

ÉCOLE NORMALE SUPÉRIEURE DE LYON

Concours d'admission session 2021

Filière universitaire : Second concours

## COMPOSITION DE BIOLOGIE-BIOCHIMIE

Durée : 3 heures

L'utilisation de tout document ou de calculatrice n'est pas autorisée pour cette épreuve.

\* \* \*

Ce livret comprend 13 pages numérotées de 1 à 13.

Le sujet comprend 24 questions, dont une première question de synthèse pour laquelle le temps à consacrer sera plus important. La progression des questions suit un ordre logique, il est donc recommandé de les traiter dans l'ordre prévu, mais de nombreuses questions peuvent être traitées indépendamment les unes des autres.

NB : Une brève analyse des résultats expérimentaux présentés suivie d'une conclusion dûment justifiée est attendue pour chaque question.

# Le système CRISPR-Cas

## **Introduction**

La française Emmanuelle Charpentier et l'américaine Jennifer Doudna ont reçu en 2020 le prix Nobel de Chimie pour la mise au point, en 2012, d'une méthode d'édition du génome dérivée du système CRISPR-Cas. Elles ont mis au point ce système à partir de mécanismes de défense de bactéries contre des virus appelés bactériophages ou phages, qui s'attaquent aux bactéries. Comme tous les virus, les bactériophages sont des parasites intracellulaires. Ils injectent leur ADN dans la cellule bactérienne et contraignent la bactérie à produire de nouvelles particules virales à partir de cet ADN, au détriment de son développement. Nous vous proposons de nous intéresser au système bactérien à l'origine de cette technique et à un exemple simple montrant la possibilité de l'utiliser pour modifier des cellules Eucaryotes.

## **Question de synthèse :**

**Question 1 :** En une à deux pages maximum, vous présenterez de manière illustrée les principales caractéristiques des cellules bactériennes, en mettant en évidence les différences avec celles des cellules Eucaryotes, au niveau de l'organisation cellulaire en général et de leur information génétique en particulier.

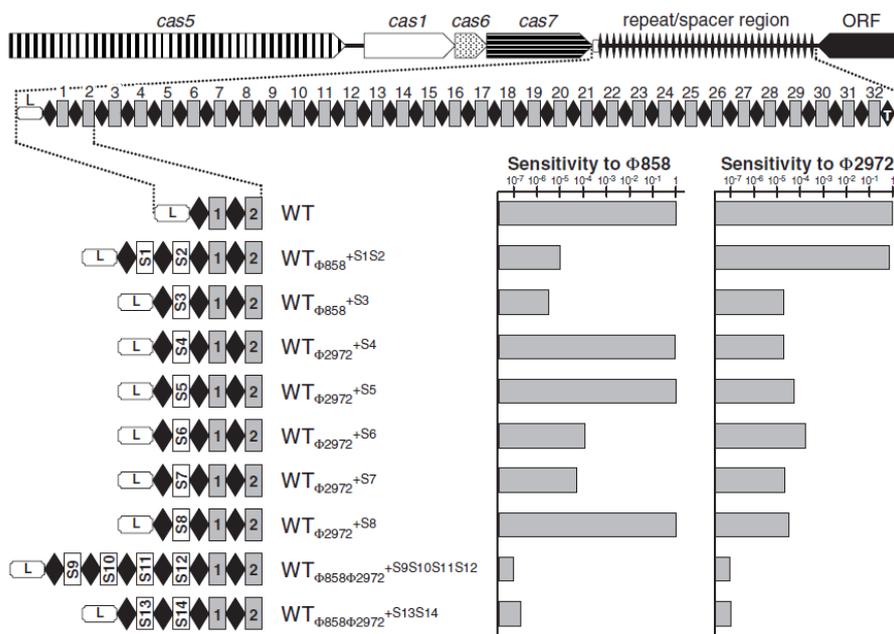
## **Partie 1 : Etude d'un mécanisme de résistance aux bactériophages**

Dans les années 1990, avec l'avènement du séquençage des génomes, des séquences inhabituellement répétées ont été découvertes dans le génome d'environ 40% des bactéries (et de manière encore plus fréquente chez les archaeobactéries) : il s'agit de répétitions de séquences plus ou moins inversées, séparées par des séquences très variables d'une trentaine de bases (appelées « spacers »). Les régions comprenant ces séquences répétées et les spacers ont été appelées CRISPR, pour Clustered Regularly Interspaced Short Palindromic Repeats. Des gènes qui codent pour des protéines sont associés à ces séquences : ils ont été appelés Cas pour CRISPR-associated.

