

Nouveau programme de BCPST1 2021
Petit condensé pour les 5/2-cubes
Génétique

I. Ce que dit le programme

(seules les différences avec l'ancien programme sont notées)

SV-F Génomique structurale et fonctionnelle

La présentation des génomes et de leur organisation est l'occasion de préciser les points communs et les différences entre Eucaryotes, bactéries et virus.

SV-F-1 Génome des cellules et des virus, transmission de l'information génétique (BCPST 1)

savoirs visés	capacités exigibles
Des techniques de séquençage permettent de déterminer la séquence d'un fragment d'ADN puis de proche en proche la séquence des génomes.	- Interpréter l'organisation des génomes à partir des résultats de séquençage.
Précisions et limites : <i>Pour le séquençage, seul le principe de la méthode de Sanger doit être connu.</i>	
Les virus ou particules virales sont des entités nucléoprotéiques comprenant un acide nucléique (sous forme d'ADN ou d'ARN) constituant le génome viral, et des protéines. On distingue des protéines à rôle structural, formant la capsid, et parfois des protéines à rôle enzymatique. Les virus sont très divers et possèdent parfois une enveloppe lipoprotéique.	- Comparer l'organisation du génome des bactéries, des Eucaryotes et des virus. - Estimer la proportion de séquences codantes et non codantes dans les génomes [...] et des virus. - Illustrer la diversité structurale et la diversité d'hôte des virus.
Précisions et limites : <i>Aucune monographie de chaque virus n'est attendue. Il s'agit de montrer la diversité structurale (organisation structurale, taille, présence ou non d'une enveloppe, nature de l'information génétique) et la diversité d'hôte à l'aide de trois exemples, sans rentrer dans les détails des cycles de multiplication : bactériophage lambda, VMT, un coronavirus zoonotique.</i>	
Les connaissances sur les ADN polymérases ont permis d'élaborer des méthodes d'amplification in vitro de l'ADN.	- En se basant sur le fonctionnement des ADN polymérases, expliquer le principe de la PCR.

SV-F-2 L'expression du génome (BCPST1)

savoirs visés	capacités exigibles
Les virus se multiplient en détournant la machinerie d'expression de l'information génétique de la cellule hôte.	
Précisions et limites : <i>On présente une vue générale de la multiplication sur un exemple au choix en se limitant aux modalités de réplication de l'information génétique, de synthèse des protéines et de formation de nouveaux virus. On se limite à distinguer la synthèse des polymérases et des protéines de la capsid, sans détailler les différentes protéines virales et les mécanismes moléculaires impliqués.</i>	

II. Ce qu'il faut savoir et savoir faire

1. Séquençage

Le séquençage est le fait de **déterminer la séquence de nucléotides** d'une molécule d'acide nucléique (généralement, de l'ADN). Diverses techniques existent. La plus ancienne, toujours utilisée, même si de nombreuses autres techniques plus performantes existent, est la **méthode de Sanger** (inventée en 1977 par Frederick Sanger, prix Nobel en 1980).

La technique est basée sur la **réplication de l'ADN**. On place dans un tube :

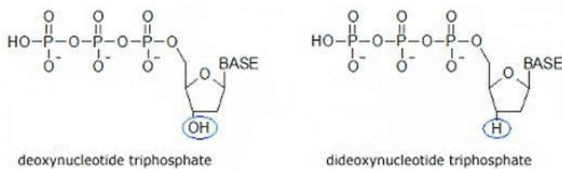
- de nombreuses copies de **l'ADN à séquencer** ;
- une **ADN-polymérase** (enzyme catalysant la réplication) ;
- les **4 désoxyribonucléotides triphosphates** (dNTP) : dATP, dTTP, dCTP et dGTP ;
- des **amorces** (permettant d'initier la polymérisation).

