_1614</i>	Inosine 5-monophosphate dehydrogenase	88,0%	195	F	Sim	Não
AKN77248	<i>recA</i>	Protein RecA	92,1%	182	L	Sim	Sim
AKN76359	<i>rpoA</i>	DNA-directed RNA polymerase subunit alpha	72,7%	165	K	Sim	Não
AKN77981	<i>dnaK</i>	Chaperone protein DnaK (Heat shock protein 70)	89,5%	164	O	Sim	Não
AKN76689	<i>eno</i>	Enolase	88,4%	158	G	Sim	Não
AKN76386	<i>groL</i>	Chaperonin	85,7%	158	O	Sim	Não
AKN77886	<i>groEl2</i>	Chaperonin (GroEL protein)	90,0%	156	O	Sim	Não
AKN76777	<i>odhA</i>	2-oxoglutarate dehydrogenase E1 component	89,7%	156	C	Sim	Não
AKN76277	<i>rpoB</i>	DNA-directed RNA polymerase subunit beta	92,2%	154	K	Sim	Sim
AKN76296	<i>fusA</i>	Elongation factor G	75,7%	153	J	Sim	Não
AKN76278	<i>rpoC</i>	DNA-directed RNA polymerase subunit beta	82,8%	152	K	Sim	Não
AKN77314	<i>rpsB</i>	30S ribosomal protein S2	86,6%	140	J	Sim	Não
AKN76294	<i>rpsL</i>	30S ribosomal protein S12	54,4%	130	J	Sim	Não
AKN76704	<i>glyA</i>	Serine hydroxymethyltransferase	92,3%	129	E	Sim	Não
AKN76313	<i>rpsC</i>	30S ribosomal protein S3	92,5%	129	J	Sim	Sim
AKN75859	<i>gyrB</i>	DNA gyrase subunit B	82,5%	126	L	Sim	Não
AKN76307	<i>rplD1</i>	50S ribosomal protein L4	90,2%	126	J	Sim	Não
AKN76308	<i>rplD2</i>	50S ribosomal protein L4	95,0%	126	J	Sim	Não
AKN76336	<i>rpsE</i>	30S ribosomal protein S5	79,2%	126	J	Sim	Não
AKN77338	<i>ffh</i>	Signal recognition particle protein (Fifty-four-like)	77,5%	125	U	Sim	Não
AKN76310	<i>rplB</i>	50S ribosomal protein L2	91,4%	124	J	Sim	Não
AKN76371	<i>rplM</i>	50S ribosomal protein L13	91,1%	123	J	Sim	Sim
AKN76358	<i>rpsD</i>	30S ribosomal protein S4	92,1%	123	J	Sim	Sim
AKN76297	<i>tuf</i>	Elongation factor	80,8%	122	J	Sim	Não
AKN75863	<i>gyrA</i>	DNA gyrase subunit A	84,1%	121	L	Sim	Não
AKN76323	<i>rplE</i>	50S ribosomal protein L5	92,9%	121	J	Sim	Sim
AKN76335	<i>rplR</i>	50S ribosomal protein L18	91,9%	121	J	Sim	Sim
AKN77334	<i>trmD</i>	tRNA (guanine-N(1)-)-methyltransferase	91,0%	120	J	Sim	Sim
AKN77313	<i>tsf</i>	Elongation factor Ts (EF-Ts)	92,4%	120	J	Sim	Sim
AKN76927	<i>rpsA</i>	30S ribosomal protein S1	85,0%	119	J	Sim	Não
AKN76264	<i>rplA</i>	50S ribosomal protein L1	93,1%	118	J	Sim	Sim
AKN76350	<i>secY</i>	Protein translocase subunit	73,8%	117	U	Sim	Não
AKN76334	<i>rplF</i>	50S ribosomal protein L6	95,1%	116	J	Sim	Sim
AKN76306	<i>rpsJ</i>	30S ribosomal protein S10	94,3%	116	J	Sim	Sim
AKN76312	<i>rplV</i>	50S ribosomal protein L22	87,0%	115	J	Sim	Não
AKN76357	<i>rpsK</i>	30S ribosomal protein S11	80,2%	115	J	Sim	Não
AKN76356	<i>rpsM</i>	30S ribosomal protein S13	76,7%	115	J	Sim	Não
AKN77268	<i>rpsO</i>	30S ribosomal protein S15	94,9%	115	J	Sim	Sim
AKN77103	<i>tpiA</i>	Triosephosphate isomerase (TIM)	95,5%	115	G	Sim	Não
AKN76295	<i>rpsG</i>	30S ribosomal protein S7	95,6%	113	J	Sim	Sim
AKN76832	<i>atpA</i>	ATP synthase subunit alpha	87,2%	112	C	Sim	Não
AKN77403	<i>ftsZ</i>	Cell division protein FtsZ	90,0%	111	D	Sim	Sim
AKN76263	<i>rplK</i>	50S ribosomal protein L11	77,6%	110	J	Sim	Não
AKN76333	<i>rpsH</i>	30S ribosomal protein S8	93,8%	109	J	Sim	Sim
AKN76314	<i>rplP</i>	50S ribosomal protein L16	89,2%	108	J	Sim	Não
AKN77667	<i>tig</i>	Trigger factor (TF)	78,7%	108	D	Sim	Não
AKN77311	<i>frr</i>	Ribosome-recycling factor (RRF) (Ribosome-releasing factor)	94,5%	107	J	Sim	Sim
AKN77123	<i>rpoZ</i>	DNA-directed RNA polymerase subunit omega (RNAP omega subunit)	91,3%	107	K	Sim	Sim
AKN76928	<i>ptsG</i>	Phosphotransferase system II Component	88,4%	106	G	Sim	Não
AKN77278	<i>ifB2</i>	Translation initiation factor IF-2	72,4%	105	J	Sim	Não
AKN77330	<i>lpS</i>	50S ribosomal protein L19	92,9%	105	J	Sim	Sim

(Continua)

(Continuação)