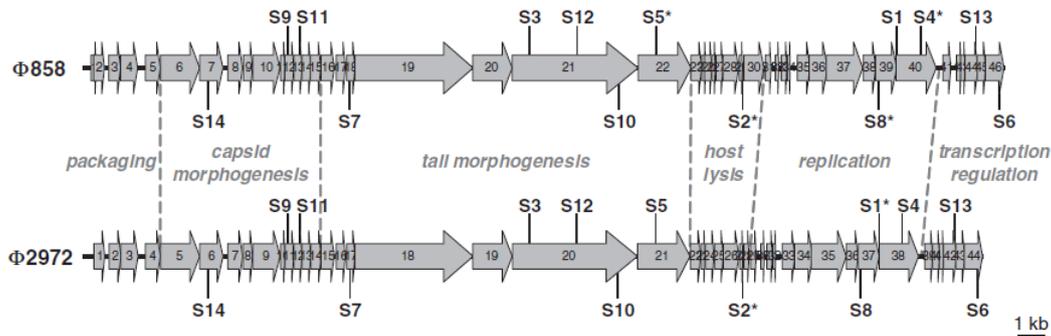
Parmi les nombreux articles décrivant le fonctionnement de ce système, un article publié en 2007 par Barrangou *et al.* dans la revue Science en présente une étude préliminaire chez la bactérie *Streptococcus thermophilus*, une bactérie d'intérêt industriel utilisée dans la production de yaourt et de fromage. Les scientifiques ayant réalisé cette étude se sont intéressés à différentes souches de *S. thermophilus* et ont remarqué, au niveau des spacers, une variabilité entre des souches très proches différant seulement dans leur capacité à résister à des bactériophages. Ils ont en particulier noté la présence de spacers supplémentaires dans les souches résistantes à des bactériophages par rapport à des souches sensibles à ces mêmes phages.

Ils ont alors mis une souche de *S. Thermophilus* en présence de deux phages appelés  $\Phi 858$  et  $\Phi 2972$ , seuls ou en combinaison et ont pu isoler 9 souches de bactéries mutantes résistantes à l'un, à l'autre ou à ces deux phages. Ils ont analysé le génome de ces souches résistantes, et ont pu constater l'apparition de spacers supplémentaires (S1 à S14) au niveau d'un des deux locus CRISPR de ces bactéries. Ils ont analysé ces séquences S1 à S14 et ont découvert qu'elles correspondent à des séquences retrouvées sur les génomes des deux phages, qui ont des génomes semblables à 93%.

**Figure 1 : Carte des loci CRISPR de la souche initiale de *S. Thermophilus* (souche WT) et des souches mutantes obtenues après traitement par le phage  $\Phi 858$  (souches  $WT_{\Phi 858}$ ), le phage  $\Phi 2972$  (souches  $WT_{\Phi 2972}$ ) ou les deux phages (souches  $WT_{\Phi 858\Phi 2972}$ ). La sensibilité de chaque bactérie aux deux phages est indiquée, une sensibilité (sensitivity) de 1 indique une souche très sensible et  $10^{-7}$  une souche très résistante à l'infection par le bactériophage indiqué.**



**Figure 2 : Carte des génomes des phages  $\Phi$ 858 et  $\Phi$ 2972 indiquant la position des séquences spacers retrouvées dans les bactéries devenues résistantes à ces phages. Un astérisque indique une différence mineure entre la séquence S du spacer et celle retrouvée dans le phage. Chaque flèche représente un gène viral, qui sont regroupés dans le génome des phages en grandes fonctions (empaquetage, morphogénèse de la capsid, de la queue du phage, lyse de l'hôte, réplication et régulation de la transcription).**



**Question 2 :** Analysez les résultats obtenus pour les deux bactéries résistantes au phage  $\Phi$ 858. Commentez non seulement la résistance au phage  $\Phi$ 858, mais aussi au phage  $\Phi$ 2972, en vous appuyant sur la complémentarité entre la séquence des spacers et celle du génome des deux phages.

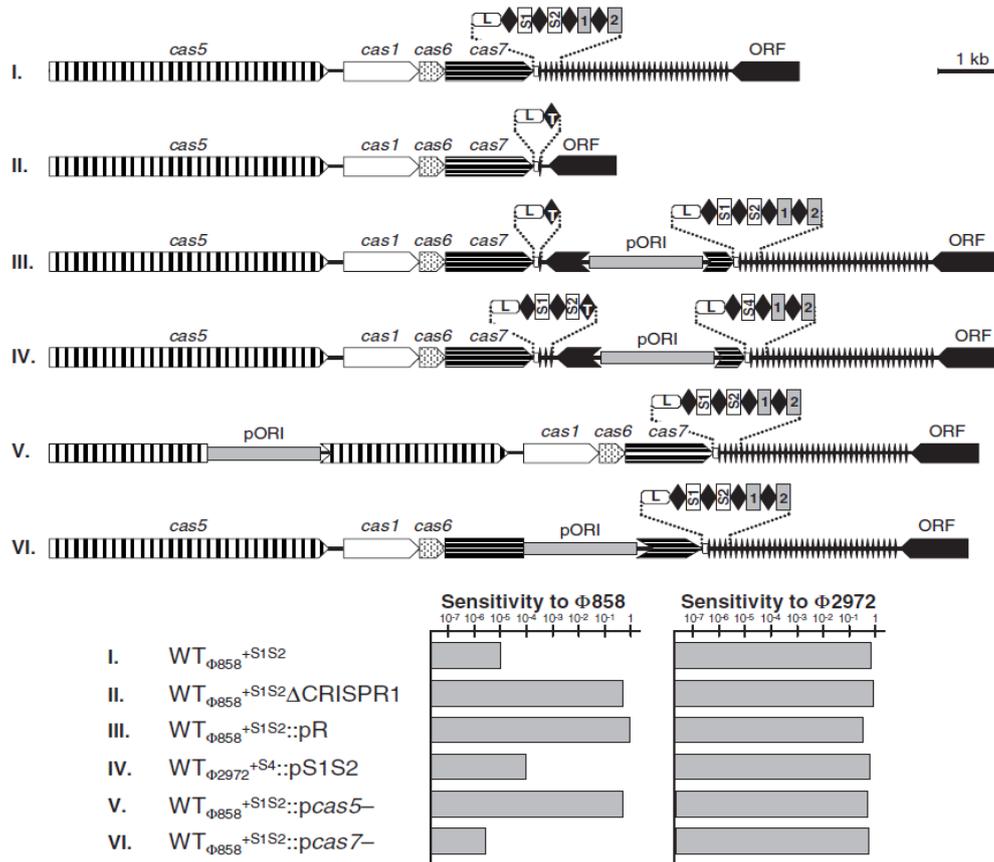
**Question 3 :** Analysez de la même façon les résultats obtenus pour les 5 bactéries résistantes au phage  $\Phi$ 2972.