Mais surtout, l'astuce essentielle du protocole est l'utilisation d'une très petite quantité d'un **didésoxyribonucléotide triphosphate (ddNTP, doc 1)**. Il s'agit de nucléotides possédant un ribose dont les deux carbones 2' **et surtout 3'** sont désoxygénés. De cette façon :

- le ddNTP peut être utilisé par l'ADN-polymérase et intégré à la molécule d'ADN en cours d'élongation, via son extrémité 5'-phosphate
- **mais, puisque le -OH de l'extrémité 3' manque, il arrête forcément la polymérisation une fois intégré.**

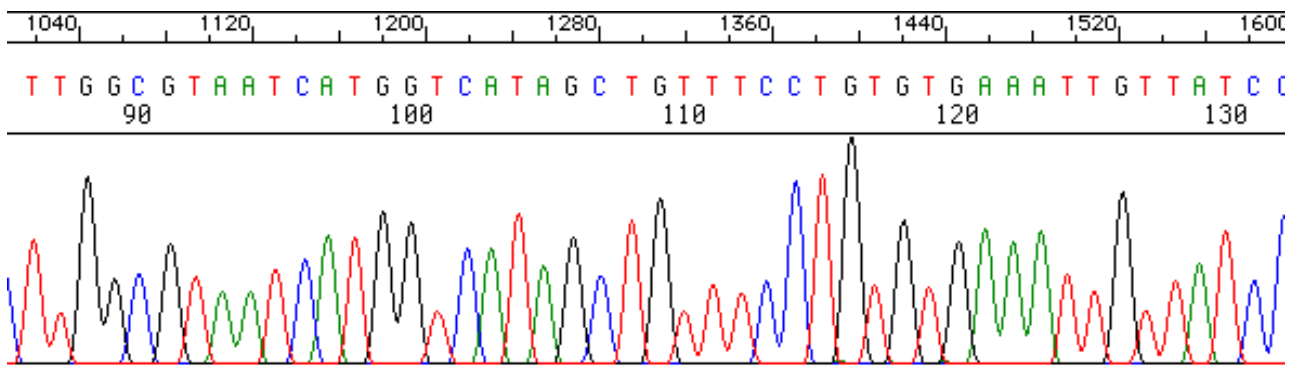
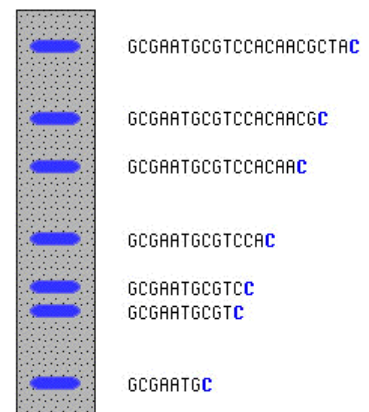
Exemple (**doc 2**) : on utilise mélange contenant les 4 dNTP et un ddCTP (ddNTP à cytosine) ; l'intégration aléatoire du ddATP est responsable de la formation de molécules d'ADN de tailles variables, mais **qui se terminent toutes par une adénosine**. Une électrophorèse à haute définition permet de déterminer la taille de ces fragments, et donc de **déterminer la position de toutes les adénines de la molécule à séquence**. Il suffit de répéter autre fois la même manipulation avec chacun des 4 ddNTP (ddATP, ddTTP, ddCTP et ddGTP).

NB : une amélioration de cette technique permet aujourd'hui de réaliser ces 4 manipulations en une seule fois, en marquant les ddNTP par des fluorochromes différents. L'électrophorèse a alors lieu dans un tube capillaire, et les molécules de taille différente sont identifiées par leur passage devant un détecteur muni d'un laser excitant la fluorescence du fluorochrome (doc 3).



Document 1: Structure d'un désoxyribonucléotide triphosphate (dNTP, gauche) et d'un didésoxyribonucléotide (ddNTP, droite), utilisé dans le séquençage de l'ADN par la méthode de Sanger.

Document 2: Résultat théorique d'une électrophorèse du produit d'une réaction de séquençage par la méthode de Sanger, avec utilisation d'un ddNTP à cytosine. La migration a lieu du haut vers le bas. La séquence déduite est donnée à droite (on ne connaît en réalité que la position des cytosines par cette seule manipulation).



