Rapport Projet Court :

Conception d’un programme de threading par double programmation dynamique

Auteurs :

Estelle Mariaux : [estelle.mariaux@hotmail.fr](mailto:estelle.mariaux@hotmail.fr)

Théo Ferreira : [theo.ferreira.med@gmail.com](mailto:theo.ferreira.med@gmail.com)

Contact :

Jean-Christophe Gelly : [jean-christophe.gelly@univ-paris-diderot.fr](mailto:jean-christophe.gelly@univ-paris-diderot.fr)

Année Universitaire : 2020 – 2021

Table des matières

Introduction

Def « threading » – état de l’art des autres méthodes alignement point fort point faible (tableau)

Img :

10.1126/science.1065659

A screenshot of a cell phone

Description automatically generated

Modélisation comparée :

Méthode de modélisation de structure protéique

* 1 Homologie, méthode comparative 🡪 recherche d’un homologue par la seq et recherche de la structure associée
  + Recherche d’une séquence homologue dont la structure est déjà connue
  + 2 homologues dérivant d’un même ancêtre ont donc surement une structure très proche
    - Annotation • Similarité de séquence • PSSM • Pro\_il HMM • par synténie

Exemple outils de modélisation : SWISSMODEL • MODELLER • COMPOSER • SEGMOD

Point faible : défaut d’alignement

* 2 Threading (‘enfilage’)
  + Recherche d’une structure ++
  + La structure est bcp plus conservée que la séquence
  + Recherche d’un homologue structural en essayant toutes les structures possibles

compatibilité entre des structures tridimensionnelles et des séquences de protéines linéaires.

Identifier une protéine qui a une structure 3D similaire avec aucune ou très lointaine homologie de séquence • Alignement de séquence à structure • Basé sur l’observation des structures ayant le même repliement avec peu ou pas d’identité de séquence

Two major drawbacks : An exhaustive library of folds • An accurate tool for aligning a sequence on a 3D structure.

Exemple : • 3D-PSSM • PHYRE • RAPTOR • MUSTER • THREADER • iTASSER

* De novo (uniquement par la séquence) ou Ab initio

S’appuie sur les connaissances des propriétés des protéines et de leur structure des protéines pour replier notre protéine

🡪 Same sequence fragments can adopt different structures depending on the environment.

Def programmation dynamique

En [informatique](https://fr.wikipedia.org/wiki/Informatique" \o "Informatique), la **programmation dynamique** est une méthode [algorithmique](https://fr.wikipedia.org/wiki/Algorithme" \o "Algorithme) pour résoudre des [problèmes d'optimisation](https://fr.wikipedia.org/wiki/Optimisation_(math%C3%A9matiques)" \o "Optimisation (mathématiques)). Le concept a été introduit au début des années 1950 par [Richard Bellman](https://fr.wikipedia.org/wiki/Richard_Bellman" \o "Richard Bellman)[1](https://fr.wikipedia.org/wiki/Programmation_dynamique#cite_note-Cormen359-1). À l'époque, le terme « programmation » signifie planification et ordonnancement[1](https://fr.wikipedia.org/wiki/Programmation_dynamique" \l "cite_note-Cormen359-1). La programmation dynamique consiste à résoudre un problème en le décomposant en sous-problèmes, puis à résoudre les sous-problèmes, des plus petits aux plus grands en stockant les résultats intermédiaires

= une méthode d’optimisation des processus de décisions séquentielles. S’appuie sur l’algorithme de Bellman

Le projet était d’implémenter une méthode de prédiction de structure tertiaire d’une protéine basé sur le « threading » par double programmation dynamique d’après le travail de Jones D. (THREADER, 1998).

Matériels & Méthodes

**Environnement informatique**

Programme codé en python 3, le programme est constitué de X fichiers (main.py, …)

Environnement Linux

Python

Env conda spécifique pour améliorer au maximum la reproductibilité

Yml requirements

**Fichiers extérieurs**

Fichiers fasta et pdb

* référence : XXX available at @
* sequence = .fasta
* structure = . pdb

Fichier dope provided by Gelly

**Travail à deux**

1. Mise en commun dans GitHub travail en parallèle
2. Utilisation **Trello** is a web-based Kanban-style list-making application gestionnaire des tache attribution des taches et allocation du temps

**Gestion du code**

Vérification Convention PEP 8

* Pep 8
* Pylint / pycodestyle

Documentation

* Doxygen

Structure du code

Tous les aa sont stockés et associé a un dope score

**Choix de la sequence a aligner** : Fasta 🡪 3 lettres

**Création bank de données**

* Utilisation structure unique pour faciliter code mais pour gain de temps et test mais banque de donnée available at @
* Pdb : parcing pdb plouf pour extraire coordonnées – CA-CA only

**Création matrice de distance pour une structure** : distance entre aa

**Calcul energie en fonction de la distance**

* Dope potentiel energie explication de 0,5 en 0,5 jusqu’à environ 15 … : only CA-CA dict res-res clef renvoyant au 30 enrgies

Matrice de bas niveau : on somme les scores trajet optimal, algorithme de Needleman-Wunsch – programmation dynamique 1

Incorporation matrice de haut niveau – programmation dynamique 2

A screenshot of text

Description automatically generated

**Résultats :**

**Pour une seq de x AA (nom) alignée sur structure (nom) on obtient x matrice avec score pour le chemin max**

**Conclusion :**