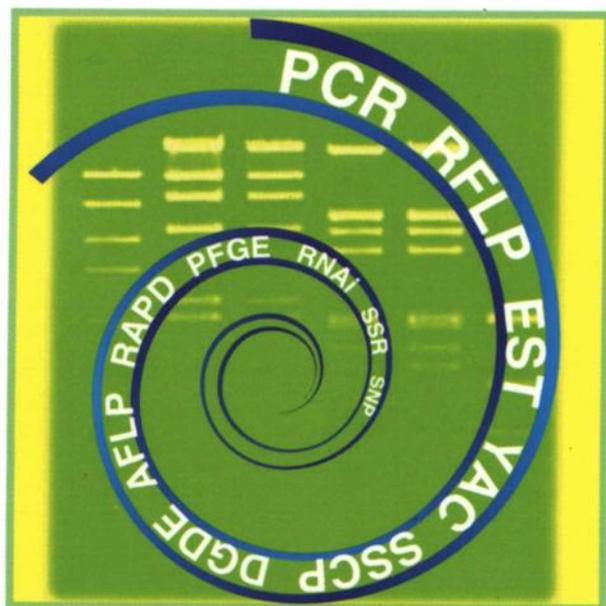


Principes des techniques de biologie moléculaire

2^e édition revue
et augmentée

D. Tagu, C. Moussard, eds



 **INRA**
EDITIONS

Principes des techniques de biologie moléculaire

2^e édition revue
et augmentée

© INRA, Paris, 2003

ISBN : 2-7380-1067-9

ISSN : 1144-7605

Le code de la propriété intellectuelle du 1^{er} juillet 1992 interdit la photocopie à usage collectif sans autorisation des ayants droit. Le non respect de cette disposition met en danger l'édition, notamment scientifique. Toute reproduction, partielle ou totale, du présent ouvrage, est interdite sans autorisation de l'éditeur et du Centre français d'exploitation du droit de copie (CFC), 20 rue des Grands Augustins, 75006 Paris.

Principes des techniques de biologie moléculaire

2^e édition revue
et augmentée

D. Tagu, C. Moussard, édés

MIEUX COMPRENDRE

Ouvrages parus dans la même collection :

Éléments de génétique quantitative

2^e édition revue et augmentée

L. OLLIVIER

2002, 184 p.

Génie génétique

Une histoire, un défi

E. HEBERLE-BORS

trad. M.L. SPIRE, R. JUDOR

2001, 304 p.

L'eau dans l'espace rural

Vie et milieux aquatiques

A. NEVEU, C. RIOU, R. BONHOMME,

P. CHASSIN, F. PAPY (éd.)

2001, 300 p.

Principes de virologie végétale

Génome, pouvoir pathogène, écologie
des virus

S. ASTIER, J. ALBOUY, Y. MAURY, H. LECOQ

2001, 488 p.

Le grain de blé

Composition et utilisation

P. FEILLET

2000, 310 p.

Biology of lactation

J. MARTINET, L.-M. HOUEBINE, H.H. HEAD

1999, 686 p.

Sol : interface fragile

Pierre STENGEL et Sandrine GELIN

1998, 222 p.

Les marqueurs moléculaires en génétique et biotechnologies végétales

Dominique DE VIENNE

1998, 200 p.

Assimilation de l'azote chez les plantes

Aspects physiologique, biochimique
et moléculaire

Jean-François MOROT-GAUDRY (éd.)

1997, 422 p.

L'eau dans l'espace rural

Production végétale et qualité de l'eau

C. RIOU, R. BONHOMME, P. CHASSIN,

A. NEVEU, F. PAPY (éd.)

1997, 414 p.

La pomme de terre

P. ROUSSELLE, Y. ROBERT

et J.C. CROSNIER (éd.)

1996, 640 p.

Vie microbienne du sol et production végétale

Pierre DAVET

1996, 380 p.

Nutrition des ruminants domestiques

R. JARRIGE, Y. RUCKEBUSH,

C. DEMARQUILLY, M.-H. FARCE

et M. JOURNET (éd.)

1995, 921 p.

Amélioration des espèces végétales cultivées

Objectifs et critères de sélection

André GALLAIS et Hubert BANNEROT

1992, 768 p.

La régression non linéaire : méthodes et applications en biologie

Sylvie HUET, Emmanuel JOLIVET

et Antoine MESSÉAN

1992, 250 p.

L'épidémiologie en pathologie végétale : mycoses aériennes

Frantz RAPILLY

1991, 318 p.

Les oligo-éléments en agriculture et en élevage

Yves Coïc et Marcel COPPENET

1989, 114 p.

Cette seconde édition est une version revue et augmentée de la précédente. Nous tenons à remercier les rédacteurs de ces fiches qui ont su donner de leur temps pour mettre à jour cet ouvrage.

SOMMAIRE

Définitions

1. Structure et expression d'un gène eucaryote codant un ARNm et une protéine 6
2. Paramètres de description d'un génome 8
3. Séquençage de génomes entiers 12

Vecteurs et clonage

4. Enzymes de restriction 14
5. Electrophorèse des acides nucléiques 16
6. Description d'un plasmide et d'un phagemide 18
7. Description d'un bactériophage et d'un cosmide 20
8. Description d'un YAC et autres grands vecteurs 22
9. Clonage moléculaire 24
10. Transformation génétique des bactéries et des levures 27

Marquage d'acides nucléiques et hybridations

11. Marquage de l'ADN 29
12. Hybridation moléculaire 32
13. Hybridation *in situ* des ARNm 36

Banques d'ADN et criblage

14. Construction d'une banque d'ADN génomique 38
15. Construction d'une banque d'ADNc 40
16. Criblage d'une banque 42
17. Criblage différentiel : banques d'ADNc soustraites, AFLP-ADNc 46
18. Criblage différentiel par DD RT-PCR : tri d'ARNm
(*differential display RT-PCR*) 50
19. Criblage différentiel par SSH : hybridation soustractive et suppressive
(*suppression subtractive hybridization*) 54
20. Criblage différentiel par RDA : analyse par différence de représentation
(*representational difference analysis*) 58

21. EST : étiquettes de gènes exprimés (<i>expressed sequence tags</i>)	62
22. Réseaux d'ADN : puces à ADN, filtres d'ADNc	65

Caractérisation d'un gène

23. Séquençage d'ADN	76
24. PCR (<i>polymerase chain reaction</i>)	80
25. RACE : amplification rapide d'extrémités d'ADNc (<i>rapid amplification of cDNAs ends</i>)	82
26. Marche génomique par PCR	86
27. RT-PCR : PCR sur ARNm (<i>reverse transcriptase PCR</i>)	88
28. Transcription <i>in vitro</i>	90
29. Détermination du site d'initiation de la transcription	92
30. Analyse fonctionnelle de promoteurs	94
31. Gel-retard	96
32. Empreinte à la DNase I (<i>footprinting</i>)	98

Transformations génétiques d'eucaryotes

33. Transformation génétique des végétaux par <i>Agrobacterium tumefaciens</i>	100
34. Transfert direct de gènes dans des protoplastes végétaux	105
35. Transfert direct de gènes par biolistique	106
36. Transformation génétique de cellules animales	107
37. Le clonage des animaux	110
38. Expression transitoire	114

Analyse de la fonction d'un gène

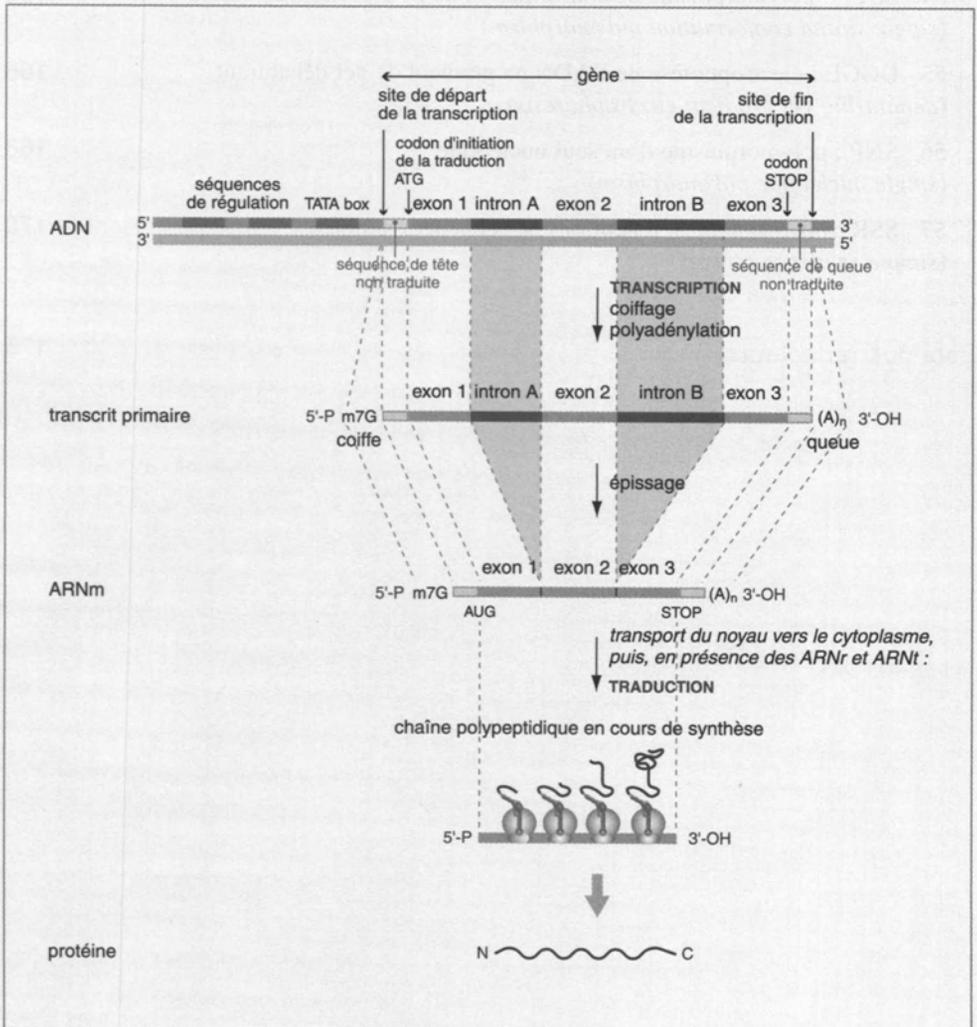
39. Protéines recombinantes	116
40. Les baculovirus d'insectes, vecteurs d'expression de gènes étrangers	120
41. Système du double hybride	124
42. Mutagenèse dirigée	128
43. Complémentation de mutation chez la Levure	132
44. Inactivation de gènes (<i>knock-out</i>)	134
45. Étiquetage moléculaire	136
46. RNAi : inactivation de gènes par interférence d'ARN (<i>RNA interference</i>)	140

Polymorphisme d'un génome	
47. Marqueurs génétiques moléculaires	143
48. Cartes génétique et physique	148
49. PFGE : électrophorèse en champs pulsés (<i>pulse field electrophoresis</i>)	152
50. RFLP : polymorphisme de longueur des fragments de restriction (<i>restriction fragment length polymorphism</i>)	154
51. RAPD : polymorphisme d'ADN par amplification aléatoire (<i>random amplified polymorphic DNA</i>)	156
52. AFLP : polymorphisme de longueur de fragments amplifiés (<i>amplified fragment length polymorphism</i>)	158
53. Les rétromarqueurs	160
54. SSCP : polymorphisme de conformation de l'ADN monocaténaire (<i>single strand conformation polymorphism</i>)	164
55. DGGE : électrophorèse de l'ADN en gradient de gel dénaturant (<i>denaturing gel gradient electrophoresis</i>)	166
56. SNP : polymorphisme d'un seul nucléotide (<i>single nucleotide polymorphism</i>)	168
57. SSR : microsatellites, répétitions de séquences simples (<i>simple sequence repeats</i>)	170
Liste des rédacteurs	175

FICHE 1

STRUCTURE ET EXPRESSION D'UN GÈNE EUCARYOTE CODANT UN ARNm ET UNE PROTÉINE

Un gène eucaryote à ARNm comporte une séquence codante bordée de séquences de régulation. Ces dernières servent de signaux de début ou de fin de transcription du gène par l'ARN polymérase II. Certaines de ces séquences (p. ex. la TATA box du promoteur) sont reconnues par des protéines appelées "facteurs de transcription généraux" car elles assistent cette enzyme dans les étapes d'initiation et d'élongation de la transcription.



D'autres séquences d'ADN¹ sont reconnues par des "facteurs de transcription spécifiques" qui modulent l'expression des gènes dans l'espace (selon le type de cellule), dans le temps (au cours du développement) et/ou sous l'effet de stress. L'agencement de ces protéines sur l'ADN déclenchera ou inhibera la transcription du gène.

La séquence codante est constituée d'exons et d'introns. Ces deux types de séquences sont transcrites (transcrit primaire), mais les introns sont éliminés lors de l'épissage. L'ARN est d'abord modifié en 5'-P (addition d'une coiffe) et en 3'-OH (addition de nucléotides à adénine), puis épissé (élimination des introns) avant d'être transporté dans le cytoplasme. Là, les ribosomes se fixent sur l'ARNm (ARN messenger) et, par l'intermédiaire des ARNt (ARN de transfert), l'ARNm est traduit en le polypeptide correspondant.

1. Elles peuvent se trouver en aval et/ou en amont de la séquence codante.

PARAMÈTRES DE DESCRIPTION D'UN GÉNOME

Paramètres¹

- Quantité d'ADN par noyau 2C : de l'ordre du picogramme ($1 \text{ pg} = 10^{-12} \text{ g}$).
- Longueur totale de l'ADN : exprimée en paire de bases (pb), en kilobp (kpb), en mégabases ($1 \text{ Mb} = 10^6 \text{ pb}$).
- Pourcentage en G + C : pourcentage de bases cytosine et guanine dans une séquence d'ADN connue.
- Diamètre de l'ADN : 20 \AA . Un morceau d'ADN de 1 \AA de longueur a une masse de $193/6 \cdot 10^{23} \text{ g}$.
- 1 paire de bases a une masse de $660/N$, soit $660/6 \cdot 10^{23} \text{ g}$.

Grandeur de différents génomes¹

- Virus (ADN ou ARN) : de 10^3 à 10^5 pb , soit de 1 à 100 \mu m déroulé.
- Bactéries : 10^6 à 10^7 pb , soit environ 1 mm .
- Levure : $1,2 \cdot 10^7 \text{ pb}$.
- Plantes supérieures :
 - noyau : de 10^8 à 10^{10} pb , soit environ 1 m . La ploïdie est généralement de 2 et l'hérédité de type mendélienne.
 - mitochondries : de $1,8$ à $6 \cdot 10^5 \text{ pb}$. La longueur est très variable. La ploïdie est non déterminée et l'hérédité est de type cytoplasmique, maternelle.
 - chloroplastes : $1,5 \cdot 10^5 \text{ pb}$, soit 50 \mu m . La ploïdie est de l'ordre de $10x$ par chloroplaste et il y a de 10 à 100 chloroplastes par cellule. L'hérédité est de type cytoplasmique, maternelle.
- Homme :
 - noyau : $3 \cdot 10^9 \text{ pb}$, soit 1 m .
 - mitochondries : 16 à $20 \cdot 10^3 \text{ pb}$.

Les trois types de séquences du génome nucléaire¹

- Séquences "uniques" : une copie ou quelques copies par génome haploïde (on parle alors d'une petite famille multigénique, p. ex. les gènes codant pour les protéines de réserve des Graminées). Ces séquences comprennent des gènes de protéines et des transposons.

- Séquences moyennement répétées : de 1 000 à 100 000 copies par génome. Elles comprennent les gènes codant pour les ARN ribosomiques (ARNr), ARN de transfert (ARNt), éventuellement des gènes de protéines.
- Séquences hautement répétées : de 100 000 à 1 000 000 de copies. Ces séquences ne sont pas des gènes et peuvent être ou non transcrites. Leur rôle est peu connu (p. ex. l'ADN satellite).

Données actuelles sur les génomes

De nombreux génomes procaryotes et eucaryotes ont été séquencés ou sont en cours de séquençage (voir Fiche 3). Le tableau suivant donne quelques exemples de paramètres physiques de génomes.

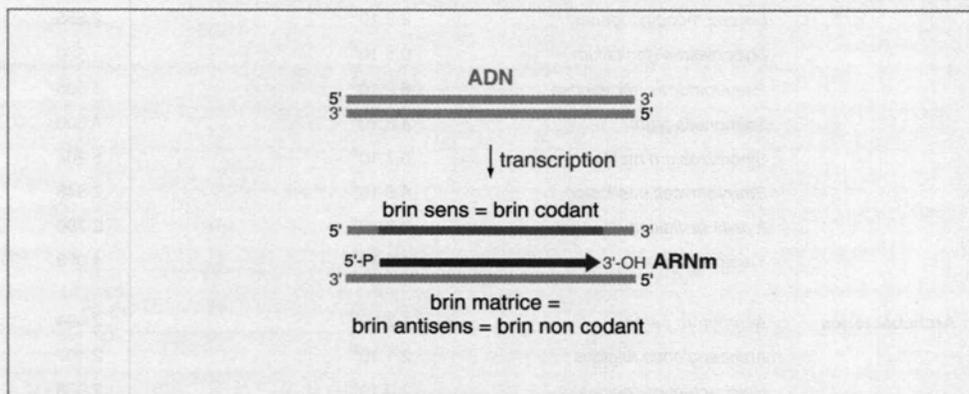
		Taille (pb)	Nombre de gènes
Eubactéries	<i>Agrobacterium tumefaciens</i>	4,8.10 ⁶	4 554
	<i>Bacillus subtilis</i>	4,2.10 ⁶	4 100
	<i>Buchnera species</i>	0,6.10 ⁶	564
	<i>Enterococcus faecalis</i>	3,2.10 ⁶	3 334
	<i>Escherichia coli</i>	4,6.10 ⁶	4 288
	<i>Haemophilus influenzae</i>	1,8.10 ⁶	1 709
	<i>Helicobacter pylori</i>	1,6.10 ⁶	1 566
	<i>Lactococcus lactis</i>	2,3.10 ⁶	2 566
	<i>Listeria monocytogenes</i>	2,9.10 ⁶	2 855
	<i>Mycoplasma genitalium</i>	0,5.10 ⁶	480
	<i>Pseudomonas aeruginosa</i>	6,2.10 ⁶	5 565
	<i>Salmonella typhi</i>	4,8.10 ⁶	4 600
	<i>Sinorhizobium meliloti</i>	6,7.10 ⁶	6 204
	<i>Streptomyces coelicolor</i>	8,6.10 ⁶	7 848
	<i>Xylella fastidiosa</i>	2,6.10 ⁶	2 766
<i>Yersinia pestis</i>	4,6.10 ⁶	4 008	
Archéobactéries	<i>Aeropyrum pernix</i>	1,7.10 ⁶	2 694
	<i>Archaeoglobus fulgidus</i>	2,1.10 ⁶	2 407
	<i>Halobacterium species</i>	2,0.10 ⁶	2 058
	<i>Methanobacterium thermoautotrophicum</i>	1,7.10 ⁶	1 869
	<i>Methanococcus jannaschii</i>	1,6.10 ⁶	1 715
	<i>Pyrobaculum aerophilum</i>	2,2.10 ⁶	2 605
	<i>Pyrococcus abyssi</i>	1,7.10 ⁶	1 765
	<i>Sulfolobus solfataricus</i>	2,9.10 ⁶	2 977
	<i>Thermoplasma acidophilum</i>	1,5.10 ⁶	1 478

		Taille (pb)	Nombre de chromosomes	Nombre de gènes (estimation)
Eucaryotes	<i>Arabidopsis thaliana</i>	125.10 ⁶	5	25 500
	<i>Caenorhabditis elegans</i>	97.10 ⁶	6	19 000
	<i>Drosophila melanogaster</i>	140.10 ⁶	4	13 600
	<i>Encephalitozoon cuniculi</i>	2.9.10 ⁶	11	1 997
	<i>Homo sapiens</i>	3 100.10 ⁶	21	35 000
	<i>Oryza sativa</i>	420.10 ⁶	12	40 000
	<i>Saccharomyces cerevisiae</i>	12.10 ⁶	16	6 100

Brin codant, brin sens : quelques définitions

L'ADN est constitué de deux brins antiparallèles et complémentaires. Lors de sa transcription en ARN, seul un des deux brins est lu et copié par l'ARN polymérase. Le brin d'ARN obtenu est donc complémentaire du brin d'ADN matrice qui a servi de copie. Par définition, ce brin d'ADN matrice est appelé antisens ; l'autre brin, qui a la même séquence que l'ARNm, est le brin sens.

L'ARNm est ensuite lu et décrypté en protéine : c'est la traduction. Par convention, la séquence d'ADN identique à l'ARN (donc le brin sens) est appelé brin codant ; l'écriture d'un gène sur le papier correspond au brin codant dans son orientation 5'-P vers l'extrémité 3'-OH.



1. D'après Francis Quétier

SÉQUENÇAGE DE GÉNOMES ENTIERS

La connaissance de la structure d'un génome dans son entier peut passer par son séquençage. Cependant, la taille des génomes étant de plusieurs millions de bases (ou mégabases, voir Fiche 2), il est nécessaire de coupler les approches de biologie moléculaire et d'informatique pour pouvoir gérer ces millions de données.

On distingue deux méthodes de séquençage des génomes entiers. Dans les deux cas, l'ADN génomique est tout d'abord fragmenté par des méthodes soit enzymatiques (enzymes de restriction), soit physiques (ultrasons) :

- la première méthode de séquençage, dite par ordonnancement hiérarchique, consiste à classer les fragments génomiques obtenus avant de les séquencer ;
- la seconde méthode, dite globale (ou *whole-genome shotgun*), néglige la phase préalable de tri et propose de séquencer dans le désordre chacun des fragments génomiques. La puissance de la bioinformatique permet dans un second temps de réordonner les fragments génomiques par chevauchement de leurs séquences communes.

La principale différence est que l'ordonnancement hiérarchique essaie d'aligner un jeu de clones de grande taille (~ 100 kb) alors que, dans la méthode globale, le génome entier est réduit en fragments de petite taille, séquencés puis alignés.

Ordonnancement hiérarchique

L'ADN génomique est, après extraction, fragmenté, généralement par des procédés de sonication, en fragments maniables (c'est-à-dire clonables dans des grands vecteurs) de 50 à 200 kb, puis cloné dans un vecteur adapté comme les chromosomes artificiels bactériens (ou BACs, voir Fiche 8). Le nombre de clones doit permettre une redondance (une représentation) de 5 à 10 fois le génome total. Le chevauchement et l'ordonnancement des clones est alors réalisé soit par hybridation de sondes spécifiques (marqueurs moléculaires), soit par analyse des profils de restriction, soit plus fréquemment par un ordonnancement après séquençage et hybridation des extrémités des BACs (500 à 600 bp). Après avoir effectué cet ordonnancement des clones (donc des fragments génomiques), les clones sélectionnés sont individuellement fragmentés, séquencés et assemblés par informatique (alignements).

Les avantages de l'ordonnancement hiérarchique sont de deux ordres : (i) une plus grande facilité d'assemblage des séquences grâce aux données des cartes physiques, génétiques et aux chevauchements des BACs ; (ii) une ouverture vers un travail en collaboration entre plusieurs laboratoires, chacun séquençant une région de chromosome. L'inconvénient majeur est la présence, à la fin, de portions d'ADN absentes de l'analyse à cause généralement des difficultés de clonage dans *E. coli* de certaines séquences répétées très fréquentes.

Méthode globale

Il s'agit d'une méthode de séquençage "au tout venant" d'ADN génomique fragmenté, mise au point par le groupe privé Celera Genomics sur des génomes bactériens, puis sur le génome de la Drosophile et enfin sur les génomes de l'Homme et de la Souris. Les étapes sont les suivantes : de deux à trois banques de fragments d'ADN aléatoires de tailles différentes sont réalisées. De nombreux clones sont séquencés puis assemblés. La séquence totale est réalisée à force de recouvrements et d'assemblages.

Les avantages par rapport au séquençage hiérarchique sont la rapidité et le faible coût. L'inconvénient reste le problème d'assemblage, notamment en raison de la présence importante de séquences répétées, particulièrement dans les génomes de mammifères.

Séquençage et assemblage

Pour les deux approches décrites précédemment, le séquençage et l'assemblage des fragments se font de la même manière. Le principe de séquençage repose sur la méthode de Sanger (voir Fiche 23). Le séquençage d'un génome entier nécessite des séquenceurs d'ADN particulièrement performants, ainsi que l'assistance de robots capables de réaliser des milliers de pipettages et réactions enzymatiques d'une manière automatisée.

Après la récupération des données brutes du séquençage, il faut assembler les séquences afin de reconstituer la séquence initiale. Ainsi, toutes les séquences obtenues doivent être ordonnées les unes par rapport aux autres par la mise en évidence de recouvrements (extrémités chevauchantes), c'est-à-dire des régions terminales présentant un enchaînement de nucléotides identiques (aux erreurs de séquençage près). Lorsque deux séquences se chevauchent, elles sont contiguës et forment ce qu'on appelle un "contig" qui désigne deux ou plusieurs fragments chevauchants. Le but du projet est d'aboutir (par recouvrements successifs de l'ensemble des contigs) à un contig unique correspondant à la séquence totale de la molécule d'ADN analysée (un chromosome). Le taux d'erreur peut approcher 1 %, sachant que certaines portions (comme les zones répétées) sont plus délicates à séquencer. L'obtention d'une séquence qui couvre 100 % d'un génome est particulièrement coûteuse. C'est pour ces raisons que l'on qualifie de "brouillon" (*working draft*) certains génomes en voie d'achèvement de séquençage.

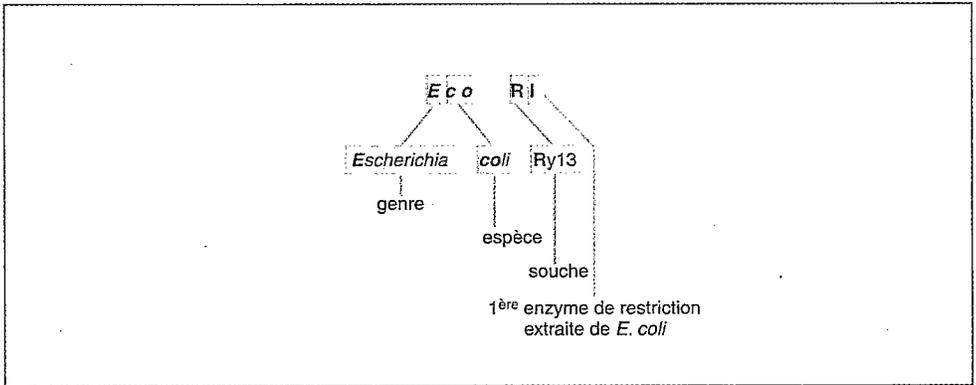
Annotation

L'annotation est l'étape finale : elle est basée sur un travail de bioinformatique à partir des séquences obtenues et alignées. Elle consiste à : (i) localiser les régions codantes (gènes) ; (ii) déterminer la direction de la transcription ; (iii) définir les protéines encodées avec des indications de structure et de fonction ; (iv) identifier les familles de gènes (paralogie¹).

1. Les gènes qui présentent un taux élevé de similarité et qui dérivent donc vraisemblablement d'un même gène ancestral sont qualifiés de paralogues.

ENZYMES DE RESTRICTION

Les enzymes de restriction sont d'origine bactérienne. Elles ont la particularité de couper les molécules d'ADN bicaténaire (ou double brin) à des sites spécifiques de la séquence : ce sont des endonucléases. Chaque enzyme de restriction reconnaît et coupe une séquence nucléotidique donnée. Le nom des enzymes de restriction provient du nom de genre et d'espèce de la bactérie dont elles ont été isolées.

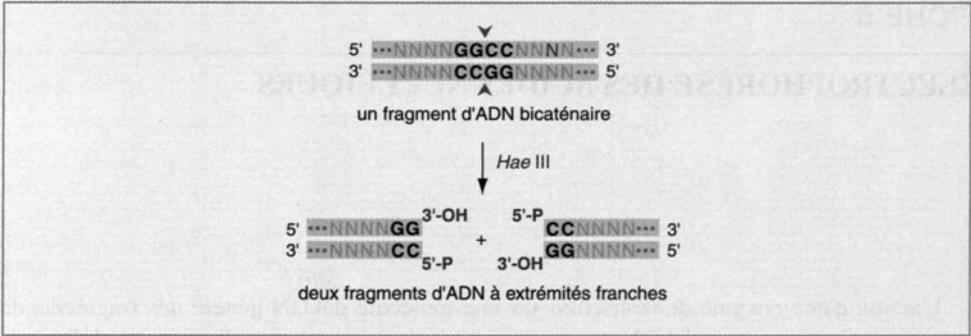


Il existe trois types d'enzymes de restriction. Les types I et III sont des protéines complexes coupant l'ADN bicaténaire en dehors de leur site de reconnaissance. Les enzymes de restriction de type II sont les outils indispensables au génie génétique. Elles reconnaissent une séquence spécifique de 4, 6 ou 8 pb et coupent à l'intérieur de cette séquence, appelée site de restriction. Une particularité de ces sites est qu'ils sont généralement palindromiques, c'est-à-dire que la même séquence est lue sur les deux brins, mais en sens inverses l'un de l'autre.

De très nombreuses enzymes de restriction sont disponibles dans le commerce. Elles sont généralement fournies avec leur tampon de réaction et la température optimale d'activité est indiquée. Il est possible de digérer un même fragment d'ADN par deux enzymes de restriction différentes. Si elles fonctionnent dans des conditions analogues (concentration en sels, pH, température...), les deux digestions se déroulent simultanément. Sinon, le fragment d'ADN doit être digéré par les deux enzymes successivement, en changeant entre les réactions enzymatiques les conditions de l'expérience.

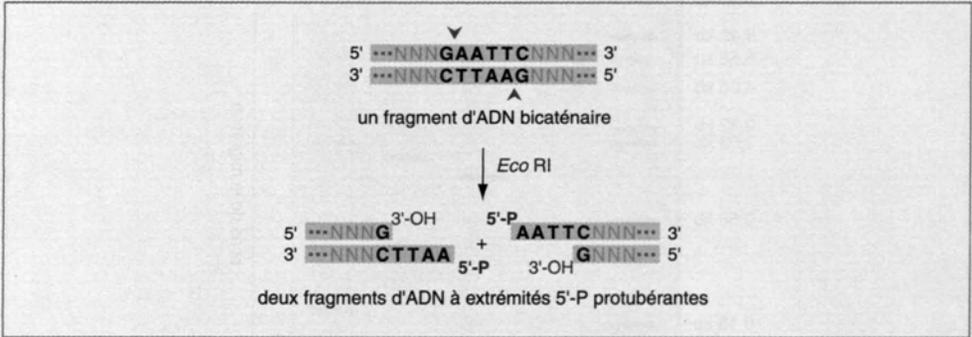
Les enzymes de restriction utilisées coupent dans la séquence selon deux modes :

- coupure franche : sont obtenues des extrémités franches par coupure au même endroit sur les deux brins. Exemple : l'enzyme *Hae* III isolée de *Haemophilus aegyptius* coupe l'ADN double brin au niveau de la séquence GG/CC.

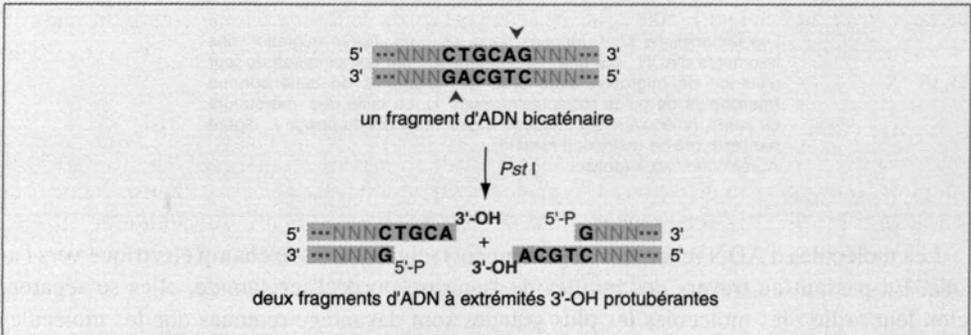


• coupure décalée sur les deux brins : sont générées des extrémités monocaténares (simple brin) ayant des séquences complémentaires appelées extrémités cohésives ou bouts collants. Exemples :

– l'enzyme de restriction *Eco* RI isolée de la bactérie *Escherichia coli* coupe l'ADN bicaténaire dans la séquence palindromique G/AATTC ;

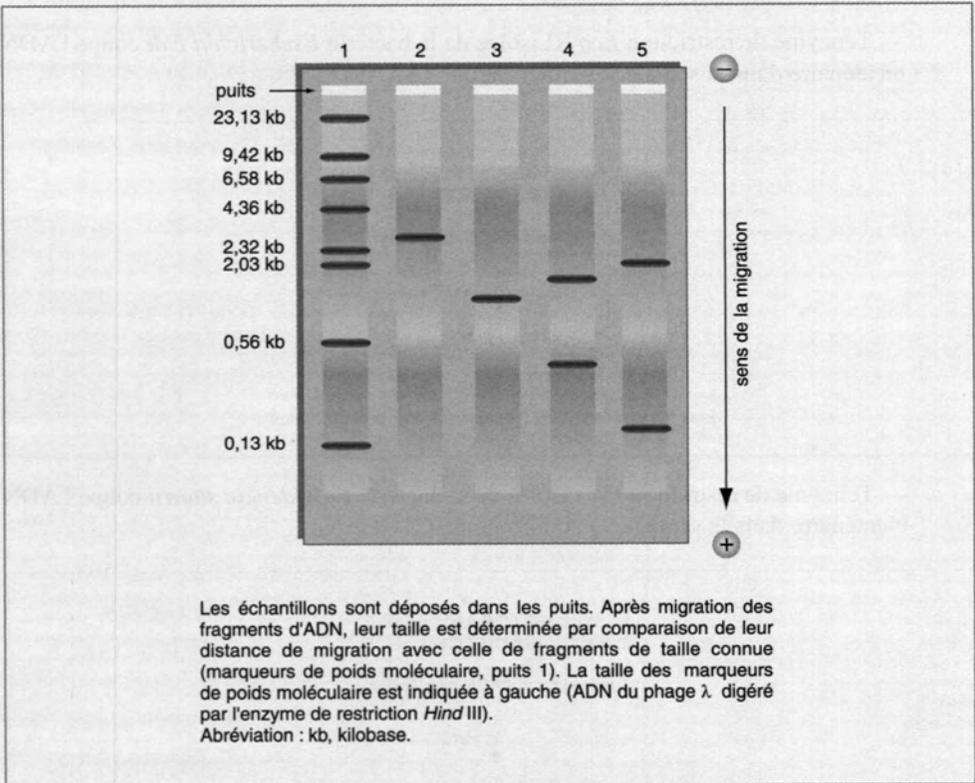


– l'enzyme de restriction *Pst* I isolée de la bactérie *Providencia stuarti* coupe l'ADN bicaténaire dans la séquence palindromique CTGCA/G.



ÉLECTROPHORÈSE DES ACIDES NUCLÉIQUES

L'action d'une enzyme de restriction sur une molécule d'ADN génère des fragments de restriction. Ces fragments d'ADN peuvent être séparés suivant leur taille sur un gel d'agarose ou d'acrylamide. Le mélange contenant l'ADN coupé par l'enzyme de restriction est déposé à une extrémité du gel est ensuite soumis à un champ électrique.



Les molécules d'ADN (chargées négativement) migrent dans le champ électrique vers l'anode. En passant au travers des mailles de l'agarose ou de l'acrylamide, elles se séparent selon leur taille : les molécules les plus grandes sont davantage retenues que les molécules plus petites et migrent moins vite et donc moins loin dans le gel. L'acrylamide a un pouvoir séparateur plus important que l'agarose. Cependant, la réalisation des gels d'acrylamide est moins aisée et ce produit est toxique.

Afin de visualiser les fragments d'ADN après électrophorèse, le gel est trempé dans une solution contenant du bromure d'éthidium¹. Cette molécule s'intercale entre les bases des aci-

des nucléiques et a la propriété d'émettre une fluorescence rouge orange lorsqu'elle est excitée par la lumière ultraviolette. Le gel est alors observé sous une lampe à U.V. et les molécules d'ADN complexées au bromure d'éthidium deviennent visibles. Comme la distance de migration est proportionnelle au logarithme du nombre de bases, il est possible de déterminer la taille des fragments de restriction obtenus en comparant leur mobilité électrophorétique à celle de fragments d'ADN de taille déterminée (marqueurs de taille). Il existe également des marqueurs permettant d'estimer pour chaque fragment la quantité d'ADN bicaténaire.

Dans certains cas, les molécules à séparer ont été préalablement marquées par l'incorporation d'un isotope radioactif, ce qui permet une détection facile par autoradiographie : les particules énergétiques émises par le radio-isotope impressionnent un film photographique placé sur le gel². Il suffit ensuite de développer le film comme un film photographique ordinaire pour voir apparaître des signaux (bandes) noirs correspondant aux molécules radioactives.

1. Le bromure d'éthidium est un dangereux mutagène et cancérigène et doit être manipulé dans des conditions particulières (gants, blouse, déchets recyclés...).
2. Les films autoradiographiques sont de plus en plus souvent remplacés par des écrans autoradiographiques dont les signaux sont révélés par numérisation.

DESCRIPTION D'UN PLASMIDE ET D'UN PHAGEMIDE

Les plasmides et les phagemides, qui se développent dans les bactéries, sont très largement utilisés pour manipuler des fragments de gène de n'importe quel organisme, grâce à l'utilisation des enzymes de restriction (voir Fiche 4), de la PCR (voir Fiche 24) et du clonage moléculaire (voir Fiche 9).

Plasmide (type pUC)

Un plasmide est une petite molécule d'ADN bicaténaire (de 3 à 10 kb) circulaire extra-chromosomique capable de se répliquer (indépendamment du chromosome bactérien) dans une cellule bactérienne et d'être transférée dans une autre. Il est possible de les purifier en grande quantité car ils se multiplient dans des bactéries hôtes.

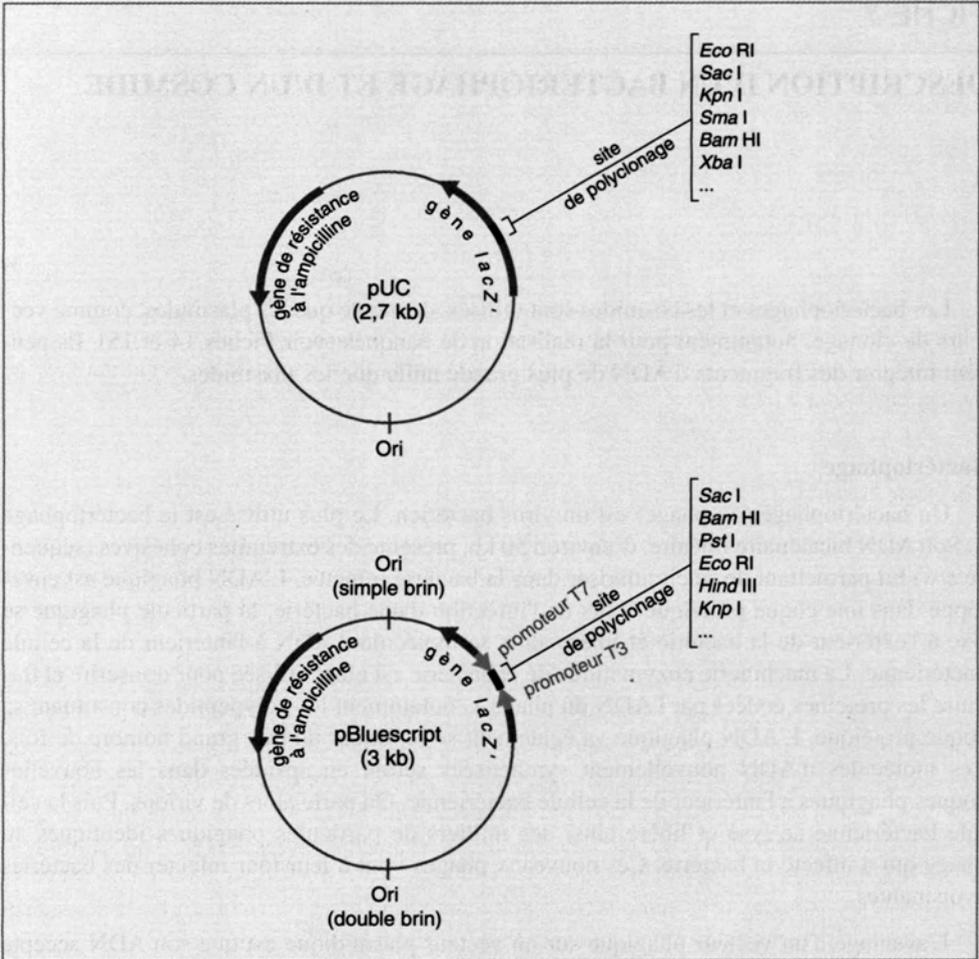
Les plasmides utilisés en génie génétique sont des plasmides naturels qui ont été largement modifiés. Leur génome comprend notamment : (i) une origine de réplication ; (ii) des marqueurs génétiques de sélection des bactéries transformées (c'est-à-dire ayant intégré un plasmide, recombinant ou vide : gène de résistance à un antibiotique) et des bactéries recombinantes (ayant intégré un plasmide recombinant : gène codant pour la β -galactosidase p. ex.) ; (iii) un site de polyclonage. Un site de polyclonage (ou *polylinker*) est une séquence constituée d'une succession de séquences reconnues et coupées par différentes enzymes de restriction. Ces sites sont uniques (ils n'apparaissent qu'à cet endroit dans toute la séquence du plasmide) et sont destinés à recevoir l'ADN étranger (voir Fiche 9).

Des molécules d'ADN exogènes de l'ordre de 4 kb peuvent facilement être intégrées dans ces vecteurs plasmidiques. Par contre, des molécules de taille plus importante sont difficilement acceptées.

Phagemide (type Bluescript)

Les phagemides sont des molécules hybrides entre un plasmide et un phage. Ce sont des molécules d'ADN bicaténaires, circulaires, qui peuvent être obtenues sous forme monocaténaire dans certaines conditions. Ils possèdent une origine de réplication, au moins un gène de résistance à un antibiotique, un site de polyclonage et une séquence provenant du phage M13 contenant l'origine de réplication qui permet d'obtenir la forme monocaténaire (par co-infection avec un phage *helper*). En général, des promoteurs reconnus par des ARN polymérases ont été introduits en amont et/ou en aval du site de polyclonage, afin de pouvoir produire des ARN par transcription *in vitro* (voir Fiche 28).

Des fragments d'ADN de l'ordre de 4 kb peuvent être introduits dans ces vecteurs, mais il est parfois possible d'intégrer des inserts de l'ordre de 10 kb.



DESCRIPTION D'UN BACTÉRIOPHAGE ET D'UN COSMIDE

Les bactériophages et les cosmides sont utilisés, de même que les plasmides, comme vecteurs de clonage, notamment pour la réalisation de banques (voir Fiches 14 et 15). Ils peuvent intégrer des fragments d'ADN de plus grande taille que les plasmides.

Bactériophage

Un bactériophage (ou phage) est un virus bactérien. Le plus utilisé est le bactériophage λ . Son ADN bicaténaire linéaire, d'environ 50 kb, présente des extrémités cohésives (séquence *cos*) lui permettant de se circulariser dans la bactérie infectée. L'ADN phagique est enveloppé dans une coque protéique. Lors de l'infection d'une bactérie, la particule phagique se fixe à l'extérieur de la bactérie et libère alors sa molécule d'ADN à l'intérieur de la cellule bactérienne. La machinerie enzymatique de la bactérie est alors utilisée pour transcrire et traduire les protéines codées par l'ADN du phage λ , notamment les polypeptides constituant sa coque protéique. L'ADN phagique va également se répliquer un très grand nombre de fois. Les molécules d'ADN nouvellement synthétisées seront encapsidées dans les nouvelles coques phagiques à l'intérieur de la cellule bactérienne. On parle alors de virions. Puis la cellule bactérienne se lyse et libère ainsi des milliers de particules phagiques identiques au phage qui a infecté la bactérie. Ces nouveaux phages vont à leur tour infecter des bactéries avoisinantes.

L'avantage d'un vecteur phagique sur un vecteur plasmidique est que son ADN accepte de plus grands fragments d'ADN exogène (20 kb en moyenne contre 4 kb pour un plasmide). Cependant, à cause de la grande taille de l'ADN, il est plus difficile de le manipuler *in vitro*.

Dans la pratique, les phages sont utilisés comme vecteurs de construction de banques d'ADNc ou génomiques (voir Fiches 14 et 15). Lorsque les clones d'intérêt ont été repérés, l'ADN phagique est purifié et les inserts extraits sont intégrés dans un vecteur plasmidique de manipulation plus aisée. On parle alors de sous-clonage (voir Fiche 9). Plus simplement, ce sous-clonage peut également être effectué par PCR (voir Fiche 24).

Cosmide

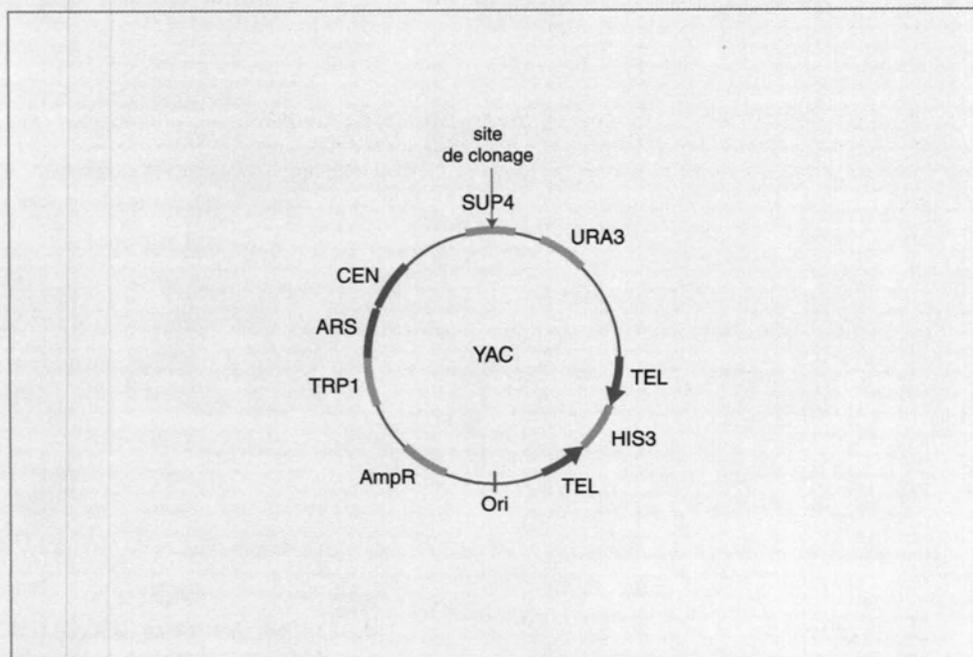
Un cosmide est un plasmide (avec son origine de répllication, ses gènes de sélection et son site de polyclonage) dans lequel on a inséré les séquences *cos* nécessaires à l'encapsidation dans des particules de bactériophage λ . Les fragments d'ADN exogène introduits dans les cosmides doivent avoir une taille comprise entre 35 et 45 kb. Ainsi, après ligature d'un cosmide préalablement linéarisé et d'un fragment d'ADN étranger de taille convenable (voir [Fiche 9](#)), l'ADN recombinant peut être encapsidé comme celui d'un phage, puis utilisé pour

“infecter” des bactéries. Une fois dans la bactérie, le cosmide se réplique comme un plasmide, car il ne possède pas les gènes du phage nécessaires à la production de nouvelles particules phagiques. Les cosmides sont des vecteurs utilisés pour construire des banques d'ADN génomique (voir Fiche 14).

DESCRIPTION D'UN YAC ET D'AUTRES GRANDS VECTEURS

Les recherches visant à caractériser la structure de génomes entiers nécessitent l'utilisation de vecteurs de clonage permettant l'insertion de grands fragments d'ADN. Cela permet de manipuler un nombre restreint de vecteurs recombinants tout en couvrant le génome entier. Les YACs et les BACs font partie de ces vecteurs acceptant de grands fragments.

Les YACs (*Yeast Artificial Chromosome* ou chromosome artificiel de Levure, *Saccharomyces cerevisiae*) sont des vecteurs construits à partir de séquences d'ADN chromosomique de Levure. Ils possèdent une origine de réplication (ARS, *Autonomous Replicating Sequence*), un site centromérique (CEN) et des séquences télomériques (TEL) leur permettant de se comporter comme un chromosome une fois introduit sous forme linéarisée dans la Levure. Plusieurs gènes de sélection codent pour des enzymes permettant de sélectionner les Levures ayant incorporé un vecteur viable (p. ex. TRP1 et URA3¹). Un site de clonage unique, situé dans le gène SUP4², permet l'introduction d'ADN étranger ainsi que la sélection des vecteurs recombinants. En effet, l'insertion d'un fragment d'ADN interrompt le gène SUP4, ce qui permet la détection des colonies ayant incorporé un vecteur recombinant par apparition d'une coloration rose. Ces vecteurs sont propagés dans la Levure (voir Fiche 10).



De très grands fragments d'ADN exogène (de 150 à 1 000 kb) peuvent être introduits dans les YACs. Actuellement, le plus grand insert cloné a une taille de 2 900 kb. Ces vecteurs sont donc de bons outils pour l'analyse des génomes complexes mais sont sujets à des instabilités.

D'autres vecteurs permettent également le clonage de grands fragments d'ADN : les *Bacterial Artificial Chromosome* (BAC) ou les *P1-derived Artificial Chromosome* (PAC). Ces vecteurs sont capables, tout comme les YACs, d'incorporer de grands fragments d'ADN. Les BACs se comportent comme des plasmides de très grande taille alors que les PAC se manipulent comme des bactériophages de type I. Les BACs sont de plus en plus utilisés : par rapport aux YACs, le clonage de l'ADN y est plus aisé et efficace, et les clones sont plus stables dans *E. coli* que dans la Levure. Cependant, la taille des fragments intégrés ne dépasse pas 300 kb.

Vecteurs	Forme du vecteur	Hôte	Taille de l'insert	Usage majeur
plasmides	ADN db circulaire	<i>E. coli</i>	0,1 à 10 kb	banque d'ADNc, sous-clonage
bactériophages λ	virus, ADN db linéaire	<i>E. coli</i>	10 à 20 kb	banque génomique et d'ADNc
cosmides	ADN db circulaire	<i>E. coli</i>	40 kb	banque génomique
phagemides	hybride plasmide/phage	<i>E. coli</i>	4 à 10 kb	banque d'ADNc, sous-clonage
YAC	Levure, chromosome artificiel	Levure	200 à 1 000 kb	banque génomique
BAC	chromosome artificiel de bactérie	<i>E. coli</i>	100 à 500 kb	banque génomique
PAC	virus, ADN circulaire	<i>E. coli</i>	80 kb	banque génomique

1. Le gène TRP1 codant pour la N-(5'-phosphoribosyl)-anthranilate isomérase est utilisé comme marqueur de sélection : les cellules possédant ce gène sont capables de se développer sur milieu sans tryptophane. Le gène URA3 codant pour l'orotidine-5'-phosphate décarboxylase permet de sélectionner les cellules possédant ce gène sur milieu sans uracile.

2. Le gène SUP4 code pour l'ARN-tyrosine.

CLONAGE MOLÉCULAIRE

Le clonage repose sur l'insertion d'un fragment d'ADN exogène dans un vecteur (plasmide, phagemide, bactériophage, cosmide, YAC, BAC, PAC...). Le vecteur de clonage est coupé par une enzyme de restriction qui reconnaît un site unique (généralement placé au site de polyclonage). Ce site étant unique, le vecteur est donc linéarisé et possède à chaque extrémité une partie de la séquence d'ADN reconnue par l'enzyme de restriction.

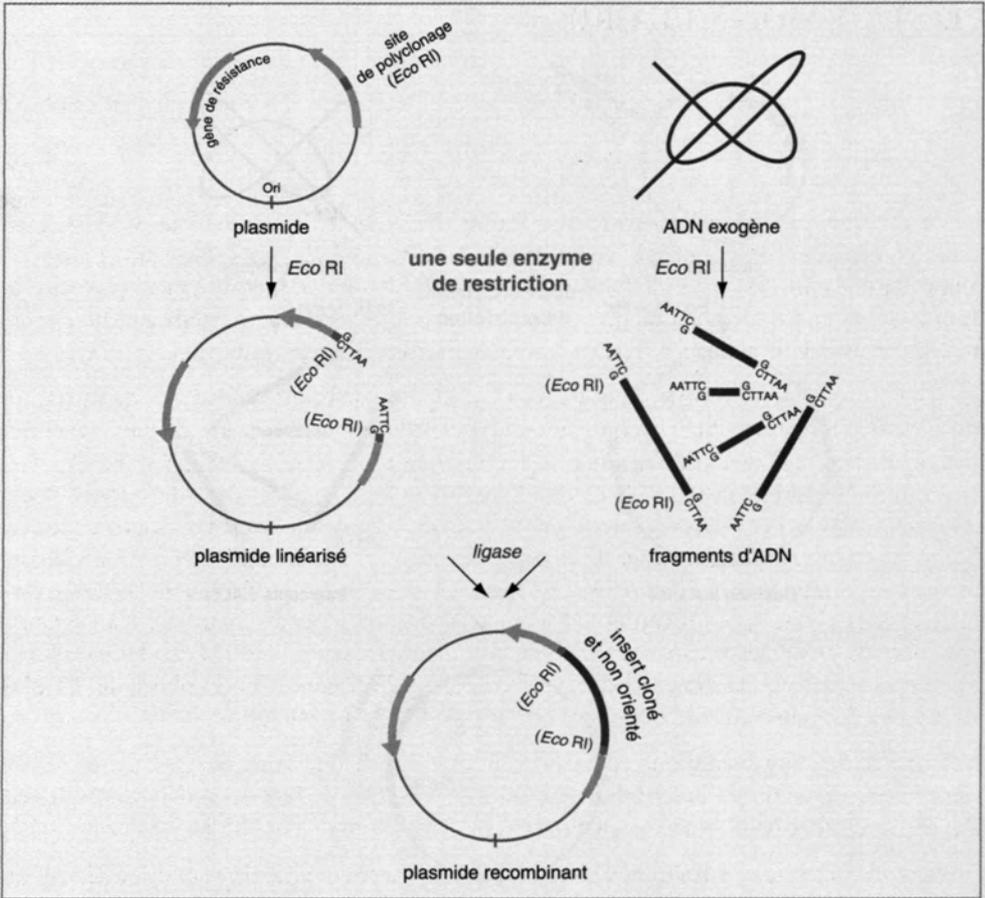
L'ADN exogène de l'organisme donneur est digéré par la même enzyme de restriction que celle utilisée pour la linéarisation du vecteur : les différents fragments obtenus possèdent donc également à leurs extrémités une partie de la séquence d'ADN reconnue par l'enzyme de restriction. Lorsque le plasmide ouvert, linéarisé, et les fragments de l'ADN donneur sont confrontés, il y a hybridation (formation de liaisons hydrogène) des extrémités cohésives complémentaires, suivant les lois de complémentarité des bases (A/T, C/G). Une enzyme, une ligase, permettant la formation d'une liaison covalente entre ces fragments d'ADN hybridés, est alors ajoutée pour ligaturer le fragment d'ADN étranger au plasmide. L'action de la ligase est de créer des liaisons phosphodiester entre les extrémités 5'-P et 3'-OH rendues adjacentes par l'hybridation des extrémités cohésives du fragment et du plasmide. Le plasmide obtenu, de nouveau circulaire, est dit recombinant s'il a intégré un insert.

Lorsque l'ADN plasmidique et l'insert sont digérés par une seule enzyme de restriction, l'insert peut s'intégrer au hasard dans les deux orientations. Il faut alors déterminer le sens d'insertion du fragment en séquençant par exemple le plasmide recombinant (voir Fiche 23).

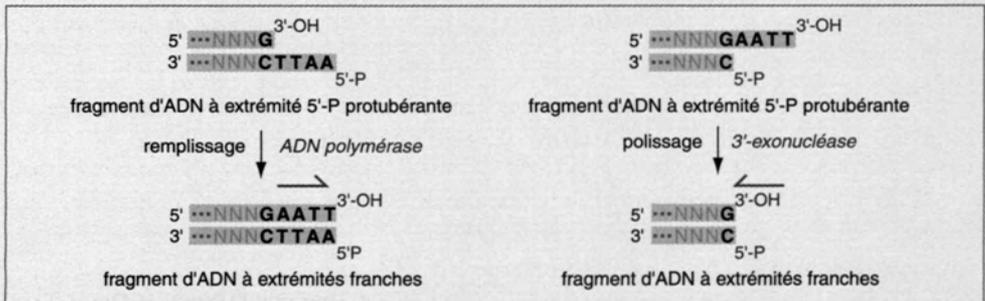
Lors d'un clonage réalisé après coupure par une seule enzyme de restriction, deux événements de ligature peuvent se produire : la ligature d'une molécule de plasmide avec une molécule d'insert ou la recircularisation du plasmide sur lui-même, sans insert. Cette dernière possibilité étant souvent majoritaire, il est nécessaire de tenter de s'en affranchir. Afin de favoriser l'insertion du fragment d'ADN étranger et donc d'enrichir la population en plasmide recombinant, la recircularisation du vecteur sur lui-même est empêchée par traitement à la phosphatase alcaline : les groupements phosphate présents aux extrémités 5'-P du vecteur linéarisé sont éliminés si bien que la ligase ne peut plus catalyser la formation de la liaison phosphodiester entre les deux extrémités déphosphorylées du plasmide. Par contre, elle peut catalyser la liaison entre une extrémité déphosphorylée du plasmide et une extrémité phosphorylée de l'insert.

Des cas moins favorables peuvent se présenter. Lorsque plasmide et fragment d'ADN étranger ont été coupés par une enzyme de restriction générant des bouts francs (voir Fiche 4), le principe est le même, mais l'efficacité de ligature est réduite. Néanmoins l'avantage de cette approche est de permettre la ligature de n'importe quelles extrémités, même si elles ne possèdent pas de séquences complémentaires. C'est le cas notamment pour le clonage en plasmide de fragments d'ADN obtenus par PCR¹ (Fiche 24). Dans le cas d'un clonage "à bouts francs", l'insertion du fragment d'ADN exogène peut avoir lieu dans les deux

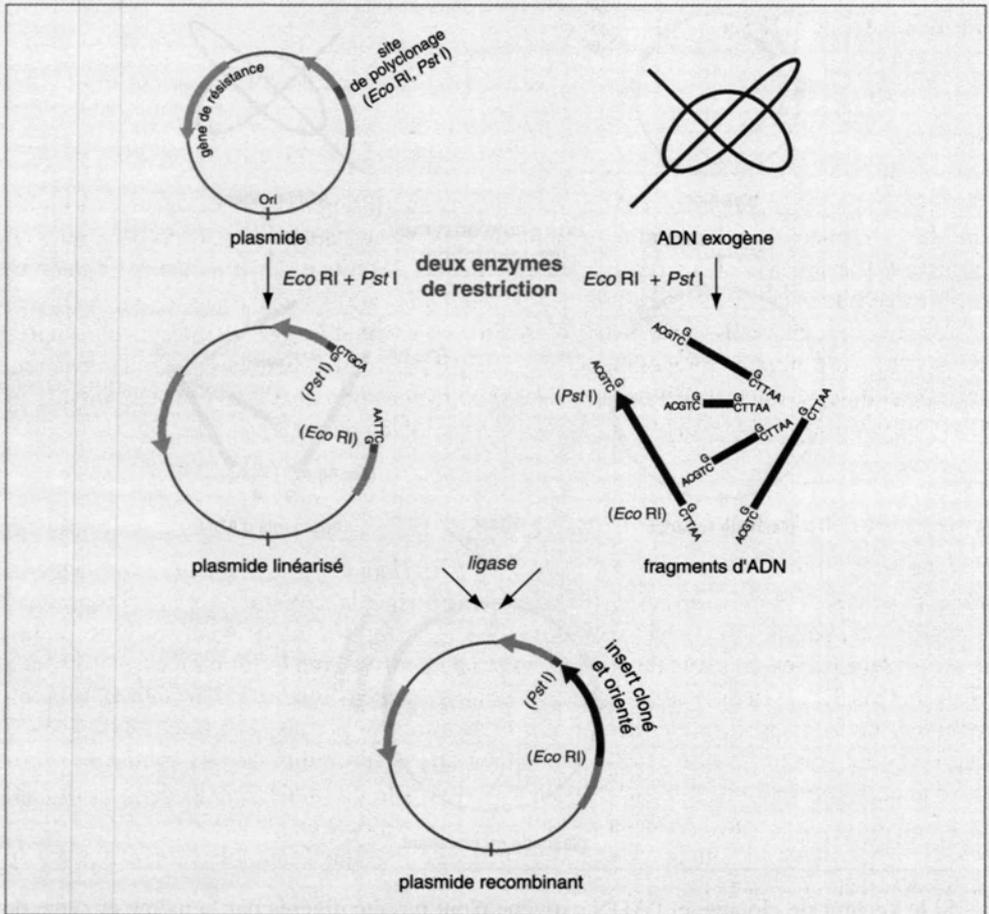
orientations au hasard. Il faudra alors déterminer le sens d'insertion du fragment en séquencant par exemple le plasmide recombinant (voir Fiche 23).



