

GeneDatabase: Plataforma Web para Comparação de Estruturas de Proteínas com Mutações Associadas à Resistência a Fármacos em *Mycobacterium tuberculosis*

Veridiana P. Richter¹, Abdirasak Mohamed Osman¹, Karina S. Machado¹

¹Laboratório de Biologia Computacional (COMBI-LAB)
Universidade Federal do Rio Grande (FURG)
Rio Grande – RS – Brazil

veri.piva@gmail.com, abdirasakm0@gmail.com, karinaecomp@gmail.com

Abstract. *The GeneDatabase is a web platform for comparing protein structure predictions of single-point mutations associated with drug resistance in Mycobacterium tuberculosis. The platform was developed using React, TypeScript, and Vite, and integrates three-dimensional structures computationally predicted by eight protein structure prediction tools (AlphaFold3, ColabFold, Swiss-Model, MODELLER, Phyre2, I-TASSER, trRosetta, and OmegaFold). In addition to providing the models, the platform also integrates quality metrics. The interface allows researchers to explore genes, select structural models, view comparative quality metrics, and analyze three-dimensional molecular structures using 3Dmol.js with WebGL rendering. Data are stored as static JSON and PDB files, deployed via GitHub Pages, eliminating the need for backend infrastructure. Features include gene search, links to the NCBI, TBDB, and MycoBrowser databases where the original data were obtained, and bulk download of PDB files of the predicted models. This resource supports molecular surveillance and structural analysis of drug resistance-associated variants without requiring specialized software.*

Resumo. *A GeneDatabase é uma plataforma web para comparação de predições de estruturas de proteínas com mutações pontuais associadas a resistência a fármacos em Mycobacterium tuberculosis. Essa plataforma foi desenvolvida com React, TypeScript e Vite e integra estruturas tridimensionais determinadas computacionalmente utilizando oito programas de predição de estrutura de proteínas (AlphaFold3, ColabFold, Swiss-Model, MODELLER, Phyre2, I-TASSER, trRosetta e OmegaFold). Além de disponibilizar os modelos, a plataforma integra métricas de qualidade. Essa interface permite aos pesquisadores explorar genes, selecionar modelos estruturais, visualizar métricas comparativas de qualidade e analisar estruturas moleculares tridimensionais por meio do 3Dmol.js com renderização via WebGL. Os dados são armazenados como arquivos estáticos em formato JSON e PDB, disponibilizados via GitHub Pages, eliminando a necessidade de infraestrutura de backend. Entre as funcionalidades, destacam-se a busca por genes, links para as bases NCBI, TBDB e MycoBrowser onde os dados originais foram obtidos, e o download em lote de arquivos PDB dos modelos determinados. Este recurso oferece suporte à vigilância molecular e à análise estrutural de variantes associadas à resistência a fármacos, sem a necessidade de softwares especializados.*

1. Introdução

Doenças tropicais negligenciadas (DTNs) são um grupo diverso de doenças que incluem principalmente infecções de origem viral, protozoária, helmíntica e bacteriana, juntamente com doenças transmitidas por vetores ou que envolvem reservatórios animais [da Saúde 2024, Milanez 2026]. A Organização Mundial da Saúde (OMS)¹ reconhece 20 condições prioritárias como DTNs, entre elas estão malária, dengue, esquistossomose, tuberculose, hanseníase, entre outras. Elas afetam um bilhão de pessoas, predominantemente, mas não exclusivamente, pessoas em zonas climáticas tropicais e subtropicais, estando presentes em 149 países [Winkler et al. 2018, Wang et al. 2021].

As DTNs são frequentemente descritas como uma “pandemia crônica” e estão intimamente ligadas à desvantagem socioeconômica. Existe uma clara associação entre a ocorrência e o Índice de Desenvolvimento Humano (IDH), com maior prevalência em regiões onde a pobreza é generalizada, como em áreas rurais e comunidades urbanas empobrecidas, particularmente em países de baixa e média renda. Essa relação é evidente ao examinar seus padrões de distribuição global [Lindoso and Lindoso 2009, Rocha et al. 2023]. Além de serem uma preocupação de saúde pública, as doenças negligenciadas também representam uma barreira para o desenvolvimento humano e econômico das nações [Hotez et al. 2006].