**Question 4 :** Commentez les résultats obtenus pour les bactéries résistantes qui avaient été exposées aux deux phages en même temps.

**Question 5 :** Quelle(s) hypothèse(s) pouvez-vous faire quant au lien entre les spacers et la résistance des bactéries à l'attaque des phages ?

La figure suivante présente les analyses complémentaires réalisées par les auteurs pour comprendre d'une part le rôle des séquences spacers et d'autre part celle des gènes *cas*. Ils ont pour ce faire modifié génétiquement les bactéries résistantes qu'ils avaient obtenues.

**Figure 3 : Cartes des génomes des bactéries résistantes modifiées expérimentalement et résultat de l'analyse de leur sensibilité aux phages  $\Phi 858$  et  $\Phi 2972$ .** Ces modifications dans le génome des bactéries résistantes aux phages ont été réalisées à l'aide de plasmides introduisant une séquence pORI, dont la présence, si elle n'est pas dans un gène, n'interfère pas avec le fonctionnement de la bactérie.



**Question 6 :** Bactérie II. WT $_{\Phi 858}$ <sup>+S1S2</sup>  $\Delta$ CRISPR1. Les scientifiques ont tout d'abord supprimé l'ensemble de la séquence CRISPR et tous les spacers (La bactérie I. WT $_{\Phi 858}$ <sup>+S1S2</sup> est la bactérie résistante non modifiée). Analysez le résultat obtenu.

**Question 7 :** Bactérie III. WT $_{\Phi 858}$ <sup>+S1S2</sup> :: pR. Les scientifiques ont rajouté à la Bactérie II., à distance de son emplacement original sur le chromosome bactérien, la séquence CRISPR comprenant tous les spacers y compris S1 et S2. Analysez le résultat obtenu.

**Question 8 :** Bactérie IV. WT $_{\Phi 2972}$ <sup>+S4</sup> :: pS1S2. Dans des bactéries WT $_{\Phi 2972}$ <sup>+S4</sup> initialement résistantes au phage  $\Phi 2972$ , les scientifiques ont déplacé les séquences CRISPR contenant S4 et ajouté les séquences CRISPR contenant S1 et S2 à proximité des gènes cas. Analysez le résultat obtenu.

**Question 9 :** Bactéries V./VI. WT  $\Phi_{858}^{+S152}$  :: *pcas5*<sup>-</sup> / *pcas7*<sup>-</sup>

Les scientifiques ont inactivé le gène *cas5* ou le gène *cas7* en insérant une séquence pORI dans la séquence codante de chaque gène. Analysez les résultats obtenus.

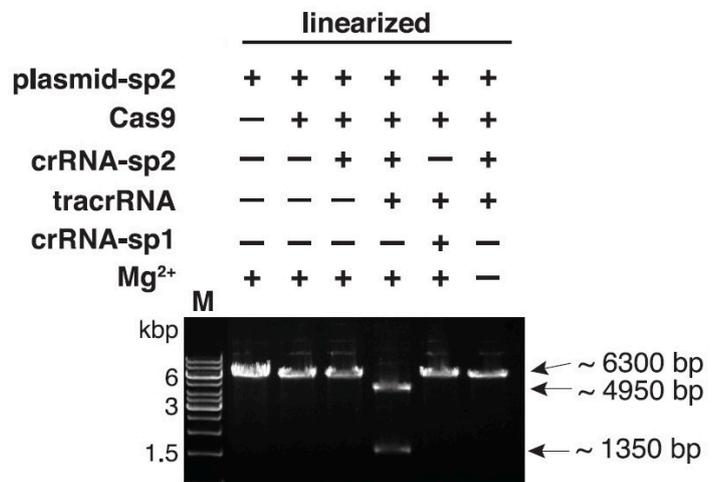
**Question 10 :** Cette équipe a, de la même façon, obtenu des bactéries n'exprimant plus le gène *cas1*. En termes de sensibilité aux phages, elle a obtenu des résultats similaires à ceux obtenus avec *cas7*, mais n'a pas pu, en traitant cette souche modifiée avec d'autres phages, obtenir de nouvelles résistances à ces phages. Quel pourrait être le rôle de la protéine Cas1 dans la résistance aux phages ?

