10 AlphaFold e a busca pelo Santo Graal da Biologia Molecular

Autores 10.1

Diego Mariano

Revisão: Wylerson Guimarães Nogueira

Resumo 10.1

Recentemente, o uso de técnicas de inteligência artificial tem proporcionado um grande avanço para a solução de um problema da biologia estrutural estudado há mais de 50 anos, trazendo novas perspectivas para a compreensão de doenças e para a descoberta de novos medicamentos.

No começo de 2020, o mundo científico foi surpreendido com a possível solução de um dos maiores desafios da biologia moderna: o enovelamento de proteínas [1]. Munidos de técnicas de inteligência artificial, a DeepMind, uma empresa britânica adquirida pela Google em 2014, afirmou ter conseguido resultados excepcionais no desafio de modelagem de estruturas 3D de proteínas.

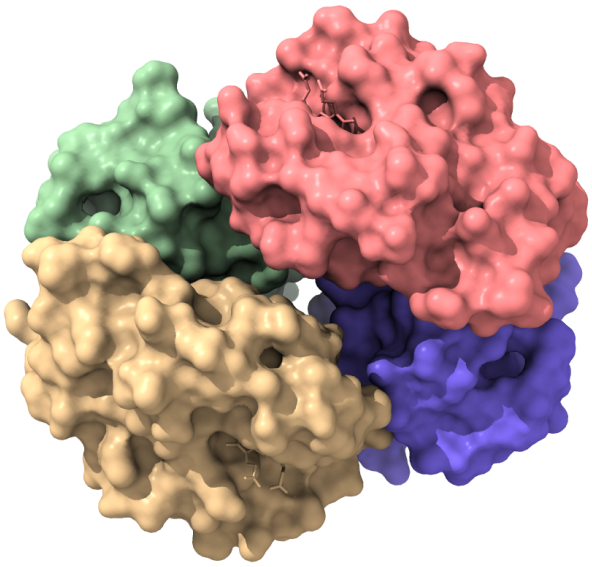

Figura 10.1: Estrutura 3D da proteína hemoglobina, responsável por transportar oxigênio dos pulmões para o corpo. Fonte: Próprio autor (figura gerada com ChimeraX [2]).

Proteínas são moléculas essenciais para a vida. Nos seres vivos, elas têm variadas funções, desde o controle do processo de respiração até a proteção do nosso organismo contra invasores. Como a forma e composição de uma proteína está diretamente ligada à sua função, estabelecer a estrutura 3D de uma proteína

pode ser essencial, por exemplo, para avaliar o impacto de mutações e auxiliar na descoberta de novos medicamentos. No entanto, obter a estrutura por meio de experimentos em laboratório é bastante custoso e demorado. Para se ter uma noção, o banco de dados de proteínas PDB armazena cerca de 180 mil estruturas resolvidas experimentalmente, uma ínfima porção se levarmos em consideração as bilhões de proteínas existentes. Assim, o uso de simulações computacionais se tornou essencial para obtenção de estruturas 3D em larga escala. Mas a tarefa se mostrou muito mais complexa do que o imaginado, sendo até mesmo chamada de a busca do Santo Graal da biologia molecular [3], em referência ao cálice sagrado supostamente usado por Cristo na última ceia.

À medida que proteínas são codificadas nas células humanas (ou de outros organismos), partes delas se dobram em diferentes posições até adquirir uma conformação mais estável, na qual a proteína estará apta a desempenhar suas funções. Esse processo é denominado **enovelamento da proteína**. O desafio computacional aqui está em prever qual será a forma final dessa proteína com base em sua sequência (que nos dias de hoje pode ser mais facilmente obtida devido a evolução das tecnologias de sequenciamento de DNA).

Desde 1994, uma competição realizada a cada dois anos, denominada CASP (uma sigla que poderia ser traduzida como "avaliação crítica de previsão de estrutura de proteínas"), busca incentivar cientistas a desenvolver algoritmos de programas de computador para prever a estrutura final. Entretanto, pouco sucesso havia sido obtido até 2018, quando a DeepMind apresentou o AlphaFold, um programa para modelagem de estruturas 3D usando aprendizagem profunda, uma técnica da inteligência artificial.

