

SESSION 2005

Filière : 2^{ème} concours

BIOLOGIE - BIOCHIMIE

(Epreuve commune aux ENS Ulm et Lyon)

Durée : 3 heures

L'usage des calculatrices électroniques de poche à alimentation autonome, sans imprimante et sans document d'accompagnement est autorisé. Cependant, une seule calculatrice à la fois est admise sur la table ou le poste de travail. Aucun échange n'est permis entre les candidats.

Tournez la page S.V.P.

Note : les problèmes 1 et 2 peuvent être traités de manière indépendante. Le nombre de points attribués à chaque question est donné à titre indicatif. Ce barème est susceptible d'être ajusté à la correction.

Problème 1
(5 points)

1 – (0,75 point)

Bengkala est un village reculé de Bali dans lequel 2% de la population présente une surdité (47 individus sur 2185 en 1994). Du fait de son isolement géographique, un degré élevé de consanguinité (c'est à dire d'endogamie) est observé au sein de la population de ce village. La généalogie d'un groupe familial est présentée sur la Figure 1.

Ce syndrome de surdité est-il monogénique ou polygénique ?

La(es) mutation(s) responsable(s) de ce syndrome est(sont)-elle(s) récessive(s), dominante(s) ou semi-dominante(s) ?

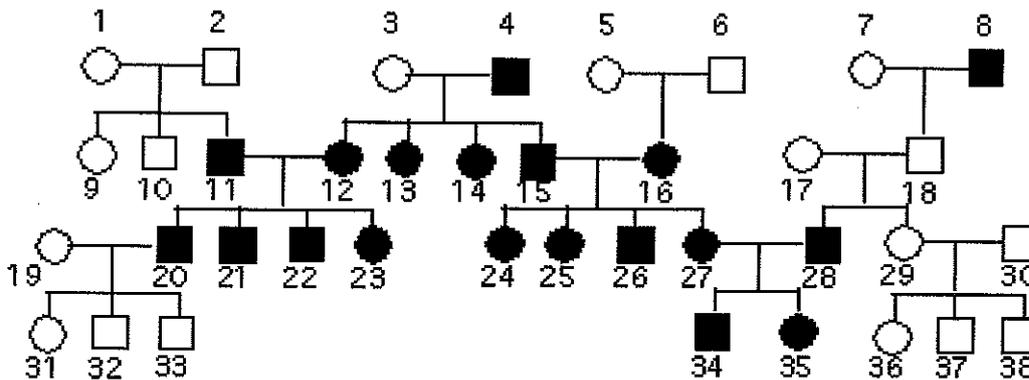


Figure 1

Généalogie d'un groupe familial. Chaque individu répertorié dans cette étude est numéroté. Le genre (femme : cercle ; homme : carré) et le phénotype (blanc : aucun défaut auditif ; noir : malentendant) sont indiqués.

2 – (0,75 point)

Au sein de la population mondiale, la quasi-totalité des croisements inter-ethniques entre deux personnes malentendantes conduit à une descendance qui ne présente pas de surdité. Quelles hypothèses pouvez-vous faire pour expliquer cette observation ? Quel est l'intérêt d'étudier une population isolée au sein de laquelle les croisements intra-ethniques entre deux personnes malentendantes conduit à une descendance présentant une surdité.

3 – (0,75 point)

Proposez une stratégie expérimentale pour identifier le(s) chromosome(s) porteur(s) du(es) gène(s) impliqué(s) dans ce syndrome congénital.

4 – (0,75 point)

Le gène (ou un des gènes) impliqué(s) dans ce syndrome congénital a été cartographié dans une région péricentromérique du chromosome 17 d'une taille de 12 centimorgans. Cette région chromosomique correspond à la région du chromosome 11 de la souris : ces deux régions sont dites synténiques car elles codent des gènes homologues disposés selon le même

ordre et peuvent donc être alignées l'une avec l'autre. La mutation Shaker2 de souris, qui est associée à une surdité congénitale et à des défauts morphologiques des cellules de l'oreille interne, a été cartographiée dans cette région du chromosome 11.

Quelle hypothèse simple pouvez-vous faire sur la base de la synténie entre ces deux régions ? Discutez la validité de votre hypothèse en estimant le nombre de gènes codés par cette région de 12 centimorgans (cM) du chromosome 17?

Pour faire cette estimation, vous ferez l'hypothèse que le génome humain code 40 000 gènes pour une taille de $3,2 \cdot 10^9$ paires de bases. Le chromosome 17 représente 3% de l'ADN génomique. Enfin, une distance génétique de 1 cM couvre en moyenne 10^6 paires de bases.

5 – (0,75 point)

Sachant que le gène Shaker2 a été cloné chez la souris, proposez une stratégie expérimentale pour identifier le(un des) gène(s) impliqué(s) dans le syndrome congénital observé dans la population de Bengkala.

