AKE Franz-Arnold
Université Paris Diderot – Paris 7
Master Biologie – Informatique / Master 2
Septembre 2018

REPLIEMENT D'UN MODÈLE SIMPLIFIE DE PROTÉINE PAR UN ALGORITHME DE MONTE CARLO

Thachuk C, Shmygelska A, Hoos HH. A replica exchange Monte Carlo algorithm for protein folding in the HP model. BMC Bioinformatics. 2007 Sep

Introduction

Les méthodes de repliement des protéines à partir des séquences protéiques primaires s'appuient principalement dans leurs résolutions sur la minimisation d'une fonction d'énergie associé à ces séquences. Notre travail a consisté à implémenter un algorithme de recherche de Monte Carlo sur des modèles de séquences particulières, les séquences Hydrophobes / Polaires à partir desquels une conformation pourra être établi par l'intermédiaire d'une grille sur 2 dimensions, et l'implémentation de types de mouvements différents pouvant être réalisés par chacun des résidus dans des conditions particulières. Ainsi, l'exécution répétée des déplacements aléatoires de ces résidus pendant une nombre d'étapes donnés permettra d'observer un ensemble large d'états conformationnels associés à la séquence Hydrophobe / Polaire.

Matériels et Méthodes

Le modèle de séquence sur lequel s'exécutera l'algorithme est un modèle dit « *Hydrophobe _ Polaire »* préalablement introduit par *Dill & all* en 1985, où les acides aminés sont classifiées en 2 catégories « hydrophobe » et « polaire ». La séquence HP est ainsi, un enchaînement de résidus H et P, auxquel des coordonnées géométriques sont attribuées afin d'être intégré à une grille à 2 dimensions pour permettre d'observer sa conformation (ex : *HPPPHPHPPHHHPPH*).
Ces informations sont stockées dans un fichier au format *txt* pour chaque séquence HP (*répertoire Data/sequencesHP*). L'énergie de conformation associée à la séquence correspond au nombre de contact entre résidus hydrophobes voisins dans la grille 2D (*contribution négative de -1 pour chaque contact*). Pour chaque résidu de la séquence HP, plusieurs types de mouvement sont possible suivant les contraintes géométrique associées à la conformation de la séquence dans la grille ;
Il y a un mouvement de **type *VHSD*** pouvant concerner le déplacement d'un résidus unique (*End_moves*) ou la rotation de plusieurs résidus (*Crankshaft_moves*) et le mouvement de ***type PULL*** pouvant concerner le déplacement d'une suite de résidus de la séquence.

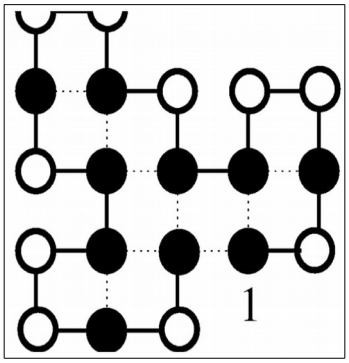
L'exécution de l'algorithme de repliement de Monte Carlo pour le repliement de la séquence HP implique plusieurs paramètres devant être préalablement définis ; Tout d'abord le nombre d'étapes de recherche de l'algorithme de Monte Carlo sur la séquence HP (*nb_step*), puis le type de mouvement que devra subir un résidu choisi aléatoirement parmi la séquence à chaque étape.
Ces paramètres seront à passer en ligne de commande dans un terminal UNIX, au programme REMC qui retournera à l'issue de l'exécution du programme, un fichier pdb décrivant la nouvelle conformation.

Résultats

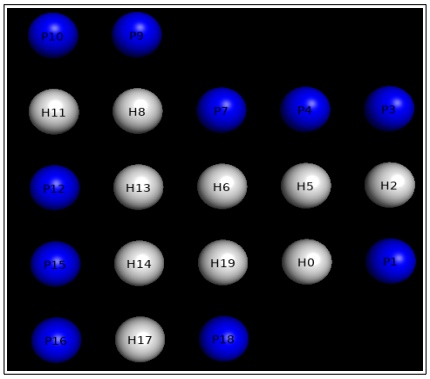
Nous avons réalisé sur un ensemble de 2 séquence HP issues de l'article, l'algorithme REMC avec pour chacune des séquences HP des paramètres spécifiques.

- *Séquence 1* : HPHPPHHPHPPHPHHPPHPH (taille = 20, énergie de départ = -9)
  - 1er essai : *nombre d'étapes = 500* *temps d'exécution = 0m1,672s / énergie de sortie -7* ***mouvement de type « pull »***
  - 2eme essai : *nombres d'étapes = 500* *temps d'exécution = 0m1,079s / énergie de sortie -1* ***mouvement de type « vhsd »***

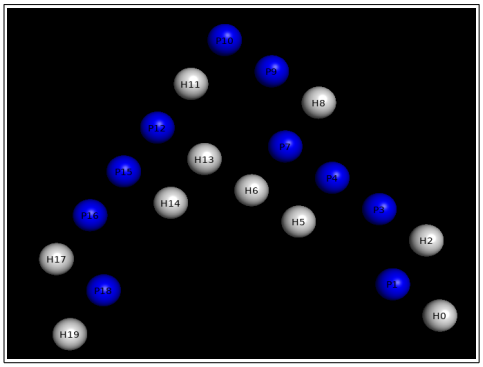
- *Séquence 2* : PPPHHPPPPPPHHPPP (taille = 16, énergie de départ = -2)
  - 1er essai : *nombre d'étapes = 500* *temps d'exécution = 0m1,212s / énergie de sortie -2* ***mouvement de type « pull »***
  - 2eme essai : *nombres d'étapes = 500* *temps d'exécution = 0m0,869s / énergie de sortie 0* ***mouvement de type « vhsd »***



