



UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PRÓ-REITORIA DE GRADUAÇÃO
DEPARTAMENTO DE BIOCÊNCIAS
CURSO DE BIOTECNOLOGIA

JULIANA NILSE CUNHA RODRIGUES

**Caracterização *in silico* das proteínas rCp01850 e rCp09720 de
*Corynebacterium pseudotuberculosis***

MOSSORÓ

2024

JULIANA NILSE CUNHA RODRIGUES

**Caracterização *in silico* das proteínas rCp01850 e rCp09720 de
*Corynebacterium pseudotuberculosis***

Monografia apresentada a Universidade Federal Rural do Semi-Árido como requisito para obtenção do título de Bacharel em Biotecnologia.

Orientador: Profa. Dra. Isadora Louise Alves da Costa Ribeiro Quintans

MOSSORÓ

2024

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

R696c Rodrigues, Juliana Nilse Cunha.
Caracterização in silico das proteínas rCp01850 e
rCp09720 de *Corynebacterium pseudotuberculosis* /
Juliana Nilse Cunha Rodrigues. - 2024.
39 f. : il.

Orientadora: Isadora Louise Alves da Costa
Ribeiro Quintans.
Monografia (graduação) - Universidade Federal
Rural do Semi-árido, Curso de Biotecnologia,
2024.

1. Linfadenite caseosa . 2. Bioinformática. 3.
Modelagem. I. Quintans, Isadora Louise Alves da
Costa Ribeiro, orient. II. Título.

Ficha catalográfica elaborada por sistema gerador automático em conformidade
com AACR2 e os dados fornecidos pelo(a) autor(a).
Biblioteca Campus Mossoró / Setor de Informação e Referência
Bibliotecária: Keina Cristina Santos Sousa e Silva
CRB: 15/120

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

JULIANA NILSE CUNHA RODRIGUES

**Caracterização *in silico* das proteínas rCp01850 e rCp09720 de
*Corynebacterium pseudotuberculosis***

Monografia apresentada a Universidade Federal Rural do Semi-Árido como requisito para obtenção do título de Bacharel em Biotecnologia.

Defendida em: 23 / 10 / 2024.

BANCA EXAMINADORA

Profa. Dra. Isadora Louise Alves da Costa Ribeiro Quintans (UFERSA)
Presidente

Ana Karolina Melo Oliveira Barros (UFERSA)
Membro Examinador

Mônica Ellen da Costa Soares (UFERSA)
Membro Examinador

AGRADECIMENTOS

Agradeço a Deus em primeiro lugar, por estar sempre comigo me dando forças e clareza e iluminando meu caminho na faculdade. Também agradeço a Nossa Senhora que me deu o conforto que meu coração necessitava e me ajudava todas as vezes que eu intercedia por ela.

Agradeço a minha família, em especial meus pais, José Nilson e Maria José, e minha irmã Greicy, por todos os conselhos, amor, acolhimento, discernimento e suporte durante toda a vida, e em especial essa jornada. Agradeço à minha sobrinha Fernanda e aos meus primos Luana e Willy por terem me apoiado com tanto amor e carinho em várias situações vividas nesses últimos cinco anos. Amo muito todos vocês.

Agradeço à minha orientadora, Isadora, por me acolher em seu laboratório, por me aconselhar e me orientar nesse projeto. E por ter sido uma professora que me ensinou muito ao longo da graduação, em especial me lembrando de não desistir dos meus sonhos.

Agradeço a todos os professores que eu já tive o prazer em assistir suas aulas, todos me ensinaram muito além do conteúdo proposto pela disciplina. Agradeço especialmente aos que tornaram a sala de aula um local acolhedor para mim e para meus colegas e criaram conosco um vínculo de amizade e companheirismo.

Agradeço à Banca Examinadora por ter aceitado fazer parte desse momento tão especial em minha vida e por todas as correções feitas no presente trabalho.

E por fim, agradeço aos meus amigos, aos que estão comigo desde a época de escola, Guedes e Raj. Aos amigos que fiz ao longo do caminho, em especial Lenise que viveu comigo toda a experiência da graduação, passando noites em claro fazendo trabalhos ou estudando pra provas, celebrando cada pequena ou grande conquista, rindo em nossos tempos livres e chorando juntas quando a dificuldade vinha; e Chagas, Karol, Bia, Mônica, Paulo, Gustavo e Adrielle por todos os momentos que passamos juntos. Aos meus colegas de laboratório que me deram suporte e deixaram o ambiente de trabalho incrível. Aos meus amigos que fizeram muita diferença na minha vida nos últimos cinco anos: Fernando, Renata, Pedro, Bruno e Alessandra. Aos meus queridos amigos da EJNS, que não só comemoraram comigo, como me abraçaram e me acolheram, me deram conselhos e sempre me lembraram que eu tenho uma segunda família ao meu lado e que Deus sempre está comigo.

RESUMO

A Linfadenite Caseosa (LC) é uma doença causada pela bactéria *Corynebacterium pseudotuberculosis* (CP) que acomete pequenos ruminantes, causando abscessos na pele e/ou linfonodos dos animais, podendo ser ou não detectável exteriormente. A LC tem impacto significativo na economia, registrando um prejuízo anual de cerca de trinta e cinco milhões de reais na agropecuária brasileira, afetando principalmente os produtores de caprinos e ovinos do Nordeste do Brasil. E apesar de existir vacina, esta doença é por muitas vezes subnotificada. Neste contexto, a bioinformática emerge como ferramenta fundamental para o desenvolvimento de vacinas e testes diagnósticos da LC. Esse estudo visa caracterizar e modelar tridimensionalmente duas proteínas recombinantes de CP, a rCp01850 e a rCp09720 usando ferramentas de bioinformática. Os principais resultados obtidos nesse estudo foram que ambas as proteínas são termo resistentes, estáveis, hidrofílicas e tem seu ponto isoelétrico próximo a pH 6,0. Além disso, foi observado que essas proteínas são insolúveis em água sugerindo que outro solvente seja utilizado neste caso. Adicionalmente, estas proteínas foram consideradas potenciais antígenos não alergênicos, indicando seu potencial na produção de vacinas recombinantes contra a LC. Por fim, as estruturas tridimensionais foram preditas e os modelos apresentaram alta confiabilidade (80%), podendo ser usadas na realização de ancoragem com complexos de interesse e simulações moleculares. Em conclusão, os resultados destacam o potencial dessas proteínas como alvo promissor na produção de vacinas contra a linfadenite caseosa.

Palavras-chave: Linfadenite caseosa. Bioinformática. Modelagem.

ABSTRACT

Caseous lymphadenitis (CLA) is a disease caused by the bacterium *Corynebacterium pseudotuberculosis* (CP) that affects small ruminants, causing abscesses in the skin and/or lymph nodes of the animals, which may or may not be detectable externally. CLA has a significant impact on the economy, with an annual loss of approximately thirty-five million reais in Brazilian agriculture, mainly affecting goat and sheep producers in the north-east of Brazil. And although there is a vaccine, this disease is often underreported. In this context, bioinformatics is emerging as a fundamental tool for the development of vaccines and diagnostic tests for CLA. This study aims to characterize and create three-dimensional models of two recombinant CP proteins, rCp01850 and rCp09720, using bioinformatics tools. The main results obtained here were that both proteins are thermo-resistant, stable, hydrophilic and have their isoelectric point close to pH 6.0. Furthermore, it was observed that these proteins are insoluble in water, suggesting that another solvent should be used in this case. In addition, these proteins were considered potential non-allergenic antigens, indicating their potential in the production of recombinant vaccines against CLA. Finally, the three-dimensional structures were predicted and the models showed high reliability (80%), which can be used to perform docking with complexes of interest and molecular simulations. In conclusion, the results highlight the potential of these proteins as promising targets to produce vaccines against caseous lymphadenitis.

Keywords: Caseous lymphadenitis. Bioinformatics. 3D modeling.

LISTA DE FIGURAS

Figura 1 - Diagrama de caracterização de proteínas.....	18
Figura 2 - Fluxograma da metodologia	23
Figura 3 – Estrutura tridimensional da proteína rCP01850	31
Figura 4 – Estrutura tridimensional da proteína rCP09720.....	32

LISTA DE GRÁFICOS

Gráfico 1 – Resultados de solubilidad.....	29
--	----

LISTA DE QUADROS

Quadro 1	–	Transformação sequência de aminoácidos para sequência FASTA	26
Quadro 2	–	Caracterizações físico-químicas da proteína	27

LISTA DE ABREVIATURAS

CP	<i>Corynebacterium pseudotuberculosis</i>
LC	Linfadenite Caseosa
3D	Tridimensional
ELISA	Enzyme-Linked Immunosorbent assay
GRAVY	Grand Average of Hydropathicity Index
pLDDT	predicted Local Distance Difference Test
pI	Ponto Isoelétrico

