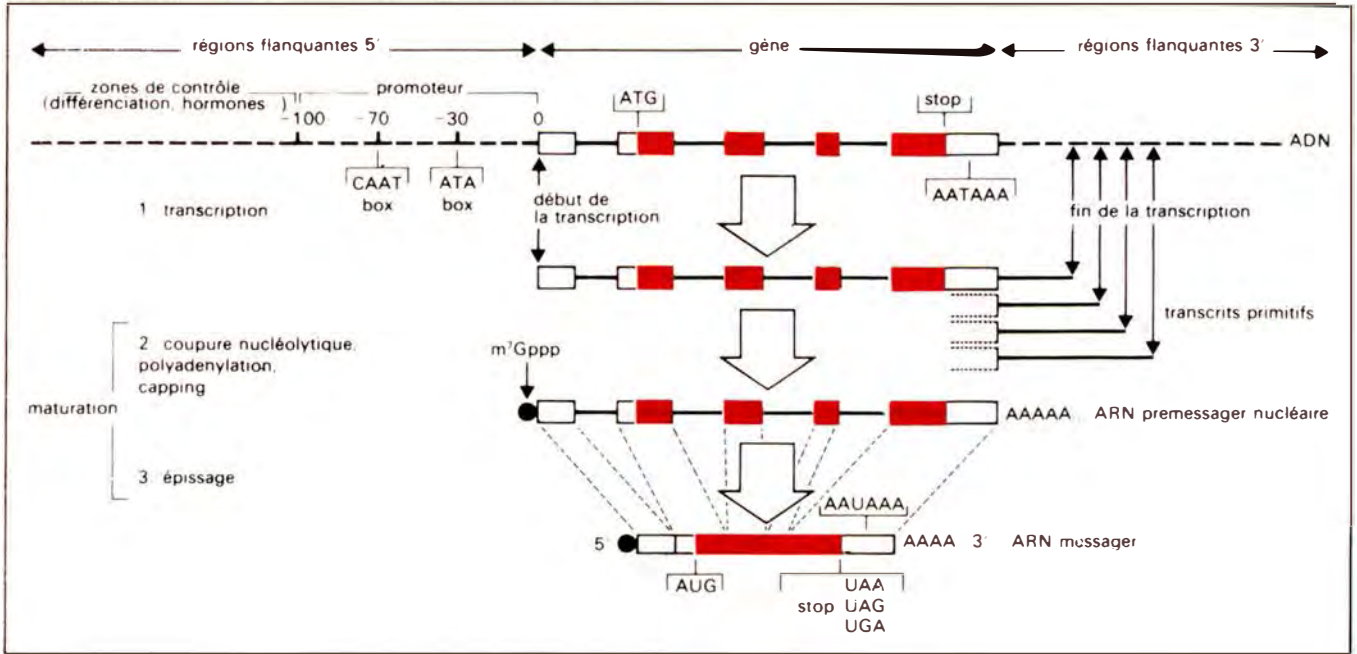


# Structure des gènes chez les eucaryotes



**Exon** : la partie du gène qui persiste dans l'ARN messager. En amont du signal de début de traduction (AUG) et en aval de celui de fin de traduction (UAA, UAG ou UGA), les exons ne sont pas traductibles en protéine et sont appelés « extrémités non codantes 5' et 3' du messenger ». Les exons contenant le code traduit en protéine sont représentés en rouge.

**Intron** : segment du gène qui sera excisé (processus d'épissage) lors de la maturation des ARN pré-messagers nucléaires en ARN messager.

**Régions intergéniques**, avec le promoteur et les régions régulatrices en amont du gène (régions flanquantes 5'), les hypothétiques signaux de fin de transcription en aval (régions flanquantes 3'). Le gène débute, par définition, au site d'initiation de la transcription (début du premier exon), et se termine à la fin du dernier exon.

Le promoteur est la zone de fixation de l'ARN polymérase, enzyme catalysant la transcription du gène, c'est-à-dire son recopiage en ARN. Il comporte des séquences très conservées, parmi lesquelles la « boîte ATA » (ATA box), située environ 30 bases en amont du site d'initiation de la transcription (marqué « O » sur le schéma), et, à un moindre titre, la « boîte CAAT » (CAAT box), située plus en amont (position -70 sur le schéma).

Les régions flanquantes 5' contiennent des séquences qui sont reconnues par les hormones et leur récepteur et par des facteurs diffusibles intervenant dans l'expression ou l'extinction d'un gène au cours de la différenciation cellulaire.

Les régions flanquantes 3' contiennent peut-être des signaux, probablement multiples, au niveau desquels cesse la transcription, en aval du gène.

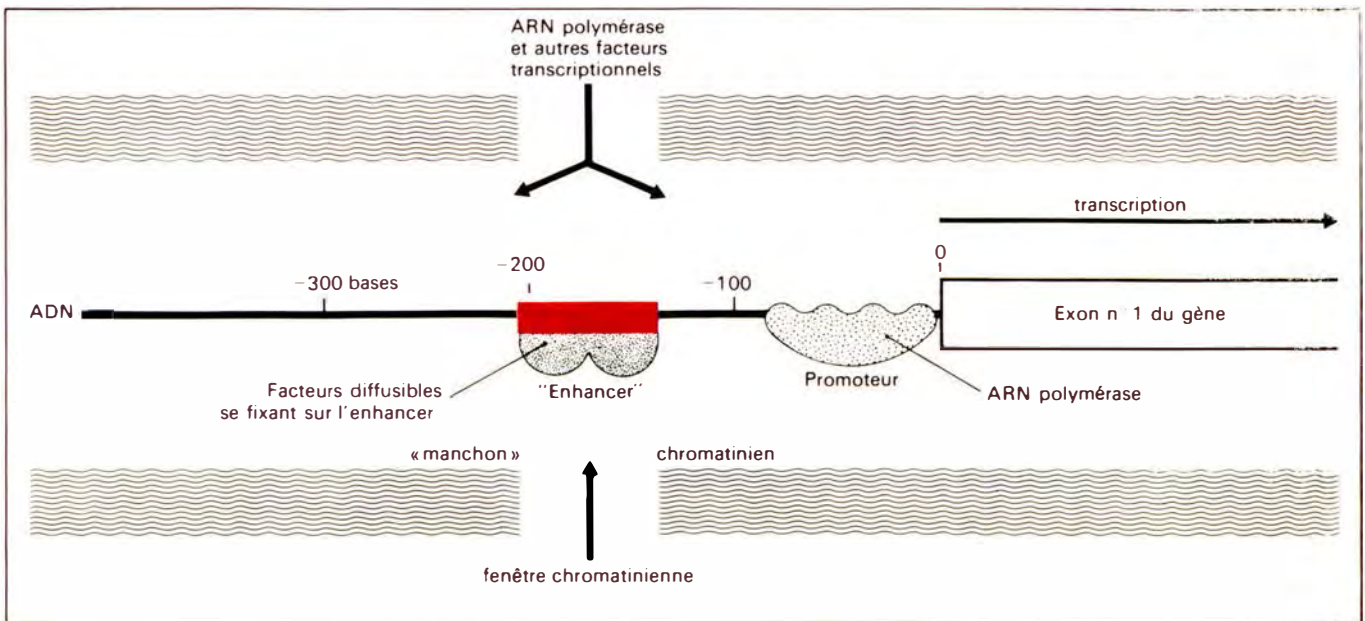
Les transcrits primitifs sont les ARN initiaux débutant au site d'initiation (O) et s'interrompant en différents sites en aval du gène.

La coupure nucléolytique et la polyadénylation, ces transcrits sont ensuite coupés environ 18-20 bases après un signal AAUAAA (c'est-à-dire AATAAA au niveau de l'ADN), l'extrémité 3' ainsi formée étant allongée par des résidus d'acide adénylique (une centaine).

Le « capping » : très précocement l'extrémité 5' du transcrit (c'est-à-dire le site d'initiation de la transcription) est bloquée par un « chapeau » (cap) formé d'un acide guanylique méthylé sur un azote en position 7.

Épissage, ensemble des phénomènes aboutissant à l'excision des introns. Les jonctions introns-exons ont une structure très conservée : sites reconnus d'épissage. **A.K.**

## *Les séquences stimulatrices de la transcription ou « enhancers »*



Dans les cellules, L'ADN n'est jamais « nu », mais complexé à de nombreuses protéines qui forment la structure chromatinienne, sorte de manchon entourant les gènes. Pour que l'ARN polymérase et les autres facteurs agissant sur la transcription aient accès aux régions de contrôle du gène et à son promoteur, il faut que cette structure chromatinienne soit « relâchée ».

Les séquences stimulatrices (*enhancers*, en anglais) semblent capables de modifier la chromatine, créant en elle une sorte de fenêtre d'accès que peuvent emprunter l'ARN polymérase et d'autres facteurs transcriptionnels. La polymérase cheminerait ensuite le long du brin d'ADN jusqu'à rencontrer un promoteur au niveau duquel elle déclencherait la transcription (initiation) du gène adjacent. Les séquences stimulatrices pourraient aussi créer des contraintes de torsion dans la dou-

ble hélice d'ADN, facilitant la dissociation des brins au cours de la transcription; la différence importante entre les deux types d'activité est que la première, c'est-à-dire la création d'une « fenêtre » chromatinienne, persiste lorsque l'enhancer est situé à grande distance du promoteur, en amont aussi bien qu'en aval, alors que cette deuxième activité qu'est la contrainte physique exige que la distance entre l'enhancer et le promoteur ne dépasse pas quelques centaines de bases.

Pour exercer son influence, il semble que la séquence stimulatrice doive fixer un facteur diffusible, probablement protéique, qui peut être spécifique de la différenciation tissulaire, ou bien représenter un récepteur hormonal lié à l'hormone correspondante.

Les propriétés fonctionnelles de ces séquences sont d'être actives à

grande distance du promoteur, en amont aussi bien qu'en aval de lui, et de n'être pas « polarisées », ce qui signifie que le brin d'ADN les portant peut être retourné sans perte de l'activité stimulatrice.

On trouve de tels *enhancers* dans les régions de contrôle des virus à ARN (au niveau des LTR, ou *Long Terminal Repeats* qui entourent l'ADN proviral), ou à ADN; dans le grand intron des gènes réarrangés des immunoglobulines, en aval du promoteur; en amont de gènes cellulaires dont l'expression dépend de la différenciation cellulaire.

Les séquences stimulatrices constituent donc, en résumé, des éléments fondamentaux de la régulation de l'expression du génome des eucaryotes, leur importance fonctionnelle venant de leur propriété d'être régulées par des facteurs spécifiques de la différenciation tissulaire, voire par des hormones. **A. K.**

## Régulation positive et négative des « enhancers »

Le groupe de Pierre Chambon, à Strasbourg, a joué et continue de jouer un rôle déterminant dans la « dissection moléculaire » des phénomènes de régulation de la transcription des gènes eucaryotes. Dans deux articles récents, il précise et développe un concept qu'il a lui-même proposé il y a deux ans [1], celui de la régulation négative des séquences stimulatrices de la transcription, ou enhancers (voir le Lexique de médecine/sciences n° 2, vol. 1, p. 105). Ces enhancers, d'abord découverts dans des virus, sont en fait largement répandus dans le génome d'une grande variété d'êtres vivants où ils jouent un rôle essentiel dans le contrôle de l'expression des gènes au cours du développement, de la différenciation ou des processus de régulation hormonale. Ils peuvent en effet être « activés » lorsqu'ils sont liés à des facteurs diffusibles, probablement protéiques, caractéristiques d'un type et d'un stade particulier de différenciation tissulaire ou d'une stimulation hormonale (voir Lexique de médecine/sciences n° 6, vol. 1, p. 335). La découverte que ces enhancers, et par conséquent la transcription des gènes qu'ils contrôlent, peuvent être inhibés par certaines substances telle la protéine  $E_1A$  codée par un gène d'expression précoce de l'adénovirus [1], était donc déjà très importante puisqu'elle permettait de faire l'hypothèse que, dans les cellules, une telle régulation négative pouvait affiner le contrôle de l'expression des gènes. René Hen et coll. viennent de démontrer que de telles substances cellulaires, homologues fonctionnels d' $E_1A$ , sont présentes dans les cellules embryonnaires et expliquent très probablement l'impossibilité d'infecter ces cellules par des virus dont les enhancers sont en effet inhibés par  $E_1A$ . Le virus du polyome, par exemple, possède un enhancer régulé négativement par  $E_1A$  et ne peut pas se répliquer dans des cellules embryonnaires. Un mutant de ce virus qui a acquis la propriété d'infecter des cellules embryonnaires a, en revanche,

perdu la propriété d'être inhibé par  $E_1A$  [2].

Emiliana Borelli et coll. ont étendu cette notion de régulation négative à un enhancer cellulaire, celui responsable de la stimulation du gène des immunoglobulines dans le lymphocyte B (voir le Lexique de médecine/sciences n° 8, vol. 1, p. 442). Un gène placé sous le contrôle d'un tel enhancer s'exprime peu dans une cellule non lymphocytaire comme des fibroblastes de souris. Une expérience de « compétition » (figure 1) permet de démontrer que cette faiblesse de la transcription du gène introduit dans

les cellules fibroblastiques est due, au moins en partie, à la présence de facteurs inhibant la séquence stimulatrice d'immunoglobuline [3]. Le principe de cette expérience est d'introduire dans la cellule le gène contrôlé par la séquence stimulatrice étudiée et un excès de cette séquence non couplée au gène. S'il existe des facteurs inhibiteurs, ceux-ci vont être « consommés » par l'excès de séquences libres, et le enhancer couplé au gène analysé sera déréprimé (figure 1). La protéine  $E_1A$ , dont nous avons vu plus haut qu'elle inhibait certains enhancers, par exemple

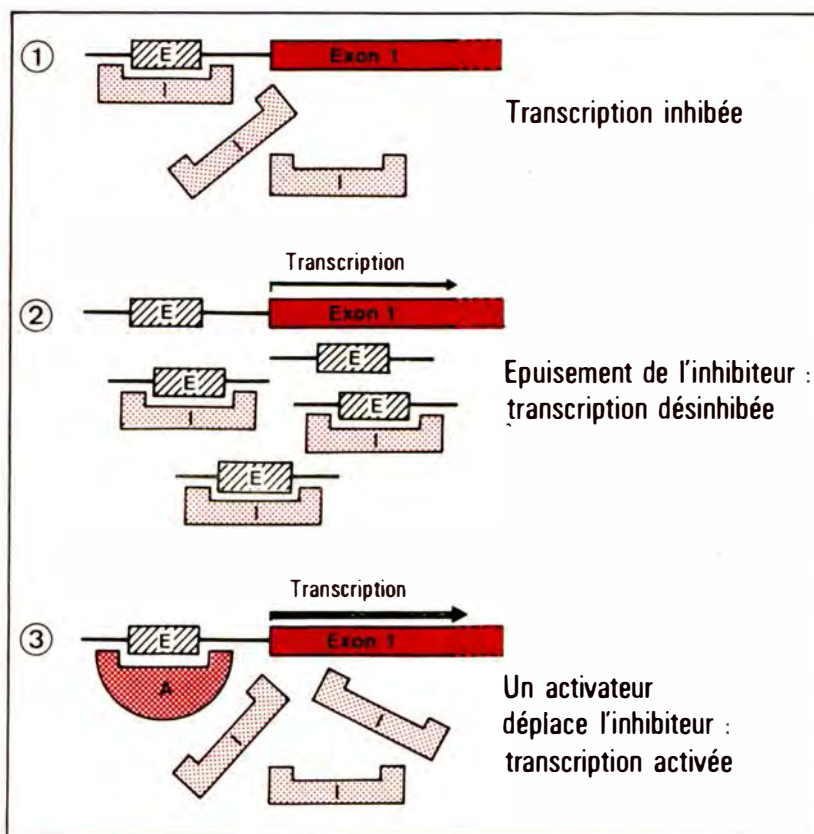


Figure 1. **Inhibition, désinhibition et activation des « enhancers ».** En (1), une séquence stimulatrice (E) (ou enhancer) est inhibée par un inhibiteur (I): le gène n'est pas transcrit. En (2), un excès de la séquence (E) consomme l'inhibiteur (I) qui ne peut plus inhiber le enhancer (E) contrôlant la transcription du gène utilisé comme témoin: ce gène est transcrit. En (3), un activateur (A) se fixe sur le enhancer (E), déplace l'inhibiteur et stimule la transcription.

celui du virus polyome, agit comme un activateur sur la séquence stimulatrice du gène d'immunoglobuline transfectée dans des fibroblastes de souris, probablement en levant l'inhibition par les facteurs de régulation négative (figure 1). Une même protéine peut donc être inhibitrice pour certains enhancers et activatrice pour d'autres. Un tel modèle, s'il était généralisé, permettrait de simplifier un peu les mécanismes impliqués dans la régulation des gènes spécifiques d'un tissu différencié : les facteurs « activateurs » des séquences stimulatrices de gènes exprimés dans ces tissus pourraient être aussi les « inhibiteurs » des enhancers d'autres gènes dont l'expression n'est pas spécifique du tissu.

L'équipe de W. J. Rutter vient d'obtenir, sur un autre modèle, des résultats très proches de ceux d'E. Borelli. Un gène dont l'expression est contrôlée par les séquences régulatrices du gène de l'insuline ne s'exprime pas dans des cellules fibroblastiques de

singe, sauf si ces cellules contiennent en même temps un large excès du enhancer du gène de l'insuline, qui est supposé épuiser les facteurs inhibiteurs selon le schéma de la figure 1 [4].

Enfin, deux autres équipes ont rapporté que les séquences stimulatrices pouvaient être inhibées par un autre mécanisme que la fixation à leur niveau de facteurs diffusibles inhibiteurs. Il existe en amont des gènes *c-myc* [5] et de l'insuline de rat [6] des « silencers » (voir médecine/sciences n° 6, vol. 1, p. 335) inhibant en cis (c'est-à-dire lorsqu'ils sont localisés sur le même double brin d'ADN) l'activité stimulatrice de enhancers. Le rôle de ces séquences pourrait être, notamment, d'isoler les gènes de l'influence de séquences stimulatrices appartenant à des gènes adjacents.

Les mécanismes résumés ici pourront sembler bien complexes au lecteur, et, de fait, ils le sont. Ils constituent cependant, touche après touche,

l'ébauche d'une des principales avancées scientifiques de notre siècle : le matériel génétique, ce qu'il est et la manière dont il fonctionne.

A. K.

1. Borelli E, Hen R, Chambon P. Adenovirus-2 E<sub>1</sub> A products repress enhancer-induced stimulation of transcription. *Nature*, 1984; 312 : 608-12.
2. Hen R, Borelli E, Fromental C, Sassone Corsi P, Chambon P. A mutated polyoma virus enhancer which is active in undifferentiated embryoma carcinoma cells is not repressed by adenovirus-2 E<sub>1</sub> A products. *Nature* 1986; 321 : 249-51.
3. Borelli E, Hen R, Wasyluk C, Wasyluk B, Chambon P. The immunoglobulin heavy chain enhancer is stimulated by the adenovirus type-2 E<sub>1</sub> A products in mouse fibroblasts. *Proc Natl Acad Sci USA* 1986; 83 : 2846-9.
4. Nir U, Walker MD, Rutter, WJ. Regulation of rat insulin 1 gene expression: Evidence for negative regulation in nonpancreatic cells. *Proc Natl Acad Sci USA* 1986; 83: 3180-4.
5. Remmers EF, Yang JQ, Marcu KB. A negative transcriptional control element located upstream of the murine *c-myc* gene. *EMBO J* 1986; 5: 899-904.
6. Laimins L, Holmgren M, Khoury G. Transcriptional silencer element in rat repetitive sequences associated with the rat insulin 1 gene locus. *Proc Natl Acad Sci USA*. 1986; 83: 3151-5.

## Enhancers et Silencers

**M**édecine/sciences expliquait récemment (glossaire du n° 1, p. 11), ce qu'étaient les « enhancers » ou séquences stimulatrices de la transcription découvertes en 1981. Depuis la parution de cet article, un autre type d'élément régulateur de la transcription des gènes a été identifié: il s'agit des séquences extinctrices (ou inhibitrices) appelées « silencers » en anglais [1]. Comme leur nom l'indique, ces séquences agissent en maintenant silencieux des gènes qui doivent être réprimés.

Exactement comme les « enhancers », ces « silencers » sont actifs à distance des gènes (jusqu'à plusieurs milliers

de base des promoteurs) en amont aussi bien qu'en aval d'eux et sont dépourvus de polarité, ce qui signifie qu'ils demeurent pleinement actifs lorsqu'ils sont excisés et réinsérés dans l'orientation opposée sur le double brin d'ADN. Comme dans le cas des « enhancers », leur action, c'est-à-dire leur pouvoir de réprimer les gènes, dépend de la présence de facteurs diffusibles, probablement protéiques, qui pourraient se lier à eux, et ainsi les activer.

La différenciation cellulaire étant une suite d'activation de certains gènes et de répression d'autres, la découverte de ces séquences extinctrices et de leur régulation par des facteurs diffusibles « extincteurs » consti-

tue probablement une pièce maîtresse de notre connaissance des divers éléments du contrôle de l'expression du génome.

A. K.

1. Brand A H, Breeden L, Abraham J, Sternglanz M, Rand Nasmyth K. Characterization of a « Silencer » in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* 1985; 41: 41-8.



John Libbey  
EUROTEXT  
LONDON · PARIS

Paris-Londres  
EDITIONS MEDECINE-SCIENCES

A PARAÎTRE  
1990

### Circulation cérébrale et vieillissement

A. BÈS, G. GÉRAUD  
Collection *Current Problems  
in Neurology*

### La tontine

M. LELART  
Collection *Sciences en Marche*  
Co-édition AUPELF/UREF/  
John Libbey Eurotext

### L'entrepreneuriat en Afrique francophone

G. HÉNAULT, R. M' RABET  
Collection *Actualité Scientifique*  
Co-édition AUPELF/UREF/  
John Libbey Eurotext

### Communications placentaires : aspects biochimique morphologi- que et cellulaire

L. CÉDARD, E. ALSAT,  
J.C. CHALLIER, G. CHAOUAT,  
A. MALASSINÉ  
Co-édition INSERM/  
John Libbey Eurotext

### Stratégie thérapeutique dans la maladie de Hodgkin

M.H. AMAR, H. SANCHO-  
GARNIER  
Co-édition INSERM/  
John Libbey Eurotext

6, rue Blanche, 92120 Montrouge  
Tél. : (1) 47.35.85.52.

## **E**nhancer **et activation transcriptionnelle : modèle de la clé ou du bouton ?**

L'activation de la transcription d'un gène par un *enhancer* exige très probablement un contact direct entre les protéines fixées sur le *enhancer* et celles (dont l'ARN polymérase II) participant à la constitution du complexe d'initiation (voir *mini-synthèses*, m/s n° 7, vol. 3, p. 428, n° 8, vol. 3, p. 487 et n° 9, vol. 3, p. 546). Lorsque ces deux complexes nucléoprotéiques sont éloignés l'un de l'autre, il est probable que l'ADN peut faire une boucle, permettant ainsi un rapprochement entre eux. Le problème se pose alors des mécanismes par lesquels est activée la transcription. Schématiquement, deux modèles peuvent être opposés, celui de la clé et celui du bouton. Dans le premier modèle, il doit y avoir une complémentarité étroite entre la protéine activatrice et la cible, comme entre une enzyme et un substrat, un récepteur et un ligand, une clé et une serrure. On peut concevoir que la protéine activatrice soit, de fait, une enzyme de modification de l'un des composants du complexe d'initiation de la transcription. L'autre modèle postule au contraire qu'il n'y a pas plus de complémentarité nécessaire entre l'activateur et la cible qu'entre ce qui pousse un bouton et le bouton ! C'est vers ce type de mécanisme qu'orientent plusieurs articles récents, utilisant tous les possibilités remarquables de la génétique de la levure. Un premier groupe de résultats montre que, chez la levure, l'exigence structurale pour qu'une protéine se liant à un *enhancer* active la transcription est faible. Les gènes actifs dans le catabolisme du galactose par *Saccharomyces cerevisiae* sont sous le contrôle d'une protéine activatrice GAL 4. En l'absence de galactose, cette protéine est complexée à une

protéine GAL 80 et n'est pas fixée à l'ADN ; en présence de galactose, le complexe GAL 80-GAL 4 se dissocie, GAL 4 se fixe sur un élément d'ADN particulier (un UAS, *upstream activator sequence*) et active la transcription des gènes du catabolisme du galactose. L'activation transcriptionnelle est en fait indépendante de la nature précise du site de liaison à l'ADN et il est possible, par des expériences de recombinaison génétique, de remplacer la région N-terminale de liaison à l'ADN de la protéine GAL 4 par une autre région protéique se fixant à un élément d'ADN différent : si ce dernier élément est présent en amont du gène contrôlé par GAL 4, la protéine hybride gardera son potentiel d'activer la transcription. Il est aussi possible de considérablement modifier les 780 acides aminés du côté COOH (la protéine a 881 résidus) sans supprimer l'activation transcriptionnelle. En fait, il suffit de fusionner au nouveau domaine de liaison à l'ADN une chaîne polypeptidique acide pour reconstituer, avec une grande fréquence, un activateur transcriptionnel [1, 2] (*figure 1*).

Deux autres articles plus récents [3, 4] montrent que la protéine GAL 4 de levure est parfaitement capable de stimuler, dans des cellules de mammifères, la transcription d'un gène en amont duquel a été placé l'élément d'ADN UAS sur lequel elle se fixe. Ce résultat témoigne aussi de la conservation entre les espèces des systèmes d'activation transcriptionnelle. L'expérience consiste à introduire dans une cellule de mammifère un gène de GAL 4 placé sous le contrôle d'un promoteur actif, et un gène « rapporteur » (par exemple celui de la chloramphénicol acétyltransférase) placé sous le contrôle de séquences de contrôle d'un gène de

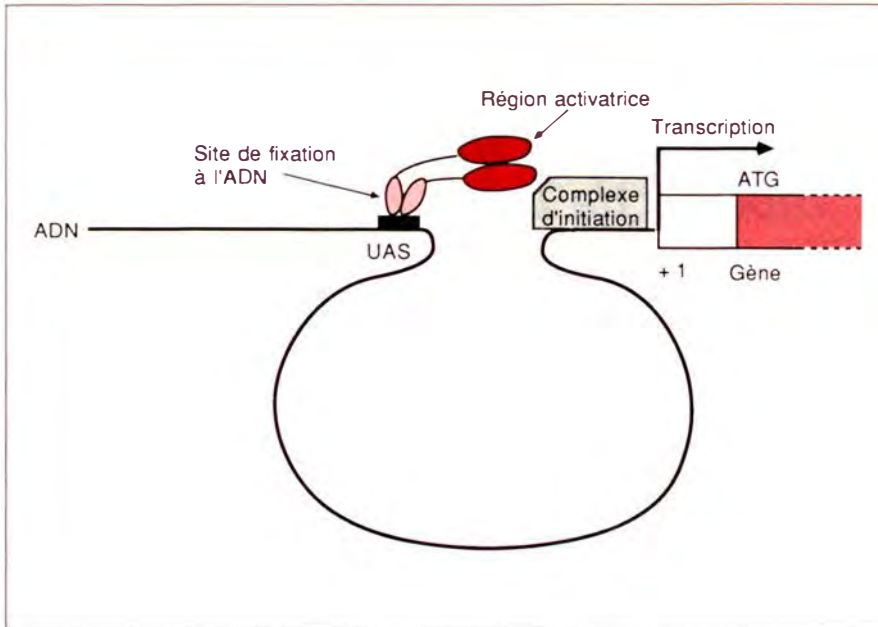


Figure 1. **Modèle d'activation de la transcription par des protéines fixées à un site enhancer.** Une protéine est fixée à un site enhancer (UAS = upstream activating sequence). Sa région N-terminale contient le domaine de fixation à l'ADN, sa partie C-terminale « touchant » le complexe d'initiation de la transcription et l'activant. Il n'y a aucune spécificité de l'UAS et du domaine de fixation à l'ADN de la protéine pour un gène donné : il suffit, que, par un domaine quelconque, la protéine se fixe en un site localisé à proximité relative du promoteur pour que l'activation ait lieu. De très nombreuses séquences protéiques greffées à un domaine de fixation à l'ADN sont capables de reconstituer une protéine activatrice, à condition que dominant les résidus acides.

mammifère, séquences auxquelles a été ajouté l'élément UAS. Il est aussi possible de remplacer la région E du récepteur des œstrogènes (région de fixation de l'hormone jouant également un rôle dans l'activation transcriptionnelle) par la région C-terminale acide de GAL4 tout en conservant l'activation transcriptionnelle des gènes en amont desquels se fixe cette molécule hybride [4]. Il est ainsi clair que la spécificité structurale requise pour que la protéine fixée à un site *enhancer* active la transcription est faible, ressemblant bien plus à celle du bouton qu'à celle de la clé.

On peut rapprocher de ces résultats ceux récemment rapportés de l'activation de promoteurs de levures par les produits des oncogènes *fos* et *myc* fixés en un site d'ADN en amont du promoteur [5]. Là encore des protéines hybrides sont fabriquées, comportant l'essentiel de la séquence des oncogènes, associée à un site de fixation à l'ADN. Compte tenu de la

grande diversité des séquences protéiques capables, dans de telles conditions, d'activer la transcription [3, 4], il n'est d'ailleurs pas évident que ces observations soient transposables au mode d'action réel des oncogènes *fos* et *myc* !

