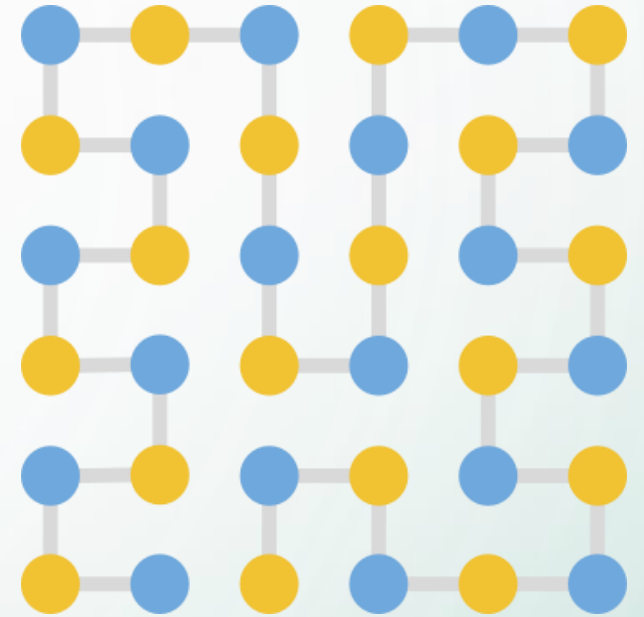


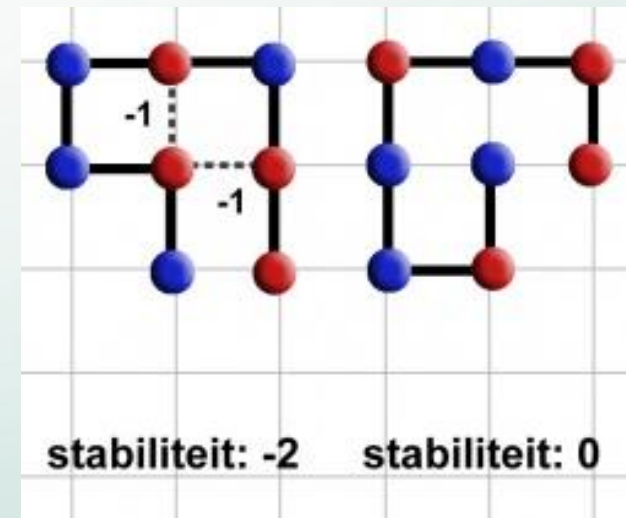
PROTEIN POW(D)ER

MAINAH FOLKERS, MILAN SCHOLTEN, RIANNE SCHOON



PROTEIN POW(D)ER: OPTIMALE VOUWING

- Eiwit: keten van aminozuren (H/P/C)
- H-H score -1 , H-C score -1 , C-C score -5
- Geen overlap!
- Vouw 2D / 3D
- Toestandsruimte upper bound:
 - 2D 3^{n-2}
 - 3D 5^{n-2}



PROTEIN POW(D)ER: DE KETENS

- Kort / lang
- H/P of H/P/C
- HHPHHPH (8)
- HHPHHPHPHHPH (14)
- HHPHPHPHPHHHPHPPHPPPHPPPPHPPPHPPPHHPHHHPHPHPHPHH (50)
- HCPHPHPHCHHHHPCCPPHPPHPPPPCPPPHPPPHHPHHHCHHPHPHPHH (50)



ALGORITMES – RANDOM SAMPLER

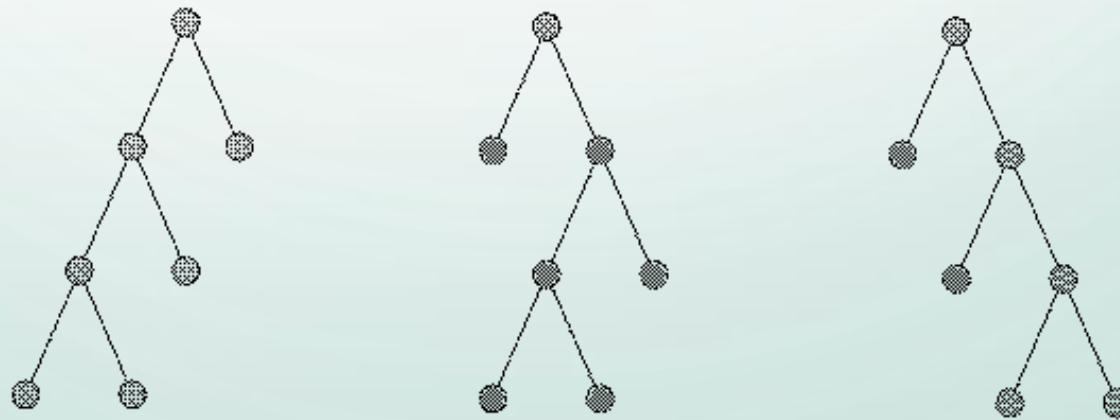
- Vouw eiwit:
 - Eerste binding naar rechts (symmetrie)
 - Tweede binding naar rechts / boven
 - Verder: steeds random richting (niet terug!)