SUMÁRIO

1. INTRODUÇÃO	14
2. GESTÃO DO CONHECIMENTO	16
2.1 LINFADENITE CASEOSA E <i>CORYNEBACTERIUM PSEUDOTUBERCULOSIS</i>	16
2.2 VACINAS E PROTEÍNAS RECOMBINANTES	17
2.3. CARACTERIZAÇÃO FÍSICO-QUÍMICA	18
3. OBJETIVOS	22
3.1 OBJETIVO GERAL.....	22
3.2 OBJETIVOS ESPECÍFICOS	22
4. METODOLOGIA.....	23
4.1 ESCOLHA DAS SEQUÊNCIAS E TRANSFORMAÇÃO EM FORMATO FASTA.....	23
4.2 CARACTERIZAÇÃO DAS PROTEÍNAS RECOMBINANTES	23
4.2.1 Caracterização físico-química	24
4.2.2 Teste de solubilidade.....	24
4.2.3 Teste de alergenicidade.....	24
4.2.4 Análise do potencial antigênico	24
4.3 MODELAGEM DA ESTRUTURA TRIDIMENSIONAL.....	24
4.3.1 Captura de imagem das estruturas	25
4.4. ANCORAGEM.....	25
5. RESULTADOS E DISCURSÃO	26
5.1 SEQUÊNCIA FASTA.....	26
5.2 CARACTERIZAÇÃO DAS PROTEÍNAS RECOMBINANTES	27
5.2.1 Caracterização físico-química	27
5.2.2 Solubilidade.....	28
5.2.3 Potencial alergênico	30
5.2.4 Potencial Antigênico	30
5.3 ESTRUTURAS TRIDIMENSIONAIS	30
5.3.1. I-Tasser.....	30
5.3.2 AlphaFold.....	31

5.4. ANCORAGEM.....	33
6. CONSIDERAÇÕES FINAIS.....	34
7. REFERÊNCIAS	35

1. INTRODUÇÃO

A Linfadenite Caseosa é uma doença que acomete pequenos ruminantes e é causada pela bactéria *Corynebacterium pseudotuberculosis*. A LC é caracterizada pelo aparecimento de abscessos nos linfonodos do animal, gerando lesões internas ou externas. Quando os abscessos se rompem, liberam pus, podendo contaminar outros animais e objetos de forma direta ou indireta (Alves; Pinheiro; Pires, 1997).

A doença é bastante significativa a nível mundial, e no Brasil, o Nordeste é a região mais afetada. A LC causa grande impacto econômico, resultando na diminuição na produção de lã, carne e leite. Os animais infectados são abatidos e têm suas carcaças descartadas, inviabilizando seu uso (Faccioli-Martins; Alves; Pinheiro, 2014).

O diagnóstico inicial da doença é clínico, com base em exame físico que consiste no toque e observação de nódulos e/ou feridas. Posteriormente, podem ser realizados testes sorológicos, como o ELISA (Enzyme Linked Immuno Sorbent Assay), que detecta a presença de anticorpos específicos no sangue do animal (Aydin, 2015; Baird; Fontaine, 2007).

Atualmente as principais medidas de profilaxia são a separação dos animais contaminados e o uso de vacinas. Neste contexto, as vacinas de subunidade recombinante surgem como uma possível ferramenta eficaz e segura no controle da doença (Faccioli-Martins; Alves; Pinheiro, 2014; Guimarães *et al.*, 2011).

Com isso, as ferramentas de bioinformática podem acelerar e otimizar o desenvolvimento de imunobiológicos. O uso de testes *in silico*, como predição de epítomos, descoberta de sítios de interesse e a sequência do genoma de antígenos, antes da etapa experimental em bancada, economiza tempo e recursos, já que a permite o conhecimento de dados básicos para tal. (Guissoni; Cardoso, 2019).

A busca *in silico* por epítomos que possam ser utilizados em construções vacinais requer a caracterização das proteínas alvo selecionadas quanto ao seu peso molecular, meia-vida, nível de instabilidade, ponto isoelétrico, solubilidade, entre outros. Além disso, é possível propor modelos tridimensionais (3D) das proteínas que ainda não foram preditas experimentalmente, facilitando posteriormente a predição de epítomos. Pode-se também realizar a ancoragem das proteínas com ligantes de vacina ou anticorpos, verificando a resposta imunológica daquela sequência de epítopo (Droppa-Almeida; Franceschi; Padilha, 2018).

A modelagem tridimensional de proteínas é uma técnica da bioinformática que utiliza uma série de proteínas já conhecidas, e que já teve sua forma e localização de aminoácidos definidos de forma experimental, para determinar, por meio de programas computacionais, a localização dos átomos da proteína de interesse análoga (Alencastro, 2003).

Com isso, temos que o uso das ferramentas de bioinformática é um aliado de suma importância na descoberta de novos alvos para imunobiológicos e ao ser associada a experimentos *in vitro* acelera o desenvolvimento da pesquisa de forma eficiente e com menos custos.

Sendo assim, o presente visa a caracterização *in silico* das proteínas rCp01850 e rCp09720 de *Corynebacterium pseudotuberculosis*, que são proteínas previamente preditas por integrantes do nosso grupo de pesquisa como candidatas relevantes para a construção vacinal, visto que estas podem estar envolvidas em mecanismos de virulência e/ou sobrevivência bacteriana. Desta forma, a caracterização físico-química *in silico*, teste de potencial alergênico e antigênico e a modelagem tridimensional, são ferramentas essenciais para uma análise detalhada de características destas proteínas que irão influenciar na qualidade da vacina desenvolvida.

2. GESTÃO DO CONHECIMENTO

2.1 Linfadenite Caseosa e *Corynebacterium pseudotuberculosis*

A Linfadenite Caseosa é uma doença que afeta principalmente ovinos e caprinos, sendo causada pela bactéria *Corynebacterium pseudotuberculosis* (CP). A doença provoca lesões cutâneas e nos linfonodos, que podem ser visíveis externamente ou manifestar-se na forma visceral, que dificulta o diagnóstico (Fontaine; Baird, 2008).

A *C. pseudotuberculosis* que é uma bactéria pleomórfica, não móvel, intracelular facultativa e gram-positiva, pertencente à família *Actinomycetaceae* (Bettini *et al.*, 2022). Sua temperatura ideal de crescimento é em torno de 37°C, podendo proliferar tanto em condições aeróbias quanto anaeróbias (Baird; Fontaine, 2007; Dorella *et al.*, 2006).

A bactéria apresenta dois biovars que acometem diferentes grupos de animais: o biovar *ovis* e o biovar *equi*. Biovar *ovis*, acomete ovelhas e cabras, causando a linfadenite caseosa, caracterizada pela formação de abscessos tanto na pele quanto em órgãos internos, como pulmões. Já a biovar *equi*, afeta equinos e bovinos, causando a linfangite ulcerosa, caracterizada pela formação de abscessos nos linfonodos e na região abdominal dos animais infectados (Odhah *et al.*, 2022).

A identificação de um animal doente é feita inicialmente por exames clínicos de visualização e toque, avaliando a presença de lesões externas ou na camada subcutânea. Além disso, o diagnóstico pode ser complementado através de exames sorológicos, como o teste de ELISA (teste ELISA (*Enzyme-Linked Immunosorbent assay* ou Ensaio de Imunoabsorção Enzimática), especialmente quando o animal não apresenta lesões visíveis. Contudo, em alguns casos essa identificação só é possível após a morte do animal, quando a necropsia revela lesões internas em seus órgãos e tecidos (Alves; Pinheiro; Pires, 1997).

Quando um animal diagnosticado para LC morre, sua carne, pele e carcaça não podem ser aproveitadas, devido aos riscos à saúde pública, causando assim prejuízos importantes para o agronegócio (Sá *et al.*, 2018). Em 2017, a linfadenite caseosa gerou perdas econômicas de aproximadamente 35 milhões de reais para os produtores de ruminantes no Brasil, sendo o nordeste a região mais afetada (Barnabé, 2018).

Por se tratar de uma enfermidade bacteriana, a LC se espalha rapidamente, podendo contaminar não apenas os animais, mas também o ambiente ao redor, como o solo, bebedouros e as áreas de alimentação e descanso. Até mesmo moscas podem ser veículos de contaminação da doença, uma vez que entrem em contato com o pus ou dejetos contaminados (Guimarães *et al.*, 2011).

O controle da LC geralmente envolve a separação dos animais doentes dos saudáveis, mas essa medida pode não ser totalmente eficaz, já que um animal assintomático pode estar infectado. Por tanto, o diagnóstico precoce é essencial. O tratamento inclui cauterização das lesões externas para reduzir a liberação da secreção infecciosa, diminuindo a propagação da doença no rebanho, embora não cure o animal (Alves; Santiago; Pinheiro, 2007).

A profilaxia do rebanho é feita por meio da vacinação. A primeira vacina testada para a doença foi na década de 60 e desde então, vários estudos têm sido desenvolvidos, em âmbito nacional e internacional, visando o desenvolvimento de vacinas mais eficazes (Faccioli-Martins; Alves; Pinheiro, 2014). Atualmente, há três vacinas brasileiras em comercialização e todas precisam de revacinação anual. A Glanvac® 6 tem como principal componente toxóide de CP, e a LinfoVac® e a Vacina 1002® tem como principal componente bactérias vivas atenuadas (Faccioli-Martins; Yoshida, 2014). Entretanto, estas vacinas por muitas vezes não oferecem proteção completa e necessitam de doses de reforço. Estudos com proteínas recombinantes tem demonstrado um grande potencial na produção de vacinas contra LC (Faccioli-Martins; Yoshida, 2014). Sendo assim, as pesquisas com a CP devem avançar para a produção de vacinas mais eficientes.

