

1.1. Introduction

Si le génome est identique dans chacune des cellules d'un organisme donné, en revanche, les gènes, eux, peuvent avoir une expression génique spécifique différenciée dans le temps (propre à un stade du développement), dans l'espace (propre à un type cellulaire, tissulaire ou organique) ou/et caractéristique d'un état donné (normal, pathologique ou en réponse à un stimulus particulier). L'étude de l'expression génique consiste à caractériser et quantifier les produits d'expression de l'ADN (les ARN messagers : le transcriptome) de manière à identifier, dans un tissu, dans un état et à un moment donné du développement, les séquences actives et donc à révéler ainsi le niveau d'expression des gènes dont elles sont issues. Les études d'expression globale peuvent être utilisées pour classer les gènes selon leur profil d'expression spacial et cinétique.

De nouveaux éléments de régulation peuvent être identifiés en comparant les séquences de régulation des gènes de même classe. La détermination de la cinétique d'expression d'un gène, combinée aux profils d'expression de mutants pour des gènes de régulation connus, peut permettre l'étude des réseaux d'expression. Les profils d'expression analysés peuvent également être utiles pour caractériser les défauts des lignées de mutants dont le développement est anormal. Il existe actuellement deux stratégies dominantes pour effectuer une analyse quantitative et qualitative en série de l'expression génétique : les méthodes de séquençage d'étiquettes (méthode SAGE) et celles d'hybridation (puces à ADN).

1.2. Méthode SAGE (Serial Analysis of genes expression)

1.2.1. Définition

Technique d'analyse quantitative et qualitative en série de l'expression génétique d'un grand nombre de transcrits. Elle consiste à réaliser un inventaire des transcrits (ARNm présents dans un échantillon de cellules, tissus ou organes et convertis en ADNc) en isolant à partir de chacun d'eux un courts fragment spécifique de 9 à 14 pb appelés sequence tags. Tous les tags (jusqu'à une cinquantaine) sont ensuite concaténés en une longue séquence unique qui est séquencée, révélant ainsi l'identité de chaque tag (et donc du gène dont il est issu). En générant ainsi un grand nombre de séquences à partir d'un échantillon donné, il est possible de mesurer la fréquence des différents tags (= mesure quantitative du profil d'expression).

1.2.2. Principe

La méthode SAGE, extension logique du séquençage par EST, consiste à réaliser un inventaire des transcrits au moyen de courts fragments d'ADNc de 9 à 14 pb appelés sequence tags. Dans un premier temps, un pool d'ADNc est synthétisé à partir des ARNm d'un échantillon (cellules, tissus, organes...). Un traitement enzymatique permet ensuite d'isoler à partir des extrémités 3'de chacune des molécules d'ADNc, une courte séquence unique : le tag dont la taille est suffisante pour pouvoir identifier le gène dont elle dérive. Tous les tag (jusqu'à une cinquantaine) sont ensuite concaténés pour former un fragment d'ADN qui puisse être manipulé et finalement séquencé. Les gènes, à l'origine des tags, peuvent alors être identifiés par comparaison de la séquence avec les banques. En faisant un grand nombre de séquences, il est possible de mesurer la fréquence des différents tags dans un échantillon donné. En comparant les données issues de plusieurs échantillons, on peut donc avoir une idée du profil d'expression d'un grand nombre de gènes.

Toutefois, cette technique n'est pas adaptée à l'étude de nombreux échantillons et de transcrits rares. De plus, elle requiert l'existence de bases de données de séquences génomiques complètes pour les espèces étudiées. La méthode SAGE a déjà fait l'objet d'applications chez l'homme, chez la levure et chez les plantes.

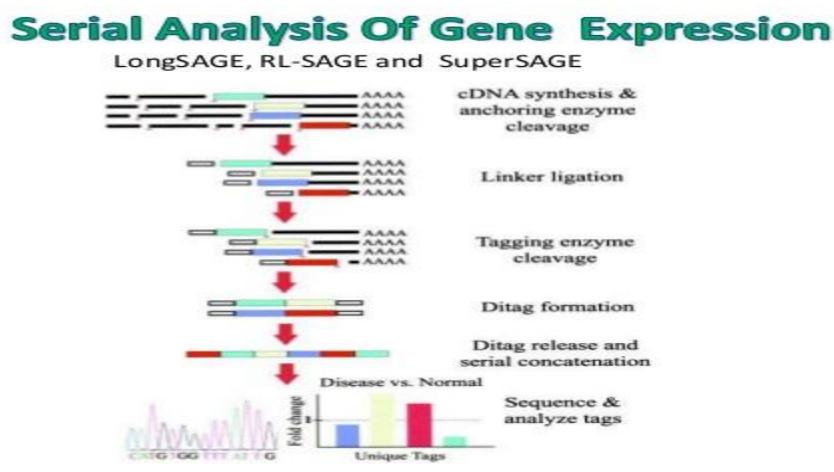


Fig.01. Méthode SAGE (Serial Analysis of genes expression)

