

Protein Pow(d)er



The legend of Cysteine

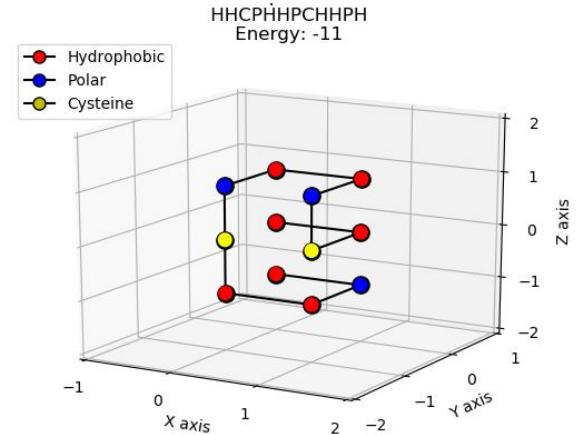
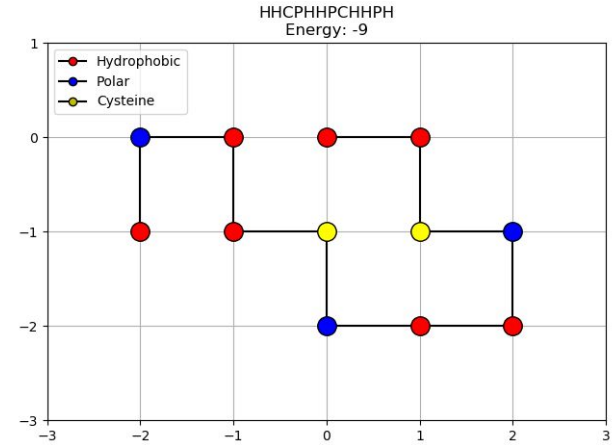
Ruby Bron

Michael Stroet

Sophie Stiekema

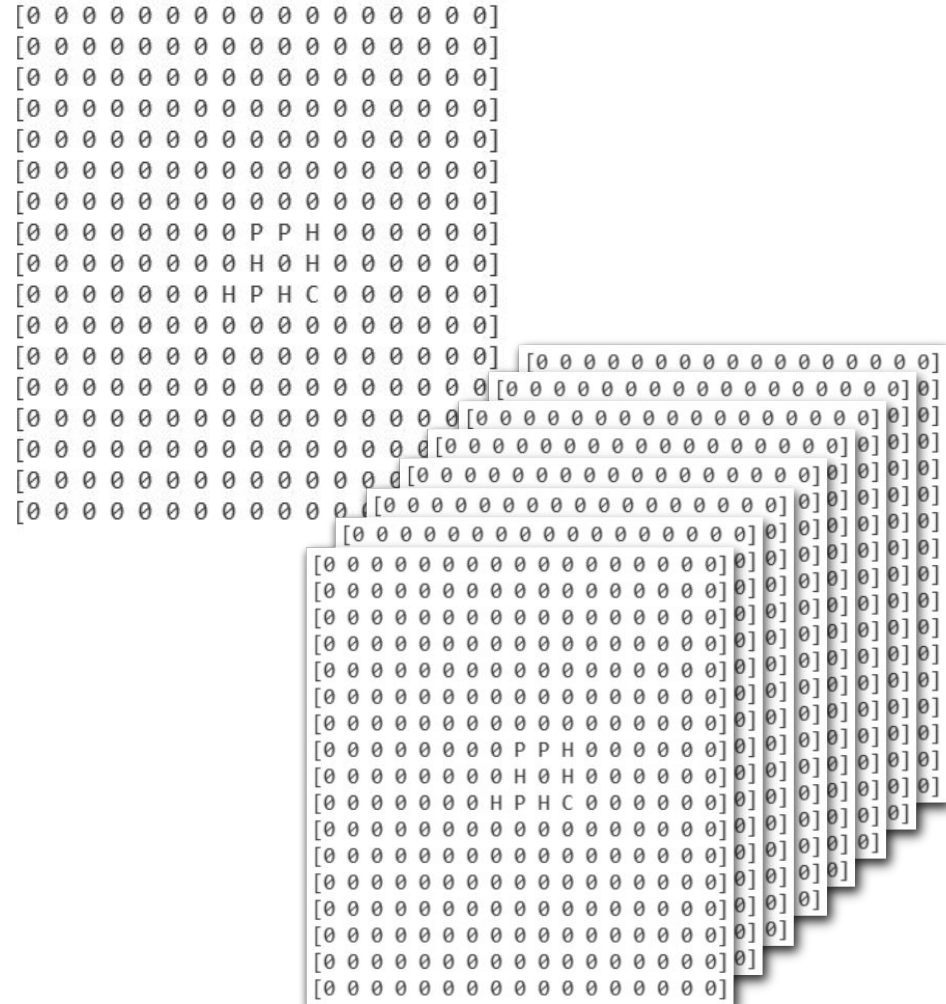
De case

- Eiwitten vouwen
- Aminozyuren op een 2D / 3D rooster
 - Hydrofoob
 - Polair
 - Cysteïne
- Minimaliseren van energie
 - P → 0
 - H - H → -1
 - H - C → -1
 - C - C → -5



