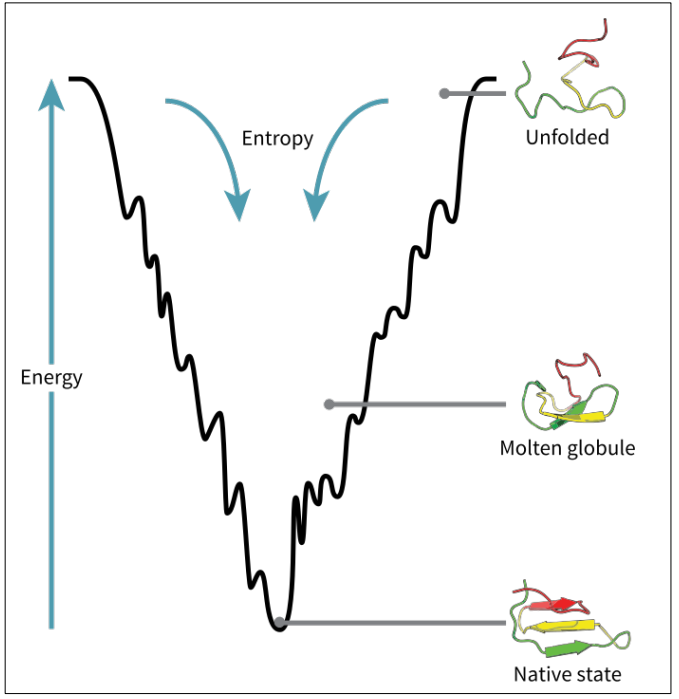
AKE Franz-Arnold
Master 2 Bioinformatique
Université Paris Diderot / Paris 7
Projet Court Python

A MONTE CARLO ALGORITHM FOR SIMPLIFIED PROTEIN MODEL FOLDING

Based on the article, ***A replica exchange Monte Carlo algorithm for protein folding in the HP model, BMC Bioinformatics 2007***
Chris Thachuk, Alena Shmygelska and Holger H

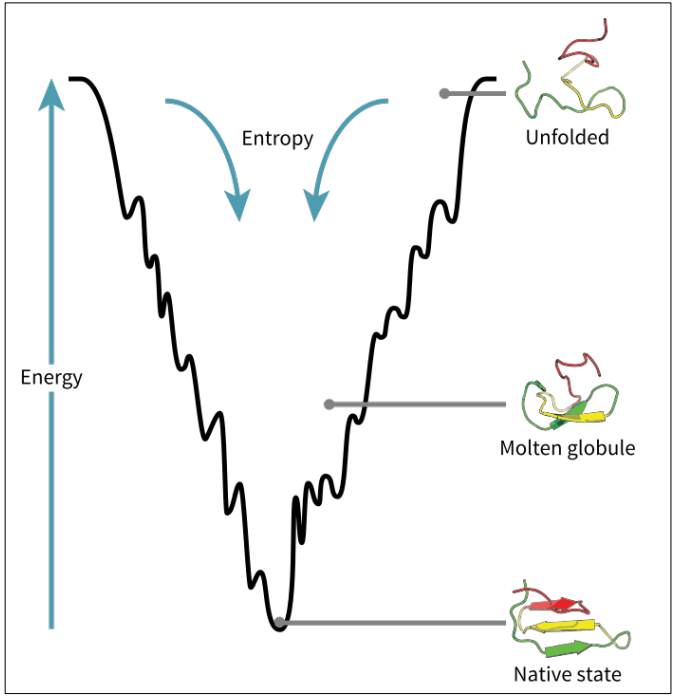
Introduction

- Ab initio Protein folding problem
  - Predicting protein tertiary structure from given amino acid sequence
- Ab initio protein folding problem is NP-HARD
- Approximation and Heuristique Methods

