

50376  
1985  
135

50376  
1985  
135

UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

N° D'ordre : 19

## THÈSE

présentée pour l'obtention du grade de  
**DOCTEUR DE L'UNIVERSITÉ DES SCIENCES  
ET TECHNIQUES DE LILLE**

option : **Biochimie**

par

**Catherine HENRY-DUTHOIT**

# L'ONCOGENE *erbA* : DETERMINATION DE LA STRUCTURE PRIMAIRE DU GENE VIRAL ET ETUDE DES EQUIVALENTS CELLULAIRES CHEZ LE POULET ET CHEZ L'HOMME

soutenue le **8 Novembre** devant la commission d'examen  
**1985**

Président : Monsieur J. MONTREUIL  
Rapporteurs : Mademoiselle B. DEBUIRE  
Monsieur J. KREMBEL  
Monsieur C. LARSEN  
Examineurs : Monsieur G. BISERTE  
Monsieur D. STEHELIN



DOYENS HONORAIRES DE L'ANCIENNE FACULTE DES SCIENCES

MM. H. LEFEBVRE, M. PARREAU.

PROFESSEURS HONORAIRES DES ANCIENNES FACULTES DE DROIT  
ET SCIENCES ECONOMIQUES, DES SCIENCES ET DES LETTRES

MM. ARNOULT, Mme BEAUJEU, BONTE, BROCHARD, CHAPPELON, CHAUDRON, CORDONNIER, DECUYPER, DEHEUVELS, DEHORS, DION, FAUVEL, FLEURY, P. GERMAIN, GLACET, GONTIER, KOURGANOFF, LAMOTTE, LASSERRE, LELONG, Mme LELONG, LHOMME, LIEBAERT, MARTINOT-LAGARDE, MAZET, MICHEL, PEREZ, ROIG, ROSEAU, ROUELLE, SCHILTZ, SAVARD, ZAMANSKI.

PROFESSEUR EMERITE

M. A. LEBRUN.

ANCIENS PRESIDENTS DE L'UNIVERSITE DES SCIENCES ET  
TECHNIQUES DE LILLE

MM. M. PARREAU, J. LOMBARD, M. MIGEON.

PRESIDENT DE L'UNIVERSITE DES SCIENCES ET TECHNIQUES  
DE LILLE

M. J. CORTOIS.

PROFESSEURS - CLASSE EXCEPTIONNELLE

M. CONSTANT Eugène	Electronique
M. FOURET René	Physique du Solide
M. GABILLARD Robert	Electronique
M. MONTREUIL Jean	Biochimie
M. PARREAU Michel	Analyse
M. TRIDOT Gabriel	Chimie appliquée
M. VIVIER Emile	Biologie cellulaire
M. WERTHEIMER Raymond	Physique atomique et moléculaire

PROFESSEURS - 1ère CLASSE

M. BACCHUS Pierre	Astronomie
M. BEAUFILS Jean Pierre	Chimie physique
M. BIAYS Pierre	Géographie
M. BILLARD Jean	Physique du solide
M. BOILLY Bénoni	Biologie

M. BOUGHON Pierre	Algèbre
M. BOURIQUET Robert	Biologie Végétale
M. BREZINSKI Claude	Analyse numérique
M. CELET Paul	Géologie générale
M. CHAMLEY Hervé	Géotechnique
M. COEURE Gérard	Analyse
M. CORDONNIER Vincent	Informatique
M. DEBOURSE Jean Pierre	Gestion des entreprises
M. DYMENT Arthur	Mécanique
M. ESCAIG Bertrand	Physique du solide
M. FAURE Robert	Mécanique
M. FOCT Jacques	Métallurgie
M. GRANELLE Jean Jacques	Sciences Economiques
M. GRUSON Laurent	Algèbre
M. GUILLAUME Jean	Microbiologie
M. HECTOR Joseph	Géométrie
M. LABLACHE-COMBIER Alain	Chimie organique
M. LACOSTE Louis	Biologie végétale
M. LAVEINE Jean-Pierre	Paléontologie
M. LEHMANN Daniel	Géométrie
Mme LENOBLE Jacqueline	Physique atomique et moléculaire
M. LHOMME Jean	Chimie organique biologique
M. LOMBARD Jacques	Sociologie
M. LOUCHEUX Claude	Chimie physique
M. LUCQUIN Michel	Chimie physique
M. MAILLET Pierre	Sciences économiques
M. MIGNOT Fulbert	Analyse numérique
M. PAQUET Jacques	Géologie générale
M. PROUVOST Jean	Minéralogie
M. ROUSSEAU Jean-Paul	Physiologie animale
M. SALMER Georges	Electronique
M. SEGUIER Guy	Electrotechnique
M. SIMON Michel	Sociologie
M. STANKIEWICZ François	Sciences économiques
M. TILLIEU Jacques	Physique théorique
M. VIDAL Pierre	Automatique
M. ZEYTOUNIAN Radyadour	Mécanique

PROFESSEURS - 2ème Classe

M. ANTOINE Philippe	Analyse
M. BART André	Biologie animale
Mme BATTIAU Yvonne	Géographie
M. BEGUIN Paul	Mécanique
M. BELLET Jean	Physique atomique et moléculaire
M. BERZIN Robert	Analyse
M. BKOCHE Rudolphe	Algèbre
M. BODARD Marcel	Biologie végétale
M. BOSCOQ Denis	Probabilités
M. BRASSELET Jean-Paul	Géométrie et topologie

M. BRUYELLE Pierre	Géographie
M. CAPURON Alfred	Biologie animale
M. CARREZ Christian	Informatique
M. CAYATTE Jean-Louis	Sciences économiques
M. CHAPOTON Alain	Electronique
M. COQUERY Jean-Marie	Psychophysiologie
Mme CORSIN Paule	Paléontologie
M. CORTOIS Jean	Physique nucléaire et corpusculaire
M. COUTURIER Daniel	Chimie organique
M. CROSNIER Yves	Electronique
M. CURGY Jean-Jacques	Biologie
Mlle DACHARRY Monique	Géographie
M. DAUCHET Max	Informatique
M. DEBRABANT Pierre	Géologie appliquée
M. DEGAUQUE Pierre	Electronique
M. DELORME Pierre	Physiologie animale
M. DELORME Robert	Sciences économiques
M. DE MASSON D'AUTUME Antoine	Sciences économiques
M. DEMUNTER Paul	Sociologie
M. DENEL Jacques	Informatique
M. DE PARIS Jean-Claude	Analyse
Mlle DESSAUX Odile	Spectroscopie de la réactivité chimique
M. DEVRAINNE Pierre	Chimie minérale
M. DHAINAUT André	Biologie animale
Mme DHAINAUT Nicole	Biologie animale
M. DORMARD Serge	Sciences économiques
M. DOUKHAN Jean-Claude	Physique du solide
M. DUBOIS Henri	Spectroscopie hertziennne
M. DUBRULLE Alain	Spectroscopie hertziennne
M. DUBUS Jean-Paul	Spectrométrie des solides
M. DUPONT Christophe	Vie de la firme (I.P.A.)
M. FAKIR Sabah	Algèbre
M. FONTAINE Hubert	Dynamique des cristaux
M. FOUQUART Yves	Optique atmosphérique
M. FRONTIER Serge	Ecologie numérique
M. GAMBLIN André	Géographie urbaine, industrielle et démographi
M. GLORIEUX Pierre	Physique moléculaire et rayonnements atmosphé- riques
M. GOBLOT Rémi	Algèbre
M. GOSSELIN Gabriel	Sociologie
M. GOUDMAND Pierre	Chimie physique
M. GREGORY Pierre	I.P.A.
M. GREMY Jean-Paul	Sociologie
M. GREVET Patrick	Sciences économiques
M. GUILBAULT Pierre	Physiologie animale
M. HENRY Jean-Pierre	Génie mécanique
M. HERMAN Maurice	Physique spatiale
M. JACOB Gérard	Informatique
M. JACOB Pierre	Probabilités et statistiques
M. JACQUILLAT Bertrand	Gestion
M. JEAN Raymond	Biologie des populations végétales
M. JOFFRE Patrick	Vie de la firme (I.P.A.)
M. JOURNEL Gérard	Spectroscopie hertziennne

M. KREMBEL Jean  
M. LANGRAND Claude  
M. LATTEUX Michel  
Mme LECLERCQ Ginette  
M. LEFEVRE Christian  
Mlle LEGRAND Denise  
Mlle LEGRAND Solange  
Mme LEHMANN Josiane  
M. LEMAIRE Jean  
M. LE MAROIS Henri  
M. LHENAFF René  
M. LOCQUENEUX Robert  
M. LOSFELD Joseph  
M. LOUAGE Francis  
M. MACKE Bruno

M. MAIZIERES Christian  
M. MESSELYN Jean  
M. MESSERLIN Patrick  
M. MONTEL Marc  
Mme MOUNIER Yvonne  
Mme N'GUYEN VAN CHI Régine  
M. PARSY Fernand  
M. PASZKOWSKI Stéphan  
Mlle PAUPARDIN Colette  
M. PERROT Pierre  
M. PERTUZON Emile  
M. PONSOLLE Louis  
M. PORCHET Maurice  
M. POVY Lucien  
M. RACZY Ladislas  
M. RAOULT Jean-François  
M. RICHARD Alain  
M. RIETSCH François  
M. ROBINET Jean-Claude  
M. ROGALSKI Marc  
M. ROY Jean-Claude  
M. SCHAMPS Joël  
Mme SCHWARZBACH Yvette  
M. SLIWA Henri  
M. SOMME Jean  
Mlle SPIK Geneviève  
M. STAROSWIECKI Marcel  
M. STERBOUL François  
M. TAILLIEZ Roger  
Mme TJOTTA Jacqueline  
M. TOULOTTE Jean-Marc  
M. TURREL Georges  
M. VANDORPE Bernard  
M. VAST Pierre  
M. VERBERT André  
M. VERNET Philippe  
M. WALLART Francis  
M. WARTEL Michel  
M. WATERLOT Michel  
Mme ZINN Justin Nicole

Biochimie  
Probabilités et statistiques  
Informatique  
Catalyse  
Pétrologie  
Algèbre  
Algèbre  
Analyse  
Spectroscopie hertzienne  
Vie de la firme (I.P.A.)  
Géographie  
Physique théorique  
Informatique  
Electronique  
Physique moléculaire et rayonnements atmosphé-  
riques

Automatique  
Physique atomique et moléculaire  
Sciences économiques  
Physique du solide  
Physiologie des structures contractiles  
Géographie  
Mécanique  
Analyse numérique  
Biologie physiologie végétales  
Chimie appliquée  
Physiologie animale  
Chimie physique  
Biologie animale  
Automatique  
Electronique  
Géologie structurale  
Biologie animale  
Physique des polymères  
E.U.D.I.L.  
Analyse  
Psychophysiologie  
Spectroscopie moléculaire  
Géométrie  
Chimie organique  
Géographie  
Biochimie  
Informatique  
Informatique  
Génie alimentaire  
Mathématiques  
Automatique  
Spectrochimie infrarouge et Raman  
Chimie minérale  
Chimie inorganique  
Biochimie  
Génétique  
Spectrochimie infrarouge et Raman  
Chimie inorganique  
Géologie générale  
Algèbre

**A MA FAMILLE**

L'aboutissement de ce travail nous donne l'occasion de témoigner notre reconnaissance à ceux qui en ont permis la réalisation :

Nous remercions en particulier le Docteur **B. DEBUIRE** qui a guidé nos débuts dans la recherche. Ses judicieux conseils, sa compréhension et son dynamisme m'ont été une aide quotidienne. Je tiens à lui exprimer ma très vive reconnaissance.

Ce travail doit beaucoup à Monsieur le Professeur **G. BISERTE**, pour la confiance qu'il nous a accordée en nous accueillant à l'Institut de Recherches sur le Cancer de Lille et la bienveillance qu'il nous a toujours témoignée.

Monsieur le Docteur **D. STEHELIN** a suivi ce travail avec un chaleureux intérêt : qu'il nous soit permis aujourd'hui d'exprimer notre admiration pour la qualité des travaux qu'il dirige sur les oncogènes. Son groupe de recherches et en particulier Monsieur le Docteur **S. SAULE** rendent notre collaboration fructueuse et agréable.

Je remercie Monsieur le Professeur **J. MONTREUIL** de présider le jury de cette thèse et Monsieur le Professeur **J. KREMBEL** ainsi que Monsieur le Docteur **C.J. LARSEN** d'avoir accepté de juger ce travail.

Je remercie Madame le Docteur **M.H. LOUCHEUX** de nous avoir intégrée au sein de l'Unité INSERM (U. 124) qu'elle dirige.

Madame **M. BENAÏSSA** nous a toujours apporté une aide technique efficace et Mademoiselle **M. LESTIENNES** a dactylographié ce mémoire : qu'elles en soient vivement remerciées.

L'accomplissement de ce travail n'aurait pas été possible sans l'aide financière que nous ont accordée le Ministère de l'Industrie et de la Recherche et la Ligue Nationale Française contre le Cancer.

Les résultats exposés dans ce mémoire ont fait l'objet des publications et communications suivantes :

#### PUBLICATIONS

---

1. DEBUIRE B., HENRY C., BENAÏSSA M., BISERTE G., CLAVERIE J.M., SAULE S., MARTIN P. et STEHELIN D. (1984)  
Sequencing the erbA gene of Avian Erythroblastosis Virus reveals a new type of oncogene.  
**Science**, 224, 1456-1459.
2. HENRY C., DEBUIRE B., BENAÏSSA M. et BISERTE G.  
SAULE S., MARTIN P. et STEHELIN D. (1984).  
The nucleotide sequence of the erbA gene of Avian Erythroblastosis Virus appears unrelated to other known oncogenes  
Proceedings of the International Symposium "**Retroviruses and Human Pathology**" Lerici (La Spezia) Italie. Sept. 1984.
3. HENRY C., COQUILLAUD M., SAULE S., STEHELIN D. et DEBUIRE B. (1985).  
The four c-terminal amino-acids of the v-erbA polypeptide are encoded by an intronic sequence of the v-erbB oncogene.  
**Virology**, 140, 179-182.

#### COMMUNICATIONS

---

1. DEBUIRE B., HENRY C., BENAÏSSA M., BISERTE G., SAULE S., MARTIN P. et STEHELIN D.  
Sequencing the erbA gene of Avian Erythroblastosis Virus reveals a new type of oncogene.  
Présentation au Cold Spring Harbor Meeting "**RNA Tumor Viruses**" 22-27 Mai 1984. Cold Spring Harbor, New York, USA.
2. HENRY C., DEBUIRE B., BENAÏSSA M., BISERTE G., SAULE S., MARTIN P. et STEHELIN D.  
Sequencing the erbA gene of Avian Erythroblastosis Virus reveals a new type of oncogene.  
Présentation au Symposium International "**Retroviruses and Human pathology**". 25-26 Septembre 1984, Lerici (La Spezia), Italie.

## TRAVAUX ANNEXES

### PUBLICATIONS

1. MARTIN P., HENRY C., FERRE F., BECHADE C., BEGUE A., CALOTHY C., DEBUIRE B., STEHELIN D. et SAULE S. (1985).

Characterization of a myc containing retrovirus generated by propagation of the MH2 viral subgenomic RNA.

**J. Virol.**, soumis pour publication.

2. MARTIN P., FERRE F., GHYSDAEL J., STEHELIN D. et SAULE S., HENRY C., COQUILLAUD M. et DEBUIRE B. (1985).

Transformation of quail embryo fibroblasts by a retrovirus carrying the normal human c-myc gene, en préparation.

### COMMUNICATIONS

1. HENRY C., COQUILLAUD M., MARTIN P., FERRE F., SAULE S., STEHELIN D., BECHADE C., CALOTHY G. et DEBUIRE B.

Characterization of a myc containing retrovirus generated by propagation of the MH2 subgenomic RNA.

Présentation au Cold Spring Harbor Meeting "RNA Tumor Viruses" 21-26 Mai 1985. Cold Spring Harbor, New York, USA.

2. HENRY C., MARTIN P., FERRE F., COQUILLAUD M., GHYSDAEL J., DEBUIRE B., STEHELIN D. et SAULE S.

Recombinant avian retroviruses containing cellular human c-myc sequences.

Présentation au **First annual meeting on oncogenes**. 10-13 Juillet 1985. Frederik, Maryland, USA.

3. HENRY C., SAULE S., STEHELIN D. et DEBUIRE B.

The erbA oncogene, homology with the carbonic anhydrase gene family.

**13th International Congress of Biochemistry**. Communication lors de la réunion satellite sur les anhydrases carboniques organisée par le Professeur R. TASHIAN. 28 Septembre 1985. Amsterdam, Pays-Bas.

## TABLE des MATIERES

---

INTRODUCTION .....	P. 1
<hr style="border-top: 1px dashed black;"/>	
<u>CHAPITRE I : LES RETROVIRUS, VECTEURS DE GENES CELLULAIRES</u> .....	P. 3
I. <u>Rappel sur les rétrovirus de type C</u> .....	P. 4
II. <u>Caractères généraux des c-onc identifiés à partir des v-onc.</u> .....	P. 6
1. <u>c-onc</u> actuellement identifiés	
2. <u>Conservation des c-onc</u>	
3. <u>Localisation chromosomique et structure des c-onc.</u>	
4. Les produits des oncogènes :	
a. Les gènes codant pour des protéines kinases phosphorylant les tyrosines.	
b. La famille <u>ras</u> .	
c. Le gène <u>sis</u> et les gènes produisant des facteurs de croissance.	
d. Les oncogènes dont les produits ont une localisation nucléaire.	
III. <u>Mécanismes d'activation des oncogènes cellulaires au cours de la carcinogénèse.</u> .....	P. 12
1. Carcinogénèse par augmentation de l'expression des <u>c-onc</u> .	
a. La LTR activation.	
b. L'amplification génique.	
2. Carcinogénèse par altérations qualitatives des <u>c-onc</u> .	
3. Les deux théories ne sont pas mutuellement exclusives.	
4. Les translocations chromosomiques.	
<u>CHAPITRE II : METHODES DE DETERMINATION DE LA STRUCTURE PRIMAIRE DE L'ADN</u> .....	P. 16
I. <u>La méthode des terminaisons de chaînes de SANGER, NICKLEN et COULSON.</u>	
1. La méthode "plus-moins" de SANGER et COULSON	
a. Le système "moins".	
b. Le système "plus".	
2. La méthode des terminaisons de chaînes de SANGER, NICKLEN et COULSON.	
3. Clonage du fragment d'ADN à séquencer dans le bactériophage M 13.	
a. Cycle évolutif du bactériophage M 13.	
b. Caractéristiques du bactériophage M 13 utilisé en séquence.	
c. La méthode dite "M 13 dideoxy sequencing".	
II. <u>La méthode des dégradations chimiques de MAXAM et GILBERT.</u>	
<u>CHAPITRE III : PRESENTATION DU TRAVAIL.</u> .....	P. 31

**CHAPITRE IV : DETERMINATION DE LA STRUCTURE PRIMAIRE  
DE L'ONCOGENE v-erbA ET ORGANISATION DES GENES v-erbA et  
v-erbB DANS AEV. .... P. 37**

1. Stratégie de séquence.
2. Résultats et discussion.
  - a. Séquence de v-erbA.
  - b. Architecture du génome viral : organisation des gènes erbA et erbB dans AEV.
  - c. Le polypeptide erbA.
    - caractéristiques,
    - l'anhydrase carbonique,
    - homologie du polypeptide erbA avec les anhydrases carboniques.

**CHAPITRE V : ETUDE DES EQUIVALENTS CELLULAIRES CHEZ LE  
POULET ET CHEZ L'HOMME DE erbA-domaine 2 (HOMOLOGUE  
A L'ANHYDRASE CARBONIQUE)..... P. 53**

**I. Etude du locus c-erbA-domaine 2 de l'ADN cellulaire de poulet.**

1. Carte de restriction.
2. Séquence d'un fragment d'ADN cellulaire de poulet comportant le premier exon de c-erbA-domaine 2.

**II. Etude de la partie c-erbA-domaine 2 de l'ADN humain.**

1. Carte de restriction.
2. Séquence du domaine 2 du locus c-erbA complet.

**CONCLUSION ..... P. 58**

**APPENDICE TECHNIQUE ..... P. 62**

**BIBLIOGRAPHIE ..... P. 71**

**ARTICLES..... P. 87**

**INTRODUCTION**

Des études réalisées au cours de ces dernières années ont montré que la transformation cellulaire implique une cascade d'évènements conduisant chez l'animal de la cellule normale à la cellule cancéreuse. Cette cascade d'évènements peut être visualisée par une série de critères biologiques et plus récemment la transformation cellulaire a pu être reliée à un évènement génétique précis, l'activation de gènes dits oncogènes. Bien qu'on les ait d'abord observés lors de travaux sur les rétrovirus, ces gènes impliqués dans le développement des cancers ne sont pas d'origine virale. En fait, ils ne sont pas non plus spécifiques des cellules cancéreuses puisqu'on les trouve dans les cellules saines où on les appelle proto-oncogènes. Ils sont phylogénétiquement conservés, ce qui prouve le rôle important qu'ils jouent très certainement dans la croissance et/ou la différenciation cellulaire.

L'étude des rétrovirus a largement contribué aux progrès réalisés en cancérogénèse car ils représentent un outil de choix pour l'approche des mécanismes de cancérisation de la cellule.

L'isolement par Peyton Rous, en 1911, du premier virus tumorigène à partir d'un sarcome spontané du poulet a marqué le début des recherches sur les rétrovirus ou rétroviridae. Ces rétrovirus sont responsables de tumeurs variées chez de nombreuses espèces animales ; agents oncogènes extrêmement puissants, ils peuvent induire des tumeurs solides telles que des sarcomes ou des carcinomes ainsi que de nombreux types de leucémies.

Le présent mémoire relate l'étude structurale de l'oncogène v-erbA du virus de l'érythroblastose aviaire (AEV). Ce virus provoque l'apparition de sarcomes et leucémies chez le poulet et transforme les fibroblastes et les érythroblastes *in vitro*. Notre travail a révélé l'homologie du polypeptide erbA avec la famille des anhydrases carboniques, homologie intéressante puisque l'on connaît le rôle primordial de l'anhydrase carbonique dans le métabolisme érythrocytaire, les érythroblastes étant la principale cible de AEV.

Des précisions en ce qui concerne l'architecture du génome viral ont été apportées en étudiant un fragment du locus c-erbB de l'ADN cellulaire de poulet homologue à la partie 5' de v-erbB. Enfin, nous avons complété notre étude en déterminant la structure des équivalents cellulaires de v-erbA chez le poulet et chez l'homme.

Après une mise au point sur les rétrovirus, vecteurs de gènes cellulaires et les méthodes de détermination de la structure primaire de l'ADN, nous présenterons notre travail en essayant d'en dégager les points essentiels et les perspectives qui s'y rattachent.

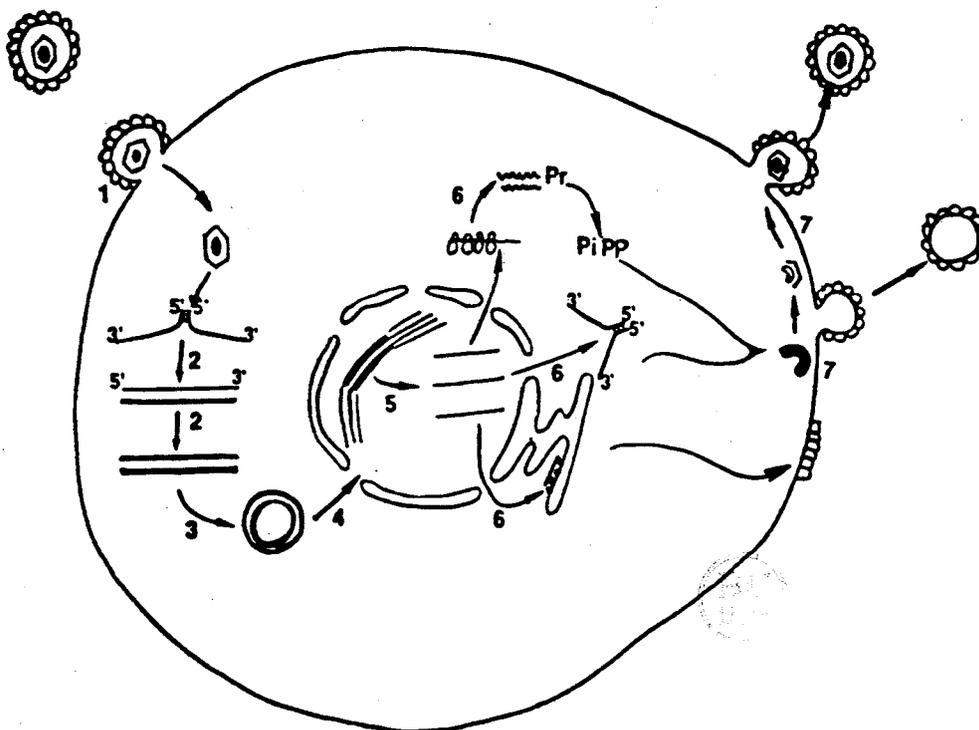
CHAPITRE I

LES RETROVIRUS, VECTEURS DE GENES CELLULAIRES

Une vingtaine de gènes humains normaux dont on peut penser qu'ils jouent un rôle dans le processus de cancérisation sont actuellement connus. Ces gènes sont appelés oncogènes cellulaires (c-onc) ou proto-oncogènes. La plupart ont été découverts grâce à leur remarquable similitude avec certains gènes viraux, les v-onc, caractéristiques des rétrovirus de type C et directement impliqués dans leur pouvoir transformant.

### I. Rappel sur les rétrovirus de type C : notions de v-onc et de c-onc.

Les rétrovirus ont un génome à ARN diploïde et possèdent la particularité de dépendre d'un intermédiaire ADN pour l'établissement de l'infection et pour leur propagation. Cet ADN, qui s'intègre à l'état double brin dans les chromosomes de l'hôte, est synthétisé à partir du génome viral par une ADN polymérase ARN dépendante ou rétropolymérase (TEMIN et BALTIMORE, 1972 ; VERMA, 1977) d'où le nom de rétrovirus (voir ci-dessous schéma du cycle réplcatif des rétrovirus. D'après ROUSSEL, 1982).



Cycle réplcatif des rétrovirus

1. Adsorption et pénétration du virus.
2. Synthèse d'ADN viral linéaire.
3. Circularisation de l'ADN viral.
4. Intégration de l'ADN de l'hôte.
5. Transcription de l'ADN proviral.
6. Traduction de l'ARN messenger.
7. Assemblage des particules virales.

Parmi ces rétrovirus, les oncovirus ont été classés en plusieurs catégories, les particules de type A, B, C et D, en fonction de leur structure et des modalités de leur maturation (BERNHARD, 1960 ; DALTON *et al.*, 1974). Les particules de type C, les plus répandues dans la nature, se forment à la surface des cellules par bourgeonnement de la membrane plasmique.

Connus chez de nombreuses espèces de vertébrés et particulièrement bien étudiés chez le poulet, la souris et le chat, les rétrovirus de type C ont des pouvoirs pathogènes que l'on peut diviser en trois classes.

Une première contient des virus apparemment dépourvus de pouvoir pathogène. Une seconde est faite de virus capables de déclencher des leucémies de types cytologiques très divers : les virus HTLV I et II responsables de lymphomes T chez l'homme appartiennent à cette classe. Une troisième classe de rétrovirus de type C possède un pouvoir transformant rapide, l'infection étant suivie en quelques jours à quelques semaines de l'apparition de sarcomes au point d'inoculation. Ils ne sont leucémogènes que dans quelques cas chez le poulet. Leur pouvoir pathogène est facilement détecté *in vitro*.

L'étude du génome viral révèle une différence majeure entre ces virus et ceux des deux autres classes. Les virus apparemment dépourvus de pouvoir pathogène et les virus à pouvoir pathogène d'expression lente ont en commun une même structure génomique. L'ADN complémentaire obtenu dans la cellule grâce à la rétopolymérase s'intègre dans le génome de l'hôte pour y former un provirus contenant trois gènes principaux gag, pol, env qui codent respectivement pour les protéines de structure du noyau, la rétopolymérase et les glycoprotéines d'enveloppe. A ces gènes s'ajoutent parfois de courtes séquences codantes comme ORF-X dans HTLV I (SEIKI *et al.*, 1983). Deux séquences identiques, les LTR (Long Terminal Repeats) situées aux extrémités 5' et 3' contiennent les promoteurs responsables de la transcription des messages viraux.

Les rétrovirus de type C à pouvoir transformant rapide sont caractérisés par la présence d'un gène appelé v-onc responsable de l'activité transformante *in vitro* et *in vivo* et qui se localise en des points très variables du génome viral. Le v-onc prend la place d'un ou plusieurs gènes de structure viraux normaux. Le virus qui le porte est donc défectif pour la répllication, c'est-à-dire qu'il ne peut se répliquer sans l'aide d'un virus auxiliaire complet qui lui fournit les protéines manquantes. Depuis le travail initial de STEHELIN *et al.* en 1976, nous savons que les v-onc dérivent de gènes cellulaires normaux accidentellement incorporés par les rétrovirus au cours de leur répllication. Les

séquences des v-onc diffèrent le plus souvent de celles des gènes cellulaires dont elles dérivent par diverses substitutions ou délétions mais elles en restent cependant très proches, si bien que les sondes radioactives préparées à partir de v-onc ont permis l'isolement des c-onc d'abord dans le génome de l'espèce porteuse du virus puis dans les autres espèces grâce à la conservation des c-onc au cours de l'évolution.

## II. Caractères généraux des c-onc identifiés à partir des v-onc.

### 1. c-onc actuellement identifiés.

Le tableau I dresse la liste des c-onc actuellement identifiés.