6 – (1,25 point)

Le gène Shaker2 code une myosine non-musculaire, la myosine XVa. Cette myosine est présente dans les cellules ciliées de l'oreille interne qui sont directement impliquées dans la réception des signaux auditifs. Ces cellules spécialisées portent à leur surface apicale de longues microvillosités appelées cils. Ces cils apicaux contiennent de nombreux microfilaments qui sont responsables de l'organisation structurale de ces cils et qui jouent un rôle essentiel dans le transport intracellulaire de différentes protéines synthétisées dans le cytoplasme et transportées vers l'extrémité apicale des cils. A l'aide d'anticorps qui reconnaissent de manière spécifique la myosine XVa, il a été montré que cette myosine est accumulée dans la partie la plus apicale des cils.

Sur la base de cette localisation subcellulaire spécifique, quelle(s) hypothèse(s) pouvez-vous faire quand au rôle de cette myosine dans la physiologie des cellules ciliées ?

Proposez une expérience pour tester votre hypothèse sur le rôle de cette myosine ?

Problème 2

(15 points)

Note préliminaire:

Les parties 1 et 2 du problème 2 peuvent être traitées de manière indépendante.

Introduction

La transcription du rétrovirus HIV-1 conduit à la synthèse d'un ARNm primaire (ou transcrit primaire). Lors de la phase précoce d'infection virale, cet ARNm primaire est épissé pour produire un ARNm mature codant trois protéines régulatrices, dont la protéine Rev (Figure 1 ; par simplicité, seule la protéine Rev est représentée dans la figure 1). Lors d'une phase plus tardive, l'épissage de l'ARNm primaire est inhibé et un ARNm partiellement épissé est exporté du noyau. La traduction de cet ARNm partiellement épissé conduit à la synthèse des protéines de l'enveloppe du virus (Figure 1). La transition entre les phases d'infection précoce et tardive repose donc sur l'inhibition de l'épissage de l'ARNm primaire. Cette inhibition dépend en partie de la protéine Rev.

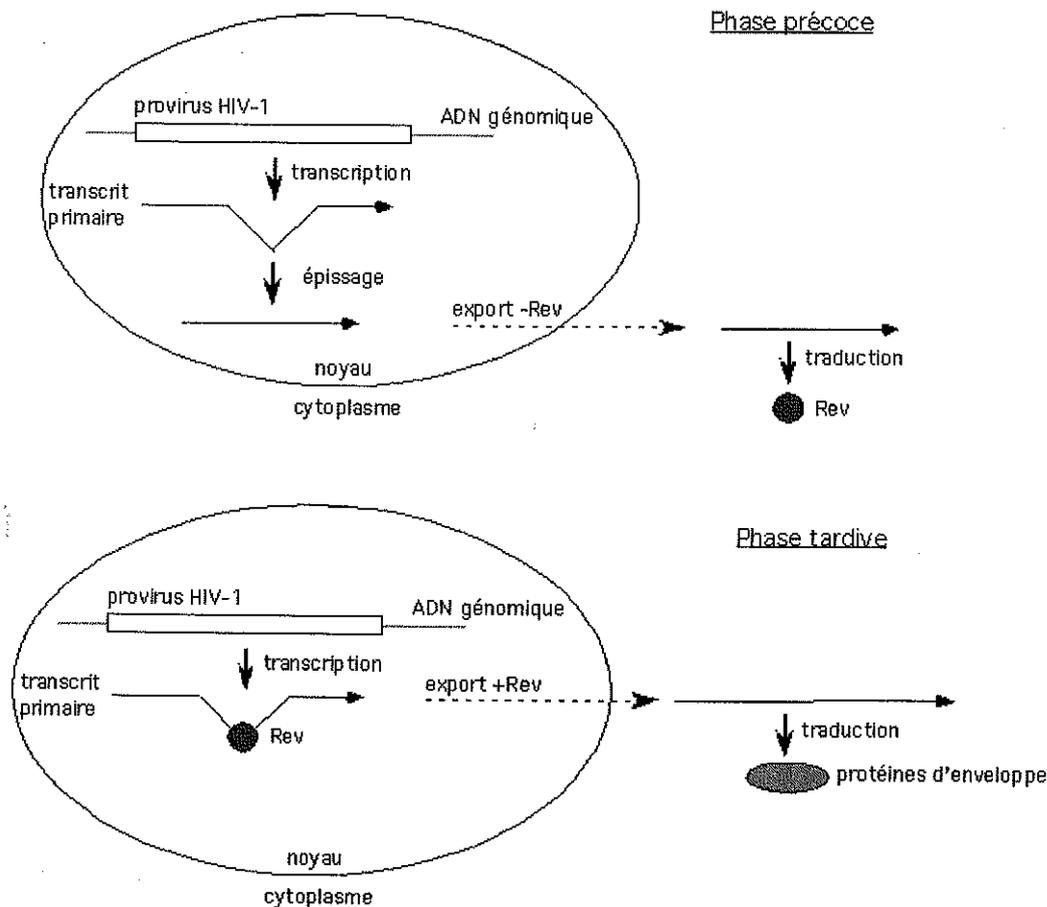


Figure 1

Représentation schématique de la régulation de l'inhibition de l'épissage de l'ARNm primaire du provirus HIV-1 par la protéine Rev