1.2. Puces à ADN

1.2.1. Définition

C'est une technique d'hybridation permettant une analyse génomique comparative de l'expression d'un grand nombre de patterns d'ARNm. Immobilisés sur un support solide (matrice), des oligonucléotides (simples brins) spécifiques de différents gènes ou ADNc connus constituent les sondes dont le rôle est de détecter des cibles marquées complémentaires, présentes dans le mélange complexe à analyser (ARNm extraits de cellules, tissus ou organismes entiers et convertis en ADNc). Les sondes sont soit greffées sur le support, soit synthétisées *in situ* (unité d'hybridation = plot). Les signaux d'hybridation sont détectés selon le type de marquage, radioactivité ou fluorescence, par mesure radiographique ou par fluorescence, et quantifiés.

1.2.2. Principe

Le concept de biopuce (= puce à ADN) est né dans les années 1990. Il repose sur une technologie pluridisciplinaire intégrant la micro-électronique, la chimie des acides nucléiques, l'analyse d'images et la bioinformatique.

Le principe de la puce à ADN est basé sur la technique d'hybridation. Immobilisés sur un support solide (matrice), des oligonucléotides (simples brins) spécifiques de différents gènes ou ADNc connus constituent les sondes dont le rôle est de détecter des cibles marquées complémentaires, présentes dans le mélange complexe à analyser (ARNm extraits de cellules, tissus ou organismes entiers et convertis en ADNc). Les sondes sont soit greffées sur le support, soit synthétisées *in situ* (unité d'hybridation = plot). Les signaux d'hybridation sont détectés selon le type de marquage, radioactivité ou fluorescence, par mesure radiographique ou par fluorescence, et quantifiés (**Fig.02**).

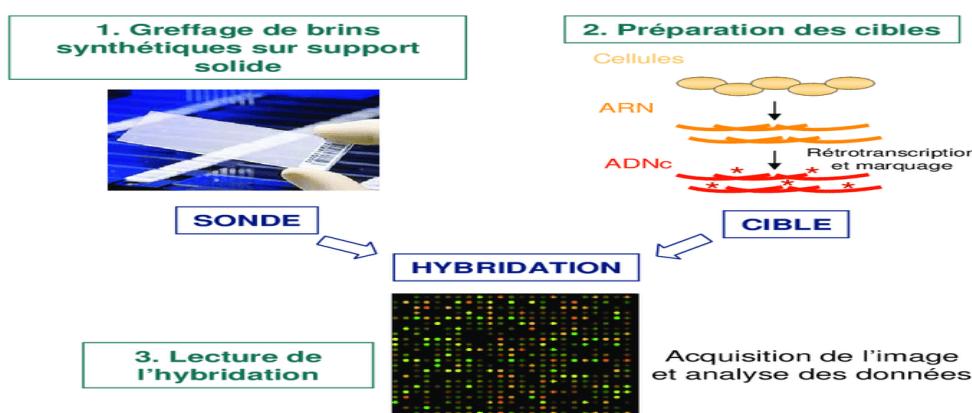


Fig.02. Principe d'une puce à ADN

Le support sur lesquels sont fixées les sondes se présente sous la forme d'une surface (matrice) inférieure à 1cm², plane ou poreuse (percées de trous) et composés de matériaux tels que le verre, les polymères, le silicium ou l'or et le platine. L'élément principal de la puce à ADN est l'unité d'hybridation appelée **Plot** sur lequel sera greffée une sonde d'ADN synthétique. Les plots, présents en un grand nombre d'exemplaires, sont répartis régulièrement sur toute la surface de la puce.

