

Licence-Master Bioinformatique
Contrôle continu 06/03/06

Correction

-« Vraies » questions de cours

-« fausses » questions de cours: questions pour voir si pouviez imaginer une réponse crédible qui n'était pas de façon claire dans le cours

-« Vrais » exercices dont la réponse était trouvable si le cours était assimilé ou de logique pure

-« faux » exercices, d'anticipation par rapport au cours à venir, mais faisant appel à des notions de biologie très médiatisées, permettant de voir quel est votre recul maintenant par rapport à ces notions

Thème 1 : Les protéines

On donne la séquence d'une protéine :

Ile-Ala-His-Thr-Tyr-Gly-Pro-Phe-Glu-Ala-Ala-Met-Cys-Lys-Trp-Glu-Glu-Glu-Pro-
Asp-Gly-Met-Glu-Cys-Ala-Phe-His-Arg

Question 1 : Définissez une protéine. Quelle est l'importance de connaître sa séquence ?

Vraie question de cours

Question 2 : Comment peut-on obtenir la séquence d'une protéine ?

On peut faire un séquençage direct par des méthodes chimiques basées sur les propriétés chimiques différentes des acides aminés que l'on peut ainsi identifier: non présenté en cours, deviné par un étudiant.

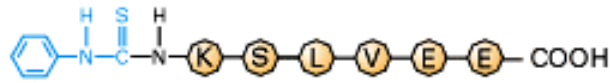
On peut purifier l'ARNm, on le rétrotranscrit en cDNA, on séquence le cDNA, on fait une traduction « in silico » en appliquant le code génétique. Tout le monde a peu près dit cela

MAIS il faut être capable de purifier le bon ARNm, donc de l'identifier. Pour cela

- il faut un indice: un fragment de séquence pour pouvoir faire un Northern-blot.
- On peut aussi faire des banques d'expression

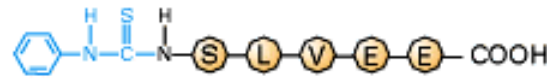
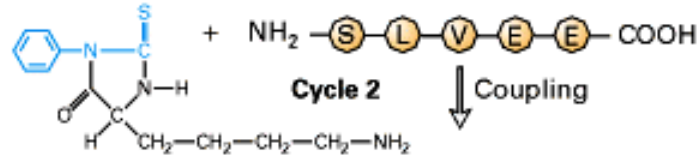


Cycle 1 ↓ Coupling (c1ccc(cc1)N=C=S PITC)