Document 3: résultat d'un séquençage de Sanger avec ddNTP fluorescents. Les quatre courbes colorées représentent l'intensité de la fluorescence (T en rouge, G en noir, C en bleu et A en vert). Echelle en haut : temps de migration ; échelle en bas : position des nucléotides ; la séquence déduite est inscrite au dessus des courbes.

Ce qu'on peut vous demander au concours :

- **expliquer le principe de la méthode de Sanger**
- **exploiter un résultat d'électrophorèse permettant de déterminer la séquence d'une molécule d'ADN**
- **exploiter un résultat de séquençage de Sanger sur tube capillaire avec ddNTP fluorescent**

2. La PCR et la RT-PCR

La PCR (*polymerisation chain reaction*) est une technique d'identification de séquence, qui est elle aussi **basée sur la réplification de l'ADN**. Elle consiste à réaliser une réaction de polymérisation, en sélectionnant un couple de **deux amorces** (une pour un brin, une pour l'autre brin), qui encadrent la séquence d'intérêt. L'ADN polymérase polymérise l'ADN à partir de ces deux amorces, et copie l'ADN lors d'un **cycle de réplification**. Il est constitué des étapes suivantes :

- chauffage à 95°C : **dénaturation de l'ADN**
- refroidissement à 60°C : **hybridation des amorces sur leur complémentaire**
- chauffage à 72°C : **polymérisation**

Une présentation sur youtube à ce sujet est très claire : <https://www.youtube.com/watch?v=iQsu3Kz9NYo>.

Un cycle est répété une trentaine de fois, et on peut montrer qu'à chaque cycle, le nombre de molécules d'ADN est multiplié par deux. Le nombre de molécules au bout de 30 cycles est donc de 2^{30} , soit environ 10^9 . Ces molécules sont détectées par électrophorèse. On en déduit trois intérêts majeurs :

- concernant le nombre de molécule d'ADN initial : la quantité d'ADN présente initialement peut être **extrêmement faible** (en théorie, une seule molécule), et non détectable. Le résultat de la PCR est positif **si et seulement si cet ADN initial est amplifié** (= répliqué) et qu'il devient détectable par électrophorèse.
- concernant les amorces : s'il n'existe pas, dans la séquence présente initialement dans le tube, une séquence complémentaire à une des amorces, **la réaction de polymérisation n'a pas lieu**, et aucun ADN ne sera détectable par électrophorèse à l'issue de la réaction. C'est ce qui fait que cette technique est bien une **technique d'identification de séquence**.
- concernant la taille de l'ADN amplifié : l'ADN amplifié a une taille imposé par le choix des amorces, et cette taille est déterminée par électrophorèse. Cette taille est un élément important d'identification de la séquence.

La PCR est un outil de choix pour l'identification d'une séquence présente à l'état de trace (identification d'un microorganisme dans un prélèvement sanguin, d'un individu sur une scène de crime avec seulement cheveu ou une trace de sang). Elle a désormais quasiment totalement remplacé le *Southern blot* pour ces raisons.

La RT-PCR est une technique dérivée de la PCR et adaptée à la détection des ARN. « RT » signifie « reverse-transcription » ; il s'agit d'une étape de transcription inverse, catalysée par une reverse-transcriptase, isolée à partir de rétrovirus (comme le VIH, virus du SIDA) : un ADN double brin est synthétisé à partir d'un ARN simple brin. Cet ADN subit ensuite une PCR.

La RT-PCR est utilisée pour détecter des ARN génomique de virus, avec l'application désormais connue de tous qu'est le « test PCR de dépistage du covid » qui est en réalité une RT-PCR de recherche de l'ARN du SARS-CoV2, le virus du covid. Elle est aussi un outil de choix pour étudier l'expression génétique des cellules vivantes, en parallèle du northern blot et de l'hybridation in-situ (voir cours de BCPST1, ancien et nouveau programme inchangé).

NB : l'utilisation d'une polymérase particulière, la taq-polymérase, permet de travailler à des températures qui dénatureraient normalement la plupart des enzymes. La taq-polymérase est initialement extraites de bactéries thermophiles, vivant dans des sources hydrothermales à une température d'environ 100°C.

