

1

Introdução

As proteínas são moléculas que desempenham um papel importante no funcionamento dos organismos vivos. Nos seres humanos, por exemplo, os anticorpos são proteínas responsáveis pelo reconhecimento de “invasores”. Os diversos hormônios também são proteínas. Existem diversas outras proteínas responsáveis pelo controle de funções vitais do organismo. Apesar dos mecanismos de produção e ação das proteínas não serem totalmente compreendidos ainda, sabe-se que elas são produzidas no interior das células. No mais baixo nível, é uma molécula presente em praticamente todas as células do organismo e denominada ácido desoxirribonucleico (DNA) a responsável por determinar o conteúdo e, conseqüentemente, a função das proteínas produzidas pela célula. Por isto, pesquisas sobre o DNA são tão importantes para compreender o funcionamento dos organismos vivos.

Uma molécula de DNA é composta por outras moléculas denominadas nucleotídeos e pode ser representada por uma palavra sobre o alfabeto {A, C, G, T}, onde cada letra representa um dos quatro tipos de nucleotídeos presentes no DNA. Esta palavra define a seqüência em que os nucleotídeos estão organizados na molécula de DNA. O seqüenciamento de DNA é uma tarefa cujo objetivo é determinar a seqüência de nucleotídeos de uma dada molécula de DNA. O seqüenciamento por hibridação (SBH) é uma técnica ainda em desenvolvimento para realizar o seqüenciamento de DNA. O SBH pode superar as técnicas utilizadas hoje em dia em termos de tempo e custo. Ele consiste de duas fases. Na primeira fase realiza-se um experimento bioquímico com a molécula de DNA que se deseja estudar (DNA alvo). Este experimento é capaz de revelar as subseqüências de nucleotídeos presentes no DNA alvo que casam com um determinado padrão. A segunda fase do SBH consiste em determinar a ordem em que estas subseqüências aparecem na seqüência de nucleotídeos do DNA alvo. Este problema é denominado problema de montagem e pode ser formulado como um problema de otimização combinatória que é um caso particular do problema do caixeiro viajante com coleta de prêmios.

O principal objetivo deste trabalho consiste em desenvolver heurísticas para solucionar o problema de montagem utilizando a formulação citada acima. As heurísticas propostas nesta dissertação são métodos multi-partida com algumas melhorias. Uma heurística multi-partida utiliza um algoritmo construtivo aleatorizado para gerar várias soluções e retorna a melhor delas. O algoritmo construtivo aleatorizado proposto aqui parte de uma solução vazia e a cada iteração seleciona um novo elemento para entrar na solução em construção dentre uma lista dos melhores elementos disponíveis. Desta forma, o algoritmo é capaz de construir soluções diferentes a cada execução e, ao mesmo tempo, realizar uma intensificação na região do espaço de soluções que apresenta-se mais promissora.

Dois métodos propostos para melhorar o desempenho de heurísticas para problemas de otimização combinatória, denominados memória adaptativa e construção de vocabulário, foram adaptados e combinados com a heurística multi-partida. A memória adaptativa é baseada em um conjunto de soluções de elite que armazena as melhores soluções encontradas no decorrer da heurística multi-partida. O algoritmo construtivo aleatorizado é modificado para intensificar suas construções, de maneira a privilegiar os elementos mais freqüentes nas soluções do conjunto elite. O método de construção de vocabulário consiste em gerar soluções combinando partes de outras soluções. As partes utilizadas são aquelas mais comuns nas melhores soluções construídas anteriormente na heurística multi-partida.

As heurísticas propostas neste trabalho alcançaram resultados expressivos. Para diversos grupos de instâncias estas heurísticas superaram os resultados dos melhores algoritmos encontrados na literatura, tanto em termos da qualidade das soluções obtidas como do tempo computacional. Estes resultados contribuem para o desenvolvimento da segunda fase do SBH, que é imprescindível para seu sucesso. Os métodos de memória adaptativa e construção de vocabulário ainda são pouco explorados na literatura. Eles são particularmente indicados para problemas de escalonamento e seqüenciamento, onde as soluções ótimas geralmente são compostas de blocos de tarefas ou de elementos que aparecem com freqüência em boas soluções. As implementações propostas para os dois métodos contribuem para sua validação, assim como para um melhor entendimento do seu comportamento. Os resultados alcançados destacam a capacidade destes métodos melhorarem o desempenho de heurísticas multi-partida.

No próximo capítulo apresenta-se a técnica de seqüenciamento por hibridação, o problema combinatório envolvido e a formulação utilizada. Também discutem-se as dificuldades envolvidas nesta técnica e outras de

suas aplicações. No Capítulo 3 são apresentados uma revisão bibliográfica dos algoritmos para o problema de montagem e as novas heurísticas propostas neste trabalho. Resultados de experimentos computacionais realizados com as heurísticas propostas são discutidos no Capítulo 4. Neste capítulo os resultados destas heurísticas também são comparados com os resultados obtidos pelos melhores algoritmos disponíveis na literatura. Finalmente, no Capítulo 5 apresentam-se conclusões e possíveis trabalhos futuros.