Les six premiers gènes de ce tableau correspondent à des séquences cellulaires différentes mais ont pourtant été isolés à partir de virus sarcomatogènes du poulet. Il en est de même pour les gènes isolés chez le chat qui proviennent tous de virus sarcomatogènes félines ayant des pouvoirs pathogènes semblables. Les gènes myb, ets (LEPRINCE *et al.*, 1983), erbA, erbB, myc, (ROUSSEL *et al.*, 1979) et mil (COLL *et al.*, 1983a) ont été isolés à partir de virus aviaires qui transforment respectivement les myéloblastes, les érythroblastes et les macrophages tandis que abl a été isolé à partir d'un virus inducteur de lymphomes pré-B chez la souris (GOFF *et al.*, 1982), le virus d'ABELSON. Les gènes isolés à partir de virus de souris et de rat l'ont été à partir de virus de type MSV (Murine Sarcoma Virus) induisant des sarcomes chez ces espèces animales. Le gène sis a été isolé d'un virus de primate également inducteur de sarcomes. Dans trois cas, le virus a incorporé deux gènes cellulaires différents, erbA et erbB dans le virus de l'érythroblastose aviaire, myb et ets dans le virus E26, myc et mil dans le virus MH2.

### 2. Conservation des c-onc

Les c-onc ont été remarquablement conservés au cours de l'évolution, ce qui prouve le rôle essentiel qu'ils jouent probablement dans la cellule. Le gène src initialement décrit chez le poulet est retrouvé dans toutes les espèces de vertébrés, y compris l'homme et les poissons (BARNEKOW *et al.*, 1982) mais aussi chez les invertébrés jusqu'à la drosophile (SHILO *et al.*, 1981) et même l'éponge ; de même myc, abl et fes sont retrouvés chez la drosophile (SIMON *et al.*, 1983 ; WEN-HWA *et al.*, 1982). Seul les gènes lil et rel n'ont pas été mis en évidence chez l'homme et les vertébrés : lil semble propre au genre gallus (BOCCARA *et al.*, 1982) et rel aux oiseaux (WONG et LAI, 1981).

c-onc	Virus contenant le <u>v-onc</u> correspondant		Espèces dans laquelle la transduction <u>c-onc</u> → <u>v-onc</u> a eu lieu
	Pouvoir pathogène	Souches	
src	Sarcomes	Rous	Poulet
fps (= fes)	Sarcomes	Fujinami, PRCII URI	Poulet
yes	Sarcomes	U 73, ESH	Poulet
ros	Sarcomes	UR 2	Poulet
ski	Sarcomes	SKV	Poulet
lil *	Sarcomes	BR 77	Poulet
myb	Leucémie myéloblastique	AMV, E26	Poulet
ets	Leucémies érythroblastiques	E26	Poulet
erbA *	Leucémies érythroblastiques	AEV	Poulet
erbB }		AEV	Poulet
myc	Leucémie myélo-monocytaire	MC29, MH2, CMII, OK10	Poulet
mil (ou mht)*	Leucémie myélo-monocytaire	MH2	Poulet
rel	Lymphomes T et B	REV-T	Dinde
fos	Sarcomes	FBI-MSV	Souris
mos	Sarcomes	M-MSV	Souris
raf	Sarcomes	3611-MSV	Souris
bas	Sarcomes	BALB/C	Souris
Ha-ras-1	Sarcomes	RASHEED-MSV	Rat
Ha-ras-1	Sarcomes	HARVEY-MSV	Rat
Ha-ras-1	Sarcomes	Balb/MSV	Souris
Ki-ras-2	Sarcomes	Kirstein-MSV	Rat
abl	Lymphomes B	ABELSON L.V.	Souris
abl	Sarcomes	HZ-2-FeSV	Chat
fes (= fps)	Sarcomes	ST-FeSV, GA-FeSV	Chat
fms	Sarcomes	SM-FeSV	Chat
fgr	Sarcomes	GR-FeSV	Chat
kit	Sarcomes	HZ4	Chat
sis	Sarcomes	PI-FeSV	Chat
sis	Sarcomes	SSV	Singe Lagothrix
<u>c-onc</u> sans équivalents viraux connus isolés dans l'ADN tumoral transformant NIH-3T3			
N-ras			
Blym			

\* Le pouvoir transformant de ce gène n'est pas établi.

### TABLEAU I

Les c-onc (LEVY, 1984).

### 3. Localisation chromosomique et structure des c-onc.

Chez l'homme, les c-onc existent sous forme d'un gène unique localisé dans de nombreux cas (Tableau II). Pour c-myc cependant l'existence d'autres loci a été suggérée ; il s'agit probablement de pseudogènes non fonctionnels (DALLA-FAVERA *et al.*, 1982), un autre gène inconnu apparenté, amplifié dans certaines tumeurs a été également signalé (MONTGOMERY *et al.*, 1982).

Les c-onc sont des gènes en mosaïque faits d'exons codants et d'introns non codants répartis en général sur 12 à 35 kilobases d'ADN. Cette taille, très supérieure à celle des v-onc correspondants dépourvus d'introns, montre que les c-onc ne sont pas des gènes viraux intégrés dans le génome des mammifères et s'y maintenant. Les v-onc ayant gardé la totalité ou la plus grande partie de l'information génétique traduite en protéine des c-onc sont au contraire dérivés de ceux-ci à partir de leurs transcrits ARN.

Il faut noter cependant quelques exceptions à la structure discontinue des c-onc : c-mos est constitué d'une séquence unique colinéaire au v-mos (JONES *et al.*, 1980 ; VAN BEVEREN *et al.*, 1983), c-Ha-ras 2 est également dépourvu d'introns (O'BRIEN *et al.*, 1983) de même que c-Ki-ras 1 (Mc GRATH *et al.*, 1983).

### 4. Les produits des oncogènes

Le mode d'action des oncogènes est encore inconnu mais l'étude des facteurs impliqués dans la mitose permet de penser que les oncogènes pourraient agir de façon similaire pour transformer la cellule (HELDIN *et al.*, 1984). Les données récentes concernant les produits des oncogènes permettent de les classer en plusieurs familles (Tableau III).

#### a. Les gènes codant pour des protéines kinases phosphorylant les

tyrosines.

Le modèle de cette famille est le gène src qui donne naissance à une phosphoprotéine de masse moléculaire avoisinant 60 000 daltons (pp 60 v-src, pp 60 c-src).

Le dénominateur commun des protéines issues de ces gènes (fps, yes, ros, erbB, fgf, fes, abl) consiste en leur activité tyrosine kinase (HELDIN *et al.*, 1984 ; PAWSON *et al.*, 1980a). Dans la cellule, cet évènement est relativement rare mais semble constant dans l'interaction entre les facteurs de

Gène	:	Chromosomes
	:	
	:	
	:	
c-myb	:	6
c-mos	:	8
c-myc	:	8
c-fes	:	15
c-fos	:	
c-sis	:	22
c-src	:	20
c-abl	:	9
c-fms	:	5
c-Ha-ras 1	:	11
c-Ha-ras 2	:	X
c-Ki-ras 1	:	6
c-Ki-ras 2	:	12
c-N-ras	:	1
c-raf	:	3
c-erbA	:	17
c-erbB	:	7
c-ski	:	1
c-fos	:	2
c-ets	:	11
c-Blym	:	1
	:	



TABLEAU II

Localisation chromosomique des c-onc (GILDEN et al., 1984).

NOM DE L'ONCOGENE :		PROTEINES ONCOGENIQUES			
RETROVIRUS :	TUMEUR :	LOCALISATION CELLULAIRE :	FONCTION :	CLASSE :	
src :	SARCOME DE POULET :	- :	MEMBRANE PLASMIQUE :		
yes :	SARCOME DE POULET :	- :	MEMBRANE PLASMIQUE (?) :		
fgf :	SARCOME DE CHAT :	- :	(?) :	CLASSE 1 (PROTEINES	
abl :	LEUCEMIE DE SOURIS :	LEUCEMIE HUMAINE :	MEMBRANE PLASMIQUE :	PROTEINES KINASES : KINASES CYTOPLAS-	
fps :	SARCOME DE POULET :	- :	CYTOPLASME (MEMBRANE PLASMIQUE ?) :	MIQUE D'UN RECEPTEUR DE LA TYROSINE : MIQUES SPECIFIQUES DE LA TYROSINE)	
fes :	SARCOME DE CHAT :	- :	CYTOPLASME (CYTOSQUELETTE ?) :		
ros :	SARCOME DE POULET :	- :	(?) :		
erb-B :	LEUCEMIE DE POULET :	- :	MEMBRANES PLASMIQUE ET INTERNES :	REGION PROTEINE : KINASE CYTOPLASMIQUE DU RECEPTEUR DE L'EGF, SPECIFIQUE DE LA TYROSINE	
fms :	SARCOME DE CHAT :	- :	MEMBRANES PLASMIQUE ET INTERNES :	REGION CYTOPLASMIQUE D'UN RECEPTEUR DE FACTEUR DE CROISSANCE (?) : APPARENTEES A LA CLASSE 1 (PROTEINES KINASES POTENTIELLES)	
mil :	CARCINOME DE POULET :	- :	CYTOPLASME :	(?) :	
raf :	SARCOME DE SOURIS :	- :	CYTOPLASME :	(?) :	
mos :	SARCOME DE SOURIS :	LEUCEMIE DE SOURIS :	CYTOPLASME :	(?) :	
sis :	SARCOME DE SINGE :	- :	SECRETEE :	FACTEUR DE CROISSANCE SEMBLABLE AU PDGF : CLASSE 2 (FACTEURS DE CROISSANCE)	
Ha-ras :	SARCOME DE RAT :	CARCINOME HUMAIN : CARCINOME DE RAT :	MEMBRANE PLASMIQUE :		
Ki-ras :	SARCOME DE RAT :	CARCINOME, LEUCEMIE ET SARCOME HUMAINS :	MEMBRANE PLASMIQUE :	CLASSE 3 (CYTOPLASMIQUES, SE FIXENT SUR LE GTP)	
N-ras :	- :	LEUCEMIE ET CARCINOME HUMAINS :	MEMBRANE PLASMIQUE :		
fos :	SARCOME DE SOURIS :	- :	NOYAU :	(?) :	
myc :	LEUCEMIE DE POULET :	LYMPHOME HUMAIN :	NOYAU :	FIXATION SUR L'ADN :	
myb :	LEUCEMIE DE POULET :	LEUCEMIE HUMAINE :	NOYAU :	FIXATION SUR L'ADN : CLASSE 4 (NUCLEAIRES)	
B-lym :	- :	LYMPHOME DE POULET, LYMPHOME HUMAIN :	NOYAU (?) :	(?) :	
ski :	SARCOME DE POULET :	- :	NOYAU (?) :	(?) :	
rel :	LEUCEMIE DE DINDE :	- :	(?) :	(?) :	
erb-A :	LEUCEMIE DE POULET :	- :	CYTOPLASME :	(?) : NON CLASSEES	
ets :	LEUCEMIE DE POULET :	- :	(?) :	(?) :	

TABLEAU III

LES FAMILLES DE PROTEINES CODEES PAR LES ONCOGENES D'APRES "POUR LA SCIENCE", OCTOBRE 1984. (CE TABLEAU A ETE REACTUALISE EN FONCTION DES ETUDES REALISEES DEPUIS OCTOBRE 1984).

croissance et leurs récepteurs (HELDIN *et al.*, 1984). Dès lors, on peut imaginer que ces oncogènes constituent une classe de récepteurs de facteurs de croissance ; la confirmation de cette hypothèse a été obtenue récemment avec la mise en évidence d'une similitude entre un gène de cette famille (erbB) et le récepteur d'un facteur de croissance, le récepteur de l'EGF (Epidermal Growth Factor) (DOWNWARD *et al.*, 1984).

#### b. La famille ras.

La seule caractéristique biochimique du produit de ces gènes, actuellement mise en évidence, consiste en la fixation du GTP (guanosine triphosphate). Ce groupe de gènes, également présent chez la levure serait impliqué dans la sporulation des microorganismes (TATCHELL *et al.*, 1984).

#### c. Le gène sis et les gènes produisant des facteurs de croissance.

Le produit du gène sis présente une très forte homologie avec la chaîne bêta d'un facteur de croissance impliqué dans la réparation tissulaire (WATERFIELD *et al.*, 1983) et produit ainsi que d'autres, par les plaquettes : le PDGF (Platelet Derived Growth Factor).

Connaissant la parenté qui existe dans le mode d'action du PDGF et des facteurs de croissance cellulaire tels que l'EGF normal ou les TGF (Transforming Growth Factors) produits par des cellules tumorales (COOPER *et al.*, 1981), on peut imaginer que c-sis intervienne dans la transformation par l'intermédiaire d'un facteur de croissance produit en excès ou qualitativement anormal et responsable, en conjonction éventuelle avec d'autres facteurs physiologiques, d'une croissance accrue des cellules tumorales.

#### d. Les oncogènes dont les produits ont une localisation nucléaire

Les protéines issues de ces gènes (exemple myc et myb) pourraient transformer la cellule par leur interaction avec l'ADN, interaction qui a été prouvée pour le produit des deux gènes myc et myb (ABRAMS *et al.*, 1982 ; MOELLING *et al.*, 1985). Dans l'ordre des événements intervenant au cours de la transformation cette classe de gènes pourrait prendre le relai du signal

mitogène délivré lors de l'interaction facteur de croissance-récepteur. A l'appui de cette hypothèse, KELLY et al., (1983) ont montré une activation très rapide du taux de transcription du gène myc en réponse des cellules à l'addition de PDGF dans le milieu de culture.

Il reste à savoir si les oncogènes non encore classés appartiennent à l'une ou l'autre de ces familles ou en révéleront de nouvelles.

### III MECANISMES D'ACTIVATION DES ONCOGENES CELLULAIRES AU COURS DE LA CARCINOGENESE.

Schématiquement, deux hypothèses, l'une qualitative, l'autre quantitative, peuvent expliquer qu'un gène normal devienne responsable de cancers.

Selon l'**hypothèse quantitative**, le c-onc normal fonctionne de façon excessive et l'accumulation de son produit est responsable de la transformation maligne. Ceci implique que le proto-oncogène de la cellule normale ne soit pas différent de l'oncogène des cellules tumorales et que la protéine produite soit identique. De plus, le gène viral transformant, dérivé du gène cellulaire par transduction, doit être identique à celui-ci (mises à part les mutations ponctuelles liées aux divergences évolutives et qui ne modifient pas la fonction du produit) : le v-onc est identique au c-onc et leurs produits sont les mêmes.

Au contraire, dans l'**hypothèse qualitative** de la carcinogénèse, celle-ci est due à une mutation du c-onc modifiant la fonction de la protéine. Le c-onc des cellules malignes et les protéines produites dans les deux cas se distinguent par au moins une mutation significative régulièrement retrouvée dans les tissus anormaux. Dans cette hypothèse la transduction d'un c-onc dans un virus donne naissance à un v-onc transformant à la suite de remaniements intervenant au cours de cette transduction : le v-onc est différent du c-onc.

#### 1. Carcinogénèse par augmentation de l'expression des c-onc.

##### a. La LTR activation.

Un ALV (Avian Leukosis Virus), virus ne contenant pas d'oncogène, peut être à l'origine d'une carcinogénèse lente (lymphomes B aviaires) dont le mécanisme a été explicité par les travaux de HAYWARD et al., (1981) et

PAYNE *et al.*, (1982). Dans les lymphomes B, le virus s'intègre en amont du gène c-myc et son LTR 3', seul nécessaire à la transformation, permet une lecture accrue du message c-myc avec formation d'ARN messagers mixtes contenant des séquences virales et cellulaires.

De même une érythroleucémie du poulet peut être induite lorsqu'un ALV s'intègre au voisinage du gène c-erbB (FUNG *et al.*, 1983). KUNG *et al.* viennent de mettre en évidence quelques faits surprenant concernant cette LTR activation de c-erbB (RAINES *et al.*, 1985) : le provirus utilise le LTR 5' plutôt que le LTR 3' pour traduire un message, dit message "readthrough", comprenant les séquences virales et les séquences c-erbB situées en 3' ; les transcrits primaires sont remaniés par deux phénomènes d'épissage (gag-env et env-erb) pour donner une protéine commençant à l'AUG de gag et comprenant 53 acides aminés de env puis des séquences homologues à v-erbB. De plus, ce produit de c-erbB activé comporte à son extrémité C-terminale 32 amino-acides homologues au récepteur de l'EGF et absents dans la protéine gp74 erbB.

#### b. L'amplification génique.

Dans certaines tumeurs, la production accrue de messagers est liée à l'amplification de gènes. Ceci a été observé pour c-myc dans des tumeurs humaines telles que la leucémie promyélocytaire HL 60 (COLLINS *et al.*, 1982) ou la tumeur endocrine COLO 320 (ALITALO *et al.*, 1983). Les gènes additionnels peuvent se trouver sur des chromosomes double-minutes ou, de façon stable, sur des régions homogènes hypercolorées.

Dans la leucémie myéloïde K 562, l'amplification de c-abl en même temps que celle de gènes de chaînes lambda d'immunoglobulines a été observée (COLLINS *et al.*, 1983).

## 2. Carcinogénèse par altérations qualitatives des c-onc.

L'augmentation de l'expression des c-onc n'est pas toujours observée dans les tumeurs, par ailleurs un c-onc normal lié à un LTR n'est pas toujours transformant (exemple c-fos et c-N-ras) (DUESBERG, 1983). Les v-onc et c-onc sont différents dans de nombreux cas et donnent naissance à des produits différents. Une modification qualitative d'un gène cellulaire au cours de son insertion dans un virus aboutit souvent à la création d'un gène transformant à

partir d'un proto-oncogène qui ne l'était probablement pas. La démonstration d'une telle mutation somatique a été apportée dans le cas d'une lignée de tumeur de vessie, maintenue *in vitro*, T<sub>24</sub>. Cette tumeur est connue pour posséder un gène c-Ha-ras 1 activé. La seule différence entre le gène normal et le gène de la tumeur T<sub>24</sub> est une mutation sur le douzième codon aboutissant au remplacement de la glycine normale par une valine (REDDY, 1983 ; TABIN *et al.*, 1982 ; TAPAROWSKY *et al.*, 1982). Une anomalie du même douzième codon est retrouvée dans les tumeurs SW 480 (valine 12) et CALU 1 (cystéine 12) (CAPON *et al.*, 1983) au niveau de c-Ki-ras 2 et dans les séquences des divers v-ras (DHAR *et al.*, 1982 ; TSUCHIDA *et al.*, 1982).

### 3. Les deux théories ne sont pas mutuellement exclusives.

Dans certaines tumeurs des gènes peuvent être à la fois amplifiés et mutés (Mc COY *et al.*, 1983 ; SCHWAB *et al.*, 1983). De même le gène c-ras peut être activé aussi bien par simples couplages avec un LTR que par des mutations ponctuelles, telles celles observées dans T<sub>24</sub>. Donc les arguments solides à la fois en faveur d'une carcinogénèse qualitative et en faveur d'une carcinogénèse quantitative ne sont pas nécessairement contradictoires.

### 4. Les translocations chromosomiques.

Dans les tumeurs de BURKITT humaines, on note l'existence habituelle d'une translocation chromosomique 8 : 14 portant plus précisément sur les régions 8q24 et 14q32, régions qui sont précisément celles comprenant normalement le gène c-myc (8q24) et les gènes des chaînes lourdes d'immunoglobulines (14q32) (ZHANG *et al.*, 1982 ; DALLA FAVERA *et al.*, 1983). D'autres translocations chromosomiques plus rares, de type 8 : 2 et 8 : 22, affectent des régions chromosomiques qui portent respectivement les gènes des chaînes kappa et lambda des immunoglobulines.

Un modèle semblable existe dans les plasmocytomes murins qui présentent le plus souvent une translocation 15 : 12 affectant le chromosome 15 où se trouve le gène c-myc et le chromosome 12 où ont été localisés les gènes des chaînes lourdes des immunoglobulines (CALAME *et al.*, 1982 ; HARRIS *et al.*, 1982). Dans un cas de plasmocytome murin, on a observé que la translocation habituelle 15 : 12 entraînait l'activation de c-mos et non de c-myc (RECHAVI *et al.*, 1982).

Il a été montré récemment que la translocation 9 : 22 caractéristique du chromosome Philadelphie mettait en cause les régions porteuses de c-abl sur le chromosome 9 et c-sis sur le chromosome 22 (GROFFEN *et al.*, 1983 ; DE KLEIN *et al.*, 1982). Il semblerait également que c-fes soit transloqué du chromosome 15 au chromosome 17 dans un cas de leucémie aiguë à promyélocyte (SHEER *et al.*, 1983).

#### CONCLUSION :

-----

Il est acquis que les oncogènes jouent un rôle déterminant dans la carcinogénèse chez l'homme aussi bien que chez les autres espèces animales. Cependant la séquence des phénomènes en cause reste encore incertaine et le nombre réel d'oncogènes capables d'intervenir est inconnu bien que vraisemblablement limité. La connaissance des fonctions cellulaires normales des c-onc, encore très peu avancée, et des altérations de ces fonctions permettra d'appréhender le détail de ces phénomènes.

Une théorie générale de la carcinogénèse a été énoncée. Il est en effet très probable que les mêmes gènes soient en cause dans la carcinogénèse virale et dans les carcinogénèses non virales, qu'elles soient dues à des radiations ionisantes, à des carcinogènes chimiques ou à des mutations spontanées de quelque nature qu'elles soient.

**CHAPITRE II**

**METHODES DE DETERMINATION DE LA STRUCTURE PRIMAIRE DE L'ADN**

Pour des raisons d'ordre technique, les scientifiques ont abordé l'analyse structurale des génomes après celle des protéines et de l'ADN, c'est-à-dire dans un ordre strictement inverse de celui du passage de l'information au sein de la cellule.

Jusqu'en 1972, manquant de d'enzymes suffisamment spécifiques pour couper les gènes en oligonucléotides, les biochimistes étaient dans l'impossibilité d'analyser la structure primaire de l'ADN.

En 1974, utilisant une nouvelle enzyme de restriction, l'endonucléase IV, GALIBERT *et al.* décrivaient une première méthode permettant l'analyse de la structure primaire de l'ADN du phage  $\phi$ x 174 (ZIFF *et al.*, 1973 ; GALIBERT *et al.*, 1974). Toutefois cette méthode complexe et laborieuse n'aurait permis que difficilement l'analyse de gènes entiers.

Aujourd'hui plusieurs méthodes rapides d'analyse de l'ADN ont été publiées et, l'on est passé en quelques années des premiers résultats correspondant à des fragments de quelques dizaines de nucléotides à la connaissance de génomes entiers, tels que, ceux du phage  $\phi$ x 174 (SANGER *et al.*, 1977a), du virus de l'hépatite B (GALIBERT *et al.*, 1979) et plus récemment celui du virus LAV responsable du SIDA (syndrome immuno-déficitaire acquis) (WAIN-HOBSON *et al.*, 1985).

Deux faits principaux expliquent cette évolution. Tout d'abord, la découverte et le prodigieux essor des enzymes de restriction (SANGER *et al.*, 1977b) puis l'idée apparemment toute simple de F. SANGER selon laquelle il est possible de fractionner des oligonucléotides en fonction de leur taille par électrophorèse en gel d'acrylamide.

Depuis 1975, plusieurs méthodes de séquence de l'ADN ont été décrites, mais deux d'entr'elles sont principalement utilisées :

- une méthode enzymatique dite méthode des terminaisons de chaînes de F. SANGER, S. NICKLEN et A.R. COULSON (1977) qui constitue une remarquable amélioration de leur méthode "plus-moins" publiée en 1975 (SANGER et COULSON, 1975) ; et,
- une méthode chimique, la méthode des dégradations chimiques de MAXAM et GILBERT (1977).

En respectant un ordre chronologique, nous décrirons tout d'abord la méthode des terminaisons de chaînes de SANGER, NICKLEN et COULSON puis la méthode des dégradations chimiques de MAXAM et GILBERT.

## I. La méthode des terminaisons de chaînes de SANGER, NICKLEN et COULSON.

Elle constitue une amélioration de la méthode "plus-moins" de SANGER et COULSON (1975).

### 1. La méthode "plus-moins" de SANGER et COULSON.

Le fragment d'ADN monocaténaire à séquencer n'est pas séquencé directement mais répliqué *in vitro* par l'ADN polymérase d'*E. Coli* à partir d'une séquence amorce ou "primer" (courte séquence nucléotidique appariée à l'ADN modèle étudié) en présence de quatre désoxyribonucléotides (dXTP) dont l'un est marqué au P<sup>32</sup>, le plus souvent il s'agit de dATP (figure 1).

Cette synthèse doit être aussi désordonnée que possible, de façon telle qu'un maximum d'oligonucléotides de différentes longueurs mais commençant tous à la séquence "amorce" soient formés. Ce mélange réactionnel est ensuite purifié sur colonne d'agarose afin d'éliminer l'excès de dXTP et l'éluat, divisé en deux lots, est traité comme suit :

#### a. Le système "moins".

Le mélange d'oligonucléotides encore hybridé à la molécule modèle est incubé à nouveau avec l'ADN polymérase I en présence de 3 dXTP ; si, par exemple, le dATP est le dXTP manquant (système A-) chaque chaîne se terminera en 3' par le dXTP précédant un résidu A.

Quatre réactions séparées sont donc réalisées, chacune en l'absence de l'un des dXTP.

Les quatre mélanges réactionnels sont dénaturés afin de séparer les chaînes nouvellement synthétisées de la molécule modèle et soumis à une électrophorèse en gel d'acrylamide en présence d'urée 8M. Les différents oligonucléotides, ayant tous la même extrémité 5' (l' "amorce") mais des extrémités 3' différentes seront donc séparés en fonction de leur taille.

Idéalement, chaque oligonucléotide doit être séparé de son voisin comportant un résidu supplémentaire et l'autoradiographie du gel permet une lecture directe de la séquence complémentaire de celle de la molécule modèle.

Cependant, ce système seul n'est généralement pas suffisant pour déterminer une séquence, et un second système similaire est normalement utilisé conjointement.

Matrice 3'-----A T G C T G-----5'

19

Amorce 5'-----3'

DNA-polymérase + 4 triphosphates  
dont 1 est marqué au <sup>32</sup>P.

Matrice 3'-----A T G C T G-----

-----T A C G A C

-----T A C G A

Copies de différentes-----T A C G

longueurs marquées -----T A C

au <sup>32</sup>P. -----T A

-----T

a

Système -A

Système +A

-----A T G C T G-----

-----A T G C T G-----

-----T A C G A C-----

-----T A C G A

-----T A C G A C

-----T A C G A

-----T A C G

-----T A

-----T A C G

-----T A

-----T A C G

-----T A

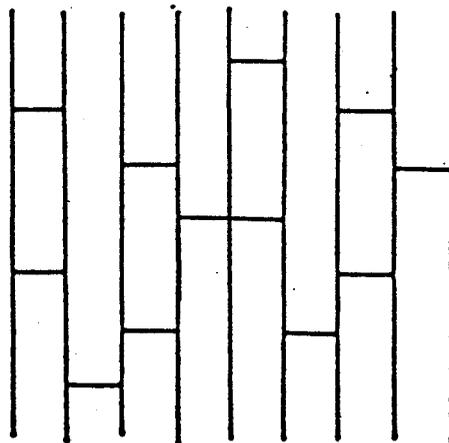
-----T

-----T A

b

électrophorèse

-C -T -A -G +C +T +A +G



c

Figure 1 : Principe de la méthode "plus-moins".

### b. Le système "plus".

Ce système se réfère à la méthode d'ENGLUND (1971, 1972) qui montre qu'en présence d'un seul dXTP, l'ADN polymérase du **phage T4** infectant **E. Coli** (T4-polymérase) dégrade l'ADN double-brin à partir de son extrémité 3', mais cette action exonucléasique est arrêtée aux résidus A, par exemple, si le dATP est présent.

Cette méthode est appliquée au mélange réactionnel obtenu en a dans la figure 1. Les échantillons sont incubés avec la T4-polymérase et un seul dXTP, puis fractionnés par électrophorèse parallèle en gel d'acrylamide.

Donc, dans le système A+, seul le dATP est présent et toutes les chaînes se termineront par un A. La position des résidus A sera indiquée par les bandes de l'autoradiographie.

En général, dans le système +, les oligonucléotides ont un résidu de plus que dans le système -, mais s'il y a plusieurs A consécutifs, la distance entre les bandes des systèmes A- et A+ indique le nombre de résidus présents.

Ni la méthode plus, ni la méthode moins, ne sont complètement satisfaisantes et, de façon à déterminer une séquence, les deux méthodes doivent être utilisées et des vérifications complémentaires sont parfois nécessaires.

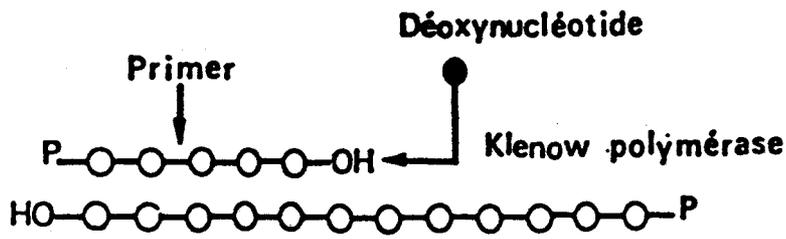
SANGER *et al.* améliorèrent considérablement la méthode "plus-moins" en publiant en 1977, la méthode des terminaisons de chaînes.

## 2. La méthode des terminaisons de chaînes de SANGER, NICKLEN et COULSON.

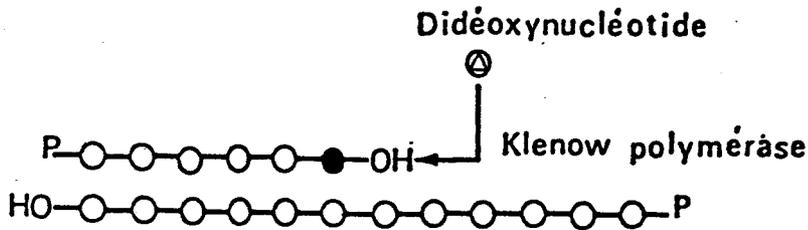
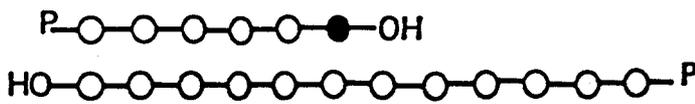
Cette méthode reprend le principe de la méthode de SANGER.

Il a été démontré que l'activité inhibitrice du 2'-3' didéoxythymidine triphosphate (ddTTP) sur l'ADN polymérase I repose sur son incorporation dans la chaîne nucléotidique synthétisée à la place de l'acide thymidilique (dT). Le ddTTP ne comportant pas d'hydroxyle en 3', la chaîne ne peut s'allonger et la synthèse se termine chaque fois qu'un dT devrait être ajouté.