Si le vecteur de clonage et l'ADN exogène n'ont pas été digérés par la même enzyme de restriction et ne possèdent pas d'extrémités franches ou compatibles, il est alors indispensable de créer des extrémités franches pour insérer l'ADN exogène dans le vecteur. Il s'agit, par action d'une ADN polymérase, de remplir les extrémités 5'-P protubérantes, grâce à son activité 5' → 3' polymérase, ou de polir les extrémités 3'-OH protubérantes grâce à l'activité 3' → 5' exonucléase de l'enzyme.



Dans certains cas, l'insert et le plasmide sont digérés chacun par deux enzymes différentes. Le clonage est alors directionnel puisqu'un seul sens d'intégration de l'insert est possible.



1. La PCR génère des fragments d'ADN bicaténaires avec une base A protubérante en 5'-P. Il existe donc des vecteurs de clonage adaptés, linéarisés et possédant une base T (complémentaire de A) protubérante en 3'-OH. Le rendement du clonage en est fortement augmenté.

TRANSFORMATION GÉNÉTIQUE DES BACTÉRIES ET DES LEVURES

Après clonage de fragments d'ADN dans un vecteur, il faut intégrer ce vecteur recombinant dans un micro-organisme afin de l'amplifier, de le purifier ou d'exprimer le gène qu'il contient. Cette intégration se fait par transformation génétique. Le principe est d'intégrer dans une cellule bactérienne (généralement *Escherichia coli*) ou eucaryote (p. ex. la levure) les vecteurs recombinants, c'est-à-dire contenant un insert d'ADN exogène, obtenus après clonage. Lorsqu'un vecteur recombinant est intégré par transformation génétique dans une nouvelle cellule, il est capable de s'y répliquer de façon autonome. Ce faisant, il réplique son ADN, donc l'insert d'ADN qu'il contient au site de polyclonage.

Transformation des bactéries

Afin de faciliter la pénétration des molécules d'ADN au travers de la paroi et de la membrane plasmique, les bactéries en phase exponentielle de croissance sont "fragilisées" (comme p. ex. par un traitement au chlorure de calcium à 4 °C). Ces bactéries ainsi préparées, appelées bactéries compétentes, sont mises en contact avec la solution de plasmides à intégrer. Puis la membrane plasmique est momentanément rendue perméable (formation transitoire de micropores) soit par un choc thermique, soit par un choc électrique (électroporation¹). Les bactéries sont alors mises en culture sur un milieu gélosé contenant l'antibiotique correspondant au gène de résistance porté par le plasmide (voir Fiche 6). Par l'application de cet agent sélectif, seules les bactéries ayant intégré un plasmide peuvent se développer. Le plasmide recombinant va alors se multiplier dans la cellule, amplifiant par la même occasion la séquence d'ADN clonée.

Lorsqu'une bactérie est transformée par un plasmide ou un phagémide recombinant, elle se développe sur un milieu gélosé en formant une colonie. Chaque colonie correspond à un ensemble de bactéries identiques provenant de la division d'une seule bactérie. Le maintien du plasmide dans la bactérie recombinante est assuré par la pression de sélection exercée par la présence de l'antibiotique dans le milieu de culture.

Transformation des levures

L'introduction d'une molécule d'ADN hybride dans la Levure *Saccharomyces cerevisiae* peut s'effectuer de deux manières. D'une part, des sphéropastes (cellules libérées de leur paroi) sont préparés, puis incubés en présence de l'ADN plasmidique recombinant, de polyéthylène glycol (PEG) et de chlorure de calcium. D'autre part, des cellules entières, avec leur paroi, sont traitées par une solution saline alcaline comme l'acétate de lithium qui a pour effet de fragiliser les membranes, puis incubées avec l'ADN et le PEG. La pénétration de l'ADN dans les levures ainsi traitées est stimulée par un choc thermique. L'ADN étranger peut être

maintenu dans la cellule transformée sous une forme intégrée à l'ADN chromosomique par recombinaison homologue ou bien sous une forme répliquative indépendante du chromosome. Cette capacité de répliquaison indépendante lui est conférée par les séquences appelées ARS. Les molécules d'ADN ne possédant pas cette séquence ARS doivent s'intégrer au génome par recombinaison homologue pour que des levures transformées de façon stable soient obtenues.

1. L'électroporation donne des rendements supérieurs d'efficacité de transformation.

MARQUAGE DE L'ADN

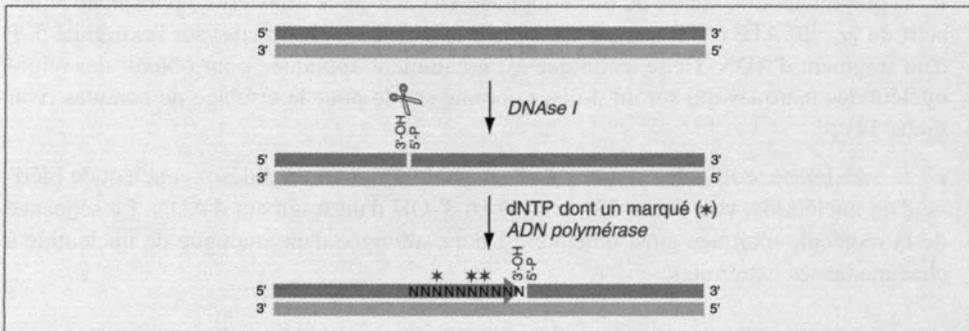
L'hybridation des acides nucléiques est un outil fondamental en biologie moléculaire pour détecter une séquence d'ADN d'intérêt. Elle est très fréquemment utilisée soit lors de criblages de banques d'ADNc ou génomique, soit lors d'études de l'organisation de régions spécifiques du génome (par *Southern blot*), soit lors de l'analyse de l'accumulation des transcrits dans les cellules. Le succès de ces techniques dépend de la possibilité d'obtenir des sondes d'ADN marquées à l'aide de nucléotides radioactifs ou modifiés chimiquement. Les techniques de marquage permettent d'obtenir des sondes marquées soit uniformément (marquage interne), soit à leurs extrémités (marquage aux extrémités).

Marquage interne

Marquage par translation de coupure (nick translation)¹

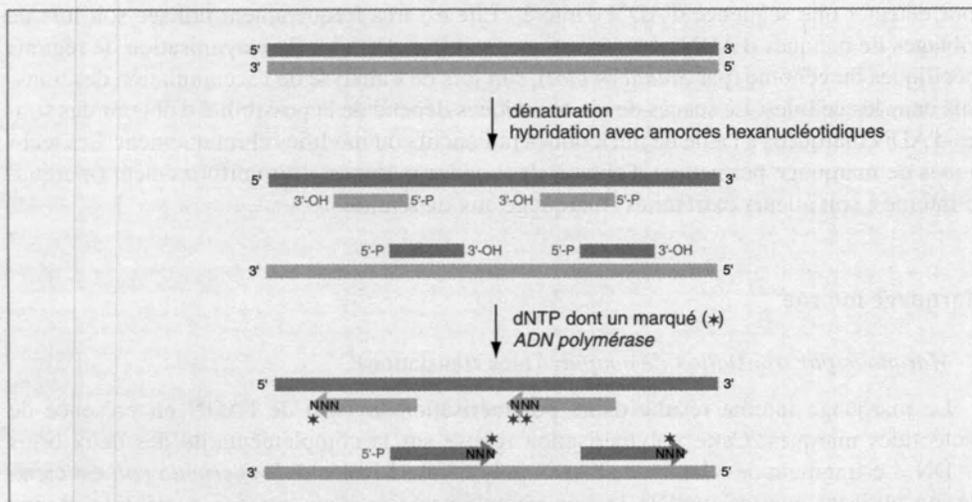
Le marquage interne résulte d'une polymérisation *in vitro* de l'ADN en présence de nucléotides marqués. Cette polymérisation repose sur la complémentarité des deux brins d'ADN. Le fragment de Klenow de l'ADN polymérase I isolée d'*Escherichia coli* est capable, en utilisant comme modèle le brin complémentaire, d'ajouter des nucléotides à une extrémité 3'-OH d'ADN, qui apparaît lorsque l'un des brins d'une molécule d'ADN est coupé. Cette enzyme présente également une activité 5'-exonucléasique qui élimine des nucléotides du côté 5'-P de la coupure. Ces deux actions conjuguées provoquent un mouvement du site de coupure (translation de coupure) le long de l'ADN. Avant l'action de la polymérase, les coupures sont réalisées par la DNase I. Cette endonucléase provoque des coupures simple brin et au hasard sur l'ADN.

Si l'on utilise des nucléotides marqués au ^{32}P à haute radioactivité spécifique, on obtient des molécules d'ADN bicaténares dont certaines portions sont fortement marquées (de l'ordre de 10^7 cpm/ μg).



Marquage par amorçage aléatoire (random priming)

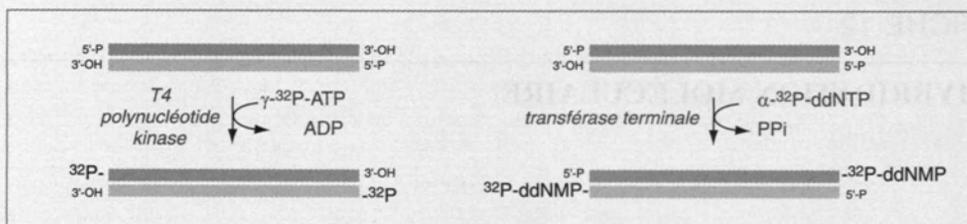
Un mélange d'hexanucléotides de séquence aléatoire est utilisé comme source d'amorces pour la synthèse *in vitro* d'ADN à partir d'une molécule d'ADN bicaténaire préalablement dénaturée (les deux brins sont séparés par la chaleur). Si le mélange d'hexanucléotides est suffisamment hétérogène, il se forme des hybrides sur n'importe quelle séquence de l'ADN matrice rendu monocaténaire. Ces structures hybrides bicaténares servent de point d'ancrage à une ADN polymérase (le fragment de Klenow de l'ADN polymérase I d'*E. coli* ou de plus en plus souvent une *Taq* polymérase) qui copie le brin complémentaire. L'utilisation de nucléotides marqués permet la synthèse d'une molécule d'ADN radioactive. Les sondes obtenues ont une radioactivité spécifique très élevée de l'ordre de 10^8 à 10^{10} cpm/ μ g.



Marquage aux extrémités

Ce type de marquage est utilisé par exemple pour de petites molécules de quelques dizaines de nucléotides, comme un oligonucléotide. Cette technique est également utile lorsque l'on désire une polarité du marquage comme pour les empreintes à la DNase I (voir Fiche 32). La limite de ce marquage est l'introduction d'un seul atome marqué par molécule d'ADN, d'où une activité spécifique assez faible. Deux enzymes différentes permettent de réaliser le marquage de chacune des extrémités :

- la polynucléotide kinase du bactériophage T4 catalyse le transfert du phosphate radioactif du $[\gamma\text{-}^{32}\text{P}]$ ATP (ATP marqué au ^{32}P sur le troisième phosphate) sur l'extrémité 5'-P d'un fragment d'ADN. Cette technique est notamment appliquée pour obtenir des oligonucléotides marqués qui seront utilisés comme sonde pour le criblage de banques (voir Fiche 14) ;
- la transférase terminale est quant à elle capable d'ajouter un didésoxynucléotide (dérivé d'un nucléotide, voir Fiche 23) marqué en 3'-OH d'un fragment d'ADN. La séquence de la molécule marquée ainsi obtenue est donc allongée d'un analogue de nucléotide à chacune de ses extrémités.



Marquage non radioactif

De plus en plus, les nucléotides radioactifs sont remplacés par des nucléotides marqués à l'aide de molécules non radioactives (d'où le terme de "sondes froides", par opposition aux "sondes chaudes"). Différents types de marquage existent, avec parfois une sensibilité de détection supérieure à celle des sondes radioactives, notamment grâce à l'utilisation de fluorochromes. Les différentes molécules associées aux nucléotides sont, entre autres, la digoxigénine, la biotine et la fluorescéine ou tout autre fluorochrome. Les principes de marquage sont les mêmes que ceux décrits dans les paragraphes précédents ; seules les techniques de détection diffèrent d'un type de marquage froid à l'autre. Par exemple, la digoxigénine est détectée à l'aide d'anticorps spécifiques et la biotine à l'aide de la streptavidine.

1. Ce type de marquage est de plus en plus remplacé par l'amorçage aléatoire.

HYBRIDATION MOLÉCULAIRE

Les techniques d'hybridation reposent sur les propriétés physico-chimiques des acides nucléiques : (i) la complémentarité des bases constitutives de l'ADN (A/T, G/C) et de l'ARN (A/U, G/C) ; (ii) la réversibilité du processus de séparation des deux brins d'une molécule d'ADN (dénaturation) et de ré-association des deux brins (re-naturation). Les dénaturations se traduisent par une dissociation des brins d'ADN (à 100 °C ou à pH alcalin) due à la rupture des liaisons hydrogène qui unissent les deux brins d'ADN. Lorsqu'il y a re-naturation (reformation des liaisons hydrogène entre les deux brins), on parle d'hybridation.

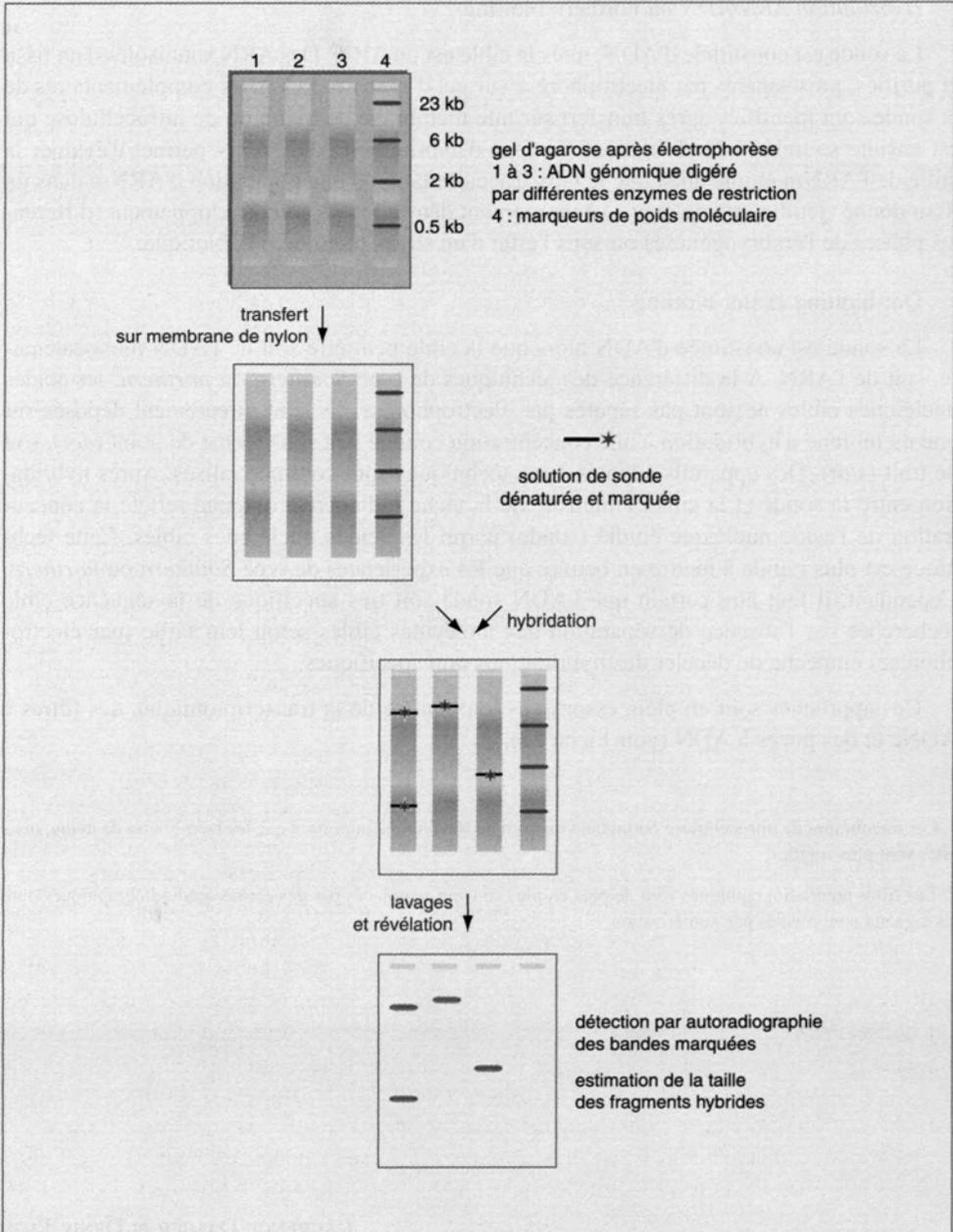
- Si les deux brins qui s'hybrident présentent exactement les mêmes séquences (100 % de similarité), l'hybride sera stable dans des conditions de contrainte importante (voir Fiche 2).
- Si les deux brins ne présentent pas de similarité dans leurs séquences, l'hybridation n'aura pas lieu.
- En situation intermédiaire (similarité partielle entre les deux brins), l'hybride formé sera instable en conditions de contrainte forte, mais stable en conditions de contrainte faible. Une condition de contrainte forte est par exemple un milieu proche de l'eau (concentration en sel très faible) et une température élevée (> à 65 °C). La température à laquelle les deux brins d'ADN se séparent est propre à chaque séquence (T_m). Plus la molécule étudiée est riche en bases G et C, plus la molécule bicaténaire sera stable, car l'appariement G/C est constitué de trois liaisons hydrogène, alors que l'appariement A/T n'en possède que deux.

L'hybridation moléculaire permet de repérer ou de rechercher une molécule d'ADN ou d'ARN spécifique dans une population hétérogène. Le principe est de marquer par un traceur radioactif ou chimique (voir Fiche 11) la séquence d'ADN connue et purifiée que l'on veut repérer dans une population. Ce fragment monocaténaire marqué constitue la sonde (voir Fiche 12). Les molécules cibles rendues monocaténaires (population hétérogène d'acides nucléiques dénaturés) sont au préalable fixées sur une membrane d'hybridation en nylon ou en nitrocellulose¹. La sonde monocaténaire et la cible dénaturée, donc également monocaténaire, sont mises en contact, la membrane étant incubée dans une solution contenant la sonde, dans des conditions permettant une hybridation. Lorsque la molécule sonde reconnaît son homologue dans la population de molécules cibles, il y a hybridation et l'hybride devient marqué par la présence de la sonde. Des lavages appropriés des membranes permettent d'éliminer alors toute hybridation non spécifique et de ne garder que les hybrides sonde/cible recherchés. Dans le cas d'un marquage radioactif, les hybrides sont repérés en plaçant la membrane au contact d'un film autoradiographique qui est ensuite développé comme un film photographique². Dans le cas d'un marquage chimique, l'hybride moléculaire est souvent révélé par un test colorimétrique sur membrane (voir Fiche 11).

Les différentes techniques d'hybridation

Hybridation ADN/ADN ou Southern blotting

La sonde est constituée d'une molécule d'ADN. La cible est également une molécule d'ADN comme un ADN nucléaire, un ADN plasmidique, un ADN phagique, une banque d'ADNc ou encore des fragments d'ADN cible séparés selon leur taille par électrophorèse sur gel d'agarose.



Ces ADN cibles sont dénaturés puis transférés par capillarité sur une membrane de nitrocellulose ou de nylon. Après mise en contact de la membrane avec la sonde marquée dans des conditions favorables à l'hybridation, des lavages sont effectués afin de faire apparaître après autoradiographie des bandes marquées correspondant à des fragments d'ADN complémentaire de la sonde. Une des applications du *Southern blot* est, par exemple, le diagnostic des maladies génétiques dans le cadre d'études de polymorphisme de fragments de restriction (voir Fiche 50).

Hybridation ARN/ADN ou northern blotting

La sonde est constituée d'ADN, mais la cible est un ARN. Des ARN sont isolés d'un tissu et purifiés, puis séparés par électrophorèse sur gel d'agarose. Les ARN complémentaires de la sonde sont identifiés après transfert sur une membrane de nylon ou de nitrocellulose qui est ensuite soumise à l'hybridation. Ce type d'hybridation ADN/ARN permet d'évaluer la taille de l'ARNm étudié ainsi que le taux d'accumulation d'une population d'ARNm dans un tissu donné (feuille, racine, tige...) à un moment déterminé de son développement (différentes phases de l'embryogenèse) ou sous l'effet d'un stress biotique ou abiotique.

Dot-blotting et slot-blotting

La sonde est constituée d'ADN alors que la cible peut être soit de l'ADN monocaténaire, soit de l'ARN. A la différence des techniques de type *Southern* ou *northern*, les acides nucléiques cibles ne sont pas séparés par électrophorèse. Ils sont directement déposés sur une membrane d'hybridation à une concentration connue soit sous forme de point (*dot*), soit de trait (*slot*). Des appareils adaptés à ces techniques sont commercialisés. Après hybridation entre la sonde et la cible, l'intensité de la tache radioactive obtenue reflète la concentration de l'acide nucléique étudié (sonde) parmi les acides nucléiques cibles. Cette technique est plus rapide à mettre en oeuvre que les expériences de type *Southern* ou *northern*. Cependant, il faut être certain que l'ADN sonde soit très spécifique de la séquence cible recherchée car l'absence de séparation des molécules cibles selon leur taille (par électrophorèse) empêche de déceler des hybridations non spécifiques.

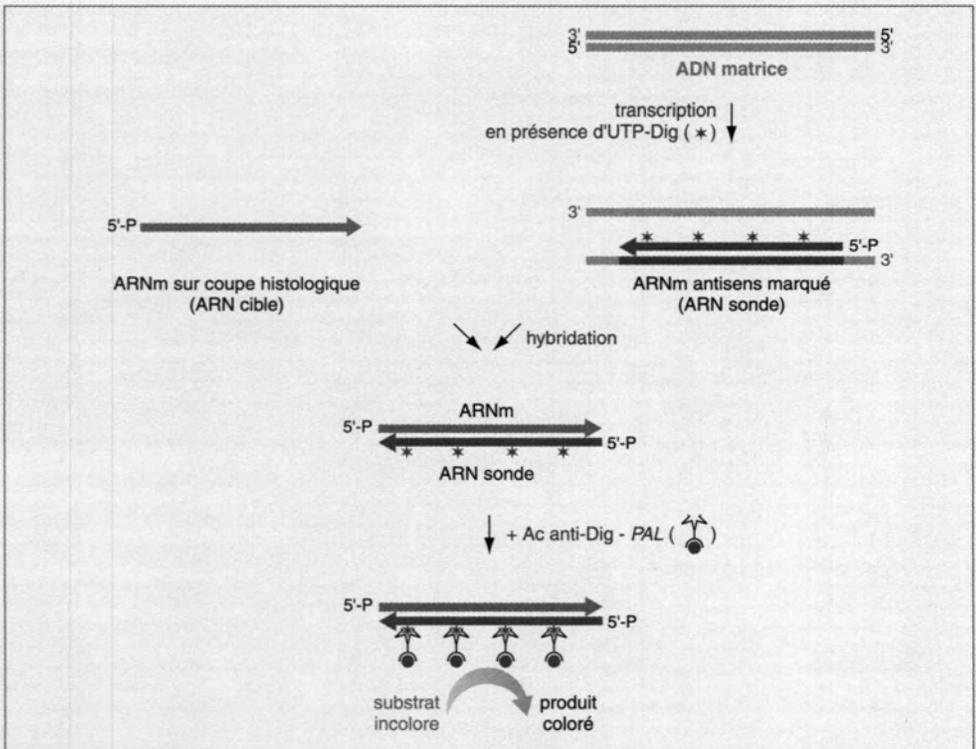
Ces approches sont en plein essor dans le domaine de la transcriptomique, des filtres à ADNc et des puces à ADN (voir Fiche 22).

1. Les membranes de nitrocellulose permettent un bruit de fond moins important que les membranes de nylon, mais elles sont plus fragiles.
2. Les films autoradiographiques sont de plus en plus souvent remplacés par des écrans autoradiographiques dont les signaux sont révélés par numérisation.

HYBRIDATION *IN SITU* DES ARNm

Un tissu est composé de plusieurs types de cellules, chacun exprimant un lot de gènes. Ainsi, deux cellules adjacentes dans un même tissu peuvent ne pas exprimer le même gène. Il est parfois nécessaire de connaître la localisation du transcrite du gène étudié afin d'affiner la connaissance de son mode d'expression.

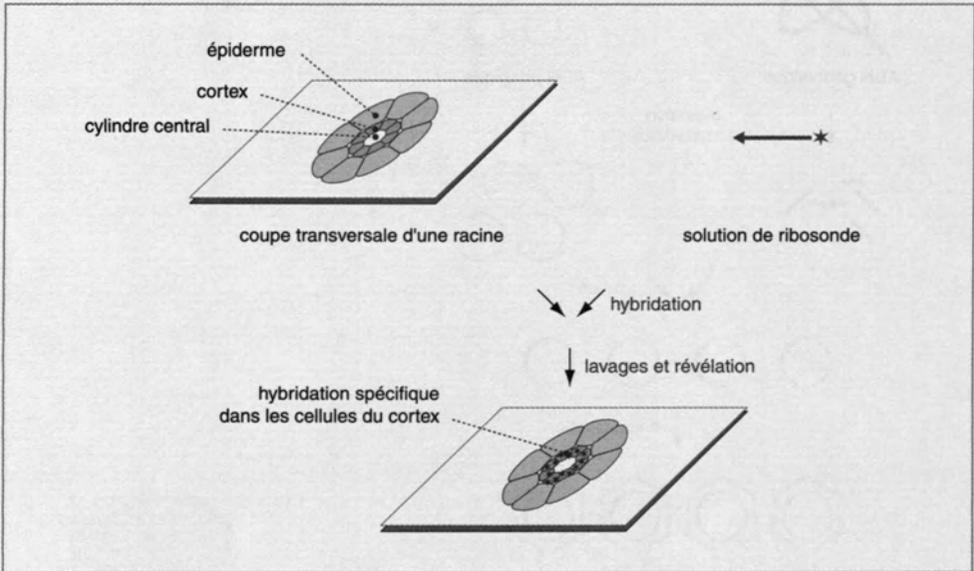
Le principe de l'hybridation *in situ* des ARNm consiste à déterminer sur coupe les cellules d'un tissu accumulant une population d'ARNm donnée. Cette technique repose sur une hybridation moléculaire. Les molécules cibles sont les ARNm présents dans les cellules. Le tissu ou l'organe est d'abord fixé chimiquement, inclus dans de la paraffine, débité en coupes fines (de 5 à 10 µm d'épaisseur environ), puis mis au contact de la sonde. Les sondes utilisées sont principalement des ARN marqués (ribosondes). Les ribosondes sont réalisées par transcription *in vitro* (voir Fiche 28) du brin complémentaire de l'ARNm cible en présence d'UTP marqué. Cet UTP est marqué soit radioactivement (généralement au $[^{35}\text{S}]$ UTP et plus récemment au $[^{33}\text{P}]$), soit chimiquement (digoxigénine).



Dans le cas d'une sonde radioactive, une émulsion photographique liquide et translucide est alors coulée sur la coupe après hydratation et lavages. Cette émulsion translucide va se solidifier en suivant tous les contours de la coupe. La radioactivité peut imprimer localement le film et la coupe est examinée en microscopie optique. Il y a superposition des structures cellulaires et des précipités de grains d'argent.

Le marquage chimique des ribosondes se fait généralement à l'aide d'UTP couplé à la digoxigénine (Dig), un haptène stéroïde. Après hybridation et lavages, l'hybride ARN/ARN est détecté à l'aide d'anticorps anti-digoxigénine couplés à la phosphatase alcaline (PAL). L'activité enzymatique portée par l'anticorps est alors détectée par addition d'un substrat formant un précipité coloré. L'observation des lames au microscope permet de visualiser ce précipité au niveau cellulaire et d'en déduire la localisation du transcrite.

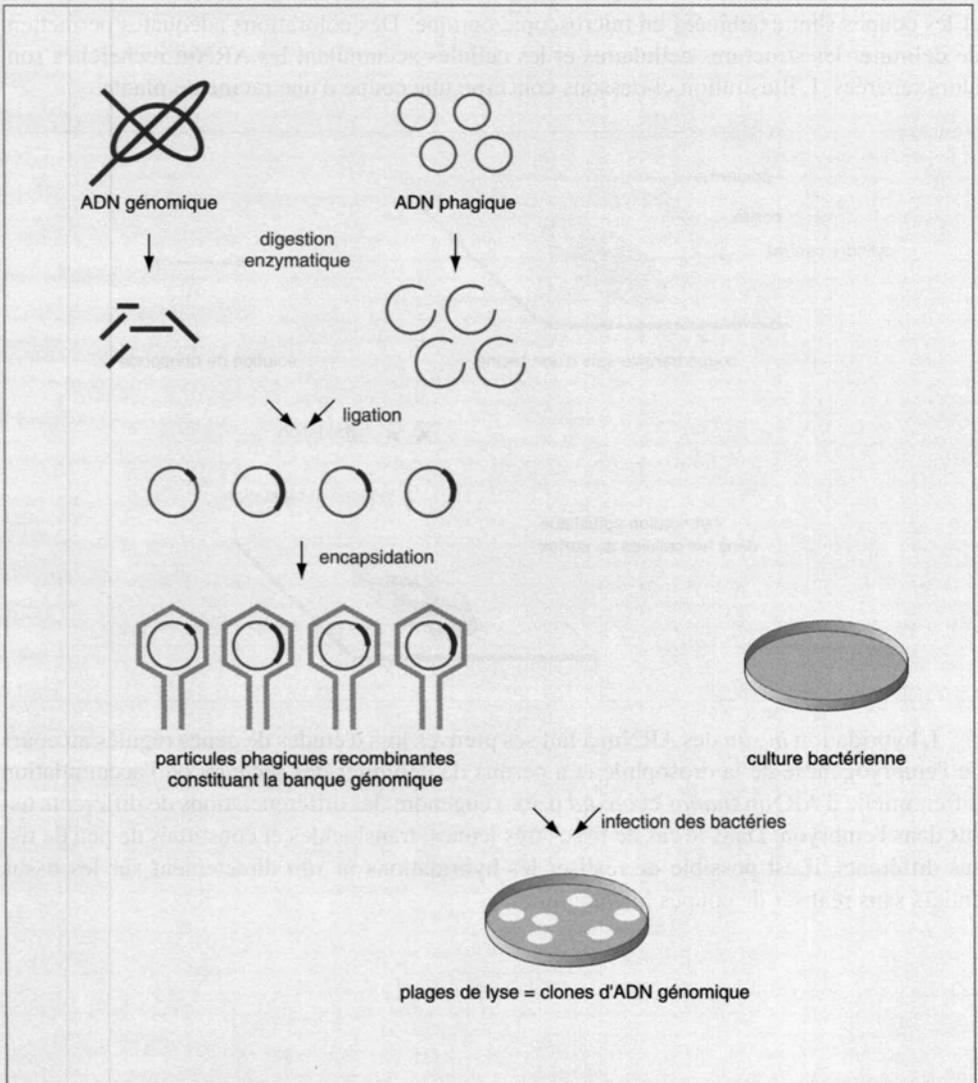
Comme lors des hybridations moléculaires sur filtre (voir Fiche 12), des lavages appropriés permettent d'éliminer toute hybridation non spécifique. Les hybrides sont alors révélés et les coupes sont examinées en microscopie optique. Des colorations adéquates permettent de délimiter les structures cellulaires et les cellules accumulant les ARNm recherchés sont alors repérées. L'illustration ci-dessous concerne une coupe d'une racine de plante.



L'hybridation *in situ* des ARNm a fait ses preuves lors d'études de gènes régulés au cours de l'embryogenèse de la drosophile et a permis de délimiter des secteurs où l'accumulation différentielle d'ARNm (*nanos* et *bicoid* p. ex.) engendre des différenciations de différents tissus dans l'embryon. Dans le cas de tissus très jeunes, translucides et constitués de peu de tissus différents, il est possible de réaliser les hybridations *in situ* directement sur les tissus entiers sans réaliser de coupes histologiques.

CONSTRUCTION D'UNE BANQUE D'ADN GÉNOMIQUE

L'obtention d'une banque génomique est nécessaire lorsque l'on veut par exemple caractériser une séquence génomique particulière, établir la carte physique d'un génome ou séquencer à grande échelle des portions ou la totalité d'un génome...



Une banque d'ADN génomique est un ensemble de vecteurs recombinants contenant chacun un morceau différent du génome de l'espèce étudiée. Le principe consiste à : (i) isoler l'ADN génomique de l'espèce d'intérêt ; (ii) digérer cet ADN par une enzyme de restriction permettant de libérer des fragments de restriction de taille compatible avec le vecteur choisi¹ ; (iii) cloner ces fragments dans un vecteur de clonage ; (iv) intégrer ces vecteurs recombinants dans un micro-organisme afin de les multiplier. Les vecteurs phagiques (voir Fiche 7) seront mieux adaptés au clonage d'une séquence génomique particulière alors que les cosmides (voir Fiche 7) et les YACs ou BACs (voir Fiche 8) seront choisis dans un but de caractérisation d'un génome. La banque obtenue est ainsi constituée de millions de clones recombinants.

L'ADN génomique étant considéré comme identique dans toutes les cellules d'un même individu, la provenance tissulaire de l'ADN n'a pas d'importance. Par exemple, une banque génomique faite à partir d'ADN de racines d'un individu est théoriquement identique à une banque génomique fabriquée à partir d'ADN de feuilles du même individu.

1. Il est parfois nécessaire d'éliminer les fragments d'ADN trop petits pour ne pas saturer la banque de clones de petite taille. Pour ce faire, on réalise une digestion ménagée de l'ADN génomique afin d'obtenir des fragments de taille satisfaisante.

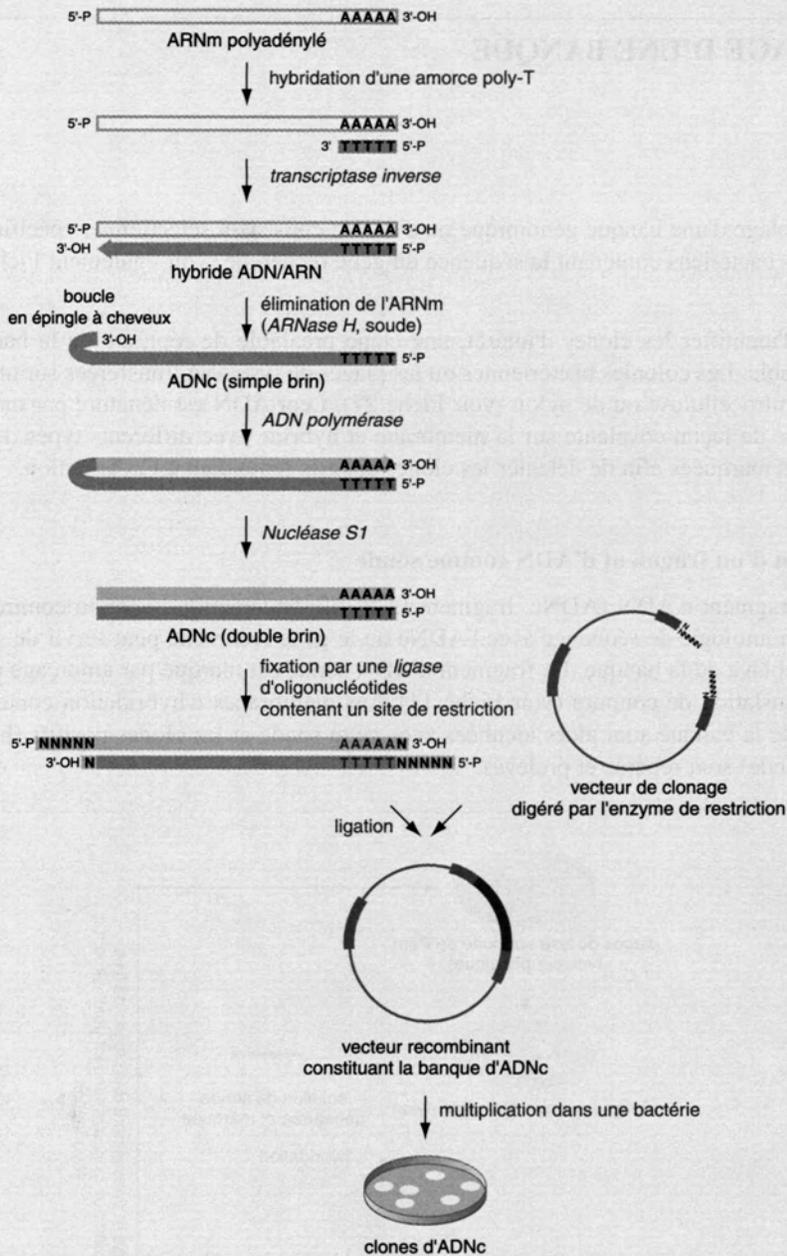
CONSTRUCTION D'UNE BANQUE D'ADNc

Les populations d'ARNm accumulées dans un tissu donné sont représentatives de ce tissu. Il est donc de première importance de pouvoir caractériser les gènes exprimés dans un organe déterminé. Mais la manipulation des ARNm est délicate (faible quantité, sensibilité aux nucléases) et il est nécessaire de transformer ces ARNm en ADN bicaténaires aisément manipulables et adaptés à des clonages dans des vecteurs bactériens. Une banque d'ADNc (ADN complémentaire) est donc représentative des populations d'ARNm présentes dans un tissu donné à un stade déterminé de son développement. Contrairement à la banque génomique, une banque d'ADNc sera spécifique d'un tissu. Une banque d'ADNc peut être considérée comme "une photographie instantanée" des populations d'ARNm représentées dans un organe.

La construction d'une banque d'ADNc nécessite de nombreuses étapes. Le principe repose sur : (i) l'extraction d'ARN et parfois la purification des ARNm polyadénylés de l'organe (par exemple par chromatographie d'affinité sur une colonne polyT)¹ ; (ii) la copie de ces ARNm en ADN complémentaires par l'action d'une transcriptase inverse ; (iii) l'élimination spécifique des ARNm par l'ARNase H ou la soude ; (iv) la synthèse du second brin d'ADN par une ADN polymérase ; (v) la ligature d'oligonucléotides (adaptateurs) pour créer des sites de restriction ; (vi) la ligature dans un vecteur de clonage (plasmide ou phage) ; (vii) l'intégration des vecteurs recombinants dans une bactérie. Les vecteurs de clonage utilisés pour la construction de banques d'ADNc sont soit les plasmides, soit les phages (voir Fiches 6 et 7). Les étapes de synthèse du second brin d'ADN et de ligature d'adaptateurs sont de plus en plus souvent réalisées à l'aide de la PCR¹ (voir Fiche 24).

La synthèse du brin d'ADNc est amorcée généralement par fixation d'une courte séquence polyT sur l'extrémité polyA de l'ARNm. Une autre approche consiste à utiliser comme amorce pour la synthèse du brin d'ADNc par la transcriptase inverse un mélange d'hexanucléotides synthétiques correspondant à de nombreuses séquences différentes et s'hybridant au hasard sur l'ARNm (voir Fiches 17 et 51). Cependant, comme par définition l'amorçage de la synthèse de l'ADNc se fait sur des parties internes des ARNm, les ADNc obtenus ne correspondent qu'à des fragments d'ARNm.

1. Dans les cas pour lesquels il est difficile d'extraire de grandes quantités d'ARNm, la construction d'une banque d'ADNc peut se faire à partir d'ARN totaux. Il faut savoir que les ARNm ne représentent que quelques % des ARN cellulaires constitués essentiellement d'ARN ribosomiques. La PCR est particulièrement adaptée pour la construction de banques d'ADNc à partir de très petites quantités d'ARN.



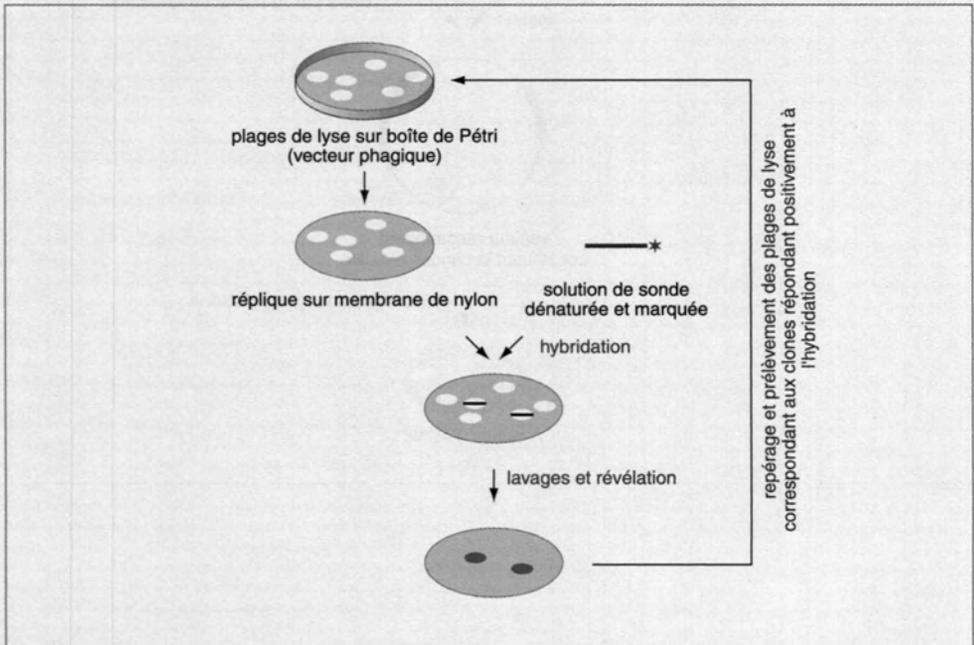
CRIBLAGE D'UNE BANQUE

Le criblage d'une banque génomique ou d'ADNc consiste à sélectionner spécifiquement des clones bactériens contenant la séquence du gène recherché (voir également Fiches 17 et 20).

Afin d'identifier les clones d'intérêt, une étape préalable de réplique de la banque est indispensable. Les colonies bactériennes ou les plages de lyse sont transférées sur une membrane de nitrocellulose ou de nylon (voir Fiche 27). Leur ADN est dénaturé par un bain de soude, fixé de façon covalente sur la membrane et hybridé avec différents types de sondes nucléiques marquées afin de détecter les clones positifs répondant à l'hybridation.

Utilisation d'un fragment d'ADN comme sonde

Tout fragment d'ADN (ADNc, fragment PCR, oligonucléotide...) connu comme possédant une homologie de séquence avec l'ADNc ou le gène recherché peut servir de sonde en vue du criblage de la banque. Le fragment d'ADN sonde est marqué par amorçage aléatoire ou par translation de coupure (voir Fiche 11). Les membranes d'hybridation contenant les vecteurs de la banque sont alors incubées avec cette sonde et les clones positifs (hybridant avec la sonde) sont repérés et prélevés.

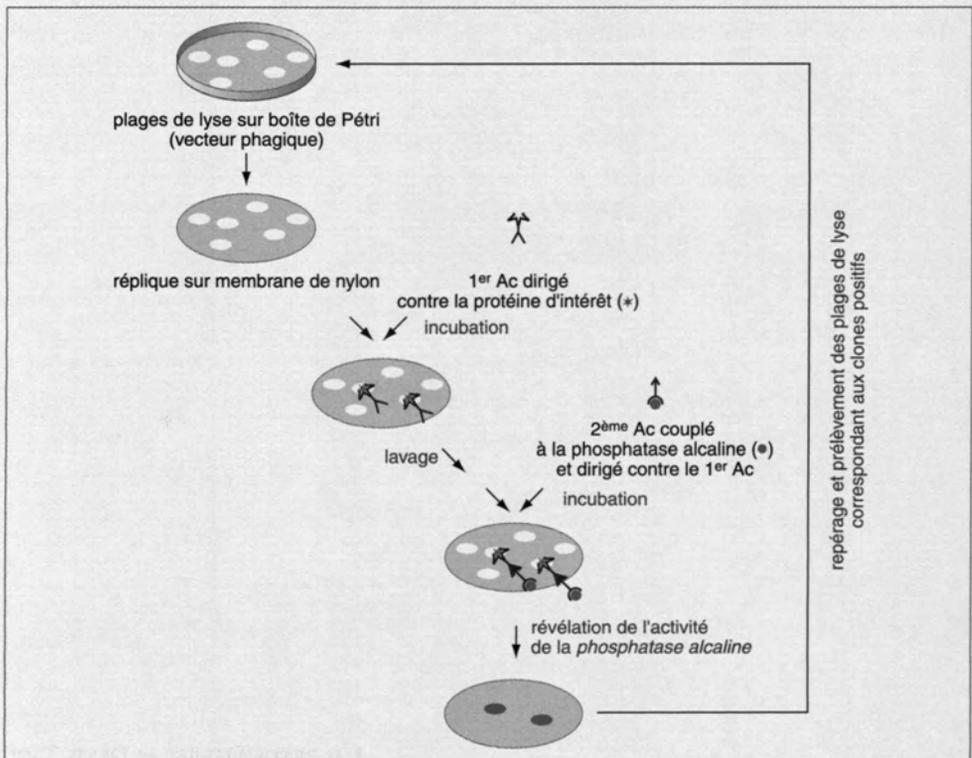


Utilisation d'un oligonucléotide de synthèse comme sonde

Connaissant la séquence de la protéine dont on veut isoler le gène, il est possible de concevoir une séquence nucléotidique correspondante en suivant les règles du code génétique. Du fait de la dégénérescence du code génétique¹, plusieurs séquences nucléotidiques sont obtenues. Le mélange d'oligonucléotides (généralement de l'ordre de 20 bases) est synthétisé, marqué radioactivement (voir Fiche 11) et hybridé avec les clones fixés sur le filtre.

Criblage par anticorps

Parfois aucune connaissance sur la séquence de la protéine ou du gène à cloner n'est disponible. Une approche biochimique d'un phénomène peut aboutir à la purification d'une protéine et l'obtention d'anticorps spécifiques de cette protéine. Ces anticorps peuvent être utilisés pour isoler un ADNc correspondant. Ceci suppose que la banque d'ADNc a été construite dans un vecteur d'expression : le site d'insertion des ADNc dans le vecteur se trouve en aval d'un promoteur bactérien. Ainsi, toutes les bactéries de la banque contenant ces vecteurs recombinants sont capables de synthétiser la protéine correspondant à la séquence d'ADNc². Les membranes d'hybridation, contenant la banque d'ADNc, sont incubées avec les anticorps dirigés contre la protéine dont on veut cloner l'ADNc. Les bactéries qui possèdent l'ADNc correspondant vont pouvoir synthétiser la protéine qui sera reconnue par les anticorps. Pour cela, il faut que l'ADNc soit dans une bonne phase de lecture afin que les codons soient lus correctement. Le complexe stable antigène-anticorps est ensuite révélé par une réaction colorimétrique à l'aide d'un second anticorps couplé à la phosphatase alcaline. Les clones répondant à cette hybridation sont alors repérés sur la boîte de Pétri et prélevés.



Recherche de facteurs de transcription

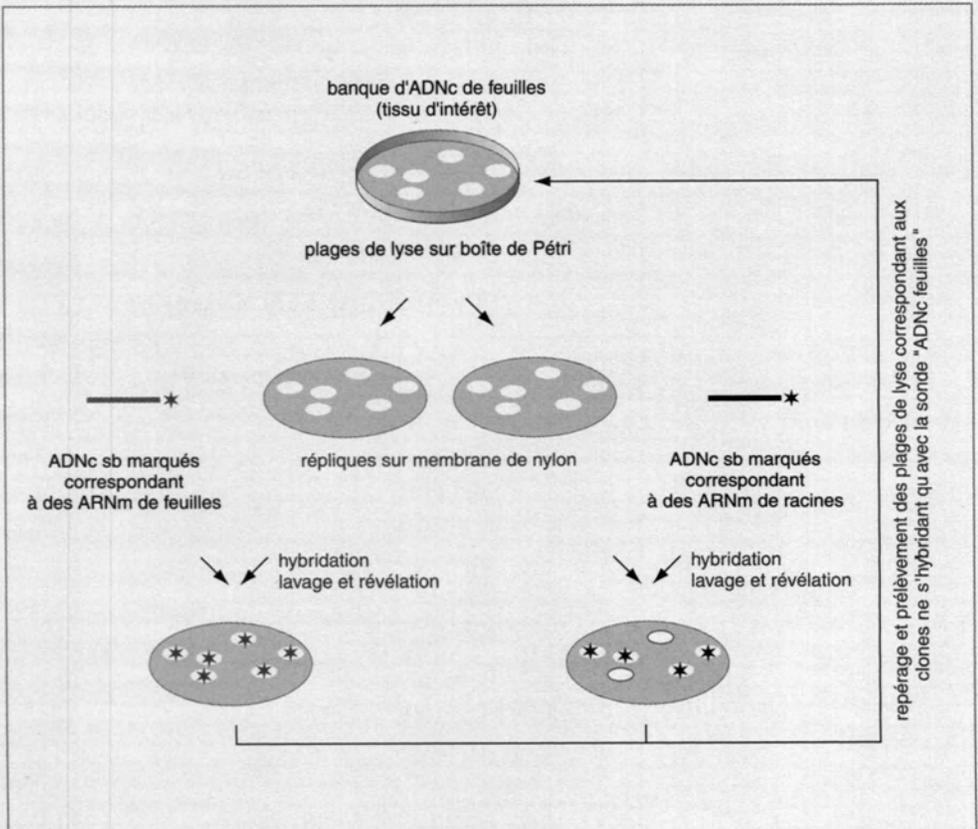
L'obtention de banques d'ADNc dans des vecteurs d'expression a permis de développer des techniques afin de cloner des gènes codant pour des protéines affines de séquences d'ADN régulatrices. Ces protéines sont des facteurs de régulation de la transcription. La banque d'expression est criblée avec un fragment d'ADN marqué radioactif, dont il a été démontré au préalable qu'il pouvait se complexer à des protéines nucléaires (voir Fiches 31 et 32). Si une bactérie de la banque contient un vecteur recombinant exprimant une protéine affine de ce fragment d'ADN, elle sera repérée par l'hybridation. Les ARNm correspondant à des protéines de régulation sont souvent accumulés en faible quantité dans les tissus. Les ADNc de ces ARNm sont donc faiblement représentés dans les banques et il est nécessaire de cribler un grand nombre de clones avant d'obtenir un résultat (environ 10^6 clones).

1. Un acide aminé est codé par plusieurs codons différents : c'est la dégénérescence du code génétique. Exemple : l'alanine est codée par les quatre codons GCU, GCC, GCA et GCG.

2. Il faut que la synthèse de la protéine se fasse dans la bonne phase de lecture. Les vecteurs d'expression utilisés sont généralement des phages de type λ (λ gt11). Le code génétique étant lu 3 bases par 3, il y a donc 3 phases de lecture possibles. Mais l'ADNc peut être cloné dans le bon sens (début du brin codant vers le promoteur bactérien) ou dans le mauvais sens (début du brin codant à l'opposé du promoteur). Au total, on a donc 1 chance sur 6 que l'ADNc soit dans la bonne phase de lecture. Si le clonage des ADNc est directionnel (les sites de restriction bordant les extrémités 5' et 3' de l'ADNc sont différents et permettent de s'assurer de la bonne orientation de l'ADNc par rapport au promoteur bactérien), nous avons alors 1 chance sur 3 que l'ADNc soit en phase de lecture.

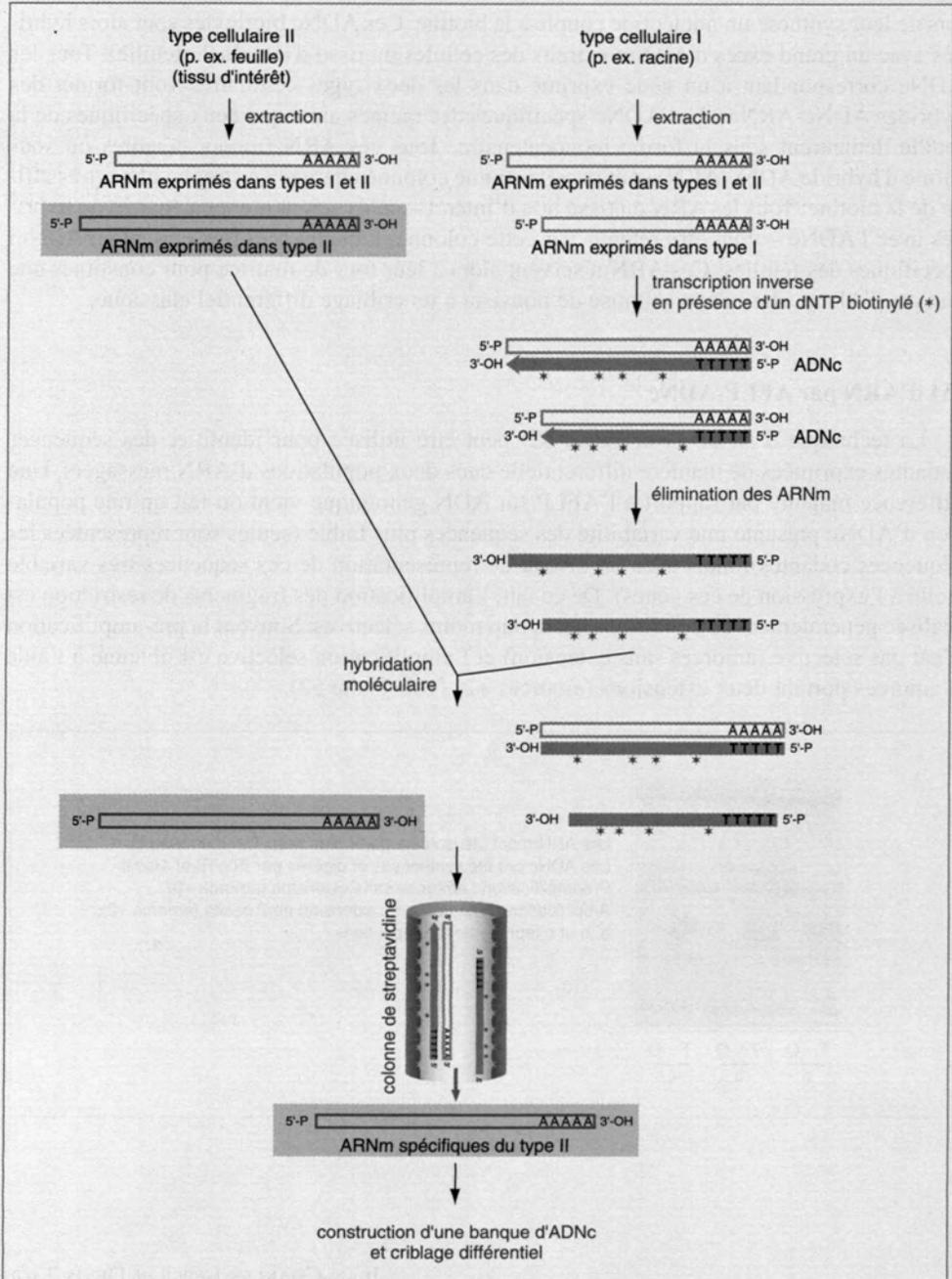
CRIBLAGE DIFFÉRENTIEL : BANQUES SOUSTRAITES, AFLP-ADNc

Le criblage différentiel est une technique qui permet de cloner des ADNc correspondant à des ARNm spécifiques d'un tissu donné, d'un stade de développement déterminé ou sous l'effet d'un stress. Le principe consiste tout d'abord à construire une banque d'ADNc à partir d'ARNm du traitement dont on veut isoler des ADNc spécifiques (p. ex. feuilles). Puis, des ADNc monocaténaire (simple brin ou sb) du tissu d'intérêt et d'un autre tissu (p. ex. racines) sont synthétisés par transcription inverse des ARN extraits des deux types de tissus. Ces ADNc sont ensuite marqués pour servir de sonde. La banque est alors criblée parallèlement avec les deux sondes. Certains clones d'ADNc de la banque vont s'hybrider avec les deux sondes : ils correspondent à des ARNm accumulés dans les deux types de tissu. Par contre, certains ADNc ne vont s'hybrider qu'avec la sonde d'ADNc du tissu d'intérêt : ils correspondent à des ARNm spécifiques de ce tissu. Ces clones sont alors purifiés et caractérisés.



Cette technique de criblage différentiel n'est efficace que pour isoler les ARNm les plus abondants. Plusieurs variantes ont été développées permettant d'isoler des ARNm rares, comme les banques soustraites (voir aussi Fiches 18 à 20).

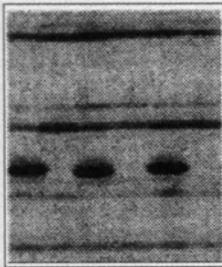
Utilisation d'une banque d'ADNc soustraite



La présence d'ARNm abondants dans un tissu masque les ARNm rares qui peuvent cependant jouer des rôles prépondérants dans le mécanisme étudié. Afin d'augmenter les chances de cloner des ADNc correspondant à des ARNm rares, on tente, avant d'effectuer le criblage différentiel, d'éliminer de la banque les ARNm abondants. C'est ce que l'on appelle un épuisement ou une soustraction d'ADNc. Plusieurs approches sont possibles. Par exemple, on peut produire les ADNc monocaténaire des ARNm des cellules d'un tissu différent (la racine) de celui d'intérêt (la feuille). Ces ADNc sont modifiés, en ajoutant par exemple lors de leur synthèse un nucléotide couplé à la biotine. Ces ADNc biotinylés sont alors hybridés avec un grand excès d'ARNm extraits des cellules du tissu d'intérêt (la feuille). Tous les ADNc correspondant à un gène exprimé dans les deux types cellulaires vont former des hybrides ADNc-ARNm. Les ADNc spécifiques des racines ainsi que ceux spécifiques de la feuille demeurent sous la forme monocaténaire. Tous ces ARN (monocaténaire ou sous forme d'hybride ADNc/ARN) sont passés sur une colonne chargée en streptavidine, très affine de la biotine. Tous les ARN du tissu non d'intérêt – qu'ils soient monocaténaire ou hybridés avec l'ADNc – vont être retenus par cette colonne. L'éluant sera très enrichi en ARNm spécifiques des feuilles. Ces ARNm servent alors à leur tour de matrice pour constituer une banque d'ADNc elle-même soumise de nouveau à un criblage différentiel classique.

Tri d'ARN par AFLP-ADNc

La technique d'AFLP (voir Fiche 52) peut être utilisée pour identifier des séquences codantes exprimées de manière différentielle dans deux populations d'ARN messagers. Une différence majeure par rapport à l'AFLP sur ADN génomique vient du fait qu'une population d'ADNc présente une variabilité des séquences plus faible (seules sont représentées les séquences codantes), mais avec un niveau de représentation de ces séquences très variable (lié à l'expression de ces gènes). De ce fait, l'amplification des fragments de restriction est réalisée généralement en conditions beaucoup moins sélectives. Souvent la pré-amplification n'est pas sélective (amorces sans extension) (amorces +0) et l'amplification sélective est obtenue à l'aide d'amorces portant deux extensions (amorces +2) (voir Fiche 52).



Les ARNm ont été isolés à partir d'un tissu T et d'un tissu O.
 Les ADNc ont été synthétisés et digérés par *Eco* RI et *Mse* I.
 Préamplification : amorces sans extension (amorces +0).
 Amplification : amorces avec extension de 2 bases (amorces +2).
 a, b et c représentent 3 répétitions.

