

Introdução

Utilizando a metaheurística de evolução diferencial com adaptação de parâmetros busca-se determinar a estrutura tridimensional de dissacarídeos. Espera-se mostrar um método alternativo à dinâmica molecular, alcançando um resultado equivalente e com um tempo computacional reduzido. São considerados dissacarídeos formados por hexoses (monossacarídeos de 6 átomos de carbono) de cadeia cíclica.

Dissacarídeos

Dissacarídeos são moléculas orgânicas pertencentes ao grupo dos carboidratos, formadas pela ligação de dois monossacarídeos. Os monossacarídeos fazem parte do grupo mais simples de carboidratos, tendo forma geral $C_n(H_2O)_n$, havendo diversos isômeros para uma mesma fórmula. A formação do dissacarídeo ocorre por meio de uma ligação glicosídica entre grupos hidroxilas de cada monossacarídeo. Existe uma grande diversidade de dissacarídeos, uma vez que para um mesmo par de monossacarídeos existem diversas opções de ligação.

Estrutura dos dissacarídeos

A estrutura tridimensional de uma molécula é de grande importância na compreensão de suas propriedades e funções. Estão sendo considerados os seguintes monossacarídeos na formação dos dissacarídeos: talose, allose, galactose, glicose, idose e altrose. Cada um possui dois isômeros, α e β . No total são 16 monossacarídeos. Para cada par de monossacarídeos temos 5 (monossacarídeos idênticos) ou 25 (distintos) possíveis ligações, ocorrendo entre os átomos C1, C2, C3, C4 e C6. Uma estrutura é descrita pelos seguintes parâmetros: comprimento das ligações, ângulos entre 3 átomos adjacentes, ângulos diedrais de átomos adjacentes, ângulos diedrais impróprios e conformação dos anéis de cada monossacarídeo que o compõe (exemplos na Figura 1).

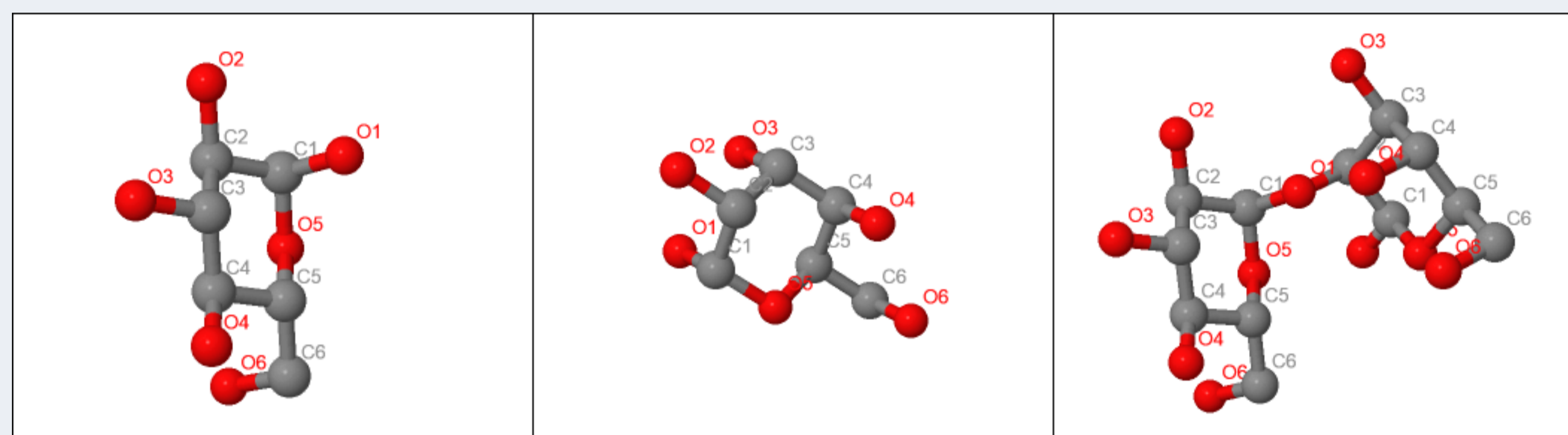


Figura 1: Monossacarídeos α – Glicose (esq.) e α – Talose. Dissacarídeo (dir.) resultante da ligação dos carbonos C1 e C2.

