

Repliement d'un modèle simplifié de protéine par un algorithme de Monte-Carlo et échange de répliques

Lien du Git : <https://github.com/lisabdj/Repliement-prot-MC.git>

Introduction

Pour connaître la structure tridimensionnelle d'une protéine, il existe des méthodes très coûteuses et chronophages comme la cristallographie à rayon X et la résonance magnétique nucléaire. Pour pallier ces problèmes de coût et de temps, l'informatique est utilisée pour prédire la structure tridimensionnelle d'une protéine. Grâce à des algorithmes qui résolvent ce problème de repliement *ab initio*, il est possible de replier une protéine jusqu'à obtenir son état natif. Pour prédire cet état fonctionnel de la protéine, les algorithmes utilisent la séquence d'acides aminés et cherchent à minimiser une fonction d'énergie. Dans l'article de Thachuk & al. (2007) [1], ils ont implémenté un algorithme de Monte-Carlo afin de prédire la structure tertiaire de protéine.

L'objectif de ce projet est de réaliser un programme reprenant la méthode décrite dans l'article [1]. L'algorithme de Monte Carlo a été implémenté en python.

Matériel et méthodes

Librairies

Pour réaliser le programme, nous avons utilisé la version 3.9 de Python. Nous avons eu besoin d'installer les librairies Matplotlib et Numpy.

Traduction de protéine en modèle HP

Dans un premier temps, nous avons traduit notre séquence protéique entrée en code à 1 lettre par l'utilisateur en acide aminé *polaire* ou *hydrophobe*. Grâce à cette traduction, nous obtenons un modèle simplifié de la protéine : le modèle hydrophobe polaire (HP). Cela nous permet de travailler uniquement sur les propriétés d'hydrophobicité des acides aminés.

Matrice bidimensionnelle

Nous avons représenté notre protéine sur une matrice bidimensionnels. Cette matrice 2D a été réalisé grâce à la librairie Numpy. Les coordonnées des acides aminés sont stockées.

VSHD mouvement

Dans ce projet, nous avons utilisé les mouvements VSHD utilisés pour la première fois par Verdier et Stockmayer [2]. Il y a 3 types de mouvements VSHD : les mouvements de fin, les mouvements de coin et les mouvements du vilebrequin à deux résidus.

Le mouvement de fin est effectué sur un résidu à l'extrémité de la séquence HP. Le résidu pivote par rapport à son voisin (cf. Figure 1 ci-contre), afin de former un coin. S'il existe plusieurs position valide, nous choisissons une position uniformément au hasard.

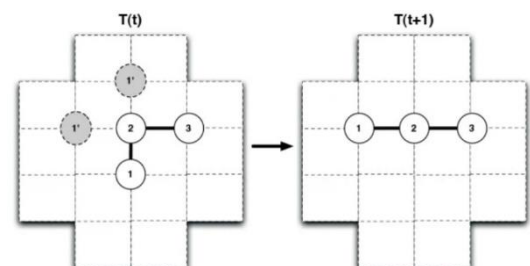


Figure 1: Mouvement de fin (figure de l'article [1])

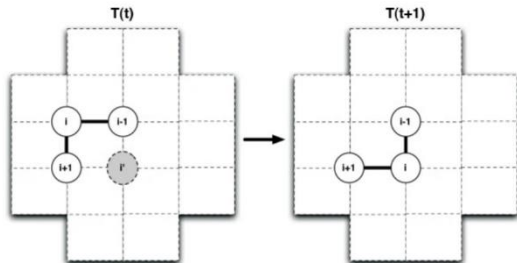


Figure 2: Mouvement de coin (figure de l'article [1])

Le mouvement de coin est effectué sur n'importe quel résidu, sauf les résidus à l'extrémité de la séquence HP. Pour que ce mouvement soit possible (cf. Figure 2), les voisins connectés du résidu qui nous intéresse doit être mutuellement adjacents à une autre position libre sur la matrice 2D.

Le mouvement du vilebrequin à deux résidus est effectué si deux résidus forment un coude en forme de U. Le mouvement implique une rotation de 180° (cf. Figure 3), si les positions en face sont vides.

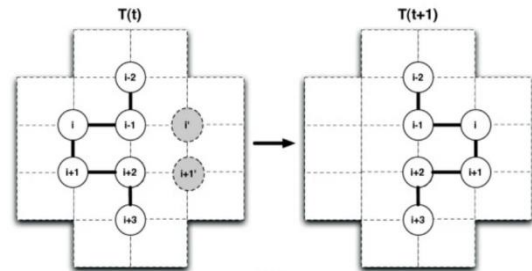


Figure 3: Mouvement du vilebrequin (figure de l'article [1])

Calcul d'énergie

L'énergie d'une conformation est calculée comme le nombre de contacts d'acides aminés hydrophobes non connectés comme deux résidus dans la chaîne protéique. Chaque contact d'acides aminés non connectés H-H ajoute -1 à l'énergie totale de la conformation. Les autres contacts n'ajoutent rien à l'énergie totale.

