

RÉGULATION DE L'EXPRESSION DES GÈNES

Comme nous l'avons vu, le programme de transcription n'est pas fixe. La cellule sait adapter ce programme aux conditions extérieures, au mieux de son économie. Chez les Eucaryotes, l'environnement peut être représenté par les cellules voisines. Le développement et la différenciation exigent l'expression régulée de gènes différents dans les diverses cellules.

La question posée dans ce chapitre est celle du choix des portions du génome devant être exprimées à un moment donné dans un environnement donné.

Dans une première partie, nous allons utiliser l'étude de la cellule procaryotique pour donner quelques exemples de contrôle de l'expression de gènes et dégager des notions fondamentales de régulation.

1 - LES PROCARYOTES

La cellule bactérienne est capable d'adapter très rapidement son métabolisme aux modifications de l'environnement. Prenons une suspension de cellules d'*Escherichia coli*. Une partie de la suspension est cultivée dans un milieu minimum contenant du lactose (disaccharide composé d'une molécule de glucose et d'une de galactose) comme seule source de carbone (milieu A). Une autre partie est cultivée dans le même milieu minimum, mais contenant du glucose comme source de carbone et auquel on a ajouté les 20 acides aminés (milieu B). On constate que les cellules du milieu A croissent (prolifèrent), mais moins vite que celles du milieu B. Elles doivent en effet assurer le clivage du lactose en deux monosaccharides (glucose et galactose) assimilables et toute la biosynthèse des acides aminés.

L'étude biochimique et génétique de l'utilisation du lactose par *E. coli* va nous montrer comment l'économie réalisée par les cellules cultivées dans le milieu B (de génotype identique à celles cultivées sur le milieu A) se fait, par l'adaptation du programme d'expression des gènes, dès le niveau transcriptionnel. Ces résultats sont le fruit des travaux de Jacob, Monod et leurs collaborateurs dans les années 60.

Par des méthodes biochimiques appropriées, il est possible de doser l'activité de la β -galactosidase, enzyme catalysant la dégradation du lactose. Cette activité est présente dans les cellules du milieu A, mais absente (à la limite de la détection) dans les cellules du milieu B. Si des cellules «B»,

sont transférées dans le milieu A (ne contenant plus de glucose mais du lactose), une activité β -galactosidase apparaît très rapidement et s'amplifie d'un facteur 1000 en quelques 20 minutes. L'activité enzymatique semble donc induite par la présence de lactose dans le milieu.

De plus, si l'on transfère ces cellules ayant développé une activité β -galactosidase dans un milieu contenant à nouveau du glucose, l'activité enzymatique disparaît donc semble réprimée.

Ainsi, la présence de lactose paraît déclencher la synthèse d'enzymes spécifiques. L'analyse biochimique montre l'apparition simultanée de trois enzymes : la β -galactosidase, codée par le gène *lacZ*, mais également la perméase au lactose (codée par le gène *lacY*), capable d'accélérer la pénétration du lactose dans la cellule et une thiogalactoside acétyl transférase (codée par le gène *lacA*), qui joue un rôle dans le métabolisme d'autres galactosides. Ces trois activités enzymatiques sont co-régulées (elles apparaissent en réponse à une même induction).

Des méthodes très fines de cartographie montrent que ces gènes sont adjacents dans une région appelée «*lac*» du chromosome bactérien.

Remarque : L'apparition brutale d'une activité enzymatique ne suffit pas à prouver une régulation au niveau de la transcription des gènes. En effet, une enzyme peut être présente sous une forme inactive et l'induction représenter en fait l'activation de protéines préexistantes dans le cytoplasme. La preuve directe d'un contrôle au niveau transcriptionnel a été apportée plus tard grâce aux techniques permettant de doser des ARN messagers spécifiques. Un ADN complémentaire de l'ARN messager de la β -galactosidase a été inséré dans un vecteur de clonage. Après amplification, cette séquence a servi de sonde pour détecter, dans des préparations brutes d'ARN (extraits avant et après le passage au lactose), l'ARN messager de la β -galactosidase (seul capable de s'hybrider avec la sonde) et le quantifier. Les résultats indiquent que l'accumulation de ce messager débute dès l'induction et qu'elle précède bien celle de l'enzyme.

L'induction porte sur la transcription du gène lui-même.

Remarque : Le lactose fait partie de ce que l'on appelle des molécules effectrices, c'est-à-dire des

molécules capables de refléter l'environnement cellulaire et de transmettre un signal permettant un choix transcriptionnel. Il faut noter que, dans cet exemple particulier, le lactose est loin d'être le meilleur inducteur de la transcription des gènes impliqués. D'autres petites molécules glycosidiques, naturelles ou artificielles (telles que l'isopropyl-thiogalactoside ou IPTG) sont beaucoup plus efficaces et utilisées au laboratoire.