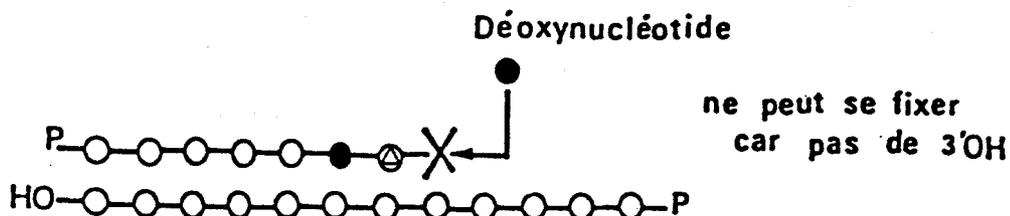
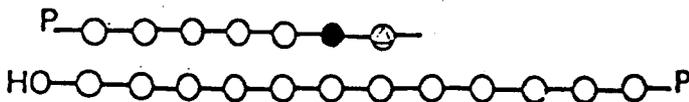
Si une "amorce" et une molécule d'ADN simple-brin sont incubés avec l'ADN polymérase I en présence d'un mélange de ddTTP et dTTP ainsi que des trois autres dXTP (l'un d'entre eux marqué au P<sup>32</sup>), des fragments ayant tous la même extrémité 5' et des résidus ddT en 3' sont obtenus (figure 2).



**Extension de la chaîne :**  
incorporation du déoxynucléotide



**Terminaison de la chaîne :**  
incorporation du didéoxynucléotide



**Figure 2 :** Principe de la méthode des terminaisons de chaînes.

Quand cette population de molécules est fractionnée sur gel d'acrylamide dénaturant (contenant de l'urée), la répartition des bandes en autoradiographie montre la distribution des dT dans l'ADN synthétisé.

En utilisant le même principe pour les autres nucléotides au cours d'incubations séparées, puis en faisant migrer les échantillons en parallèle sur gel d'acrylamide, une succession de bandes est obtenue. Elle permet de lire directement la séquence complémentaire de celle de la molécule modèle.

Le point délicat de toutes les méthodes de séquence d'ADN est l'obtention d'ADN simple-brin à partir de l'ADN bicaténaire initial.

Dans la méthode des terminaisons de chaînes, le clonage du fragment d'ADN à séquencer dans le bactériophage M 13 simplifie cette étape.

### 3. Clonage du fragment d'ADN à séquencer dans le bactériophage M 13.

Cette méthode (MESSING *et al.*, 1977) utilise comme vecteurs de clonage des bactériophages dérivés du bactériophage M 13.

#### a. Cycle évolutif du bactériophage M 13.

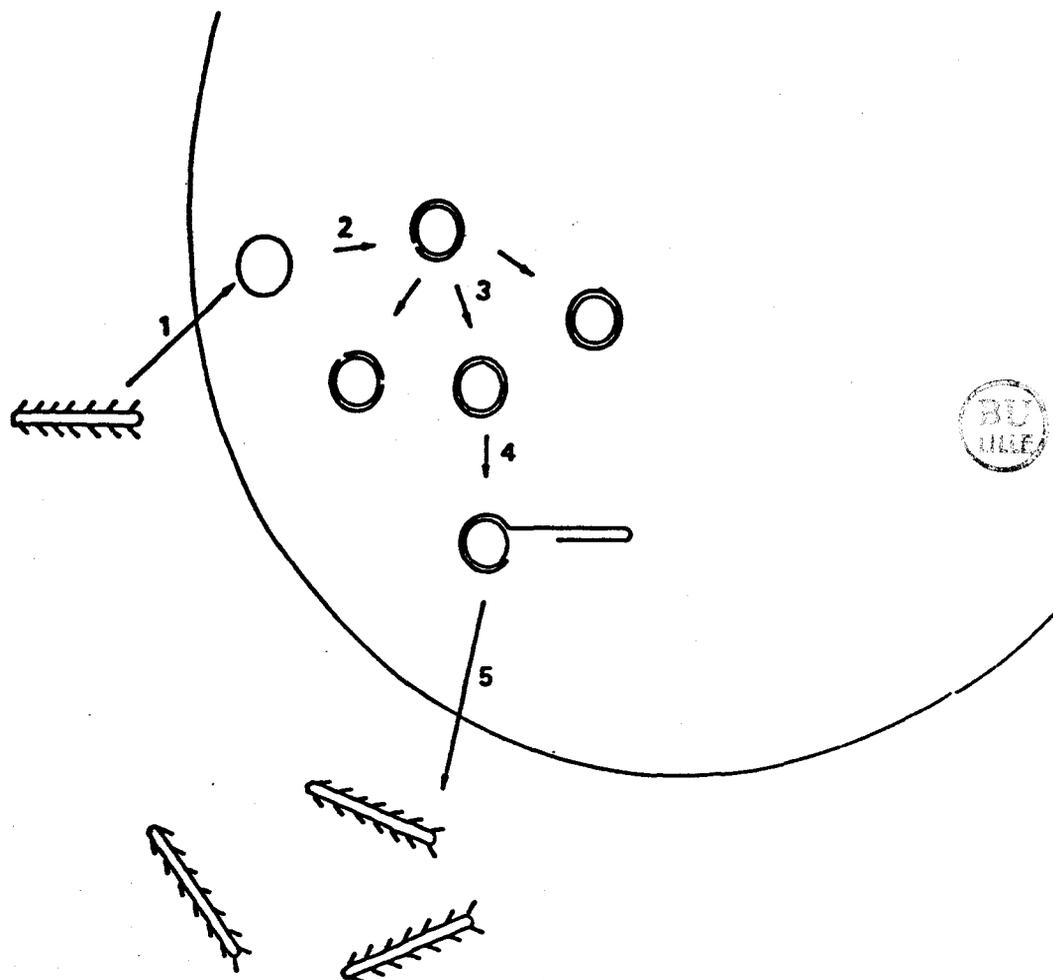
M 13 est un phage filamenteux dont l'ADN monocaténaire et circulaire est constitué de 6500 nucléotides. Ce phage infecte *E. Coli*, mais il n'infecte que les bactéries mâles car il s'attache à l'épisome de la bactérie.

Après pénétration dans la bactérie, l'ADN phagique simple-brin est converti en forme répliquative (RF) double-brin qui peut être isolée et utilisée comme vecteur de clonage.

Lorsque 100 à 200 copies de forme répliquative (par bactérie) ont été synthétisées, la synthèse de M 13 devient asymétrique produisant en abondance l'un des deux brins (le brin +). Puis des phages à ADN monocaténaire sont excrétés dans le milieu de culture sans lyse de la bactérie (figure 3).

#### b. Caractéristiques du bactériophage M 13 utilisé en séquence.

Tout le génome du phage M 13 contient des informations nécessaires à la répliquaison à l'exception d'une région intergénique de 500 nucléotides. Dans les dérivés de M 13 construits comme véhicules de clonage, deux types de séquence sont insérés :



1. Pénétration de l'ADN du phage dans la bactérie.
2. ADN phagique monocaténaire converti en forme réplivative double-brin RF.
3. Réplication de l'ADN (environ 100 copies par cellule).
4. Synthèse d'un ADN simple-brin.
5. ADN simple brin enfermé dans une capside virale et expulsé de la cellule.

**Figure 3** : Cycle évolutif du phage M 13.

- un fragment de l'opéron lactose d'*E. Coli* contenant la région régulatrice et l'information codant pour les 146 amino-acides N-terminaux de la bêta-galactosidase. Ce fragment amino-terminal de la bêta-galactosidase complémente la bêta-galactosidase défective de la bactérie hôte (on utilise une souche d'*E. Coli* particulière, défective pour la bêta-galactosidase, JM 101 ou JM 103). Ceci produit une bêta-galactosidase active qui en présence de l'inducteur isopropylthiogalactoside (IPTG) peut hydrolyser le chromogène qu'est le 5 bromo-4-chloro-3-indolyl-bêta-galactoside (BGIG). Le bromochloro-indole libéré confère alors une coloration bleue aux plages de phages.

- le second type de séquence inséré est un petit fragment d'ADN contenant plusieurs sites de restriction servant au clonage (figure 4). Ce fragment est inséré en 5' du gène bêta-galactosidase mais n'empêche pas la bêta-galactosidase du phage de complémenter celle d'*E. Coli*. Cependant, l'insertion dans l'un des sites de restriction de l'ADN à séquencer détruit la complémentation. L'hydrolyse du BCIG en présence d'IPTG est alors impossible et les plages de phages sont incolores.

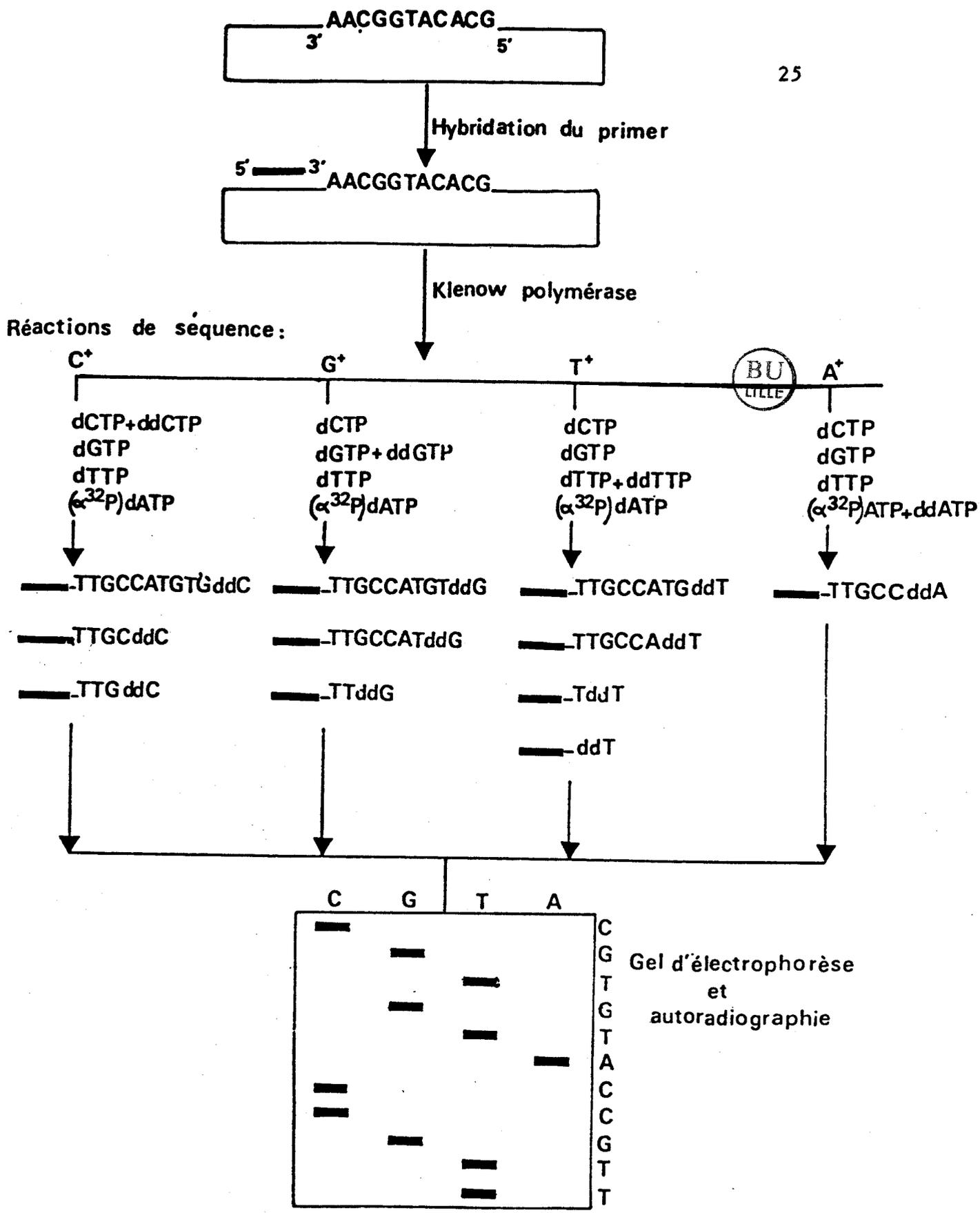
### c. La méthode dite "M 13 dideoxy sequencing" (PONCZ et al., 1982).

A l'ADN phagique contenant le fragment d'ADN à séquencer est hybridé un oligonucléotide de synthèse simple-brin ("amorce" universelle) qui sert de point de départ à la réplication par la Klenow polymérase (large fragment de l'ADN polymérase I) (protocole explicité dans la figure 4).

Les phages M 13 mp 8 et mp 9 contiennent les mêmes sites de restriction servant au clonage mais dans des orientations inversées (figure 5a). Donc le clonage d'un même fragment d'ADN dans M 13 mp 8 et mp 9 permettra d'obtenir la séquence de chacun des deux brins du fragment en question (figure 5b).

## II La méthode des dégradations chimiques de MAXAM et GILBERT.

Cette méthode (MAXAM et GILBERT, 1977) détermine la séquence nucléotidique d'une molécule d'ADN simple-brin marquée à son extrémité 5' en la modifiant et la coupant par des agents chimiques en A, G, C, et T ; l'ADN est donc séquencé directement et non plus répliqué *in vitro*.



Séquence lue : 5'-TTGCCATGTGC-3'

Figure 4 : Principe des réactions de séquence.

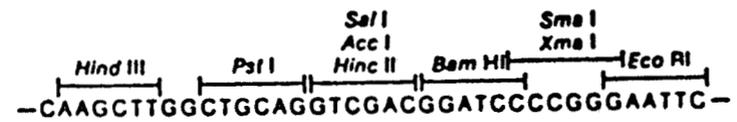
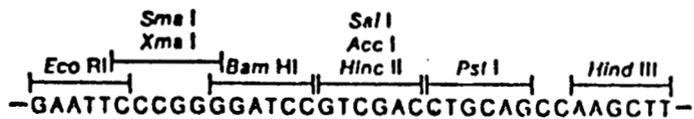
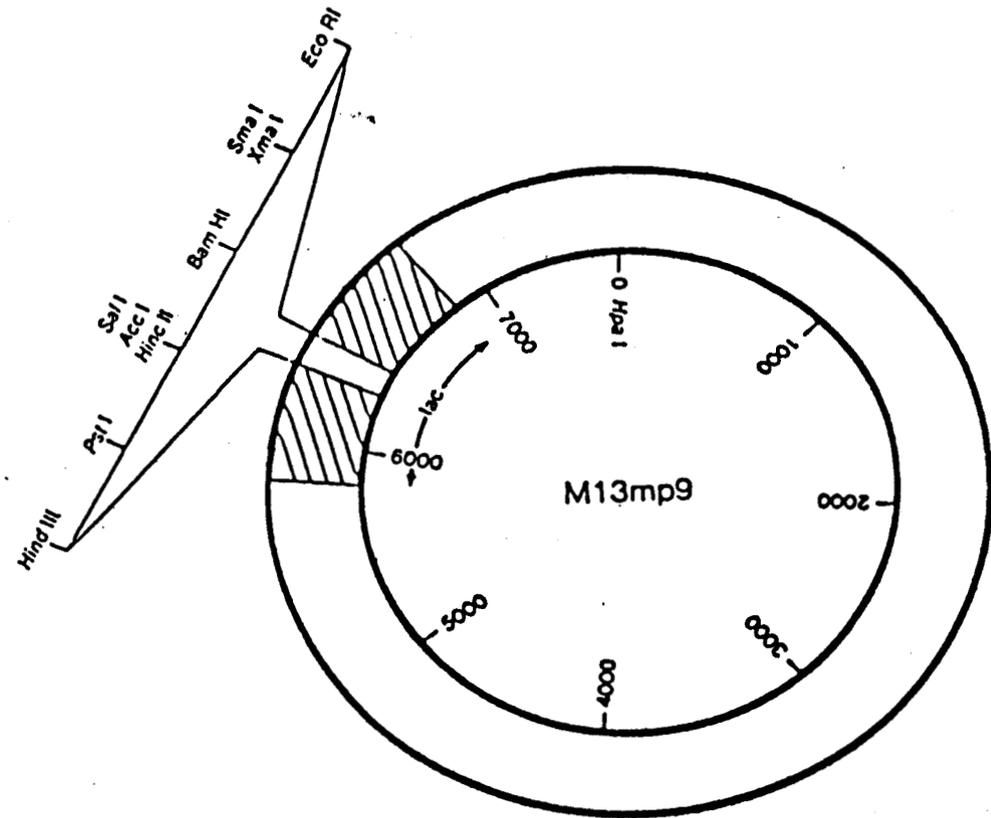
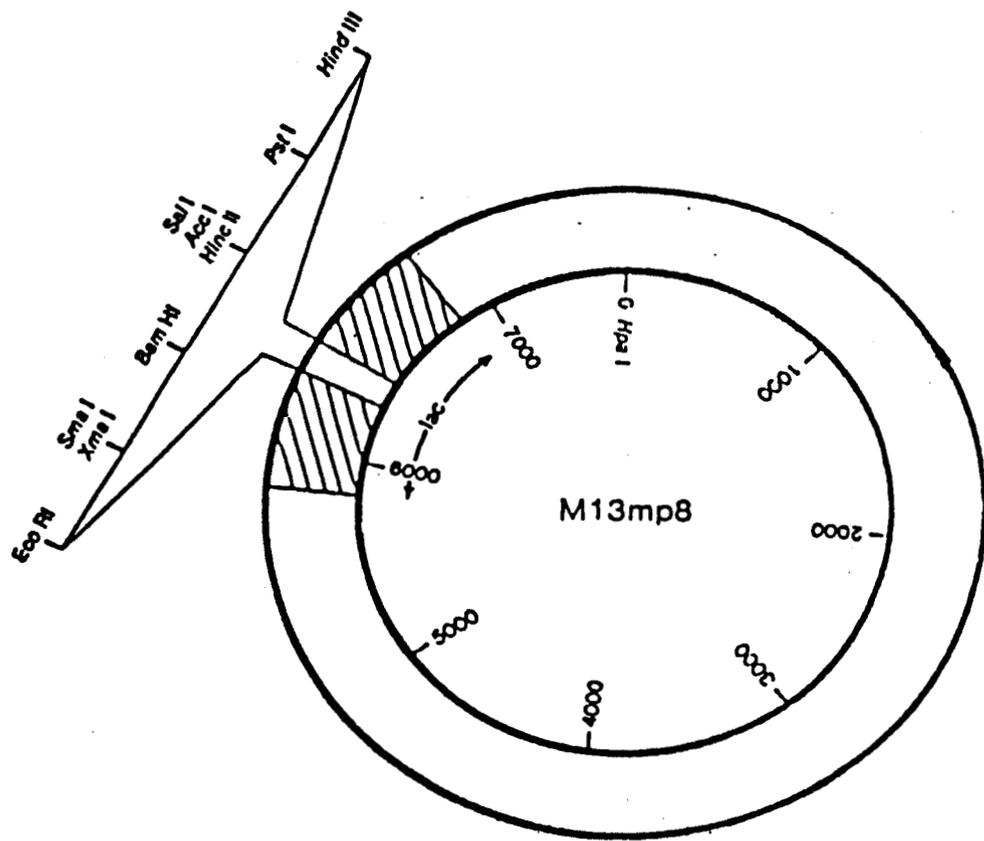
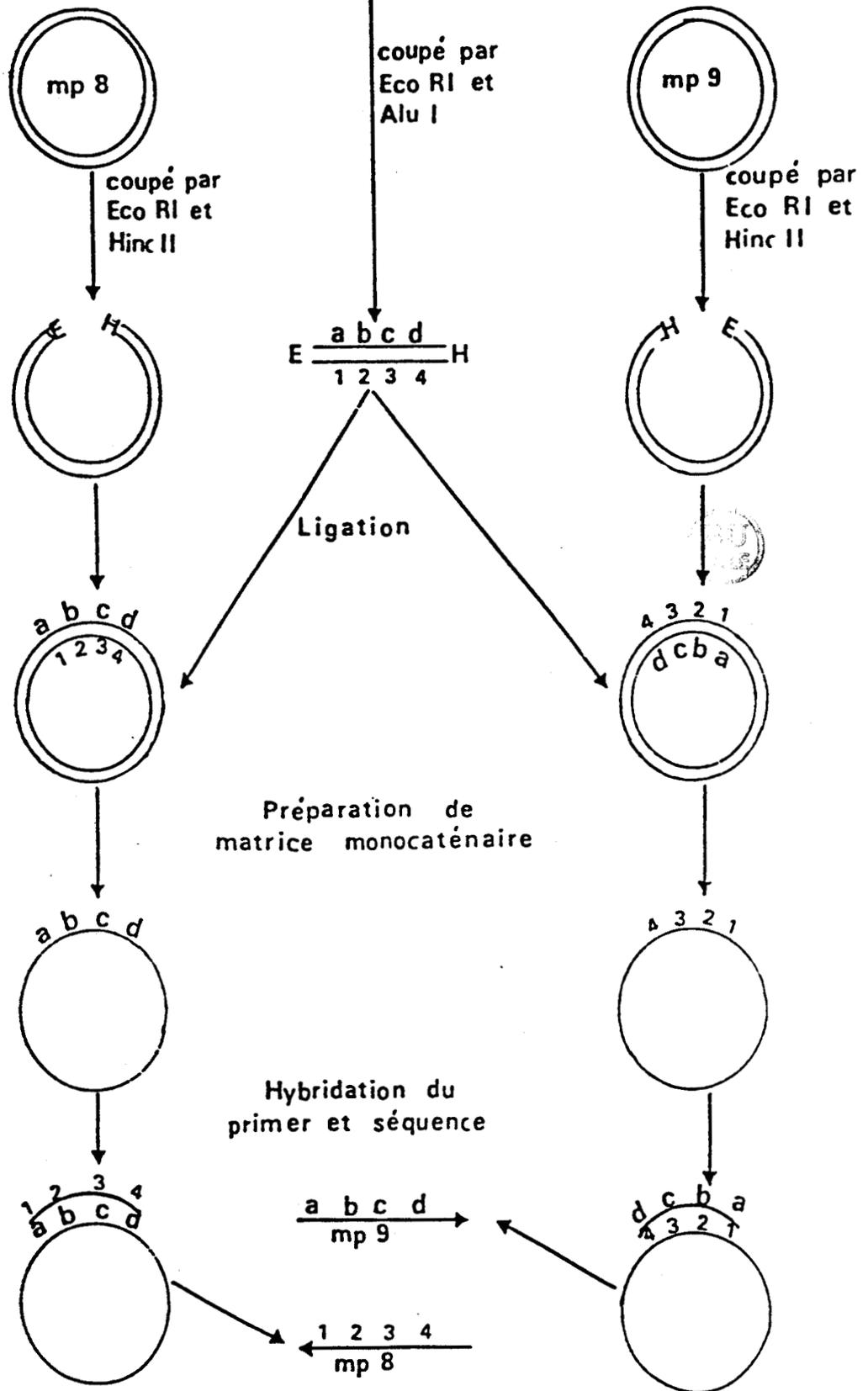


Figure 5a

ADN à séquencer



**Figure 2b :** Exemple d'utilisation de M13 mp8 et mp9 pour séquencer les deux brins d'ADN d'un même fragment.

Le fragment d'ADN double-brin dont la structure primaire doit être déterminée est tout d'abord marqué à ses extrémités 5' au P<sup>32</sup>. Le marquage comprend deux étapes : l'action de la phosphatase alcaline d'*E. Coli* qui élimine les 5' phosphates, puis leur remplacement par un phosphate radioactif grâce à l'action de la polynucléotide kinase du **phage T4** en présence d'ATP-  $\gamma$ -P<sup>32</sup>.

Cette molécule d'ADN, marquée à ses extrémités 5' est ensuite dénaturée par la chaleur en deux molécules d'ADN simple-brin.

Cet ADN simple-brin est alors attaqué dans des conditions particulières par des réactifs qui modifient spécifiquement les bases (figure 6). La modification de celle-ci entraîne une rupture de la liaison N-glycosidique laissant ainsi le sucre exposé ; ce dernier représente un point de fragilité dans l'enchaînement sucre-phosphate qui est facilement rompu par une amine, en général la pipéridine (figure 6a).

Environ une base sur cinquante ou cent, distribuée de façon aléatoire le long de la chaîne d'ADN, est modifiée. Par contre la rupture de la liaison sucre-phosphate est totale.

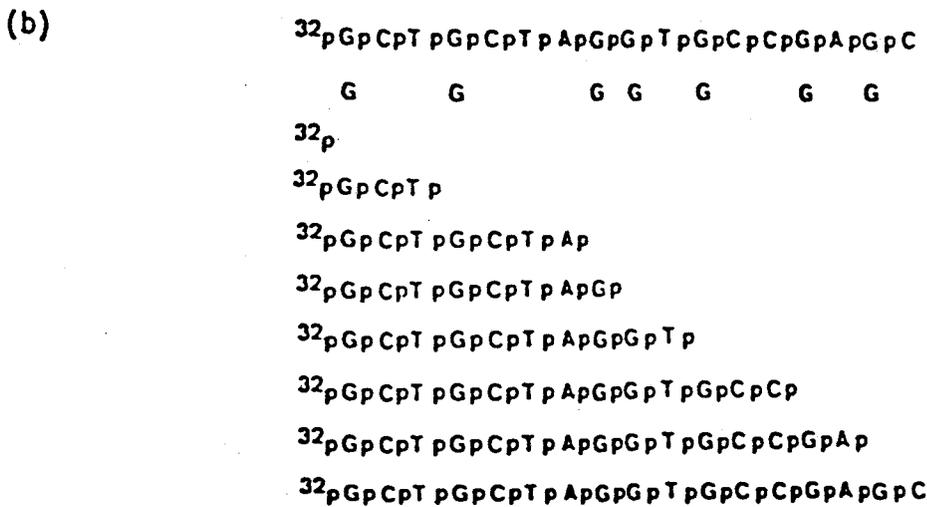
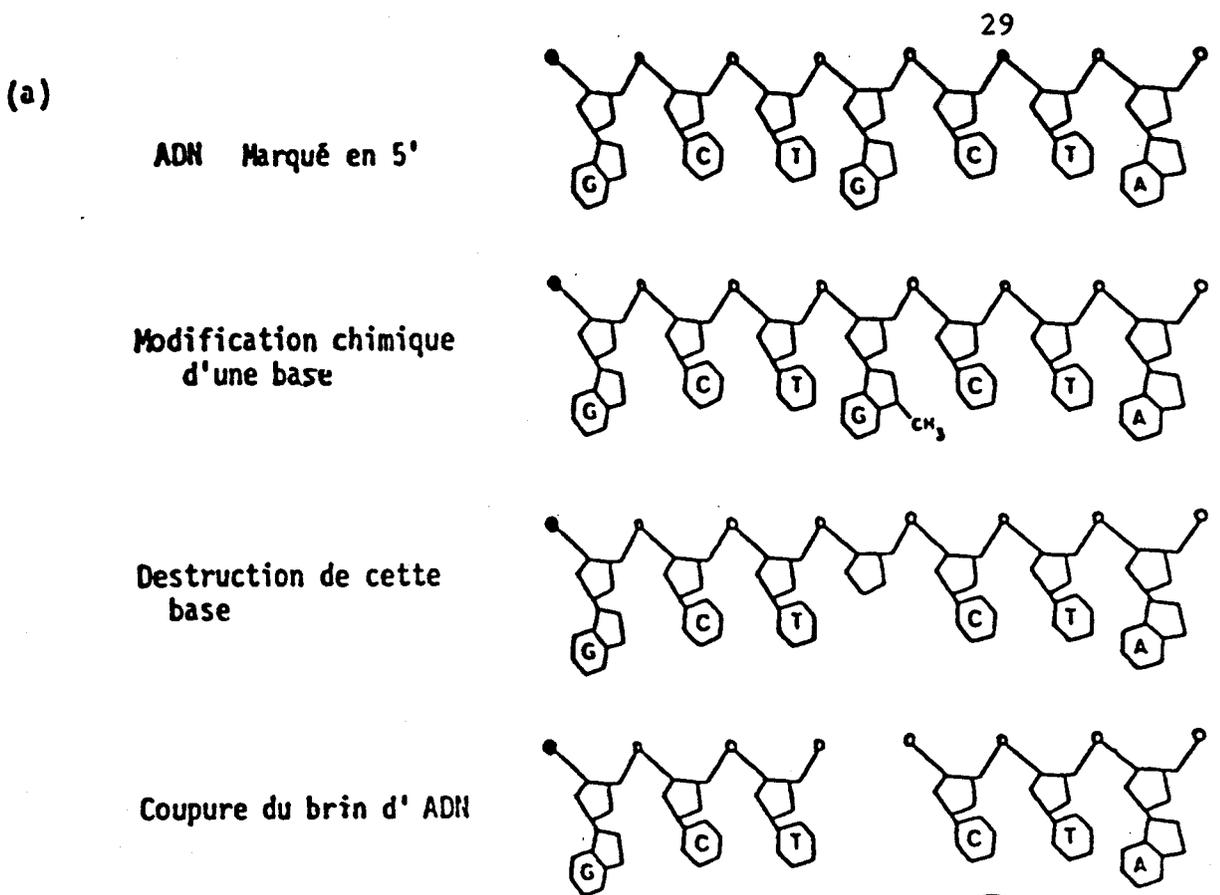
De toutes les molécules ainsi produites, seules celles commençant à l'extrémité radioactive sont détectées par autoradiographie (figure 6b). Le diméthylsulfate est le réactif spécifique des purines tandis que l'hydrazine est le réactif spécifique des pyrimidines.

Des modalités opératoires permettent de discriminer A de G d'une part, et C de T d'autre part. Ainsi dans le cas de réactions à l'hydrazine, la présence de NaCl 5M limitera la réaction aux cytosines. La liaison N-glycosidique de l'adénine, moins stable que celle de la guanine, sera facilement rompue par un acide dilué. En général, cinq réactions sont effectuées : G, AG, CT, C, A/C.

Le fractionnement en parallèle du produit de ces cinq réactions permet, après autoradiographie de lire directement la séquence nucléotidique du fragment étudié (figure 7).

L'utilisation de gels de concentration différente en acrylamide (25 %, 16 %, 8 %, 6 %) donne l'enchaînement de 200 à 250 nucléotides à partir de l'extrémité 5' marquée.