Rev est une protéine de 116 acides aminés se liant à l'ARN. C'est une « RNA-binding protein ». Rev se fixe de manière spécifique à une région intronique de l'ARNm primaire codé par HIV-1. La région intronique de l'ARNm HIV-1 qui est requise pour la fixation de Rev est appelée RRE (pour « Rev Responsive Element »). La taille de ce RRE est de 234 nucléotides. Six molécules Rev se fixent de manière coopérative à une copie du RRE. La fixation de Rev à son RRE inhibe l'épissage de l'ARNm HIV-1. Cette inhibition conduit à l'export nucléaire d'un ARNm codant pour les protéines d'enveloppe. Dans ce problème, nous étudierons comment plusieurs molécules REV se lient à une copie de l'ARNm de HIV-1.

Partie 1 (6,5 points):

analyse de la structure secondaire de la région RRE de l'ARNm primaire de HIV-1

1 – (0,5 point)

Donner une définition de la structure primaire d'un ARN. Donner une définition de la structure secondaire d'un ARN.

2 – (0,5 point)

Dessiner la structure des paires de ribonucléotides A-U et G-C en indiquant les liaisons de type Watson-Crick qui stabilisent les interactions entre ribonucléotides.

3 – (1 point)

La figure 2 ci-dessous montre la structure secondaire prédite d'une sous-région du RRE de l'ARNm de HIV-1. Cette sous-région, appelée domaine II, est suffisante pour lier *in vitro* une molécule de Rev avec une très haute affinité. Le domaine II est long de 66 nucléotides.

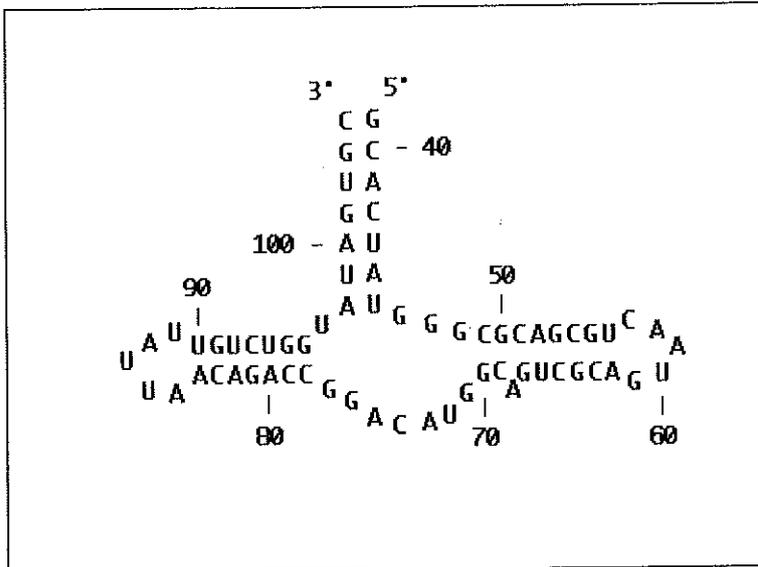


Figure 2

Structure prédite du domaine II du RRE de l'ARNm de HIV-1

La validité de ce modèle de structure secondaire est testée par une approche génétique décrite ci-dessous. Cette approche est basée sur l'idée que l'effet d'une mutation qui abolit une interaction entre deux nucléotides peut être compensé par une seconde mutation qui restaure cette interaction. Justifier brièvement cette idée.

Un oligoribonucléotide correspondant au domaine RRE est produit par synthèse chimique. Un taux fixe de mutations est introduit lors de la synthèse chimique de cet oligoribonucléotide de telle sorte que, à chacune des 66 positions du domaine II, un ribonucléotide identique à celui trouvé dans le domaine II sauvage est introduit avec une fréquence de 65%, un ribonucléotide différent de celui trouvé dans le domaine II sauvage est introduit avec une fréquence de 30%, et une délétion est introduite avec une fréquence de 5%.

Calculez la probabilité de synthétiser une molécule de séquence identique à celle du domaine II sauvage ?

Calculez la probabilité pour que la paire (G53,C65) soit conservée?

4 – (1,5 point)

Un mélange de 10^{13} molécules est obtenu par synthèse chimique. Ce mélange d'oligoribonucléotides est incubé avec la protéine Rev purifiée. Les oligoribonucléotides se liant *in vitro* à Rev avec une affinité similaire à celle observée pour le domaine II sauvage sont sélectionnés. Une copie ADN des oligoribonucléotides ainsi sélectionnés est ensuite obtenue par une réaction de transcription inverse et d'amplification PCR (« polymerase chain reaction »). La séquence de 29 oligoribonucléotides ainsi sélectionnés a été déterminée.

L'alignement de ces 29 séquences permet d'identifier 9 bases strictement conservées (Figure 3).