A tuberculose causada por *Mycobacterium tuberculosis* continua a representar um desafio significativo de saúde pública em todo o mundo [Machado et al. 2026, Schwalb et al. 2026]. A segunda edição do “Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance”² serve como um guia abrangente que relaciona variantes genéticas específicas à resistência a fármacos. Ao consolidar extensos dados genômicos e fenotípicos de parceiros de pesquisa globais, estabelece um padrão confiável para a vigilância molecular e a interpretação clínica de testes genéticos. Em última análise, esse recurso permite que profissionais de saúde alcancem diagnósticos mais precisos e implementem tratamentos de tuberculose altamente direcionados e eficazes [Organization 2021]. Este catálogo serviu como base fundamental para a pesquisa que levou ao desenvolvimento da nossa plataforma web Gene Database.

A versão atual do GeneDatabase inclui 15 genes associados à resistência em *M. tuberculosis*, compreendendo um total de 384 variantes de ponto único modeladas (*pncA*: 204; *rpoB*: 103; *embB*: 13; *gyrA*: 10; *ethA*: 9; *gid*: 9; *Rv0678*: 8; *gyrB*: 8; *atpE*: 6; *katG*: 5; *rpsL*: 4; *tlyA*: 2; *ddn*: 1; *inhA*: 1; *rplC*: 1). Cada variante é representada por até oito modelos preditivos independentes, resultando em aproximadamente 3.000 modelos estruturais individuais, todos avaliados por meio de seis métricas de qualidade.

Algumas dessas variantes já tem estrutura tridimensional determinada experimentalmente e depositada em base de dados como o *Protein Data Bank* (PDB). Contudo, a maioria dessas mutações ainda não foi estruturalmente caracterizada e compreendê-las é fundamental para o desenvolvimento de novos fármacos voltados ao tratamento da tuberculose resistente. Em um trabalho prévio do grupo [Richter et al. 2026], foram modelados e comparados 204 variantes para o gene *pncA* do *Mycobacterium tuberculosis*. Durante essa pesquisa percebeu-se a necessidade de organização

¹ <https://www.who.int/health-topics/neglected-tropical-diseases>

² <https://www.who.int/publications/i/item/9789240082410>

e disponibilização dos dados que estavam sendo gerados. Assim, este artigo tem por objetivo propor o GeneDatabase, uma plataforma web para comparação de modelos de estruturas de proteínas com mutações pontuais associadas a resistência a fármacos em *Mycobacterium tuberculosis*. Para modelar as estruturas foram considerados oito diferentes algoritmos de predição: AlphaFold3 [Jumper et al. 2021], ColabFold [Mirdita et al. 2022], Swiss-Model [Waterhouse et al. 2018], MODELLER [Webb and Sali 2016], Phyre2 [Kelley et al. 2015], I-TASSER [Roy et al. 2010], trRosetta [Jianyi et al. 2020], e OmegaFold [Wu et al. 2022]. Os modelos foram gerados para variantes do Catálogo com mutações pontuais associadas a resistência confirmadas para diferentes genes e todos os modelos foram validados para diferentes métricas. As avaliações dos modelos também estão disponíveis na plataforma permitindo que o usuário possa escolher o modelo mais conveniente do seu gene de interesse.

Os recursos públicos existentes oferecem perspectivas complementares, porém limitadas: o *AlphaFold Protein Structure Database*³ disponibiliza um único modelo predito por entrada UniProt, sem cobertura de variantes; o *Protein Data Bank*⁴ (PDB) contém estruturas determinadas experimentalmente, existentes para poucas variantes de resistência; e recursos específicos para tuberculose, como TBDB e MycoBrowser, concentram-se em anotações genômicas e funcionais, sem disponibilizar modelos estruturais. O GeneDatabase complementa esses recursos ao fornecer, para cada mutação pontual do Catálogo da OMS, até oito modelos estruturais preditos e seis métricas comparativas de qualidade em uma única interface interativa.

2. Materiais e Métodos

2.1. Projeto do Sistema e Desenvolvimento do Frontend

A plataforma GeneDatabase foi desenvolvida como uma plataforma web para facilitar a comparação de modelos de estruturas de proteínas associadas à resistência a fármacos em *Mycobacterium tuberculosis*. O *frontend* foi implementado utilizando *React* e *TypeScript*, suportando o desenvolvimento modular e melhorando a confiabilidade do código à medida que o projeto evoluiu.

O *Vite* foi utilizado para gerenciar o processo de *build* e otimizar o fluxo de desenvolvimento. A interface foi organizada em quatro componentes principais: uma tabela de navegação de genes, um painel de seleção de modelos, uma visualização comparativa de métricas de qualidade e um visualizador de estruturas tridimensionais. Essa organização modular permitiu que diferentes partes do sistema fossem desenvolvidas e testadas de forma independente.

A estilização foi implementada utilizando *Tailwind* CSS para garantir um *layout* consistente e responsivo. A navegação entre as visualizações é realizada no lado do cliente utilizando *React Router*.

