

# **Systèmes d'expression procaryotes : Production de protéines recombinantes par *E. Coli***

## **Introduction**

Les méthodes de biologie moléculaire et biologie cellulaire sont maintenant d'un usage courant dans les laboratoires de recherche pour la production de peptides et de protéines d'intérêt thérapeutique ou la création d'outils pour étudier, par exemple, le métabolisme des médicaments. Actuellement, plus de 200 protéines et peptides d'intérêt thérapeutique dont les séquences codantes sont clonées, ont été produits par les méthodes de biologie moléculaire. Le plus connu de ces produits recombinants est l'insuline humaine, mais des molécules comme l'albumine, certaines protéines de coagulation, des facteurs de croissance et des interleukines ont été produites et sont soit déjà utilisées en thérapeutique, soit en cours d'essais cliniques.

Auparavant, la plupart de ces protéines d'intérêt pharmacologique ou biologique étaient obtenues après purification à partir de tissus animaux ou humains. Des problèmes plus récents liés à la possible contamination par des virus comme celui du SIDA ou celui de l'hépatite.

**Une protéine recombinante (ou protéine hétérologue)** est une protéine produite par une cellule dont le matériel génétique a été modifié. Un gène codant une protéine d'intérêt est introduit dans le génome de l'espèce productrice (bactéries, cellules mammifères en culture, animaux transgéniques, etc.). Les protéines recombinantes peuvent être purifiées et utilisées à des fins thérapeutiques, industrielles ou bien encore dans les activités de recherche.

## **Stratégie pour la production de protéines recombinantes**

La biosynthèse d'une protéine hétérologue dans une cellule-hôte se déroule en plusieurs étapes qui doivent correspondre à celles qui se produisent normalement dans les cellules originales. Les stratégies d'expression des protéines doivent également permettre d'obtenir une protéine recombinante la plus proche possible de la protéine naturelle et particulièrement en ce qui concerne ses propriétés physicochimiques, catalytiques et immunologiques.

L'ARNm codant la protéine d'intérêt est isolé, transformé en l'ADNc, amplifié par PCR puis cloné dans un vecteur (chromosome artificiel bactérien, chromosome artificiel de levure, plasmide bactérien, etc.) contenant les séquences régulatrices.

Au sens large, un système de production adapté à la fabrication d'une protéine recombinante donnée, est un processus biotechnologique qui s'appuie principalement sur :

- l'emploi d'un vecteur d'expression (en général un plasmide ou un virus -pour les vecteurs eucaryotes-), jouant le rôle de transporteur génétique du gène d'intérêt codant pour la protéine recherchée ;
- l'utilisation d'une cellule hôte, chargée d'exécuter les instructions fournies par le gène d'intérêt qui lui est inséré, dans l'objectif de synthétiser la protéine recherchée ;

- une phase de production proprement dite permettant de fabriquer les volumes de protéines souhaités ;
- enfin, une séparation et extraction de la protéine du milieu de culture, suivie par une purification de celle-ci.

Au sens restreint, un système de production de protéines recombinantes est caractérisé par un couple constitué d'un vecteur d'expression et d'un hôte (cellule ou organisme).

### **Vecteurs d'expression : les plasmides, caractéristiques communes**

Un vecteur d'expression doit composer des séquences suivantes :

1. Promoteur (constitutif ou inductible)
2. Site de fixation des ribosomes
3. Polylinker/Multiple cloning site (MCS)
4. Signal de terminaison de transcription
5. Origine de réplication dans les bactéries
6. Gène de résistance à un antibiotique

### **Système d'expression procaryote : les bactéries (*Escherichia coli* comme un système de choix)**

Actuellement, 30% des protéines recombinantes utilisées en thérapeutique sont produites dans *Escherichia coli*. Dans les systèmes procaryotes, l'expression par la bactérie *E. coli* est la plus connue puisque c'est l'un des tout premiers organismes utilisés pour la production de protéines recombinantes. *Bacillus* ou *Streptomyces* sont deux autres types bactériens utilisés pour la production de protéines recombinantes mais de façon plus marginale.

Trois voies d'expression, en fonction de l'adressage de la protéine recombinante, sont classiquement exploitées en système procaryote. Les bactéries Gram- comme *E.coli* possèdent deux membranes et délimitant trois espaces. La membrane plasmique (MP) est la plus interne et délimite l'espace cytoplasmique. La membrane externe (ME) sépare l'environnement bactérien de l'espace extracellulaire. Enfin, l'espace periplasmique est localisé entre ces deux membranes.