Une grande diversité de protéines Cas et leur fonctionnement ont depuis été décrits, en particulier celui de la protéine Cas9 de la bactérie *Streptococcus pyogenes*, jouant un rôle similaire à celui de la protéine Cas5 vu précédemment. Chez cette bactérie, il a été démontré que deux ARN qui s'associent à la protéine Cas9 sont synthétisés à partir du locus CRISPR : un long ARN *tracrRNA* et un ARN court *crRNA* synthétisé à partir des séquences des spacers.

La souche de *S. pyogenes* étudiée ici peut produire différents ARN *crRNA* à partir des spacers du locus CRISPR, dont *sp1* et *sp2*. L'effet de la protéine Cas9 sur un ADN double brin linéaire de 6300 paires de bases (bp) contenant la séquence du spacer *sp2* a été étudié : cet ADN a été mis en présence des éléments indiqués, puis la taille de l'ADN obtenu a été analysée par électrophorèse sur un gel d'agarose.

**Figure 4 :** Analyse de l'effet d'un traitement par la protéine Cas9 d'un fragment d'ADN double brin linéaire contenant la séquence *sp2* (plasmid-*sp2*) en présence + ou absence - de différents éléments :

- la protéine Cas9
- le *crRNA-sp2* transcrit à partir du spacer *sp2*
- le *tracrRNA* transcrit par le locus CRISPR
- le *crRNA-sp1* transcrit à partir du spacer *sp1*
- du  $Mg^{2+}$



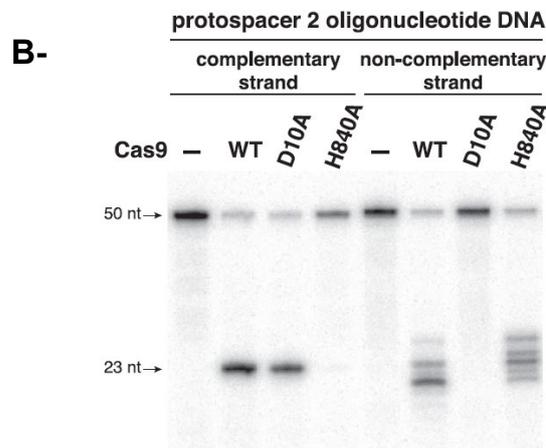
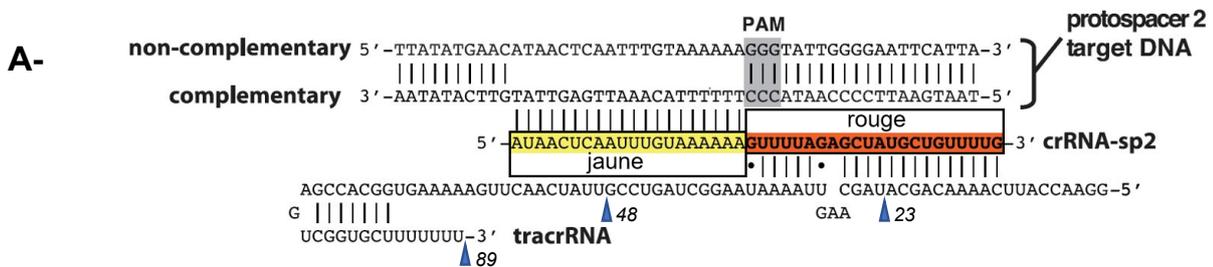
**Question 11 :** Interprétez ce résultat en précisant les éléments nécessaires pour que la protéine Cas9 ait un effet sur l'ADN. Quel est selon vous le rôle des éléments *crRNA* ?

La protéine Cas9 possède deux domaines endonucléases distincts homologues à des domaines endonucléases connus (des protéines RuvC1 et HNH). Afin de préciser le rôle respectif de ces deux domaines, des mutants inactivant l'un ou l'autre domaine ont été

réalisés, et l'effet de ces mutations sur l'activité des protéines Cas9 ainsi obtenues a été analysé sur de courts fragments d'ADN correspondant à une courte séquence cible de la protéine Cas9. Cette séquence sera appelée protospacer par la suite.

