

Rapport

Replica Exchange Monte Carlo Algorithm

JEAN MARIE Roude
Étudiant en Master 2 Bioinformatique

année scolaire 2023/2024

Responsable :
Catherine Etchebest
Jean-Christophe GELLY

1. Introduction

Contexte et objectif

Au sein de l'organisme, les protéines contribuent au bon fonctionnement du métabolisme par le biais d'interactions intermoléculaires. Celles-ci sont composées d'un embranchement de plus petites molécules appelées 'résidus' qui forment des interactions intra-moléculaires contribuant à stabiliser la conformation de la protéine (dont certaines interactions inter-moléculaires). En effet, les propriétés des atomes entre eux vont régir le repliement de la protéine et cela aura un impact sur la conformation finale de la protéine. Malgré l'aspect dynamique des protéines, on les retrouve sous certaines conformations plus ou moins natives. Les interactions qui existent entre les atomes de la molécule participent à la stabilisation de la protéine vers un état conformationnel. De par ce principe, les protéines possèdent des états dits "natifs" correspondant à des conformations plus ou moins stables et donc à un niveau d'énergie faible. Sans être totalement rigide, c'est cette stabilité qui va permettre la bonne interaction de la protéine avec d'autres molécules de l'organisme.

De ce fait, la capacité à représenter une conformation à l'état native d'une protéine demeure être un enjeu important pour permettre une meilleure étude des interactions entre les molécules de l'organisme. Différents algorithmes ont à ce jour été conçus pour résoudre cette problématique, avec quelques avancées notamment avec celle d'AlphaFold. Mais d'autres algorithmes ont été développés tel que la méthode du Ant Colony Optimisation (ACO) ainsi que PERM qui est une méthode basée sur les méthodes de Monte Carlo mais ces méthodes possédaient des difficultés pour le repliement au sein d'un plan 3D ou certains types de structures (PERM repliait difficilement les protéines à coeurs hydrophobes). S'en est suivi de la publication de l'algorithme du Replica exchange Monte Carlo (REMC) en 2007 par Thachuk et al. qui su montrer de bonnes performances principalement en termes de temps d'exécution pour approximer des conformations que pourrait prendre la protéine d'intérêt.

Ces modèles se basent sur le placement des résidus au sein d'une grille 2D ou 3D, avec pour objectif la recherche de la "meilleure" conformation au sein de la grille. Ce projet a pour objectif la reproduction du REMC, les résultats ainsi que les limites pourront être commentées.

2. Matériels et Méthode

Modèle HP et Grille 2D

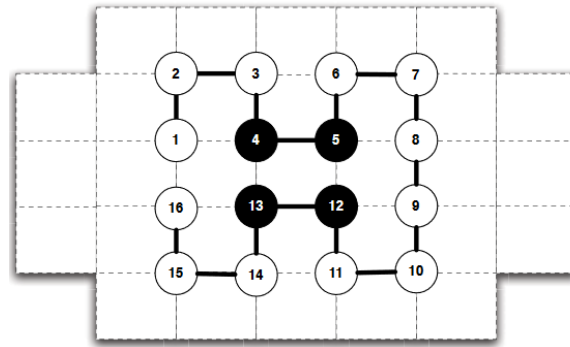


Figure 1: Conformation d'une protéine au sein d'une grille 2D.

Une grille correspond à un ensemble de points discrets sur lesquels seront placés les résidus de la protéine d'intérêt (voir Figure 1). Ces résidus pourront être placés linéairement ou aléatoirement au sein de la grille, tout en conservant la propriété que le résidu $\{i\}$ est relié au résidu $\{i+1\}$ (voisins adjacents). Le de représentation de nos résidus est le modèle HP. Ce modèle a été conçu par Dill, et correspond à une représentation simplifiée des résidus en 2 classes. La classe des résidus hydrophobes, ainsi que la classe des résidus polaires. Ainsi, on considère les résidus V, I, F, L, M, C et W comme hydrophobes, caractérisés par la lettre H. Et les autres résidus D, E, K, R, H, Y, S, T, N, Q, G, A, P polaires (ou "non hydrophobes"). Sur la Figure 1, les résidus hydrophobes sont représentés par des cercles pleins noir.

Au sein de ce modèle, on considère que l'énergie de la conformation diminue à chaque contact entre résidus hydrophobes non reliés (voisins topologiques), au sein de la Figure 1 l'énergie de la conformation est de -2.