2.2 Vacinas e Proteínas Recombinantes

A primeira vacina criada surgiu no século XVIII, contra a varíola, uma doença que assolava a população mundial na época. Desde então, inúmeros avanços tecnológicos ocorreram e o desenvolvimento de vacinas se tornou uma prioridade. Atualmente, um dos principais avanços no desenvolvimento de vacinas é a tecnologia do DNA recombinante (Pinto; Da-Cruz, 2011).

As proteínas recombinantes são obtidas a partir da tecnologia DNA recombinante, uma técnica da biologia molecular em que um fragmento do material genético de interesse é isolado utilizando enzimas de restrição, sendo posteriormente inserido em um vetor de expressão, geralmente um plasmídeo. Se o vetor recombinante for de origem plasmidial, essa construção gênica será inserida no organismo hospedeiro, no caso, uma bactéria através da técnica de

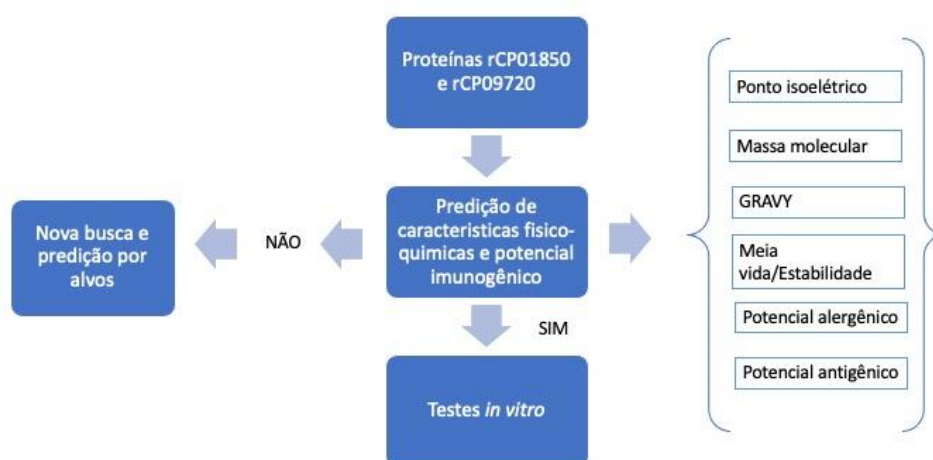
transformação. As bactérias transformadas irão produzir a proteína recombinante de interesse (Rodrigues, 2003). Essas proteínas também podem apresentar epítopos, também conhecidos como determinantes antigênicos, que são fragmentos específicos de uma molécula capazes de desencadear uma resposta imune (Montassier, 2018).

A predição de epítopos capazes de conferir respostas imunes eficazes para a construção de vacinas baseadas na tecnologia do DNA recombinante deve ser fundamentada em um amplo conhecimento e caracterização das sequências alvo. Nesse contexto, a bioinformática é uma importante aliada, com ferramentas como a caracterização físico-química, análise de alergenidade, e o docking molecular desempenhando papéis essenciais, que serão discutidos nos próximos tópicos.

2.3. Caracterização físico-química

A caracterização físico-química de uma proteína realizada por predição *in silico*, envolve o uso de ferramentas para cálculo de parâmetros cruciais para o desenho de construções vacinais, como hidrofobicidade, meia vida, ponto isoelétrico, entre outros. Cada uma destas propriedades vai influenciar na viabilidade e eficácia da vacina, para a produção de uma resposta imune duradoura. O conhecimento prévio sobre as propriedades das sequências vai auxiliar na seleção de alvos promissores e viáveis para os experimentos de bancada. A figura 1 mostra o fluxo de trabalho da caracterização de uma proteína recombinante à suas aplicações *in vitro*.

Figura 1: Diagrama de caracterização de proteínas



A análise da hidrofobicidade é crucial para identificar proteínas adequadas para o desenvolvimento para vacinas e fármacos, uma vez que as formulações exigem que essas proteínas permaneçam solúveis, em água ou outro solvente. A hidrofobicidade também pode estar relacionada com a estrutura tridimensional da proteína e sua interação com o sistema imune. Sendo assim, um melhor design vacinal irá priorizar a otimização da distribuição das regiões hidrofóbicas da proteína, otimizando sua solubilidade, sem comprometer seu potencial imunogênico. A análise de hidrofobicidade é baseada no índice GRAVY (Grand Average of Hydropathicity Index ou Média do índice de hidropaticidade), onde valores positivos indicam que a proteína é hidrofóbica (não tem afinidade com água), enquanto valores negativos indicam que a proteína é hidrofílica (possui afinidade com água) (Garbis; Lubec; Fountoulakis, 2005).

Outro parâmetro importante no desenvolvimento de uma vacina de proteína recombinante, é sua meia vida *in vitro*, ou seja, o tempo que ela permanece ativa. Proteínas com meia vida longa, podem ter sua eficiência de resposta aumentada, evitando que a construção vacinal precisar ser re-administrada com frequência. Sendo assim, a determinação do tempo de meia-vida de uma proteína pode intervir na escolha de um epítopo, já que quanto maior a meia-vida da proteína, mais eficiente será o imunobiológico (Kontermann, 2011). Além disso, pequenas modificações na sequência da proteína podem aumentar sua meia vida, sem interferir na sua atividade (Vanderschuren *et al.*, 2022).

A análise do nível de instabilidade é essencial no desenvolvimento de vacinas de proteínas recombinantes, pois esta ferramenta avalia diversos tipos de instabilidade, como desnaturação, desnovelamento, degradação, entre outros. Esse parâmetro é utilizado para escolher epítopos que sejam viáveis e estáveis, aumentando a probabilidade de sucesso no desenvolvimento de um produto biotecnológico. Quanto mais estável o epítopo, maior a chance de que a pesquisa resulte em um produto viável (Deller; Kong; Rupp, 2016).

A determinação do peso molecular também pode auxiliar na escolha de alvos para a produção de vacinas recombinantes, visto que desempenha um papel importante na expressão, imunogenicidade, estabilidade, e no processamento da proteína pelo sistema imune (Lampe *et al.*, 2020). Estudos já indicaram que um peso molecular aceitável para a produção e purificação de uma proteína recombinante, deve ser inferior a 110 kDa (Barh *et al.*, 2013).

Por fim, a identificação do ponto isoelétrico (pI) é um parâmetro relevante na formulação de vacinas, pois influencia na solubilidade, interação com adjuvantes e

componentes do sistema imune. O pI é o valor de pH no qual a carga líquida da proteína é igual a zero, ou seja, está em equilíbrio (Righetti, 2004).

Concluindo, a predição das características físico-químicas das proteínas alvo para o desenvolvimento de vacinas baseadas em proteínas recombinantes é essencial, visto que facilita a escolha de epítomos com parâmetros “ótimos”, ou, permite que modificações na sequência codificante sejam feitas para otimizar a funcionalidade e estabilidade dos alvos selecionados, tanto nos experimentos *in vitro* quanto *in vivo*.

2.4. Análise de alergenicidade e Potencial antigênico

A análise da alergenicidade permite prever se a proteína pode causar reações adversas, como intolerância ou alergia, uma vez que essas reações produzem anticorpos diferentes dos gerados na resposta imunológica (Knippels; Penninks, 2005). O potencial antigênico é uma ferramenta eficiente na busca por um antígeno de interesse, já que o software prediz se a proteína estudada é um possível antígeno sem a necessidade de testes *in vitro* com o patógeno. Os testes de potencial antigênico escolhem os melhores epítomos capazes de gerar boas respostas imunológicas, podendo ser usados como antígenos em vacinas (Dalsass *et al.*, 2019). Quando associados, torna-se possível determinar se a proteína com potencial antigênico pode ser alergênica, evitando assim seu uso.

2.5. Modelagem tridimensional de proteínas

A modelagem tridimensional (3D) de proteínas é uma técnica crucial para predição da estrutura da proteína usando ferramentas de bioinformática. Esse método é importante pois a determinação das estruturas usando métodos experimentais é um processo caro e laborioso. A estrutura 3D de uma proteína está diretamente relacionada à sua função, e sua determinação é essencial para compreender sua biologia, mecanismos de ação, e também de interação com outras moléculas, como por exemplo anticorpos, sendo essencial para o desenvolvimento de vacinas. O método mais preciso para determinar a estrutura de uma proteína é a cristalografia de raio X, porém essa técnica é muito cara e complexa. A modelagem 3D surge para prever a estrutura da proteína a partir de softwares específicos. Essa técnica pode ser dividida em dois tipos: modelagem por homologia ou *ab initio*. (Alencastro, 2003).

A modelagem por homologia, ou comparativa, utiliza uma proteína de estrutura tridimensional definida experimentalmente, e depositada em um banco de dados, como molde para proteína de interesse. As proteínas comparadas (molde e alvo) devem apresentar

homologia em suas sequências, indicando assim uma boa correspondência entre suas estruturas tridimensionais. Os passos para a modelagem por homologia são: 1) escolha da proteína de interesse e da proteína molde; 2) alinhamento global entre as proteínas molde e alvo; usando o software escolhido; 3) construção do modelo 3D; e 4) avaliação do modelo criado (Santos Filho; Alencastro, 2003). Se a estrutura estiver compatível com as leis da física e da química, o modelo é validado. Caso contrário, o processo é repetido até que uma estrutura satisfatória seja obtida. (DINIZ; CANDURI, 2017).

