



# Bioinformatique en biologie moléculaire



## Travaux des Actions Académiques Mutualisées

Niveau

Terminale STL Biotechnologie

Thème du programme

- enseignement de Biotechnologies:
- **Initiation à la biologie moléculaire et au génie génétique** (rechercher et exploiter des informations dans des banques de données; déterminer la taille de fragments de restriction par une digestion virtuelle à l'aide d'un logiciel de bio-informatique adapté)

Situations pédagogiques

- Analyse d'un gène humain et clonage dans un vecteur d'expression au sein d'un laboratoire de recherche

Liens internet

- <http://www.ncbi.nlm.nih.gov/gene/>
- <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- <http://tools.neb.com/NEBcutter2/index.php>

Compétences B2i

- **Domaine 4 : Acquérir, transformer, produire de l'information**  
**utiliser les outils et logiciels adaptés à un projet de production**

Matériels TICE

- Un poste PC par binôme
- Une connexion internet
- Logiciel de traitement de texte et d'images

Mots clés

- Gène, séquence, clonage, plasmide



*Votre avis nous intéresse, merci de répondre à notre enquête concernant ce scénario.*

Elève, cliquer [ici](#).

Professeur, cliquer [ici](#).



## CONTEXTE

*La connexine 43 (Cx43) est une protéine impliquée dans la formation des Jonctions Communicantes ou Gap Junctions. Ces jonctions permettent une communication directe entre deux cellules adjacentes.*

*La connexine se retrouve en particulier au niveau du cœur et des testicules. Une diminution de sa quantité dans les cellules cardiaques est une cause de cardiomyopathie. La surexpression de cette protéine serait liée à des cancers des testicules. Elle serait également impliquée dans des cancers de la prostate.*

***Afin d'étudier les rôles de cette protéine, un modèle animal surexprimant cette protéine doit être créé.***

*Dans un laboratoire de recherche, un étudiant récupère le plasmide pcDNA3.1-Cx43 d'un autre laboratoire. Il correspond à un plasmide appelé pcDNA3.1 dans lequel a été intégré le gène codant la connexine 43, noté Cx43. Ce plasmide doit être introduit à l'intérieur de cellules embryonnaires de rat. Cependant un manque d'information fait que l'étudiant ignore si le gène intégré correspond au gène d'origine humaine ou de rat.*

***Pour répondre à cette question, l'étudiant va, dans un premier temps, rechercher les différences entre les deux séquences du gène Cx43 puis déterminer les profils attendus pour chaque gène après digestion enzymatique, un moyen rapide d'analyser la séquence d'un gène.***



## Activité n°1 Récupération des séquences géniques

### Objectifs

- Utiliser une méta-base de données professionnelle pour obtenir des séquences géniques



### Consignes

*Afin de répondre à cette question nous allons exploiter les nombreux outils d'analyse disponibles sur Internet.*

1-Dans un premier temps, pour récupérer les séquences des gènes humain et de rat de la Cx43 , suivre le lien suivant : <http://www.ncbi.nlm.nih.gov/gene/>

*Le site NCBI est une métabase de données regroupant de nombreuses informations scientifiques dont des séquences de gènes, de protéines, des références d'articles...*

2- Utiliser « *connexin 43* » comme critère de recherche puis cliquer sur les liens  
*GJA1 – gap junction protein, alpha 1, 43kDa [Homo sapiens]* pour le gène humain  
*Gja1 – gap junction protein, alpha 1 [Rattus norvegicus]* pour le gène de rat

3-Trouver la séquence du gène en cliquant sur le lien *go to nucleotide FASTA*

4-Identifier le chromosome porteur de la séquence du gène dans chacun des cas

5-Copier-coller dans un fichier Word les deux séquences en les identifiant clairement et rappeler à quoi correspondent chaque lettre dans ces séquences et à quoi correspond l'association de trois lettres



## Activité n°2 Identification des cadres ouverts de lecture

### Objectifs

-Identifier un cadre ouvert de lecture dans une séquence génique



### Consignes

*Les séquences trouvées sur NCBI ne correspondent pas à la portion du gène qui sera traduite en protéine et que l'on nomme ORF pour Open Reading Frame ou cadre ouvert de lecture*

1-Afin de repérer l'ORF dans chacune des deux séquences d'ADN, chercher et marquer la séquence suivante correspondant au début de la protéine **ATGGGTGACT**

2-Indiquer s'il est commun de trouver les trois premières lettres ATG au début de la séquence d'un gène. Donner le nom du codon correspondant.

3-Enfin, chercher et marquer les séquences terminatrices qui sont respectivement pour la séquence de rat **GAGATTTAA** et pour la séquence humaine **GAGATCTAG**

4-A l'aide du code génétique, indiquer à quel codon correspondent les trois dernières lettres de chaque séquence

**Remarque** : pour les recherches de séquence, faire bien attention à supprimer les espaces

**Les portions de séquences ainsi délimitées correspondent aux ORF du gène de Cx43 que nous allons utiliser dans la suite du travail et qui ont potentiellement été intégrées dans le plasmide pcDNA3.1. Vous pouvez donc les copier-coller plus loin dans le document word en les identifiant clairement.**



### Activité 3: Différences entre connexine-43 humaine et connexine-43 de rat

#### Objectifs

- Comparer des séquences géniques
- Calculer des pourcentages d'homologies



#### Consignes

Afin de vérifier la similarité entre ces séquences, nous allons utiliser l'application BLAST. BLAST est un logiciel en ligne permettant d'aligner une séquence donnée avec toute une banque de données de séquences c'est-à-dire un ensemble de séquences déjà identifiées. Pour cela il suffit de donner la séquence à aligner au logiciel et de lancer l'alignement. Les résultats se présentent sous la forme de barre de couleur. La couleur rouge indique un fort pourcentage d'identité. Il est possible de savoir d'où proviennent les séquences qui sont le plus homologues en cliquant sur la séquence correspondante.