Cette énergie est calculé à l'aide de la formule suivante:

$$E(c_i) = \sum_{j=1}^{n-1} \sum_{k=j+1}^n N_{jk}, \text{ with}$$

$$N_{jk} = \begin{cases} -1 & \text{if } j \text{ and } k \text{ are both H residues} \\ & \text{and topological neighbours;} \\ 0 & \text{otherwise.} \end{cases}$$

Monte Carlo Search Algorithm

Cet algorithme vise à minimiser l'énergie d'une conformation de départ, en lui appliquant un ensemble de mouvements. Les VSHD moves (voir Figure 2) sont introduits par Gurler et al. qui combine 3 types de mouvements : les end moves, les corner moves ainsi que les crankshaft moves. Les Pull moves (voir Figure 3) ont plus tard aussi été introduits par Lesh et al. aussi dans le cas des repliement des protéines sur une grille 2D. L'algorithme consiste à

appliquer un mouvement sur notre conformation, et la mesure de l'énergie de la nouvelle conformation permet avec une certaine probabilité de remplacer la conformation actuelle par la nouvelle. Dans le cas où l'énergie de la nouvelle est plus favorable que l'ancienne, on la conserve, sinon il existe une probabilité non nulle de permettre de remplacer la conformation actuelle par une conformation plus défavorable. En appliquant ces changements de conformations pendant plusieurs itérations, on s'attend à observer une diminution de l'énergie (qui reste dépendante de la température).

$$Pr[c \rightarrow c'] := \begin{cases} 1 & \text{if } \Delta E \leq 0, \\ e^{\frac{-\Delta E}{T}} & \text{otherwise.} \end{cases}$$

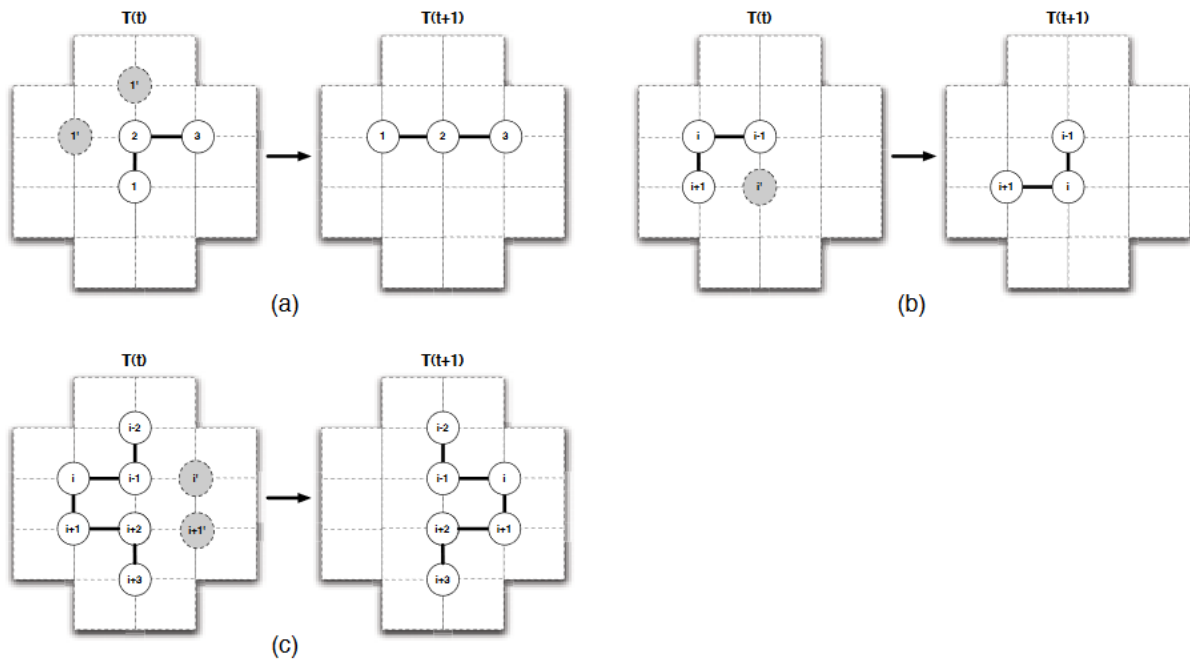


Figure 2: VSHD moves. Les cercles remplis correspondent aux cases pour lesquelles le déplacement est possible. a) Application du End move pour le résidu 1. b) Application du corner move. c) Application du crankshaft move.