**Figure 5 : Analyse du rôle respectif des deux domaines endonucléases de la protéine Cas9.**

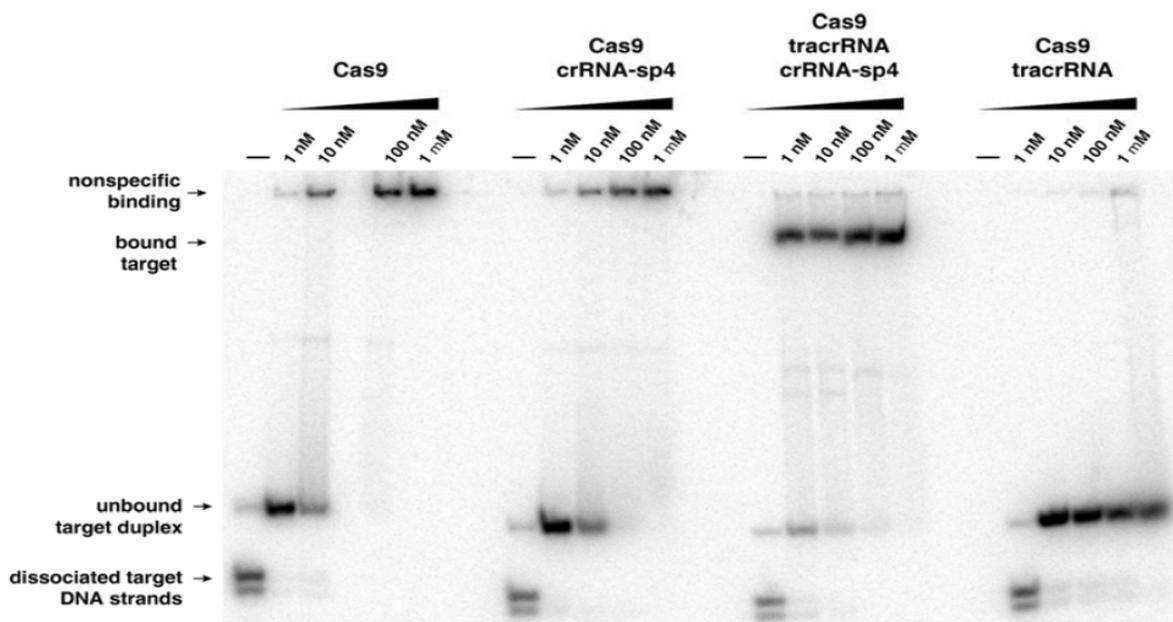
**A-** Schéma présentant les séquences de tracrRNA, crRNA-sp2 et de la séquence cible protospacer 2 de 50 nucléotides utilisée dans cette analyse. La séquence PAM est une séquence dont la présence a été démontrée comme essentielle à l'action de la protéine Cas9. Les séquences complémentaires et supposées s'apparier sont reliées par des traits (ou un point pour l'appariement possible entre un G et un U). (Les nucléotides du brin complémentaire du protospacer 2 indiqués comme pouvant s'apparier avec la séquence du crRNA-sp2 sont aussi complémentaires de la séquence du brin non-complémentaire indiqué au-dessus). Les séquences du crRNA-sp2 complémentaires au protospacer sont surlignées/encadrées en « jaune » et celles complémentaires au tracrRNA en « rouge ». **B-** De la protéine Cas9 associée aux ARNs crRNA-sp2 et tracrRNA a été mise en présence de l'ADN double brin de protospacer 2 représenté en A. Cet ADN avait été marqué radioactivement au préalable soit en 5' du brin complémentaire (complementary strand) au crRNA-sp2, soit en 5' du brin non complémentaire (non-complementary strand). Le signe "-" correspond à l'absence de protéine Cas9, WT au résultat obtenu avec la protéine non mutée, D10A à une protéine Cas9 portant une mutation inactivant le domaine RuvC1 et H840A à une protéine Cas9 portant une mutation inactivant le domaine HNH. L'ADN marqué a été analysé par migration sur gel. Remarque : Les positions des nucléotides 23, 48 et 89 du tracrRNA mentionnées à la figure 7 sont indiquées par un chiffre en italique à côté d'une tête de flèche.



**Question 12 :** Analysez cette expérience. Quel rôle respectif pour les deux domaines endonucléases pouvez-vous proposer ? Illustrez l'effet de chacun d'entre eux sur un schéma simplifié reprenant la figure 5A. .

Les scientifiques ont ensuite cherché à caractériser le rôle de l'ARN tracrRNA en réalisant une expérience de retard de migration sur gel. Cette technique consiste à analyser la migration de fragments d'ADN (ici marqués radioactivement sur les deux brins) sur un gel en conditions non dénaturantes (ne dissociant pas les éventuels complexes moléculaires formés).