CRIBLAGE DIFFÉRENTIEL PAR DD RT-PCR : TRI D'ARNm (*Differential Display RT-PCR*)

La technique de tri d'ARNm (ou *differential display*) est une variante du criblage différentiel adaptée à l'identification d'ARNm rares (voir Fiche 17). Elle est fondée sur trois étapes majeures : (i) transcription inverse à partir d'ARNm aboutissant à l'obtention d'ADNc monocaténares (voir Fiche 27) ; (ii) amplification par PCR de ces ADNc ; (iii) séparation électrophorétique des fragments d'ADNc bicaténares avant leur clonage. Ces trois étapes sont réalisées simultanément et en parallèle sur des ARNm isolés de différents tissus ou traitements. La comparaison des gels électrophorétiques permet d'identifier des bandes d'ADNc différenciellement accumulés.

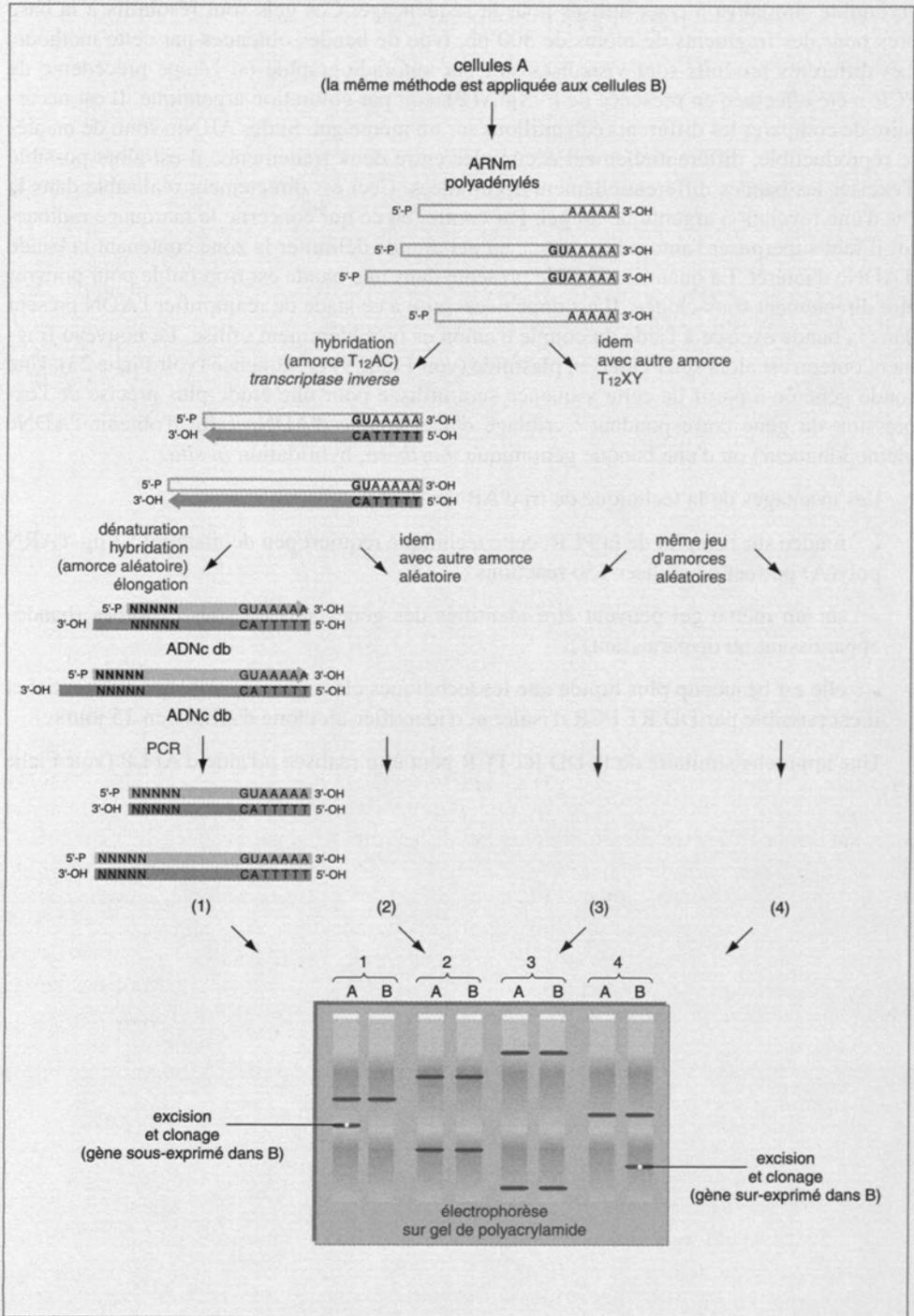
Transcription inverse

Après extraction des ARN totaux et éventuellement une sélection des ARNm polyadénylés (dans la figure à partir des cellules A et B que l'on veut comparer), une première étape consiste à synthétiser des ADNc monocaténares. Cette transcription inverse est effectuée en utilisant une amorce de type $T_{12}XY$, complémentaire à la queue poly(A). X peut correspondre à n'importe quelle base sauf une thymidine et Y à toute base. Ainsi, il existe 12 combinaisons possibles d'amorces $T_{12}XY$, toutes complémentaires de familles d'ARNm différents. Par exemple, l'amorce $T_{12}AC$ sera complémentaire de tous les ARNm se terminant par GUA_{12} , alors que l'amorce $T_{12}AG$ sera complémentaire de tous les ARNm se terminant par CUA_{12} . C'est le choix des amorces qui déterminera le tri des ARNm analysés. Ce tri est nécessaire pour améliorer la résolution du gel d'électrophorèse par la suite. Si l'on n'opérait pas de tri (utilisation d'une amorce T_{12} complémentaire de tous les ARNm) la quantité d'ADNc obtenus serait trop grande et la résolution du gel d'électrophorèse serait mauvaise, trop de bandes apparaissant. Cette étape de tri permet donc de sélectionner une population correspondant en moyenne au douzième de la population d'ARN poly(A) présente dans le type cellulaire étudié. Lorsqu'une amorce est choisie, c'est la même qui est utilisée sur les ARNm isolés des différents tissus ou traitements étudiés.

Amplification par PCR

Au cours de cette étape, les ADNc monocaténares préalablement obtenus par transcription inverse sont amplifiés par PCR (voir Fiche 24). L'amplification nécessite deux amorces : la première, complémentaire de l'extrémité 3'-OH de l'ADNc, sera l'amorce $T_{12}XY$ employée lors de la transcription inverse et la seconde, complémentaire de l'extrémité 5'-P de l'ADNc, correspondra à une amorce aléatoire de 10 paires de bases (voir Fiche 51). L'amorce aléatoire s'hybride au hasard dans la partie 5'-P et, par rapport à la queue poly(A), à une certaine distance qui varie pour chaque ADNc sélectionné au cours de la première

étape. Ceci conduit à l'obtention d'ADNc partiels bicaténaires, de tailles différentes, qui renferment la partie 3'-OH de l'ARNm correspondant.



Séparation électrophorétique et clonage des ADNc

Les ADNc bicaténaire issus de l'étape d'amplification PCR sont séparés sur gels d'acrylamide similaires à ceux utilisés pour le séquençage. Ces gels sont résolutifs à la base près pour des fragments de moins de 500 pb, type de bandes obtenues par cette méthode. Les différents produits sont visualisés soit par autoradiographie (si l'étape précédente de PCR a été effectuée en présence de [³⁵S]dATP), soit par coloration argentique. Il est nécessaire de comparer les différents échantillons sur un même gel. Si des ARNm sont, de manière reproductible, différenciellement accumulés entre deux traitements, il est alors possible d'exciser les bandes différenciellement accumulées. Ceci est directement réalisable dans le cas d'une révélation argentique du gel. Par contre, en ce qui concerne le marquage radioactif, il faut superposer l'autoradiogramme au gel afin de délimiter la zone contenant la bande d'ADNc d'intérêt. La quantité d'ADNc présente dans une bande est trop faible pour pouvoir être directement sous-clonée. Il est donc nécessaire à ce stade de réamplifier l'ADN présent dans la bande excisée à l'aide du couple d'amorces précédemment utilisé. Le nouveau fragment obtenu est alors sous-cloné en plasmide (voir Fiche 9) et séquencé (voir Fiche 23). Une sonde générée à partir de cette séquence sera utilisée pour une étude plus précise de l'expression du gène correspondant : criblage d'une banque d'ADNc (afin d'obtenir l'ADNc pleine longueur) ou d'une banque génomique, *northern*, hybridation *in situ*...

Les avantages de la technique de tri d'ARNm sont nombreux :

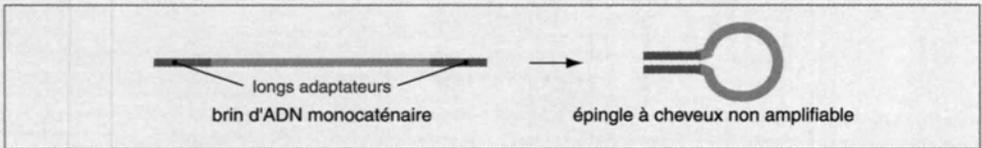
- fondée sur l'emploi de la PCR, cette technique requiert peu de matériel : 1 µg d'ARN poly(A) permet de réaliser 150 réactions ;
- sur un même gel peuvent être identifiés des gènes sur- ou sous-exprimés (bandes apparaissant ou disparaissant) ;
- elle est beaucoup plus rapide que les techniques classiques de criblage différentiel et il est possible par DD RT-PCR d'isoler et d'identifier un clone d'ADNc en 15 jours.

Une approche similaire de la DD RT-PCR peut être réalisée à l'aide d'AFLP (voir Fiche 17).

**CRIBLAGE DIFFÉRENTIEL PAR SSH :
HYBRIDATION SOUSTRACTIVE ET SUPPRESSIVE**

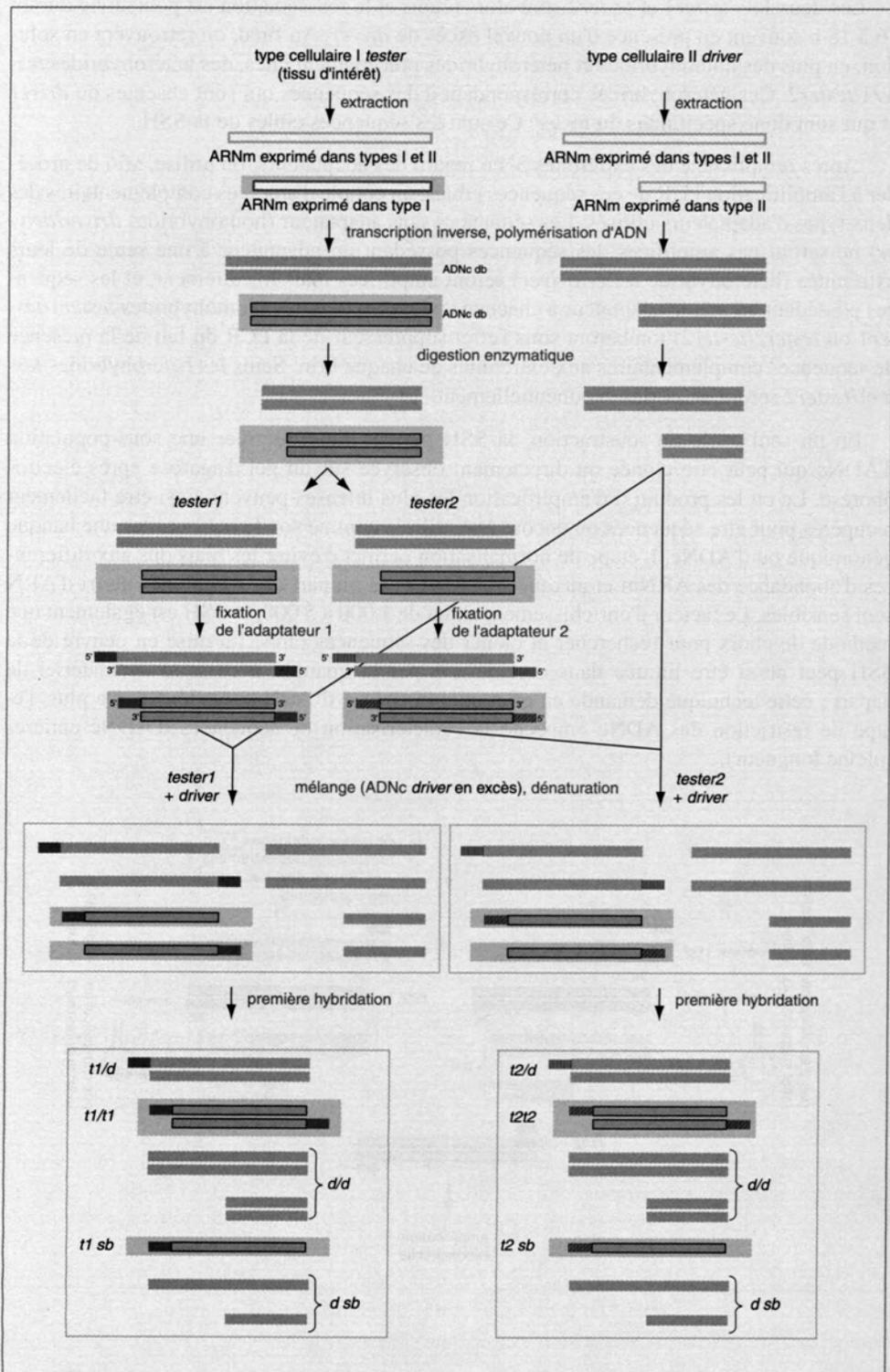
(Suppression Subtractive Hybridization)

L'hybridation soustractive et suppressive (ou SSH, *Suppression Subtractive Hybridization*) est une méthode de criblage différentiel couplée à la PCR. L'objectif est de supprimer les ARNm communs à deux échantillons ou deux types cellulaires afin de créer des sous-populations simplifiées ne contenant que les ARNm exprimés différemment entre les échantillons de départ. Très proche de la RDA (voir Fiche 20), la SSH s'en distingue essentiellement par la mise en oeuvre d'une étape de normalisation entre les séquences abondantes et celles au contraire très rares qui a pour but de favoriser la détection des ARNm peu abondants. Elle est fondée à la fois sur la cinétique de réassociation des brins d'ADN (les fragments abondants s'hybrideront plus vite que les fragments rares) et sur l'existence de l'effet suppresseur de la PCR. Cet effet est dû au fait que de longs adaptateurs inversés (complémentaires) placés de part et d'autre d'un brin monocaténaire auront plus d'affinité l'un pour l'autre que pour une amorce d'amplification plus courte. Lorsque ces longs adaptateurs inversés se reconnaissent, ils forment des structures en épingle à cheveux non amplifiables.



Comme dans la plupart des techniques de soustraction, on désigne par le terme de *tester* l'échantillon d'intérêt et par le terme de *driver* l'échantillon témoin auquel est comparé le *tester*. Les ARN sont extraits des deux types de tissus, puis transformés en ADNc monocaténaire par transcription inverse et en ADNc bicaténaires par polymérisation d'ADN. Les ADNc *tester* et *driver* sont tout d'abord digérés par une même enzyme afin d'obtenir des fragments de petites tailles, plus favorable aux étapes suivantes d'hybridation. Les ADNc *tester* ainsi digérés sont divisés en deux pools égaux (*tester1* et *tester2*) auxquels sont liés de longs adaptateurs (44 pb) différents entre *tester1* et *tester2*. Aucun adaptateur n'est ligaturé aux ADNc bicaténaires *driver* digérés.

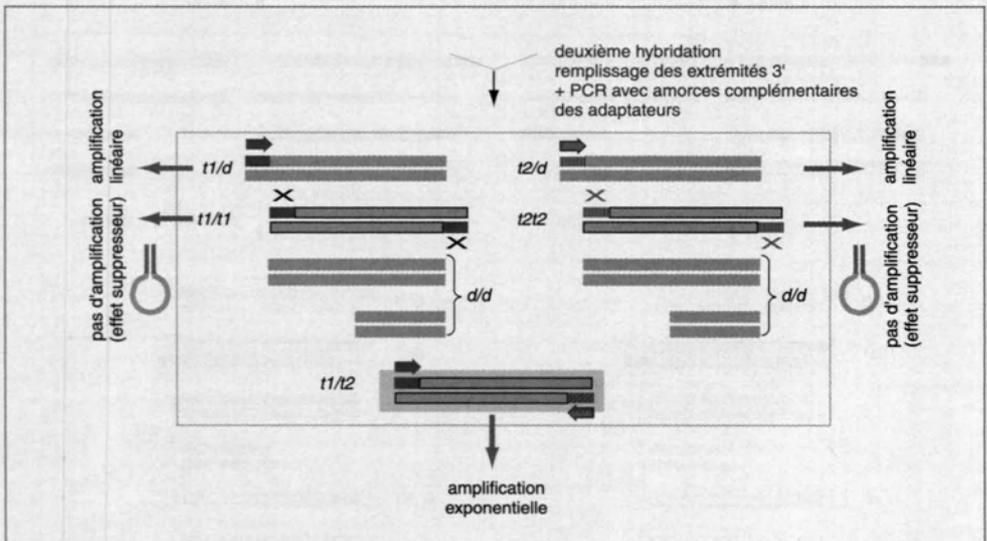
Les trois lots d'ADNc (*driver*, *tester1* et *tester2*) sont dénaturés. Les ADNc *driver* sont mélangés en excès à chacun des lots *tester*. Le mélange d'ADNc est laissé à la réassociation. Plusieurs cas de figure apparaissent : on trouvera en solution dans chaque lot des séquences monocaténaires (simple brin ou sb) qui n'ont pas encore trouvé leur brin complémentaire, des homohybrides *tester/tester* ou *driver/driver* (effet normalisateur de la cinétique d'hybridation) et des hétérohybrides *tester/driver* (effet soustractif de l'hybridation). Toutes ces molécules homo- ou hétérohybrides ne peuvent donc plus participer à la suite de la réassociation.



Les deux lots *tester1* et *tester2* sont alors réunis et la réassociation est poursuivie durant 16 à 18 h souvent en présence d'un nouvel excès de *driver*. Au final, on retrouvera en solution, en plus des homohybrides et hétérohybrides précédemment cités, des hétérohybrides *tester1/tester2*. Ces hétérohybrides correspondent à des séquences qui sont absentes du *driver* et qui sont donc spécifiques du *tester*. Ce sont les séquences cibles de la SSH.

Après remplissage des extrémités 3' en regard des adaptateurs, on utilise, afin de procéder à l'amplification PCR de ces séquences cibles, un couple d'amorces complémentaires des deux types d'adaptateurs utilisés. Les séquences sans adaptateur (homohybrides *driver/driver*) ne seront pas amplifiées, les séquences possédant un adaptateur à une seule de leurs extrémités (hétérohybride *tester/driver*) seront amplifiées mais linéairement, et les séquences possédant un même adaptateur à chacune de leurs extrémités (homohybrides *tester1/tester1* ou *tester2/tester2*) tomberont sous l'effet suppresseur de la PCR du fait de la présence de séquences complémentaires aux extrémités de chaque brin. Seuls les hétérohybrides *tester1/tester2* seront amplifiés exponentiellement.

En un seul cycle de soustraction, la SSH permet donc de créer une sous-population d'ADNc qui peut être clonée ou directement observée sur un gel d'agarose après électrophorèse. Le ou les produit(s) d'amplification les plus intenses peuvent ainsi être facilement récupérés pour être séquencés ou encore être utilisés comme sonde pour cribler une banque génomique ou d'ADNc. L'étape de normalisation permet d'éviter les biais dus aux différences d'abondance des ARNm et auxquels la RDA et la plupart des techniques de tri d'ARN sont sensibles. Le facteur d'enrichissement allant de 1 000 à 5 000, la SSH est également une méthode de choix pour rechercher et cloner des séquences rares. La mise en oeuvre de la SSH peut aussi être limitée dans certains cas par la quantité nécessaire de matériel de départ : cette technique demande en effet au moins 1 µg d'ARN polyadénylé. De plus, l'étape de restriction des ADNc empêche la caractérisation de séquences d'ADNc entières (pleine longueur).



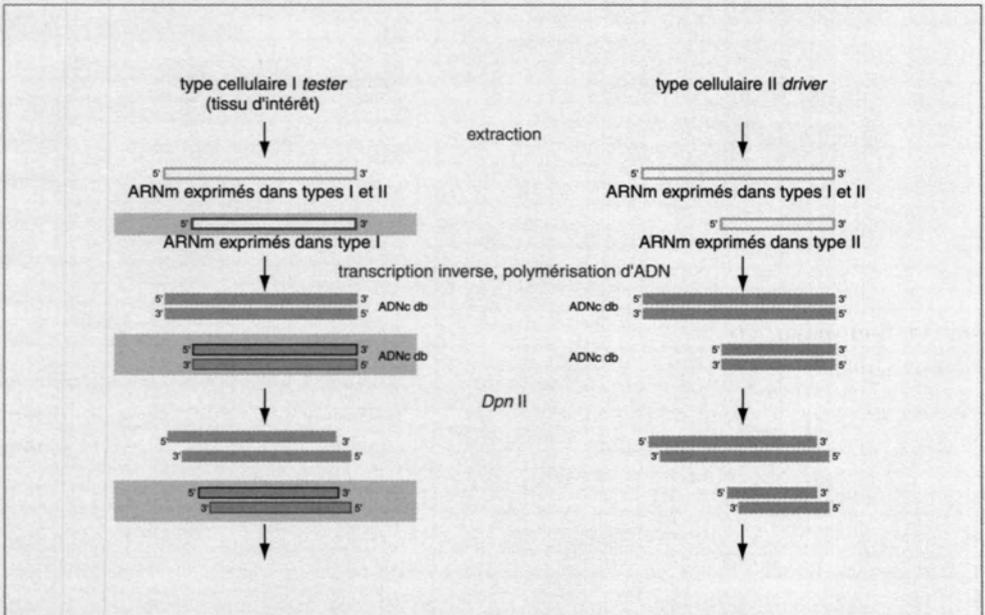
**CRIBLAGE DIFFÉRENTIEL par RDA :
ANALYSE PAR DIFFÉRENCE DE REPRÉSENTATION**

(Representational Difference Analysis)

La soustraction d'ADNc par RDA est une technique de criblage différentiel assistée par PCR permettant d'identifier des gènes exprimés spécifiquement dans un tissu cible par rapport à un tissu contrôle. La soustraction comporte les étapes suivantes : (i) synthèse d'ADNc bicaténaires et leur digestion par une enzyme de restriction ; (ii) leur multiplication par amplification PCR ; (iii) hybridation soustractive. Plusieurs cycles d'hybridations soustractives sont effectués de manière à enrichir le produit obtenu en gènes différentiellement exprimés.

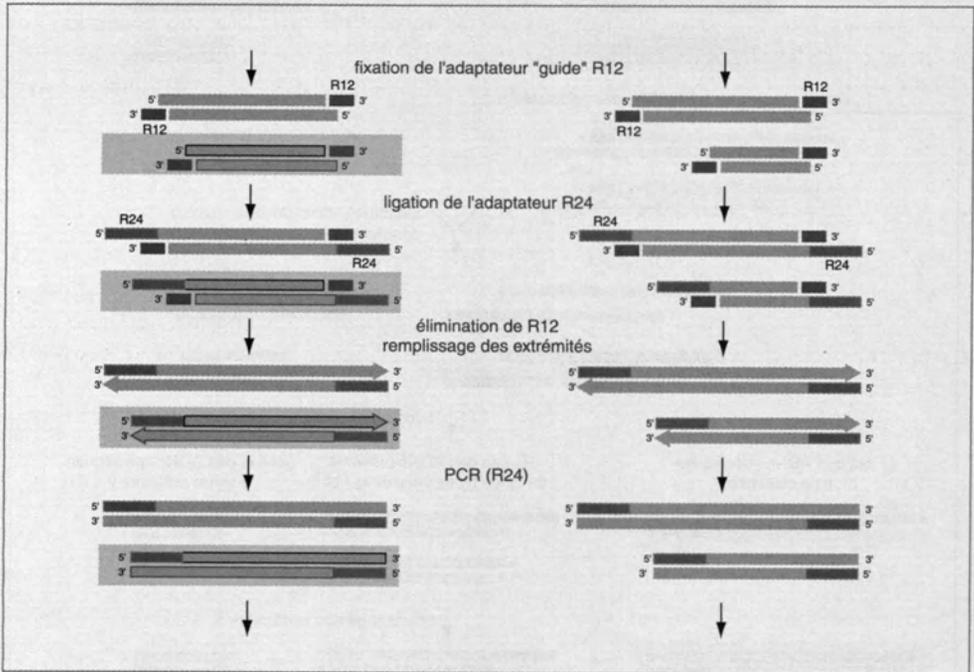
Synthèse des ADNc bicaténaires et digestion

Des ARN sont extraits à partir des tissus cible (ou *tester*) et contrôle (ou *driver*). Les ARNm polyadénylés sont sélectionnés et les ADNc bicaténaires (double brin ou db) correspondants sont synthétisés. Ces ADNc sont ensuite digérés par l'enzyme de restriction *Dpn II* dont le site de reconnaissance est de 4 pb et qui produit des extrémités cohésives 5'-P protubérantes.



Multiplication par PCR

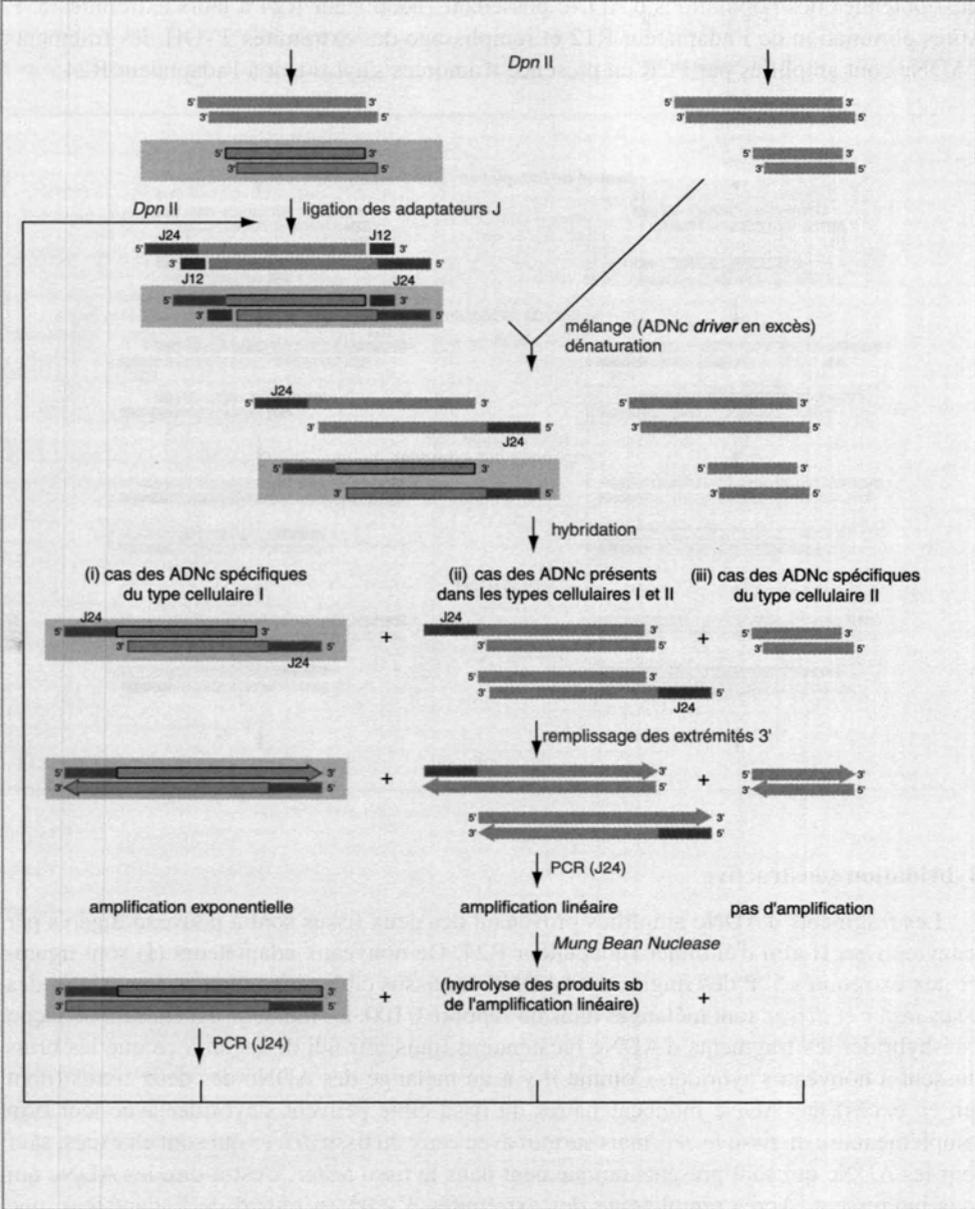
Deux oligonucléotides de synthèse sont hybridés aux extrémités 5'-P et 3'-OH des fragments d'ADNc provenant des deux tissus, de façon à produire des adaptateurs après ligature. L'oligonucléotide R12 possède une partie complémentaire de l'extrémité cohésive du site *Dpn* II et va se placer sur ce brin d'ADN. Cependant, du fait de l'absence d'un groupement phosphate à son extrémité 5'-P, il ne pourra pas former de liaison covalente avec le brin d'ADNc. Il va en réalité servir de guide au second oligonucléotide R24 qui lui est partiellement complémentaire. Cet oligonucléotide R24 se fixe de façon covalente à l'ADNc. Sont ainsi obtenues des populations d'ADNc possédant l'adaptateur R24 à leurs extrémités 5'-P. Après élimination de l'adaptateur R12 et remplissage des extrémités 3'-OH, les fragments d'ADNc sont amplifiés par PCR en présence d'amorces s'hybridant à l'adaptateur R24.



Hybridation soustractive

Les fragments d'ADNc amplifiés provenant des deux tissus sont à nouveau digérés par l'enzyme *Dpn* II afin d'éliminer l'adaptateur R24. De nouveaux adaptateurs (J) sont ligaturés aux extrémités 5'-P des fragments d'ADNc des tissus cibles uniquement. Les ADNc des tissus *tester* et *driver* sont mélangés dans un rapport 1/100. Le mélange est chauffé de façon à déshybrider les fragments d'ADNc bicaténaire, puis refroidi de façon à ce que les brins puissent à nouveau s'hybrider. Comme il y a un mélange des ADNc des deux tissus (dont l'un en excès), les ADNc monocaténaire du tissu cible peuvent s'hybrider avec leur brin complémentaire du tissu *tester*, mais surtout avec ceux du tissu *driver* qui sont en excès, sauf pour les ADNc qui sont présents uniquement dans le tissu *tester*, c'est-à-dire les ADNc qui nous intéressent. Après remplissage des extrémités 3'-OH en regard de l'adaptateur, une amplification PCR est réalisée et trois cas de figure se présentent : (i) lorsque le gène n'est

exprimé que dans le tissu *tester*, le duplex comportera un adaptateur J sur chaque brin et la PCR produira une amplification exponentielle ; (ii) lorsque le gène est exprimé dans les deux tissus, le duplex ne comportera qu'un seul brin possédant un adaptateur J qui subira une amplification linéaire ; (iii) lorsque le gène n'est exprimé que dans le tissu *driver*, le duplex formé ne comportera pas d'adaptateur J et ne sera pas amplifié. Les produits PCR sont digérés par l'exonucléase *Mung Bean Nuclease* spécifique de l'ADN simple brin, de façon à éliminer les produits de l'amplification linéaire. Les produits de la digestion sont à nouveau amplifiés par PCR à l'aide des amorces J.



De façon à effectuer une seconde hybridation soustractive sur les produits PCR obtenus, les amorces J sont éliminées par l'enzyme *DpnII* et de nouveaux adaptateurs sont ligaturés en 5'-P. Une nouvelle hybridation soustractive identique à la première est appliquée de façon à enrichir le produit de la RDA en gènes spécifiquement exprimés dans le tissu *tester*. Cet enrichissement peut être poursuivi à volonté par des hybridations soustractives successives. Par la suite, la population d'ADNc enrichie en séquences spécifiques du tissu *tester* peut être clonée dans des vecteurs plasmidiques pour conservation et analyse.

Cette technique présente les mêmes avantages que l'hybridation soustractive suppressive (SSH, voir Fiche 19), mais elle est reconnue comme générant beaucoup moins de faux positifs qu'elle. Par contre, aucun kit commercial n'est disponible et la RDA est moins rapide à mettre en oeuvre que la SSH.

EST : ÉTIQUETTES DE GÈNES EXPRIMÉS (*Expressed Sequence Tags*)

Le génome d'un organisme eucaryote possède de nombreux gènes mais chaque type de cellules n'en exprime qu'une fraction selon les fonctions remplies. Comme nous l'avons vu (voir Fiche 15), une banque d'ADNc obtenue à partir d'un tissu déterminé donne une "photographie instantanée" des gènes exprimés dans ce tissu au moment de l'extraction des ARNm. Théoriquement, les milliers de clones phagiques ou plasmidiques de la banque recouvrent la presque totalité des séquences exprimées sous forme d'ARNm. Les gènes fortement exprimés sont plus représentés que ceux qui le sont faiblement (ARNm rares).

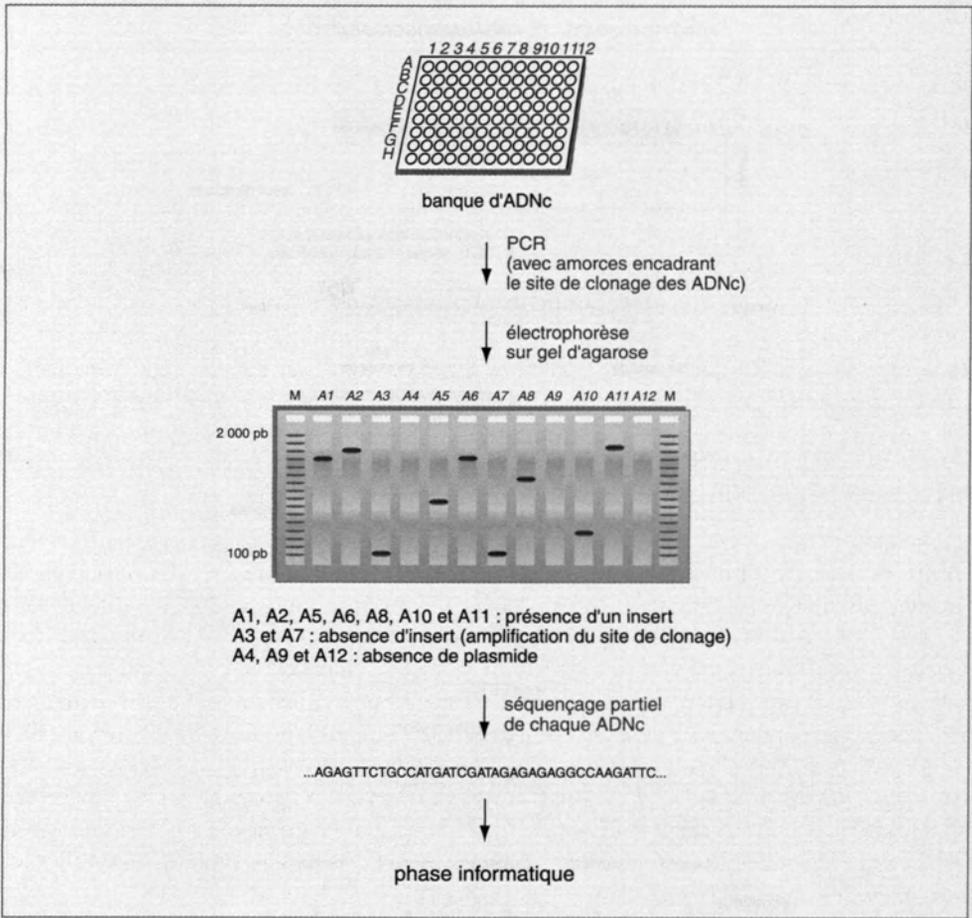
La détermination d'étiquettes de gènes exprimées (EST, *Expressed Sequence Tags*) a pour objectif de caractériser une partie de la séquence de la majeure partie des ARNm accumulés dans un tissu donné. Elle implique le séquençage partiel de chacun des ADNc présents dans une banque d'ADNc. Ces séquences sont ensuite comparées une à une à celles qui sont stockées dans les banques de données internationales (p. ex. GenBank *via* Internet). Les informations obtenues sur l'homologie de séquences sont dans certains cas suffisantes pour assigner une fonction potentielle au gène correspondant. On peut faire l'hypothèse que les EST qui ne présentent pas d'homologie avec des gènes connus correspondent à des gènes nouveaux, spécifiques du système étudié (espèce, tissu ou état physiologique). Cette recherche d'exhaustivité (séquencer tous les gènes exprimés) se fait en favorisant le nombre par rapport à la qualité. Cependant les informations obtenues permettent le plus souvent d'identifier les gènes et leur niveau d'expression dans le tissu et l'état physiologique considéré.

Cette approche EST permet ainsi l'élaboration d'un catalogue de gènes exprimés dans un tissu particulier. Elle peut alors être couplée à des études d'expression en masse : chacun des ADNc identifiés est déposé sur une membrane ou une lame de verre pour constituer des filtres à haute densité comprenant des centaines, voire des milliers d'ADNc (voir Fiche 22). Ces membranes ou lames peuvent alors être hybridées avec des sondes correspondant aux ARNm de n'importe quel autre tissu ou cellule soumis à différents stress biotiques ou abiotiques : l'étude simultanée de l'expression des centaines de gènes déposés sur les filtres est alors possible. Les EST sont aussi utilisés pour l'annotation des génomes (localisation des exons) et dans la recherche de polymorphisme pour des études de génétique.

Méthode

Un projet EST comporte une phase de biologie moléculaire et une phase bioinformatique.

- Étapes de biologie moléculaire : (i) extraction d'ARN et sélection facultative de la fraction poly(A) ; (ii) construction de la banque d'ADNc (voir Fiche 15) ; (iii) séquençage en masse sans criblage préalable et sans chercher à corriger les erreurs.



- Étapes informatiques (traitement des données de séquence) :
 - Création de contigs : (i) élimination des zones de qualité insuffisante ; (ii) élimination des séquences de vecteurs et d'adaptateurs ; (iii) élimination des EST correspondants à des contaminants (ARN ribosomiques, séquences bactériennes, mitochondriales, etc.) ; (iv) masquage des séquences répétées ; (v) "clusterisation" : il s'agit de regrouper les EST codant pour le même gène, sur la base d'une forte similarité de séquence ; (vi) assemblage ("contigage") : il s'agit de créer pour un même gène une séquence "consensus" à partir de plusieurs EST couvrant des parties différentes du gène, c'est-à-dire le contig le plus long possible à partir des fragments chevauchants du cluster.
 - Annotations : cette étape consiste à assigner une fonction et à caractériser les contigs et les EST isolés (singletons). Elle est fondée sur l'interrogation de bases de données de séquences (nucléiques et protéiques), de motifs, de profils de spectrométrie de masse de protéines, etc.. L'utilisation de méthodes de prédiction (de fonction, de localisation cellulaire, etc.) peut s'avérer utile au cours de cette étape.

RÉSEAUX D'ADN : PUCES À ADN, FILTRES D'ADNc

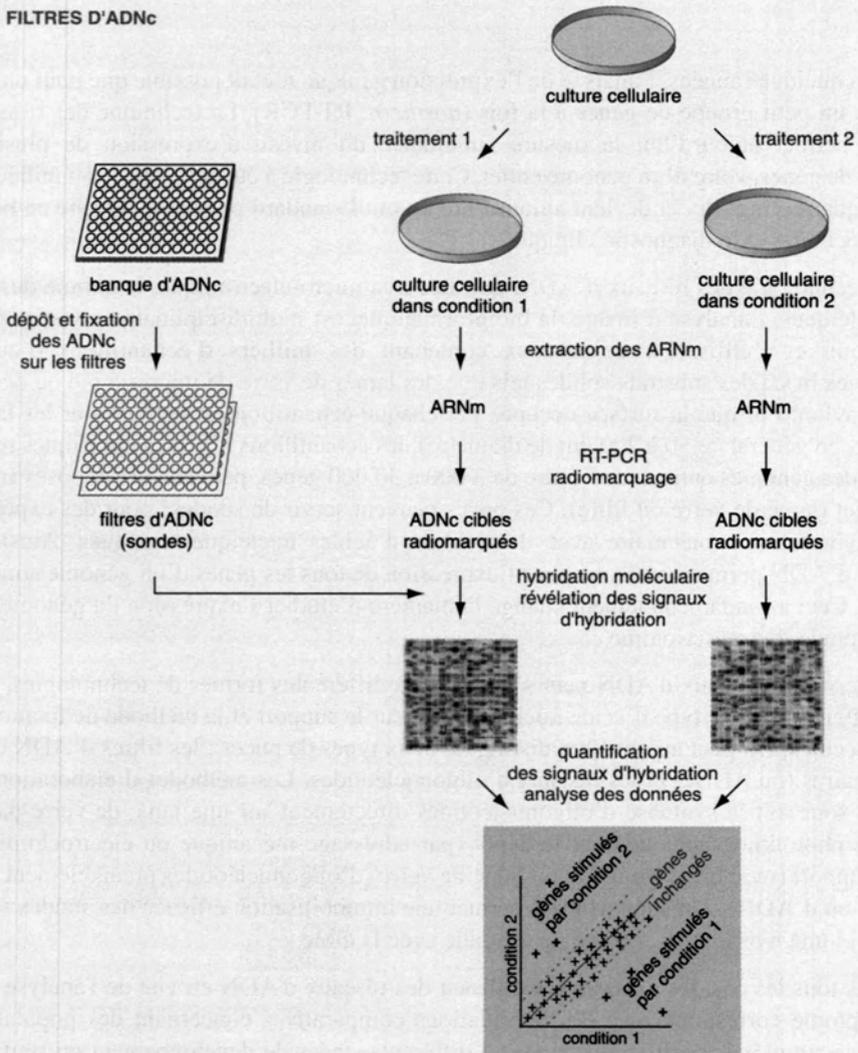
Il y a quelques années, l'analyse de l'expression génique n'était possible que pour un seul gène ou un petit groupe de gènes à la fois (*northern*, RT-PCR). La technique des réseaux¹ d'ADN permet aujourd'hui la mesure simultanée du niveau d'expression de plusieurs milliers de gènes, voire d'un génome entier. Cette technologie a été développée au milieu des années quatre-vingt-dix et devient aujourd'hui un outil standard pour la recherche en biologie moléculaire et le diagnostic clinique.

La technologie des réseaux d'ADN, qui, liée à la micro-électronique, la chimie des acides nucléiques, l'analyse d'image, la bioinformatique, est multidisciplinaire, consiste en la fabrication et l'utilisation de réseaux contenant des milliers d'échantillons d'acides nucléiques liés à des substrats solides tels que des lames de verre de microscopie ou des filtres de nylon. Puisque la surface occupée par chaque échantillon est petite (pour les lames de verre, en général de 50 à 200 μm de diamètre), les échantillons d'acides nucléiques représentant des génomes entiers, de l'ordre de 3 000 à 30 000 gènes, peuvent être déposés sur un seul objet (lame de verre ou filtre). Ces puces peuvent servir de sondes² pour des expériences d'hybridation moléculaire avec des cibles² d'acides nucléiques marqués. Ainsi, les réseaux d'ADN permettent de mesurer l'expression de tous les gènes d'un génome simultanément. Ceci a fondamentalement changé la manière d'étudier l'expression du génome : on parle d'étude du transcriptome.

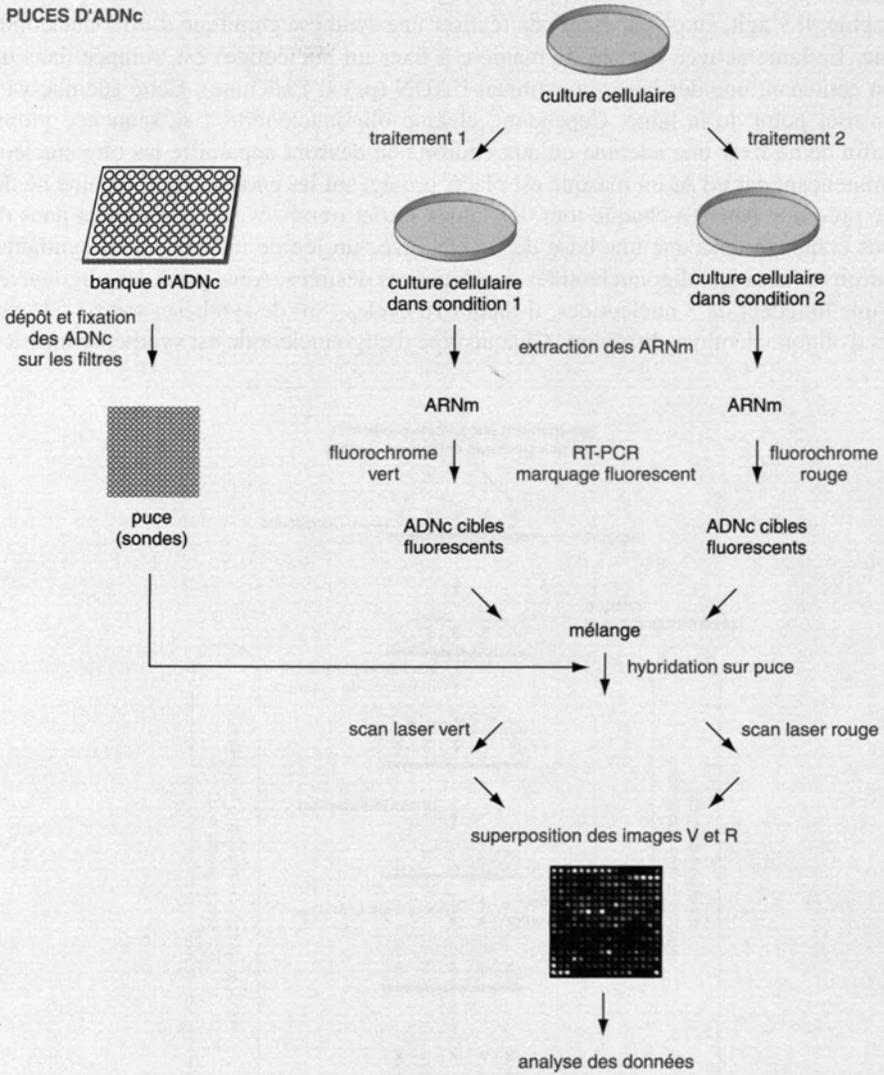
Le terme de réseaux d'ADN peut s'appliquer à différentes formes de technologies, chacune différant dans le type d'acide nucléique fixé sur le support et la méthode de fixation de ces molécules. On peut aujourd'hui distinguer deux types de puces : les filtres d'ADN complémentaires (ou ADNc) et les réseaux d'oligonucléotides. Les méthodes d'élaboration des réseaux sont soit la synthèse d'oligonucléotides directement sur une lame de verre par un procédé photolithographique, soit le dépôt (par adressage mécanique ou électrochimique) sur le support (membrane de nylon ou lame de verre) d'oligonucléotides préalablement synthétisés ou d'ADNc. Un support idéal permet une immobilisation efficace des sondes sur la surface et une hybridation robuste de la sonde avec la cible.

Dans tous les cas, les expériences utilisant des réseaux d'ADN en vue de l'analyse d'un transcriptome correspondent à des hybridations comparatives concernant des populations d'ARN accumulés dans différents tissus, à différents stades de développement ou soumis à différents traitements. Les différentes étapes concernant ces approches sont décrites dans les figures ci-dessous.

FILTRES D'ADNc



PUCES D'ADNc

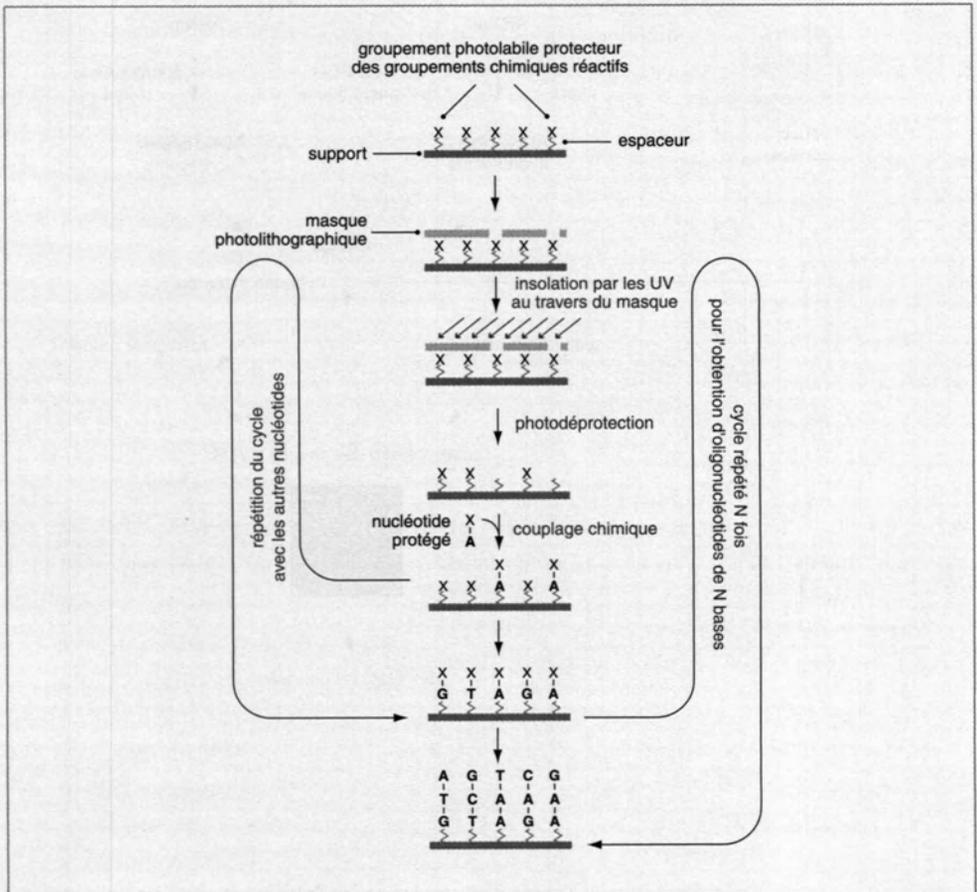


Élaboration des filtres et des lames de verre

Dans le cas des filtres, ce sont des ADNc qui sont déposés. En ce qui concerne les lames de verre, il est possible de déposer soit des ADNc, soit de synthétiser directement sur les lames des oligonucléotides représentatifs des gènes de l'organisme étudié.

Synthèse d'oligonucléotides sur lames de verre

Certaines puces à ADN sont fabriquées par synthèse *in situ* d'oligonucléotides par photolithographie. Il s'agit, étape par étape, de réaliser une synthèse chimique d'oligonucléotides sur lame. La lame activée (traitée de manière à fixer un nucléotide) est trempée dans une solution contenant une des bases constituant l'ADN (p. ex. l'adénine). Cette adénine va se fixer en tout point de la lame. Cependant, chaque oligonucléotide a sa séquence propre. Ainsi, afin de ne fixer une adénine qu'aux endroits où devront apparaître les oligonucléotides commençant par un A, un masque est placé, protégeant les endroits où l'adénine ne doit pas être présente. Ainsi, à chaque tour, les lames seront trempées successivement dans des solutions contenant chacune une base de l'ADN, avec un jeu de masques pour synthétiser aux endroits voulus les oligonucléotides de séquences désirées. Ainsi pour des oligonucléotides d'une longueur de 8 nucléotides, il faudra 16 cycles (8h) de synthèse, soit 65 536 possibilités d'oligonucléotides différents. Chaque type d'oligonucléotide est synthétisé en 1 à 10



millions de copies sur une surface variant de 20 à 50 μm^2 . Ces puces posent un problème majeur, celui de leur incapacité à détecter et à corriger les erreurs éventuelles introduites lors de la synthèse *in situ* sur la surface. D'autre part, la longueur des ADN qu'il est ainsi possible de synthétiser est faible (généralement 20 nucléotides).

Dépôts d'ADNc

Les ADNc sondes que l'on désire déposer (sur filtre ou sur lame de verre) sont amplifiés par PCR à partir de banques bactériennes ou plasmidiques, puis ordonnés dans des plaques de micro-titration de 96, 384 ou 1 586 puits. Des standards pour le contrôle de la qualité de l'hybridation peuvent aussi être intégrés à cette collection d'ADNc sondes.

Les ADNc sont ensuite transférés sur la matrice constituant le support. Le plus souvent, la concentration d'ADNc sondes est normalisée avant le dépôt afin de limiter des écarts de quantité d'ADN d'un puits à l'autre. Les quantités déposées dépendent de la taille de la surface du dépôt et varient de l'ordre du nanogramme dans le cas de dépôt manuel jusqu'à quelques picogrammes dans le cas de dépôts réalisés par des robots. Dans le cas des lames de verre, il est nécessaire de les fonctionnaliser au préalable pour permettre l'accrochage des molécules d'ADN. Pour les principaux types actuellement utilisés, on peut citer les revêtements époxy, poly-L-lysine, silane et aldéhyde. Dans les cas des revêtements poly-L-lysine et silane, l'ADNc est déposé puis lié de manière covalente au support par une exposition aux UV. Dans le cas d'un revêtement aldéhyde, les amorces utilisées pour la PCR sont aminées, ce qui permet la formation de liaisons covalentes avec le support. Après fixation des ADNc, les lames subissent un traitement qui est fonction du revêtement et qui permet de saturer les éventuels sites réactifs de la lame. A titre d'exemple, les dépôts sont réalisés par un robot qui charge environ 500 nanolitres de produits de PCR purifiés prélevés par une aiguille creuse dans les puits et dépose 5 nanolitres de chaque échantillon par impact de l'aiguille sur chacune des lames. Un robot peut, avec un même prélèvement, faire jusqu'à une centaine de dépôts. Les dépôts sont en général d'environ 50-200 μm de diamètre et sont espacés de 180-350 μm .

Dans le cas des filtres d'ADNc, il s'agit essentiellement de filtres de nylon ou de nitrocellulose présentant une composition de qualité identique à ceux employés pour les hybridation de type *Southern* et *northern* (voir Fiche 12). Cependant, de nouveaux supports en matière plastique sont proposés, mais leur utilisation reste toutefois encore limitée. Les systèmes manuels permettent d'effectuer des dépôts de quelques mm^2 à 1 cm^2 pour une surface totale de filtre d'environ 50 à 200 cm^2 . De la surface unitaire de dépôt d'un ADNc, et donc du système de dépôt manuel, dépendra la densité finale d'ADNc sur le filtre. Dans le cas de dépôts automatisés, un bras mécanique muni de fines aiguilles prélève et dépose par capillarité les ADNc. Ces dépôts sont de l'ordre du picolitre au nanolitre. La limitation de la densité tient aux propriétés physiques du filtre, mais aussi à la longueur d'émission du radioélément composant l'ADN cible qui est utilisé pour révéler la présence d'une hybridation (voir plus loin). Des miniaturisations avancées existent tout de même pour ce type de support sans égaler la densité d'ADNc des lames de verre. Les dépôts sur filtres sont suivis d'une étape de cuisson ou d'exposition aux UV pour fixer définitivement et d'une manière covalente l'ensemble des ADNc sondes au support.

Afin d'être accessibles par la suite à l'hybridation avec les ADN cibles, les ADNc sondes doivent être dénaturés (rendus monocaténares). La soude en faible concentration (NaOH 0,1 - 0,5 N) est principalement employée pour dénaturer les sondes après leur dépôt.

Méthodes

Marquage et hybridation

Les ADN cibles sont préparés à partir des ARN extraits des tissus dont on veut comparer les contenus en gènes exprimés. Ces ARN sont copiés en ADNc monocaténares à l'aide d'une transcriptase inverse (voir Fiche 27). Ces ADNc sont alors marqués radioactivement (cas des filtres d'ADNc) ou par fluorescence (cas des lames de verre), leur synthèse se déroulant en présence d'une base nucléotidique modifiée (voir Fiche 11). Cette étape de transcription inverse nécessite une quantité assez importante d'ARN (plusieurs microgrammes). Dans le cas d'un matériel biologique disponible en de faibles quantités, il est possible de coupler cette transcription inverse à des amplifications des ADNc monocaténares par PCR (RT-PCR). Dans ce cas, on obtient une population d'ADNc bicaténares que l'on peut alors marquer par les techniques traditionnelles (voir Fiche 11).

En ce qui concerne l'hybridation des filtres à ADNc, les conditions d'expérience reprennent les protocoles utilisés pour les hybridations de type *Southern* ou *northern* (voir Fiche 12), les filtres étant préhybridés afin de saturer les sites aspécifiques. La première série de filtres est hybridée avec les ADN cibles de la population 1 et la seconde série de filtres (copie identique à la première) est hybridée avec les ADN cibles de la population 2. Pour l'hybridation, les ADNc mono- ou bicaténares cibles (marqués radioactivement) sont mis en présence du filtre dans un milieu clos (tube, boîte, sachet) dans un tampon d'hybridation (voir Fiche 12). Cette étape est suivie d'une série de lavages de plus en plus contraignants qui permet d'éliminer des hybridations aspécifiques entre les ADN sondes et cibles.

L'une des différences majeures entre l'utilisation des filtres à ADNc et des lames de verre est que, dans le second cas, une même lame (portant les ADN sondes) est hybridée avec un mélange d'ADNc cibles marqués provenant des deux conditions que l'on veut comparer : les populations d'ADNc cibles des deux situations sont marquées chacune par un fluorochrome différent et le signal obtenu correspondra au rapport de la fluorescence des cibles de la population 1 et des cibles de la population 2. Chaque sonde pourra s'hybrider avec sa cible provenant soit de la population 1 (marquée par un fluorophore), soit de la population 2 (marquée par un autre fluorophore). Si la cible est plus abondante dans la population 1, elle aura plus de chance de s'hybrider sur la sonde correspondante qui sera alors davantage chargée en fluorophore de la cible 1. Les cibles fluorescentes sont concentrées, reprises dans une solution d'hybridation et mélangées en concentrations équimolaires avant d'être mises en contact avec la lame. Les lames sont ensuite recouvertes d'une lamelle et incubées 16-24h à une température d'hybridation pouvant varier entre 42 et 65 °C. Les cibles non liées aux sondes sont éliminées par lavage. Les lames sont ensuite séchées par centrifugation.

Dans tous les cas, des séries de contrôles sont nécessaires afin de vérifier que les hybridations aspécifiques ont été éliminées.

Acquisition des données

Dans le cas des lames de verre, les signaux obtenus sont des émissions de fluorescence aux points d'hybridation. Ces lames sont d'abord numérisées dans des analyseurs d'images (Imageurs) qui détectent séparément chaque longueur d'onde émise spécifiquement par chaque fluorophore. Des images colorisées sont ainsi obtenues. Les numérisateurs à lames de verre sont des microscopes liés à un système de laser d'émission double ou multiple qui excite les fluorophores liés à chaque cible et réceptionne la fluorescence émise. L'intensité

du signal de chaque spot pour les deux longueurs d'onde d'émission des deux fluorophores est ensuite mesurée par un logiciel de quantification, ce qui permet de déterminer pour chaque ADN sonde le rapport entre ses ARNm présents dans la population 1 et dans la population 2.

Dans le cas des filtres à ADNc, la lecture des signaux radioactifs peut être obtenue par simple contact des filtres sur des films autoradiographiques. Cependant, ces films ne sont souvent pas assez sensibles pour détecter des signaux de faible intensité. On utilise alors des numériseurs qui lisent des écrans photosensibles sur lesquels les filtres sont placés dans une cassette à l'obscurité. Les émissions du radioélément utilisé pour marquer les ADN cibles vont laisser une empreinte par transfert d'énergie sur l'écran. La révélation de ces empreintes est réalisée à l'aide d'un "Imageur" qui permet l'obtention d'une image des signaux d'hybridation. Chaque couple de filtres est numérisé séparément. Comme pour les lames de verre, un logiciel de quantification permet d'apprécier le nombre de pixels associé à chaque signal d'hybridation, ce qui permet de déterminer pour chaque ADN sonde le rapport des signaux d'hybridation entre le premier filtre (population 1) et le second (population 2).

