L'examen comporte trois exercices devant être traités sur des feuilles séparées.
Les notes de cours et transparents de CASM sont autorisés.
Le barème n'est fourni qu'à titre **indicatif**, i.e. sujet à modifications.

Problème 1 : Réarrangements chromosomiques

1. [2 pt] Définir précisément le tri par reversion.

2. [4 pt] Trier par reversion la suite `10 6 5 3 4 1 9 2 7 8` en détaillant la suite des opérations effectuées.

3. [3 pt] Quelle structure de données suggérez-vous pour effectuer efficacement ces opérations ?

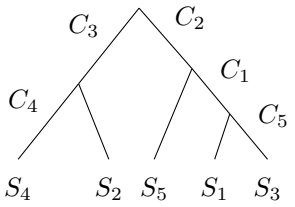
Problème 2 : Phylogénies parfaites

On considère dans ce problème un modèle idéalisé pour la construction d'arbres phylogénétiques ; Les génomes sont des séquences de $\{0,1\}^m$, donc de longueur m ; l'ancêtre commun a la séquence 0^m et au cours de l'évolution, si une position (SNP) passe de 0 à 1, elle me pourra plus jamais repasser à 0. Chaque position de la séquence est appelée un **caractère**.
Un ensemble de n génomes est donc représenté par une matrice M de dimensions $n \times m$ dont les colonnes représentent les caractères. Une phylogénies *parfaite* est un arbre enraciné T **binaire** à n feuilles étiquetées chacune par un génome et dont m arêtes sont étiquetées par un caractère unique.

Exemple : $m = 5$, $n = 5$

$$\begin{array}{rcl} S_1 & = & \\ S_2 & = & \\ S_3 & = & \\ S_4 & = & \\ S_5 & = & \end{array} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} = M$$
$$\begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \end{array}$$

$T =$



4. [2 pt] Expliquer les caractéristiques de l'arbre T de l'exemple ci-dessus à partir de propriétés des séquences.

5. [2 pt] Montrer qu'on peut l'obtenir simplement en triant les colonnes par ordre lexicographique. Quelle est la complexité de l'algorithme ?

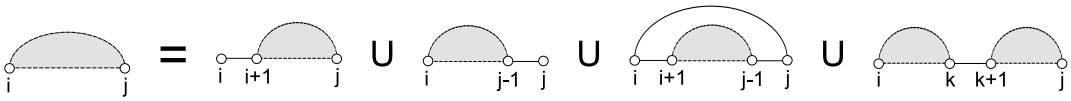
6. [2 pt] Montrer que si on ajoute 1 à S_2 et S_5 en C_5, cette construction n'est plus possible.

7. [2 pt] Proposer un arbre phylogénétique pour les séquences de la question 6. Donner une méthode générale.

8. [2 pt] Expliciter les critères selon lesquels les arbres des questions 5 et 7 peuvent être appelés optimaux. À quelle condition peuvent-ils représenter le temps entre les mutations ?

Problème 3 : Qualité d'une décomposition

La formulation historique de l'algorithme de Nussinov repose sur la décomposition :



On se place dans le modèle d'énergie de Nussinov le plus simple, où l'énergie d'une structure ayant k paires de bases est de $-k$ KCal.mol^{-1}. En d'autres termes, chaque paire de base contribuera pour -1 KCal.mol^{-1} à l'énergie libre. On autorise l'appariement des bases complémentaires Watson-Crick (A/U, G/C) sauf celles directement consécutives $(i, i+1)$. Ceci correspond au cas $\theta = 1$ pour l'algorithme vu en cours.
On disposera une **fonction de test** ψ, qui prend une séquence d'ARN r et deux positions (i, j) en entrée, et renvoie 1 si l'appariement des bases r_i et r_j est possible, et 0 sinon.

9. [3 pt] Écrire la récurrence à laquelle obéit l'énergie minimale $E(i, j)$ d'une structure repliée sur le sous-intervalle $[i, j]$ d'une séquence d'ARN r.

10. [1 pt] Transformer *naïvement* cette équation de minimisation en une équation de récurrence sur le nombre $\mathcal{N}(i, j)$ de structures secondaires compatibles avec r sur l'intervalle $[i, j]$.

11. [2 pt] Utiliser l'équation obtenue pour compter les structures secondaires compatibles avec la séquence $r :=$ `GGACC`.

On rappelle que, dans la décomposition vue en cours et admise non-ambiguë, le nombre de structures compatibles obéit à

$$\begin{aligned}
\mathcal{M}(i,i) &= \mathcal{M}(i,i+1) = \mathcal{M}(i,i-1) = 1, \quad \forall i \in [1,n] \\
\mathcal{M}(i,j) &= \mathcal{M}(i+1,j) + \sum_{k=i+2}^{j} \psi(r,i,j) \cdot \mathcal{M}(i+1,k-1) \cdot \mathcal{M}(k+1,j), \forall 1 \leq i < j \leq n
\end{aligned}$$

12. [2 pt] Calculer le nombre de structures compatibles avec $r :=$ `GGACC` selon cette nouvelle formule.

13. [1 pt] Qu'en déduisez vous sur la décomposition historique de Nussinov ? Prouver votre affirmation.