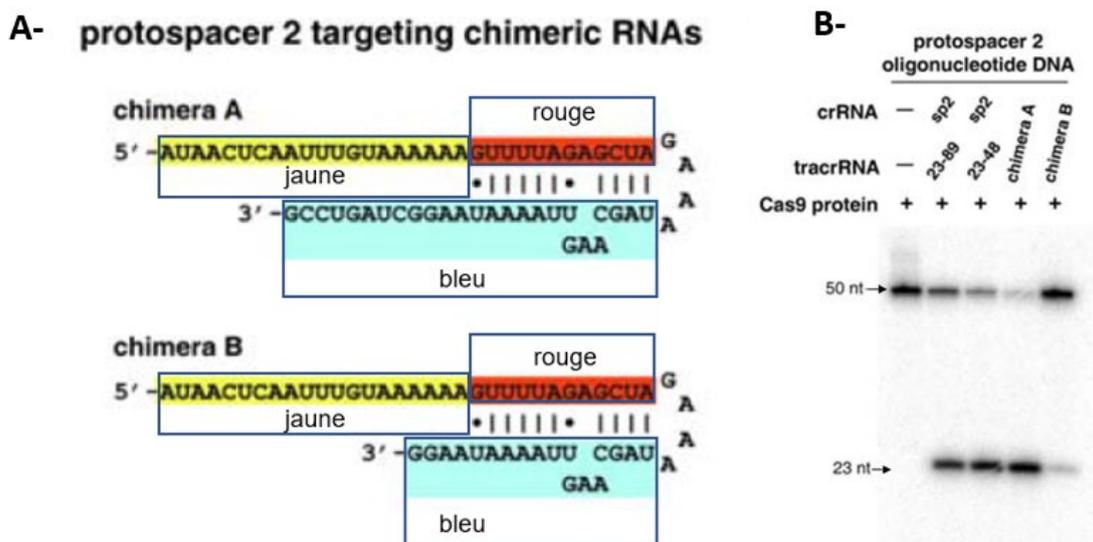
**Figure 6 : Analyse du rôle de l'ARN tracrRNA.** Des expériences de retard de migration sur gel d'un ADN double brin marqué radioactivement correspondant au protospacer 4 ont été réalisées en présence d'une quantité croissante de protéine Cas9 doublement mutée D10A et H840A (de 1nM à 1mM), en absence de tout ARN, en présence de l'ARN crRNA-sp4, de cet ARN crRNA-sp4 et du tracrRNA ou du tracrRNA seul. Les flèches indiquent l'identité des différentes bandes : l'ADN double brin du protospacer 4 libre (unbound target duplex) et ce même ADN dont les brins se sont dissociés (dissociated target DNA strand), un complexe correspondant à une fixation non spécifique (nonspecific binding) sur le protospacer 4 de la protéine Cas9, un complexe correspondant à la fixation spécifique de la protéine Cas9 sur le protospacer 4 (bound target).



**Question 13 :** Pourquoi les auteurs ont-ils réalisé cette expérience avec la protéine Cas9 doublement mutée en D10A et H840A ? Analysez les résultats obtenus, concluez notamment sur le(s) rôle(s) du tracrRNA.

Les auteurs ont ensuite réalisé des ARN chimériques comprenant des séquences issues de crRNA (en « jaune » et « rouge » sur la Figure 7) et des séquences issues du tracrRNA (en « bleu » sur la Figure 7) et testé leur capacité à remplacer les crRNA et tracrRNA dans le système CRISPR-Cas9.

**Figure 7 : Utilisation d'ARN chimériques crRNA-tracrRNA en association avec la protéine Cas9.** A- Séquence des ARN chimériques (à comparer avec la figure 5A). Les séquences du crRNA-sp2 complémentaires du protospacer 2 sont surlignées/encadrées en « jaune » et celles complémentaires du tracrRNA en « rouge ». Les séquences issues du tracrRNA ont été surlignées/encadrées en « bleu ». B- L'activité de la protéine Cas 9 sur l'ADN double brin du protospacer 2 en présence de crRNA-sp2 et d'une portion de tracrRNA (séquence 23-89 ou 23-48), ou en présence des ARN chimériques A ou B représentés en A- ont été analysés selon le même protocole que celui décrit dans la figure 5B (seul le brin complémentaire est analysé). Les positions des nucléotides 23, 48 et 89 sur le tracrRNA sont indiqués sur la figure 5A.



**Question 14 :** Analysez cette figure. Quel est l'intérêt de réaliser des ARN chimériques ?

**Question 15 :** La protéine Cas9 portant les mutations D10A et H840A est appelée dCas9 (pour deadCas9). Cette protéine Cas9, portant ou non des modifications complémentaires, est de plus en plus utilisée pour diverses applications en génie génétique. Quelle activité intéressante cette protéine conserve t'elle malgré la mutation de ses deux domaines endonucléases ? Quels sont les limites de l'utilisation de la protéine Cas9 non mutée qui pourraient justifier l'utilisation de cette dCas? Quelles sont les utilisations que vous pourriez imaginer pour cette dCas9 ?

## **Conclusion de la partie 1 : retour sur l'origine du système CRISPR-Cas**

**Question 16 :** Résumez par un schéma ce qui se passe au niveau du locus CRISPR la première et la deuxième fois qu'une bactérie rencontre un même phage.

**Question 17 :** Ce système est souvent décrit comme un « système immunitaire » bactérien, quelles sont les points communs et les différences entre ce système et ce que vous connaissez du système immunitaire chez l'Homme ?