Le travail que nous présentons dans ce mémoire a été réalisé par la méthode des dégradations chimiques de MAXAM et GILBERT.



**Figure 6** : (a) Réactions coupant la molécule d'ADN au niveau d'une guanine.

(b) Après coupure au niveau de la guanine, on obtient une série de molécules ayant une extrémité radioactive donc détectables par autoradiographie.

Réactions	G	AG	CT	C	A/C
<b>G</b> →	—	—			
<b>A</b> →	—				—
<b>T</b> →			—		
<b>C</b> →			—	—	.....



**Figure 7** : Principe de la lecture d'un gel de séquence en fonction du profil électrophorétique.

**CHAPITRE III.**

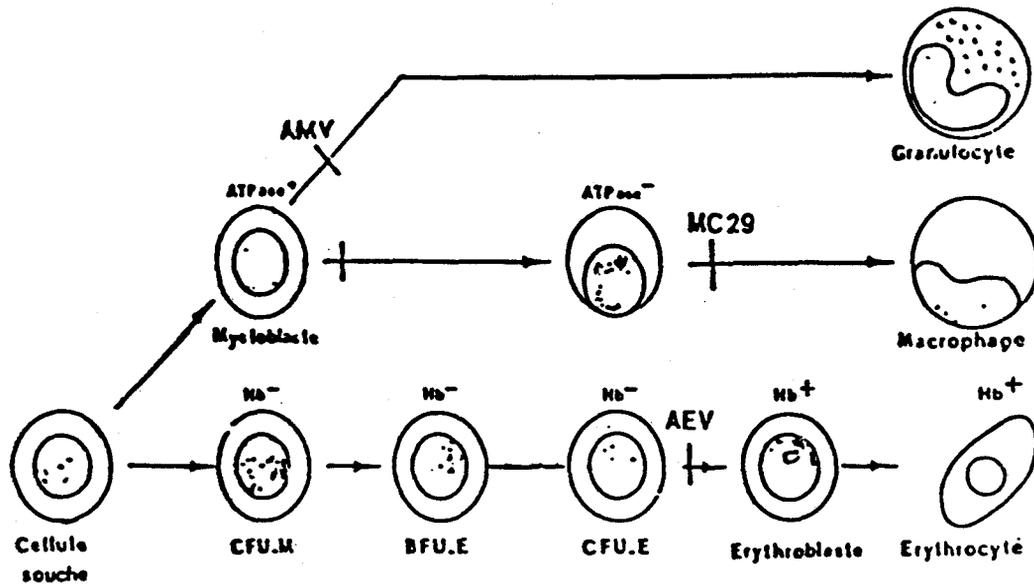
**PRESENTATION DU TRAVAIL**

Les virus défectifs des leucémies aviaires (DLV) possèdent la propriété de transformer différentes cellules cibles du système hématopoïétique entraînant la mort de l'animal en quelques semaines. Ces DLV, composés de sept isolats indépendants provenant de leucémies naturelles du poulet, ont été subdivisés en trois groupes en fonction du type de néoplasme induit et du phénotype de différenciation des cellules hématopoïétiques transformées (BEUG *et al.*, 1979 ; GRAF et STEHELIN, 1982). Ils comprennent le virus de l'érythroblastose aviaire (AEV), les virus de la myélocytomatose (MC 29, MH2, CM II et OK 10) et les virus de la myéloblastose (AMV et E26) qui transforment respectivement les érythroblastes, les macrophages et les myéloblastes (figure 8).

Un intérêt considérable concerne la découverte récente du fait que deux oncogènes distincts peuvent agir en synergie/complémentation pour transformer des fibroblastes embryonnaires de rat (LAND *et al.*, 1983). Certains rétrovirus, tels que AEV, MH2 et E26 (VENNSTROM et BISHOP, 1982 ; COLL *et al.*, 1983a ; LEPRINCE *et al.*, 1983) sont intéressants de ce point de vue puisqu'ils contiennent deux séquences spécifiques (oncogènes) d'origine cellulaire, qui ont probablement conduit à leur sélection comme virus hautement transformants

Le virus de l'érythroblastose aviaire (AEV), objet de nos travaux, est un virus défectif des leucémies qui induit des érythroblastoses et des sarcomes chez le poulet et transforme les érythroblastes et les fibroblastes *in vitro* (GRAF et BEUG, 1978 ; ROYER-POKORA, 1978 ; SAMARUT et GAZZOLO, 1982). Le génome de ce virus contient deux oncogènes indépendants v-erbA et v-erbB (LAI *et al.*, 1979 ; BISTER et DUESBERG, 1979) exprimés dans les cellules transformées sous forme de deux ARN messagers (ARNm) (ANDERSON *et al.*, 1980 ; SCHEINESS *et al.*, 1981).

Un ARNm génomique de 5,4 kilobases (kb) code pour une protéine de fusion P 75 gag-erbA (masse moléculaire 75 000 daltons) de localisation cytoplasmique (LAI *et al.*, 1980 ; PAWSON *et al.*, 1980b). Un ARNm sous-génomique de 3,5 kb produit une glycoprotéine membranaire gp74 erbB (masse moléculaire 74 000 daltons) (PRIVALSKY et BISHOP, 1982 ; PRIVALSKY *et al.*, 1983 ; HAYMAN *et al.*, 1983 ; HAYMAN et BEUG, 1984).



**Figure 8 :** Voies de différenciation des cellules hématopoétiques chez les oiseaux et cellules cibles des rétrovirus aviaires (d'après GRAF et BEUG, 1978 ; GAZZOLO et al., 1980).

→ sens de la différenciation  
 + → blocage de la différenciation

CFU.M - "colony forming unit, marrow"  
 BFU.E - "burst forming unit, erythroid"  
 CFU.E - "colony forming unit, erythroid".

Des expériences impliquant des mutants de délétion en chacun des gènes v-erbA et v-erbB (FRYKBERG *et al.*, 1983) ont montré que v-erbB est responsable des principales propriétés transformantes du virus : un mutant v-erbA<sup>+</sup>B<sup>-</sup> est incapable de transformer les cellules en culture. Par contre, un mutant v-erbA<sup>-</sup>B<sup>+</sup> transforme les érythroblastes qui présentent alors un large spectre de différenciation pouvant aller jusqu'au stade érythrocyte (SAMARUT et GAZZOLO, 1982). Le virus de type sauvage (v-erbA<sup>+</sup>B<sup>+</sup>) transforme les érythroblastes en culture et l'on note que leur différenciation est étroitement bloquée au stade CFU-E (GAZZOLO *et al.*, 1980) (Tableau IV). On peut donc en conclure que v-erbA, bien que non transformant, potentialise l'action de v-erbB en bloquant la différenciation des érythroblastes à un stade immature (GAZZOLO *et al.*, 1980). Des expériences réalisées par SAULE *et al.* (communication personnelle) ont montré que les exigences des cellules infectées par AEV, en matière de pH et de concentration en ions sodium, sont moins strictes que celles de cellules infectées par des mutants ne comportant que le gène v-erbB. La protéine erbA pourrait donc être impliquée dans le contrôle intracellulaire du pH et des flux ioniques.

En 1983, une équipe japonaise (YAMAMOTO *et al.*, 1983a) déterminait la séquence nucléotidique de l'oncogène v-erbB contenu dans un mutant de AEV, AEV-H, dépourvu du gène erbA. La séquence de la protéine erbB déduite de la séquence nucléotidique présente une homologie avec la protéine pp60<sup>src</sup>. La protéine erbB fait partie de la famille des protéines oncogéniques ayant une activité tyrosine-kinase (activité mise en évidence depuis peu pour la protéine gp 74 erbB, GILMORE *et al.*, 1985 ; KRIS *et al.*, 1985). Par ailleurs, cette protéine présente une très forte homologie avec la moitié C-terminale du récepteur du facteur de croissance de l'épiderme (EGF : Epidermal Growth Factor) (DOWNWARD *et al.*, 1984 ; ULLRICH *et al.*, 1984).

Les gènes v-erbA et v-erbB sont homologues à des séquences cellulaires c-erbA et c-erbB (SAULE *et al.*, 1981 ; SERGEANT *et al.*, 1982 ; VENNSTROM *et al.*, 1982) que l'on trouve chez les oiseaux et les mammifères. Ces gènes sont phylogénétiquement conservés et pourraient jouer un rôle important dans la différenciation des cellules rouges (COLL *et al.*, 1983b). Dans l'ADN cellulaire du poulet, le gène c-erbB comporte 12 exons et s'étend sur 20 kilobases d'ADN (SERGEANT *et al.*, 1982). Chez l'homme, ce locus a été assigné au chromosome 7 (SPURR *et al.*, 1984).

L'équivalent cellulaire du gène v-erbA, chez le poulet et chez l'homme, se présente sous forme de deux loci indépendants mis en évidence à l'aide de sondes virales radioactives (JANSSON *et al.*, 1983 ; SAULE *et al.*,

(	:	:	)
(	:	: Type de cellules transformées <b>in vitro</b>	
(	:	:	)
(	Virus	Erythroblastes	Fibroblastes
(	:	:	)
(	-----	-----	-----
(	:	:	)
(	v-erbA+B <sup>-</sup>	-	-
(	:	:	)
(	:	:	)
(	v-erbA <sup>-</sup> B <sup>+</sup>	+	+
(	:	: large spectre de	
(	:	: différenciation	
(	:	:	)
(	:	:	)
(	v-erbA+B <sup>+</sup>	+	+
(	(AEV)	: blocage au stade	
(	:	CFU-E	:
(	:	:	)
(	:	:	)

TABLEAU IV

communication personnelle). Chez l'homme, ces loci ont été localisés sur le chromosome 17 (SPURR *et al.*, 1984).

Les particularités du gène erbA, sa fonction dans le virus et son organisation sous forme de deux loci indépendants dans l'ADN cellulaire du poulet et de l'homme nous ont incités à en déterminer la structure primaire. Dans le chapitre "résultats", nous ferons donc figurer les séquences nucléotidiques du gène v-erbA et de ses équivalents cellulaires chez le poulet et chez l'homme. Nous nous sommes également intéressés à l'organisation du génome viral en étudiant la partie 5' du gène v-erbB et son équivalent cellulaire chez le poulet.

Notre travail, qui relève d'une collaboration avec le laboratoire d'Oncologie moléculaire de l'Institut Pasteur de Lille (Directeur : D. STEHELIN) représente donc une contribution à l'étude de la structure fine des oncogènes viraux et de leurs équivalents cellulaires normaux, travail qui a été résumé dans deux articles que nous incluons à la fin de ce mémoire.

#### CHAPITRE IV

DETERMINATION DE LA STRUCTURE PRIMAIRE DE L'ONCOGENE v-erbA ET  
ORGANISATION DES GENES v-erbA ET v-erbB DANS AEV.

## 1. Stratégie de séquence.

Nous avons isolé le gène v-erbA à partir d'un clone moléculaire de AEV, pAEV<sub>11</sub>, biologiquement actif (VENNSTROM **et al.**, 1980). Trois sous-clones de ce gène ont été obtenus ; ils couvrent la partie de gag adjacente à v-erbA, l'oncogène v-erbA lui-même ainsi que la partie 5' du gène v-erbB. Ces trois sous-clones (AvaI - AvaI : 1 kbp, SalI - SacI : 0,5 kbp, SacI - BamHI : 0,7 kbp) sont schématisés dans la figure 9A. Nous avons déterminé leur séquence nucléotidique par la méthode des dégradations chimiques de MAXAM et GILBERT, décrite dans le chapitre B. La stratégie de séquence est résumée dans la figure 9A.

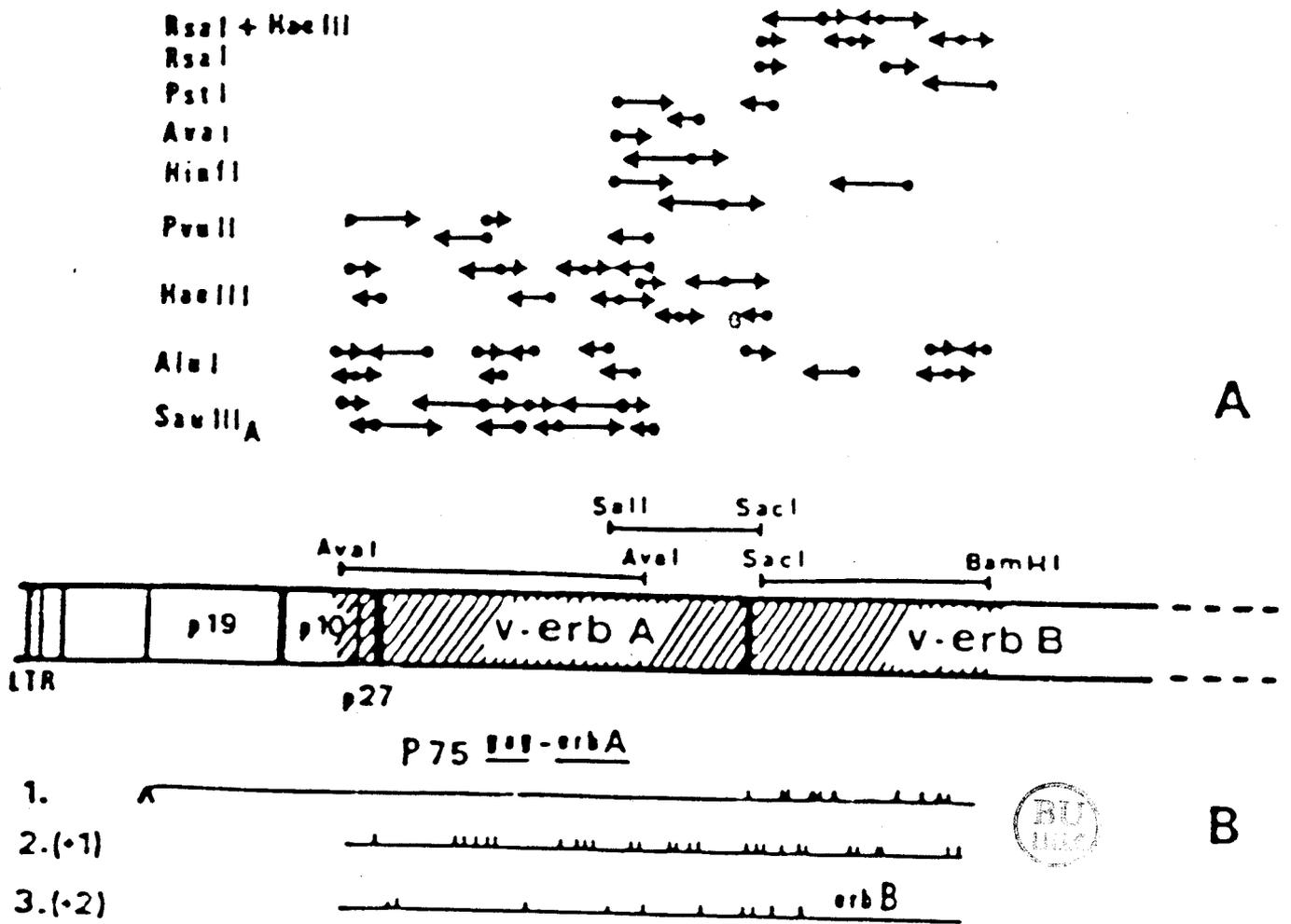
## 2. Résultats et discussion.

### a. Séquence de v-erbA.

L'extrémité 5' de v-erbA a été définie par comparaison de notre séquence (figure 10) avec la séquence nucléotidique du virus du sarcome de ROUS (RSV) (SCHWARTZ **et al.**, 1983).

Le site AvaI, à l'extrémité 5' de notre séquence se situe dans la région codant pour la protéine virale P10 gag. Les séquences de RSV et AEV sont virtuellement identiques de ce site (-171) jusqu'au nucléotide + 1 de AEV où une complète divergence des deux séquences marque le point d'insertion de erbA. Cette insertion se situe dans la région codant pour la protéine virale P27 gag (SCHWARTZ **et al.**, 1983). Le cadre de lecture de v-erbA est le même que celui de gag jusqu'au nucléotide 1195 où un codon ambre (TAG) marque la fin de la protéine codée P75 gag-erbA. Il est peu probable que les deux autres cadres de lecture soient utilisés, car ils sont fréquemment fermés par des codons de terminaison (figure 9 B). Le codon de terminaison de la protéine P75 gag-erbA est suivi de séquences non codantes jusqu'à un site d'épissage accepteur potentiel (MOUNT, 1982) (Sa dans la figure 10, nucléotide 1370) TTCCTTTTTGCAG.G probablement utilisé par le virus pour engendrer l'ARN sous-génomique de v-erbB.

Les séquences en amino-acides du polypeptide v-erbA et de la protéine v-erbB sont indiquées sous la séquence nucléotidique dans la figure 10. Le cadre de lecture utilisé par le virus pour engendrer la protéine v-erbB est différent du cadre de lecture de la protéine de fusion P75 gag-erbA.



**Figure 9 :** A - Stratégie de séquence de l'oncogène v-erb A.

B - Cadres de lecture de la séquence nucléotidique.

(-171) CCG CCG CCG GAA CAG CCA ACC GCG CAG CCA CCG CAC CCG CAG CAG CCG CCG -121  
 Arg Gly Ala Glu Gln Pro Arg Ala Glu Pro Gly His Ala Gly Gln Ala Pro  
 -50

**GAG** ←

UUG CCG CCG CUG AUC UAC UGG CCA AUG ATC ACU CAG CAG UTT CCG AGU AUC UGU CCG CCG GTG GTA UUU ATG CCG GTA GTG ATT AAG ACA CAG UUA CCG UCU UUU ACC CCG CCG CCG CCG CCG -1  
 Gly Pro Ala Leu Thr Asp Trp Ala Arg Ile Arg Glu Glu Leu Ala Ser Thr Gly Pro Pro Val Val Ala Met Pro Val Val Ile Lys Thr Glu Gly Pro Ala Trp Thr Pro Leu Glu Pro  
 -40

→ **erba**

GAG CAC ACT CCG UGG CUG GAT CCG AAG CAC AAA AGA AAG ACC ACC CAA TGT TTC CUG AAG ACC ACC ATG TCA CCG TAC ATC CCG AUC UUC CUG CAC AAA CAT GAA CAG UUC CUG CUG TGT 120  
 Glu Asp Thr Arg Trp Leu Asp Gly Lys His Lys Arg Lys Ser Ser Gln Cys Leu Val Lys Ser Ser Met Ser Gly Tyr Ile Pro Ser Cys Leu Asp Lys Asp Glu Gln Cys Val Val Cys  
 10 20 30 40

UUG CAC AAA CCG ACC CCG TAC CAC TAC CCG UCC ATC ACC UUC CAG CCG UCC UCC AAG ACC TTT TTC CCG UUG ACC ATC CAG AAG AAC CUG UAC CCG ACC TAC UCC UCC AUG TAC GAT CCG UCC 240  
 Gly Asp Lys Ala Thr Gly Tyr His Tyr Arg Cys Ile Thr Cys Glu Gly Cys Lys Ser Phe Phe Arg Arg Thr Ile Gln Lys Asn Leu His Pro Thr Tyr Ser Cys Thr Tyr Asp Gly Cys  
 50 60 70 80

UUC CUC ATC CAC AAG ATC ACC CCG AAC CAG UUC CAG CUG UCC CUC TTC AAG AAG TTC ATC UCC CUG UUC ATG CCG ATG CAC CUG CUG CUG CAC UAC UGG AAG CCG GTA CCG AAG CCG AAG 360  
 Cys Val Ile Asp Lys Ile Thr Arg Asn Gln Cys Gln Leu Cys Arg His Lys Lys Lys Ile Ser Val Gly Met Ala Met Asp Leu Val Leu Asp Asp Ser Lys Arg Val Ala Lys Arg Lys  
 90 100 110 120

CTG ATC CAG CAG AAC CCG CAG UUG CCA CCG AAG CAG CAG ATG ATC AAA UCC CUG CAG CAC CCG CCG AUC CCG AUC CCA CAG CAG CCG CAG CUG ATC CAC CUG CUG AUC CAG CCG CAC CCG 480  
 Leu Ile Glu Glu Asn Arg Glu Arg Arg Arg Lys Glu Glu Met Ile Lys Ser Leu Gln His Arg Pro Ser Pro Ser Ala Glu Glu Trp Glu Leu Ile His Val Val Thr Glu Ala His Arg  
 130 140 150 160

ACC ACC AAC CCG CAG CCG ACC CAC UUG AAG CAG AUG AAG AAA TTC CUG CTC GAA GAT ATC CCG CAG UCG CCG ATG CCG UCC ATG CTT GAC CCG CAC AAA CUG CAC CUG CAG CCG TTC ACC 600  
 Ser Thr Asn Ala Gln Gly Ser His Trp Lys Gln Arg Arg Lys His Leu Leu Glu Asp Ile Gly Gln Ser Pro Met Ala Ser Met Leu Asp Gly Asp Lys Val Asp Leu Glu Ala Phe Ser  
 170 180 190 200

**Domain1** ← **Domain2**

GAG TTT ACA AYA ATC ATC ACC CCG CCG ATC ACC CUC CUG CUC CAC TTT UCC AAA AAC CUG CCG ATG TTC CCG CAG CUG CCG UCC CAG CAT CAG ATC ATC CUG CUG AAG CCG UCC UUC ATG 720  
 Glu Phe Thr Lys Ile Ile Thr Pro Ala Ile Thr Arg Val Val Asp Phe Ala Lys Asn Leu Pro Met Phe Ser Glu Leu Pro Cys Glu Asp Gln Ile Ile Leu Leu Lys Gly Cys Cys Met  
 210 220 230 240

GAG ATC ATG UCG CUG CCG CCG CCG CUG UUC TAC CAC CCG CAG ACC GAA ACC CUG ACC CUG ACC CCG GAA ATG CCG CUC AAA CCG CAG CAG TTG AAG AAC CCA CCG CUG CCG CUC CUG TCT 840  
 Glu Ile Met Ser Leu Arg Ala Ala Val Arg Tyr Asp Pro Glu Ser Glu Thr Leu Thr Leu Ser Gly Glu Met Ala Val Lys Arg Glu Gln Leu Lys Asn Gly Gly Leu Gly Val Val Ser  
 250 260 270 280

GAT CCG ATC TTC CAC CTC CCG AAG UUG CUG TCT UCC TTC AAC CUG CAC CAC ACC CAG CUG CCG CCG CUG CCG CUG CUG CUG CUG CUG CUG ATT CCG CCA CAC CCG ACC CCG CUG ATC UCC CUG CAT 960  
 Asp Ala Ile Phe Asp Leu Gly Lys Ser Leu Ser Ala Phe Asn Leu Asp Asp Thr Glu Val Ala Leu Leu Gln Ala Val Leu Leu Met Ser Ser Asp Arg Thr Gly Leu Ile Cys Val Asp  
 290 300 310 320

AAG ATA CAG AAG UCC CAG CAG UGG TAC CUG CUG CCG TTC CAG CAC TAC ATC AAC TAC CCG AAA CAC AAC ATT CCG CAC TTC UGG UCC AAG CUG CUG ATG AAG CUG CCG CAC CUG CCG ATG 1080  
 Lys Ile Glu Lys Cys Gln Glu Ser Tyr Leu Leu Ala His Glu His Tyr Ile Asn Tyr Arg Lys His Asn Ile Pro His His Trp Ser Lys Leu Leu Met Lys Val Ala Asp Leu Arg Met  
 330 340 350 360

ATC CCG CCG TAC CAC CCG ACC CCG TTC CUG CAC ATG AAG CUG CAG UCC CCG ACC CAG CUG CCG CCG ACC ACC UCG TGT ACA CCG CTA CAG ATT TTG CCA AGT ATA TTC CCG TTT GTA TAG UCC 1200  
 Ile Gly Ala Tyr His Ala Ser Arg His Leu His Met Lys Val Glu Cys Pro Thr Glu Leu Pro Pro Arg Arg Lys Arg Ala Leu Gln Ile Leu Gly Ser Ile Leu Pro Phe Val Asn  
 370 380 390 400

ATCCTACAGCAAAAACCTGTAGCATTTCAGTACCAAGTAATCAAAACAACAATCCAGAAATCCGACAGTCAATCAACATATTTATGTTCTGTGTTATCCCGAAGTGTAGCAATGTAGCAAAATCATACATTAGTTCTTT 1359

→ **erbB**

CGT TTT CCG CCG CAC CAC UCC ATG AAG TGT CCG CAT TTT ATA GAT UGT CCG CAC TGT CUG AAG CUC CCG 1481  
 Gly Pro Asp His Cys Met Lys Cys Ala His His Ile Asp Gly Pro His Cys Val Lys Gly Pro Ala Pro Leu Gly Ser Trp Val Arg Met Ile Pro Trp Ser Gly Ser Met  
 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200

CAG ATG CCA ATG CUG TTT CCG ACC TCT CCG ATC CAA ACT GTA CAC CAG TGT CCA ACC CAG CAG CUG CUG AAG GAT ACC CAA ATG CCG CCA AAA CTC CAT CTA UCG CCG CUG CUG CUG CCG 1601  
 Gln Met Pro Met Leu Phe Ala Ser Ser Ala Ile Gln Thr Val His Glu Gly Ala Arg Asp Gln Val Leu Lys Asp Ala Gln Met Ala Pro Lys Leu His Leu Ser Arg Leu Val Leu Ser  
 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200

CCA CCA CTC CUG UCC CUG GTT CUG GTT CCG CTA CCG ATC CCG CTT TAC CUG CCG CCA CCG CAT ATC CUG CCG AAG CCG ACC CUG CCG ACC CUG CUG CAA CAG ACC CAG CTT CUC GAA CCA 1721  
 Gly Gly Leu Leu Cys Leu Val Val Val Gly Leu Gly Ile Gly Leu Tyr Leu Arg Arg Arg His Ile Val Arg Lys Arg Thr Leu Arg Arg Leu Leu Gln Glu Arg Glu Leu Val Glu Pro  
 80 90 100 110 120 130 140 150 160 170 180 190 200

CTG ACA CCG ACT CCG CAG CCA CCA AAC CAG CCG CAC CUG AGA ATT TTA AAG GAA ACA GAA TTT AAA AAG CUC AAA GTT TTG CCG TTT CCA UCT TTT CCG ACT CTT TAT AAG CCA CTT UCG 1841  
 Leu Thr Pro Ser Gly Glu Ala Pro Asn Gln Ala His Leu Arg Ile Leu Lys Glu Thr Glu Phe Lys Lys Val Lys Val Leu Gly Phe Gly Ala Phe Gly Thr Val Tyr Lys Gly Leu Trp  
 120 130 140 150 160 170 180 190 200

Figure 10



Si le site d'épissage accepteur mentionné ci-dessus est bien utilisé par le virus pour engendrer la protéine v-erbB, cette protéine commence à l'AUG de gag tout comme la protéine P75 gag-erbA (si l'on admet que AEV et RSV utilisent le même site d'épissage donneur dans gag) ; les deux protéines (P75 gag-erbA et gp74 erbB) auraient donc quelques amino-acides identiques à leur extrémité N-terminale. Il est également possible d'imaginer que v-erbB ait son propre codon d'initiation et que son produit commence à l'AUG en position 1386 de notre séquence. La structure de v-erbB est virtuellement identique, dans la région que nous avons étudiée, à celle décrite pour un autre isolat indépendant de AEV, AEV-H, dépourvu du gène v-erbA (YAMAMOTO *et al.*, 1983b).

La protéine gp74 erbB présente une homologie avec la protéine pp60 src de RSV (YAMAMOTO *et al.*, 1983a) et une quasi similitude avec la moitié C-terminale du récepteur de l'EGF (DOWNWARD *et al.*, 1984 ; ULLRICH *et al.*, 1984).

b. Architecture du génome viral : organisation des gènes erb-A et erb-B dans AEV.  
-----

Nous avons comparé notre séquence à celle des gènes gag, pol, env de RSV et mis en évidence une homologie significative (82 %) entre la partie 3' de v-erbA (nucléotides 1151 à 1183) et une partie du gène env de RSV (nucléotides 5804 à 5836) (SCHWARTZ *et al.*, 1983) (figure 11). Le codon de terminaison de la protéine de fusion P75 gag-erbA (TAG en position 1195) n'appartient pas à cette séquence homologue. Ainsi, chose surprenante, les quatre amino-acides C-terminaux du polypeptide erbA sont codés par des nucléotides qui n'appartiennent pas au gène v-erbA.

Afin de définir l'origine de ces nucléotides et de façon à déterminer avec précision la limite exacte entre les deux oncogènes v-erbA et v-erbB dans AEV, nous avons isolé et séquencé un fragment d'ADN cellulaire de poulet correspondant à la partie 5' de v-erbB (NO<sub>4</sub> dans SERGEANT *et al.*, 1982) (figure 12). Nous avons comparé cette séquence à la séquence de v-erb (figure 13).

