

Ant Colony Optimization Algorithm for the 2D HP Protein Folding Problem

Gustavo Alves Kuabara

Victor Gava Senema

Fernando de Moraes Faquim

Problema

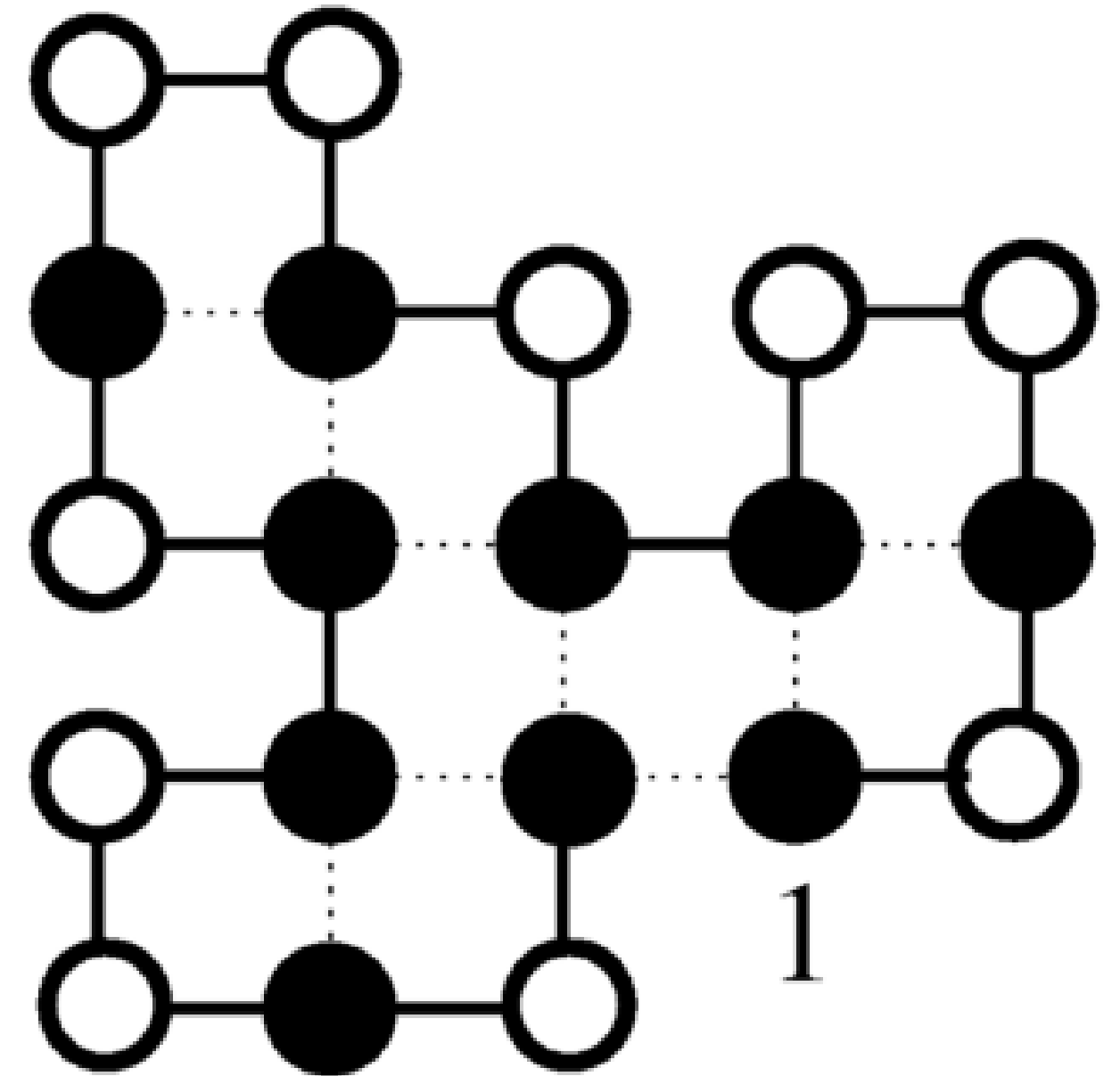
Dobramento de proteínas

- Descobrir o formato de uma proteína pela sua cadeia de aminoácido
- A proteína assume o formato de menor energia
- Estruturas corretas são essenciais para que a proteína desempenhe sua função biológica
- O estudo do protein folding é crucial para desenvolvimento de medicamentos, engenharia de proteínas e simulações computacionais que buscam prever estruturas proteicas

Problema

Dobramento de proteínas no modelo HP bidimensional (2D HP)

- Representação em grade 2D
- P – Polar , H – Hidrofóbico
- Busca aumentar a quantidade de ligações hidrofóbicas não consecutivas