1.2.3. Fixation des sondes (lecture de l'hybridation)

Il existe deux méthodes pour "greffer" les sondes (dont on connaît la séquence) sur les plots :

A. fixation d'oligonucléotides déjà synthétisés

Ce mode opératoire permet de fixer des sondes relativement longues, atteignant 40 à 60 bases et s'effectue selon deux types d'adressage possibles :

- par adressage mécanique, grâce à une micropipette robotisée
- par adressage électrochimique : activation d'une microélectrode spécifique sur laquelle est réalisée une électro-polymérisation.

B. synthèse in situ

(Affymetrix par exemple). Dans ce cas, la construction des sondes se fait par dépôt de couches successives des quatre bases de l'ADN sur un support en verre. C'est un **masque** lithographique, dont la configuration varie à chaque dépôt d'une couche, qui permet aux bases de s'empiler correctement. Avec ce procédé, les sondes comportent au maximum 30 bases. Ce procédé permet de fabriquer indifféremment telle ou telle sorte de biopuce. L'adjonction successive de bases à la sonde s'effectue par adressage photochimique.

C. Hybridation sonde-cible

Une séquence d'ADN cible est marquée par fluorescence (fluorochrome devenant fluorescent lorsqu'il est excité par un faisceau lumineux) et préparée en solution et mise en contact avec la biopuce.

Le rôle de chaque sonde est de reconnaître, dans le mélange appliqué à la surface de la biopuce, cette séquence d'ADN cible. La phase d'hybridation est réalisée dans une sorte d'incubateur (station fluidique) et suivie d'un lavage destiné à débarrasser la puce des cibles nucléiques non hybridées.

D. Détection des hybridations

Il s'agit ici de déterminer les sondes qui ont été effectivement hybridées, c'est à dire de repérer à quelles adresses les sondes sont complémentaires des cibles de l'échantillon test. Une lecture optique permet de révéler les sondes d'ADN devenues fluorescentes (hybridées avec les cibles marquées). Les données récoltées sont analysées par un logiciel-traitement d'images.

L'utilisation de 2 fluorochromes différents permet la détermination des signaux d'hybridation de deux sondes distinctes au cours d'une seule expérience.

1.2.4. Possibilités d'utilisation

Les puces à ADN permettent des tests plus rapides, plus sensibles et plus spécifiques. En évitant certaines étapes préliminaires telle que la culture, il permet d'obtenir un résultat en quelques heures là où plusieurs jours étaient nécessaires.

- Analyse de mutations

A côté de la séquence sauvage (sonde de référence), on utilise plusieurs sondes identiques à une mutation près. Ces sondes vont couvrir l'ensemble des mutations potentielles (substitution, délétion, addition). De manière générale, l'analyse de substitutions (pour une position donnée) implique donc l'utilisation de quatre sondes, une pour la séquence sauvage et trois pour chacun des nucléotides pouvant se substituer au nucléotide original. Cela permet ainsi de repérer des séquences mutantes dans un échantillon test.

- Séquençage par hybridation

L'analyse procède par lecture de petits blocs. L'analyse porte sur des sous-séquences chevauchantes qui sont lues et rassemblées au moyen d'un programme informatique qui reconstitue la séquence étudiée.

-Diagnostic des maladies génétique

les Immuno-essais (hormone et cancer), en microbiologie (antigènes/anticorps), les facteurs aggravants des maladies cardio-vasculaires et du vieillissement. Etude d'oncogènes spécifiques.

-Diagnostic des maladies infectieuses

Détection de microorganismes (présence, mutations, capacité à résister à certains antibiotiques ou à des inhibiteurs d'enzymes).

- Pharmacogénomique

Identification de cibles pour la recherche thérapeutique.

-Toxicogénomique

Dans le cas de l'analyse de la réponse toxicologique à l'aide de puces à ADN, on doit au départ faire l'hypothèse que l'effet toxicologique, direct ou indirect, d'une substance résulte de la modification de l'expression de certains gènes, ou du moins qu'il lui est associé.

Les mesures consistent alors à évaluer l'expression différentielle de gènes de deux échantillons, grâce à un marquage par fluorescence en deux couleurs (une par échantillon). Cela permet de comparer les profils d'expression génique de deux tissus différents (normal/pathologique ou traité/non traité par une substance).

-Agro-alimentaire

Contrôle des microorganismes utilisés dans certaine fabrication (ferments lactiques, levures, mycélium, etc....), détection des séquences provenant d'organismes génétiquement modifiés (OGM) dans les semences ou dans certains aliments, détection d'agents infectieux dans l'alimentation comme la Salmonelle ou la Listeria.