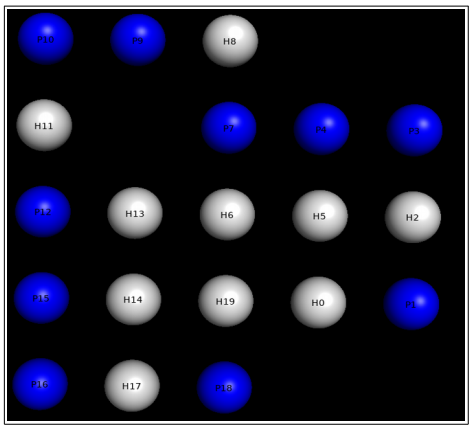
a) Conformation de la séquence 1 avant l'algorithme REMC dans la grille (résidu Hydrophobe Noirs ;résidu Polaires en blanc



b) Conformation de la séquence 1 avant l'algorithme REMC ; Energie de conformation : -9 . Residu Hydrophobe en blanc et Polaire en bleu



c) Conformation de la séquence 1 après 500 runs de l'algorithme REMC associé au mouvement vhsd ; Energie de conformation : -1 Residu Hydrophobe en blanc et Polaire en bleu



d) Conformation de la séquence 1 après 500 runs de l'algorithme REMC associé au mouvement pull ; Energie de conformation : -7. Residu Hydrophobe en blanc et Polaire en bleu

Fig 1. Exemple d'Observation de la conformation de la séquence 1 avant l'application de l'algorithme REMC dans la grille (a) et observation de la conformation via pymol par le fichier pdb retourné par le programme (b) ; Augmentation importante de l'énergie de conformation de la séquence 1 après 500 runs de l'algorithme REMC avec le mouvement_vhsd avec une variation d'énergie de 8 (c) ;Faible augmentation de l'energie de conformation de la séquence 1 après 500 runs de l'algorithme REMC avec le mouvement_pull (d).

On peut observer que l'algorithme de recherche de Monte Carlo associée aux mouvements vhsd à une plus grande action sur le repliement de la protéine avec une rupture du coeur hydrophobe comparativement aux mouvements pull qui modifient peu la conformation de départ de la protéine.

Annexe

Dans le but d'implémenter l'algorithme de Recherche de Monte Carlo, le code python a été construit suivant l'organisation suivante :

- Création d'une ***classe Résidue*** associée à chaque résidu de la séquence HP et contenant le type du résidu (H ou P) et les coordonnées géométriques X et Y

- Création d'une ***classe Conformation***, correspondant à un ensemble d'objets Résidus contenue dans une collection (OrderedDict) auquel est associée une énergie.

- Création d'une ***classe Grille***, contenant une grille à 2 dimensions (numpy.array) sur lequel l'on pourra projeter la conformation grâce aux coordonnées des résidus.
Celle-ci contient des fonctions propres, permettant de vérifier des conditions préalable sur la grille initialisé avec une conformation (fonction ~ maj_grille(), etc.) pour des opérations de mouvements des résidus

- Création des ***classes de mouvements VHSD et Pull***, associées aux mouvements possible pour les résidus de la séquence. Celles-ci prennent en entrée un résidu, effectuent les mouvements associés à la classe et retournent la conformation modifiée à l'issue du mouvement.

- Création de la ***classe Mc_Search,*** prenant en entrée un paramètre *np_step* correspond au nombre de runs de l'algorithme effectuant un déplacement aléatoire des résidus, une conformation, une grille, et un mouvement d'entrée.

Procedure `MCsearch`(ϕ, c, ν)

Input: ϕ – the number of search steps to perform, c – the current conformation, and ν the search neigbourhood
Output: c' – the modified conformation
for $i \leftarrow 1 \dots \phi$ **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

L' Algorithme MC search tel que présenté dans l'article et implémenté dans le code personnel.

La classe de mouvement VHSD peut réaliser potentiellement 3 mouvements qui sont *les mouvement_end , les mouvements_corners, et les mouvements crankshafts*. Dans le cas ou l'on exécute la classe VHSD dans l'algorithme de Monte carlo, un de mouvement est choisi aléatoirement et réalisé sur le résidu d'entrée.

Si le mouvement est non réalisable du fait des contraintes liées à la conformation dans la grille, les classes mouvements renvoie la classe conformation d'entrée sans aucune modification.

La réalisation d'un mouvement consiste simplement à vérifier grâce à la classe grille initialisé avec la conformation si les conditions sont réunies pour la réalisation du mouvement et modifier les coordonnées du résidu soumis dans la classe conformation.
Le paramètre T lié à la température est fixée dans le code à une valeur de 160.

- Création d'une ***fonction get_PDB()*** permettant d'obtenir un fichier PDB contenant des pseudos-atomes auxquels sont attribués les coordonnées géométriques de chaque résidu afin de visualiser la conformation dans le logiciel pymol (figures 1.b,c,d)