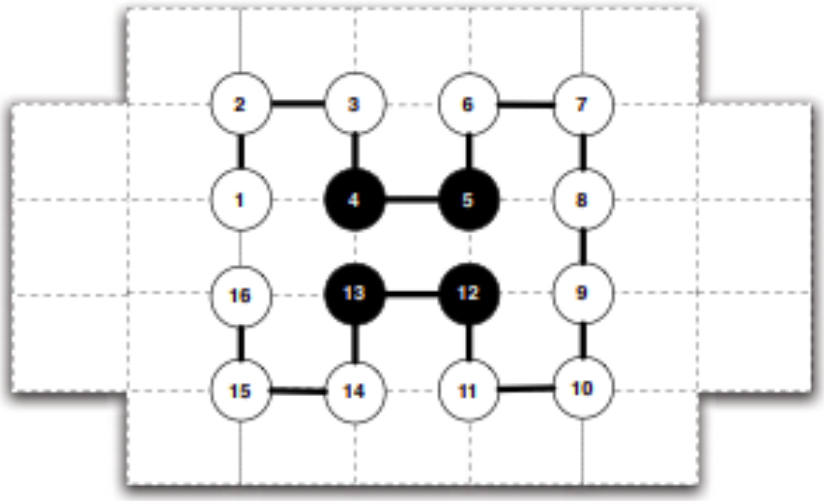


State of the Art

- Somes approaches to solving the prediction of protein folds
  - Chain growth algorithm
    - **PERM** (Monte Carlo based)
  - Ant colony optimization
    - **ACO-HPPF3** (Population based stochasticed search)



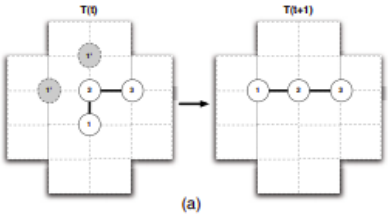
Sequences HP and lattice representation

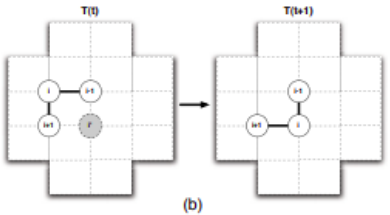


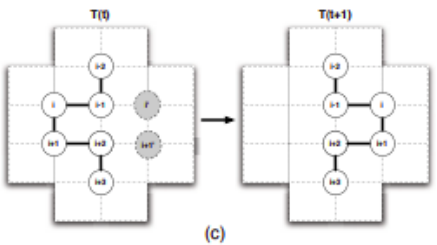
- Hydrophobic Polar (HP) Model
  - Ex : sequence ~ P3H2P6H2P3

  - Energy of conformation is determined by the number oh H-H contacts between topological neighbours

Moves Types







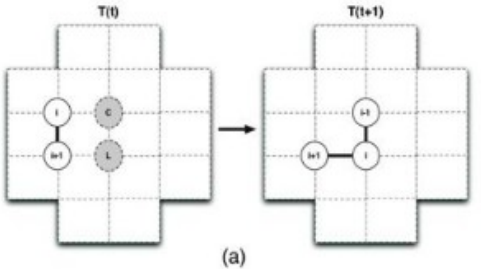
➔ **VHSD Moves**

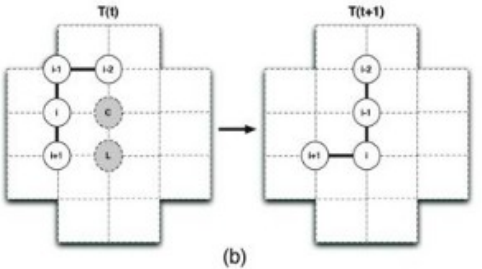
~ end moves

~ corner moves

~ crankshaft moves

➔ **Pull Moves**





Algorithme de recherche de monte carlo

Procedure `MCsearch`$(\phi,\ c,\ \nu)$

Input: ϕ – the number of search steps to perform, c – the current conformation, and ν the search neigbourhood

Output: c' – the modified conformation

for $i \leftarrow 1 \ldots \phi$ **do**
  $c' \leftarrow c$;
  $k \leftarrow \widehat{\mathcal{U}}(1, n)$;
  $c' \leftarrow \mathcal{M}(c', k, \nu)$;
  $\Delta E \leftarrow E(c') - E(c)$;
  if $\Delta E \leq 0$ **then**
    $c \leftarrow c'$;
  else
    $q \leftarrow \mathcal{U}(0, 1)$;
    if $q > e^{\frac{-\Delta E}{T}}$ **then**
      $c \leftarrow c'$;
    endif
  endif
endfor

- *Random selection of a residue k among the sequence*
  - Make the moves specified for residue k
  - Check the difference beetween the new conformation and the old
- *Save the conformation if the energy decrease or if a random uniform selection of a q number is greater than a certain parameter (temp ; energy variation)*
- *Repeat the operation n steps*

Python Implementation

Conformation class
Ordered Dict of Residue object
Energy (int)

& functions

Grille class
Taille (int)