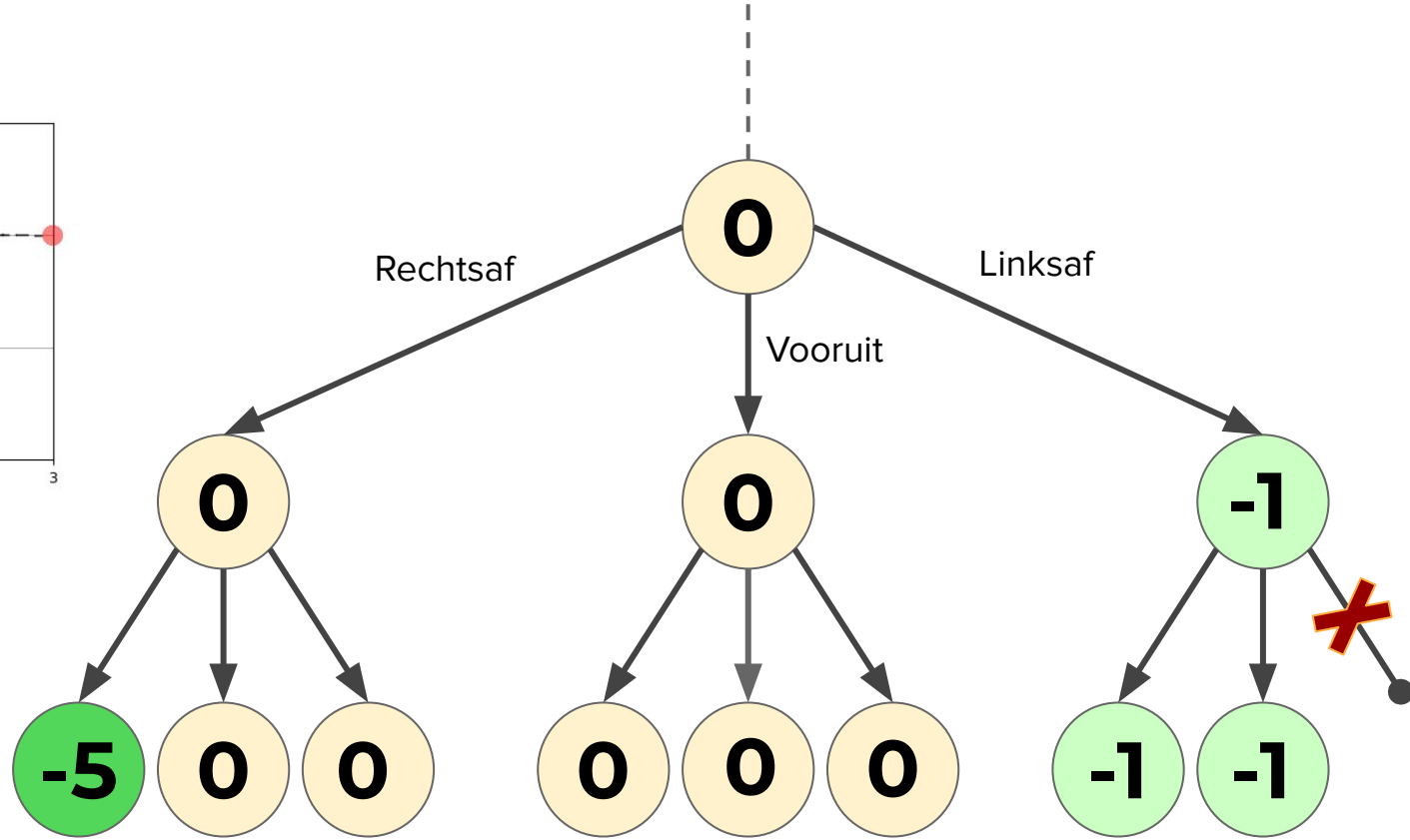
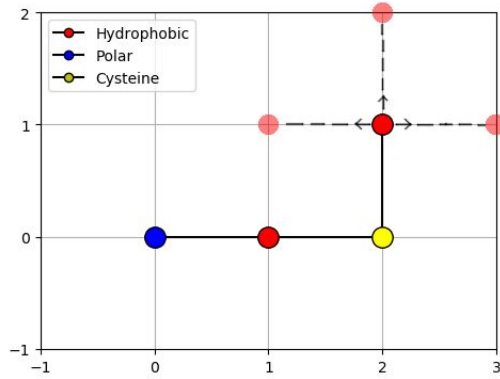
Eiwitten vouwen

- 2D: Matrix (lijst van lijsten)
- 3D: lijst van matrices
- Matrix grootte varieerbaar