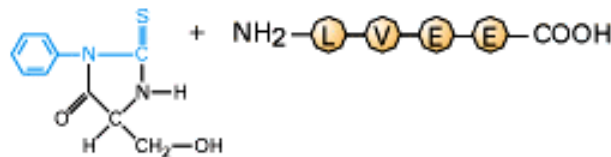
↓ Acid cleavage

PTH lysine



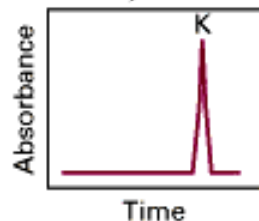
↓ Cleavage

PTH serine

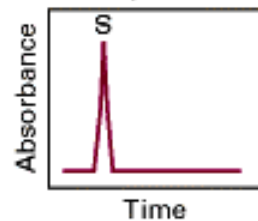


↓

Cycle 1



Cycle 2



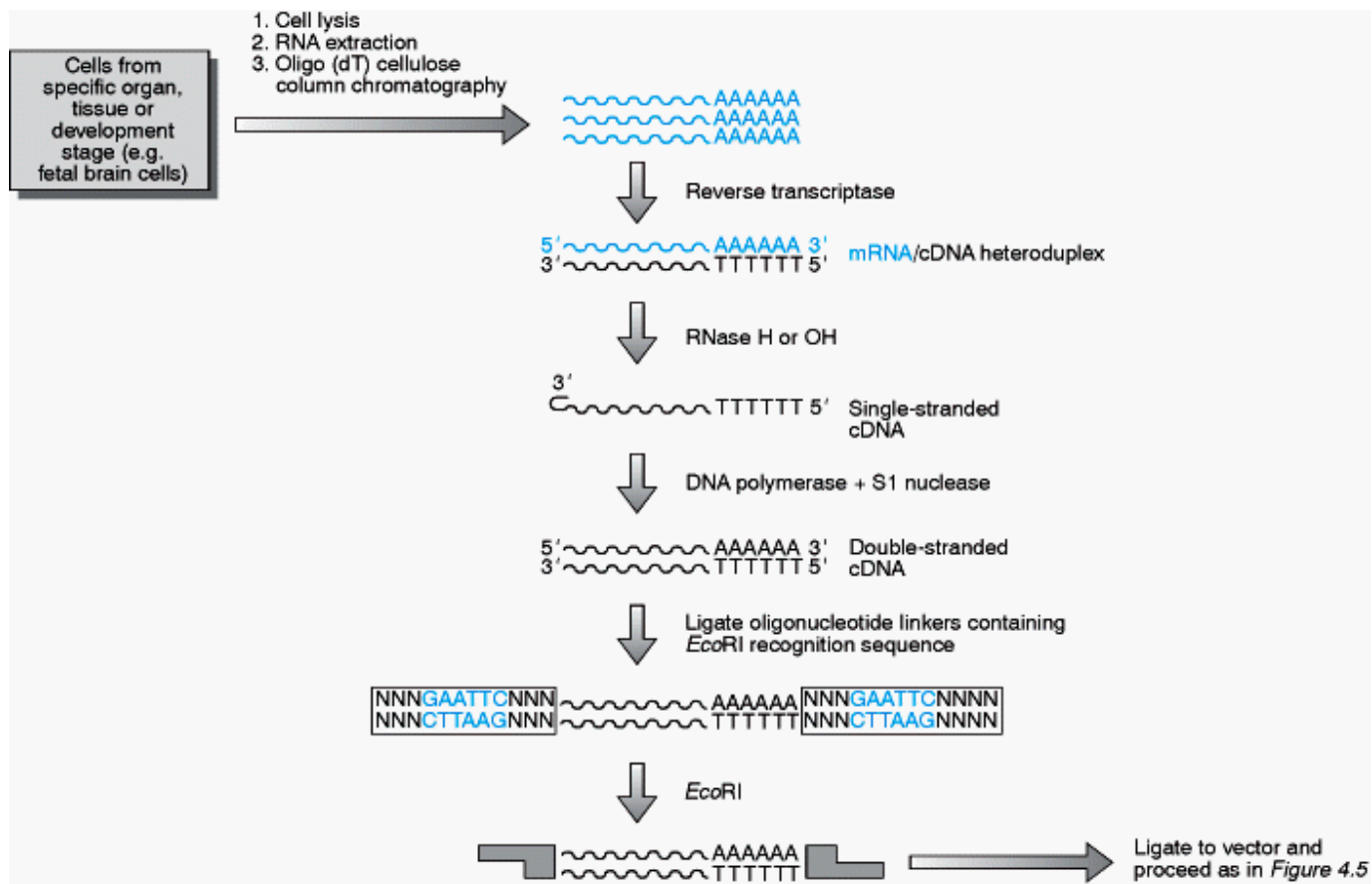
Analysis

Détermination chimique de la séquence d'une protéine par la méthode de la dégradation d'Edman.