Objetivo

Objetivo Geral: Previsão da melhor configuração de uma proteína a partir de sua sequência de aminoácidos

- Maximizar o contato hidrofóbico não consecutivo, minimizando a energia

$$E(c^*) = \min\{E(c) \mid c \in C\}$$

- Função de Energia: E ,
- Conformação de aminoácidos: c
- Conformação ótima de aminoácidos: c^*
- Combinação de sequências válidas: C

O Algoritmo

```
aco_protein_folding_2DHP(proteina) {  
    Inicializar feromônios  
  
    Para cada iteração {  
        Para cada formiga {  
            conformação ← []  
  
            Enquanto conformação não estiver completa {  
                sucesso ← crescer_conformação(conformação)  
                Se não sucesso {  
                    Jogar fora metade da conformação já criada  
                }  
            }  
  
            busca_local(conformação)  
        }  
  
        Atualizar feromônios com base nas conformações encontradas  
    }  
  
    Retornar melhor conformação encontrada  
}
```

○ Algoritmo: crescimento

- Escolhe uma direção aleatória com base na probabilidade:

$$p_{i,d} = \frac{(\tau_{i,d})^\alpha (\eta_{i,d})^\beta}{\sum_{e \in \{L,R,S\}} (\tau_{i,e})^\alpha (\eta_{i,e})^\beta}$$

$$\eta_{i,e} = E(c') - E(c) + 1$$

- c é a conformação base
- c' é c após colocar o aminoácido i na direção e
- Não cresce para becos sem saída
- Rebobina se não for possível crescer

○ Algoritmo: busca local

- Tenta melhorar a conformação encontrada
- **Point Mutation**
 - Tenta alterar todos os pontos individualmente em uma ordem aleatória
 - Aplica a alteração se houver melhora
- **Macro Mutation Neighbourhood**
 - Seleciona uma fatia aleatória da conformação e modifica todos os valores simultaneamente
 - Aplica a alteração se houver melhora
- Repete esse processo até que não haja melhora por **NoImprMax** interações

O Algoritmo: atualização de feromônio

$$\tau_{i,d} \leftarrow (1 - p)\tau_{i,d} + \Delta_{i,d,c}$$

$$\Delta_{i,d,c} = \begin{cases} \frac{E(c)}{E^*}, & \text{se } c_i = d, \\ 0, & \text{caso contrário.} \end{cases}$$

Mudanças em relação ao artigo original

- Crescimento unilateral
- Mudança na normalização de feromônio
- Taxa de mutação neutra

Resultados

Configuração		id	Tamanho	E*	Nopt/Nruns	Melhor Encontrado	Média	Média/E*	% suc
Formigas	40	1	20	9	54/54	9	9,0000	1,0000	1,0000
NoImprMax	20	2	24	9	54/54	9	9,0000	1,0000	1,0000
Evaporação	0,9	3	25	8	0/54	7	7,0000	0,8750	0,0000
Alpha	1	4	36	14	9/54	14	13,1667	0,9405	0,1667
Beta	2	5	48	23	0/54	20	18,4630	0,8027	0,0000
Mutação neutra	0,5	6	50	21	0/53	20	18,0189	0,8580	0,0000
		7	59	36	0/51	33	30,7255	0,8535	0,0000
		8	64	42	0/51	32	29,9216	0,7124	0,0000
		9	20	10	0/51	9	9,0000	0,9000	0,0000

<i>Instances</i>			<i>ACO + Local Search</i>				<i>Local Search Only</i>			
Seq. No.	Length	E^*	sq	n_{opt}/n_{runs}	% suc.	t_{avg}	sq	n_{opt}/n_{runs}	% suc.	t_{avg}
1	20	-9	-9	711/711	100.0	23.90	-9	100/258	38.7	111.43
2	20	-9	-9	596/596	100.0	26.44	-9	8/113	7.0	162.15
3	25	-8	-8	120/120	100.0	35.32	-8	44/129	34.1	125.42
4	36	-14	-14	21/128	16.4	4746.12	-14	5/72	6.9	136.10
5	48	-23	-23	1/151	0.6	1920.93	-21	1/20	5.0	1780.74
6	50	-21	-21	18/43	41.9	3000.28	-20	3/18	16.7	1855.96
7	60	-36	-34	1/119	0.8	4898.77	-33	2/20	10.0	1623.21
8	64	-42	-32	1/22	4.5	4736.98	-33	2/9	22.2	1441.88
9	24	-10	-10	247/247	100.0	43.48	-10	5/202	25.0	134.57