1.1. ANALYSE GÉNÉTIQUE DE LA RÉGULATION

L'analyse génétique des procaryotes, qui a permis d'élucider les mécanismes fondamentaux de la régulation, relève d'une gageure : il s'agit de cellules haploïdes tout au long de leur cycle, sans reproduction sexuée, nous privant, apparemment des deux outils de base que sont la **recombinaison** et la **complémentation**. En fait, dès 1946, Lederberg et Tatum utilisaient d'une façon très élégante les possibilités de mélange de matériel génétique chez les bactéries, liées au phénomène de **conjugaison**. Voir en annexe «La Conjugaison bactérienne».

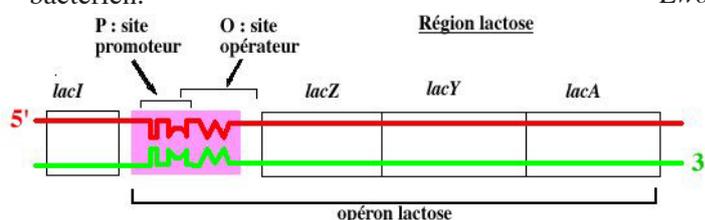
1.1.1 EXEMPLE DE L'ADAPTATION A L'UTILISATION DU LACTOSE

Plusieurs types de mutation peuvent interférer avec l'utilisation du lactose.

Des mutations dans le gène *lacZ* empêchent le catabolisme du lactose par absence de β -galactosidase fonctionnelle. Des mutations dans le gène *lacY* empêchent la pénétration active du lactose à l'intérieur de la cellule. Dans ces deux cas, le phénotype, symbolisé par Lac^- , c'est-à-dire l'impossibilité d'utiliser le lactose, résulte d'un défaut enzymatique.

D'autres mutations perturbent d'une façon tout à fait différente l'utilisation du lactose : elles affectent la production de trois enzymes en même temps (β -galactosidase, perméase et acétyl transférase). Plusieurs de ces mutations entraînent un phénotype **constitutif**, c'est-à-dire **production d'enzymes en absence d'inducteur**.

Etant donné que les trois gènes sont adjacents sur le chromosome, on peut supposer que leur expression est régulée par un même système et que les mutations constitutives affectent non pas une production d'enzyme mais un élément de contrôle. Cet ensemble forme ce que l'on appelle un **opéron** bactérien.



Les premières mutations constitutives étudiées ont été appelées *I* et sont localisées dans le gène *lacI*, en amont du gène *lacZ*. Par la suite on a caractérisé des mutations constitutives *O^c*, ces mutations sont localisées au niveau du site opérateur.

L'analyse de ces mutants par des expériences très élégantes faisant appel à des **diploïdes partiels** a permis au groupe de Monod et Jacob, dans les années 60, d'élaborer le célèbre modèle de **régulation de l'opéron lactose par répression de la transcription**.

Le tableau ci-dessous rappelle les caractéristiques phénotypiques des mutants utilisés par rapport au phénotype sauvage.

génotype	phénotype
$I^+ O^+ Z^+$	sauvage (inductible)
$I^+ O^+ Z^-$	non inductible
$I^- O^+ Z^+$	constitutif
$I^+ O^c Z^+$	constitutif

La construction de diploïdes partiels a été réalisée par conjugaison, entre des souches Hfr (dont le facteur F est inséré à proximité de l'opéron *lac*), sensibles à la streptomycine (Sm^S) sont utilisées comme cellules donneuses, les souches receveuses sont résistantes à la streptomycine (Sm^R). Après contact assez bref, les cellules sont étalées sur un milieu contenant de la streptomycine et contenant ou non un inducteur (ici de l'IPTG). Le tableau suivant résume les principaux «croisements» réalisés (le génotype est celui des diploïdes partiels) et indique les phénotypes résultants.

génotype	phénotype
$I^+ O^+ Z^-$ ----- $I^+ O^+ Z^+$	sauvage (inductible)
$I^+ O^+ Z^-$ ----- $I^- O^+ Z^+$	sauvage (inductible)
$I^+ O^+ Z^-$ ----- $I^+ O^c Z^+$	constitutif

Remarque : Ce tableau est à analyser avec beaucoup d'attention, l'interprétation de tels résultats a valu le prix Nobel à F. Jacob, J. Monod et A. Lwoff.

Nous allons reprendre leur interprétation dans les paragraphes suivants.