Cette comparaison a permis les observations suivantes :

- l'homologie entre v-erbB et c-erbB commence au nucléotide 1181 de la séquence virale. Ceci définit la limite 5' des séquences apparentées à c-erbB dans AEV ; par conséquent les quatre amino-acides C-terminaux de la

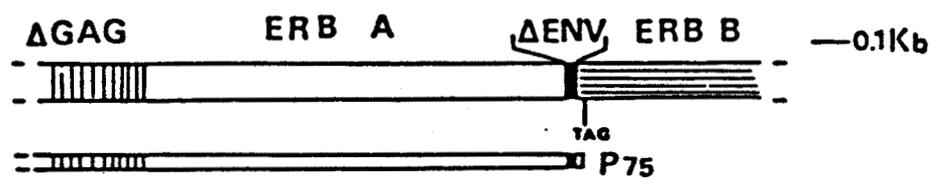
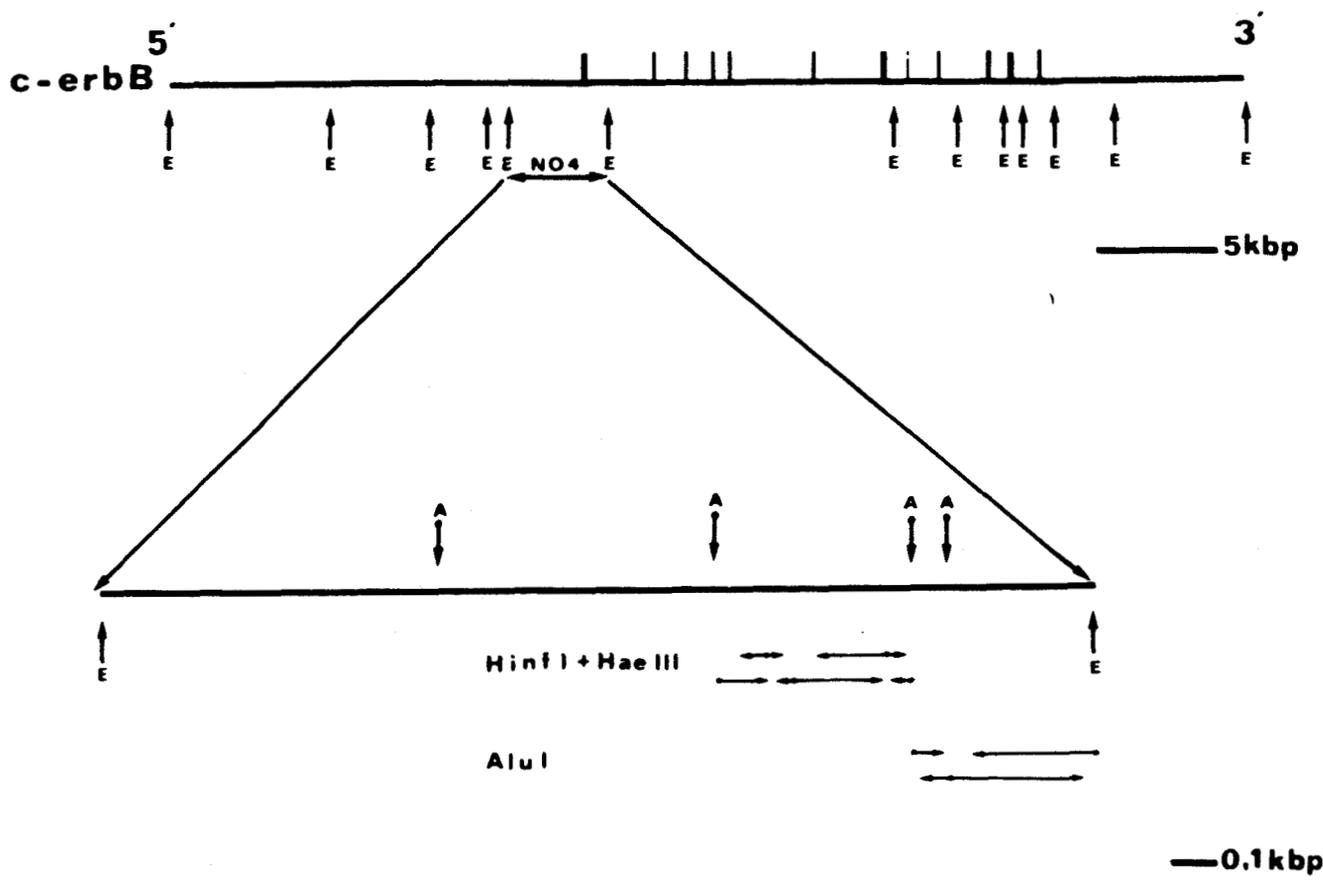


Figure 11





**Figure 12 :** Stratégie de séquence de la partie du locus *c-erb B* d'ADN cellulaire de poulet homologue à l'extrémité 5' de *v-erb B*. A la partie supérieure de la figure, les barres verticales représentent les exons du locus *c-erb B* (l'épaisseur de chacune de ces barres est proportionnelle au nombre de nucléotides de l'exon).

↑ E : EcoRI  
 ↓ A : AluI

```

      * * * * *
c-erbB  cttgaccatcagggcagggcaggaaaaaagagcagaagggaatcacgcagcaataatcccaacttgatgaaccattagtttaaacatca
v-erb 1289 CTACCACGCCAGCCGCTTCCCTGCACATGAAGGTGGAGTGCCECCACGGAGCTCCCCCAAGGAGGTGTAGAGCCCTACAGATTTGGGAAG
      Sac1

      1181
      **|
c-erbB  ggtattgccctttgtatagtcacatcctacaagagcaaaaacttgaagcatttcaggtagcaagtaatgaaccacaaaatgctgaaaa
v-erb 1179 TATATGCCCTTTGTATagtcacatcctacaagagcaaaaacttgaagcatttcaggtagcaagtaatgaaccacaaaatgctgaaaa
      fin de la protéine P75

      1306
      AEV-H|
c-erbB  tgctccacagttgaatgcacagatatttatgttctgtgtttattcctgaagtgttaagggaatgttaagggaatcatagattagttgtttctt
v-erb 1269 tgctccacagttgaatgcacagatatttatgttctgtgtttattcctgaagtgttaagggaatgttaagggaatcatagattagttgtttctt
REH ADN c .....TGCCTCAGGCCATGAACATCAC
      * * * * *

      1370
      * Sa Apa1
c-erbB  ttcttttttgcagGGCCCAGACCACTGCATGAAGTGTGCCCATTTTATAGATGGTCCCCACTGTGTGAAGGCCCTGCCCGCTGGGGTCTT
v-erb 1359 tccttttt-gcagGGCCCAGACCACTGCATGAAGTGTGCTCATTTTATAGATGGTCCCCACTGTGTGAAGGCCCTGCCCGCTGGGGTCTT
REH ADN c CTGCACAGGACGGGGACCAGACAACCTGTATCCAGTGTGCCCACTACATTGACGGC.....
      * * * * *

      1531
      Sd *****
c-erbB  GGGTGAGAATGATACCCTGGTCTGGAAGTATGCAGATGCCAATGCTGTTTCCAGCTCTGCCATCCAAACTGTACACGAGGgtgagtaaa
v-erb 1449 GGGTGAGAATGATACCCTGGTCTGGAAGTATGCAGATGCCAATGCTGTTTCCAGCTCTGCCATCCAAACTGTACACGAGGGTGAACGGG
REH ADN c .....CTGTGCCATCCAAACTGCACCTACGGGATGCCTG
      * * * * *

      * * * * *
c-erbB  accttcttccaaatggagagggaaaaagtgaaaagcagaggaacag
v-erb 1539 ACCAGGCTCTGAAGGATGCCCA---AATGGCTCCAAACTCCATC
REH ADN c GCCAGGCTCTGAAGGCTGTCCAACCAATGGGCCTAAGATCCCGTC
      * * * * *

```

**Figure 13 :** Séquence nucléotidique de la partie du locus c-erb B de l'ADN cellulaire de Poulet homologue à la région 5' de v-erb B : comparaison à la séquence de v-erb ainsi qu'à la séquence de l'ADN<sub>c</sub> du REH.

\* Les séquences codantes sont écrites en majuscules et les séquences non codantes en minuscules.

\* Sa et Sd indiquent les sites d'épissage (accepteur et donneur).

\* Les astérisques indiquent les différences entre les séquences virale et cellulaire (ligne supérieure) et entre v-erb et l'ADN<sub>c</sub> du REH (ligne inférieure).

\* La flèche en position 1181 indique le point de recombinaison entre c-erb B et AEV.

\* La flèche en position 1306 indique le point de recombinaison entre c-erb B et AEV-H (YAMAMOTO et al, 1983).

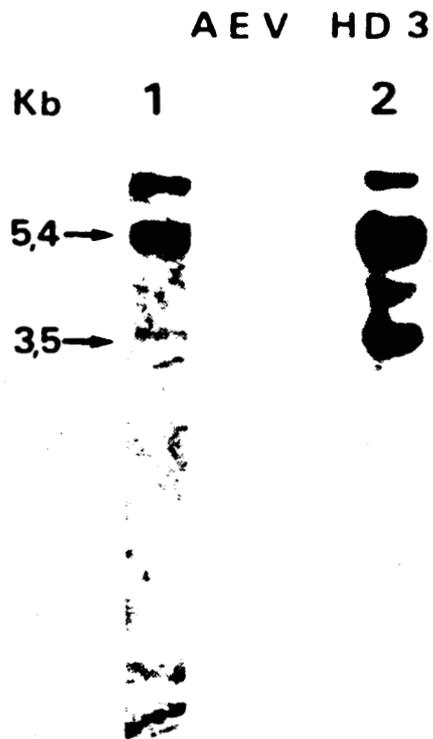
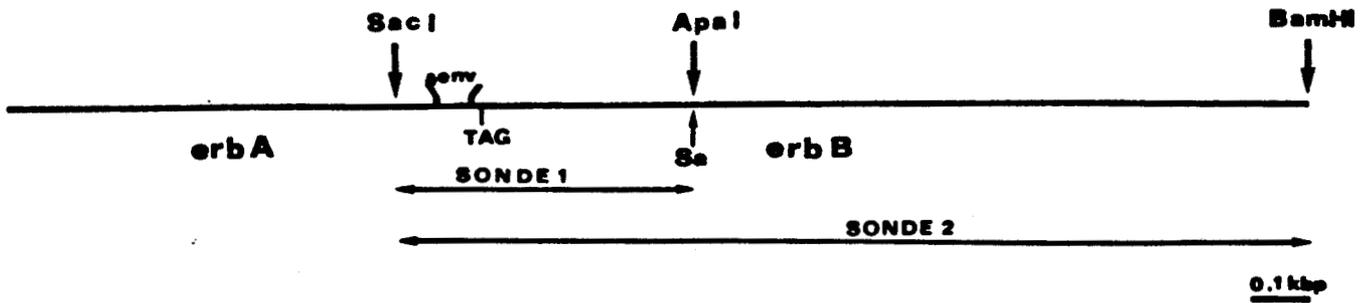
protéine P75 gag-erbA sont codés par des nucléotides dérivés de c-erbB qui produit également le codon de terminaison de cette protéine (nucléotides 1195-1197, figure 13), indiquant que AEV a acquis un locus c-erbA partiel.

- La séquence virale des nucléotides 1181 à 1370 correspond à une partie d'un intron de c-erbB fermé dans les trois cadres de lecture et se terminant par un site d'épissage accepteur TTTTCTTTTGCAG.G (MOUNT, 1982 (Sa dans la figure 13). Ce site d'épissage accepteur est probablement utilisé par le virus pour donner naissance à l'ARNm sous-génomique v-erbB. Cet ARNm n'hybride pas en "Northern blot" (figure 14, colonne 1, sonde 1) avec une sonde SacI-ApaI (sites de restriction indiqués dans la figure 13), synthétisée à partir d'un fragment d'ADN du provirus AEV (nucléotides 1139 à 1374 de v-erb). Comme contrôle (figure 14, colonne 2, sonde 2), nous montrons que cet ARNm hybride avec une sonde SacI-BamHI (nucléotides 1139 à 1841 de v-erb).

- Au nucléotide 1371 de v-erbB commence une forte homologie avec l'ADN complémentaire (ADNc) du récepteur de l'EGF humain (ADNc du REH) (nucléotide 1918, ULLRICH *et al.*, 1984). L'exon de c-erbB commençant au site d'épissage accepteur en position 1370 se termine par un site d'épissage donneur G.GTGAG (MOUNT, 1982) en position 1528 puisqu'en aval de ce site, la séquence de c-erbB diffère de celles de v-erbB et de l'ADNc du récepteur de l'EGF (en aval du nucléotide 2067, ULLRICH *et al.*, 1984).

Nous concluons que le gène v-erbB est une version tronquée de c-erbB commençant par un intron à l'extrémité 5'. La présence de cet intron de c-erbB dans le virus nous laisse supposer que la recombinaison entre AEV et le gène cellulaire eut lieu au stade ADN ou ARN prémessager ; mais comment et quand AEV a-t-il acquis le gène erbA ? Cela reste à déterminer : est-ce erbB avant erbA, ou erbA avant erbB, ou les deux gènes en même temps à partir d'une cellule où les deux gènes avaient été réunis au cours d'une translocation ? L'homologie détectée entre l'extrémité 3' de v-erbA et le gène env de RSV permet d'exclure la dernière hypothèse et suggère que deux événements de recombinaison successifs conduisirent à la capture des gènes erb par le virus.

Un mécanisme possible de capture de c-erbB par un ALV pourrait résulter de l'insertion d'un LTR en amont de ce gène comme cela a été décrit par FUNG *et al.* (1983). Une telle activation pourrait engendrer un rétrovirus tel que



**Figure 14**

- La sonde 1 ne met pas en évidence l'ARNm sous génomique de AEV (photographie, colonne 1)

- La sonde 2 met en évidence les deux ARNm de AEV (photographie, colonne 2).

AEV-H (YAMAMOTO *et al.*, 1983b) contenant erbB uniquement. Un second évènement de recombinaison, au stade ADN ou ARN, pourrait alors intervenir et aboutir à la capture de erbA. Finalement, un fait marquant de l'activation de c-erbB lié à la transformation paraît impliquer l'exon de c-erbB que nous avons séquencé et qui marque l'origine de l'homologie avec la moitié C-terminale du récepteur de l'EGF.

Les trois exemples connus d'activation de c-erbB : dans AEV, AEV-H et dans une érythroleucémie du poulet ("LTR-activation") concernent tous des insertions dans l'intron que nous avons décrit plus haut. Récemment, RAINES *et al.* (1985) ont montré que dans les érythroleucémies du poulet par LTR-activation de c-erbB, l'insertion du provirus se fait dans une région spécifique (de 0,5 kbp) située immédiatement en amont de l'exon où commence l'homologie avec v-erbB. Ceci vient corroborer nos travaux et conduit à penser que cet intron constitue un site privilégié de recombinaison aboutissant à l'activation de c-erbB. On peut imaginer que le récepteur de l'EGF tronqué que constitue la protéine v-erbB synthétisée dans les cellules transformées échappe à toute régulation en perdant le site de fixation de l'EGF (domaine extracellulaire) (figure 15).

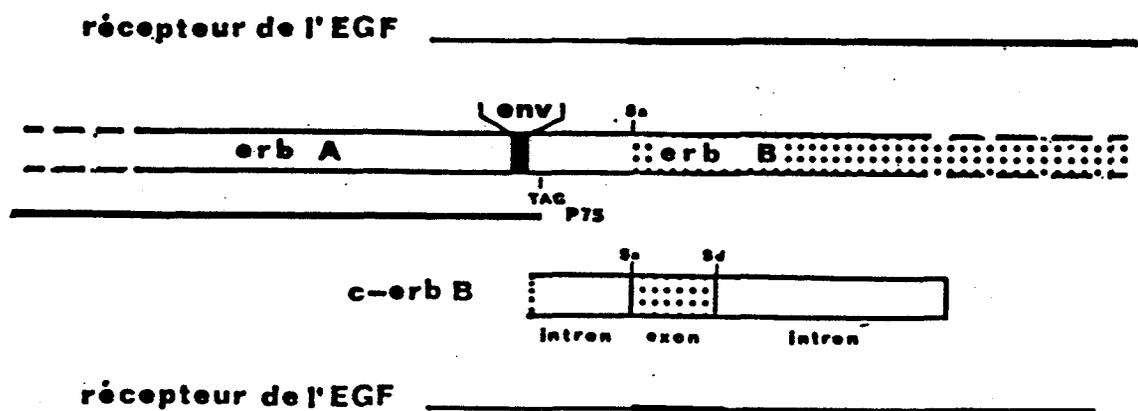
### c. Le polypeptide erbA

#### \* Caractéristiques

Le polypeptide erbA représente 398 résidus d'acides-amino de la protéine P75 gag-erbA.

Les protéines kinases AMP-cyclique dépendantes phosphorylent les sérines (Ser) et les thréonines (Thr) appartenant à des séquences X-Y-Z-(Ser ou Thr) (KREBS *et al.*, 1979) où X et Y sont des résidus basiques ; deux résidus Ser de ce type sont présents dans le polypeptide erbA en position 14 et 15. On ne dénote pas de site de glycosylation Asn-X-(Thr/Ser) (MARSHALL, 1974), il se pourrait donc que le polypeptide erbA soit phosphorylé mais non glycosylé.

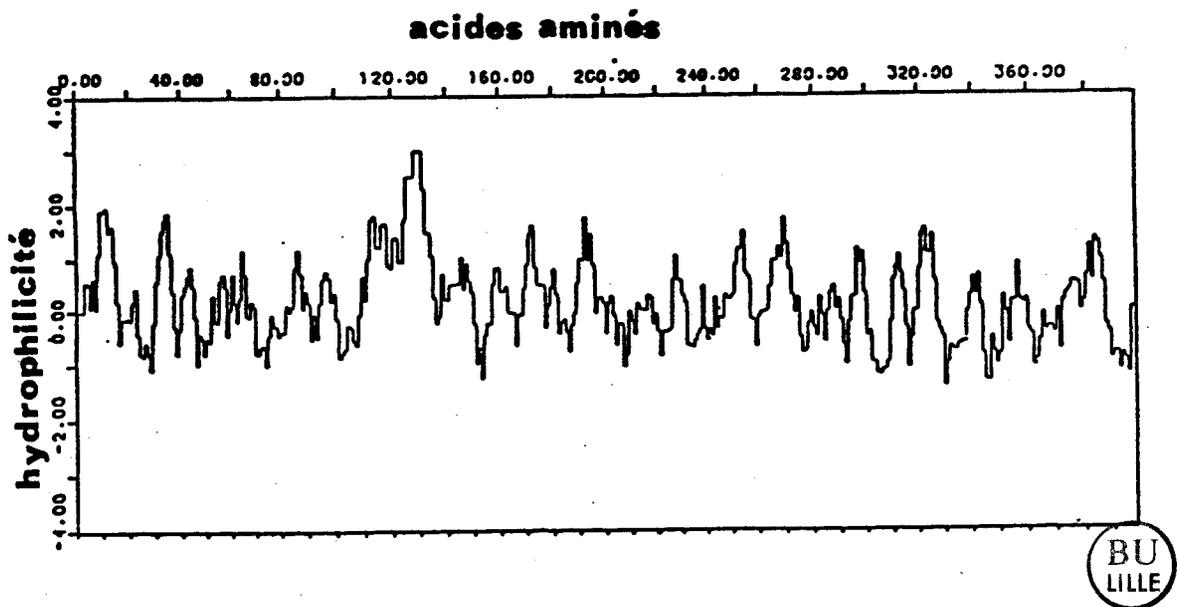
La courbe d'hydrophilicité (figure 16) déterminée par la méthode de Hopp-Woods (1981) montre que la protéine est relativement hydrophile (la valeur globale pour le polypeptide erbA est 0,1 alors que la valeur moyenne pour une protéine est de 0,07) ce qui est normal pour une protéine cytoplasmique (HAYMAN *et al.*, 1983).



100N



Figure 15



**Figure 16** : Courbe d'hydrophilicité du polypeptide v-erb A.

Nous avons comparé la séquence du polypeptide erbA à celle de toutes les autres protéines contenues dans deux banques de données : NEWAT (DOOLITTLE, 1981) et NBRF (DAYHOFF, 1983) et n'avons relevé aucune homologie avec d'autres oncogènes connus. Par contre la moitié C-terminale du polypeptide erbA présente une homologie avec la famille des anhydrases carboniques (EC 4.2.1.1) ; ce polypeptide peut donc être divisé en deux domaines :

- le domaine 1 (acides aminés 1 à 209 dans la figure 10) n'est homologue à aucune protéine connue mais présente une particularité structurale : il est riche en cystéine et en acides aminés basiques (10 % et 21 % respectivement, des acides aminés 1 à 131).

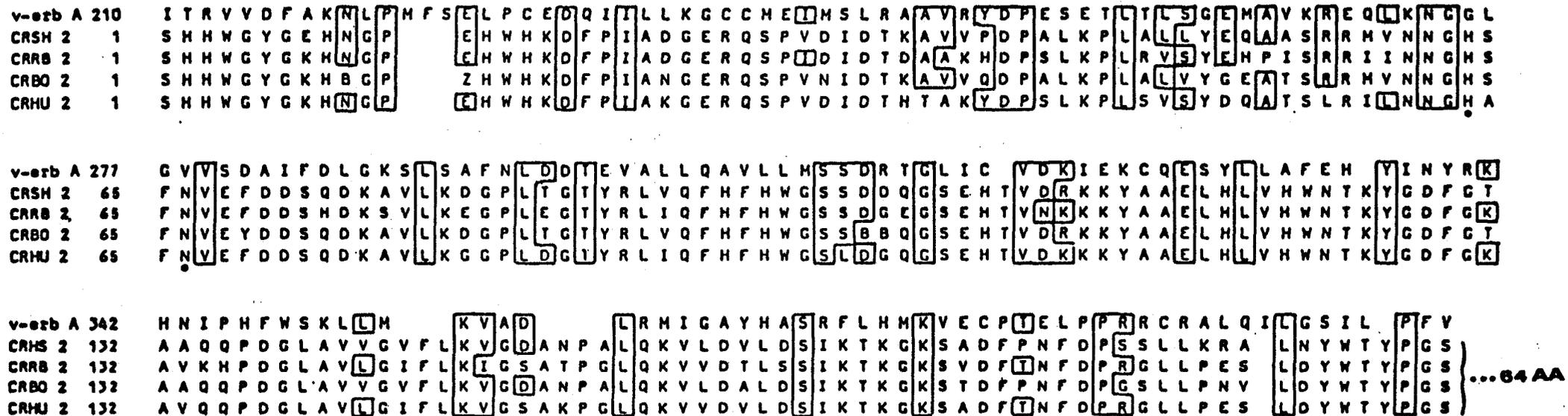
- le domaine 2 (acides aminés 210 à 398 dans la figure 10) sans particularités structurales présente une homologie avec la famille des anhydrases carboniques (figure 17). Cette homologie a été détectée en utilisant un programme d'alignement maximum de séquences protéiques décrit par STADEN *et al.*, (1979) et adapté par J.M. CLAVERIE (1984). Les séquences en amino-acides du polypeptide erbA et des anhydrases carboniques de mouton (CRSH<sub>2</sub>, TANIS *et al.*, 1974), de lapin (CRRB<sub>2</sub>, FERREL *et al.*, 1978), de boeuf (CRBO<sub>2</sub>, SCIAKI *et al.*, 1976) et de l'anhydrase carbonique humaine (CRHU<sub>2</sub>, LILJAS *et al.*, 1972) ont été comparées (figure 10). L'alignement de séquences obtenu est supérieur de 3,8 écarts types à ce qui est attendu par chance uniquement, ce qui est significatif puisque des valeurs supérieures à 3 écarts types représentent d'authentiques homologies (BARKER *et al.*, 1972). Cette vérification objective de l'alignement de séquences permet à notre avis de parler de façon prudente mais inattaquable de cette homologie de erbA avec les anhydrases carboniques.

\* L'anhydrase carbonique (EC 4.2.1.1)

L'anhydrase carbonique catalyse la réaction :



qui peut avoir lieu spontanément mais très lentement (VAN SLYKE *et al.*, 1930). A 38°C, l'anhydrase carbonique accélère 7500 fois la vitesse de cette réaction.



...64 AA

**Figure 17 :** Comparaison de la séquence protéique déduite de *erb A*-domaine 2 à celles de 4 anhydrases carboniques de type 2 (MOUTON : CRSH<sub>2</sub> (TANIS et al., 1974) ; LAPIN : CRRB<sub>2</sub> (FERREL et al., 1978) ; BOEUF : CRBO<sub>2</sub> (SCIACKY et al., 1976) ; HOMME : CRHU<sub>2</sub> (LILJAS et al., 1972)) contenues dans les banques de données NEWAT (DOOLITTLE 1981) et NBRF (DAYHOFF et al., 1983).

\* Le code une lettre a été employé pour les acides aminés : A, alanine ; R, arginine ; N, asparagine ; D, acide aspartique ; C, cystéine ; Q, glutamine ; E, acide glutamique ; G, glycine ; H, histidine ; I, isoleucine ; L, leucine ; K, lysine ; M, méthionine ; F, phénylalanine ; P, proline ; S, sérine ; T, thréonine ; W, tryptophane ; Y, tyrosine ; V, valine ; B, asparagine ou acide aspartique ; Z, glutamine ou acide glutamique.



\* Les résidus identiques sont encadrés.

\* Les points indiquent deux des amino-acides importants dans le site actif des anhydrases carboniques.

## 1. Rôle dans l'organisme

. Dans le système sanguin, l'anhydrase carbonique existe exclusivement dans les globules rouges où elle est étroitement associée au maintien de l'équilibre acido-basique (figure 18). Le  $\text{CO}_2$  qui passe des tissus dans le sang, pénètre dans les cellules rouges où il forme de l'acide carbonique sous l'action de l'anhydrase carbonique. Une petite partie de l'acide carbonique est dirigée vers le plasma, le reste est transformé en bicarbonates ( $\text{HCO}_3^-$ ) et en ion  $\text{H}^+$ . Les bicarbonates passent dans le plasma (où ils s'échangent avec des ions chlore) pour former du bicarbonate de sodium. Les ions chlore libérés pénètrent dans les érythrocytes où ils sont neutralisés par les ions potassium.

Toutes ces réactions sont réversibles et, dans le poumon, lorsque le sang devient artériel, les ions chlore retournent dans le plasma libérant ainsi du potassium intracellulaire qui tamponne l'oxyhémoglobine nouvellement formée. Dans le plasma, les ions chlore neutralisent le sodium libéré par l'élimination du  $\text{CO}_2$  des bicarbonates pendant la respiration (CHINARD et al., 1960 ; WISTRAND, 1981).

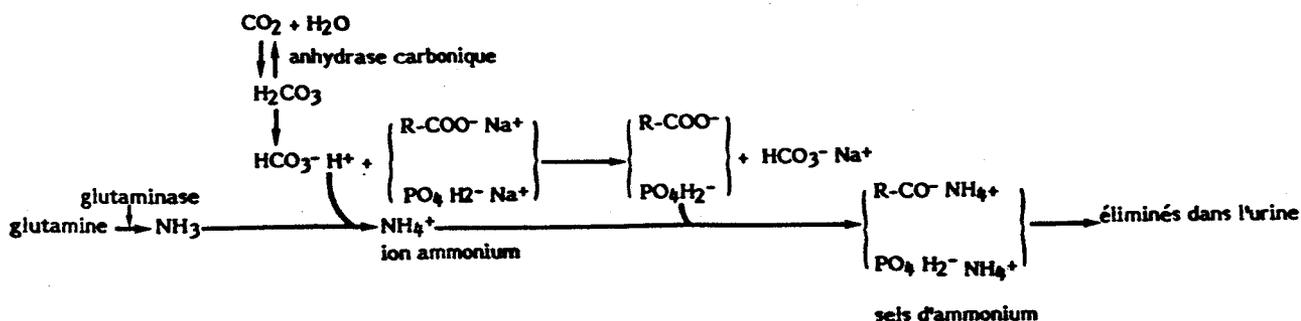
Des concentrations élevées d'anhydrase carbonique existent également dans la muqueuse gastrique et le cortex rénal (LONNERHOLM, 1984).

. dans l'estomac, l'anhydrase carbonique intervient dans le mécanisme de sécrétion de l'acide chlorhydrique du suc gastrique. La réaction globale peut s'écrire :

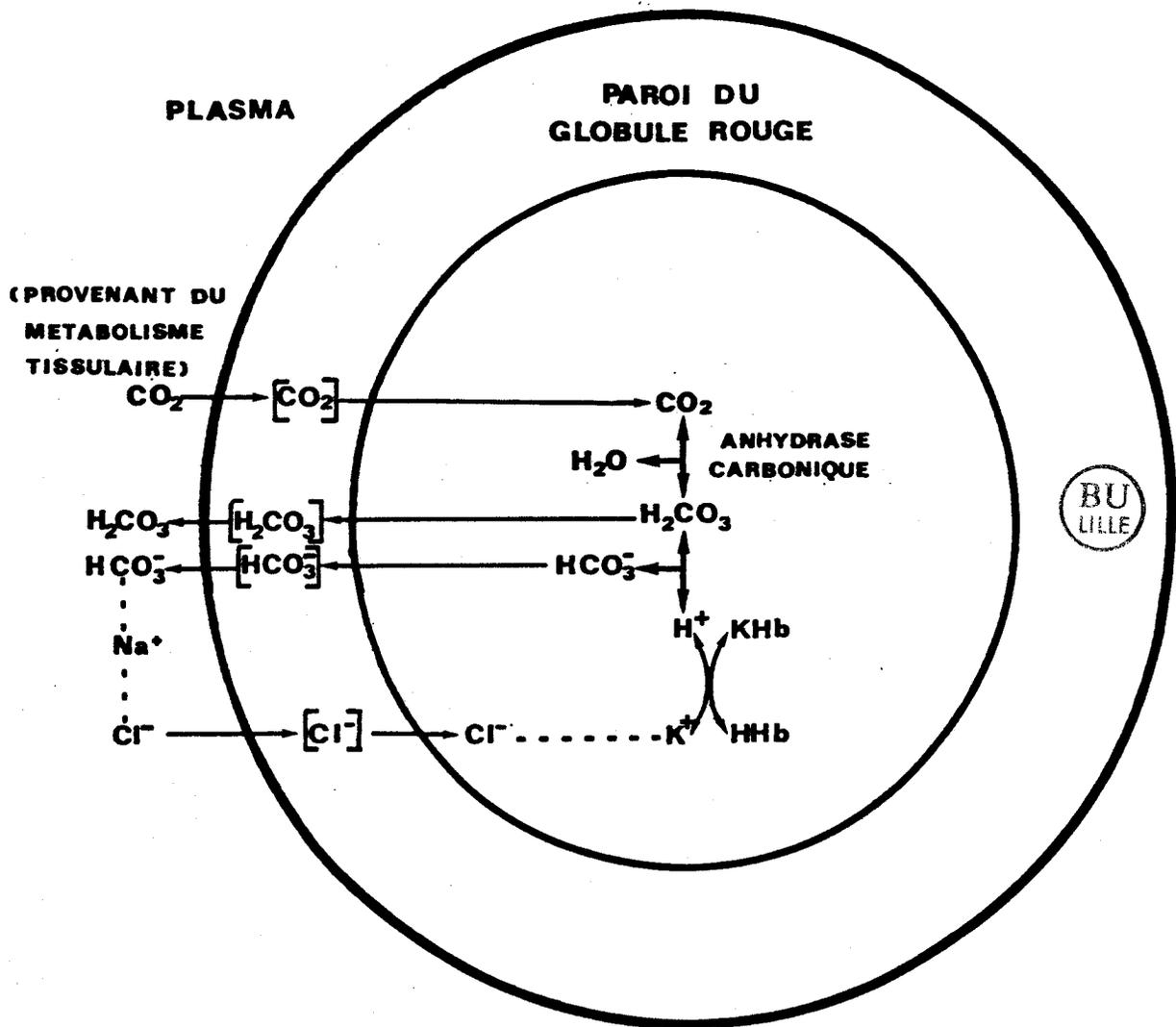


. au niveau du rein, elle joue un rôle important dans l'élimination urinaire. La glutaminase des cellules tubulaires désamine la glutamine et l'ammoniaque formé se substitue aux ions sodium pour neutraliser les acides à éliminer (figure 19).