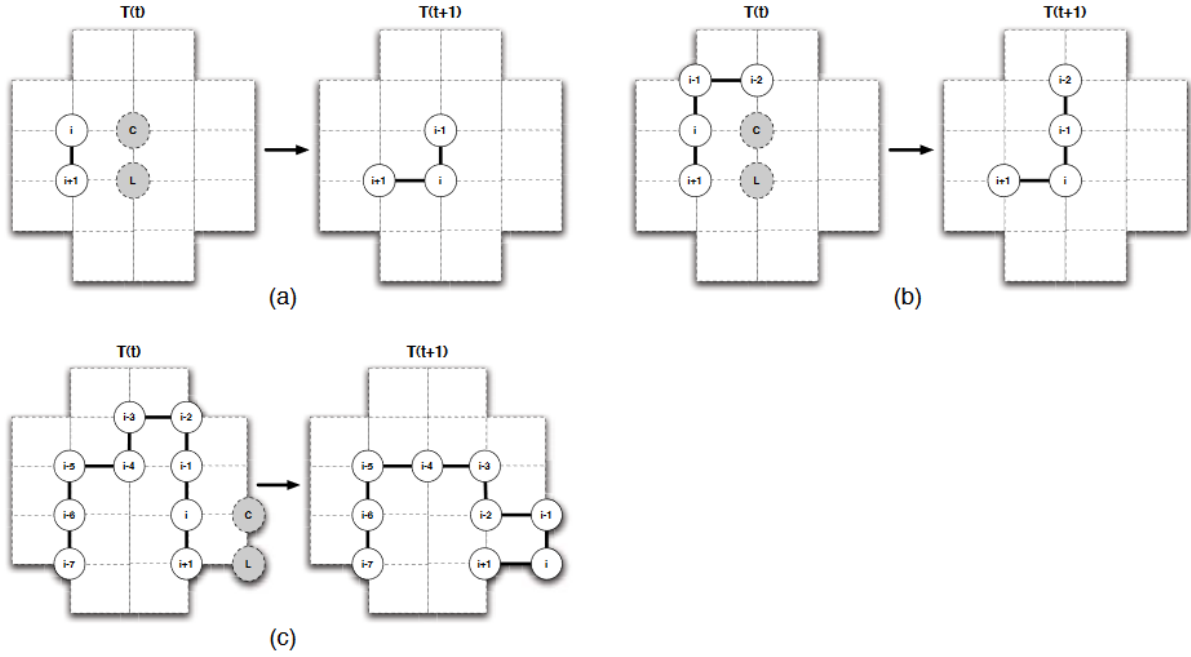


Figure 3: Pull moves. a) Cas où le résidu $\{i-1\}$ occupe C. b) Dans le cas où le résidu $\{i-1\}$ n’occupe pas la position C. c) Lorsque $\{i-2\}$ n’est pas adjacent à C.

Replica Exchange Monte Carlo search

Cet algorithme consiste pour un ensemble n de répliques, dont les températures des répliques sont définies de sorte à ce que les températures des répliques allant de 1, 2, ..., n suivent la propriété que $T_1 < T_2 < \dots < T_n$. L’objectif consiste alors à appliquer l’algorithme du Monte Carlo search sur ces différents répliques, puis d’échanger la température entre répliques adjacents avec une probabilité définie par la formule suivant :

$$\begin{aligned}
 Pr[c \rightarrow c'] &:= Pr[l(c_i) \leftrightarrow l(c_j)] & \Delta &:= (\beta_j - \beta_i)(E(c_i) - E(c_j)). \\
 &:= \begin{cases} 1 & \Delta \leq 0 \\ e^{-\Delta} & \text{otherwise.} \end{cases} & \text{where } \beta_i &= \frac{1}{T_i} \text{ is the inverse of the temperature of replica } i.
 \end{aligned}$$

On applique alors cette échange plusieurs fois au cours de l’algorithme, ce qui permet le passage de conformations “dynamiques” (pour lesquelles la température est élevée) à des conformations plus ou moins “stables” dans le sens où les changements de conformations seront moins permissifs, et vice-versa.

