

## **Nova abordagem eletrônica de construção de modelos de QSAR para o planejamento de inibidores de GSK-3 $\beta$ no âmbito da Doença de Alzheimer**

Daniel S. de Sousa<sup>1</sup>; Aldineia P. da Silva<sup>2</sup>; Albérico B. F. da Silva<sup>1</sup>

<sup>1</sup> Instituto de Química de São Carlos (IQSC), Universidade de São Paulo (USP), 13560-970, São Carlos, SP, Brasil;

<sup>2</sup> Escola de Artes, Ciências e Humanidades (EACH), Universidade de São Paulo (USP), São Paulo, 03828-000, Brasil.

**Palavras-Chave:** QSAR, Alzheimer, GSK-3 $\beta$ .

### **Introdução**

A doença de Alzheimer (DA) é a principal causa de demência no mundo, afetando atualmente cerca de 55 milhões de pessoas, com uma projeção de crescimento para 131,5 milhões de casos até 2050. Dada a sua enorme relevância, há uma necessidade urgente de tratamentos eficazes. Diversos estudos e estratégias estão sendo desenvolvidos, baseados em diferentes teorias sobre a etiologia da doença, que ainda não é totalmente compreendida. Pesquisas mostram que a hiperfosforilação da proteína tau, facilitada pela GSK-3 $\beta$ , está diretamente ligada à formação dos emaranhados neurofibrilares, uma das principais características patológicas da DA. Além disso, a GSK-3 $\beta$  também está envolvida na regulação da produção e deposição do beta-amiloide, outro componente fundamental das placas encontradas nos cérebros de pacientes com Alzheimer. Esses achados sugerem que a inibição da GSK-3 $\beta$  pode ser uma estratégia promissora para o tratamento da DA. Diante desse cenário, nossa pesquisa explora uma nova abordagem na descoberta de fármacos para a DA, utilizando modelos de QSAR baseados em Redes Neurais artificiais.

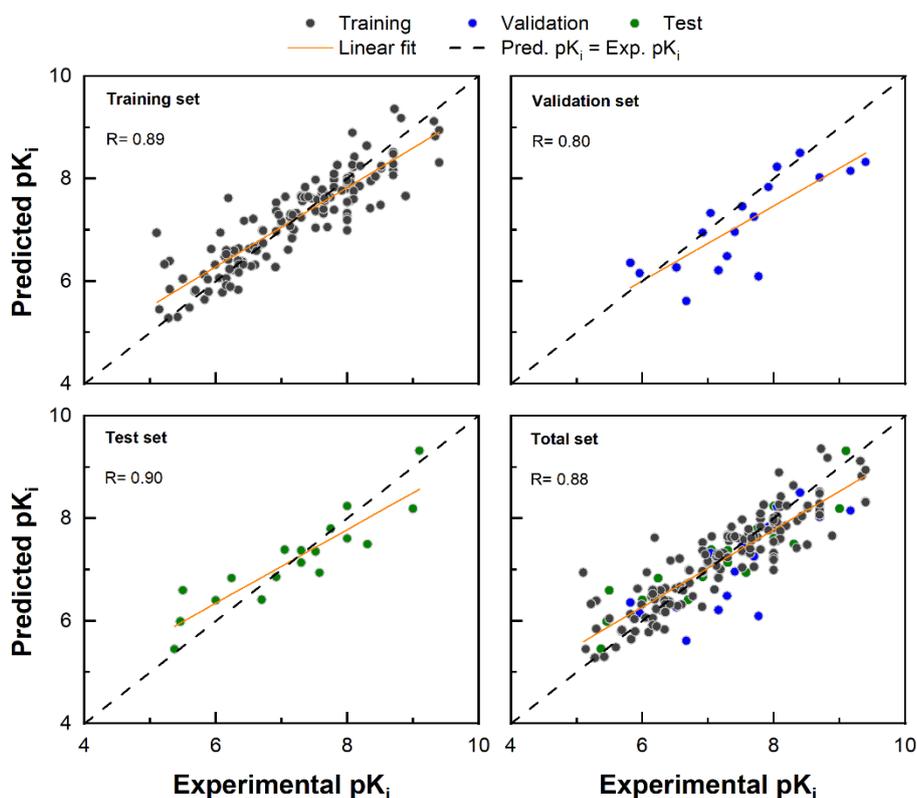
### **Material e Métodos**

Um conjunto de 180 compostos ativos de GSK-3 $\beta$  foi extraído do banco de dados ChEMBL e, posteriormente, as moléculas foram otimizadas tridimensionalmente utilizando a Teoria do Funcional da Densidade (DFT) com o software Gaussian16, empregando o funcional B3LYP e o conjunto de funções de base 6-311++G(d,p). Mais de 50 descritores eletrônicos, como HOMO, LUMO, dureza, polarizabilidade, afinidade eletrônica e energia de ionização, foram extraídos e submetidos a uma análise de seleção usando o coeficiente de correlação de Pearson para evitar redundâncias. Com os descritores eletrônicos selecionados, foram construídos modelos de QSAR por meio de redes neurais artificiais (RNA) no MatlabR2022b, ajustando parâmetros para otimizar a performance. O modelo final foi validado internamente por métodos como validação cruzada e y-randomization, e externamente com um conjunto de teste independente.