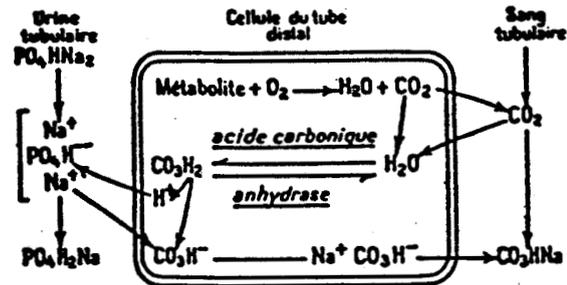
Les ions  $\text{Na}^+$  libérés se fixent sur les ions bicarbonates formés à partir de l'acide carbonique (voir ci-dessous) :



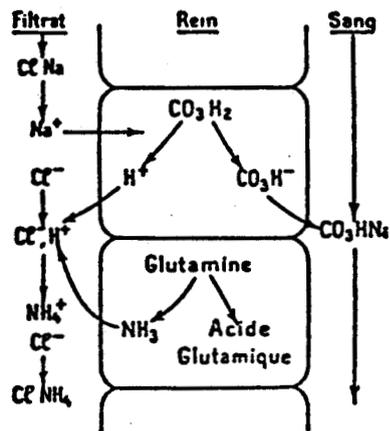
Le rôle joué par l'anhydrase carbonique est donc très important.



**Figure 18** : Rôle de l'anhydrase carbonique dans les érythrocytes.



BU  
LILLE



**Figure 19** : Rôle de l'anhydrase carbonique dans les cellules rénales (d'après BOULANGER et al., 1968).

## 2. Propriétés

- La masse moléculaire de cet enzyme est de 30 000 daltons. C'est un métalloenzyme dont le métal est le zinc (teneur 0,2 % soit un atome de zinc par molécule) (COLEMAN, 1984). Son point isoélectrique est de 5,3. De nombreux inhibiteurs ont été décrits (agents oxydants, métaux lourds, cyanures, azides). Les dérivés sulfamidés dont le groupement-CO-NH<sub>2</sub> est libre, inhibent également l'anhydrase carbonique. Cette inhibition explique certains troubles de la fonction respiratoire au cours de traitements sulfamidés. L'un des inhibiteurs les plus actifs est l'acétazolamide ou 2 acétylamino 1-3-4 thiodiazol 5 sulfonamide qui est utilisé comme diurétique (TOMASHEFSKI *et al.*, 1974).

- Les structures primaires et tridimensionnelles de plusieurs anhydrases carboniques ont été déterminées (LILJAS *et al.*, 1972 ; TANIS *et al.*, 1974 ; SCIAKI *et al.*, 1976 ; FERREL *et al.*, 1978 ; WENDORFF *et al.*, 1985).

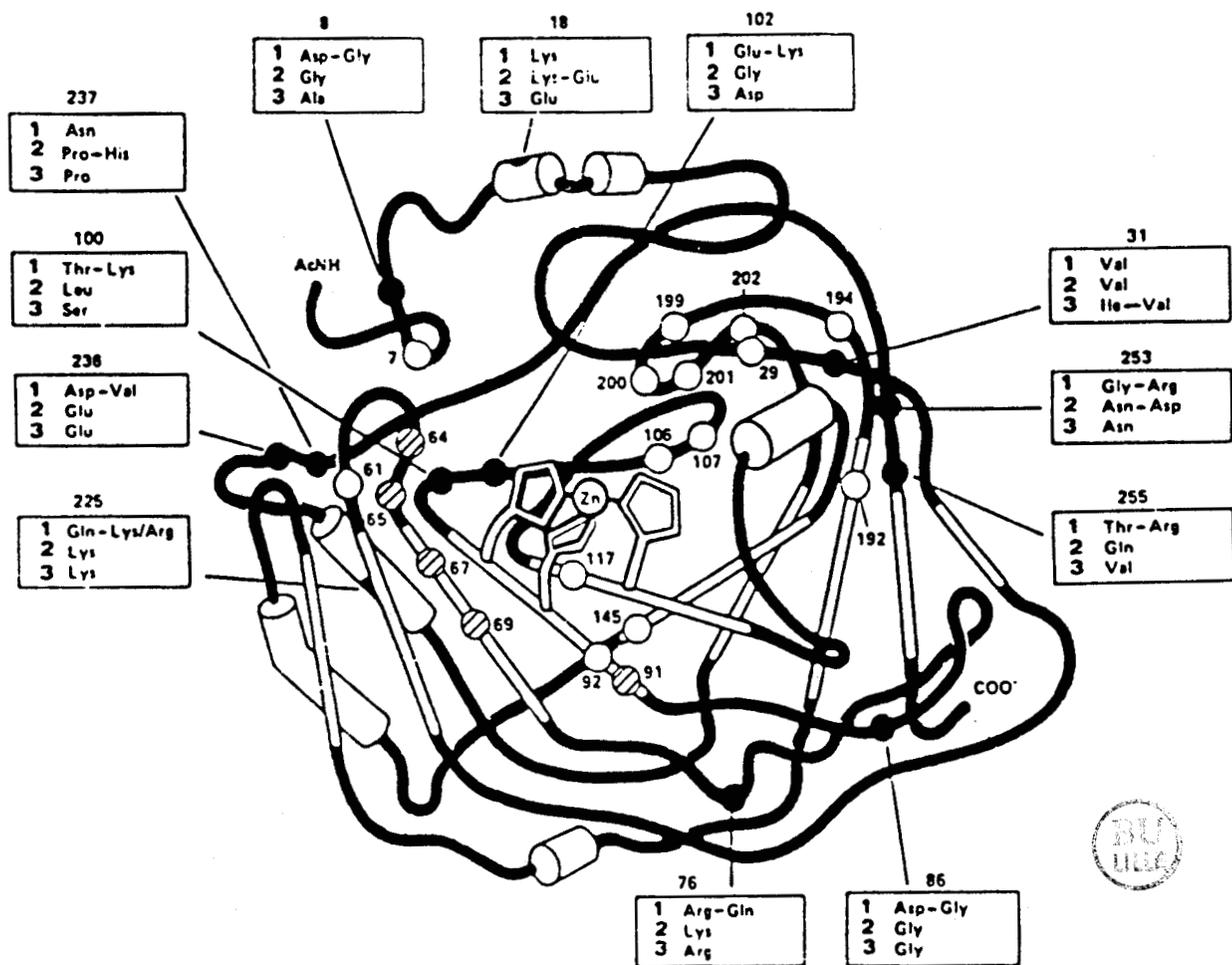
Les anhydrases carboniques comportent 259 acides aminés et, dans toutes les espèces animales étudiées, existent sous formes de plusieurs isoenzymes : les anhydrases carboniques 1 et 2 correspondent dans l'ancienne nomenclature aux anhydrases carboniques B et C. L'anhydrase carbonique A représentait une forme mineure quasiment identique à la forme B (TASHIAN *et al.*, 1975).

Les isoenzymes 1 et 2 ont des structures tridimensionnelles très semblables bien que 60 % seulement de leurs acides aminés soient identiques (figure 20).

Le zinc se trouve au fond d'une cavité de 15 angströms de profondeur, ses ligands sont trois histidines (94, 96, 119) et une molécule d'eau. La cavité du site actif est formée de 50 résidus d'acides-amino qui sont pour moitié hydrophobes.

Dans le globule rouge l'anhydrase carbonique 2 est l'isoenzyme actif ; l'anhydrase carbonique 1 est également présente en quantité abondante bien qu'elle soit inhibée par les ions chlore aux concentrations physiologiques (pourquoi cet isoenzyme persiste-t-il dans la cellule alors qu'il aurait pû être éliminé par mutation sans dommage pour l'organisme ? a-t-il un autre rôle ?).

Une anhydrase carbonique 3 a été découverte dans le muscle et d'autres tissus (TASHIAN *et al.*, 1975) ; son pouvoir catalytique est faible, environ 4 % de celui de l'anhydrase carbonique 2. Dans le site actif de cet isoenzyme, deux histidines essentielles sont remplacées par Lys et Arg, ce qui expliquerait cette perte d'activité. Il se pourrait également que l'enzyme fût présent dans un autre but : l'anhydrase carbonique 3 a une action nitrophénylphosphatase et pourrait donc jouer le rôle d'une phosphatase.



**Figure 20 :** Représentation schématique de la structure tridimensionnelle de l'anhydrase carbonique humaine 2 (cette structure est similaire à celle de l'anhydrase carbonique 1).

Dans les encadrés nous avons fait figurer les substitutions intervenant dans différents variants génétiques et les résidus homologues en cette position dans les autres isoenzymes.

Les segments de structure bêta sont représentés en blanc et les segments d'hélice alpha sous forme de cylindres.

L'ion zinc est lié aux histidines 94, 96 et 119.

Schéma d'après TASHIAN et al, 1984.

\* Homologie du polypeptide *erbA* avec les anhydrases carboniques.

L'homologie du polypeptide *erbA* avec les anhydrases carboniques de type 2 commence au résidu d'acide aminé 219 du polypeptide *erbA* (résidu 10 pour les anhydrases carboniques) et s'étend jusqu'à l'extrémité C-terminale de la molécule (résidu 398 de *erbA* ou résidu 195 des anhydrases carboniques qui en comportent 259). L'homologie globale est de 27 % sur 180 amino-acides, (une homologie entre protéines est significative à partir de 20 %). Elle est plus élevée en certaines régions de la molécule ; par exemple, au voisinage de deux acides aminés impliqués dans le site actif des anhydrases carboniques (LILJAS *et al.*, 1972) (His 63 et Asn 66) (astérisques dans la figure 17), l'homologie atteint 51 % (*erbA* des résidus 248 à 274), bien que ces deux résidus ne soient pas retrouvés dans *erbA*. 18 autres résidus du site actif des anhydrases carboniques (Tableau V) peuvent être comparés au polypeptide *erbA* et parmi ceux-ci bien que les Tyr 131 et Leu 143 montrent quelque homologie avec des résidus du site actif, seul le résidu Glu 117 est homologue a un résidu complètement invariant dans les anhydrases carboniques. De plus, les résidus d'His (94, 96, 119) qui forment les ligands avec l'ion zinc essentiel à l'activité des anhydrases carboniques sont absents dans *erbA*-domaine 2. Cependant, parmi les 44 résidus du site actif invariants dans les trois isoenzymes, 10 (soit 22 %) sont retrouvés dans *erbA*-domaine 2. Des résultats de séquence partiels concernant l'anhydrase carbonique de poulet (TASHIAN *et al.*, communication personnelle), il ne ressort pas que le polypeptide *erbA* soit plus homologue à l'anhydrase carbonique de poulet qu'à celle des mammifères.

Il est donc très probable que l'homologie que nous avons mise en évidence soit une homologie structurelle mais non fonctionnelle. A ce jour aucune activité anhydrase carbonique de la protéine P75 *gag-erbA* n'a été détectée (CONSCIENCE *et al.*, communication personnelle).

Il serait bien sûr extrêmement intéressant de savoir comment le gène cellulaire *erbA* s'est comporté au cours de l'évolution et quel est le rôle de la protéine *c-erbA* dans le métabolisme cellulaire normal. C'est ce qui nous a incités à étudier le gène *c-erbA* chez le poulet et chez l'homme.

Position des acides aminés à partir de l'extrémité N-terminale															
Isoenzymes	7	29	61	64	65	67	69	91	92	94	96	106	107	117	119
(v-erbA-domaine 2		Glu	Lys	Gly	Leu	Val	Ser	Leu	Leu	Ala	Leu	Ile	Cys	Glu	Tyr
( Homme 1	Tyr	Ser	Asn	His	Ser	His	Asn	Phe	Gln	His	His	Glu	His	Glu	His
( Rhésus 1	Tyr	Ser	Asn	His	Ser	His	Asn	Phe	Gln	His	His	Glu	His	Glu	His
( Boeuf 1	Tyr	Ser	Asn	His	Ser	His	Asn	Phe	Gln	His	His	Glu	His	Glu	His
( Cheval 1	Tyr	Ser	Asn	His	Ser	Gln	Lys	Val	Gln	His	His	Glu	His	Glu	His
( Tortue 1	Tyr	Ser	Asn	His	Ser	His	Asn	His	Gln	His	His	Glu	His	Glu	His
( Homme 2	Tyr	Ser	Asn	His	Ala	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His
( Boeuf 2	Tyr	Ser	Asn	His	Ser	Asn	Glu	Val	Gln	His	His	Glu	His	Glu	His
( Cheval 2	Tyr	Ser	Asx	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His
( Lapin 2	Tyr	Ser	Asn	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His
( Souris 2	Tyr	Ser	Asn	His	Ser	Asn	Glu	Ile	Gln	His	His	Glu	His	Glu	His
( Poulet 2		Ser	Asn	His	Ser	Asn	Glu								
( Homme 3		Tyr	Ser	Asn	Lys	Thr	Arg	Val	Arg	Gln	His	His	Glu	Glu	His
( Boeuf 3	Tyr	Ser	Asn	Lys	Thr	Arg	Val	Arg	Gln	His	His	Glu	His	Glu	His
( Cheval 3	Tyr	Ser	Asn				Val		Gln	His	His	Glu	His	Glu	His
( Poulet 3		Ser					Val		Gln	His	His				
(v-erbA-domaine 2	121	131	141	143	145	192	194	199	200	201	202	204	206	207	211
( Homme 1	Ala	Leu	Leu	Val	Gly	Trp	Tyr	Thr	His	Pro	Pro	Tyr	Ser	Val	Ile
( Rhésus 1	Val	Leu	Leu	Val	Gly	Trp	Tyr	Thr	His	Pro	Pro	Tyr	Ser	Val	Ile
( Boeuf 1	Val	Phe	Leu	Leu	Gly	Trp	Tyr	Thr	His	Pro	Pro	Leu	Ser	Val	Ile
( Cheval 1	Val	Phe	Leu	Ile	Gly	Trp	Tyr	Thr	His	Pro	Pro	Tyr	Ser	Val	Val
( Tortue 1	Val	Ile	Leu	Leu			Tyr	Thr	His	Pro	Pro	Ser	Ser	Val	Ile
( Homme 2	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val
( Boeuf 2	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Ser	Val	Val
( Cheval 2	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val
( Lapin 2	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val
( Souris 2	Val	Phe	Leu	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Leu	Cys	Val	Val
( Poulet 2															
( Homme 3		Val	Phe				Trp	Tyr	Thr	Thr	Pro	Pro			
( Boeuf 3	Val	Ile	Ile	Val	Gly	Trp	Tyr	Thr	Thr	Pro	Pro	Cys	Cys	Ile	Leu
( Cheval 3	Val	Tyr				Trp	Tyr	Thr	Thr	Pro	Pro	Glu	Cys	Ile	Leu

**TABLEAU V** : Résidus d'acides aminés impliqués dans le site actif des isoenzymes 1, 2, 3 de l'anhydrase carbonique. (d'après TASHIAN et al., 1985).

CHAPITRE V

ETUDE DES EQUIVALENTS CELLULAIRES CHEZ LE POULET ET CHEZ L'HOMME  
DE  $\gamma$ -erbA-domaine 2 (HOMOLOGUE A L'ANHYDRASE CARBONIQUE).

## I Etude du locus c-erbA-domaine 2 de l'ADN cellulaire de poulet.

### 1. Carte de restriction.

L'ADN cellulaire d'érythrocytes de poulet a été digéré par l'enzyme de restriction EcoRI et les fragments clonés dans le phage lambda-CHARON. Les phages recombinants ont été ensuite hybridés avec deux sondes virales (AvaI-SalI ; SalI-SacI) correspondant respectivement aux domaine 1 et 2 de v-erbA. La sonde 2 hybridait avec deux phages recombinants contenant des inserts de 20 kbp et 5 kbp représentant respectivement le locus c-erbA complet et un locus indépendant correspondant au domaine 2 uniquement.

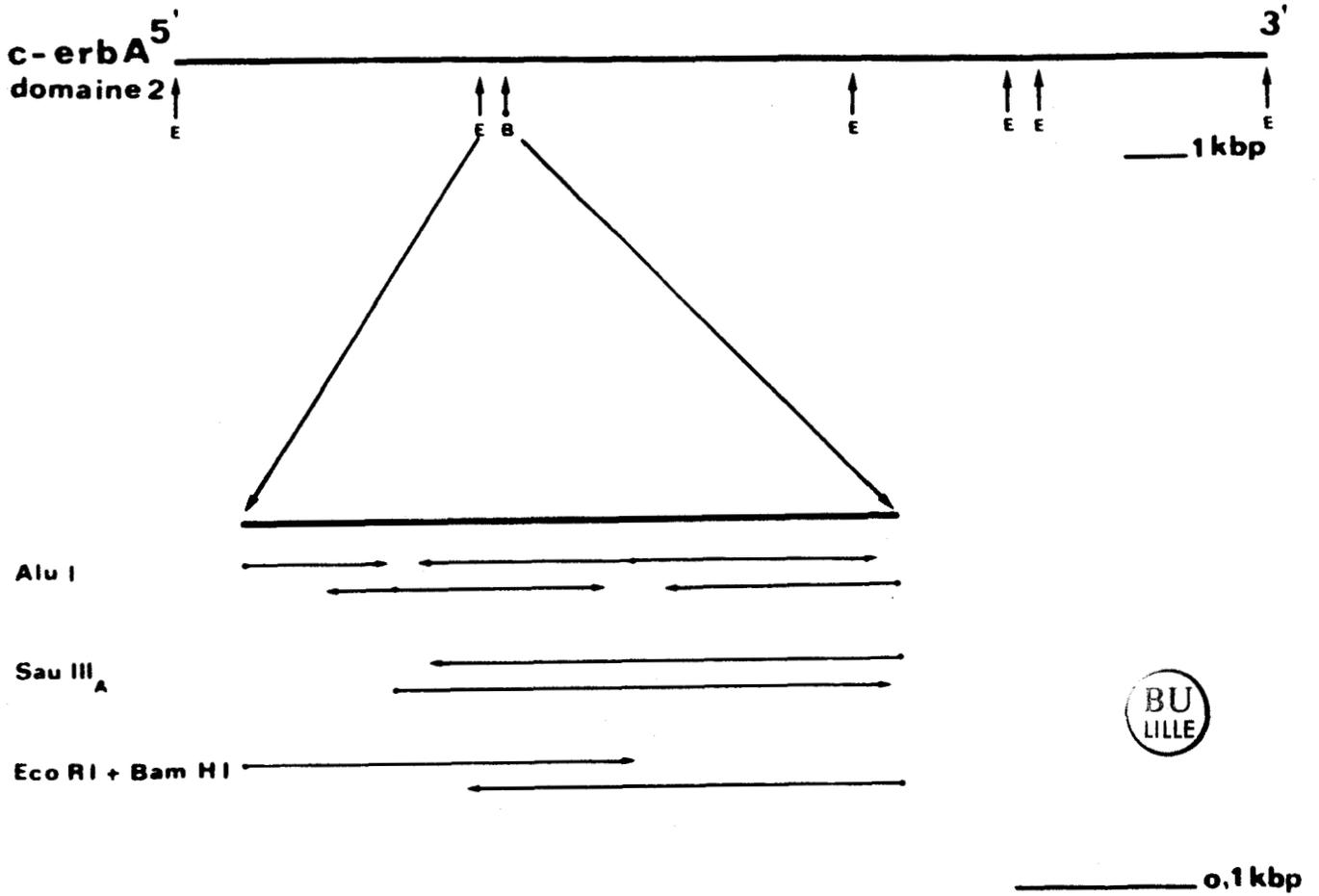
La carte de restriction de ce locus indépendant est schématisée dans la figure 21.

### 2. Séquence d'un fragment d'ADN cellulaire de poulet comportant le premier exon de c-erbA domaine 2.

De façon à déterminer avec précision le début du deuxième domaine de erbA dans AEV, nous avons séquencé un fragment EcoRI-BamHI (0,37 kbp) d'ADN cellulaire de Poulet comportant l'extrémité 5' du locus c-erbA-domaine 2.

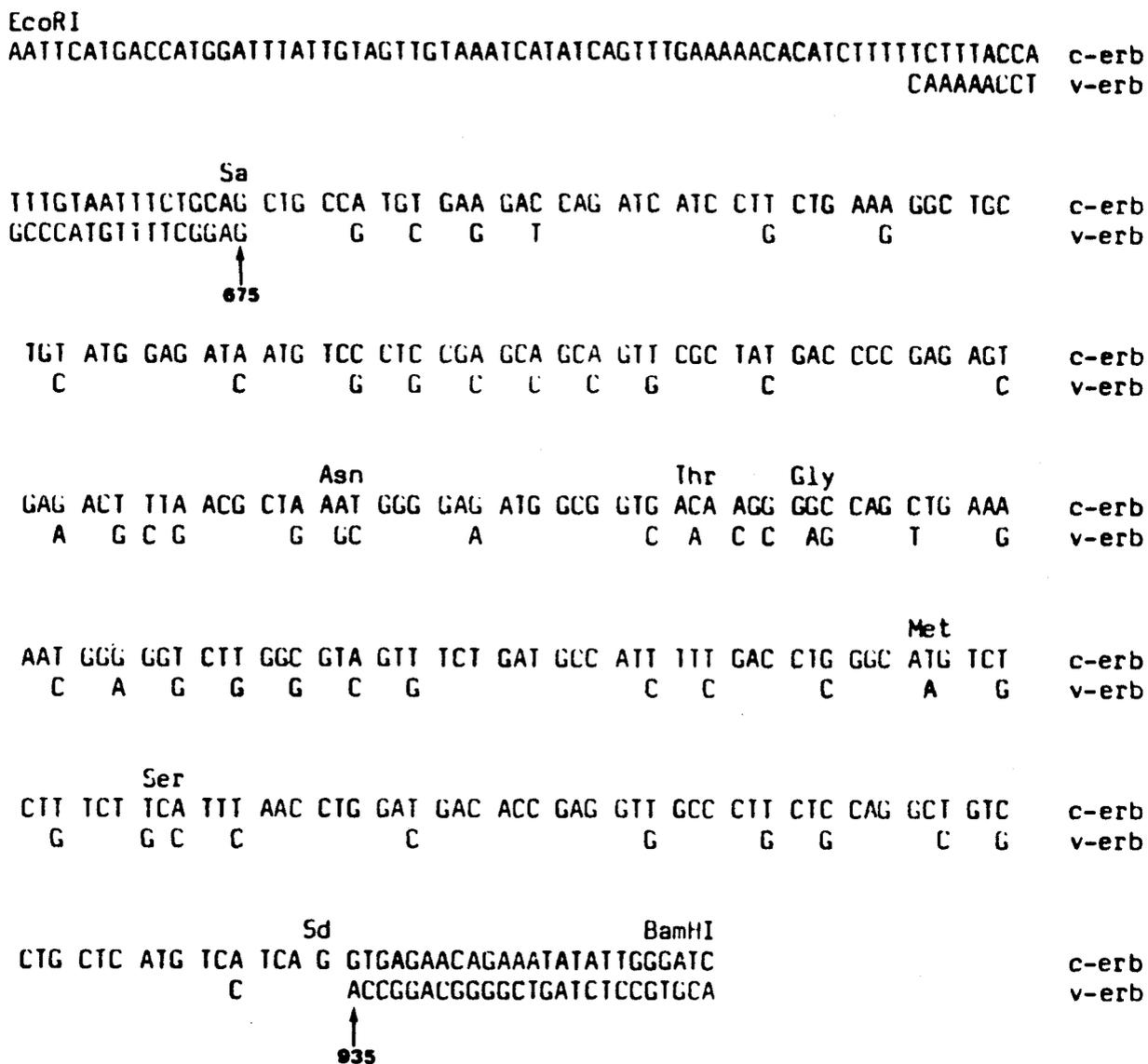
La figure 22 montre la comparaison entre les séquences virales et cellulaires. L'homologie commence au nucléotide 676 de v-erbA et s'étend sur 258 nucléotides jusqu'au nucléotide 934 de v-erbA où une complète divergence des deux séquences indique la fin du premier exon de c-erbA-domaine 2. Un site d'épissage accepteur TGTAATTTCTGCAG.C (Sa dans la figure 22) et un site d'épissage donneur CAG.CT (Sd dans la figure 22) marquent le début et la fin de cet exon. L'homologie entre les séquences nucléotidiques, virale et cellulaire, est de 78 %, de 94 % en ce qui concerne les séquences protéiques déduites, la large majorité des mutations impliquant le troisième nucléotide des codons.

La présence de sites d'épissage, accepteur et donneur (MOUNT, 1982), aux extrémités de cet exon montre que le locus c-erbA-domaine 2 a une structure classique exon-intron. SAULE et al. ont mis en évidence la présence d'au moins deux exons dans chacun des domaines 1 et 2 de erbA. L'existence d'un locus indépendant permet de penser qu'il existe une famille de gènes ayant en commun le domaine 2 de erbA, homologue à l'anhydrase carbonique,



**Figure 21** : Stratégie de séquence du fragment d'ADN cellulaire de poulet comportant le 1er exon de c-erbA-domaine 2.

↑ E = EcoRI  
 ↑ B = BamHI.

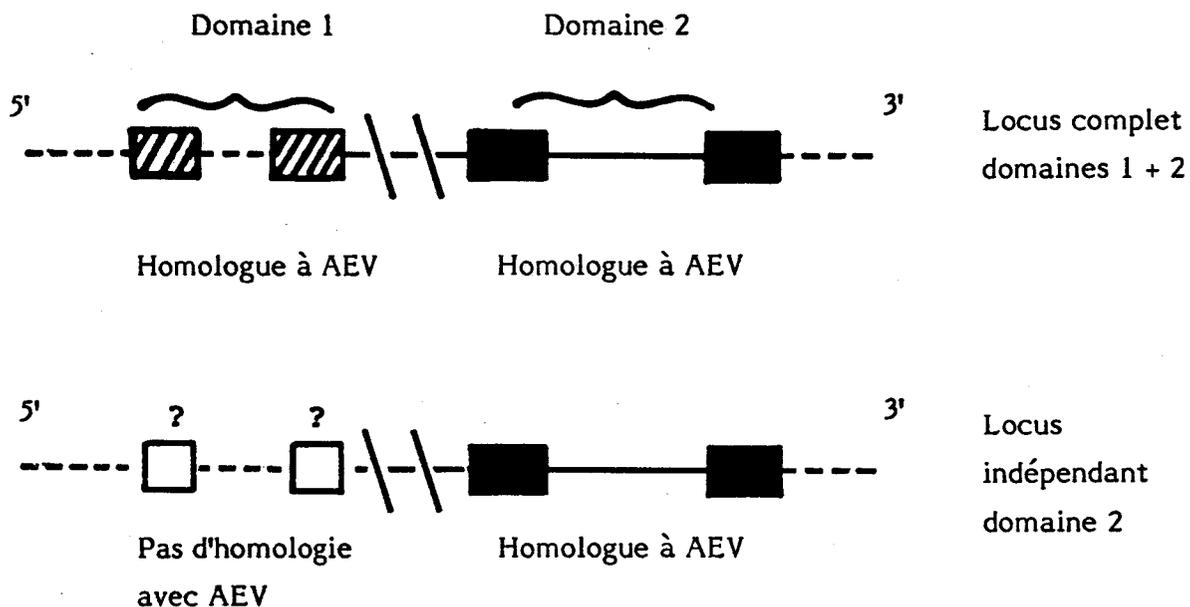


**Figure 22** : Comparaison de la séquence nucléotidique de v-erb et de celle d'un fragment EcoRI-BamHI d'ADN cellulaire de poulet correspondant au premier exon de erba-domaine 2.

\* Sa et Sd indiquent les sites d'épissage (accepteur et donneur).

\* Dans la séquence de v-erb (correspondant à l'exon) seuls les nucléotides qui diffèrent de c-erb sont indiqués. Les acides aminés de c-erb qui diffèrent de ceux de v-erb sont indiqués au-dessus des triplets nucléotidiques correspondants.

et une partie 5' variable, tout comme les immunoglobulines qui possèdent une région variable et des domaines constants (PUTNAM et al., 1960) :



## II Etude de la partie c-erbA-domaine 2 du locus complet d'ADN cellulaire humain.

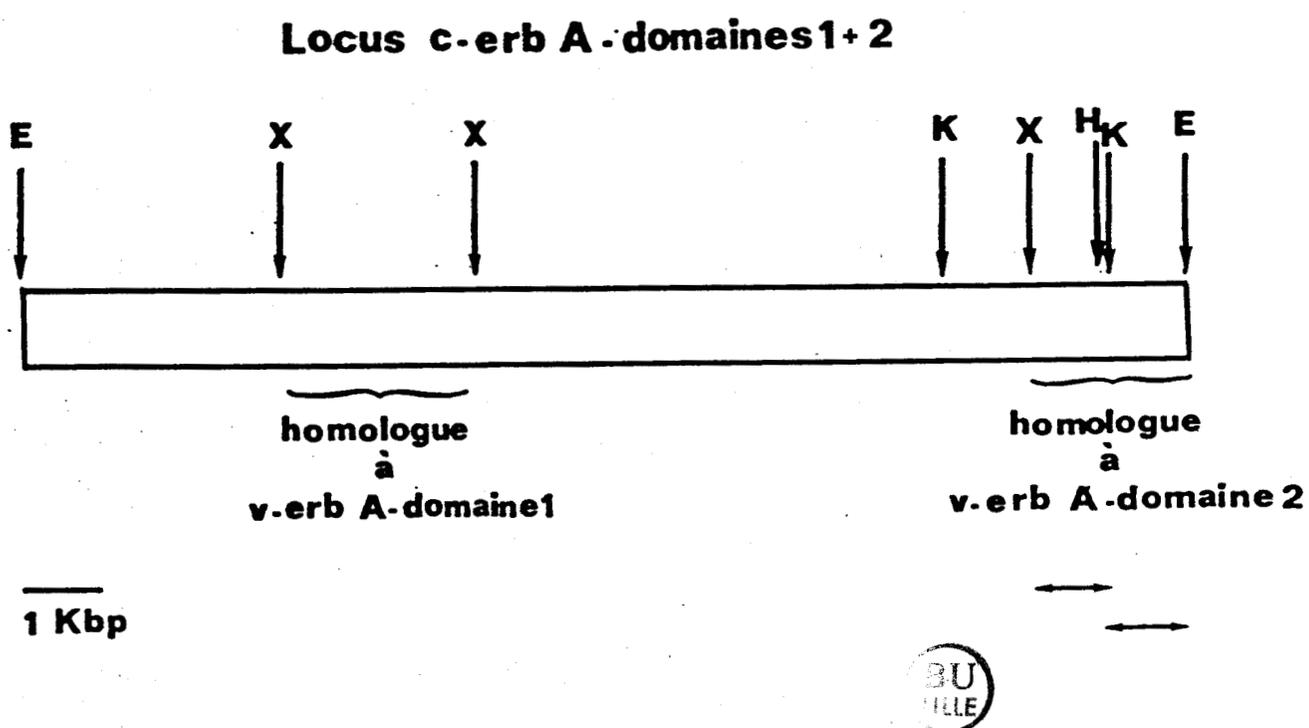
### 1. Carte de restriction.