A.K.

1. Ma J, Ptashne M. A new class of yeast transcriptional activators. *Cell* 1987; 51 : 113-9.
2. Gill G, Ptashne M. Mutants of GAL4 protein altered in an activation function. *Cell* 1987; 51 : 121-6.
3. Kakidani H, Ptashne M. GAL4 activates gene expression in mammalian cells. *Cell* 1988; 52 : 161-7.
4. Webster N, Jin JR, Green S, Hollis M, Chambon P. The yeast UAS<sub>6</sub> is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 transactivator. *Cell* 1988; 52 : 169-78.
5. Lech K, Anderson K, Brent R. DNA-bound Fos proteins activate transcription in yeast. *Cell* 1988; 52 : 179-84.

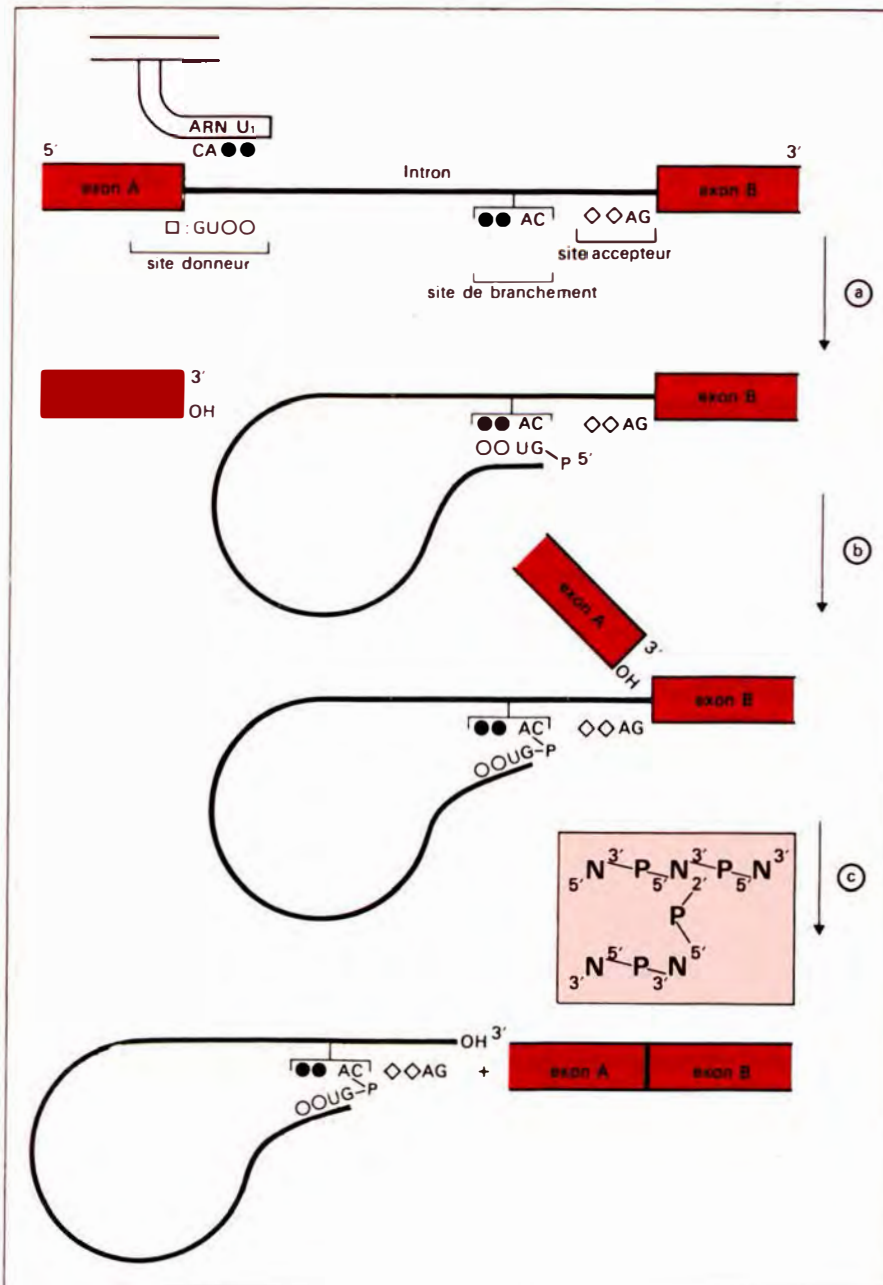
## Epissage des ARN messagers nucléaires

L'épissage, ou splicing, des ARN messagers dans le noyau correspond à l'excision précise des introns, suivi de la ligature ordonnée des exons entre eux, le processus aboutissant au messager cytoplasmique traductible. Le mécanisme exact de ces opérations compliquées n'est pas encore connu; cependant, ses principales étapes ont été récemment élucidées.

1. Sites importants pour un épissage correct. Ils sont au nombre de trois :

Le site donneur, ou site 5', de l'épissage correspond à la jonction exon-intron en position 5' de l'intron. Les deux premières bases de l'intron sont G,T sur l'ADN (et G,U sur l'ARN, U étant l'uracile), les bases environnantes présentant des variations plus ou moins importantes par rapport à une séquence consensus qui est celle de l'extrémité 5' du petit ARN nucléaire U<sub>1</sub>. Le site accepteur, ou site 3', de l'épissage correspond à la jonction intron-exon en position 3' de l'intron. Les deux dernières bases sont AG, précédées de bases pyrimidiques (on peut, là aussi, parler de « séquence consensus » de tous les sites accepteurs).

Le site de branchement, situé environ à 30-40 bases en amont du site accepteur, est complémentaire du site donneur, cette complémentarité n'étant cependant jamais parfaite.



## 2. Les différentes étapes de l'épissage: formation du « lasso »

(a) Le site donneur est clivé, une hybridation avec le petit ARN U<sub>1</sub> pouvant jouer un rôle essentiel dans la reconnaissance du site de clivage. L'extrémité 5' de l'intron est dès lors libre, et peut s'hybrider à la séquence complémentaire du site de branchement.

(b) Une liaison covalente phosphodiester s'établit entre l'hydroxyle en position 2' de l'adénosine du site de branchement et le phosphate de l'extrémité 5' libre de l'intron, établissant la structure branchée schématisée dans le cadre. Dans les acides nucléiques, la succession des nucléosides se fait par l'intermédiaire de ponts phosphodiester entre les hydroxyles 5' et 3' des riboses adjacents. Le lasso se forme par établissement de cette liaison inhabituelle 5'-2'.

(c) Le site accepteur d'épissage est précisément clivé, une liaison s'établissant entre l'hydroxyle libre 3' de l'exon A et le phosphate 5' de l'exon B.

Les produits de la réaction sont l'intron avec sa structure branchée en lasso, qui sera probablement dégradé, et les deux exons correctement liés.

### 3. Rôle (s) et anomalies de l'épissage: le rôle des introns est inconnu, mais il est impératif qu'ils soient excisés de l'ARN, car ils ne peuvent pas être traduits en protéine et, s'ils étaient présents dans le messager cytoplasmique, ils introduiraient des arrêts précoces de la traduction, et donc de la synthèse des protéines. Des mutations peuvent modifier les signaux indispensables à l'épissage, ou encore faire apparaître dans les exons ou les introns de nouveaux sites de coupure. Les conséquences de ces anomalies peuvent être une dégradation rapide de l'ARN incapable de subir sa maturation normale, ou l'arrivée dans le cytoplasme d'ARN qui ne peuvent être traduits en la protéine biologiquement active

A.K.

Les symboles vides ou pleins représentent des bases appariées.  
*N* = Nucléoside; *P* = Phosphate; *OH* = radical hydroxyle;  
*A* = Adénine; *C* = Cytosine; *G* = Guanine; *T* = Thymine.

Le cadre en c montre un détail schématisé de la liaison entre l'extrémité 5' phosphate de l'intron et l'hydroxyle 2' de l'adénosine du site de branchement.

## Épissage des ARN et ribozymes

Le nom de ribozyme est donné aux molécules d'ARN douées d'activité catalytique, se comportant donc comme des enzymes (*cf m/s 1985; 1: 107*). Le concept est d'une extrême importance théorique, notamment pour les hypothèses concernant l'origine de la vie. Une même molécule d'ARN, capable tout à la fois de porter une information génétique « codée » et de catalyser des réactions biologiques, a pu constituer le premier système auto-replicatif... et donc le premier embryon de la vie. Les deux premiers exemples connus de tels ribozymes sont ceux de l'ARN constituant l'élément actif de la ribonucléase P d'*Escherichia coli* [1] et de l'intron de l'ARN ribosomique de *Tetrahymena thermophila* [2]. Dans ce dernier cas, Cech *et al.* montraient, qu'en présence de GTP ou de guanosine, l'intron était capable de s'autoexciser, puis de se cycliser, selon le schéma de la *figure 1*. Stricto-sensu, une modification « autocatalytique » d'une molécule ne signifie pas qu'elle est un « catalyseur » dont la définition est d'accélérer une réaction sans être lui-même modifié au terme de celle-ci.

Par une succession de traitements alcalins ménagés, il est cependant possible de transformer l'intron cyclisé de *Tetrahymena* en un ARN linéaire, incapable de se cycliser, mais doué lui-même d'une activité de dégradation de l'ARN [3, 4]. Le mécanisme chimique de cette action « ribonucléasique » est une « attaque nucléophile »\* d'un pont phosphodiester par un radical hydroxyle libre libérant une extrémité 5' phosphate et 3' OH. Il s'agit d'une « trans-estérification », sans modification du nombre total de liaisons phosphodiester dans le milieu de réaction, et ne nécessitant

pas d'apport extérieur d'énergie (*figure 1*). Au contraire, les autres ribonucléases connues libèrent des extrémités 5' OH et 3' phosphate... à l'exception de la ribonucléase P dont l'activité est due à sa partie ARN. Les ARN de plusieurs systèmes ont donc une activité catalytique vraie dont le mécanisme est une trans-estérification suivant l'attaque nucléophile par un radical hydroxyle libre de ponts phosphodiester entre le phosphate 5' et l'hydroxyle 3' (*figure 1*).

Des études ultérieures devaient montrer que plusieurs introns mitochondriaux appartenant, comme l'intron ribosomique de *Tetrahymena*, à la classe 1 (cette classe d'introns est définie par de nombreuses similitudes structurales) étaient excisés selon le même modèle que l'intron de *Tetrahymena* [5]. L'épissage des introns nucléaires est cependant différent, passant par la formation d'un « lasso » formé par la liaison de l'acide guanylique en 5' de l'intron avec un résidu adénylique situé au niveau du « site de branchement » à quelques dizaines de bases de l'extrémité 3' de l'intron (*figure 2 et m/s 1985; 1: 158-9*). Cette liaison phosphodiester est inhabituelle en ce qu'elle s'établit entre les hydroxyles 5' et 2' (et non 5' -3') des résidus liés. Plusieurs équipes ont récemment démontré que certains introns mitochondriaux dits de classe 2 (eux aussi définis par leurs caractéristiques structurales) étaient excisés selon ce même modèle, et que l'ensemble constitué de l'excision, l'épissage des exons et la formation du lasso était autocatalytique [6-8]. Contrairement aux réactions intéressant l'intron de *Tetrahymena*, la présence de guanosine est inutile. L'hypothèse formulée [4, 5, 8] est que la première étape de l'épissage pourrait être ici l'attaque nucléophile de la liaison exon-intron 5' par l'hydroxyle 2' de l'acide adénylique du site de branchement (*figure 2*). On sait cependant que les phénomènes d'épissage des

introns nucléaires et d'autres introns mitochondriaux de classe 2 ne semblent pas autocatalytiques et nécessitent la présence de nombreux facteurs cellulaires, protéines et petits ARN. Les réactions d'excision des introns et d'épissage pourraient néanmoins présenter une profonde unité, impliquant chaque fois l'intervention catalytique de l'intron lui-même. Cette action de type « ribozymique » serait cependant insuffisante dans de nombreux cas, ceux notamment des introns complexes des ARN nucléaires; dans ces cas, il faudrait de plus que des facteurs cellulaires divers interviennent sur la conformation de l'intron pour permettre son excision. La modulation de la conformation ainsi imposée par des facteurs cellulaires pourrait expliquer les phénomènes d'épissages alternatifs (*m/s 1985; 1: 442-3*). Ainsi Solnick a-t-il pu récemment démontrer que l'on pouvait, in vivo et in vitro, reproduire de tels épissages alternatifs en imposant à des transcrits des conformations en « épingles à cheveux » dans lesquelles étaient séquestrés des exons [9].

Il semble ainsi que les ARN soient des molécules douées de propriétés catalytiques qui pourraient être à la base de l'épissage des introns de différentes classes.

A. K.

1. Guerrier-Tadaka C, Gardiner K, Marsh T, *et al.* The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 1983; 35 : 849-57; *Science* 1984; 223 : 285-6.
2. Zaug A J, Grabowski P J, Cech T R. Autocatalytic cyclization of an excised intervening sequence is a cleavage-ligation reaction. *Nature* 1983; 301 : 578-83.
3. Zaug A J, Cech T R. The intervening sequence of *Tetrahymena* is an enzyme. *Science* 1986; 231 : 470-4.
4. Westheimer F H. Polyribonucleic acids as enzymes. *Nature* 1986; 319 : 534-6.
5. Cech T R. The generality of self-splicing RNA : relationship to nuclear mRNA splicing. *Cell* 1986; 44 : 207-10.
6. Arnberg A C, Van Der Horst G, Tabak H F. Formation of lariats and circles in self-splicing of the precursor to the large ribosomal RNA of yeast mitochondria. *Cell* 1986; 44 : 235-42.

\* Nucléophile : Est « nucléophile » un groupe chimique possédant un doublet d'électron à la recherche d'un noyau déficient en électron avec lequel il va réagir. Une « attaque nucléophile » se réfère à l'action d'un tel résidu, par exemple un radical OH.



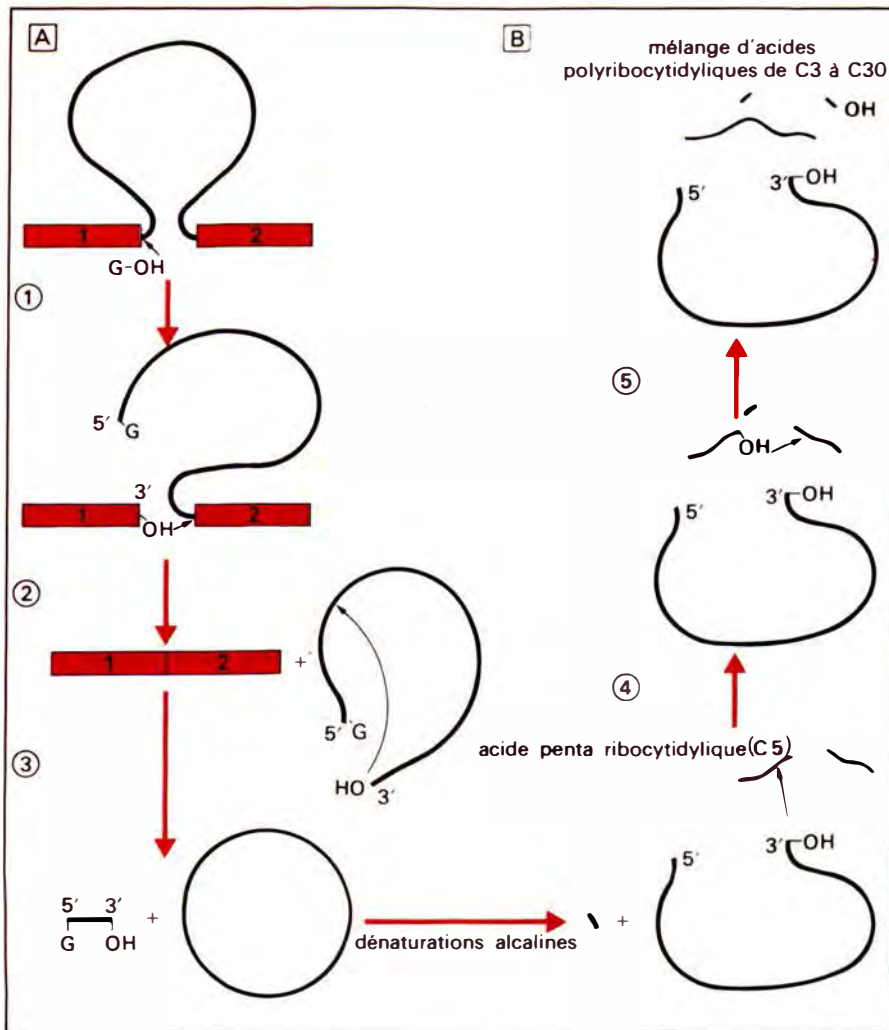


Figure 1. A : Auto-épissage de l'intron ribosomique de *Tetrahymena* et des introns mitochondriaux de type 1. G = Guanosine ou acide guanylique. Rectangles rouges : exons; lignes épaisses : introns. Les flèches partant des radicaux hydroxyles (OH) montrent les points d'attaque nucléophile des liaisons phosphodiester de molécules d'ARN.

Étape 1 : Attaque nucléophile de la liaison exon-intron 5' par l'hydroxyle 3' d'un résidu guanosine, fixation de la guanosine en 5' de l'intron.

Étape 2 : Attaque nucléophile de la liaison intron-exon 3' par l'hydroxyle 3' de l'exon 1. Épissage des exons.

Étape 3 : Attaque nucléophile de l'intron excisé par son hydroxyle 3'. Cyclisation de l'intron. Au terme des étapes 1 à 3, les 3 liaisons phosphodiester rompues (5' et 3' de l'intron, clivage de l'intron) ont été remplacées par 3 nouvelles liaisons (fixation de la guanosine, liaison entre les exons, et cyclisation de l'intron). Le bilan global est celui d'une « trans-estérification ».

B : Rôle catalytique de l'intron dégradé de *Tetrahymena*. Les lignes fines représentent des molécules d'acides polyribocytidyliques, c'est-à-dire de ribonucléotides à cytidine. Deux phases successives de dénaturations alcalines ménagées de l'intron cyclisé le dégradent partiellement et le transforment en une forme linéaire qui, mise en présence de molécules d'acide pentaribocytidylique (C<sub>5</sub>) :

Étape 4 : ...provoque une attaque nucléophile des molécules...

Étape 5 : ...et, par trans-estérifications en cascade, la synthèse de molécules hétérogènes d'acide polycytidylique.

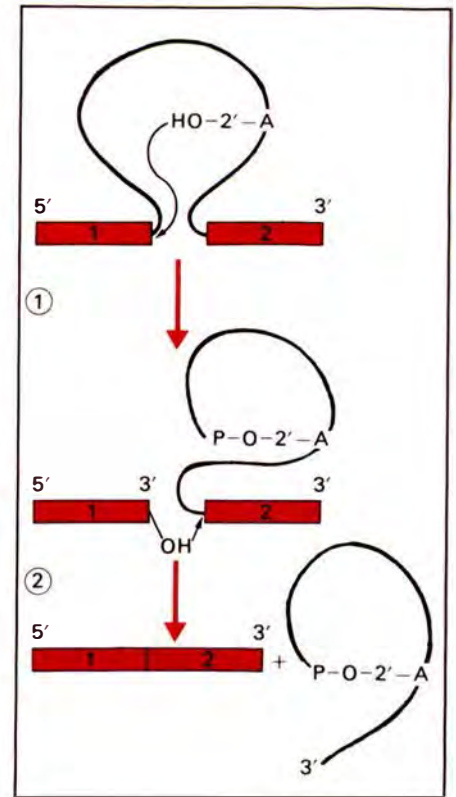


Figure 2. Modèle présumptif de l'auto-épissage des introns mitochondriaux de classe 2 et, peut-être, des introns nucléaires. Les symboles sont les mêmes que dans la Figure 1.

Étape 1 : L'hydroxyle 2' de l'acide adénylique du site de branchement est responsable d'une attaque nucléophile de la liaison phosphodiester exon-intron en 5' de l'intron; après clivage l'extrémité 5' phosphate libre de l'intron se lie par un pont phosphodiester 5' -2' avec l'acide adénylique sus-cité.

Étape 2 : Une attaque nucléophile par l'hydroxyle 3' du premier exon hydrolyse la liaison entre l'intron et le deuxième exon. Les deux exons sont alors épissés précisément l'un à l'autre. Il faut noter que cet ensemble de réaction équivaut là encore à une « trans-estérification », un nombre identique de liaisons phosphodiester étant créé et détruit.

7. Peebles C L, Perlman P S, Mecklenburg K L, et al. A self-splicing RNA excises an intron lariat. *Cell* 1986; 44 : 213-23.

8. Van Der Veen R, Arnberg A C, Van Der Horst G, et al. Excised group II introns in yeast mitochondria are lariats and can be formed by self splicing in vitro. *Cell* 1986; 44 : 225-34.

9. Solnick D. Alternative splicing caused by RNA secondary structure. *Cell* 1985; 43 : 667-76.

## Des ribozymes à la demande

La découverte des ribozymes, et les perspectives qu'elle révèle quant au « monde de l'ARN », est bien décidément un des faits scientifiques majeurs des dernières années (*m/s n° 2, vol. 1, p. 107, et suppl. au n° 7, vol. 7, p. 30*).

Une équipe australienne de Canberra vient maintenant d'élucider certaines règles fondamentales définissant les ARN à activité catalytique (endonucléasique) et leur cible ([1] et figure 1). Le point de départ de leurs travaux a consisté à isoler l'une de l'autre des régions d'un ARN autocatalytique de plante, l'une étant responsable de l'activité catalytique et l'autre fournissant le site « substrat » pour la coupure endoribonucléasique.

L'ARN d'un virus de tabac (mais aussi d'autres virus auxiliaires ou de viroïdes) se réplique pour donner des enchaînements (concatémères) de l'unité de transcription qui doivent être secondairement clivés pour libérer les unités indépendantes.

La réaction est autocatalytique, se fait à pH neutre ou légèrement alcalin en présence de cations divalents. Elle engendre des extrémités 5' hydroxyles et 2'-3' phosphate cycliques.

Il est possible, en combinant diverses expériences de mutagenèse dirigée et de sous-clonages de fragments d'ADN complémentaires de l'ARN viral, d'obtenir des oligonucléotides de respectivement 19 et 24 bases dont les copies ARN possèdent les activités catalytiques (ARN de 19 bases) et de substrat (ARN de 24 bases).

La comparaison entre la structure de ces oligonucléotides et celles de sites correspondants d'autres ARN qui subissent une réaction d'auto-clivage a permis aux auteurs de prédire les caractéristiques essentielles conférant à un ARN, d'une part la propriété catalytique de cliver un autre ARN,

d'autre part d'être substrat de cette réaction.

La figure 1 montre que le clivage survient en 3' d'un triplet GUC ou d'une séquence apparentée. Cela constitue la seule similitude de séquence entre différents sites de clivage.

L'ARN ribozymatique a une structure conservée comportant deux boucles, l'une étant fermée par une tige appariée. La continuité de la boucle ne serait d'ailleurs pas nécessaire à l'activité.

De plus, des séquences d'ARN de part et d'autre du site de clivage

sur le brin substrat et du motif ribozymatique sur le brin catalytique sont étroitement complémentaires.

Afin de tester ce modèle, Jim Hasehoff et Wayne L. Gerlach ont synthétisé des ribozymes conçus pour cliver un ARN procaryotique : le motif ribozymatique a donc été entouré de séquences complémentaires des bases situées de part et d'autre de triplets GUC sur l'ARN à cliver.

Les ARN transcrits *in vitro* à partir de constructions ADN codant pour ces caractéristiques se révèlent avoir en effet une action catalytique d'endoribonucléase spécifique de site : ils

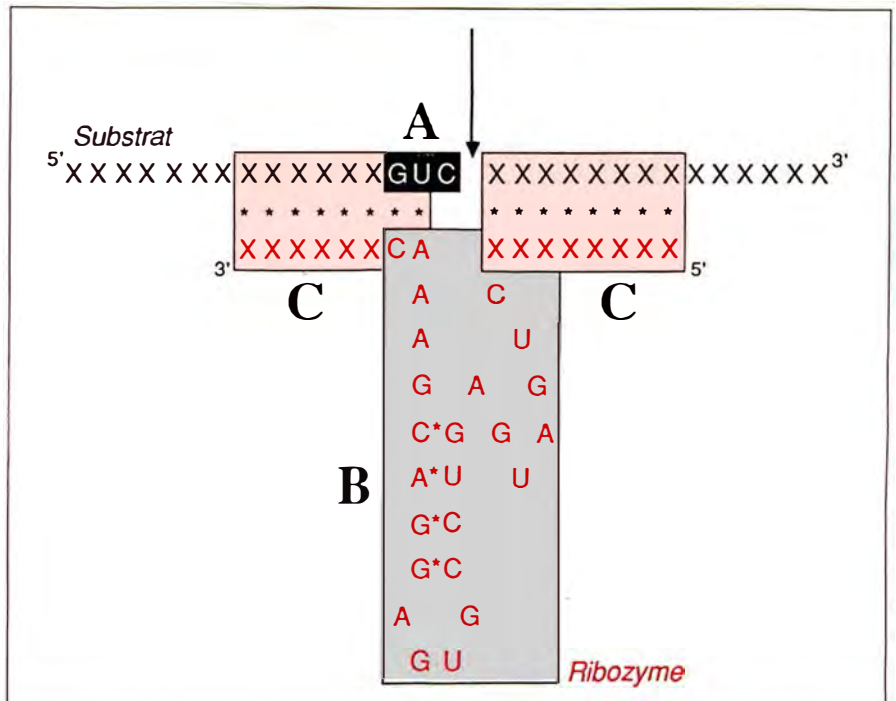


Figure 1. **Modèle d'un ribozyme et de son substrat.** L'ARN substrat est représenté en noir, avec en lettres blanches dans un rectangle noir (motif A) le triplet conservé GUC en 3' duquel se fait le clivage. L'ARN ribozymatique est représenté en rouge. Le rectangle gris (motif B) délimite le motif conservé du ribozyme, formant probablement son centre actif. Les rectangles roses (motif C) délimitent les zones appariées entre les ARN substrat et catalytique. Les astérisques indiquent les paires de bases appariées (d'après [1]).

digèrent en une heure à 50° C un excès molaire d'ARN substrat d'environ 10 fois.

Le mécanisme catalytique précis reste inconnu ; il pourrait s'agir de la déprotonation (déplacement d'un proton) d'un radical 2' hydroxyle exposé qui réagirait à son tour par attaque nucléophile du phosphate appartenant au pont phosphodiester le plus proche, aboutissant donc à un clivage engendrant une extrémité 5' OH libre du fragment adjacent et une extrémité cyclique 2'-3' phosphate du fragment portant le triplet GUC.

Le « centre catalytique » du ribozyme pourrait être constitué par des métaux divalents précisément complexés au site actif ribozymatique, ou bien par des groupes actifs de l'ARN lui-même.

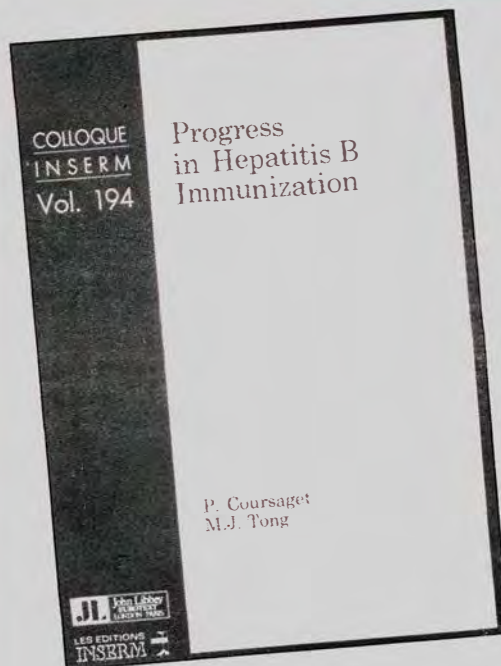
Quoiqu'il en soit, cette découverte doit être considérée comme tout à fait essentielle. Elle généralise les notions acquises ces dernières années sur la catalyse par l'ARN (« ribozymatique »). Elle ouvre la voie à la biotechnologie de l'ARN, les ribozymes ainsi synthétisés pouvant être considérés comme les premières enzymes de restriction de l'ARN (un « ribozyme de restriction », en quelque sorte). Elle suggère enfin une nouvelle voie pour réaliser des « phénocopies » d'altération génique que les ARN antisens permettent rarement d'obtenir chez la souris : un transgène transcrit en un « ribozyme de restriction » spécifique du messenger d'un gène dont on désire supprimer les effets pourrait peut-être, du fait du potentiel catalytique du ribozyme (dont une molécule a le potentiel d'en digérer un grand nombre), aboutir à la destruction totale du messenger cible. Puisque la science consiste aussi à rêver au possible, sinon au probable, l'on peut aussi imaginer que les ribozymes se révéleront demain de redoutables agents antiviraux, inhibant l'expression des gènes de virus à ADN ou de provirus, voire clivant les formes répliquatives des virus à ARN.

**A. K.**

Haseloff J, Gerlach WL. Simple RNA enzymes with new and highly specific endoribonuclease activities. *Nature* 1988 ; 334 : 585-91.