Os parâmetros de maior interesse são os ângulos diedrais formados na ligação glicosídica entre os monossacarídeos, designados ϕ (phi) e ψ (psi). A conformação do anel pode assumir infinitas configurações, todas representáveis através do globo conformacional (Figura 2). As mais comuns são as encontradas próximo dos polos (chair) e equador (boat).

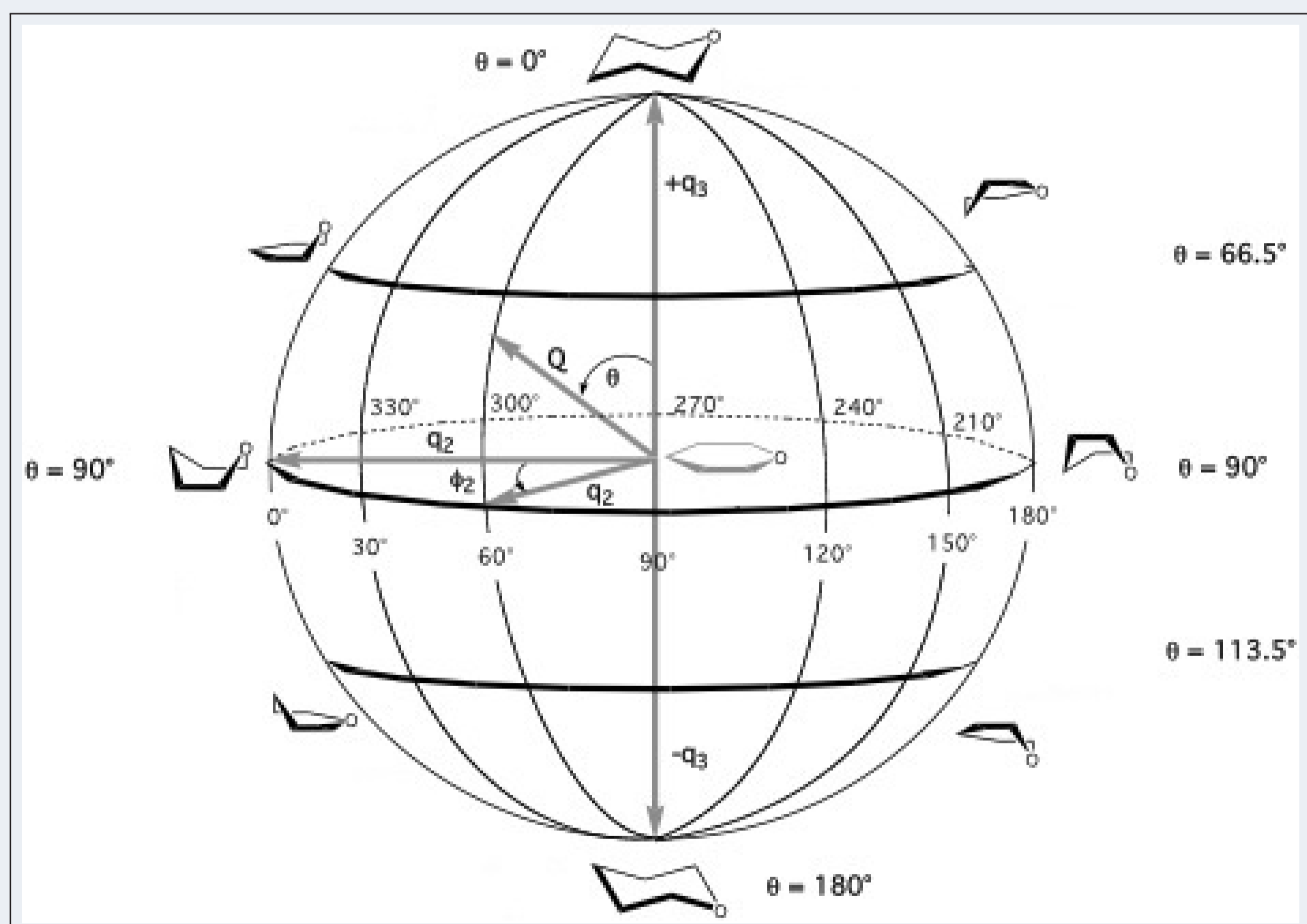
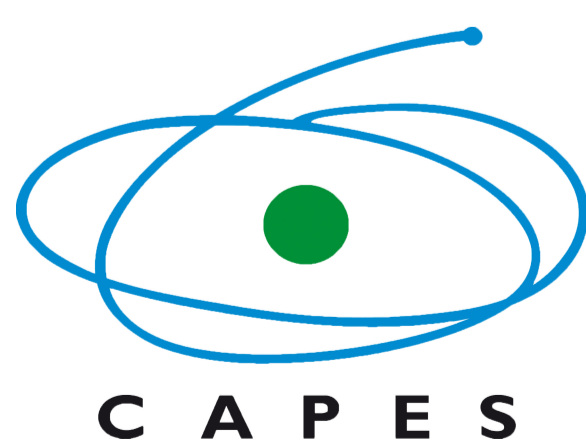