& functions

Residue class
Type the Residue (String)
Coord X, Y (int)

& functions

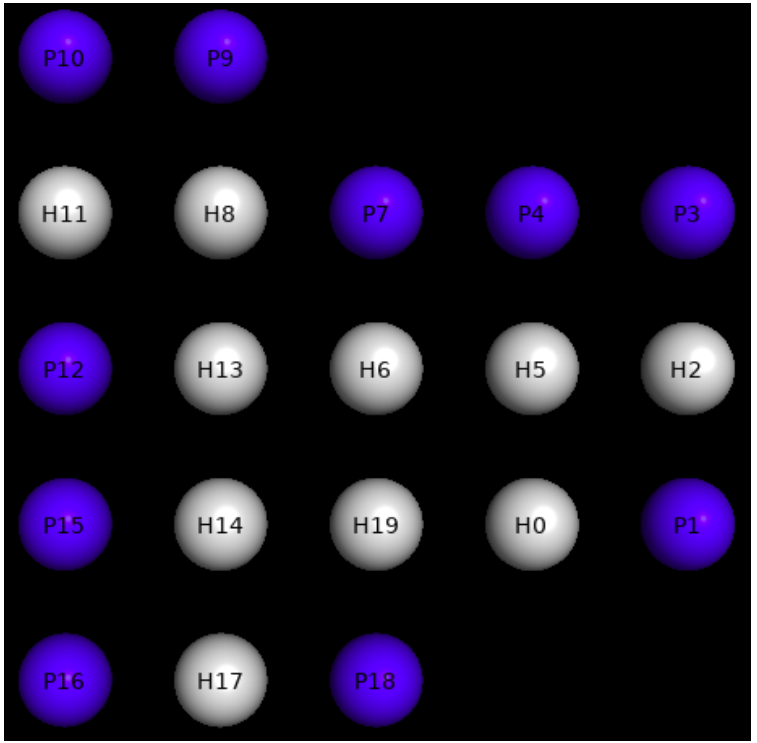
Pull class
& functions

pull_mouvement function

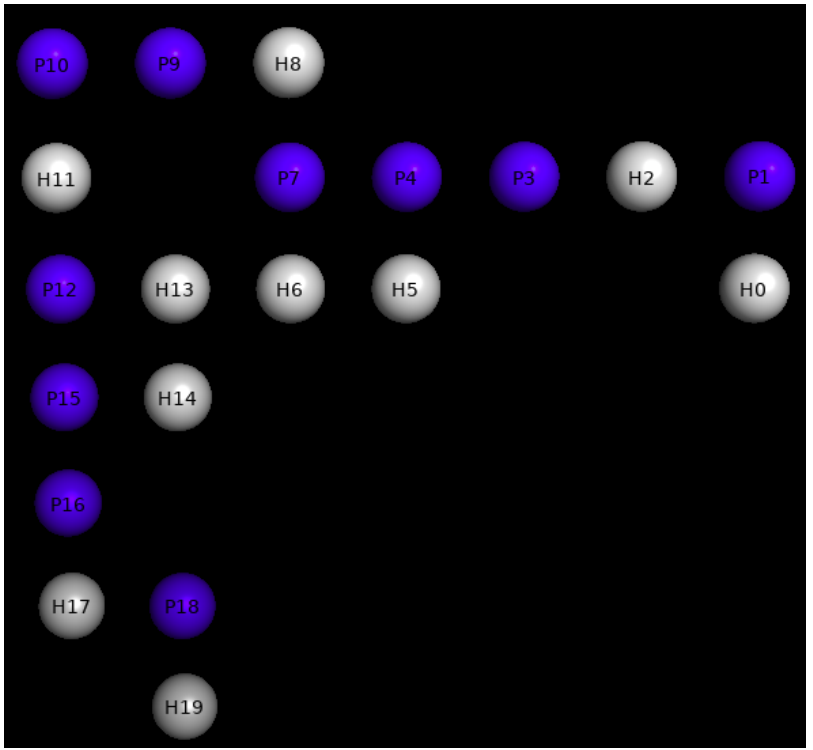
vhsd class
& functions

end_mouvement