³<https://alphafold.com/>

⁴<https://www.rcsb.org/>

2.2. Dados

A plataforma integra predições de estruturas de proteínas geradas por oito ferramentas computacionais: AlphaFold3⁵, ColabFold⁶, SWISS-MODEL⁷, MODELLER⁸, Phyre2⁹, I-TASSER¹⁰, trRosetta¹¹ e OmegaFold¹². Além dos modelos estruturais, métricas de avaliação de qualidade foram incluídas para dar suporte à análise comparativa.

Todos os dados foram armazenados como recursos estáticos. As informações dos genes e as métricas associadas foram organizadas em formato JSON, enquanto as estruturas de proteínas foram armazenadas como arquivos PDB. Esses recursos são incorporados à aplicação e carregados dinamicamente no navegador quando necessário. Essa abordagem simplifica a arquitetura do sistema ao eliminar a necessidade de serviços de *backend*, ao mesmo tempo em que facilita a implantação e a manutenção.

Todos os dados são versionados no repositório GitHub, e cada versão é identificada com a data e um *changelog* listando genes, variantes ou modelos recém-adicionados. A plataforma exibe uma indicação de "última atualização" no rodapé de todas as páginas. Novas variantes são incorporadas sempre que (i) uma nova edição do Catálogo da OMS for publicada, ou (ii) novas mutações associadas à resistência forem caracterizadas e modeladas pelo grupo de pesquisa. Os usuários podem acompanhar as atualizações do repositório por meio do sistema de notificações do GitHub.

2.3. Interface e Visualização

A interface segue um fluxo de trabalho direto. Os usuários iniciam navegando pela lista de genes e selecionando um alvo para análise posterior. Após a seleção, a plataforma fornece opções para comparar diferentes modelos de predição estrutural.

A visualização tridimensional foi implementada utilizando 3Dmol.js, permitindo que as estruturas de proteínas sejam renderizadas diretamente no navegador por meio de WebGL. Isso possibilita a exploração interativa, incluindo rotação, zoom e inspeção estrutural, sem a necessidade de softwares adicionais.

Cada entrada de gene inclui links para bases de dados externas, como *NCBI*, *TBDB* e *MycoBrowser*, permitindo acesso rápido a informações complementares. A plataforma também suporta o download em lote de arquivos PDB para um gene selecionado em formato compactado, facilitando a análise *offline*.

2.4. Implantação

A plataforma é implantada como um site estático utilizando *GitHub Pages*. A aplicação é construída utilizando *Vite*, que gera arquivos otimizados para produção.

Os artefatos de *build* são então publicados em um repositório, permitindo que atualizações sejam implantadas de forma eficiente sem a necessidade de infraestrutura de servidor dedicada.

⁵<https://alphafoldserver.com/>

⁶<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

⁷<https://swissmodel.expasy.org/>

⁸<https://salilab.org/modeller/>

⁹<https://www.sbg.bio.ic.ac.uk/phyre2/>

¹⁰<https://www.aideepmed.com/I-TASSER/>

¹¹<https://yanglab.qd.sdu.edu.cn/trRosetta/>

¹²<https://github.com/HeliXonProtein/OmegaFold>

A Figura 1 resume a metodologia utilizada para o desenvolvimento da plataforma GeneDatabase. A plataforma está disponível em ¹³