## **Partie 2 : Utilisation du système CRISPR-Cas en cellule Eucaryote**

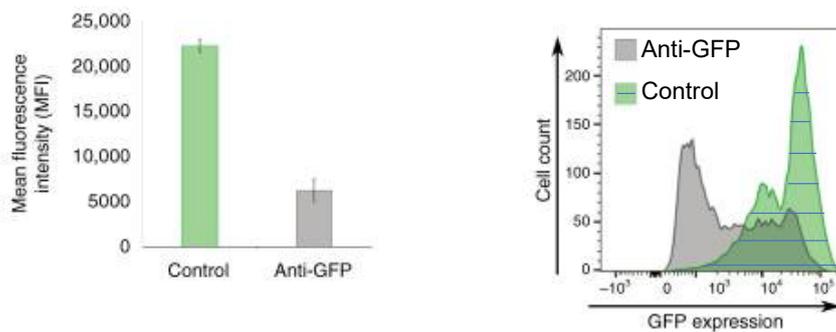
Dès la caractérisation du système, les personnes qui se sont vu attribuer le prix Nobel ont initié le développement de ciseaux moléculaires reposant sur l'utilisation de la protéine Cas9 dirigée vers sa cible par un petit ARN modifié, l'ADN chimérique vu précédemment que nous appellerons maintenant ARN guide.

L'une des difficultés pour utiliser les systèmes CRISPR-Cas dans différents organismes, notamment Eucaryotes, est de faire parvenir la protéine Cas appropriée et le(s) ARN guide(s) à l'intérieur des cellules dont on souhaite modifier le génome. Différents systèmes ont été mis au point, dont les Nanoblades auxquels nous allons nous intéresser maintenant. Les scientifiques ont mis au point ce système dérivé d'un lentivirus pour réaliser efficacement la « livraison » du système CRISPR-Cas comprenant une protéine Cas et un ou plusieurs ARN guide(s) dans des cellules de Mammifères connues pour être difficiles à modifier. Dans leur publication de 2019 dont sont extraites les données suivantes, l'efficacité de leur système de « livraison » dans différents types cellulaires est évaluée, notamment dans des cellules de moelle osseuse de Souris, qu'il est possible de différencier *in vitro* en différentes cellules du système immunitaire (comme des macrophages et des cellules dendritiques). Des cellules de moelle osseuse issues de souris transgéniques exprimant une protéine fluorescente verte, la GFP, ont été traitées avec des Nanoblades comprenant la protéine Cas9 et un ARN guide ciblant le gène codant la GFP.

**Question 18 :** Quel est l'intérêt d'avoir utilisé les cellules de ces souris en particulier ?

La différenciation des cellules de moelle osseuse en macrophages a ensuite été induite et les macrophages obtenus ont été analysés. La cytométrie en flux est une technique qui permet de mesurer individuellement la fluorescence de plusieurs milliers de cellules. Les résultats obtenus sont présentés sous deux représentations graphiques dans la figure ci-dessous.

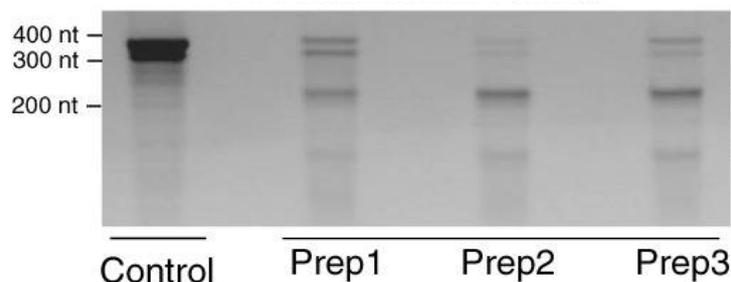
**Figure 5 : Fluorescence de macrophages issus de souris GFP après traitement des cellules de moelle osseuse par des Nanoblades ciblant le gène codant la GFP.** Le panneau de gauche présente l'intensité moyenne de fluorescence mesurée sur des macrophages non traités et sur des macrophages traités avec le système Cas dirigé contre le gène codant la GFP (Anti-GFP). Le panneau de droite représente l'intensité de fluorescence en abscisses, et en ordonnées le nombre de cellules présentant ce niveau de fluorescence. Note : une intensité de fluorescence inférieure à  $10^3$  correspond à un niveau d'autofluorescence des cellules indépendant de la GFP.



**Question 19 :** Interprétez ces deux figures, en quantifiant si possible l'effet obtenu et en faisant ressortir les informations différentes qu'elles vous apportent.

L'effet du traitement CRISPR-Cas9 sur l'ADN des cellules a été étudié avec une technique dite de test à l'endonucléase T7.

**Figure 6 : Analyse de l'efficacité de la modification des macrophages par les Nanoblades par test à l'endonucléase T7.** Pour réaliser cette analyse, de l'ADN génomique de cellules traitées et non traitées par les Nanoblades en quantités égales a été extrait, et une région de 400 paires de bases du gène codant la GFP autour du site de liaison de l'ARN guide a été amplifiée par la technique de PCR (Polymérisation en Chaîne) afin de disposer de suffisamment d'ADN pour pouvoir le visualiser sur un gel d'agarose. Cette technique amplifie avec la même efficacité les séquences originales et les séquences portant d'éventuelles mutations causées par le traitement CRISPR-Cas. L'ADN ainsi obtenu est dénaturé (séparation des brins) et réapparié au hasard puis soumis au traitement par l'endonucléase T7 qui coupera tout ADN non correctement apparié. La piste « Control » correspond à un traitement CRISPR-Cas9 avec un ARN guide dont la cible est en dehors du gène codant la GFP et Prep1, 2 et 3 à trois expériences différentes. L'ADN a été analysé par électrophorèse sur gel d'agarose.