corner_mouvement
crankshaft_mouvement

Benchmark



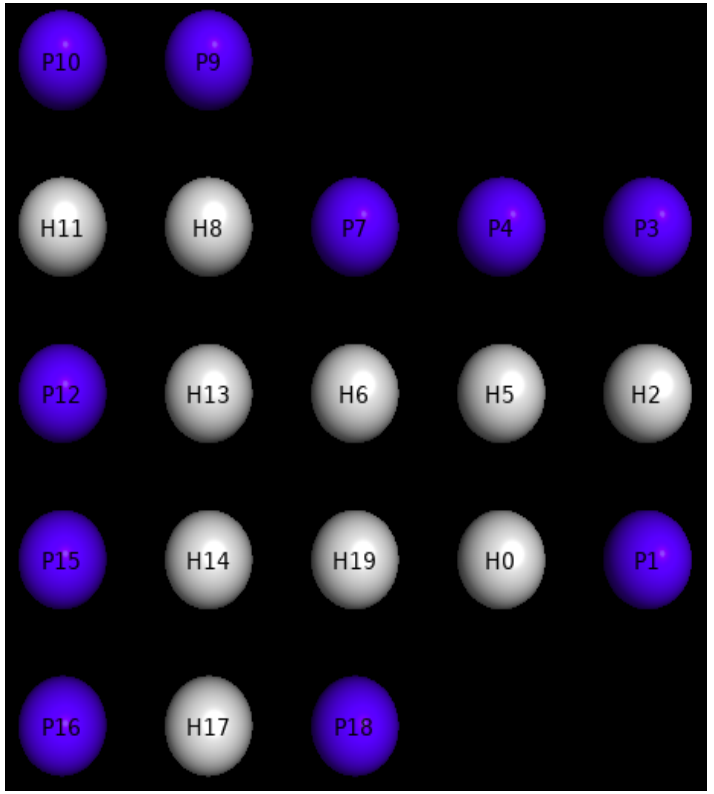
nb_steps = 500
« VHSD »
Temp = 160
Time ~
0m1,079s



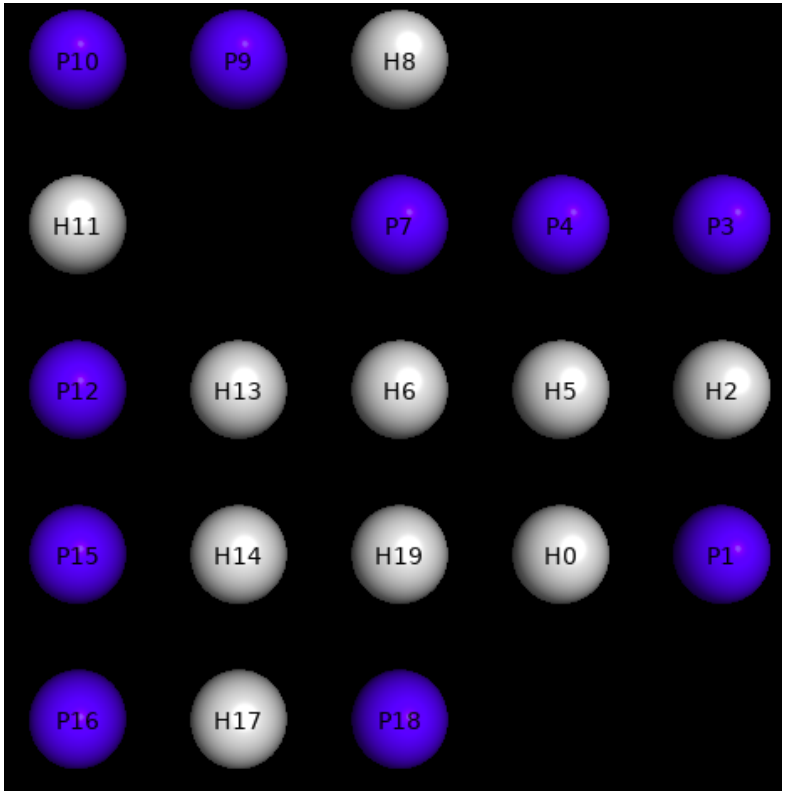
Input sequence :
HPHPPHHPHPPHPHHPPHPH
Start energy : -9

