

Repléments d'ARN

Épreuve pratique d'algorithmique et de programmation

Concours commun des écoles normales supérieures

Durée de l'épreuve: 3 heures 30 minutes

Juin Juillet 2012

ATTENTION !

N'oubliez en aucun cas de recopier votre u_0
à l'emplacement prévu sur votre fiche réponse

Important.

Sur votre table est indiqué un numéro u_0 qui servira d'entrée à vos programmes. Les réponses attendues sont généralement courtes et doivent être données sur la fiche réponse fournie à la fin du sujet. À la fin du sujet, vous trouverez en fait deux fiches réponses. La première est un exemple des réponses attendues pour un \tilde{u}_0 particulier (précisé sur cette même fiche et que nous notons avec un tilde pour éviter toute confusion!). Cette fiche est destinée à vous aider à vérifier le résultat de vos programmes en les testant avec \tilde{u}_0 au lieu de u_0 . Vous indiquerez vos réponses (correspondant à votre u_0) sur la seconde et vous la remettrez à l'examineur à la fin de l'épreuve.

En ce qui concerne la partie orale de l'examen, lorsque la description d'un algorithme est demandée, vous devez présenter son fonctionnement de façon schématique, courte et précise. Vous ne devez en aucun cas recopier le code de vos procédures !

Quand on demande la complexité en temps ou en mémoire d'un algorithme en fonction d'un paramètre n , on demande l'ordre de grandeur en fonction du paramètre, par exemple: $O(n^2)$, $O(n \log n)$,...

Il est recommandé de commencer par lancer vos programmes sur de petites valeurs des paramètres et de *tester vos programmes sur des petits exemples que vous aurez résolus préalablement à la main ou bien à l'aide de la fiche réponse type fournie en annexe*. Enfin, il est recommandé de lire l'intégralité du sujet avant de commencer afin d'effectuer les bons choix de structures de données dès le début.

1 Génération de brins

L'acide ribonucléique (ARN) est une molécule que l'on retrouve dans la grande majorité des organismes vivants. Elle est formée d'une séquence de nucléotides. Chaque nucléotide porte une base azotée choisie parmi les quatre suivantes : adénine (A), guanine (G), cytosine (C) et uracile (U). Un brin d'ARN sera décrit par la suite des bases qui le composent.

Considérons la suite d'entiers $(u_k)_{k \in \mathbb{N}}$ définie par

$$u_{k+1} = 15091 \times u_k \pmod{64007}.$$

Le brin B_ℓ de longueur ℓ est généré aléatoirement en calculant les valeurs v_1, \dots, v_ℓ de la suite $(v_k)_k$ définie par $v_k = \lfloor u_k / 16002 \rfloor$. Les entiers entre 0 et 3 s'interprètent en bases de la façon suivante :

$$0 \mapsto A \quad 1 \mapsto C \quad 2 \mapsto G \quad 3 \mapsto U.$$

Question 1 Quelles sont les trois dernières bases du brin **a)** B_{20} ? **b)** B_{400} ? **c)** B_{10000} ?

2 Appariements de bases

Laissé libre, un brin d'ARN a tendance à se replier sur lui-même, les bases complémentaires s'attirant les unes les autres. On parle alors d'appariement entre deux bases. Dans le modèle de Watson-Crick, les seuls appariements possibles sont A avec U et G avec C.

Question 2 Combien d'appariements au maximum peuvent se former simultanément en repliant le brin **a)** B_{20} ? **b)** B_{400} ? **c)** B_{10000} ?

On dit qu'il y a un empilement de taille n entre les positions p et q (avec $p < q$) si v_{p-i} est apparié à v_{q+i} pour tout i tel que $0 \leq i < n$.

Question 3 Quelle est la longueur du plus grand empilement qu'il est possible de créer en repliant le brin **a)** B_{20} ? **b)** B_{400} ? **c)** B_{10000} ?

Question à développer pendant l'oral : Quelle est la complexité en temps de votre algorithme ?

3 Structure secondaire

Lors de son repliement, un brin d'ARN cherche à maximiser le nombre d'appariements tout en respectant des contraintes imposées par les lois physiques. Afin de modéliser les contraintes géométriques, l'hypothèse suivante est faite : les appariements ne peuvent pas se croiser. Formellement, si (v_i, v_j) et (v_p, v_q) sont deux appariements et si $i < p < j$, alors $i < q < j$. On appelle structure secondaire un repliement qui respecte cette règle (par opposition à la structure primaire qui est simplement la séquence des bases du brin). On peut représenter l'ensemble des appariements d'un brin replié par une suite $(r_i)_{1 \leq i \leq n}$ définie par

$$r_i = \begin{cases} j - i & \text{si } i \leq j \text{ et } v_i \text{ et } v_j \text{ sont appariées} \\ 0 & \text{sinon} \end{cases}$$

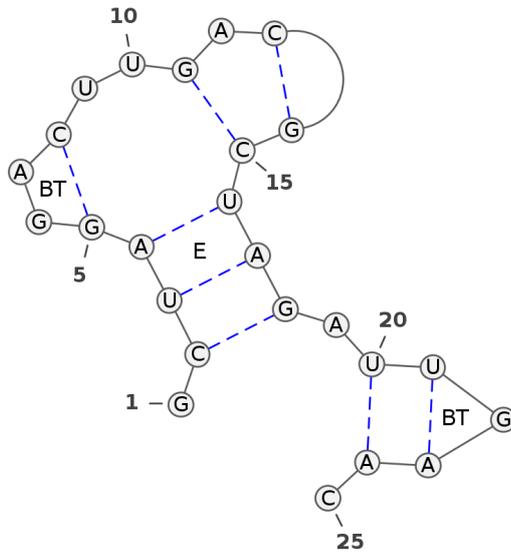


FIGURE 1 – Brin de 25 bases replié pour former 8 appariements sans croisement. Il contient 3 bases à l’intérieur de boucles terminales (BT) et son plus grand empilement (E) à une taille de 3.