*****UGGC*CAG*GU*****A**CUG**GUACA*****U*U*****

bases strictement conservées (caractère gras et souligné):

G46, G47, C49, C51, U66, G67, G70, A73, C74

bases conservées (dans au moins 24 clones sur 29):

U45, G48, A50, G51, G53, G55, U56, A62, C65, G71, U72, A75, U87, U89

Figure 3

Positions conservées au sein des 29 oligoribonucléotides sélectionnés par liaison in vitro à la protéine Rev. On note que la séquence du domaine II sauvage donné en Figure 1 diffère de la séquence consensus donnée ici aux positions 50 et 51.

Calculez la probabilité de trouver par hasard la même base à une position donnée entre ces 29 oligoribonucléotides ?

Calculez la probabilité de trouver par hasard la même base à neuf positions données entre ces 29 oligoribonucléotides ?

La séquence consensus présentée en Figure 2 vous paraît-elle significative ?

Cet alignement permet également de définir deux régions conservées, entre les positions 45-53 et 65-75. Pensez-vous que cette observation confirme la prédiction de structure secondaire montrée en Figure 1 qui met en évidence trois zones d'appariements qui couvrent les bases aux positions 39-46, 49-56, 62-70, 78-84, 90-96 et 98-104 ? Justifiez votre réponse.

5 – (1 point)

Les données de la Figure 2 indiquent que les bases G46, C74 sont strictement invariantes. Par ailleurs, les bases U45, A75 sont invariantes, à l'exception du clone 52 qui présente la paire C45, G75. On observe également que les bases A44, G76 sont fréquemment changées en A44, U76 ou en C44, G76. Enfin, dans 10 clones, les bases U43, G77 sont changées en C43, G77 ou U43, A77.

Sur la base de ces données, dessinez un modèle de structure secondaire pour le domaine II du RRE.

6 – (1 point)

Afin de tester l'importance des deux régions conservées (45-53 et 65-75) du domaine II, des oligoribonucléotides plus courts sont obtenus par synthèse chimique.

Le premier, RBE1, correspond à une sous-région du domaine II allant des positions U46 à A74, précédée de GG à l'extrémité 5'.

Le second, RBE2, correspond à RBE1 avec une extension CC à l'extrémité 3'.

Le troisième, appelé RBE3, correspond à RBE2 à l'exception de la boucle dont la séquence GUCAAUGAC a été remplacée par la séquence UUUCG.

La structure prédite de ces oligoribonucléotides est présentée dans la Figure 4.

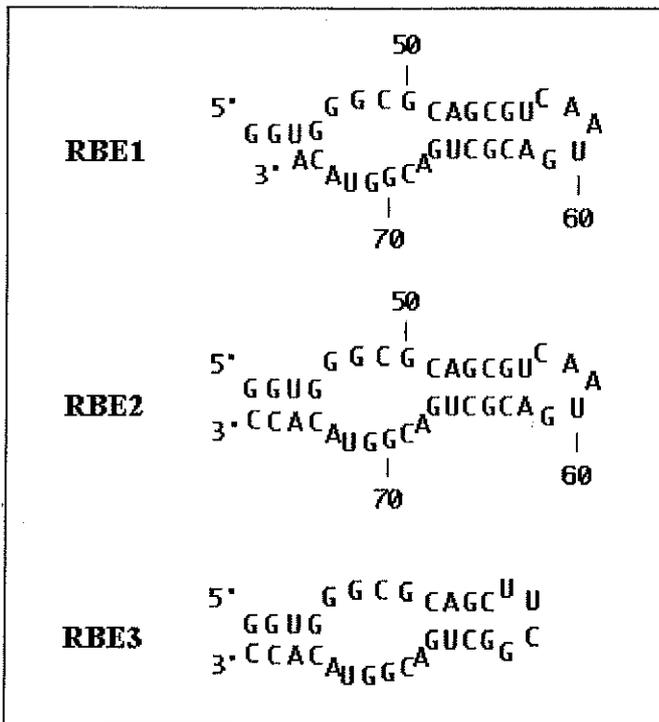


Figure 4 Structure prédite de RB1, RB2 et RB3.

Ces oligoribonucléotides sont comparés à l'oligoribonucléotide correspondant au domaine II du RRE sauvage pour leur capacité à lier la protéine Rev in vitro. Les résultats de cette expérience sont présentés dans la Table 1 (lignes 1 à 4 uniquement).

	Oligoribonucléotide testé :	Activité de liaison à Rev
1	RRE	1
2	RBE1	0.1
3	RBE2	0.9
4	RBE3	0.9
5	RBE2-C48	0.2
6	RBE2-C71	0.2
7	RBE2-C48C71	1.5

Table 1

Que pouvez-vous en déduire quand à la séquence minimale reconnue par la protéine Rev ?