AKN76942	pheS	Phenylalanine--tRNA ligase alpha subunit	77,5%	105	J	Sim	Não
AKN76372	rpsI	30S ribosomal protein S9	68,9%	105	J	Sim	Não
AKN77184	thrS	Threonine--tRNA ligase	86,5%	104	J	Sim	Não
AKN77712	fas	Fatty acid synthase	92,4%	103	IQ	Sim	Sim
AKN77121	metK	S-adenosylmethionine synthase (AdoMet synthase)	92,5%	103	H	Sim	Não
AKN77650	obgE	GTPase Obg (GTP-binding protein Obg)	90,1%	103	S	Sim	Sim
AKN76321	rplN	50S ribosomal protein L14	87,6%	103	J	Sim	Não
AKN77574	dnaG	DNA primase	90,3%	102	L	Sim	Sim
AKN77870	ftsH	ATP-dependent zinc metalloprotease FtsH	85,3%	102	O	Sim	Não
AKN76338	rplO	50S ribosomal protein L15	89,7%	102	J	Sim	Não
AKN76268	rplJ	50S ribosomal protein L10	95,4%	100	J	Sim	Sim
AKN76626	rpsR	30S ribosomal protein S18	72,1%	100	J	Sim	Não
AKN76311	rpsS	30S ribosomal protein S19	74,0%	100	J	Sim	Não
AKN77211	sigB	RNA polymerase sigma factor	91,6%	100	K	Sim	Sim
AKN76834	atpD	ATP synthase subunit beta	81,6%	99	C	Sim	Não
AKN77204	sigA	RNA polymerase sigma factor RpoD	81,7%	98	K	Sim	Não
AKN77361	pyk	Kinaseinase	84,4%	97	G	Sim	Não
AKN76821	prfA	Peptide chain release factor 1	79,9%	96	J	Sim	Não
AKN76360	rplQ	50S ribosomal protein L17	92,2%	96	J	Sim	Sim
AKN76316	rpsQ	30S ribosomal protein S17	89,3%	96	J	Sim	Não
AKN77664	pabB	Para-aminobenzoate synthase component I	86,6%	95	EH	Sim	Não
AKN75856	dnaN	DNA polymerase III subunit beta	88,0%	93	L	Sim	Não
AKN77633	lepA	Elongation factor 4 (EF-4)	80,3%	93	M	Sim	Não
AKN76309	rplW	50S ribosomal protein L23	91,8%	93	J	Sim	Sim
AKN76337	rpmD	50S ribosomal protein L30	96,2%	93	J	Sim	Sim
AKN78128	rpsF	30S ribosomal protein S6	95,3%	92	J	Sim	Sim
AKN76235	hemE	Uroporphyrinogen decarboxylase	85,7%	91	H	Sim	Não
AKN76322	rplX	50S ribosomal protein L24	90,6%	91	J	Sim	Sim
AKN76627	rpsN	30S ribosomal protein S14	88,6%	91	J	Sim	Não
AKN77267	pnp	Polyribonucleotide nucleotidyltransferase	75,3%	89	J	Sim	Não
AKN77435	qcrA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	80,7%	89	C	Sim	Não
AKN76939	rplT	50S ribosomal protein L20	94,9%	89	J	Sim	Sim
AKN77337	rpsP	30S ribosomal protein S16	94,7%	89	J	Sim	Sim
AKN76721	ychF	Ribosome-binding ATPase	80,9%	89	J	Sim	Não
AKN77341	ftsY	Signal recognition particle receptor FtsY (SRP receptor)	77,1%	88	U	Sim	Não
AKN76542	prfB	Peptide chain release factor 2	91,3%	88	J	Sim	Sim
AKN77130	pyrB	Aspartate carbamoyltransferase	88,0%	88	F	Sim	Não
AKN76315	rpmC	50S ribosomal protein L29	98,3%	88	J	Sim	Sim
AKN77660	valS	Valine--tRNA ligase	78,9%	88	J	Sim	Não
AKN78126	rplI	50S ribosomal protein L9	90,1%	87	J	Sim	Sim
AKN77658	CuIFRC58_1804	Hypothetical protein	95,5%	86	S	Não	Sim
AKN76548	smpB	SsrA-binding protein	69,0%	86	O	Sim	Não
AKN77652	rpmA	50S ribosomal protein L27	73,5%	85	J	Sim	Não
AKN77271	truB	tRNA pseudouridine synthase B	88,7%	85	J	Sim	Não
AKN77136	efp	Elongation factor P (EF-P)	90,6%	84	J	Sim	Sim
AKN77105	gapA	Glyceraldehyde-3-phosphate dehydrogenase	84,0%	84	G	Sim	Não
AKN76269	rplL	50S ribosomal protein L7/L12	92,9%	84	J	Sim	Sim
AKN77975	aldH	Aldehyde dehydrogenase	90,0%	83	C	Sim	Não
AKN77656	ndk	Nucleoside diphosphate kinase (NDK) (NDP kinase)	87,6%	82	F	Sim	Não
AKN77348	rnc	Ribonuclease 3	91,2%	82	J	Sim	Sim
AKN76134	topA	DNA topoisomerase I	84,0%	82	L	Sim	Não
AKN76377	CuIFRC58_0523	Abhydrolase domain-containing protein	85,6%	81	Q	Não	Não
AKN76573	glTA	Citrate synthase	90,9%	80	C	Sim	Sim
AKN77449	aceF	Dihydrolipoamide acyltransferase	90,2%	79	C	Sim	Sim
AKN78146	leuS	Leucine--tRNA ligase	78,4%	79	J	Sim	Não
AKN77115	ribD	Riboflavin biosynthesis protein RibD	79,9%	79	H	Sim	Não
AKN76494	secA1	Protein translocase subunit	88,7%	78	U	Sim	Não
AKN76611	csm	Chorismate mutase	98,6%	77	E	Não	Sim

(Continua)

(Continuação)