Já a modelagem *ab initio* é a técnica que dispensa o uso de uma proteína molde, baseando-se apenas em softwares computacionais. Essa ferramenta é muito útil para regiões que não podem ser modeladas por homologia. Os programas de modelagem *ab initio* utilizam um sistema de pontuação para selecionar o melhor modelo criado. No entanto, esta técnica exige um alto processamento computacional, e o sistema de pontuação pode não ser eficaz e sensível o suficiente para prever estruturas que sejam biologicamente ativas (Stein; Mosca; Aloy, 2011).

A partir das estruturas preditas é possível realizar diferentes tipos de simulações computacionais e experimentos, como o docking molecular, otimizando os experimentos em bancada. Dessa forma, essa técnica acelera o desenvolvimento de vacinas e imunobiológicos.

3. Objetivos

3.1 Objetivo geral

Caracterizar as sequências das proteínas rCp01850 e rCp09720 da *Corynebacterium pseudotuberculosis*, usando ferramentas de bioinformática.

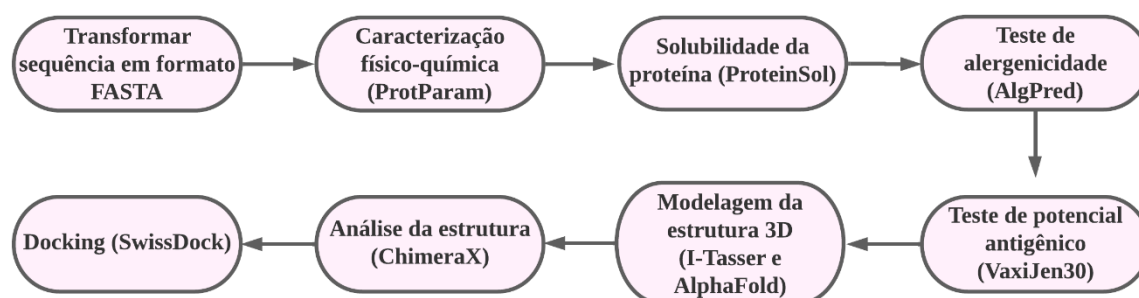
3.2 Objetivos específicos

- 1) Caracterizar as proteínas rCp01850 e rCp09720 quanto hidrofobicidade, meia-vida, peso molecular, nível de instabilidade, ponto isoelétrico e solubilidade utilizando ferramentas de bioinformática;
- 2) Analisar a alergenicidade e o potencial antigênico das proteínas recombinantes;
- 3) Realizar modelagem tridimensional das proteínas recombinantes.

4. Metodologia

A metodologia utilizada no presente trabalho foi a metodologia adaptada de (Silva *et al.*, 2021) e está descrita no diagrama da figura 2.

Figura 2: Diagrama da metodologia.



Fonte: Autoria própria com auxílio do programa Lucidchart

4.1 Escolha das sequências e transformação em formato FASTA

As sequências das proteínas rCp01850 e rCp09720 da *Corynebacterium pseudotuberculosis*, utilizadas neste trabalho, foram gentilmente cedidas pelo Professor Dr. Francisco Silvestre Brilhante Bezerra, responsável pelo Laboratório de Vacinologia e Imunologia Aplicada (LAVIA) na Universidade Federal Rural do Semi Árido.

As sequências foram recebidas em formato de sequência peptídica de três letras, sendo necessário fazer a conversão para formato FASTA, que é o formato utilizado nos programas escolhidos para caracterização. O programa utilizado para a conversão foi o SMS Three To One (https://www.bioinformatics.org/sms2/three_to_one.html) (Stothard, 2000).

4.2 Caracterização das proteínas recombinantes

Com as sequências no formato FASTA, foi realizado o processo de caracterização das proteínas recombinantes. A caracterização foi dividida em quatro experimentos, sendo eles: análise físico-química, teste de solubilidade, teste de alergenicidade e teste de antigenicidade, que serão descritos a seguir.

4.2.1 Caracterização físico-química

O programa escolhido para a caracterização físico-química foi o ProtParam (<https://web.expasy.org/protparam/>) (Wilkins *et al.*, 1998). Nesse experimento foi predito o peso molecular, ponto isoelétrico (pI), tempo de meia-vida, nível de instabilidade e hidrofobicidade.

4.2.2 Teste de solubilidade

O programa Protein-Sol (<https://protein-sol.manchester.ac.uk/>) (Hebditch *et al.*, 2017) foi o programa escolhido para a determinação da solubilidade das proteínas alvo. Para esse experimento, a sequência da proteína rCP01850 precisou ser adaptada para conter apenas aminoácidos canônicos, utilizando os dados da universidade de Birkbeck (https://www.cryst.bbk.ac.uk/education/AminoAcid/the_twenty.html).

4.2.3 Teste de alergenicidade

O teste de alergenicidade foi realizado pelo software AlgPred (<http://crdd.osdd.net/raghava/algpred/>) (Saha; Raghava, 2006), utilizando o modo “SVM Method Based on Amino Acid Composition”.

4.2.4 Análise do potencial antigênico

Nessa análise foi utilizado o programa VaxiJen30 (<https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (Doytchinova; Flower, 2007) no modo “Bacteria”.

4.3 Modelagem da estrutura tridimensional

Foram comparados dois programas de modelagem de proteína. A primeira modelagem foi realizada no I-Tasser (<https://zhanggroup.org/I-TASSER/>) (YANG; ZHANG, 2015). E segunda modelagem foi feita no programa AlphaFold Collab (<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>) (Jumper *et al.*, 2021), disponível na plataforma do Google Collab.

4.3.1 Captura de imagem das estruturas

Utilizando o programa UCSF ChimeraX (Pettersen *et al.*, 2004) versão 1.7.1 foram feitas as imagens das estruturas preditas pelo AlphaFold. Além disso, foram também editadas as cores da proteína para estar na paleta de confiabilidade do AlphaFold, disponibilizada no próprio programa, utilizando o comando “color bfactor palette alphafold”.

4.4. Ancoragem

Utilizando o SwissDock (<http://old.swissdock.ch/>) (Grosdidier; Zoete; Michielin, 2011), as duas proteínas recombinantes foram ancoradas ao MHC ovino, recuperado no UniProt (<https://www.uniprot.org/>), sob o código “Q30838_SHEEP”.

5. Resultados e Discussão

5.1 Sequência FASTA

As sequências de aminoácidos de três letras das proteínas recombinantes rCP01850 e rCP09720 de *Corynebacterium pseudotuberculosis* foram transformadas para formato FASTA. E ambas as sequências podem ser observadas no quadro 1 abaixo.

Quadro 1: Sequência de aminoácidos original e sequência do tipo FASTA gerada pelo SMS Three to One. Autoria própria.