7 – (1 point)

Dans l'expérience de mutagenèse décrite aux questions 3 et 4, une co-variation est observée aux positions G48 et G71 (cf. numérotation donnée en Figure 2). Sur les 29 séquences analysées, G71 est conservée dans 26 clones. Dans les trois autres clones, G71 est changé en C71. Dans ces trois clones, le G en position 48, G48, est changé en C48 ou A48. Cette co-variation est remarquable car G48 est également une position hautement conservée : G48 est conservé dans 24 clones sur 29.

L'importance fonctionnelle de ces mutations a été testée dans le contexte de l'oligoribonucléotide court RBE2. Les variants suivants de RBE2 ont donc été synthétisés :

RBE2-C48 : G48 muté en C

RBE2-C71 : G71 muté en C

RBE2-C48C71 : G48 et G71 mutés en C

Ces oligoribonucléotides sont comparés à l'oligoribonucléotide RRE, qui couvre le domaine II du RRE sauvage, pour leur capacité à lier la protéine Rev in vitro. Les résultats de cette expérience sont présentés dans la Table 1 (lignes 5-7).

Interprétez ces résultats. Quelle est la structure du domaine II du RRE reconnue par la protéine Rev ?

Partie 2 (8,5 points):
analyse de la structure de la protéine Rev

Dans cette seconde partie, nous analyserons les acides aminés de la protéine Rev qui sont importants pour la reconnaissance de la région RRE de l'ARNm primaire de HIV-1.

La protéine Rev synthétisée chez *E. coli* est capable de se lier à la séquence ARN U45-A81 du domaine II du RRE dans la bactérie. Nous rappelons ici qu'une seule molécule Rev se fixe à cette séquence d'ARN. L'ADN codant cette séquence est placé dans la région 5' non traduite d'un gène rapporteur, en amont du codon d'initiation du gène lacZ codant pour la β -galactosidase. Ce gène rapporteur est appelé LacZ-RRE. La liaison de Rev à l'ARNm du gène lacZ-RRE bloque sa traduction. Une simple mesure de l'activité enzymatique β -galactosidase permet alors de mesurer la répression traductionnelle du gène lacZ, et donc l'activité de liaison de la protéine Rev à sa cible ARN. Ce test d'activité de liaison de Rev à la séquence RRE est schématisé dans la Figure 4.

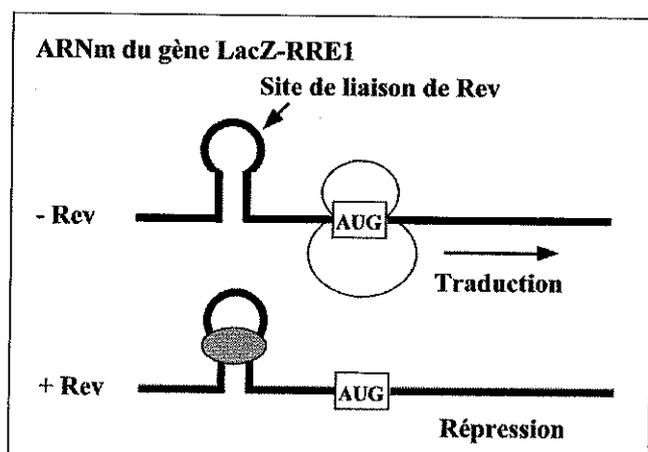


Figure 5

Test de répression traductionnelle du gène lacZ-RRE par la protéine Rev chez *E. coli*.

8 – (1,5 points)

Ce test de répression est utilisé pour mesurer la capacité de la protéine Rev à se fixer à différents mutants du domaine II du RRE (le mutant G47A correspond à la mutation suivante : G47 est muté en A47).

Les résultats de cette expérience sont présentés dans la Table 2 :

Gène rapporteur	taux de répression	Kd (nM)
LacZ-RRE	34	4,8
LacZ-RRE G47A	3	280
LacZ-RRE G71U	1,5	990
LacZ-RRE G46A, C74U	9,5	27
LacZ-RRE A73G	1,3	12 000
LacZ-RRE C49U, G70A	1,2	20 000

Table 2

En parallèle, la constante de liaison à l'équilibre (Kd ; encore appelée constante de dissociation) est mesurée in vitro dans un test de retard de migration électrophorétique (Figure 6). Dans cette expérience, des concentrations croissantes de protéine Rev purifiée est incubée avec une quantité fixe d'ARN radiomarqué (0,5 nM). Le mélange est ensuite séparé par électrophorèse sur gel d'acrylamide. L'ARN libre migre rapidement, tandis que l'ARN complexé à Rev a une migration retardée (la protéine Rev retarde la migration de l'ARN dans le gel). Après électrophorèse, le gel est séché et autoradiographié. L'autoradiographie de ce gel est présentée en Figure 6A. La radioactivité présente dans le complexe Rev-ARN et l'ARN libre est mesurée après séparation électrophorétique .

Ces données sont corrélées à celles obtenues dans le test de répression (Figure 6B).