AKN76414	folD	Bifunctional protein	90,0%	77	F	Sim	Não
AKN77575	glmS	Glutamine--fructose-6-phosphate aminotransferase isomerizing	69,1%	77	M	Sim	Não
AKN77128	carA	Carbamoyl-phosphate synthase small chain	85,6%	76	F	Sim	Não
AKN77859	gbsA	Betaine aldehyde dehydrogenase	92,2%	76	C	Sim	Não
AKN76885	leuB	3-isopropylmalate dehydrogenase	89,3%	76	CE	Sim	Não
AKN76456	pccB1	Propionyl CoA carboxylase beta chain 1	85,1%	75	I	Sim	Não
AKN76457	pccB2	Propionyl-CoA carboxylase beta chain 2	85,1%	75	I	Sim	Não
AKN77127	carB	Carbamoyl-phosphate synthase large chain	89,2%	74	F	Sim	Não
AKN76973	engA	GTPase	78,5%	74	F	Sim	Não
AKN77124	gmk	Guanylate kinase	92,9%	74	F	Sim	Não
AKN77653	rplU	50S ribosomal protein L21	91,8%	74	J	Sim	Sim
AKN78112	trxA2	Thioredoxin	93,4%	74	O	Sim	Não
AKN77397	ileS	Isoleucine--tRNA ligase	76,6%	73	J	Sim	Não
AKN76196	lpd	Dihydrolipoamide dehydrogenase	87,7%	73	C	Sim	Não
AKN77231	ptsF	PTS system fructose-specific EIIABC component	85,6%	73	GT	Sim	Não
AKN78084	sodA	Superoxide dismutase	89,8%	73	C	Sim	Não
AKN76434	CuIFRC58_0580	Hypothetical protein	94,1%	72		Não	Sim
AKN76071	leuA	2-isopropylmalate synthase	90,4%	72	E	Sim	Sim
AKN76943	pheT	Phenylalanine--tRNA ligase beta subunit	88,5%	72	J	Sim	Não
AKN76629	rpmB	50S ribosomal protein L28	59,4%	72	J	Sim	Não
AKN77170	ruvB	Holliday junction ATP-dependent DNA helicase RuvB	95,4%	72	L	Sim	Sim
AKN78206	trxA3	Thioredoxin	87,8%	72	O	Sim	Não
AKN76937	infC	Translation initiation factor IF-3	83,5%	71	J	Sim	Não
AKN78081	ldh2	L-lactate dehydrogenase (L-LDH)	91,3%	71	C	Sim	Não
AKN77661	mdh	Malate dehydrogenase	93,2%	70	C	Sim	Não
AKN76677	rplY	50S ribosomal protein L25	92,9%	70	J	Sim	Sim
AKN76938	rpmI	50S ribosomal protein L35	59,3%	70	J	Sim	Não
AKN76351	adk	Adenylate kinase	96,1%	69	F	Sim	Sim
AKN76833	atpG	ATP synthase gamma chain	92,9%	69	C	Sim	Sim
AKN76610	gabD	Succinate-semialdehyde dehydrogenase	91,0%	69	C	Sim	Não
AKN76892	gltX	Glutamate--tRNA ligase	92,1%	69	J	Sim	Sim
AKN76925	polA	DNA polymerase	78,1%	69	L	Sim	Não
AKN76831	atpH	ATP synthase subunit delta	88,3%	68	C	Sim	Não
AKN77855	lysS	Lysine--tRNA ligase	87,2%	68	J	Sim	Não
AKN77335	rimM	Ribosome maturation factor RimM	80,0%	67	J	Sim	Não
AKN76953	tyrS	Tyrosine--tRNA ligase	91,7%	67	J	Sim	Não
AKN77296	CuIFRC58_1442	Mycothione/glutathione reductase	83,9%	66	C	Sim	Não
AKN77720	CuIFRC58_1866	Hypothetical protein	70,4%	65	V	Sim	Não
AKN78067	CuIFRC58_2213	Septum formation	76,9%	65	S	Não	Não
AKN75999	embB	Arabinosyl transferase	85,0%	65	M	Sim	Não
AKN77158	hisS	Histidine--tRNA ligase (Histidyl-tRNA synthetase)	88,4%	65	J	Sim	Não
AKN76879	ilvC	Ketol-acid reductoisomerase	93,5%	65	EH	Sim	Sim
AKN76441	lpdA	Flavoprotein disulfide reductase	84,6%	65	C	Sim	Não
AKN76262	nusG	Transcription termination/antitermination protein	72,4%	65	K	Sim	Não
AKN76679	prs	Ribose-phosphate pyrophosphokinase	91,5%	65	F	Sim	Não
AKN77635	rspT	30S ribosomal protein S20	98,8%	65	J	Sim	Sim
AKN76951	argH	Argininosuccinate lyase	92,9%	64	E	Sim	Não
AKN76998	deaD2	DEAD-box helicase	90,5%	64	L	Sim	Sim
AKN77353	gdh	Glutamate dehydrogenase	90,2%	64	E	Sim	Sim
AKN76906	leuD	3-isopropylmalate dehydratase small subunit	90,5%	64	E	Sim	Sim
AKN77049	mshC	L-cysteine:1D-myo-inositol 2-amino-2-deoxy-alpha-D-glucopyranoside ligase	89,6%	64	J	Sim	Não
AKN76609	pgi	Glucose-6-phosphate isomerase	90,4%	64	G	Sim	Não
AKN78009	aspC	Aminotransferase AlaT	86,2%	63	E	Sim	Não
AKN76509	CuIFRC58_0655	Hypothetical protein	81,9%	63	S	Sim	Não
AKN76744	CuIFRC58_0890	Hypothetical protein	89,7%	63	S	Não	Não
AKN77934	CuIFRC58_2080	Hypothetical protein	83,0%	63	I	Sim	Não

(Continua)

(Continuação)