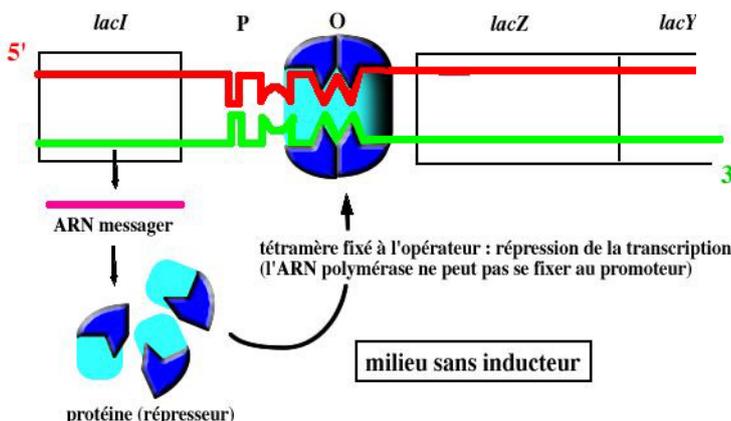
1.1.2. MODÈLE DE RÉGULATION PAR RÉPRESSION DE LA TRANSCRIPTION

La présence de β -galactosidase chez un diploïde partiel Z^-/Z^+ indique qu'une complémentation est possible. L'allèle Z^+ de la cellule donneuse est exprimé dans le cytoplasme de la cellule receveuse (on peut dire que l'allèle sauvage Z^+ , est dominant par rapport à l'allèle muté Z^-).

De même, la présence de l'allèle I^+ (dans le contexte O^+) rétablit le contrôle normal (inductible) de l'expression du gène *lacZ*. On peut en conclure que ce gène est exprimé en une protéine capable d'agir sur le fonctionnement de l'opéron. Monod et Jacob proposent que cette protéine (produit du gène *lacI*) soit un répresseur de la transcription lorsque l'inducteur est absent (l'induction serait en fait une levée de la répression).

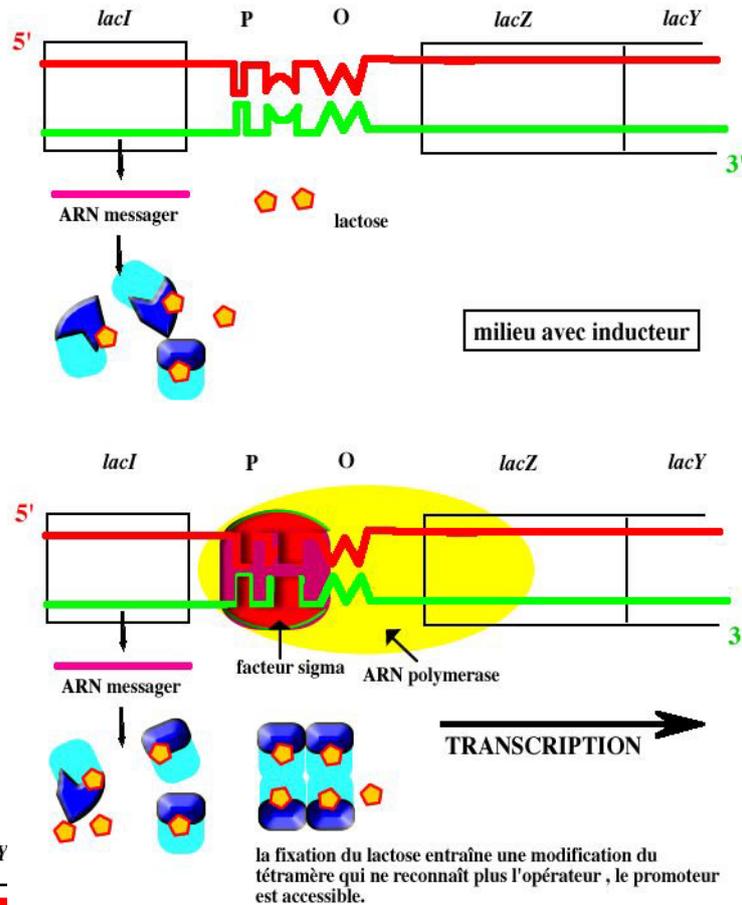
Les résultats obtenus avec les diploïdes partiels O^+/O^c sont très différents, il n'y a pas complémentation, l'allèle muté semble dominant par rapport au sauvage. Ceci ne peut s'expliquer que si le locus *O* n'est pas un gène exprimé, mais une séquence particulière d'ADN qui est appelée **site opérateur**.

Le schéma de fonctionnement normal de l'opéron serait alors le suivant : en absence d'inducteur, le répresseur produit par le gène *lacI* reconnaît spécifiquement le site *O* et s'y fixe. L'encombrement de ce complexe (on découvrira plus tard que c'est en fait un tétramère qui se fixe) est tel que l'ARN polymérase est incapable de se fixer au site promoteur. Le promoteur étant unique pour les trois gènes *lacZYA*, on comprend que la répression bloque la transcription de l'ensemble.



