

BIM/ARN : TP 2

YANN PONTY

1. TRAVAUX PRATIQUES

On commencera par **terminer le TP1** si celui-ci n'a pas été fait.

1.1. RNASubopts. `RNASubopts` est un programme qui permet la génération de structures sous-optimales comme décrit en cours. On lui fournit une séquence d'ARN ω ainsi qu'une tolérance Δ sur la sous-optimalité, et il renvoie la liste de toutes les structures secondaires pour ω ayant une énergie dans l'intervalle $[mfe, mfe + \Delta]$, où mfe est l'énergie la plus faible sur l'ensemble de Boltzmann.

Codez une fonction `runRNASubopts(ω, Δ)` qui invoque `RNASubopts` avec les arguments correspondants, et renvoie sous la forme d'une liste structure/énergie les structures obtenues.

1.2. RNAForester. Faites de même pour le programme `RNAForester`, aussi décrit en cours et qui compare deux couples séquence/structure d'ARN. La fonction `runRNAForester($\omega_1, S_1, \omega_2, S_2$)` invoquera `RNAForester` avec les paramètres par défaut, et retournera uniquement le score d'alignement.

1.3. Entropie. De nombreuses mesures ont été proposées pour évaluer la stabilité du repliement d'un ARN. On en propose ici une nouvelle, basée sur la distance moyenne dans l'ensemble de Boltzmann. Afin d'approximer cette mesure pour un ARN ω , on procédera comme suit :

- On calculera avec `RNAFold -p` la valeur Z_ω de la fonction de partition.
- On engendrera avec `RNASubopts` l'ensemble S_Δ des structures à $\Delta = 2 \text{ KCal.mol}^{-1}$ de la structure optimale.
- Pour chacune d'entre elles, on calculera le facteur de Boltzmann $B_S = e^{-E/RT}$, puis la probabilité de Boltzmann associée $P_S = B/Z_\omega$.
- On répétera alors M fois la séquence d'opérations suivante, partant d'un ensemble d'échantillons $E = \emptyset$:
 - Tirer au hasard (avec la fonction `python random.sample`) un couple de structures (S_1, S_2) , de poids associé $W_{S_1, S_2} = P_{S_1} \cdot P_{S_2}$.
 - Calculer avec `RNAForester` la distance d'alignement D_{S_1, S_2} .
 - Ajouter le couple $(W_{S_1, S_2}, D_{S_1, S_2})$ à l'ensemble des échantillons E .
- Calculer et renvoyer la distance moyenne pondérée

$$D_\omega^* = \frac{\sum_{(w,d) \in E} w \cdot d}{\sum_{(w,d) \in E} w}$$

1.4. Clustering. Proposez une méthode (heuristique) pour classer un ensemble de structures en clusters, en se basant sur le critère de distance deux à deux utilisé ci-dessus.

E-mail address: `yann.ponty@polytechnique.fr`