

Projetando uma Insulina Mais Estável: Algoritmos Genéticos e Dinâmica Molecular Aplicados à Engenharia de Proteínas

Gabriel Fruet, Mateus Ribeiro

15 de fevereiro de 2025

1 Introdução

A insulina é um hormônio essencial no controle da glicemia e seu uso terapêutico é indispensável para milhões de pessoas com diabetes ao redor do mundo. No entanto, a estabilidade térmica da insulina representa um desafio significativo, especialmente em regiões de clima quente ou com acesso limitado à refrigeração. A degradação da insulina em temperaturas elevadas compromete sua eficácia, exigindo soluções que aumentem sua estabilidade sem comprometer sua funcionalidade biológica.

Os algoritmos genéticos (AGs) têm se destacado como uma abordagem promissora para problemas de otimização em bioinformática e engenharia de proteínas. Inspirados na seleção natural, esses algoritmos simulam processos evolutivos para encontrar soluções otimizadas dentro de um espaço de busca complexo. No contexto da engenharia de proteínas, os AGs podem ser empregados para identificar mutações que conferem maior estabilidade térmica à insulina, preservando sua atividade biológica.

Este trabalho explora a aplicação de algoritmos genéticos na identificação de mutações que aumentem a termoestabilidade da insulina. Para isso, utilizamos modelos computacionais para simular mutações e avaliar seu impacto na estabilidade da proteína. A metodologia proposta visa acelerar o desenvolvimento de variantes mais estáveis da insulina, contribuindo para avanços na terapia do diabetes e na biotecnologia farmacêutica.

2 Materiais e Métodos

2.1 Bibliotecas e Tecnologias

Os experimentos foram realizados em computador com uma NVIDIA GeForce RTX 3060 com 12GB de VRAM e um processador Ryzen 5 5600, para aceleração das predições estruturais via ESMFold, foi utilizado CUDA. O pipeline foi implementado em Python, integrando as bibliotecas PyGAD, PyTorch, BioPython, OpenMM e outras dependências.

2.2 População Inicial

A sequência primária da insulina foi utilizada como ponto de partida para a geração de variantes mutantes. Cada indivíduo na população inicial corresponde a uma variante da insulina onde apenas um único aminoácido é alterado. As sequências geradas são representadas em formato FASTA para facilitar a manipulação computacional.

2.3 Cálculo da Estrutura Conformacional

Ao trabalharmos com proteínas, por vezes se faz necessário ter o conhecimento da sua estrutura conformacional. Porém, o conhecimento prévio desse dado nem sempre está disponível, especialmente quando trabalhos com proteínas mutantes.

Atualmente, temos diversos modelos computacionais que conseguem prever a estrutura conformacional de uma proteína, tais como AlphaFold [?], OmegaFold [?], RoseTTAFold [?], ESMFold [?]. No entanto, a maioria desses modelos exigem um grande poder computacional, como o AlphaFold, que exige em torno de 80GB de memória VRAM, além de disco rígido na ordem de terabytes para armazenar os banco de dados de proteínas usados no *multiple sequence alignments* (MSA).

Com esse problema em vista, o ESMFold foi proposto. O principal diferencial do ESMFold é a quantidade menor de parâmetros, no nosso caso, 3 bilhões, necessitando de um hardware menos potente e a predição sem utilizar MSA. Isso faz com que o ESMFold consiga ser executado na ordem de poucos segundos em GPUs de entrada.

2.4 Cálculo da Termoestabilidade

A termoestabilidade de proteínas pode ser avaliada por métodos como a energia livre de dobramento ΔG , a temperatura de desnaturação (T_m), a flutuação quadrática média dos átomos (RMSF) e a fração de contatos nativos (QQ). A escolha do método depende do sistema estudado, dados experimentais e objetivos da análise.

A ΔG é uma medida direta da estabilidade, mas requer simulações avançadas. O T_m é utilizado em estudos experimentais, mas precisa de simulações em várias temperaturas. A RMSF mede a flexibilidade local, mas não a estabilidade global.

Optamos pela fração de contatos nativos (QQ) devido à sua simplicidade, eficiência e capacidade de quantificar a preservação da estrutura nativa em simulações de dinâmica molecular. QQ reflete a proporção de contatos atômicos mantidos em relação à estrutura nativa, sendo robusta a variações nas condições de simulação.

O cálculo de QQ foi feito com a biblioteca MDTraj [?], utilizando a estrutura nativa como referência e o esquema "closest-heavy" para contatos atômicos. Três réplicas independentes foram simuladas para cada sistema, com os valores de QQ reportados como média \pm desvio padrão. As simulações usaram o campo de forças AMBER14 [?] e o modelo de água TIP3P [?].