### **Resultados e Discussão**

Como principal resultado, em termos de desempenho, o modelo demonstrou características desejáveis em diferentes conjuntos de dados. Os coeficientes de correlação de Pearson, denotados por R, foram usados para medir o desempenho do modelo. Esses

coeficientes fornecem uma medida da correlação linear entre os valores preditos e os reais. Para os conjuntos de teste, treinamento e validação, os valores de  $R$  foram 0,90, 0,89 e 0,80, respectivamente. Esses valores estão bastante próximos do cenário ideal, onde o  $pK_i$  predito é igual ao  $pK_i$  experimental. Isso sugere que as previsões do modelo são não apenas robustas, mas também confiáveis, como mostrado na Figura 1.



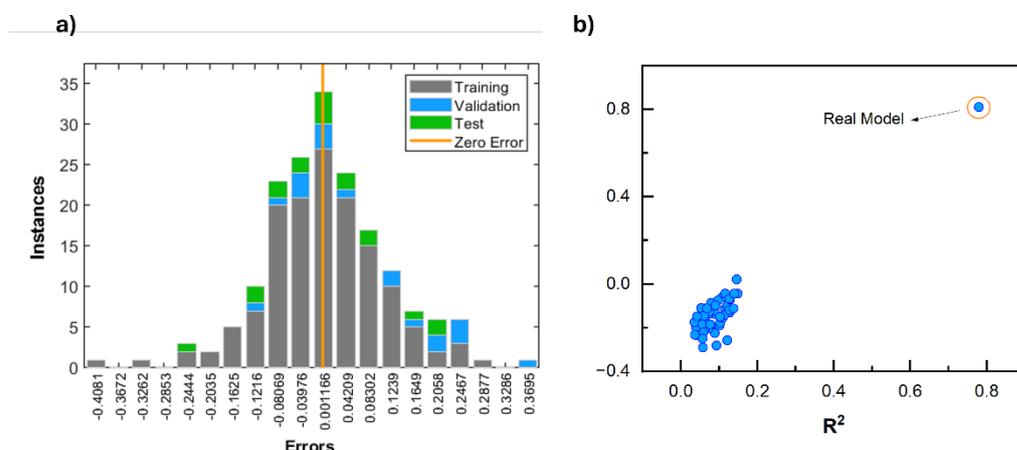
**Figura 1** – Capacidade preditiva do modelo de RNA obtido. Exp.  $pK_i$  vs Pred.  $pK_i$  do conjunto de treinamento (cinza escuro), validação (azul), teste (verde) e do conjunto total.

A análise dos erros gerados pelo modelo, mostrada na Figura 2a, revela uma distribuição bilateral em torno de zero para todos os conjuntos (treinamento, validação e teste). Esse padrão sugere uma semelhança com a distribuição normal, o que é um sinal positivo. Em uma distribuição normal, a maior parte dos dados está concentrada em torno da média, que, neste caso, é zero, indicando que a maioria dos erros gerados pelo modelo são pequenos, enquanto erros de maior magnitude são menos frequentes e aparecem nas extremidades da distribuição. Isso indica que o modelo tem um bom desempenho na maioria dos casos, com apenas alguns resultando em erros maiores.

A análise de  $y$ -randomization, apresentada na Figura 2b, envolveu a geração de 50 modelos com uma distribuição aleatória de  $y$ . Todos os modelos obtidos apresentaram métricas inferiores às de um modelo válido, distantes dos altos valores de  $R^2$  e  $Q^2$  do modelo real. Isso sugere que o modelo obtido não foi gerado por acaso.

Na fase de validação externa, foi observada uma correlação substancial de 0,90 no conjunto de teste, indicando que o modelo é eficaz em prever resultados com precisão em dados

independentes não utilizados durante a fase de treinamento. Isso confirma a capacidade robusta de generalização do modelo, reforçando sua utilidade e confiabilidade para aplicações futuras.



**Figura 2** – Validações do modelo: a) Histograma dos erros dos conjuntos de teste (verde), treinamento (cinza) e validação (ciano). 20 intervalos. b) Gráfico de y-Randomization com 50 modelos y-random avaliados pelas relações  $R^2$  e  $Q^2$ .

## Conclusões

Este estudo resultou na criação de um modelo preditivo altamente preciso para compostos ativos contra GSK-3 $\beta$ , evidenciando excelente desempenho em diferentes fases de validação. A otimização molecular via DFT e a seleção criteriosa de descritores eletrônicos com base na correlação de Pearson garantiram a robustez dos modelos de QSAR. Redes neurais artificiais foram ajustadas para maximizar o desempenho, alcançando coeficientes de correlação elevados ( $R \approx 0,90$ ) nos conjuntos de teste, treinamento e validação. Os resultados indicam que o modelo possui grande potencial para o design racional de novos compostos, acelerando o desenvolvimento de fármacos voltados à inibição da GSK-3 $\beta$ .

## Agradecimentos

Este estudo foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Bolsas 140894/2021-7 e 152105/2022-0). Os autores agradecem o apoio da FAPESP. A pesquisa foi realizada utilizando os recursos computacionais do Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI), que foi financiado pela FAPESP (bolsa 2013/07375-0).

## Referências

Alzheimer's & Dementia 2023, 19, 1598–1695.

CHAUHAN, Neha et al. GSK-3 $\beta$  and its inhibitors in Alzheimer's disease: a recent update. **Mini reviews in medicinal chemistry**, v. 22, n. 22, p. 2881-2895, 2022.

DA SILVA, Aldineia P. et al. Drug design of new 5-HT<sub>6</sub>R antagonists aided by artificial neural networks. **Journal of Molecular Graphics and Modelling**, v. 104, p. 107844, 2021.