-Environnement

Analyse bactérienne de l'eau de consommation, détection des agents infectieux dans l'alimentation, l'air ou l'eau (Salmonella, Listeria, Legionella).

1.3. Analyse protéome

1.3.1 Introduction

L'information sur les taux d'ARNm n'est pas suffisante pour étudier et analyser la régulation de l'expression d'un gène dans la cellule. Les données sur l'expression des protéines (protéome) sont souvent plus informatives. Ce qui compte, en effet, dans une cellule, c'est le degré d'activité des protéines, qui se trouvent fréquemment modulées et régulées par des modifications post-traductionnelles.

Ainsi, le défi de la protéomique consiste à trouver des moyens de mesurer les protéines actives à l'échelle du protéome et de connaître avec quels composants de la cellule elles interagissent (ADN, ARN, autres protéines et molécules). Alors que le génome est constant (aux mutations somatiques près) dans toutes les cellules d'un organisme, le protéome (tout comme le transcriptome) varie d'un type cellulaire à l'autre et selon la situation physiologique et de développement dans laquelle se trouve chaque cellule. Ainsi, un même individu possède en fait, non pas un protéome, mais plusieurs.

Une grande partie des variations de fonctionnement d'une cellule s'expliquerait au moins autant par la modulation de la quantité des protéines que par la présence ou l'absence de certaines d'entre elles. Avant l'annonce du décryptage presque complet du génome humain, on expliquait souvent les différences entre les cellules par le jeu de l'"allumage" et de l'"extinction" (on/off) de certains gènes parmi les 100000 ou 120000 estimés. Compte tenu du plus petit nombre de gènes humains désormais attendu (environ 30000), cette explication devient insuffisante : ce sont plus vraisemblablement les quantités de protéines présentes dans une cellule donnée, ainsi que les modifications post-traductionnelles qui les affectent, qui vont différencier cette cellule des autres (**Fig.03**).

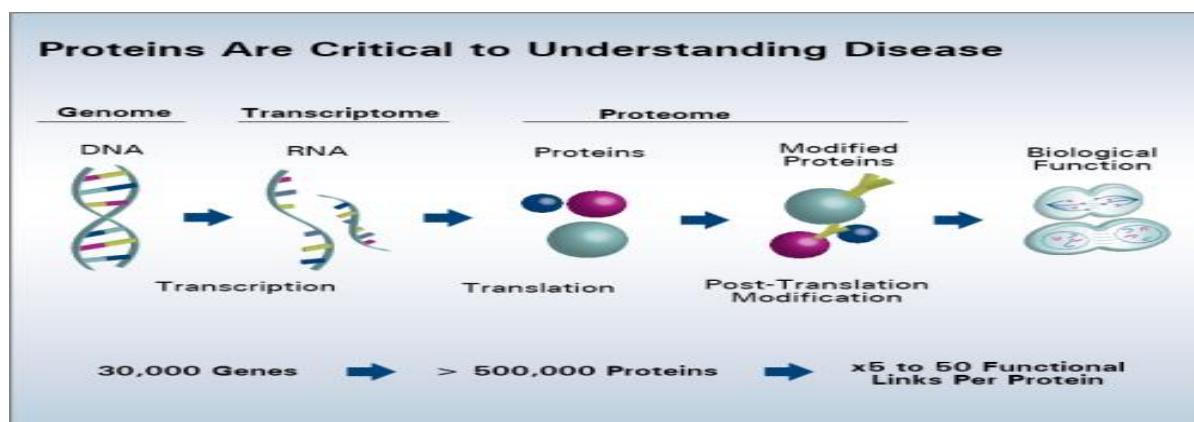


Fig.03. Le génome et le protéome

Les méthodes actuellement utilisées pour analyser les protéines n'ont pas la capacité de haut débit, ni la reproductibilité et l'automatisation nécessaires pour être aujourd'hui associées de manière acceptable dans une perspective de génomique à grande échelle.

1.3.2. Electrophorèse bidimensionnelle sur gel

Permet de séparer des protéines contenues dans les extraits cellulaires. Elle est pour le moment la seule méthode fiable pour étudier l'abondance et les modifications post-traductionnelles de plusieurs centaines de protéines en parallèle. Récemment, la résolution et la reproductibilité du système de séparation ont été considérablement améliorées, et des logiciels pour quantifier automatiquement les taches ou spots correspondant aux protéines ont été développés.