Analyse des données

- **Traitement des données**

Les données brutes obtenues après quantification des signaux d'hybridation sont souvent corrigées avant de procéder à une analyse plus fine. Différents types de corrections peuvent être apportées. Le plus souvent, il s'agit d'une soustraction du bruit de fond et d'une normalisation de l'ensemble des niveaux d'expression mesurés dans chacune des conditions. La normalisation a pour but d'équilibrer les intensités d'hybridation obtenues entre les deux populations d'ADNc cibles, ainsi que de comparer les résultats entre les différentes répétitions d'une expérience. La normalisation peut être basée sur des standards disposés parmi les ADNc sondes ou encore en intégrant la somme de tous les signaux mesurés dans chacune des conditions et en rapportant chaque valeur d'expression à cette somme des signaux pour chaque condition. Dans le cas des lames de verre, un biais peut être dû aux fluorophores : ce biais est bien visualisable dans une expérience où deux échantillons identiques d'ARN sont marqués avec deux fluorophores différents puis hybridés sur la même lame. Dans cette situation, il est rare d'avoir des intensités égales, et donc les ratios de 1 attendus. Ces biais ont plusieurs origines comme des propriétés physiques différentes des fluorophores, des efficacités d'incorporation non égales ou la variabilité expérimentale.

Les données ainsi corrigées peuvent servir directement pour procéder aux analyses des profils d'expression. Elles sont aussi très souvent réduites par différents modes de transformation (logarithmique ou en rang p. ex.) avant analyse.

- **Reproductibilité**

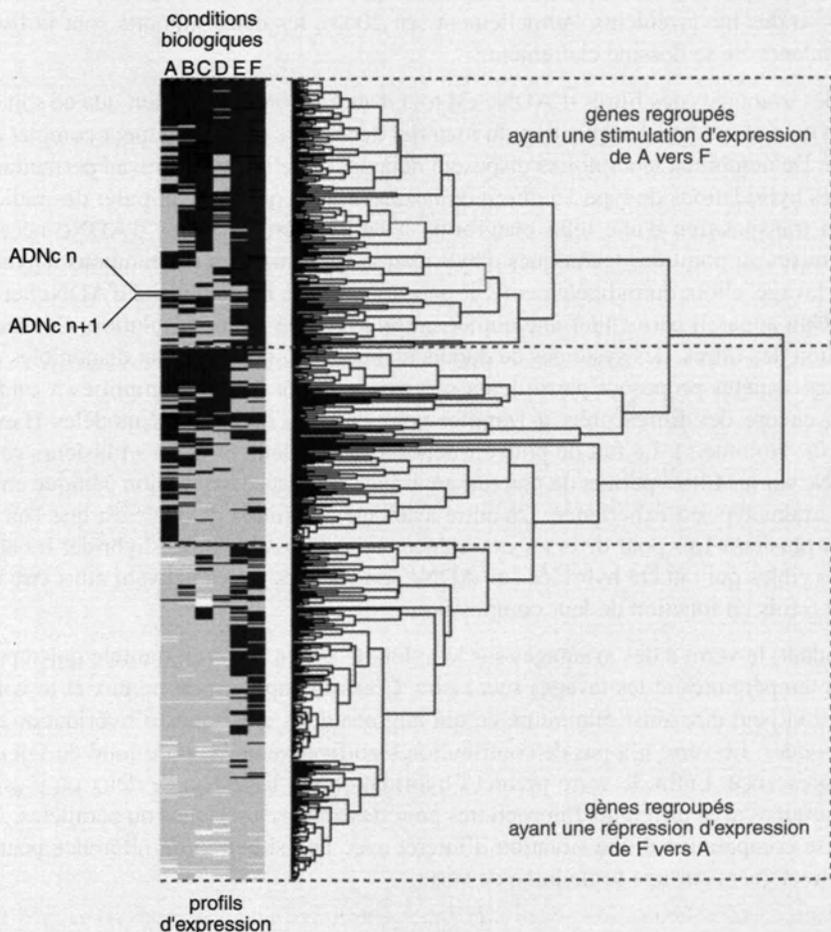
Les technologies basées sur les filtres d'ADNc ou les lames de verre pour quantifier le niveau d'expression de gènes sont très performantes. Toutefois, elles présentent dans bien des cas une variabilité non négligeable entre plusieurs répétitions d'expériences. En effet, différentes étapes de la technique peuvent induire des erreurs dans l'appréciation du niveau des transcrits : le dépôt des sondes (ADNc ou oligonucléotides), les étapes de transcription inverse, de PCR et de marquage des ADNc cibles, l'étape d'hybridation et les lavages. Il est alors nécessaire d'appliquer des méthodes d'analyse statistique à ce style de données, comme l'a-

analyse de variance (ANOVA), l'analyse en composante principale (ACP) ou un test *t* de Student.

- Méthodes d'analyse

Le cas le plus simple est la comparaison des niveaux de transcrits entre une condition expérimentale et une condition non traitée dite de référence. Les valeurs numériques corrigées obtenues pour chacun des gènes peuvent être placées sur un graphique en nuage de points. Ce mode de représentation présente l'avantage de donner pour chaque gène le niveau de l'expression dans chacune des conditions (condition A sur un axe et condition B sur l'autre), mais aussi de déterminer le niveau de la régulation entre les deux états. Les gènes dont les points sont placés le long de la bissectrice du graphe ont des niveaux d'expression identiques dans chacune des conditions ($x = y$), alors que ceux qui s'éloignent de cette droite sont régulés. Cette méthode d'analyse est très pratique pour identifier les gènes les plus fortement régulés dans un état donné. Toutefois, lorsque le nombre d'expériences augmente (p. ex. une cinétique d'expression), ce type de représentation graphique ne convient plus et nécessite d'autres modes de visualisation des données.

Dans le cas d'une comparaison de plusieurs conditions entre elles, une masse très importante de données est vite obtenue (plusieurs milliers de gènes, plusieurs conditions, plusieurs répétitions). Il est donc nécessaire de trier ces données afin de faciliter leur interprétation. Le premier tri concerne l'analyse statistique qui permet d'éliminer les gènes qui ne montrent pas de régulations dans les conditions testées. En ce qui concerne les gènes régulés (qui ont "passé" le test statistique), il existe de nombreuses méthodes de tri qui permettent de regrouper les gènes suivant leurs profils d'expression. Différents algorithmes sont disponibles pour effectuer ces tris, mais ils reposent tous sur un même principe : une analyse de distance entre les profils d'expression des gènes. Les gènes sont assimilés à des vecteurs dont les coordonnées correspondent aux niveaux d'expression ou aux rapports d'expression obtenus dans différentes expériences. Ces "vecteur-gènes" sont replacés dans un espace comportant autant de dimensions que d'expériences prises en compte. Les analyses de distance regroupent alors les gènes par paires en fonction des plus faibles distances mesurées dans les différentes dimensions (regroupement hiérarchisé ou *hierarchical clustering*), ou bien encore définissent plusieurs partitions regroupant des gènes proches (centres mobiles ou *k-means*, cartes auto-organisatrices ou *Self-Organizing Maps*). Ces regroupements peuvent ensuite être hiérarchisés sur des clades placés sur un arbre consensus qui rend compte des distances entre les différents profils d'expression. Ces arbres sont le plus souvent associés à une matrice colorée qui indique le niveau d'expression ou le niveau de régulation de chaque gène. Dans le cas de matrices de régulation, le code couleur attribue une couleur donnée aux stimulations de l'expression génique (rapport d'expression > 1) et une seconde couleur aux répressions de l'expression (rapport d'expression < 1). Les gènes dont l'expression ne varie pas sont le plus souvent représentés en noir.



Représentation graphique d'un regroupement de profils d'expression de gènes par tri hiérarchisé. Le résultat d'une analyse en regroupement hiérarchisé se décompose en deux parties : un arbre (à droite de la figure) qui regroupe les gènes de proche en proche en fonction de la similarité de leurs profils d'expression, et une matrice colorée (ici en niveaux de gris, à gauche sur la figure) qui décrit ces profils d'expression. Sur cette représentation, chaque ligne correspond à un gène (= 1 ADNc présent sur le filtre ou la lame de verre) et chaque colonne correspond au niveau d'expression du gène dans une condition donnée (A, B, C, D, E ou F). Plus exactement, il s'agit du rapport du niveau d'expression de ce gène entre la condition testée et la condition de référence. Lorsque l'expression du gène est stimulée (rapport > 1), une couleur noire est attribuée, alors que lorsque l'expression du gène est réprimée (rapport < 1), une couleur blanche est attribuée. Le code couleur attribué est toujours représenté dans un encart de la figure. L'arbre sur la droite est le résultat d'une analyse de distance entre les profils d'expression de ces gènes, ce qui permet de classer les gènes suivant leur profils d'expression et d'effectuer des regroupements.

Filtres d'ADNc ou lames de verre ?

Les filtres d'ADNc et les lames de verre ont des particularités d'utilisation présentant des avantages ou des inconvénients. Actuellement (en 2003), les deux supports sont utilisés et aucune tendance ne se dessine clairement.

L'un des avantages des filtres d'ADNc est tout d'abord d'ordre financier, que ce soit dans la mise en oeuvre ou dans l'acquisition du matériel nécessaire au déroulement complet de la procédure. De nombreux laboratoires disposent déjà des plateformes de travail permettant de réaliser des hybridations de type *Southern* ou *northern* ainsi que de manipuler des radioéléments. La transposition d'une telle plateforme à l'utilisation de filtres d'ADNc nécessite quelques mises au point des techniques d'hybridation (optimisation des tampons d'hybridation et de lavage, choix du radioélément ^{33}P pour les filtres à haute densité d'ADNc) et l'acquisition d'un appareil permettant une numérisation avec une bonne résolution. Concernant la confection des filtres, des systèmes de dépôts manuels peu onéreux sont disponibles et de nombreuses sociétés proposent parmi leurs services le dépôt d'ADNc amplifiés à l'aide de robots ou encore des filtres prêts à l'emploi pour certains organismes modèles (*Levure*, *Arabidopsis*, Homme...). Le fait de pouvoir déposer manuellement jusqu'à plusieurs centaines d'ADNc sur les filtres permet de pouvoir envisager des tests d'expression génique en routine sur certains types d'expérience. Un autre avantage des filtres d'ADNc est que l'on peut les utiliser plusieurs fois pour diverses expériences : il est possible de déshybrider les acides nucléiques cibles qui ont été hybridés aux ADNc sondes. Ces filtres peuvent ainsi être utilisés de 2 à 6 fois en fonction de leur composition.

Cependant, le verre a des avantages sur le nylon. C'est un matériel durable qui supporte les hautes températures et les lavages successifs. C'est un support non poreux et le volume d'hybridation peut être ainsi minimum, ce qui augmente les cinétiques d'hybridation entre cibles et sondes. Le verre n'a pas de contribution significative au bruit de fond du fait de sa faible fluorescence. Enfin, le verre permet l'hybridation sur le réseau de deux ou plusieurs sondes marquées avec différents fluorophores pour des analyses sérielles ou parallèles. Ainsi une analyse comparative d'une situation d'intérêt avec une situation de référence peut être réalisée directement sur une seule lame de verre.

Applications

Profils d'expression génique

L'application principale de cette technique est l'étude de l'expression des gènes. Ainsi sont générés des profils d'expression ou des matrices de régulation lorsque différentes conditions expérimentales sont comparées à un matériel biologique de référence. Cette technique est extrêmement puissante en termes de diagnostic et permet de suivre par exemple l'effet de stress biotiques ou abiotiques sur un matériel biologique donné. De même, des cinétiques d'expression peuvent être réalisées.

Détection de régulateurs

Le regroupement des gènes suivant leur profil d'expression permet de supposer que les gènes regroupés présentent une expression coordonnée. Ils répondent de la même manière aux différentes conditions biologiques testées et peuvent constituer ce qu'on appelle des

régulons (groupe de gènes soumis à l'action d'un même facteur transcriptionnel). Cela suppose la présence de séquences régulatrices communes à proximité de ces gènes (promoteurs...).

Ces regroupements en régulons permettent également d'entrevoir la fonction ou la catégorie cellulaire potentielle d'un gène dont la fonction codée est inconnue, si ce gène est regroupé avec un ensemble de gènes dont la fonction est connue (p. ex. une même voie métabolique).

Autres applications

Les réseaux d'ADN peuvent avoir des utilisations pour l'analyse des génomes en corrélation avec la découverte de nouvelles molécules médicamenteuses (médecine) ou insecticides (agronomie). Par exemple, des réseaux d'oligonucléotides sont utilisables pour la recherche de polymorphisme ou de mutations dans un génome entier, notamment de SNP (voir Fiche 56). Les puces à ADN peuvent intervenir à différents niveaux dans la recherche de nouvelles molécules actives, en permettant l'approfondissement de voies métaboliques données grâce à l'identification de séries de gènes co-régulés. Lorsqu'une nouvelle molécule doit entrer en phase de test, les puces à ADN peuvent servir à étudier les profils d'expression génique sous l'effet de cette molécule afin d'identifier ses éventuels effets non désirés. Ce domaine en expansion a récemment été appelé "pharmacogénomique".

1. La terminologie concernant ces techniques n'est pas encore totalement fixée. Les Anglo-Saxons parlent de "macro-" ou "micro-arrays", de "DNA chips". Le terme de "puces à ADN" ou de "bio-puces" peut être considéré comme générique. Comme il existe deux procédés majeurs de fabrication de puces, on peut distinguer : (i) les macro-réseaux (sur filtres d'un dm² et plus et contenant moins de 100 spots par cm²) et les micro-réseaux (sur lames de verre, de quelques cm² et contenant plus de 500 spots par cm²) avec dépôt direct des molécules d'ADNc sur le support ; (ii) les puces à oligonucléotides avec synthèse *in situ* des sondes oligonucléotidiques sur le support.

2. La terminologie "sonde" et "cible" dans les expériences d'hybridation moléculaire est très précise. Les cibles correspondent aux molécules d'acides nucléiques (ici les ARNm) que l'on veut détecter : ces cibles se trouvent dans un mélange complexe. Les sondes sont les molécules d'acides nucléiques qui vont servir à piéger par hybridation les molécules ciblées : les sondes sont des molécules purifiées. Dans le cas des réseaux d'ADN, les sondes sont fixées au support et ne sont pas marquées, alors que les cibles sont en solution et sont marquées. Dans le cas des hybridations de type *northern*, les cibles sont fixées sur les membranes de nylon et ne sont pas marquées, alors que la sonde est en solution et marquée. On peut comparer les réseaux d'ADN à un "*northern blot* inversé".

SÉQUENÇAGE D'ADN

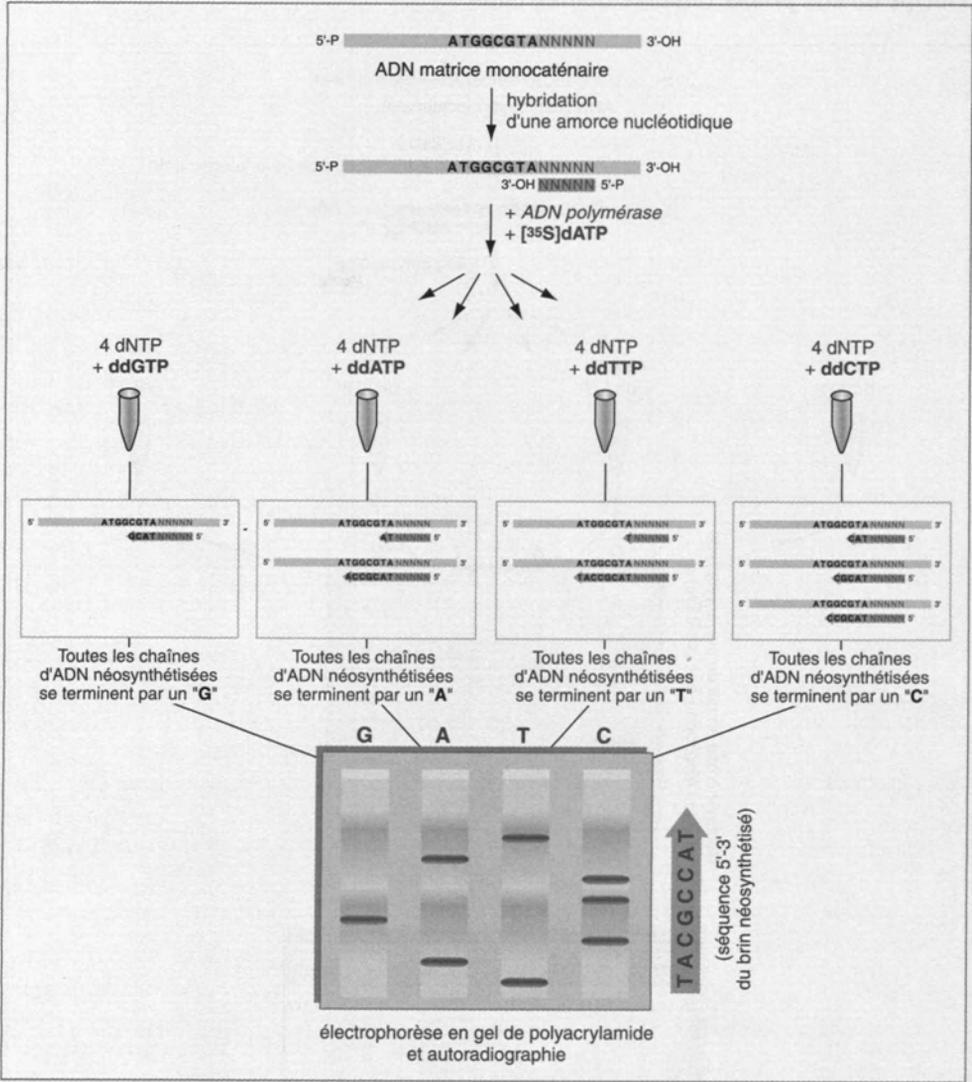
La caractérisation fine d'un gène passe par son séquençage, c'est-à-dire par la connaissance du nombre, de la nature et de l'ordre des nucléotides qui le composent. Cet agencement permet de connaître par exemple l'emplacement des différents sites de restriction du gène afin de mieux le manipuler. Enfin, la traduction *in silico* (par l'ordinateur) de la séquence nucléotidique en séquence d'acides aminés permet éventuellement de confirmer ou de proposer une fonction de la protéine codée par le gène.

Plusieurs techniques de séquençage d'ADN existent, mais nous ne décrivons ici que le principe d'un séquençage enzymatique par incorporation de didésoxynucléotides. Les didésoxynucléotides (ddNTPs) sont des désoxynucléotides modifiés, capables de s'intégrer dans une chaîne d'ADN en synthèse, mais empêchant l'incorporation du nucléotide suivant.

Principe

Une amorce nucléotidique est hybridée au fragment d'ADN monocaténaire dont on veut déterminer la séquence. A partir de l'extrémité 3'-OH de l'amorce appariée, une ADN polymérase synthétise le brin complémentaire de l'ADN matrice en présence de désoxynucléotides. L'un des désoxynucléotides doit être marqué (radioactif ou fluorescent). Le mélange est ensuite réparti dans 4 tubes marqués A, C, G et T. Chacun de ces tubes contient, en plus des dNTP, le ddNTP correspondant. Dans chaque tube, le ddNTP ajouté est incorporé dans les fragments en cours d'élongation.

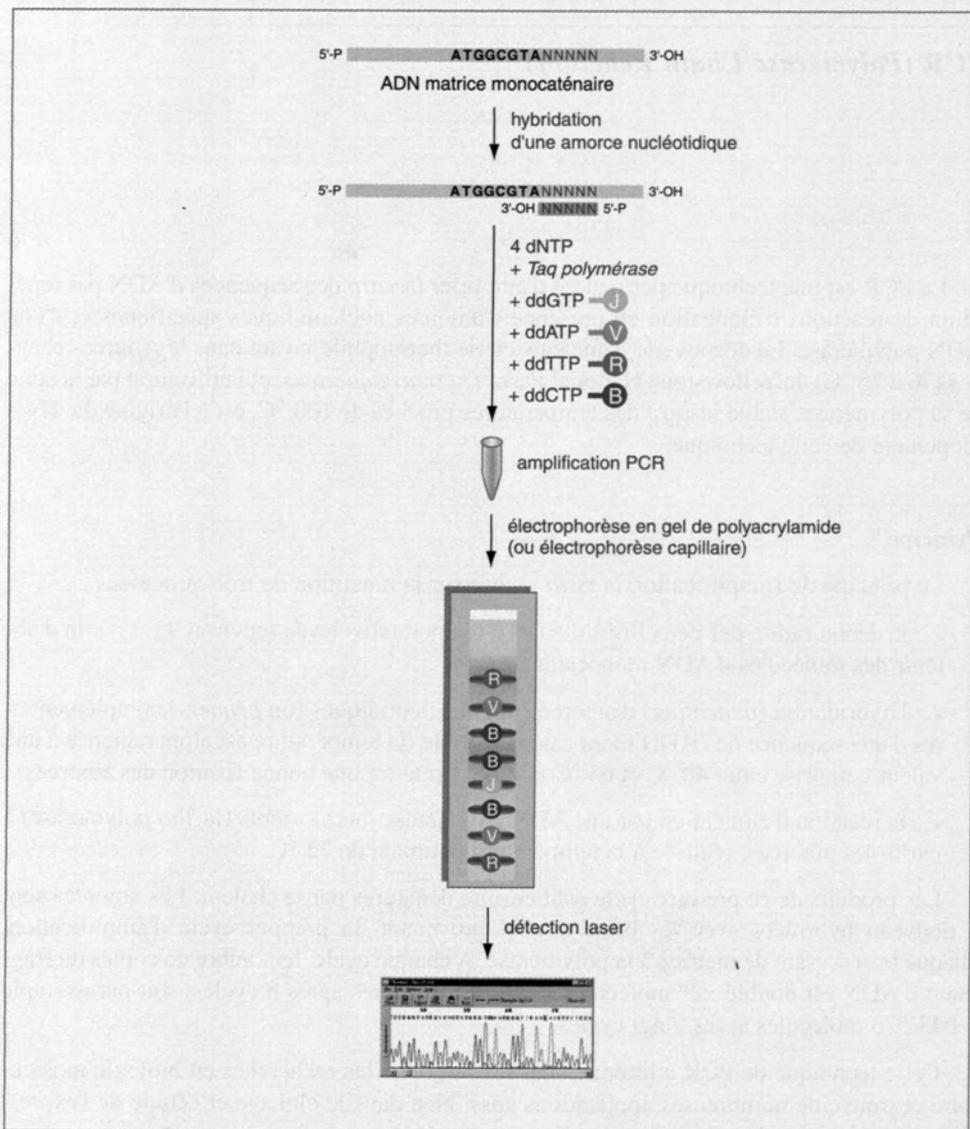
Chaque fois qu'un ddNTP est incorporé à une position, l'élongation de la chaîne est stoppée, ce qui génère un ensemble de molécules de tailles différentes, mais se terminant toutes par le même ddNTP. Si le dNTP et le ddNTP correspondant sont ajoutés en quantités adéquates, tous les fragments d'ADN synthétisés *de novo* et se terminant par ce ddNTP seront représentés. Le contenu des 4 tubes A, C, G et T est ensuite analysé par électrophorèse en gel d'acrylamide en conditions dénaturantes (ou en électrophorèse capillaire). Les différents fragments marqués néosynthétisés migrent en fonction de leur taille et sont séparés à la base près. Une autoradiographie (cas des nucléotides radioactifs) ou une lecture en fluorescence (cas des nucléotides fluorescents) est réalisée après migration des molécules dans le gel. La séquence 5'-P → 3'-OH du brin néosynthétisé peut être lue du bas vers le haut, en comparant les positions relatives des bandes dans les 4 pistes.



Séquençage automatique

Lors de ces analyses, la lecture du gel et l'acquisition des données sont automatiques. Dans ce cas, le marquage des molécules se fait à l'aide de marqueurs fluorescents. Lors de la migration sur gel, les échantillons sont détectés à leur sortie par un laser qui identifie la molécule marquée. De nouveaux systèmes fondés sur l'électrophorèse capillaire permettent une meilleure automatisation du système : les échantillons sont préparés manuellement puis déposés sur un chargeur automatique. L'appareil prélève automatiquement l'échantillon et le place sur le capillaire. Il n'est plus nécessaire de préparer des gels d'acrylamide. Les séquenceurs automatiques peuvent analyser en quelques heures jusqu'à 96 échantillons. L'automatisation peut également concerner la réaction de séquençage qui se fait de plus en plus par des robots couplés à des machines PCR.

Principe du *Dye terminator* (nucléotides fluorescents)



Dans la méthode dite du *Dye terminator*, ce sont les ddNTP qui sont marqués chacun avec un fluorophore spécifique. Tout se déroule selon le même principe que pour la méthode dite du *Dye primer* à la différence près que la réaction se fait dans un seul tube. L'avantage de la méthode dite du *Dye terminator* est que l'on peut utiliser n'importe quelle amorce pour la réaction de séquence. Il est donc possible de séquencer un long fragment d'ADN cloné en choisissant l'amorce suivante à partir de la fin de la réaction de séquence précédente. C'est la méthode actuellement la plus utilisée.

PCR (*Polymerase Chain Reaction*)

La PCR est une technique permettant d'amplifier *in vitro* des séquences d'ADN par répétition de réactions d'élongation en présence d'amorces nucléotidiques spécifiques et d'une ADN polymérase. La découverte d'une eubactérie thermophile vivant dans les sources chaudes (70 à 75 °C) du Yellowstone National Park, *Thermus aquaticus*, et l'utilisation par la suite de sa polymérase, stable jusqu'à des températures proches de 100 °C, est à l'origine du développement de cette technique.

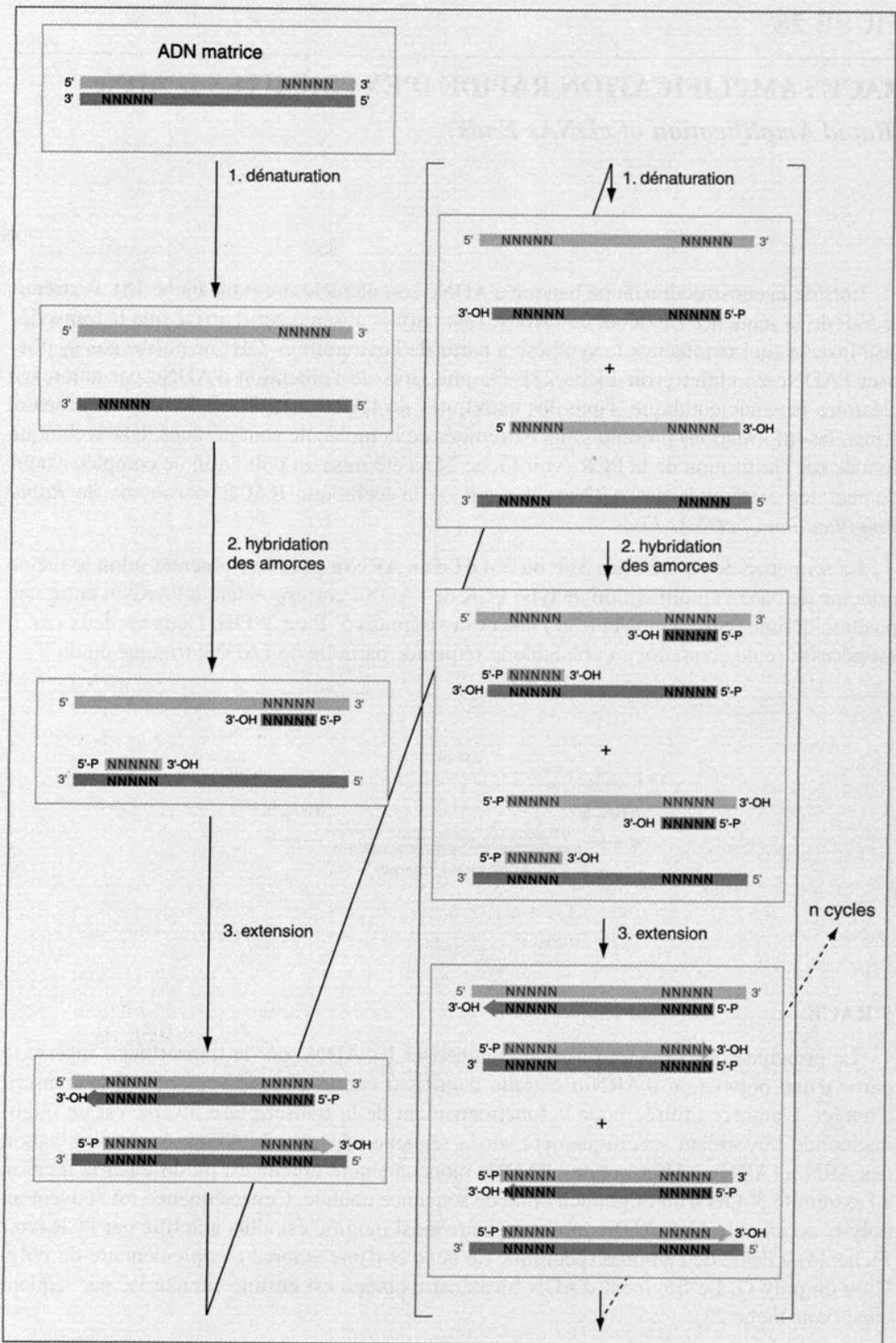
Principe

Le principe de l'amplification *in vitro* repose sur la répétition de trois processus :

- la dénaturation des deux brins d'ADN à température élevée (environ 95 °C) afin d'obtenir des molécules d'ADN monocaténaïres ;
- l'hybridation (*annealing*) d'amorces oligonucléotidiques (ou *primers*) complémentaires d'une séquence de l'ADN monocaténaire cible (la température est alors ramenée à une valeur comprise entre 40 °C et 65 °C afin de permettre une bonne fixation des amorces) ;
- la réaction d'élongation par une ADN polymérase thermostable (la *Taq* polymérase) à partir des amorces, réalisée à la température optimale de 72 °C.

Les produits de ce premier cycle sont ensuite dénaturés par la chaleur. Les amorces sont à nouveau hybridées avec les brins d'ADN provenant du premier cycle d'amplification, chaque brin servant de matrice à la polymérase. A chaque cycle, le nombre de copies du fragment d'ADN est doublé : 2^n molécules sont ainsi obtenues après n cycles, soit par exemple 1 048 576 molécules après vingt cycles.

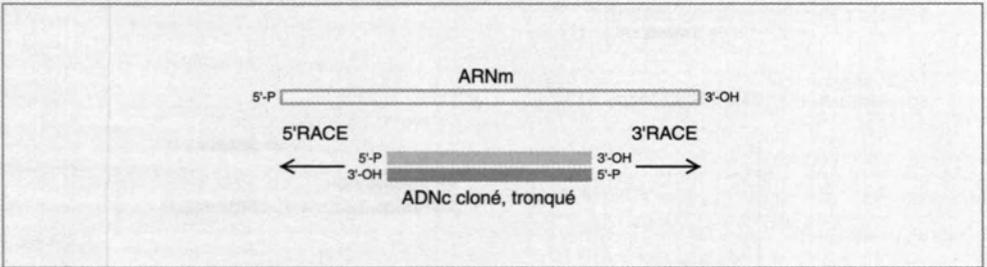
Cette technique de PCR a littéralement révolutionné les recherches en biologie moléculaire et trouve de nombreuses applications aussi bien dans le clonage et l'étude de l'expression des gènes que dans la recherche d'un polymorphisme génétique.



RACE : AMPLIFICATION RAPIDE D'EXTRÉMITÉS D'ADNc
(Rapid Amplification of cDNAs Ends)

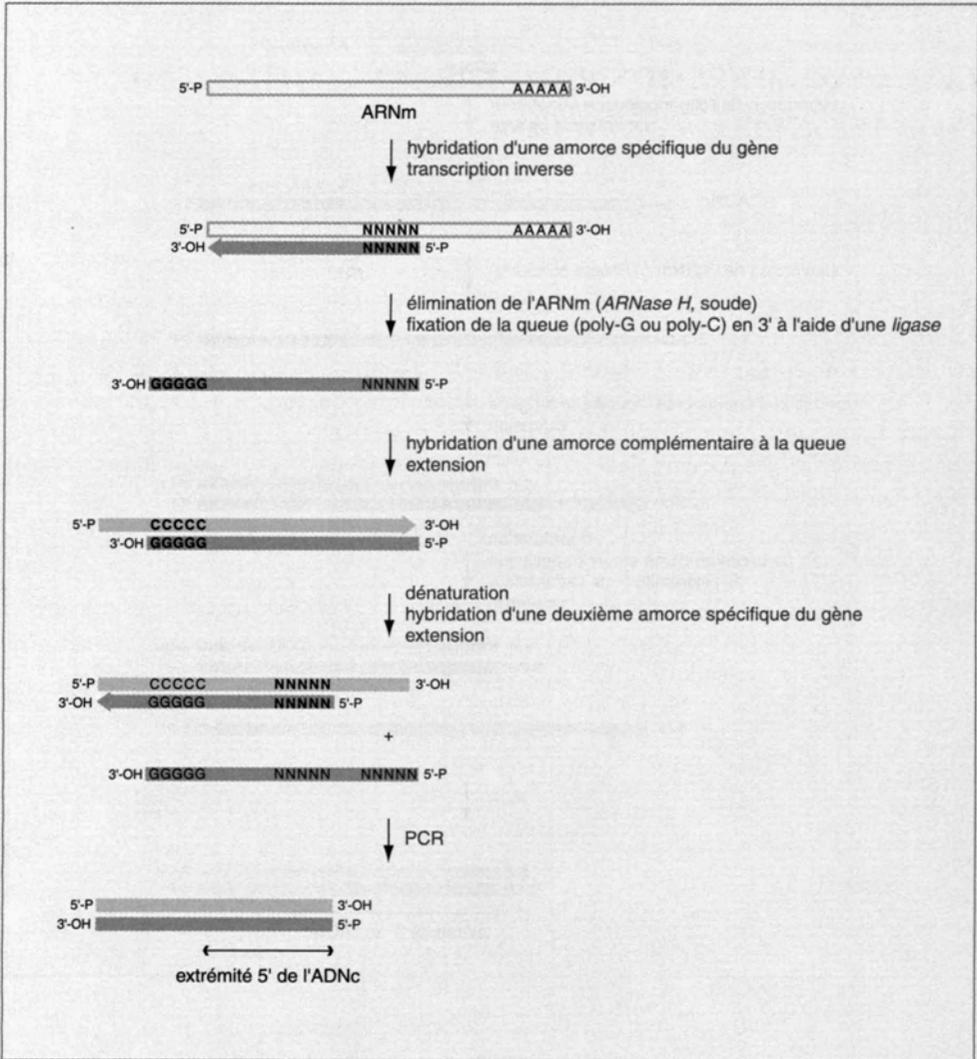
Lors de la construction d'une banque d'ADNc complémentaire (voir Fiche 15), l'extrémité 5'-P de la séquence (le début de l'ADNc) est parfois absente car il arrive que la transcriptase inverse (qui commence la synthèse à partir de l'extrémité 3'-OH) ne puisse pas synthétiser l'ADNc en entier (voir Fiche 27). De plus, lors de l'obtention d'ADNc par amorçage aléatoire hexanucléotidique, l'une des extrémités de l'ARNm (5'-P ou 3'-OH) est absente. Ainsi, les informations présentes aux extrémités de la molécule sont perdues. Une technique fondée sur l'utilisation de la PCR (voir Fiche 24) a été mise au point afin de compléter rapidement les extrémités des ARNm. Il s'agit de la technique RACE, acronyme de *Rapid Amplification of cDNAs Ends*.

La séquence des extrémités 5'-P ou 3'-OH d'un ARNm peut être obtenue selon le même principe de base : amplification de type PCR de l'ADNc correspondant à l'ARNm entre une position définie dans la molécule et l'une des extrémités 5'-P ou 3'-OH. Dans les deux cas, il est nécessaire de connaître au préalable la séquence partielle de l'ADNc tronqué étudié.



5'RACE

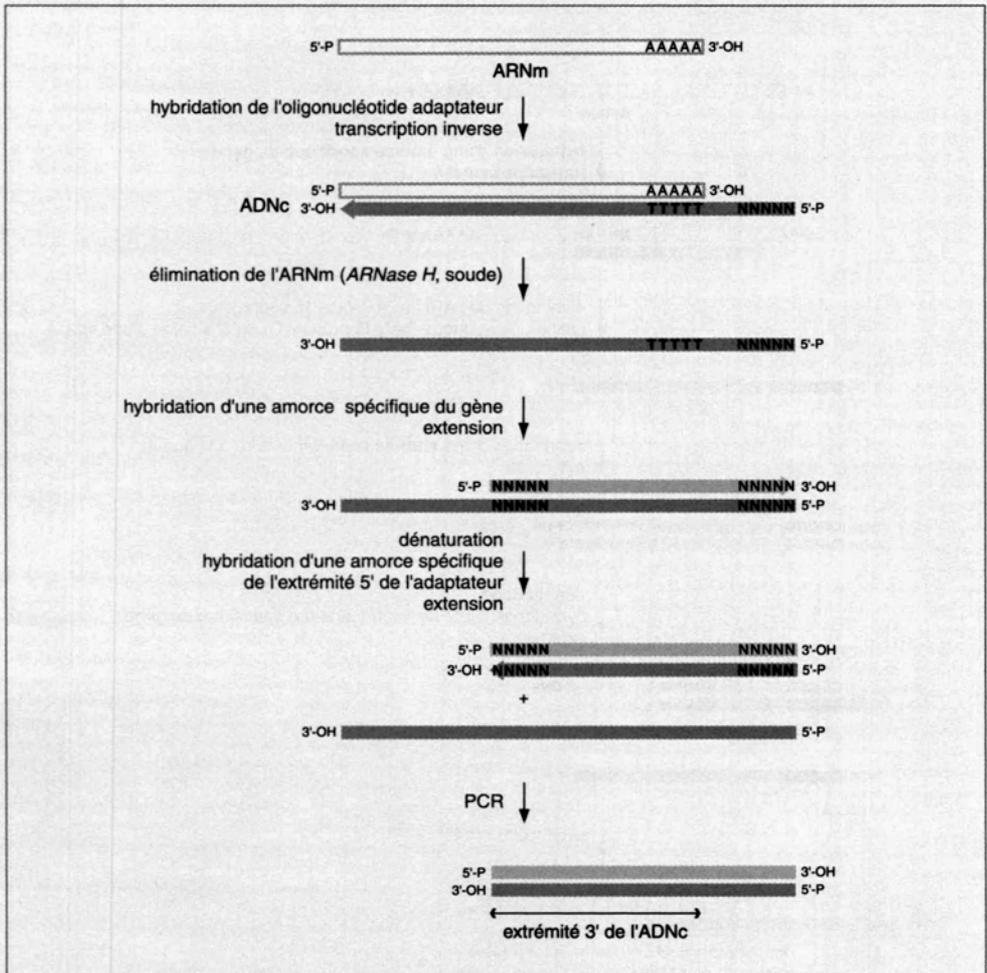
Le principe consiste à tout d'abord synthétiser les ADNc par la transcriptase inverse, à partir d'une population d'ARNm extraite d'un tissu connu comme accumulant le transcrit d'intérêt. L'amorce utilisée pour le fonctionnement de la transcriptase inverse est un oligonucléotide s'hybridant spécifiquement sur la séquence du gène à cloner. Après élimination des ARNm (ARNase H ou soude), l'ADNc monocaténaire obtenu est modifié par la fixation à l'extrémité 3'-OH d'un oligonucléotide de séquence connue. Cette séquence est souvent un poly-C ou un poly-G. L'ADNc monocaténaire ainsi modifié est alors amplifié par PCR (voir Fiche 24) à l'aide de l'amorce spécifique du gène et d'une amorce complémentaire du poly-C ou du poly-G. Le fragment d'ADN bicaténaire obtenu est ensuite caractérisé par séquençage (voir Fiche 23).



3'RACE

Le principe consiste à :

- synthétiser les brins d'ADNc par la transcriptase inverse en utilisant comme amorce un oligonucléotide poly-T contenant à son extrémité 5'-P une séquence connue supplémentaire (oligonucléotide adaptateur) et s'hybridant sur l'extrémité poly-A de l'ARNm ;
- amplifier par PCR (voir Fiche 24) un fragment d'ADNc à l'aide d'une première amorce s'hybridant sur la séquence supplémentaire de l'oligonucléotide adaptateur et d'une seconde amorce s'hybridant spécifiquement sur le gène. Le fragment d'ADN bicaténaire obtenu est ensuite caractérisé par séquençage.



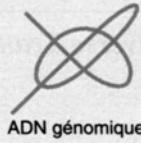
MARCHE GÉNOMIQUE PAR PCR

La marche génomique par PCR permet de cloner les séquences génomiques flanquant une séquence d'ADN connue. Elle est utile par exemple dans le cas d'un gène étiqueté par un transposon (ou, pour les plantes, une insertion d'ADN-T) (voir Fiche 45), ou encore pour cloner le promoteur d'un gène dont on ne possède que l'ADNc. Cette technique permet de caractériser aussi bien la séquence située en amont (5'-P) que la séquence située en aval (3'-OH) de la séquence connue. C'est une approche rapide, très sensible, spécifique et qui nécessite peu d'ADN génomique. Elle présente l'avantage de pouvoir travailler avec l'ADN d'un individu (p. ex. un mutant) pour lequel une banque génomique n'est pas toujours disponible.

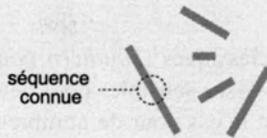
L'ADN génomique est digéré par une enzyme de restriction laissant des bouts francs. Un adaptateur présentant un brin plus long que l'autre est ligaturé à chaque extrémité¹. Cet ADN ainsi couplé est utilisé comme matrice dans une réaction de PCR. La première amorce est spécifique de la séquence d'ADN connue. Selon son orientation (sens ou antisens), son extension se fera vers l'amont ou vers l'aval de cette séquence (vers l'amont dans la figure). La seconde amorce a la même séquence que le brin long de l'adaptateur. Cette amorce ne pourra s'hybrider que si un brin d'ADN a préalablement été recopié à partir de la première amorce spécifique de la séquence connue. Au cours des cycles suivants, on aura donc une amplification spécifique de la séquence comprise entre l'amorce spécifique et l'adaptateur. La longueur de cette séquence va dépendre de la position du site de restriction (de l'enzyme utilisée pour digérer l'ADN) le plus proche. On utilise en général 4 ou 5 enzymes différentes afin d'avoir des produits PCR de tailles différentes.

A partir de cette nouvelle séquence, on peut dessiner une nouvelle amorce et recommencer la marche afin d'aller plus loin. On réutilise alors la même population de fragments d'ADN génomique couplés aux adaptateurs, qui constitue ainsi une sorte de banque génomique pour la marche.

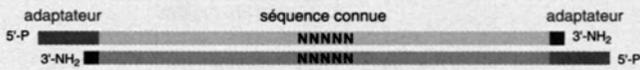
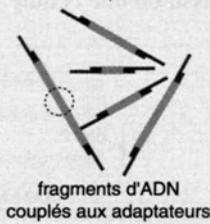
1. Le 3'-OH du brin court de l'adaptateur est bloqué par un groupement amine afin d'éviter au cours de la PCR une élongation non spécifique de ce brin en regard du brin long.



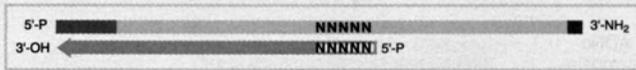
enzyme de restriction



ligature des adaptateurs



dénaturation
hybridation de l'amorce spécifique
élongation



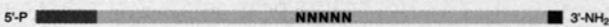
+

3'-NH₂

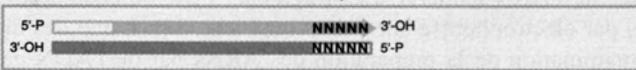
NNNNN

5'-P

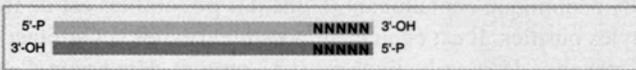
dénaturation
hybridation de l'amorce de l'adaptateur
élongation



+



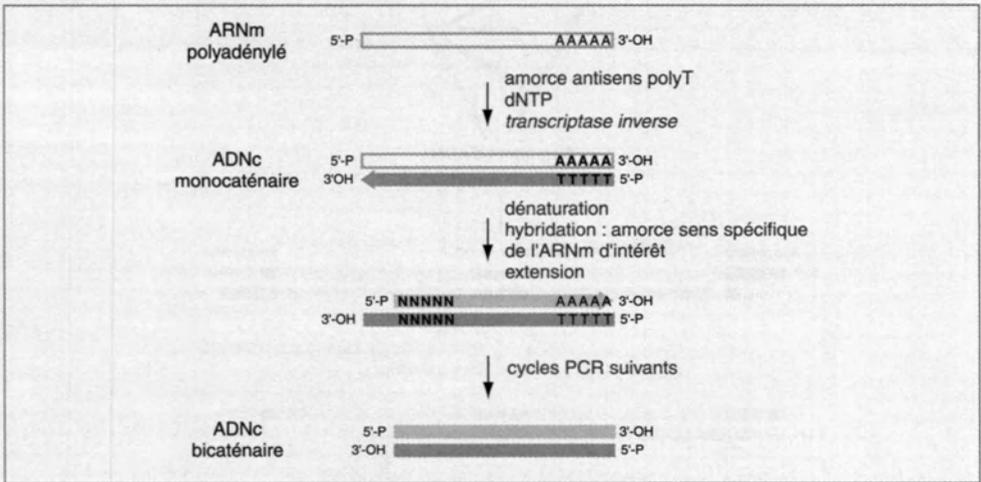
cycles suivants de PCR



produit PCR final contenant une partie de la séquence connue
et la séquence génomique en amont

RT-PCR : PCR SUR ARNm (*Reverse Transcriptase PCR*)

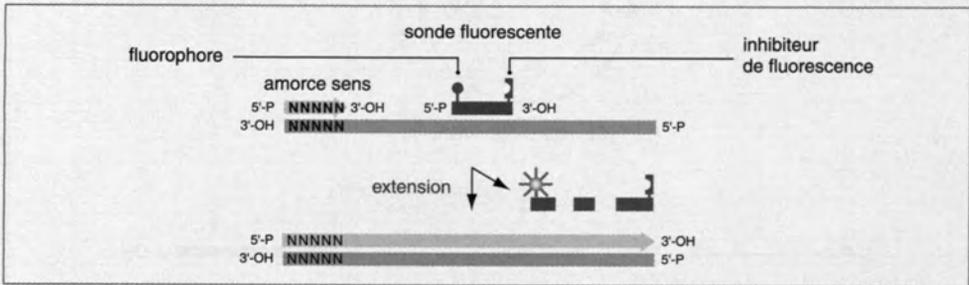
Les hybridations moléculaires dites classiques (*Southern* pour l'ADN et *northern* pour l'ARN, voir Fiche 12) ne sont parfois pas assez sensibles pour détecter des quantités infimes d'ADN ou d'ARN cible. C'est notamment le cas pour de nombreux ARNm qui ne sont présents dans les cellules qu'en très petit nombre. L'usage de la PCR permet de travailler avec de toutes petites quantités d'acides nucléiques (voir Fiche 24). Une technique, appelée RT-PCR, a été développée pour permettre la détection et la mise en évidence de l'accumulation d'un ARNm rare dans un organe, un tissu ou une cellule.



Le principe consiste à extraire les ARN totaux des tissus étudiés et les copier *in vitro* en ADNc monocaténaire, grâce à l'action de la transcriptase inverse. Les molécules d'ADNc obtenues servent alors de matrice à une réaction PCR utilisant un couple d'amorces spécifiques de la séquence de l'ARN d'intérêt. Les fragments PCR obtenus après les cycles de PCR sont visualisés par électrophorèse sur gel¹ (voir Fiche 24). L'une des difficultés de la technique est la contamination de la préparation des ARNs par de l'ADN génomique : en effet, les amorces nucléotidiques se fixent tout aussi bien sur l'ADN monocaténaire issu des ARNs que sur l'ADN génomique contaminant. L'une des possibilités est de travailler à partir d'ARN polyadénylés purifiés. Il est également possible d'éviter cet artéfact en traitant les échantillons d'ARN par une désoxyribonucléase (très pure et dépourvue d'activité ribonucléase susceptible de dégrader les ARN) afin d'éliminer toute trace d'ADN génomique. D'autre part, lorsque c'est possible, on peut choisir un couple d'amorces qui encadre une séquence intronique : si de l'ADN génomique contamine la préparation, le fragment d'amplification obtenu aura une taille plus importante que celle du produit d'amplification obtenu à partir de l'ARNm.

Une modification de cette technique de base permet d'évaluer la quantité relative d'un transcrit donné dans plusieurs échantillons et ainsi d'évaluer le niveau d'accumulation de ce transcrit à la suite de différents traitements ou dans différents tissus. La difficulté de l'interprétation vient du fait qu'il faut s'assurer : (i) que l'on part d'une même quantité d'ARN pour chaque échantillon que l'on veut comparer ; (ii) que l'on est dans une phase linéaire de l'amplification. De nombreux standards internes sont nécessaires, comme par exemple la co-amplification du gène d'intérêt et d'un gène contrôle non régulé par le processus étudié ; dans ce cas, l'amplification se déroule en présence de deux couples d'amorces, l'un reconnaissant le gène d'intérêt et l'autre reconnaissant le gène contrôle.

Il est aussi possible d'évaluer la quantité d'ARN présente dans un échantillon donné : c'est la RT-PCR quantitative. Il faut pour cela utiliser un marqueur fluorescent permettant de détecter les produits PCR au fur et à mesure de leur synthèse. La technique la plus élaborée et la plus spécifique consiste à utiliser comme marqueur une sonde fluorescente complémentaire de l'ARN cible. Cette sonde de 20 à 30 nucléotides est couplée de façon covalente à un fluorophore à l'extrémité 5'-P et à un inhibiteur de fluorescence à l'extrémité 3'-OH. De plus, cette sonde doit s'hybrider à l'ARN cible entre les deux amorces utilisées classiquement au cours d'une PCR classique. Lors de la phase de polymérisation de l'ADN, la sonde est dégradée (déplacée par la molécule d'ADN en synthèse) et il y a émission de fluorescence car la dégradation entraîne sa fragmentation et donc la séparation entre le fluorophore et l'inhibiteur de fluorescence.

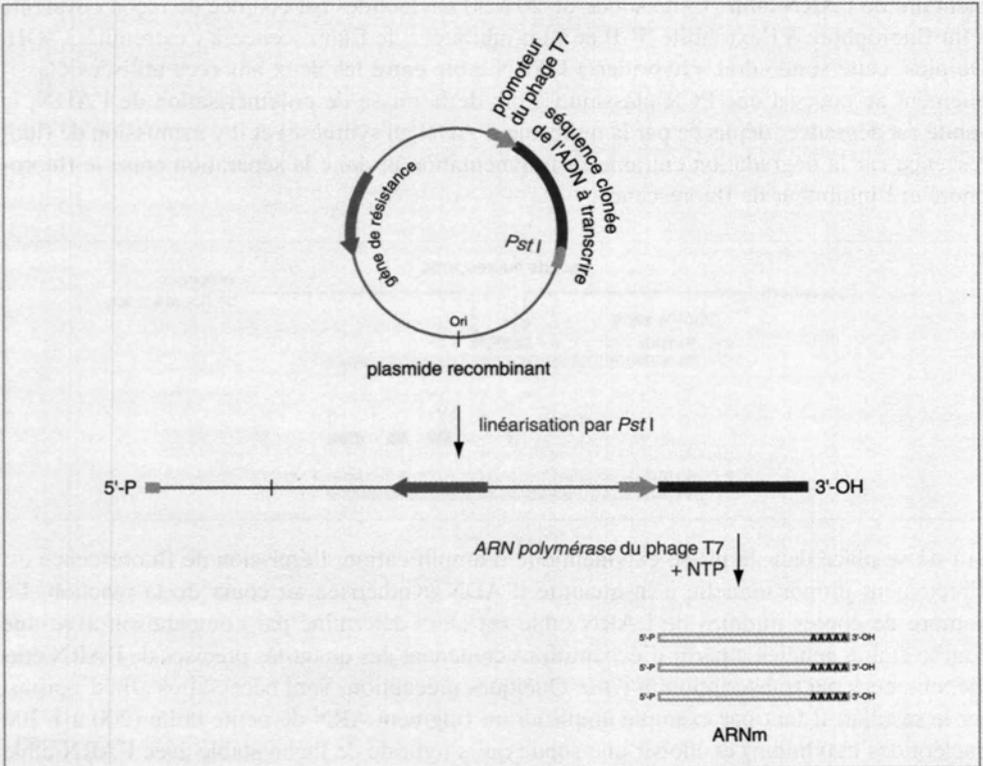


Si l'on se place dans la phase exponentielle d'amplification, l'émission de fluorescence est directement proportionnelle à la quantité d'ADN synthétisée au cours de la réaction. Le nombre de copies initiales de l'ARN cible est alors déterminé par comparaison avec une courbe étalon générée à partir d'échantillons contenant des quantités précises de l'ARN étudié, obtenues par transcription *in vitro*. Quelques précautions sont nécessaires afin d'optimiser le résultat. Il faut par exemple amplifier un fragment ARN de petite taille (200 à 1 200 nucléotides maximum) et choisir une sonde qui s'hybride de façon stable avec l'ARN cible dans les conditions de PCR utilisées. Par ailleurs, il est indispensable de tester deux, voire trois dilutions de l'échantillon à analyser afin d'avoir un résultat compris dans les limites de la gamme étalon. Grâce à la spécificité de cette méthode et en utilisant différents fluorophores, il est possible de suivre le devenir de plusieurs sondes au cours d'une même réaction PCR. L'une des applications de cette possibilité est d'utiliser des sondes allèle-spécifiques pour mettre en évidence le polymorphisme génétique.

1. Le gel peut également par la suite être transféré sur une membrane de nylon et soumis à une hybridation moléculaire pour s'assurer que les produits d'amplification obtenus correspondent à la séquence étudiée.

TRANSCRIPTION *IN VITRO*

La production de molécules d'ARNm pures en grande quantité est parfois utile afin d'étudier par exemple le mécanisme d'épissage des ARN ou d'utiliser des ARN comme sondes d'hybridation.

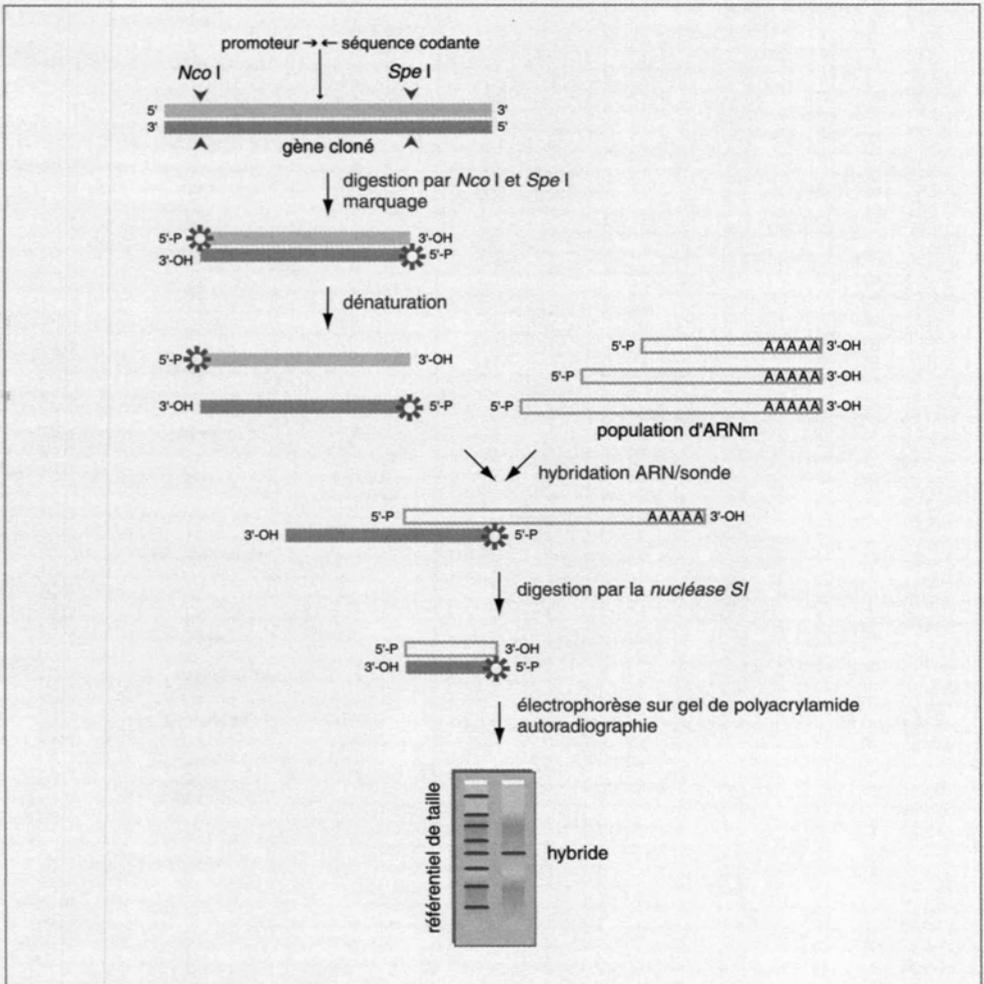


La séquence d'ADN codant pour l'ARN à étudier est clonée dans un plasmide ou un phagemide en aval d'un promoteur utilisé par une ARN polymérase de phage (p. ex. T7, T3 ou SP6). L'ADN servant de matrice est d'abord linéarisé à l'extrémité de la séquence à transcrire par coupure avec une enzyme de restriction. L'ARN polymérase initie ensuite la transcription au promoteur et synthétise le brin d'ARN complémentaire de la séquence d'ADN clonée en présence de ribonucléotides. Si l'un des ribonucléotides ajoutés est marqué (radioactivement ou chimiquement), un ARN marqué est alors produit. La transcription *in vitro* s'arrête lorsque l'ARN polymérase atteint l'extrémité de la molécule de plasmide linéarisée.

LAURENCE DAMIER

DÉTERMINATION DU SITE D'INITIATION DE LA TRANSCRIPTION

Lorsque la séquence d'un gène est connue et clonée dans un vecteur, il est possible de déterminer le site d'initiation de la transcription de ce gène par cartographie à la nucléase S1. Déterminer ce site consiste à identifier la première base d'ARN synthétisée par l'ARN polymérase à partir de la matrice d'ADN (voir Fiche 1). Cette information sur la localisation du début de l'ARNm d'un gène est importante car elle permet de délimiter avec précision (à la base près) la limite entre les séquences promotrices et la séquence codante.



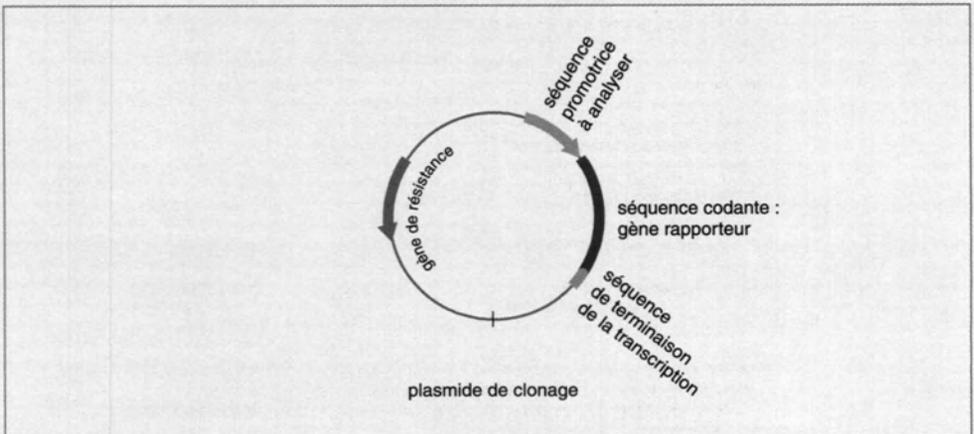
Le principe de cette technique repose sur l'hybridation en solution entre l'ARNm du gène étudié (à partir d'un mélange d'ARNm provenant du tissu étudié) et un fragment de l'ADN génomique de ce gène, comprenant le début de la partie codante et une portion des séquences promotrices. Cette séquence d'ADN est tout d'abord obtenue par restriction enzymatique utilisant des sites de restriction qui encadrent la jonction promoteur/séquence codante. Dans l'exemple donné, il s'agit de *Nco* I et de *Spe* I. C'est en analysant la séquence de cet ADN cloné que l'on détermine le couple d'enzymes approprié. Ce fragment d'ADN isolé est marqué radioactivement à son extrémité 5'-P (voir Fiche 11). L'hybride obtenu se fera entre l'ARNm et la partie codante du gène. Les séquences promotrices, quant à elles, resteront monocaténares. L'utilisation d'une nucléase spécifique des régions monocaténares (nucléase S1) permet d'éliminer ces séquences et de ne garder que l'hybride bicaténaire. La taille de cet hybride est déterminée sur un gel de type séquençage (voir Fiche 23) afin d'identifier sa première base nucléotidique qui représente le site d'initiation de la transcription du gène étudié.