Figura 2: Globo conformacional da hexose. Três parâmetros definem todas conformações possíveis.

Agradecimentos:



Função de energia

Para avaliar a estabilidade de uma estrutura usa-se como critério o valor de sua energia. Quanto menor este valor, mais estável a estrutura. A configuração real de uma molécula deve ser aquela de menor energia possível.

Uma implementação da função de energia do campo de força GROMOS está sendo usada neste trabalho para este fim. Sua forma geral é:

$$U = \sum_{n=1}^{n_{bonds}} \frac{1}{4} k_b n (r - r_{0n})^2 + \sum_{n=1}^{n_{angles}} \frac{1}{2} k_{a_n} (\cos \theta - \cos \theta_{0n})^2 + \sum_{n=1}^{n_{impropers}} \frac{1}{2} k_{i_n} (\xi - \xi_{0n})^2 + \sum_{n=1}^{n_{dihedrals}} \frac{1}{2} k_{d_n} (1 + \cos(\delta_d) \cos(m_d \phi)) + \sum_{n=1}^{n_{pairs}} \frac{C12_{ij}}{r_{ij}^{12}} - \frac{C6_{ij}}{r_{ij}^6} + \sum_{n=1}^{n_{pairs}} \frac{q_i q_j}{4\pi \epsilon_0 \epsilon_1 r_{ij}}$$

Os termos da função representam a energia devido às distâncias de ligação, ângulos entre três átomos adjacentes, ângulos impróprios, ângulos diedrais de 4 átomos adjacentes, termo de Lennard-Jones e Coulomb, respectivamente. Os dois últimos termos são calculados para pares de átomos não ligados, sendo que os demais dizem respeito à interações entre átomos com alguma conexão entre si.

Evolução Diferencial (DE)

A evolução diferencial (DE) é um algoritmo evolutivo usado para a otimização de funções objetivo de parâmetros reais $f: \mathbb{R}^N \rightarrow \mathbb{R}$. Inicia-se com uma população de soluções candidatas aleatórias e iterativamente cada solução $\mathbf{x}_i \in \mathbb{R}^N$ passa pelos processos de mutação, combinação e seleção:

■ Mutação: segue a fórmula geral $\mathbf{v}_i = \mathbf{x}_{base_i} + \sum_{k=1}^K F_{k_i} (\mathbf{x}_{1_{k_i}} - \mathbf{x}_{2_{k_i}})$, onde \mathbf{x}_{base_i} e os pares $\mathbf{x}_{1_{k_i}}, \mathbf{x}_{2_{k_i}}$ são escolhidos conforme a estratégia de mutação utilizada (ex.: rand, best, curr-to-best) e F_{k_i} é um fator de escala no intervalo $[0, 1]$. Geralmente usa-se $K = 2$.