Des bases de données contiennent les résultats d'expériences d'electrophorèse bidimensionnelles. Une application de la protéomique consiste aussi à étudier la fonction des gènes à partir de mutants d'inactivation ou de surexpression, et à analyser les changements des profils protéiques sur gel bidimensionnel.

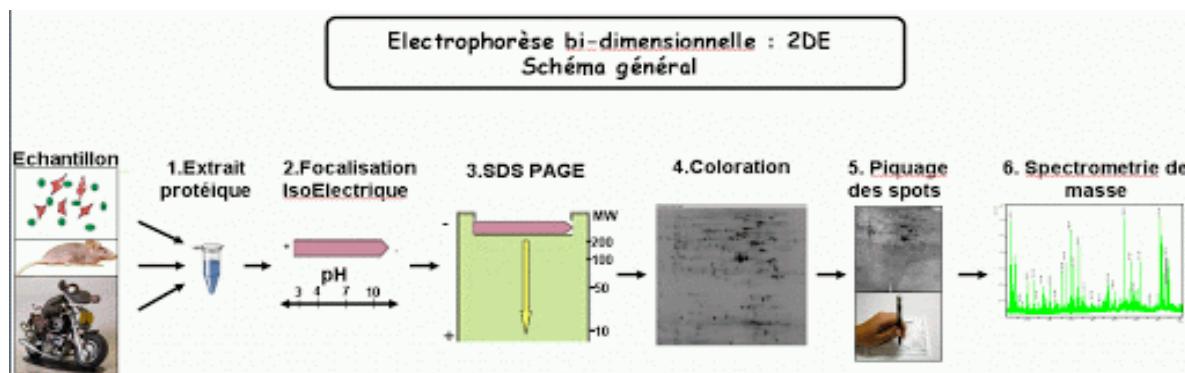


Fig.04. Electrophorèse bidimensionnelle.

Le traitement protéasique par une enzyme telle que la trypsine (digestion enzymatique) appliqué aux quantités infimes de protéines extraites des gels libèrent des fragments protéiques caractéristiques de chacune des protéines.

Enfin la spectrométrie de masse permet d'identifier une protéine et ses modifications en mesurant la masse des peptides (fragments protéiques) issus de la digestion. Les résultats sont stockés dans les bases de données analysables par des programmes spécifiques. Les bases de données de séquences permettent de prédire la masse théorique des fragments de protéolyse d'une protéine. Ainsi, lorsqu'une base de données complète d'un organisme est disponible, il est possible d'identifier une protéine d'après le spectre de masse de ses fragments de protéolyse.

1.3.3. Microséquençage N-terminal

Séquençage automatisé par la méthode d'Edman de très petites quantités de protéines entières ou de fragments protéiques qui permet ainsi, aussi, d'identifier les séquences correspondantes. La dégradation d'Edman permet d'enlever un par un les résidus en commençant par l'extrémité N-terminale. On arrive ainsi à déterminer la séquence de fragments d'environ 20 résidus.

1.3.4. Puces à protéines

Le fonctionnement de ces systèmes, qui permettraient idéalement d'identifier en parallèle plusieurs protéines au sein d'un mélange, est fondé sur la réaction antigène-anticorps entre les protéines à identifier et des anticorps spécifiques fixés à la surface de la puce. Les protéines sont des acteurs majeurs dans la vie cellulaire. Les puces à protéines sont utilisées à la fois pour étudier le niveau d'expression des protéines et pour l'analyse des interactions de certaines protéines avec d'autres composants (**Fig.05**).