AKN77831	cysS	Cysteinyl-tRNA synthetase	88,4%	63	J	Sim	Não
AKN76373	glmM	Phosphoglucosamine mutase	83,6%	63	G	Sim	Não
AKN77332	manP	PPTS system mannose-specific EIIBCA component	89,1%	63	GT	Sim	Não
AKN77944	fbaA	Fructose-bisphosphate aldolase class 2	92,6%	62	G	Sim	Sim
AKN76385	groS	Chaperonin	88,9%	62	O	Sim	Não
AKN76392	guaB2	Inosine-5'-monophosphate dehydrogenase	90,1%	62	F	Sim	Sim
AKN76355	infA	Translation initiation factor 1	77,0%	62	J	Sim	Não
AKN76650	metG	Methionine--tRNA ligase	81,0%	62	J	Sim	Não
AKN76617	purH	Bifunctional purine biosynthesis protein	92,4%	62	F	Sim	Não
AKN78065	serS	Serine--tRNA ligase	89,4%	62	J	Sim	Não
AKN77096	tkt	Transketolase	87,4%	62	G	Sim	Não
AKN77481	glnA2	Glutamine synthetase	80,3%	61	E	Sim	Não
AKN76877	ilvB	Acetolactate synthase	91,3%	61	H	Sim	Sim
AKN77104	pgk	Phosphoglycerate kinase	90,5%	61	F	Sim	Não
AKN77645	proA	Gamma-glutamyl phosphate reductase (GPR)	92,9%	61	E	Sim	Não
AKN76431	trpS	Tryptophanyl-tRNA ligase	93,4%	61	J	Sim	Não
AKN76362	truA	tRNA pseudouridine synthase A	89,7%	61	J	Sim	Não
AKN76575	accDA	Acetyl-CoA carboxylase carboxyl transferase	85,1%	60	I	Sim	Não
AKN76813	argS	Arginyl-tRNA ligase	82,2%	60	J	Sim	Não
AKN76513	CuIFRC58_0659	Hypothetical protein	81,3%	60	S	Não	Não
AKN76667	gap	Glyceraldehyde-3-phosphate dehydrogenase A	81,9%	60	G	Sim	Não
AKN77464	glnA1	Glutamine synthetase	89,0%	60	E	Sim	Não
AKN77366	hisF	Imidazole glycerol phosphate synthase subunit HisF	91,1%	60	E	Sim	Sim
AKN76874	ilvD	Dihydroxy-acid dehydratase	90,8%	60	EG	Sim	Sim
AKN77743	nrdE	Ribonucleoside-diphosphate reductase	87,1%	60	F	Sim	Não
AKN76510	rhIE	ATP-dependent RNA helicase	92,2%	60	L	Sim	Não
AKN77746	rpmJ	Ribosomal protein L36	64,7%	60	J	Sim	Não
AKN78205	trxB	Thioredoxin reductase	81,8%	60	C	Sim	Não
AKN77110	uvrC	UvrABC system protein C (Protein UvrC) (Excinuclease ABC subunit C)	90,4%	60	L	Sim	Sim
AKN76751	dapC	N-succinyldiaminopimelate aminotransferase	83,8%	59	E	Sim	Não
AKN77118	fmt	Methionyl-tRNA formyltransferase	89,2%	59	J	Sim	Não
AKN76866	pfkA	6-phosphofructo kinase	94,3%	59	F	Sim	Não
AKN77116	rpe	Ribulose-phosphate 3-epimerase	85,1%	59	G	Sim	Não
AKN76936	uvrA2	UvrABC system protein A	81,9%	59	L	Sim	Não
AKN76501	aroA	3-phosphoshikimate 1-carboxyvinyltransferase	87,5%	58	E	Sim	Não
AKN76830	atpF	ATP synthase subunit b	94,5%	58	C	Sim	Sim
AKN76950	argG	Argininosuccinate synthase	91,5%	57	E	Sim	Não
AKN77665	clpP1	ATP-dependent Clp protease proteolytic subunit	92,9%	57	O	Sim	Não
AKN77374	hisD	Histidinol dehydrogenase (HDH)	92,4%	57	E	Sim	Sim
AKN77369	hisH	Imidazole glycerol phosphate synthase subunit HisH	88,4%	57	E	Sim	Não
AKN77666	clp2	ATP-dependent Clp protease proteolytic subunit	92,8%	56	O	Sim	Não
AKN77025	cobL	Precorrin-6Y C5,15-methyltransferase	86,4%	56	H	Sim	Não
AKN77350	CuIFRC58_1496	Hypothetical protein	93,9%	56	D	Sim	Sim
AKN75855	dnaA	Chromosomal replication initiator protein DnaA	80,5%	56	L	Sim	Não
AKN77299	mapB	Methionine aminopeptidase (MAP) (MetAP)	90,0%	56	E	Sim	Não
AKN77815	poxB	Pyruvate dehydrogenase	86,8%	56	EH	Sim	Não
AKN76199	sdhA	Succinate dehydrogenase flavoprotein subunit	82,2%	56	C	Sim	Não
AKN76772	tatB	Sec-independent protein translocase protein	83,1%	56	U	Não	Não
AKN77098	zwf	Glucose-6-phosphate 1-dehydrogenase (G6PD)	88,7%	56	G	Sim	Não
AKN77708	CuIFRC58_1854	Hypothetical protein	93,5%	55	S	Não	Sim
AKN76791	deaD	ATP-dependent RNA helicase	86,6%	55	JKL	Sim	Não
AKN77451	gcvP	Glycine cleavage system P protein	89,5%	55	E	Sim	Não
AKN76239	gpmB1	Phosphoglycerate mutase	89,0%	55	G	Sim	Não
AKN76814	lysA2	Diaminopimelate decarboxylase	92,2%	55	E	Sim	Sim
AKN76352	map	Methionine aminopeptidase	90,0%	55	E	Sim	Sim
AKN76630	rpmE	50S ribosomal protein L31 type B	26,7%	55	J	Sim	Não
AKN78215	rpmH	Ribosomal protein L34	94,7%	55	J	Sim	Não

(Continua)

(Continuação)

