

Aplicação de Técnicas de Inteligência Artificial e Mineração de Dados no Design de Proteínas

Rafael K. Andrades, Márcio Dorn, Daniel S. Farenzena e Luis C. Lamb
Instituto de Informática, UFRGS
Financiamento: CNPq e CAPES

Uma proteína é um polímero formado por aminoácidos ligados por meio de uma ligação peptídica. A cadeia polipeptídica de uma proteína assume uma estrutura tridimensional (3D) que associa funções biológicas à proteína, tais como: transporte, síntese e catálise. As atuais técnicas laboratoriais para o estudo da estrutura de proteínas possuem limitações. Estas limitações justificam a aplicação de técnicas computacionais para o estudo destas importantes moléculas. O problema da predição da estrutura 3D de proteínas a partir da seqüência de aminoácidos e o problema do design de proteínas, cujo objetivo é deduzir a seqüência de aminoácidos correspondentes a uma determinada estrutura 3D são problemas de alta complexidade computacional. A indústria farmacêutica tem grande interesse no desenvolvimento de novos métodos computacionais para a síntese de novas proteínas. Neste trabalho, técnicas de Mineração de Dados e Inteligência Artificial são combinadas para reduzir o espaço-solução do problema de design de proteínas. O objetivo desta pesquisa é desenvolver novos algoritmos computacionais para a determinação de seqüências de aminoácidos de biomoléculas cujas estruturas 3D e suas respectivas funções são conhecidas. O método proposto possui cinco etapas: (1) Extração de informações 3D da proteína-alvo; (2) Seleção e agrupamento das informações obtidas a partir de modelos do Protein Data Bank (PDB); (3) Codificação do espaço-solução em uma Forma Normal Conjuntiva (FNC) para representação ótima do espaço-solução; (4) Restrição do espaço conformacional da proteína através do método Dead-End Elimination (DEE); e (5) Geração das sequências-solução da FNC para determinação da seqüência da proteína-alvo. Para avaliação do método proposto foram realizados testes com diferentes proteínas-alvo: PDB ID: 2OMM, 1OEH, 3DI0, 1ZDD e 1J9I. Os resultados obtidos revelam uma taxa de sucesso de 50% a 85% na determinação das seqüências de aminoácidos. Os resultados também apresentam performance computacional compatível ou superior a outros métodos tipicamente utilizados e descritos na literatura.