Comme pour l'ADN cellulaire de poulet, la carte de restriction du locus c-erbA complet de l'ADN cellulaire humain a été établie par SAULE et al. (figure 23).

### 2. Séquence du domaine 2 du locus c-erbA complet.

Nous avons déterminé par la méthode des dégradations chimiques de MAXAM et GILBERT la séquence de 1850 nucléotides couvrant la totalité du domaine 2. Nous avons séquencé deux fragments de restriction HindIII-EcoRI (0,95 kbp) et XhoI-KpNI (0,9 kbp). La stratégie de séquence est indiquée dans la figure 23.

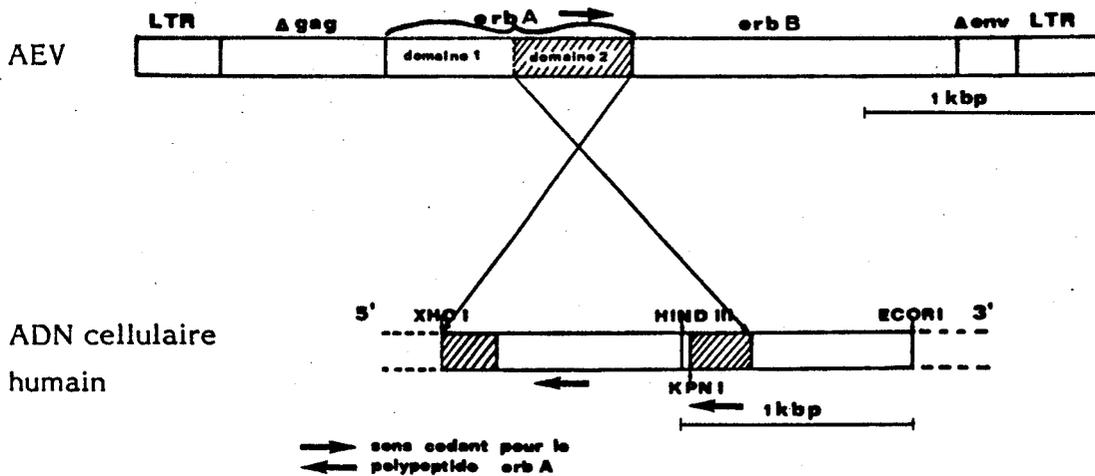
Nos résultats confirment une constatation surprenante qui émanait des cartes de restriction : dans le locus c-erbA complet de l'ADN cellulaire humain, l'orientation du domaine 2 est inversée par rapport au sens viral.



**Figure 23** : Carte de restriction du locus c-erb A complet de l'ADN cellulaire humain.

Les flèches situées à la partie inférieure de la figure représentent les fragments d'ADN séquencés.

↓ E : EcoRI  
 ↓ X : XhoI  
 ↓ K : KpnI  
 ↓ H : HindIII



Le fragment Hind III-EcoRI dont la séquence est indiquée dans la figure 24 comporte l'exon correspondant à l'exon de l'ADN cellulaire de poulet décrit dans la figure 22. L'homologie entre les séquences nucléotidiques humaine et virale est de 83 %. Cette homologie, comme pour l'ADN cellulaire de poulet, commence au nucléotide 676 de v-erbA et s'étend sur 258 nucléotides.

Le fragment XhoI-KpnI (séquence indiquée dans la figure 24) comporte le deuxième exon de c-erbA-domaine 2. L'homologie entre les séquences nucléotidiques humaine et virale est de 88 %. Cette homologie commence au nucléotide 935 de v-erbA et s'étend sur 199 nucléotides.

La séquence en amino-acides déduite de la séquence nucléotidique des deux exons, dans le même cadre de lecture que celui du virus (tous les autres cadres sont fermés par des codons de terminaison), présente une homologie avec le polypeptide erbA viral de 95 % pour le premier exon et de 87 % pour le deuxième exon. Les homologies de la séquence humaine avec l'anhydrase carbonique humaine sont conservées mais pas sensiblement augmentées, ce qui prouve que le polypeptide erbA, bien qu'apparenté à l'anhydrase carbonique n'est pas l'enzyme lui-même.



CONCLUSION

Nous avons contribué à l'étude des oncogènes viraux et cellulaires en déterminant la séquence nucléotidique de l'oncogène viral v-erbA et de ses équivalents cellulaires chez le poulet et chez l'homme.

Nous avons montré qu'une trentaine de nucléotides de la partie 3' du gène v-erbA présentent une forte homologie (82 %) avec le gène env de RSV et que les quatre amino-acides C-terminaux du polypeptide erbA sont codés par des nucléotides dérivés du locus c-erbB.

Une région non codante de v-erbB sépare le codon de terminaison de la protéine P75 gag-erbA (position 1195) du site d'épissage accepteur (position 1370) probablement utilisé par le virus pour engendrer l'ARN messager sous-génomique erb-B.

Deux domaines peuvent être définis dans le polypeptide erbA. Le domaine 1 est riche en cystéine et en acides aminés basiques (Lys + Arg). Le domaine 2, sans particularités structurales, présente une homologie avec la famille des anhydrases carboniques et plus particulièrement avec l'isoenzyme de type 2, isoenzyme actif. Cette homologie est globalement de 27 % sur 180 résidus d'acides aminés. Elle est plus élevée en certaines régions de la molécule, en particulier, au voisinage de deux amino-acides du site actif des anhydrases carboniques bien que l'on ne retrouve pas ces deux acides aminés dans le polypeptide erbA. Les trois histidines servant à lier le zinc indispensable à l'action des anhydrases carboniques sont également absentes dans le polypeptide erbA. Ceci pourrait impliquer une homologie structurale mais non fonctionnelle, ce qui semble vérifié par le fait qu'aucune activité anhydrase carbonique de la protéine P75 gag-erbA n'ait pu être mise en évidence.

Dans l'ADN cellulaire du poulet et de l'homme, le gène erbA existe sous forme de deux loci : un locus c-erbA complet (domaines 1 + 2) et un locus indépendant correspondant au domaine 2 uniquement. Ces loci ont une structure intron-exon classique et chaque domaine comporte au moins deux exons.

Il est possible que les gènes c-erbA correspondent à une famille de gènes ayant tous la même extrémité 3' (le domaine 2) et une partie 5' (le domaine 1) différente. De façon surprenante, le domaine 2 du locus c-erbA complet de l'ADN cellulaire humain est dans un sens inversé par rapport au sens viral.

Nous nous interrogeons maintenant sur le sens de ce retournement. Nous savons que dans le sens viral le gène erbA bloque la différenciation des érythroblastes à un stade immature. La protéine erbA agissant en compétition avec l'anhydrase carbonique (peut-être pour la fixation du substrat en raison de

leur homologie structurelle) pourrait intervenir sur l'un des contrôles du métabolisme comme le pH et les flux ioniques et réguler par ce biais les phosphorylations que l'on sait très sensibles à ce facteur. Dans les cellules infectées par le virus, ce contrôle se répercuterait sur la différenciation cellulaire et maintiendrait les cellules dans un état où elles prolifèrent sous l'action de v-erbB.

On peut aussi imaginer que l'équivalent cellulaire de v-erbA, dans le sens viral, bloque la différenciation des cellules hématopoïétiques et que le fonctionnement de ce gène soit requis pour remplir un compartiment de cellules rouges immatures (BFU-E par exemple). Le retournement du domaine 2 que l'on observe chez l'homme pourrait être l'un des mécanismes par lesquels le gène erbA régule son expression au cours de la différenciation. Ce gène apparaît en définitive comme un gène de différenciation et nous voyons combien il serait intéressant d'avoir accès à la protéine c-erbA. Connaissant la séquence nucléotidique du locus-c-erbA domaine 2 que nous avons déterminée, un chercheur du laboratoire d'Oncologie moléculaire de l'Institut Pasteur de LILLE, Y. GOLDBERG a inséré dans un vecteur d'expression bactérien (et dans sa phase de lecture correcte) un fragment d'ADN cellulaire humain correspondant au premier exon de c-erbA domaine 2 (séquence figure 24). La protéine synthétisée par le vecteur recombinant a été isolée, purifiée et injectée à des lapins afin d'obtenir l'antisérum correspondant. Cet antisérum s'est révélé efficace en précipitant la protéine P75 gag-erbA dans des cellules infectées par AEV. Nous allons donc étudier la protéine c-erbA humaine ; nous espérons pouvoir expliquer son action au cours de l'ontogénèse et de la différenciation des cellules hématopoïétiques et l'isoler en quantité suffisante pour réaliser son étude biochimique directe.

Comme nous venons de le voir, les méthodes de détermination de la structure primaire de l'ADN permettent une remarquable approche de certains problèmes biologiques. Le travail que nous avons résumé en constitue un exemple et nous devons préciser que pendant les quatre années que nous avons consacrées à l'étude de ce gène, nous avons mené à bien, en parallèle, deux autres projets de recherche.

Le premier sujet était consacré à la détermination de la séquence nucléotidique du gène c-myc humain conférant des propriétés transformantes à un rétrovirus aviaire.

Le gène c-myc semble impliqué dans les lymphomes de BURKITT chez l'homme dans la mesure où les cellules tumorales présentent toujours une translocation du chromosome 8q24 point précis de la localisation de cet

oncogène (ROBERTSON, 1983). Ces cellules montrent à la fois des modifications du taux de transcription et des mutations du gène c-myc (RABBITS et al., 1984). Par ailleurs, ces cellules contiennent également d'autres oncogènes activés (DIAMOND et al., 1983). Il est dans ces conditions très difficile de savoir, d'une part si le gène myc transloqué a joué un rôle de par lui-même dans la transformation et, d'autre part, quelle est l'altération de ce gène qui serait responsable de son activation. Nous avons abordé ce problème en collaboration avec les chercheurs de l'Unité INSERM 186 qui ont développé un système biologique dans lequel le gène c-myc est actif et induit l'apparition de foyers de cellules transformées. Ces chercheurs ont introduit le gène cellulaire normal dans un rétrovirus aviaire qui s'est par la suite révélé transformant. Le clonage moléculaire de ce virus a été réalisé (MARTIN et al., 1985) et nous avons déterminé la séquence nucléotidique de la fraction c-myc du provirus intégré. Nos résultats ont montré qu'une surexpression du gène c-myc humain normal est suffisante pour induire la transformation de fibroblastes embryonnaires de caille puisque nous n'avons trouvé aucune mutation dans la séquence codante du gène myc inséré dans le virus.

Nous avons également participé à la caractérisation d'un mutant du virus MH2 ne contenant que le gène myc (MARTIN et al., 1985). Nos résultats ont contribué à montrer que ce mutant correspond en fait à la propagation de l'ARN sous-génomique de MH2 (2,8 kb). Ce virus synthétise la protéine myc (61-63 kilodaltons) également synthétisée par MH2.

L'efficacité des méthodes de détermination de séquence nucléotidique n'est donc plus à démontrer et le fascicule "RNA tumor viruses" (Cold Spring Harbor Laboratory, 1985) qui compile toutes les séquences nucléotidiques d'oncogènes viraux et cellulaires élucidées en moins de dix ans (la première méthode efficace de détermination de la structure primaire de l'ADN est apparue en 1977), consacre l'essor de ce domaine de la biologie moléculaire et concrétise les efforts de chercheurs dont la tâche ardue a abouti à la découverte de données essentielles en cancérologie.

APPENDICE TECHNIQUE

## I. Préparation du matériel génétique (préparation du plasmide).

Isolement : Une souche d'*Escherichia coli* contenant le plasmide pKH 47, est mise en culture dans un milieu LB (bactotryptone 10 g, extrait de levure 5 g, NaCl 5 g, H<sub>2</sub>O qsp 1000 ml pH 7,2-7,3.) afin d'obtenir 500 ml de culture. Le plasmide est amplifié par le chloramphénicol à une dilution au 1/200<sup>ème</sup> pendant une nuit à 37°C.

Récolte : Les bactéries sont ensuite récoltées par centrifugation (6000 t/mn - 10 minutes) et lavées dans 100 ml de tampon STE (NaCl 0,1 M, Tris 10 mM, EDTA 1 mM pH 8).

Lyse : Après centrifugation et solubilisation dans 9 ml de tampon GTE (Glucose 50 mM, Tris 25 mM, EDTA 10 mM), les bactéries sont lysées par 1 ml de lysozyme (10 mg/ml GTE) 5 à 10 secondes à 4°C puis par 20 ml d'une solution (Na OH 0,2 N, SDS 1 %) 4 minutes à 4°C et enfin 10 ml de tampon KAC (Acétate K 5M, acide acétique 2,5 M pH8) 10 minutes à 4°C.

Précipitation : Le lysat est centrifugé (5000 t/mn - 15 minutes) et filtré sur laine de verre ; on récupère le filtrat que l'on précipite par 25 ml d'isopropanol (15 minutes à - 20°C). On centrifuge (5000 t/mn - 10 minutes). Le précipité est redissous dans 4 ml de tampon TE (Tris 40 mM, EDTA 2mM, pH8) et 100 µl de Tampon (Tris 3M, pH8).

Purification : A la solution précédente ; on ajoute 50 µl de bromure d'éthidium (BET) et 5 g de chlorure de caesium (CsCl<sub>2</sub>) le mélange est centrifugé à 63 000 t/mn - 4 heures à 18°C.

La bande de plasmide est récupérée à la seringue et le BET éliminé par extraction à l'isopropanol. On dialyse une nuit à + 4°C contre du tampon TE. Le dialysat est précipité par un volume d'acétate de sodium et deux volumes d'éthanol à - 20°C pendant 48 heures.

## II. La méthode des dégradations chimiques de MAXAM et GILBERT.

### 1. Marquage.

Les fragments d'ADN double-brin sont hydrolysés par une enzyme de restriction et ensuite marqués au P<sup>32</sup> à chacune de leurs extrémités 5'.

Le marquage comporte deux étapes :

a. Action de la phosphatase alcaline d'E. coli.

On effectue une incubation d'une heure à 65° C en tampon Tris 0,05 M ; MgCl<sub>2</sub> 0,05 M ; ph 8,5 en présence de la phosphatase alcaline. Après destruction de l'excès de phosphatase par un tampon NaOH 1,5 M ; EDTA 1 mM ; les fragments d'ADN sont précipités par 3 volumes d'éthanol absolu en milieu acétate de sodium 0,3 M.

b. Marquage proprement dit.

Le culot de centrifugation est repris dans un tampon Tris 20 mM ; spermidine 1 mM ; EDTA 0,1 mM ; ph 9,5 et l'ADN légèrement dénaturé à 92°5 C dans ce tampon spermidine. (1 minute). On refroidit dans la glace fondante, on ajoute du tampon MgCl<sub>2</sub> 100 mM ; Tris 500 mM, dithiothréitol 50 mM ; ph 9,5 puis l'ATP lyophilisé (1 millicurie) et enfin la polynucléotide kinase. On laisse incuber une heure à 37°C.

2. Séparation des fragments en fonction de leur taille.

Les fragments radioactifs sont séparés sur gel d'acrylamide à 6 % en tampon acétate de sodium (Tris 0,04 M ; acétate de sodium  $5 \times 10^{-3}$  M, EDTA  $10^{-3}$  ; ph 7,8).

Composition du gel à 6 % d'acrylamide :

- Acrylamide	14,5 g
- Bis-acrylamide (1/30e)	0,5 g
- Tampon acétate de sodium	25 ml
- Temed (N,N,N',N' - tétraméthyl éthylène diamine)	125 µl
- Persulfate d'ammonium (10 %)	2,5 ml
H <sub>2</sub> O distillée qsp	250 ml

Le gel est un gel vertical de 300/400/1,2 mm.

La migration est de 16 heures à 150 volts.

On utilise comme témoin de migration un mélange de bleu de bromophénol (BB) (0,05 %) et de xylène cyanol FF (0,05 %).

### 3. Récupération des fragments d'ADN par électroélution.

L'autoradiographie du gel permet de visualiser les bandes correspondant aux différents fragments.

La distance de migration des fragments est une fonction logarithmique de leur taille. Chacune des bandes de gel est donc ensuite découpée et électroéluée en tampon acétate de sodium, 2 x 2 heures. Puis les fragments d'ADN sont précipités par 3 volumes d'éthanol en milieu acétate de sodium 0,3 M, une nuit à - 70°C.

### 4. Essais de dénaturation.

Les fragments d'ADN correspondant à chacune des bandes sont récupérés par centrifugation (10000 t/mn - 1 heure) et leur radioactivité mesurée au compteur à scintillation.

En général, 1/100e de la quantité obtenue est utilisée pour les essais de dénaturation. Il faut en effet pour chaque fragment obtenir séparément les deux brins d'ADN. L'ADN est dénaturé en milieu diméthylsulfoxyde (DMSO), deux minutes à 92°C et refroidi dans la glace fondante.

#### - Tampon de dénaturation.

- \* DMSO            15  $\mu$ l
- \* EDTA 0,1 M    3  $\mu$ l    par échantillon
- \* H<sub>2</sub>O distillée 5  $\mu$ l
- \* Colorant        10  $\mu$ l

Chaque échantillon est ensuite mis en électrophorèse sur deux gels d'acrylamide (300/400/1,2 mm) l'un à 5 %, l'autre à 8 %, en tampon borate (Tris 1 M, acide borique 1 M ; EDTA 0,02 M).

- Composition du gel à 5 % d'acrylamide :

.. Acrylamide	12,25 g
.. Bis-acrylamide	0,25 g
.. Tampon borate	12,25 ml
.. Temed	125 $\mu$ l
.. Persulfate d'ammonium	2,5 ml
.. H <sub>2</sub> O qsp	250 ml

- Composition du gel à 8 % d'acrylamide :

.. Acrylamide	19,6 g
.. Bis-acrylamide	0,4 g
.. Tampon borate	12,25 ml
.. Temed	125 $\mu$ l
.. Persulfate d'ammonium	2,5 ml
.. H <sub>2</sub> O qsp	250 ml

La migration est de 16 heures à 140 volts et les gels sont laissés en autoradiographie pendant 48 heures. Les autoradiogrammes permettent de choisir le gel de dénaturation donnant la meilleure séparation des deux brins d'ADN. Une nouvelle dénaturation est alors effectuée avec la totalité du fragment d'ADN.

### 5. Dénaturation totale.

L'opération est la même que précédemment en prenant soin de respecter le rapport quantité d'ADN/tampon de dénaturation.

### 6. Purification.

Chacun des brins d'ADN obtenu doit être purifié avant d'être soumis aux réactions de séquence.

On découpe les bandes de gel correspondant aux deux brins d'ADN et on les place au sommet d'un gel à 8 % d'acrylamide Biorad (300/400/1,2 mm). La migration est de 16 heures à 140 volts en tampon Tris-Borate.

## 7. Les réactions de séquence.

Après autoradiographie du gel de purification, les bandes correspondant à l'ADN monobrin sont électroéluées et l'ADN précipité par l'éthanol absolu en milieu acétate de sodium 0,3 M.

La radioactivité est mesurée au compteur à scintillation. Cet ADN très pur est alors prêt pour les réactions de séquence dont nous résumons le protocole :

### Réactifs

#### .. Tampon DMS :

+ cacodylate de sodium	50 mM
+ MgCl <sub>2</sub>	10 mM
+ EDTA	0,1 mM
	pH 8



#### .. Tampon Stop CT :

+ ARN entraîneur 1 mg/ml	20 µl
+ Acétate de sodium 0,3	1 ml

#### Solution a

+ mercapto éthanol	1 M
+ Tris	1 M pH 7,5
+ Acétate de sodium	1,5 M
+ Acétate de magnésium 0,05	M

#### Solution b

+ ARN entraîneur 1 mg/ml	10 µl
+ H <sub>2</sub> O distillée	15 µl

#### Tampon Stop AG

+ Solution a	900 µl
+ Solution b	100 µl

Les réactions de séquence sont résumées dans le tableau VI.

SOLUTION	REACTIF	INCUBATION	ARRET	
G	50 µl tampon DMS 1 µl ADN en- traîneur 5 µl échantil- lon	7 mn à 20°C	0°C 50 µl de tampon stop AG Agiter	750 µl d'éthanol absolu 7 mn à -70°C Centrifuger 10 mn (10000 t/ mn)
AG	10 µl H <sub>2</sub> O 1 µl ADN entraîneur (1/10) 9 µl échantil- lon	1 heure à 20°C	Congeler Lyophiliser	3 lavages H <sub>2</sub> O 50 µl, 30 µl, 30 µl
CT	15 µl H <sub>2</sub> O 1 µl ADN entraîneur 10 µl échan- tillon	8 mn à 20°C	0°C 200 µl de tampon stop CT Agiter	750 µl d'éthanol absolu 7 mn à -70°C Centrifu- ger 10 mn (10000 t/mn)
C	20 µl NaCl 5M, 1 µl ADN Entraîneur 5 µl échantil- lon	9 mn à 20°C	0°C 200 µl de tampon stop CT Agiter	750 µl d'éthanol absolu 7 mn à -70°C Centrifu- ger 10 mn (10000 t/mn)
A/C	20 µl NaOH 1,5 M EDTA 1 mM 1 µl ADN entraî- neur 7 µl échantillon	6 mn à 92°C	20 µl HCl N 250 µl de tampon stop CT Agiter	750 µl éthanol 5 mn à -70°C Centrifuger 10 mn (10000 t/mn)



TABLEAU VI : REACTIONS DE SEQUENCE

100 µl de pipéridine 10%  
Lyophilisation une nuit.  
3 lavages, H<sub>2</sub>O, 50 µl, 30 µl, 30 µl.  
Congeler, Lyophiliser.

### 8. Gels de séquence.

Le produit des cinq réactions est mis à migrer sur des gels d'acrylamide (concentration en acrylamide de 25 % à 6 %).

Tous ces gels contiennent de l'urée, nous donnons leur composition ci-après.

#### - Gel 25 %

=====

- Acrylamide	36,25 g
- Bis-acrylamide 1/30e	1,25 g
- Urée 7 M	63 g
- Tampon Tris-borate	15 ml
- Temed	60 µl
- Persulfate 10 %	0,6 ml
- H <sub>2</sub> O qsp	150 ml

#### - Gel 16 %

=====

- Acrylamide	23,2 g
- Bis-acrylamide (1/30e)	0,8 g
- Tampon Tris-borate	15 ml
- Urée 50 %	75 g
- Temed	60 µl
- Persulfate 10 %	0,6 ml
- H <sub>2</sub> O qsp	150 ml

#### - Gel 8 %

=====

- Acrylamide	30,4 g
- Bis-acrylamide (1/20e)	1,2 g
- Tampon Tris-borate	40 ml
- Urée 50 %	200 g
- Temed	160 µl
- Persulfate 10 %	2 ml
- H <sub>2</sub> O qsp	400 ml



- Gel 6 %  
=====

- Acrylamide	22,8 g
- Bis-acrylamide	1,2 g
- Tampon Tris-borate	40 ml
- Urée 50 %	200 g
- Temed	160 µl
- Persulfate 10 %	2 ml
- H <sub>2</sub> O            qsp	400 ml

Les conditions de migration sont :

- Gel 25 %  
=====

- Prémigration	600 volts	15 heures
- Migration	1300 volts	5 heures



- Gel 16 %  
=====

- Prémigration	1000 volts	1 heure
- Migration	600 volts	15 heures

- Gel 8 % et 6 %  
=====

- Prémigration	1800 volts	2 heures
- Migration	1800 volts	24 heures

Les quantités d'ADN déposées sont fonction de la radioactivité de chaque échantillon.

Ces quantités sont ajustées de façon à avoir pour chaque brin d'ADN un même nombre de coups par base.

La méthode de MAXAM et GILBERT permet une lecture directe de la séquence nucléotidique à partir de l'extrémité 5' marquée :

- un gel à 25 % d'acrylamide (300/400/0,8 mm) permet de lire les 30 à 35 premières bases,

- un gel à 16 % d'acrylamide (300/400/0,8 mm) permet de lire les bases de la position 25 à la position 80 environ donc d'obtenir un recoupement avec le gel à 25 %,

- un gel à 8 % d'acrylamide (300/800/0,8 mm) donne les bases de la position 70 à la position 150-170 environ,

- un gel à 6 % d'acrylamide (300/800/0,8 mm) permet dans le meilleur des cas de lire la séquence jusqu'à la position 300.

**BIBLIOGRAPHIE**

---

- ABRAMS H.D., ROHRSCHEIDER L.R. and EISENMAN R.N. (1982).  
Nuclear location of the putative transforming protein of avian myelocytomatosis virus.  
**Cell**, 29, 427-439.
- ALITALO K., SCHWAB M., LIN C.C., VARMUS H.E. and BISHOP J.M. (1983).  
Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cell from a human colon carcinoma.  
**Proc. Natl. Acad. Sci. USA**, 80, 1701-1711.
- ANDERSON S.M., HAYWARD W.S., NEIL B.G. and HANAFUSA H. (1980).  
Avian erythroblastosis virus produces two mRNAs.  
**J. Virol.** 36, 676-683.
- BARNEKOW A., SCHARTI M., ANDERS F. and BAUER H. (1982).  
Identification of a fish protein associated with a kinase activity and related to Rous sarcoma virus transforming protein.  
**Cancer Res.**, 42, 2429-2433.
- BARKER W.C. and DAYHOFF M.O. (1972).  
**Atlas of protein sequence and structure**, (National Biomedical Research Foundation, Washington, D.C.), Vol. 5, 105-110
- BERNHARD W. (1960).  
The detection and study of tumor viruses with the electron microscope.  
**Cancer Res.** 20, 712-727.
- BEUG H., KIRCHBACH A., DODERLEIN G., CONSCIENCE J.F. and GRAF T. (1979).  
Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation.  
**Cell**, 18, 375-390.
- BISTER K. and DUESBERG P. (1979).  
Structure and specific sequences of avian erythroblastosis virus RNA : Evidence for multiple classes of transforming genes among avian tumor viruses.  
**Proc. Natl. Acad. Sci. USA**, 76, 5023-5027.
- BOCCARA M., PLUQUET N., COLL J., ROMMENS C. and STEHELIN D. (1982).  
Characterization of c-lil, a chicken cellular sequence associated with a stock B 77 RSV.  
**J. Virol.**, 43, 925-931.
- BOULANGER P., POLONOVSKI J., TAYEAU F., MANDEL P. et BISERTE G. (1968).  
Biochimie médicale.  
**Masson et Cie éditeurs**, Fascicule III, page 269.

- CALAME K., KIM S., LALLEY P., HILL R., DAVIS M. and HOOD L. (1982).  
Molecular cloning of translocations involving chromosome 15 and the immunoglobulin C $\alpha$  gene from chromosome 12 in two murine plasmacytomas.  
**Proc. Natl. Acad. Sci. USA**, 79, 6994-6998.
- CAPON D.J., SEEBURG P.H., MC GRATH J.P., HAYTLICK J.S., EDMAN V., LEVINSON A.D. and GOEDDEL D.V. (1983).  
Activation of ki-ras-2 gene in human colon and lung carcinomas by two different point mutations.  
**Nature (London)**, 304, 507-513.
- CHINARD F.P., ENNS T. and NOLAN M.F. (1960).  
Contributions of bicarbonate in and out of dissolved CO<sub>2</sub> to expired CO<sub>2</sub> in dogs.  
**Am. J. Physiol.**, 198, 78-88.
- CLAVERIE J.M. (1984).  
A common philosophy and FORTRAN 77 software package for implementing and searching sequence databases.  
**Nucleic Acids Research**, 1, 1984, 397-403.
- COLEMAN J.E. (1984).  
Carbonic anhydrase : zinc and the mechanism of catalysis.  
**Ann. N. Y. Acad. Sci.**, 249, 26-48.
- COLL J., RIGHI M., DE TAISNE C., DISSOUS C., GEGONNE A. and STEHELIN D. (1983a).  
Molecular cloning of the avian acute transforming retrovirus MH2 reveals a novel cell-derived sequence (v-mil) in addition to the myc oncogene.  
**The Embo J.**, 2, 2189-2194.
- COLL J., SAULE S., MARTIN P., RAES M.B., LAGROU C., GRAF T., BEUG H., SIMON I.E. and STEHELIN D. (1983b).  
The cellular oncogenes c-myc, c-myb and c-erb are transcribed in defined types of avian hematopoietic cells.  
**Exp. Cell. Research**, 149, 151-162.
- COLLINS S.J. and GROUDINE M.T. (1982).  
Amplification of endogenous myc related DNA sequences in a human myeloid leukaemia cell line.  
**Nature (London)**, 298, 679-681.
- COLLINS S.J. and GROUDINE M.T. (1983).  
Rearrangement and amplification of c-abl sequences in the human chronic myelogenous leukemia cell line K-562.  
**Proc. Natl. Acad. Sci. USA**, 80, 4813-4817.

