



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIX SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2025

Inteligência artificial, *deep learning* e construção incremental para o desenvolvimento de novas entidades químicas com ação sobre o *Mycobacterium tuberculosis*

Késsia Souza Santos¹; Manoelito Coelho dos Santos Junior²

1.Késsia Souza Santos, PIBITI/CNPq, FARMACIA, e-mail: kelsouzs.uefs@gmail.com

2.Manoelito Coelho dos Santos Junior, DEPARTAMENTO DE SAÚDE, e-mail: manoelito@uefs.br

PALAVRAS-CHAVE: *Deep Learning*. *Mycobacterium tuberculosis*. Construção Incremental. KAST.

INTRODUÇÃO

A tuberculose (TB) permanece como uma ameaça significativa à saúde pública, sendo uma das principais causas de mortalidade por agentes infecciosos (OMS, 2023). O surgimento de cepas resistentes aos tratamentos convencionais evidencia a necessidade urgente de novos fármacos dirigidos a alvos específicos (Brasil, 2022a).

O desenvolvimento tradicional de medicamentos é longo e custoso, o que estimula a adoção de metodologias inovadoras. Nesse cenário, técnicas *in silico*, integrando inteligência artificial, triagem virtual e acoplamento molecular, surgem como abordagem promissora para o planejamento e descoberta de novos compostos com potencial inibitório contra *Mycobacterium tuberculosis* (Chen *et al.*, 2023). Dessa forma, este estudo empregou uma abordagem integrativa, combinando inteligência artificial, *deep learning*, construção incremental e métodos computacionais clássicos para identificar novos inibidores do *Mycobacterium tuberculosis*.

METODOLOGIA

A seleção de alvos moleculares de *M. tuberculosis* foi realizada por revisão de literatura assistida por Processamento de Linguagem Natural (PLN), considerando critérios de essencialidade, baixa similaridade com proteínas humanas, validação estrutural e *druggability*, enquanto alvos com número insuficiente de inibidores conhecidos foram excluídos (Jurafsky; Martin, 2008). Para o alvo selecionado, foi construído um banco de dados de inibidores, extraíndo IC₅₀ de bases públicas e literatura, classificando as moléculas como ativas (IC₅₀ ≤ 2.000 nM) ou inativas (IC₅₀ > 10.000 nM), e particionando o conjunto em treino (70%) e teste (30%) (Liu *et al.*, 2024).

O modelo preditivo foi desenvolvido em Python, baseado em rede neural *feedforward*, utilizando *Circular Fingerprints* como descritores moleculares. Técnicas de validação cruzada e análise de curvas de aprendizagem foram aplicadas para assegurar generalização e minimizar *overfitting*. O desempenho foi avaliado pelo conjunto de teste, destacando-se AUC e Fator de Enriquecimento (EF) (Oliveira *et al.*, 2023).

Hits preditos em uma biblioteca de produtos naturais foram otimizados por construção incremental com vBROOD, gerando análogos bioisotéricos (Nicholls *et al.*,

2010). Estes foram submetidos a acoplamento molecular com a estrutura cristalográfica do alvo, utilizando o programa DOCK. O composto com melhor afinidade foi analisado quanto às interações intermoleculares via PLIP.

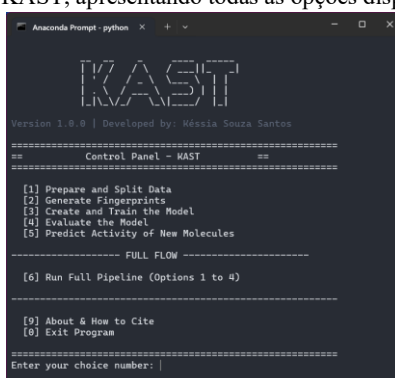
RESULTADOS E DISCUSSÃO

A prospecção inicial por NLP em bases científicas resultou em 90 artigos (Jurafsky; Martin, 2008), reduzidos a 23 após curadoria, dos quais emergiram 11 alvos moleculares distintos. A aplicação de filtros de essencialidade, ausência no hospedeiro, validação estrutural e relevância metabólica permitiu priorizar oito candidatos. Dentre eles, a poliketide sintase 13 domínio tioesterase (PKS13-TE) foi selecionada, destacando-se pela validação estrutural e pela ampla disponibilidade de compostos ativos reportados ($n = 695$), aspecto determinante para a construção de modelos preditivos (Kim, 2021).