Comment se fait l'induction ? Une propriété très intéressante de certaines protéines est l'**allostérie** : la fixation d'une molécule particulière (ligand) provoque la modification globale de la structure tridimensionnelle d'une protéine réceptrice. C'est ce qui se produit ici. Le lactose a une

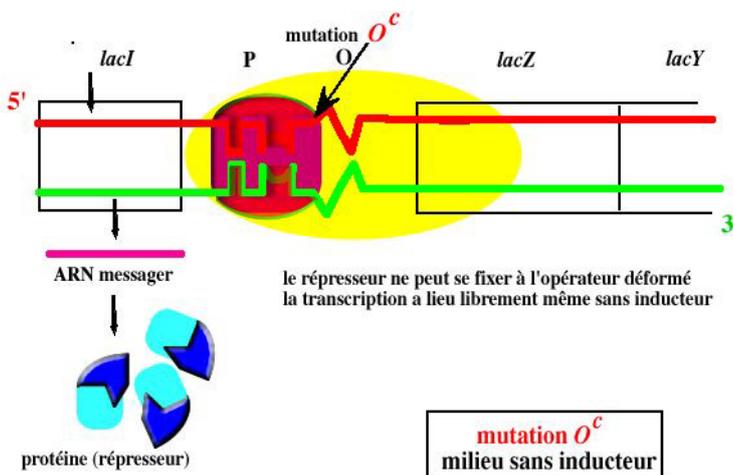
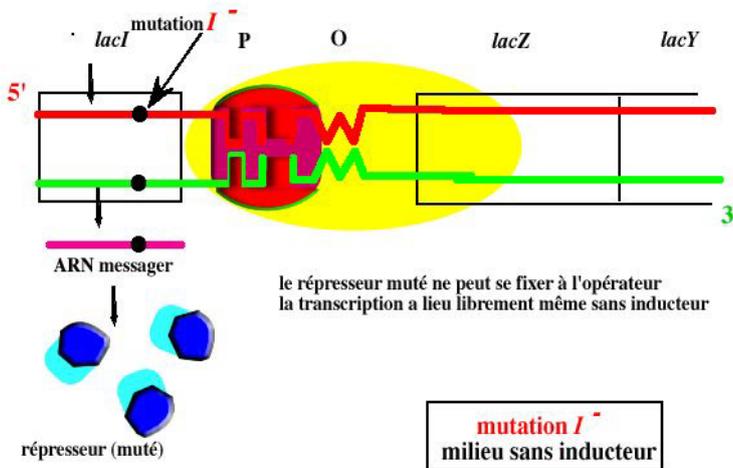
affinité pour la protéine répresseur LacI et sa liaison provoque une transition allostérique de celui-ci. Si l'on tient compte du fait que l'interaction d'une protéine avec une séquence d'ADN nécessite l'établissement de liaisons hydrogène entre des atomes précis d'acides aminés particuliers et des atomes précis de bases données, la déformation de la protéine ne permettra plus cette interaction et le complexe répresseur-inducteur sera incapable de se fixer à l'opérateur.



Le principe de l'interaction entre une protéine (en tant que séquence spécifique d'acides aminés conditionnant sa structure tridimensionnelle) et une séquence d'ADN permet de comprendre l'effet des mutations constitutives I^- et O^c .

Ainsi, une mutation dans le gène *I* conduit à une altération de la structure du répresseur voire à une absence de la protéine (**allèle nul**). Dans tous les cas, une liaison répresseur - opérateur ne peut s'établir et l'ARN polymérase peut se fixer au promoteur. Si le répresseur possède une structure correcte, mais que la séquence opératrice est altérée par une mutation (O^c par exemple), le résultat est le même : aucune possibilité

de former un complexe répresseur - opérateur ; la transcription de l'opéron s'effectue en permanence (voir les figures ci-dessous).



1.1.3 LE RÉPRESSEUR

Une confirmation du mode de régulation par répression a été apportée par des mutations I^S , signifiant «super-réprimé». Ces mutants sont incapables d'utiliser le lactose car l'opéron est réprimé en permanence, le lactose n'induit pas la transcription. Les mutations affectent le gène *lacI* dans une région importante pour la formation du complexe répresseur - inducteur. La région essentielle pour la liaison du répresseur à la séquence opératrice restant intacte, l'état réprimé est stable.