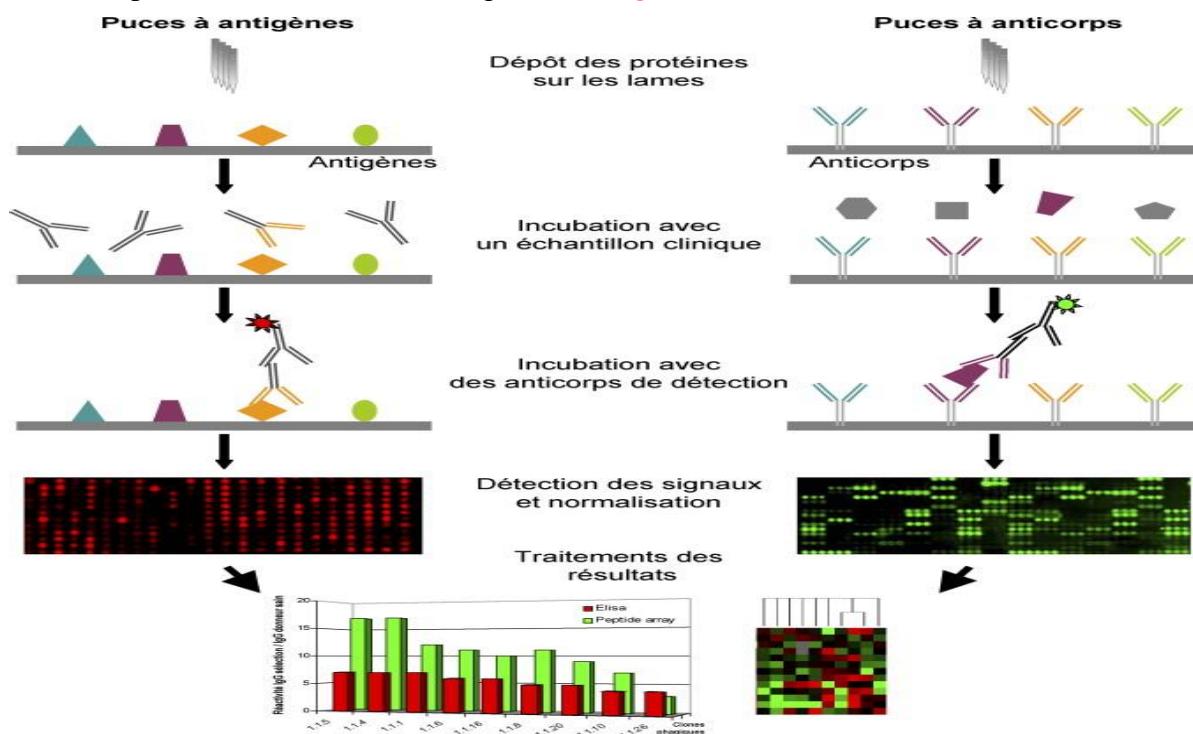


Fig.05. Puces à protéines et perspectives d'applications médicales

Les puces à protéines sont également utilisées pour étudier les interactions entre protéines. La connaissance de ces interactions est fondamentale pour le ciblage actif des tumeurs. Cette technique consiste à envelopper les nanoparticules d'une molécule destinée à se lier spécifiquement à un antigène ou à un récepteur exprimé à la surface des cellules tumorales. Le ciblage actif est utilisé aussi bien pour l'adressage de médicament (vectorisation) que pour

l'illumination des tumeurs (par Quantum Dots ou nanoparticule d'or) ou encore pour l'élimination des cellules cancéreuses.

1.3.5. Système double hybride

Une autre stratégie de la protéomique vise à établir des cartes d'interactions protéiques de manière à reconstituer des voies métaboliques. De telles cartes permettent de guider sur le rôle d'une protéine totalement inconnue, pour peu qu'elle interagisse avec une autre dont on connaît la fonction.

Pour cela, des plate-formes technologiques se sont développées en se fondant sur la technique du double hybride. L'analyse bioinformatique des homologies de séquences protéiques avec d'autres protéines avec lesquelles on sait qu'elles interagissent permet d'établir une liste des interactions susceptibles de se produire. Pour vérifier et valider *in vivo* cette interaction, on utilise la méthode du double hybride. Celle-ci consiste à étudier, dans un organisme (notamment une bactérie), la capacité d'une protéine connue (la cible) à interagir avec une protéine inconnue (la proie) *in vivo*. Dans ce système, l'interaction est détectée par la formation d'un complexe moléculaire qui active l'expression d'un gène rapporteur.

Un tel système permet de définir de nouvelles cibles thérapeutiques. Le peptide "proie", qui s'exprime dans la bactérie, entre en compétition avec la protéine endogène qui lui correspond, sans pouvoir remplir la fonction de cette dernière : on peut ainsi montrer quel effet provoquerait l'inhibition de l'interaction sur la bactérie. Il reste alors à mettre au point des molécules capables de moduler cette interaction.