Procédure de Monte Carlo

Nous avons implémenté une procédure de recherche Monte Carlo. La procédure est effectuée sur une séquence protéique HP et un nombre prédéterminé d'étapes. Nous choisissons uniformément au hasard un acide aminé sur lequel un mouvement VSHD est possible. Puis, nous calculons la différence d'énergie entre la nouvelle et l'ancienne conformation de la protéine. Si la différence est inférieure ou égale à 0, nous gardons la nouvelle conformation.

Sinon, nous calculons une valeur $e^{-\frac{\text{différence d'énergie}}{\text{temperature}}}$, et nous tirons au hasard un nombre q entre 0 et 1. Si q est supérieure à cette valeur, nous gardons la nouvelle conformation.

La procédure de Monte Carlo ne contient que les mouvements VSHD. Par manque de temps, la programmation des mouvements de tractions n'a pu être possible.

Résultat

Nous avons fait tourner le programme avec 100 itérations sur la séquence protéique « TGLMFWGVYGATT » avec une température de repliement égale à 0.3°C. Nous obtenons la variation d'énergie suivante :

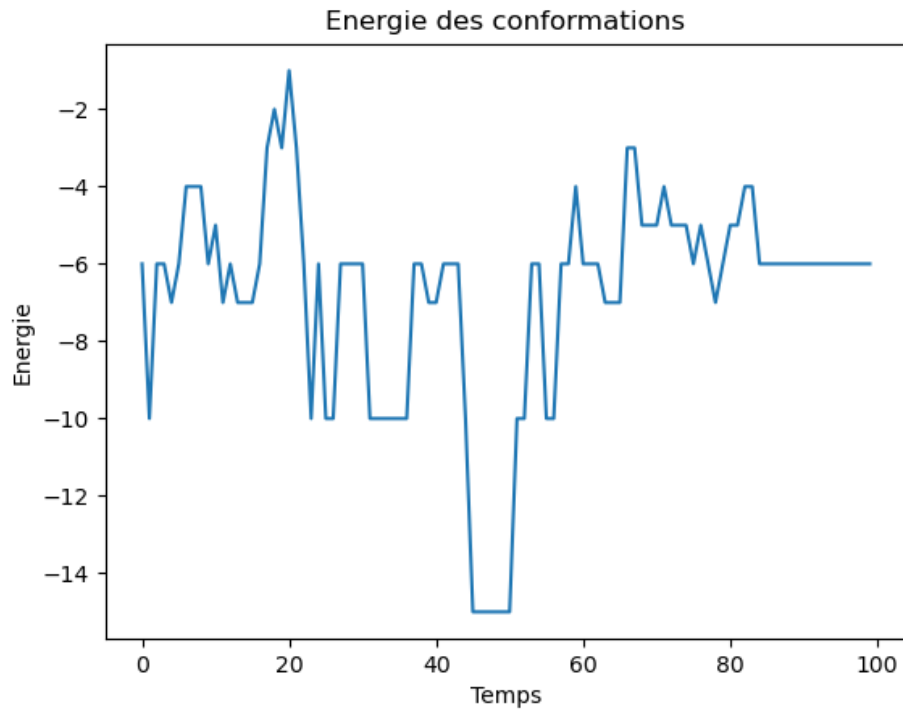


Figure 4: Variation des énergies des conformations en fonction du temps ($T=0.3$)

Sur la même séquence en acides aminés et avec une température=0.6°C, nous obtenons la variation d'énergie suivante :