Proteínas	Sequência de aminoácidos	Sequência FASTA
rCP01850	Met Arg Gly Ser His His His His His His Leu Glu Gly Ser Gln Pro Val Leu Gln Val Pro Ser Asp Ser Ser Val Gly Ala Ser Leu Ala Ser Pro Ser Val Leu His Pro Gly Ala Pro Ile Pro Gln Pro Phe Gly Pro Ala Asp Tyr Ile Gly Tyr Ile Ser Asp Ile Ser Ser Tyr Ser Gly Gly Ile Tyr Tyr Asp Val Val Ser Asn Phe Glu Asp Leu Arg Gln Asn His Pro Lys Ala Met Glu Glu Asn Ser Ala Thr Ser Ile Arg Ile Asn Asn Glu Ala Thr Ala Glu Ser Val Ala Arg Ala Gln Arg Asp Ala Leu Thr Glu His Asp Gly Val Leu Val Ala Phe Ser Asp Ala Leu Gly Pro Arg Leu Gly Glu His Leu Arg Thr Ala Leu Ala Glu Asn Arg Leu Pro Lys Leu Gln Met Leu Leu Gly Gly Trp Thr Ala Arg Ala Gly Gly Leu Ala Ser Ser Thr Phe Val Glu Lys Asn Ile Phe Arg Asn Pro Arg Pro Tyr Ile Ala His Ala Gly Ser Ile Arg Gln Tyr Arg Val Ser His His Asp Leu Tyr Pro Thr Ser Gly Ser Phe Pro Ser Gly His Thr Asn Gln Ala Thr Trp Thr Thr Ala Leu Trp Ala Tyr Met Leu Pro Glu Phe Ser Ala Gln Ile Leu Asp Arg Gly Ser Glu Ala Gly Phe Asn Arg Val Val Leu Gly Val His Tyr Pro Leu Asp Val Ile Gly Gly Arg Met Thr Gly Asn Ala Ala Ala Ala Asp Arg Leu Asn Asp Pro Lys Met Arg Ile Ala Leu Asp Ala Ala Ala Ala Glu Leu Arg Ala Glu Leu Glu Trp Arg Cys Gly Lys Ser Leu Leu Glu Cys Ala Thr Asp Thr Arg Ser Lys Glu Gln Ala Val Gln Lys Xaa	MGRSHHHHHHLEGSQP VLQVPSDSSVGASLASPS VLHPGAPIQPFGPADYI GYISDISSYSGGIYYDVV SNFEDLRQNHPKAMEEN SATSIRINNEATAESVAR AQRDALTEHDGVLVAFS DALGPRLGEHLRTALAE NRLPKLQMLLGGWTAR AGGLASSTFVEKNIFRNP RPYIAHAGSIRQYRVSHH DLYPTSGSFPSGHTNQAT WTTALWAYMLPEFSAQI LDRGSEAGFN RVVLGVH YPLDVIGGRMTGNAAAA DRLNDPKMRIAALDAAAA ELRAELEWRCGKSLLEC ATDTRSKEQAVQKX
rCP09720	Met His His His His His His Leu Glu Gly Ser Gly Pro Arg Asp Trp Leu Arg Pro Asp Ala Thr Gly Thr Cys Glu Trp Asp Ala Ala Asn Tyr Trp Val Gln Arg Cys Asp Val Ala Ser Pro Ala Met Gly His Asn Val Thr Val Gln Ile Gln Pro Ala Gly Arg Gly Gly Asn Ala Ala Leu Tyr Leu Leu Asp Gly Ala Arg Ala Ser Asp Ala Ala Asn Ala Tyr Thr Thr Asp Ala Ser Ala Gln Ser Met Phe Val Asp Asn Asn Ile Thr Leu Val Met Pro Val Gly Gly Ala Gly Ser Phe Tyr Thr Asp Trp Ala Ala Pro Ser Gly Ile Gly Ser Val Gly Gln Val Phe Lys Trp Glu Thr Phe Leu Thr Gln Glu Leu Pro Ala Tyr Leu Glu Ala His Phe Gly Val Ala Arg Asn Asn Tyr Ser Val Gly Gly Leu Ser Met Gly Ala Thr Ala Ala Met Asn Leu Ala Ala Leu His Pro Glu Gln Phe Arg Gln Val Leu Ser Tyr Ser Gly Tyr Leu Ala Met Thr Ala Pro Gly Met Tyr Ser Leu Leu Gly Leu Ala Leu Leu Glu Val Gly Gly Leu Ser Ile Asn Ser Met Tyr Gly Ser Phe Phe Ser Pro Arg Arg Ile Glu Leu Asp Pro Leu Phe Asn Met Arg Gly Leu Asp Gly Lys Asp Val Tyr Val Ser Ala Ala Ser Gly Ala Trp Gly Gly Pro Asp Ile Glu Lys Tyr Ser Phe Lys Asp Arg Val Asn Gly Ser Ile Leu Glu Val Ala Ser Arg Ile Ser Thr Lys Ile Trp Glu Thr Arg Ala Lys Gly Tyr Gly Leu Asn Tyr Thr Ala Asp Tyr Pro Ala Val Gly Val His Asn Trp Leu Gln Trp	MHHHHHHHLEGSGPRDW LRPDATGTCEWDAANY WVQRCDVASPAMGHNV TVQIQPAGRGGNAALYL LDGARASDAANAWTFD ASAQSMFVDNNITLVMP VGGAGSFYTDWAAPSGI GSVGQVFKWETFLTQEL PAYLEAHFGVARNNYSV GGLSMGATAAMNLAAL HPEQFRQVLSYSGYL AMTAPGMYSLGLALLE VGGLSINSMYGSFFSPRR IELDPLFNMRGLDGKDV YVSAASGAWGGPDIEKY SFKDRVNGSILEVASRIS TKIWETRAKGYGLNYTA DYPAVGVHNWLQWNY QLAITRDRILNVMNAW

	Asn Tyr Gln Leu Ala Ile Thr Arg Asp Arg Ile Leu Asn Val Met Asn Ala Trp	
--	--	--

Quadro 1: Sequência de aminoácidos original e sequência do tipo FASTA gerada pelo SMS Three to One. Autoria própria.

5.2 Caracterização das proteínas recombinantes

5.2.1 Caracterização físico-química

No quadro 2 é possível observar os resultados da caracterização físico-química das proteínas recombinantes realizadas pelo ProtParam.

Quadro 2: Caracterizações gerais das proteínas rCP01850 e rCP09720 previstas com o programa de bioinformática ProtParam. Autoria própria.

Caracterizações	rCP01850	rCP09720
Número de aminoácidos	308	315
Peso molecular	33494.68 Da	34408.73 Da
pI	6.42	5.90
Principais resíduos	Alanina (12,0%); Leucina (9,1%) e Serina (9,1%)	Alanina (12.1%); Glicina (10.5%) e Leucina (8.9%)
Meia vida	Célula mamífera: 30h; Levedura: 20h; <i>E. Coli</i> : 10h	Célula mamífera: 30h; Levedura: 20h; <i>E. Coli</i> : 10h
Índice de instabilidade	36.26	30.45
Índice alifático	78.99	78.76
Média geral da hidropatia (gravy)	-0.365	-0.122

As duas proteínas alvo têm características físico-químicas semelhantes entre si, e se assemelham em muitas dessas características a outras proteínas de CP, diferindo apenas no peso molecular e tamanho da sequência (Araújo *et al.*, 2020, Droppa-Almeida; Franceschi; Padilha, 2018). Esta diferença pode ser devido as proteínas rCp01850 e a rCp09720 terem suas sequencias modificadas. Estas proteínas podem ser comparadas a proteína CP40 (Droppa-Almeida; Franceschi; Padilha, 2018) de *Corynebacterium pseudotuberculosis*, que foi utilizada com sucesso na prospecção de seis epítomos com potencial para produção de vacinas contra a linfadenite caseosa.

O ponto isoelétrico da proteína rCp01850 é em pH 6.4 e da rCp09720 é em pH 5.9, indicando que a estabilidade das proteínas é alcançada em meio levemente ácido. Esses dados se assemelham aos encontrados por Araújo *et al.*, 2020 e Droppa-Almeida; Franceschi; Padilha, 2018, que mostram proteínas alvo de CP preditas com pI semelhante, em pH 6.4. Desse modo, para a produção de vacina, o pH deve se manter estável e de acordo com o pI da proteína escolhida como componente ativo, já que mudanças no pH podem prejudicar sua estabilidade, conformação e consequentemente sua atividade.

A estabilidade de uma proteína é a capacidade de manter sua estrutura tridimensional, e consequentemente sua função, em condições naturais ou experimentais. Três parâmetros adicionais são fundamentais para essa capacidade: tempo de meia-vida, índice de instabilidade e índice alifático. O tempo de meia-vida predito para as proteínas recombinantes foi de 30h em célula mamífera, 20h em levedura e 10h em *E. coli* (Quadro 2). Esse resultado mostra que essas proteínas se mostram estáveis, mantendo tempo de meia-vida hábil para realização de experimentos testes e para a produção de vacinas duradouras. Além disso, o baixo índice de instabilidade (36.26 e 30.45) indica que a proteína possui uma baixa probabilidade de desnóvelar ou desnaturar, sendo comparáveis a outras proteínas (Araújo *et al.*, 2020; Droppa-Almeida; Franceschi; Padilha, 2018). O índice alifático predito para as proteínas foi de 78, sendo considerado alto. Este índice é responsável pela resistência em altas temperaturas, o que é um fator crucial para desenvolvimento, armazenamento e distribuição de vacinas, especialmente em regiões com pouca infraestrutura.

Por fim, em uma vacina deve haver a interação entre seus componentes, que podem ser polares ou apolares. Desta forma, foi predita a hidropatia da proteína pela escala GRAVY, cujo valores foram -0.365 e -0.122, para as proteínas rCp01850 e rCp019720, respectivamente. Estes valores negativos indicam que a proteína é hidrofílica, ou seja, podem possuir interação com a água e outras moléculas polares.

Diante de tudo isso, é possível observar que, de acordo com suas características físico-químicas, a rCp01850 e a rCp09720 também são boas candidatas para o desenvolvimento de uma vacina já que são estáveis, são hidrofílicas e termo resistentes.

5.2.2 Solubilidade

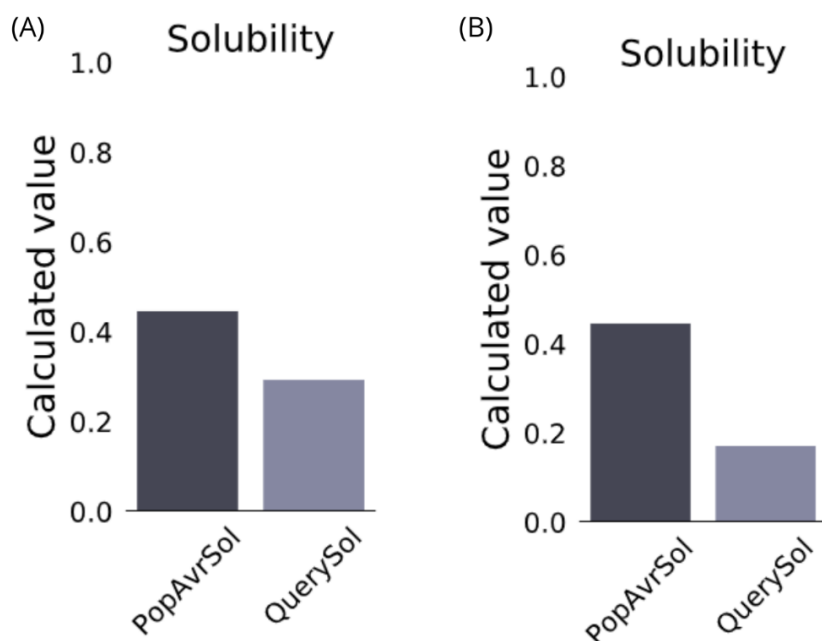
Nesse experimento, a sequência da proteína rCp01850 precisou ser adaptada para conter apenas aminoácidos canônicos. Desse modo, a nova sequência utilizada foi:

>rCP01850

MRGSHHHHHHLEGSQPVLQVPSDSSVGASLASPSVLHPGAPIQPFGPADYIGYISDISSYSGGIYYD
VVSNFEDLRQNHPKAMEENSATSIRINNEATAESVARAQRDALTEHDGVLVAFSDALGPRLGEHLRTA
LAENRLPKLQMLLGWTARAGGLASSTFVEKNIFRNPRPYIAHAGSIRQYRVSHHDLYPTSGSFPSGHT
NQATWTTALWAYMLPEFSAQILDRGSEAGFNRVVLGVHYPLDVIGGRMTGNAAAADRLNDPKMRL
DAAAELRAELEWRCGKSLLECATDTRSKEQAVQK

O teste de solubilidade mostrou que a proteína rCP01850 tem níveis de solubilidade de cerca de 0.3 e a rCP09720 de cerca de 0.2, como é possível observar no gráfico 1. Estes resultados indicam que essas proteínas são insolúveis em água, visto que o ponto de corte sugerido pelo programa ProteinSol é de 0.45 (Gráfico 1).

Gráfico 1: Resultados de solubilidade através do ProteinSol. (A) Resultados da proteína rCP01850. (B) Resultados da proteína rCP09720. Autoria própria.



O gráfico 1A mostra o índice de solubilidade da proteína rCp01850 (QuerySol) em comparação com o padrão de referência do programa (PopAvrSol). O gráfico 1B mostra o índice de solubilidade da proteína rCp09720 (QuerySol) em comparação com o padrão de referência do programa (PopAvrSol). Segundo (Hebditch *et al.*, 2017), o PopAvrSol indica a solubilidade média das proteínas de *E. coli* em água (0.45), onde um valor inferior ao padrão indica que as proteínas alvo são consideradas insolúveis em água.

A solubilidade de uma proteína deve ser considerada ao se desenvolver uma vacina, já que a água é um dos diluentes mais utilizados (World Health Organization, 2020), entretanto no caso das proteínas alvo deste estudo a diluição com outro solvente seria mais indicada.

5.2.3 Potencial alergênico

O teste de potencial alergênico (AlgPred) obteve que as proteínas alvo são não alergênicas, com resultado para a proteína rCP01850 de -1.9832729 e para a proteína rCP09720 de -0.92114026. Esse valor é calculado pela composição dos aminoácidos da proteína, onde um valor inferior a -0.4 indica que a proteína é não alergênica. Este teste possui precisão de 85% (Saha; Raghava, 2006).

Na busca por epítomos para o desenvolvimento de vacinas, estudos recentes como o de Barral *et al.*, (2022) e Droppa-Almeida, Franceschi, Padilha (2018) mostram que proteínas não alergênicas devem ser escolhidas, para que uma resposta imune diferente da esperada não seja ativada.

5.2.4 Potencial Antigênico

O potencial antigênico das proteínas rCp01850 e rCp09720 foi de 0.6079 e 0.5465, respectivamente. Esse valor indica que estas proteínas podem ser consideradas possíveis antígenos, visto que o programa determina um valor de antigenicidade mínimo de 0.4, ou seja, acima desse valor a proteína pode ser considerada um possível antígeno. O índice de confiabilidade do programa VaxiJen30 é entre 70% e 89% (Doytchinova; Flower, 2007).

Os antígenos são uma parte fundamental no desenvolvimento de uma vacina, já que eles são responsáveis pela ativação da resposta imune do organismo. Na busca por possíveis epítomos para produção de vacina, de acordo com a literatura, são escolhidas apenas as proteínas que possuem potencial antigênico (Araújo *et al.*, 2020, Droppa-Almeida; Franceschi; Padilha, 2018).

5.3 Estruturas tridimensionais

5.3.1. I-Tasser

A predição da estrutura tridimensional das proteínas foi inicialmente realizada pelo programa I-Tasser (Yang; Zhang, 2015). Os resultados foram avaliados de acordo com o C-Score, onde os valores obtidos devem estar mais próximos a 2, de acordo com Yang e Zhang

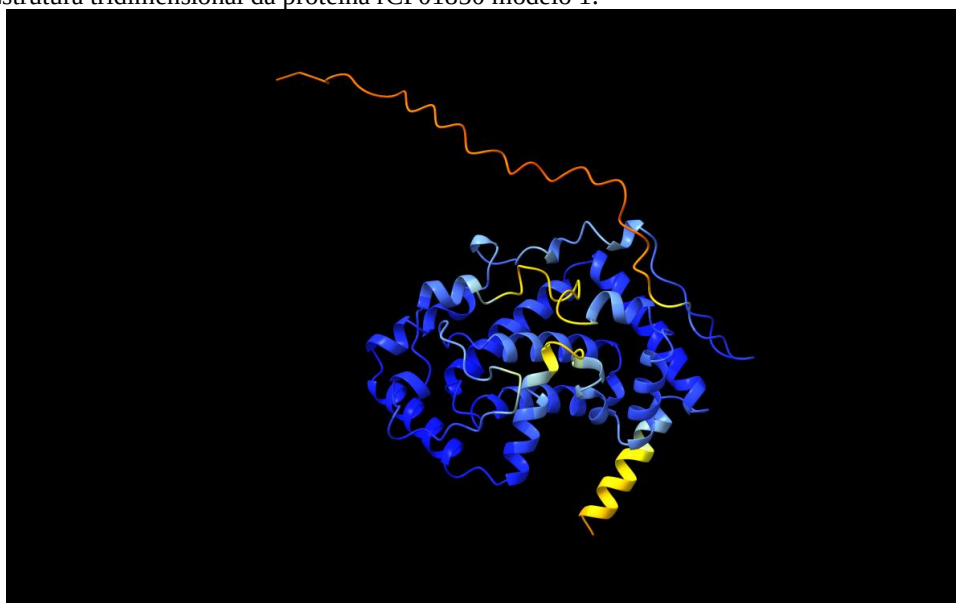
(2015). Porém, de acordo com a predição, o resultado mais favorável foi de -2,41 e 0,35 para a rCP01850 e a rCp09720, respectivamente. Sendo assim, estas estruturas não foram utilizadas para análises seguintes.

5.3.2 AlphaFold

As estruturas foram corretamente preditas utilizando o programa AlphaFold. Na figura 3 é possível observar o melhor modelo para a estrutura tridimensional da rCp01850, que apresentou um índice de confiabilidade de 80%, de acordo com a escala de cores do pLDDT (predicted Local Distance Difference Test). Na figura 4 é possível observar o melhor modelo para a estrutura tridimensional da rCp09720, que apresentou um índice de confiabilidade de 90%, de acordo com a escala de cores do pLDDT. A escala de cores do pLDDT está relacionada a com a precisão da modelagem em relação a estrutura real.

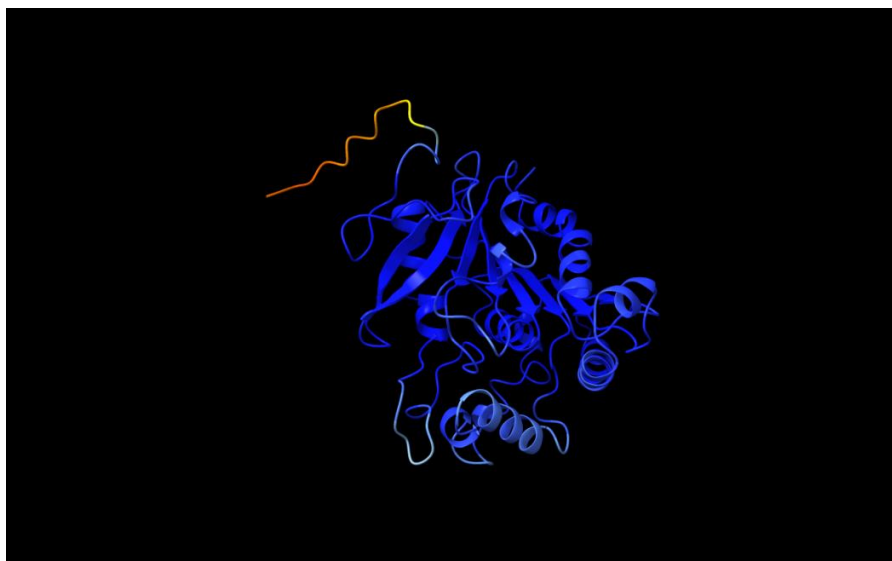
O AlphaFold utiliza principalmente o pLDDT para mostrar a confiabilidade da modelagem realizada. Esses valores podem ser observados através das cores ao longo da estrutura da proteína, por exemplo. A distribuição de cores nas estruturas obtidas com o AlphaFold para as proteínas rCp01850 e rCp09720 foi predominantemente azul, indicando um índice de confiabilidade acima de 80% (Figuras 3 e 4) (Jumper *et al.*, 2021).

Figura 3: Estrutura tridimensional da proteína rCP01850 modelo 1.