Resultados

Configuração		id	Tamanho	E*	Nopt/Nruns	Melhor Encontrado	Média	Média/E*	% suc
Formigas	20	1	20	9	61/63	9	8,9683	0,9965	0,9683
NoImprMax	20	2	24	9	63/63	9	9,0000	1,0000	1,0000
Evaporação	0,9	3	25	8	0/63	7	7,0000	0,8750	0,0000
Alpha	1	4	36	14	3/63	14	12,8730	0,9195	0,0476
Beta	3	5	48	23	0/63	20	18,3492	0,7978	0,0000
Mutação neutra	0,5	6	50	21	0/63	19	17,8889	0,8519	0,0000
		7	59	36	0/63	33	30,6032	0,8501	0,0000
		8	64	42	0/63	33	29,9841	0,1739	0,0000
		9	20	10	0/63	9	9,0000	0,9000	0,0000

<i>Instances</i>			<i>ACO + Local Search</i>				<i>Local Search Only</i>			
Seq. No.	Length	E^*	sq	n_{opt}/n_{runs}	% suc.	t_{avg}	sq	n_{opt}/n_{runs}	% suc.	t_{avg}
1	20	-9	-9	711/711	100.0	23.90	-9	100/258	38.7	111.43
2	20	-9	-9	596/596	100.0	26.44	-9	8/113	7.0	162.15
3	25	-8	-8	120/120	100.0	35.32	-8	44/129	34.1	125.42
4	36	-14	-14	21/128	16.4	4746.12	-14	5/72	6.9	136.10
5	48	-23	-23	1/151	0.6	1920.93	-21	1/20	5.0	1780.74
6	50	-21	-21	18/43	41.9	3000.28	-20	3/18	16.7	1855.96
7	60	-36	-34	1/119	0.8	4898.77	-33	2/20	10.0	1623.21
8	64	-42	-32	1/22	4.5	4736.98	-33	2/9	22.2	1441.88
9	24	-10	-10	247/247	100.0	43.48	-10	5/202	25.0	134.57

Resultados

Configuração		id	Tamanho	E*	Nopt/Nruns	Melhor Encontrado	Média	Média/E*	% suc
Formigas	20	1	20	9	50/51	9	8,9804	0,9978	0,9804
NolmprMax	40	2	24	9	51/51	9	9,0000	1,0000	1,0000
Evaporação	0,9	3	25	8	0/51	7	7,0000	0,8750	0,0000
Alpha	1	4	36	14	6/51	14	12,8824	0,9202	0,1176
Beta	2	5	48	23	0/51	19	18,2549	0,7937	0,0000
Mutação neutra	0,5	6	50	21	0/51	19	17,9804	0,8562	0,0000
		7	59	36	0/51	33	30,6667	0,8519	0,0000
		8	64	42	0/51	32	29,7451	0,7082	0,0000
		9	20	10	0/51	9	9,0000	0,9000	0,0000

<i>Instances</i>			<i>ACO + Local Search</i>				<i>Local Search Only</i>			
Seq. No.	Length	E^*	sq	n_{opt}/n_{runs}	% suc.	t_{avg}	sq	n_{opt}/n_{runs}	% suc.	t_{avg}
1	20	-9	-9	711/711	100.0	23.90	-9	100/258	38.7	111.43
2	20	-9	-9	596/596	100.0	26.44	-9	8/113	7.0	162.15
3	25	-8	-8	120/120	100.0	35.32	-8	44/129	34.1	125.42
4	36	-14	-14	21/128	16.4	4746.12	-14	5/72	6.9	136.10
5	48	-23	-23	1/151	0.6	1920.93	-21	1/20	5.0	1780.74
6	50	-21	-21	18/43	41.9	3000.28	-20	3/18	16.7	1855.96
7	60	-36	-34	1/119	0.8	4898.77	-33	2/20	10.0	1623.21
8	64	-42	-32	1/22	4.5	4736.98	-33	2/9	22.2	1441.88
9	24	-10	-10	247/247	100.0	43.48	-10	5/202	25.0	134.57

Conclusão

- Em geral, o algoritmo implementado apresentou soluções ótimas em alguns casos e soluções próximas da ótima
- Em comparação a busca local o método se sobrepõe e demonstra ser útil para previsão da estrutura de proteínas

Referências

- Shmygelska, A., Aguirre–Hernández, R., & Hoos, H. H. (2002). An Ant Colony Optimization Algorithm for the 2D HP Protein Folding Problem. In M. Dorigo, G. Di Caro, & M. Sampels (Eds.), *Ant Algorithms (ANTS 2002)* (Lecture Notes in Computer Science, Vol. 2463, pp. 40–52). Springer.
- Krasnogor, N., Hart, W. E., Pelta, D., & Lopez, P. (1999). Protein structure prediction with evolutionary algorithms. Sandia National Laboratories; University of the West of England; Universidad de Granada.