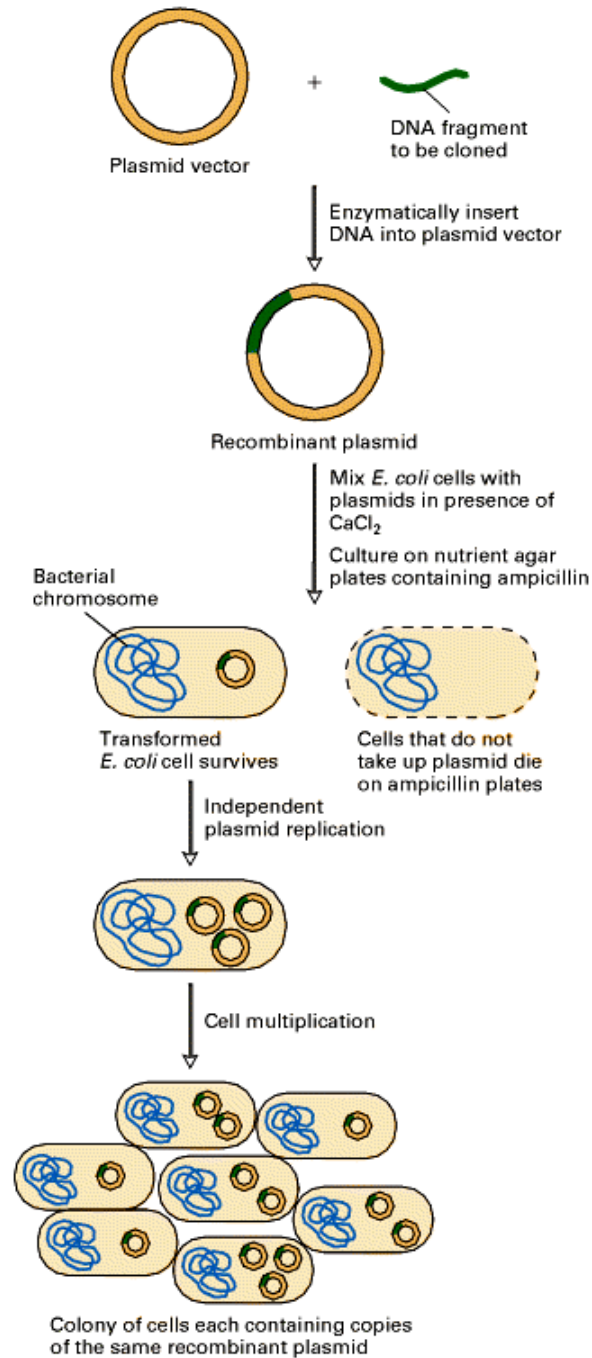
Dans une première étape, on fait réagir l'extrémité N-terminale de la protéine avec du phenylisothiocyanate (PITC), un composé fluorescent. L'extrémité COOH ne peut pas réagir avec ce composé.

Dans une deuxième étape, on fait subir à la protéine marquée une hydrolyse acide légère. Dans cette condition, Le PITC a tendance à se cycliser et ainsi à se cliver du reste de la protéine. Le produit cyclisé s'appelle le phenylthiohydantoin (PTH). L'hydrolyse acide génère donc un PTH associé à l'acide aminé N-terminal et une protéine raccourcie de son acide aminé N-terminal et présentant une nouvelle extrémité N-terminale. L'acide aminé fixé au PTH est ensuite identifié par chromatographie liquide, sur la base de ses propriétés chimiques.

Le restant de la protéine est soumise de nouveau aux deux étapes précédentes. Ainsi les acides aminés N-terminaux sont identifiés successivement générant la séquence de la protéine.



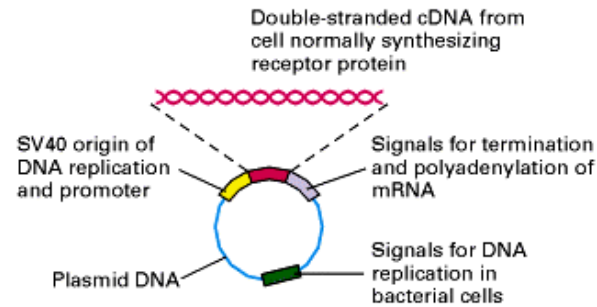
Faire une banque de cDNA. Les ARNm totaux sont extraits et purifiés de cellules en culture ou d'organes. La retrotranscription par la reverse transcriptase (une enzyme virale capable de polymériser de l'ADN à partir d'une matrice ARN) est amorcée à l'aide d'un primer oligo(dT). La RNase H (une enzyme qui dégrade uniquement l'ARN dans les hybrides ADN/ARN) est ensuite utilisée pour éliminer l'ADN. L'extrémité 3' end du cDNA résultant a tendance à former une petite épingle à cheveux; cette petite épingle sert de primer « naturel » pour amorcer la polymérisation du 2^{ème} brin d'ADN. La nucléase S1 permet ensuite de couper la liaison interbrin. Le cDNA double brin ainsi généré est ligué à ses deux extrémités à un linker, c'est-à-dire un petit fragment d'ADN contenant un site de restriction pour EcoRI. Une digestion EcoRI permettra de l'insérer dans un vecteur plasmidique ouvert par EcoRI. Le clonage des bactéries puis le séquençage individuel de chaque insert plasmidique permettra de déterminer les gènes exprimés dans les cellules d'origine.