On ordonne les différents repliements d’un brin à l’aide de l’ordre lexicographique : (r_i) est strictement plus petit que (r'_i) s’il existe j tel que $r_i = r'_i$ pour $1 \leq i < j$ et $r_j < r'_j$.

Question 4 Combien y a-t-il d’appariements dans le repliement le plus grand par ordre lexicographique du brin **a)** B_{20} ? **b)** B_{400} ? **c)** B_{10000} ?

Question à développer pendant l’oral : Quelle est la complexité en temps de votre algorithme ?

On dit qu’une séquence de bases $v_i, v_{i+1} \dots v_j$ forme une boucle terminale si aucune de ces bases n’est appariée et si les bases v_{i-1} et v_{j+1} sont appariées entre elles.

Question 5 Combien y a-t-il de bases dans l’ensemble des boucles terminales du repliement le plus grand par ordre lexicographique du brin **a)** B_{20} ? **b)** B_{400} ? **c)** B_{10000} ?

4 Énergie libre

Le modèle de Watson-Crick est trop simple pour modéliser les repliements. En effet, des appariement G–U peuvent se produire en pratique, même s’ils sont moins probables que les autres. Le modèle de Nussinov-Jacobson prend cela en compte en associant à un brin replié une énergie libre dont la valeur dépend des appariements. Plus précisément,

l'énergie libre d'un repliement est définie par

$$-3n_{C-G} - 2n_{A-U} - n_{G-U}$$

avec n_{X-Y} le nombre d'appariements entre les bases X et Y . Dorénavant, les repliements autoriseront donc les appariements C-G, A-U et G-U.

Question 6 *Quelle est l'énergie libre du repliement le plus grand par ordre lexicographique du brin* **a)** B_{20} ? **b)** B_{400} ? **c)** B_{10000} ?

Lors du repliement, un brin cherche à minimiser son énergie libre, tout en respectant la contrainte selon laquelle les appariements ne peuvent pas se croiser. Pour un brin donné, on notera E_m l'énergie libre correspondant au repliement d'énergie minimale. On notera $E_m(i, j)$ l'énergie minimale qu'il est possible d'atteindre en ne repliant que la sous-séquence $v_i, v_{i+1} \dots v_j$. Autrement dit, $E_m = E_m(1, n)$ pour un brin de longueur n .

Question à développer pendant l'oral : Expliquez comment exprimer $E_m(i, j)$ en fonction des valeurs $E_m(i', j')$ pour $[i', j'] \subset [i, j]$.

Question 7 *Que vaut E_m pour le brin* **a)** B_{20} ? **b)** B_{150} ? **c)** B_{1000} ?

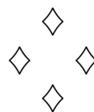
Question à développer pendant l'oral : Quelle est la complexité en temps et en espace de votre algorithme ?

Pour résoudre les questions suivantes, on pourra s'inspirer de l'approche utilisée pour l'algorithme précédent.

Question 8 *Combien y a-t-il de repliements différents d'énergie libre minimale pour le brin* **a)** B_{10} ? **b)** B_{20} ? **c)** B_{50} ?

Même si les brins ont tendance à adopter le repliement qui minimise l'énergie libre, cela reste un phénomène statistique : tous les autres repliements existent naturellement, ils sont juste moins répandus que ceux d'énergie minimale. La probabilité de voir apparaître un repliement d'énergie $E \geq E_m$ pour un brin B est donnée par le quotient $\exp(-0.6 \times E)/S$ avec $S = \sum_R \exp(-0.6 \times E_R)$ la somme des exponentielles de l'énergie E_R de chaque repliement R possible de B .

Question 9 *Quelle est la probabilité d'obtenir un repliement d'énergie minimale pour le brin* **a)** B_{10} ? **b)** B_{20} ? **c)** B_{50} ? *Les valeurs seront données avec 4 chiffres significatifs.*



Fiche réponse type: Repléments d'ARN

$\widetilde{u}_0 : 3$

Question 1

a)

b)

c)

Question 2

a)

b)

c)

Question 3

a)

b)

c)

Question 4

a)

b)

c)

Question 5

a)

b)

c)

Question 6

a)

b)

c)

Question 7

a)

b)

c)

Question 8

a)

b)

c)

Question 9

a)

b)

c)



Fiche réponse: Repléments d'ARN

Nom, prénom, u₀:

Question 1

a)

b)

c)

Question 2

a)

b)

c)

Question 3

a)

b)

c)

Question 4

a)

b)

c)

Question 5

a)

b)

c)

Question 6

a)

b)

c)

Question 7

a)

b)

c)

Question 8

a)

b)

c)

Question 9

a)

b)

c)