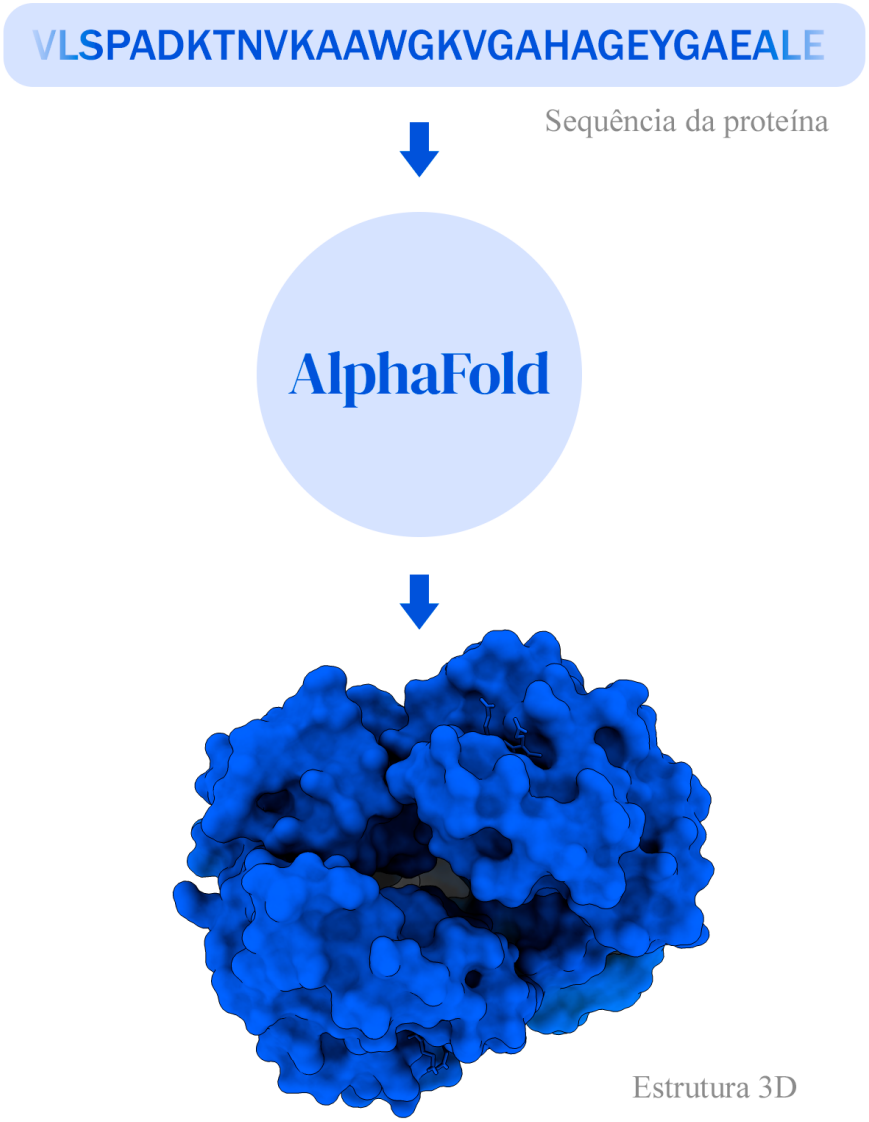

Figura 10.2: Esquema básico de funcionamento do AlphaFold. Parte da sequência de uma proteína (obtida por alguma das várias tecnologias de sequenciamento disponíveis) é usada como entrada. A seguir, AlphaFold utiliza seu algoritmo para predição da conformação 3D da proteína. Fonte: próprio autor (figura gerada com ChimeraX [2])

Na última edição do CASP (2020), a segunda versão do AlphaFold obteve em média 87 pontos na categoria modelagem livre (numa escala que vai de 0 a 100). Na prática, isso indica que o programa foi capaz de reconstruir as estruturas 3D de uma série de proteínas com base apenas em suas sequências com uma taxa de

acertos nunca antes vista. Isso representou um salto surpreendente até mesmo para o resultado do próprio AlphaFold na edição anterior.

