



Ciências Exatas e da Terra

USO DA LINGUAGEM “R” PARA ANÁLISE ESTATÍSTICA DE PARÂMETROS BIOLÓGICOS DE SIMULAÇÕES COMPUTACIONAIS

Gabriel Almeida Escodino da Silva, Jorge Hernandez Fernandez

R é uma linguagem de programação e um ambiente voltado para computação estatística que permite a realização de análises de dados complexos, seja através da geração de gráficos, relatórios, realizações de cálculos, dentre outras ferramentas matemáticas. A linguagem vem se mostrando poderosa para ser utilizada na resolução de problemas específicos de áreas de estudo que envolvem análise, mineração e tratamento de informações. Neste contexto, surgem as aplicações da linguagem na bioinformática, área onde constantemente surge a necessidade de ferramentas computacionais com integridade e confiabilidade para facilitar o tratamento de dados obtidos em experimentos biológicos. Esta pesquisa terá como objetivo construir ferramentas para análise estatística de dados obtidos em experimentos de dinâmica molecular, o que aperfeiçoará o entendimento dos processos estudados e auxiliará o avanço nas técnicas de experimentação “in silico” no LQFPP-UENF. Considerando que o projeto se encontra em fase inicial, foi feita uma pesquisa bibliográfica relacionada aos aspectos fundamentais de estatística e aos conceitos básicos de funcionamento de R. Após esse embasamento teórico, foi feita a instalação do R para estudos iniciais, familiarização com a linguagem e realização de testes de geração de gráficos simples utilizando as funções do R-base, além da busca por pacotes da linguagem que poderiam possuir aplicação prática na pesquisa. O primeiro resultado preliminar foi a criação de um script para geração de Ramachandran Plot a partir da leitura de dados inseridos em um arquivo de texto no padrão gerado pelo programa de dinâmica molecular GROMACS 4.2. Atualmente tentamos correlacionar áreas de maior densidade de pontos no Ramachandran Plot gerado com qualidade de estruturas proteicas. Em uma etapa posterior, as ferramentas de análise desenvolvidas serão organizadas e distribuídas para testes e uso contínuo com usuários da PGBB-CBB-UENF.

Palavras-chave: “R”, Bioinformática, Análise estatística de dados

Instituição de fomento: FAPERJ
UENF