Ces observations permettent d'aborder un aspect plus général des protéines de régulation, celui de **domaines fonctionnels** spécialisés. L'analyse détaillée de la protéine LacI, après clonage du gène *lacI* dans un vecteur d'expression et surproduction par des clones bactériens transformés (voir le chapitre correspondant) confirmera ce concept. Les mutations I^- , I^S et d'autres ne sont pas disposées de façon aléatoire dans le gène. Leur cartographie reflète le fait qu'une partie de la protéine codée est essentielle dans la reconnaissance du site opérateur, une autre région est nécessaire pour la fixation de l'inducteur et la transition allostérique, une autre permet la formation d'un tétramère. Cette notion de **protéines de régulation modulaires**, séparables en domaines fonctionnels, se retrouvera chez les Eucaryotes.

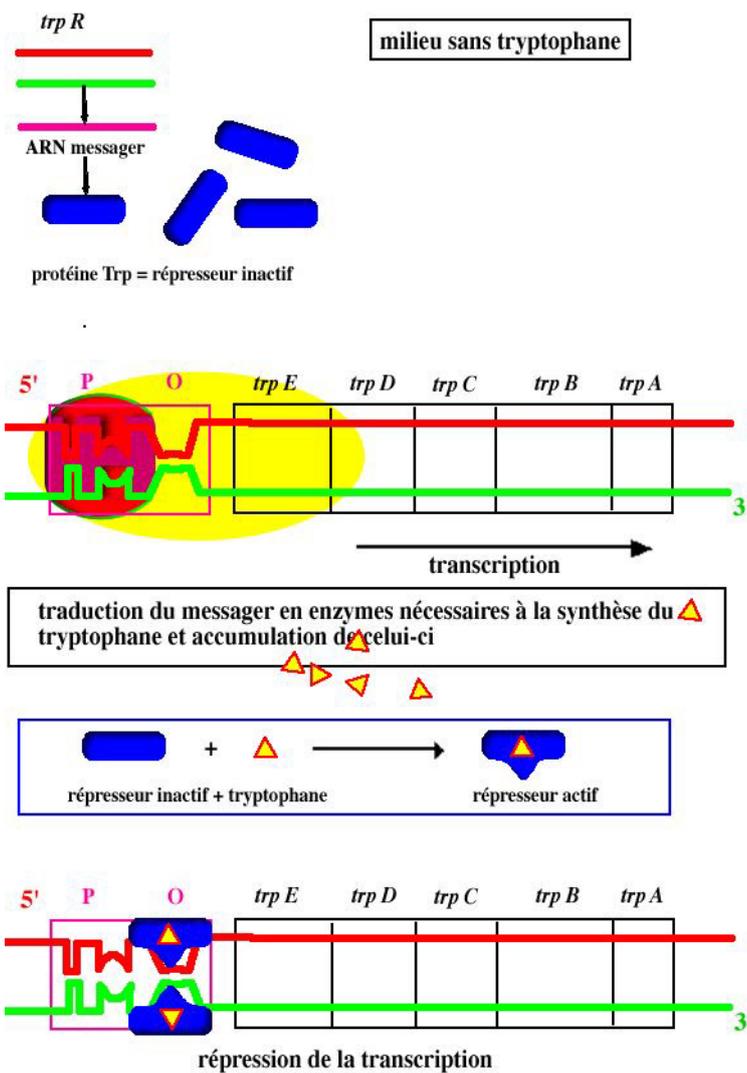
1.1.4 OPERON INDUCTIBLE ET OPERON «REPRESSIBLE»

L'opéron lactose ne fonctionne que s'il est induit par une molécule effectrice capable de lever l'effet du répresseur, il est dit inductible. Pour d'autres opérons, c'est l'inverse, la molécule effectrice provoque la répression de la transcription, l'opéron est dit **répressible**.

Un exemple typique nous est fourni par un opéron dont les produits interviennent dans la biosynthèse du tryptophane.

Remarque : l'opéron lactose intervient dans le catabolisme, l'opéron tryptophane intervient dans l'anabolisme, en règle générale, les voies cataboliques sont régulées par répression et les opérons intervenant dans l'anabolisme par induction.

Cet opéron comporte cinq gènes groupés codant pour des enzymes impliquées dans la synthèse du tryptophane (*trpEDCBA*). Ces gènes sont sous la dépendance d'un seul système promoteur - opérateur. Le gène *trpR*, qui ne fait pas partie de l'opéron, code pour une protéine qui est le répresseur spécifique de l'opéron tryptophane. Cependant, cette protéine est incapable de se lier au site opérateur, et par conséquent est inactive, **tant qu'elle n'est pas complexée avec une molécule effectrice** : le tryptophane lui-même. Il agit donc comme un **corépresseur** dans ce mécanisme de régulation en retour par le produit final de la chaîne métabolique de l'opéron. (voir la figure suivante)