Para a construção do banco de dados, o ChEMBL foi a principal fonte, complementado por 128 moléculas extraídas da literatura. Após padronização em SMILES e processamento no MarvinSketch, aplicou-se curadoria automatizada com remoção de duplicatas e filtros de atividade ($IC_{50} \leq 2000$ nM para ativos e ≥ 10.000 nM para inativos), consolidando um conjunto final de 495 moléculas (Liu *et al.*, 2024). A adoção do método *Scaffold Splitter* foi estratégica, pois assegura que os esqueletos químicos presentes no conjunto de teste sejam inéditos ao modelo, garantindo um cenário mais realista para a descoberta de novos compostos (Yang *et al.*, 2019).

O modelo preditivo foi desenvolvido com algoritmos de *deep learning* e *machine learning*, implementados em Python com o suporte de bibliotecas especializadas. Denominado KAST (*K-atalytic Automated Screening TaskFlow*), o sistema foi estruturado em módulos integrados por uma interface interativa (figura 1), permitindo a execução de todo o fluxo de trabalho, desde a preparação dos dados até a validação e aplicação dos modelos (Harris *et al.*, 2020; Landrum, 2025).

Figura 1 – Tela inicial (menu) do KAST, apresentando todas as opções disponíveis.



Fonte: o autor (2025).

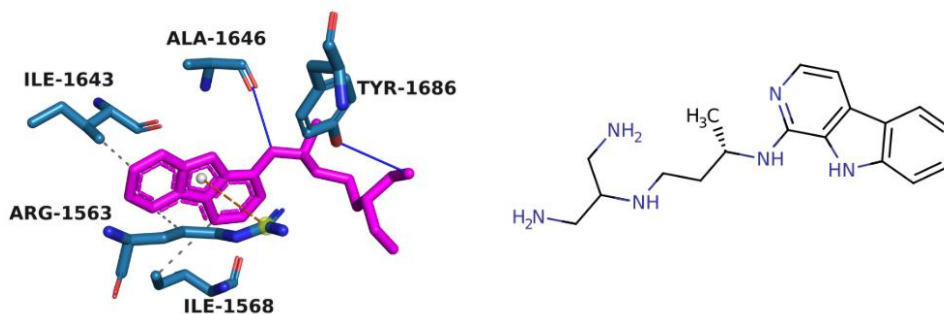
A arquitetura do KAST foi baseada em uma rede neural *Feedforward* aplicada a *fingerprints* moleculares. Cada composto foi convertido em vetor numérico, processado pelas camadas ocultas e, na saída, a função *Softmax* gerou o K-Predictive Score (KPS), variando entre 0 e 1, como estimativa probabilística de atividade biológica (Goodfellow, Bengio e Courville, 2016). A arquitetura do KAST foi baseada em uma rede neural *Feedforward* aplicada a *fingerprints* moleculares, com validação robusta utilizando métricas como AUC-ROC, Fator de Enriquecimento (EF) e validação cruzada com *scaffold splitting* (Oliveira *et al.*, 2023). O modelo apresentou desempenho notável, com

AUC de 0,99, acurácia de 0,97, precisão de 0,89, sensibilidade de 0,94, especificidade de 0,98 e F1-score de 0,91. A robustez foi confirmada por validação cruzada (AUC médio de $0,977 \pm 0,021$), e pelas curvas de aprendizagem, que indicaram bom ajuste sem evidência de *overfitting*.

Na aplicação em larga escala, o KAST avaliou o banco COCONUT, identificando 577 compostos como *hits* de alta probabilidade ($KPS > 0,7$). O composto KAST_402049 ($KPS = 0,99$) foi priorizado e otimizado via vBROOD 4.1, resultando em 1392 análogos, posteriormente submetidos a acoplamento molecular. A estrutura cristalográfica da PKS13-TE (PDB ID: 8Q17) foi preparada no UCSF Chimera 1.19. Os parâmetros estabelecidos no DOCK demonstraram o composto 95 (- 68,5) apresentou maior afinidade predita.

Delineando o mapa de interações intermoleculares (figura 2), nota-se que no domínio tioesterase (TE) da PKS13 é estabilizado por interações específicas. O núcleo aromático derivado de carbazol forma três contatos hidrofóbicos com ARG1563, ILE1568 e ILE1643, além de uma interação cátion- π com ARG1563. Complementarmente, o anel piperazínico estabelece duas pontes de hidrogênio com ALA1646 e TYR1686, contribuindo para a ancoragem e estabilidade conformacional do complexo.

Figura 2 – Representação 3D das interações intermoleculares da PKS13-TE com a molécula 95, acompanhada da representação 2D do mesmo composto.