ANALYSE FONCTIONNELLE DE PROMOTEURS

La transcription d'un gène est finement régulée dans le temps et dans l'espace : un ARNm donné ne sera synthétisé que dans certains types cellulaires, sous différents stress biotiques ou abiotiques et à une certaine étape du développement de l'individu. Les informations nécessaires à cette régulation sont portées à la fois par des protéines dites de régulation et par des séquences nucléotidiques placées en amont et en aval de la séquence codante (voir Fiche 1). C'est l'interaction entre les protéines de régulation et ces séquences d'ADN qui décide de la transcription d'un gène. L'analyse fonctionnelle de promoteurs permet de disséquer les régions de l'ADN intervenant dans une régulation spatio-temporelle de l'expression d'un gène.

Principe

L'analyse fonctionnelle d'un promoteur nécessite de connaître la séquence nucléotidique de cette région régulatrice. Une construction chimérique est réalisée, constituée : (i) du promoteur soumis à l'analyse fonctionnelle ; (ii) d'un gène rapporteur¹ ; (iii) des signaux permettant la terminaison de la transcription. Ces gènes chimères sont construits *in vitro* dans un plasmide par des séries de coupure grâce à des enzymes de restriction, des ligatures et des clonages dans *E. coli*.



Le plasmide recombinant portant la construction chimérique est utilisé pour la transformation de cellules soit dans des expériences de transformations stables et d'obtentions d'organismes transgéniques (voir Fiches 34 à 38), soit dans des expériences d'expression transitoire (voir Fiche 38). L'activité enzymatique de la protéine codée par le gène rapporteur est

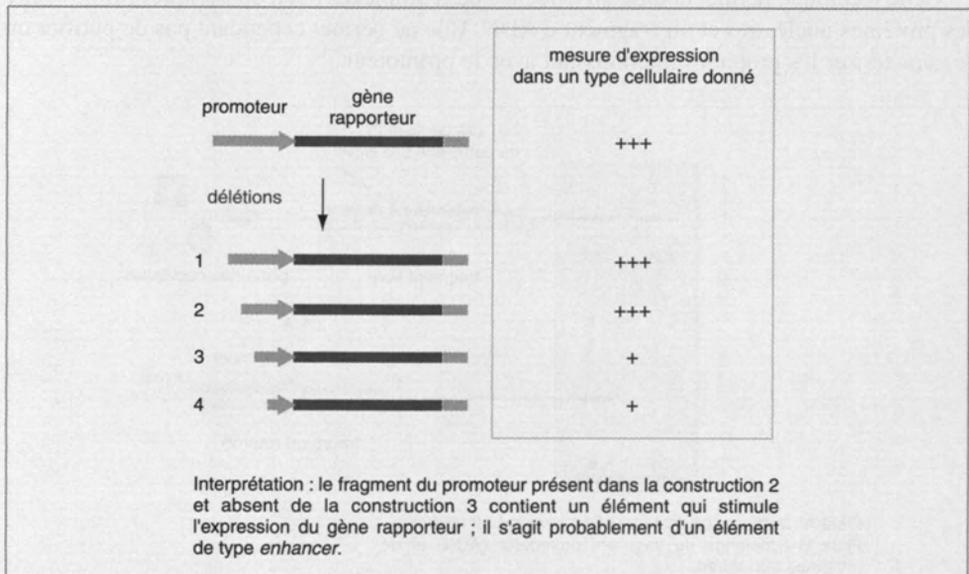
alors mesurée dans les différents tissus ou cellules transformées. Les résultats sont ensuite interprétés de la façon suivante :

- si une activité enzymatique est mesurable dans un tissu donné, cela signifie que la protéine codée par le gène rapporteur a été synthétisée, donc que le gène rapporteur a été transcrit sous le contrôle de la partie promotrice. Ce promoteur est donc actif dans les conditions d'analyse : il contient toutes les séquences nucléotidiques nécessaires à la transcription d'un gène dans le tissu analysé ;
- si aucune activité enzymatique n'est détectée, la protéine codée par le gène rapporteur n'a pas été synthétisée et le gène n'a pas été transcrit. Le promoteur analysé n'est pas fonctionnel dans le tissu transformé. Soit des séquences nucléotidiques font défaut à son bon fonctionnement (séquence de type *enhancer*), soit certaines séquences nucléotidiques présentes dans ce promoteur empêchent une transcription active (séquence de type *silencer*).

En comparant ainsi le fonctionnement d'un promoteur dans différents tissus, à différents stades de développement sous différentes contraintes extérieures, il est possible de caractériser les modalités d'expression spatio-temporelle du promoteur étudié.

Analyse par délétions

Lorsqu'un promoteur dirigeant la transcription d'un gène rapporteur est fonctionnel après intégration dans un tissu donné, il est possible de délimiter plus finement les séquences d'ADN régulant la transcription. Pour ce faire, le promoteur de départ est réduit en taille et chaque nouvelle construction chimérique ainsi obtenue est testée par transformation.



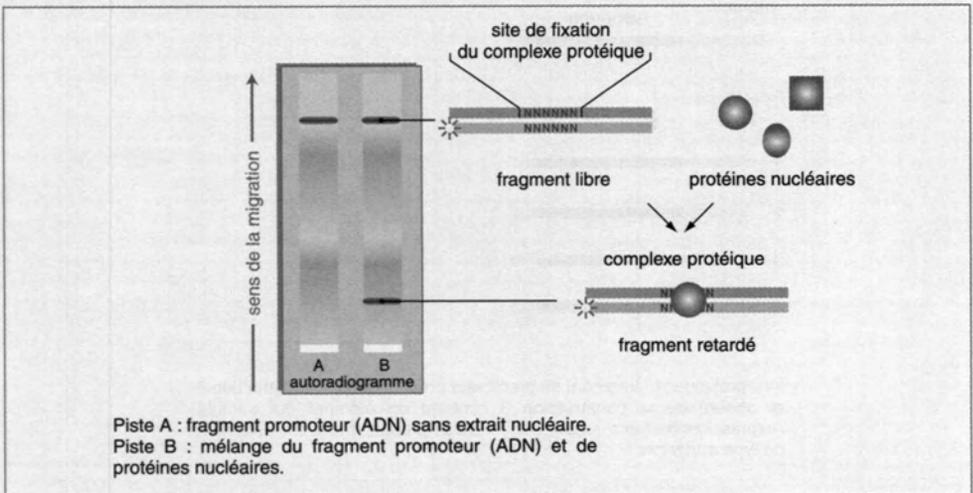
1. Pour une définition d'un gène rapporteur, voir Fiche 38.

GEL-RETARD

L'expression d'un gène dépend de la fixation sur la région promotrice de protéines dites de régulation (facteurs de transcription, voir Fiche 1). La technique de transformation génétique alliée à l'analyse fonctionnelle de promoteurs permet de délimiter des régions nucléotidiques du promoteur susceptibles d'intervenir dans la régulation de l'expression du gène (voir Fiche 30). La technique de gel-retard permet de mettre en évidence les protéines ou les complexes protéiques qui se fixent sur ces régions promotrices.

Le principe consiste à : (i) extraire des protéines à partir des noyaux¹ de différents tissus (généralement un tissu exprimant ce gène et un tissu de référence ne l'exprimant pas) ; (ii) hybrider ces protéines avec le fragment d'ADN radioactif portant la séquence promotrice d'intérêt ; (iii) déposer sur gel non dénaturant le mélange d'hybridation pour séparer les fragments d'ADN par électrophorèse. Si le fragment s'est hybridé avec une protéine nucléaire, un complexe se forme et sa migration sera ralentie par rapport au fragment libre non hybridé à des protéines. Le gel est autoradiographié et la mobilité électrophorétique des fragments d'ADN complexés ou non est révélée.

Cette technique permet la mise en évidence des complexes nucléoprotéiques formés entre des protéines nucléaires et un fragment d'ADN. Elle ne permet cependant pas de purifier ou de caractériser les protéines interagissant avec le promoteur.



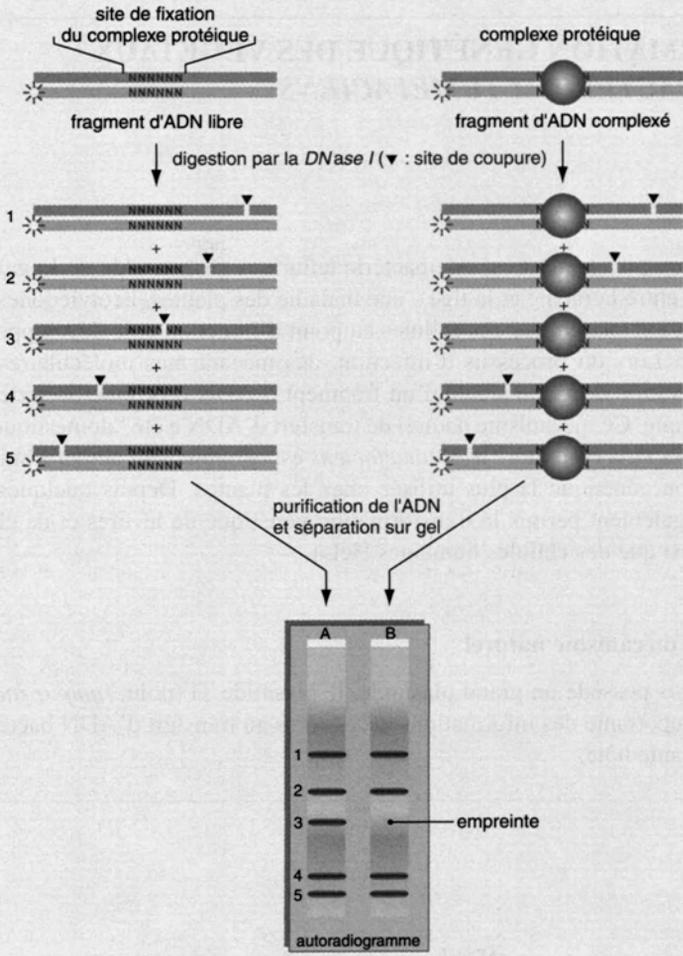
1. Bien que les facteurs de régulation soient nucléaires, il est possible d'utiliser également des protéines totales.

EMPREINTE A LA DNase I (*Footprinting*)

Les expériences d'empreinte à la DNase I font généralement suite aux expériences de "gel retard" (Fiche 31). Elles permettent de définir avec précision les bases du promoteur qui interagissent physiquement avec un facteur protéique nucléaire.

Le fragment contenant la séquence du promoteur – et marqué radioactivement à une de ses extrémités – est hybridé avec des extraits protéiques nucléaires comme précédemment (voir Fiche 31). Puis le mélange est soumis à un traitement modéré à la DNase I qui est une endonucléase capable de couper les liaisons phosphodiester à l'intérieur des chaînes d'ADN sur un seul des deux brins. Il en résulte une collection de fragments de tailles différentes, dont certains sont marqués à l'une de leurs extrémités. La présence d'un complexe protéique cache un site éventuel de coupure par la DNase I et empêche l'action de cette enzyme. Les produits de digestion sont alors séparés par électrophorèse sur un gel d'acrylamide à grand pouvoir séparateur (de même type que les gels de séquence), puis détectés par autoradiographie. En comparant les bandes obtenues pour un fragment d'ADN coupé, en présence ou en absence de facteurs protéiques, on met en évidence une région sans bande correspondant au site de fixation du facteur protéique, qui par sa liaison à l'ADN a protégé celui-ci des coupures à la DNase I. En réalisant un séquençage du fragment étudié et en faisant migrer les produits de réaction parallèlement à ceux obtenus lors d'une expérience d'empreinte à la DNase I, il sera alors possible de déterminer la séquence du site de fixation du facteur protéique.

D'autres techniques (non développées ici) permettent de confirmer et d'affiner les résultats obtenus par les empreintes à la DNase I (empreintes chimiques, protection par la méthylation...). Cependant, les expériences qui mettent en évidence la fixation de protéines *in vitro* ne reflètent pas obligatoirement la situation *in vivo*. Au sein de la cellule, les concentrations des protéines régulatrices et de leur cible d'ADN sont très différentes de celles utilisées *in vitro* et il y a notamment un large excès d'ADN non ligand. De plus, l'ADN est fortement empaqueté dans le noyau sous forme de chromatine et beaucoup d'autres protéines peuvent interférer avec la fixation d'une protéine particulière sur sa séquence d'ADN cible. Ainsi, des techniques dites *in vivo* permettent d'analyser la fixation de protéines régulatrices sur des portions d'ADN, sans trop perturber l'intégrité de la cellule. Les méthodes consistent à modifier l'ADN *in situ* par U.V., chimiquement ou par des enzymes (nucléases) dans des conditions douces. L'ADN est ensuite isolé et les nucléotides modifiés sont détectés sur gel de séquence.



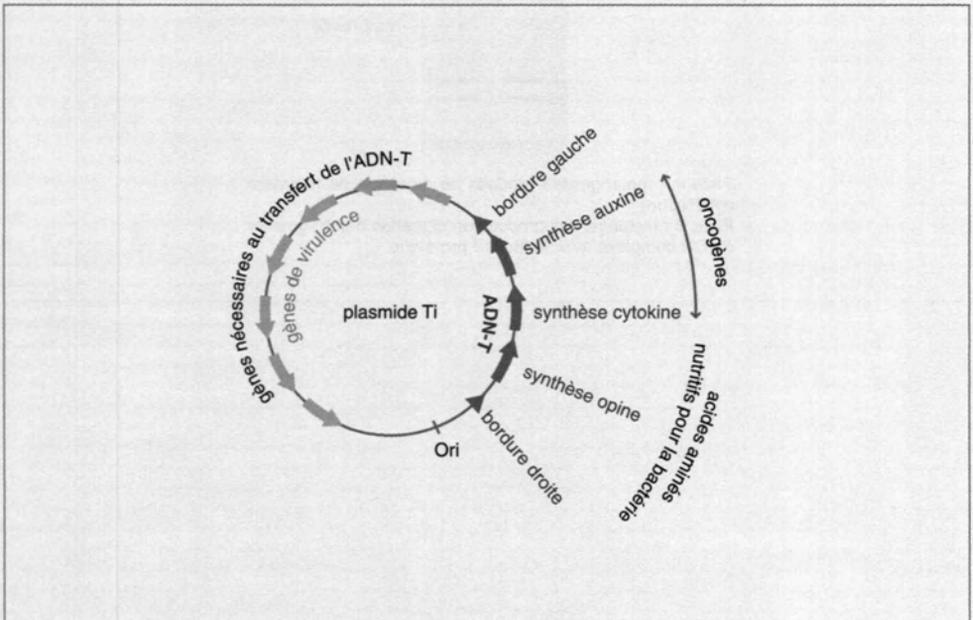
Piste A : mélange des produits de digestion du fragment d'ADN libre
 Piste B : mélange des produits de digestion des fragments d'ADN complexé avec le facteur protéique

TRANSFORMATION GÉNÉTIQUE DES VÉGÉTAUX PAR *AGROBACTERIUM TUMEFACIENS*

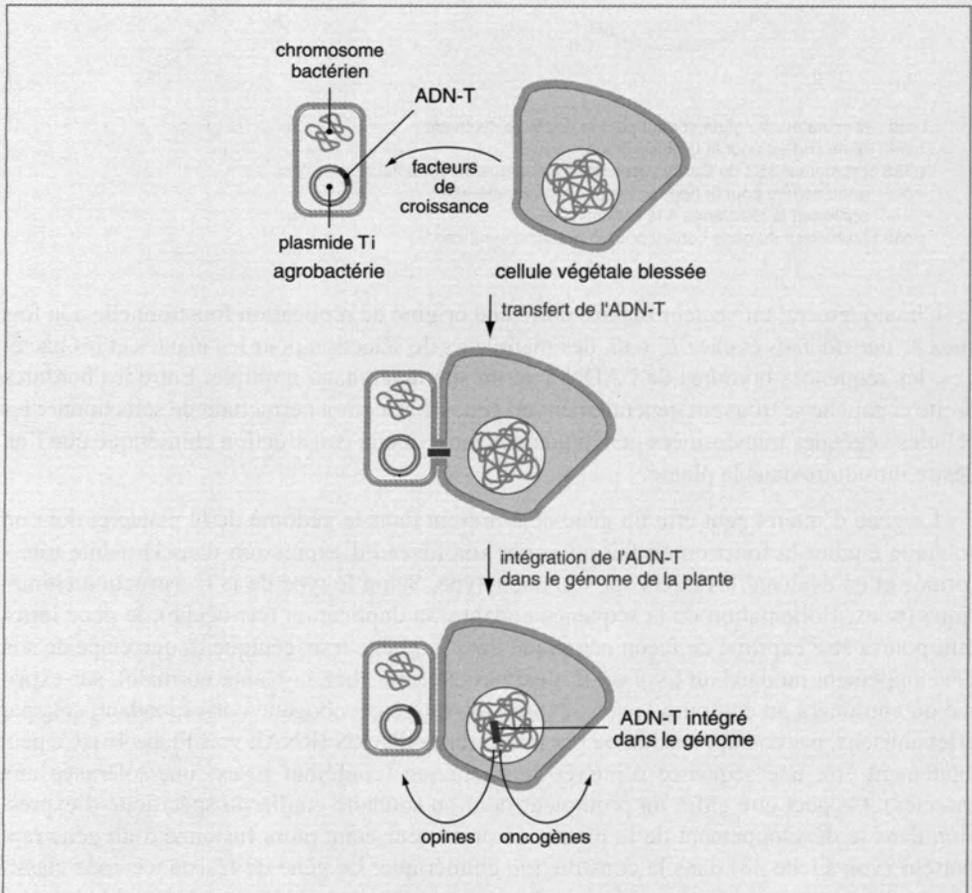
Agrobacterium tumefaciens est une bactérie tellurique responsable de la galle du collet (zone de liaison entre la racine et la tige), une maladie des plantes dicotylédones, qui se traduit par une division anarchique des cellules au point d'infection et le développement d'une tumeur végétale. Lors du processus d'infection, des mécanismes moléculaires complexes mènent à l'intégration de façon stable d'un fragment d'ADN d'origine bactérienne dans le génome de la plante. Ce mécanisme naturel de transfert d'ADN a été "domestiqué" à des fins biotechnologiques et la co-culture d'*A. tumefaciens* est certainement aujourd'hui la méthode de transformation génétique la plus utilisée chez les plantes. Depuis quelques années, *A. tumefaciens* a également permis la transformation génétique de levures et de champignons filamenteux, ainsi que des cellules humaines HeLa.

Description du mécanisme naturel

A. tumefaciens possède un grand plasmide, le plasmide Ti (pour *Tumour inducing*), qui porte une part importante des informations nécessaires au transfert d'ADN bactérien dans le génome de la plante hôte.

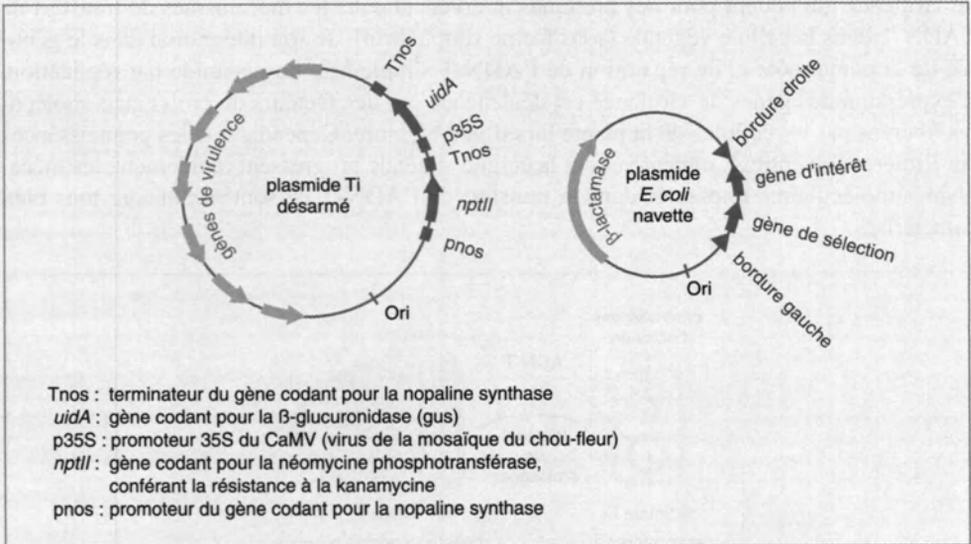


La région T, présente sur ce plasmide Ti, correspond à la séquence d'ADN qui sera transférée dans le génome végétal sous le nom d'ADN-T. Cette région est délimitée par 2 répétitions directes de 25 paires de bases, les bordures. Entre ces bordures sont localisés : (i) les gènes induisant le développement de la tumeur (oncogènes), dont l'expression entraîne des déséquilibres dans la balance hormonale des cellules transformées, changeant ainsi leurs caractéristiques de croissance (gènes des phytohormones auxines et cytokines, responsables de la multiplication incontrôlée des cellules végétales) ; (ii) les gènes qui dirigent la synthèse des opines, acides aminés liés à un sucre utilisables par la bactérie et non par la plante (comme l'octopine, l'agropine et la nopaline). Il a été montré que les séquences d'ADN placées entre les deux bordures sont transférées dans le génome des cellules végétales sous l'action des protéines codées par les gènes de la région *vir*. La région *vir* porte une série de gènes de virulence qui codent pour des protéines intervenant dans les mécanismes de transfert de l'ADN-T dans la cellule végétale (sous forme simple brin), de son intégration dans le génome de la plante hôte et de réparation de l'ADN-T simple brin du plasmide par réplication. L'expression des gènes de virulence est déclenchée par des facteurs de croissance, molécules libérées par les cellules de la plante lors d'une blessure. Cependant, si les connaissances sur l'interaction entre *A. tumefaciens* et la cellule végétale progressent rapidement, les mécanismes moléculaires impliqués dans le transfert de l'ADN-T ne sont pas encore tous bien caractérisés.



Utilisation biotechnologique de *A. tumefaciens*

Ainsi, l'interaction entre *A. tumefaciens* et la plante hôte constitue un mécanisme naturel de transfert d'ADN chez les plantes. Ce mécanisme a été modifié pour devenir un outil versatile de transformation génétique. En effet, seules les bordures de l'ADN-T sont nécessaires à son transfert dans le génome végétal. De ce fait, le plasmide Ti peut être désarmé par délétion des oncogènes présents sur l'ADN-T de la souche sauvage : la nouvelle souche perd alors sa capacité d'induire des tumeurs. Les gènes d'intérêt ainsi que les séquences nécessaires à la réalisation de la transformation génétique sont placés entre les bordures de l'ADN-T. Le plus souvent, cet ADN-T modifié est placé sur un second petit plasmide appelé vecteur binaire, tandis que les fonctions *vir* sont fournies en *trans* par le plasmide Ti désarmé.



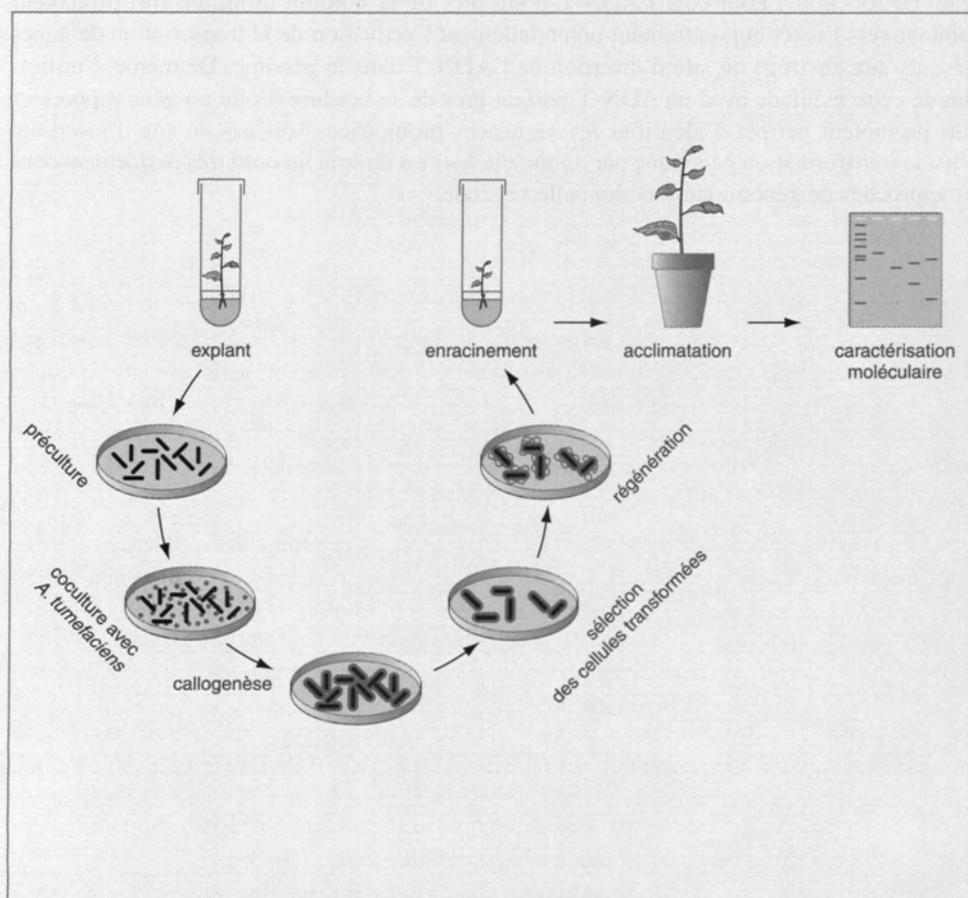
Classiquement, un vecteur binaire porte une origine de réplication fonctionnelle à la fois chez *A. tumefaciens* et chez *E. coli*, des marqueurs de sélection pour les plantes et les bactéries, les séquences bordures de l'ADN-T et un site de clonage multiple. Entre les bordures droite et gauche se trouvent généralement un gène de sélection permettant de sélectionner les cellules végétales transformées génétiquement ainsi que la construction chimérique que l'on désire introduire dans la plante.

Le gène d'intérêt peut être un gène déjà présent dans le génome de la plante et dont on souhaite étudier la fonction en faisant varier son niveau d'expression dans la plante transformée et en évaluant les effets sur son phénotype. Selon le type de la construction chimérique (p. ex. l'orientation de la séquence codante, sa duplication tête-bêche), le gène introduit pourra être exprimé de façon ectopique dans la plante transgénique (à une étape de son développement ou dans un tissu où il n'est pas exprimé chez la plante normale), sur-exprimé ou entraînera au contraire la sous-expression du gène endogène correspondant, cela par effet antisens, par co-suppression ou par interférence d'ARN (RNAi, voir Fiche 46). Ce peut également être une séquence d'intérêt agronomique (conférant p. ex. une tolérance aux insectes). Ce peut être enfin un promoteur dont on souhaite étudier la spécificité d'expression dans le développement de la plante, ce promoteur étant alors fusionné à un gène rapporteur (voir Fiche 38) dans la construction chimérique. Le gène de résistance code classiquement pour une protéine conférant une résistance à un antibiotique. Aujourd'hui, appa-

raissent de nouveaux vecteurs qui utilisent d'autres agents de sélection (résistance aux herbicides, résistance aux métaux lourds...) ou encore des mécanismes permettant l'excision du gène de résistance à l'antibiotique après la phase de sélection (p. ex. le système recombinase). Cette dernière stratégie permet d'envisager la réalisation de plusieurs transformations successives avec un même agent de sélection.

Transformation

Pour transformer des cellules végétales, des explants de la plante (disques foliaires, entre-nœuds de tiges...) sont mis en contact avec la culture diluée d'*A. tumefaciens* en conditions contrôlées. C'est pendant cette étape de co-culture que l'ADN-T modifié est transféré dans la cellule végétale et s'intègre à son génome. L'étape de décontamination consiste à se débarrasser par lavages successifs du maximum des agrobactéries désormais inutiles. Les explants végétaux sont ensuite maintenus sur un milieu approprié contenant des bactériostatiques empêchant la multiplication des agrobactéries et surtout contenant l'antibiotique permettant la sélection des cellules végétales transformées génétiquement. Les cellules transformées sont ensuite multipliées et placées sur des milieux permettant leur régénération en plantes entières, par bourgeonnement adventif ou par embryogenèse somatique.



Les plantes transgéniques sont ensuite caractérisées au niveau moléculaire afin, d'une part, de vérifier l'intégration de l'ADN-T dans le génome et d'estimer le nombre de copies transférées (hybridation de *Southern*) et, d'autre part, d'évaluer le profil d'expression du ou des gènes introduits au niveau des ARNm (analyse de type *northern*), des protéines (analyse de type *western*) ou de leur activité biologique. Le phénotype des plantes transgéniques est comparé à celui des plantes témoins non modifiées. Des croisements contrôlés sont réalisés afin d'obtenir dans la descendance des lignées homozygotes pour le transgène. Des plantes multi-transgéniques sont obtenues lorsque des plantes transformées avec des transgènes différents sont croisées ensemble.

Une technique de transformation génétique à haut débit a été développée chez la plante modèle *Arabidopsis thaliana*. Cette méthode, qui s'affranchit de toutes les étapes de culture *in vitro* présentes dans les protocoles classiques, consiste en un trempage bref de plantes ayant des bourgeons floraux dans la solution d'agrobactéries. La sélection s'effectue ultérieurement au niveau de la germination des graines produites par ces plantes. Cette méthode peut être utilisée classiquement pour introduire un gène dans une plante, mais surtout elle a été utilisée comme méthode de mutagenèse au hasard par étiquetage de gènes avec l'ADN-T : plusieurs dizaines de milliers de lignées transgéniques ont déjà été produites. En utilisant des vecteurs binaires adaptés, cette méthode a également permis d'étiqueter des mutants "gain de fonction". Pour cela l'ADN-T porte près de la bordure droite un fort promoteur pointant vers l'extérieur, entraînant potentiellement l'activation de la transcription de gènes présents aux environs du site d'insertion de l'ADN-T dans le génome. De même, l'utilisation de cette méthode avec un ADN-T portant près de sa bordure droite un gène rapporteur sans promoteur permet d'identifier les séquences promotrices voisines au site d'insertion. Ainsi la transformation génétique par *A. tumefaciens* est devenu un outil très performant dans les approches de génomique fonctionnelle végétale.

TRANSFERT DIRECT DE GÈNES DANS DES PROTOPLASTES VÉGÉTAUX

Le principe du transfert direct de gènes consiste à mélanger des cellules végétales avec de l'ADN en solution, sans que les gènes que l'on désire introduire soient obligatoirement portés par des vecteurs de transformation. On parle parfois alors d' "ADN nu". Cependant, la paroi cellulaire présente chez les végétaux (et les champignons) est un obstacle à l'intégration de l'ADN dans les cellules. Aussi est-il nécessaire de préparer des protoplastes par digestion de la paroi cellulosique qui seront mélangés avec l'ADN en solution. L'application de conditions physico-chimiques appropriées sur ce mélange provoque une perméabilisation temporaire et réversible de la membrane plasmique qui permet à des molécules d'ADN de pénétrer dans les protoplastes. Ces molécules d'ADN atteignent le noyau – par un mécanisme encore inconnu – et s'intègrent par recombinaison à un chromosome de la cellule végétale.

Méthode

Le mélange d'ADN utilisé peut être constitué : (i) d'un plasmide portant un gène dont l'expression confère une résistance à un agent sélectif afin de sélectionner les cellules transformées ; (ii) d'un plasmide portant la construction chimérique d'intérêt. On parle alors de co-transformation. Les protoplastes ainsi traités sont cultivés d'abord sur un milieu non sélectif afin de leur permettre de synthétiser une nouvelle paroi et d'entamer les premières divisions cellulaires, puis sur un milieu contenant l'agent sélectif afin d'éliminer les cellules végétales non transformées. Ces cellules sont alors soumises au processus de régénération afin d'obtenir des plantes transgéniques. L'analyse des transformants se fait d'une façon analogue à celle décrite pour les plantes obtenues par transformation avec *A. tumefaciens*.

Trois techniques principales permettent d'introduire l'ADN dans les protoplastes : (i) l'utilisation de polyéthylène glycol (PEG) ; (ii) l'électroporation ; (iii) la biolistique (voir Fiche 35).

Le polyéthylène glycol est un polymère qui provoque des fusions entre membranes plasmiques. Il permet le rapprochement physique des molécules d'ADN et du plasmalemme, et des déstabilisations locales et réversibles de la membrane plasmique. Il en résulte la pénétration de molécules d'ADN dans le cytoplasme des protoplastes.

L'électroporation consiste à appliquer au mélange ADN-protoplastes un choc électrique qui déstabilise également de façon réversible la structure du plasmalemme, ce qui permet l'intégration de l'ADN dans les cellules. Ce protocole, dans certaines conditions, augmente l'efficacité de la transformation d'un facteur 100 à 1 000.

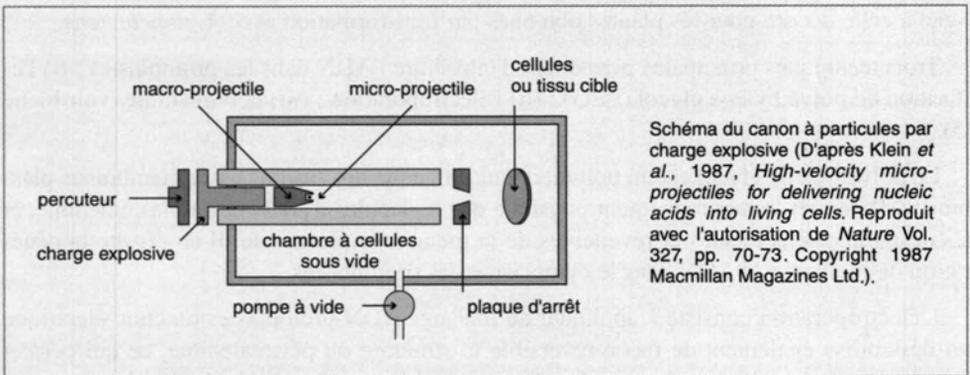
TRANSFERT DIRECT DE GÈNES PAR BIOLISTIQUE

Principe

Cette méthode a été développée dans le but d'introduire des gènes dans des cellules entières afin de s'affranchir des protoplastes et des difficultés de régénération rencontrées pour certaines espèces végétales. Elle est également applicable aux cellules animales. Son principe consiste à délivrer en force de l'ADN sur des tissus par propulsion. La force de propulsion est obtenue soit par explosion d'une poudre dans une balle, soit par décompression d'un gaz (l'hélium le plus souvent).

Méthode

L'ADN nu à introduire est tout d'abord précipité sur des billes de tungstène ou d'or d'environ 1 µm de diamètre. Ces micro-projectiles sont propulsés à l'aide d'un macro-projectile sur les cellules. Après arrêt du macro-projectile pour éviter qu'il ne détruise les cellules, certaines des micro-billes de tungstène vont pénétrer dans les cellules, transportant avec elles l'ADN. Cet ADN, comme dans le cas du transfert direct de gènes dans des protoplastes, doit atteindre le noyau, s'intégrer dans une partie du génome et s'y exprimer. Les explants de la plante, après "bombardement", sont mis à cultiver sur un milieu sélectif afin de régénérer des plantes transgéniques. Cette approche a permis d'obtenir de nombreuses plantes transgéniques pour des espèces jusqu'alors réfractaires, notamment le maïs et le blé. Plusieurs types d'appareils sont disponibles dans le commerce, soit à décharge de poudre, soit à décharge d'hélium. Des miniaturisations ont été réalisées afin de permettre notamment l'application de la biolistique à la transformation de tissus animaux *in vivo* sous anesthésie.



TRANSFORMATION GÉNÉTIQUE DES CELLULES ANIMALES

L'introduction d'une séquence d'ADN dans une cellule animale permet d'analyser *in vivo* le rôle de cette séquence, de déterminer la fonction d'un gène ou d'identifier des séquences régulatrices contrôlant l'expression de ce gène. On choisira alors la lignée cellulaire la plus appropriée à la production d'une protéine donnée ou à l'étude que l'on veut mener, ainsi que le vecteur le plus adapté à la lignée cellulaire choisie. Le transfert de gènes peut permettre son expression transitoire à partir des plasmides non intégrés (voir Fiche 38). Il est également possible de sélectionner les cellules dans lesquelles le gène transféré s'est intégré. Certains vecteurs peuvent se maintenir à l'état d'épisomes autonomes (molécules d'ADN libre dans le cytoplasme). C'est le cas des vecteurs dérivés du virus SV40 dans les cellules COS et de ceux dérivés du virus d'Epstein Barr dans les cellules de primates.

Tous les vecteurs utilisés doivent posséder les éléments essentiels suivants :

- une origine de réplication bactérienne et un marqueur de résistance à un antibiotique qui permettent l'amplification du plasmide dans des bactéries ;
- les séquences eucaryotes qui contrôlent l'initiation de la transcription (promoteurs) et la maturation des transcrits (signaux d'épissage, de polyadénylation...) dérivées de gènes cellulaires ou viraux ;
- un marqueur de sélection permettant la survie des cellules eucaryotes recombinantes si elles sont cultivées dans un milieu sélectif ;
- le gène à exprimer sous forme d'un ADN complémentaire ou bien d'un ADN génomique.

Deux approches expérimentales ont été développées pour introduire de l'ADN dans des cellules eucaryotes : la transfection d'un vecteur d'expression d'une part sous la forme d'ADN nu et d'autre part par un processus d'infection virale. Dans ce dernier cas, la technique d'infection dépend du type de virus et de cellules hôtes. Les principaux vecteurs viraux dérivent des adénovirus et des rétrovirus, mais aussi d'autres virus comme le *Semliki Forest Virus*.

Les techniques employées

Nous présentons ici plusieurs méthodes de transfection de cellules animales par de l'ADN nu. L'efficacité de pénétration et d'expression de l'ADN étranger peut varier de 1 à 100 (voire plus) selon le type des cellules utilisées, la nature, la quantité et l'état de l'ADN recombinant et différents paramètres physico-chimiques d'introduction de l'ADN. Le choix de l'une ou l'autre de ces techniques dépend notamment de la nature des cellules employées (p. ex. la méthode utilisant le phosphate de calcium est inefficace sur des lymphocytes) et du

type d'expression souhaitée (un taux élevé d'expression stable est obtenu par électroporation). Très souvent, l'ADN à introduire est associé à des molécules chargées positivement (calcium, polymères cationiques) de manière à favoriser son rapprochement avec la membrane plasmique chargée négativement à sa surface.

Utilisation du phosphate de calcium

Un mélange d'ADN, de phosphate et d'un sel de calcium conduit à la coprécipitation de phosphate de calcium et d'ADN qui se complexe alors au calcium. Ce précipité entre spontanément dans les cellules mais cette étape est grandement améliorée par un choc osmotique appliqué aux cellules à l'aide de glycérol ou de diméthylsulfoxyde (DMSO). Cette méthode est beaucoup moins utilisée depuis l'utilisation des polymères polycationiques.

Utilisation des polymères cationiques

Des molécules hybrides comportant une région lipidique et une région cationique cumulent les avantages des deux types de groupement. De nombreuses formules sont commercialisées. L'utilisation des lipides cationiques est aisée. Leur efficacité étant variable selon les lignes cellulaires, le choix doit donc être fait le plus souvent après avoir procédé à des tests ou en fonction des résultats publiés. Certains lipides cationiques sont coûteux et d'une stabilité limitée.

- Le DEAE dextrane est un polycation qui facilite l'absorption de l'ADN sur la membrane plasmique et favorise son entrée à l'intérieur des cellules. De plus il protège l'ADN contre l'action des nucléases et augmente ainsi son taux d'expression dans la cellule. Il est aussi utilisé pour augmenter l'efficacité des infections virales. Les cellules contiennent alors un grand nombre de vecteurs dans leur noyau, mais peu de copies sont intégrées dans le génome de la cellule. Cependant le DEAE dextrane peut être toxique pour les cellules. L'optimisation des conditions de transfection est fonction de la sensibilité des cellules hôtes à ce polymère. Néanmoins, l'efficacité de transfection est élevée puisqu'on peut obtenir jusqu'à 25 % de cellules transfectées.
- La polyéthylèneimine (PEI) est un polycation qui s'associe aisément à l'ADN : le complexe pénètre alors spontanément dans les cellules. L'efficacité du PEI dépend en grande partie de sa taille. Il est stable, peu coûteux, fiable et efficace dans un grand nombre de types cellulaires. Son utilisation est aisée. L'Exgen 500 est une forme optimisée du PEI. D'autres polycations sont également utilisables mais non décrits ici.

Utilisation des liposomes

La technique consiste à piéger l'ADN étranger dans des vésicules membranaires artificielles. Les vésicules pénètrent dans la cellule par endocytose. Le pouvoir infectieux de virus à ADN comme SV 40 peut être ainsi augmenté de près de 200 fois. Néanmoins, cette technique n'est pas plus efficace que celles utilisant d'autres agents transfectants et elle est difficile à mettre en oeuvre.

La fusion des protoplastes

Le principe consiste à fusionner des cellules animales avec des bactéries contenant le gène à introduire. Au préalable, des protoplastes sont préparés à partir de bactéries contenant le plasmide recombinant. Les protoplastes et les cellules animales sont mis en contact. La

fusion des membranes cellulaires, favorisée par l'addition de polyéthylène glycol, permet l'incorporation du plasmide dans la cellule. Cette méthode est très peu utilisée.

Électroporation

La transfection par électroporation consiste en un transfert direct de l'ADN dans le cytoplasme de la cellule hôte. Des pores transitoires sont créés dans la membrane des cellules hôtes lorsqu'elles sont soumises à une brève impulsion électrique, permettant ainsi le passage de l'ADN. Cette technique présente l'avantage de pouvoir être utilisée sur tous les types de cellules. Elle nécessite cependant un appareillage particulier, mais son efficacité est nettement supérieure puisque, dans certains cas, près de 90 % des cellules intègrent l'ADN recombinant dans leur cytoplasme. L'électroporation est requise pour certains types cellulaires comme les cellules ES (*Embryonic Stem*). Elle est également très souhaitable pour les gros fragments d'ADN (BAC, YAC...).

Biolistique

La miniaturisation du canon à particules sous forme d'un pistolet permet de transformer des cellules ou des tissus d'animaux vivants (voir Fiche 35).

Microinjection

La microinjection de gènes dans le noyau des cellules est la méthode la plus efficace. Elle peut conduire à l'obtention de 20 % de cellules dans lesquelles le gène étranger est intégré. Elle nécessite un appareillage particulier et n'est justifiée que pour obtenir des clones stables de cellules pour lesquelles la transfection est très difficile (p. ex. certaines cellules primaires).

Transfection et apoptose

Le transfert d'ADN induit parfois une apoptose des cellules, ce qui réduit considérablement l'efficacité de la transfection. L'addition d'agents anti-apoptose peut améliorer cette situation.

LE CLONAGE DES ANIMAUX

Le clonage est par définition la reproduction d'organismes génétiquement identiques. Ceci implique en pratique un mode de reproduction non sexuée. Le clonage est le mode naturel de reproduction chez des organismes unicellulaires (bactéries, levures). Chez les organismes pluricellulaires, le clonage implique un phénomène de dé-différenciation. Les cellules d'un tissu sont le plus souvent spécialisées dans une fonction particulière. Pour ce faire, elles expriment un lot de gènes spécifiques. Ces cellules sont dites différenciées. En se spécialisant, ces cellules ont perdu leur capacité à exprimer d'autres fonctions et elles sont figées dans leur spécialisation. On dit qu'elles sont incapables de se dé-différencier. La dé-différenciation est donc la capacité pour une cellule appartenant à un tissu donné de donner naissance par division cellulaire à des cellules filles capables de produire toutes les fonctions cellulaires. Chez les végétaux, le clonage peut être obtenu par marcottage, bouturage ou dé-différenciation *in vitro* des cellules somatiques. Les cellules ainsi dé-différenciées ont les caractéristiques d'un embryon et chacune peut à son tour donner une plante normale. Rien de tel n'a jamais été observé chez les animaux.

Le clonage par clivage d'embryon

Les cellules de l'embryon très précoce sont considérées comme totipotentes : chacune d'elles peut redonner un embryon normal si elle est replacée dans une zone pellucide vide. Cette opération peut donc engendrer des clones.

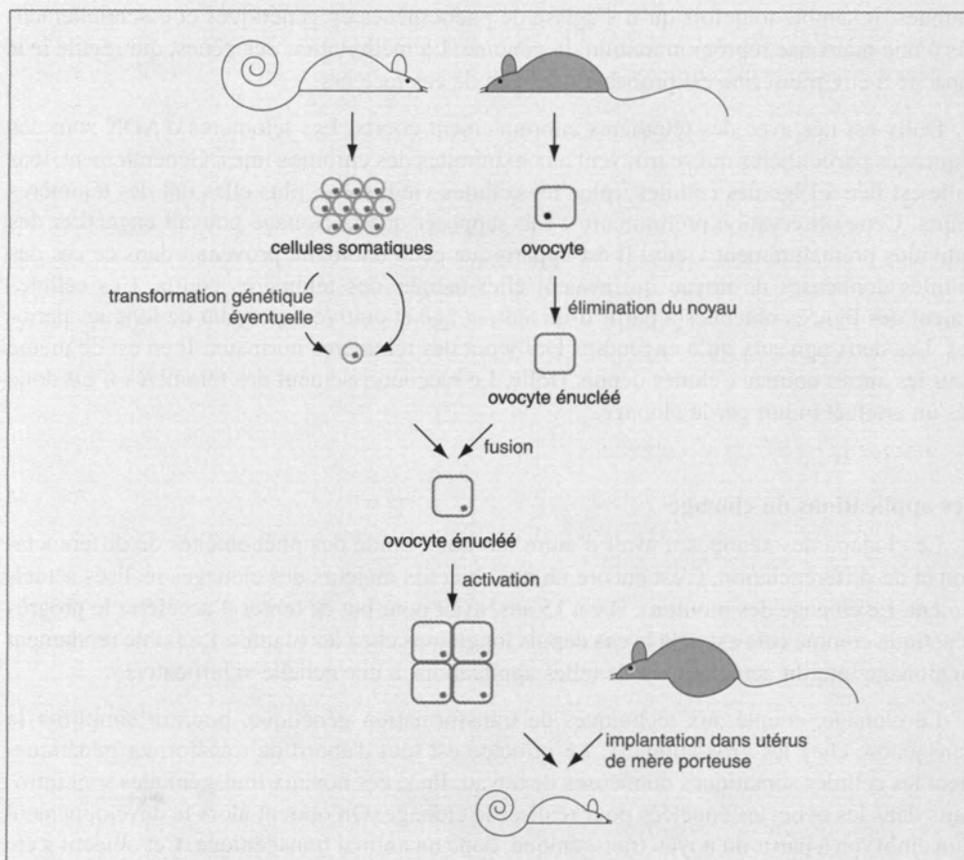
Les embryons de mammifère jusqu'au stade blastocyste peuvent être mécaniquement clivés. Les deux moitiés d'embryon peuvent dans certaines conditions se développer et donner naissance à des jumeaux vrais. Dans les conditions naturelles, chez les mammifères, la naissance des vrais jumeaux résulte d'un développement précoce anormal qui donne l'équivalent d'un clivage artificiel d'embryon. Les vrais jumeaux sont donc des clones.

Le clonage par transfert de noyau

Après la fécondation, le noyau du spermatozoïde, dans lequel aucun ou très peu de gènes sont actifs, se décondense pour donner un pro-noyau morphologiquement peu différent de celui de l'ovocyte. Quelques heures ou jours après, les gènes provenant du spermatozoïde sont activés pour permettre le développement de l'embryon. Le cytoplasme de l'ovocyte doit donc posséder les éléments qui permettent la programmation du génome du spermatozoïde. Le principe du clonage est d'utiliser cette propriété du cytoplasme de l'ovocyte pour ré-activer les gènes de la différenciation à partir de n'importe quel type de noyaux, notamment ceux des cellules somatiques rendus inactifs par la différenciation. Au préalable, il est nécessaire d'éliminer le noyau de l'ovocyte.

Le transfert mécanique de noyau dans le cytoplasme d'ovocytes énucléés a ainsi pu donner naissance à des xénopes puis des moutons, des vaches, des chèvres et des porcs clonés à partir de cellules peu différenciées provenant d'embryons précoces puis de cellules somatiques totalement différenciées. Le premier xénope cloné a été obtenu il y a 40 ans, le premier mouton il y a 15 ans à partir de noyaux de cellules embryonnaires non cultivées. La brebis Dolly est le premier animal né à partir d'une cellule d'adulte différenciée.

En pratique, le noyau est apporté sous forme de cellule entière elle-même placée entre la zone pellucide de l'ovocyte énucléé et sa membrane plasmique. Un choc électrique contrôlé fait fusionner les membranes et active le nouvel embryon ainsi formé en induisant entre autres des flux de calcium. Chez la souris, ce procédé s'est avéré très peu efficace. Le transfert de noyau isolé dans le cytoplasme de l'ovocyte énucléé est nécessaire.



Le rendement du clonage

Le rendement du clonage depuis ses débuts est faible. Il est de quelques pour cent d'animaux clonés nés par rapport au nombre de noyaux transférés. Ce rendement est d'autant plus faible que les cellules donneuses de noyau sont plus différenciées et ont été maintenues plus longtemps en culture. Certains types cellulaires se prêtent au clonage (fibroblastes foetaux, cellules du cumulus) et d'autres non ou très peu (cellules nerveuses).

La méthode utilisée pour engendrer Dolly a reposé sur la mise au repos des cellules donneuses de noyau en retirant tous les facteurs de croissance de leur milieu de culture. Il a alors été postulé que le stade G0 du cycle cellulaire (phase du cycle de division des cellules pendant lesquelles ces dernières sont au repos) était essentiel pour obtenir un clonage à partir d'une cellule différenciée. Cette hypothèse ne s'est pas vérifiée dans la mesure où plusieurs groupes ont indépendamment obtenu des animaux clonés avec un rendement identique sans que les cellules donneuses aient au préalable été amenées au stade G0. Les échecs se traduisent par un non-développement des embryons reconstitués jusqu'au stade blastocyste. Ces blastocystes réimplantés se développent de manière variable conduisant souvent à des gestations interrompues très tardivement. Un foetus ou nouveau-né sur deux meurt dans la période périnatale avec des symptômes divers mais reproductibles. L'un d'eux est l'excès de poids. Les raisons pour lesquelles le rendement de clonage est aussi faible ne sont pas bien connues. Il semble toutefois qu'il s'agisse de phénomènes épigénétiques et essentiellement dûs à une mauvaise reprogrammation du génome. La méthylation des gènes, qui régule leur capacité à être transcrits, est probablement l'un de ces facteurs.

Dolly est née avec des télomères anormalement courts. Les télomères d'ADN sont des séquences particulières qui se trouvent aux extrémités des chromosomes. Généralement, leur taille est liée à l'âge des cellules : plus les cellules vieillissent, plus elles ont des télomères courts. Cette observation préliminaire a fait supposer que le clonage pouvait engendrer des individus prématurément vieux. Il est apparu que cette anomalie provenait dans ce cas des cellules donneuses de noyau qui avaient elles-mêmes des télomères courts. Ces cellules étaient des lignées obtenues à partir d'un animal âgé et cultivées pendant de longues périodes. Les deux agneaux qu'a engendrés Dolly ont des télomères normaux. Il en est de même pour les autres animaux clonés depuis Dolly. Le raccourcissement des télomères n'est donc pas un artefact induit par le clonage.

Les applications du clonage

Le clonage des xénopes n'avait d'autre but que l'étude des phénomènes de différenciation et dé-différenciation. C'est encore un des objectifs majeurs des clonages réalisés actuellement. Le clonage des moutons, il y a 15 ans, avait pour but de tenter d'accélérer le progrès génétique comme cela est déjà le cas depuis longtemps chez les plantes. Le faible rendement du clonage interdit actuellement de telles applications à une échelle significative.

Le clonage, couplé aux techniques de transformation génétique, pourrait simplifier la transgénèse chez les gros animaux. Le principe est tout d'abord de transformer génétiquement les cellules somatiques donneuses de noyau. Puis, ces noyaux transgéniques sont introduits dans les ovocytes énucléés pour réaliser le clonage. On obtient alors le développement d'un embryon à partir du noyau transgénique, donc un animal transgénique. Cet objectif a été atteint chez le Mouton, la Chèvre, la Vache, le Porc et la Souris. Des gènes étrangers ont été ajoutés dans des cellules foetales en culture par transfection. Les cellules porteuses des gènes étrangers ont été utilisées comme source de noyaux pour engendrer des animaux clonés transgéniques. Chez le Mouton et la Souris, cette même approche a permis de procéder à des remplacements de gènes par recombinaison homologe. Le rendement du clonage, qui se trouve diminué par la longue période de culture nécessaire à la sélection des clones cellulaires dans lesquels a eu lieu le remplacement de gène, rend cette opération délicate.

L'addition et le remplacement de gènes sont actuellement exploités pour produire des protéines recombinantes d'intérêt pharmaceutique dans le lait d'animaux transgéniques, pour adapter les organes de porc dans le but de les greffer à des patients et plus généralement pour étudier des maladies humaines. Les applications du clonage pour l'amélioration des élevages sont en principe possibles. L'extension de ces techniques est attendue chez le Rat, le Poulet, le Lapin, certains poissons et d'autres espèces encore. Le clonage reproductif chez l'Homme (obtenir un individu clonal identique à l'original) est actuellement impensable pour des raisons strictement techniques indépendamment des problèmes éthiques. Le clonage thérapeutique (appliqué à des cellules humaines, voire des embryons) qui été envisagé pourrait consister à reconstituer un embryon à partir des cellules somatiques d'un patient. Cet embryon pourrait être la source de cellules souches totipotentes puis des cellules souches d'organes obtenues par différenciation des cellules totipotentes *in vitro*. Une telle opération conduirait donc en principe à une régénération d'organe par autogreffe. Cette approche est actuellement théorique et elle soulève des problèmes éthiques qui la rend difficilement acceptable. Ceci est d'autant plus vrai que les cellules souches d'un organe peuvent expérimentalement se transformer en cellules souches d'un autre organe et permettre ainsi sa régénération.

EXPRESSION TRANSITOIRE

Par les techniques de transformation génétique classiques (voir Fiches 33 à 37), les cellules transformées sont sélectionnées sur un milieu sélectif inhibant la croissance des cellules non transformées. Les cellules, tissus ou plantes ainsi obtenus sont ensuite analysés sur l'expression du ou des gènes introduits. L'obtention d'organismes transgéniques est sans aucun doute un outil fondamental pour l'étude de l'expression des gènes, mais les procédures d'obtention de ces organismes sont souvent longues, voire difficiles pour certaines espèces. L'expression transitoire permet l'analyse rapide de l'expression d'un gène sans sélection des cellules transformées. Ces études peuvent être par la suite complétées par l'analyse de ces promoteurs dans des organismes transgéniques (voir Fiche 30).

Principe

Les cellules transformées sont cultivées sans pression de sélection, récoltées de 24 à 72 heures après transfection (période durant laquelle les gènes portés par le plasmide sont transcrits et traduits avant leur intégration dans les chromosomes), puis broyées afin d'étudier l'expression du gène introduit. En général, des gènes dits rapporteurs, dont l'expression peut être facilement mise en évidence, sont utilisés dans ces expériences.

Définition d'un gène rapporteur et exemples

Le produit du gène rapporteur (ou *reporter*) doit être absent des cellules soumises à la transformation. C'est pour cette raison qu'en ce qui concerne les cellules eucaryotes, les gènes rapporteurs sont la plupart du temps d'origine bactérienne. La protéine codée par le gène rapporteur doit être facilement détectable ; c'est souvent une enzyme dont l'expression est visualisable par simple coloration ou mesurable par un test d'activité enzymatique. Les gènes rapporteurs les plus utilisés sont :

- le gène de la chloramphénicol acétyl transférase (CAT). L'activité de la CAT, catalysant le transfert d'un groupement acétyle sur le chloramphénicol, est simple à mesurer en présence de chloramphénicol ou d'acétyl-coenzyme A marqué au ^{14}C ;
- le gène *lacZ* de la β -galactosidase et le gène *gus* de la β -glucuronidase. L'expression de la β -galactosidase ou de la β -glucuronidase est détectable par coloration bleue des cellules en présence respectivement de X-Gal ou de X-Gluc¹ ;
- le gène de la luciférase. En présence de luciférine et d'ATP, la luciférase exprimée dans les cellules est capable d'induire l'émission de lumière facilement repérable et mesurable ;

- le gène de la protéine fluorescente verte (GFP). La protéine GFP exprimée émet une lumière verte (à 509 nm) visualisable par fluorescence sous UV. La fluorescence est intrinsèque à la protéine et sa détection ne nécessite aucune infiltration des tissus par un substrat puisqu'il ne s'agit pas d'une enzyme. Ce gène rapporteur peut donc être utilisé pour des expériences *in vivo*.

Le choix du gène rapporteur dépend du type de tissu étudié et de l'usage. Par exemple, les gènes *gus* et *gfp* sont appropriés pour la localisation *in situ* des activités des promoteurs étudiés. Le gène *cat* est plus adapté aux mesures quantitatives d'expression à partir d'extraits cellulaires bruts.

1. La β -galactosidase coupe la liaison glycoside de son substrat, l'X-Gal. Le groupement chromophorique X est donc libéré, ce qui induit une coloration bleue. Il en est de même avec la β -glucuronidase, qui catalyse la coupure de la liaison glucuronide de son substrat, l'X-Gluc.

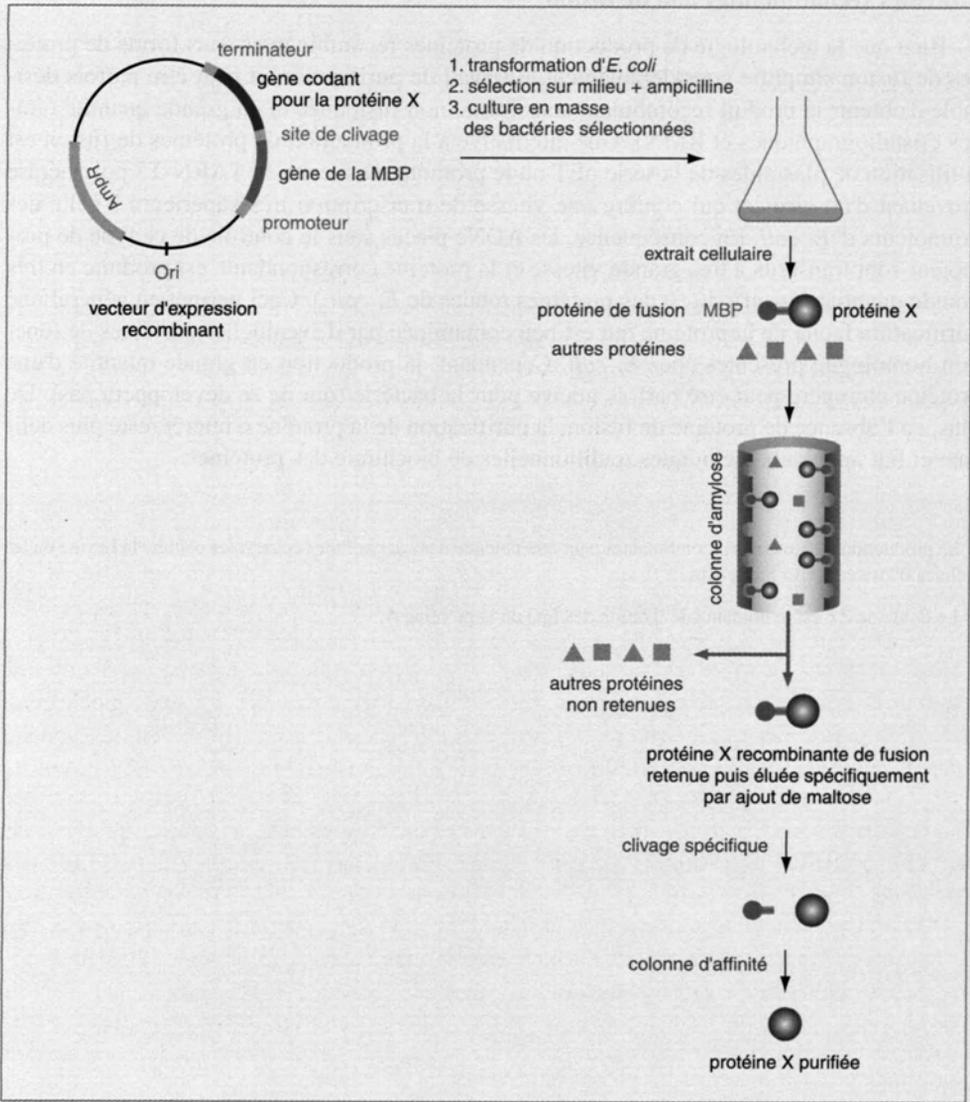
PROTÉINES RECOMBINANTES

L'étude des protéines passe par leur production en grande quantité et leur purification. La stratégie développée ici consiste à faire produire massivement par des bactéries (généralement *E. coli*¹) la protéine d'intérêt sous forme d'une protéine de fusion : celle-ci est constituée d'un fragment de protéine bactérienne auquel est associée la protéine eucaryote désirée (protéine X) par l'intermédiaire d'une courte séquence. La protéine de fusion, souvent plus stable que la protéine X seule, peut être produite en grande quantité.