AKN76518	uvrD3	DNA helicase II	79,1%	55	L	Sim	Não
AKN77139	aroB	3-dehydroquinate synthase	88,0%	54	E	Sim	Não
AKN77021	cobJ	Precorrin-3B C17-methyltransferase	88,2%	54	H	Sim	Não
AKN76007	CulFRC58_0153	GtrA integral membrane protein	88,8%	54	S	Sim	Não
AKN77117	fmu	Ribosomal RNA small subunit methyltransferase B	88,0%	54	JK	Sim	Não
AKN76048	gluQ	Glutamyl-Q tRNA synthetase	88,5%	54	J	Sim	Não
AKN76481	manB2	Phosphomannomutase	87,5%	54	G	Sim	Não
AKN77763	murA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	91,5%	54	M	Sim	Sim
AKN77391	rulD	Pseudouridine synthase	87,1%	54	J	Sim	Não
AKN76931	uvrB2	UvrABC system protein B	93,5%	54	L	Sim	Sim
AKN77146	alaS	Alanine--tRNA ligase (Alanyl-tRNA synthetase)	86,7%	53	J	Sim	Não
AKN76240	ccsX	Thiol-disulfide isomerase/thioredoxin	80,0%	53	CO	Sim	Não
AKN77458	CulFRC58_1604	FecCD transport family protein	90,1%	53	P	Sim	Sim
AKN76554	fecC	ABC transporter iron III	89,5%	53	P	Sim	Não
AKN76444	pyc	Pyruvate carboxylase	89,4%	53	C	Sim	Não
AKN77933	purA	Adenylosuccinate synthetase (AMPSase) (AdSS)	88,5%	52	F	Sim	Não
AKN76200	sdhB	Succinate dehydrogenase iron-sulfur subunit	77,1%	52	C	Sim	Não
AKN76989	secA2	Protein translocase subunit	88,5%	52	U	Sim	Não
AKN77663	clpX	ATP-dependent Clp protease ATP-binding subunit ClpX	86,0%	51	O	Sim	Não
AKN76538	CulFRC58_0684	Methylmalonyl-CoA carboxyltransferase 1.3S subunit	79,8%	51	I	Sim	Não
AKN78202	CulFRC58_2348	Hypothetical protein	88,6%	51	NU	Sim	Não
AKN77373	hisC	Histidinol-phosphate aminotransferase	90,8%	51	E	Sim	Sim
AKN77253	pspA	Phage shock protein A (IM30)	94,7%	51	KT	Não	Sim
AKN77298	cobQ2	Cobyric acid synthase	81,1%	50	H	Sim	Não
AKN76227	CulFRC58_0373	Glutaredoxin domain protein	75,0%	50	O	Não	Não
AKN76680	glmU	Bifunctional protein	86,0%	50	M	Sim	Não
AKN76905	leuC	3-isopropylmalate dehydratase large subunit	88,4%	50	E	Sim	Não
AKN78091	recB2	RecB family nuclease	89,8%	50	S	Não	Não
AKN77097	tal	Transaldolase	93,3%	50	H	Sim	Sim
AKN77843	clpB2	ATP-dependent Clp protease ATP-binding subunit	81,5%	49	O	Sim	Não
AKN77052	lppL	Prolipoprotein	80,7%	49	S	Sim	Não
AKN76959	recN	DNA repair protein	84,8%	49	L	Não	Não
AKN78017	CulFRC58_2163	Aldehyde dehydrogenase	90,0%	48	C	Sim	Não
AKN76713	fumC	Fumarate hydratase class II	91,2%	48	C	Sim	Não
AKN76684	mfd	Transcription-repair-coupling factor	80,7%	48	L	Sim	Não
AKN77414	mraW	Ribosomal RNA small subunit methyltransferase H	87,4%	48	J	Sim	Não
AKN77741	nrdF1	Ribonucleoside-diphosphate reductase subunit beta	98,1%	48	F	Sim	Sim
AKN76058	recR	Recombination protein	89,8%	48	L	Sim	Não
AKN77399	ag84	Antigen 84	94,4%	47	D	Sim	Sim
AKN77148	CulFRC58_1294	Phosphotransferase enzyme family protein	83,6%	47	S	Não	Não
AKN77462	CulFRC58_1608	Hypothetical protein	83,3%	47	S	Sim	Não
AKN77589	CulFRC58_1735	Isoprenyl transferase	93,8%	47	H	Sim	Não
AKN77727	CulFRC58_1873	Hypothetical protein	88,2%	47	S	Não	Não
AKN77446	ilvE	Branched-chain-amino-acid aminotransferase	90,7%	47	EH	Sim	Não
AKN77409	mraY	Phospho-N-acetylmuramoyl-pentapeptide-transferase	92,0%	47	M	Sim	Sim
AKN76076	pmmC	Phosphoglucosamine mutase	81,7%	47	G	Sim	Não
AKN76603	thyA	Thymidylate synthase	87,7%	47	F	Sim	Não
AKN78192	trpB	Tryptophan synthase beta chain	93,0%	47	E	Sim	Sim
AKN76948	argF	Ornithine carbamoyltransferase	82,1%	46	E	Sim	Não
AKN76733	CulFRC58_0879	Hypothetical protein	86,3%	46	G	Não	Não
AKN76854	CulFRC58_1000	Hypothetical protein	83,3%	46	KT	Não	Não
AKN77091	CulFRC58_1237	Hypothetical protein	90,2%	46	S	Sim	Sim
AKN77488	CulFRC58_1634	Bifunctional RNase H/acid phosphatase	90,4%	46	GL	Sim	Sim
AKN76075	deoC1	Deoxyribose-phosphate aldolase	94,8%	46	F	Sim	Sim
AKN77597	dnaJ1	Chaperone protein DnaJ	89,2%	46	O	Sim	Não
AKN77790	purF	Amidophosphoribosyltransferase (ATase)	87,5%	46	F	Sim	Não
AKN77171	ruvA	Holliday junction ATP-dependent DNA helicase RuvA	86,3%	46	L	Sim	Não
AKN76726	uvrA1	UvrABC system protein A	85,6%	46	L	Sim	Não

(Continua)

(Continuação)