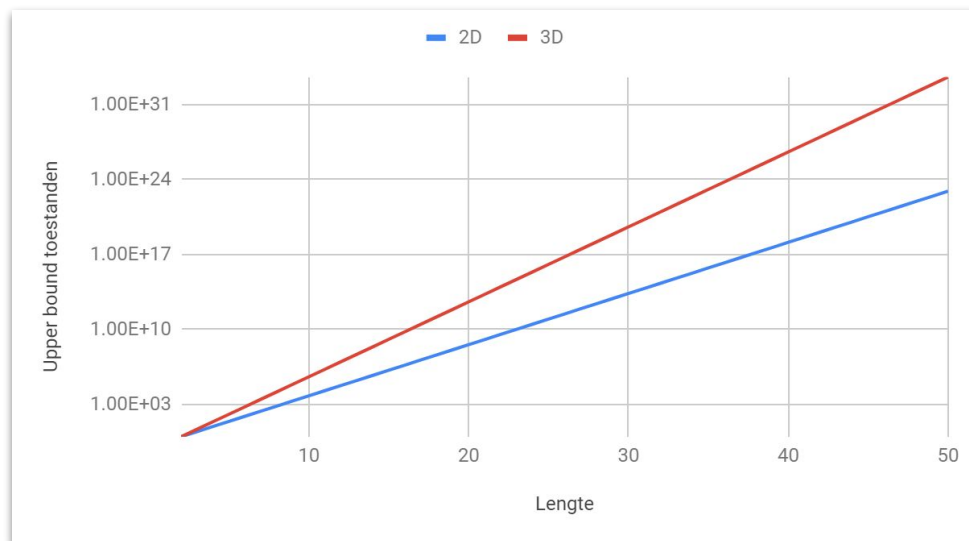
Toestandsruimte

→ P H C H H C



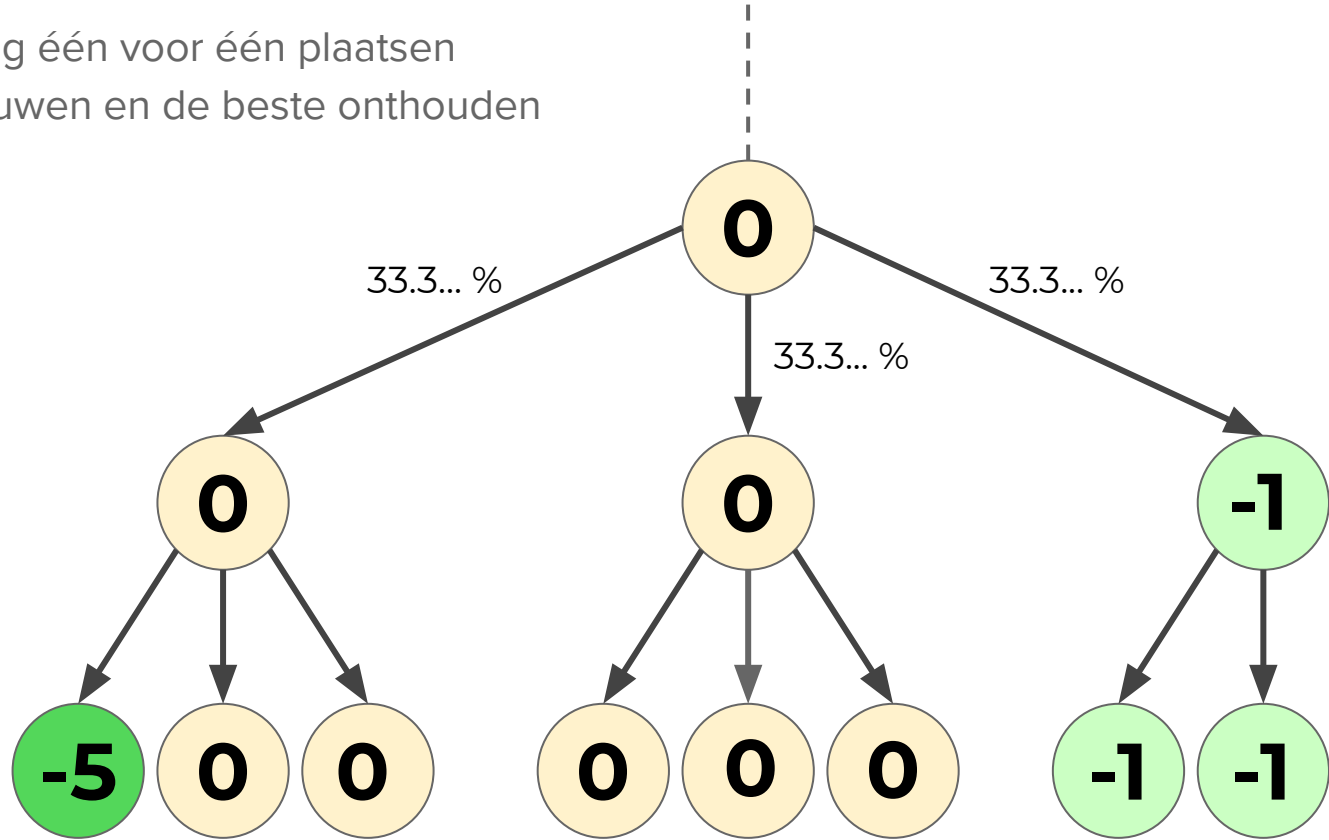
Toestandsruimte

- 2D: $3^{\text{length} - 2}$ length 50: $3^{48} \approx 8.0 * 10^{22}$
- 3D: $5^{\text{length} - 2}$ length 50: $5^{48} \approx 3.6 * 10^{33}$
- Toestandsruimte verkleinen:
 - Symmetrie
 - Matrix grootte



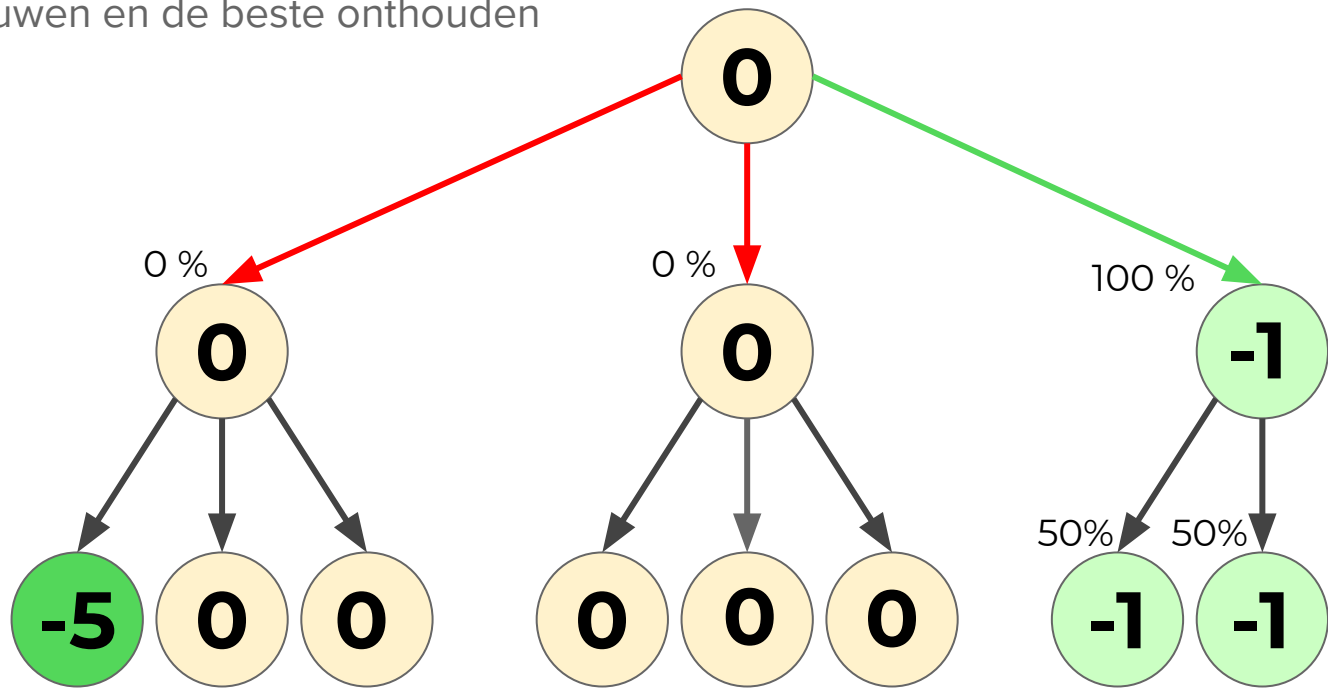
Random Walk

- Aminoszuren willekeurig één voor één plaatsen
- N aantal proteïnen vouwen en de beste onthouden



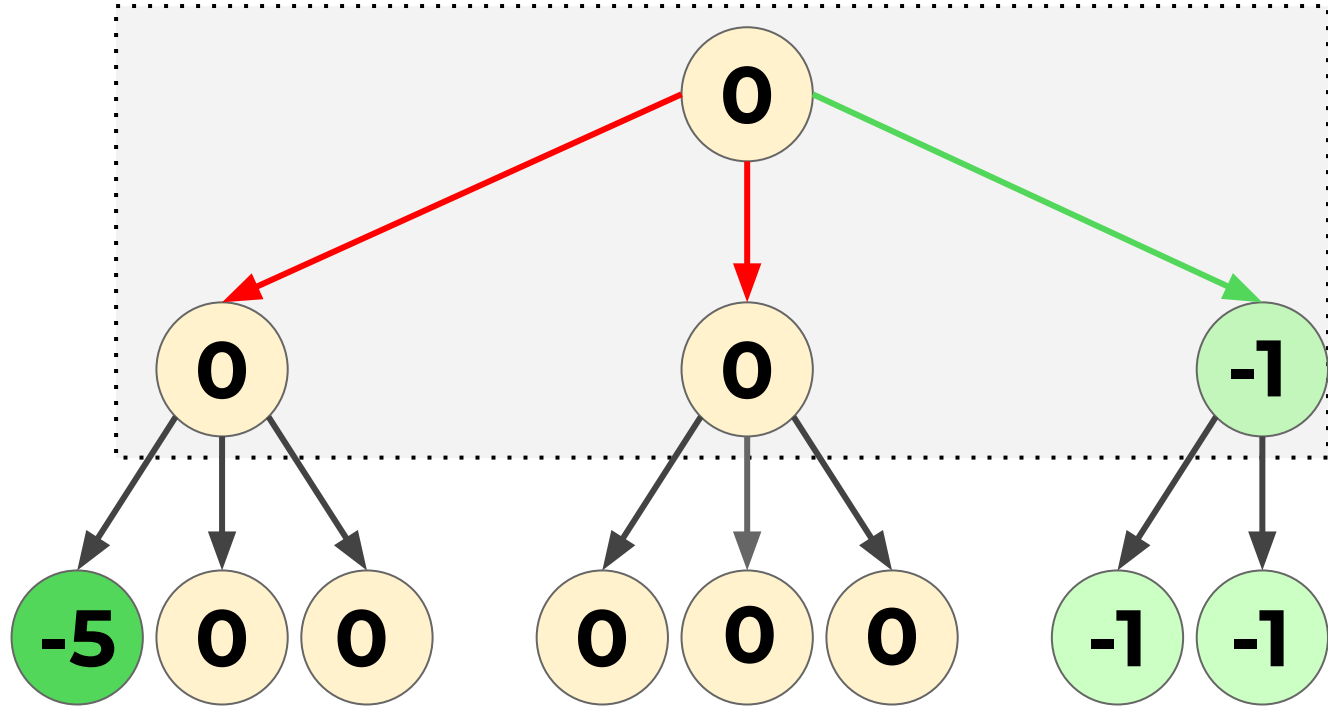
Greedy (look-ahead)

