

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.

Exercice : Enumeration de mots

Soit $\mathcal{H} = \{acatacataca, caggttggcc, cattggcacc\}$.

- 2 pt Dessiner le graphe de recouvrement *OvGraph*.
- 2 pt Ecrire le polynôme d'autocorrélation $A_{H_1, H_1}(z)$ et le polynôme de corrélation $A_{H_2, H_3}(z)$.

Problème : Interactions ARN/ARN

On souhaite prédire l'interaction de deux ARN de séquences α et β . Pour cela, on étend les structures secondaires en un modèle de *conformations jointes*, assimilées à des triplets $(S_\alpha, S_\beta, S_{\alpha+\beta})$ tels que :

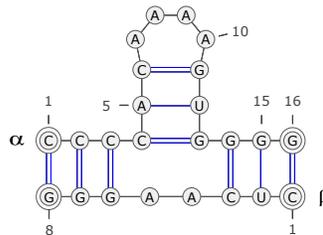
- S_α est l'ensemble des paire de bases, ou **liaisons internes** à α .
- S_β est l'ensemble des paire de bases, ou **liaisons internes** à β .
- $S_{\alpha+\beta}$ est l'ensemble des paire de bases reliant α et β , ou **liaisons externes**.

L'énergie d'une conformation $S = (S_\alpha, S_\beta, S_{\alpha+\beta})$ est alors simplement donnée par :

$$E_{\alpha, \beta, S} = \sum_{(a,b) \in S_\alpha} E_{a,b} + \sum_{(a,b) \in S_\beta} E_{a,b} + \sum_{(a,b) \in S_{\alpha+\beta}} E'_{a,b}$$

avec $E_{G,C} = E_{C,G} = E'_{G,C} = E'_{C,G} = -3$, $E_{A,U} = E_{U,A} = E'_{A,U} = E'_{U,A} = -2$ et $E_{G,U} = E_{U,G} = E'_{G,U} = E'_{U,G} = -1$ (kcal.mol⁻¹).

- 1 pt Donner la conformation jointe $S = (S_\alpha, S_\beta, S_{\alpha+\beta})$ et l'énergie libre $E_{\alpha, \beta, S}$ correspondant à la figure ci-dessous :



Solution:

$$S_\alpha = \{(4, 13), (5, 12), (6, 11)\}$$

$$S_\beta = \emptyset$$

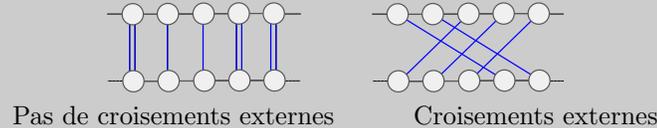
$$S_{\alpha+\beta} = \{(1, 8), (2, 7), (3, 6), (14, 3), (15, 2), (16, 1)\}$$

$$E_{\alpha, \beta, S} = -8[\alpha] + 0[\beta] + -16[\alpha + \beta] = -24 \text{ kcal.mol}^{-1}$$

Interactions sans repliement : On se place tout d'abord dans un modèle très simple de conformations n'autorisant que des paires de base entre α et β ($\Rightarrow S_\alpha = S_\beta = \emptyset$). Par ailleurs, on interdit les **croisements externes**, c'est à dire des paires (i, j) et (i', j') telles que $i < i'$ et $j < j'$.

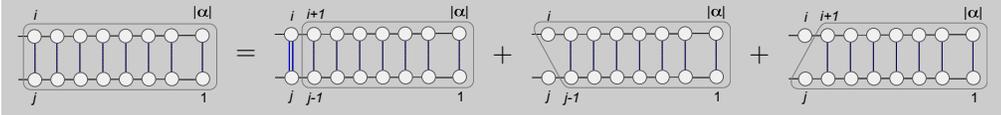
4. 1 pt Dessiner deux interactions sans repliement, l'une présentant des croisements externes et l'autre non.

Solution:



5. 2 pt Proposer un schéma de décomposition (faire un dessin) de l'ensemble des conformations sans repliement. Votre décomposition devra impérativement être complète, c'est à dire engendrer **toutes** les conformations sans repliement. *S'inspirer de la décomposition sous-jacente aux algorithmes Smith-Waterman et/ou Needleman-Wunsch vus (et revus) en cours.*

Solution:



6. 1 pt Adapter ce schéma de décomposition en une équation calculant l'énergie de la conformation sans repliement d'énergie libre la plus faible. On dénotera par α_i (resp. β_j) la i -ième base dans α (resp. la j -ième base dans β).