AKN76086	asd	Aspartate-semialdehyde dehydrogenase	87,2%	45	E	Sim	Não
AKN77495	cobD	Cobalamin biosynthesis protein CobD	87,5%	45	H	Sim	Não
AKN77443	cobU	Bifunctional adenosylcobalamin biosynthesis protein CobU	88,0%	45	H	Sim	Não
AKN76210	CuIFRC58_0356	Hypothetical protein	85,6%	45	S	Não	Não
AKN77290	CuIFRC58_1436	Cobalamin adenosyltransferase	85,5%	45	H	Sim	Não
AKN77368	hisA	1-(5-phosphoribosyl)-5-(5-phosphoribosylamino)methylideneamino imidazole-4-carboxamide isomerase	88,8%	45	E	Sim	Não
AKN77387	ilvA	Threonine dehydratase	75,5%	45	E	Sim	Não
AKN77793	nrdF2	Ribonucleoside-diphosphate reductase beta chain 2	96,6%	45	F	Sim	Sim
AKN77166	oppA	Oligopeptide-binding protein	84,6%	45	E	Não	Não
AKN76613	pcrA	ATP-dependent DNA helicase	87,9%	45	L	Sim	Não
AKN77947	pyrE	Orotate phosphoribosyltransferase (OPRT) (OPRTase)	96,1%	45	F	Sim	Sim
AKN76453	rbsK	Ribokinase	74,7%	45	H	Sim	Não
AKN76947	argD	Acetylornithine aminotransferase	86,0%	44	E	Sim	Não
AKN77957	clpB	ATP-dependent chaperone protein ClpB	83,3%	44	O	Sim	Não
AKN77289	cobB	Cobyrinic acid A,C-diamide synthase	81,4%	44	H	Sim	Não
AKN76121	CuIFRC58_0267	Hypothetical protein	94,4%	44	S	Sim	Sim
AKN77349	CuIFRC58_1495	Hypothetical protein	100,0%	44	S	Sim	Sim
AKN78203	CuIFRC58_2349	MviN-like protein	91,7%	44	S	Sim	Sim
AKN76816	hom	Homoserine dehydrogenase	90,5%	44	E	Sim	Sim
AKN77329	lepB	Signal peptidase I	84,2%	44	U	Sim	Não
AKN77226	lexA	LexA repressor	87,8%	44	K	Sim	Não
AKN76747	lpqW	Monoacyl phosphatidylinositol tetramannoside-binding protein	88,4%	44	E	Sim	Não
AKN76084	lysC	Aspartokinase	87,0%	44	E	Sim	Não
AKN76021	pat	Phenylalanine aminotransferase	86,8%	44	E	Sim	Não
AKN77910	pknG	Serine/threonine protein kinase	89,6%	44	KLT	Sim	Não
AKN78212	yidC1	Inner membrane protein translocase component YidC	94,7%	44	U	Sim	Sim
AKN76946	argB	Acetylglutamate kinase	84,9%	43	E	Sim	Não
AKN77444	cobT	Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase (NN:DBI PRT)	91,8%	43	H	Não	Sim
AKN77265	dapB	4-hydroxy-tetrahydrodipicolinate reductase (HTPA reductase)	89,0%	43	E	Sim	Não
AKN76008	glfT1	UDP-galactofuranosyl transferase	83,2%	43	S	Sim	Não
AKN77758	pgm	Phosphoglucomutase	89,9%	43	G	Sim	Não
AKN77803	purB	Adenylosuccinate lyase	91,0%	43	F	Sim	Não
AKN77131	pyrR	Bifunctional protein PyrR	88,6%	43	F	Não	Não
AKN76701	uppS1	Isoprenyl transferase	90,5%	43	H	Sim	Sim
AKN77412	ftsI	Penicillin-binding protein	87,3%	42	M	Sim	Não
AKN77283	proS	Proline--tRNA ligase	91,0%	42	J	Sim	Sim
AKN77141	aroC	Chorismate synthase (5-enolpyruvylshikimate-3-phosphate phospholyase)	90,0%	41	E	Sim	Sim
AKN77240	dapF	Diaminopimelate epimerase	88,0%	41	E	Sim	Não
AKN77912	eutD	Phosphotransacetylase	91,5%	41	C	Sim	Sim
AKN77659	folC	Bifunctional protein	91,0%	41	H	Sim	Sim
AKN76208	pflB	Formate acetyltransferase	88,8%	41	C	Sim	Não
AKN75899	pknB	Serine/threonine protein kinase	85,3%	41	KLT	Sim	Não
AKN76500	rsgA	Ribosome biogenesis GTPase	79,0%	41	S	Sim	Não
AKN76945	argJ	Arginine biosynthesis bifunctional protein	87,4%	40	E	Sim	Não
AKN77122	coaBC	Coenzyme A biosynthesis bifunctional protein (DNA/pantothenate metabolism flavoprotein)	88,5%	40	H	Sim	Não
AKN77430	glk	RGlucose kinase	84,8%	40	G	Sim	Não
AKN77045	hisE	Phosphoribosyl-ATP pyrophosphatase	97,4%	40	E	Sim	Sim
AKN76491	lpqB	Lipoprotein	78,9%	40	S	Sim	Não
AKN77241	miaA	tRNA dimethylallyltransferase	89,6%	40	J	Sim	Não

(Continua)

(Continuação)