- Aminoszuren één voor één plaatsen bij de beste energie
 - Bij gelijke energie waarde, willekeurige keuze tussen deze richtingen
- N aantal proteïnen vouwen en de beste onthouden



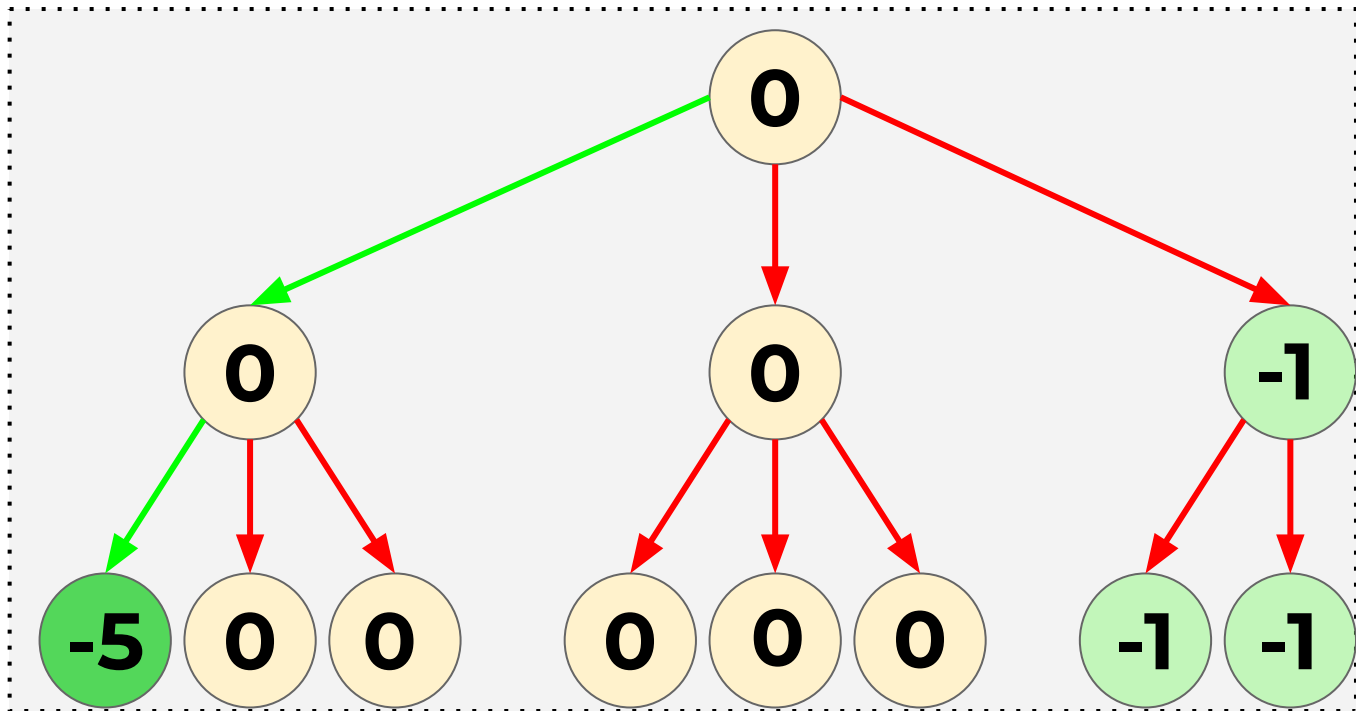
Greedy (look-ahead)

→ Zonder look-ahead



Greedy (look-ahead)