**John Libbey  
EUROTEXT**

**Paris-Londres  
ÉDITIONS MÉDECINE-SCIENCES**



**P. COURSAGET  
M.J. TONG**

Colloque INSERM  
Volume 194  
1990, broché  
600 pages  
520 FF

**Co-édition INSERM/  
John Libbey Eurotext**

*Un ouvrage, destiné à tous les professionnels de la Santé, où les plus grands spécialistes internationaux font le point sur les avancés les plus récentes concernant la vaccination contre l'hépatite B*

- Vaccins plasmatiques et recombinants
- Protection à long terme et doses de rappel
- Vaccination des enfants et des nouveau-nés
- Stratégies et programmes de vaccination.

#### BON DE COMMANDE

NOM \_\_\_\_\_ Prénom \_\_\_\_\_

Adresse \_\_\_\_\_

Désire recevoir l'ouvrage de P. Coursaget et M.J. Tong, au prix de 520 FF + 30 F de frais de port, soit 550 F.

Ci-joint mon règlement à l'ordre de **John Libbey Eurotext**

6, rue Blanche - 92120 MONTROUGE - France - Tél. : (1) 47.35.85.52

# Gènes, pseudogènes, familles multigéniques et évolution

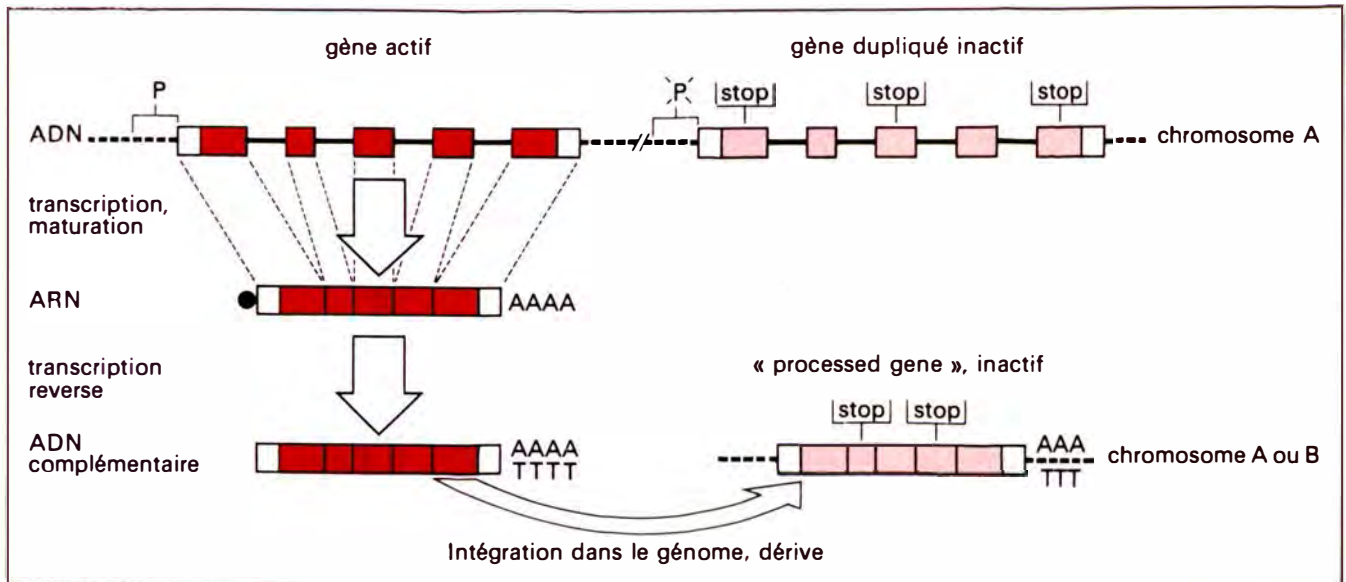


Figure 1. Formation de pseudogènes par duplication ou par transcription reverse d'ARN mature suivie de réintégration dans le génome.

Les pseudogènes n'étant pas soumis à une pression sélective s'opposant à leurs mutations, les changements de séquence s'y accumulent, créant notamment l'apparition de nombreux codons stop (*stop*)

dans les exons et altérant la structure du promoteur (P)

La couleur rose des exons appartenant aux pseudogènes indique qu'ils ne sont plus codants.

Pour une définition plus complète des symboles, se reporter à la page 2 de ce document.

## A. Duplications géniques et dérive génétique (figure 1)

Au cours de l'évolution surviennent des phénomènes de duplication de gènes préexistants, aboutissant à de nouvelles copies dont le destin est variable. Parfois la nouvelle copie reste active, et code pour le même message que le gène initial; des différences entre les deux gènes s'accumulent cependant dans les régions non codantes, introns et portions non codantes des exons. Ce modèle s'applique parfaitement aux deux gènes de la chaîne  $\alpha$  de l'hémoglobine. Dans d'autres cas les gènes dupliqués vont, tout en restant actifs, évoluer pour leur propre compte, aboutissant à la synthèse de protéines fonctionnellement différentes et parfois contrôlées différemment au cours de développement et de la différenciation. Ainsi en est-il probablement des isoenzymes, de l'albumine et de l' $\alpha$  foeto-protéine ou des chaînes  $\beta$ ,  $\gamma$  et  $\delta$  de l'hémoglobine. Souvent enfin, les nouvelles copies de gènes vont être rapidement modifiées dans leurs séquences codantes et régions de

contrôle, ce qui aboutit à des pseudogènes non fonctionnels (non transcrits) dans lesquels, en l'absence de la pression sélective qui tend à éliminer les mutations d'un gène risquant d'en altérer la fonction, les modifications de séquence vont s'accumuler au cours de l'évolution. Certains de ces pseudogènes inactifs pourraient représenter de véritables « laboratoires » où se prépare l'acquisition d'une fonction nouvelle, le pseudogène précédant ici le gène actif et pouvant jouer un rôle essentiel dans l'évolution des espèces.

## B. « Rétrogènes » (*processed genes* figure 1)

Certains pseudogènes semblent être des copies ADN d'ARN messagers ayant subi une maturation normale. De telles copies se réintègreraient au hasard dans le génome, donnant des pseudogènes caractérisés par l'absence de promoteur, d'introns et la présence d'une extension d'acide polyadénylique.

La signification et le mécanisme de formation de ces séquences restent parfaitement énigmatiques.

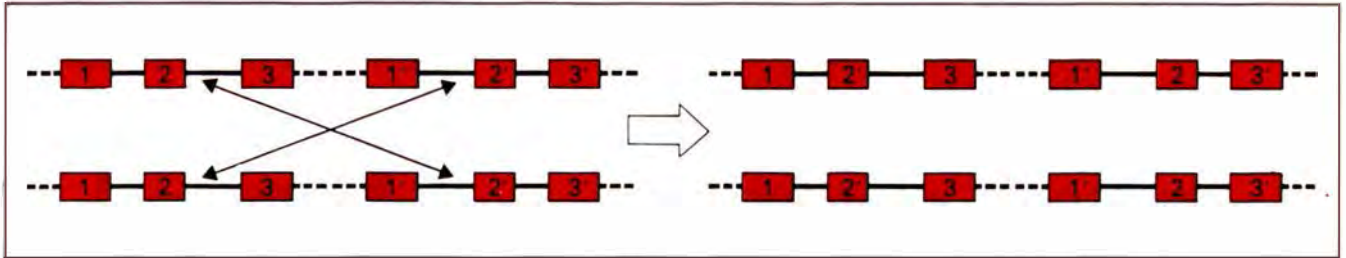
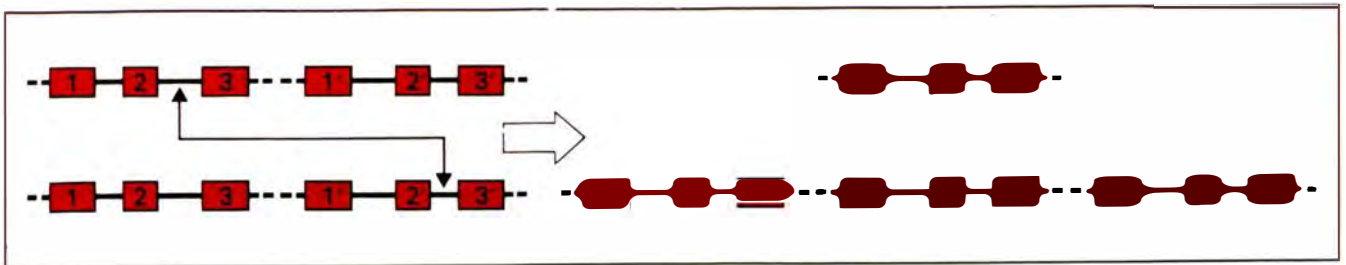


Figure 2. *Conversion génique : échange d'information entre deux gènes non alléliques sans perte d'information ni modification de la structure globale des gènes. Les mécanismes possibles en restent hypothétiques.*



### C. Familles multigéniques : conversions géniques et « crossing over » non équationnels

Ces familles sont composées d'un nombre, pouvant varier de quelques unités à quelques centaines, de gènes et pseudogènes partiellement homologues, localisés en une même région chromosomique ou dispersés sur tous les chromosomes. Dans le cas d'une famille multigénique « géographiquement » localisée, des appariements entre des séquences partiellement homologues non alléliques (c'est-à-dire ne correspondant pas aux deux gènes parentaux d'un même locus) aboutissent à des échanges d'information génétique, responsables d'une évolution

concertée d'un groupe de gènes qui, en l'absence de ces échanges, divergeraient rapidement (figure 2) (cas par exemple des gènes du complexe majeur d'histocompatibilité, ou des régions variables des immunoglobulines). Ce phénomène est dénommé « conversion génique »; son mécanisme reste controversé.

Par ailleurs ces répétitions de gènes partiellement homologues en une même zone du chromosome favorisent les échanges non équationnels entre chromatides lors de la méiose, ou « crossing over », à l'origine de délétions géniques particulièrement bien illustrées dans de nombreux types de thalassémie (figure 3).

A.K.

Figure 3. *Crossing over non équationnel entre deux gènes non alléliques partiellement ou totalement homologues. Le résultat net en est un réarrangement des gènes, avec perte d'information au niveau d'une des chromatides et gain au niveau de l'autre. Le mécanisme probable en est un appariement entre les séquences homologues non alléliques de deux chromatides au moment de la méiose, avec cassure-réparation des chromatides ainsi appariées.*

# Génétique

## et particules d'information : redistribution des exons au cours de l'évolution

La physique des particules étudie les unités élémentaires dont est composée la matière : elles sont identiques d'un atome et d'une molécule à l'autre, seul leur agencement étant différent. Plus simplement, les enfants créent des formes diverses selon la manière dont ils associent les pièces de leur *Mécano* ou de leur *Légo*. C'est un procédé semblable que semble utiliser « l'évolution » pour créer des gènes nouveaux en associant variablement des « particules d'information » préexistantes. Ces particules sont des exons dont on trouve des homologues très conservés dans plusieurs gènes, dans un environnement variable. Un bon exemple de ce phénomène est la structure du gène codant pour l'activateur tissulaire du plasminogène [1] (voir schéma). Ses 5 exons 3' sont homologues de ceux des protéases à sérine, telle la trypsine. Les 4 exons précédents codent pour la structure protéique dite « Kringle » retrouvée dans d'autres facteurs de la coagulation tels la thrombine, la plasmine et l'urokinase. L'exon 4 est homologue d'une séquence du gène codant pour le facteur de croissance EGF. L'exon 3 code pour un peptide comportant des ponts disulfures qui maintiennent une conformation « Finger » (en doigt) très caractéristique de la fibronectine et retrouvée dans le facteur IX (antihémophilique B) et la thrombine.

La constitution du gène de l'activateur tissulaire du plasminogène semble donc être le résultat d'une « redistribution » d'exons préexistant dans d'autres gènes, conférant au domaine « protéase à sérine » de nouvelles spécificités. Le même phénomène est observé pour d'autres facteurs de la coagulation, des protéases, ainsi que pour des récepteurs membranaires tel celui des lipoprotéines de faible densité [2].

Le mécanisme d'une telle redistribution des exons au cours de l'évolution reste obscur. Il pourrait impliquer des phénomènes de duplication et réarrangement de gènes (voir *Lexique en pages 6 et 7 de ce document*). Un autre mécanisme dont la possibilité a été rapportée très récemment [3, 4] est l'association entre eux d'exons issus de gènes différents au cours de l'épissage des transcrits nucléaires (voir *Lexique en pages 4 et 5 de ce document*). De nouvelles espèces d'ARN pourraient être ainsi créées, copiées en ADN par transcription-reverse et réinsérées dans le génome, comme le sont les « processed genes » ou rétrogènes (voir *Lexique pages 6 et 7*).

Resterait alors à comprendre le phénomène de la réapparition d'introns avant que ces nouveaux gènes ne soient réexprimés sous la forme d'une nouvelle protéine.

A. K.

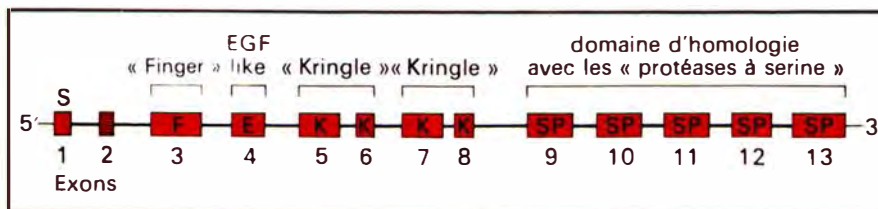
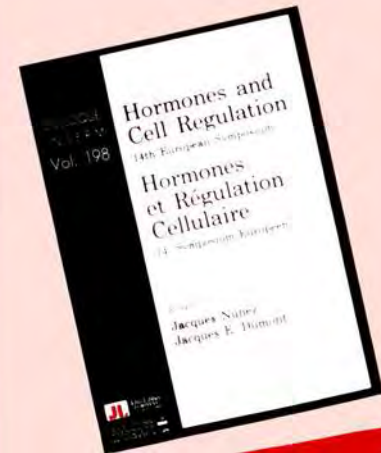


Figure 1. Structure du gène codant pour l'activateur tissulaire du plasminogène. S = peptide signal, partie hydrophobe N-terminale des protéines sécrétées qui permet leur passage à travers la membrane. La signification des homologies « Finger », EGF-like et Kringle » notée au-dessus de certains exons est donnée dans le texte.



John Libbey  
EUROTEXT  
LONDON PARIS

Paris - Londres  
EDITIONS MEDECINE-SCIENCES



COLLOQUES INSERM

J. NUNEZ  
J.E. DUMONT

### Hormones et Régulation Cellulaire n° 14

- Récepteurs de surface et cyclases
- Régulation de l'expression génique
- Facteurs de croissance
- Canaux ioniques

Co-édition John Libbey  
Eurotext/INSERM, vol. 198

1989, broché, 144 pages  
240 FF

#### BON DE COMMANDE

NOM ..... Prénom .....  
Adresse .....

Désire recevoir **Hormones (n° 14)** au prix de 240 FF + 30 FF de frais de port, soit 270 FF. Ci-joint mon règlement à l'ordre de **John Libbey Eurotext** 6, rue Blanche, 92120 Montrouge  
Tél. : (1) 47.35.85.52.

---

# *Le code génétique :*

## *variations sur un thème universel*

**D**e récents travaux, qui ont bénéficié d'un grand retentissement, ont sérieusement ébranlé le dogme de l'universalité du code génétique édicté voici vingt-cinq ans. Pour évaluer les limites comme les implications de ces découvertes, il est nécessaire de rappeler les données de base.

### ● Le code génétique

Il fut élucidé en 1960, principalement par M. Nirenberg, à l'aide d'oligonucléotides de synthèse fonctionnant dans un système de synthèse protéique provenant du colibacille. L'ARN messager est découpé en triplets, suites de trois nucléotides; les quatre bases, combinées trois à trois, peuvent former  $4^3$ , soit 64 combinaisons différentes. Le code ainsi défini présente plusieurs caractères :

(a) le code est universel « du colibacille à l'éléphant ». Il est le témoin de l'unicité du vivant;

(b) le code est « dégénéré ». Sur les 64 triplets possibles, 61 codent pour un acide aminé, et un seul. Mais, puisque il n'y a que 20 acides aminés, la plupart d'entre eux (en fait, tous sauf méthionine et tryptophane) sont sous la dépendance de plusieurs codons, jusqu'à six pour leucine, sérine et arginine (*figure 1*);

(c) trois triplets TAA, TAG, TGA — ou, en dialecte ARN, UAA, UAG, UGA — ne codent pour

aucun acide aminé et leur présence sur le messenger signifie « stop » ou « fin de synthèse ». Un seul codon AUG (méthionine) permet l'initiation de la synthèse protéique;

(d) il n'y a pas de correspondance chimique entre codon et acide aminé. Il faut entre eux un adaptateur, l'ARN de transfert. Ces ARN, qui comptent de 73 à 93 nucléotides, ont une structure secondaire en feuille de trèfle. Chacun d'eux est doté d'une double spécificité : à l'une de ses extrémités il peut se charger en un acide aminé déterminé; en une autre partie de sa molécule il possède un triplet appelé anticodon, complémentaire du codon correspondant du messenger. Le nombre des anticodons est inférieur à celui des codons. Seules les deux premières bases de chaque codon ont une signification rigoureuse; la troisième est, partiellement ou totalement selon les cas, indifférente. Pour en rendre compte Crick a formulé en 1966 sa théorie des oscillations (*wobble hypothesis*). La base de l'anticodon qui se trouve en face de la troisième base du codon est capable d'osciller et d'accepter ainsi des appariements qui sortent des règles classiques. Par exemple, U peut s'accorder avec G en plus de A, et G avec U en plus de C. On trouve en outre, dans certains anticodons, des bases modifiées : c'est ainsi que la base I, hypoxanthine, résultant de la désamination de l'adénine, peut reconnaître U et A. Un même

ARNt peut ainsi se fixer sur 2 ou même 3 codons ayant une troisième base différente. Le calcul a montré que le nombre minimum d'ARNt différents est de 32. Par contre, une mutation qui affecterait une base de l'anticodon correspondant à une des deux premières bases du codon provoquerait un changement du code.

### ● Mutations non-sens et supprimeurs

Des mutations, reconnues d'abord chez les bactéries, mais identifiées récemment en pathologie humaine (certaines thalassémies) font apparaître un triplet « stop » au milieu d'un messenger, bloquant sa traductibilité. Ces mutants « non-sens » ont été appelés ocre pour UAA, ambre pour UAG et opale pour UGA. On a découvert chez les bactéries et la levure des ARNt modifiés capables de « lire » ces triplets. Ce sont les ARNt supprimeurs; la protéine peut à nouveau être synthétisée, mais l'acide aminé inséré grâce au supprimeur n'est pas nécessairement celui qui était présent dans la protéine initiale. Très récemment des ARNt supprimeurs ont été découverts chez des mammifères, dont l'homme.

On sait également, depuis peu, en fabriquer à volonté au laboratoire par mutagenèse dirigée. Un supprimeur rend inefficace le triplet de terminaison qu'il « supprime » et donc change le code génétique.

## ● L'ADN des mitochondries

C'est le premier contrevenant au code génétique. Les mitochondries sont génétiquement des hybrides. La plupart de leurs protéines sont importées par elles après synthèse d'après des gènes du noyau. Elles ont cependant gardé un ADN propre, qui se charge de la biosynthèse de quelques protéines. L'ADN mitochondrial a une taille qui peut varier de plus de 100 000 à moins de 20 000 paires de bases. Le génome humain, un des plus succincts, se limite à 16 569 paires de bases, et cette concision a soulevé l'enthousiasme des auteurs (small is beautiful!). S'y trouvent comprimés les gènes de 2 ARN de ribosomes, 22 ARN de transfert, et 13 protéines dont certaines sont encore en cours d'identification. Aucun intron n'y a de place, alors que la levure, à l'aise dans ses 80 000 paires de bases, en possède [1, 2].

Le système de synthèse n'utilise que les ARNt propres aux mitochondries; or il n'en existe que 22 à 24 au lieu des 32 nécessaires au code nucléaire. La constitution et la conformation des ARNt sont quelque peu différentes de celles des ARNt habituels. Un seul ARNt est capable de lire les 4 codons qui ne diffèrent que par la troisième base. De plus, la règle de Crick (oscillations) s'applique à l'anticodon du tryptophane, qui peut par conséquent reconnaître aussi bien UGA (codon stop officiel) que UGG, et à celui de la méthionine qui reconnaît dans certains organismes AUA comme AUG (voir figure 2). Ces particularités expliquent les entorses, découvertes dès 1979, que se permettent les mitochondries vis-à-vis du code. Il n'y a pas même constance du code entre mitochondries d'espèces différentes : UGA signifie tryptophane et non « fin de chaîne » chez l'homme et la levure, mais pas chez le maïs. Le codon AUA vaut méthionine et sert, comme AUG, d'initiateur chez l'homme, mais pas dans la levure. L'évolution des mitochondries n'est donc pas régie par des règles aussi strictes que celles qui règnent dans le noyau, où de telles variations auraient sans doute des conséquences catastrophiques.

uniques		doublets		quartet	sextet
A	U	C C	C C	A A A A	C C C C   A A
U	G	A A	A A	C C C C	G G G G   G G
G	G	A G	C U	A G C U	A G C U   A G
Met	Trp	Gln	His	Thr	Arg

Figure 1. Le code génétique et sa dégénérescence. Les triplets se lisent de haut en bas.

La figure montre des exemples (non exhaustifs) de plusieurs possibilités.

3<sup>e</sup> base totalement significative : codons uniques.

3<sup>e</sup> base organisée en doublets : AG (purines) d'un côté, CU (pyrimidines) de l'autre.

3<sup>e</sup> base indifférente : 4 codons pour un même acide aminé : quartets.

Sextet : un quartet plus un doublet pour un même acide aminé. Dans ce cas, quartet et doublet diffèrent par une de leurs deux premières bases, et utilisent donc au moins deux ARN de transfert différents.

L'ensemble du code pour les 20 acides aminés, constituants primaires des protéines, comprend 3 sextets, 5 quartets, 9 doublets, 1 trio (isoleucine) et 2 codons uniques. Il rend compte des 61 codons. Trois triplets « stop » signifient « terminaison de chaîne » et ne codent normalement pour aucun acide aminé.

sens de lecture	code universel	mitochondrie	m. capricolum	ciliés
←	tryptophane			
←	anticodons	A C C	A C U	A C C
→	codons	U G G	U G A W	U G G
←	méthionine			
←	anticodons	U A C	U A U	A C U
→	codons	A U G	A U A W	U G A
→	glutamine			
→	codons	C A A	C A G	C A A
		C A G	C A G	C A G
			U A A	U A A
			U A G	U A G

Figure 2. Essai d'interprétation des exceptions au code génétique.

Deux mécanismes sont proposés

(a) duplication d'un gène d'ARN de transfert suivi d'une mutation sur l'anticodon du gène dupliqué. Ce mécanisme pourrait s'appliquer quand le code nucléaire est changé (Ciliés, *M. capricolum*); tout récemment en effet, un gène de l'ARN de transfert de glutamine correspondant à un codon UAA a été trouvé chez un cilié [9].

(b) dans les mitochondries, altération de conformation d'un ARNt avec ou sans mutation d'une base de l'anticodon; elle permet l'oscillation (wobble) de la première base de l'anticodon sur certains ARNt qui n'ont pas cette propriété lorsqu'ils sont codés par des gènes du noyau, et qui peuvent acquérir les propriétés d'un ARNt suppresseur.

Sont indiquées en rouge souligné les particularités communes aux mitochondries de l'homme et de la levure; en rouge encadré les particularités propres au génome humain

W = wobble (oscillation de la troisième base du codon).

+ = duplication.



## ● Les Protozoaires Ciliés

C'est de leur étude, développée depuis un an, qu'est venue la mise en cause de l'universalité du code. Ces organismes unicellulaires, dont la paramécie est le type, ont un cycle de vie original. Ils possèdent deux noyaux, le macronucleus et le micronucleus. Le micronucleus n'intervient qu'au cours de la phase sexuelle du cycle : il s'organise en chromosomes, se divise par méiose avant de subir une conjugaison. Dans la cellule nouvelle, une copie du micronucleus va générer un macronucleus, à la suite de remaniements très amples de l'ADN. Dans la phase végétative, seul fonctionne le macronucleus; il ne forme pas de chromosomes et se divise sans mitose. Toutes les recherches sur les séquences d'ADN des gènes et les protéines dont ils gouvernent la synthèse portent sur les gènes du macronucleus [3-7].

De nombreux travaux, dont ceux de l'équipe française de Caron et Meyer, ont abouti à une même conclusion. Dans quatre espèces de ciliés, dont deux variétés de paramécies et sur des gènes aussi variés que ceux d'histones, de tubulines, ou d'antigènes de surface, on ne trouve qu'un seul triplet « stop », UGA. Les deux autres, UAA et moins souvent UAG, sont devenus signifiants et paraissent coder pour l'acide aminé glutamine, qui garde cependant ses codons réguliers CAA et CAG. On conçoit que le changement de sens de la première base d'un codon, et son extension apparente à toutes les protéines de l'organisme, prennent une signification beaucoup plus ample que les observations antérieures faites sur les mitochondries.

Mais voici [8] que des divergences avec le code classique sont signalées chez un procaryote, *Mycoplasma capricolum*. Un triplet UGA code pour le tryptophane, exactement comme dans les mitochondries, bien que le mécanisme soit différent (figure 2). Il a été démontré, et c'est d'importance capitale, que *M. capricolum* possède deux gènes différents d'ARNt de tryptophane dont les anticodons sont distincts. Il s'est donc probablement produit une duplication d'un gène d'ARNt.

Quand et comment les divergences par rapport au code standard sont-elles apparues? La séparation des ciliés d'avec les autres eucaryotes a-t-elle été très précoce; a-t-elle précédé la fixation définitive du code, ou l'a-t-elle suivie? On ne possède pas actuellement de réponse à cette question.

Quant au mécanisme, on en trouve probablement la clef au niveau des ARN de transfert; deux modalités au moins peuvent être invoquées (explicitées dans la figure 2).

Finalement il ne nous semble pas que l'universalité du code soit réellement en cause. On conçoit qu'un grain de sable dans les rouages de l'évolution puisse altérer le code pour un acide aminé dans une espèce ou un groupe d'espèces, sans remettre en question les principes généraux.

J.-C.D.

1. Borst P, Grivell LA. Small is beautiful - portrait of a mitochondrial genome. *Nature* 1981; 290 : 443-4.

2. Heckman JE, Sarnoff J, Alzner-Deweerd, et al. Novel features in the genetic code and codon reading patterns in *Neurospora crassa* mitochondria based on sequences of six mitochondrial tRNAs. *Proc Natl Acad Sci USA* 1980; 77 : 3159-63.

3. Fox TD. Diverged genetic codes in protozoans and a bacterium. *Nature* 1985; 314 : 132-3.

4. Caron F, Meyer E. Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature* 1985; 314 : 185-8.

5. Preer JR, Preer LB, Rudman BM, Barnett AJ. Deviation from the universal code shown by the gene for surface protein in *Paramecium*. *Nature* 1985; 314 : 188-90.

6. Helftenbein E. Nucleotide sequence of a macronuclear DNA molecule coding for alpha tubulin from the Ciliate *Stylonychia lemnae*. Special codon usage: TAA is not a translation termination codon. *Nucl Ac Res* 1985; 13 : 415-33.

7. Horowitz S, Gorovsky MA. An unusual genetic code in nuclear genes of *Tetrahymena*. *Proc Natl Acad Sci USA* 1985; 82 : 2452-55.

8. Yamao F, Muto A, Kawauchi Y, et al. UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc Natl Acad Sci USA* 1985; 82 : 2006-9.

9. Kuchino Y, Hanyu N, Tashiro F, Nishimura S. *Tetrahymena thermophila* glutamine tRNA and its gene that corresponds to UAA termination codon. *Proc Natl Acad Sci USA* 1985; 82 : 4758-62.

John Libbey  
EUROTEXT  
LONDON PARIS

Paris - Londres  
EDITIONS MEDECINE-SCIENCES



COLLECTION

G. BENZI

### Advances in Myochemistry (2)

- Processus de peroxydation et fonction musculaire
- Protéines contractiles et muscles squelettiques
- Inter-relations système nerveux - muscles
- Dystrophies musculaires, myopathie de Duchenne

1989, broché, 392 pages

375 FF

#### BON DE COMMANDE

NOM ..... Prénom .....

Adresse .....

Désire recevoir **Advances in Myochemistry (2)** au prix de 375 FF + 30 FF de frais de port, soit 405 FF.

Ci-joint mon règlement à l'ordre de **John Libbey Eurotext**  
6, rue Blanche, 92120 Montrouge  
Tél. : (1) 47.35.85.52.