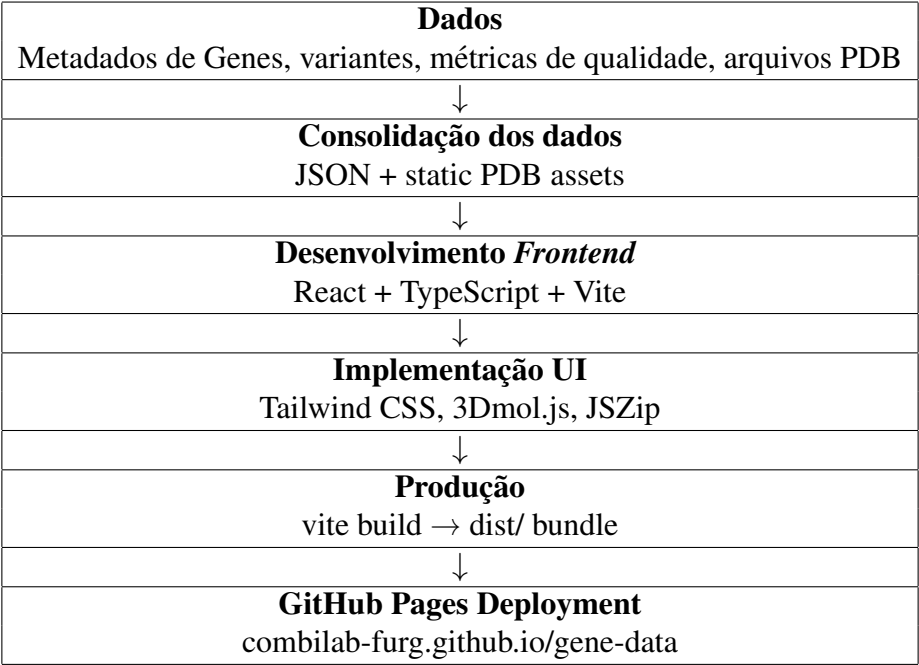


Figura 1. Visão geral do fluxo de desenvolvimento da GeneDatabase, desde a preparação e consolidação dos dados até a implementação do *frontend*, geração do *build* e implantação como uma aplicação web estática.

2.5. Reprodutibilidade e Parâmetros

Para garantir a reprodutibilidade, a Tabela 1 lista as versões dos softwares e os parâmetros utilizados em cada uma das oito ferramentas de predição. As ferramentas foram executadas em um cluster de alto desempenho, ou, quando a execução local não era viável (Phyre2, SWISS-MODEL, trRosetta), por meio de seus servidores web oficiais com parâmetros padrão, acessados entre 2024 e 2025. Para cada variante, apenas o modelo correspondente à sequência mutante foi gerado. Os alinhamentos de múltiplas sequências (MSAs), quando necessários, foram gerados a partir do banco de dados UniRef30 (versão 2023_02) utilizando o pipeline padrão MMseqs2 do ColabFold. Templates foram habilitados para todas as ferramentas que os suportam, utilizando a estrutura PDB listada na coluna PDB WT da Figura 2 como template primário, quando disponível. A plataforma está implantada como um site estático e está disponível em <https://combilab-furg.github.io/gene-data>.

2.6. Limitações e Decisões de Projeto

A versão atual da GeneDatabase é um recurso somente de leitura: os usuários podem explorar, visualizar e baixar os modelos estruturais pré-computados, mas não podem submeter sequências próprias para predição. Essa é uma decisão deliberada de projeto: ao restringir o conjunto de dados às variantes curadas do Catálogo da OMS e servir todos os modelos como arquivos estáticos, a plataforma dispensa infraestrutura de servidor, opera

¹³combilab-furg.github.io/gene-data

Tabela 1. Ferramentas de predição de estrutura, versões e parâmetros utilizados.

Ferramenta	Versão	Parâmetros principais
AlphaFold3	v3.0.0	num_recycles=10, seeds=5, modelo: ranked_0
ColabFold	v1.5.5	num_recycles=3, model=alphafold2_ptm, MSA=mmseqs2_uniref.env
SWISS-MODEL	2024-12	Template selecionado automaticamente do PDB; parâmetros padrão
MODELLER	v10.5	num_models=5, modelo selecionado pelo escore DOPE
Phyre2	v2	Modo intensivo; parâmetros padrão
I-TASSER	v5.2	Parâmetros padrão; modelo selecionado pelo C-score
trRosetta	2024-12	Parâmetros padrão
OmegaFold	v1.1.0	num_recycles=10, num_cycles=1

de forma gratuita via GitHub Pages e garante que todos os modelos exibidos foram validados com as mesmas seis métricas de qualidade sob condições controladas. A extensão da GeneDatabase para aceitar variantes submetidas pelo usuário está prevista como trabalho futuro e exigiria uma arquitetura híbrida combinando o frontend estático atual com um serviço de predição em backend.

3. Resultados

A plataforma web GeneDatabase fornece uma interface intuitiva para explorar genes, mutações e informações estruturais associadas. Conforme ilustrado na Figura 2, a página inicial apresenta uma barra de busca centralizada que permite aos usuários localizar rapidamente genes, identificadores ou variantes de interesse. O painel principal apresenta uma tabela estruturada contendo atributos-chave para cada gene, incluindo seu identificador, disponibilidade de estruturas de proteínas do tipo selvagem (*wild-type*) (PDB WT), comprimento da sequência em aminoácidos e o número de variantes modeladas. Adicionalmente, são fornecidos *links* diretos para bases de dados externas, como National Center for Biotechnology Information (NCBI)¹⁴, TB Database (TBDB)¹⁵ e Mycobrowser¹⁶, facilitando investigações adicionais. Ao selecionar um gene ou mutação, os usuários podem acessar análises detalhadas das estruturas tridimensionais modeladas e visualizações interativas dessas estruturas, apoiando uma exploração abrangente da variação genética e suas implicações estruturais.