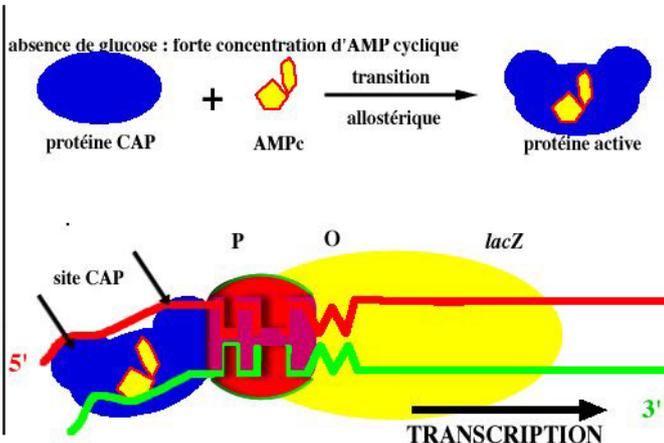
1.1.5 CONTRÔLE NÉGATIF ET CONTRÔLE POSITIF

Les gènes soumis à un contrôle négatif ne sont pas transcrits si un répresseur est lié au site opérateur. C'est le cas des opérons lactose ou tryptophane.

Les gènes soumis à un contrôle positif ne sont transcrits efficacement que si une protéine régulateur favorise l'initiation.

L'opéron lactose est également soumis à un contrôle positif. Il a été précisé au début de ce chapitre que la transcription est induite si l'on remplace le glucose par du lactose comme seule source de carbone dans le milieu minimum. Cependant, si l'on ajoute du lactose, l'opéron n'est pas transcrit tant que le glucose n'est pas épuisé. La présence de glucose ne permet pas une transcription efficace de l'opéron lactose. Ce phénomène, qui concerne de nombreux opérons du catabolisme est appelé «effet glucose» ou encore répression catabolique. Il repose, comme les autres modes de régulation sur une protéine à régulation allostérique et une molécule effectrice. Deux ensem-

bles de mutations abolissent l'effet glucose : le premier concerne le gène de l'adénylate cyclase impliquée dans la formation de l'adénosine monophosphate cyclique (AMPc), le second implique un gène codant pour une protéine appelée «CAP» pour Protéine Activatrice du Catabolisme. Il a été montré que le glucose freine la production d'AMPc à partir de l'ATP et maintient un très faible niveau d'AMPc. Lorsque le glucose diminue, la concentration en AMPc augmente, or cette molécule peut former un complexe spécifique avec la protéine CAP. Il en résulte une modification de la structure tridimensionnelle de cette protéine et le complexe est capable de se fixer sur l'ADN, au niveau d'une séquence particulière appelée site CAP, située en amont du promoteur. La liaison entraîne une contrainte topologique de la double hélice d'ADN qui favorise l'initiation de la transcription.



1.2 CONCLUSION

Les quelques exemples choisis (il existe d'autres mécanismes qui n'ont pas été décrits ici) montrent que quel que soit le mode de contrôle, positif ou négatif, quelles qu'en soient les modalités dans le détail, le schéma de base est le même.

Les gènes tels que *lacZ*, *lacY*, *trpA* ...codant pour des enzymes sont appelés des gènes de structure : leur produit participe directement à la structure de la cellule ou à son métabolisme (les protéines membranaires, les protéines des ribosomes, enzymes etc...) sont codées par de tels gènes.

A côté de cela, d'autres éléments informatifs interviennent dans le contrôle de l'expression des gènes de structure.

Au niveau transcriptionnel, on envisage des gènes de régulation codant pour des protéines sans fonction enzymatique, des protéines de régulation (le répresseur LacI, la protéine CAP en sont des exemples), ils agissent en *trans* (sur un site pou

vant être éloigné) en se fixant spécifiquement à une séquence précise d'ADN que l'on appellera **région de régulation**. Ces sites assurent une régulation en *cis* sur l'efficacité du promoteur adjacent, situé sur la molécule d'ADN. La relation avec l'environnement intra ou extra cellulaire est assurée par des **molécules effectrices** (le lactose, l'AMPc en sont des exemples).