→ Met look-ahead



Constructief: Beam Search

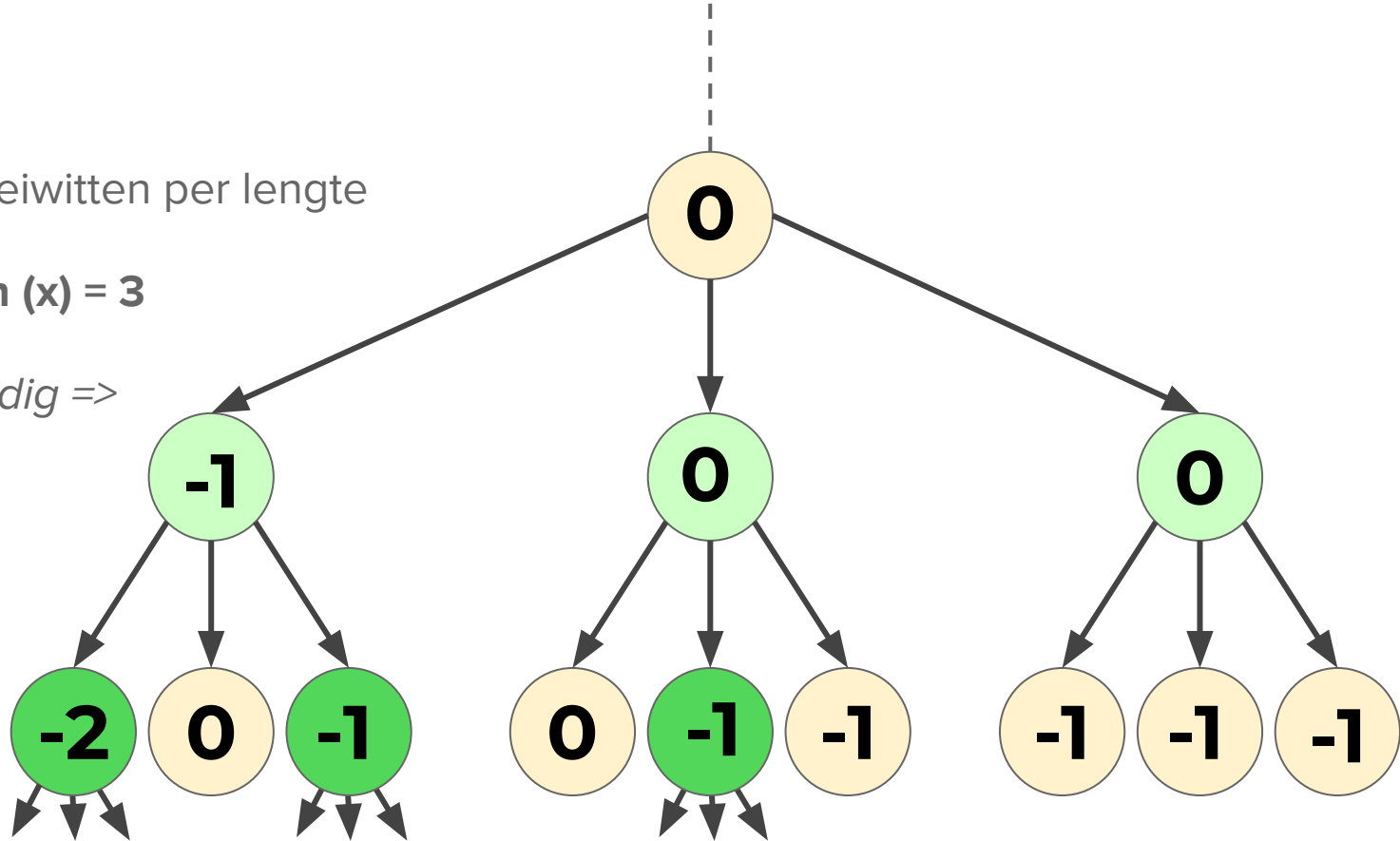
Breadth-first

Onthoudt x beste eiwitten per lengte

→ Beam width (x) = 3

Beam width oneindig =>

Exact algoritme



Constructief: Branch 'n Bound

Depth-first, recursief, non-stack en probability based

Aminozuren een voor een plaatsen en energie bijhouden

Laagste energie tot nu toe?

- Verder gaan

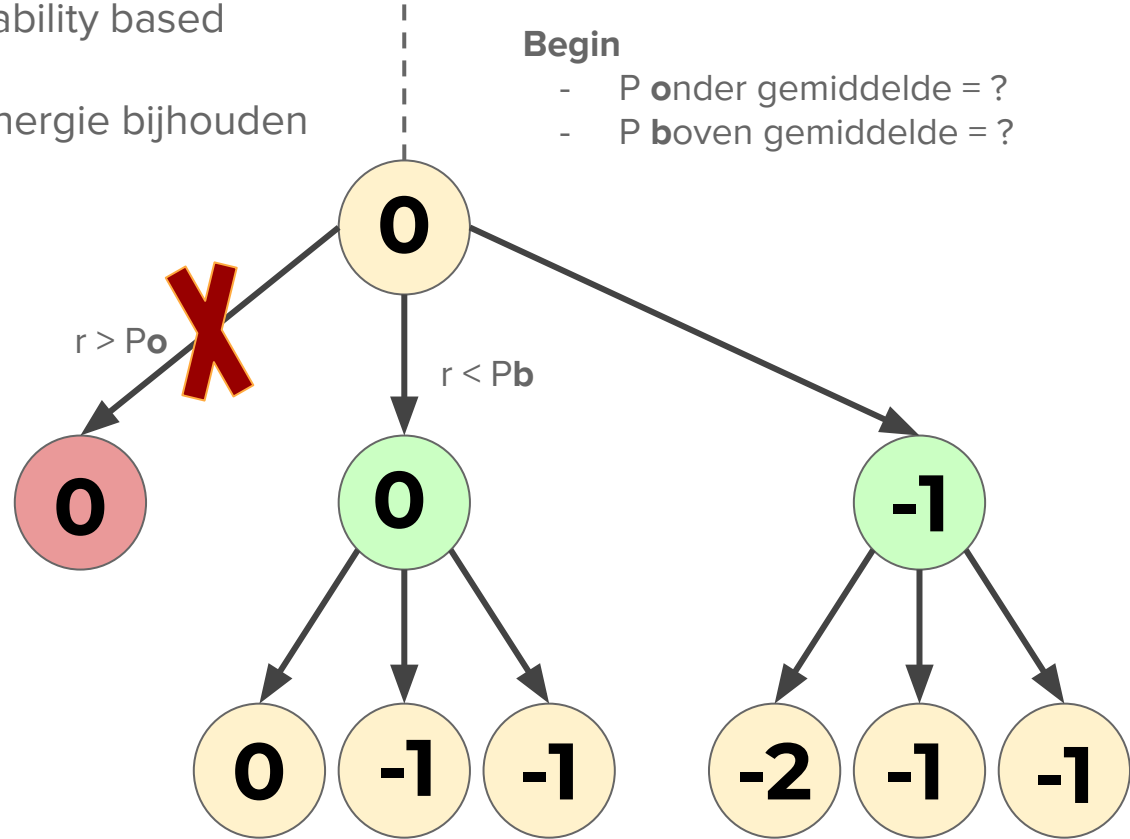
Lagere energie dan gemiddelde?

- Kans om verder te gaan

Hogere energie dan gemiddelde?

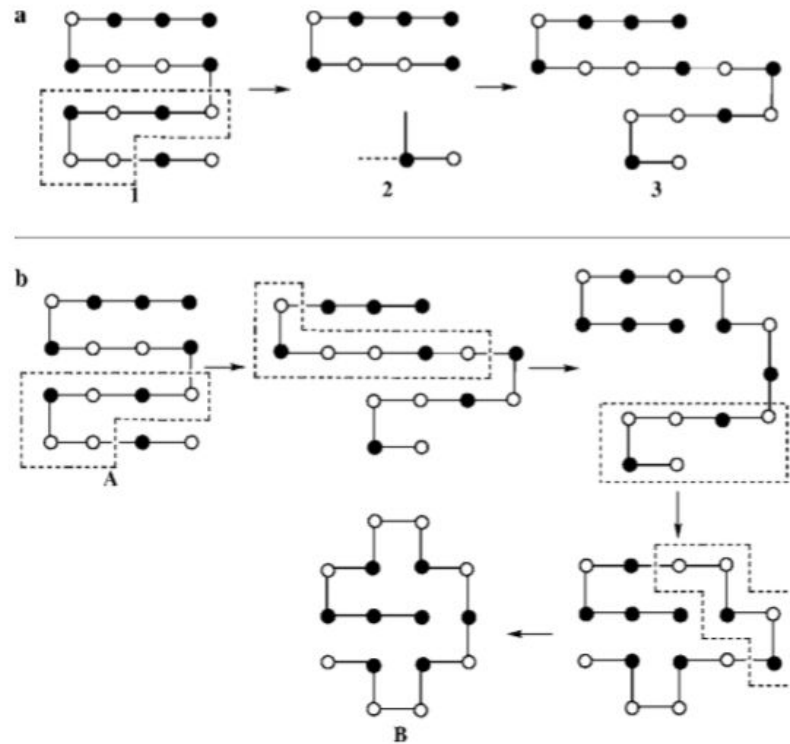
- Kleinere kans om verder te gaan

Kansen 1 & 1 => Exact algoritme



Iteratief: Hill Climber (Local Search)