ITERATIEF: HILLCLIMBER / SIMULATED ANNEALING

- Hillclimber
 - Begin met random vouwing
 - Verleg een (paar) binding(en)
 - Betere score? Daarmee verder!
- Simulated annealing
 - Accepteert ook verslechtingen
 - Koelingsschema lineair / hyperbolisch
 - Random kans

CONSTRUCTIEF: DEPTH FIRST SEARCH

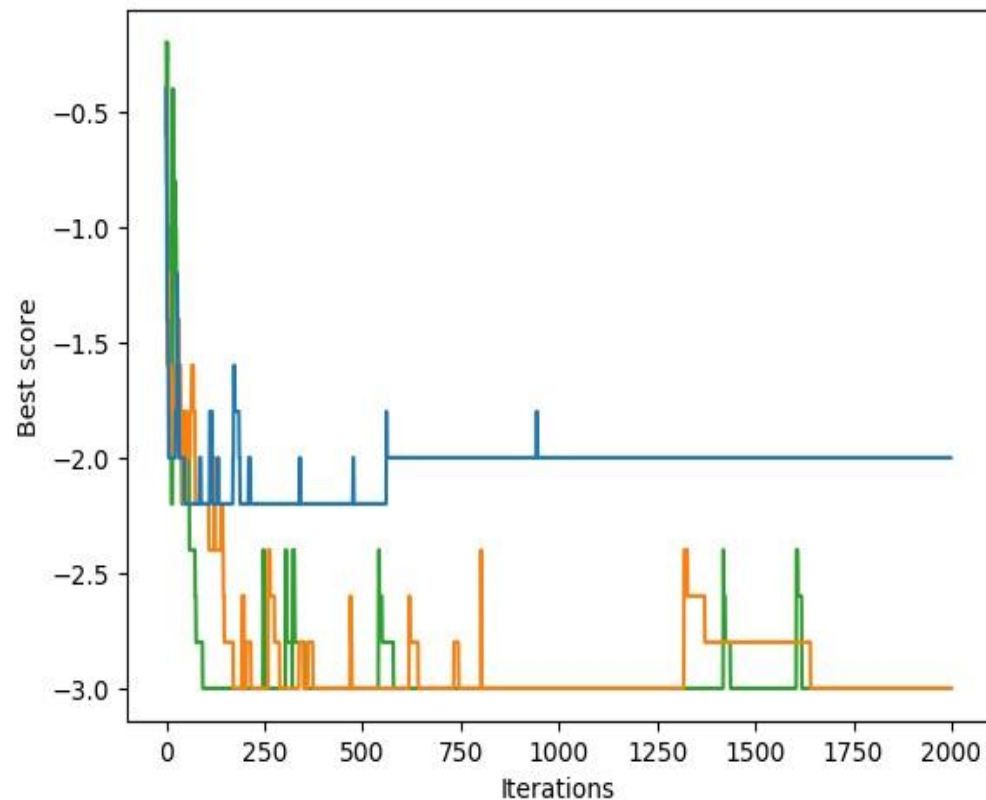
- Eerste twee bindingen vast (rechts – boven/rechts)
- Daarna drie keuzes (behalve overlap!)
- Heul Veul mogelijkheden: greedy keuzes
- Ketting > 10 : selectie tak met beste score