AKN77126	pyrF	Orotidine 5'-phosphate decarboxylase	90,1%	40	F	Não	Sim
AKN77172	ruvC	Crossover junction endodeoxyribonuclease RuvC (Holliday junction nuclease RuvC) (Holliday junction resolvase RuvC)	81,8%	40	L	Sim	Não
AKN76193	slpA	Surface layer protein A	88,0%	40	S	Sim	Não
AKN76488	tmk	Thymidylate kinase	90,2%	40	F	Sim	Sim
AKN76437	upp	Uracil phosphoribosyltransferase	88,6%	40	F	Sim	Não
AKN77913	CulFRC58_2059	Ferredoxin/ferredoxin-NADP reductase	89,9%	39	E	Sim	Não
AKN77901	def2	Peptide deformylase (PDF)	88,9%	39	J	Sim	Não
AKN76214	gpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	91,5%	39	G	Sim	Não
AKN76231	hemD	Uroporphyrinogen-III synthase	79,6%	39	H	Sim	Não
AKN77365	hisl	Phosphoribosyl-AMP cyclohydrolase (PRA-CH)	85,4%	39	E	Sim	Não
AKN76439	manB1	Phosphoglucomutase/phosphomannomutase	81,6%	39	G	Sim	Não
AKN76477	manC	Mannose-1-phosphate guanylyltransferase	76,3%	39	JM	Sim	Não
AKN77406	murG	UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N- acetylglucosamine transferase	91,7%	39	M	Sim	Sim
AKN77014	rbpA	RNA polymerase-binding protein	82,6%	39	K	Sim	Não
AKN76515	uvrD2	DNA helicase, UvrD/REP type	87,3%	39	L	Sim	Não
AKN77491	cobC	Cobalamin-biosynthesis related aminotransferase	87,2%	38	E	Sim	Não
AKN77201	CulFRC58_1347	Hypothetical protein	81,5%	38	S	Não	Não
AKN77312	CulFRC58_1458	Amino acid kinase family	86,9%	38	F	Sim	Não
AKN78069	CulFRC58_2215	Prephenate dehydratase	89,7%	38	E	Sim	Não
AKN77263	dapA	4-hydroxy-tetrahydrodipicolinate synthase (HTPA synthase)	91,1%	38	E	Sim	Sim
AKN77869	folE	GTP cyclohydrolase 1	89,6%	38	H	Sim	Não
AKN75902	ftsW1	Cell division protein	90,2%	38	D	Sim	Sim
AKN77980	grpE	Protein GrpE (HSP-70 cofactor)	81,5%	38	O	Sim	Não
AKN76421	hmuU	Hemin import ATP-binding protein	88,7%	38	P	Sim	Não
AKN76850	mnmA	tRNA-specific 2-thiouridylase	87,8%	38	J	Sim	Não
AKN77411	murE	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2, 6- diaminopimelate ligase	88,7%	38	M	Sim	Não
AKN76907	mutT2	7,8-dihydro-8-oxoguanine-triphosphatase	88,0%	38	FG	Não	Não
AKN77280	nusA	Transcription termination protein NusA	83,6%	38	K	Sim	Não
AKN78164	oppCD2	Oligopeptide transport ATP-binding protein OppD	81,4%	38	P	Sim	Não
AKN77300	pbpB2	Secreted penicillin-binding protein	82,8%	38	M	Sim	Não
AKN77425	pknL	Serine/threonine protein kinase	86,6%	38	KLT	Sim	Não
AKN77649	proB	Glutamate 5-kinase	85,2%	38	E	Sim	Não
AKN77436	qcrC	Ubiquinol-cytochrome C reductase cytochrome C subunit	79,9%	38	C	Sim	Não
AKN77113	ribA	Riboflavin biosynthesis protein RibBA	82,3%	38	H	Sim	Não
AKN77363	trpC2	Indole-3-glycerol phosphate synthase	89,8%	38	E	Sim	Não
AKN76506	whiB3	Transcription regulator	93,5%	38	K	Não	Sim
AKN77219	ahpC	Alkyl hydroperoxide reductase subunit C	80,5%	37	O	Sim	Não
AKN76378	alr	Alanine racemase	86,8%	37	M	Sim	Não
AKN76780	aroE1	Shikimate 5-dehydrogenase	88,7%	37	E	Sim	Não
AKN77143	aroE2	Shikimate dehydrogenase	88,1%	37	E	Sim	Não
AKN76981	ciuC	Iron ABC transporter permease	88,8%	37	P	Sim	Não
AKN78048	cmtB	Trehalose corynomycyl transferase B	85,1%	37	S	Sim	Não
AKN77445	cobS	Cobalamin synthase	80,0%	37	H	Não	Não
AKN77979	dnaJ2	Chaperone protein DnaJ	85,1%	37	O	Sim	Não
AKN76137	dnaX2	DNA polymerase III subunit gamma/tau	80,7%	37	L	Sim	Não
AKN76640	galU	UTP--glucose-1-phosphate uridylyltransferase	92,7%	37	M	Sim	Sim
AKN77246	miaB	tRNA-2-methylthio-N(6)-dimethylallyl adenosine synthase	76,5%	37	J	Sim	Não
AKN77423	mptA	Alpha mannopyranosyltransferase	88,4%	37	S	Sim	Não
AKN75901	pbpA	Penicillin-binding protein A	88,4%	37	M	Sim	Não

(Continua)

(Continuação)