- Start eiwit:
 - Eiwit van random_walk
- Verwijderen en plaatsen aminozuren:
 - Tussen start & end
 - Range: n
 - Aantal iteraties
 - Verbinding leggen
 - Energie testen
 - Lager: accepteren
 - hoger/gelijk: weigeren
- Paper:
 - Simulated annealing
 - Zang, Kou & Liu (2007)



Beste resultaten

	Proteïne	Lengte	2D	3D
1	HHPHHPH	8	-3	-3
2	HHPHHPHPHHPH	14	-6	-7
3	HPHPPHHPHPPHHPHPPH	20	-9	-11
4	PPPHHPHPPPPHHHHHHHPHHPPPPHHPHPP	36	-14	-18
5	HHPHHPHHPHHHPHPPPHPPPHPPPHPPPHHPHHHPHHPHHPH	50	-21	-30
6	PPCHHPCHPPPPCHHHHCHHPPHHPPPHHPHPP	36	-25	-35
7	CPPCHPPCHPPCPPHHHHHHCCPCHPPCPCHPPHPC	36	-38	-58
8	HCPHPCPHPCCHPHPPPHPPPHPPPHPCPHPPPHPHHHCHCHCHCHH	50	-29	-47
9	HCPHHPHCHHHHPCCPPHPPPHPPPPCPPPHPPPHPHHHHCHPHPHPHH	50	-34	-54

Energien per algoritme (2D, lengte/2, ~10 minuten)						
	<u>Lengte</u>	<u>Random walk</u> 1.000.000	<u>Greedy</u> 12.500 3	<u>Beam Search</u> 10.000	<u>Branch 'n bound</u> ~ 10 min	<u>Hill Climber</u> Iteraties cut
1	8	-3	-3	-3	-3	-
2	14	-6	-6	-6	-6	-
3	20	-8	-8	-9	-9	-
4	36	-12	-13	-13	-13	-
5	50	-14	-20	-21	-20	-
6	36	-21	-25	-23	-21	-
7	36	-32	-37	-38	-36	-
8	50	-24	-28	-29	-28	-
9	50	-27	-33	-34	-33	-

Energiën per algoritme (3D, lengte/3, ~20 minuten)

	<u>Lengte</u>	<u>Random walk</u> 1.000.000	<u>Greedy</u> 15.000 2	<u>Beam Search</u> 6000	<u>Branch 'n bound</u> ~ 20 min	<u>Hill Climber</u> Iteraties cut
1	8	-3	-3	-3	-3	-
2	14	-7	-7	-7	-7	-
3	20	-9	-11	-11	-10	-
4	36	-14	-18	-18	-18	-
5	50	-18	-30	-30	-	-
6	36	-26	-35	-35	-33	-
7	36	-38	-58	-57	-	-
8	50	-35	-47	-47	-	-
9	50	-27	-50	-53	-	-

Alleen cysteine met alle algoritmes
Greedy onderling? + random?
Bnb beam in percentages?
Andere vergelijkingen?

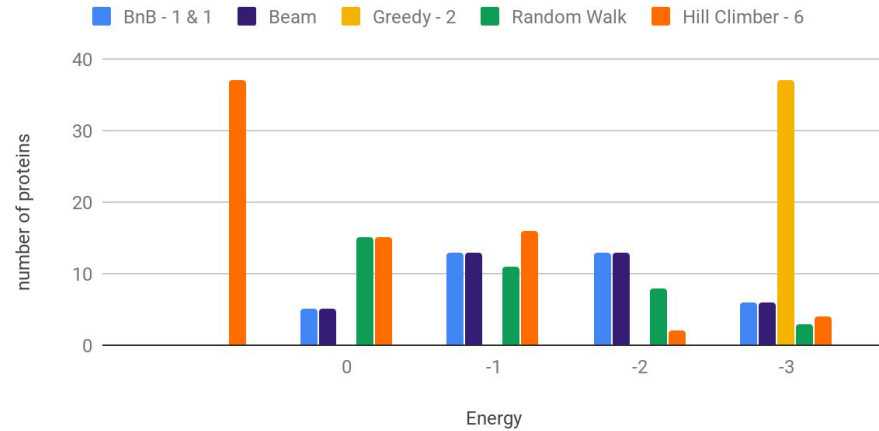
**Afhankelijk van timing 10 min voor
presentatie**

Vergelijking - Kort

2D

HHPHHHPH (2D)

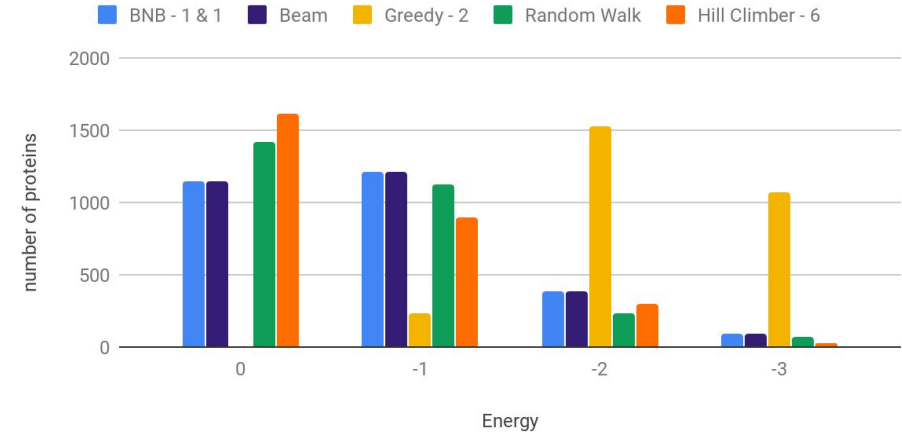
37 Proteins Folded



3D

HHPHHHPH (3D)