## Le code génétique et la spécificité des ARN de transfert

L'ARN de transfert (ARNt) est un adaptateur universel entre l'ARN messager et les acides aminés dont la succession constituera la protéine. L'importance de sa structure secondaire et tertiaire est attestée par leur très grande conservation au cours de l'évolution (figure 1). La figure 1 montre la structure secondaire classique en trèfle d'un ARNt avec, aux extrémités opposées, l'anticodon complémentaire du codon de l'ARNm et la branche-accepteur sur laquelle se fixe l'acide aminé. Dans l'espace, cette molécule est repliée en L, l'anticodon et le site accepteur étant situés aux extrémités du L et les boucles T et D (figure 1) au niveau de la pliure. Cette molécule d'ARNt possède de nombreuses bases appariées, maintenant la stabilité des tiges double-brin. De nombreux nucléotides sont chimiquement modifiés. Un ARNt possède une double spécificité : pour le codon de l'ARNm d'une part, pour l'acide aminé fixé au site accepteur d'autre part. Les mécanismes de la complémentarité entre le codon et l'anticodon est, dans l'ensemble, bien compris, puisqu'il repose sur la règle générale d'appariement des bases, avec cette particularité déjà longuement discutée dans *médecine/sciences (suppl. au n° 7, vol. 2, p. 8)* que la première position de l'anticodon (correspondant au troisième nucléotide du codon) obéit à des règles plus flexibles que celles gouvernant l'appariement de deux brins d'ADN : un U peut reconnaître un A ou un G, un G peut reconnaître un C ou un U. De plus, certaines modifications des bases de l'anticodon modifient leur spécificité. C'est ainsi que l'inosine

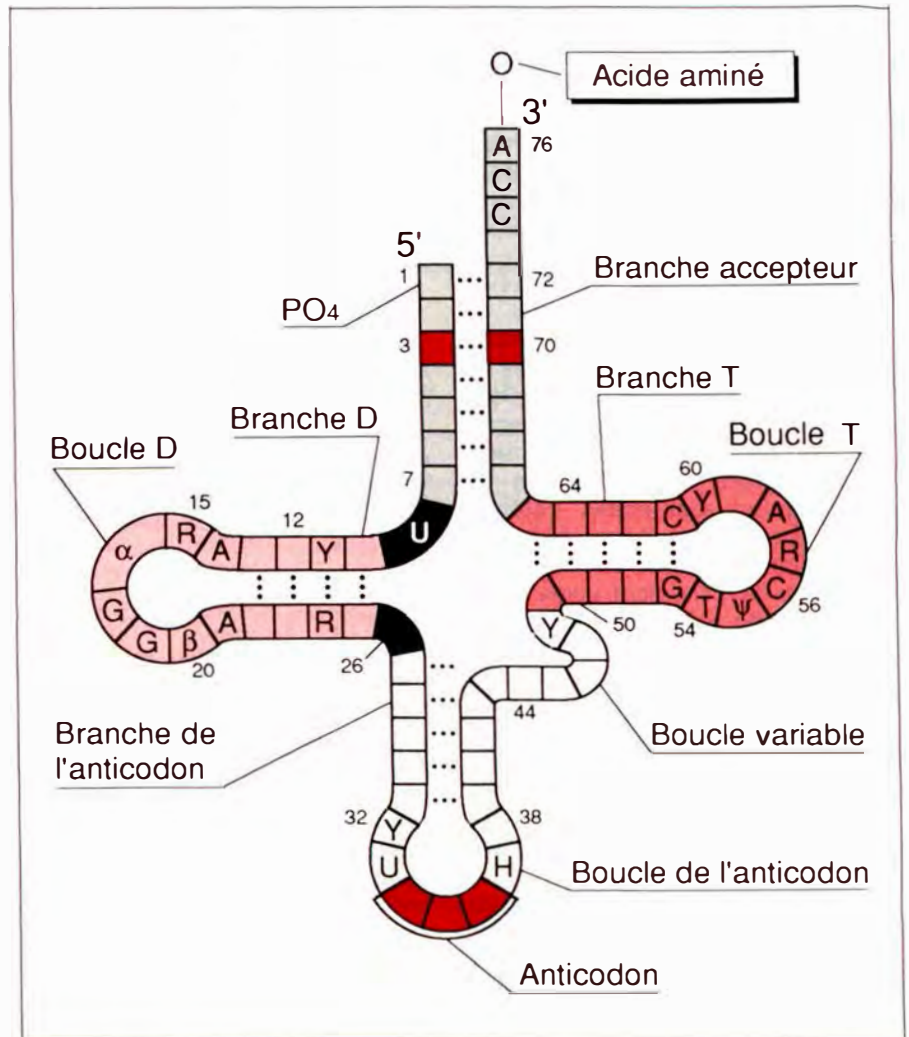


Figure 1. **Structure secondaire en feuille de trèfle d'une molécule d'ARN de transfert.** A = adénosine ; G = guanine ; C = cytosine ; U = uracile ; R = base purique, adénine ou guanine ; Y = base pyrimidique, cytosine ou uracile ; T = ribothymidine ; U = pseudo-uridine ; H = adénine ou guanine modifiées. Les positions invariantes ont été indiquées, de même que les appariements de paires de base (barreaux).

en première position (issue d'une modification de l'adénine) peut s'apparier à U, C et A. Chez *E. coli*, il existe deux ARNt<sup>Ile</sup> spécifiques de l'isoleucine ; dans l'espèce majoritaire (ARNt<sup>Ile-1</sup>) l'anticodon  ${}_{3'}\overline{\text{UAG}}_{5'}$  reconnaît les codons  ${}_{5'}\overline{\text{AUC}}_{3'}$  et  $\overline{\text{AUU}}$ . Dans l'espèce minoritaire (ARNt<sup>Ile-2</sup>), l'anticodon  $\overline{\text{UAL}}$  reconnaît le codon  $\overline{\text{AUA}}$ . L est la lysidine, dérivée de la cytosine par fixation d'une lysine (amino-acylation). Le gène de cet ARNt<sup>Ile-2</sup> minoritaire code donc pour un anticodon  $\overline{\text{UAC}}$  qui est transformé en  $\overline{\text{UAL}}$  par amino-acylation.  $\overline{\text{UAC}}$  non modifié devrait reconnaître le codon méthionine  $\overline{\text{AUG}}$ ..., la transformation de C en L étant responsable de la reconnaissance en troisième position du codon d'un A au lieu d'un G. Pour prouver cela, une équipe japonaise de Tokyo a modifié une molécule d'isoleucyl-ARNt minoritaire pour transformer l'anticodon  $\overline{\text{UAL}}$  en  $\overline{\text{UAC}}$  (réalisant donc l'inverse de la modification survenant spontanément dans la cellule). Le résultat attendu était que l'ARNt ainsi modifié reconnaisse le codon méthionine tout en restant chargé à son site accepteur par une isoleucine dans la séquence protéique. En réalité, le résultat de la modification fut double : d'une part, la spécificité de l'anticodon fut bien changée comme il était attendu (d'isoleucine en méthionine). D'autre part, l'ARNt modifié devint un mauvais substrat pour l'isoleucyl ARNt synthétase... et un très bon substrat pour la méthionyl ARNt synthétase. Les ARNt synthétases sont les enzymes qui chargent les acides aminés sur le site accepteur des ARNt et sont donc à la base du deuxième type de spécificité de l'ARNt, celle qui aboutit à ce que des ARNt<sup>Met</sup> ou ARNt<sup>Ile</sup> ne soient chargés que par, respectivement, la méthionine et l'isoleucine. Le résultat de l'équipe japonaise indique par conséquent que le changement de la spécificité de l'anticodon (Ile → Met) change également la spécificité pour l'amino-acyl ARNt synthétase. Une conclusion du même ordre a été tirée par une équipe américaine (L. Schulman et H. Pelka, cités dans [2]) qui a démontré que la modification

de l'anticodon  $\overline{\text{CAU}}$  de l'ARNt<sup>Val</sup> (reconnaissant le codon  $\overline{\text{GUA}}$  de la valine) en  $\overline{\text{UAC}}$  (anticodon méthionine) amenait également l'ARNt<sup>Val</sup> à être chargé par la méthionine au lieu de la valine. Inversement, la mutation de l'anticodon méthionine en anticodon valine amène l'ARNt<sup>Met</sup> à être chargé par la valine. Ainsi, au moins pour les méthionyl, isoleucyl et valyl ARNt synthétases, apparaît-il que la première base de l'anticodon est un déterminant essentiel de spécificité ou, en d'autres termes, qu'un événement unique change de manière coordonnée la spécificité de l'anticodon et du site accepteur de l'acide aminé. Ce phénomène est évidemment très avantageux pour la cellule bactérienne puisqu'une modification de l'anticodon (ou, dans le cas de l'ARNt<sup>Ile-2</sup>, une transformation insuffisante de  $\overline{\text{UAC}}$  en  $\overline{\text{UAL}}$ ) va en même temps changer la reconnaissance du codon et le chargement par l'acide aminé ; de ce fait, l'ARNt<sup>Ile-2</sup> incomplètement transformé va, par exemple, s'apparier avec un codon méthionine... ce qui n'est pas grave puisqu'il sera alors chargé par une méthionine ! Les bases moléculaires de cet effet de l'anticodon sur la reconnaissance par l'amino-acyl ARNt synthétase sont inconnues ; elles impliquent probablement plus une reconnaissance de la structure globale de l'ARNt que la séquence particulière de l'anticodon. Les mécanismes de reconnaissance par les amino-acyl ARNt synthétases ne sont, de plus, probablement pas univoques. En mai 1988, en effet, W.H. McClain et K. Foss, de Madison dans le Wisconsin (USA) rapportaient qu'une modification de bases à la troisième position de la tige double-brin du bras accepteur de l'ARNt<sup>Phe</sup> (spécifique de la phénylalanine) (*figure 1*) modifiait la reconnaissance par la synthétase, sans toucher naturellement à l'appariement avec le codon. Dans l'ARNt<sup>Ala</sup>, cette position (bases 3 et 70) est occupée par la paire variante G-U (les bases G-C échangent entre elles trois liaisons, hydrogènes, G et U, une seule), alors qu'elle est CG dans l'ARNt<sup>Phe</sup>. Le changement de CG en GU dans ce dernier ARNt<sup>Phe</sup> l'amène à être

maintenant chargé par l'alanine de préférence à la phénylalanine. Il faut remarquer qu'un tel remplacement d'une paire à trois liaisons hydrogènes (GC) en une paire à une seule liaison peut en effet avoir d'importantes conséquences conformationnelles qui pourraient être le véritable signal de cette modification de spécificité. Ainsi est-ce lentement — mais, maintenant, sûrement — que, près de 30 ans après l'élucidation du code génétique que l'on pourrait appeler de stockage de l'information (les codons...), on en arrive à comprendre quelques-unes des règles et des lois du «second code génétique», celui par lequel chaque ARNt est reconnu par une amino-acyl ARNt synthétase particulière qui le charge avec l'acide aminé correct à l'exclusion de tout autre.

**Axel Kahn**

## RÉFÉRENCES

1. Muramatsu T, Nishikawa K, Nemoto F, *et al.* Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* 1988 ; 336 : 179-81.
2. RajBhandary UL. Genetic code : modified bases and amino-acylation. *Nature* 1988 ; 336 : 112-3.
3. McClain WH, Foss K. Changing the identity of a tRNA by including a G-U wobble pair near the 3' acceptor end. *Science* 1988 ; 240 : 793-6.

# Code génétique et séléno-protéines

Le sélénium est un oligo-élément jouant un rôle dans l'activité d'une classe d'enzymes, les séléno-enzymes, qui comprend notamment la glutathion peroxydase des eucaryotes et la formate déshydrogénase des procaryotes. Chez *E. Coli*, aussi bien que chez les mammifères, il a été montré que l'incorporation de sélénium dans les sélénopeptides, sous la forme de sélénocystéine, était un processus co-translationnel et était commandée par un codon stop UGA dans la phase ouverte de lecture [1-3]. Afin d'élucider le mécanisme de cette utilisation d'un codon stop pour incorporer la sélénocystéine dans les protéines, une équipe franco-allemande a analysé différentes souches de *E. Coli* (obtenues par traitement avec un mutagène) déficientes en l'activité de la formate déshydrogénase [4]. Quatre types de mutants (c'est-à-dire quatre groupes de omplémentation) ont été obtenus. L'un correspond à une altération du gène d'un ARN de transfert de la sérine, minoritaire dans la cellule. La sérine est chargée sur cet ARN<sup>t</sup> sous l'action de l'enzyme amino-acyl ARN<sup>t</sup>-synthétase habituelle. Du fait des caractéristiques structurales particulières de l'ARN<sup>t</sup>, le séryl-amino-acyl ARN<sup>t</sup> obtenu est alors le substrat spécifique d'enzymes de modification encore mal connues, transformant la sérine en sélénocystéine. L'anticodon de l'ARN<sup>t</sup> reconnaît alors le codon UGA (se comportant ainsi comme un suppresseur de codon stop UGA, aussi dénommé codon opale), incorporant la sélénocystéine dans la protéine au niveau de ce qui, ailleurs, aura la signification d'une fin de traduction. Il existe, chez les eucaryotes, un ARN<sup>t</sup> suppresseur de codon opale dont la fonction est inconnue ; il est chargé par un résidu sérine phosphorylé. Cet ARN<sup>t</sup><sup>OP-Ser</sup> suppresseur pourrait permettre l'incorporation directe de phosphosérine dans des protéines ; il pourrait aussi être un intermédiaire réactionnel entre l'ARN<sup>t</sup><sup>OP-Ser</sup> et l'ARN<sup>t</sup><sup>OP-Cys</sup> [4, 5].

La nature de l'information déterminant qu'un codon UGA sera utilisé pour incorporer de la sélénocystéine et non pour terminer une protéine reste absolument mystérieuse. Le contexte de séquence dans lequel s'insère cet UGA pourrait être en cause. Ces résultats soulèvent également des questions sur la signification originelle du codon opale UGA : était-il un codon stop ayant acquis, pour les besoins de l'évolution, une signification nouvelle ? Était-il, au contraire, le codon sélénocystéine des organismes primitifs vivant en anaérobiose ? La grande réactivité du sélénium à l'oxygène aurait, alors que se développait sur la terre la vie aérobie, éliminé les organismes comportant de nombreuses séléno-enzymes et rendu inutile, voire potentiellement dangereuse, l'utilisation générale d'UGA pour fabriquer des séléno-protéines. Le rôle de terminateur de traduction du codon opale serait alors un phénomène secondaire, son rôle maintenant exceptionnel de codon sélénocystéine étant un « fossile » de sa fonction initiale. Peut-être l'étude approfondie du code génétique de certaines archéobactéries anaérobies permettra-t-elle de répondre à ces questions.

A.K

1. Zinoni F, Birkmann A, Stadman TC, Böck A. Nucleotide sequence and expression of the selenocysteine containing polypeptide of formate dehydrogenase from *Escherichia coli*. *Proc Natl Acad Sci USA* 1986 ; 83 : 4650-4.
2. Chambers I, Frampton J, Goldfarb P, Alfara N, McBain W, Harrison PA. The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the termination codon, TGA. *EMBO J* 1986 ; 5 : 1221-7.
3. Sukenaga Y, Ishida K, Takeda T, Takagi K. cDNA sequence coding for human glutathione peroxidase. *Nucleic Acids Res* 1987 ; 15 : 71-78.
4. Leinfelder W, Zehelein E, Mandrand-Berthelot MA, Böck A. Gene for a novel tRNA species that accepts L-serine and co-translationally inserts selenocysteine. *Nature* 1988 ; 331 : 723-5.
5. Söll D. Genetic code: enter a new amino-acid. *Nature* 1988 ; 331 : 662-3.

John Libbey  
EUROTEXT  
LONDON PARIS

Paris - Londres  
EDITIONS MEDECINE-SCIENCES



## Médicaments anti-cancéreux

H. TAPIERO, J.M. ROBERT  
Th. J. LAMPIDIS

Co-édition INSERM/  
John Libbey Eurotext

1989, broché, vol. 191, 360 pages  
360 FF, 30 £, 58 \$ US

- Une confrontation entre cliniciens et chercheurs impliqués dans la pharmacologie et la chimiothérapie des médicaments anti-cancéreux
- Une meilleure compréhension des mécanismes fondamentaux responsables des échecs ou de l'efficacité de la chimiothérapie anti-cancéreuse.

### BON DE COMMANDE

NOM ..... Prénom .....

Adresse .....

Désire recevoir « Médicaments anti-cancéreux » au prix de 360 FF + 30 FF de frais de port, soit 390 FF.

Ci-joint mon règlement à l'ordre de  
**John Libbey Eurotext**  
6, rue Blanche,  
92120 Montrouge, France

# Régulation de l'expression des gènes et différenciation

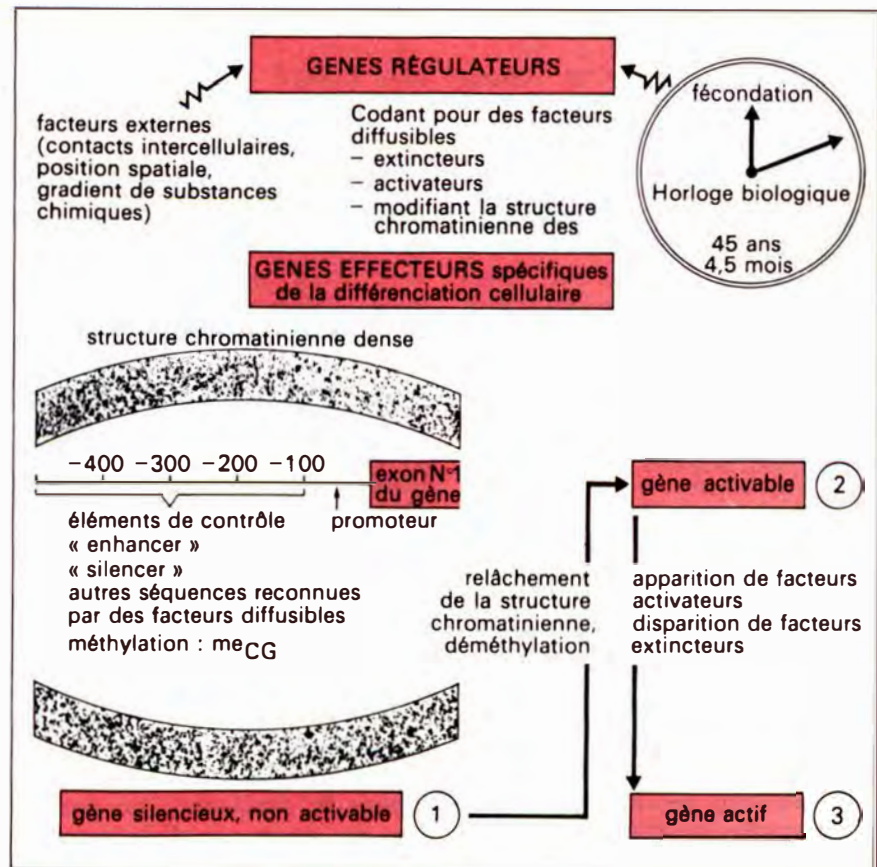
## Mécanismes transcriptionnels

On peut schématiquement opposer d'une part les gènes effecteurs et, parmi eux, ceux dont l'expression est spécifique d'un type donné de différenciation tissulaire et, d'autre part, les gènes régulateurs. Les premiers assurent une fonction structurale, métabolique, ou bien sont impliqués dans le cycle cellulaire et leur contrôle commence à être élucidé. Les seconds gouvernent l'expression des premiers et restent absolument inconnus chez les eucaryotes supérieurs, même si leur existence apparaît certaine au vu d'arguments fonctionnels et génétiques.

On suppose que des facteurs externes combinés avec une horloge biologique de nature et fonctionnement complètement inconnus, régulent l'expression de ces gènes. Ces facteurs externes peuvent être par exemple : la position spatiale au cours de l'embryogenèse, les contacts intercellulaires, ou des gradients de substances chimiques s'établissant dans l'embryon et jouant le rôle de différenciateurs.

Les meilleures illustrations de l'horloge biologique sont les phénomènes de puberté, ménopause, sénescence ainsi que le cycle cellulaire. Les gènes régulateurs coderaient pour des facteurs diffusibles, protéines, ou certaines classes d'ARN, qui seraient susceptibles de contrôler toute une catégorie de gènes activés à une phase donnée du développement d'un tissu particulier. Ces facteurs actifs en « trans » (1) reconnaîtraient des séquences régulatrices de la transcription des gènes qui sont elles-mêmes actives en « cis » (1); il pourrait s'agir d'activateurs stimulant les *enhancers* ou inhibant les *silencers* ou encore d'extincteurs ayant les effets inverses (voir les deux premiers Lexiques de médecine/sciences, pages 2 et 3).

D'autres substances (ou peut-être les mêmes, agissant à une concentration ou à une phase cellulaire dif-



férente) modifieraient la conformation des structures chromatinienne entourant les gènes.

Le mécanisme de l'activation d'un gène effecteur spécifique comporterait ainsi plusieurs stades comme le montre le schéma ci-dessus.

1. Dans un premier temps, le gène se trouve dans une structure chromatinienne dense qui ne permet ni sa transcription ni un accès facile de ses régions régulatrices aux facteurs diffusibles. De nombreux dimères CG (cytosine-guanine) sont méthylés sur le C<sup>m</sup>CG. Le gène n'est ni actif ni rapidement activable.

2. Au cours de la différenciation, la structure chromatinienne change, se relâche, rendant le gène accessible aux facteurs diffusibles et permet-

tant la transcription. Parallèlement, sans qu'il soit possible de déterminer avec certitude ce qui est cause et conséquence, on assiste à une déméthylation des doublets <sup>m</sup>CG. 3. En présence des facteurs activateurs (et/ou en l'absence des facteurs extincteurs) spécifiques de la différenciation tissulaire, et parfois après une stimulation hormonale appropriée, le gène peut alors être transcrit. A. K.

(1) Un effet est dit en « cis » lorsqu'il ne s'exerce que sur le brin d'ADN portant l'effecteur alors qu'il est dit en « trans », lorsqu'il s'exerce sur un gène situé à distance, éventuellement sur un autre brin d'ADN, via la synthèse de substances régulatrices diffusibles. Les *enhancers* ont une action en « cis » alors que les gènes codant pour des substances diffusibles régulatrices ont des actions en « trans ».

## *L'hétérocaryon, système de reprogrammation du génome cellulaire*

C'est à Ephrussi que l'on doit les premiers travaux sur l'emploi des hybrides somatiques dans la différenciation cellulaire. Dans les syncaryons obtenus par exemple entre fibroblastes et hépatocytes, il n'y a qu'un seul noyau et les hybrides poursuivent leurs divisions. Dans ce type d'hybride somatique, les résultats les plus habituels sont une extinction des fonctions codées par les chromosomes dérivés des cellules parentales différenciées (ici, les hépatocytes), ce phénomène d'extinction étant lié à un locus génétique précis du génome des cellules parentales indifférenciées (ici, les fibroblastes). Cependant, les cellules hybrides perdent progressivement des chromosomes, si bien que le développement ou l'extinction de fonctions spécifiques sont souvent difficiles à interpréter. Si l'on parvient à fusionner des cellules dont les noyaux restent distincts, on obtient des hétérocaryons. Dans ces conditions, toute action activatrice ou inhibitrice d'un type de cellules sur les fonctions de l'autre est nécessairement médiée par des substances qui ont diffusé dans le cytoplasme et agissent en trans sur le génome du noyau de l'autre type. Deux exemples ont démontré le pouvoir de cette technique. Le premier est celui des cellules musculaires. Ce modèle a été développé depuis 1983 par

Helen Blau (Stanford), qui a pu former des hétérocaryons stables entre des myocytes de souris et soit des cellules peu différenciées comme les amniocytes et des fibroblastes humains, soit des cellules très différenciées comme les hépatocytes. Les méthodes destinées à identifier les produits musculaires spécifiques de la souris et de l'homme ont évolué au cours des années depuis l'analyse électrophorétique [1] (créatine kinase, chaînes légères de la myosine) jusqu'à des sondes d'ADN spécifiques des  $\alpha$  actines de type musculaire et cardiaque, en passant par des anticorps dirigés contre des antigènes de surface caractéristiques des myoblastes ou des myotubes [2, 3]. Un fait fondamental ressort de ces travaux : les fibroblastes humains subissent une activation de leurs propres gènes musculaires et les protéines musculaires retrouvées dans le cytoplasme sont à la fois d'origine humaine et murine.

L'emploi récent de sondes d'actine a permis d'affiner l'analyse en quantifiant les ARN synthétisés. Le mode d'accumulation des transcrits humains dans les hétérocaryons s'est montré distinct de celui des cellules musculaires de souris et très semblable à celui de cellules musculaires humaines. La fusion de cellules musculaires de souris avec des fibroblastes humains a donc

déclenché chez ces derniers une activation de leurs gènes musculaires ; les cellules humaines agissent ensuite pour leur propre compte, tendant à se comporter comme d'authentiques cellules musculaires humaines, dont les facteurs régulateurs vont même pouvoir modifier l'évolution dans le temps des protéines de type souris.

Dans les expériences de Blau, le rapport du nombre de noyaux dérivés des cellules musculaires et non musculaires varie dans des limites assez étroites de 2/1 à 1/2. L'effet de dose ne s'est pas révélé important dans le cas des fibroblastes ; lorsque les cellules humaines étaient des hépatocytes, au contraire, le pourcentage des noyaux d'origine musculaire au départ exerçait une influence décisive. Les facteurs « d'extinction » semblaient donc beaucoup plus actifs lorsque la fusion avait lieu entre deux types de cellules différenciées.

On pouvait s'attendre à des difficultés pour étendre à d'autres cellules le modèle des hétérocaryons : les myocytes sont des cellules multinucléées, dont la forme assure un contact aisé avec les fibroblastes qui leur sont accolés. C'est pourquoi il faut souligner l'importance du travail de Baron et Maniatis (Harvard) qui ont réussi à former des hétérocaryons, de durée limitée à quel-

ques jours, à partir de cellules hémopoïétiques. Leur modèle fusionne des cellules K 562 d'origine humaine avec des cellules d'érythroleucémie de souris (MEL). Les K 562 synthétisent uniquement des globines embryonnaires et fœtales et non la  $\beta$  globine adulte, dont le gène est cependant présent. Les cellules MEL sont des précurseurs érythroïdes adultes transformés par le virus de Friend, qui produisent de très faibles quantités de globine  $\alpha$  et  $\beta$ . Traitées par le diméthylsulfoxyde (DMSO), elles subissent une différenciation terminale comportant une forte induction des messagers des globines  $\alpha$  et  $\beta$ . Les expériences ont été faites en 2 temps, distinguant l'effet de cellules induites ou non. Lorsqu'on fusionne des cellules MEL non induites avec des K 562, on observe uniquement l'apparition d'ARNm et de globine  $\epsilon$  de souris, précurseur embryonnaire de  $\beta$  (la souris ne possède pas d'équivalent des gènes  $\gamma$  humains). Les cellules fœto-embryonnaires humaines ont donc activé le gène embryonnaire de la souris. Si l'on utilise, pour la fusion, des cellules MEL induites par le DMSO, on constate la production de globine  $\beta$  adulte humaine en même temps que de globine  $\epsilon$  de souris. Il y a donc eu, dans ce cas, activation réciproque des gènes de globine. L'apparition des globines nouvelles est rapide, elle se fait en moins de vingt-quatre heures. Il n'existe aucune indication en faveur de l'existence de plusieurs étapes dans cette activation.

Les auteurs ont ensuite étendu leurs expériences en hybridant le partenaire humain K 562 avec des

cellules murines de lignées non hémato-poïétiques. Une expression des gènes de globine de souris  $\alpha$  et  $\epsilon$  a été obtenue, non seulement avec des fibroblastes en culture primaire — ce que l'on peut réaliser aussi par la méthode des syncaryons — mais aussi avec plusieurs lignées établies, alors que les syncaryons avaient toujours fourni une réponse négative. Dernière donnée importante : la réponse n'est pas identique pour toutes les cellules, ni pour tous les

gènes de globine ; le gène embryonnaire de type alpha murin, appelé  $\zeta$ , n'est jamais activé ; certaines lignées expriment à la fois  $\alpha$  et  $\epsilon$ , d'autres seulement  $\alpha$ .

De ces recherches émergent plusieurs conclusions importantes. La fusion de cellules spécialisées avec d'autres cellules, spécialisées ou non, provoque une reprogrammation du génome induisant la production de régulateurs spécifiques capables de déterminer l'expres-

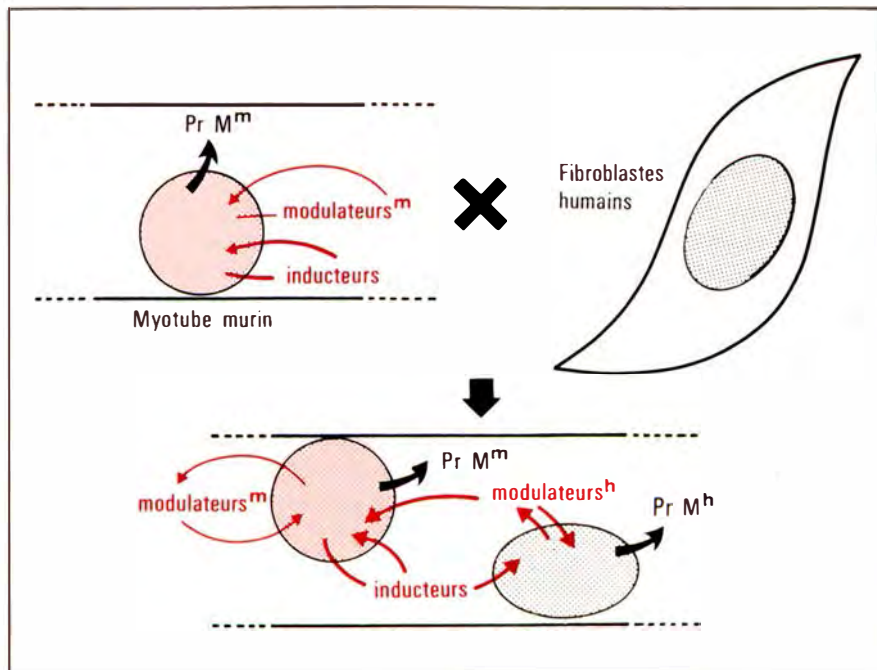


Figure 1. **Reprogrammation de l'expression des gènes codant pour des protéines musculaires dans les hétérocaryons entre des myotubes murins et des fibroblastes humains.** Modulateurs<sup>m</sup> = modulateurs murins, régulant le rapport et la chronologie de l'expression des gènes codant pour des protéines musculaires ; modulateurs<sup>h</sup> = modulateurs humains ; inducteurs = substances responsables de l'induction de l'expression des gènes codant pour des protéines musculaires ; Pr M<sup>m</sup> = protéines musculaires murines ; Pr M<sup>h</sup> = protéines musculaires humaines. Dans l'hétérocaryon, les inducteurs de souris induisent les gènes humains codant pour les « modulateurs humains » et les protéines musculaires humaines. Les modulateurs<sup>h</sup> agissent sur le noyau de souris et l'emportent sur les modulateurs<sup>m</sup>, l'expression du « programme musculaire » par le noyau de myotube murin devenant ainsi de type « humain ».

