

pour le prochain TD ECR.

## UE de Licence « Expression Génique et Régulation » Travaux Dirigés N° 2

Suite au séquençage systématique des génomes, la recherche des gènes codant pour une protéine se fait dans un premier temps par la recherche des ORF.

\_ Comment pourriez-vous définir une ORF ?

Dans la séquence du génome de levure le gène NCE101 code pour une protéine de fonction inconnue (www.yeastgenome.org).

La séquence de la protéine est la suivante :

NH<sub>2</sub>-M<sup>V</sup>Q<sup>Y</sup>Ä<sup>P</sup>F<sup>L</sup>L<sup>G</sup>K<sup>F</sup>S<sup>D</sup>P<sup>L</sup>L<sup>A</sup>I<sup>M</sup>V<sup>G</sup>Ç<sup>L</sup>S<sup>Y</sup>Y<sup>V</sup>Y<sup>È</sup>R<sup>K</sup>M<sup>G</sup>R<sup>P</sup>O<sup>G</sup>H<sup>H</sup>L<sup>H</sup>E<sup>L</sup>I<sup>K</sup>K<sup>R</sup>W<sup>D</sup>D<sup>R</sup>K<sup>·</sup>-COOH

La séquence de la région où se localise le gène est la suivante (orientée 5'→3') :  
(la séquence d'un seul brin est donnée, les valeurs à gauche de la séquence servent uniquement de repères)

```

1  TGCTGCATTA TGTAATTCC TATCAATGAT TAATTATATT ACTCTATATA
51  GGTATGAAGT GCGGAAATAA AAGACTGTTC ACATATTAGA TTATTCAAGA
101 TTACGCGGGC TCTATTTTTT CCTGTTAGCC ACACGCACAC AAGGGGCATA
151 CAAATGTTCA CTCCATTTAT AAGAGGATTC CACACAAGAT CATCGATAAC
201 ATGGTTCAGT ACGCTCCCTT TCTATTAGGA AAGTATGTTT ATTCTTCATA
251 CGCTATTCTA TTTTTTTTTT TTCACACTT ATTCTTTTAC TAACGCATAG
301 TTAACCATTT GCTGTTTCGA AGAGGCAACA TCAAACAAGT CGATGTGCTC
351 TTTCCÇAÄGÇ TGGTCTCTAT TGCAGATTTT CTGATCCGTT ATTGGCAATT
401 ATGGTAGGCT GTCTCAGCTA CTATGTGTAC GAAAGGAAAA TGGGGCGTCC
451 ACAAGGTCAT CATTTACACG AGCTCATAAA GAAGCGATGG GACGATCGCA
501 AATAATAATC CGGCGTGAGG GGAGTATATA TAAGTTTCTG GCGATATGTA
551 AATACCATAA TAATTAATCC TAGGATCTAC CTTTATACGC ATTTATCCTA
601 TAAAAACCCA ATCTTTCAGA TCCACAGTAT TTTTATATGT GTGATTTCTT
651 TGATT

```

TAA  
TAG  
TGA

3 possibilités (bleu et vert)  
et jamais  
le reste :  
séparer tête et pied  
introns : GT-AG.

\_ Localisez la séquence protéique sur la séquence ADN. Vous avez une chance sur deux de faire cet exercice en pure perte, pourquoi?

\_ La séquence ADN présentée est-elle celle du brin codant ?

\_ Calculez approximativement la masse moléculaire théorique de la protéine codée par le gène NCE101 ?

Pour analyser la fonction de ce gène, cette séquence d'ADN va être amplifiée par PCR et clonée dans un plasmide qui servira à transformer une souche haploïde de levure

site donneur dans l'intron (bleu)  
et accepteur (à la fin de l'intron)  
sont conservés



séquençage systématique : toutes les bases, sans exception, sans réfléchir.

ORF: cadre ouvert de lecture : séquence potentiellement codante d'un gène

on cherche un ORF d'un stop à un stop : une raison : tous les CDS ne commencent pas tous par un ATG.  
3x2 cadres de lecture, les 2 amino codants

→ Est-ce le bon codant du gène? ✓

donc matrice / bon codant  
même séquence que  
gen

→ Probablement pas ces petits ORF.

→ le principe de la PCR est d'amplifier la séquence d'un gène : de réaliser des copies multiples de celui-ci.

moins : 5' → 3' 1ère amorce

pour 1<sup>er</sup> bin  
ACGACGVAATACAATAAGG  
TGCATGCAATTAATGTAATTC ✓

pour 2<sup>ème</sup> bin  
UCCUUUAGUCUGUAUAUUUU ✓

AGGAAATCACACATATAAAA ✓

- une copie structurée linéaire autorépliquante d'ADN double brin portant des deux bouts des amorce des la PCR.

réplique à l'identique une séquence d'ADN / réplique à l'identique des cellules.

- clonage : multiplication à l'identique d'une bactérie. Transformation : entrée d'ADN simple brin dans une bactérie (processus naturel ou artificiel) qui va ensuite se recombiner avec des séquences homologues du génome bactérien. (contient cellule et établissement de celui-ci (naturelle ou artificielle) au cours des générations.)

X - autorépliquant, transmissible moyen de sélection.

log d'ampicilline : transformation







- Cela nous permet de ne regarder que la copie qui est sur le plasmide.
- Cela nous permet de dire que le gène n'est pas essentiel à la survie de l'organisme.
- Il y a un intron !! Il faut l'enlever d'abord!

TATA box des 32.

+ 2 promoteurs : promoteurs eucaryotes et procaryotes différents : la TATA box n'est pas située à la même position : - eucaryotes : -35.

- procaryotes : -10.

+ Ribosome box en -35 des procaryotes.

+ terminateur différent!

+ chez eucaryotes, ribosome se fixe sur la coiffe, se déplace 5' à 3' et s'arrête au premier AUG, et initie la traduction.

- chez procaryotes, séquence RBS à côté de l'AUG, homologue de l'ARN 16S de la 0-11 du ribosome. Shine et Dalgarno.

Pour passer un gène des procaryotes aux eucaryotes (ou inversement), on ne garde que la séquence codante, en enlevant (et remplaçant) promoteur et terminateur.