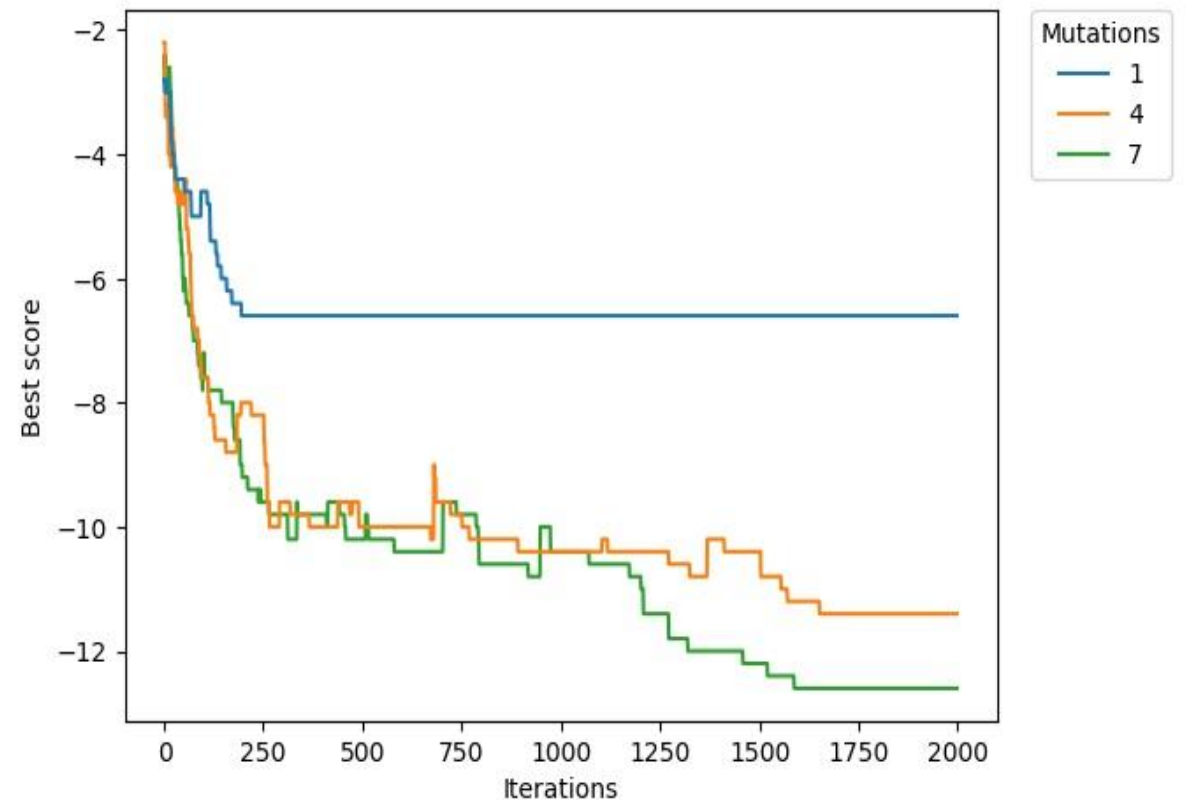
MUTATIEFREQUENTIE (CLIMBER/SIMULATED ANNEALING)