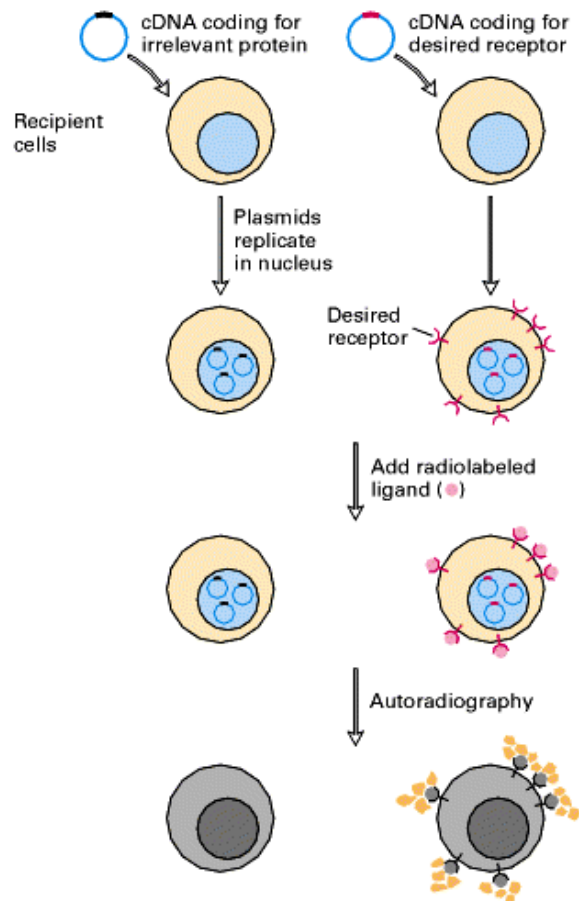
Principe général de clonage d'un fragment d'ADN dans un vecteur plasmidique.

Un plasmide est un petit chromosome bactérien, autonome, qui contient des gènes de résistance aux antibiotiques (non montré). On sait extraire les plasmides des bactéries et les purifier. On y insère in vitro le fragment d'ADN à cloner. On réintroduit le plasmide recombinant dans des bactéries dépourvues de plasmide par transfection. Les bactéries effectivement transfectées sont sélectionnées grâce à leur résistance aux antibiotiques. Les bactéries transfectées sont ensuite mise en culture de façon individuelle. Chaque bactérie se multiplie sous forme d'un clone. Le fragment d'ADN inséré a donc été amplifié.

(a) **Plasmid expression vector**



(b) **Initial screening of cDNA pools**



Identification et isolement d'un cDNA codant une protéine d'intérêt (par exemple un récepteur membranaire) dans une banque d'expression.

Les ARNm totaux sont extraits de cellules qui expriment le récepteur membranaire d'intérêt. Ils sont rétrotranscrits en cDNA.

(a) La totalité des cDNAs sont insérés dans des plasmides modifiés de tel sorte que le cDNA puisse être transcrit et traduit dans des cellules eucaryotes (ajout d'un promoteur eucaryote et de signaux de polyadénylation).

(b) Les plasmides sont transfectés dans une population population de cellules en culture eucaryotes qui n'expriment pas le récepteur d'intérêt. Donc seules les cellules qui ont reçu le cDNA codant pour le récepteur l'expriment. Pour repérer ces cellules, on incube la population cellulaire avec le ligand de ce récepteur marqué radioactivement.

Cette approche peut-être généralisée à l'isolement de cDNA exprimant n'importe quelle protéine, dans la mesure où l'on dispose d'anticorps permettant le repérage de ces protéines.

Question 3 : Quelle partie de cette protéine vous attendez-vous à trouver sous la configuration d'une hélice alpha ?

Présence relative des acides aminés dans les différentes structures secondaires

Acide aminé	Hélice alpha	Feuillet beta	Tour
Ala	1.29	0.90	0.78
Cys	1.11	0.74	0.80
Leu	1.30	1.02	0.59
Met	1.47	0.97	0.39
Glu	1.44	0.75	1.00
Gln	1.27	0.80	0.97
His	1.22	1.08	0.69
Lys	1.23	0.77	0.96
Val	0.91	1.49	0.47
Ile	0.97	1.45	0.51
Phe	1.07	1.32	0.58
Tyr	0.72	1.25	1.05
Trp	0.99	1.14	0.75
Thr	0.82	1.21	1.03
Gly	0.56	0.92	1.64
Ser	0.82	0.95	1.33
Asp	1.04	0.72	1.41
Asn	0.90	0.76	1.28
Pro	0.52	0.64	1.91
Arg	0.96	0.99	0.88

La connaissance de ces fréquences permet de faire de la prédiction de structure secondaire à partir de la séquence

