

APLICAÇÃO DE EVOLUÇÃO DIFERENCIAL EM GPU PARA O PROBLEMA DE PREDIÇÃO DE ESTRUTURA DE PROTEÍNAS COM MODELO 3D AB OFF-LATTICE¹

André Eduardo Pacheco Dias², Rafael Stubs Parpinelli³.

¹ Vinculado ao projeto “Predição de estrutura de proteínas utilizando algoritmos bio-inspirados”

² Acadêmico (a) do Curso de Ciência da Computação – CCT– Bolsista PROBIC

³ Orientador, Departamento de Ciência da Computação –CCT– rafael.parpinelli@udesc.br

Proteínas são macromoléculas compostas por cadeias de aminoácidos presentes em todos os organismos vivos, sendo responsáveis por diversas funções celulares como transporte, regulação, proteção, mobilidade, suporte estrutural, entre outras. A função que uma proteína exerce está diretamente relacionada com a sua estrutura tridimensional e cada proteína possui uma estrutura nativa onde a energia livre de Gibbs é mínima. Tal estrutura pode ser descoberta por métodos experimentais como Difração de Raios X e Espectroscopia por Ressonância Magnética Nuclear, mas estes métodos são caros e demorados. Por isso, métodos computacionais são empregados para tal finalidade e estes podem ser divididos em métodos que utilizam informações de estruturas de proteínas já conhecidas (*homology* e *threading*) ou que buscam representar matematicamente as propriedades físico-químicas envolvidas no processo de dobramento da proteína (*ab-initio*). O presente trabalho utiliza o modelo 3D AB *Off-Lattice* que considera a polaridade dos átomos para avaliar possíveis conformações de proteínas, a partir da observação que os aminoácidos apolares são hidrofóbicos e tendem a se concentrar no interior da proteína enquanto os hidrofílicos na parte mais externa. Para otimizar a equação de energia livre proposta pelo modelo um algoritmo evolutivo que busca imitar o processo de seleção natural onde os indivíduos mais bem adaptados ao ambiente sobrevivem para as próximas gerações foi implementado. Este algoritmo denominado DSMDE é uma variante do DE (Evolução Diferencial) e insere o conceito de espécies, agrupando indivíduos similares e considerando a espécie de cada um na etapa de geração de novos indivíduos de cada geração. Além disso, algumas rotinas auxiliares foram implementadas: uma estratégia de autoajuste dos parâmetros de controle; uma estratégia de *crowding* para a etapa de seleção que evita com que a população se concentre em uma mesma região do espaço de busca; uma técnica de busca local em vizinhança que desloca aminoácidos a fim de encontrar soluções superiores; uma estratégia de reinicialização parcial da população em caso de estagnação da melhor solução. Todo o algoritmo é escrito em CUDA C++ para ser executado em unidades de processamento gráfico de propósito geral (GPGPU), aproveitando o paralelismo massivo presente nestes dispositivos.

Para validar o algoritmo implementado, denominado como cuDSMjDE, duas proteínas reais foram otimizadas e comparadas com os resultados dos trabalhos estado-da-arte da otimização do modelo 3D AB *Off-Lattice*. O protocolo de experimentação foi de dez execuções com um critério de parada de três horas de execução, enquanto os demais trabalhos comparados otimizaram cada proteína por trinta execuções com um tempo limite de quatro dias para cada execução. Os resultados alcançados, exibidos na Tabela 1, foram superiores aos algoritmos DE_{pfo} e DE_{lsrc} e competitivo com o DE_{2L}. A maior proteína otimizada (2EWH) de 98 aminoácidos

conseguiu alcançar um valor de energia livre superior aos demais algoritmos. Uma comparação do tempo de execução do algoritmo elaborado com a sua versão escrita em C++ e executada completamente em CPU é exibido na Tabela 2. É possível observar que a taxa de crescimento do tempo de execução em CPU é significativamente superior ao crescimento do tempo em GPU, alcançando um *Speed Up* de até 708 vezes.

Tabela 1. Comparação dos resultados de energia livre para as proteínas otimizadas.

Proteína	L	DE _{pfo}	DE _{lsrc}	DE _{2L}	cuDSMjDE
1CRN	46	\bar{x}	-86.0390	-89.8223	-93.7138
		s	1.45	0.65	0.55
		f^*	-89.2001	-92.9853	-95.3159
2EWH	98	\bar{x}	-203.6810	-240.2247	-250.2833
		s	7.18	2.14	3.18
		f^*	-225.0968	-245.5190	-257.0741

Tabela 2. Speed Up do algoritmo para as proteínas otimizadas.

Proteína	L	Tempo [s]			Speed Up [\times]
		CPU	GPU		
1CRN	46	382.6000	2.1213		180.36
2EWH	98	1550.7017	2.1878		708.78

Palavras-chave: Predição de Estrutura de Proteínas. Algoritmos Evolutivos. Otimização Numérica.