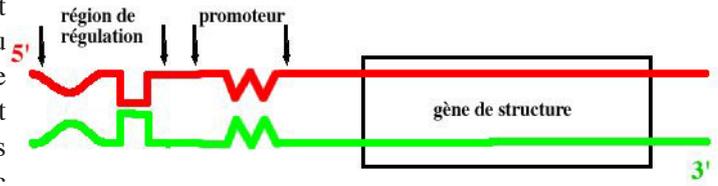
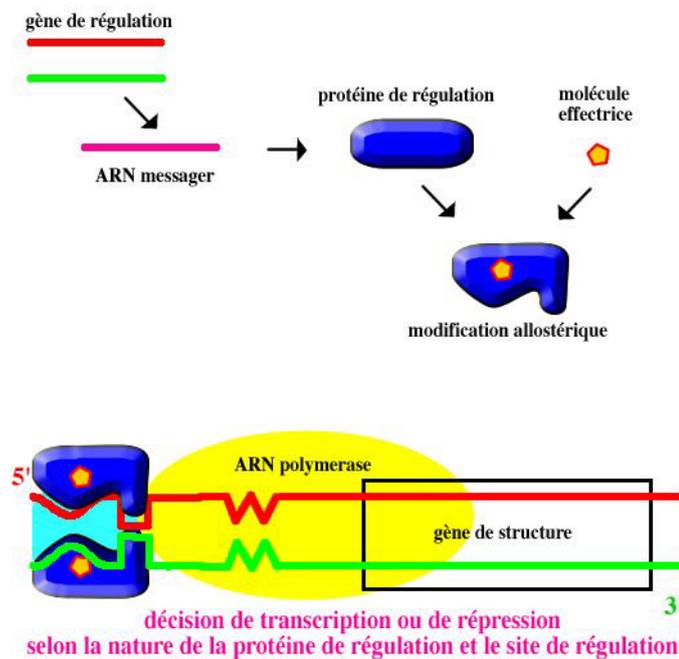


SCHÉMA DE SYNTHÈSE



La souplesse de la régulation est liée aux propriétés de modifications allostériques des protéines de régulation qui, selon leur conformation, assurent ou n'assurent pas leur fonction.

ANNEXE : LA CONJUGAISON BACTÉRIENNE

Certaines souches bactériennes, en raison de leur génotype particulier, sont capables d'émettre des prolongements cytoplasmiques jusqu'à des cellules d'un génotype différent et de faire circuler des éléments génétiques par rapprochement et fusion des membranes. Les éléments génétiques ainsi transférés d'une **cellule «donneuse»** à une **cellule «receveuse»** (l'échange n'est pas réciproque) sont de nature variée.

Le premier élément «échangeable» identifié a été baptisé **facteur F** (comme fertilité).

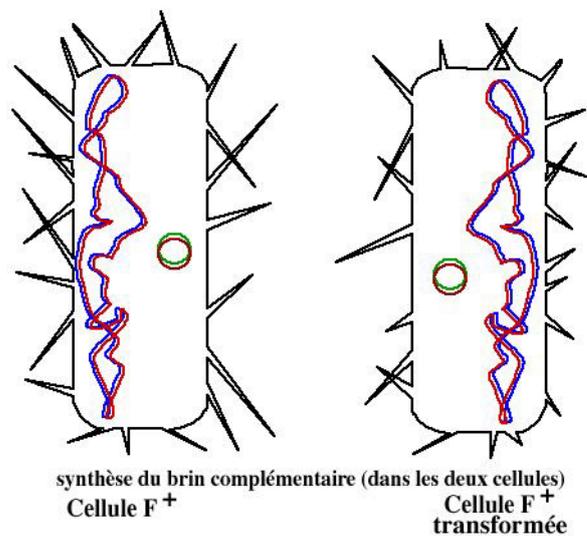
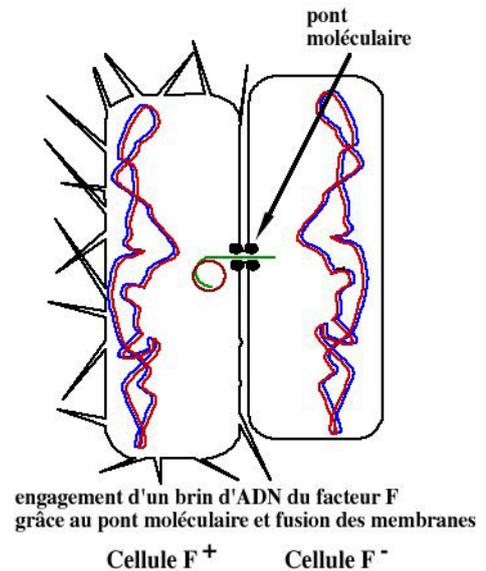
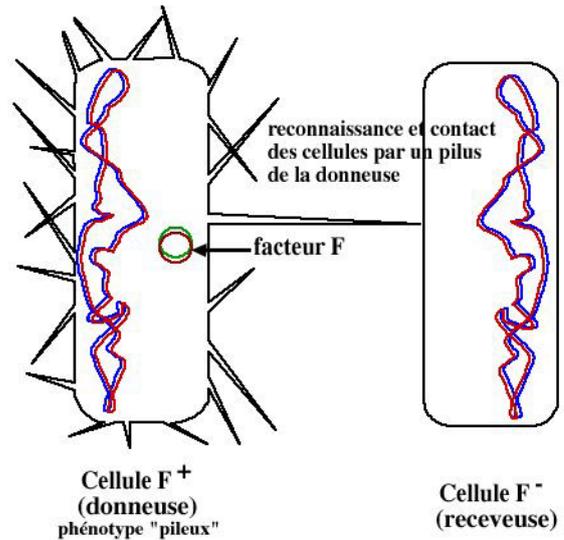
Il s'agit de l'une des «pièces détachées» du génome bactérien que sont les petites molécules d'ADN circulaire, extrachromosomiques, appelées **plasmides**. Dans le cas du facteur F, il s'agit en fait d'un **épisode** qui peut se comporter soit comme un plasmide soit, par un mécanisme de **recombinaison**, s'intégrer au chromosome bactérien. Dans ce dernier cas, son transfert, par conjugaison, peut entraîner un morceau du chromosome. Ce phénomène (rare) permet la formation de **diploïdes partiels**, c'est-à-dire que la bactérie receveuse possède des informations génétiques homologues provenant d'une autre cellule. Le petit fragment de génome ainsi ajouté à l'ensemble résident est appelé **exogénote**.