■ Combinação: é feita pelo crossover uniforme entre o vetor \mathbf{x}_i e \mathbf{v}_i , $u_{i,d} = \begin{cases} v_{i,d} & \text{rand}(0, 1) < CR, \\ x_{i,d} & \text{caso contrário} \end{cases}$

■ Seleção: para determinar a solução candidata da próxima iteração é feita a seleção elitista com o resultado do crossover,

$$\mathbf{x}_i = \begin{cases} \mathbf{v}_i & f(\mathbf{v}_i) < f(\mathbf{x}_i) \\ \mathbf{x}_i, & \text{caso contrário} \end{cases} \text{ (supondo minimização de } f)$$

Algoritmo SHADE

O algoritmo SHADE (Success-History based Adaptive DE) é uma das versões de evolução diferencial que apresentam bons resultados em diversos benchmarks, assim como suas extensões. Suas principais características e modificações são:

■ Mutação curr-to-pbest: $\mathbf{v}_i = \mathbf{x}_i + F_i (\mathbf{x}_{p\text{-best}} - \mathbf{x}_i) + F_i (\mathbf{x}_{r_1} - \mathbf{x}_{r_2})$, onde p-best é um índice sorteado aleatoriamente entre as p melhores soluções candidatas, \mathbf{x}_{r_1} e \mathbf{x}_{r_2} são índices aleatórios.

■ Arquivo externo: memória de tamanho fixo que armazena os \mathbf{x}_i bem sucedidos, que podem ser sorteados para serem usados como \mathbf{x}_{r_2} na mutação.

■ Memória de parâmetros: memória de tamanho fixo H que armazena tuplas das médias dos parâmetros F e CR bem sucedidos das últimas H iterações.

■ Geração de parâmetros: cada \mathbf{x}_i tem F_i e CR_i gerados amostrando, respectivamente, uma distribuição de Cauchy e uma distribuição normal, com médias selecionadas da memória de parâmetros.

O grande número de dissacarídeos, cada qual possuindo de certa forma sua própria função objetivo, e a necessidade de se escolher corretamente os parâmetros F e CR para a evolução diferencial funcionar satisfatoriamente torna o uso de parâmetros adaptativos uma boa escolha para tratar este problema.

Resultados

A função objetivo foi implementada e validada por comparação com resultados do GROMACS. Para a conformação dos dissacarídeos foram alteradas as distâncias de ligação, ângulos entre três átomos, ângulos diedrais e conformação do anel. Apesar do bom desempenho do SHADE nos casos testados (baixo desvio padrão entre diversas execuções e baixo valor de energia final) muitos valores finais de ϕ e ψ não corresponderam aos equivalentes obtidos pela dinâmica molecular. Uma das hipóteses para esta diferença seria a ausência de solvente na função objetivo utilizada pela metaheurística.

Referências

- 1 R. Storn and K. Price, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," J. Global Optimiz., vol. 11, no. 4, pp. 341–359, 1997.
- 2 R. Tanabe and A. Fukunaga, "Success-History Based Parameter Adaptation for Differential Evolution," in IEEE CEC, 2013, pp. 71–78.
- 3 R. Tanabe and A. Fukunaga, "Evaluating the performance of SHADE on CEC 2013 benchmark problems," in IEEE CEC, 2013, pp. 1952–1959.
- 4 R. Tanabe and A. Fukunaga, "Success-History Based Parameter Adaptation for Differential Evolution," submitted for publication, 2014.
- 5 R. Tanabe and A. Fukunaga, "Improving the search performance of SHADE using linear population size reduction," in 2014 IEEE Congress on Evolutionary Computation (CEC2014). IEEE, 2014, pp. 1658–1665.
- 6 J. Brest, M. S. Maučec and B. Bošković, "iL-SHADE: Improved L-SHADE algorithm for single objective real-parameter optimization," 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, 2016, pp. 1188–1195.
- 7 Borguesan, B.; Torbes, A.R.; Dorn, M.; Verli, H. CarbM: a web tool to build 3-D structures of carbohydrates. Third International Society for Computational Biology Latin America X-Meeting on Bioinformatics with BSB and SoBio (ISCB-Latin America), 2014, Belo Horizonte, Brazil.
- 8 D. Cremer, J. A. Pople. J. Am. Chem. Soc., 1975, 97 (6), pp 1354–1358. "A General Definition of Ring Puckering Coordinates".