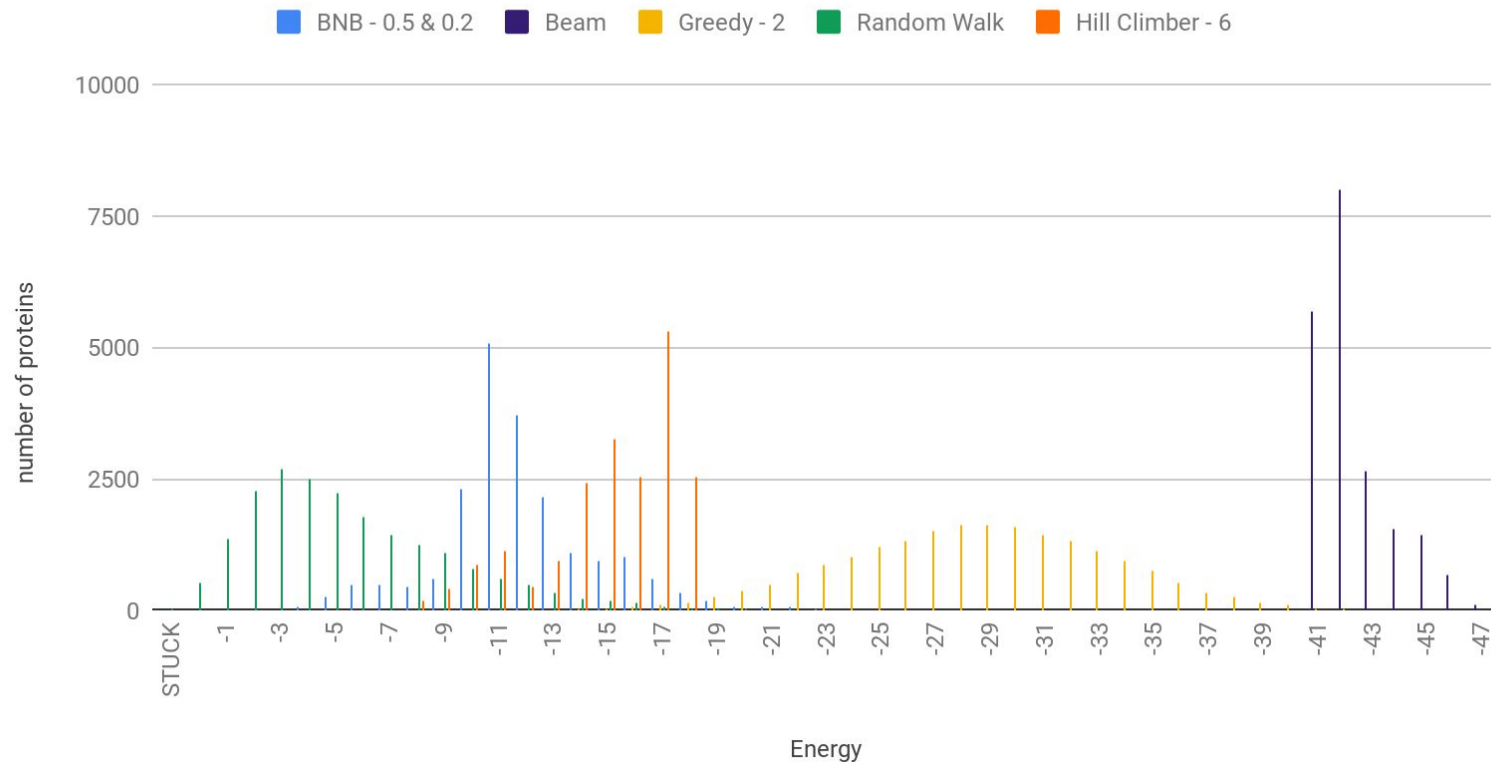
2834 Proteins Folded



Vergelijking - Lang

HCPHPCPHPCCHCHPHPPPHPPPHPPPHPCPHPPPHPHHHCCCHCHCHHH (3D)