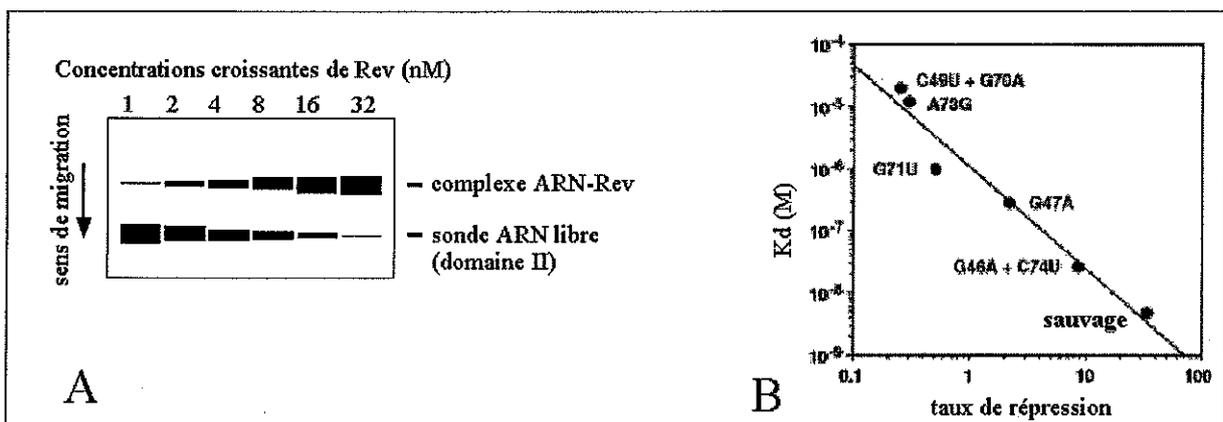


Figure 6

A : test de retard de migration sur gel du complexe Rev-ARN.

B : Relation entre valeurs de Kd et taux de répression

Qu'est ce qu'une constante de liaison à l'équilibre (Kd) ?

Comment la Kd peut-elle être mesurée expérimentalement à partir des données obtenues dans le test de retard de migration sur gel ?

Les valeurs obtenues dans le test de répression (Table 2) vous paraissent-elles en accord avec les données présentées dans la partie 1 de ce problème ?

Que pouvez-vous déduire de la linéarité observée entre les mesures du taux de répression et de valeurs de Kd pour les différents RRE testés ?

9 – (0,75 points)

Afin de tester simultanément un grand nombre de mutations affectant la séquence de la protéine Rev, la séquence codant la protéine Rev a été mutée in vitro en utilisant une méthode (non décrite ici) permettant d'introduire de manière aléatoire des mutations au sein de la séquence de Rev. Cette mutagenèse permet de produire une banque d'ADN contenant environ

75 000 séquences différentes codant des versions mutées de la protéine Rev. Chacune de ces 75 000 protéines mutantes diffère de la protéine Rev sauvage à 1 ou 2 acides aminés.

Un nombre important de ces molécules mutantes ne sont plus capables de réprimer la traduction des gènes lacZ-RRE, lacZ-RRE G47A ou lacZ-RRE A73G.

Quelles peuvent être les bases moléculaires de ce défaut ?

10 – (1 point)

Trois clones mutants de Rev présentent une activité de répression de la traduction des ARNm mutants lacZ-RRE G47A et lacZ-RRE A73G qui est plus forte que celle observée pour Rev sauvage (Table 3). Chacun de ces clones porte une seule mutation : Q36G, E47A et Q51R (le code à une lettre des acides aminés est donné en Annexe 1). Les trois protéines mutantes, dénommées Rev Q36G, Rev E47A et Rev Q51R, sont accumulées en quantité comparable à la protéine sauvage chez *E. coli*. Les différences mesurées dans le test de répression ne sont donc pas dues à des différences d'expression chez *E. coli*.

Protéine Rev	Taux de répression pour différentes séquences RRE :		
	lacZ-RRE	lacZ-RRE G47A	lacZ-RRE A73G
Rev	34	3	1
Rev Q36G	31	18	3
Rev E47A	48	8	3
Rev Q51R	35	8	3

Table 3

Interprétez ces données. Comment peut-on expliquer le taux élevé de répression observé pour ces trois protéines mutantes avec les gènes rapporteurs contenant un RRE muté ?

11 – (1 point)

L'analyse de la structure secondaire de la protéine Rev en solution indique que la région de la protéine allant de l'acide aminé 34 à 51 forme une hélice alpha. D'autre part, il est établi que les acides aminés R35, R39, R42, R43, R46, R50 contactent de manière directe la séquence RRE.

Quel est le nombre d'acides aminés par tour d'hélice alpha ?

Que pouvez-vous en déduire pour la localisation des acides aminés Q36, E47 et Q51 dans le complexe Rev-RRE? Même question pour les acides aminés R35, R39, R42, R43, R46, R50 ?

Quel peut-être le rôle des acides aminés de type Arginine (R) dans l'interaction avec l'ARN ?