- Hyperbool koelingschema simulated annealing

8 aminozuren



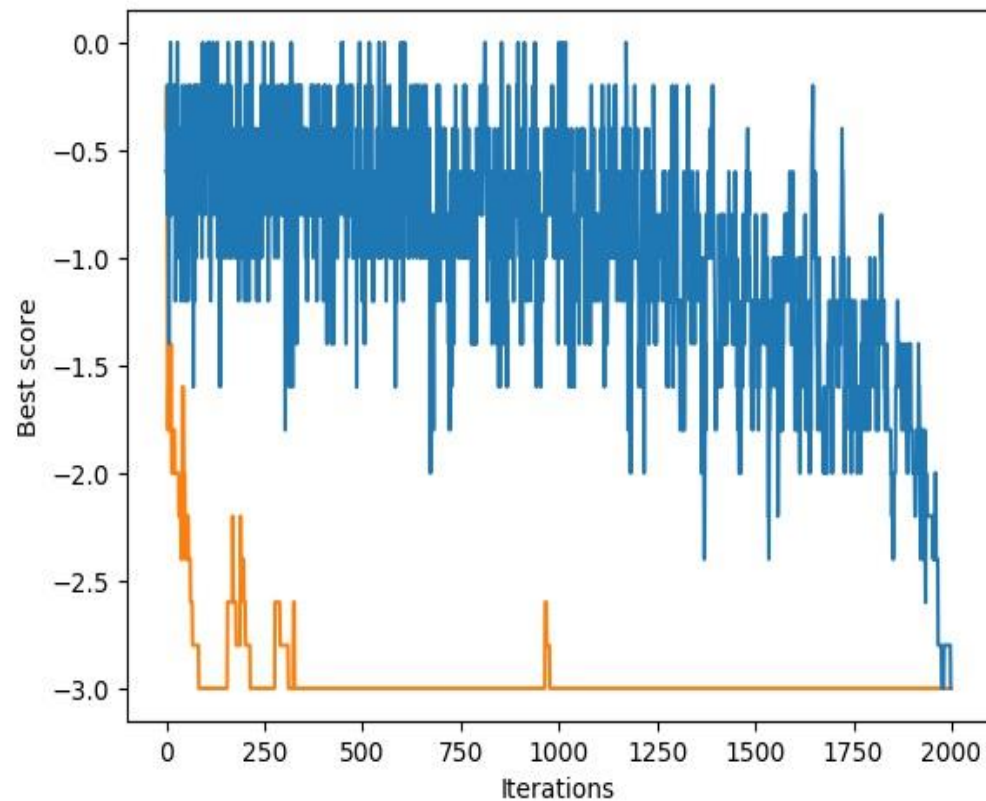
50 aminozuren



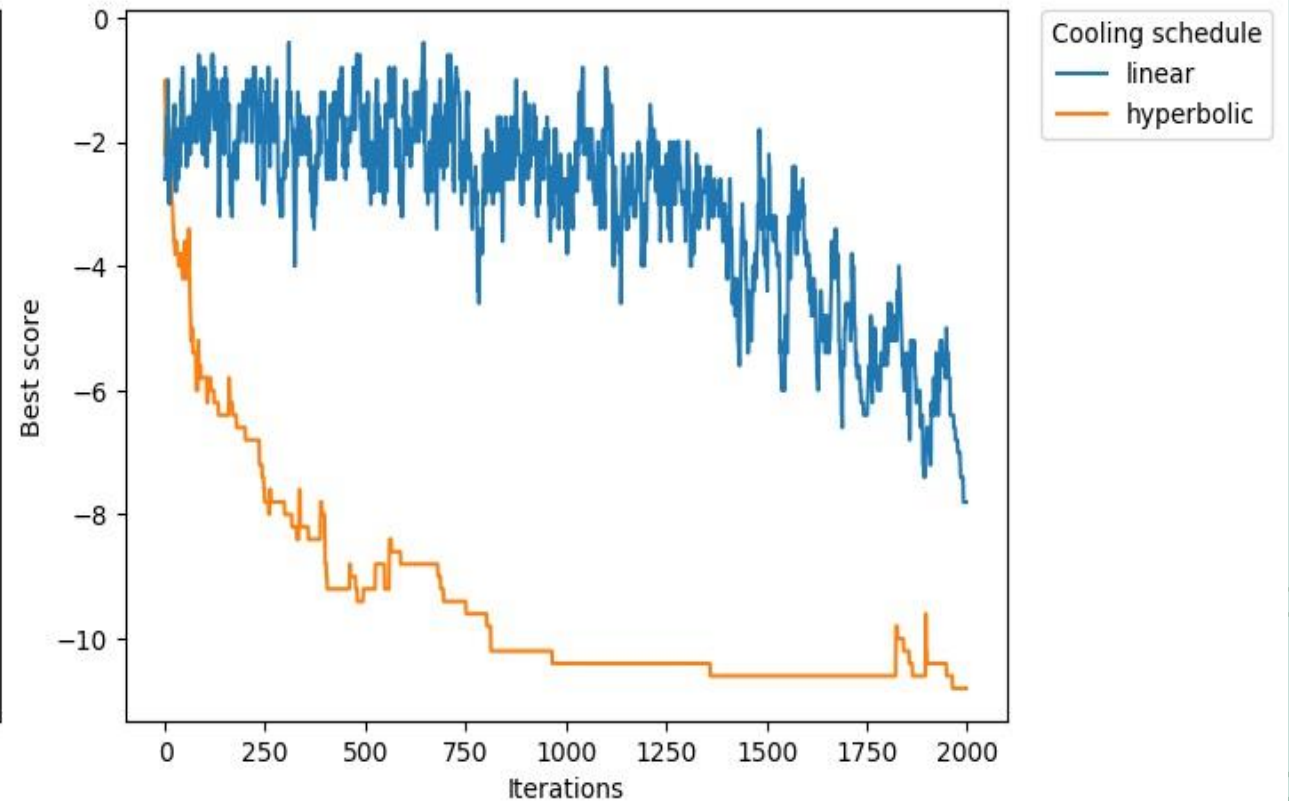
KOELINGSSHEMA SIMULATED ANNEALING

- Hyperbool

8 aminozuren



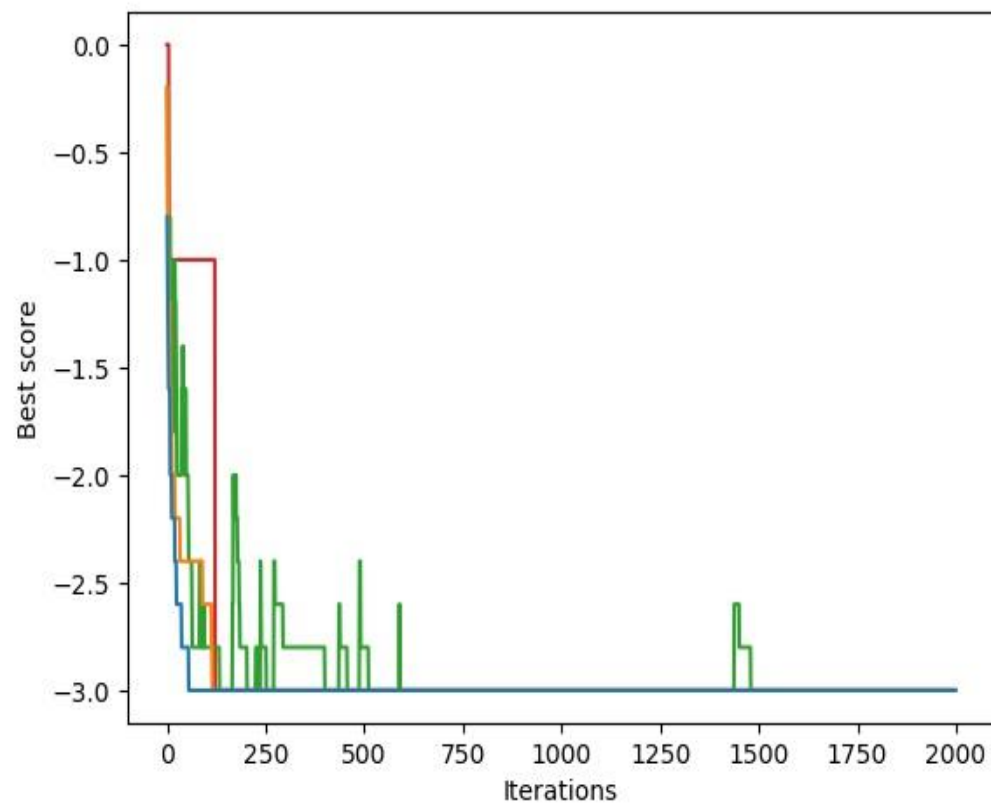
50 aasdkjuren



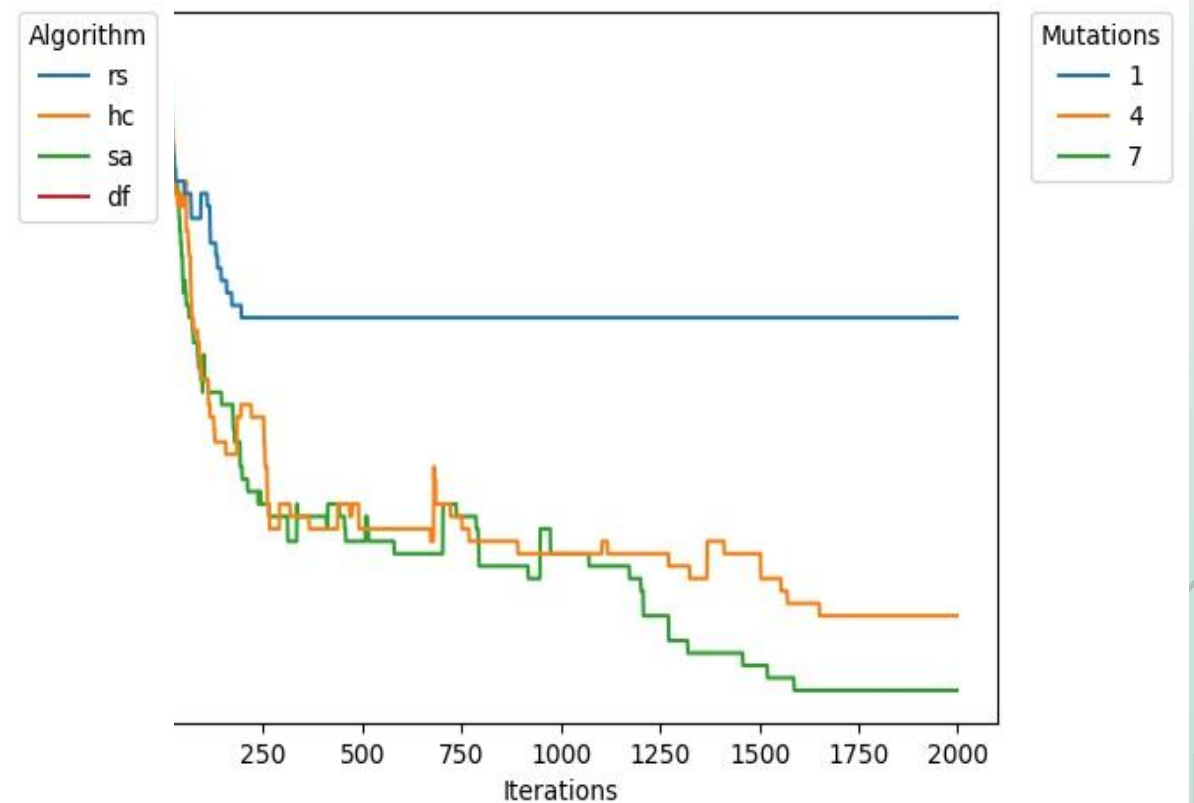
VERGELIJKING PRESTATIES ALGORITMES

- Hyperbool

8 aminozuren



50 aminozuren



CONCLUSIES ALGORITMES 2D

- Mutatiefrequentie:
 - Korte keten: 7 mutaties
 - Lange keten: 7 mutaties (meer?)
- Koelingsschema: hyperbool
- Algoritme prestatie:
 - Random sampler snelst laagste score

CONCLUSIE: SCORE EIWITTEN 2D

- aklsdj

Keten	Score
HHRHHRPH (8)	- 3
HHRHHRPHRHHHRPH (14)	- 6
HHRPHRPHRHHHHHRPHPPPHPPPHPPPHPPPHPPPHRHHHHHRPHRPHRPHH (50)	- 17
HCRPHRPHCHHHHRCCPPHPPPHPPPHCPPPHPPPHRHHHHCHRPHRPHH (50)	- 20