Após selecionar um gene na tabela principal do GeneDatabase, a plataforma atualiza dinamicamente a interface para refletir a escolha do usuário, conforme mostrado na Figura 3. O gene selecionado é destacado visualmente em um tom de azul mais escuro para manter o contexto, enquanto componentes interativos adicionais são apresentados para orientar os próximos passos da análise. Em particular, um painel de seleção de modelos torna-se disponível, permitindo aos usuários escolher entre diferentes modelos estruturais ou preditivos associados ao gene de interesse. Esse design facilita um fluxo de trabalho contínuo, transitando da exploração dos dados para uma análise detalhada.

Conforme ilustrado na Figura 4, ao interagir com o controle *Select model*, a interface se expande e exibe um menu suspenso contendo múltiplas abordagens computacionais para predição de estruturas de proteínas. Esse menu permite aos usuários filtrar e selecionar métodos específicos de predição estrutural para análise posterior dos modelos computacionais disponíveis, incluindo AlphaFold3, ColabFold, I-TASSER, MO-

¹⁴<https://www.ncbi.nlm.nih.gov/protein/>

¹⁵<http://tbdb.bu.edu/>

¹⁶<https://mycobrowser.epfl.ch/>

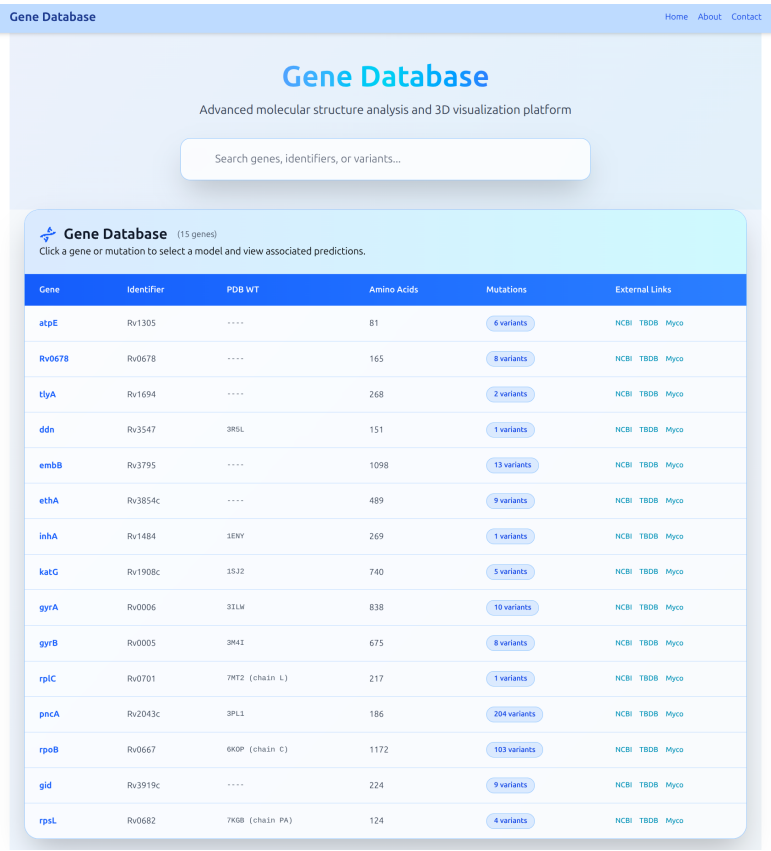


Figura 2. Overview da página inicial da plataforma GeneDatabase.

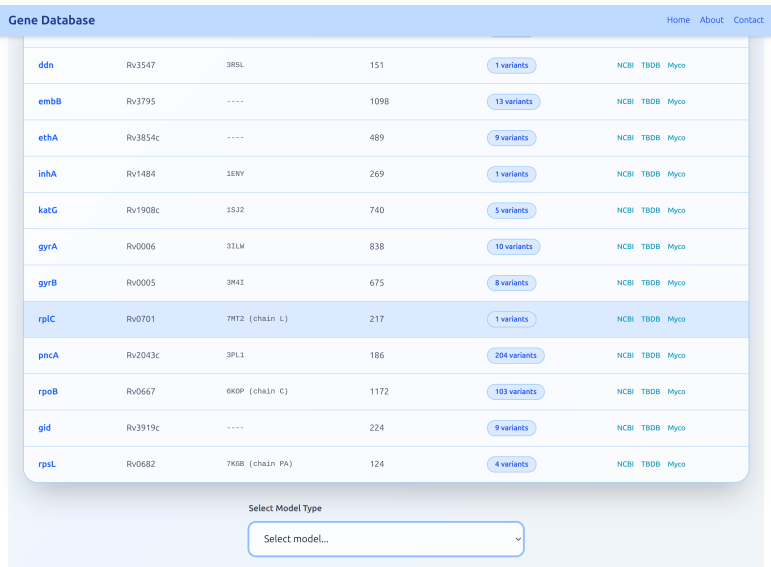


Figura 3. Interface de seleção específica por gene após a interação do usuário. Ao selecionar um gene na tabela do banco de dados, a interface destaca a entrada escolhida e fornece controles adicionais para a seleção de modelos.