### **Expression cytoplasmique**

#### **a) Corps d'inclusions**

Lors d'une expression cytoplasmique, la localisation finale des protéines recombinantes est le cytoplasme. Une expression trop forte ou l'expression de protéines ne parvenant pas à atteindre leur conformation native (ce qui est souvent le cas des protéines humaines nécessitant des ponts disulfures et/ou des glycosylations) entraîne une agrégation dans le cytoplasme et la formation de corps d'inclusions. On estime que 70% des protéines hétérologues produites se retrouvent dans les corps d'inclusions.

#### **b) Protéines solubles**

Certaines protéines exprimées dans le cytoplasme restent solubles et la renaturation n'est alors pas nécessaire. Pour augmenter les taux d'expression il est possible d'intervenir sur des paramètres extérieurs comme la température. En effet une baisse de la température lors de la production permet de diminuer le taux de synthèse protéique, évitant ainsi non seulement une accumulation dans le cytoplasme, mais limitant également les interactions hydrophobes qui initient des agrégats et

aboutissent à la formation des corps d'inclusion. La co-expression de protéines chaperonnes permet également d'accroître le taux de protéines correctement repliées. Cependant, cette technique, en sollicitant « la machinerie de synthèse » de l'hôte, peut également diminuer le rendement de production.

### **Expression périplasmique**

Le périplasma est localisé entre la membrane externe et la membrane plasmique de la bactérie. Ce compartiment constitue un milieu oxydant favorable au repliement des protéines, notamment grâce aux protéines Dsb. Afin d'adresser la protéine vers le périplasma une séquence signal est ajoutée en N-ter de la protéine.

### **Expression extracellulaire**

Les protéines sécrétées passent par le périplasma avant d'être libérées dans le milieu de culture. Cette voie confère les avantages suivants :

- les avantages de l'expression périplasmique décrits ci-dessus (milieu oxydant, protéines solubles et fonctionnelles).
- très peu de protéines endogènes sont excrétées, ce qui permet d'effectuer une purification plus simple et d'obtenir une bonne qualité.
- une faible activité protéase propre au milieu extracellulaire permet d'éviter toute dégradation excessive.
- l'excrétion a lieu en continu, évitant l'accumulation des protéines recombinantes surexprimées et la formation des corps d'inclusion.
- la production de protéines dont l'expression cytoplasmique ou périplasmique serait nocive ou toxique pour la cellule se trouve ainsi délocalisée.

### **Les avantages de système d'*Escherichia coli***

- Génétique très bien connue (le mieux connu)
- Nombreuses souches utilisées
- Nombreux outils de clonage commerciaux (large choix de vecteurs)
- Facile à cultiver (fermentation)
- Croissance rapide
- Expression élevée (plusieurs grammes / L de protéine recombinante)
- Possibilité de sécrétion dans l'espace périplasmique afin de simplifier la purification
- Possibilité de production en corps d'inclusion

### **Les inconvénients**

Le système procaryote n'est pas adapté pour les protéines glycosylées, même si certaines protéines eucaryotes produites sans glycosylation dans ce système s'avèrent être actives, comme le tPA (pour tissu Plasminogen Activator) humain (Demain and Vaishnav, L'absence de modifications post-traductionnelles est souvent à l'origine des problèmes de solubilité de certaines protéines recombinantes. Autre inconvénients sont :

- Faible potentiel de sécrétion
- Conséquences de la présence du vecteur (taux de croissance cellulaire, consommation d'énergie)
- Surexpression de la protéine recombinante (toxicité) (amélioration par les systèmes inductibles)
- Pas de modifications post-traductionnelles (activité biologique différente de la protéine native)
- Corps d'inclusion: protéine insoluble, mal repliée (co-expression de protéines chaperonnes (foldases))
- Protéines sécrétées dans l'espace péri plasmique: rendement moins important (co-expression de translocases)

#### **Autres bactéries (*Bacillus subtilis*, *Streptomyces*...)**

- Fort potentiel de sécrétion
- Génétique moins connue
- Niveaux de production inférieurs

#### **Exemple de la production d'une protéine d'intérêt thérapeutique par ce système : l'insuline humaine**

Le premier médicament produit par génie génétique pour lequel un brevet a été déposé est l'insuline humaine. Hormone importante qui régule le métabolisme des sucres (voir la figure 23-1).

Malgré de bons rendements et un faible coût, ce système reste peu adapté à la production de protéines complexes telles que les anticorps. En thérapie, parmi les 47 anticorps ou dérivés ayant obtenus une autorisation de mise sur le marché, seuls deux Fab sont produits chez *E.coli*.