Principe

Le gène codant pour la protéine d'intérêt X est cloné dans un vecteur d'expression. Ce vecteur est un plasmide possédant un promoteur bactérien (le plus souvent inductible) suivi d'une partie de la séquence codant pour une protéine bactérienne, d'une séquence correspondant à un site de clivage par une protéase et d'un site de polyclonage au niveau duquel le gène d'intérêt sera introduit, en respectant la phase de lecture pour la traduction en acides aminés. Après transformation, les bactéries ayant intégré un vecteur recombinant sont sélectionnées sur milieu sélectif, cultivées en grande quantité, puis récoltées et lysées de façon à obtenir un extrait cellulaire brut contenant toutes les protéines synthétisées dont la protéine de fusion. A partir de cet extrait brut, la purification de la protéine de fusion est réalisée par chromatographie d'affinité sur colonne : des groupements, appelés ligands, fixés sur la colonne et interagissant spécifiquement avec la protéine fusion, retiennent cette dernière, alors que les autres protéines sont éliminées. La protéine de fusion est récupérée après élution spécifique et l'on induit la coupure entre la protéine bactérienne et la protéine X par incubation en présence de la protéase adéquate. Une deuxième chromatographie d'affinité peut être réalisée pour éliminer la protéine bactérienne : seule celle-ci est retenue sur la colonne alors que la protéine X est éluee.

Différents systèmes sont actuellement disponibles. Le choix de l'un d'eux doit tenir compte de différents paramètres tels que la stabilité de la protéine de fusion synthétisée, sa solubilité, les conditions d'élution... Quelques exemples (dont la liste n'est pas exhaustive) sont indiqués dans le tableau ci-dessous. Régulièrement de nouveaux ligands sont proposés. Par exemple, un type de fusion de plus en plus utilisé consiste à greffer l'ADNc codant pour une protéine thermostable (la thiorédoxine) à la séquence d'ADN que l'on veut traduire en acides aminés : l'addition de la thiorédoxine peut dans la plupart des cas augmenter la solubilité de la protéine désirée ou/et sa stabilité thermique.



Peptide ou protéine
bactérienne de fusion

Ligand

Agent permettant
l'élution spécifique

Maltose Binding Protein (MBP)

amylose

maltose

glutathion S-transférase

glutathion

agent réducteur

6 histidines (His-Tag)

Ni⁺⁺ ou Zn⁺⁺

imidazole, gradient de pH

peptide d'affinité pour la calmoduline (CBP)

calmoduline

EGTA

domaine ZZ²

IgG

faible pH

intéine

chitine

agent réducteur

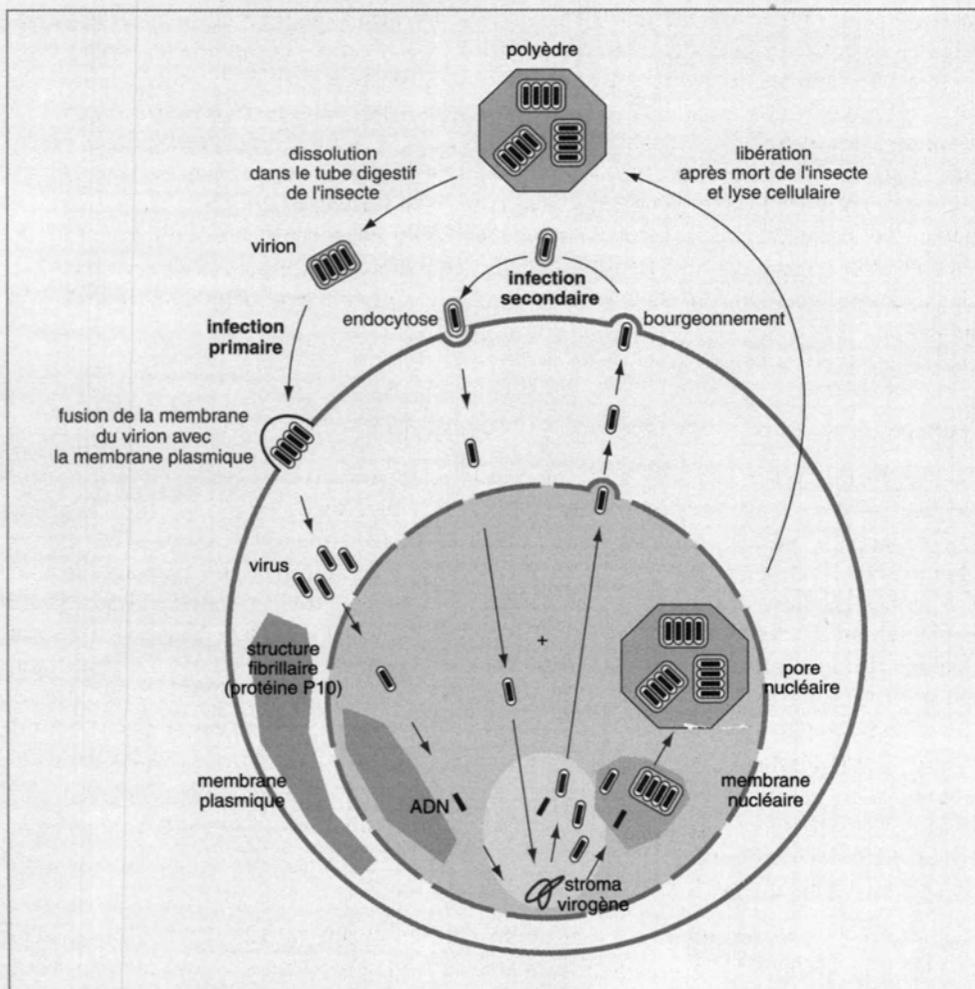
Protéines recombinantes non de fusion

Bien que la technologie de production de protéines recombinantes sous forme de protéines de fusion simplifie considérablement le travail de purification, il peut être parfois désirable d'obtenir le produit recombinant sous forme non fusionnée et en grande quantité (études cristallographiques et RMN). Une alternative à la production de protéines de fusion est l'utilisation de plasmides de la série pET où le promoteur est celui de l'ARN-T7 polymérase provenant d'un virus et qui confère une vitesse de transcription très supérieure à celle des promoteurs d' *E. coli*. En conséquence, les ADNc placés sous le contrôle de ce type de promoteur sont transcrits à très grande vitesse et la protéine correspondante est produite en très grande quantité (jusqu'à 50 % des protéines totales de *E. coli*). Ceci permet en général une purification facile de la protéine qui est peu contaminée par d'éventuelles protéines de fonction homologues présentes chez *E. coli*. Cependant, la production en grande quantité d'une protéine étrangère peut être parfois nocive pour la bactérie (qui ne se développera pas). De plus, en l'absence de protéine de fusion, la purification de la protéine d'intérêt reste plus délicate et fait appel aux techniques traditionnelles de biochimie des protéines.

1. La production de protéines recombinantes peut être obtenue dans des cellules eucaryotes comme la Levure ou les cellules d'insectes (voir Fiche 40).
2. Le domaine ZZ est le domaine de fixation des IgG de la protéine A.

LES BACULOVIRUS D'INSECTES, VECTEURS D'EXPRESSION DE GÈNES ÉTRANGERS

Les Baculovirus (du latin *baculum*, petit bâtonnet) constituent un groupe de virus très vaste. Ils sont présents uniquement chez les Arthropodes et ont été décrits chez plus de 600 espèces d'insectes ainsi que chez quelques crustacés. Ce sont des virus enveloppés en forme de bâtonnets, d'une longueur d'environ 350 nm et d'un diamètre moyen de 50 nm. Ils contiennent un génome constitué d'une molécule d'ADN bicaténaire, circulaire, superenroulée, dont la taille varie entre 88 et 153 kpb.



Un des baculovirus les plus étudiés aujourd'hui est celui qui a été à l'origine isolé des larves du papillon *Autographa californica*. Ce virus connu sous le sigle AcMNPV (pour *Autographa californica Multiple Nuclear Polyedrosis Virus*) se multiplie dans les noyaux des cellules de plus de 30 espèces différentes de lépidoptères. Dans ce cas, plusieurs particules virales sont incluses dans une même enveloppe (dans d'autres cas, il n'y a qu'une seule particule par enveloppe).

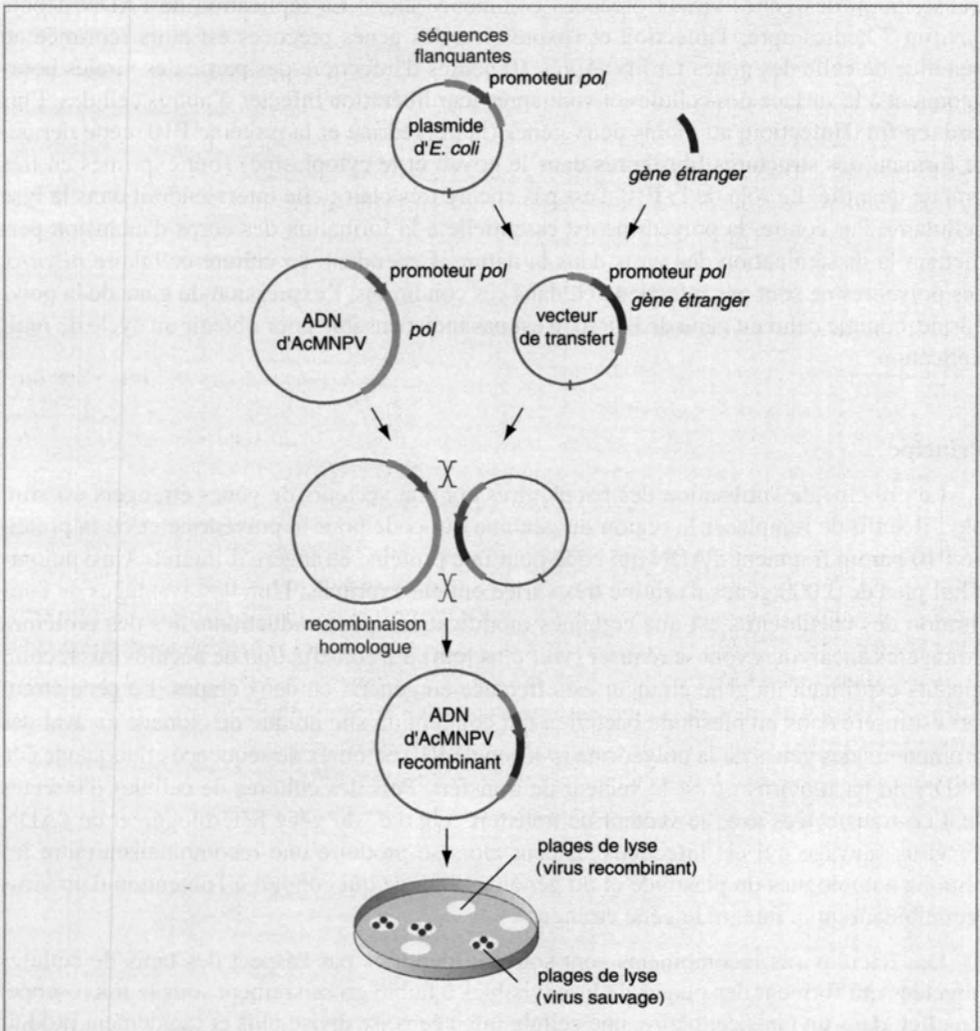
Dans la nature, les virus sont présents à l'intérieur de polyèdres, structures cristallines facilement observables en microscopie photonique, qui forment de véritables "corps d'inclusion". La matrice protéique de ces corps d'inclusion est constituée par la polyédrine qui est une protéine codée par le génome viral. Le virus, protégé des facteurs de l'environnement par ce corps d'inclusion, est ainsi disséminé à la mort de l'insecte. Après ingestion par un autre insecte, les polyèdres sont dissous dans le suc intestinal, libérant les particules virales qui peuvent pénétrer dans les cellules de l'épithélium intestinal par fusion membranaire ou par endocytose. Après passage de l'épithélium intestinal, les virus infectent les tissus de l'hôte. Dans les cellules, la nucléocapside est amenée jusqu'au noyau où elle libère son ADN. La transcription des gènes viraux précoces commence alors. La réplication de l'ADN débute environ 7 heures après l'infection et l'expression des gènes précoces est alors réprimée au bénéfice de celle des gènes tardifs. Après 10 heures d'infection, des particules virales bourgeonnent à la surface des cellules et vont après leur libération infecter d'autres cellules. Plus tard, en fin d'infection, au moins deux gènes (la polyédrine et la protéine P10, cette dernière formant des structures fibrillaires dans le noyau et le cytoplasme) sont exprimés en très grande quantité. Le rôle de la P10 n'est pas encore très clair ; elle interviendrait dans la lyse cellulaire. Par contre, la polyédrine est essentielle à la formation des corps d'inclusion permettant la dissémination des virus dans la nature. Cependant, en culture cellulaire *in vitro*, ces polyèdres ne sont pas infectieux et, dans ces conditions, l'expression du gène de la polyédrine, comme celui du gène de la P10, n'est pas indispensable pour obtenir un cycle de multiplication.

Principe

Le principe de l'utilisation des baculovirus comme vecteurs de gènes étrangers est simple : il suffit de remplacer la région du génome qui code pour la polyédrine et/ou la protéine P10 par un fragment d'ADN qui code pour une protéine étrangère d'intérêt. Ainsi aujourd'hui plus de 3 000 gènes d'origine très variée ont été exprimés. L'un des avantages de l'utilisation des baculovirus est que certaines modifications post-traductionnelles des protéines étrangères eucaryotes vont se réaliser (voir plus loin). La construction de baculovirus recombinants exprimant un gène étranger est effectuée en général en deux étapes. Le gène étranger est inséré dans un plasmide bactérien qui contient un site unique de clonage en aval des promoteurs des gènes de la polyédrine (*pol*) ou de P10 entourés de séquences flanquantes de l'ADN du baculovirus : c'est le vecteur de transfert. Puis des cultures de cellules d'insectes sont co-transfectées avec le vecteur de transfert "chargé" du gène hétérologue et de l'ADN du virus sauvage qui est infectieux. Il peut alors se produire une recombinaison entre les régions homologues du plasmide et du génome viral, ce qui conduit à l'obtention d'un virus recombinant qui a intégré le gène étranger.

Les Baculovirus recombinants sont souvent identifiés par l'aspect des tapis de cellules infectées qui forment des plages de lyse, visibles à faible grossissement sous le microscope. En effet, dans un tapis cellulaire, une cellule infectée ne se divise plus et rapidement produit

des particules virales bourgeonnantes. Celles-ci vont infecter les cellules voisines. Pour éviter la dispersion de ces virions une couche d'agarose est placée sur le tapis cellulaire. Au bout de 4 à 5 jours, de nombreuses cellules qui entourent la cellule initiale sont également infectées et meurent. Ainsi en utilisant un colorant vital comme le rouge neutre, il est très facile de distinguer ces plages de lyse qui forment des "trous" non colorés dans le tapis rouge des cellules vivantes. Toutes les plages examinées au microscope vont présenter soit un phénotype "présence de polyèdre" correspondant à une infection par un virus sauvage qui n'a pas recombiné, soit un phénotype "absence de polyèdre" correspondant à un virus recombinant qui possède le gène étranger à la place de celui de la polyédrine. Ces plages peuvent alors aisément être prélevées, puis sous-clonées une ou deux fois pour s'assurer de la pureté du recombinant. Plusieurs recombinants indépendants sont prélevés, purifiés et testés pour l'expression de la protéine recombinante. On pourra alors produire et étudier la protéine étrangère. Il est évident que l'on peut utiliser soit le promoteur de la polyédrine, soit celui de la P10, voire les deux, ce qui permet alors l'expression de protéines constituées de plusieurs sous-unités comme des immunoglobulines.



Modifications post-traductionnelles des protéines produites dans les cellules d'insectes

Les protéines exprimées dans les cellules d'insectes présentent, dans la grande majorité des cas, des activités biologiques semblables à celles des protéines natives : activité enzymatique, antigénicité, immunogénicité...

Les cellules d'insectes possèdent les équipements enzymatiques leur permettant d'effectuer les modifications post-traductionnelles semblables à celles des Vertébrés. Elles reconnaissent les peptides signaux homologues et hétérologues et les protéines matures sont excrétées dans le milieu de culture. De la même façon, les signaux d'adressage aux différents compartiments cellulaires sont reconnus : adressage nucléaire, cytoplasmique ou membranaire. Les polyprotéines sont clivées la plupart du temps de façon similaire à celle de la protéine native. Les protéines produites sont phosphorylées, myristylées, palmitylées, farnesylées, carboxy-méthylées, amidées ou acétylées correctement. L'assemblage en dimères, en complexe oligomérique et la formation de ponts disulfures se font de manière correcte.

Un des points importants concerne la glycosylation des protéines. Une analyse détaillée des glycoprotéines exprimées dans des cellules d'insectes a montré des différences dans la taille et la structure des oligosaccharides par rapport à celles des glycoprotéines de mammifères ou de plantes. Les glycoprotéines de mammifères exprimées dans les cellules de *Spodoptera frugiperda* (Sf9) sont souvent plus petites et les chaînes oligosaccharidiques sont plus courtes. Dans de très nombreux cas, cette différence n'affecte pas, ou peu, l'activité biologique des protéines *in vitro*. Ce problème de la glycosylation semble encore aujourd'hui l'une des seules modifications post-traductionnelles que les cellules de *S. frugiperda* ne font pas de façon identique aux cellules de mammifères.

La culture de cellules d'insectes

Un grand nombre de lignées de cellules d'insectes est aujourd'hui disponible pour étudier la multiplication du baculovirus d'*A. californica* : les plus utilisées sont les cellules Sf21 et le clone Sf9, qui proviennent du tissu ovarien de la nymphe du papillon. La lignée TN-368, dérivée de *Trichoplusia ni*, ainsi que la lignée Mb, dérivée de *Mamestra brassicae*, permettent également une très bonne multiplication du baculovirus.

La plupart des cellules de lépidoptères ont, pour leur croissance, un optimum de température compris entre 25 et 30 °C. Contrairement aux cellules de mammifères, elles ne nécessitent pas l'emploi d'une étuve à CO₂ et l'ajout de rouge de phénol ; il suffit d'utiliser un tampon efficace à pH 6,2. Le temps de doublement des populations de cellules est de 16 à 24 heures et dans des conditions normales la densité cellulaire varie de 2.10⁶ à 5.10⁶ cellules/mL. Cette densité peut être augmentée dans des bioréacteurs. Enfin, la plupart des lignées de cellules d'insectes peuvent croître en tapis cellulaire ou en suspension. La culture des cellules d'insectes dans des bioréacteurs de grand volume commence à être de mieux en mieux maîtrisée et des bioréacteurs de 5 à 100 litres sont fréquemment rencontrés dans des installations pilotes.

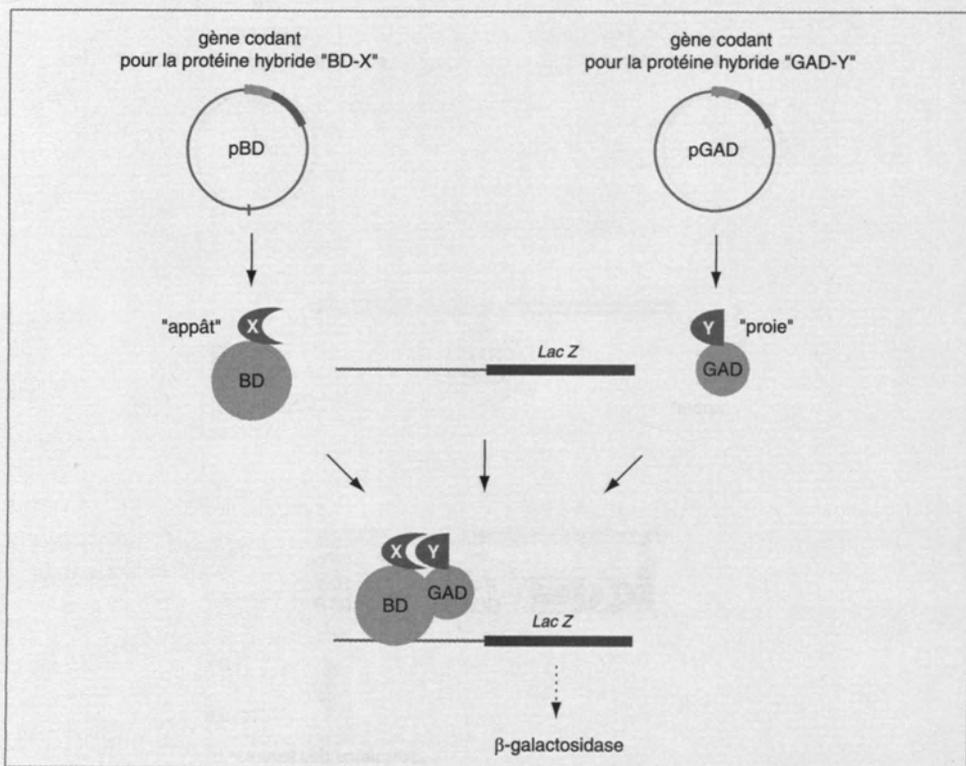
SYSTÈME DU DOUBLE HYBRIDE

Le système du double hybride a été développé pour mettre en évidence des interactions protéine-protéine. Ce test repose sur le fait que beaucoup de protéines, dont les facteurs de transcription, sont composées de plusieurs domaines fonctionnels indépendants. Lorsque ces domaines sont physiquement éloignés, ils n'ont pas d'activité biologique. Mais si ces domaines sont amenés à proximité les uns des autres par des interactions non covalentes, une protéine fonctionnelle peut alors être reconstituée. La protéine Gal4, activatrice de la transcription des gènes du métabolisme du galactose chez la levure *Saccharomyces cerevisiae*, est composée de deux domaines fonctionnels distincts et séparables : un domaine de liaison à l'ADN (domaine appelé DBD pour *DNA Binding Domain*) et un domaine activateur de la transcription (domaine appelé AD pour *Activation Domain*).

Principe

Soit deux protéines X et Y pour lesquelles on veut tester une éventuelle interaction. Chacun des deux domaines de la protéine Gal4 peut être fusionné séparément à chacune des protéines d'intérêt. Ainsi, la protéine X sera liée au domaine DBD et la protéine Y au domaine AD. Si les protéines X et Y interagissent, les domaines DBD et AD de la protéine Gal4 vont se rapprocher et une protéine Gal4 active sera alors reconstituée. La fonction biologique de la protéine Gal4 étant l'activation transcriptionnelle de gènes, la protéine Gal4 reconstituée pourra activer la transcription d'un gène dit rapporteur ou *reporter* (voir Fiche 38).

Afin de réaliser ce test, les gènes X et Y sont clonés en aval des séquences codant le domaine BD ou le domaine AD de la protéine Gal4. Ces deux plasmides sont utilisés pour transformer une souche de levure qui contient le gène rapporteur *LacZ* codant la β -galactosidase sous le contrôle d'un promoteur activable par la protéine Gal4 (promoteur Gal1). Le gène rapporteur ne sera transcrit que si les deux protéines X et Y, en interagissant entre elles, reconstituent une protéine Gal4 fonctionnelle. Le produit du gène rapporteur *LacZ* pourra alors être détecté par mesure de son activité enzymatique ou par coloration bleue des colonies de levures cultivées sur des milieux gélosés en présence d'un substrat qui donnera une réaction colorée (X-Gal, voir Fiche 38).

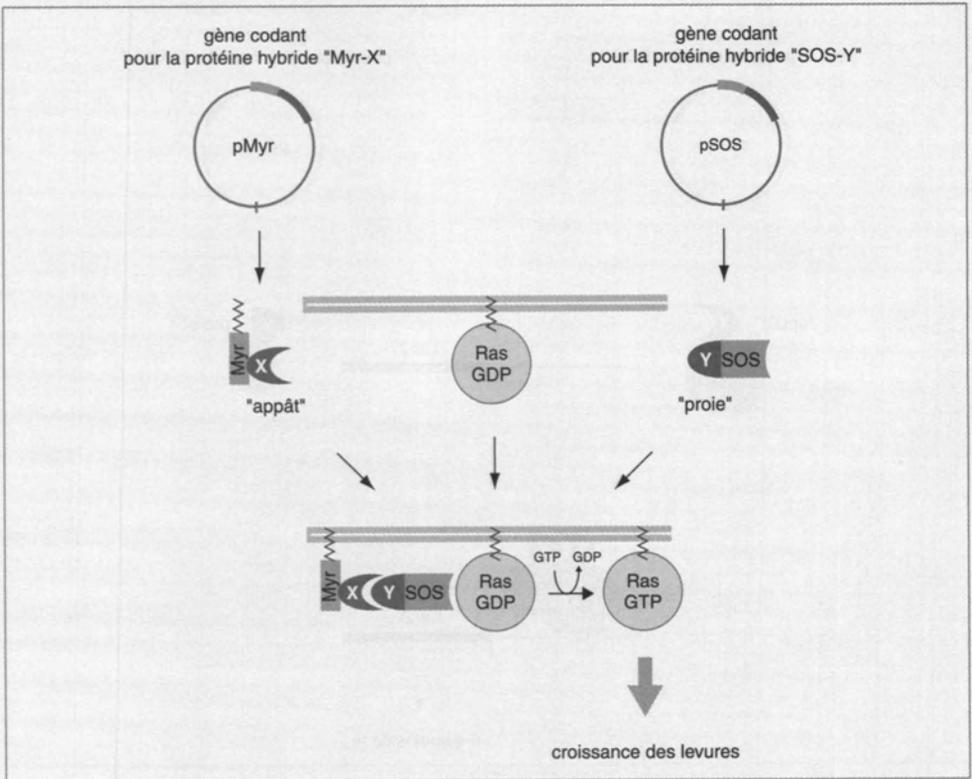


Applications

Ce système permet d'étudier de manière simple et peu coûteuse des interactions entre protéines provenant de n'importe quel organisme eucaryote. Cependant, la protéine Gal4 est active dans le noyau puisqu'elle se fixe sur l'ADN. Donc, ce test de double hybride n'est valable que si l'interaction entre les deux protéines X et Y peut avoir lieu dans le noyau de la cellule. Différentes stratégies ont donc été développées pour étudier des interactions entre deux protéines membranaires ou deux protéines cytosoliques.

Protéines membranaires

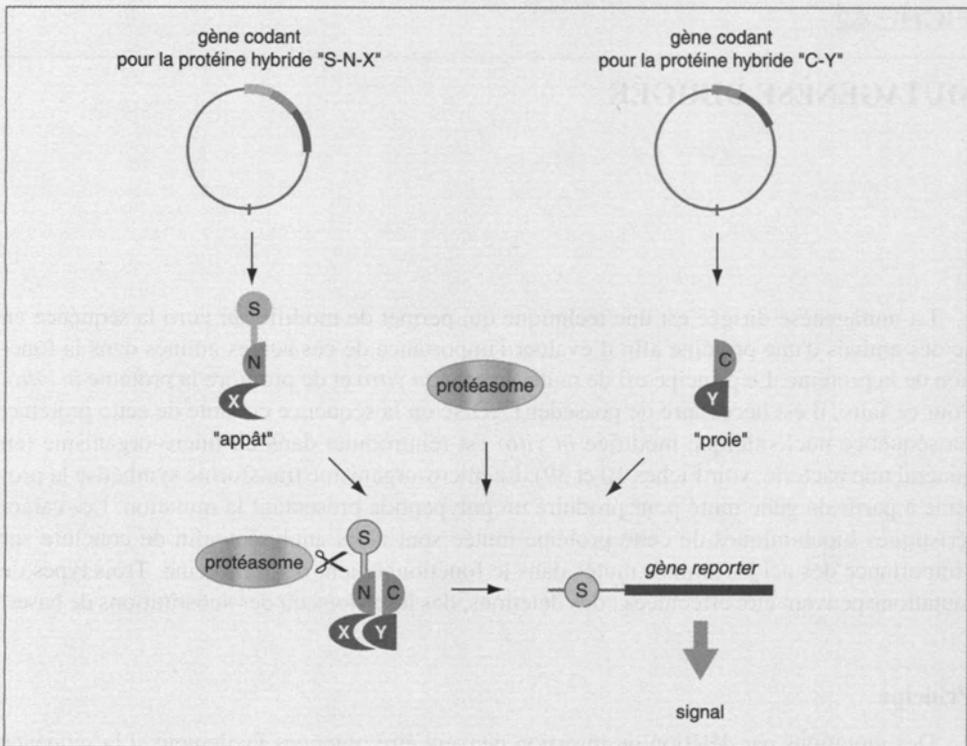
Dans l'exemple présenté ici, nous voulons savoir si la protéine X (l'appât) interagit avec la protéine Y (la proie) au niveau membranaire. Pour réaliser cet essai, il est nécessaire de travailler avec une souche de levure mutante qui n'exprime pas la protéine GEF (*Guanine nucleotide Exchange Factor*). En absence de GEF, la protéine Ras n'est pas activée et la levure ne peut se développer¹. Le gène de la protéine X est fusionné à un segment d'ADN codant pour un signal de myristylation (Myr) qui permet sa localisation membranaire. Les levures exprimant ce gène auront donc la protéine X ancrée dans la membrane plasmique vers le cytosol. Le gène de la protéine Y est fusionné à celui d'une protéine appelée SOS (l'équivalent de GEF chez les mammifères). L'interaction de X et Y permet la localisation membranaire de la protéine SOS. SOS active alors la protéine Ras (par échange de son GDP contre un GTP) et la levure peut pousser à la température restrictive, alors qu'en l'absence d'interaction entre X et Y, aucune croissance n'est possible, donc aucune colonie n'est visible.



Protéines cytosoliques

Dans les cellules eucaryotes, les protéines fixées au peptide ubiquitine sont reconnues et dégradées dans le cytosol par le complexe protéique du protéasome. Pour savoir si les protéines X et Y interagissent dans le cytosol, le gène de la protéine X est fusionné au fragment du gène de l'ubiquitine codant la partie N-terminale de ce peptide (noté N sur la figure), lui-même fusionné au gène codant pour un facteur de transcription (noté S) qui contrôle un gène *reporter*. En parallèle, le gène de la protéine Y est fusionné au fragment du gène de l'ubiquitine codant la partie C-terminale (noté C). Si X et Y interagissent, l'ubiquitine est reconstituée et le système est reconnu par le protéasome. Le clivage obtenu à la suite de l'interaction entre la protéine X et Y permet la libération du facteur de transcription qui à son tour active un gène *reporter* dans le noyau, ce qui génère un signal. Par conséquent, l'interaction entre X et Y pourra être détectée par sélection des souches par un test colorimétrique.

Le système double hybride est également une technique très efficace pour rechercher non pas un partenaire, mais tous les partenaires pouvant interagir avec la protéine d'intérêt X. Dans ce cas, une banque d'expression d'ADNc est construite (voir Fiche 15) dans un vecteur permettant de fusionner à chaque ADNc de la banque le domaine AD. La protéine X connue est quant à elle fusionnée au domaine BD. La banque d'ADNc est criblée par incubation des bactéries exprimant les protéines Y fusionnées à AD avec la protéine X fusionnée à BD. Cependant, cette technique génère encore un nombre important de "faux positifs".



Interactome

Le séquençage de génomes complets a radicalement changé la recherche de partenaires cellulaires par double hybride. Dans ce cas, il s'agit de faire un catalogue de toutes les interactions protéine-protéine possibles. La levure contient environ 6 000 gènes. Afin de réaliser l'interactome de la levure, il faut fusionner chacun de ces 6 000 gènes avec le domaine AD et une collection de 6 000 levures contenant chacune une construction différente sera obtenue. Parallèlement, une collection identique est construite en fusionnant chacun des 6 000 gènes avec le domaine BD. On aura pris soin de construire les deux collections dans des souches de levures haploïdes et sexuellement compatibles. Tous les croisements deux à deux seront réalisés et les résultats analysés. Cette approche nécessite une grande automatisation et une analyse informatique des matrices obtenues. De grands efforts restent à faire si un interactome du génome humain (plus de 30 000 gènes) doit être réalisé.

1. La mutation est thermosensible, ce qui permet de conserver les souches mutées à une température permissive pour laquelle la mutation est muette.

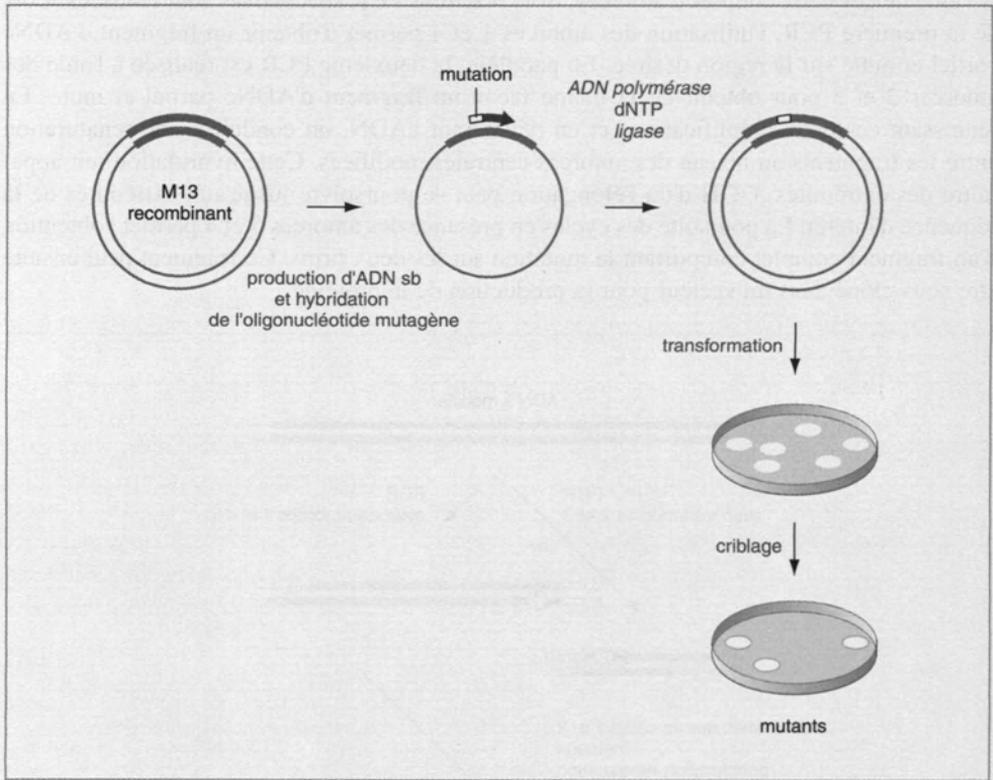
MUTAGENÈSE DIRIGÉE

La mutagenèse dirigée est une technique qui permet de modifier *in vitro* la séquence en acides aminés d'une protéine afin d'évaluer l'importance de ces acides aminés dans la fonction de la protéine. Le principe est de muter le gène *in vitro* et de produire la protéine *in vitro*. Pour ce faire, il est nécessaire de posséder l'ADNc ou la séquence codante de cette protéine. La séquence nucléotidique modifiée *in vitro* est réintroduite dans un micro-organisme (en général une bactérie, voir Fiches 10 et 39). Le micro-organisme transformé synthétise la protéine à partir du gène muté pour produire un polypeptide présentant la mutation. Les caractéristiques biochimiques de cette protéine mutée sont alors analysées afin de conclure sur l'importance des acides aminés mutés dans le fonctionnement de la protéine. Trois types de mutations peuvent être effectuées : des délétions, des insertions ou des substitutions de bases.

Principe

Des mutations par délétion ou insertion peuvent être obtenues facilement si la séquence du gène contient des sites de restriction appropriés. Par digestion à l'aide des enzymes de restriction adéquates, il est alors possible d'éliminer une partie de la séquence ou d'introduire un fragment d'ADN exogène à une position définie. Néanmoins, cette approche est limitée car elle ne permet pas une mutation ciblée à un petit nombre d'acides aminés. Pour remédier à ces inconvénients, une technique de mutagenèse dirigée *in vitro* en présence d'oligonucléotides mutagènes de séquence connue a été développée. Cette méthode permet d'effectuer les modifications désirées à une position choisie dans un fragment d'ADN de séquence connue. Par exemple, le changement d'un seul acide aminé d'une protéine, dont la séquence du gène correspondant est connue, peut être réalisé.

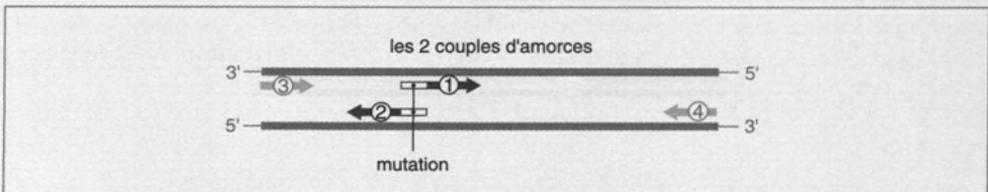
Le principe de base repose sur la synthèse d'un oligonucléotide de séquence complémentaire à celle de la région du gène à muter, mais possédant la mutation désirée. Le gène à muter est introduit dans un vecteur phagique de type M13 et l'ADN est purifié sous sa forme monocaténaire. L'oligonucléotide mutagène est alors hybridé avec l'ADN cible monocaténaire. Le duplex ainsi formé est utilisé comme amorce par l'ADN polymérase qui synthétise le brin complémentaire. Les molécules obtenues correspondent à un plasmide bicaténaire dont l'un possède la mutation et l'autre non. Des bactéries sont transformées avec cette molécule : certaines vont posséder le brin muté, d'autres le brin sauvage. Un séquençage d'un petit nombre de clones est alors nécessaire pour identifier les phages possédant la séquence mutée. Les protéines mutantes sont ensuite produites dans *E. coli* afin d'analyser l'effet de la modification de la séquence modifiée.



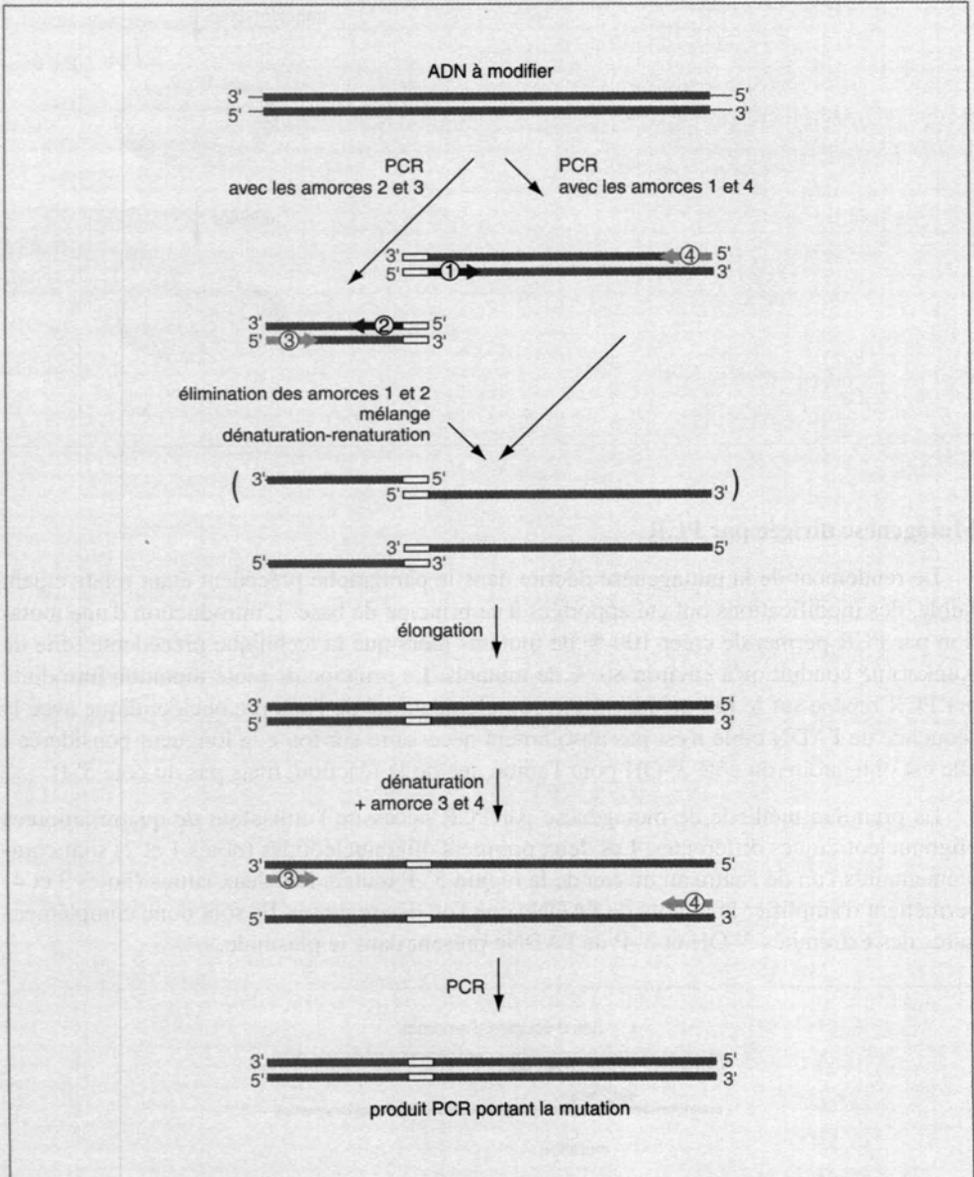
Mutagenèse dirigée par PCR

Le rendement de la mutagenèse décrite dans le paragraphe précédent étant relativement faible, des modifications ont été apportées à ce principe de base. L'introduction d'une mutation par PCR permet de créer 100 % de mutants alors que la technique précédente (dite de Kunkel) ne conduit qu'à environ 80 % de mutants. Le principe de toute mutation introduite par PCR repose sur le fait qu'une stricte complémentarité de l'amorce nucléotidique avec la séquence de l'ADN cible n'est pas absolument nécessaire sur toute la longueur considérée : elle est obligatoire du côté 3'-OH pour l'amorçage de la réaction, mais pas du côté 5'-P.

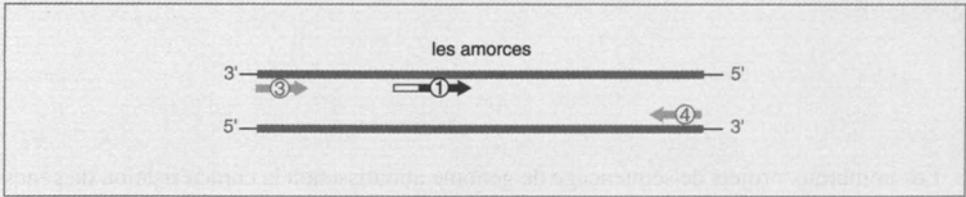
La première méthode de mutagenèse par PCR nécessite l'utilisation de quatre amorces oligonucléotidiques différentes. Les deux premiers oligonucléotides (notés 1 et 2) sont complémentaires l'un de l'autre au niveau de la région 5'-P mutée. Les deux autres (notés 3 et 4) permettent d'amplifier la totalité de l'ADNc que l'on désire muter. Ils sont donc complémentaires des extrémités 3'-OH et 5'-P de l'ADNc présent dans le plasmide.



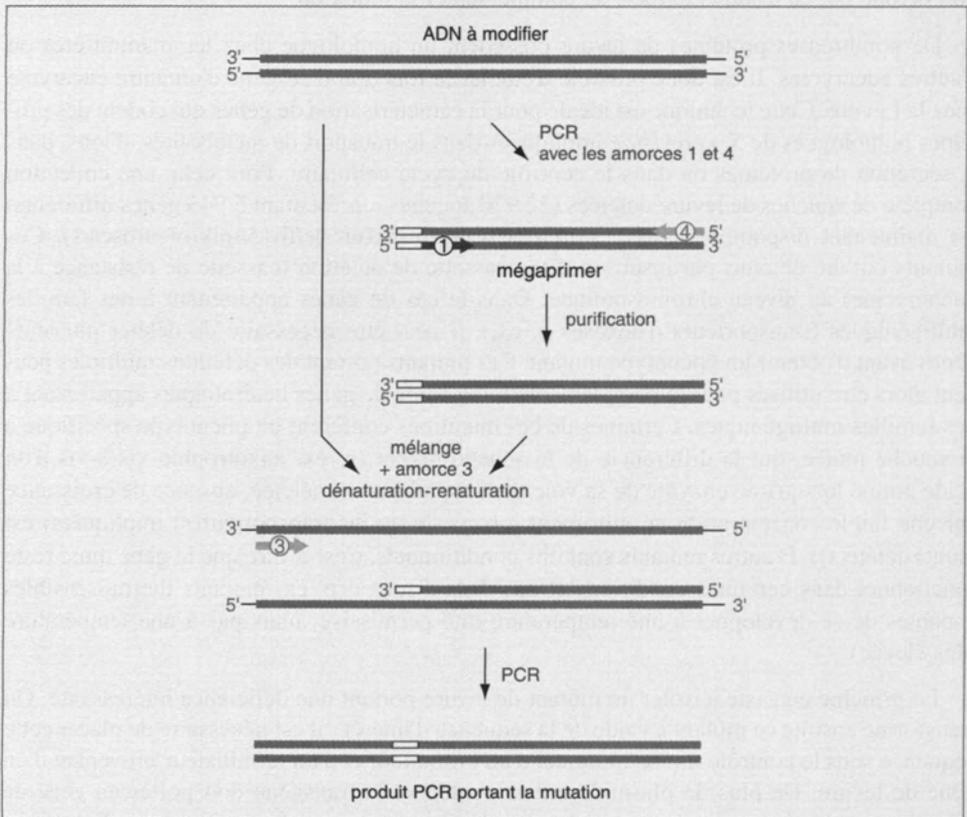
A l'aide de ces deux couples d'amorces, trois réactions PCR successives sont réalisées. Lors de la première PCR, l'utilisation des amorces 1 et 4 permet d'obtenir un fragment d'ADNc partiel et muté sur la région désirée. En parallèle, la deuxième PCR est réalisée à l'aide des amorces 2 et 3 pour obtenir de la même façon un fragment d'ADNc partiel et muté. En réunissant ces deux amplifications et en dénaturant l'ADN, on conduit à une renaturation entre les fragments au niveau des amorces centrales modifiées. Cette hybridation fait apparaître des extrémités 3'-OH d'où l'élongation peut se poursuivre jusqu'aux extrémités de la séquence d'intérêt. La poursuite des cycles en présence des amorces 3 et 4 permet l'obtention d'un fragment complet comportant la mutation sur les deux brins. Ce fragment peut ensuite être sous-cloné dans un vecteur pour la production de la protéine.



La seconde méthode de mutagenèse par PCR est dite du "mégaprimer". Elle nécessite l'utilisation de trois amorces oligonucléotidiques. La première porte la mutation que l'on désire introduire (notée 1) et est complémentaire de l'un des brins de l'ADN cible. Les deux autres (notées 3 et 4), comme précédemment, correspondent aux extrémités de l'ADNc.



Dans la première réaction de PCR, l'utilisation des amorces 1 et 4 permet l'obtention d'un fragment PCR portant la mutation et correspondant à une séquence tronquée de l'ADNc matrice. Ce fragment est purifié et va servir d'amorce (mégaprimer) dans une deuxième réaction de PCR avec l'oligonucléotide 3. De cette façon, on génère un fragment complet et muté.



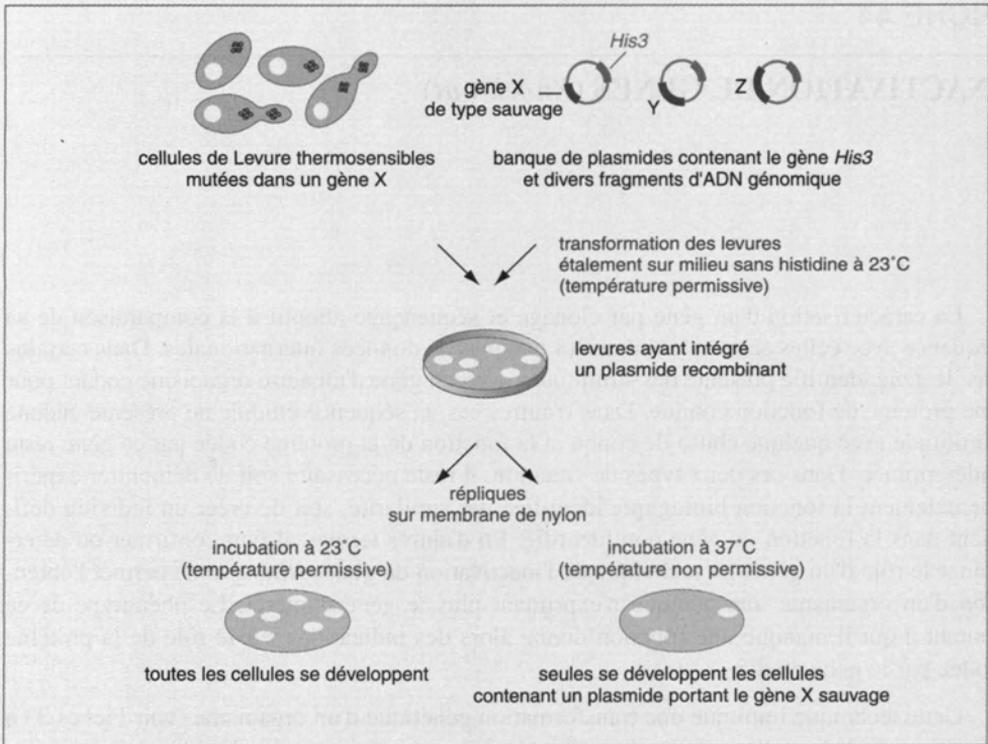
L'un des inconvénients de la technique de mutagenèse par PCR, ce sont les erreurs de lecture de la *Taq* polymérase. Ce problème est aisément surmonté puisque l'on travaille en utilisant un nombre réduit de cycles (< 30) et que de nouvelles générations de *Taq* polymérase réputées fidèles sont disponibles dans le commerce.

COMPLÉMENTATION DE MUTATION CHEZ LA LEVURE

Les nombreux projets de séquençage de génome aboutissent à la caractérisation de gènes à fonction inconnue ou identifiée par simple homologie de séquence. L'une des façons de démontrer la fonction biologique de ces gènes est de compléter une mutation de la fonction étudiée chez la Levure. La transformation génétique permet l'intégration du gène d'intérêt dans la levure mutée ayant perdu une fonction connue, puis le phénotype de la levure transformée est déterminé afin de savoir si le gène introduit est capable de récupérer la fonction perdue ou, en d'autres termes, de compléter la mutation.

De nombreuses protéines de levure possèdent un homologue chez les mammifères ou d'autres eucaryotes. Il est donc possible d'étudier la fonction d'un gène d'un autre eucaryote chez la Levure. Cette technique est idéale pour la caractérisation de gènes qui codent des protéines homologues de *S. cerevisiae* impliquées dans le transport de métabolites, d'ions, dans la sécrétion de protéines ou dans le contrôle du cycle cellulaire. Pour cela, une collection complète de souches de levure délétées (22 970 souches représentant 5 943 gènes différents) est maintenant disponible (URL : <http://www.uni-frankfurt.de/fb15/mikro/euroscarf>). Ces mutants ont été obtenus par insertion d'une cassette de délétion (cassette de résistance à la kanamycine) au niveau chromosomique. Dans le cas de gènes appartenant à des familles multigéniques (transporteurs d'hexoses p. ex.), il peut être nécessaire de déléter plusieurs gènes avant d'obtenir un phénotype mutant. Ces mutants portant des délétions multiples peuvent alors être utilisés pour la complémentation à l'aide de gènes hétérologues appartenant à des familles multigéniques. Certaines de ces mutations confèrent un phénotype spécifique à la souche mutée, qui la différencie de la souche parent (p. ex. auxotrophie vis-à-vis d'un acide aminé lorsqu'une enzyme de sa voie de biosynthèse est délétée, absence de croissance sur une faible concentration en nutriment lorsque le ou les transporteur(s) impliqué(s) est (sont) délété(s)). D'autres mutants sont dits conditionnels, c'est-à-dire que le gène muté reste fonctionnel dans certaines conditions et pas dans d'autres (p. ex. mutants thermosensibles capables de se développer à une température dite permissive, mais pas à une température plus élevée).

Le principe consiste à isoler un mutant de levure portant une déficience intéressante. On transforme ensuite ce mutant à l'aide de la séquence d'intérêt : il est nécessaire de placer cette séquence sous le contrôle transcriptionnel d'un promoteur et d'un terminateur provenant d'un gène de levure. De plus, le plasmide utilisé contenant le transgène doit porter un gène de sélection pour la levure, par exemple le gène *His3*¹. Après transformation, les cellules sont étalées sur boîte de Pétri dans des conditions ne permettant que le développement des levures transformées par un plasmide. Des répliques sur membrane de nylon sont ensuite réalisées et utilisées pour sélectionner les colonies contenant le gène de type sauvage permettant de restaurer l'activité déficiente. Le plasmide contenu dans une telle colonie est isolé et analysé.



L'application de cette technique peut être élargie à la recherche dans un génome donné d'un gène présentant une fonction particulière, s'il existe un mutant de levure pour cette fonction. Dans ce cas, une banque d'ADNc ou génomique est construite dans un vecteur permettant l'expression de ces séquences dans la Levure (même type de plasmide que celui utilisé précédemment). Le mutant d'intérêt est alors transformé à l'aide de la totalité de la banque d'ADN et les cellules complémentées sélectionnées. Une fois que les transformants avec le phénotype approprié ont été obtenus, il est essentiel de confirmer que la complémentation est liée au plasmide. En effet, la complémentation d'une déficience peut être parfois due à un suppresseur dont la surexpression peut contrecarrer l'effet de la mutation. Un test de perte de plasmide, par croissance de la souche sur milieu riche par exemple, est nécessaire. Les levures ayant perdu le plasmide doivent être incapables de se développer sur le milieu sélectif. Après ce test, le plasmide portant le gène d'intérêt peut être isolé de la souche de levure, multiplié chez *E. coli* et réintroduit dans la levure déficiente afin de confirmer la complémentation de la déficience.

1. Le gène *His3* code pour l'imidazole glycérol phosphate déshydrogénase. Il permet de sélectionner les levures exprimant ce gène par croissance sur milieu sans histidine.

INACTIVATION DE GÈNES (*Knock-out*)

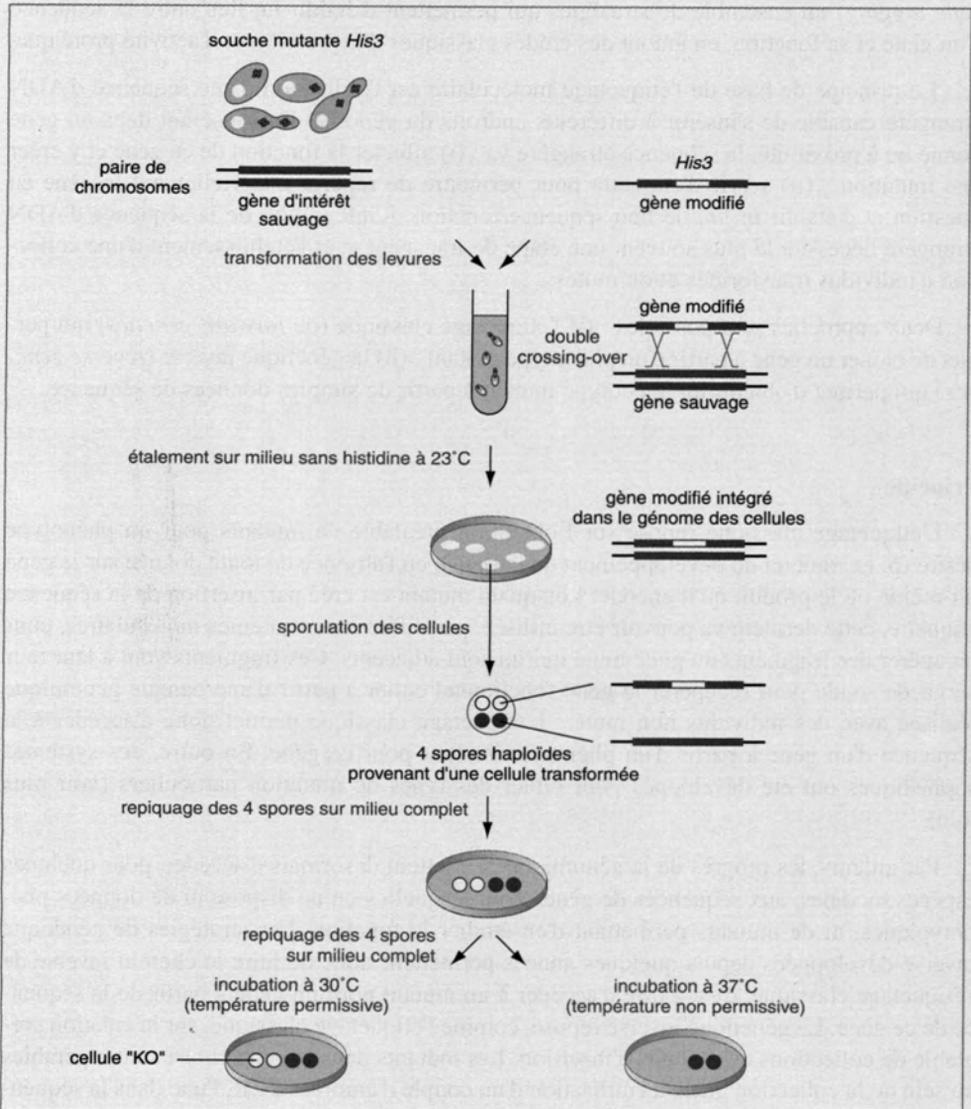
La caractérisation d'un gène par clonage et séquençage aboutit à la comparaison de sa séquence avec celles accumulées dans les banques de données internationales. Dans certains cas, le gène identifié présente des similitudes avec un gène d'un autre organisme codant pour une protéine de fonction connue. Dans d'autres cas, la séquence étudiée ne présente aucune similitude avec quelque chose de connu et la fonction de la protéine codée par ce gène reste indéterminée. Dans ces deux types de situation, il reste nécessaire soit de démontrer expérimentalement la fonction biologique identifiée par similarité, soit de créer un individu déficient dans la fonction du gène non identifié. En d'autres termes, il faut confirmer ou déterminer le rôle d'un gène. La technique de l'inactivation de gènes (*Knock-out*) permet l'obtention d'un organisme transgénique n'exprimant plus le gène d'intérêt. Le phénotype de ce mutant à qui il manque une fonction donne alors des indications sur le rôle de la protéine codée par le gène étudié.

Cette technique implique une transformation génétique d'un organisme (voir Fiches 33 à 37) au cours de laquelle un transgène est intégré dans la séquence du gène d'intérêt et non au hasard dans le génome. Pour ce faire, on modifie *in vitro*, à l'aide d'enzymes de restriction (voir Fiche 4), un gène d'intérêt en y plaçant la séquence codant pour un gène de sélection. Ainsi, la traduction du gène d'intérêt donnera une protéine incomplète et inactive. Cette construction est alors intégrée dans l'organisme par transformation. Les cellules transformées sont sélectionnées sur l'agent sélectif présent dans le milieu. Lors de l'événement de transformation, le transgène peut soit s'intégrer au hasard dans le génome, soit s'intégrer par recombinaison homologue au locus visé. Il faut alors trier les cellules qui ont subi la recombinaison homologue. Chez la Levure et la plupart des champignons, la sélection se fait *a posteriori*, en analysant le site d'intégration du transgène dans le génome des cellules résistantes à l'agent sélectif. Il suffit de digérer l'ADN génomique par des enzymes de restriction qui permettent de discriminer le gène endogène non touché par l'insertion et le gène endogène modifié par le transgène. Les organismes ainsi obtenus deviennent incapables de synthétiser la protéine d'intérêt et son phénotype peut être étudié. La recombinaison homologue, donc la technique d'inactivation de gènes, n'est pas applicable pour tous les organismes eucaryotes. La technique reste à démontrer, notamment chez les plantes.