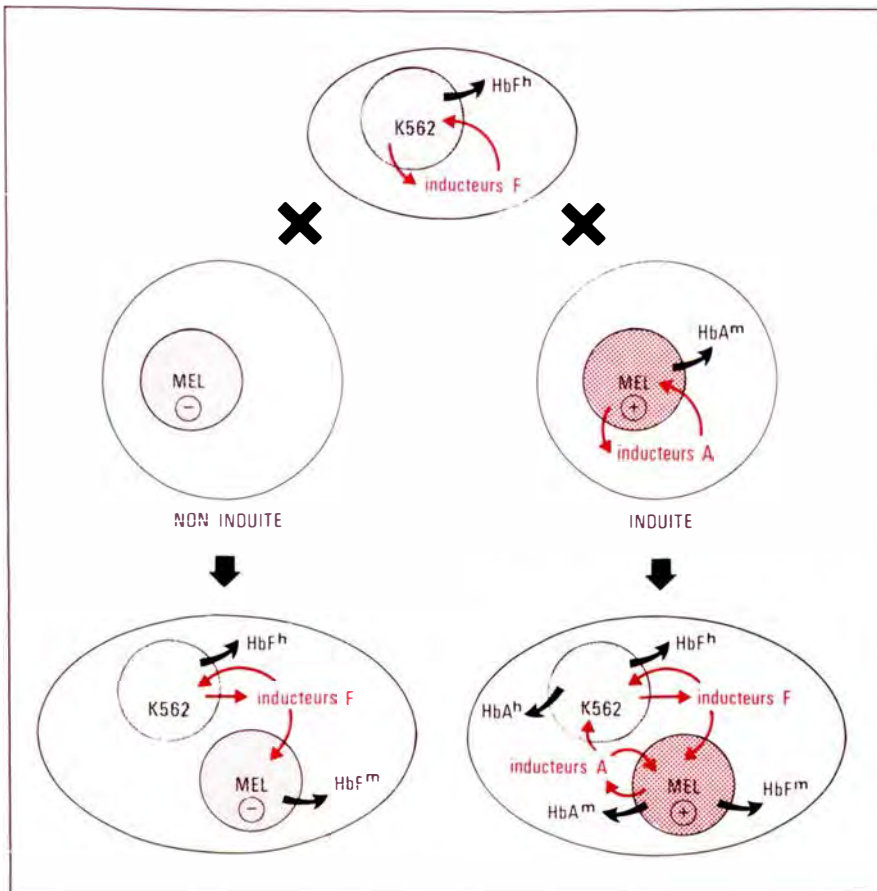


Figure 2. **Modulation respective de l'expression des gènes globines par les noyaux de cellules érythroleucémiques murines et de la lignée K 562 humaine dans des hétérocaryons.** HbF<sup>h</sup> = hémoglobines fœtales humaines ; HbF<sup>m</sup> = hémoglobines fœtales murines ; HbA<sup>h</sup> = hémoglobines adultes humaines ; HbA<sup>m</sup> = hémoglobines adultes murines ; inducteurs F = facteurs transactivateurs, programmés par les noyaux K 562 ; inducteurs A = facteurs transactivateurs adultes, programmés par les noyaux MEL induits ; MEL (-) = cellule MEL non traitée ; MEL (+) = cellule MEL traitée par le DMSO. Les flèches en noir indiquent la nature des hémoglobines synthétisées sous le contrôle des différents noyaux, celles en rouge indiquent l'origine et les points d'impact des facteurs trans-activateurs.

d'un équilibre dynamique entre les facteurs activateurs et extincteurs, les résultats résumés ici indiquent que la richesse en ce type de facteurs diffère selon les cellules. Ainsi, les facteurs « activateurs musculaires » pourraient être particulièrement abondants ou actifs, l'emportant, en cas de fusion myocyte/hépatocyte, sur les facteurs « activateurs hépatocytaires ». On ne comprend pas bien, dans l'état actuel de nos connaissances, la raison pour laquelle la constitution de syncaryons permet surtout de mettre en évidence les facteurs extincteurs alors que ce sont les phénomènes d'activation de fonctions différenciées qui prédominent dans les hétérocaryons. Enfin, pour un type donné de cellules, certains gènes sont plus facilement activés que d'autres ; tandis que, pour un gène donné, l'activation est plus facile à déclencher dans certains types cellulaires que dans d'autres.

J.-C.D.

sion d'un phénotype, par exemple myogénique ou érythroïde. Cette reprogrammation dépend de l'état des cellules « inductrices », puisque seules des cellules MEL traitées par le DMSO sont capables de provoquer la synthèse d'hémoglobine humaine adulte

dans des cellules fœtales humaines. Si, comme de très nombreux travaux actuels semblent l'indiquer et selon le modèle présenté ici même récemment (*m/s* n° 7, vol. 2, p. 410), l'expression des gènes spécifiques de l'état de différenciation est régulée en fonction

1. Blau HM, Chiu CP, Webster C. Cytoplasmic activation of human nuclear genes in stable heterokaryons. *Cell* 1983 ; 32 : 1171-80.
2. Hardeman EC, Chiu CP, Minty A, Blau HM. The pattern of actin expression in human fibroblast and mouse muscle heterokaryons suggests that human muscle regulatory factors are produced. *Cell* 1986 ; 47 : 123-30.
3. Blau HM, Pavlath GK, Hardeman EC *et al.* Plasticity of the differentiated state. *Science* 1985 ; 230 : 758-66.
4. Baron MH, Maniatis T. Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell* 1986 ; 46 : 591-602.



## ADN, protéines et transcription

L'activation d'un gène, c'est-à-dire le déclenchement de sa transcription, nécessite la constitution d'un complexe nucléoprotéique comportant l'ARN polymérase et de nombreuses autres protéines. La première étape de la constitution de ce complexe, dénommé « complexe d'initiation de la transcription », est probablement la fixation sur le promoteur de protéines reconnaissant les motifs « TATA box » ou d'autres séquences situées à proximité immédiate du site d'initiation de la transcription [1, 2]. Puis, correctement positionnée grâce à la

mise en place préalable des facteurs précédents, l'ARN polymérase se fixe, suivie d'autres protéines qui pourraient être requises pour l'activation de cette enzyme [1, 2]. Certains des partenaires du complexe d'initiation sont des protéines reconnaissant des séquences particulières d'ADN, d'autres se lient à ces protéines fixées à l'ADN plutôt qu'à l'ADN lui-même (figure 1). La formation ou l'activation de ce complexe sont soumises à l'influence régulatrice d'autres protéines se liant soit à proximité immédiate, soit à distance sur un

élément *enhancer* (lexique m/s, supplément au n° 7, vol. 1, p. 3 et m/s n° 7, vol. 2, p. 410). Dans ce dernier cas (figure 2), il semble que les deux complexes nucléoprotéiques (complexe d'initiation et complexe *enhancer*) puissent néanmoins interagir par contact direct, via la constitution de boucles d'ADN [3]. L'interaction entre des protéines liées à l'un et à l'autre de ces complexes, éventuellement entre molécules de la même protéine s'il existe un site de fixation équivalent au niveau du promoteur et du *enhancer*, peut constituer le mécanisme stabilisant

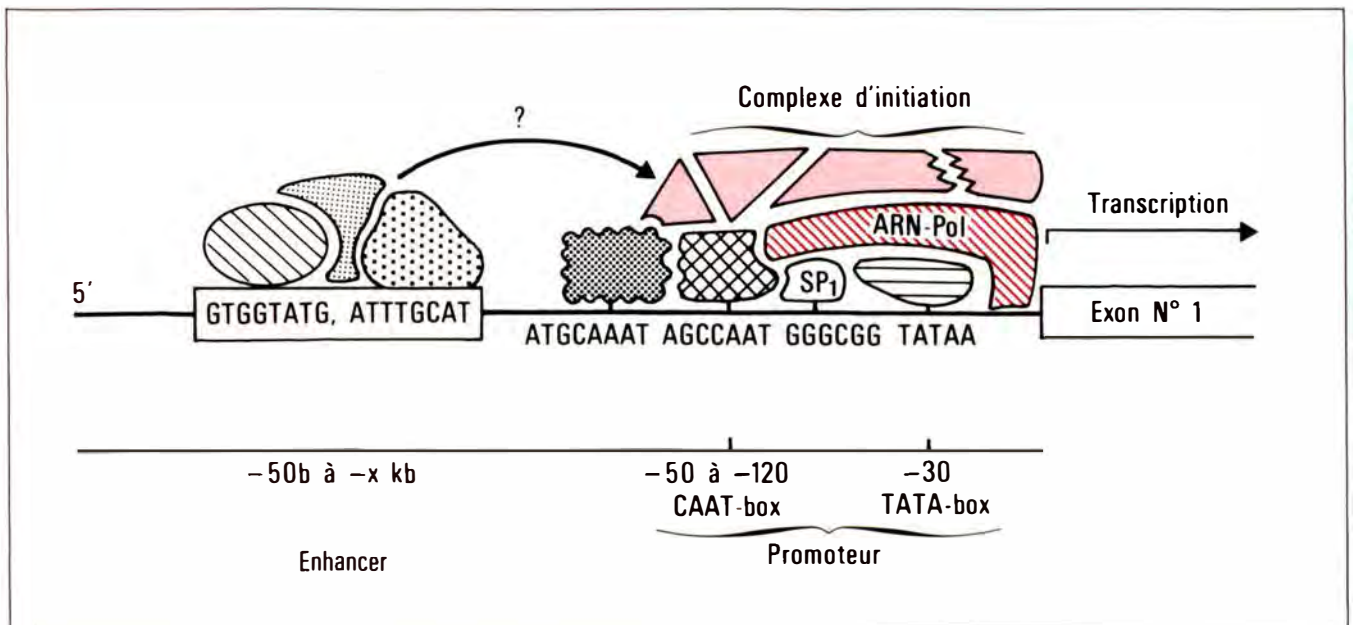


Figure 1. **Complexe d'initiation de la transcription et complexe enhancer.** Les distances sont exprimées, sous le schéma, en nombre de bases en amont du site d'initiation de la transcription. Le enhancer est situé à une distance très variable du site d'initiation, de quelques bases à plusieurs kilobases. Outre la « TATA box » (TATAA), la « CAAT box » (AGCCAAT), d'autres éléments fixant des facteurs transcriptionnels ont été représentés : le facteur SP1 se fixe sur une « GC box » (GGGCGG) ; le facteur NF-A reconnaît la séquence ATGCAAAT du promoteur et son inverse ATTTGCAT localisé au niveau du enhancer ; la séquence GTGGTATG du enhancer est un motif (le core enhancer) très fréquemment trouvé au niveau de ce type d'élément. Les mécanismes de l'interaction restent hypothétiques.

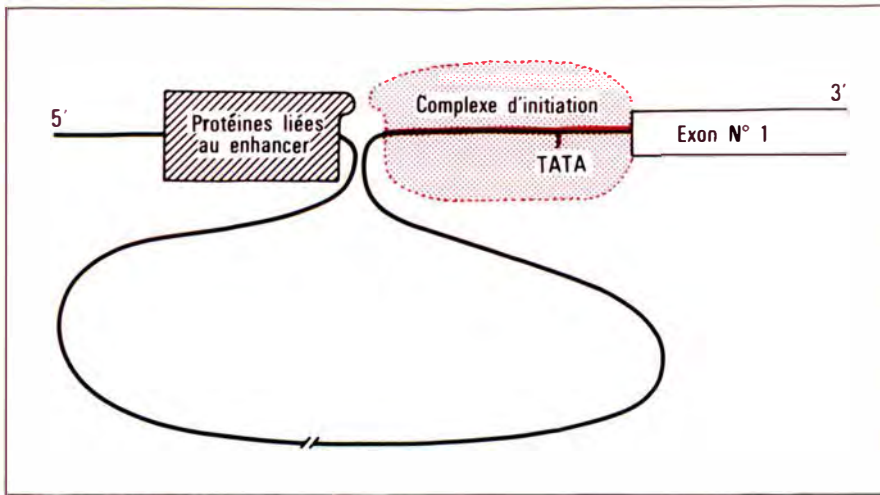


Figure 2. **Schéma de l'interaction directe** entre les éléments du complexe enhancer et ceux du complexe d'initiation.

ces boucles, comme nous l'avons rapporté pour le répresseur du phage lambda (*m/s n° 7, vol. 3, p. 428*). Une telle hypothèse constituerait une explication rationnelle de l'efficacité particulière de l'enhancer du gène des chaînes lourdes d'immunoglobuline sur les promoteurs situés en amont des segments codant pour les parties variables de ces gènes. Ces deux éléments, promoteurs et enhancers, contiennent en effet, dans l'un ou l'autre sens (*figure 1*), un motif dénommé octamère qui lie une protéine spécifique NF-A (il existerait une forme ubiquitaire de cette protéine [NF-A<sub>1</sub> ou IgNF-A<sub>1</sub>] et une forme spécifique des lymphocytes T [IgNF-A<sub>2</sub>]). Deux molécules de NF-A, fixée l'une sur le promoteur et l'autre sur le enhancer pourraient ainsi conduire à la formation de boucles d'ADN rapprochant les complexes enhancer et d'initiation [4, 6]. Une autre question concerne les relations existant entre une protéine fixée à l'ADN et la stimulation (ou l'inhibition) transcriptionnelle qui s'ensuit. Dans un certain nombre de cas, particulièrement bien illustrés par les protéines GAL-4 et GCN-4 de levure, on a pu démontrer que les domaines protéiques responsables de la fixation

à l'ADN et de l'activation transcriptionnelle étaient différents [7-9]. A condition que la séquence reconnue soit insérée à proximité du promoteur contrôlé, il est possible de remplacer GAL-4 par une protéine hybride dont la partie N-terminale est constituée du domaine de liaison à l'ADN d'un répresseur d'*E. coli*, seule la partie C-terminale appartenant à GAL-4. Des expériences de mutagenèse dirigée ont permis de démontrer que les contraintes de séquence pour que les protéines GAL-4 et GCN-4 restent activatrices étaient extrêmement faibles, nécessitant seulement la présence d'une séquence composée d'acides aminés acides dans la moitié C-terminale des molécules. Cette grande variabilité des structures peptidiques susceptible d'assurer l'activation transcriptionnelle suggère que le mécanisme n'est pas ici de nature catalytique (action de topo-isomérase, nucléase, méthylase, etc.) mais implique bien, plutôt, une interaction physique directe entre macromolécules via des liaisons ioniques, hydrogène ou hydrophobes (*figure 3, partie 2*). Dans le cas des récepteurs des hormones stéroïdes (*m/s n° 3, vol. 3, p. 172*), plusieurs auteurs ont démontré qu'il était

possible de déléter largement les régions N-terminales (domaines A et B) et C-terminales (régions D et E) de la protéine. Le récepteur tronqué, réduit à moins de 100 résidus codés par la région C, a alors une activité « constitutive », c'est-à-dire indépendante de l'hormone [10, 11]. Ces résultats indiquent que la région E correspond probablement à un domaine inhibiteur de la fixation du récepteur à l'ADN, l'inhibition étant levée lorsque l'hormone se fixe à son site situé dans ce domaine E. Les domaines A et B ont un rôle encore mal compris ; ils pourraient augmenter la spécificité de la fixation du récepteur aux séquences d'ADN responsables de la réponse transcriptionnelle à l'hormone, ou encore participer à l'activation transcriptionnelle. Quoi qu'il en soit, il n'est pour l'instant pas possible de séparer fonctionnellement les régions du récepteur responsables de la fixation à l'ADN de celles indispensables à la stimulation de la transcription. Une tierce protéine, différente du récepteur, pourrait constituer un intermédiaire transmettant l'information activatrice depuis la région du récepteur interagissant avec l'ADN jusqu'au complexe d'initiation (*figure 3, partie 2*). Un autre mécanisme, d'ailleurs non exclusif du précédent, pourrait impliquer des modifications de la torsion et de la courbure de l'ADN secondaires à la fixation d'une ou de plusieurs molécules de récepteurs. Le caractère coopératif de la réponse transcriptionnelle, lorsque plusieurs sites sont occupés [12], pourrait militer en faveur de l'intervention d'un tel phénomène, puisque de telles fixations multiples sont susceptibles d'augmenter les contraintes induites dans la double hélice (*m/s n° 7, vol. 3, p. 428*). Dans cette hypothèse, schématisée dans la *figure 3, partie 3*, la stimulation de la transcription serait transmise à distance, via ces modifications topologiques de l'ADN. La période actuelle connaît une

activité fébrile de purification et de caractérisation des facteurs transcriptionnels spécifiques de motifs particuliers d'ADN, reconnaissant des séquences localisées dans les promoteurs ou les *enhancers* et parfois dans l'un et l'autre. Le groupe de Tjian à Berkeley (Californie, USA) est probablement le plus actif dans ce domaine ; il a notamment déjà cloné l'ADN complémentaire codant pour le facteur SP1 qui se fixe sur des régions riches en GC

des promoteurs (« GC boxes », *figure 1*) ([13] et résultats non publiés). Au moins six autres facteurs ont été purifiés totalement par ce groupe ou par d'autres et il est probable que leurs gènes spécifiques seront bientôt isolés. Tous ces résultats permettent enfin d'entrevoir le mécanisme du contrôle de l'expression des gènes au cours de la différenciation, ainsi que nous l'envisagerons dans une prochaine nouvelle.

Axel Kahn

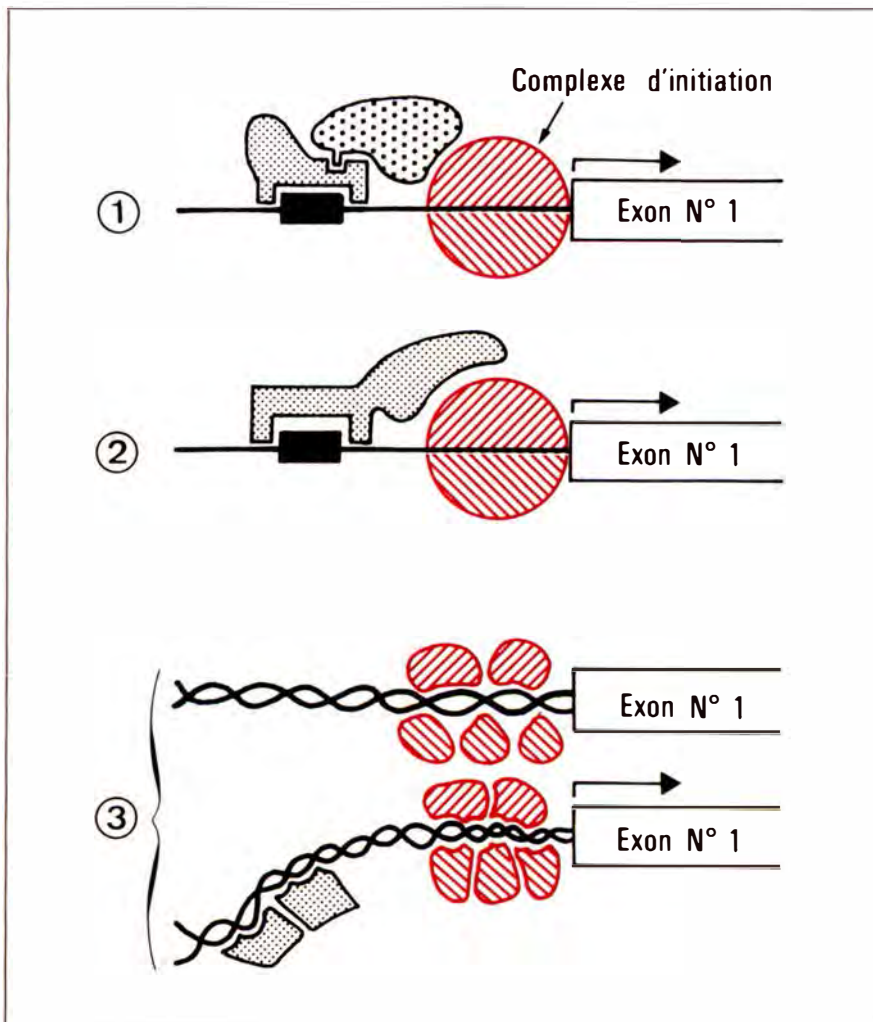


Figure 3. **Modèle de la transmission du message d'activation transcriptionnelle entre une protéine liée à l'ADN et le complexe d'initiation.** (1) transmission par l'intermédiaire d'une tierce protéine. (2) facteur transcriptionnel bifonctionnel, un domaine du facteur assurant la liaison à l'ADN alors qu'un autre domaine est responsable de l'effet transcriptionnel. (3) modification topologique de l'ADN (courbure, modification de l'hélicité) induite par la fixation d'une protéine. Cette modification de la courbure et de l'hélicité transmet, à distance, le message au complexe d'initiation.

## RÉFÉRENCES

1. Reinberg D, Rooder RG. Factors involved in specific transcription by mammalian RNA polymerase II. *J Biol Chem* 1987 ; 262 : 3310-37.
2. Hawley DK, Rooder RG. Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. *J Biol Chem* 1987 ; 262 : 3452-61.
3. Ptashne M. Gene regulation by proteins acting nearby and at a distance. *Nature* 1986 ; 322 : 697-701.
4. Staudt LM, Singh H, Sen R, Wirth T, Sharp PA, Baltimore D. A lymphoid specific protein binding to the octamer motif of immunoglobulin genes. *Nature* 1986 ; 323 : 640-3.
5. Doyen N, Leblond-Francmillard M, Holm I, Dreyfus M, Rougeon F. Analysis of promoter and enhancer cell type specificity and the regulation of immunoglobulin gene expression. *Gene* 1986 ; 50 : 321-31.
6. Dreyfus M, Doyen N, Rougeon F. The conserved decanucleotide from the immunoglobulin heavy chain promoter induces a very high transcriptional activity in B-cells when introduced into an heterologous promoter. *EMBO J* 1987 ; 6 : 1685-90.
7. Brent R, Ptashne M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 1985 ; 43 : 729-36.
8. Ma J, Ptashne M. Deletion analysis of GAL-4 define two transcriptional activating segments. *Cell* 1987 ; 48 : 847-53.
9. Struhl K. Promoters, activator proteins and the mechanism of transcriptional initiation in yeast. *Cell* 1987 ; 49 : 295-7.
10. Hollenberg SM, Giguere V, Segui P, Evans RM. Co-localization of DNA binding and transcriptional activation functions in the human glucocorticoid receptor. *Cell* 1987 ; 49 : 39-46.
11. Miesfeld R, Godowski PJ, Maler BA, Yamamoto K. Glucocorticoid receptor mutants that define a small region sufficient for enhancer activation. *Science* 1987 ; 236 : 423-7.
12. Jantzen HM, Strähle U, Gloss B, et al. Cooperativity of glucocorticoid response element located far upstream of the tyrosine aminotransferase gene. *Cell* 1987 ; 49 : 29-38.
13. Briggs MR, Kadonaga JT, Bell SP, Tjian R. Purification and biochemical characterization of the promoter specific transcription factor SP1. *Science* 1986 ; 234 : 47-52.

## La riche union de l'ADN et des protéines

La régulation de la transcription des gènes implique, dans toutes les cellules vivantes, des interactions entre l'ADN des séquences régulatrices et des protéines qui contrôlent, *in fine*, la fixation sur le promoteur de l'ARN polymérase. Plus généralement, d'ailleurs, de tels contacts entre des protéines et des acides nucléiques interviennent aussi dans la réplication de l'ADN, les réactions d'excision-épissage, la maturation des extrémités 5' et 3' des transcrits, le stockage et la traduction des ARN messagers, la formation des ribosomes et des autres particules ribonucléoprotéiques... c'est-à-dire dans la totalité des réactions mettant en jeu des acides nucléiques.

La mieux étudiée des interactions entre une protéine et un double brin d'ADN concerne le répresseur et la protéine Cro du phage lambda et leurs sites opérateurs de fixation [1-5]. La fixation alternative de ces deux protéines aux sites opérateurs du phage contrôle soit le maintien de l'ADN phagique à l'état intégré dans le chromosome bactérien (phénomène de lysogénie), soit le déclenchement du cycle lytique.

La figure 1 montre un schéma des interactions entre un dimère de répresseur et la double hélice d'ADN : chaque monomère de répresseur est composé de quatre hélices  $\alpha$  reliées par des feuillettes  $\beta$  (hélice  $\alpha$  et feuillettes  $\beta$  correspondent à deux des conformations usuelles des protéines). Une hélice  $\alpha$  se positionne dans le grand sillon de la double hélice d'ADN, parallèlement à ses berges, établissant des contacts étroits et spécifiques avec certaines des bases. Une autre hélice  $\alpha$ , perpendiculaire à la précédente, établit des contacts avec le squelette de ponts phosphodiester de l'ADN. Les deux monomères sont fixés de façon identique sur la même face de l'ADN, à un tour de spire de distance de la dou-

ble hélice (c'est-à-dire 10 paires de bases) (figures 1 et 2A).

Des dimères liés à des sites proches sur la même face de l'ADN établissent entre eux de puissantes liaisons hydrogènes et hydrophobes, de telle sorte que leur fixation sur l'ADN est hautement coopérative (figure 2B) : la fixation d'une molécule de répresseur interagissant avec une autre molécule augmente la concentration locale de ces protéines et donc l'affinité pour elles du site d'ADN non encore occupé. Cette coopérativité se manifeste encore lorsque les sites de reconnaissance sont éloignés l'un de l'autre, à condition cependant qu'ils soient toujours situés sur une même face

de l'ADN, c'est-à-dire qu'ils restent séparés d'un nombre entier de tours de spire.

La figure 2C montre que la conséquence d'une telle fixation est une courbure imposée au double brin d'ADN, amenant un « baillement » du grand sillon à l'extérieur de la courbure et son pincement à l'intérieur [1]. Une telle courbure de l'ADN n'est d'ailleurs pas forcément la conséquence de la fixation d'une protéine : elle peut au contraire la précéder et la faciliter [2, 5]. Le site d'ADN reconnu par la protéine est alors un double brin incurvé. Les séquences séparant deux points de contact de la protéine et de l'ADN, séquences

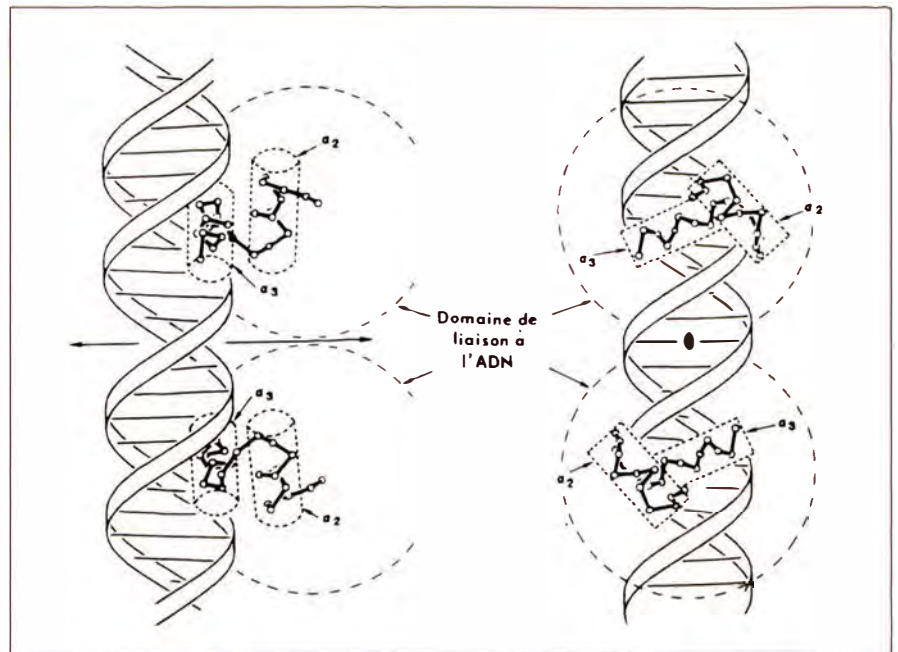


Figure 1. **Interactions entre le répresseur et un opérateur du phage lambda : modèle « hélice-tour-hélice ».** Deux des quatre hélices  $\alpha$  de chaque monomère,  $\alpha_2$  et  $\alpha_3$ , établissent des contacts avec, respectivement, le squelette de ponts phosphodiester et les bases du grand sillon du double brin d'ADN. Deux monomères sont ainsi fixés sur une même face de l'ADN, liés à des séquences situées à un tour d'hélice de distance. La flèche (schéma de gauche) et l'ovale (schéma de droite) indiquent le centre de symétrie des sites de fixation. Les deux monomères sont liés entre eux par des interactions hydrogènes et hydrophobes.

n'ayant elles-mêmes aucune interaction directe avec la protéine, peuvent ainsi influencer sur l'affinité pour cette protéine (figure 2) [2, 5]. Il existe probablement une correspondance assez étroite entre les acides aminés et les bases au niveau des sites de reconnaissance et, dans des cas très privilégiés, on a pu rétablir l'affinité d'un site muté pour sa protéine spécifique en mutant également celle-ci au niveau du résidu interagissant avec la base modifiée de l'ADN [6].