Ce qu'on peut vous demander au concours :

- **expliquer le principe de la PCR et de la RT-PCR**
- **analyser un résultat de PCR (ou de RT-PCR) en validant la détection de la séquence recherchée par le choix des amorces**

3. Les virus

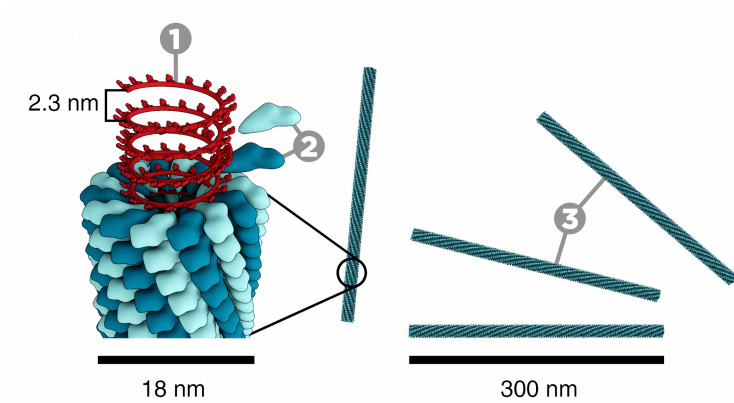
Un virus est une particule constituée d'un acide nucléique (ARN ou ADN) enfermé dans une coque protéique (la capsid), et éventuellement protégé par une enveloppe lipoprotéique, de structure analogue à celle d'une membrane. Les virus ne sont pas des cellules vivantes, mais partagent des caractéristiques qui font que le débat de leur appartenance à la vie est aussi passionnant que stérile.

a) Structure

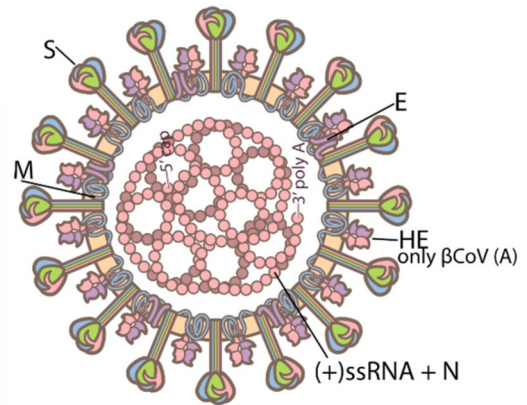
De nombreuses morphologies virales existent. Parmi les virus au programme :

- Le virus de la mosaïque du Tabac (VMT) est un virus à capsid hélicoïdale non enveloppé (doc 4)

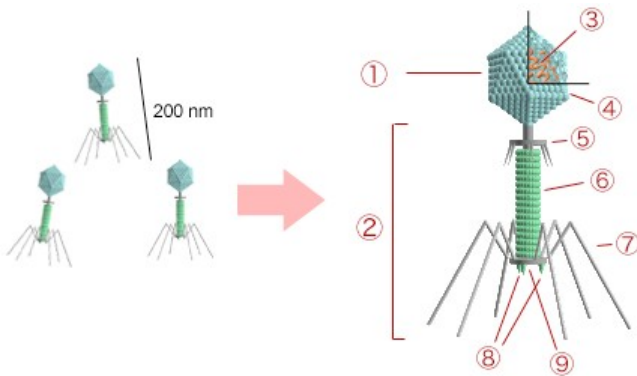
- Le virus du covid (SARS-CoV2) est un virus enveloppé (qui possède une enveloppe lipoprotéique) (doc 5)
- Le phage λ (virus infectant la bactérie *Escherichia coli*) est un virus à capsid complexe (doc 6)



Document 4: Structure du virus de la mosaïque du tabac. 1 : ARN ; 2 : (+)ssRNA + N : ARN simple brin de polarité positive avec nucléocapside ; 3 virions.



Document 5: Structure d'un virus enveloppé : le SARS-CoV2. S : spike ; E : protéine d'enveloppe ; M : protéine membranaire ; (+)ssRNA + N : ARN simple brin de polarité positive avec nucléocapside.



