TP Bioinformatiquen°2:

Clonage du gène de la peroxydase de radis

# Pour permettre une production industrielle de la peroxydase, il est envisagé d’extraire le gène de la peroxydase de l’ADN végétal et l’insérer dans une cellule hôte qui peut se multiplier rapidement et donc produire en grande quantité l’enzyme.

Pour atteindre cet objectif de production, il est nécessaire de réaliser un plasmide recombinant à intégrer dans la cellule hôte.

Pour ce faire, il est nécessaire :

* d’obtenir la séquence codante du gène de la peroxydase
* de choisir les outils pour insérer le gène d’intérêt dans le vecteur d’expression

# **Obtention de la séquence du gène de la peroxydase**

Nous allons nous concentrer sur la peroxydase P7 issue du navet. Des bases de données en libre accès permettent d’obtenir les séquences de nombreux gènes.

* Aller sur le site <https://www.ncbi.nlm.nih.gov/gene/>
* Rechercher gene « peroxidase P7 *Brassica rapa* » comme critère de recherche

dans “[Genomic regions, transcripts, and products](https://www.ncbi.nlm.nih.gov/gene/?term=peroxidase+P7+brassica+rapa)”,



* Dans cette fenêtre, sous “Genes, NCBI [...], cliquer sur un rectangle vert contenant une flèche. D’autres lignes apparaissent en-dessous, une bleue et une rouge. Placer votre curseur de souris sur chacune des lignes.
* Vous trouverez des éléments de vocabulaire sur le site suivant :

<http://www.scfbio-iitd.res.in/contact/aboutus.htm>

1. A quoi correspond chacune de ces 3 lignes de rectangles vert, bleu et rouge ?
2. Rappeler à quoi correspondent chaque lettre de la séquence et à quoi correspond l’association de trois lettres.
3. Combien de nucléotides composent le gène de la peroxydase P7 ?
4. Combien d’acides aminés composent la peroxydase P7 (TD n°1) ?
5. Est-ce que tous ces nucléotides vont coder un acide aminé ?
6. A l’aide des informations données dans la partie “[Genomic regions, transcripts, and products](https://www.ncbi.nlm.nih.gov/gene/?term=peroxidase+P7+brassica+rapa)”, déterminer le nombre d’introns et le nombre d’exons pour ce gène.
7. Réaliser un schéma du gène avec les séquences codantes et non codantes.
* En haut à droite de cette fenêtre vous trouverez : “Go to nucleotide”, cliquer sur “FASTA”.
* Copier-coller la séquence de nucléotides dans un fichier de traitement de texte. Éliminer les sauts de ligne.
* Utiliser l’outil statistiques du logiciel traitement de textes pour comptabiliser le nombre de nucléotides de cette séquence.
1. Combien de nucléotides composent cette séquence ? A quoi correspond-elle sur votre schéma du gène ?

Les séquences trouvées sur NCBI ne correspondent pas à la portion du gène qui sera traduite en protéine et que l’on nomme CDS (CoDing Sequence) qui se situe dans un ORF pour Open Reading Frame ou cadre ouvert de lecture.

* Sur la page NCBI, en plaçant le curseur de la souris sur la ligne rouge, une fenêtre apparaît. Cliquer sur “Blast genomic”. Une nouvelle page s’ouvre, cliquer tout en bas à gauche sur Blast. Dans une nouvelle fenêtre apparaîtra une liste de séquences, choisir :

|  |
| --- |
| **Select seq XM\_009127350.1**[PREDICTED: Brassica rapa peroxidase P7 (LOC103850580), mRNA](https://blast.ncbi.nlm.nih.gov/Blast.cgi#alnHdr_685269168) |

* Cliquer sur Sequence ID : [XM\_009127350.1](https://www.ncbi.nlm.nih.gov/nucleotide/XM_009127350.1?report=genbank&log$=nuclalign&blast_rank=5&RID=DAAKVJ7M015)
* Vous trouverez sur cette page la séquence d’ADN correspondant à l’ARNm. Si vous cliquez sur “Fasta” en haut à gauche, sur cette nouvelle page, vous pourrez copier-coller la séquence sans les numéros sur la gauche.
* Copier-coller la séquence dans votre fichier texte. Éliminer les sauts de ligne si vous en avez.