DELLER, OmegaFold, Phyre2, SWISS-MODEL e trRosetta. A plataforma integra um conjunto diversificado de métodos de modelagem amplamente utilizados, permitindo que os usuários escolham uma técnica específica ou visualizem os resultados em todos os modelos disponíveis.

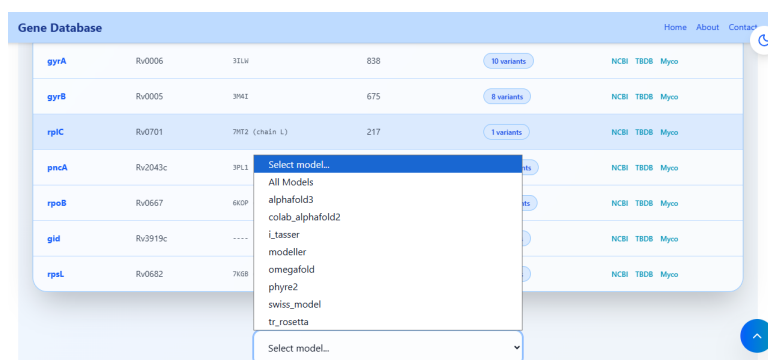


Figura 4. Menu suspenso de seleção de modelos na plataforma GeneDatabase.

Quando a opção *All Models* é selecionada, a plataforma exibe uma tabela de análise comparativa das previsões estruturais geradas por diferentes métodos computacionais, bem como métricas de qualidade para cada método de predição aplicado à variante escolhida, conforme mostrado na Figura 5. Neste exemplo, para o gene selecionado *inhA*, que contém uma única variante (Ser94Ala), a interface apresenta uma tabela detalhada resumindo múltiplas métricas de avaliação de qualidade, incluindo os escores ERRAT [Colovos and Yeates 1993], Verify3D [Bowie et al. 1991, Lüthy et al. 1992], MolProbity [Williams et al. 2018], VoroMQA [Olechnovič and Česlovas Venclovas 2019], QMEAN-Disco [Studer et al. 2020] e QMEAN [Benkert et al. 2011]. Isso permite que os usuários avaliem de forma sistemática a confiabilidade e a consistência dos modelos gerados por diferentes algoritmos.

Em paralelo, uma visualização tridimensional interativa é exibida, fornecendo uma representação do modelo selecionado e facilitando a inspeção qualitativa. Ao combinar métricas quantitativas de avaliação com visualização em tempo real, a plataforma apoia a tomada de decisão informada e uma análise mais aprofundada das variantes estruturais de proteínas.

3.1. Caso de Uso: Variante Ser94Ala do Gene *inhA*

Para ilustrar um fluxo de trabalho típico, descrevemos a análise da variante Ser94Ala no gene *inhA*, cujo produto — a enoil-ACP redutase (InhA) — é o principal alvo molecular da isoniazida. Esta variante está documentada no Catálogo da OMS como associada à resistência a esse fármaco de primeira linha.

Na página inicial (Figura 2), o pesquisador localiza *inhA* pela barra de busca e seleciona a opção *All Models* (Figura 4), carregando a tabela comparativa (Figura 5) com as oito previsões e suas métricas. A análise revela que nenhuma ferramenta domina todas as métricas simultaneamente: o OmegaFold obtém o maior escore ERRAT (100,0), o SWISS-MODEL apresenta o melhor MolProbity (0,78) e o AlphaFold3 se destaca no VoroMQA (0,478). Essa ausência de um modelo universalmente superior justifica a decisão de projeto da plataforma de expor todas as métricas, permitindo que o pesquisador esco-

lha o modelo mais adequado à sua aplicação — por exemplo, MolProbity para docagem molecular e VoromQA para dinâmica molecular.

O visualizador 3Dmol.js integrado (Figura 5, painel inferior) permite inspecionar interativamente a região de ligação do cofator onde a substituição Ser94Ala se localiza, sem necessidade de softwares adicionais. O modelo selecionado pode ser baixado para uso em ferramentas como AutoDock, GROMACS ou PyMOL, posicionando a GeneDatabase como filtro racional para triagem virtual de fármacos antes de experimentos *in vitro* e *in vivo*.