NOTA: a linha azul contínua - ligações de hidrogênio; linha amarela tracejada - interações iônicas; linha cinza tracejada - interações hidrofóbicas; linha tracejada com esferas nas extremidades - interações do tipo π .

Fonte: o autor (2025).

Essas observações reforçam a relevância de explorar tais contatos no design racional de novos compostos, visando bloquear a biossíntese de ácidos micólicos essenciais à viabilidade do bacilo (Uppumavuluri *et al.*, 2023).

CONCLUSÃO

O estudo demonstrou que a integração de IA e métodos computacionais permitiu identificar a molécula 95 como hit promissor contra a PKS13-TE do *Mycobacterium tuberculosis*, com alta afinidade sustentada por interações hidrofóbicas, cátion- π e pontes de hidrogênio, validando seu potencial para desenvolvimento de fármacos antituberculose.

REFERÊNCIAS

CHEN, W. *et al.* Artificial intelligence for drug discovery: Resources, methods, and applications. **Molecular Therapy Nucleic Acids**, [s. l.], v. 31, p. 691–702, 2023.

GAULTON, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. **Nucleic Acids Research**, [s. l.], v. 45, n. D1, p. D945–D954, 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning. **Adaptive computation and machine learning**, [s. l.], 2016.

HARRIS, C. R. *et al.* Array programming with NumPy. **Nature**, [s. l.], v. 585, n. 7825, p. 357–362, 2020.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. Prentice Hall, 2008.

KIM, S. Exploring Chemical Information in PubChem. **Current Protocols**, [s. l.], v. 1, n. 8, p. e217, 2021.

LIU, T. *et al.* BindingDB in 2024: a FAIR knowledgebase of protein-small molecule binding data. **Nucleic Acids Research**, [s. l.], v. 53, n. D1, p. D1633–D1641, 2024.

MCKINNEY, W. Data Structures for Statistical Computing in Python. **Proceedings of the 9th Python in Science Conference**, [s. l.], v. 445, p. 51–56, 2010.

MOREIRA, A. da S. R.; KRITSKI, A. L.; CARVALHO, A. C. C. Determinantes sociais da saúde e custos catastróficos associados ao diagnóstico e tratamento da tuberculose. **Jornal Brasileiro de Pneumologia**, [s. l.], v. 46, p. e20200015, 2020.

NASCIMENTO, I. J. dos S.; AQUINO, T. M. de; SILVA-JÚNIOR, E. F. da. The New Era of Drug Discovery: The Power of Computer-aided Drug Design (CADD). **Letters in Drug Design & Discovery**, [s. l.], v. 19, n. 11, p. 951–955, 2022.

NEVES CRUZ, J.; OTHERS. Modelagem Molecular de Novos Inibidores da PKs13 do *Mycobacterium tuberculosis*. In: , 2018, Rio de Janeiro, Brasil. **58º Congresso Brasileiro de Química**. Rio de Janeiro, Brasil: [s. n.], 2018.

NICHOLLS, A. *et al.* Molecular replacement in ligand design: the BROOD program. **Journal of Chemical Information and Modeling**, [s. l.], v. 50, n. 7, p. 1161–1173, 2010.

OMS, O. M. da S. **Global Tuberculosis 2023 Report**. Geneva: WHO, 2023. Disponível em: <https://iris.who.int/bitstream/handle/10665/373828/9789240083851-eng.pdf?sequence=1>. Acesso em: 14 abr. 2025.

PEDREGOSA, F. *et al.* Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, [s. l.], v. 12, p. 2825–2830, 2011.

RAMSUNDAR, B. *et al.* **Deep Learning for the Life Sciences**. O'Reilly Media, 2019.

ROGERS, D.; HAHN, M. Extended-connectivity fingerprints. **Journal of Chemical Information and Modeling**, [s. l.], v. 50, n. 5, p. 742–754, 2010.

UPPUMAVULURI, N. T. *et al.* Pks13 inhibitors—a promising target for future antitubercular agents. **Medicinal Chemistry Research**, [s. l.], v. 32, p. 1574–1588, 2023.

VAN ROSSUM, G. Python Tutorial. **Centrum voor Wiskunde en Informatica (CWI), Amsterdam**, [s. l.], v. GVR, n. 10, 1995.

YANG, K. *et al.* Analyzing Learned Molecular Representations for Property Prediction. **Journal of Chemical Information and Modeling**, [s. l.], v. 59, n. 8, p. 3370–3388, 2019.

ZHANG, W. *et al.* Design and synthesis of mycobacterial pks13 inhibitors: Conformationally rigid tetracyclic molecules. **European Journal of Medicinal Chemistry**, [s. l.], v. 213, p. 113202, 2021.