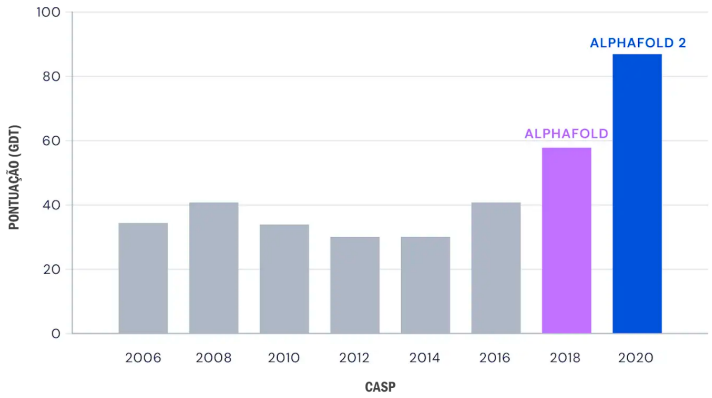

Figura 10.3: Maior pontuação média na categoria de modelagem livre nas edições do CASP (2006-2020). Observe que AlphaFold apresenta um resultado muito superior aos vencedores das edições passadas. Fonte: traduzido do blog da DeepMind [4].

Em 15 de julho de 2021, a DeepMind publicou um artigo na prestigiada revista Nature[5], descrevendo detalhes da implementação do AlphaFold. Os autores ainda disponibilizaram uma base de dados com mais de 350 mil estruturas do proteoma humano (cerca de 98,5% das proteínas humanas) e de outros organismos modelo [6]. Além disso, a DeepMind ainda disponibilizou o código fonte do AlphaFold, permitindo que qualquer cientista do mundo consiga usá-lo em seus experimentos.

Isso pode trazer resultados imediatos para diversas pesquisas em andamento. Como por exemplo, os estudos de mutações em proteínas do coronavírus e seu impacto na ligação com células humanas. Ou ainda no estudo de diversas doenças hereditárias, como Alzheimer, Parkinson, doenças cardíacas e diversos tipos de câncer, trazendo esperança de novos tratamentos.

De fato, estamos diante de uma das maiores revoluções na ciência moderna. Um triunfo da computação sobre os desafios da biologia. Mas ainda há um longo caminho a ser percorrido. A modelagem computacional apenas fornece meios para realização de estudos baseados em dados. Contudo, graças às

novas tecnologias, estratégias e programas computacionais para análise de dados biológicos, já podemos vislumbrar um futuro promissor.

Bloco 10.1

Este artigo está disponível em https://bioinfo.com.br/alphafold-e-a-busca-pelo-santo-graal-da-biologia-molecular/

10.1 Referências

[1] GOMES, H. 2020 também será lembrado por salto da inteligência artificial que pode transformar a medicina. Folha de S. Paulo. Disponível em https://www1.folha.uol.com.br/colunas/henrique-gomes/2020/12/2020-tambem-sera-lembrado-por-salto-da-inteligencia-artificial-que-pode-transformar-a-medicina.shtml ↑

[2] UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. Protein Sci. 2021 Jan;30(1):70-82. ↑

[3] WHIPPLE, T. Deepmind finds biology's 'holy grail' with answer to protein problem. The Times. Disponível em https://www.thetimes.co.uk/article/deepmind-finds-biology-s-holy-grail-with-answer-to-protein-problem-htg6s7qlq ↑

[4] The AlphaFold team. AlphaFold: a solution to a 50-year-old grand challenge in biology. Blog DeepMind. November 30, 2020. Disponível em https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology ↑

[5] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2. Disponível em https://www.nature.com/articles/s41586-021-03819-2 ↑

[6] Callaway, Ewen. DeepMind's AI predicts structures for a vast trove of proteins. Nature 595, 635 (2021). Disponível em https://www.nature.com/articles/d41586-021-02025-4 ↑