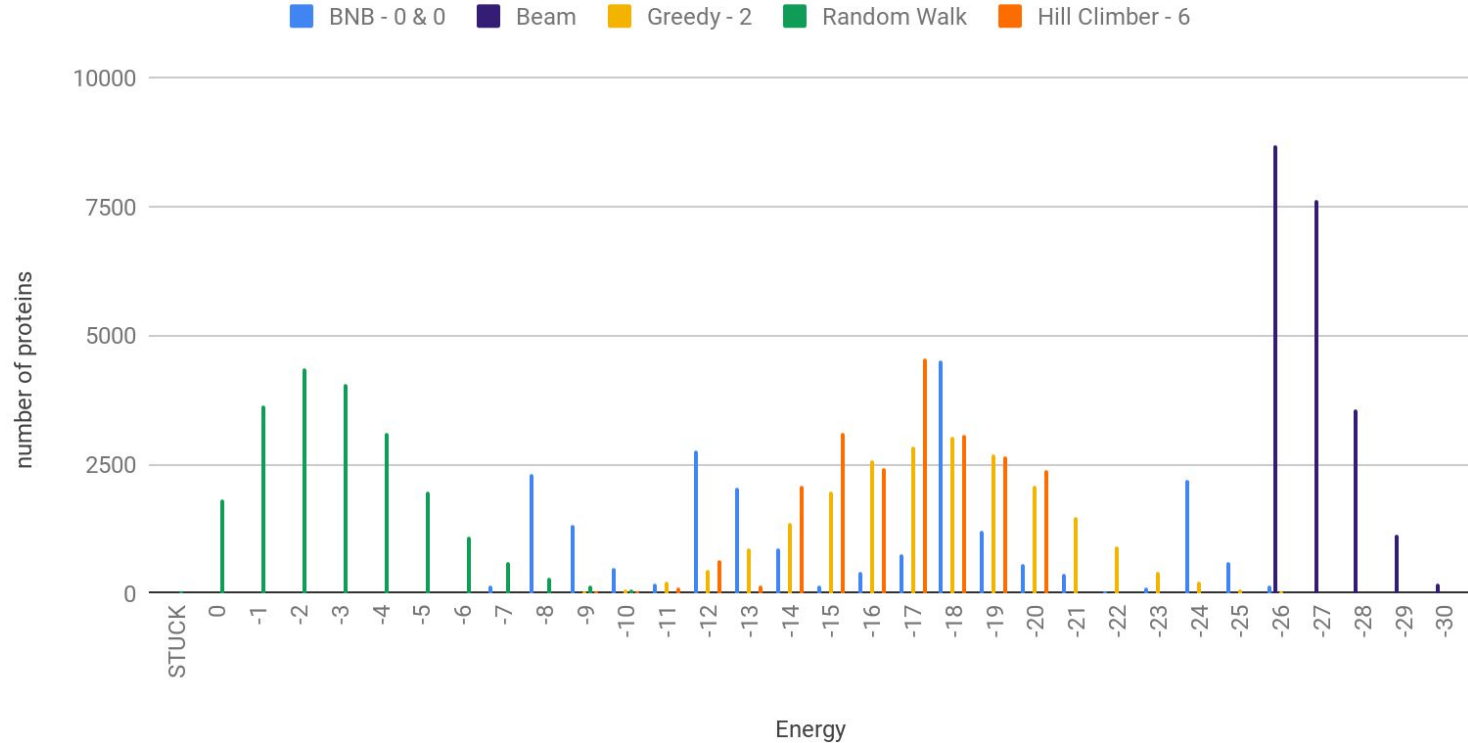
20111 Proteins Folded



Vergelijking - Cysteine

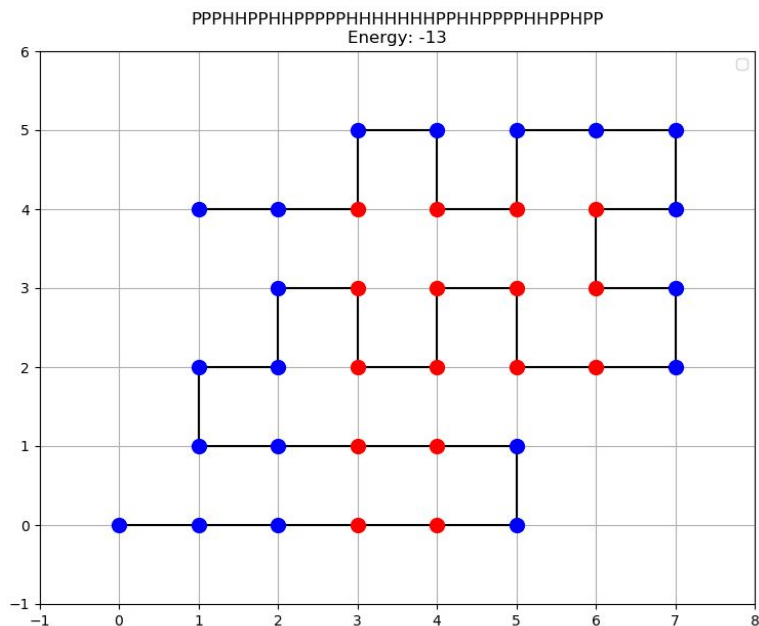
HHHPHPHPHHHPHPPHPPPHPPPHPPPHPPPHPPHHPHHHPHPHPHPHH (3D)

21216 Proteins Folded

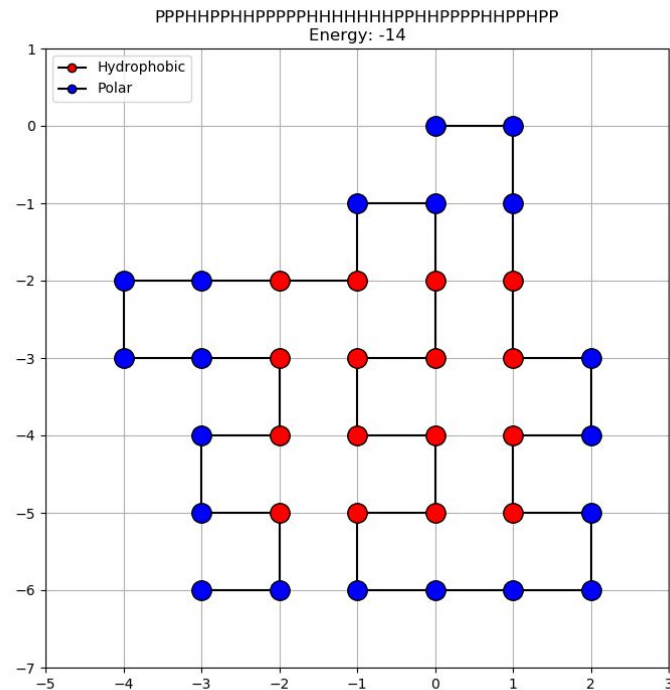


Vouwing proteïne 4

Branch 'n bound, 0.75 & 0.25, 00:14:22,
Matrix size: $2 * \text{lengte} - 1$

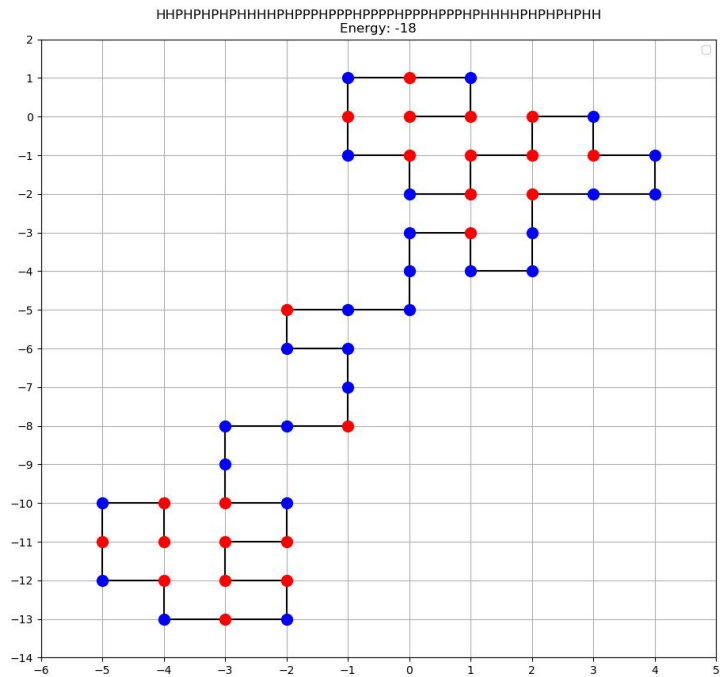


Look-ahead 2, 10.000 proteïnes, 00:02:04,
Matrix size: $\text{lengte} / 2$

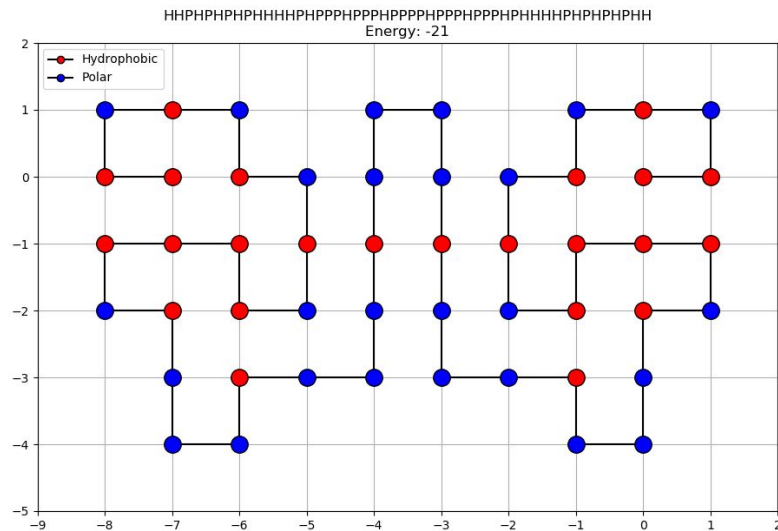


Vouwing proteïne 5

Greedy, 100.000 protéines, 00:15:54,
Matrix size: lengte



Look-ahead 5, 100.000 protéines, 07:18:20,
Matrix size: lengte / 2



Referenties

Engstler, J., & Giovambattista, N. (2018). Comparative Study of the Effects of Temperature and Pressure on the Water-Mediated Interactions between Apolar Nanoscale Solutes. *The Journal of Physical Chemistry B*, 123(5), 1116-1128.

Chen, M. & Huang, W. (2005). A Branch and Bound Algorithm for the Protein Folding Problem in the HP Lattice Model. *Genomics, Proteomics & Bioinformatics*, 3(4), 225-230.

Selkoe, D. J. (2003). Folding proteins in fatal ways. *Nature*, 426(6968), 900.

Zhang, Jinfeng, Samuel C. Kou, and Jun S. Liu. (2007). Biopolymer structure simulation and optimization via fragment regrowth Monte Carlo. *Journal of Chemical Physics* 126(22): 225101.