Output sequence :
HPHPPHHPHPPHPHHPPHPH
Final energy : -1

Benchmark



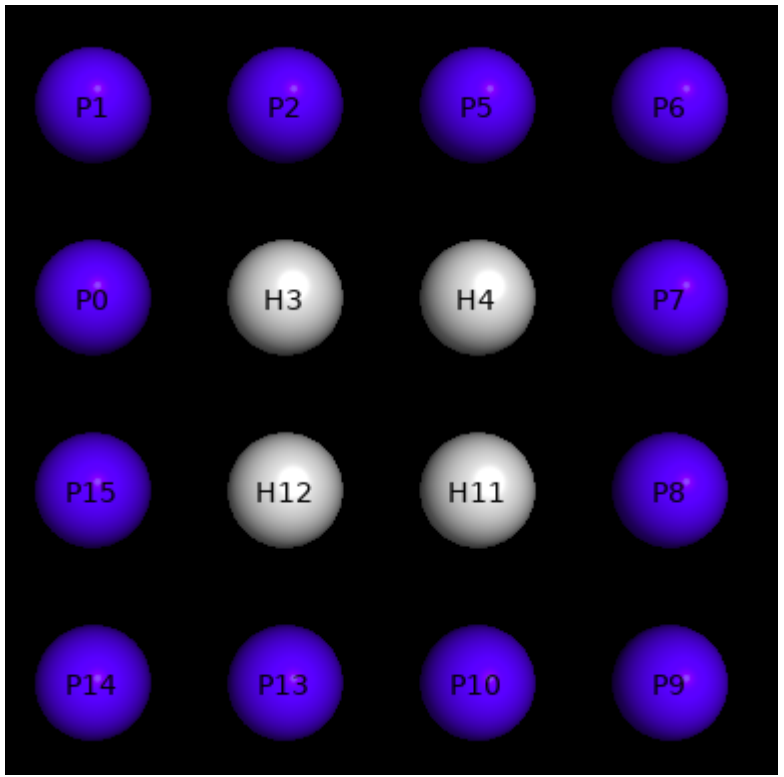
nb_steps = 500
« Pull »
Temp = 160
Time ~
0m1,672s



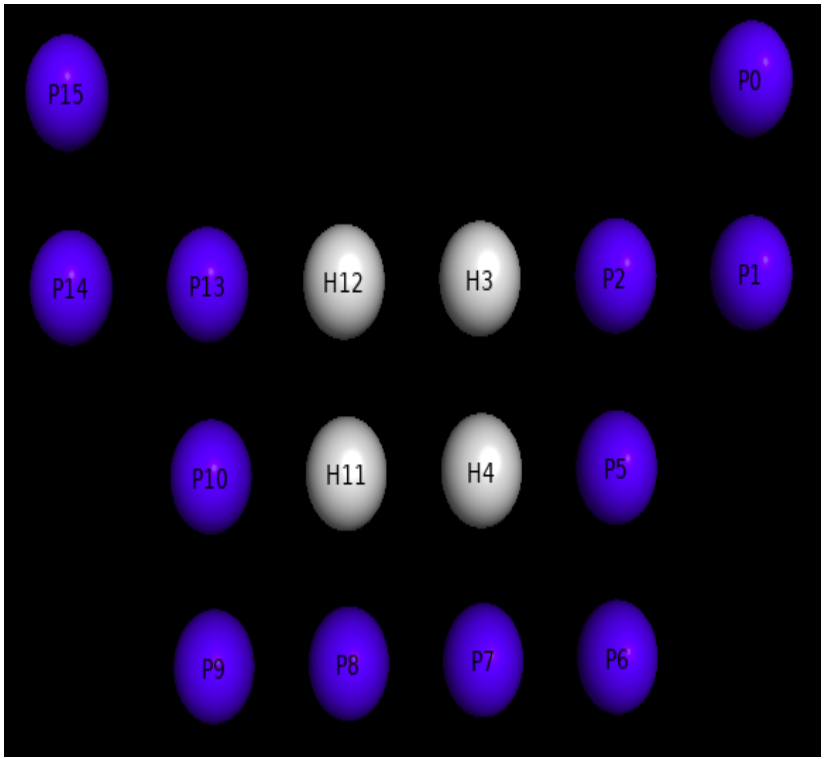
Input sequence :
HPHPPHHPHPPHPHHPPHPH
Start energy : -9

Output sequence :
HPHPPHHPHPPHPHHPPHPH
Final energy : -7

Benchmark



nb_steps = 500
« VHSD »
Temp = 160
Time ~
0m0,869s



Input sequence :
PPPHHPPPPPPHHPPP
Start energy : -2

Output sequence :
PPPHHPPPPPPHHPPP
Final energy : -2

Conclusion

- Bad performances for minimise the energy of conformation

- Bad performances on protein with hydrophobic cores