3. Résultats

L'algorithme du REMC a été appliqué avec le VSHD moves, le Pull move et avec les deux (hybrid move). Voici les résultats obtenus avec les deux moves possibles.

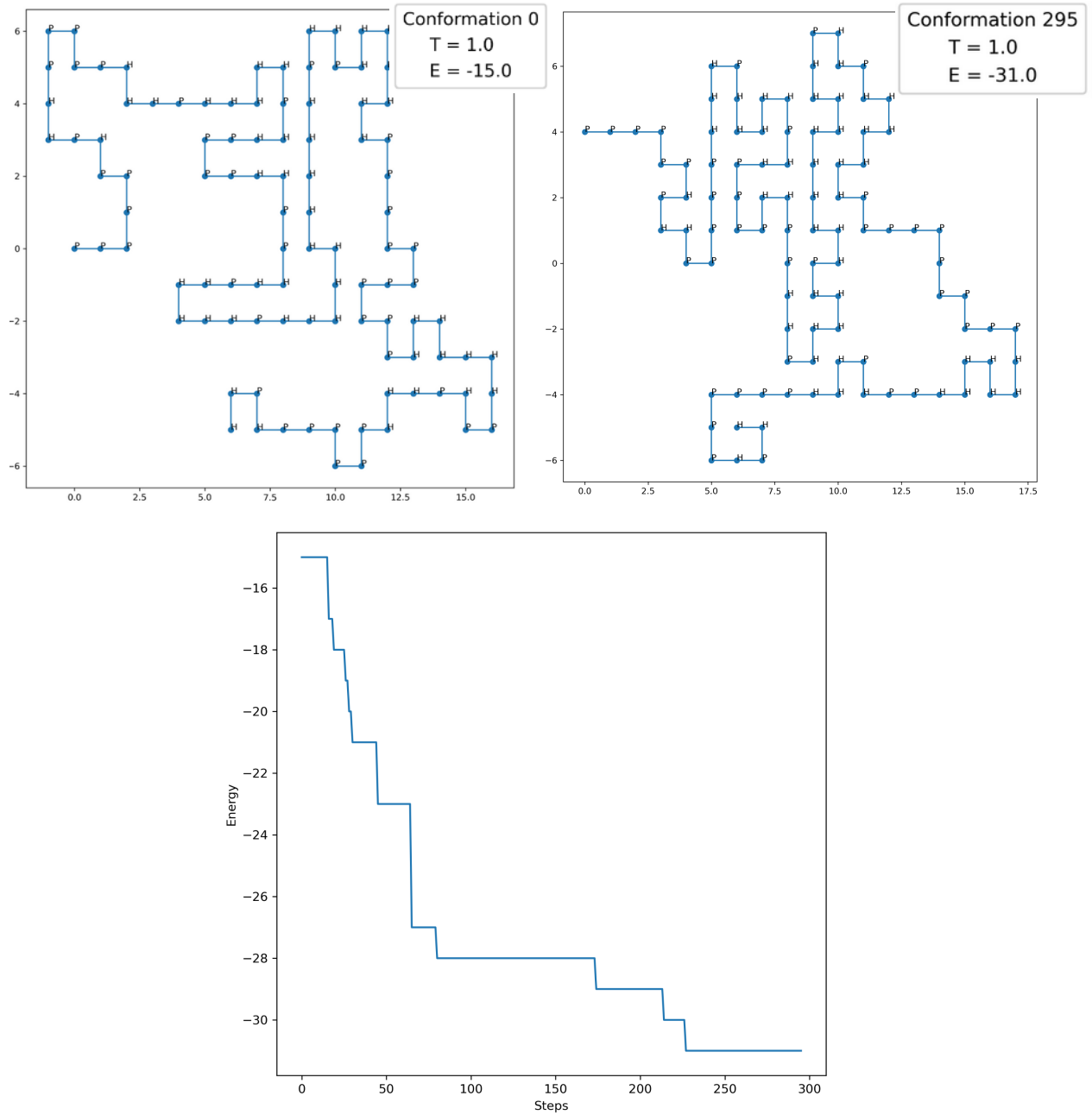


Figure 4: Changement de la conformation de la séquence S1_n100_e48 d'énergie optimale -48, et de taille 100. L'évolution de l'énergie est représentée