Autor: Rafael Cauduro Oliveira Macedo

Título: On the Analysis of Remd Protein Structure Prediction Simulations for Reducing Volume of Analytical Data.

Ingresso: 2015/2

Defesa:30/08/2017, ata nr. 520 **Homologação:** xx/xx/xxxx ata nr./

Linha de Pesquisa: Bioinformática e Computação Bioinspirada.

Orientador: Prof. Dr. Osmar Norberto de Souza

Comissão Examinadora: Prof. Dr. Duncan Dubugras Alcoba Ruiz (PPGCC/PUCRS) e Prof. Dr. Rafael Andrade Caceres (UFCSPA).

Resumo:

Proteínas executam um papel vital em todos os seres vivos, mediando uma série de processos necessários para a vida. Apesar de que existam maneiras de determinar a composição dessas moléculas, ainda falta-nos conhecimentos suficiente para determinar de uma maneira rápida e barata a sua estrutura 3D, que desempenha um papel importante na suas funções. Um dos principais métodos computacionais aplicados ao estudo das proteínas e o seu processo de enovelamento, o qual determina a sua estrutura, é Dinâmica Molecular. Um aprimoramento deste método, conhecido como Replica Exchange Molecular Dynamics (ou REMD), é capaz de produzir resultados muito melhores, com o revés de significativamente aumentar o seu custo computacional e gerar um volume muito maior de dados. Esta dissertação apresenta um novo método de otimização deste método, intitulado Filtragem de Dados Analíticos, que tem como objetivo otimizar a análise pós-simulação filtrando as estruturas preditas insatisfatórias através do uso de métricas de qualidade absolutas. A metodologia proposta tem o potencial de operar em conjunto com outras abordagens de otimização e também cobrir uma área ainda não abordada por elas. Adiante, a ferramenta SnapFi é apresentada, a qual foi designada especialmente para o propósito de filtrar estruturas preditas insatisfatórias e ainda operar em conjunto com as diferentes abordagens de otimização do método REMD. Um estudo foi então conduzido sobre um conjunto teste de simulações REMD de predição de estruturas de proteínas afim de elucidar uma séries de hipóteses formuladas sobre o impacto das diferentes temperaturas na qualidade final do conjunto de estruturas preditas do processo REMD, a eficiência das diferentes métricas de qualidade absolutas e uma possível configuração de filtragem que utiliza essas métricas. Foi observado que as temperaturas mais altas do método REMD para predição de estruturas de proteínas podem ser

descartadas de forma segura da análise posterior ao seu término e também que as métricas de qualidade absolutas possuem uma alta variância (em termos de qualidade) entre diferentes simulações de predições de estruturas de proteínas. Além disso, foi observado que diferentes configurações de filtragem que utilize tais métricas carrega consigo esta variância.