Document 6: Structure d'un phage. 1. tête ; 2 queue ; 3 : ADN ; 4 : capsid ; 5 : col ; 6 : anneau contractile ; 7 : fibres caudales ; 8 : spicules ; 9 : plaque terminale.

Contrairement aux cellules vivantes pour lesquelles le support de l'information génétique est toujours un ADN double brin, l'information génétique des virus peut être présent sous de très nombreuses formes. Voir le tableau (doc 7) pour plus d'information.

Le génome des virus est généralement de très petite taille (à comparer aux 10^6 pb des bactéries, et 10^9 pb des mammifères) :

- VMT : 6395 nucléotides
- SARS-CoV2 : 29 903 nucléotides
- Le phage λ : 48 502 paires de bases

Ce génome est quasiment totalement codant, et code un très petit nombre de protéines (généralement, pas plus de quelques dizaines, à comparer aux dizaines à centaines de milliers de protéines produites par les cellules). Ces protéines sont les protéines de capsid, et parfois quelques enzymes (polymérases permettant la répllication de leur génome, intégrase permettant l'intégration de leur génome dans celui de la cellule hôte...)

Chez de nombreux virus, un gène peut coder plusieurs protéines différentes, par utilisation de plusieurs cadres de lecture, ou par des gènes partiellement chevauchants (voir doc 7 pour un exemple chez le SARS-CoV2).

le support de l'information génétique peut être...	exemples
...de l'ADN	cellules eucaryotes, eubactériennes et archéobactériennes, virus de la varicelle (Herpesviridés), virus du typhus félin (Parvoviridés)...
...de l'ARN	virus du SIDA (retroviridae), virus de la grippe (Orthomyxoviridés), virus du covid (Coronaviridés)
...double brin	cellules eucaryotes, eubactériennes et archéobactériennes, virus de la varicelle (Herpesviridés),
...simple brin	virus du SIDA (retroviridae), virus du covid (Coronaviridés), virus du typhus félin (Parvoviridés)
...circulaire	cellules eubactériennes et archéobactériennes, plastes et mitochondries des cellules eucaryotes, papillomavirus humain (Papillomaviridés)
...linéaire	génome nucléaire des cellules eucaryotes, virus de la varicelle (Herpesviridés), virus du covid (Coronaviridés)
...en une seule molécule	génome principal des cellules eubactériennes, virus de la varicelle (Herpesviridés), virus du covid (Coronaviridés)
...en plusieurs molécules (segmenté)	cellules eucaryotes, virus de la grippe (Orthomyxoviridés)

Document 7: Diversité des génomes chez les cellules et les virus.

Dans le cas de molécules d'ARN génomiques simple brin, l'acide nucléique est dit à polarité positive si elle est directement

b) Reproduction

Document 8: Le cycle de reproduction du SARS-CoV2, le virus à l'origine de la pandémie de Covid-19. Il s'agit d'un virus à ARN simple brin positif. D'après Biologie-Géologie BCPST1 Tout-en-un, Dunod, 2021

Les virus sont des parasites obligatoires, qui se reproduisent en exploitant la machinerie d'expression génétique de l'hôte. Certains virus exploitent l'ARN polymérase de l'hôte de façon à transcrire leur matériel génétique quand il est présent sous forme d'ADN ; de nombreux virus produisent cependant eux-même leur propre réplicase (SARS-CoV2 et VMT, par exemple). En revanche, **la totalité des virus** utilisent la machinerie de traduction (c'est-à-dire les ribosomes) pour produire leurs protéines. Cette exploitation de processus cellulaires par les virus est une des conséquences du dysfonctionnement, parfois mortel, des cellules infectées.

On donne ici (doc 8) le cycle de reproduction du SARS-CoV2. On conseille de connaître un cycle, parmi celui du SARS-CoV2, **ou** du VMT, **ou** du phage λ .

Ce qu'on peut vous demander au concours :

- **comparer les génomes de virus et des cellules vivantes (nature de la molécule support de l'information génétique, taille du génome, part codante...)**
- **dans une synthèse (orale ou écrite) sur l'expression génétique, intégrer la notion d'exploitation de l'expression génétique de la cellule par le virus**