Lors d'une recombinaison homologue, deux événements de recombinaison, un à chaque extrémité du gène, sont nécessaires pour remplacer de façon précise le gène chromosomique par le gène muté. C'est un événement rare et la fréquence de recombinaison homologue est faible. De plus, l'inactivation d'un gène aboutit à une mutation récessive (perte de fonction) qui est donc silencieuse dans les individus diploïdes : l'événement de recombinaison homologue ne touche qu'un des deux allèles du locus touché, l'autre pouvant s'exprimer et donner une protéine sauvage active. Il est donc nécessaire d'isoler des individus haploïdes portant la mutation. Chez la Levure, par exemple, on induit une sporulation qui produit par méiose qua-

tre spores haploïdes (équivalentes aux gamètes) : deux d'entre elles portent le gène muté inactivé ainsi que le gène de sélection, et deux d'entre elles ont le gène sauvage actif et sont privées du gène de sélection. Ces spores sont mises à germer sur milieu sélectif afin de sélectionner les colonies provenant des spores transformées.

Cette technique peut également être appliquée au modèle animal en utilisant des cellules embryonnaires de mammifères et un gène modifié portant un marqueur de sélection adéquat. Dans ce cas, des animaux transgéniques hétérozygotes sont obtenus à la première descendance. Ils doivent être croisés entre eux pour obtenir des individus homozygotes dont le phénotype sera étudié. Cependant, la fréquence de la recombinaison homologe reste très faible : elle est par exemple 100 fois plus faible chez la Souris que chez la Levure.



ÉTIQUETAGE MOLÉCULAIRE

On regroupe sous le terme général d'étiquetage moléculaire (ou étiquetage génétique, *gene tagging*) un ensemble de stratégies qui permettent d'établir un lien entre la séquence d'un gène et sa fonction, en amont des études classiques d'expression et d'activité protéique.

Le principe de base de l'étiquetage moléculaire est l'utilisation d'une séquence d'ADN étrangère capable de s'insérer à différents endroits du génome. En s'insérant dans un gène donné ou à proximité, la séquence étrangère va : (i) affecter la fonction de ce gène et y créer une mutation ; (ii) servir d'étiquette pour permettre de repérer matériellement le gène en question et d'établir *in fine* le lien séquence/fonction. L'intégration de la séquence d'ADN étrangère nécessite le plus souvent une étape de transgénèse et l'établissement d'une collection d'individus transformés et/ou mutés.

Deux approches sont possibles : (i) l'étiquetage classique (ou *forward genetics*) qui permet de cloner un gène à partir d'un phénotype mutant ; (ii) la génétique inverse (*reverse genetics*) qui permet d'obtenir un phénotype mutant à partir de simples données de séquence.

Principe

L'étiquetage classique repose sur l'obtention préalable de mutants pour un phénotype désiré (p. ex. mutant du développement de la fleur), en l'absence de toute donnée sur le gène lui-même ou le produit qu'il encode. Lorsqu'un mutant est créé par insertion de la séquence étiquette, cette dernière va pouvoir être utilisée, par différentes stratégies moléculaires, pour récupérer des fragments du gène muté qui lui sont adjacents. Ces fragments vont à leur tour servir de sonde pour récupérer le gène fonctionnel entier à partir d'une banque génomique réalisée avec des individus non mutés. L'étiquetage classique permet donc d'accéder à la séquence d'un gène à partir d'un phénotype mutant pour ce gène. En outre, des systèmes sophistiqués ont été développés pour cibler des types de mutation particuliers (voir plus loin).

Par ailleurs, les progrès de la génomique permettent désormais d'accéder, pour quelques espèces modèles, aux séquences de gènes pour lesquelles on ne dispose ni de données phénotypiques, ni de mutants permettant d'en étudier la fonction. Les stratégies de génétique inverse développées depuis quelques années permettent donc de faire le chemin inverse de l'étiquetage classique, c'est à dire d'accéder à un mutant pour un gène à partir de la séquence de ce gène. La génétique inverse repose, comme l'étiquetage classique, sur la création préalable de collections de mutants d'insertion. Les mutants dans le gène choisi sont repérables au sein de la collection grâce à l'utilisation d'un couple d'amorces PCR, l'une dans la séquence du gène, l'autre dans l'étiquette : l'amplification entre ces deux amorces ne pourra avoir lieu que si la lignée porte une insertion dans le gène choisi et il sera alors possible de faire

correspondre le gène avec son phénotype mutant associé. Toute la complexité de la stratégie repose sur la nécessité de cribler un grand nombre de lignées : on les regroupe en une pyramide de "pools" et de "sous-pools" qui sont analysés de façon séquentielle, jusqu'à l'identification de lignées individuelles. La génétique inverse est principalement utilisée pour des organismes largement séquencés, comme la plante modèle *Arabidopsis thaliana*, la Levure ou la Drosophile.

Nature de l'étiquette

Deux types d'étiquette peuvent être utilisées : (i) les transposons ; (ii) l'ADN-T (uniquement chez les plantes). L'ADN-T¹ est un fragment d'ADN plasmidique originaire de la bactérie *Agrobacterium tumefaciens*, capable de se transférer à un génome de plante infectée (voir Fiche 33). La génération de collections de mutants d'insertion par l'ADN-T nécessite donc une étape de transformation, mais les insertions sont stables.

Les transposons sont des éléments génétiques capables de s'insérer dans le génome grâce à des fonctions qui leur sont propres (tout au moins dans le cas des versions autonomes). Ils se différencient en deux classes : (i) les éléments à ADN, comme les éléments Ac du Maïs ou P de la Drosophile, qui transposent en s'excisant hors du locus d'origine ; (ii) les rétroéléments, qui transposent de manière répliquative par l'intermédiaire d'un ARN. Ces derniers ont l'avantage de créer des insertions stables. A l'heure actuelle, les éléments les plus utilisés chez les plantes et chez la Drosophile sont des éléments à ADN. Les rétroéléments sont essentiellement utilisés pour l'étiquetage chez la Levure et la Souris ; leur utilisation a également commencé chez les plantes, notamment chez le Riz.

Chez les plantes, les transposons sont utilisés pour étiqueter des gènes, soit dans l'espèce qui les héberge (situation homologue), soit après leur introduction dans le génome d'une autre espèce (situation hétérologue). La situation hétérologue est largement favorite, d'une part parce qu'elle permet de s'abstraire des problèmes que crée la présence de nombreuses copies homologues endogènes, mais aussi parce qu'elle permet d'utiliser des étiquettes plus sophistiquées, modifiées de façon à orienter le type de mutation ou le type de fonction recherché (voir plus loin). Elle requiert toutefois de disposer d'un élément autonome fonctionnel et qui reste mobile dans l'espèce hétérologue, ce qui n'est toujours le cas. Notons que le développement récent de stratégies PCR d'analyse des séquences génomiques flanquant une insertion a largement facilité l'utilisation en situation homologue d'éléments présents en de nombreuses copies (voir plus loin).

Le principal avantage des transposons réside en leur capacité à bouger par eux-mêmes, contrairement à l'ADN-T qui ne s'insère que lors de l'étape de transformation. De plus, en situation hétérologue, il est possible de contrôler leur mobilité en utilisant des versions défectives, qui ne bougent que si une version autonome est présente en même temps. Par des jeux de croisements génétiques habiles, il est donc possible de faire bouger puis d'immobiliser l'étiquette à volonté, et de la remobiliser ultérieurement. De cette façon, il est également possible de s'affranchir du principal inconvénient des éléments transposables à ADN, qui est le risque de perdre la mutation d'insertion si l'élément s'excise hors du locus qu'il a étiqueté.

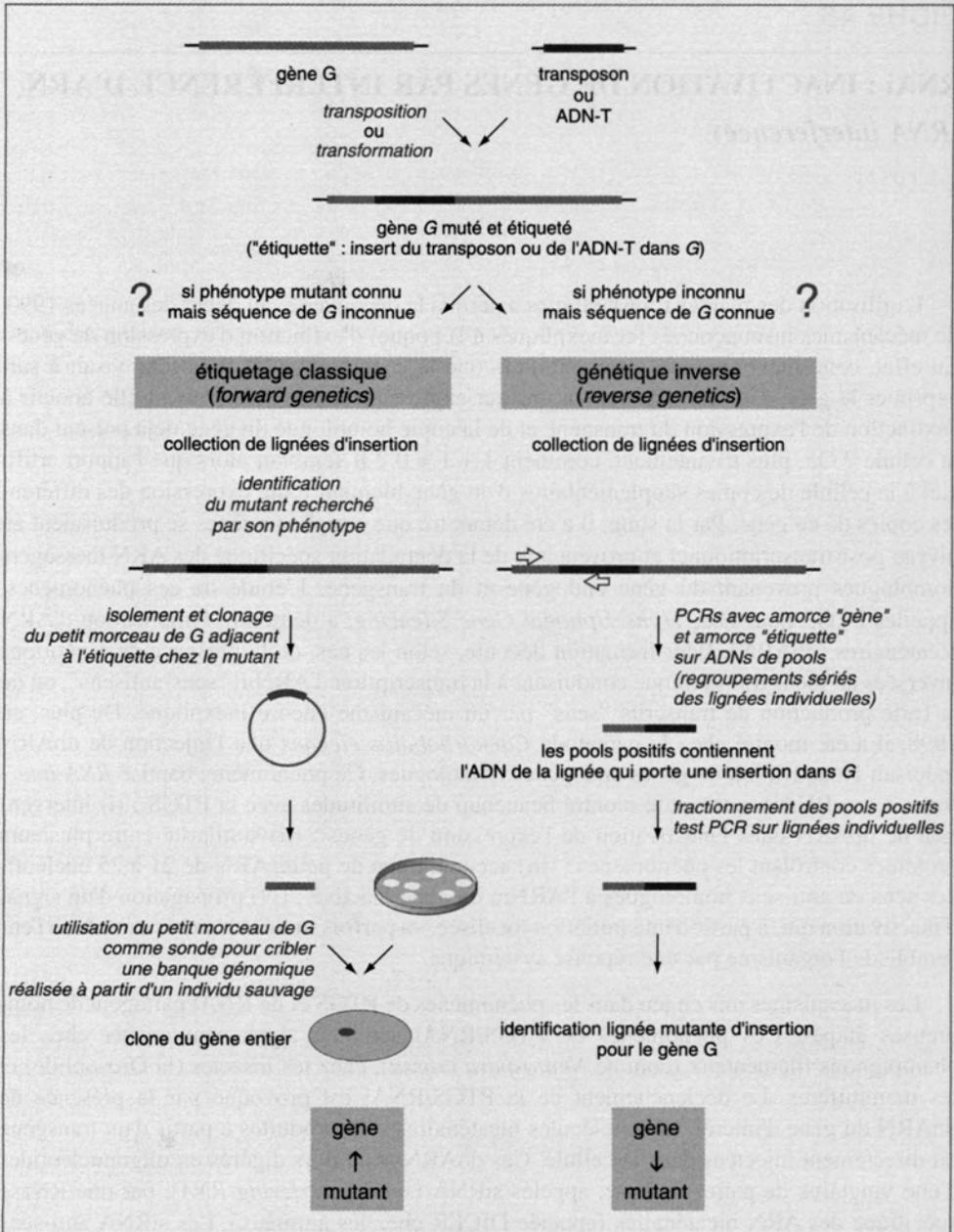
Identification des séquences génomiques flanquant l'insertion (étiquetage classique)

L'étape fondamentale de l'étiquetage classique est l'obtention des fragments du gène flanquant l'étiquette, lesquels vont ultérieurement servir à cloner le gène entier à partir d'une banque génomique réalisée avec des individus non mutés. Cette étape a longtemps fait appel à la construction de banques génomiques à partir du mutant et au criblage de cette banque avec une sonde "étiquette". Les clones positifs sont ensuite analysés pour repérer puis sous-cloner le fragment génomique partiel flanquant l'étiquette. Ce type de clonage "standard" est relativement aisé lorsque l'on travaille en système hétérologue ou avec un transposon naturellement présent en petit nombre de copies (ce qui est rarement le cas), et pour lequel on peut aisément suivre la co-ségrégation du phénotype mutant et d'un fragment identifiable par la technique de *Southern* par exemple. Plus récemment, des stratégies PCR comme l'iPCR ou la tail-PCR, développée avec l'essor de l'utilisation de l'ADN-T, ont permis de faciliter grandement le clonage des régions flanquant les insertions.

Dans les situations où le transposon est présent en grand nombre de copies, il est plus délicat de distinguer la copie qui sert d'étiquette de toutes les autres copies présentes. Un progrès très sensible a été réalisé récemment grâce à la mise au point de techniques fines permettant d'amplifier simultanément un grand nombre de sites d'insertion et de les visualiser ensemble, avec une grande résolution, sur un gel d'acrylamide. Ces techniques, nommées *Transposon Display* ou SSAP (voir Fiche 53), consistent en une amplification de type AFLP (voir Fiche 52) ancrée dans l'insertion et permettent de suivre la ségrégation d'une bande particulière au sein d'un grand nombre de copies présentes (plusieurs dizaines, voire plusieurs centaines).

Systèmes d'étiquetage modifié

L'une des conditions clés du succès d'une stratégie d'étiquetage, qu'elle soit classique ou inverse, consiste en une réduction maximale des populations d'individus à screener avant d'obtenir l'insertion à l'endroit désiré. Grâce à l'utilisation d'étiquettes modifiées, porteuses de séquences et de fonctions diverses, qui permettent de cibler voire de sélectionner pour un type de mutation donnée, on peut améliorer sensiblement les chances de succès. Il est possible de construire des étiquettes qui permettent de repérer plus facilement une insertion dans un gène, ou dans un promoteur, par exemple parce qu'elles sont porteuses d'un gène rapporteur dont l'expression ne sera assurée qu'après insertion dans une séquence codante (*gene trap*) ou dans un promoteur (*enhancer trap*). Les gènes rapporteurs les plus utilisés sont des marqueurs colorimétriques qui permettent du même coup d'étudier le mode d'expression du gène étiqueté. Mais il est également possible d'utiliser des marqueurs de sélection (résistance à un antibiotique p. ex.), ce qui permet alors de sélectionner directement les individus porteurs de ce type d'insertion, et non plus seulement de les repérer au sein de la collection. Un pas supplémentaire dans la sophistication est l'association des deux types de marqueurs au sein de la même étiquette. Inversement, il est possible de rechercher des mutations de gain de fonction, plutôt que de perte de fonction, en insérant dans l'étiquette une séquence promotrice forte qui va permettre d'activer des séquences adjacentes (*activation tagging*). Naturellement, ces améliorations ne sont possibles qu'en utilisant des éléments modifiés réintroduits dans le génome (ainsi que l'ADN-T dans le cas des plantes), c'est-à-dire pour l'essentiel en situation hétérologue.

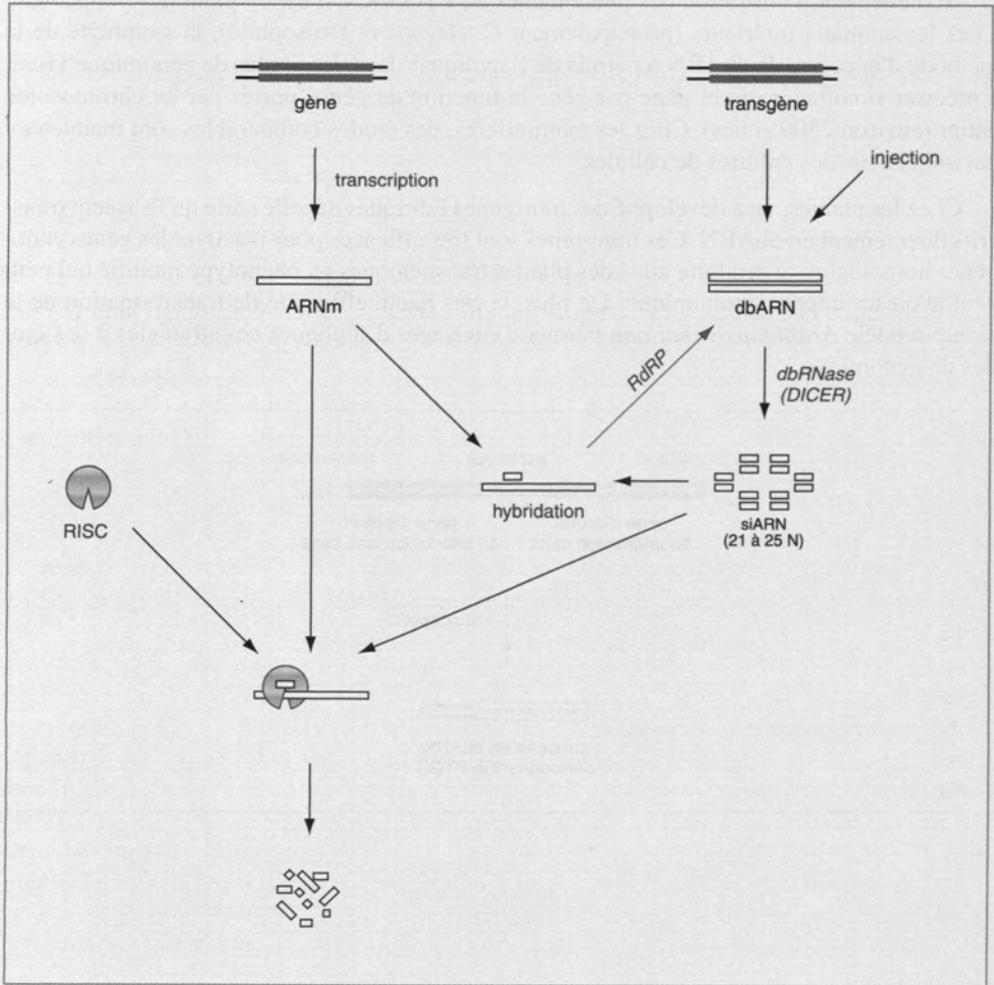


1. La transformation génétique par *Agrobacterium tumefaciens* (porteuse de l'ADN-T) s'étend à d'autres organismes, notamment les champignons.

RNAi : INACTIVATION DE GÈNES PAR INTERFÉRENCE D'ARN (*RNA interference*)

L'utilisation des plantes transgéniques a permis la découverte, au début des années 1990, de mécanismes insoupçonnés (et inexpliqués à l'époque) d'extinction d'expression de gènes. En effet, cette inexpression s'observait alors que la construction du transgène visait à sur-exprimer le gène d'intérêt ! Comment une sur-expression recherchée pouvait-elle aboutir à l'extinction de l'expression du transgène et de la copie homologue du gène déjà présent dans la cellule ? Ou, plus trivialement, comment $1 + 1 = 0$? Il semblait alors que l'apport artificiel à la cellule de copies supplémentaires d'un gène bloquait toute expression des différentes copies de ce gène. Par la suite, il a été démontré que ces phénomènes se produisaient au niveau post-transcriptionnel et provenaient de la dégradation spécifique des ARN messagers homologues provenant du gène endogène et du transgène. L'étude de ces phénomènes, appelés PTGS pour *Post-Transcriptional Gene Silencing*, a démontré l'implication d'ARN bicaténares (dbARN). Leur formation découle, selon les cas, de la présence de répétitions inversées au locus transgénique conduisant à la transcription d'ARNm "sens-antisens", ou de la forte production de transcrits "sens" par un mécanisme encore inexpliqué. De plus, en 1998, il a été montré chez le nématode *Caenorhabditis elegans* que l'injection de dbARN induisait l'inactivation de gènes endogènes homologues. Ce phénomène, baptisé *RNA interference* (ou RNAi), a très vite montré beaucoup de similitudes avec la PTGS : (i) intervention de dbARN dans l'inactivation de l'expression de gènes ; (ii) similarité entre plusieurs protéines contrôlant les phénomènes ; (iii) accumulation de petits ARN de 21 à 25 nucléotides sens ou anti-sens homologues à l'ARNm du gène inactivé ; (iv) propagation d'un signal d'inactivation qui, à partir d'une initiation localisée, va parfois induire l'inactivation dans l'ensemble de l'organisme par une réponse systémique.

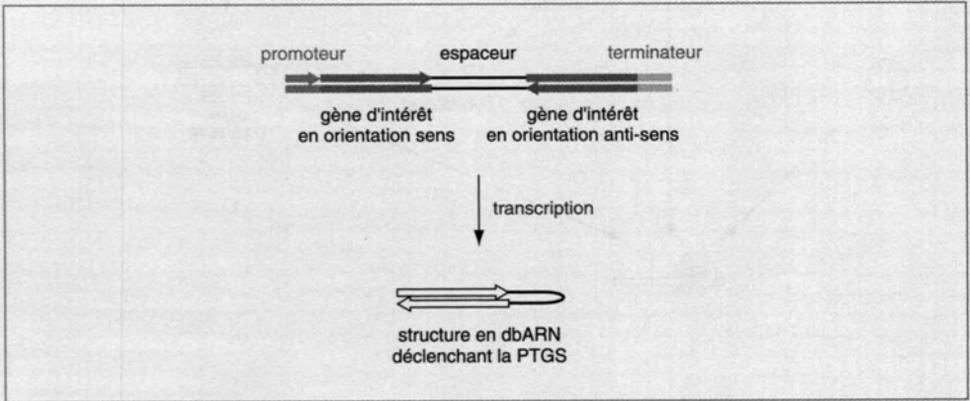
Les mécanismes mis en jeu dans les phénomènes de PTGS et de RNAi partagent de nombreuses étapes. Ces phénomènes de PTGS/RNAi semblent également exister chez les champignons filamenteux (comme *Neurospora crassa*), chez les insectes (la *Drosophile*) et les mammifères. Le déclenchement de la PTGS/RNAi est provoqué par la présence de dbARN du gène d'intérêt. Ces molécules bicaténares sont produites à partir d'un transgène ou directement injectées dans la cellule. Ces dbARN sont alors digérés en oligonucléotides d'une vingtaine de paires de base, appelés siRNA (*small interfering RNA*), par une RNase spécifique des ARN bicaténares (appelée DICER chez les animaux). Les siRNA anti-sens issus de cette coupure vont servir de guide à un complexe enzymatique (*RNA-Induced Silencing Complex* ou RISC) qui va dégrader les ARNm provenant du gène endogène. Cette disparition de transcrit bloque alors la traduction de la protéine correspondante. De plus, chez *C. elegans*, il a été montré que les siRNA anti-sens pouvaient servir d'amorce à une ARN polymérase ARN dépendante (RdRP) pour la synthèse d'ARN complémentaire à partir de l'ARNm du gène endogène, enrichissant alors la cellule en dbARN et amplifiant ainsi le signal de RNAi.



Ainsi, la RNAi peut devenir une technique d'inactivation de gènes relativement aisée à mettre en oeuvre, une bonne alternative à l'inactivation de gènes par recombinaison homologe (voir Fiche 44) qui est limitée à un nombre très réduit d'espèces. La RNAi a très vite été appliquée à de nombreux animaux inférieurs. Par exemple, chez *C. elegans*, il suffit de nourrir ces animaux par des bactéries contenant des dbARN des gènes d'intérêt pour bloquer de façon transitoire et quasi systémique l'expression de ce gène dans presque toutes les cellules du ver. Chez les mammifères, la présence de dbARN dans la cellule induit habituellement des réponses anti-virales de type PKR et/ou interféron, conduisant à l'inhibition généralisée de la traduction dans la cellule. C'est pourquoi les tentatives initiales d'utilisation de la RNAi, du fait de l'induction de ces réponses non-spécifiques après l'injection des dbARN, se sont généralement soldées par des échecs. Seules les cellules ou tissus dépourvus de ce type de réponse (zygotes uni-cellulaires, cultures de cellules non différenciées) ont répondu positivement à la RNAi. Cependant, on a récemment réussi à contourner cet obstacle en limitant la taille des dbARN à 21 nucléotides de structure bien déterminée qui sont capables de provoquer la RNAi sans induire de réponse anti-virale.

Actuellement, l'utilisation des phénomènes de PTGS/RNAi est en plein développement. Chez les animaux inférieurs (principalement *C. elegans* et *Drosophile*), la simplicité de la méthode d'injection de dbARN a permis de l'appliquer dans des études de génomique visant à préciser simultanément et gène par gène la fonction de gènes portés par un chromosome entier (environ 2500 gènes). Chez les mammifères, des études comparables sont maintenant envisagées sur des cultures de cellules.

Chez les plantes, on a développé des transgènes fabriqués de telle sorte qu'ils soient transcrits directement en dbARN. Ces transgènes sont très efficaces pour inactiver les gènes endogènes homologues et produire ainsi des plantes transgéniques au phénotype modifié qui peuvent avoir un intérêt agronomique. De plus, la très haute efficacité de transformation de la plante modèle *Arabidopsis thaliana* permet d'envisager d'appliquer ces stratégies à des études de génomique.



MARQUEURS GÉNÉTIQUES MOLÉCULAIRES

La recherche de polymorphisme d'un génome est une approche en plein essor qui voit chaque année de nouvelles techniques se développer. Chacune d'elle présente des avantages et des inconvénients et le choix d'utilisation dépend essentiellement de la nature du problème génétique à traiter¹.

Qu'est ce qu'un marqueur génétique ?

Un marqueur génétique est un caractère mesurable à hérédité mendélienne. On oppose traditionnellement les marqueurs morphologiques (couleur, forme...) aux marqueurs moléculaires (au niveau de l'ADN) et aux marqueurs biochimiques (isozymes, protéines, métabolites secondaires comme les terpènes). Les caractéristiques d'un marqueur idéal peuvent être résumées comme suit :

- polymorphe : c'est-à-dire variable entre individus ;
- discriminant : c'est-à-dire permettant de différencier des individus très proches ;
- multiallélique : c'est-à-dire possédant plusieurs allèles (donc des séquences d'ADN différentes) sur un même locus ;
- codominant : un individu hétérozygote présente simultanément les caractères des 2 parents homozygotes. Il peut donc être distingué de chacun des parents ;
- non épistatique : le génotype du marqueur peut être déterminé à partir de son phénotype, quel que soit le génotype aux autres locus. En d'autres termes, la lecture du marqueur étudié est indépendante de l'expression des autres marqueurs du génome ;
- indépendant du milieu : le génotype du marqueur peut être déterminé à partir de son phénotype quel que soit le milieu ;
- neutre : quel que soit l'allèle présent au locus marqueur, la valeur sélective de l'individu est la même ;
- réparti uniformément dans tout le génome : afin d'évaluer le polymorphisme en de multiples endroits du génome ;
- reproductible d'une expérience à une autre ;
- économique et manipulable à grande échelle.

Les marqueurs morphologiques sont pour la plupart peu polymorphes et en général dominants. Ils peuvent être influencés par le milieu et interférer avec d'autres caractères. La limitation des marqueurs biochimiques est le faible nombre de loci susceptibles d'être révélés ainsi qu'une certaine spécificité d'organe et/ou de stade de développement. Au contraire, les

marqueurs au niveau de l'ADN sont en nombre quasiment illimité et sont indépendants du stade de développement ou de l'organe analysé. Certaines techniques de marquage révèlent des marqueurs moléculaires de façon "individuelle" : une réaction produit alors un seul marqueur (microsatellites p. ex.). D'autres techniques révèlent des marqueurs en "masse" : une réaction fournit alors plusieurs marqueurs (RAPD ou AFLP p. ex., voir Fiches 51 et 52).

Principales sources de marqueurs moléculaires

Polymorphisme de séquence

Le séquençage est la seule méthode exhaustive de recherche de polymorphisme dans un fragment d'ADN. Malgré les progrès de l'automatisation, l'utilisation en routine de cette technique reste peu développée, à cause du nombre de loci et des grands effectifs généralement nécessaires dans les études de génétique. On a donc recours à des méthodes indirectes, non exhaustives, mais beaucoup plus rapides et moins coûteuses.

- Marqueurs codominants révélés individuellement ; les techniques sont fondées sur la détection de :
 - différences au niveau de sites d'enzymes de restriction : techniques RFLP (voir Fiche 50) et CAPS ;
 - différences de conformation : technique SSCP (voir Fiche 54) ;
 - différence de stabilité : technique DGGE/TGGE (voir Fiche 55).
- Marqueurs dominants révélés en masse : les techniques sont fondées sur la détection de :
 - différences de site d'hybridation d'une amorce arbitraire (une mutation sur le site de fixation d'une amorce empêche sa fixation et il en résulte la disparition d'une bande) : technique de MAPP (RAPD, AP-PCR, DAF) ;
 - différences de sites de restriction et de sites d'hybridation d'une amorce arbitraire : AFLP et tecMAAP.

Polymorphisme de nombre d'unités de répétition

- Les microsatellites

Les microsatellites (ou SSR) sont constitués de répétition en tandem de motifs mono-, di-, tri- ou tétranucléotidiques, par exemple (A)_n, (TC)_n, (TAT)_n, (GATA)_n, la valeur de n pouvant varier de quelques unités à plusieurs dizaines. Les microsatellites sont caractérisés par un polymorphisme extrêmement élevé. Il s'agit d'un polymorphisme de nombre d'unités de répétition. La technique PCR est utilisée pour révéler les loci microsatellites qui sont des marqueurs codominants et multialléliques. Si un microsatellite n'est pas spécifique d'un locus, les régions flanquantes, par contre, le sont. Une paire d'amorces spécifiques de ces régions flanquantes amplifiera donc ce seul microsatellite, dont le polymorphisme de longueur est révélé en gel très résolusif d'acrylamide (voir Fiche 57).

- Les IMA

Une exploitation plus récente des microsatellites consiste à les révéler en masse, en s'inspirant du principe de la RAPD. On utilise pour cela une amorce constituée pour partie d'une séquence de microsatellite et pour partie de bases arbitraires.

- Les minisatellites

De même structure générale que les microsattellites, les minisatellites s'en distinguent par la longueur de l'unité de répétition qui peut être de plusieurs dizaines de bases.

Quels marqueurs pour quoi faire ?

L'une des applications des marqueurs génétiques est l'établissement de cartes génétiques (voir Fiche 48) permettant de repérer les régions chromosomiques qui contrôlent un caractère quantitatif d'intérêt (QTL, *Quantitative Trait Loci*). L'utilisation de ces données dans un programme d'amélioration génétique constitue l'une des facettes de la "sélection assistée par marqueurs" (SAM).

- Les marqueurs dominants MAAP et AFLP permettent de générer des marqueurs en masse et sont bien adaptés aux objectifs nécessitant de nombreux marqueurs. Ils sont donc utilisés :
 - pour saturer rapidement une carte génétique chez les espèces peu travaillées où des sondes RFLP ou des marqueurs STS et autres marqueurs codominants n'ont pas été développés ;
 - en génétique des populations, pour des études de diversité : structuration intra- ou interpopulations, distances génétiques, classifications, etc. ;
 - pour le clonage positionnel, où ils permettent d'obtenir une forte densité de marqueurs au voisinage d'un gène à cloner (technique de marche ou atterrissage sur le chromosome).
- Les marqueurs codominants, spécifiques d'un locus sont aussi très utilisés. En effet :
 - la codominance permet de mesurer la dominance au QTL ;
 - s'ils sont multialléliques, ils seront informatifs dans de nombreux pedigrees et pourront être utilisés pour construire une carte génétique consensus au niveau de l'espèce ;
 - s'ils correspondent à des séquences codantes (EST, voir Fiche 21), ils pourront être utilisés pour la construction de cartes de gènes ;
 - s'ils peuvent être en plus utilisés dans des espèces, voire des genres phylogénétiquement proches, ils pourront permettre de faire de la cartographie comparée.

marqueurs	neutralité	codominance		polymorphisme		technicité	séquence codante		
		nombre	spécificité	de locus ¹	stade, organe ²		coût		
morphologiques	non	limité	rare	oui	faible	oui	faible	faible	?
isozymes	oui	< 40	oui	oui	faible	oui	faible	faible	oui
protéines (2D)	oui	< 100	oui	oui	faible	oui	moyenne	moyen	oui
RFLP	oui	illimité	oui	oui	élevé	non	élevée	moyen	oui / non
MAAP	oui	illimité	non	non	très élevé	non	faible	faible	oui / non
AFLP	oui	illimité	non	non	très élevé	non	moyenne	moyen	oui / non
microsatellites	oui	quelques	oui	oui	très élevé	non	élevée	très élevé ³	oui / non
IMA	oui	illimité	oui	non	très élevé	non	faible	faible	non
STS	oui	quelques	oui	oui	moyen	non	élevée	moyen	oui

1. Dans cette colonne, on distingue les techniques qui révèlent les loci marqueurs individuellement de celles qui les révèlent en masse.

2. Marqueurs liés au stade de développement ou à l'organe étudié.

3. Surtout pour ce qui est des étapes nécessaires au développement des microsattellites.

1. Se référer à l'ouvrage : *Les marqueurs moléculaires en génétique et biotechnologies végétales*, 1998, D. de VIENNE, INRA Editions, Paris, 204 p.

			DESCRIPTION	REFERENCE
AFLP	<i>Amplification Fragment Length Polymorphism</i>	Polymorphisme de longueur des fragments d'amplification	Amplification PCR d'ADN génomique après digestion par deux enzymes de restriction et ligation d'un adaptateur d'environ 20 paires de bases. Les amorces sont constituées de l'adaptateur et de trois bases aléatoires en 3'.	Vos <i>et al.</i> (1995) <i>Nucl Acids Res</i> 23 : 4407-4414 (voir Fiche 52)
AP-PCR	<i>Arbitrary Primed PCR</i>	PCR avec des amorces arbitraires	Amplification par PCR de séquences inconnues en utilisant une seule amorce de 20 bases arbitraires, avec des températures d'hybridation faibles	Welsh, McClelland (1990) <i>Nucl Acids Res</i> 19 : 861-866
CAPS	<i>Cleaved Amplified Polymorphic Sequence</i>	Polymorphisme de séquence obtenu après digestion de produit d'amplification	Un fragment d'ADN génomique amplifié (STS ou SCAR) est digéré par des enzymes de restriction, puis le polymorphisme de longueur des fragments générés est révélé par électrophorèse.	Konieczny, Ausubel (1993) <i>Plant J</i> 4 : 403-410
DAF	<i>DNA Amplification Fingerprinting</i>	Empreintes génétiques de produits d'amplification	RAPD en utilisant des amorces plus courtes (5 à 10 bases)	Caetano-Anollès <i>et al.</i> (1991) <i>Bio/technology</i> 9 : 553-557
DGGE	<i>Denaturing Gradient Gel Electrophoresis</i>	Electrophorèse en gel à gradient de dénaturation	Electrophorèse dans un gel d'acrylamide contenant un gradient linéaire d'urée/formamide. Le fragment d'ADN, initialement double brin, est partiellement dénaturé pour former une structure "branchée" (en "Y") dont la mobilité est quasi nulle. Des différences de séquence dans le domaine de fusion le moins stable entraînent des différences de stabilité, donc de niveau de migration.	Myers <i>et al.</i> (1987) <i>Methods Enzymol</i> 155 : 501-527 (voir Fiche 55)
IMA	<i>Inter Microsatellite Amplification</i>	Amplification inter-microsatellite	Amplification PCR sur l'ADN génomique avec des amorces qui portent des motifs répétés dinucléotidiques et quelques bases arbitraires	Zietkiewicz <i>et al.</i> (1994) <i>Genomics</i> 20 : 176-183 (voir Fiche 57)
MAAP	<i>Multiple Arbitrary Amplicon Profiling</i>	Profils d'amplifications arbitraires multiples	Terme générique qui désigne toutes les techniques d'amplification PCR sur ADN génomique utilisant des amorces arbitraires et générant des profils électrophorétiques complexes et polymorphes (RAPD, AP-PCR, DAF, AFLP, IMA...)	Caetano-Anollès (1994) <i>Plant Mol Biol</i> 25 : 1011-1026
tecMAAP	<i>Template Endonuclease Cleavage MAAP</i>	MAAP après digestion de l'ADN cible	Amplification PCR de type MAAP sur ADN génomique préalablement digéré par une enzyme de restriction. La technique AFLP entre dans cette catégorie.	Caetano-Anollès <i>et al.</i> (1993) <i>Mol Gen Genet</i> 241 : 57-64
Microsatellite ou SSR	<i>Simple Sequence Repeat</i>	Répétition de séquences simples	Répétitions en tandem de motifs mono-, di-, tri- ou tétranucléotidiques à différents loci. De tels motifs sont très abondants et très polymorphes dans le génome des eucaryotes.	Tautz (1989) <i>Nucleic Acids Res</i> 17 : 6463-6471 (voir Fiche 57)
RAPD	<i>Random Amplified Polymorphic DNA</i>	Polymorphisme de l'ADN amplifié au hasard	Amplification par PCR de fragments d'ADN en utilisant une amorce de 10 bases arbitraires	Williams <i>et al.</i> (1990) <i>Nucl Acids Res</i> 18 : 6531-6535 (voir Fiche 51)
RFLP	<i>Restriction Fragment Length Polymorphism</i>	Polymorphisme de longueur de fragments de restriction	Hydrolyse de l'ADN par une endonucléase de restriction puis polymorphisme révélé après hybridation d'une sonde marquée sur les fragments séparés par électrophorèse sur gel et transférés sur une membrane (technique de Southern)	Botstein <i>et al.</i> (1980) <i>Am J Human Genet</i> 32 : 314-331 (voir Fiche 50)
SCAR	<i>Sequence-Characterized Amplified region</i>	Caractérisation de produit d'amplification	Fragment d'ADN génomique amplifié par PCR avec des amorces spécifiques (14 à 20 bases). Ces amorces sont définies grâce à la connaissance de la séquence d'un fragment RAPD intéressant isolé d'un gel, cloné et séquencé	Paran, Michelmore (1993) <i>Theor Appl Genet</i> 85 : 985-993 (voir Fiche 51)
SSCP	<i>Single Strand Conformation Polymorphism</i>	Polymorphisme de conformation du simple brin d'ADN	Fragment d'ADN génomique amplifié par PCR avec des amorces spécifiques. L'ADN double brin est dénaturé (94 °C, 10 minutes), rapidement refroidi afin d'empêcher la réassociation des simples brins et chargé sur un gel d'acrylamide non dénaturant.	Orita <i>et al.</i> (1989) <i>Proc Natl Acad Sci USA</i> 86 : 2766-2770 (voir Fiche 54)
STS	<i>Sequence Tagged Site</i>	Séquence ciblée	Amplification PCR avec des amorces spécifiques d'une séquence	Olson <i>et al.</i> (1989) <i>Science</i> 254 : 1434-1435

CARTES GÉNÉTIQUE ET PHYSIQUE

Une carte génétique est une représentation simplifiée du génome qui consiste à jaloner l'ensemble des chromosomes de points de repère (appelés marqueurs) et à déterminer les distances qui les séparent. La métrique utilisée (le centimorgan, cM, ou unité de distance génétique) est calculée à partir de la fréquence de recombinaison entre loci (marqueurs génétiques) pris deux à deux. Dans le cas d'une carte physique, la molécule d'ADN est jalonnée de marqueurs moléculaires constitués par des séquences particulières et les distances sont exprimées en nombre de paires de bases (pb).

Établissement d'une carte génétique

Sur le plan pratique, l'établissement d'une carte génétique est fondée sur le phénomène du crossing-over à la méiose. La création d'une carte génétique nécessite de disposer : (i) d'une descendance issue d'un croisement obtenu par reproduction sexuée ; (ii) de la détermination du génotype des différents marqueurs pour les individus de cette descendance. L'utilisation d'outils d'analyse statistique et de méthodes de calculs mathématiques permet alors d'estimer les distances génétiques entre les marqueurs (p. ex. absence ou présence d'une bande RAPD) pour un même groupe de liaison, puis de les ordonner à l'intérieur de chaque groupe de liaison.

C'est dans le développement des cartes de liaison génétique que les marqueurs moléculaires ont trouvé une de leur plus retentissante utilisation. En effet, les cartes génétiques saturées permettent des études génétiques détaillées, une décomposition des caractères complexes en leurs composantes discrètes (QTL : *Quantitative Trait Loci*) et l'utilisation de la liaison entre marqueur et QTL pour la sélection assistée par marqueurs. Dès la fin des années 1980, les cartes RFLP (voir Fiche 50) d'espèces agronomiques majeures ont été établies. Durant ces dix dernières années, l'avance technologique dans la détection des marqueurs moléculaires (principalement l'avènement des marqueurs issus de la PCR), ainsi que le développement de l'analyse informatique des données issues du marquage, ont entraîné une véritable explosion dans le domaine de la cartographie génétique. Ainsi, les cartes existantes ont été complétées jusqu'à saturation (autant de groupes de liaison identifiés que de chromosomes). Mais surtout, ces cartes permettent d'étudier des espèces sur lesquelles aucune information génétique n'existait ou pour lesquelles aucun pedigree type ne permettait d'envisager la cartographie (espèces ayant une longue période de génération, comme les arbres forestiers).

Etablissement d'une carte physique

L'établissement d'une carte physique consiste à reconstruire un génome en un ensemble ordonné de fragments d'ADN comme ils se trouvent sur les chromosomes. Les vecteurs de clonage utilisés sont de type cosmide, BAC ou YAC (voir Fiche 8) puisqu'ils permettent l'intégration de grands fragments d'ADN. La cartographie physique est souvent une étape préalable indispensable au séquençage génomique. En pratique, la construction d'une carte physique passe par l'étiquetage de chaque clone par des marqueurs moléculaires (séquences d'ADN), si possible les mêmes que ceux utilisés pour l'obtention de la carte génétique. La difficulté consiste ensuite à ordonner les clones entre eux afin de retrouver leur agencement initial sur le chromosome. Les clones présentant des parties communes, révélées par des marqueurs communs, sont regroupés en "contigs", c'est-à-dire en ensembles de clones chevauchants. L'ensemble des "contigs" permet alors de reconstruire l'enchaînement des fragments d'ADN constituant les chromosomes.

Relation entre carte génétique et carte physique

L'intégration de la carte physique et de la carte génétique d'un même génome consiste à placer de nombreux points de repère communs sur les deux cartes : par exemple des marqueurs RFLP ou microsatellites localisés sur une carte génétique peuvent être facilement séquencés et repérés sur une carte physique. La difficulté d'intégration est liée au fait que les deux échelles utilisées (le cM d'une part et le nombre de pb d'autre part) ne sont pas colinéaires. Chez *Arabidopsis*, 1 cM correspond à 220 kb en moyenne, mais varie, suivant les régions du génome, de 30 à plus de 1 000 kb. Ainsi, même si l'ordre des marqueurs est respecté entre les deux types de carte, leurs distances relatives sont parfois très différentes. Ceci est dû au fait que la fréquence de recombinaison n'est pas constante sur tout le génome. Il existe des régions qui recombinent beaucoup (points chauds de recombinaison) alors que les distances physiques y sont petites. Inversement, il existe de très longues portions d'ADN (régions centromériques et télomériques) où le phénomène de recombinaison est réprimé par la présence d'hétérochromatine ou d'ADN hautement répété. Ces zones montrent par conséquent des distances génétiques plus faibles entre les marqueurs qui tendent à être regroupés en "cluster". En d'autres termes, dans une région du génome peu recombinée, 1 cM correspondra à un grand nombre de bases, alors que dans une région fortement recombinée, 1 cM correspondra à moins de bases.

Si le rapport distance physique/distance génétique varie au niveau intra-chromosomique, il varie plus généralement lorsque l'on compare des espèces à contenus d'ADN différents. Le tableau ci-dessous présente la relation entre distance physique et génétique chez quelques espèces végétales pour lesquelles des cartes génétiques saturées sont disponibles. Il montre que :

- la taille génétique d'un génome est peu variable (variant du simple au double) et se situe approximativement autour de 150 cM (= 1,5 Morgan) par chromosome. Il faut donc se représenter la méiose comme un processus qui fractionne peu le génome avec en moyenne 1,5 crossing-over par chromosome ;
- la faible variation des longueurs des cartes génétiques entre espèces contraste avec les considérables variations de la quantité d'ADN par cellule. Comme le nombre de gènes exprimés doit être du même ordre de grandeur au sein des végétaux supérieurs, ces différences de quantité d'ADN seraient essentiellement dues à des variations très importan-

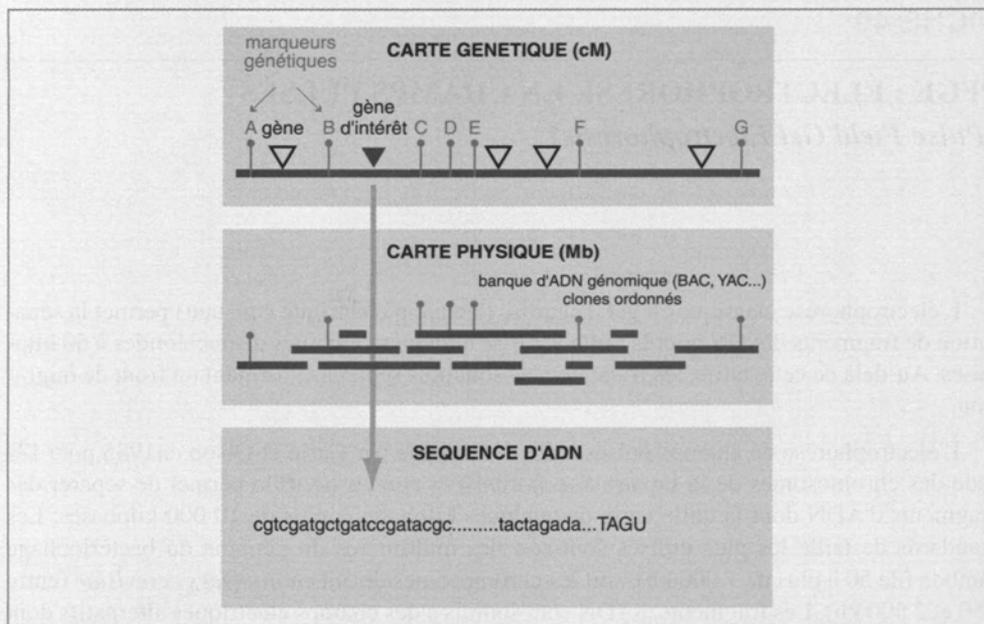
tes dans la quantité des ADN répétés non transcrits. Ces variations de la taille physique des génomes semblent indépendantes des tailles génétiques. Ainsi, il apparaît que la fréquence des crossing-over par unité de longueur physique décroît lorsque la taille du génome augmente.

Disposer à la fois de cartes génétique et physique d'un même génome permet :

- à partir de la position d'un locus intéressant sur la carte génétique, un accès facilité aux gènes correspondants (clonage positionnel) ;
- ou, à l'inverse, d'obtenir la position génétique d'une séquence d'ADN quelconque.

La cartographie physique nécessite des moyens importants et elle n'est envisagée pour l'instant que sur des espèces pilotes à "petit génome" (*Arabidopsis*, Riz...) chez lesquelles des cartes génétiques sont également disponibles. Cependant, grâce aux études de cartographie comparée, on peut espérer transférer une partie des connaissances obtenues chez des espèces modèles vers des espèces d'intérêt agronomique, par exemple d'*Arabidopsis* au Colza, du Riz au Blé...

Espèce	Taille physique (Mb)	Taille génétique (cM)	Nombre de chromosomes (n)	Taille génétique moyenne par chromosome (cM)	Mb / cM
<i>Phage T4</i>	0,16	800	/	/	0,0002
<i>Escherichia coli</i>	4,6	1750	/	/	0,0026
<i>Saccharomyces cerevisiae</i>	14	4200	16	262,5	0,0033
<i>Caenorhabditis elegans</i>	100	320	11/12	/	0,312
<i>Arabidopsis thaliana</i>	150	675	5	135	0,22
<i>Drosophila melanogaster</i>	170	280 F	4	70	0,607
<i>Prunus persica</i>	300	712	8	90	0,42
<i>Oryza sativa</i>	450	1 490	12	125	0,30
<i>Eucalyptus grandis</i>	600	1 370	11	125	0,43
<i>Brassica rapa</i>	650	1 850	10	185	0,35
<i>Quercus robur</i>	900	1 200	12	100	0,75
<i>Lycopersicon esculentum</i>	980	1 280	12	107	0,76
<i>Solanum tuberosum</i>	1 540	1 120	12	93	1,37
<i>Zea mays</i>	2 500	1 860	10	186	1,34
<i>Lactuca sativa</i>	2 730	1 950	9	217	1,40
<i>Homo sapiens</i>	3 000	2 800 H 4 800 F	23	121,7 H 208,7 F	1,071 H 0,625 F
<i>Mus musculus</i>	3 000	1 700	20	100	1,764
<i>Triticum tauschii</i>	4 200	1 330	7	190	3,15
<i>Hordeum vulgare</i>	5 500	1 250	7	178	4,40
<i>Triticum aestivum</i>	16 000	3 500	21	167	4,57
<i>Pinus pinaster</i>	25 500	2 000	12	167	12,75



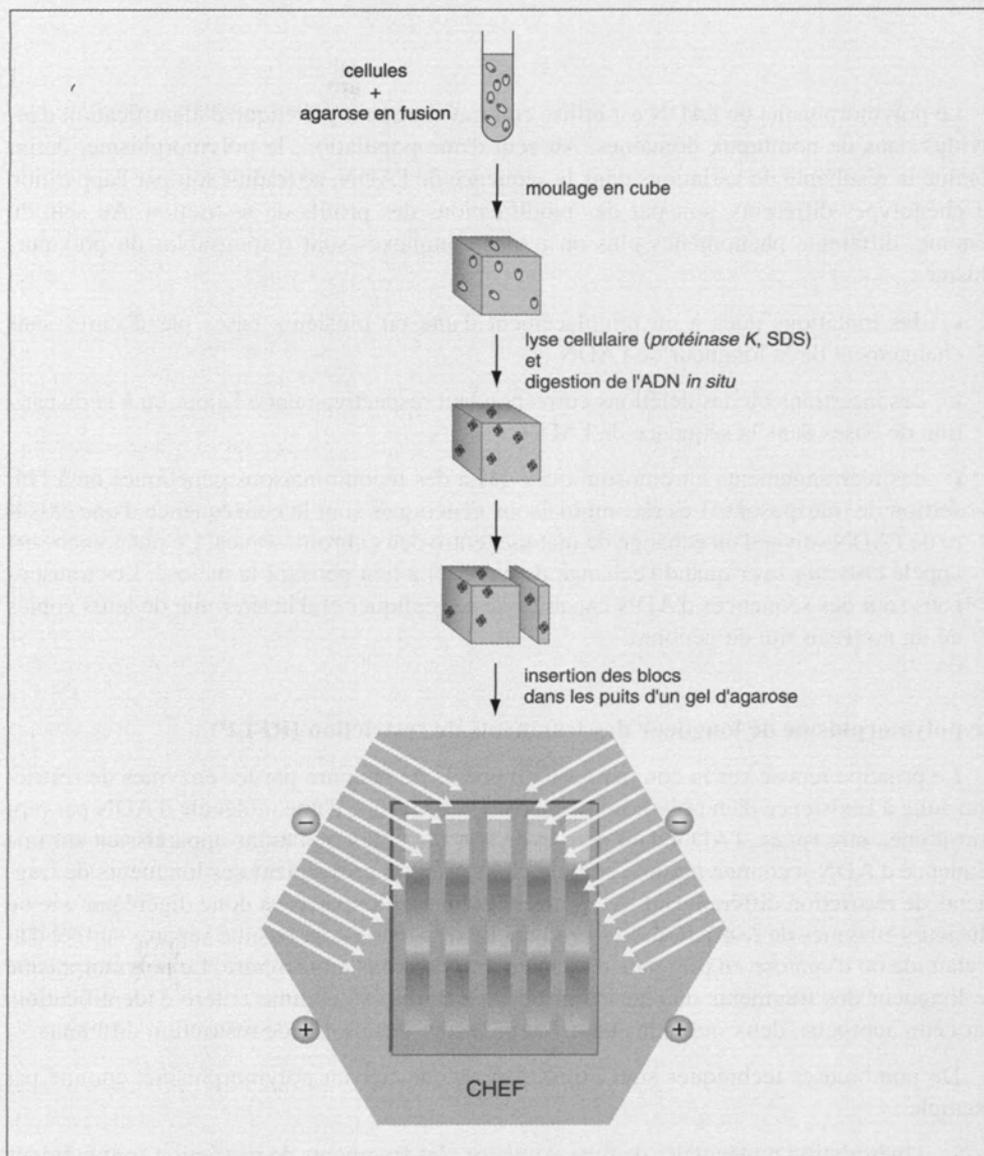
PFGE : ÉLECTROPHORÈSE EN CHAMPS PULSÉS (Pulse Field Gel Electrophoresis)

L'électrophorèse classique en gel d'agarose (à champ électrique constant) permet la séparation de fragments d'ADN dont la taille varie de quelques centaines de nucléotides à 50 kilobases. Au-delà de cette taille, les fragments ne sont plus séparés et forment un front de migration.

L'électrophorèse en champs pulsés a été développée par Carle et Olson en 1985 pour l'étude des chromosomes de la Levure *Saccharomyces cerevisiae*. Elle permet de séparer des fragments d'ADN dont la taille varie de quelques kilobases à plus de 10 000 kilobases. Les standards de taille les plus utilisés sont soit des multimères du génome du bactériophage lambda (de 50 à plus de 1 000 kb), soit les chromosomes de *Saccharomyces cerevisiae* (entre 250 et 2 500 kb). Les fragments d'ADN sont soumis à des champs électriques alternatifs dont l'orientation relative peut varier selon les systèmes de 90 à 180°. La configuration actuellement la plus utilisée est le CHEF (*Contour-clamped Homogeneous Electric Fields*) où l'angle entre les deux champs électriques est de 120°. Le principe de séparation est fondé sur la vitesse différentielle de réorientation des molécules au moment de l'alternance des deux champs électriques. Ainsi, les grandes molécules mettront plus de temps à s'orienter selon la direction du second champ électrique que celles de plus petite taille. L'accumulation du retard à chaque alternance aboutit à la séparation des molécules en fin de migration. Différents paramètres comme le pourcentage en agarose, la tension appliquée ou encore la température permettent d'optimiser la résolution. Cependant, le paramètre principal est le temps de pulsation, c'est-à-dire le temps d'application du champ électrique dans chaque orientation : il permet de favoriser la séparation des fragments dans une gamme de taille donnée.

L'électrophorèse en champs pulsés nécessite un procédé spécifique de purification de l'ADN. Ainsi, réaliser un caryotype (ensemble des chromosomes d'un organisme) ou un profil de macrorestriction d'un ADN génomique (fragments d'ADN générés par une endonucléase de restriction à faible fréquence de coupure et permettant la visualisation de l'ensemble du génome) nécessite la préparation d'un ADN de haut poids moléculaire. La préparation des échantillons interdit les purifications classiques par extraction phénol/chloroforme qui favorisent les cassures aléatoires de l'ADN. C'est pourquoi les cellules sont empaquetées dans une matrice d'agarose à bas point de fusion, puis lysées *in situ* (par le lysozyme pour les bactéries ou d'autres enzymes comme la lyticase pour les levures) afin de dégrader l'enveloppe cellulaire externe. Puis l'action conjointe d'un détergent et d'une protéinase libère le contenu cellulaire et dégrade les protéines cellulaires (dont les protéines liant l'ADN) qui vont diffuser à l'extérieur de la matrice d'agarose. L'ADN reste prisonnier de l'agarose. De la même façon, la coupure par les endonucléases de restriction (choisies en fonction de leur fréquence de coupure sur l'ADN étudié) sera réalisée à l'intérieur du bloc d'agarose protégeant l'ADN.

Les applications de l'électrophorèse en champs pulsés sont multiples : identification et suivi des micro-organismes (applications à l'épidémiologie, domaine agro-alimentaire...), cartographie de fragments des chromosomes humains, étude de la structure et de l'évolution des chromosomes bactériens. Dans ce dernier domaine, la réalisation de cartes physiques chromosomiques (plus de 100 cartes construites depuis l'avènement de la PFGE) a ouvert la voie au séquençage systématique des génomes.



RFLP : POLYMORPHISME DE LONGUEUR DES FRAGMENTS DE RESTRICTION

(*Restriction Fragment Length Polymorphism*)

Le polymorphisme de l'ADN est utilisé comme marqueur génétique d'identification d'individus dans de nombreux domaines. Au sein d'une population, le polymorphisme, défini comme la résultante de variations dans la séquence de l'ADN, se traduit soit par l'apparition de phénotypes différents, soit par des modifications des profils de restriction. Au sein du génome, différents phénomènes plus ou moins complexes sont responsables du polymorphisme :

- des mutations dues à un remplacement d'une ou plusieurs bases par d'autres sans changement de la longueur de l'ADN ;
- des insertions ou des délétions correspondant respectivement à l'ajout ou à la disparition de bases dans la séquence de l'ADN ;
- des réarrangements chromosomiques dûs à des recombinaisons génétiques ou à l'insertion de transposons. Les recombinaisons génétiques sont la conséquence d'une cassure de l'ADN suivie d'un échange de matériel entre deux chromosomes. Ce phénomène est appelé *crossing-over* quand l'échange de matériel a lieu pendant la méiose. Les transposons sont des séquences d'ADN capables de se répliquer et d'insérer une de leurs copies en un nouveau site du génome.

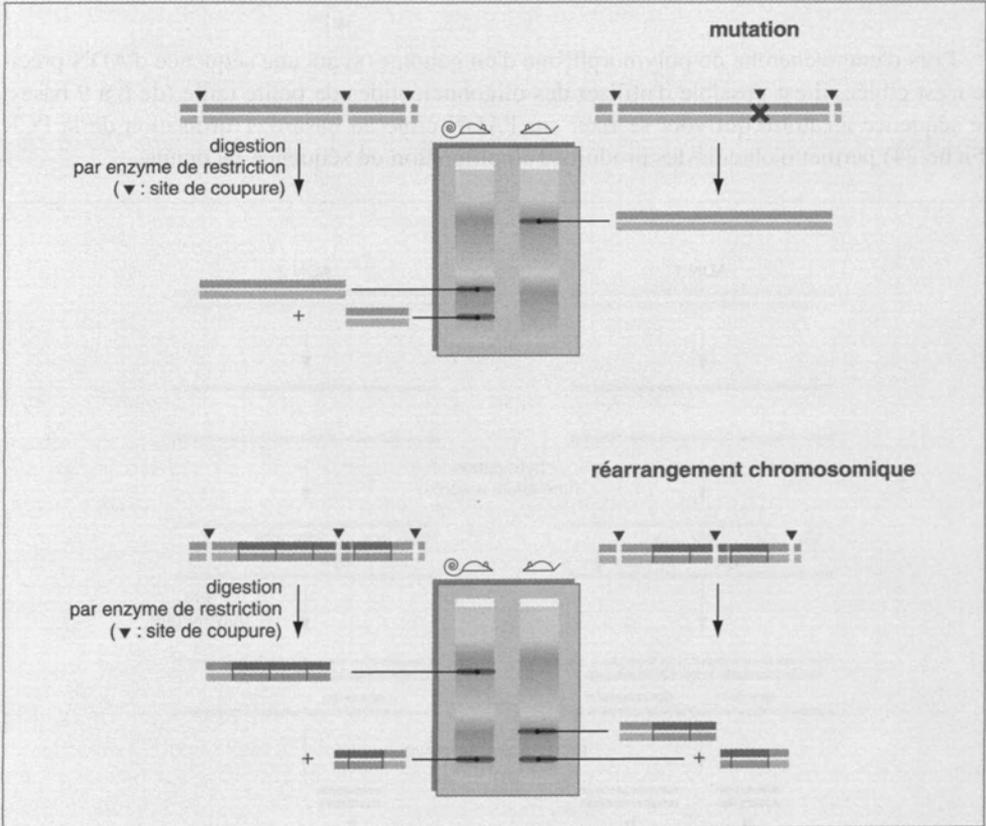
Le polymorphisme de longueur des fragments de restriction (RFLP)

Le principe repose sur la comparaison de profils de coupure par les enzymes de restriction suite à l'existence d'un polymorphisme dans la séquence d'une molécule d'ADN par rapport à une autre (p. ex. l'ADN d'un père et de son fils). Des mutations apparaissant sur une séquence d'ADN reconnue par une enzyme de restriction provoquent des longueurs de fragments de restriction différentes. L'ADN des individus à comparer est donc digéré par une ou plusieurs enzymes de restriction. Les produits de restriction sont ensuite séparés sur gel d'acrylamide ou d'agarose en présence d'un marqueur de poids moléculaire. Le polymorphisme de longueur des fragments de restriction observé est utilisé comme critère d'identification. Par cette approche, deux individus pourront présenter des profils de restriction différents.