Fonte: Autoria própria utilizando ChimeraX com os dados do AlphaFold Collab.

Figura 4: Estrutura tridimensional da proteína rCP09720 modelo 1.



Fonte: Autoria própria utilizando ChimeraX com os dados do AlphaFold Collab

O programa PyMOL foi utilizado para determinar a composição estrutural da proteína rCp01850, que é majoritariamente formada por alfa-hélices, possuindo cento e noventa e três alfa-hélices. A proteína rCp01850 é uma fosfatase ácida, que tem como função a hidrólise de fosfato (Bezerra *et al.*, 2020). Essa função acontece no centro do sítio ativo da proteína que é formado por arginina com histidina, formando uma zona de carga positiva ideal para interação com o grupo fosfato (Araujo; Vihko, 2013). A grande presença de alfa-hélices na estrutura faz com que a proteína se torne mais flexível e promova mais interação entre a proteína e o ligante (Abd Rashid *et al.*, 2022).

A composição estrutural da proteína rCp09720 foi determinada pelo programa PyMOL, e é formada por cento e quarenta e uma alfa-hélices e quarenta e seis beta folhas. A proteína rCp09720 é uma esterase (Rezende *et al.*, 2016), que tem como função enzimática de catalisar éster. Seu sítio ativo é composto por serina, histidina e glutamina, e ele é o responsável por transformar éster em álcool (Smith *et al.*, 2008).

Atualmente, não existem muitos estudos de predição de estrutura da fosfatase ácida ou da esterase, principalmente no que se refere às proteínas de *C. pseudotuberculosis*. Além disso, as proteínas estudadas no presente trabalho são proteínas recombinantes, que apesar de serem utilizadas para pesquisas de vacinas *in vitro* e/ou *in vivo*, ainda não haviam sido caracterizadas e modeladas tridimensionalmente. Desse modo, essas proteínas e suas estruturas poderão servir de base para pesquisas com a CP e a linfadenite caseosa, principalmente na área de desenvolvimento de vacinas já que as proteínas são alvos ideais.

5.4. Ancoragem

Foi realizada a tentativa de ancoragem das proteínas rCp01850 e rCp09720 ao ligante MHC ovino, porém ocorreram erros durante a execução do experimento causados pelo não reconhecimento de algumas estruturas, e dessa forma, não foram produzidos resultados.

6. CONSIDERAÇÕES FINAIS

A linfadenite caseosa é uma doença de significativo impacto socioeconômico e, no cenário atual, ainda há necessidade de avanços em pesquisas em diversas áreas para suprir a demanda de testes, tratamentos e vacinas. Desse modo, a bioinformática desempenha um papel crucial na criação de novas tecnologias e produtos, que após análise *in silico* serão desenvolvidos e validados *in vitro* ou *in vivo*.

A utilização das ferramentas de bioinformática permitiu que as proteínas rCp01850 e rCp09720 fossem caracterizadas, demonstrando que ambas são escolhas promissoras para o desenvolvimento de imunobiológicos, em especial vacinas, devido à sua estabilidade, resistência térmica, baixa alergenicidade e potencial antigênico. Além disso, foi possível prever a estrutura tridimensional dessas proteínas, que poderão ser utilizadas para novos experimentos *in silico*.

Por fim, futuras pesquisas podem utilizar os dados gerados neste trabalho, bem como os modelos de estruturas 3D das proteínas para realizar estudos de ancoragem com diferentes ligantes, como potenciais fármacos e vacinas, além de realizar simulações moleculares para analisar interações proteína-complexo. Dessa forma, será possível avançar com mais eficácia na busca por vacinas e tratamentos para linfadenite caseosa.

7. REFERÊNCIAS

- ABD RASHID, Syaida Anati *et al.* A COMBINATION OF IN SILICO SUBTRACTIVE AND REVERSE VACCINOLOGY APPROACHES REVEALS POTENTIAL VACCINE TARGETS IN *Corynebacterium pseudotuberculosis*. **JOURNAL OF SUSTAINABILITY SCIENCE AND MANAGEMENT**, [s. l.], v. 17, n. 1, p. 99–109, 2022. Disponível em: <https://jssm.umt.edu.my/wp-content/uploads/sites/51/2023/07/7.-A-Combination-Of-In-Silico-Subtractive-And-Reverse-Vaccinology-Approaches-Reveals-Potential-Vaccine-Targets-In-Corynebacterium-Pseudotuberculosis-.pdf>. Acesso em: 18 out. 2024.
- ALENCASTRO, Ricardo Bicca de. Modelagem de proteínas por homologia. **Química Nova**, [s. l.], v. 26, p. 253–259, 2003. Disponível em: <http://www.scielo.br/j/qn/a/phHhSHJNSJCTDYfGRzgfktt/?lang=pt>. Acesso em: 23 nov. 2022.
- ALVES, F. S. F.; PINHEIRO, R. R.; PIRES, P. C. Linfadenite caseosa: patogenia, diagnóstico, controle. **EMBRAPA-CNPQ. Documentos**, [s. l.], 1997. Disponível em: <http://www.infoteca.cnptia.embrapa.br/bitstream/doc/514831/1/UMTDoc27.pdf>. Acesso em: 23 nov. 2022.
- ALVES, Francisco Selmo Fernandes; SANTIAGO, Lauana Borges; PINHEIRO, Raymundo Rizaldo. Linfadenite Caseosa: o Estado da Arte. [s. l.], p. 57, 2007.
- ARAÚJO, Carlos Leonardo *et al.* In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar ovis. **PeerJ**, [s. l.], v. 8, p. e9643, 2020. Disponível em: <https://peerj.com/articles/9643>. Acesso em: 18 out. 2024.
- ARAUJO, César L.; VIHKO, Pirkko T. Structure of Acid Phosphatases. In: MILLÁN, José Luis (org.). **Phosphatase Modulators**. Totowa, NJ: Humana Press, 2013. (Methods in Molecular Biology). v. 1053, p. 155–166. Disponível em: https://link.springer.com/10.1007/978-1-62703-562-0_11. Acesso em: 18 out. 2024.
- AYDIN, Suleyman. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. **Peptides**, [s. l.], v. 72, p. 4–15, 2015.
- BAIRD, G. J.; FONTAINE, M. C. *Corynebacterium pseudotuberculosis* and its Role in Ovine Caseous Lymphadenitis. **Journal of Comparative Pathology**, [s. l.], v. 137, n. 4, p. 179–210, 2007. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0021997507001211>. Acesso em: 23 nov. 2022.
- BARH, Debmalaya *et al.* Exoproteome and Secretome Derived Broad Spectrum Novel Drug and Vaccine Candidates in *Vibrio cholerae* Targeted by Piper betel Derived Compounds. **PLoS ONE**, [s. l.], v. 8, n. 1, p. e52773, 2013. Disponível em: <https://dx.plos.org/10.1371/journal.pone.0052773>. Acesso em: 16 out. 2024.
- BARNABÉ, Nathanael Natércio da Costa. Avaliação de perda econômica por linfadenite caseosa em caprinos abatidos no semiárido brasileiro. [s. l.], 2018. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/8858>. Acesso em: 25 nov. 2022.
- BARRAL, Thiago Doria *et al.* Immunoprophylactic properties of the *Corynebacterium pseudotuberculosis*-derived MBP:PLD:CP40 fusion protein. **Applied Microbiology and Biotechnology**, [s. l.], 2022. Disponível em: <https://doi.org/10.1007/s00253-022-12279-1>. Acesso em: 23 nov. 2022.

BETTINI, Astrid *et al.* A Seroepidemiological Survey of *Corynebacterium pseudotuberculosis* Infection in South Tyrol, Italy. **Pathogens**, [s. l.], v. 11, n. 11, p. 1314, 2022. Disponível em: <https://www.mdpi.com/2076-0817/11/11/1314>. Acesso em: 23 nov. 2022.

BEZERRA, Francisco Silvestre Brilhante *et al.* The combination of Brazilian red propolis and recombinant protein rCP01850 in the immunoprophylaxis of *Corynebacterium pseudotuberculosis* infection in mice. **Microbial Pathogenesis**, [s. l.], v. 149, p. 104354, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0882401020307208>. Acesso em: 18 out. 2024.

DALSASS, Mattia *et al.* Comparison of Open-Source Reverse Vaccinology Programs for Bacterial Vaccine Antigen Discovery. **Frontiers in Immunology**, [s. l.], v. 10, 2019. Disponível em: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00113>. Acesso em: 25 nov. 2022.

DELLER, M. C.; KONG, L.; RUPP, B. Protein stability: a crystallographer's perspective. **Acta Crystallographica Section F: Structural Biology Communications**, [s. l.], v. 72, n. 2, p. 72–95, 2016. Disponível em: <https://scripts.iucr.org/cgi-bin/paper?en5571>. Acesso em: 25 nov. 2022.

DORELLA, Fernanda *et al.* *Corynebacterium pseudotuberculosis*: Microbiology, biochemical properties, pathogenesis and molecular studies of virulence. **Veterinary research**, [s. l.], v. 37, p. 201–18, 2006.

DOYTCHINOVA, Irini A.; FLOWER, Darren R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. **BMC Bioinformatics**, [s. l.], v. 8, n. 1, p. 4, 2007. Disponível em: <https://doi.org/10.1186/1471-2105-8-4>. Acesso em: 16 out. 2024.