Nous vous proposons à présent d’aligner la séquence du gène et de l’ARNm :

* Aller sur <https://www.ebi.ac.uk/Tools/msa/clustalo/>
* Dans le menu déroulant, choisir DNA.
* Copier-coller les séquences du gène et de l’ARNm en plaçant au-dessus “>gène”pour la séquence du gène et ou“>ARNm” en début de chaque séquence. Réaliser l’alignement des séquences en cliquant sur SUBMIT. Copier cet alignement dans votre fichier texte.
1. Qu’observez-vous ? Est-ce cohérent avec les observations précédentes ?

Afin de repérer l’ORF dans la séquence d’ARNm copiée précédemment, nous vous proposons de traduire cette séquence.

* Aller sur <https://web.expasy.org/translate/>
* Copier la séquence d’ARNm et cliquer sur “translate sequence”.
1. Sur le site de traduction, vous obtenez plusieurs résultats, pourquoi ? Lequel correspond à la protéine peroxydase P7 de *Brassica rapa* ? Justifiez.
2. Indiquer s’il est commun de trouver une méthionine en début de protéine traduite. A quel codon cela correspond-il ?

Pour localiser la séquence codante sur le gène, nous vous proposons de rechercher et marquer les séquences permettant d’obtenir les premiers acides aminés : ATGGCTTCA et les derniers acides aminés de la protéine : AAAACAAACTGA.

* Sur votre traitement de texte, avec l’outil rechercher, localiser les séquences ATGGCTTCA et AAAACAAACTGA dans le gène et l’ARNm.
* Surligner dans les séquences de nucléotides tous les nucléotides appartenant à la séquence codante.
1. Combien de nucléotides composent la séquence codante dans la séquence de l’ARNm ? Est-ce cohérent avec le nombre d’acides aminés de la peroxydase déterminé en Q4 ? Comment expliquez-vous la différence ?

# **Construction du vecteur d’expression**

Nous souhaitons utiliser le vecteur PURExpr\_DHFR\_ctrl pour produire un plasmide recombinant contenant le gène de la peroxydase P7 de *Brassica rapa*. Pour ce faire, il faut intégrer le gène dans le plasmide.

1. Réaliser un schéma expliquant les étapes nécessaires à cette intégration.
* Ouvrir le lien suivant : <http://tools.neb.com/NEBcutter2/index.php>
* Copier-coller la séquence de nucléotides obtenue complète du gène de la peroxydase P7.
* Cliquer sur “submit”. Vous obtenez une carte de restriction avec un grand nombre d’enzymes capables de couper le fragment d’ADN à un endroit spécifique.
* Parmi les enzymes de restriction proposées, sélectionner les enzymes qui possèdent un site de restriction en amont et en aval de la séquence codante déterminée précédemment.
* Cliquer sur “custom digest”, cocher les enzymes de restriction choisies dans la liste.
* Cliquer sur “Enzymes & sites”
1. Comment coupe l’enzyme BmrI ? comment appelle-t-on ce type de coupure ?
2. Comment coupe l’enzyme HpaI ? Comment appelle-t-on ce type de coupure ?
* Cliquer sur View gel.
1. Quelle technique permet d’obtenir le résultat schématisé apparaissant à l’écran ?
2. Quelle est la ou les bandes contenant le fragment d’ADN codant pour notre protéine ?
* Pour vérifier qu’il est possible d’utiliser le plasmide PURExpr\_DHFR\_ctrl, revenir à la page d’accueil (cliquer sur “New DNA” pour travailler sur la séquence du plasmide. Pour cela, il faut le choisir dans le menu déroulant sur la droite de la fenêtre et soumettre (“submit”). En réalisant de nouveau l’action “custom digest” avec les enzymes de restriction choisies pour le gène d’intérêt, vérifier que des sites de restriction BmrI et HpaI sont présents sur le plasmide.
1. A votre avis, la construction plasmidique envisagée est-elle possible ? Justifier.
2. Donner les caractéristiques de ce plasmide.
3. Proposer un protocole permettant de mettre en évidence la bonne insertion du gène d’intérêt dans le vecteur puis du vecteur dans la bactérie hôte.