



Evento	Salão UFRGS 2015: SIC - XXVII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2015
Local	Porto Alegre - RS
Título	Desenvolvimento de Algoritmo Genético Distribuído para a Predição da Estrutura Tridimensional de Proteínas
Autor	JONAS DA SILVEIRA BOHRER
Orientador	MARCIO DORN

Desenvolvimento de Algoritmo Genético Distribuído para a Predição da Estrutura Tridimensional de Proteínas

*Autor: Jonas da S. Bohrer, Orientador: Márcio Dorn
Instituto de Informática, Universidade Federal do Rio Grande do Sul*

O conhecimento sobre a estrutura tridimensional (3D) de proteínas é, atualmente, um tópico extremamente importante em vários campos da ciência. Proteínas são macromoléculas biológicas constituídas por uma ou mais cadeias de resíduos de aminoácidos. Estas macromoléculas exercem importantes papéis na natureza e sua função está diretamente relacionada a sua estrutura 3D. Desta forma, ao conhecer a estrutura tridimensional é possível desenvolver compostos químicos (fármacos) que possam inibir ou ativar a sua função no organismo. Entre os métodos experimentais mais comuns utilizados na determinação da estrutura 3D de uma proteína, destacam-se a Cristalografia de raios-X e a Ressonância Magnética Nuclear (RMN). Apesar da qualidade das estruturas obtidas, o alto custo e tempo demandados por estas técnicas impossibilita que se obtenha, em alguns casos, a estrutura 3D de uma proteína. Em função disso, atualmente existe uma enorme lacuna entre o número de sequências de proteínas e o número de estruturas tridimensionais que são conhecidas. A análise e tentativa de resolução desses problemas através de outras técnicas, especialmente por meios computacionais, é uma saída para superar as limitações das técnicas experimentais.

Na Bioinformática, a predição da estrutura 3D de Proteínas é um problema complexo que vêm sendo estudado há décadas e que ainda possui muito a ser explorado. Dentre as diversas estratégias existentes que abordam o problema, destaca-se a utilização de algoritmos evolutivos. A ideia básica desses algoritmos é a de manter uma população de indivíduos, representando soluções candidatas para o problema, que evolui ao longo de gerações através de um processo de competição, onde os mais aptos (segundo um padrão definido pelo problema) têm maiores chances de sobreviver e se reproduzir. Neste trabalho, é proposto um algoritmo genético distribuído baseado em conhecimento para o problema da predição. Um algoritmo genético (GA) é uma técnica de busca estocástica utilizada na computação de soluções em problemas de otimização. Nestes algoritmos, utiliza-se uma população composta por um conjunto de indivíduos (soluções) que "evolui" durante um número definido de gerações ou de tempo, através de operações de recombinação (crossover) e mutação. Visando a melhora da diversidade e qualidade das soluções, também foi implementada a distribuição do algoritmo em mais de uma população, delegando diferentes populações à processos paralelos que interagem entre si, periodicamente. Essa modificação favorece a variabilidade das soluções devido ao aumento do ambiente de interações entre indivíduos. O algoritmo foi implementado na linguagem Python, e utiliza informações estruturais obtidas do RCSB Protein Data Bank (PDB). Para avaliação das soluções, foi utilizada a função de energia potencial do Rosetta, e consideram-se como mais viáveis as soluções com menor energia potencial.

O método proposto foi testado com uma sequência de 8 proteínas com estruturas já disponíveis no RCSB Protein Data Bank (PDB). O algoritmo de predição foi executado 16 vezes, para cada proteína, durante 12 horas em cada execução. Os resultados obtidos foram analisados em termos de qualidade estrutural e de estereoquímica. Após estas análises, concluímos que as estruturas são similares e comparáveis às suas correspondentes experimentais. O aprimoramento do método atual através da utilização de outras variações de algoritmos evolutivos, implementação de modelos híbridos e variações nas interações entre populações e dados iniciais poderão ser estudados em trabalhos futuros.