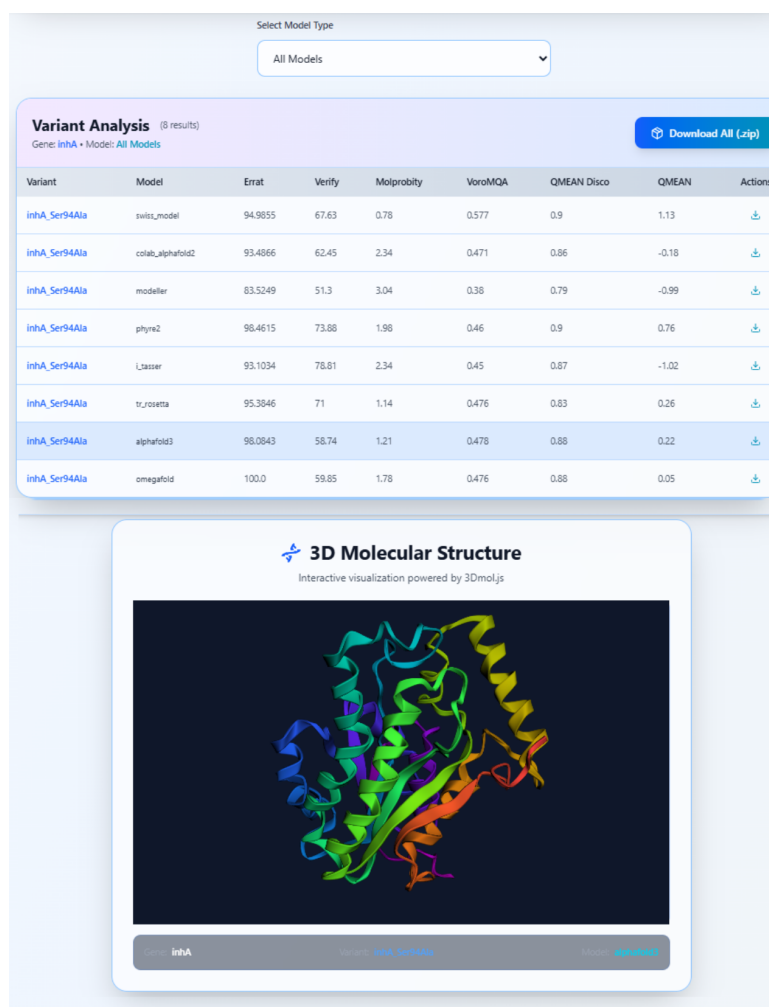


Figura 5. Análise de variantes e visualização das estruturas modeladas para o gene e modelo selecionado. No exemplo da Figura, é exibida a estrutura tridimensional gerada pelo algoritmo AlphaFold3 para a variante Ser94Ala do gene *inhA*.

4. Conclusões

Neste artigo apresentamos a plataforma GeneDatabase, uma plataforma Web projetada para impulsionar a pesquisa em bioinformática estrutural relacionada a predição de estruturas de proteínas com mutações pontuais associadas a resistência no *Mycobacterium Tuberculosis*. Ao unificar dados sobre genes, variantes, modelos tridimensionais gerados utilizando múltiplos algoritmos, métricas de validação desses modelos e visualização

interativa em uma única interface integrada, essa plataforma proposta reduz significativamente as barreiras técnicas para a investigação aprofundada de genes e variantes associadas a resistência em TB. Para *Mycobacterium tuberculosis*, um patógeno caracterizado por uma arquitetura genômica complexa e crescente resistência a múltiplos fármacos, essa integração não é apenas uma conveniência, mas uma necessidade. A capacidade de realizar análises comparativas entre diferentes abordagens de modelagem permite uma validação robusta das predições estruturais, apoiando diretamente a identificação de novos alvos terapêuticos e a caracterização de mutações associadas à resistência. Esses modelos são o ponto de partida para a busca de novos fármacos de forma racional, utilizando os modelos *in-silico* como filtros para experimentos *in-vitro* e *in-vivo* menos custosos, mais direcionados e promissores.

Além disso, ao combinar visualização interativa com métricas quantitativas, a plataforma permite que pesquisadores transitem de dados brutos para *insights* biológicos com maior precisão. No contexto das DTNs, ferramentas como a *GeneDatabase* desempenham um papel importante ao democratizar o acesso à bioinformática e dados modelados. Por fim, o GeneDatabase se apresenta como um recurso estratégico para a comunidade científica, acelerando o desenvolvimento de soluções inovadoras em saúde alinhadas aos objetivos de vigilância global sustentável e descoberta de fármacos.