Le modèle d'interaction que je viens de décrire, appelé « hélice-tour-hélice », se retrouve dans bien d'autres systèmes que celui du phage lambda, dans la levure comme chez des eucaryotes supérieurs. L'interaction de « l'homéobox » des gènes homéotiques avec l'ADN serait de ce type. La particularité des sites d'ADN participant à ce type de contact est d'être « palindromiques », c'est-à-dire d'avoir une symétrie en miroir et de comporter ainsi des séquences identiques sur les deux brins (figure 2A) interagissant avec deux monomères de la protéine spécifique, également symétriques l'un par rapport à l'autre.

Un second modèle de liaison d'une protéine à une séquence spécifique d'ADN est celui des protéines dactyles dont l'archétype est TFIII-A, un facteur de transcription se fixant à un promoteur spécifique de l'ARN polymérase III et initialement caractérisé chez une espèce d'amphibien, le xénope. Le promoteur des gènes transcrits par l'ARN polymérase III a la particularité remarquable d'être localisé dans la séquence transcrite elle-même. La séquence de TFIII-A comporte un domaine N-terminal de 30 kDa, composé de 9 répétitions de 30 acides aminés, parmi lesquels 2 cystéines et 2 histidines placées en des positions invariantes. Miller *et al.* [7] ont suggéré que ces répétitions correspondaient à des structures en doigts, stabilisées par un atome de zinc échangeant 4 liaisons de coordination, 2 avec les cystéines et 2 avec les histidines (figure 3) [7, 8]. Différentes méthodes de

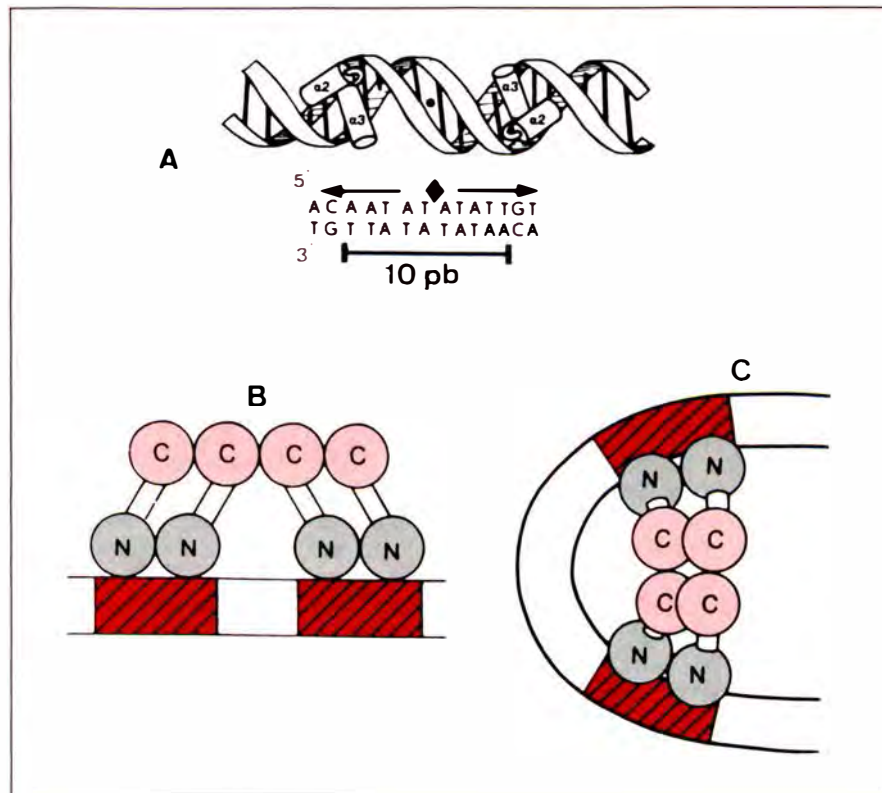


Figure 2. **Coopérativité dans la fixation du répresseur du phage lambda à l'opérateur et induction de courbures de l'ADN.** La figure 2A montre un opérateur « canonique », constitué d'une séquence palindromique, dont les centres de chaque hémi-site sont espacés de 10 paires de bases, c'est-à-dire d'un tour d'hélice. Les séquences des hémi-sites de fixation sont symétriques, de même que les monomères interagissant avec eux. Le centre de symétrie est symbolisé par un point noir au niveau du double brin et un losange au niveau de la séquence. La figure 2B montre la fixation de deux dimères à des sites proches (en rouge hachuré) du double brin d'ADN : les extrémités N-terminales des monomères forment la région de liaison à l'ADN, alors que les extrémités C-terminales interagissent entre elles, entraînant un phénomène de coopérativité dans la fixation des molécules. En 2C, si les sites opérateurs sont éloignés l'un de l'autre tout en demeurant sur une même face de l'ADN, l'interaction entre les molécules de répresseur entraînera une courbure du double brin... et la fixation restera coopérative.

protection du complexe TFIII-A-ADN contre la digestion par des nucléases ont indiqué que la zone d'interaction était de 50 bases et n'intéressait qu'un seul des deux brins, avec des points de contact toutes les 5 bases, soit tous les 1/2 tours de spire de l'ADN [9], selon l'un des deux modèles proposés par Fairall *et al.* [10] (figure 3). Les particularités de la liaison de TFIII-A à l'ADN apportent un élément d'explication à la manière dont cette protéine, fixée dans le gène lui-même, peut ne pas inter-

rompre la transcription par l'ARN polymérase III : on peut supposer que les doigts proximaux se détachent de leur site lors du « passage » de la polymérase, puis se fixent à nouveau, alors que la dissociation des doigts distaux laisse le champ libre à cette polymérase. L'intérêt pour les protéines dactyles a été considérablement renforcé par la découverte récente de possibles structures en doigts au niveau de nombreuses protéines nucléaires [11-13], notamment des produits de gènes

intervenant dans le développement de la drosophile et au niveau des récepteurs des hormones stéroïdes et de facteurs transcriptionnels ubiquitaires, tel Sp1 [14].

Il existe, cependant, d'importantes différences entre ces dernières protéines et TFIII-A, touchant au nombre et à la nature des « doigts ». De plus, les récepteurs des hormones stéroïdes se fixent en général en amont des gènes transcrits par l'ARN polymérase II et n'ont pas à être « déplacés » par le passage de la polymérase, si bien que leur ressemblance

structurale et fonctionnelle avec TFIII-A pourrait être en fait assez lointaine. Le fait que ces récepteurs, contrairement à TFIII-A, interagissent avec les deux brins d'ADN, va dans le même sens. De ce qui précède, il faut retenir qu'il existe des interactions très spécifiques entre l'ADN et des protéines jouant un rôle essentiel de régulation de la réplication et de la transcription. L'anatomie des complexes ADN-protéine est mieux connue dans le cas du modèle « hélice-tour-hélice », dont un exemple, l'interaction du

répresseur et des opérateurs du phage lambda, a pu bénéficier de l'apport de la cristallographie des complexes. Des courbures de l'ADN interviennent dans la constitution du site de reconnaissance de la protéine et sont provoquées ou amplifiées par la fixation de cette protéine. Nous envisagerons ultérieurement le mécanisme de l'effet transcriptionnel des protéines qui sont impliquées dans ce phénomène et de la régulation de leur activité.

A. K.

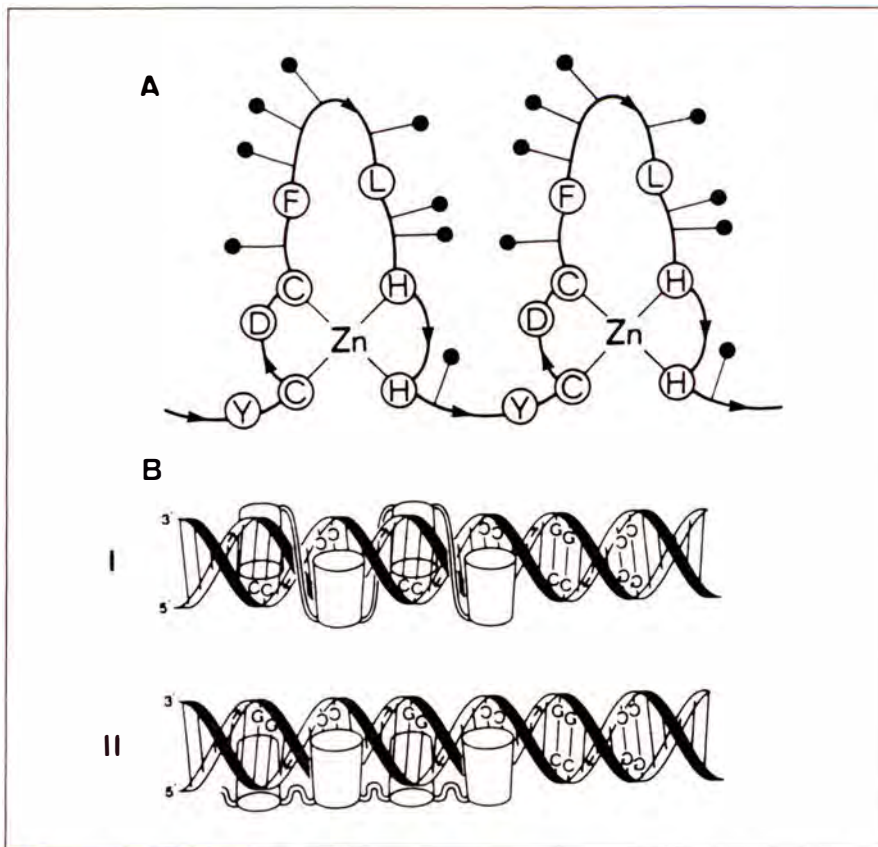


Figure 3. **Protéines dactyles et modèle de leur interaction avec l'ADN.** En 3A sont représentés, selon Miller et al. [7], deux « doigts » protéiques stabilisés par des liaisons de coordination entre un atome de zinc d'une part, 2 cystéines (C) et 2 histidines (H) d'autre part. Ces doigts sont constitués au niveau d'éléments répétitifs contenant, outre les cystéines et histidines invariantes, des acides aminés acides (D = acides aspartiques) et hydrophobes (F = phénylalanine, L = leucine, Y = tyrosine). Les points noirs indiquent les contacts probables avec l'ADN. La figure 3B présente, selon Fairall et al. [10] deux modèles possibles d'interaction entre le facteur TFIIIA et l'ADN. Les doigts sont représentés par les cylindres. Selon le modèle I, la protéine est enroulée le long d'un des brins d'ADN, les doigts établissant des contacts avec lui tous les 1/2 tours de spire. Selon le modèle II, la protéine est parallèle au double brin, les doigts établissant des contacts alternativement sur l'une et l'autre face du double brin.

1. Anderson JE, Ptachne M, Harrison SC. Structure of the repressor-operator complex of bacteriophage 434. *Nature* 1987 ; 326 : 846-52.
2. Koudelka GB, Harrison SC, Ptachne M. Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and cro. *Nature* 1987 ; 326 : 886-8.
3. Robertson M. Specificity and flexibility. *Nature* 1987 ; 327 : 464-6.
4. Takeda Y, Ohlendorf DH, Anderson WF, Matthews BW. DNA-binding proteins. *Science* 1983 ; 221 : 1020-6.
5. Zahn K, Blattner FR. Direct evidence for DNA bending at the lambda replication origin. *Science* 1987 ; 236 : 416-22.
6. Wharton RP, Ptachne M. A new-specificity mutant of 434 repressor that defines an amino acid base pair contact. *Nature* 1987 ; 326 : 888-91.
7. Miller J, Mc Lachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from xenopus oocytes. *EMBO J* 1985 ; 4 : 1609-14.
8. Berg JM. Potential metal binding domains in nucleic acid binding proteins. *Science* 1986 ; 232 : 485-7.
9. Rhodes D, Klug A. An underlying repeat in some transcriptional control sequences corresponding to half a double helical turn of DNA. *Cell* 1986 ; 46 : 123-32.
10. Fairall L, Rhodes D, Klug A. Mapping of the sites of protection on a 5S RNA gene by the xenopus transcription factor IIIA-A model for the interaction. *J Mol Biol* 1986 ; 192 : 577-91.
11. Chowdhury K, Deutsch U, Gruss P. A multigene family encoding several finger structures is present and differentially active in mammalian genomes. *Cell* 1987 ; 48 : 771-8.
12. Tautz D, Lehmann R, Schnürch H, et al. Finger protein a novel structure encoded by Hunchback, a second member of the gap class of drosophila segmentation genes. *Nature* 1987 ; 327 : 383-9.
13. Vincent A. TFIIIA and homologous genes. The « finger » proteins. *Nucleic Acids Res* 1986 ; 14 : 4385-91.
14. Briggs MR, Kadonaga JT, Bell SP, Tjian R. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* 1986 ; 234 : 47-52.

## Régulation en cis et en trans de la différenciation cellulaire

Il est maintenant parfaitement clair que l'activité des gènes est contrôlée par des protéines se fixant sur des motifs d'ADN particuliers des régions régulatrices (voir « la riche union de l'ADN et des protéines », *m/s* n° 7, vol. 3, p. 428, et « ADN, protéines et transcription », *m/s* n° 8, vol. 3, p. 487). En première analyse, cependant, cette constatation ne fait que repousser le problème de la différenciation puisque l'on peut se demander ce que sont les mécanismes contrôlant la synthèse et l'activité des protéines actives sur la transcription au cours du développement et de la différenciation. Par ailleurs, la diversité des programmes d'expression des gènes au cours du développement est extraordinaire : non seulement il existe un très grand nombre de types de différenciation cellulaire, mais encore ces types peuvent être divisés en de multiples périodes (in vivo, par exemple, périodes embryonnaire précoce, fœtale, néonatale, post-natale, adulte, etc.) caractérisées par l'expression de gènes marqueurs spécifiques. De plus, au cours d'une même « période » du développement d'un tissu donné, l'expression de ces marqueurs spécifiques est souvent asynchrone. La question des mécanismes contrôlant cette impressionnante diversité de comportement des gènes, se pose donc avec acuité.

**Au début de tout, la membrane.** La membrane plasmique est la première cible des signaux « différenciateurs » aussi bien que des signaux mitotiques (*m/s* n° 5, vol. 3, p. 174) ; c'est elle qui, précocément dans l'embryogenèse, va « percevoir » un gradient de morphogènes (*m/s* n° 5, vol. 3, p. 178), va lier sur un récepteur spécifique un facteur de différen-

ciation (*m/s* n° 6, vol. 2, p. 340), ou bien va recevoir une information par contact direct avec les cellules adjacentes.

On savait déjà que certains gènes intervenant dans le développement embryonnaire codent pour des molécules étrangement proches des facteurs de croissance (produits des gènes *notch* et *lin 12*, *m/s* n° 6, vol. 2, p. 340), d'autres pour des protéines ressemblant aux protéoglycanes de la matrice extracellulaire (gène *per*, voir *mini-synthèse sur les canaux de jonction intercellulaire dans ce numéro*). Plus récemment, l'équipe de G.M. Rubin en Californie a pu déterminer la structure du gène *sevenless* de la drosophile, impliqué dans la différenciation des photorécepteurs. Ce gène a le potentiel de coder pour un récepteur membranaire doté d'une activité tyrosine-kinase [1], comme tant de récepteurs de facteurs de croissance et de produits d'oncogène (*m/s* n° 8, vol. 3, p. 492). Les gènes du développement et les oncogènes ne sont d'ailleurs pas de nature différente et tout indique que certains oncogènes sont probablement impliqués dans la différenciation cellulaire (*m/s* n° 4, vol. 2, p. 222).

**Modifications post-traductionnelles des facteurs transcriptionnels.** Il y a peu encore, ce qui se passait entre la perception par une cellule d'un signal membranaire et la réponse transcriptionnelle au niveau du noyau, constituait un véritable « trou noir ». Le mode d'action des hormones stéroïdes suggérait cependant un modèle simple de modulation de l'activité d'un gène par activation, sous l'effet d'un inducteur externe (l'hormone), d'un facteur transcriptionnel pré-existant (le récepteur de l'hormone, *m/s* n° 3, vol. 3, p. 172). Dans ce cas,

cependant, l'hormone pénètre dans la cellule sans intervention d'un récepteur membranaire et de « seconds messagers ». Chez *E. coli*, la réponse transcriptionnelle à des stimuli extérieurs (modification du milieu de culture), comporte la synthèse d'AMP cyclique qui se fixe à un facteur transcriptionnel pré-existant (la protéine CAP) et l'active [2]. De la même manière, l'activation du répresseur de l'opéron tryptophane est due à une modification de la structure de la protéine, provoquée par la liaison d'une molécule de tryptophane [3] ; la trans-conformation ainsi induite rend le répresseur apte à reconnaître le site opérateur qui contrôle la transcription de l'opéron. Depuis quelques mois, il est apparu qu'un tel modèle de l'activation, par un inducteur, d'un facteur transcriptionnel pré-existant était probablement de portée générale. Le *enhancer* du gène codant pour la chaîne légère kappa des immunoglobulines (IgG- $\kappa$ ) n'est actif que dans les cellules lymphocytaires B ; son activation est corrélée avec la formation d'un complexe entre un motif du *enhancer* (motif B) et la protéine NF- $\kappa$ B. Cette protéine, en fait, n'est absolument pas spécifique des lymphocytes B et peut être détectée dans toutes les cellules où elle est cependant incapable de se fixer au motif B du *enhancer* IgG- $\kappa$ . Le traitement des cellules non différenciées dans la voie des lymphocytes B par des activateurs de la protéine kinase C (*m/s* n° 3, vol. 3, p. 174) active NF- $\kappa$ B, c'est-à-dire la rend capable de se fixer à son site d'ADN [4]. Deux autres facteurs transcriptionnels, probablement impliqués dans la réponse transcriptionnelle de certains gènes lors de la stimulation de la prolifération cellulaire, pour-

raient aussi être activés par la modification de molécules pré-existantes inactives : les protéines SRF, se fixant notamment sur le *enhancer* de l'oncogène *c-fos* [5, 6] et AP1 [7, 8], dont plusieurs cibles sont connues. Dans ces exemples, cependant, « l'activation » des facteurs transcriptionnels ne semble pas, au moins *in vitro*, agir sur leur capacité de liaison à l'ADN, mais plutôt sur leur aptitude à interagir avec le complexe d'initiation de la transcription.

Les facteurs transcriptionnels probablement responsables de la réponse des gènes métallothionéine aux métaux lourds (MRE-binding protein, protéine se liant à l'élément de réponse aux métaux) [9] et de la réponse à la chaleur des gènes codant pour les protéines du choc thermique (HSF, *heat shock transcriptional factor*) [10] sont également activés après leur synthèse (activation post-traductionnelle).

**Un nombre peut-être limité de facteurs...** Les travaux qui se multiplient sur la mise en évidence et la caractérisation de protéines pouvant jouer un rôle dans le contrôle de l'expression des gènes, débouchent sur deux constatations à première vue paradoxales : les mêmes facteurs sont souvent retrouvés impliqués dans le contrôle de gènes fort différents, et certains facteurs quoiqu'ils reconnaissent un motif indispensable à l'expression d'un gène spécifique d'un type particulier de différenciation, sont en fait parfaitement ubiquitaires (voir l'exemple de NF- $\kappa$ B). C'est ainsi que la protéine IgNF-A, qui reconnaît la séquence octamérique dont nous avons parlé dans la première nouvelle de cette série (*m/s n° 7, vol. 3, p. 428*) joue un rôle essentiel dans l'activation du *enhancer* des chaînes lourdes des immunoglobulines au cours de la différenciation des lymphocytes B, mais aussi dans le fonctionnement d'autres promoteurs [12]. La répartition d'IgNF-A est, de plus, ubiquitaire\*. Ces autres exemples illustrent ce phénomène de

l'intervention possible de mêmes facteurs dans des types différents de régulation ; ils concernent les facteurs SRF et AP1, dont nous avons parlé plus haut. SRF se fixe sur le *enhancer* de l'oncogène *c-fos* et en active probablement ainsi la transcription [5,6]. Françoise Phan Dinh Tuy et Adrian Minty, à Paris, ont montré que ce même facteur se liait à des séquences situées en amont des gènes d'actine musculaire et indispensables à l'activation de la transcription de ces gènes au cours de la différenciation musculaire (*communication personnelle*). Une protéine impliquée dans la réponse de gènes à l'AMP cyclique a été récemment caractérisée [13] ; elle reconnaît une séquence pratiquement identique à celle reconnue par AP1 [7, 8] et, de plus, lorsqu'un gène est contrôlé à la fois par les esters de phorbol\*\* et l'AMP cyclique, cette même séquence est indispensable aux deux types de contrôle [4]. Ces résultats suggèrent donc que la protéine AP1 et la protéine responsable de la réponse à l'AMP cyclique pourraient être identiques.

**Création de diversités.** La diversité des contrôles transcriptionnels auxquels sont soumis les gènes peut naturellement être le reflet de la diversité des facteurs régulateurs. Il faut remarquer qu'une telle solution serait peu « économique » puisqu'il faudrait alors envisager, en poussant cette hypothèse jusqu'à l'absurde, l'existence d'un gène régulateur par gène effecteur, ainsi que des gènes contrôlant les gènes régulateurs ! Nous avons, de plus, déjà signalé que les mêmes facteurs intervenaient souvent dans le contrôle de gènes différents.

Dans certains cas, cependant, des facteurs transcriptionnels semblent bien « spécifiques » de tissus don-

nés... ce qui revient à dire qu'ils y sont en concentration nettement plus élevée qu'ailleurs [15-17].

La combinaison variable d'un petit nombre de facteurs régulateurs peut, elle aussi, engendrer une considérable diversité... celle des mots innombrables composés à l'aide des lettres d'un alphabet [11, 16, 17]. De fait, « la dissection fonctionnelle » des *enhancers*, c'est-à-dire l'étude des effets de mutations ou de microdélétions en différents sites, montre qu'ils sont constitués de cassettes, chacune interagissant probablement avec un facteur protéique et ayant individuellement une spécificité qui peut être différente de celle du *enhancer* total [18-20]. Le *enhancer* du virus SV40 contrôlant l'expression du gène T est ainsi constitué : de l'octamère liant le facteur IgNF-A et conférant à des gènes en amont desquels il est inséré en plusieurs copies, une spécificité lymphoïde d'expression ; de deux autres éléments qui, analysés de la même manière que l'octamère, confèrent soit, pour l'un, une spécificité pour des cellules de rein de singe, soit, pour l'autre, une expression ubiquitaire [20].

Un même facteur peut être variablement modifié (par phosphorylation, par exemple) et interagir de façon différente selon le type de modification subie. Ainsi, avons-nous vu que des facteurs inactifs pouvaient être activés. Dans le cas d'AP1, s'il s'avérait que cette protéine était également impliquée dans la régulation par l'AMP cyclique, on pourrait imaginer qu'elle a des effets différents et (ou) interagit avec des motifs différents selon qu'elle est phosphorylée à un site sous l'action de la protéine kinase C ou à un autre site sous l'action d'une protéine kinase stimulée par l'AMP cyclique.

Enfin, dernier élément engendrant de la diversité, un même facteur peut reconnaître des motifs d'ADN voisins, mais non identiques, et contracter des interactions légèrement différentes les unes des autres selon le motif de

\* Il existerait, en fait, deux espèces d'IgNF-A, l'une ubiquitaire et l'autre caractéristique des cellules B.

\*\* Les esters de phorbol sont des « promoteurs » de la cancérogenèse. Ils agissent en activant la protéine kinase C.



liaison. Dans ce cas, le facteur « actif en *trans* » est le même pour plusieurs gènes et la diversité de réponse fonctionnelle est due aux différences de séquence des motifs reconnus ; il s'agit donc ici d'un contrôle en *cis* de la spécificité d'expression. La fixation variable de SRF en amont du gène *c-fos* ou du gène actine musculaire pourrait peut-être constituer un exemple de ce mécanisme (Françoise Phan Dunh Tuy et Adrian Minty, communication personnelle) : dans le premier cas, cette fixation conférerait la réponse à l'induction de la prolifération alors que,

dans le second cas, elle interviendrait dans l'expression au cours de la différenciation musculaire.

**Un schéma unifié.** La figure 1 présente une vue unifiée de la réponse transcriptionnelle à l'induction de la différenciation... ou de la prolifération cellulaire. Un signal perçu par la membrane (étape 1) entraîne, via la synthèse de seconds messagers, la modification post-traductionnelle d'un facteur transcriptionnel pré-existant (étape 2). Cette activation ( $F^A \rightarrow F_a^A$  dans la figure) modifie la transcription de plusieurs gènes,

dont certains codent eux-mêmes pour des facteurs transcriptionnels A, B et C. De ce fait, la proportion relative de ces facteurs se modifie (étape 3), ce qui aboutit (étape 4) à la modification des combinaisons de facteurs liés en amont des gènes qui sont, de ce fait, activés ou inactivés. Ce schéma, sûrement beaucoup trop simple (et rappelons ici que certains facteurs « spécifiques de tissus » ont été décrits), suppose l'existence de phénomènes de régulation en cascade dont un exemple pourrait être la voie à laquelle appartient *c-fos* : un phénomène d'activation post-traductionnelle de SRF explique peut-être l'augmentation de la synthèse de *c-fos* qui est, elle-même, une protéine se liant à l'ADN et pouvant contrôler l'expression d'une deuxième classe de gènes [21]. Ce schéma a, en revanche, l'intérêt de concilier l'essentiel des données récentes acquises en ce domaine : les premiers événements de l'induction de la différenciation sont membranaires ; la réponse transcriptionnelle précoce de certains gènes ne nécessite pas la synthèse de nouvelles protéines ; les mêmes facteurs variablement combinés sont souvent trouvés associés au contrôle de l'expression de gènes impliqués dans des types très différents de différenciation ; enfin, certaines des protéines codées par les gènes précocement activés au cours de ces phénomènes sont elles-mêmes des facteurs transcriptionnels jouant un rôle probable dans l'activité d'autres gènes [22].

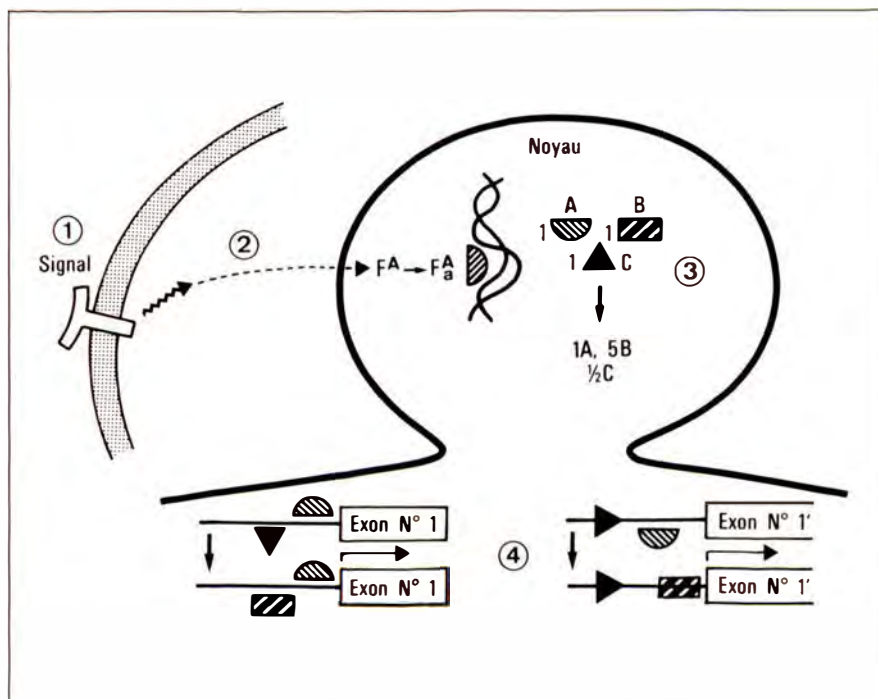


Figure 1. **Réponses transcriptionnelles en cascade à un inducteur de différenciation ou de prolifération.** Pour l'explication détaillée du schéma, voir texte, dernier paragraphe. A, B, C : facteurs transcriptionnels ;  $F_a^A$  : facteur A actif, se fixant sur la double hélice d'ADN et entraînant un effet transcriptionnel ; **étape 3** : changement quantitatif des facteurs A, B et C qui, initialement en proportion équimolaire, se retrouvent dans la proportion 1A, 5B et 1/2C, après la réponse transcriptionnelle primaire due à l'activation  $F^A \rightarrow F_a^A$  ; **l'étape 4** représente la réponse transcriptionnelle secondaire à la modification quantitative des facteurs A, B et C ; à gauche, le site de reconnaissance de B chevauche celui de C, si bien que l'augmentation du rapport B/C aboutit à un déplacement de C. A droite, B déplace A et est combiné avec les autres facteurs de façon différente de celle représentée à gauche. De plus, le facteur C peut avoir des interactions variées avec des séquences voisines mais différentes. La combinaison de tous ces phénomènes crée une grande diversité de réponses possibles à la modification d'un petit nombre de facteurs.