DROPPA-ALMEIDA, Daniela; FRANCESCHI, Elton; PADILHA, Francine Ferreira. Immune-Informatic Analysis and Design of Peptide Vaccine From Multi-epitopes Against *Corynebacterium pseudotuberculosis*. **Bioinformatics and Biology Insights**, [s. l.], v. 12, p. 1177932218755337, 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5954444/>. Acesso em: 12 out. 2024.

FACCIOLI-MARTINS, P. Y.; ALVES, F. S. F.; PINHEIRO, R. R. Linfadenite caseosa: perspectivas no diagnóstico, tratamento e controle. [s. l.], 2014. Disponível em: <http://www.infoteca.cnptia.embrapa.br/handle/doc/1007402>. Acesso em: 23 nov. 2022.

FACCIOLI-MARTINS; YOSHIDA, Patrícia. **Linfadenite caseosa: perspectivas no diagnóstico, tratamento e controle**. [S. l.: s. n.], 2014.

FONTAINE, M. C.; BAIRD, G. J. Caseous lymphadenitis. **Small Ruminant Research**, [s. l.], v. 76, n. 1, Special Issue: Current Issues in Sheep Health and Welfare, p. 42–48, 2008. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0921448807002866>. Acesso em: 23 nov. 2022.

GARBIS, Spyros; LUBEC, Gert; FOUNTOULAKIS, Michael. Limitations of current proteomics technologies. **Journal of Chromatography A**, [s. l.], v. 1077, n. 1, p. 1–18, 2005. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0021967305008496>. Acesso em: 25 nov. 2022.

GROSDIDIER, A.; ZOETE, V.; MICHIELIN, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. **Nucleic Acids Research**, [s. l.], v. 39, n. suppl,

p. W270–W277, 2011. Disponível em: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr366>. Acesso em: 18 out. 2024.

GUIMARÃES, Alessandro de Sá *et al.* CASEOUS LYMPHADENITIS: EPIDEMIOLOGY, DIAGNOSIS, AND CONTROL. **OPEN ACCESS**, [s. l.], p. 11, 2011.

GUISSONI, Ana Carla Peixoto; CARDOSO, Divina das Dores de Paula. Proteômica: uma ferramenta analítica. **Revista Eletrônica de Ciências Humanas, Saúde e Tecnologia**, [s. l.], v. 1, n. 15, p. 62–75, 2019. Disponível em: <https://revista.fasem.edu.br/index.php/fasem/article/view/193>. Acesso em: 23 nov. 2022.

HEBDITCH, Max *et al.* Protein–Sol: a web tool for predicting protein solubility from sequence. **Bioinformatics**, [s. l.], v. 33, n. 19, p. 3098–3100, 2017. Disponível em: <https://doi.org/10.1093/bioinformatics/btx345>. Acesso em: 12 out. 2024.

JUMPER, John *et al.* Highly accurate protein structure prediction with AlphaFold. **Nature**, [s. l.], v. 596, n. 7873, p. 583–589, 2021. Disponível em: <https://www.nature.com/articles/s41586-021-03819-2>. Acesso em: 17 abr. 2024.

KNIPPELS, Léon M. J.; PENNINKS, André H. Recent advances using rodent models for predicting human allergenicity. **Toxicology and Applied Pharmacology**, [s. l.], v. 207, n. 2, Supplement, Living in a Safe Chemical World. Proceedings of the 10th International Congress of Toxicology 11-15 July, 2004, Tampere, Finland, p. 157–160, 2005. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0041008X05002796>. Acesso em: 25 nov. 2022.

KONTERMANN, Roland E. Strategies for extended serum half-life of protein therapeutics. **Current Opinion in Biotechnology**, [s. l.], v. 22, n. 6, 22/6 Chemical biotechnology and Pharmaceutical biotechnology, p. 868–876, 2011. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0958166911006574>. Acesso em: 25 nov. 2022.

LAMPE, Anna T. *et al.* High- and low-molecular-weight chitosan act as adjuvants during single-dose influenza A virus protein vaccination through distinct mechanisms. **Biotechnology and Bioengineering**, [s. l.], v. 118, n. 3, p. 1224, 2020. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7897297/>. Acesso em: 16 out. 2024.

MONTASSIER, Helio José. Moléculas Reconhecidas pelo Sistema Imune:- PAMPS e Antígenos (Ag). [s. l.], p. 16, 2018.

ODHAH, Mohammed Naji *et al.* Current Review on Mycolic Acid Immunogen of *Corynebacterium pseudotuberculosis*. **Journal of Advanced Veterinary Research**, [s. l.], v. 12, n. 2, p. 177–186, 2022. Disponível em: <https://www.advetresearch.com/index.php/AVR/article/view/876>. Acesso em: 23 nov. 2022.

PETTERSEN, Eric F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, [s. l.], v. 25, n. 13, p. 1605–1612, 2004.

PINTO, Eduardo Fonseca; DA-CRUZ, Alda Maria. Vacunas: progresos y nuevos retos para el control de enfermedades prevenibles. [s. l.], p. 17, 2011.

REZENDE, Andrea de Fátima Silva *et al.* In silico identification of *Corynebacterium pseudotuberculosis* antigenic targets and application in immunodiagnosis. **Journal of Medical Microbiology**, [s. l.], v. 65, n. 6, p. 521–529, 2016. Disponível em: <https://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.000263>. Acesso em: 18 out. 2024.

RIGHETTI, Pier Giorgio. Determination of the isoelectric point of proteins by capillary isoelectric focusing. **Journal of Chromatography A**, [s. l.], v. 1037, n. 1, Estimation of Physicochemical Properties by Chromatographic and Electrophoretic Techniques, p. 491–499, 2004. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0021967303020600>. Acesso em: 25 nov. 2022.

RODRIGUES, José Antunes. TECNOLOGIA DO DNA RECOMBINANTE. **Estudos Avançados**, [s. l.], v. 8, n. 22, p. 597–604, 2003. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40141994000300091&lng=pt&tlng=pt. Acesso em: 23 nov. 2022.

SÁ, Maria *et al.* Linfadenite caseosa em caprinos e ovinos: Revisão. **Pubvet**, [s. l.], v. 12, p. 1–13, 2018.

SAHA, Sudipto; RAGHAVA, G. P. S. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. **Nucleic Acids Research**, [s. l.], v. 34, n. suppl_2, p. W202–W209, 2006. Disponível em: <https://doi.org/10.1093/nar/gkl343>. Acesso em: 12 out. 2024.

SANTOS FILHO, Osvaldo Andrade; ALENCASTRO, Ricardo Bicca de. Modelagem de proteínas por homologia. **Química Nova**, [s. l.], v. 26, p. 253–259, 2003. Disponível em: <https://www.scielo.br/j/qn/a/phHhSHJNSJCTDYfGRzgfktt/>. Acesso em: 16 out. 2024.

SILVA, Mara Thais de Oliveira *et al.* In silico analyses and design of a chimeric protein containing epitopes of SpaC, PknG, NanH, and SodC proteins for the control of caseous lymphadenitis. **Applied Microbiology and Biotechnology**, [s. l.], v. 105, n. 21, p. 8277–8286, 2021. Disponível em: <https://doi.org/10.1007/s00253-021-11619-x>. Acesso em: 4 nov. 2022.

SMITH, Adam J. T. *et al.* Structural Reorganization and Preorganization in Enzyme Active Sites: Comparisons of Experimental and Theoretically Ideal Active Site Geometries in the Multistep Serine Esterase Reaction Cycle. **Journal of the American Chemical Society**, [s. l.], v. 130, n. 46, p. 15361–15373, 2008. Disponível em: <https://pubs.acs.org/doi/10.1021/ja803213p>. Acesso em: 18 out. 2024.

STEIN, Amelie; MOSCA, Roberto; ALOY, Patrick. Three-dimensional modeling of protein interactions and complexes is going ‘omics. **Current Opinion in Structural Biology**, [s. l.], v. 21, n. 2, p. 200–208, 2011. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S09594440X11000078>. Acesso em: 23 nov. 2022.

STOTHARD, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. **BioTechniques**, [s. l.], v. 28, n. 6, p. 1102, 1104, 2000.

VANDERSCHUREN, Koen *et al.* Tuning protein half-life in mouse using sequence-defined biopolymers functionalized with lipids. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 119, n. 4, p. e2103099119, 2022. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8794819/>. Acesso em: 16 out. 2024.

WILKINS, Marc R. *et al.* Protein Identification and Analysis Tools in the ExPASy Server. In: LINK, Andrew J. **2-D Proteome Analysis Protocols**. New Jersey: Humana Press, 1998. v. 112, p. 531–552. Disponível em: <http://link.springer.com/10.1385/1-59259-584-7:531>. Acesso em: 16 out. 2024.

WORLD HEALTH ORGANIZATION. **Como são as vacinas desenvolvidas?**. [S. l.], 2020. Disponível em: <https://www.who.int/pt/news-room/feature-stories/detail/how-are-vaccines-developed>. Acesso em: 18 out. 2024.

YANG, Jianyi; ZHANG, Yang. I-TASSER server: new development for protein structure and function predictions. **Nucleic Acids Research**, [s. l.], v. 43, n. W1, p. W174–W181, 2015. Disponível em: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv342>. Acesso em: 17 abr. 2024.