Ile - Ala - His - Thr - Tyr - Gly - Pro - Phe - Glu - Ala - Ala - Met -
0.97 1.29 1.22 0.82 0.72 0.56 0.52 1.07 1.44 1.29 1.29 1.47

Cys - Lys - Trp - Glu - Glu - Glu - Pro - Asp - Gly - Met - Glu - Cys -
1.11 1.23 0.99 1.44 1.44 1.44 0.52 1.04 0.56 1.47 1.44 1.11

Ala - Phe - His - Arg
1.29 1.07 1.22 0.96

Intérêt:

- Prémices de comment construire un logiciel de prédiction
- Longueur minimale? Voir dans ce qui est connu
- Voir si hélices connues ayant un Trp au milieu

Thème 2 : ADN

On isole le matériel génétique d'un virus. Ce matériel génétique est composé d'une molécule d'ADN.

Question 4 : Expliquez ce qu'est un virus

Vu très rapidement en cours

Nouveaux éléments dans la suite

On isole le matériel génétique d'un virus. Ce matériel génétique est composé d'une molécule d'ADN.

Lors d'une première expérience, on digère cet ADN du virus par l'enzyme de restriction EcoRI. Le produit de digestion est mis à migrer sur un gel d'électrophorèse. La migration achevée, on observe une bande de masse moléculaire $3,4 \cdot 10^6$.

Question 6 : Quelle est la structure générale de cette molécule d'ADN ?

Circulaire

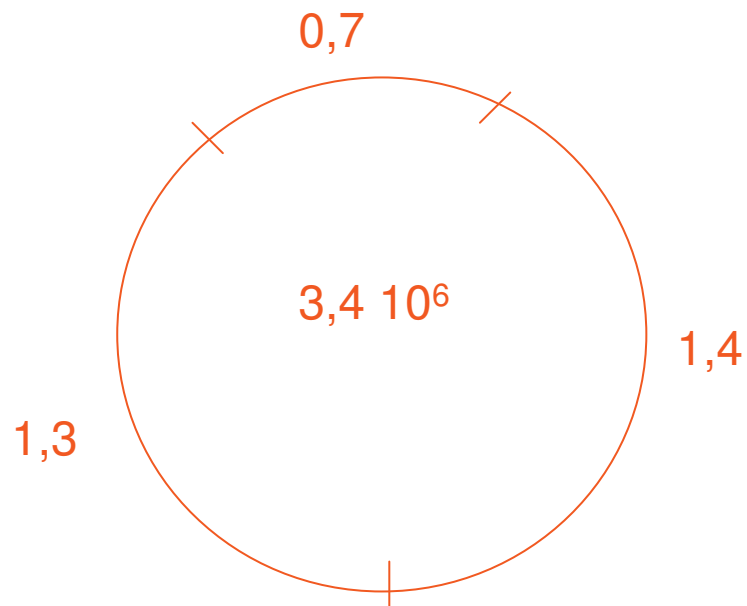
Nombreuses réponses: pas de site, sous-entendu linéaire MAIS

Question 8 : Comment pourrait-on faire pour positionner le site EcoRI par rapport aux trois sites HpaI ?

Lors d'une deuxième expérience on digère cet ADN par une autre enzyme de restriction, HpaI, on fait migrer les produits de digestion sur gel d'électrophorèse, et on observe 3 bandes de masse moléculaire $1,4 \cdot 10^6$, $0,7 \cdot 10^6$ et $1,3 \cdot 10^6$

Question 7 : Représenter schématiquement la position des trois sites de restriction sur l'ADN viral

$$1,4 \cdot 10^6 + 0,7 \cdot 10^6 + 1,3 \cdot 10^6 = 3,4 \cdot 10^6$$



Positions relatives

Question 8 : Comment pourrait-on faire pour positionner le site EcoRI par rapport aux trois sites HpaI ?

On fait les 2 digestions, une des trois bandes doit être fragmentée. Le site EcoRI est donc dans cette région. La taille des fragments indique la proximité ou pas des sites

Sur ce principe, on établit des cartes de restriction qui peuvent ensuite

- servir de marqueurs lors d'un assemblage de fragments séquencés
- Servir de sites de clonage
- ...

On connaît une deuxième souche de ce virus qui a une pathogénicité différente de la précédente. Lorsque l'on digère son ADN par HpaI, on obtient uniquement deux bandes.

Question 9 : Quelle est la plus simple explication qui rend compte de ce changement ?

Question 10 : Quel peut-être l'intérêt de connaître cette différence entre les deux souches virales ?