Axel Kahn

## RÉFÉRENCES

1. Hafen E, Basler K, Edstroem JA, Rubin GM. Sevenless, a cell-specific homeotic gene of drosophila encodes a putative transmembrane receptor with a tyrosine kinase domain. *Nature* 1987 ; 236 : 55-63.
2. Munnich A, Vaulont S, Marie J. De nouvelles fonctions pour l'AMP cyclique. *médecine/sciences* 1985 ; 1 : 192-7.
3. Zhand R-g, Joachimiak A, Lawson CL, Schevitz RN, Otwinowski Z, Sigler PB. The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature* 1987 ; 327 : 591-6.

4. Sen R, Baltimore D. Inducibility of  $\kappa$  immunoglobulin enhancer-binding protein NF- $\kappa$ B by a post-translational mechanism. *Cell* 1986 ; 47 : 921-8.

5. Treisman R. Identification of a protein binding site that mediates transcriptional response of the *c-fos* gene to serum factors. *Cell* 1986 ; 46 : 567-74.

6. Prywes R, Røder RG. Inducible binding of a factor to the *c-fos* enhancer. *Cell* 1986 ; 47 : 777-84.

7. Angel P, Imagawa M, Chin R, *et al.* Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated transacting factor. *Cell* 1987 ; 49 : 729-39.

8. Lee W, Mitchell P, Tijian R. Purified transcription factor AP 1 interacts with TPA-inducible enhancer elements. *Cell* 1987 ; 49 : 741-52.

9. Seguin C, Hamer DH. Regulation in vitro of metallothionein gene binding factors. *Science* 1987 ; 235 : 1383-7.

10. Kingston RE, Schuetz TJ, Larin Z. Heat inducible human factor that binds to a human hsp promoter. *Mol Cell Biol* 1987 ; 7 : 1530-4.

11. Sen R, Baltimore D. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 1986 ; 46 : 705-16.

12. Janson I, Bark C, Petterson U. Identification of proteins interacting with the enhancer of human U 2 small nuclear genes. *Nucleic Acids Res* 1987 ; 15 : 4997-5016.

13. Montminy MR, Bilezikjian LM. Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. *Nature* 1987 ; 328 : 175-8.

14. Comb M, Birnberg NC, Seasholtz A, Herbert E, Goodman HM. A cyclic AMP and phorbol ester-inducible DNA element. *Nature* 1986 ; 323 : 353-6.

15. Bodner M, Karin M. A pituitary-specific transacting factor can stimulate transcription from the growth hormone promoter in extracts of non-expressing cells. *Cell* 1987 ; 50 : 267-75.

16. Cereghini S, Raymondjean M, Carranca AG, Herbomel P, Yaniv M. Factors involved in control of tissue-specific expression of albumin gene. *Cell* 1987 ; 50 : 627-38.

17. Lenardo M, Pierce JN, Baltimore D. Protein-binding sites in Ig enhancers determine transcriptional activity and inducibility. *Science* 1987 ; 236 : 1573-7.

18. Ondek B, Shepard A, Herr W. Discrete elements within the SV-40 enhancer region display different cell-specific enhancer activities. *EMBO J* 1987 ; 6 : 1017-25.

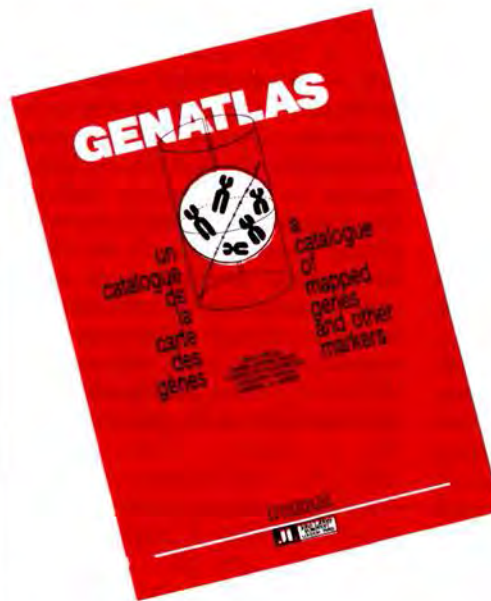
19. Rochford R, Campbell BA, Villarreal LP. A pancreas specificity results from the combination of polyomavirus and Moloney murine leukemia virus enhancer. *Proc Natl Acad Sci USA* 1987 ; 84 : 449-53.

20. Schirm S, Jiricny J, Schaffner W. The SV-40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Gene and Development* 1987 ; 1 : 65-74.

21. Distel RJ, Ro HS, Rosen BS, Groves DL, Spiegelman BM. Nucleoprotein complexes that regulate gene expression in adipocyte differentiation : direct participation of *c-fos*. *Cell* 1987 ; 49 : 835-44.

22. Robertson M. A genetic switch in drosophila morphogenesis. *Nature* 1987 ; 327 : 556-7.

# GENATLAS



J. FRÉZAL, M.S. BAULE,  
T. DE FOUGEROLLE,  
J. KAPLAN, M. LE MERRER

Co-édition INSERM/  
John Libbey Eurotext

1989, broché, 586 pages  
310 FF

- **Un catalogue des gènes et des marqueurs identifiés et localisés sur les chromosomes de l'homme jusqu'en avril 1989**, avec, en complément, une liste ordonnée de sondes anonymes localisées sur des segments définis de chromosomes.
- **Une bibliographie de 5 000 références** classées par chromosome et par marqueur.
- **Un classement par nom de marqueurs**, symbole, chromosome, catégorie.
- **Un instrument au service des médecins**, avec le répertoire des maladies dont le gène est localisé, la mention du produit de ces gènes et la liste des marqueurs qui leurs sont liés.

## BON DE COMMANDE

NOM ..... Prénom .....

Adresse .....

Désire recevoir **GENATLAS** au prix de 310 FF + 30 FF de frais de port, soit 340 FF.

Ci-joint mon règlement à l'ordre de **John Libbey Eurotext**.

**JOHN LIBBEY EUROTEXT — EDITIONS MEDECINE-SCIENCES**

6, rue Blanche, 92120 MONTROUGE Tél. (1) 47.35.85.52 - Fax (1) 46.57.10.09

## **D**omaine POU-homéo et facteurs de transcription

Les gènes homéotiques ont d'abord été décrits chez la drosophile ; leur altération modifie le schéma de développement de la mouche, avec remplacement de certaines parties du corps par d'autres (amenant par exemple au développement de pattes à la place d'antennes, ou d'un segment thoracique à la place d'un segment abdominal). Tous les gènes homéotiques possèdent une « boîte homéo » qui est un domaine protéique très conservé au cours de l'évolution et présentant des similitudes avec le domaine de fixation à l'ADN du répresseur du phage lambda, c'est-à-dire semblant correspondre au motif « hélice-tour-hélice » de ces facteurs de transcription (*m/s* n° 7, vol. 3, p. 428). D'autres gènes de développement de la drosophile ont ensuite été détectés qui, quoique ne correspondant pas à la définition d'un gène homéotique, contenaient néanmoins des « boîtes homéo » pouvant être classées en plusieurs types. La conservation au cours de l'évolution des espèces de ces différents types de « boîtes homéo » a permis, grâce à l'hybridation à l'aide de sondes d'ADN spécifiques de ce domaine, d'isoler de nombreux « homéogènes » chez les mammifères aussi bien que dans des organismes eucaryotes plus simples tel le ver *Caenorhabditis elegans*. Dans d'autres cas, c'est l'analyse de la séquence nucléotidique d'un gène cloné par d'autres moyens qui permet de retrouver des éléments ressemblant plus ou moins vaguement à des « boîtes homéo ». C'est ainsi qu'a été très récemment isolé et caractérisé le gène *unc-86* de *C. elegans* [1]. Les mutants *unc-86* ont un développement anormal de leurs précurseurs des cellules neuronales. Normalement, une cellule mère de cette lignée se différencie en deux cellules filles aux destins différents de celui de la cellule mère. Chez les mutants *unc-86*, la cellule mère donne deux cellules filles dont une lui ressemble complètement ; cette dernière aura

elle-même une de ses cellules filles qui lui sera identique. Le gène *unc-86* a été cloné par la technique de « marche sur le chromosome » qui consiste à partir d'un locus connu pour atteindre un gène inconnu situé à proximité (*m/s* n° 6, vol. 3, p. 363). Le gène *unc-86* peut coder pour une protéine de 467 acides aminés comportant, du côté carboxy-terminal, un domaine homéo de 60 bases assez différent des boîtes homéo des gènes de drosophile (de 27 à 38 % d'analogie).

Tout à fait indépendamment de ces travaux, les ADN complémentaires codant pour trois facteurs transcriptionnels de mammifères ont été isolés. GHF-1/Pit-1 (*growth hormone factor 1/pituitary specific factor 1*) a été décrit comme un facteur ayant la

capacité de se lier en des sites multiples des promoteurs des gènes de l'hormone de croissance [2] et de la prolactine [3]\*. En fait, GHF-1 et Pit-1 ont été étudiés séparément, le premier par l'équipe de M. Karin (Université de Californie, San Diego, USA) travaillant sur le gène de l'hormone de croissance et le second par l'équipe de M.G. Rosenfeld (Howard Hughes Institute, également à San Diego) travaillant sur le gène de la prolactine. Les deux équipes utilisant des techniques différentes de criblage de banque d'ADNc, semblent bien avoir isolé des clones codant pour la même protéine de 33 kDa (kilodaltons), activateur transcriptionnel des gènes possédant dans leur promoteur des sites de fixation pour cette protéine. GHF-1/Pit-1 et son

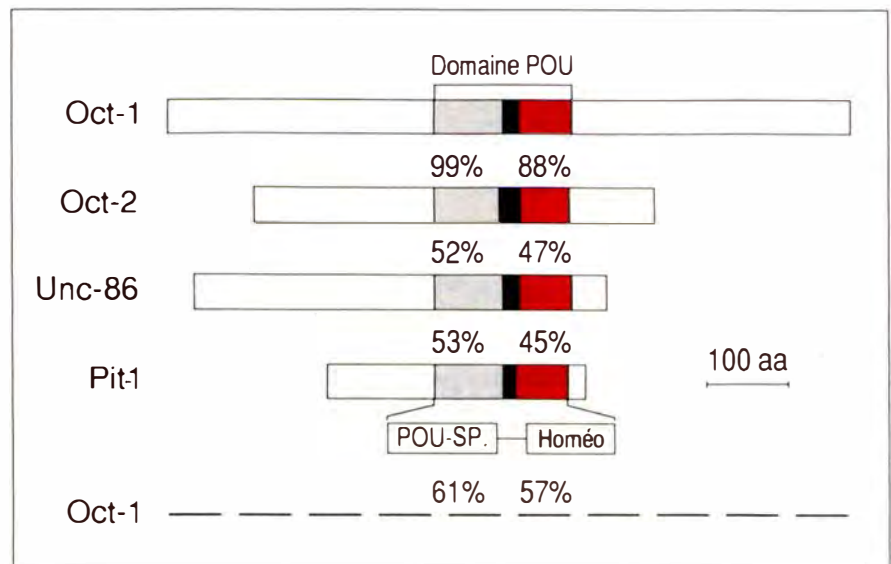


Figure 1. **Le domaine POU dans trois facteurs transcriptionnels et le produit d'un gène de développement de *C. elegans*.** En rouge, le domaine « homéo » ; en gris, le domaine « POU-spécifique ». Les pourcentages de similitudes entre les sous-domaines des protéines différentes sont indiqués entre les schémas correspondant à chacune d'entre elles (par exemple, le domaine POU-spécifique d'Oct-1 est à 99 % similaire à celui d'Oct-2 et à 61 % similaire à celui de Pit-1). (Tiré de [10].)

\* En fait, le laboratoire de Michael Karin a montré que GHF-1 et le facteur stimulant le gène de prolactine pourraient être différents. Dans un très récent article (*Science*, 10 février 1989, vol. 243, p. 814), J.-L. Castrillo et al. indiquent que M.G. Rosenfeld et ses collaborateurs [3] auraient pu cloner l'ADNc correspondant au facteur du gène de l'hormone de croissance et non à celui du gène de la prolactine... ce qui expliquerait que Pit-1 = GHF-1 !

messager ne sont trouvés que dans les cellules somatotropes et lactotropes de l'hypophyse. Des cellules indifférenciées préalablement transfectées avec l'ADNc codant pour GHF-1/Pit-1 sont capables d'utiliser les promoteurs des gènes de prolactine et d'hormone de croissance, ce qui démontre le caractère essentiel de cette protéine dans l'activation spécifique de ces gènes. Cependant, le promoteur du gène de l'hormone de croissance est inactif dans des cellules lactotropes alors même qu'elles synthétisent GHF-1/Pit-1. Ce dernier résultat suggère que les cellules lactotropes n'expriment que le gène de la prolactine, et pas celui de l'hormone de croissance, parce que ce dernier est inhibé par un facteur particulier. L'inverse explique probablement que les cellules somatotropes, quoiqu'elles contiennent le facteur GHF-1/Pit-1 aussi bien que les cellules lactotropes, expriment le gène de l'hormone de croissance et non celui de la prolactine.

Oct-1 (encore appelé NFA-1, NF-3, OTF-1 et OBP 100 1) et Oct-2 (connu sous le nom de NFA-2 et OTF-2) sont des protéines qui reconnaissent un élément d'ADN appelé octamère et présent au niveau de très nombreux promoteurs (notamment ceux des gènes d'immunoglobuline) et *enhancers* (notamment celui du gène des chaînes lourdes d'immunoglobuline).

Oct-1 est de distribution ubiquitaire alors que Oct-2 (et son messager) ne sont détectés que dans les cellules lymphocytaires B. Des répétitions de l'élément « octamère » se comportent comme des « *enhancers* » spécifiquement actifs dans les lymphocytes B. On considère ainsi que Oct-2 est un activateur transcriptionnel très spécifique des cellules B alors que Oct-1 serait un facteur transcriptionnel ubiquitaire actif en coopération avec d'autres protéines se fixant au niveau de certains promoteurs.

Les ADNc codant pour Oct-1 [4] et Oct-2 [5-9] ont été isolés en utilisant des oligonucléotides de synthèse qui soit ont été déduits de la séquence protéique, soit contenaient plusieurs versions de l'élément « octamère ». Dans le premier cas, les sondes oligonucléotidiques hybridaient avec l'ADNc cloné alors que dans le second elles se fixaient aux protéines

synthétisées par les clones bactériens contenant l'ADNc dans un « vecteur d'expression ». GHF-1/Pit-1, Oct-1 et Oct-2, comme la protéine Unc-86, ont un domaine homéo d'une soixantaine de bases qui est bien plus similaire entre ces différents gènes qu'il ne l'est à ceux des gènes de drosophile (les similitudes vont de 45 % à 88 % entre les domaines « homéo » des protéines étudiées ici). A une distance variable du domaine homéo, on trouve un autre bloc de 75 acides aminés, de 50 % à 99 % identique dans les quatre protéines (*figure 1*). L'ensemble de cette région, d'environ 150 bases, a été appelé POU (Pit-1-Oct-Unc) [10]. De nombreux arguments indiquent que le domaine POU correspond à la région de ces protéines spécialisées dans la fixation aux éléments cibles d'ADN : les analogies entre les domaines « homéo » et la région du répresseur du phage  $\lambda$  se liant à l'ADN ; la conservation presque totale de la séquence des deux domaines « homéo » et « POU spécifique » des protéines Oct-1 et Oct-2 qui reconnaissent le même élément d'ADN ; plus directement, enfin, la démonstration que la modification des deux domaines de POU dans Oct-1 altère l'aptitude de cette protéine à se lier à l'élément d'ADN « octamère » [11]. Le modèle qui émerge de ces résultats est celui d'une région protéique contactant spécifiquement l'ADN par un domaine basique reproduisant un motif « hélice-tour-hélice » (domaine « homéo ») et par un autre domaine protéique plus acide (« POU spécifique ») relié au précédent par une charnière variable et faisant contact avec l'ADN d'une manière indéterminée.

Plusieurs questions... et plusieurs voies de recherche émergent de cette masse de résultats nouveaux. Tout d'abord, comment peut-on concevoir qu'un même type de domaine protéique de liaison avec l'ADN intervienne dans des processus de régulation aussi différents que le développement embryonnaire, l'activation de gènes particuliers au cours de la différenciation cellulaire et la stimulation transcriptionnelle de gènes d'expression ubiquitaire ? Il semble que les protéines Oct soient capables de reconnaître une très

grande diversité de séquences d'ADN ne ressemblant parfois que vaguement au motif « octamère » habituel [12]. Symétriquement, ce motif octamère peut être reconnu par des protéines multiples... notamment des produits de gènes homéotiques de drosophile [13]. La rationalité de tout ceci est probablement que l'information donnée à un promoteur doit toujours être le résultat d'une « combinatoire » engendrée par les interactions variables entre des protéines actives sur la transcription. Un même facteur complexé avec des protéines différentes va transmettre au promoteur une commande différente ; il pourra aussi, selon la nature du complexe auquel il participe, reconnaître des cibles d'ADN variant légèrement. Évidemment, une famille de protéines reconnaissant le même type de séquence d'ADN mais possédant des régions d'interactions protéine-protéine et de modulation transcriptionnelle variées vont pouvoir intervenir dans des processus extrêmement différents.

Un très bon exemple de cela est constitué par la « reprogrammation » de Oct-1 lorsqu'il fixe la protéine Vmw65 codée par le virus *Herpes simplex*. Vmw65 ne se fixe pas à l'ADN par elle-même. Elle active cependant en *trans* les promoteurs des gènes d'expression précoce du virus herpes et, plus généralement, des promoteurs contenant un motif « octamère ». Cette activation se fait grâce à la formation d'un complexe Oct-1/Vmw65 [14].

Un tel modèle du contrôle de l'expression des gènes par un ensemble de protéines interactives dont la nature et l'abondance changent progressivement au cours du développement rend probablement compte de la flexibilité nécessaire au bon déroulement du programme d'expression génétique. On conçoit comment ce mécanisme peut expliquer à la fois la progressivité de certains changements (la nature du contrôle assuré par une même protéine à deux stades successifs de la différenciation est modulée par son interaction avec d'autres facteurs), et la soudaineté d'autres modifications qui sont du type « commutation » avec extinction d'un gène et activation d'un autre. Dans ce cas, l'apparition ou la disparition d'un « différenciateur »

tel HNF-1 dans le foie (*m/s n° 2, vol. 4, p. 119*), et ici, peut-être GHF-1/Pit-1 et Oct-2 est probablement en cause.

Axel Kahn

RÉFÉRENCES

1. Finney M, Ruvkun G, Horvitz HR. The *C. elegans* cells lineage and differentiation gene *unc-86* encodes a protein with a homeo-domain and extended similarities to transcription factors. *Cell* 1988 ; 55 : 757-69.
2. Bodner M, Castrillo JL, Theill LE, Deerinck T, Ellisman M, Karin M. The pituitary specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* 1988 ; 55 : 505-18.
3. Ingraham HA, Chen R, Mangalam HJ, et al. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* 1988 ; 55 : 519-29.
4. Sturm RA, Das G, Herr W. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo subdomain. *Genes Dev* 1988 ; 2 : 1582-99.
5. Clerc RG, Carcoran L, Le Bowitz JH, Baltimore D, Sharp PA. The B-cell-specific Oct-2 protein contains POU-box and homeo-box type domains. *Genes Dev*. 1988 ; 2 : 1570-81.
6. Ko HS, Fast P, McBride W, Staudt LM. A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. *Cell* 1988 ; 55 : 135-44.
7. Staudt LM, Clerc RG, Singh H, Le Bowitz JH, Sharp PA, Baltimore D. Cloning a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* 1988 ; 241 : 577-80.
8. Scheidereit C, Cromlish JA, Gerster T, et al. A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein. *Nature* 1988 ; 336 : 551-7.
9. Mülllet MM, Ruppert S, Schaffner W, Matthias P. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B-cells. *Nature* 1988 ; 336 : 544-51.
10. Herr W, Sturm RA, Clerc RG, et al. The POU domain : a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2* and *C. elegans unc-86* gene products. *Genes Dev* 1988 ; 2 : 1513-6.
11. Sturm RA, Herr W. The POU domaine is a bipartite DNA-binding structure. *Nature* 1988 ; 336 : 601-4.
12. Boumruher T, Sturm R, Herr W. OBP 100 binds remarkably degenerate octamer motif through specific interactions with flanking sequences. *Genes Dev*. 1988 ; 2 : 1400-13.
13. Thali M, Müller MM, De Lorenzi M, Matthias P, Bienz M. Drosophila homeotic genes encode transcriptional activators similar to mammalian OTF-2. *Nature* 1988 ; 336 : 598-601.
14. O'Hare P, Goding CR. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virion transactivation. *Cell* 1988 ; 52 : 435-45.

# Régulation de l'expression des gènes et différenciation

## Mécanismes pré et post-transcriptionnels

### 1. Remaniement des gènes.

Dans un petit nombre de cas, la transcription efficace d'un gène est précédée par son réarrangement. Chez les eucaryotes supérieurs, l'exemple le mieux connu est celui des immunoglobulines, où l'une des nombreuses copies codant pour les parties variables des anticorps se trouve sélectivement rapprochée des régions géniques codant pour les parties constantes, via un probable mécanisme délétionnel. Une séquence stimulatrice située dans l'intron, immédiatement en amont des gènes codant pour les parties constantes, se trouve alors suffisamment proche du promoteur du gène variable pour l'activer (*figure 1*).

### 2. Epissage alternatif.

Un même gène transcrit sous la forme d'un ARN précurseur unique peut, par épissage alternatif, donner deux ARN et deux protéines très différentes selon le tissu dans lequel il est transcrit. L'expression du gène de la calcitonine dans la thyroïde et le cerveau est le meilleur exemple de ce mécanisme (*figure 2*). Nombreux sont à présent les exemples similaires, où la diversité des ARN et parfois des protéines est due non à la transcription de gènes différents mais à la maturation (différente selon les tissus) du transcrit primitif d'un gène unique. C'est donc là un mécanisme important de la modulation de l'expression des gènes.

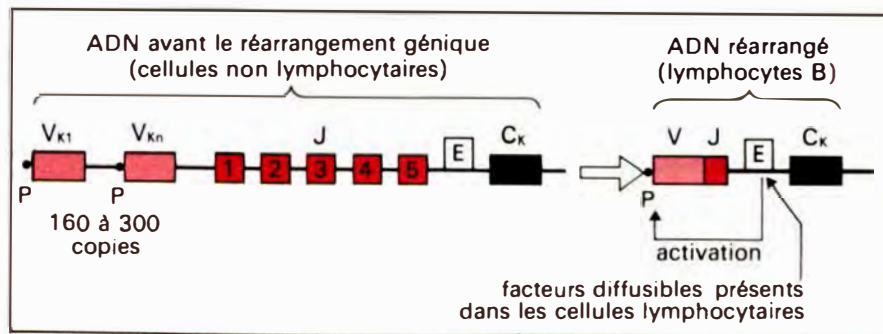


Figure 1. Réarrangement des gènes codant pour la chaîne légère K des immunoglobulines. Dans l'ADN non réarrangé, on trouve 160 à 300 copies de gènes  $V_{\kappa}$  codant pour les régions variables, 5 éléments « J » (dont 4 sont fonctionnels) et une copie du gène  $C_{\kappa}$  codant pour la partie constante. En amont de ce gène  $C_{\kappa}$  se trouve une séquence stimulatrice E (enhancer) dont l'action sur le promoteur P nécessite d'une part une relative proximité, d'autre part la présence de facteurs diffusibles spécifiques de la différenciation B lymphocytaire. Au cours de cette différenciation, un gène  $V_{\kappa}$  est réarrangé à proximité d'un élément J et ne reste séparé de  $C_{\kappa}$  que par l'intron contenant l'enhancer qui, se trouvant désormais relativement proche du promoteur, l'active.

### 3. Stabilité des ARN.

La dégradation des ARN messagers est un phénomène encore mal connu qui joue probablement un rôle significatif dans l'expression du génome. Certaines hormones qui augmentent la transcription d'un gène stabilisent également son message. La différence du niveau de synthèse de nombreuses protéines présentes dans tous les tissus, mais à des concentrations différentes (*house keeping proteins* ou protéines « de ménage ») pourrait ainsi être prioritairement non transcriptionnelle et

faire intervenir la différence de stabilité des messagers dans divers environnements cellulaires.

### 4. Régulation de la traduction des ARN messagers.

La part réellement jouée par la régulation traductionnelle reste mal connue chez les eucaryotes, à l'exception de certaines situations particulières, telle celle des œufs non fécondés du *Xenopus* (un batracien) dans lesquels sont accumulés des messagers qui ne seront traduits qu'après la fécondation.

Il ne faut pas conclure de la brève énumération ci-dessus que les niveaux de régulation sont exclusifs les uns des autres; la plupart des phénomènes qui modifient l'expression d'un gène intervenant en fait de façon coordonnée à différents niveaux. Ainsi la synthèse d'un anticorps par un lymphocyte B stimulé implique-t-elle un réarrangement génique, une activation de la transcription et un épissage alternatif aboutissant à la synthèse des formes membranaires et sécrétoires des immunoglobulines. A.K.

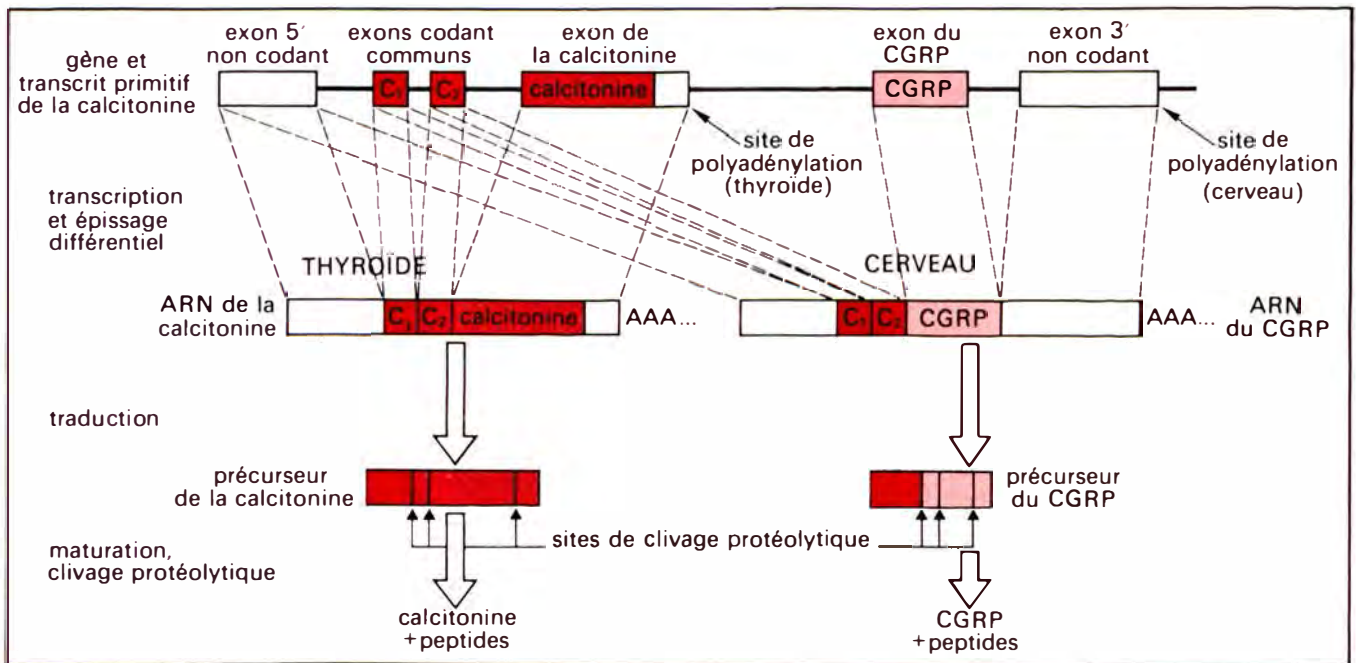


Figure 2. Epissage alternatif du transcrit primitif du gène de la calcitonine dans la thyroïde et le cerveau.

Le gène est transcrit en totalité dans les deux tissus.

Dans la thyroïde, le site de polyadénylation utilisé est situé en 3' de l'exon de la calcitonine, toute la partie aval du transcrit primitif étant probablement excisée par coupure endonucléolytique.

L'épissage donne ensuite le message de la calcitonine.

Dans le cerveau, le transcrit primitif est polyadénylé en 3' de l'exon non codant du message du CGRP (calcitonin gene related peptide), l'épissage excisant l'exon spécifique de la calcitonine.

