

008

MODELAGEM E CLASSIFICAÇÃO DE PROTEÍNAS POR CADEIAS DE MARKOV E REDES NEUROFUZZY. *Roberto de Castro Parreiras, Daniella Lima Melo Coelho, Thiago de Souza Rodrigues (orient.) (INC).*

Pesquisas na área da biologia realizadas nos últimos anos, em especial o Projeto Genoma, deram origem a uma grande massa de dados, cujo processamento está acima da capacidade humana, tornando fundamental o desenvolvimento de ferramentas para auxiliar no estudo de toda esta informação. Uma destas ferramentas é o alinhamento de seqüências, que procura determinar o grau de similaridade entre duas ou mais seqüências, ou a similaridade entre fragmentos destas, o que permiti à biologia determinar se estas são homólogas, revelando informações sobre suas funções, estrutura e características a princípio desconhecidas. O principal método de alinhamento é conhecido como alinhamento par-a-par, onde o procedimento é comparar duas ou mais seqüências através da busca de uma série de caracteres individuais ou padrões de caracteres que estejam na mesma ordem nas seqüências comparadas. Entretanto, este método de alinhamento não se mostra eficiente em alguns casos onde proteínas não apresentam homologia com nenhuma seqüência armazenada em bancos de dados públicos. Este projeto visa modelar um conjunto de seqüências através de cadeias de Markov e verificar a eficiência do modelo comparando com técnicas de alinhamento global e local, inclusive heurísticas. Posteriormente, utilizar este modelo para classificar tais proteínas através de Redes NeuroFuzzy e comparar sua eficiência com resultados de estudos anteriores. A seleção dos dados foi realizada no banco de dados público de proteínas de procariotos COG (<http://www.ncbi.nlm.nih.gov/COG>), foi implementada a cadeia de Markov de ordem 1 até 10, sendo a próxima etapa a implementação do teste estatístico Qui-quadrado e análise dos resultados.