12 – (2 points)

Les expériences présentées ci-dessus concernaient l'analyse de l'interaction d'une seule molécule Rev avec un sous-domaine de la région RRE. Or, il est établi que la région RRE entière (234 nucléotides) est capable de lier *in vitro* jusqu'à six molécules Rev et que la fixation de plusieurs molécules Rev est requise *in vivo* pour permettre le passage de la phase précoce à la phase tardive d'infection. Nous allons ici étudier le mécanisme par lequel plusieurs molécules Rev se fixent au RRE. En particulier, nous allons identifier les acides aminés de la protéine Rev important pour l'assemblage d'un multimère de Rev sur le RRE.

L'interaction entre deux molécules Rev fixées à l'ARN est étudiée grâce à un test de répression traductionnel chez *E. coli*. Ce test est dérivé de celui présenté en Figure 5. Ce nouveau test est présenté dans la Figure 6. Brièvement, deux copies du domaine II du RRE de

l'ARNm primaire de HIV-1 (de U45 à A81) sont insérées dans la région 5' non traduite du gène lacZ. Ce gène rapporteur est appelé LacZ-2xRRE. La copie la plus proche du codon d'initiation de la traduction porte une mutation qui bloque la fixation de Rev à cette version mutée du RRE (RREm). L'autre copie, située en position distale par rapport au codon d'initiation, est sauvage. Dans ce test, la protéine Rev sauvage se fixe à un site éloigné du codon d'initiation et bloque de manière moins efficace la traduction. Cependant, la liaison de Rev à ce site distal favorise par coopérativité de liaison la fixation d'une deuxième molécule Rev sur le site proximal RREm. Ceci conduit à une augmentation de la répression de la traduction. Un second gène rapporteur, appelé LacZ-1xRRE, est utilisé en contrôle. Ce gène comporte une seule copie de la version mutante du domaine II du RRE (RREm). L'activité de répression de la protéine Rev est mesurée pour LacZ-2xRRE et LacZ-1xRRE (Table 4). Ce taux de répression correspond au rapport des mesures d'activité β -galactosidase de ces rapporteurs en l'absence et en présence de la protéine Rev.

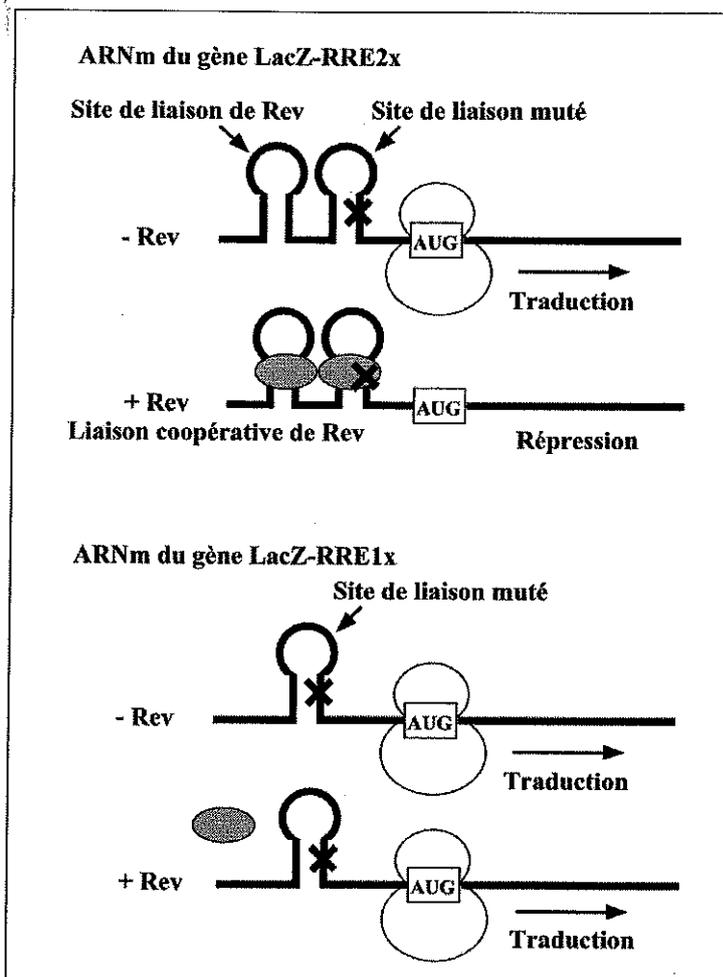


Figure 7

Test génétique permettant d'analyser la coopérativité de liaison de deux molécules Rev liées à l'ARN

Protéine Rev	Taux de répression pour :	
	LacZ-2xRRE	LacZ-1xRRE
Rev	11	1,5
Rev V16D	7	1,5
Rev L18Q	2,5	1
Rev L18R	2,5	1
Rev I55N	2,5	1
Rev L60R	7	1,5
Rev L60P	7	1,5

Table 4

Ce test est utilisé pour rechercher des protéines Rev mutantes chez lesquelles l'activité de répression de LacZ-1xRRE est inchangée mais chez lesquelles l'activité de répression de LacZ-2xRRE est diminuée. Les mutations suivantes ont été isolées: V16D, L18Q, L18R, I55N, L60R, L60P (Table 4).