AKN76958	ppnK	NAD kinase	86,2%	37	H	Sim	Não
AKN77120	priA	Primosomal protein N	86,6%	37	L	Sim	Não
AKN76914	recG	ATP-dependent DNA helicase	62,6%	37	L	Sim	Não
AKN78046	cmtC	Trehalose corynomocolyl transferase	85,3%	36	M	Sim	Não
AKN76535	CulFRC58_0681	Methylmalonyl-CoA carboxyltransferase 5S subunit	90,9%	36	C	Sim	Sim
AKN77427	CulFRC58_1573	Hypothetical protein	92,0%	36	L	Sim	Sim
AKN77766	CulFRC58_1912	Cysteine synthase	90,4%	36	E	Sim	Não
AKN77119	def1	Peptide deformylase (PDF)	88,1%	36	J	Sim	Não
AKN77195	dxs	1-deoxy-D-xylulose 5-phosphate synthase	85,4%	36	H	Sim	Não
AKN77638	holA	DNA polymerase III delta subunit	83,9%	36	L	Sim	Não
AKN76853	ligA	DNA ligase	87,9%	36	L	Sim	Não
AKN78011	lysS1	Lysyl-tRNA synthetase	88,9%	36	J	Sim	Não
AKN77047	meth	Methionine synthase	86,3%	36	E	Sim	Não
AKN77789	purM	Phosphoribosylformylglycinamide cyclo-ligase	86,3%	36	F	Sim	Não
AKN77053	pyrD	Dihydroorotate dehydrogenase	87,1%	36	F	Sim	Não
AKN75857	recF	DNA replication and repair protein RecF	81,2%	36	L	Sim	Não
AKN76746	typA	GTP-binding protein TypA/BipA	86,8%	36	T	Sim	Não
AKN76043	tyrA	Prephenate dehydrogenase	89,8%	36	E	Sim	Não
AKN77107	whiA	Sporulation transcription regulator WhiA	92,8%	36	K	Sim	Sim
AKN76478	whiB2	Transcriptional regulator	87,9%	36	K	Não	Não
AKN77316	xerC	Tyrosine recombinase XerC	89,8%	36	D	Sim	Não
AKN76963	xerD	Tyrosine recombinase	91,1%	36	D	Sim	Sim
AKN76447	accBC	Acyl coenzyme A carboxylase	81,2%	35	I	Sim	Não
AKN77164	apt	Adenine phosphoribosyltransferase (APRT)	87,6%	35	F	Sim	Não
AKN77140	aroK	Shikimate kinase (SK)	94,7%	35	F	Sim	Sim
AKN77774	cat	Succinyl-CoA Coenzyme A transferase	86,8%	35	C	Não	Não
AKN76980	ciuB	Iron ABC transporter domain-containing permease protein	89,5%	35	P	Sim	Não
AKN77310	CulFRC58_1456	Phosphatidate cytidyltransferase	89,6%	35	I	Sim	Não
AKN77041	fhf	Formate--tetrahydrofolate ligase	90,5%	35	H	Sim	Não
AKN77452	gcvT	Aminomethyltransferase	89,2%	35	E	Sim	Não
AKN77028	helY	ATP-dependent DNA helicase	90,6%	35	L	Não	Sim
AKN77372	hisB	Imidazoleglycerol-phosphate dehydratase (IGPD)	87,6%	35	E	Sim	Não
AKN76993	merR3	MerR family transcriptional regulator	95,0%	35	K	Sim	Sim
AKN77895	mrpF1	Na(+)/H(+) antiporter subunit F	95,9%	35	P	Não	Sim
AKN77644	nadD	Nicotinate-nucleotide adenyltransferase	87,6%	35	H	Sim	Não
AKN76837	nucS	Endonuclease	89,3%	35	L	Não	Não
AKN77203	ppgK	Polyphosphate glucokinase	89,8%	35	GK	Sim	Não
AKN77434	qcrB	Ubiquinol-cytochrome C reductase cytochrome B subunit	85,2%	35	C	Sim	Não
AKN78041	accD	Propionyl-CoA carboxylase beta chain 5	82,1%	34	I	Sim	Não
AKN77497	aceE	Pyruvate dehydrogenase E1 component	87,3%	34	C	Sim	Não
AKN76982	ciuD	Iron ABC transporter ATP-binding protein	87,1%	34	HP	Sim	Não
AKN76117	CulFRC58_0263	Thioredoxin-related protein	88,4%	34	CO	Sim	Não
AKN77998	CulFRC58_2144	Iron(III) dicitrate transport system permease fecD	92,3%	34	P	Sim	Sim
AKN77199	dut1	Deoxyuridine 5'-triphosphate nucleotidohydrolase	86,6%	34	F	Sim	Não
AKN77359	glgP	Phosphorylase	89,8%	34	G	Não	Não
AKN76536	mmdA	Methylmalonyl-CoA decarboxylase subunit alpha	90,2%	34	I	Sim	Não
AKN76848	nifS	Cysteine desulfurase	88,3%	34	E	Sim	Não
AKN77874	ppa	Inorganic pyrophosphatase	88,9%	34	C	Sim	Não
AKN78110	pspA1	Phage shock protein A	95,5%	34	KT	Não	Sim
AKN77778	pstB	Phosphate import ATP-binding protein PstB	88,6%	34	P	Sim	Não
AKN77328	rnhB	Ribonuclease HII	93,8%	34	L	Sim	Sim
AKN77264	thyX	Thymidylate synthase ThyX (TS) (TSase)	93,0%	34	F	Sim	Sim
AKN78187	trpE	Anthranilate synthase component I	88,5%	34	E	Sim	Não
AKN76717	xseA	Exodeoxyribonuclease 7 large subunit	87,8%	34	L	Sim	Não
AKN77440	asnB	Asparagine synthetase	85,0%	33	E	Sim	Não
AKN77042	aspA	Aspartate ammonia-lyase	90,0%	33	E	Sim	Não

(Continua)

(Continuação)

AKN76979	ciuA	Iron ABC transporter substrate-binding protein	89,1%	33	P	Não	Não
AKN77785	CuIFRC58_1931	Hypothetical protein	90,7%	33	S	Não	Sim
AKN78060	CuIFRC58_2206	ABC-type cobalamin/Fe ³⁺ -siderophores transport system, permease component	89,7%	33	P	Sim	Não
AKN76758	dapE	Succinyl-diaminopimelate desuccinylase	82,7%	33	E	Sim	Não
AKN76843	fepC1	Ferric enterobactin transport ATP-binding protein	92,4%	33	P	Sim	Sim
AKN76862	fhuC	Iron 3+-hydroxamate import ATP-binding protein	85,3%	33	HP	Sim	Não
AKN76861	fhuG	Iron 3+-hydroxamate import system permease protein	90,4%	33	P	Sim	Sim
AKN76760	folP	Dihydropteroate synthase	84,6%	33	H	Sim	Não
AKN77868	folP2	Dihydropteroate synthase	90,4%	33	H	Sim	Sim
AKN77260	ftsK	Cell division protein FtsK	89,8%	33	D	Sim	Não
AKN77584	glyS	Glycine--tRNA ligase	87,2%	33	J	Sim	Não
AKN77642	gpmB2	Phosphoglycerate mutase	82,1%	33	G	Sim	Não
AKN76211	murB	UDP-N-acetylenolpyruvoylglucosamine reductase	84,0%	33	M	Sim	Não
AKN77987	oppCD1	Oligopeptide transport system permease protein oppC	83,7%	33	P	Sim	Não
AKN77806	purD	Phosphoribosylamine--glycine ligase	88,7%	33	F	Sim	Não
AKN77925	resA	Thiol-disulfide oxidoreductase	87,9%	33	O	Sim	Não
AKN77307	rimN	Dual-specificity RNA methyltransferase RlmN	92,1%	33	J	Sim	Sim
AKN77475	thrC	Threonine synthase	83,5%	33	E	Sim	Não
AKN78188	trpG	Anthranilate synthase component II	84,6%	33	EH	Sim	Não
AKN78007	ureA	Urease subunit gamma	94,3%	33	E	Sim	Sim
AKN76514	uvrD1	DNA helicase, UvrD/REP type	86,5%	33	L	Não	Não

Fonte: Autor

Legenda: Ramachandran, porcentagem de aminoácidos em regiões permitidas; *Degree*, grau de interação; COG, categoria funcional do COG (Cluster of Orthologous Genes); DEG, quantidade de organismos Sims no DEG; *Docking*, se a proteína participou do estudo de caso.