DISCUSSIE (1): SCORES

- HHPHHPH (8)
 - Wij: -3 Dovier, Formisano, Pontelli (2009): -3
- HHPHPHPHPHHHPHPPPHPPPHPPPHPPPHPPPHHHHPHPHPHPHH (50)
 - Wij: -17 Shmygelska, Hernandez, Hoos (2002) (en vele anderen): -21
- HCPHPHPHCHHHHPCCPPHPPPHPPPPCPPPHPPPHHHHCHPHPHPHH (50)
 - Geen literatuur met C.

DISCUSSIE (2): VALIDITEIT

- Validiteit: alles minimaal 20 runs
- Iteraties: >1000 zelden score verbetering
 - Toch: meer is beter – kans!
- Toestandsruimte upper bound:
 - 2D: 3^{n-2} , 3D: 5^{n-2} (invalide vouwingen!)



SIMULATED ANNEALING WERKT HET MEEST EFFICIENT

Dank voor uw aandacht!

Speciaal bedankje naar

Bas Terwijn voor de tech assist,

Daan van den Berg voor werkgroepen en feedback!



REFERENTIES

- Zhang, J., Kou, S. C., & Liu, J. S. (2007) Biopolymer structure simulation and optimization via fragment regrowth Monte Carlo. *The Journal of chemical physics*, 126(22), 06B605
- Custódio, F. L., Barbosa, H. J., & Dardenne, L. E. (2004) Investigation of the three-dimensional lattice HP protein folding model using a genetic algorithm. *Genetics and Molecular Biology*, 27(4), 611-615.
- Dill, K. A. et al. (1995) Principles of protein folding - A perspective from simple exact models. *Protein Science*, 4, 561-602.
- Shmygelsko, A., Hernandez, R., Hoos, H. H. (2002) An ant colony optimization algorithm for the 2D HP protein folding problem. *LNCS*, 2463, 40-52.
- Dovier, A., Formisano, A., Pontelli, E. (2009) An empirical study of constraint logic programming and answer set programming solutions of combinatorial problems. *Journal of Experimental & Theoretical Artificial Intelligence* 21(2), 79-121.