Solution:

$$R(|\alpha| + 1, 0) = 0$$

$$R(i, j) = \min \begin{cases} E'(\alpha_i, \beta_j) + R(i + 1, j - 1) & \text{Si } i \leq |\alpha| \text{ et } j \geq 1 \\ R(i, j - 1) & \text{Si } j \geq 1 \\ R(i + 1, j) & \text{Si } i \leq |\alpha|. \end{cases}$$

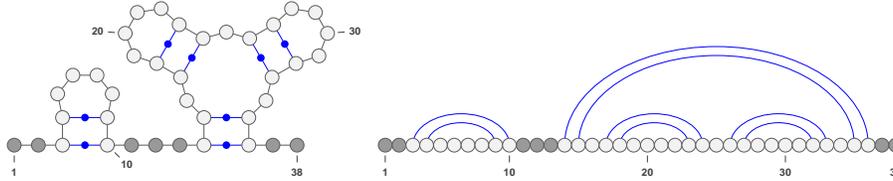
7. 1 pt Donner la complexité de cet algorithme en fonction des tailles respectives n et m des séquences α et β .

Solution: Le nombre de sous-problèmes dans la décomposition est en $\Theta(m \times n)$, donc la complexité en mémoire est en $\Theta(m \times n)$. Chaque score peut être calculé en un nombre constant d'opérations, donc la complexité générale en temps est en $\Theta(m \times n)$.

8. 1 pt Votre décomposition est-elle ambiguë? Si oui donner un exemple. Si non, donner un argument.

Solution: La décomposition proposée dans la correction est ambiguë. En effet, toute conformation mettant *face à face* deux positions non-appariées i (dans α) et j (dans β) peut être identiquement obtenue en appliquant les règles 2) puis 3), ou 3) puis 2) dans la décomposition.

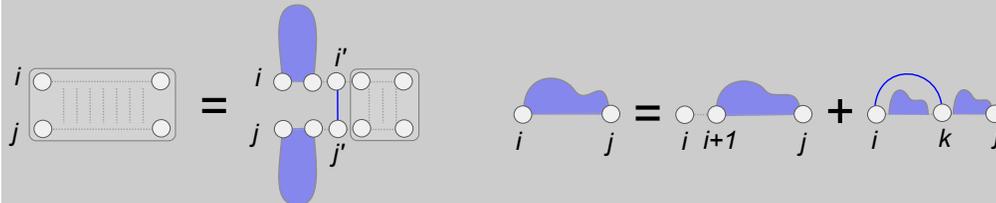
Au sein d'une structure secondaire, les **positions extérieures** sont les bases qui ne sont ni appariées, ni *survolées* par un appariement (voir-dessin ci-dessous). Par exemple, l'ARN ci-dessous présente des positions extérieures en 1, 2, 11, 12, 13, 37 et 38 (en gris foncé).



Interactions extérieures : On enrichit l'espace de conformations précédent en autorisant un repliement pour chacun des deux ARN. Chacun des deux ARN est autorisé à adopter n'importe quelle structure secondaire, mais seules les *positions extérieures* peuvent être impliquées dans des liaisons – sans croisement – externes (entre α et β).

9. 2 pt On remarque que la restriction à des liaisons externes entre positions extérieures rend indépendants les repliements de α et β .
En déduire une décomposition de l'ensemble des interactions extérieures.
Se baser sur la décomposition précédemment obtenue pour les interactions sans repliement, combinée avec la décomposition de Nussinov vue en cours.

Solution:



Remarque : Cette décomposition induit des algorithmes en $\Theta(n^2 \times m^2)$, une version en $\Theta(n \times m \times (n + m))$ peut être obtenue en séparant en deux étapes l'adjonction des structures secondaires en haut et en bas.

10. 1 pt En déduire immédiatement* les équations de récurrences permettant le calcul de la fonction de partition

$$\mathcal{Z}_{\alpha,\beta} = \sum_{S \in \mathcal{S}_{\alpha,\beta}} e^{-\frac{E_{\alpha,\beta,S}}{RT}},$$

où $\mathcal{S}_{\alpha,\beta}$ est l'ensemble des interactions extérieures.

* Vérifier au préalable la non-ambiguïté de votre décomposition !