Sauf s'il possède une **origine de réplication** et se transmet de façon autonome, cet **exogénote** n'est pas stable et se perd en quelques heures. Dans certains cas, il peut cependant s'intégrer au chromosome de la bactérie receveuse par recombinaison homologue.

Remarque : Pour des raisons mécaniques, le contact entre les deux cellules est rapidement rompu. En aucun cas le transfert d'une copie complète du chromosome de la cellule donneuse ne peut avoir lieu. D'où le terme de mérozygote parfois utilisé pour désigner le diploïde partiel.

Le terme de **zygote** rappelle qu'il s'agit de sexualité, c'est-à-dire de mise en commun d'informations génétiques d'origines différentes mais, chez les procaryotes, **la sexualité n'est pas associée à la reproduction** comme chez les Eucaryotes.

Lederberg et Tatum, en 1946, ont apporté la preuve de ce transfert de gènes entre bactéries en réalisant l'expérience consistant à mélanger deux souches porteuses de mutations différentes. Par exemple, une première souche «A» n'est pas capable de se développer dans un milieu minimum



Transfert de matériel génétique d'une cellule F⁺ vers une cellule F⁻

non supplémenté en méthionine et en biotine. Les bactéries sont auxotrophes pour ces composés par perte de fonction des gènes impliqués, l'un dans la biosynthèse de méthionine, l'autre dans la biosynthèse de la biotine. Ces mutations sont symbolisées par un génotype *met⁻ bio⁻*. Une seconde souche, la souche «B», est de génotype *thr⁻ leu⁻ thi⁻*, ici, trois gènes sont mutés et les cellules ont besoin de l'adjonction de thréonine, de leucine et de thiamine au milieu minimum pour se développer. Des bactéries des souches A et B sont mélangées et laissées environ 1 heure dans un milieu contenant tous les éléments nécessaires à la survie des deux (méthionine, biotine, thréonine, leucine et thiamine). La suspension est ensuite étalée sur un milieu minimum (non supplémenté) et, après environ 12h, on voit se développer quelques rares colonies.

La possibilité de réversion de mutation (voir le chapitre concernant la nature du matériel génétique), étant exclu pour trois gènes en même temps dans une même cellule, il faut conclure qu'un **transfert d'allèles sauvages** a été possible d'une cellule à une autre.

Deux explications, non exclusives, permettent de rendre compte du phénotype sauvage : soit l'exogénote, qui présente une grande homologie avec une région précise du chromosome de la cellule receveuse, va s'échanger avec cette région par recombinaison, soit il est transmis à la descendance (parce qu'il possède une origine de réplication par exemple) et il est capable de compléter les allèles défectueux des clones issus de la cellule receveuse. C'est ce dernier cas qui sera utilisé dans l'étude de la régulation.

*Remarque : La conjugaison représente un moyen naturel de transfert de matériel génétique. Par la suite, des souches dites **Hfr** (comme haute fréquence de recombinaison), dont le facteur F est toujours intégré dans le chromosome bactérien, ont été sélectionnées. Actuellement, la transformation, par des plasmides recombinés in vitro, apporte une grande souplesse et une grande efficacité dans la création de diploïdes partiels. De plus, des souches mutées (dans le gène *recA* notamment), incapables de recombinaison (in vivo) permettent d'étudier, sans ambiguïté, les effets de complémentation.*

Pouren savoir plus :

<http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/>
<http://nte-serveur.univ-lyon1.fr/tpgeneticmic/>