Les ARN messagers de la calcitonine et du CGRP sont traduits en polypeptides précurseurs dont le clivage protéolytique libère la calcitonine, le CGRP ainsi que d'autres peptides.

## Un nouveau mécanisme d'action pour le gène rev du virus HIV

Il existe deux protéines, codées par le génome de HIV (virus du SIDA), qui jouent un rôle essentiel dans le contrôle de l'expression de ce génome : les protéines Tat et Rev. L'une et l'autre sont produites à partir d'une espèce d'ARN messager engendrée par l'excision d'« introns » correspondant aux gènes *gag* et *pol* et à une grande partie du gène *env*. L'ARN n'ayant subi aucune excision-épissage sert de messager pour les protéines Gag et Pol alors que celui dont seul le premier « intron » a été excisé sert de messager pour la protéine d'enveloppe (Env). Alors que Tat est un activateur transcriptionnel agissant sur la séquence TAR du LTR viral

(encore que ce point ait été très discuté, certains auteurs indiquant que l'interaction TAR-Tat augmentait la traduction du messager, *m/s* n° 5, vol. 2, p. 285), Rev semble inhiber l'excision-épissage du transcrite viral, ce qui aboutit à l'accumulation des protéines structurales que sont Gag, Pol et Env. Cette étape, tardive au cours du cycle viral, est essentielle à l'assemblage des particules virales. Ainsi, l'expression du gène *tat* augmente la synthèse de toutes les protéines virales, y compris elle-même, alors que celle du gène *rev* aboutit à la diminution de la synthèse des protéines régulatrices Tat et Rev et à l'augmentation de celle des protéines structurales.

L'interprétation la plus généralement avancée pour expliquer l'action de Rev était qu'elle inhibait la réaction d'excision-épissage du transcrite viral primitif [1]. Deux laboratoires américains viennent d'infirmier cette donnée [2, 3]. Le mécanisme qui émerge de ces travaux est le suivant : le transcrite n'ayant pas subi l'excision des introns semble stabilisé par la protéine Rev [3], qui augmente également le transport de cette espèce vers le cytoplasme [2, 3]. Le transcrite complet du virus HIV peut subir trois destins : être dégradé, être transporté dans le cytoplasme... ou être intégré dans la structure « splicéosomale » qui constitue la machinerie d'excision des introns et d'épissage des exons (*figure 1*). La protéine Rev, qui semble reconnaître, directement ou indirectement, une séquence d'ADN située dans la région *env* (élément RRE, *Rev response element*), diminue la dégradation et augmente le transport du transcrite primitif qui n'est ainsi plus disponible pour la machinerie d'excision-épissage, et donc pour la traduction en protéines régulatrices. Reste maintenant à élucider le mécanisme exact de l'action de Rev, une protéine nucléaire (et peut-être même nucléolaire) sur le transport de ce messager viral vers le cytoplasme.

A. K.

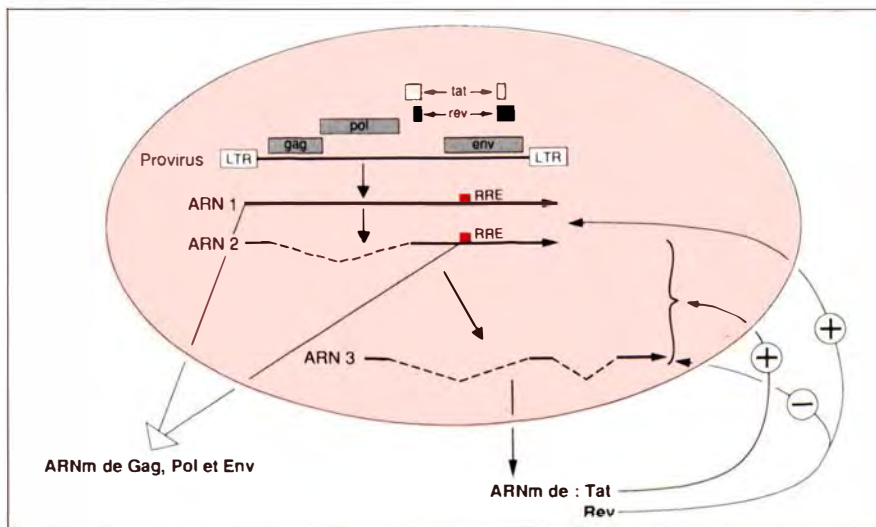


Figure 1. **Autorégulation de l'expression du génome de HIV.** Le provirus intégré avec ses deux LTR (long terminal repeats) est transcrite en un ARN qui, sans excision-épissage, est le messager des protéines Gag et Pol (ARN 1). L'excision d'un premier « intron » crée l'espèce messagère de la protéine Env (ARN 2). L'excision de deux introns aboutit à l'ARNm des protéines Tat et Rev (ARN3). Tat augmente la synthèse de toutes les espèces de messager. Rev agit sur la séquence RRE (Rev response element) des ARN 1 et 2 qu'il stabilise et dont il augmente le transport vers le cytoplasme où ils seront traduits en protéines virales de structure.

1. Green MR, Zapp ML. Revving up gene expression. *Nature* 1989 ; 338 : 200-1.
2. Malim MH, Hauber J, Le Sy, Maizel JV, Cullen BR. The HIV-1 *rev* trans-activator acts through a structural target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 1989 ; 338 : 254-7.
3. Felber BK, Hadzopoulou-Cladaras M, Cladaras C, Copeland T, Pavlakis GN. Rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci USA* 1989 ; 86 : 1495-9.

## **R**égulation par le fer de la biosynthèse de la ferritine et du récepteur de la transferrine

C'est la disponibilité du fer qui régit le rythme de synthèse d'au moins deux protéines essentielles pour la biodisponibilité de ce métal, la ferritine et le récepteur de la transferrine (TfR). La régulation de ces deux protéines s'exerce par deux mécanismes différents et en sens opposés. Le récepteur de la transferrine, nécessaire à la capture du fer, voit sa production diminuer quand le métal est abondant, augmenter en son absence, et ces variations sont consécutives à des changements de quantité de son ARN messenger [1]. A l'inverse, la ferritine, qui séquestre le fer dans le cytoplasme, est fabriquée d'autant plus vite que le métal est plus abondant ; ces changements se produisent à concentration d'ARNm constante, par redistribution du message entre cytoplasme et polysomes au bénéfice de ces derniers [2]. Cela s'applique aussi bien à la chaîne lourde H (21 000 Da) qu'à la chaîne légère L (19 000 Da) de la ferritine. Des expériences récentes ont montré qu'un même type de séquence contrôle l'action du fer sur chacune des deux protéines, malgré les différences considérables de cette action dans les deux cas. La localisation de ces séquences est toutefois très différente.

- Dans le cas de la ferritine, la régulation est traductionnelle, entièrement sous le contrôle d'une séquence du messenger située dans la zone 5' non traduite [3]. Cette séquence a été identifiée [4] en construisant une chimère dans laquelle la partie 5' du messenger de la chaîne L de ferritine de rat est fusionnée au gène marqueur bactérien CAT. Quand on introduit la chimère dans une lignée d'hépatome de rat, sa production est activée par le fer. L'activation est supprimée si on délète les 67 premiers nucléotides de la partie

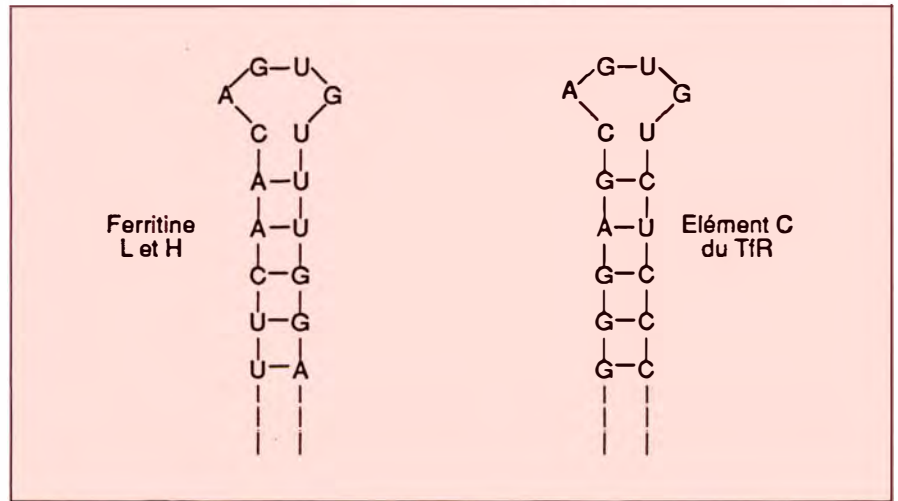


Figure 1. **Éléments répondant au fer (IRE).** Le schéma comporte une boucle qui porte les cinq mêmes nucléotides, et des tiges dont la séquence nucléotidique est variable. Deux exemples sont présentés : celui de l'élément C du TfR, et celui de la ferritine, pour lequel l'identité est parfaite dans cette zone entre les sous-unités H et L. Dans l'expérience de Casey et al. le transcrit est celui de l'hormone de croissance.

5' non traduite. L'analyse par ordinateur de cette séquence révèle une structure dite tige-boucle (*stem-loop*) (voir figure 1).

- La modulation par le fer du niveau du messenger de TfR dépend des deux régions du gène situées dans les parties 5' et 3'. C'est ici la partie 3' non traduite qui joue le rôle essentiel, et la délétion de ces séquences supprime presque complètement la régulation par le fer [5]. C'est encore la méthode d'analyse délétionnelle, utilisant des cellules murines ayant reçu une construction qui contenait la région 3' non traduite de TfR, qui a permis d'identifier une séquence de 678 nucléotides d'ADN complémentaire, indispensable à l'action du fer. Dans cette région, cinq motifs tige-boucle peuvent être identifiés. Le mécanisme d'action de cette séquence 3' n'est probablement pas transcriptionnel [6]. Il semble plutôt être une modification de la stabilité du messenger.

Quoiqu'il en soit, ces motifs isolés et transférés sur la partie 5' d'un transcrit, suffisent à lui conférer un pouvoir de régulation traductionnelle de type ferritine. Un même élément possèdera donc une activité de contrôle de la traduction ou de la stabilité de l'ARN messenger selon qu'il sera situé en 5' ou en 3' de la molécule. La figure 1 montre le schéma de ces éléments que Casey et al. [5] ont appelés IRE (*iron responsive elements*).

Comme on pouvait le supposer, ces IRE sont capables de lier spécifiquement une ou des protéines cytosoliques [7] ; les conséquences de la formation du complexe ARN-protéine seraient différentes selon sa localisation. En 3', un tel processus pourrait favoriser une coupure endonucléolytique, séparant donc la partie codante du messenger de l'extension d'acide polyadénylique. L'apparition en 3' du messenger résiduel d'une extrémité libre, sensible aux exonucléases,



## La tubuline $\beta$ : contrôle de la stabilité d'un messager par son produit de traduction

expliquerait l'instabilité ainsi induite. En 5', le complexe nucléoprotéique incluant l'élément IRE faciliterait la traduction, peut-être en modifiant la structure secondaire de la région 5' non traduite du messager. On sait que, chez les eucaryotes, la petite sous-unité ribosomale reconnaît tout d'abord des protéines fixées à la coiffe (*cap*) en 5' du messager, puis migre le long de l'ARN jusqu'au premier codon d'initiation AUG correct, au niveau duquel se constitue le complexe d'initiation de la traduction (incluant notamment la grande sous-unité ribosomale et le méthionyl-ARNt initiateur). Toute modification de la structure de l'ARN entre la coiffe 5' et l'AUG initiateur peut donc grandement moduler la traductibilité. Si ces hypothèses étaient vérifiées, il resterait alors à élucider les relations entre le fer et l'activité ou l'abondance de cette protéine censée se fixer sur les éléments IRE.

Dans leur récent article, Rouault *et al.* indiquent qu'il existe une corrélation inverse entre la protéine se liant à l'IRE et la concentration en fer cellulaire [7].

J.-C. D.  
A. K.

1. Rao KK, Shapiro D, Mattia E, Bridges K, Klausner R. Effects of alterations in cellular iron on biosynthesis of the transferrin receptor in K562 cells. *Mol Cell Biol* 1985 ; 5 : 595-600.
2. Aziz N, Munro HN. Both subunits of rat liver ferritin are regulated at a translational level by iron induction. *Nucleic Acids Res* 1986 ; 14 : 915-27.
3. Hentze MW, Rouault TA, Caughman SW, *et al.* A cis-acting element is necessary and sufficient for translational regulation of human ferritin expression in response to iron. *Proc Natl Acad Sci USA* 1987 ; 84 : 6730-4.
4. Aziz N, Munro HN. Iron regulates ferritin mRNA translation through a segment of its 5' untranslated region. *Proc Natl Acad Sci USA* 1987 ; 84 : 8478-82.
5. Casey JL, Hentze MW, Koeller DM, *et al.* Iron-responsive elements : regulatory mRNA sequences that control mRNA level and translation. *Science* 1988 ; 238 : 924-8.
6. Müllner EW, Kühn LC. A stem-loop in the 3' untranslated region mediates iron-dependent regulation of transferrin receptor mRNA stability in the cytoplasm. *Cell* 1988 ; 53 : 815-25.
7. Rouault TA, Hentze MW, Caughman SW, Harford JB, Klausner RD. Binding of a cytosolic protein to the iron-response element of human ferritin messenger RNA. *Sciences* 1988 ; 241 : 1207-10.

Les tubulines sont le constituant principal des microtubules qui jouent un rôle essentiel dans de nombreux processus cellulaires tels le maintien de la forme, la constitution des flagelles et des cils, ainsi que celle du fuseau mitotique ou méiotique. Les microtubules interviennent aussi dans l'allongement des neurites et des axones. L'expression des gènes codant pour les tubulines est contrôlée à différents niveaux. Tout d'abord, l'activation séquentielle, au cours de la différenciation, des membres d'une « petite » famille multigénique

(composée d'environ 6 gènes) codant pour les sous-unités  $\alpha$  et  $\beta$  de la tubuline. Ensuite, la modulation à court terme de la quantité disponible de messagers codant pour ces protéines. Dans la cellule, les molécules de tubuline sont soit polymérisées, formant les microtubules, soit sous forme de dimères  $\alpha$ - $\beta$  libres. Le premier état, correspondant à la construction de nouveaux microtubules, exige la synthèse de nouvelles molécules alors que le second, conséquence d'une dépolymérisation des microtubules préexistants, doit au contraire être asso-

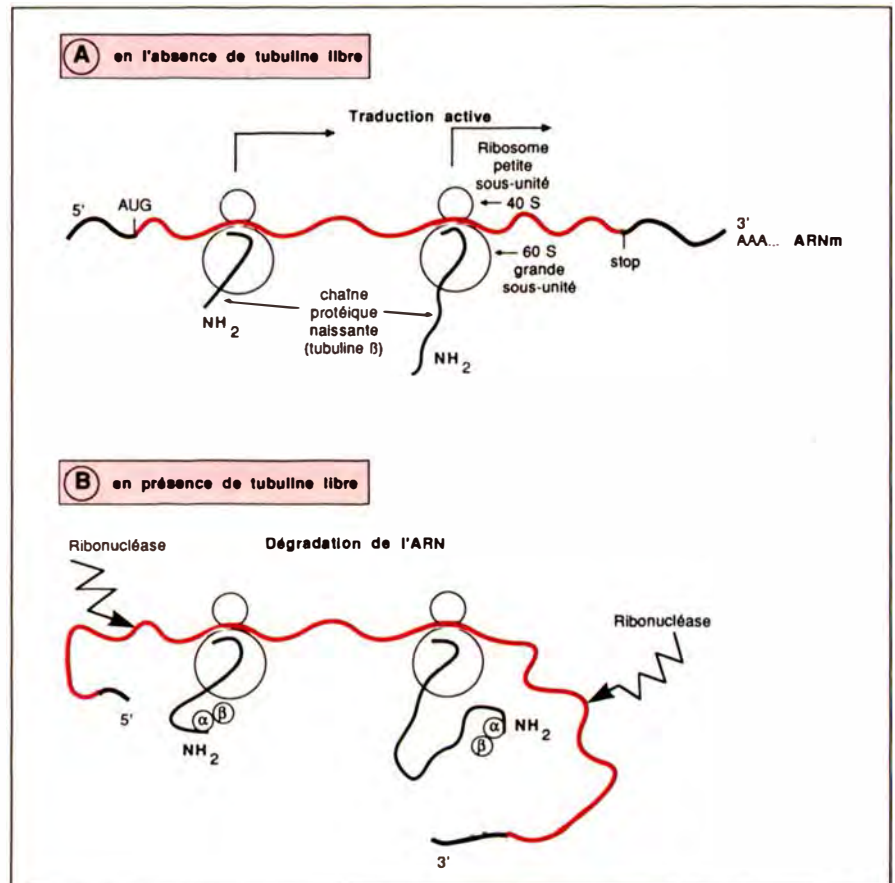


Figure 1. Schéma hypothétique du contrôle de la stabilité du messager de  $\beta$ -tubuline par les sous-unités libres ou les dimères de tubuline (pour les détails, voir le texte).

cié à la répression de la synthèse de tubuline pour éviter l'accumulation de molécules inutilisées. De fait, l'administration de drogue dépolymérisant les microtubules entraîne une diminution importante de la synthèse de nouvelles molécules de tubuline (diminution de 10 fois pour un doublement de la concentration de dimères libres  $\alpha$ - $\beta$ ). Le mécanisme de ce contrôle est une diminution de la stabilité du messenger sous l'action de la tubuline libre. De très élégantes expériences menées dans le laboratoire de Don W. Cleveland [1, 2], à Baltimore, dans le Maryland aux États-Unis, ont démontré que la cible des processus de déstabilisation activés par les sous-unités ou les dimères libres résidait dans la partie du messenger de la  $\beta$ -tubuline correspondant aux 16 premiers codons, dans le premier exon. Un messenger chimère contenant cette région du message de la  $\beta$ -tubuline en amont d'une toute autre séquence se comporte comme l'ARNm authentique de tubuline en réponse aux variations de tubuline libre. Ce type de régulation est conservé lorsque, dans la construction génétique hybride transférée dans les cellules, le promoteur d'un gène de tubuline est remplacé par un promoteur différent, viral ou cellulaire; ce dernier résultat confirme que le contrôle est totalement post-transcriptionnel. Les phénomènes en cause restent inconnus mais de récents résultats du même laboratoire suggèrent un mécanisme particulier. L'effet déstabilisateur des molécules libres de tubuline ne s'exerce que sur des molécules d'ARNm en cours de traduction, c'est-à-dire localisées dans les polysomes [3]. Ce fait, associé d'une part à l'observation que le motif responsable de la déstabilisation en présence de tubuline libre correspond à l'extrémité NH<sub>2</sub>-terminale de la protéine et, d'autre part, à l'affinité qu'ont les unes pour les autres les molécules de tubuline, milite en faveur d'une interaction entre les

sous-unités non polymérisées de tubuline et l'extrémité NH<sub>2</sub>-terminale de la chaîne protéique naissante, en cours de synthèse (figure 1, page 193). Une telle interaction pourrait modifier la conformation de l'ARN, le sensibilisant à l'action d'une ribonucléase. Également en faveur de ce mécanisme hypothétique est l'insensibilité du phénomène aux inhibiteurs de la synthèse protéique qui ne détruisent pas les polysomes (cycloheximide, par exemple). Un tel phénomène de contrôle de l'expression d'un gène par son produit est exceptionnel, l'un des rares autres exemples connus étant celui de l'antigène T du virus SV-40 qui, entre autres effets, inhibe la transcription de son gène. Une autre grande originalité du système tubuline est la localisation dans la séquence codante du motif responsable de l'instabilité du messenger en présence de sous-unités ou de dimères libres; dans la plupart des autres exemples connus de contrôle de la stabilité d'un messenger, les motifs conférant cette propriété sont localisés dans l'extension 3' non codante des ARNm [exemples des messagers *c-fos*, *c-myc*, du messenger du facteur de croissance CSF-1 (ou M-CSF)] et la déstabilisation du messenger nécessite une synthèse protéique active, probablement nécessaire à la production de l'agent « déstabilisant ».

A.K.

1. Gay DA, Yen TJ, Lau JTY, Cleveland DW. Sequences that confer  $\beta$  tubulin autoregulation through modulated mRNA stability reside within exon 1 of a  $\beta$  tubulin mRNA. *Cell* 1987; 50: 671-9.
2. Cleveland DW. The multitubulin hypothesis revisited: what have we learned? *J Cell Biol* 1987; 104: 381-3.
3. Pachter JS, Yen TY, Cleveland DW. Autoregulation of tubulin expression is achieved through specific degradation of polysomal tubulin mRNAs. *Cell* 1987; 51: 283-92.

**JL** John Libbey  
EUROTEXT  
LONDON PARIS

Paris · Londres · Rome  
EDITIONS MEDECINE-SCIENCES



VIENT DE PARAITRE

T. Barbui, A. Falanga,  
B. Minetti, S. Gorini,  
G. Tognoni, M.-B. Donati

### Infections and haemorrhage in acute leukaemia

This book assembles the presentations given at the 1988 « Bergamo Spring Conference on Haematology » and is devoted to physiopathological and clinical problems related to the two major causes of death in acute leukaemia: haemorrhage and infections.

1989, broché, 218 pages  
325 FF

BON DE COMMANDE

NOM ..... Prénom .....  
Adresse .....

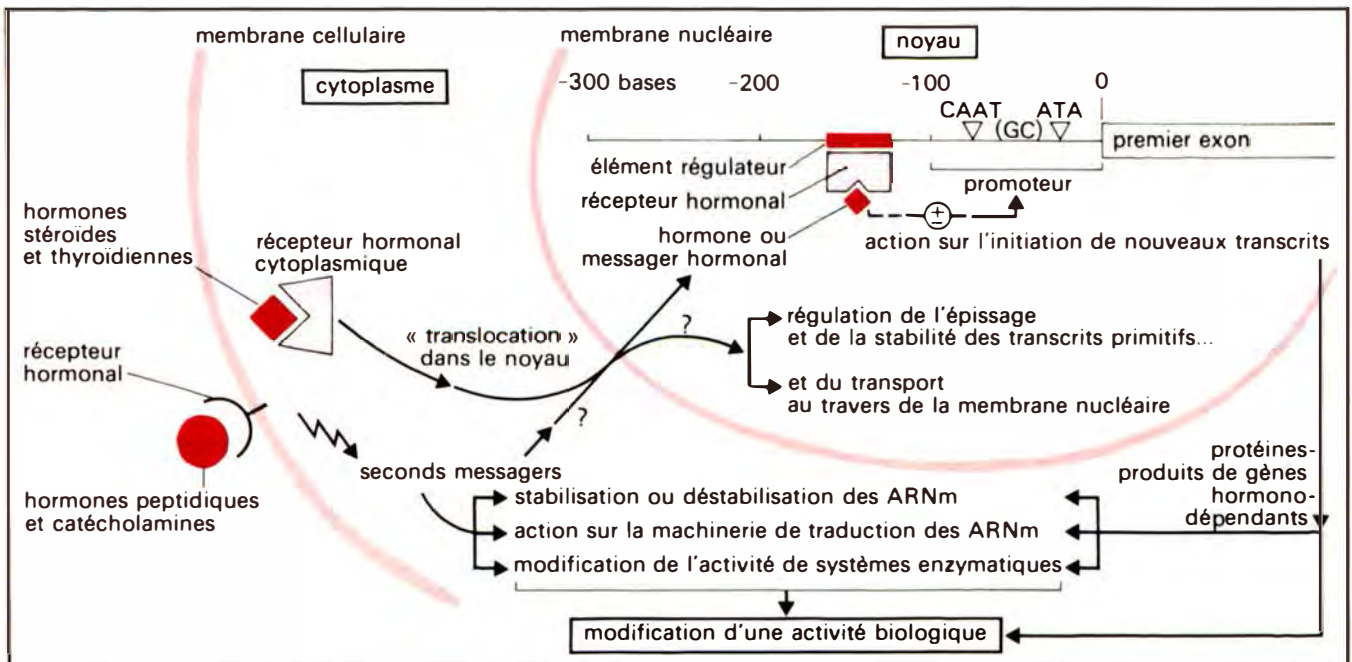
Désire recevoir **Infections and haemorrhage in acute leukaemia** au prix de 325 FF + 30 F de frais de port, soit 355 FF.  
Ci-joint mon règlement à l'ordre de **John Libbey Eurotext**  
6, rue Blanche, 92120 MONTROUGE  
Tél.: (1) 47.35.85.52.

# Contrôle hormonal de l'expression des gènes

Les hormones constituent l'élément clé de la régulation des gènes au cours de l'adaptation métabolique d'un tissu différencié et jouent un rôle essentiel dans le contrôle du développement. Elles peuvent intervenir de manière coordonnée aux différentes étapes du contrôle de l'expression d'un gène. Les hormones stéroïdes et thyroïdiennes se lient à un récepteur cytoplasmique, puis nucléaire (il peut d'ailleurs s'agir du même récepteur « transloqué » au travers de la membrane nucléaire). Le complexe « hormone-récepteur » se fixe à une séquence spécifique de l'ADN située en amont du gène, de 100 à plus de 2000 bases avant le

site d'initiation de la transcription. Cette séquence d'ADN appelée généralement « élément régulateur » est assez courte (quelques dizaines de bases) et se comporte comme un « enhancer » hormono-dépendant (voir le Lexique de médecine/sciences en page 3 de ce document) : sa liaison au complexe hormone-récepteur provoque une activation de la transcription. Dans d'autres cas, le résultat est au contraire une inhibition de la transcription, l'élément régulateur se comportant alors comme un « silencer » hormono-dépendant (voir en page 10 de ce document). Les hormones peptidiques et les catécholamines se fixent à un récep-

teur extra membranaire, cette liaison entraînant l'activation d'un système intra cellulaire qui synthétise ou libère un « second messenger » (par exemple, l'AMP cyclique ou les produits de l'hydrolyse du phosphatidyl-inositol 4-5 biphosphate (voir médecine/sciences n° 5, vol. 1, p. 255)). Ces seconds messagers peuvent agir directement sur des protéines préexistantes dont l'activité biologique peut être modifiée, par exemple par phosphorylation-déphosphorylation. Ils peuvent aussi modifier la stabilité et la traductibilité d'ARNm préexistants. Ces seconds messagers de l'action hormonale ont également une action nucléaire, dont le mécanisme reste



Le promoteur schématisé ici en amont du premier exon comporte les éléments « consensus » : « AT-A-box » et « CAAT-box » décrits dans le Lexique de médecine/sciences p. 2 de ce document. Entre ces deux séquences, des répétitions nucléotidiques riches en GC jouent également un rôle important dans la fixation et le positionnement correct de l'ARN polymérase.

mal connu mais qui pourrait être assez proche de celui des hormones stéroïdes et thyroïdiennes.

Dans le noyau, l'action hormonale peut s'exercer au niveau de la maturation et de la stabilité des transcrits primitifs des gènes, ainsi que sur le transport au travers de la membrane nucléaire de transcrits ayant subi une maturation normale. Les mécanismes en cause sont ici tout à fait inconnus.

Enfin, des protéines dont la synthèse ou l'activité ont été modifiées par l'un des processus cités plus haut peuvent, secondairement, agir elles-mêmes sur les messagers ou les protéines spécifiques de l'action hormonale.

En conclusion, l'action hormonale apparaît comme un processus étroitement coordonné et intégré.

- Les gènes sont souvent régulés par plusieurs hormones pouvant agir isolément ou en association, certaines étant activatrices et d'autres inhibitrices de l'expression d'un même gène.

- Une même hormone agit souvent de manière opposée sur des gènes codant pour les protéines régulatrices impliquées dans des voies métaboliques différentes.

- Et enfin, l'action d'une hormone sur l'expression d'un même gène fait habituellement intervenir de multiples niveaux de contrôle : transcriptionnels, post-transcriptionnels (maturation, transport et stabilité des ARNm), parfois traductionnels (efficacité de la traduction des ARNm) et post-traductionnels (régulation de l'activité biologique d'une protéine préexistante).

A. K.

1. Rogers J. Exon shuffling and intron insertion in serine protease genes. *Nature* 1985; 315: 458-9.

2. Südhof TC, Goldstein JL, Brown M, Sand Russel DW. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 1985; 228: 815-22.

3. Solnick D. Trans splicing of m RNA precursors. *Cell* 1985; 42: 157-64.

4. Konarska MM, Padgett RA, Sharp PA. Trans splicing of m RNA precursors in vitro. *Cell* 1985; 42: 165-71.



**F. ZANNAD**

1989, broché  
156 pages  
ISBN 0.86196.205.2  
180 FF

*Un ouvrage qui fait le point  
sur la physiopathologie, les conséquences et  
l'exploration de l'insuffisance cardiaque*

- Évaluation des thérapeutiques de l'insuffisance cardiaque
- Apport de la clinique
- Nouvelles méthodes d'exploration
- Analyse et critique des protocoles d'étude

BON DE COMMANDE

NOM \_\_\_\_\_ Prénom \_\_\_\_\_

Adresse \_\_\_\_\_

Désire recevoir l'ouvrage de F. ZANNAD, au prix de 180 FF + 30 F de frais de port, soit 210 FF.

Ci-joint mon règlement à l'ordre de **John Libbey Eurotext**  
6, rue Blanche - 92120 MONTROUGE - France - Tél. : (1) 47.35.85.52