De nombreuses techniques sont utilisées pour détecter un polymorphisme, comme par exemple :

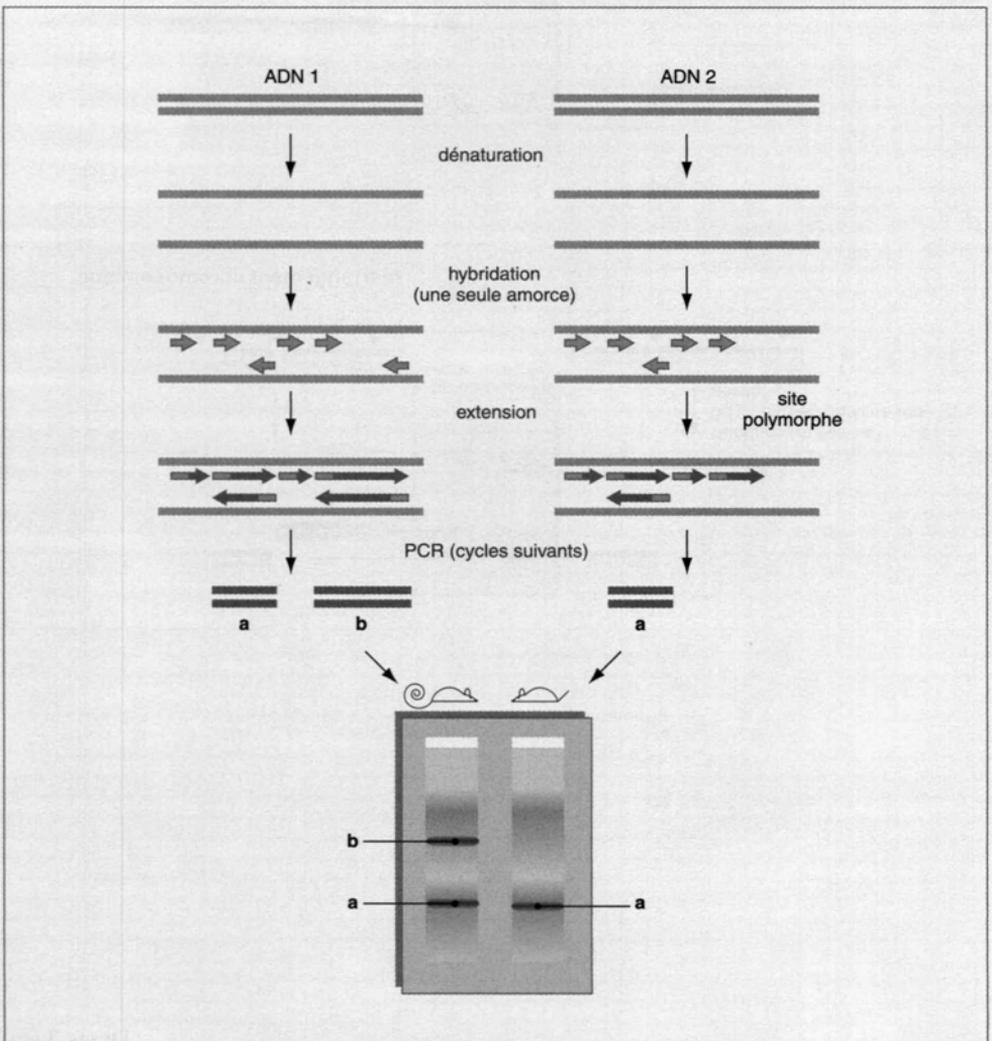
- l'hybridation moléculaire de type *Southern* : les fragments de restriction sont hybridés avec une sonde constituée d'un fragment marqué d'ADN complémentaire d'une séquence spécifique de l'ADN (voir Fiche 12) ;

- la réaction de PCR : une région spécifique de l'ADN est amplifiée, ce qui permet de visualiser les fragments de restriction de cette région, présents en grande quantité, après électrophorèse sur gel et coloration au bromure d'éthidium (voir Fiche 24) ;
- la technique RAPD : des fragments sont amplifiés au hasard et détectés après électrophorèse en gel d'agarose et coloration au bromure d'éthidium (voir Fiche 51) ;
- toute autre technique de PCR utilisant notamment les séquences répétées (transposons, microsatellites : voir Fiches 49 à 57).



RAPD : POLYMORPHISME D'ADN PAR AMPLIFICATION ALÉATOIRE
(*Random Amplified Polymorphic DNA*)

Lors d'une recherche de polymorphisme d'un génome, si aucune séquence d'ADN précise n'est ciblée, il est possible d'utiliser des oligonucléotides de petite taille (de 6 à 9 bases) de séquence aléatoire qui vont se fixer sur l'ADN cible au hasard. L'utilisation de la PCR (Fiche 24) permet d'obtenir des produits d'amplification de séquence inconnue.



Cette technique est appelée RAPD (*Random Amplified Polymorphic DNA*). A l'inverse de la PCR traditionnelle ciblée sur une séquence déterminée, une réaction de RAPD utilise un seul type d'amorce et non un couple de deux amorces différentes. Si deux amorces voisines sont en position correcte, dans la bonne orientation et pas trop éloignées l'une de l'autre (moins de 5 000 pb), la PCR génère un produit d'amplification. La RAPD permet d'obtenir rapidement des fragments variés d'ADN génomique qui sont ensuite analysés. Les conditions de préparation des échantillons et des cycles de PCR doivent être optimisées pour chaque application afin d'assurer la reproductibilité des résultats.

Les fragments d'amplification permettant de révéler un polymorphisme peuvent être sous-clonés (voir Fiche 9) afin d'être séquencés. A partir de cette séquence, il est possible de redéfinir un couple d'amorces qui aura une forte probabilité d'être spécifique de l'échantillon analysé. Il s'agit alors d'un SCAR (*Sequence Characterized Amplified Region*).

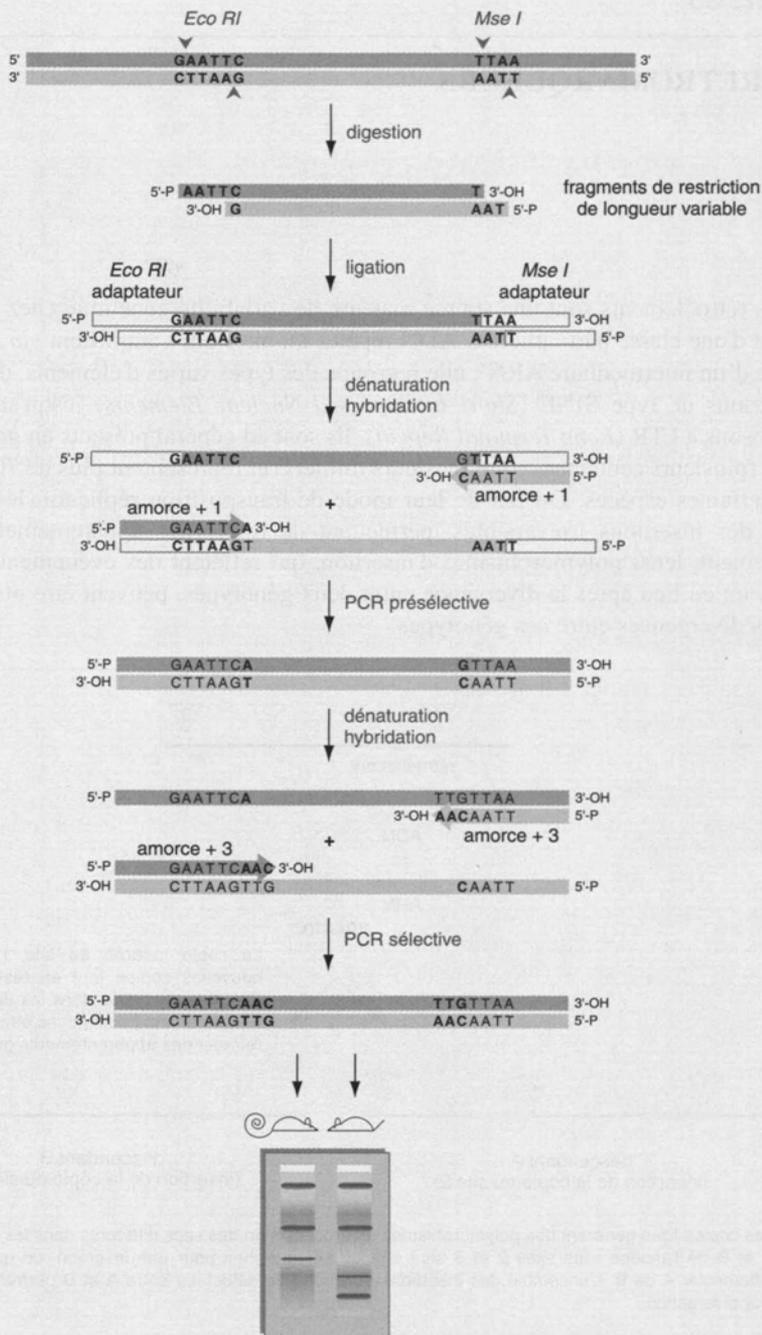
AFLP : POLYMORPHISME DE LONGUEUR DE FRAGMENTS AMPLIFIÉS

(Amplified Fragment Length Polymorphism)

La technique d'AFLP a été introduite en 1993. Son principe repose sur l'amplification sélective de fragments de restriction générés à partir d'un échantillon d'ADN génomique. Cette technique permet la recherche de polymorphisme de longueur de fragment de restriction au niveau de l'ADN. Elle est utilisée par exemple pour l'identification d'espèces, l'analyse de pedigree ou la recherche de marqueurs génétiques liés à un caractère. Elle est également utilisée pour l'identification de gènes exprimés de manière différentielle. Dans ce cas, l'analyse AFLP porte sur des ADN complémentaires (ADNc) synthétisés à partir d'ARN messagers (ARNm). Cette technique appelée AFLP-ADNc permet de visualiser de manière indirecte des sous-populations d'ARNm et de les comparer entre elles (voir Fiche 17).

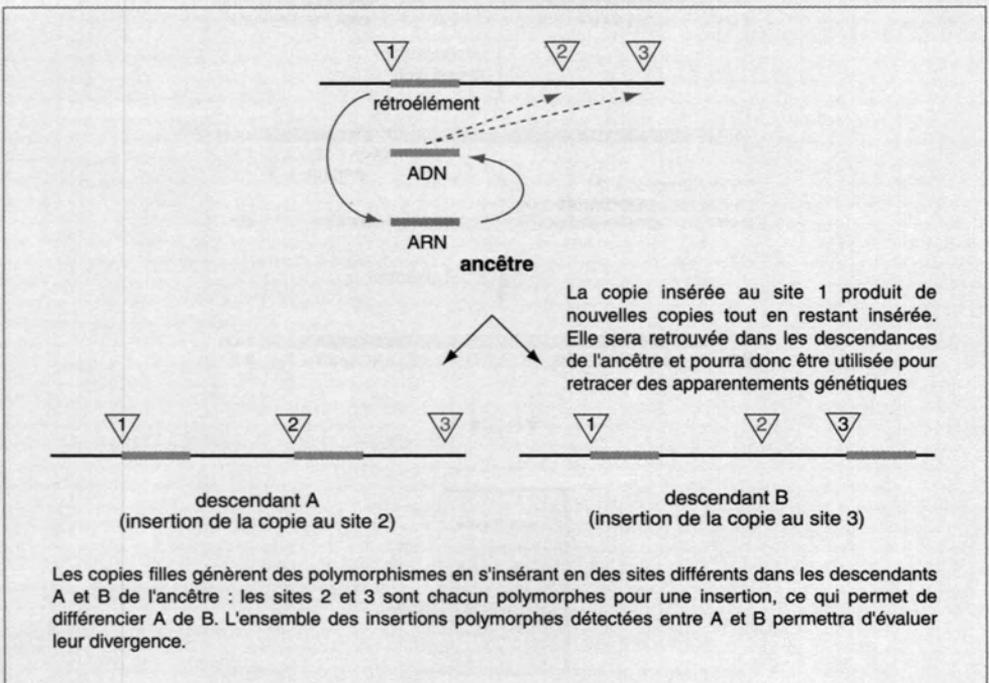
Polymorphisme de longueur de fragments amplifiés (AFLP)

L'ADN génomique (ou les ADNc) est digéré par deux enzymes de restriction coupant respectivement au niveau d'un site rare et d'un site fréquent. Puis des adaptateurs, spécifiques des sites de restriction utilisés, sont fixés aux extrémités des fragments de restriction obtenus. Les amplifications sont réalisées à l'aide d'amorces définies d'après la séquence des adaptateurs. Ces amorces portent sur leur extrémité 3'-OH des extensions aléatoires de 1 à 3 bases permettant l'amplification sélective d'une partie seulement de la population des fragments de restriction. Une pré-amplification est réalisée en premier à l'aide d'amorces présentant une extension d'une base généralement (amorces +1). Le produit de cette pré-amplification sert de matrice pour l'amplification sélective à l'aide des amorces +3. Pour un seul couple d'enzymes, l'utilisation d'amorces +3 porte à 4 096 (64 X 64) le nombre de combinaisons possibles. Les produits d'amplification sélective sont ensuite séparés sur gel de polyacrylamide en conditions dénaturantes. La détection des bandes d'AFLP est obtenue soit par coloration des gels à l'argent, soit par autoradiographie ou émission de fluorescence. Dans ce cas, une des amorces +3 d'AFLP (généralement celle correspondant au site de restriction rare) est marquée soit radioactivement (γ -³²P ou γ -³³P), soit par un chromophore fluorescent (dans le cas d'une analyse à l'aide d'un séquenceur automatique).



LES RÉTROMARQUEURS

Les rétroéléments sont une source majeure de variabilité génétique chez les eucaryotes. Il s'agit d'une classe particulière d'ADN répétés mobiles qui s'amplifient *via* la transcription inverse d'un intermédiaire ARN ; elle regroupe des types variés d'éléments, depuis les petits rétroposons de type SINE (*Short Interspersed Nuclear Elements*) jusqu'aux longs rétrotransposons à LTR (*Long Terminal Repeat*). Ils sont en général présents en grand nombre de copies (plusieurs centaines, voire plusieurs milliers) et représentent plus de 70 % du génome chez certaines espèces. Du fait de leur mode de transposition réplicatif, les rétroéléments créent des insertions irréversibles, permettant de tracer des apparentements génétiques. Inversement, leurs polymorphismes d'insertion, qui reflètent des événements de transposition ayant eu lieu après la divergence entre deux génotypes, peuvent être utilisés pour évaluer les divergences entre ces génotypes.



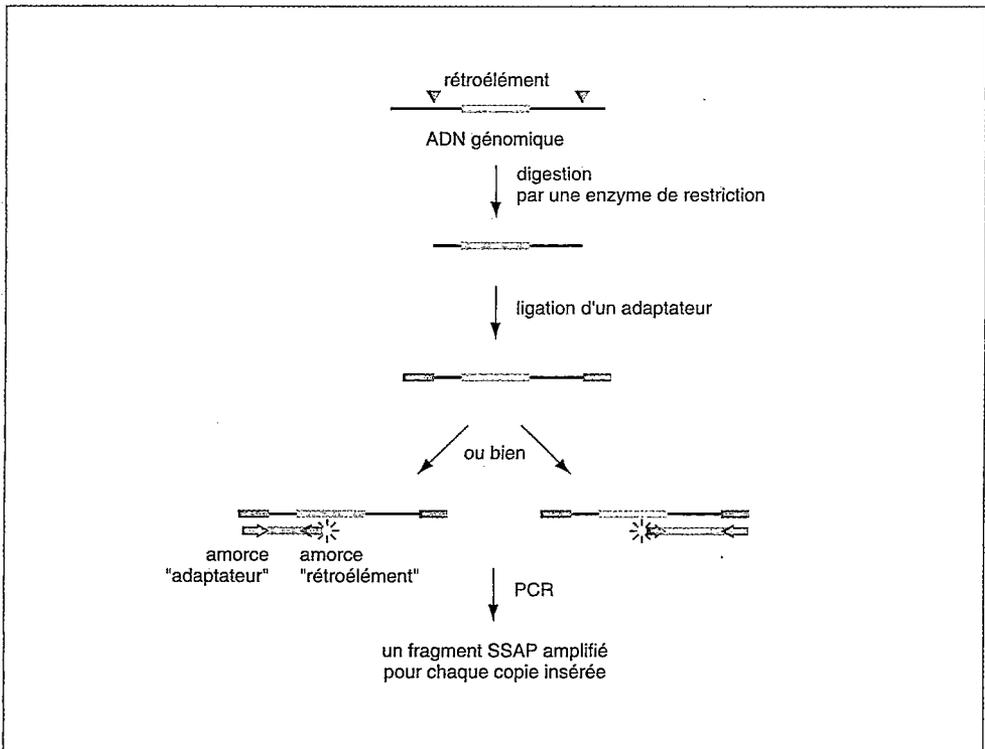
Selon le moment où les nouvelles insertions ont eu lieu, ces polymorphismes peuvent servir à identifier soit une lignée donnée, soit un groupe de lignées, voire une espèce ou un groupe d'espèces. En outre, les polymorphismes générés par l'insertion d'un rétroélément correspondent à des événements unidirectionnels (état ancien "vide", état nouveau "plein",

le retour vers l'état ancien par excision étant en théorie impossible), ce qui en fait de bons outils de systématique moléculaire, largement utilisés depuis longtemps dans les systèmes animaux et, plus récemment, chez les plantes.

Stratégies utilisables

Les stratégies "multiplex"

Elles sont fondées sur l'utilisation de la PCR et génèrent plusieurs bandes d'amplification par échantillon. Elles permettent de révéler l'ensemble (ou du moins une fraction significative) des sites d'insertion d'un élément donné. La plus répandue est la SSAP (*Sequence Specific Amplified Polymorphism*), également appelée *Transposon Display* (TD). La SSAP est une stratégie dérivée de l'AFLP (voir Fiche 52), mais utilisant, outre l'amorce "adaptateur", une amorce marquée (radioactivité, fluorescence...) correspondant à une séquence terminale du rétroélément, de façon à ne visualiser que les fragments d'ADN débutant dans le rétroélément. Le principe repose sur l'extraction d'ADN génomique suivie de sa digestion enzymatique par une enzyme de restriction. Un adaptateur est fixé aux extrémités des molécules d'ADN. Puis une PCR est réalisée en utilisant une amorce complémentaire de l'adaptateur et une amorce complémentaire de l'une des extrémités d'un rétroélément. Cela suppose donc qu'un rétroélément ait déjà été caractérisé et séquencé chez l'espèce étudiée. Les fragments d'ADN résultant de cette amplification sont analysés par exemple par électrophorèse sur des gels d'acrylamide à fort pouvoir résolutif afin de détecter un polymorphisme de taille des fragments d'amplification.

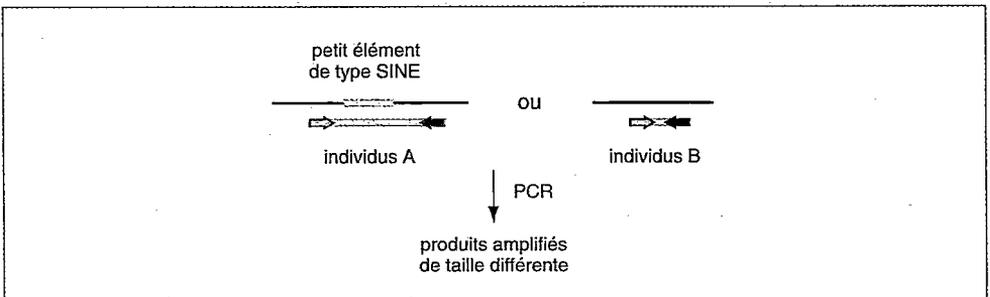


La SSAP permet de définir une carte de l'état de représentation du rétroélément dans le génotype étudié et de révéler les polymorphismes d'insertion à travers la présence/absence de fragments SSAP particuliers. Elle sert donc d'outil pour évaluer la biodiversité et établir des phylogénies ; elle s'est révélée plus informative que les marqueurs moléculaires classiques dans plusieurs cas.

D'autres stratégies "multiplex", telles que l'IRAP (*Inter-Retrotransposon Amplified Polymorphisms*) ou la REMAP (*Retrotransposon-Microsatellite Amplified Polymorphisms*) sont fondées sur des PCR directes entre rétroéléments ou entre rétroéléments et microsatellites respectivement. Elles sont plus simples à mettre en oeuvre que la SSAP, mais supposent que le nombre de copies du rétroélément soit particulièrement élevé (plusieurs dizaines de milliers de copies, ce qui est le cas du rétrotransposon Bare-1 de l'Orge pour lequel elles ont été développées).

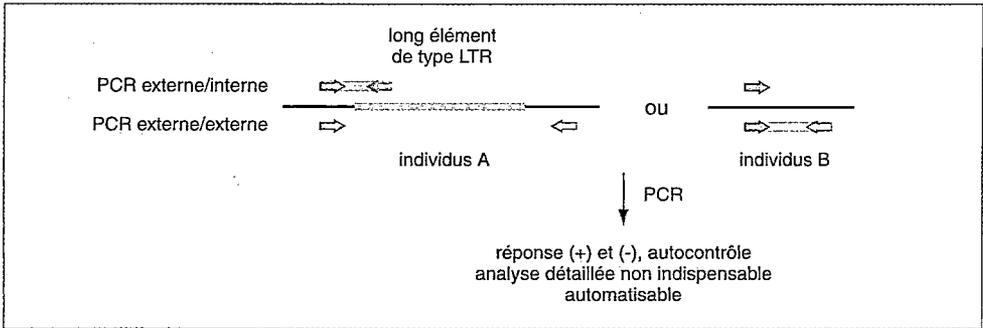
Les stratégies "site unique"

Elles sont fondées sur la détection rapide par PCR des deux états alternatifs "vide" ou "plein" d'un site donné. Elles ont d'abord été utilisées dans les systèmes animaux, puis plus récemment chez les plantes, avec de petits rétrotransposons comme les SINEs, l'état "vide" ou "plein" étant traduit par une différence dans la taille du produit PCR généré à partir de deux amorces flanquant le site d'insertion. La mise en oeuvre des stratégies "site unique" demande la connaissance préalable des régions génomiques flanquant les sites d'insertion polymorphes. Une partie des données peut être acquise par simple séquençage des bandes SSAP polymorphes extraites d'un gel, puisqu'elles contiennent les séquences flanquant un des côtés de l'élément. Les séquences flanquant l'autre côté peuvent ensuite être obtenues par diverses stratégies, comme par exemple une SSAP ancrée dans la région flanquante connue, et réalisée sur un génotype où le site est "vide".



Une autre stratégie fait appel à une simple PCR entre une amorce externe flanquante et une amorce interne du rétrotransposon ("PCR externe/interne"). En théorie, cette stratégie ne nécessite pas de connaître la taille des fragments amplifiés, et se base uniquement sur le fait qu'un produit d'amplification est présent (site "plein") ou non (site "vide"). Elle peut être mise en oeuvre par toute technique simple détectant la présence d'ADN (amorces fluorescentes, *dot-blots* avec sonde du site d'insertion), et donc être automatisée pour l'analyse d'un grand nombre d'échantillons. Une amélioration de cette stratégie a été apportée par la mise au point de la RBIP (*Retrotransposon-Based Insertion Polymorphisms*). La RBIP est fondée sur le fait que les rétrotransposons à LTR sont longs de plusieurs kilobases et ne permettent donc pas l'amplification PCR entre deux amorces flanquantes lorsque le site est "plein". La RBIP combine la PCR "externe/interne" citée plus haut à une PCR complémentaire "exter-

ne/externe” entre deux amorces flanquantes. Si le site est “plein”, seule la PCR “externe/interne” donnera un produit d’amplification, alors que si le site est “vide”, seul le couple d’amorces externes donnera un produit d’amplification. Chaque PCR est donc un contrôle de l’autre, ce qui permet de repérer les situations où l’insertion n’est pas fixée dans une population et surtout d’éviter les faux positifs ou les faux négatifs.



Une fois ces préalables réalisés, et les collections de jeux d’amorces obtenues, la RBIP représente un outil rapide et puissant, potentiellement utilisable pour la structuration de collections de ressources génétiques ou à des fins de protection variétale ou de contrôle qualité.

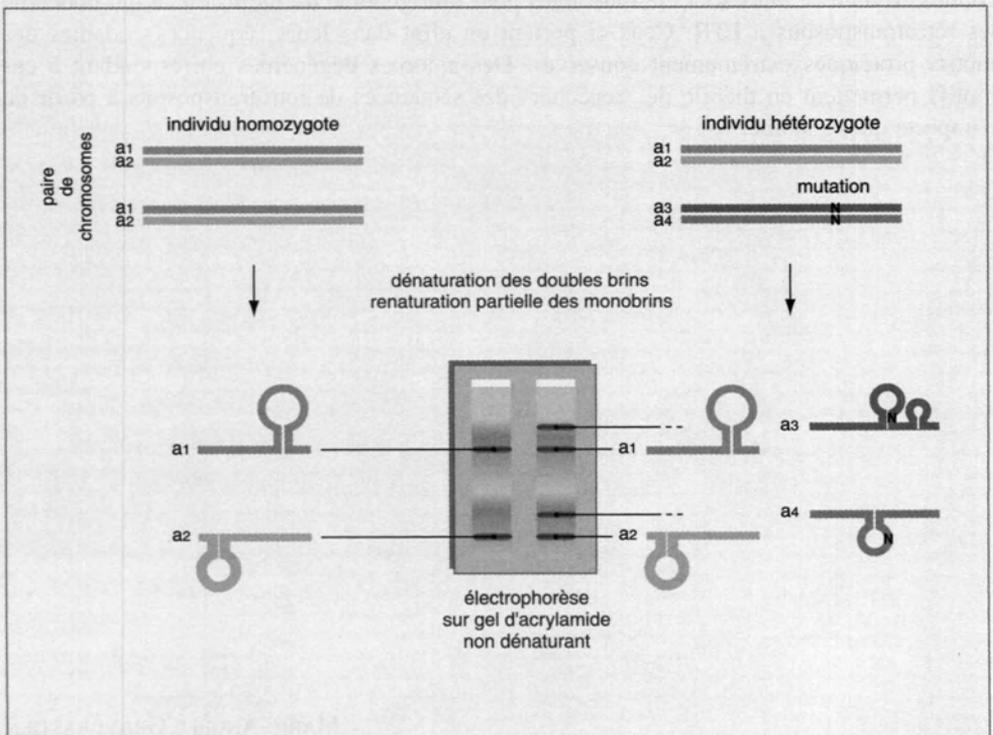
Les rétromarqueurs sont applicables à toutes les espèces

Les stratégies rétromarqueurs ont d’abord été mises au point avec des éléments déjà connus. Toutefois, elles sont envisageables pour tout génome du moins en ce qui concerne les rétrotransposons à LTR. Ceux-ci portent en effet dans leurs séquences codantes des motifs protéiques extrêmement conservés. Des amorces dégénérées correspondant à ces motifs permettent en théorie de “repêcher” des séquences de rétrotransposons à partir de n’importe quel génome.

SSCP : POLYMORPHISME DE CONFORMATION DE L'ADN MONOCATÉNAIRE

(Single Strand Conformation Polymorphism)

La technique SSCP est fondée sur le comportement électrophorétique d'un fragment d'ADN monocaténaire dans un gel d'acrylamide non dénaturant. En effet, une molécule d'acide nucléique monocaténaire peut former des structures secondaires dues à des appariements de bases au sein de la molécule. Ces structures secondaires dépendent de la séquence propre au brin d'ADN et donnent une conformation particulière à chaque type de molécule monocaténaire. Ainsi, deux séquences d'ADN très proches peuvent se différencier sur la base de la conformation de leur forme monocaténaire : deux allèles d'un même gène seront distingués. Cela permet par exemple de détecter un allèle mutant (éventuellement responsable d'une maladie génétique) chez un individu. Cette technique simple comporte deux inconvénients majeurs : le comportement électrophorétique des fragments monocaténaires est imprévisible car il est très dépendant de la température et des conditions d'électrophorèse, et, pour des fragments assez grands (> 200 pb), toutes les mutations ne semblent pas détectables (la méthode permet de détecter essentiellement les variations de séquence de type micro-insertion/micro-délétion).

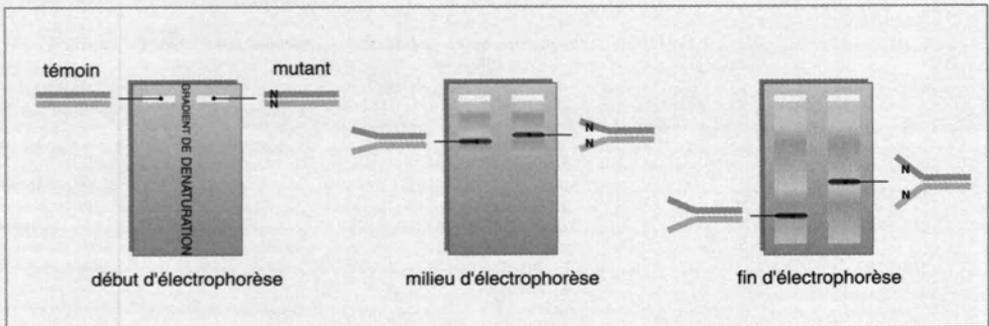


En pratique, le produit d'amplification PCR bicaténaire correspondant à la séquence d'intérêt est dénaturé par chauffage à 94 °C, puis rapidement refroidi dans de la glace : les molécules monocaténares n'ont pas le temps de se réassocier entre elles, mais forment une structure secondaire stable par des réassociations au niveau de zones de séquences complémentaires. Les fragments d'amplification réassociés sont alors séparés par électrophorèse. Chez un individu homozygote, on observe en général deux bandes, car deux molécules d'ADN monocaténares complémentaires ont des conformations secondaires légèrement différentes. Chez un individu hétérozygote, on observe en général quatre bandes. Dans quelques cas, des bandes surnuméraires peuvent apparaître en raison de l'existence de plusieurs conformations semi-stables pour un même brin.

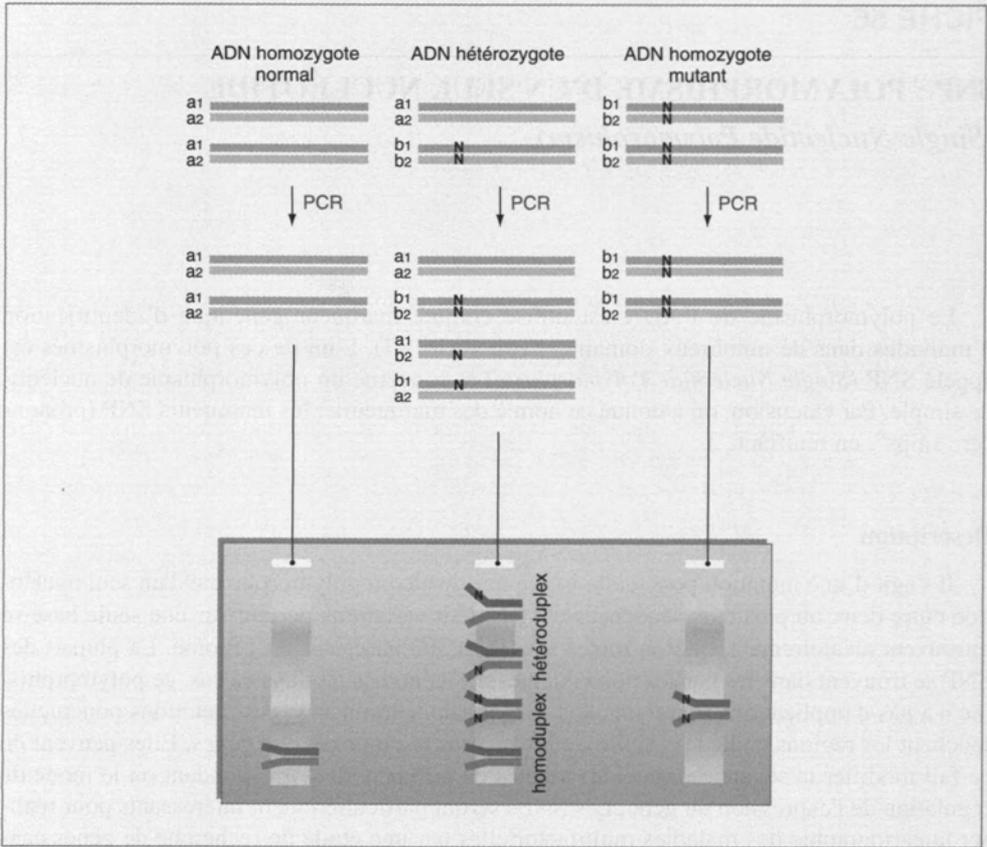
DGGE : ÉLECTROPHORÈSE DE L'ADN EN GRADIENT DE GEL DÉNATURANT

(*Denaturing Gel Gradient Electrophoresis*)

La technique DGGE permet la détection de polymorphismes ou de mutations de l'ADN très fins, même lorsqu'une seule paire de bases est substituée dans un fragment d'ADN. Elle est applicable à des fragments de plusieurs centaines de paires de bases et ne nécessite pas une connaissance préalable du site polymorphe. Cette méthode est fondée également sur les propriétés de renaturation des différents brins d'ADN. Cependant, à l'inverse de la SSCP (voir Fiche 54), la dénaturation de l'ADN ne se réalise pas avant électrophorèse, mais au cours de l'électrophorèse. Pour cela, les produits d'amplification sont déposés sur le gel qui contient soit une concentration croissante d'un agent chimique dénaturant l'ADN (urée-formamide dans le cas de la DGGE), soit une température de plus en plus élevée au cours de l'électrophorèse¹ (TGGE pour *Temperature Gel Gradient Electrophoresis*). Au cours de l'électrophorèse, l'ADN migre d'abord à l'état bicaténaire, puis rencontre des conditions qui dénaturent le domaine de fusion le moins stable de la molécule, formant une structure partiellement monocaténaire, dite branchée ou ramifiée. Cette dénaturation partielle entraîne une réduction de sa mobilité électrophorétique (le fragment ne migre quasiment plus), de sorte que sa position finale dans le gel dépend exclusivement de la température de fusion (T_m) du domaine le moins stable et par là-même de la séquence nucléotidique de ce dernier. On estime qu'environ 95 % des variations de séquence, même ponctuelles, du domaine de fusion le moins stable peuvent être détectées par des différences de migration électrophorétiques.



Par ailleurs, les individus hétérozygotes et homozygotes peuvent être distingués. Il suffit pour cela de faire une étape de dénaturation/renaturation, après le dernier cycle de la PCR : des homoduplex (deux brins d'ADN parfaitement complémentaires) et des hétéroduplex (deux brins d'ADN partiellement complémentaires) vont alors se former par réassociation des matrices alléliques. La présence de mésappariement dans les hétéroduplex va donner naissance à des bandes fortement retenues sur le gel (dénaturation rapide de l'hybride) et migrant très peu.



1. le gradient d'agents dénaturants est convertible en gradient de température : $T (^{\circ}\text{C}) = (\% \text{ dénaturant} + 182,4) / 3,2$, où le 100 % dénaturant correspond à une concentration de 40 % en formamide et 7M en urée.

SNP : POLYMORPHISME D'UN SEUL NUCLÉOTIDE

(*Single Nucleotide Polymorphism*)

Le polymorphisme de l'ADN est utilisé comme marqueur génétique d'identification d'individus dans de nombreux domaines (voir Fiche 47). L'un de ces polymorphismes est appelé SNP (*Single Nucleotide Polymorphism*) et concerne un polymorphisme de nucléotide simple. Par extension, on a donné ce nom à des marqueurs : les marqueurs SNP (prononcer "snips", en reniflant...).

Description

Il s'agit d'une mutation ponctuelle isolée qui révèle un polymorphisme d'un seul nucléotide entre deux ou plusieurs séquences d'ADN. Ces variations portant sur une seule base se retrouvent aléatoirement environ toutes les 100 à 300 bases sur un génome. La plupart des SNP se trouvent dans les parties non codantes des génomes et, dans ce cas, ce polymorphisme n'a pas d'implications fonctionnelles. Cependant, certaines de ces mutations ponctuelles touchent les régions codantes (SNPc) et les régions régulatrices des gènes. Elles peuvent de ce fait modifier la séquence en acides aminés du polypeptide correspondant ou le mode de régulation de l'expression du gène. Les SNPc seront particulièrement intéressants pour réaliser la cartographie des maladies multifactorielles par une étude de recherche de gènes candidats impliqués dans ces maladies.

Détection

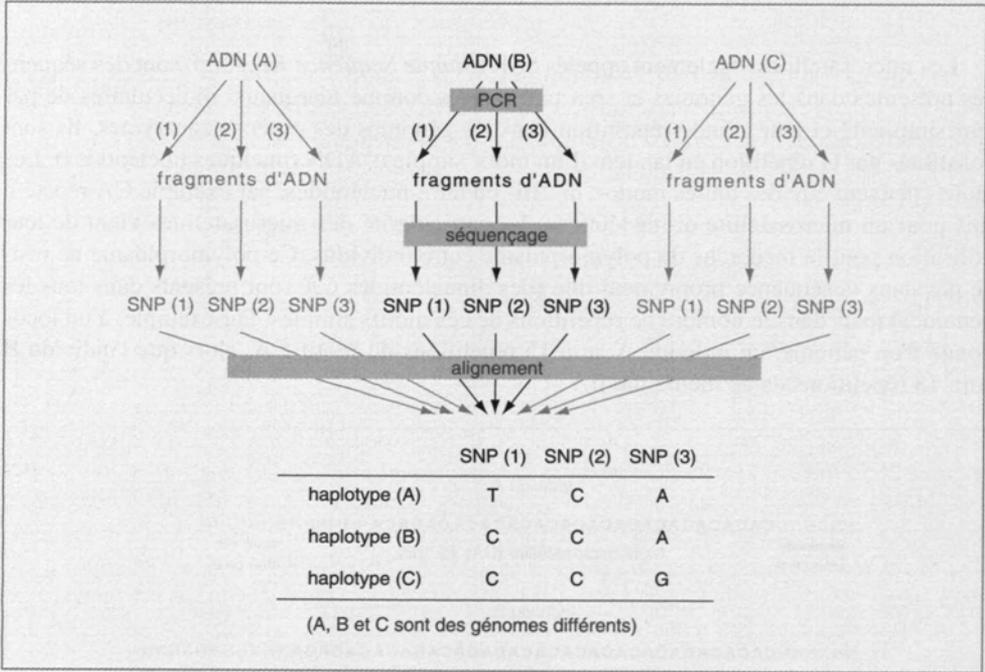
La détection des SNP nécessite tout d'abord une amplification PCR sur différents génotypes. Ces PCR peuvent être ciblées sur des séquences connues et déterminées à l'avance, mais elles peuvent être également aléatoires (AFLP). Dans un second temps, les différents produits d'amplification sont séquencés. Cette technique de SNP nécessitant le séquençage de produits de PCR est coûteuse. Après séquençage et vérification, les différentes séquences obtenues sont alignées. Les sites de polymorphisme de séquences sont alors identifiés, en particulier les polymorphismes de type substitution. Ainsi, les différents haplotypes (à partir des différentes combinaisons de SNP) peuvent être identifiés.

Intérêt

Les SNP, comme tout marqueur génétique, peuvent être exploités de différentes manières. Leur extrême précision permet des analyses très fines dans des domaines comme la recherche de parenté ou de variabilité liée à des maladies. La puissance des SNP est due essentiellement à leur présence sur l'ensemble du génome tant dans les introns que dans les exons. De plus, chaque individu d'une espèce posséderait plusieurs milliers de SNP qui sont

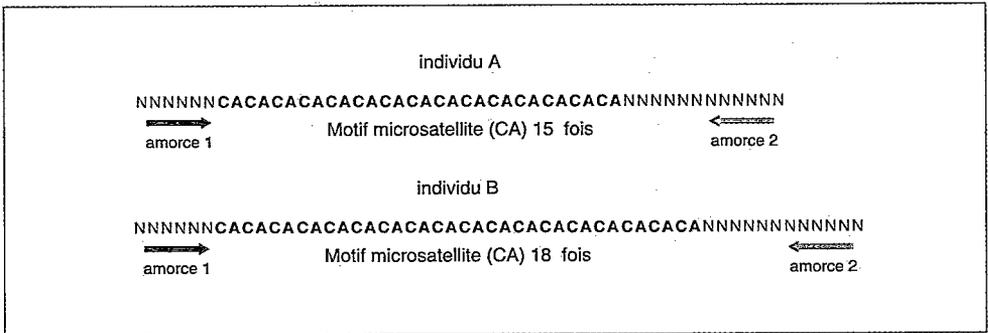
autant de variations ponctuelles dont la combinaison fait de chaque individu une "personne" unique. Le repérage de SNP peut ainsi permettre la découverte de gènes de prédisposition à de nombreuses maladies (telles que le diabète, l'hypertension artérielle ou la polyarthrite rhumatoïde chez l'Homme).

Cependant, l'utilisation des SNP pose le problème du coût et des débits d'analyses. Certaines nouvelles méthodes comme la spectrométrie de masse ou la fluorescence permettent d'envisager un génotypage à haut débit.



SSR : MICROSATELLITES, RÉPÉTITION DE SÉQUENCES SIMPLES
(Simple Sequence Repeats)

Les microsatellites, également appelés SSR (*Simple Sequence Repeats*), sont des séquences présentes dans les génomes et sont très utilisés comme marqueurs moléculaires de par leur simplicité et leur grande répartition dans les génomes des espèces eucaryotes. Ils sont constitués par la répétition en tandem d'un motif simple d'ADN (quelques nucléotides). Les motifs peuvent être des unités mono-, di-, tri- ou tétra-nucléotides, par exemple CA répété n fois pour un microsatellite di-nucléotides. La particularité des microsatellites vient de leur utilisation pour la recherche de polymorphisme entre individus. Ce polymorphisme ne réside pas dans la séquence proprement dite (des dinucléotides CA sont présents dans tous les génomes) mais dans le nombre de répétitions de ces motifs simples. Par exemple, à un locus donné d'un génome, un individu A aura 15 répétitions du motif CA, alors que l'individu B aura 18 répétitions de ce même motif.



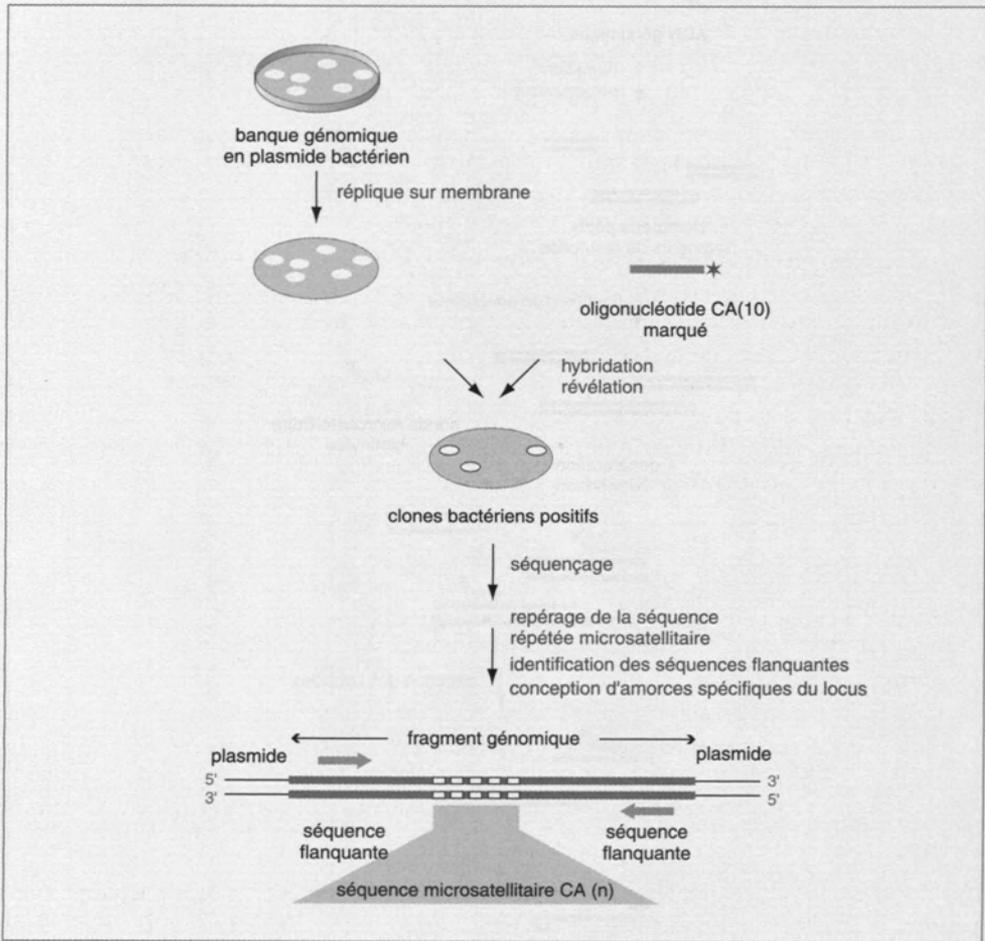
Afin de déterminer ces nombres de répétitions à un locus donné, il faut spécifiquement amplifier par PCR ce motif (à partir de l'ADN génomique des individus A et B). Il est donc nécessaire de connaître les séquences flanquantes de cette répétition afin de concevoir des amorces entourant le motif microsatellitaire. De plus, les fragments d'amplification générés ne diffèrent en taille que de quelques bases, ce qui nécessite une séparation par électrophorèse sur gel très résolutif. Le locus microsatellite est donc défini par un couple d'amorces PCR qui amplifie une région unique du génome contenant un motif répété.

Utilisation

L'utilisation des microsatellites comme marqueurs moléculaires repose sur trois étapes essentielles : isolement et caractérisation des microsatellites, conception d'amorces spécifiques et recherche de polymorphisme de taille entre individus.

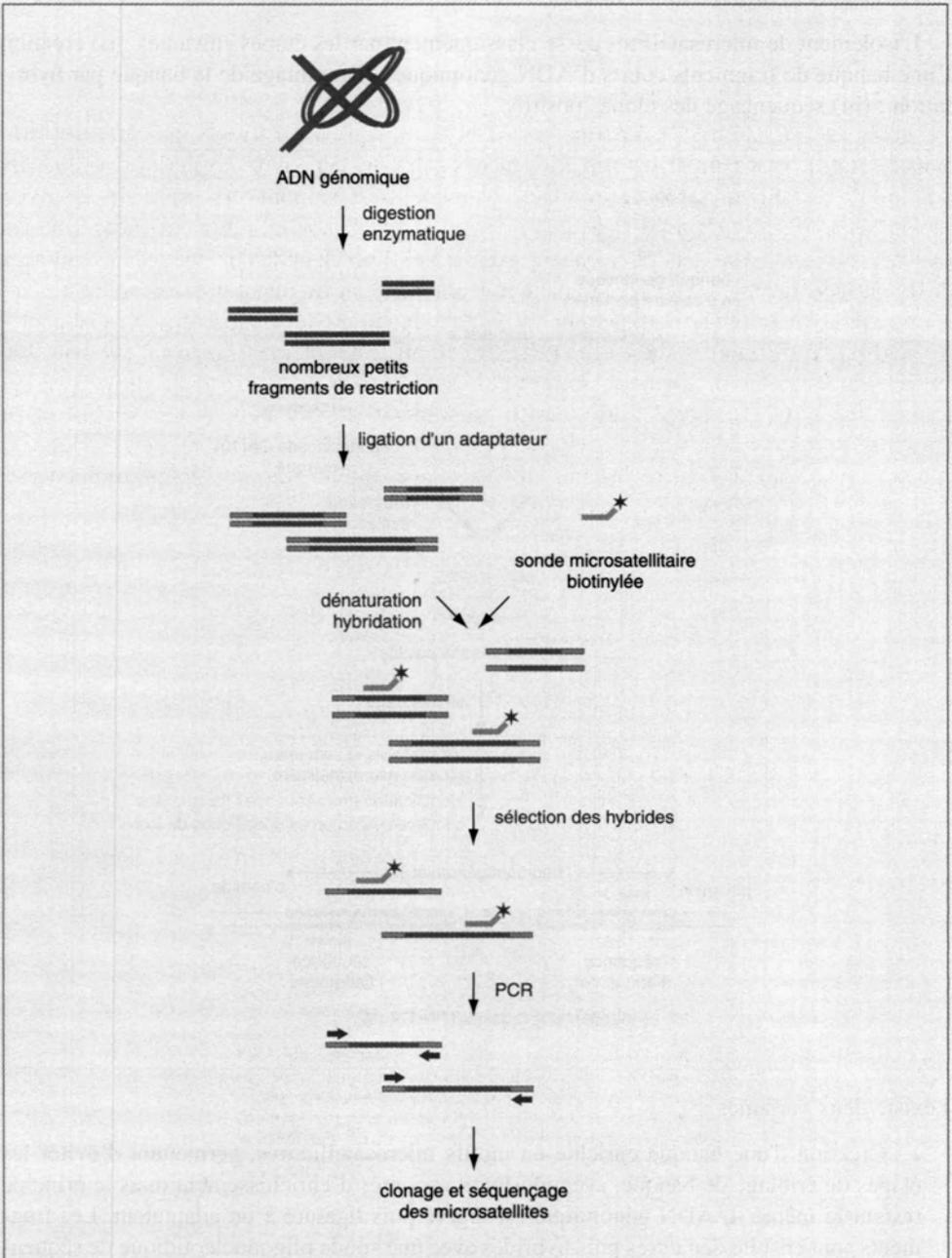
Isolement et caractérisation des microsatellites

L'isolement de microsatellites passe classiquement par les étapes suivantes : (i) création d'une banque de fragments courts d'ADN génomique ; (ii) criblage de la banque par hybridation ; (iii) séquençage des clones positifs.



Il existe deux variantes :

- Création d'une banque enrichie en motifs microsatellitaires, permettant d'éviter les étapes de criblage de banque, avec plusieurs variantes d'enrichissement mais le principe restant le même. L'ADN génomique est digéré puis ligaturé à un adaptateur. Les fragments sont ensuite dénaturés puis hybridés avec une sonde oligonucléotidique de séquence complémentaire au motif microsatellite recherché. Cette sonde est soit biotinylée, soit fixée à un support (colonne). Ainsi, les fragments d'ADN génomique s'hybridant à ces sondes microsatellitaires sont récupérés. Ces complexes purifiés sont ensuite déshybridés puis amplifiés par PCR à l'aide d'amorces complémentaires de l'adaptateur lié aux extrémités des fragments d'ADN génomique.



- Recherche de microsatellites dans les bases de données d'ADN génomique ou d'EST.

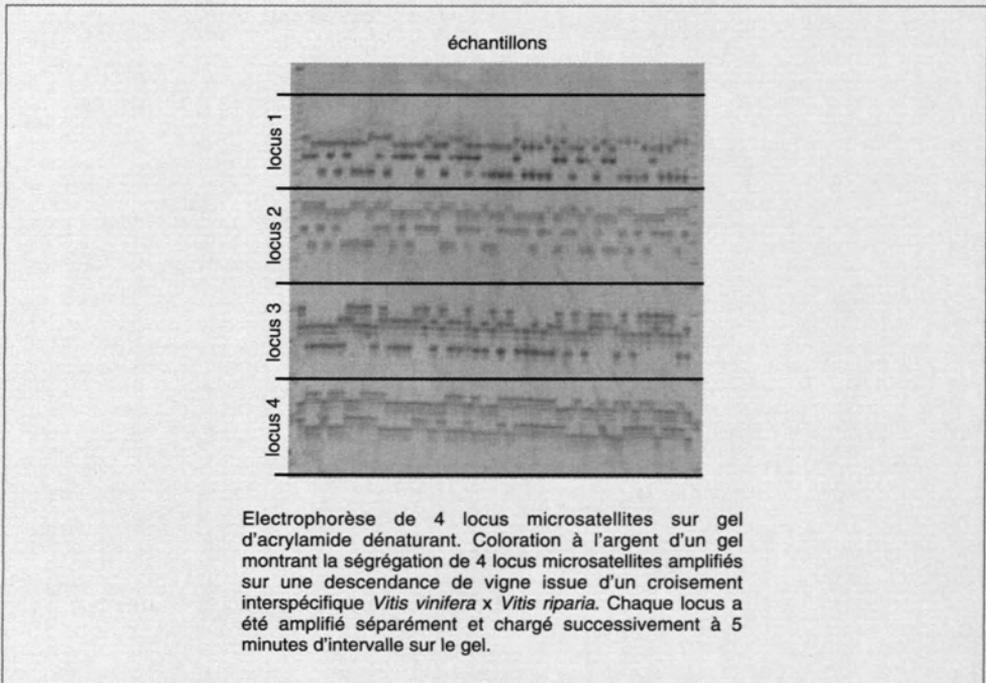
Conception des amorces

Elle peut se dérouler de deux manières différentes : (i) définition des amorces PCR directement sur la séquence clonée obtenue ; (ii) utilisation des amorces développées sur des espèces voisines. Puis il est nécessaire de réaliser des essais d'amplification pour vérifier

quelques caractéristiques du microsatellite (présence d'allèles nuls, niveau de polymorphisme...).

Révélation du polymorphisme

Après amplification PCR à partir de l'ADN génomique de différents individus, les produits d'amplification sont séparés en utilisant des gels dont le pouvoir résolutif est adapté aux différences de taille que l'on désire mettre en évidence. Les motifs tri- et tétra-nucléotides peuvent être séparés sur gel d'agarose haute résolution. Les motifs di-nucléotides qui sont les plus abondants sont généralement séparés sur gel d'acrylamide dénaturant. Les amorces PCR sont généralement choisies de manière à amplifier un fragment dont la taille est comprise entre 80 et 200 bp, ce qui permet une bonne séparation des fragments. En cas d'analyse sur séquenceur automatique, on peut choisir des tailles de fragments pouvant atteindre 400 bp. Il est possible de diminuer le nombre de réactions PCR en mélangeant plusieurs locus dans un même tube (multiplexage) mais les produits d'amplification doivent être de taille bien distincte. Le multiplexage peut être également réalisé en marquant des produits de tailles proches avec des fluorochromes différents. Les amplifications en multiplex étant parfois difficiles à mettre au point, les produits peuvent être mélangés après PCR.



RÉDACTEURS

BARROY-HUBLER Frédérique, UMR 6061, Faculté de Médecine, 2 avenue du Professeur Léon Bernard, 35043 Rennes, fhubler@univ-rennes1.fr

BÉCLIN Christophe, INRA, Biologie cellulaire, Route de Saint-Cyr, 78026 Versailles cédex, christophe.beclin@versailles.inra.fr

CERUTTI Martine, INRA, Recherches de Pathologie comparée, 30380 Saint-Christol-lès-Alès, cerutti@ales.inra.fr

CHALOT Michel, UMR INRA-UHP IaM, Biologie forestière, université Nancy I, faculté des Sciences, BP 239, 54506 Vandoeuvre cédex, michel.chalot@scbiol.uhp-nancy.fr

DAMIER Laurence, Unité des Agents antibactériens, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, ldamier@pasteur.fr

DECROOCQ Stéphane, INRA, UREFV, 71 avenue Edouard-Bourleaux, BP 81, 33883 Villenave d'Ornon cédex, sdecrooc@bordeaux.inra.fr

DEVAUCHELLE Gérard, INRA, Recherches de Pathologie comparée, 30380 Saint-Christol-lès-Alès, devauche@ales.inra.fr

DUPLESSIS Sébastien, INRA, UMR 1136 IaM, 54280 Champenoux, duplessi@nancy.inra.fr

GARCIA Virginie, INRA, Équipe de Génétique et amélioration des arbres forestiers, Centre de bioinformatique de Bordeaux, université Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux cédex, virginie.garcia@pmtg.u-bordeaux2.fr

FRIGERIO Jean-Marc, Unité de Recherches forestières, domaine de l'Hermitage, Pierroton, 33610 Cestas, jean-marc.frigerio@pierroton.inra.fr

GIBLOT Danièle, INRA Rennes, UMR BiO3P, BP 35327, 35653 Le Rheu cédex, giblot@rennes.inra.fr

GRANDBASTIEN Marie-Angèle, INRA, Biologie cellulaire, Route de Saint-Cyr, 78026 Versailles cédex, gbastien@versailles.inra.fr

GRENIER Eric, INRA Rennes, UMR BiO3P, BP 35327, 35653 Le Rheu cédex, grenier@rennes.inra.fr

HILBERT Jean-Louis, UPRES-EA 2702, Physiologie de la différenciation végétale, IFR 118, USTL, 59655 Villeneuve d'Ascq, jean-louis.hilbert@univ-lille1.fr

HOUEBINE Louis-Marie, Unité de Biologie du développement et biotechnologies, INRA, 78352 Jouy-en-Josas cédex, houebine@biotec.jouy.inra.fr

HUGOT Karine, Groupe Génomique, Centre de recherche INRA d'Antibes, 1382 route de Biot, 06560 Valbonne, hugot@antibes.inra.fr

JACQUOT Jean-Pierre, UMR INRA UHP IaM, Biochimie et biologie moléculaire végétales, université Nancy I, faculté des Sciences, BP 239, 54506 Vandoeuvre cédex, jean-pierre.jacquot@sbiol.uhp-nancy.fr

LEBLOND Pierre, Génétique et microbiologie, université Nancy I, faculté des Sciences, BP 239, 54506 Vandoeuvre cédex, pierre.leblond@sbiol.uhp-nancy.fr

LÉPLE Jean-Charles, INRA, Amélioration génétique et physiologie forestière, BP 20619, Ardon, 45166 Olivet cédex, leple@orleans.inra.fr

MASSONNEAU Agnès, Laboratoire de Reproduction et développement des plantes, ENS-Lyon, 46 allée d'Italie, 69364 Lyon cédex 07, agnes.massonneau@ens-lyon.fr

MÉREAU Agnès, UMR 6061, faculté de Médecine, 2 avenue du Professeur Léon Bernard, 35043 Rennes, agnes.mereau@univ-rennes1.fr

MONÉGER Françoise, Laboratoire de Reproduction et développement des plantes, ENS-Lyon, 46 allée d'Italie, 69364 Lyon cédex 07, francoise.moneger@ens-lyon.fr

MOUSSARD Christian, Service de Biochimie médicale, Hôpital Saint-Jacques, place Saint-Jacques, 25000 Besançon, christian.moussard@ufc-chu.univ-fcomte.fr

PILATE Gilles, INRA, Amélioration génétique et physiologie forestière, BP 20619, Ardon, 45166 Olivet cédex, pilate@orleans.inra.fr

PLOMION Christophe, Unité de recherches forestières, domaine de l'Hermitage, Pierroton, 33610 Cestas, christophe.plomion@pierroton.inra.fr

SAMSON Manuella, UMR Génétique et amélioration des plantes, BP 29, 35653 Le Rheu cédex, manusam35@yahoo.fr

SOURDILLE Pierre, UMR INRA-université Clermont II, Amélioration et santé des plantes; site de Crouël, 234 avenue du Brézet, 63039 Clermont-Ferrand cédex 2, pierre.sourdille@clermont.inra.fr

TAGU Denis, INRA Rennes, UMR BiO3P, BP 35327, 35653 Le Rheu cédex, tagu@rennes.inra.fr

Principes des techniques de biologie moléculaire

La biologie moléculaire a bouleversé les sciences du vivant. L'explosion de la génomique, qui propose des séquences de génomes entiers ainsi que des approches globales de leur fonctionnement, en est un exemple récent. L'objectif de cet ouvrage présenté sous forme de fiches n'est pas de détailler des protocoles ou des recettes toutes faites, mais d'expliquer simplement les principes théoriques des techniques de biologie moléculaire. Cette édition mise à jour propose des illustrations nouvelles et présente notamment de nombreuses techniques de génomique récemment apparues dans les laboratoires.

Cet ouvrage s'adresse à toute personne – spécialiste ou non – curieuse de connaître les bases des différentes techniques de manipulation des acides nucléiques.

Denis Tagu, docteur ès sciences, est directeur de recherches à l'INRA. Après avoir développé à Nancy un programme d'étude des mécanismes moléculaires de la symbiose mycorhizienne entre arbres et champignons, il a rejoint l'INRA de Rennes pour travailler à l'analyse moléculaire du développement des pucerons.

Christian Moussard, docteur d'état ès sciences pharmaceutiques, est maître de conférences de biochimie et de biologie moléculaire à la faculté mixte de Médecine et de Pharmacie de l'université de Franche-Comté et praticien au Centre hospitalier universitaire de Besançon.



9 782738 010674

ISSN: 1144-7605

ISBN : 2-7380-1067-9

Ref. : 01486

Prix : 22 €