1-Ouvrir le lien suivant

<http://blast.ncbi.nlm.nih.gov/>

2-Effectuer la comparaison des deux ORFs et donner le pourcentage d'homologie au niveau de la séquence génique

Grâce à un logiciel permettant de traduire la séquence nucléotidique en séquence protéique nous avons pu comparer les versions humaine et de rat dont les résultats sont les suivants :

*Alignement comparatif (en noir : séquence humaine, en bleu : séquence du rat)*

```

MGDWSALGKLLDKVQAYSTAGGKVWLSVLFIFRIILLGTAVESAWGDEQSAFRCNTQQPGCENVCYDKSFPI SHVRFWVLQII
MGDWSALGKLLDKVQAYSTAGGKVWLSVLFIFRIILLGTAVESAWGDEQSAFRCNTQQPGCENVCYDKSFPI SHVRFWVLQII
FVSVPTLLYL AHV FVYMRKEEKL NKKEEELKVAQTDGVNVDMHLKQIEIKKFKYGI EEHGKVKMRGGLLRTYIISILFKSIFE
FVSVPTLLYL AHV FVYMRKEEKL NKKEEELKVAQTDGVNVEMHLKQIEIKKFKYGI EEHGKVKMRGGLLRTYIISILFKSVFE
VAFLLIQWYIYGFSLSAVYTCKRDPCHQVDCFLSRPTEKTI FIFMLVSVLSLALNIELFYVFFKGVKDRVKGKSDPYHA
VAFLLIQWYIYGFSLSAVYTCKRDPCHQVDCFLSRPTEKTI FIFMLVSVLSLALNIELFYVFFKGVKDRVKGRSDPYHA
TS GALSPAKDCGSQKYAYFNGCSSPTAPLSPMSPPGYKLVGTDRNNSSCRNYNKQASEQNWANYSAEQNRMGQAGSTISNSHA
TTGPLSPSKDCGSPKYAYFNGCSSPTAPLSPMSPPGYKLVGTDRNNSSCRNYNKQASEQNWANYSAEQNRMGQAGSTISNSHA
QPFDFPDDNQNSKKLAAGHELQPLAIVDQRPSSRASSRASSRPRPDDLEI*
QPFDFPDDNQNAKKVAAGHELQPLAIVDQRPSSRASSRASSRPRPDDLEI*
  
```

3-Indiquer à quoi correspondent les lettres dans ces séquences

4-A l'aide des différences mises en évidence en rouge, calculer le pourcentage de similarité (il est possible d'utiliser l'outil statistique de Word pour connaître le nombre total de lettre de la séquence)

5-Ce pourcentage de similarité vous paraît-il cohérent avec une modification ou une absence de modification de la fonctionnalité protéique, les réponses possibles étant oui/non/ça dépend !

À justifier bien entendu...



## Activité n° 4 Profil de restriction et identification de l'insert

### Objectifs

- Simuler des digestions enzymatiques
- Comparer des profils de restriction



### Consignes

*Pour en revenir au but principal de notre travail, intéressons nous de nouveau aux séquences ORF potentiellement présentes dans le plasmide pcDNA3.1.*

*L'insertion de la séquence ORF a été réalisée au niveau d'un site de coupure unique EcoR I dans le plasmide. Cela signifie que la digestion du plasmide pcDNA3.1-Cx43 par cette enzyme EcoR I permet de « ressortir » l'insert correspondant à l'ORF de Cx43.*

*Grâce à une comparaison des profils de restrictions, nous allons tenter de différencier les résultats possibles pour l'ORF d'origine humaine ou de rat.*

1-Ouvrir le lien suivant

<http://tools.neb.com/NEBcutter2/index.php>

*Cette application permet de simuler la digestion de fragments d'ADN par la plupart des enzymes de restriction connues et/ou commercialisées.*

2-Déterminer l'action des enzymes Hind III, Acu I, Xmn I et EcoR I sur les deux ORF en procédant de la manière suivante :

- coller la séquence en éliminant les espaces et cliquer sur « *submit* »

**Remarque :** *Les résultats affichés sont ceux des enzymes coupant une fois, vous pouvez afficher celles coupant deux ou trois fois avec la fonction « *display* »*

- simuler la coupure par la ou les enzyme(s) de votre choix grâce à « *custom digest* »
- sélectionner la ou les enzyme(s) et cliquer sur « *digest* »
- le lien « *view gel* » représente un résultat expérimental intéressant

3-Indiquer à quel type de technique la fonction « *view gel* » correspond

4-Proposer une expérience simple suffisante pour démontrer que le gène de l'insert est celui de la Cx43 de rat et non de la Cx43 humaine et schématiser les résultats attendus

5-Dans notre analyse des résultats de fragments de restrictions, nous nous sommes focalisés sur la séquence de l'insert. Dans un cadre pratique, aurait-il été nécessaire de prendre en compte la séquence du plasmide ?