A quoi peuvent correspondre les mutants pour lesquels l'activité de répression observée avec le gène rapporteur LacZ-1xRRE est réduite ?
 Pourquoi est-il important de ne garder que les mutants pour lesquels l'activité de répression observée avec le gène rapporteur LacZ-1xRRE est inchangée ?

Formulez une hypothèse quant à l'activité de la protéine Rev altérée par ces mutations.

13 – (1,5 points)

Les protéines mutantes Rev V16D, Rev L18Q, Rev L18R, Rev I55N, Rev L60R et Rev L60P sont produites chez *E. coli*, purifiées et testées pour leur capacité à lier in vitro un ARN radiomarqué correspondant à la séquence RRE entière (234 nucléotides).

Dans cette expérience, la protéine Rev (sauvage ou mutante) est incubée avec la sonde ARN radiomarquée et le mélange est séparé par électrophorèse sur gel d'acrylamide. L'ARN libre migre rapidement, tandis que l'ARN complexé à Rev a une migration retardée sur gel. Après électrophorèse, le gel est séché et autoradiographié. L'autoradiographie de ce gel est présentée en Figure 8.

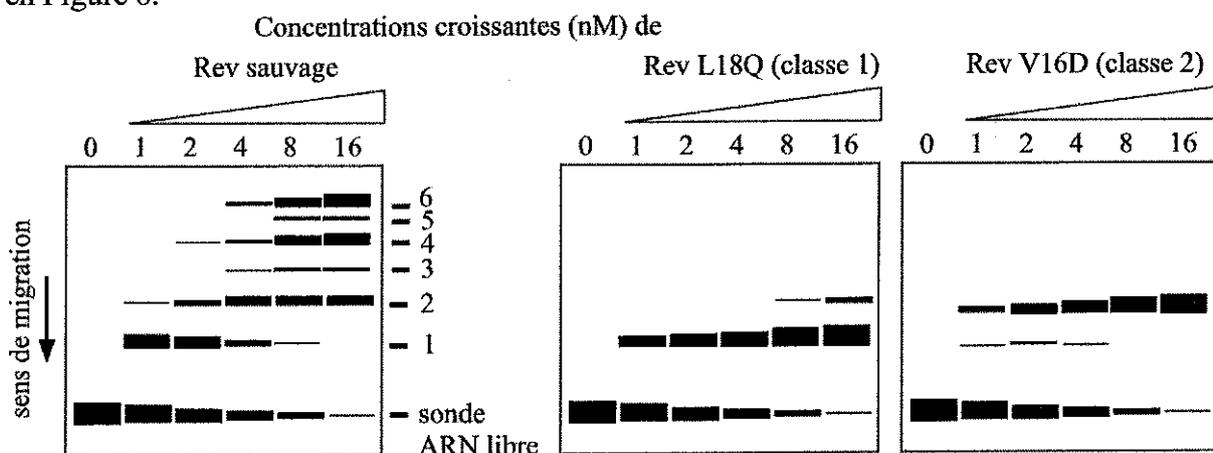


Figure 8

Analyse des complexes Rev-RRE formés in vitro dans un test de retard de migration sur gel.

Deux classes de mutants sont observées (Figure 8):

Classe 1 : l'espèce moléculaire 1 est prédominante. Les espèces moléculaires 2, 3, 4 et 5 ne sont pas détectées. Cette classe inclut les mutants affectant les positions L18 et I55.

Classe 2 : l'espèce moléculaire 2 est prédominante. Cette classe inclut les mutants affectant les positions V16 et L60.

A quels complexes correspondent les espèces moléculaires 1, 2, 3, 4, 5 et 6?

Quel est l'effet des différentes mutations de Rev testées par cette méthode ? En particulier, quel est le défaut spécifique des mutants de classe 1 ? Même question pour les mutants de classe 2 ?

14 – (0,75 point)

Les acides aminés mutés dans la protéine Rev portent tous des chaînes aliphatiques (V, L et I). Par ailleurs, ces acides aminés (V16, L18, I55 et L60) sont strictement conservés dans tous les variants naturels du virus HIV-1.

Comment interprétez-vous ces observations ?

En vous appuyant sur votre interprétation des deux classes de mutants observés ici, proposez un modèle pour l'assemblage des six molécules Rev sur la séquence RRE.

ANNEXE 1 : code à une lettre des acides aminés

A : Alanine

C : Cystéine

D : Acide aspartique

E : Acide glutamique

F : Phénylalanine

G : Glycine

H : Histidine

I : Isoleucine

K : Lysine

L : Leucine

M : Méthionine

N : Asparagine

P : Proline

Q : Glutamine

R : Arginine

S : Serine

T : Thréonine

V : Valine

W : Tryptophane

Y : Tyrosine