Solution:

$$\mathcal{X}(|\alpha| + 1, 0) = 1$$

$$\mathcal{X}(i, j) = \sum_{\substack{i \leq i' \leq |\alpha| \\ j \geq j' \geq 1}} \left(e^{-\frac{E'(\alpha_{i'}, \beta_{j'})}{RT}} \cdot \mathcal{N}^\alpha(i, i' - 1) \cdot \mathcal{N}^\beta(j' + 1, j) \cdot \mathcal{X}(i' + 1, j' - 1) \right)$$

$$\mathcal{N}^\alpha(i, i - 1) = 1, \forall i \in [1, |\alpha|]$$

$$\mathcal{N}^\alpha(i, j) = \mathcal{N}^\alpha(i + 1, j) + \sum_{k \in [i+\theta+1, j]} e^{-\frac{E(\alpha_i, \alpha_k)}{RT}} \cdot \mathcal{N}^\alpha(i + 1, k - 1) \cdot \mathcal{N}^\alpha(k + 1, j)$$

$$\mathcal{N}^\beta(i, i - 1) = 1, \forall i \in [1, |\beta|]$$

$$\mathcal{N}^\beta(i, j) = \mathcal{N}^\beta(i + 1, j) + \sum_{k \in [i+\theta+1, j]} e^{-\frac{E(\beta_i, \beta_k)}{RT}} \cdot \mathcal{N}^\beta(i + 1, k - 1) \cdot \mathcal{N}^\beta(k + 1, j)$$

$$\mathcal{Z}_{\alpha, \beta} = \mathcal{X}(1, |\beta|)$$

11. 1 pt En supposant qu'on est à l'équilibre thermodynamique, c'est à dire que chaque conformation jointe S est observée avec probabilité de Boltzmann

$$p_{\alpha, \beta, S} = \frac{e^{-\frac{E_{\alpha, \beta, S}}{RT}}}{\mathcal{Z}_{\alpha, \beta}},$$

donner la probabilité d'interaction de α et β , c'est à dire la probabilité d'observer au moins une liaison externe (paire de base entre α et β). Montrer comment calculer cette probabilité (si possible en temps constant) en fonction des termes calculés à la question précédente.

Solution: L'ensemble des structures sans liaison externe est simplement le produit cartésien des structures secondaires. En d'autres termes, soient \mathcal{S}_α et \mathcal{S}_β les ensembles de structures secondaires compatibles avec des séquences α et β , l'ensemble des structure sans liaison externe est simplement l'ensemble des triplets (a, b, \emptyset) , où $a \in \mathcal{S}_\alpha$ est n'importe lequel des repliements de α , et $b \in \mathcal{S}_\beta$ est n'importe lequel des repliements de β . La fonction de partition restreinte à ces conformations est alors donnée par

$$\begin{aligned} \sum_{(a, b) \in \mathcal{S}_\alpha \times \mathcal{S}_\beta} e^{-\frac{E_{\alpha, \beta, (a, b, \emptyset)}}{RT}} &= \sum_{(a, b) \in \mathcal{S}_\alpha \times \mathcal{S}_\beta} e^{-\frac{E_{\alpha, a} - E_{\beta, b}}{RT}} \\ &= \sum_{(a, b) \in \mathcal{S}_\alpha \times \mathcal{S}_\beta} e^{-\frac{E_{\alpha, a}}{RT}} \times e^{-\frac{E_{\beta, b}}{RT}} \\ &= \left(\sum_{a \in \mathcal{S}_\alpha} e^{-\frac{E_{\alpha, a}}{RT}} \right) \times \left(\sum_{b \in \mathcal{S}_\beta} e^{-\frac{E_{\beta, b}}{RT}} \right) \\ &= \mathcal{N}^\alpha(1, |\alpha|) \times \mathcal{N}^\beta(1, |\beta|). \end{aligned}$$

Il suffit alors de diviser cette quantité par la fonction de partition de l'ensemble des conformations $\mathcal{Z}_{\alpha, \beta}$ pour obtenir la probabilité d'absence d'interaction. La probabilité

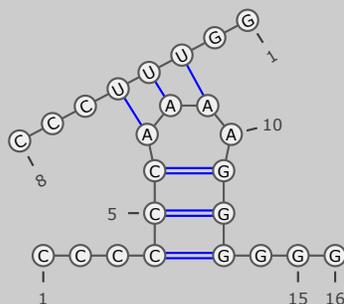
de former au moins une liaison externe est alors obtenue en complétant à 1 :

$$\Rightarrow p \text{ liais. ext.} = 1 - \frac{\mathcal{N}^\alpha(1, |\alpha|) \times \mathcal{N}^\beta(1, |\beta|)}{\mathcal{Z}_{\alpha, \beta}}.$$

Un peu plus loin ...

12. 1 pt Donner un exemple (minimal) d'interaction – sans croisement externe – exclu de la classe des interactions extérieures.

Solution:



13. 2 pt (Bonus) Proposez une décomposition parcourant toutes les interactions extérieures, enrichies d'autres interactions sans croisement. En donner la complexité.

Solution: On aurait pu proposer un algorithme basé sur l'exploration des pseudo-nœuds (algorithme d'Akutsu vu en cours), ce qui donnerait un algorithme en $\Theta(n^3 \times m^2)$.