5. Agradecimentos

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) processo 001, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) processo 406015/2025-2, e a Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) processos 25/2551-0002578-8, 22/2551-0000385-0 e 22/2551-0000390-7.

Referências

- Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27:343–350.
- Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253:164–170.
- Colovos, C. and Yeates, T. O. (1993). Verification of protein structures: Patterns of non-bonded atomic interactions. *Protein Science*, 2:1511–1519.
- da Saúde, O. P.-A. (2024). Doenças tropicais negligenciadas: dia mundial chama a atenção para o fortalecimento de ações intersetoriais para melhorar a qualidade de vida das comunidades - opas/oms — organização pan-americana da saúde.
- Hotez, P. J., Molyneux, D. H., Fenwick, A., Ottesen, E., Sachs, S. E., and Sachs, J. D. (2006). Incorporating a rapid-impact package for neglected tropical diseases with programs for hiv/aids, tuberculosis, and malaria. *PLoS Medicine*, 3:e102.
- Jianyi, Y., Anishchenkob, I., Hahnbeom, P., Pengd, Z., Ovchinnikove, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl,

- S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10:845–858.
- Lindoso, J. A. L. and Lindoso, A. A. B. (2009). Neglected tropical diseases in brazil. *Revista do Instituto de Medicina Tropical de São Paulo*, 51:247–253.
- Lüthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, 356:83–85.
- Machado, M., Jordan, A. E., Schwalb, A., Houben, R. M., Dodd, P. J., Dale, K., Schwartzman, K., and Campbell, J. R. (2026). Projected effects of changing global tuberculosis epidemiology on mycobacterium tuberculosis prevalence and immunoreactivity, 2024–2050. *Emerging Infectious Diseases*, 32:332–340.
- Milanez, B. (2026). Dia mundial das doenças tropicais negligenciadas: entenda o conceito e a importância desse controle para a saúde pública - instituto butantan.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). Colabfold: making protein folding accessible to all. *Nature Methods*, 19:679–682.
- Olechnovič, K. and Česlovas Venclovas (2019). Voromqa web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic Acids Research*, 47:W437–W442.
- Organization, W. H. (2021). *Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance*, volume 2.
- Richter, V. P., Osman, A. M., Guella, F. L., Werhli, A. V., Machado, K. S., and Osman, M. (2026). Comparative analysis of protein structure prediction tools for modeling single-point mutations in the pnca gene of mycobacterium tuberculosis. *The 41st ACM/SIGAPP Symposium on Applied Computing (SAC '26), March 23–27, 2026, Thessaloniki, Greece*, 1.
- Rocha, M. I. F., Maranhão, T. A., da Frota, M. M. C., de Araujo, T. K. A., e Silva, W. W. S. V., Sousa, G. J. B., Pereira, M. L. D., and de Araujo Filho, A. C. A. (2023). Mortalidade por doenças tropicais negligenciadas no brasil no século xxi: análise de tendências espaciais e temporais e fatores associados. *Revista Panamericana de Salud Pública*, 47:1.
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-tasser: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5:725–738.
- Schwalb, A., Dodd, P. J., Rickman, H. M., Ugarte-Gil, C. A., Horton, K. C., and Houben, R. M. G. J. (2026). Estimating the global burden of viable mycobacterium tuberculosis infection: A mathematical modelling study. *PLOS Medicine*, 23:e1004920.

- Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., and Schwede, T. (2020). Qmeandisco—distance constraints applied on model quality estimation. *Bioinformatics*, 36:1765–1771.
- Wang, A., MacNeil, A., and Maloney, S. (2021). Comparison and lessons learned from neglected tropical diseases and tuberculosis. *PLOS Global Public Health*, 1:e0000027.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A. P. D., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018). Swiss-model: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46:W296–W303.
- Webb, B. and Sali, A. (2016). Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, 54:1–55.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018). Molprobity: More and better reference data for improved all-atom structure validation. *Protein Science*, 27:293–315.
- Winkler, A. S., Klohe, K., Schmidt, V., Haavardsson, I., Abraham, A., Prodjinotho, U. F., Ngowi, B., Sikasunge, C., Noormahomed, E., Amuasi, J., Kaducu, J., Ngowi, H., Abele-Ridder, B., Harrison, W. E., and da Costa, C. P. (2018). Neglected tropical diseases – the present and the future. *Tidsskrift for Den norske legeforening*.
- Wu, R., Ding, F., Wang, R., Shen, R., Zhang, X., Luo, S., Su, C., Wu, Z., Xie, Q., Berger, B., Ma, J., and Peng, J. (2022). High-resolution de novo structure prediction from primary sequence.