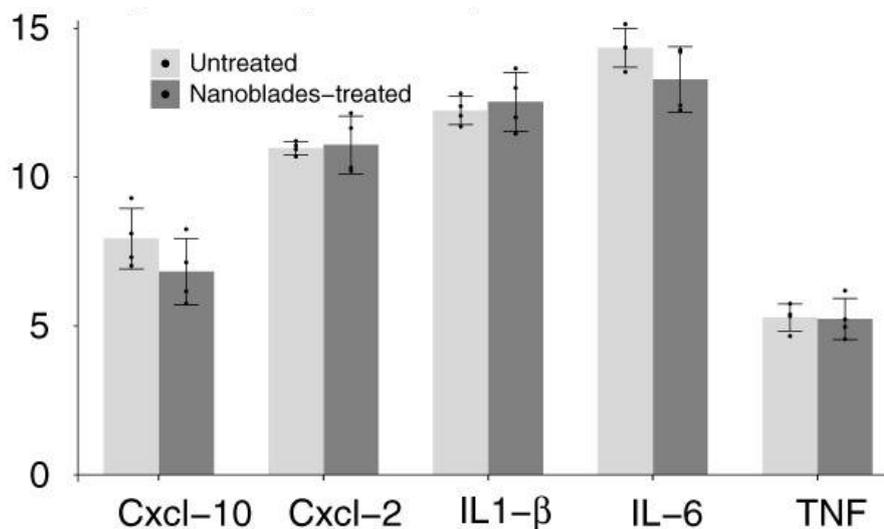
**Question 20 :** Représentez schématiquement les différentes étapes de ce protocole. Positionnez notamment les séquences des amorces de la PCR (petits fragments d'ADN d'une vingtaine de nucléotides qui délimitent la zone qui sera amplifiée) par rapport à la séquence de l'ARN guide.

**Question 21 :** La proportion de bandes de taille inférieure à 300 a été estimée à 60-65% de la quantité totale d'ADN obtenue. A quoi correspondent-elles ? Quelle hypothèse pouvez-vous faire pour expliquer que l'action de la protéine Cas9 sur une même séquence puisse générer des modifications différentes ? Analysez les résultats obtenus sans tenir compte de la présence de deux bandes pour la séquence non traitée.

**Question 22 :** Quelle(s) autre(s) technique(s) plus précise(s) aurai(en)t pu être utilisée(s) pour analyser les modifications de l'ADN dans les cellules ? Citez les avantages/inconvénients respectifs de ces techniques et de celle employée ici.

Les macrophages synthétisent différentes molécules impliquées dans la réponse immunitaire (des cytokines) lorsqu'ils sont mis en présence d'une substance d'origine bactérienne, le LipoPolySaccharide ou LPS. La capacité des macrophages traités ou non par les Nanoblades à répondre au LPS a été étudiée.

**Figure 7 :** Niveau d'expression des quelques cytokines par les macrophages modifiés par les Nanoblades (Nanoblades-treated) ou non modifiés (Untreated) après traitement au LPS. Le niveau d'expression des différentes cytokines indiquées en abscisses est exprimé en multiple leur niveau d'expression dans des macrophages non stimulés par le LPS. Les histogrammes représentent la moyenne obtenue à partir de 4 expériences indépendantes, et les barres d'erreurs représentent les écarts-types.



**Question 23 :** Analysez cette figure. Que nous apprend-elle sur les cellules modifiées par le système CRISPR-Cas9 ?

Comme illustré dans cet exemple sur des cellules immunitaires isolées, la technique CRISPR-Cas9 peut être utilisée pour modifier le génome de cellules et maintenant d'animaux de laboratoire. Une utilisation à des fins thérapeutiques est aussi envisagée chez l'Homme.

**Question 24 :** Quelles sont selon vous et/ou à votre connaissance les utilisations possibles de ce système dans différentes pathologies et les difficultés/risques d'utiliser une telle approche chez l'Homme ?

**Bibliographie :**

CRISPR provides acquired resistance against viruses in prokaryotes. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. Science. 2007 Mar 23;315(5819):1709-12.

A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. Science. 2012 Aug 17;337(6096):816-21.

Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. Mangeot PE, Risson V, Fusil F, Marnef A, Laurent E, Blin J, Mournetas V, Massouridès E, Sohier TJM, Corbin A, Aubé F, Teixeira M, Pinset C, Schaeffer L, Legube G, Cosset FL, Verhoeven E, Ohlmann T, Ricci EP. Nat Commun. 2019 Jan 3;10(1):45.