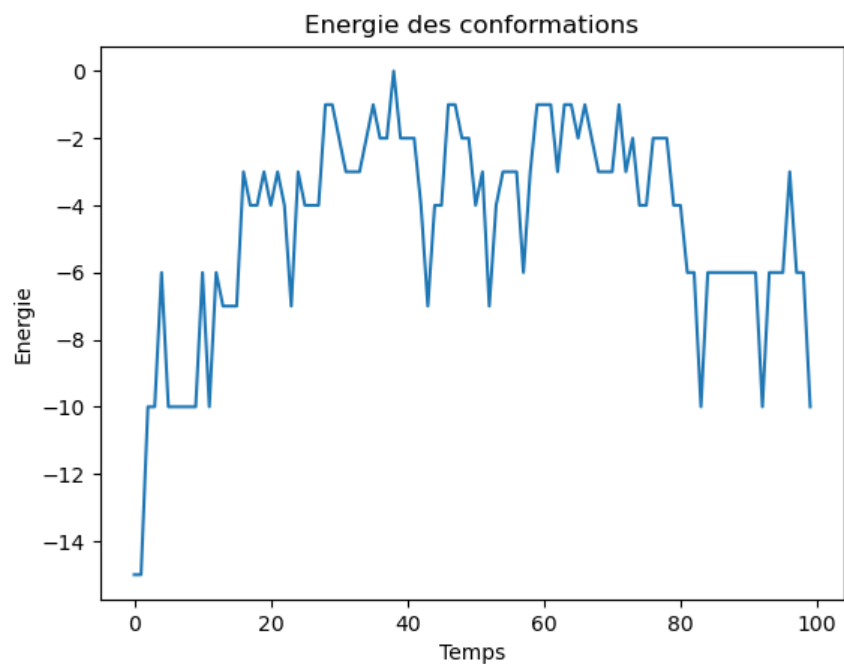


Figure 5: Variation des énergies des conformations en fonction du temps ($T=0.6$)

Sur les figures 4 et 5, nous pouvons observer que l'énergie des conformations varie beaucoup au cours du temps. La procédure Monte Carlo permet de garder les conformations qui minimisent l'énergie, mais nous gardons, quelques fois, des conformations de plus haute énergie afin de ne pas rester bloquer dans un minimum local d'énergie.

Discussion

Le programme réalisé permet, uniquement, de déplacer les résidus selon les mouvements VSHD. Les mouvements de tractions sont manquants au programme.

Nous avons choisi de faire tourner le programme sur 100 itérations. Le programme est capable de trouver les conformations qui minimisent l'énergie calculée selon les contacts d'acides aminés H-H. Grâce à ce programme, il est possible de modéliser de façon simple le repliement que la protéine pourrait avoir dans son état fonctionnel.

Conclusion

Le programme réalisé au cours de ce projet est capable de trouver un repliement simplifié de protéine qui minimise l'énergie en fonction de la procédure Monte Carlo.

Il serait intéressant de rajouter les mouvements de tractions à la procédure Monte Carlo, et de choisir uniformément au hasard entre les mouvements de tractions et VSHD afin de replier la protéine dans une conformation plus complexe encore.

Bibliographie

[1] Thachuk, C., Shmygelska, A. & Hoos, H.H. A replica exchange Monte Carlo algorithm for protein folding in the HP model. *BMC Bioinformatics* **8**, 342 (2007). <https://doi.org/10.1186/1471-2105-8-342>

[2] Verdier PH, Stockmayer WH: Monte Carlo Calculations on the Dynamics of Polymers in Dilute Solution. *The Journal of Chemical Physics*. 1962, 36: 227-235.

Annexes

Structure du programme réalisé

1. Traduction de la protéine selon le modèle HP : on utilise la fonction traduction()
2. Initialisation – Conformation C0 : grâce à la fonction init_matrice(), nous créons la conformation C0.
3. Conformation C1 : nous réalisons le premier mouvement sur l'un des deux résidus à l'extrémité de la séquence. Le résidu est choisi uniformément au hasard
4. Conformation C2 à CN : nous utilisons la fonction procedure_MC() sur notre conformation actuelle. Un acide aminé est choisi au hasard parmi les acides aminés qui peuvent être déplacés. Les acides aminés déplaçables sont sélectionnés grâce à la fonction déplacement_possible(). Après cette sélection, nous effectuons le mouvement adéquate sur l'acide aminé avec la fonction vshd_moves(). Puis, on calcule la différence d'énergie entre la nouvelle conformation et la précédente qui était stockée, grâce à la fonction energie_hydrophobe(). Si cette différence est inférieure ou égale à 0, nous gardons la nouvelle conformation. Sinon, nous la gardons si q est supérieur à $e^{-\frac{\text{différence d'énergie}}{\text{temperature}}}$. Toutes les matrices 2D, les coordonnées de chaque repliement et les énergies des conformations sont stockées au cours de la procédure.

Difficultés rencontrées

J'ai rencontré de nombreuses difficultés pour coder les mouvements VSHD, pour régler tous les cas de figures liés à ces types de déplacements. Cela m'a pris énormément de temps et je n'ai pas pu terminer la partie du programme sur les mouvements de tractions. J'ai laissé les fonctions non terminées dans mon script, mais elles ne sont jamais utilisées pour le repliement de la protéine.