- COOPER G.M. and NEIMAN P.E. (1981).  
Two distinct candidate transforming genes of lymphoid leukemia virus induced neoplasm.  
**Nature (London)**, 292, 418-421.
- DALLA-FAVERA R., GELMANN E.P., MARTINOTTI S., FRANCINI G., PAPAS T.S., GALLO R.C. and WONG-STAAAL F. (1982).  
Cloning and characterization of different human sequences related to the oncogene (v-myc) of avian myelocytomatosis virus (MC 29).  
**Proc. Natl. Acad. Sci. USA**, 79, 6497-6501.
- DALLA-FAVERA R., MARTINOTTI I.S., GALLO R.C., BRIKSON J., CROCE C.M. (1983).  
Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas.  
**Science**, 219, 963-967.
- DALTON A.J., MELNICK J.L., BAUER H., BEAUDREAU G., BENTVELZEN P., BOLOGNESI R., GALLO R., GRAFFI A., HAGUENAU F., HESTON W., HUEBNER R., TODARO G. and HEINE V.I. (1974).  
The case for a family of reverse transcriptase viruses : retroviridae  
**Intervirology**, 4, 201-206.
- DHAR R., ELLIS R.W., SHIH T.Y., OROSZLAN S., SHAPIRO B., MAIZEL J., LOWY D.R. and SCOLNICK E.M. (1982).  
Nucleotide sequence of the p21 transforming protein of Harvey murine sarcoma virus.  
**Science**, 217, 934-936.
- DIAMOND A., COOPER G.M., RITZ J. and LANE M.A. (1983).  
Identification and molecular cloning of the human Blym transforming gene activated in Burkitt's lymphomas.  
**Nature (London)**, 305, 112-116.
- DOOLITTLE R.F. (1981).  
Similar amino-acid sequences : chance or common ancestry ?  
**Science**, 214, 149-159.
- DOWNWARD J., YARDEN Y., MAYES F., SCRACE G., TOTTY N., STOCKWER L.P., ULLRICH A., SCHLESSINGER J. and WATERFIELD M.D. (1984).  
Close similarity of epidermal growth factor receptor and v-erbB oncogene protein sequence.  
**Nature (London)**, 307, 521-525.
- DUESBERG P.H. (1983).  
Retroviral transforming genes in normal cells.  
**Nature (London)**, 304, 219-226.

- ENGLUND P.T. (1971).  
Analysis of nucleotide sequences at 3' termini of duplex deoxyribonucleic acid with the use of the T<sub>4</sub>-deoxyribonucleic acid polymerase.  
**J. Biol. Chem.**, 246, 3269-3276.
- ENGLUND P.T. (1972).  
The 3' terminal nucleotide sequences of T7 DNA.  
**J. Mol. Biol.**, 66, 209-223.
- FERREL R.E., STROUP S.K., TANIS R.J. and TASHIAN R.E. (1978).  
Amino acid sequence of rabbit carbonic anhydrase II.  
**Biochim. Biophys. Acta**, 533, 1-11.
- FRYKBERG L., PALMIERI S., BEUG H., GRAF T., HAYMAN A.M. and VENNSTROM B. (1983).  
Transforming capacities of avian erythroblastosis virus mutants deleted in erbA and erbB oncogenes.  
**Cell**, 32, 227-228.
- FUNG J.K., LEWIS W.G., CRITTENDEN L.B. and KUNG H.J. (1983).  
Activation of the cellular oncogene c-erbB by LTR insertion : molecular basis for induction of erythroblastosis by avian leukosis virus.  
**Cell**, 33, 357-368.
- GALIBERT F., SEDA J.W. and ZIFF E.B. (1974).  
Direct determination of DNA nucleotide sequence : structure of a fragment of bacteriophage  $\phi$ x 174 DNA.  
**J. Mol. Biol.**, 87, 337-407.
- GALIBERT F., MANDART E., FITOUSSI F., THIOLLAIS P., CHARNAY P. (1979).  
Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in the E. Coli.  
**Nature (London)**, 281, 646-650.
- GAZZOLO L., SAMARUT J., BOUABDELLI M. and BLANCHET J.P. (1980).  
Early precursors in the erythroïd lineage are the specific target cells of avian erythroblastosis virus **in vitro**.  
**Cell**, 22, 683-691.
- GILDEN V.R., RICE N.R. and MC ALLINSTER R.M. (1984).  
Oncogenes.  
**Gene Anal. Techn.**, 1, 23-33.
- GILMORE T., DE CLUE J.E. and MARTIN S. (1985).  
Protein phosphorylation at tyrosine is induced by the v-erbB protein gene product **in vitro** and **in vivo**.  
**Cell**, 40, 609-618.

- GOFF S.P., D'EUSTACHIO P., RUDDLE F.H. and BALTIMORE D. (1982).  
Chromosomal assignment of the endogenous proto-oncogene c-abl.  
**Science**, 218, 1317-1319.
- GRAF T. and BEUG H. (1978).  
Avian leukemia viruses interaction with their target cells **in vivo** and **in vitro**.  
**Biochim. Biophys. Acta**, 516, 269-299.
- GRAF T. and STEHELIN D. (1982).  
Avian leukemia viruses : oncogenes and genome structure.  
**Biochim. Biophys. Acta**, 651, 245-271.
- GROFFEN J., HEISTERKAMP N., STEPHENSON J.R., VAN KESSEL A.G., DE KLEIN A., GROSVEDL G. and BOOTSMA D. (1983).  
c-sis translocated from chromosome 22 to chromosome 9 in chronic myelocytic leukemia.  
**J. Exp. Med.**, 158, 9-15.
- HARRIS L.J., DIEUSTACHIO P., RUDDLE P.H., MARCU K.B. (1982).  
DNA sequence associated with chromosome translocations in mouse plasmacytomas.  
**Proc. Natl. Acad. Sci. USA**, 79, 6622-6626.
- HAYMAN J.M., GARY M.R., STORN K. and KITCHENER G., GRAF T. and BEUG H. (1983).  
Identification and characterization of the avian erythroblastosis virus erbB gene product as a membrane glycoprotein.  
**Cell**, 32, 579-588.
- HAYMAN J.M. and BEUG H. (1984).  
Identification of a form of the avian erythroblastosis virus erbB gene product at the cell surface.  
**Nature (London)**, 309, 460-462.
- HAYWARD W.S., NELL B.G. and ASTRIM B.M. (1981).  
Activator of a cellular onc gene by promotor insertion in ALV-induced lymphoïd leukosis.  
**Nature (London)**, 290, 475-480.
- HELDIN C.H. and WESTERMAERK B. (1984).  
Growth factors : mechanism of fonction and relation to oncogenes.  
**Cell**, 37, 9-20.
- HOPP T.P. and WOODS K.R. (1981).  
Prediction of protein antigenic determinants from amino acid sequences.  
**Proc. Natl. Acad. Sci. USA**, 78, 3824-3828.

- JANSSON M., PHILIPSON L. and VENNSTROM B. (1983).  
Isolation and characterization of multiple human genes homologous to the oncogenes of avian erythroblastosis virus.  
**The Embo J.**, 2, 561-565.
- JONES M., BOSSELMAN R.A., HOORN F.A., BERNIS A., FAN H. and VERMA I.M. (1980).  
Identification and molecular cloning of Moloney mouse sarcoma virus specific sequences from uninfected mouse cells.  
**Proc. Natl. Acad. Sci. USA**, 77, 2651-2655.
- KELLY K., COCHRAN B.H., STILLES C.D. and LEDER P. (1983).  
Cell specific regulation of the c-myc gene by lymphocyte mitogens and platelet derived growth factors.  
**Cell**, 35, 603-610.
- DE KLEIN A., VAN KESSEL A.G., GROSVELD G., BARTRAM C.R., HAGEMEISER A., BOOTSMA D., SPURR N.K., HEISTER KAMP N., GROFFEN J. and STEPHENSON J.R. (1982).  
A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia.  
**Nature (London)**, 300, 765-767.
- KREBS E.G. and BEAVA J.A. (1979).  
Phosphorylation - dephosphorylation of enzymes.  
**Annu. Rev. Biochem.**, 48, 723-959.
- KRIS R.M., GULLICK W., WATERFIELD M.D., ULLRICH A., FRIDKIN M. and SCHLESSINGER J. (1985).  
Antibodies against a synthetic peptide as a probe for the kinase activity of the avian EGF receptor and v-erbB protein.  
**Cell**, 40, 619-625.
- LAI M.M.C., HU S.S.F. and VOGT P.K. (1979).  
Avian erythroblastosis virus : transformation specific sequences form a contiguous segment of 3,25 kb located in the middle of the 6 kb genome.  
**Virology**, 97, 366-377.
- LAI M.M.C., NEIL J.C. and VOGT P.K. (1980)  
Cell - free translation of avian erythroblastosis virus RNA yields two specific and distinct proteins with molecular weights of 75000 and 40000.  
**Virology**, 100, 475-483.
- LAND H., PARADA L.F. and WEINBERG R.A. (1983).  
Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes.  
**Nature (London)**, 304, 596-602.

LEPRINCE D., GEGONNE A., COLL J., DE TAISNE C., SCHNEEBERGER A., LAGROU C. and STEHELIN D. (1983).

A putative second cell - derived oncogene of the avian leukemia retrovirus E26.

**Nature (London)**, 306, 395-397.

LEVY J.P. (1984)

Les oncogènes.

**Nouv. Rev. Fr. Hematol.**, 26, 1-11.

LILJAS A., KANNAN K.K., BERGSTEN P.C., WAARA I., FRIDBORG K., STRANNBERG B., CARLBOM V., JARUP L., LOUGREN S. and PETER M. (1972).

Crystal structure of human carbonic anhydrase C.

**Nature (London) New Biol.**, 235, 131-137.

LONNERHOLM G. (1984).

Immunohistochemical localization of human carbonic anhydrase isoenzymes.

**Ann. N.Y. Acad. Sci.**, 429, 359-368.

MARSHALL R.D. (1974).

The nature and metabolism of the carbohydrate peptide linkage of glycoproteins.

**Biochem. Soc. Symp.**, 40, 17-26.

MAXAM A. and GILBERT W. (1977).

A new method for sequencing DNA.

**Proc. Natl. Acad. Sci. USA**, 74, 3642-3646.

MC COY M.S., TOOLE I.J., CUNNINGHAM J.M., CHANG E.H., LOWY D.R. and WEINBERG R.A. (1983).

Characterization of a human colon/lung carcinoma oncogene.

**Nature (London)**, 302, 79-81.

MC GRATH J.P., CAPON D.J., SMITH D.H., CHEN E.Y., SEEBURG P.H., GOEDDEL D.V. and LEVINSON A.D. (1983).

Structure and organization of the human ki-ras proto-oncogene and a related processed pseudogene.

**Nature (London)**, 304, 501-506.

MESSING J., GRONNENBORN B., MULLER-HILL B. and HORFSCHNEIDER P.H. (1977)

Filamentous coliphage M13 as a cloning vehicle.

**Proc. Natl. Acad. Sci. USA**, 74, 560-564.

MOUNT M.S. (1982).

A catalogue of splice junction.

**Nucleic Acid. Res.**, 10, 460-472.

MOELLING K., PFAFF E., BEUG H., HEIMLING P., BUNTE T., SCHALLER H.E. and GRAF T. (1985).

DNA binding activity is associated with purified myb proteins from AMV and E26 viruses and is temperature-sensitive for E26 ts mutants.  
**Cell**, 40, 983-990.

MONTGOMERY K.T., BIEDLER J.L., SPENGLER B.A. and MELERA P.W. (1982).

Specific DNA sequence amplification in human neuroblastoma cells.  
**Proc. Natl. Acad. Sci. USA**, 80, 5724-5728.

O'BRIEN S.J., NASH W.G., GOODWIN J.L., LOWY D.R. and CHANGE E.H. (1983).

Dispersion of the ras family of transforming genes to four different chromosomes in man.  
**Nature (London)**, 302, 839-842.

PAYNE G.S., BISHOP J.M. and VARMUS H.E. (1982).

Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas.  
**Nature (London)**, 295, 209-214.

PAWSON T., GUYDEN J., KUNG T.H., RADKE K., GILMORE T. and MARTIN G.S. (1980 a).

A strain of Fujinami sarcoma virus which is temperature sensitive in protein phosphorylation and cellular transformation.  
**Cell**, 22, 767-777.

PAWSON T. and STEVEN G.S. (1980 b).

Cell free translation of avian erythroblastosis virus RNA.  
**Journal of virology**, 34, 280-284.

PONCZ M., SOLOWIEJCZYKI D., BALLANTINE M., SCHWARTZ E. and SURREY S. (1982).

Nonrandom DNA sequence analysis in bacteriophage M13 by the dideoxychain termination method.  
**Proc. Natl. Acad. Sci. USA**, 79, 4298-4302.

PRIVALSKY M.L. and BISHOP J.M. (1982).

Proteins specified by avian erythroblastosis virus : coding region localization and identification of a previously undetected v-erbB polypeptide.  
**Proc. Natl. Acad. Sci. USA**, 79, 3958-3962.

PRIVALSKY M.L., SEALY L., BISHOP J.M., MC GRATH J.P. and LEVINSON A.D. (1983).

The product of the avian erythroblastosis virus erbB locus is a glycoprotein.  
**Cell**, 32, 1257-1267.

- PUTNAM F.W. (1960).  
The Plasma Proteins : Isolation, characterization and function.  
**Acad. Press, New-York and London, Volume 1, 241-271.**
- RABBITS T.H., FORSTER A., HAMLIN P. and BAER R. (1984).  
Effect of somatic mutation within translocated c-myc genes in Burkitt's lymphoma.  
**Nature (London), 309, 592-597.**
- RAINES M.A., LEWIS W.G., CRITTENDEN L.B. and KUNG H.J. (1985).  
c-erbB activation in avian leukosis virus induced erythroblastosis : clustered integration sites and the arrangement of provirus in the c-erb alleles.  
**Proc. Natl. Acad. Sci. USA, 82, 2287-2291.**
- RECHAVI G., GIVOL D. and CANAAVI E. (1982).  
Activation of a cellular oncogene by DNA rearrangement : possible involvement of an IS-like element.  
**Nature (London), 300, 607-611.**
- REDDY E.P. (1983).  
Nucleotide sequence analysis of the T<sub>24</sub> human bladder carcinoma oncogene.  
**Science, 220, 1061-1069.**
- ROBERTSON M. (1983).  
Paradox and paradigm : the message and meaning of myc.  
**Nature (London), 306, 733-736.**
- ROUSSEL M., SAULE S., LAGROU C., ROMMENS C., BEUG H., GRAF T. and STEHELIN D. (1979).  
Three new types of oncogene of cellular origin specific for haematopoietic cell transformation.  
**Nature (London), 281, 452-455.**
- ROYER-POKORA B., BEUG H., CLAVIEZ H., VINKHARDT H.J., FRIIS R.R. and GRAF T. (1978).  
Transformation parameters in chicken fibroblasts transformed by AEV and MC 29 avian leukemia viruses.  
**Cell, 13, 751-760.**
- SAMARUT J. and GAZZOLO L. (1982).  
Target cells infected by avian erythroblastosis virus differentiate and become transformed.  
**Cell, 28, 921-929.**
- SANGER F.K. and COULSON A.R. (1975).  
A rapid method for determining sequences in DNA by primer synthesis with DNA polymerase.  
**J. Mol. Biol, 94, 441-448.**

SANGER F.K., AIRG M., BARRELL B.G., BROWN N.L., COULSON A.R., FIDDES J.C., HUTCHISON C.A., SLUCOMBE P.M. and SMITH M. (1977a).  
Nucleotide sequence of bacteriophage  $\phi$ x 174 DNA.  
**Nature (London)**, 265, 687-690.

SANGER F.K., NICKLEN R. and COULSON A.R. (1977b).  
DNA sequencing with chain terminating inhibitors.  
**Proc. Natl. Acad. Sci. USA**, 74, 5463-5467.

SAULE S., ROUSSEL M., LAGROU C. and STEHELIN D. (1981).  
Characterization of the oncogene erb of avian erythroblastosis virus and its cellular progenitor.  
**J. Virol.**, 38, 409-419.

SCHWAB M., ALITALO K., VARMUS H.E. and BISHOP J.M. (1983).  
A cellular oncogene (c-ki-ras) is amplified, overexpressed and located within karyotypic abnormalities in mouse adrenocortical tumour cells.  
**Nature (London)**, 303, 497-501.

SCHWARTZ D.E., TIZARD R. and GILBERT W. (1983).  
Nucleotide sequence of Rous sarcoma virus.  
**Cell**, 32, 853-869.

SCIACKY M., LIMOZIN D., FILIPPI-POVEAU J., GULIAN M. et LAURENT-TABUSSE G. (1976).  
Structure primaire de l'anhydrase carbonique érythrocytaire bovine CI.  
**Biochimie**, 58, 1071-1082.

SEIKI M., HATTORI S., HIRAYAMA Y. and YOSHIDA M. (1983).  
Human adult T-cell leukemia virus : complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA.  
**Proc. Natl. Acad. Sci. USA**, 80, 3618-3622.

SERGEANT A., SAULE S., LEPRINCE D., BEGUE A., ROMMENS C. and STEHELIN D. (1982).  
Molecular cloning and characterization of the chicken DNA locus related to the oncogene erbB of avian erythroblastosis virus.  
**The Embo J.**, 1, 237-242.

SHEER D., HIORNS L.R., STANLEY K.F., GOODFELLOW .N., SWALLOW D.M., POVEY S., HEISTERKAMP N., GROTTEN J., STEPHENSON J.R. and SOLOMON E. (1983).  
Genetic analysis of the 15 : 17 chromosome translocation associated with acute promyelocytic leukemia.  
**Proc. Natl. Acad. Sci. USA**, 80, 5007-5011.

SHEINESS D., VENNSTROM B. and BISHOP J.M. (1981).  
Virus specific RNAs in cells infected by avian myelocytomatosis virus and avian erythroblastosis virus : modes of oncogene expression.  
**Cell**, 23, 291-300.

- SHILO B.Z. and WEINBERG R.A. (1981).  
DNA sequences homologous to vertebrate oncogenes are conserved in *Drosophila melanogaster*.  
**Proc. Natl. Acad. Sci. USA**, 78, 6789-6792.
- SIMON M.A., KORNBERG T.B. and BISHOP J.M. (1983).  
Three loci related to the src oncogene and tyrosine specific protein kinase activity in *Drosophila*.  
**Nature (London)**, 302, 837-839.
- SPURR N.R., SOLOMON E., JANSSON M., GROOM-FELLOW P., BODNER W.F. and VENNSTROM B. (1984).  
Chromosomal localisation of the human homologues to the oncogenes erbA and erbB.  
**The Embo J.**, 3, 159-163.
- STADEN R. (1979).  
A strategy of DNA sequencing employing computer programs.  
**Nucleic Acids. Res.**, 6, 2601-2610.
- STEHELIN D., VARMUS H.E., BISHOP J.M. and VOGT P.K. (1976).  
DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA.  
**Nature (London)**, 260, 170-173.
- TABIN C.J., BRADLEY S.M., BARGMANN C.I., WEINBERG R.A., PAPAGEORGE A.G., SCOLNICK E.M., DHAR R., LOWY D.R. and CHANG E.M. (1982).  
Mechanism of activation of a human oncogene.  
**Nature (London)**, 300, 143-149.
- TANIS R.J., FERREL R.E. and TASHIAN R.E. (1974).  
Amino acid sequence of sheep carbonic anhydrase C.  
**Biochim. Biophys. Acta.**, 371, 534-548.
- TAPAROWSKY E., SUARD Y., FASANO O., SHIMIZU K., GOLDFARB M. and WIGLER M. (1982).  
Activation of the T<sub>4</sub> bladder carcinoma transforming gene is linked to a single amino acid change.  
**Nature (London)**, 300, 762-765.
- TASHIAN R.E., GOODMAN M., TANIS R.J., FERRELL R.E. and OSBORNE R.A. (1975).  
Evolution of carbonic anhydrase isoenzymes.  
**Isozymes**, Volume 4, 207-223.

TATCHELL K., CHALEFF D.T., DE FEO-JONE S.D. and SCOLNICK E.M. (1984).

Requirement of either of a pair of ras-related genes of *Saccharomyces cerevisiae* for spore viability.

**Nature (London)**, 309, 523-527.

TEMIN H.M. and BALTIMORE D. (1972).

RNA directed synthesis and RNA tumor viruses.

**Adv. Virus Res.**, 17, 129-186.

TOMASHEFSKI J.F., CHINN H.I. and CLARK R.T. (1974).

Effect of carbonic anhydrase inhibition on respiration.

**Am. J. Physiol.**, 177, 451-454.

TSUCHIDA N., RYDER T. and OHTSUBO E. (1982).

Nucleotide sequence of the origin encoding the p21 transforming protein of Kirsten murine sarcoma virus.

**Science**, 217, 937-938.

ULLRICH A., COUSSENS L., HAYFLICK J.S., DULL T.J., GRAY A., TAMA W., LEE J., YARDEN Y., UBERMANN T.A. SCHLESSINGER J., DOWNWARD J., MAYES E.L.V., WHITTLE N., WATERFIELD M.D. and SEEBURG P.H. (1984).

Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A 431 epidermoid carcinoma cells.

**Nature (London)**, 309, 418-425.

VAN BEVEREN C., VAN STRAATEN F., CURRAN T., MULLER R. and VERMA I.M. (1983).

Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini.

**Cell**, 32, 1241-1255.

VAN SLYKE D.D. and HAWKINS J.A. (1930).

Studies of gas and electrolyte equilibria in blood. The evolution of carbon dioxide from blood and buffer solutions.

**J. Biol. Chem.**, 87, 265-279.

VENNSTROM B., FANSHIER L., MOSCOVICI C. and BISHOP J.M. (1980).

Molecular cloning of avian erythroblastosis virus genome and recovery of oncogenic virus by transfection of chicken cells.

**J. Virol.**, 36, 575-585.

VENNSTROM B. and BISHOP J.M. (1982)

Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus.

**Cell**, 28, 135-143.

- VERMA I.M. (1977).  
The reverse transcriptase.  
*Biochim. Biophys. Acta*, 473, 1-38.
- WAIN-HOBSON S., SONIGO P., DANOS O., COLE S. and ALIZON M. (1985).  
Nucleotide sequence of the AIDS virus LAV.  
*Cell*, 40, 9-17.
- WATERFIELD M.D., SCRACE G.T., WHITTLE N., STROOBANT P., JOHNSON A., WASTESON A., WESTERMARK B., HELDIN C.H., HUANG J. and DEVEL T.E. (1983).  
Plated derived growth factor is structurally related to the putative transforming protein p28<sup>sis</sup> of simian sarcoma virus.  
*Nature (London)*, 304, 35-39.
- WENDORFF K.M., NISHITA T., JABUSH R. and DEUTSH H. (1985).  
The sequence of equine muscle carbonic anhydrase.  
*J. of Biol. Chem.*, 260, 6129-6132.
- WEN-HWA L., CHIH-PING L. and DUESBERG P. (1982).  
DNA clone of avian Fujinami sarcoma virus with temperature sensitive transforming function in mammalian cells.  
*J. Virol.*, 44, 401-412.
- WISTRAND J.P. (1981).  
The importance of carbonic anhydrase B et C for the unloading of CO<sub>2</sub> by the human erythrocyte.  
*Acta Physiol. Scand.*, 113, 417-426.
- WONG T.C. and LAI M.C. (1981).  
Avian reticuloendotheliosis virus contains a new class of oncogene of turkey origin.  
*Virology*, 111, 289-293.
- YAMAMOTO T., NISHIDA T., MIYAJIMA N., KAWAY O. and TOYOSHIMA K. (1983a).  
The erbB gene of avian erythroblastosis virus is a member of the src gene family.  
*Cell*, 35, 71-78.
- YAMAMOTO T., HIHARA H., NISHIDA T., KAWAI S. and TOYOSHIMA K. (1983b).  
A new avian erythroblastosis virus AEV-H carries erbB gene responsible for the induction of both avian erythroblastosis and sarcomas.  
*Cell*, 34, 225-232.

ZHANG S., ZECH L. and KLEIN G. (1982).

High-resolution analysis of chromosome marker in Burkitt lymphoma cell lines.

**Int. J. Cancer**, 29, 153-157.

ZIFF E.B., SEDAT J.W. and GALIBERT F. (1973).

Determination of the nucleotide sequence of a fragment of bacteriophage  $\phi_x$  174 DNA.

**Nature (London)**, 241, 34-37.

ARTICLES

## **Sequencing the *erbA* Gene of Avian Erythroblastosis Virus Reveals a New Type of Oncogene**

B. Debuire\*, C. Henry, M. Benaissa, G. Biserte, J. M. Claverie, S. Saule, P. Martin, and D. Stehelin

## Sequencing the *erbA* Gene of Avian Erythroblastosis Virus Reveals a New Type of Oncogene

**Abstract.** Avian erythroblastosis virus (AEV) contains two distinct oncogenes, *erbA* and *erbB*. The *erbB* oncogene, which is homologous to a portion of the epidermal growth factor receptor, is related to the *src* family of oncogenes and efficiently transforms erythroblasts, whereas *erbA* potentiates the effects of *erbB* by blocking the differentiation of erythroblasts at an immature stage. This "potentiator" was sequenced; the amino acid sequence deduced from it was clearly different from the sequences of other known oncogene products and was related to carbonic anhydrases. These enzymes participate in the transport of carbon dioxide by erythrocytes, the precursors of which are main targets of avian erythroblastosis virus. A *src*-related oncogene such as *erbB* in synergy with an activated specific cell-derived gene such as *erbA* can profoundly affect early erythroid differentiation.

Two distinct oncogenes can act in synergy or complementation to transform normal cells (1-3). Some retroviruses contain two specific cell-derived sequences in their genome (4-6), probably leading to their selection as highly transforming viruses.

Avian erythroblastosis virus (AEV), a defective leukemia virus that transforms mainly erythroblasts (7, 8), contains two

independent oncogenes, *erbA* and *erbB* (9-14), expressed in transformed cells from two distinct messenger RNA's (mRNA's) (13, 15, 16). A 5.4-kilobase (kb) mRNA codes for a cytoplasmic P75<sup>gag-erbA</sup> fusion protein (17, 18) while a subgenomic spliced 3.5-kb mRNA produces a 65 to 68-kilodalton (kD) membrane-associated glycoprotein (19-21). Both *erbA* and *erbB* oncogenes are ho-

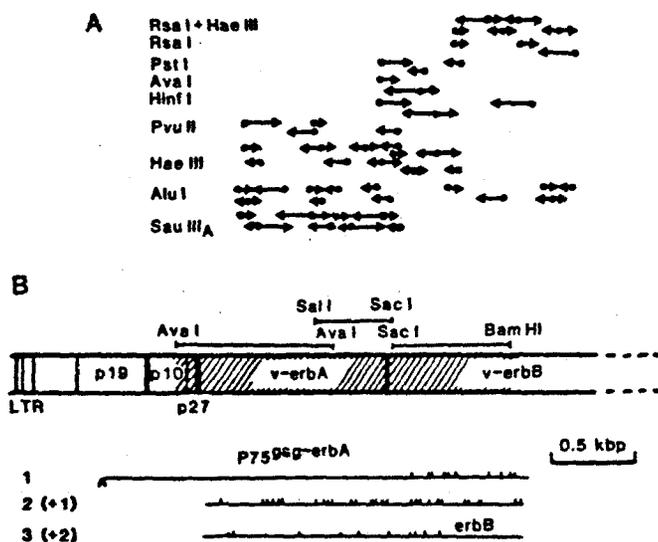


Fig. 1. (A) Sequencing strategy. The complete AEV DNA is represented linearized by Eco RI (12). The *erbA* sequence was obtained from three subclones, and solid bars indicate the *gag-erb* DNA inserts: 1-kbp Ava I-Ava I, 0.5-kbp Sal I-Sac I, and 0.7-kbp Sac I-Bam HI. Gel-purified restriction fragments used for DNA sequencing (restriction enzymes in the column at left) were labeled at their 5' termini with [ $\gamma$ -<sup>32</sup>P]adenosine triphosphate and polynucleotide kinase

and sequenced by the method of Maxam and Gilbert (31). Closed circles indicate the cleavage points inside of each insert and horizontal arrows the direction and length of the sequenced DNA strands. (B) Reading frames in the genome of AEV. The DNA sequence was determined (hatched). The recombination of *erbA* with the viral genome occurred at the sites shown as thick vertical bars in the diagram of AEV DNA. p27 denotes the remaining coding domain for the p27 protein of the *gag* gene, which was truncated during the recombination events leading to capture of the *erb* insert. The open reading frame for P75<sup>gag-erbA</sup> and the two other reading frames in *erbA* are shown below the AEV DNA. A different open reading frame was detected for *erbB*. Vertical bars denote stop codons. LTR indicates a long terminal repeat.

mologous to avian and mammalian chromosomal DNA sequences (*c-erbA* and *c-erbB*) (4, 13, 22) that may play a role in the metabolism of normal cells, possibly in the differentiation of immature red blood cells (23).

Wild-type AEV-transformed erythroblasts are tightly blocked in their maturation (at the colony-forming unit stage) (24, 25). Studies involving deletion mutants in *erbA* and *erbB* indicate that *erbB*, in vitro as well as in vivo, yields transformed erythroblast-like cells at different stages of maturation, whereas *erbA* alone induces no transformation.

Thus *erbA* potentiates the transforming activity of *erbB* and appears to be responsible for the early blockage of cell differentiation within the erythroid lineage (26-29).

Although several oncogenes have been extensively studied, "potentiators" such as *erbA* have not been studied. Therefore we analyzed the nucleotide sequence of *erbA*. Computer analysis showed that the deduced amino acid sequence of the *erbA* stretch of P75<sup>RAR-erbA</sup> is clearly different from the sequences of other reported transforming proteins. Thus *erbA* appears to represent a distinct new member of the oncogene families.

resent a distinct new member of the oncogene families.

A molecular clone of AEV (p-AEV 11) (30) with biological activity was used to sequence the *erbA* oncogene by the Maxam and Gilbert procedure (31). Figure 1A shows the strategy used to produce the nucleotide sequence (Fig. 2) defining the *erbA* boundaries (Fig. 1) between the structural *gag* gene upstream and the second oncogene, *erbB*, downstream. The left boundary of *erbA* was assessed by comparison to the nucleotide sequence of the Prague-C strain (Pr-C) of Rous sarcoma virus (32). The

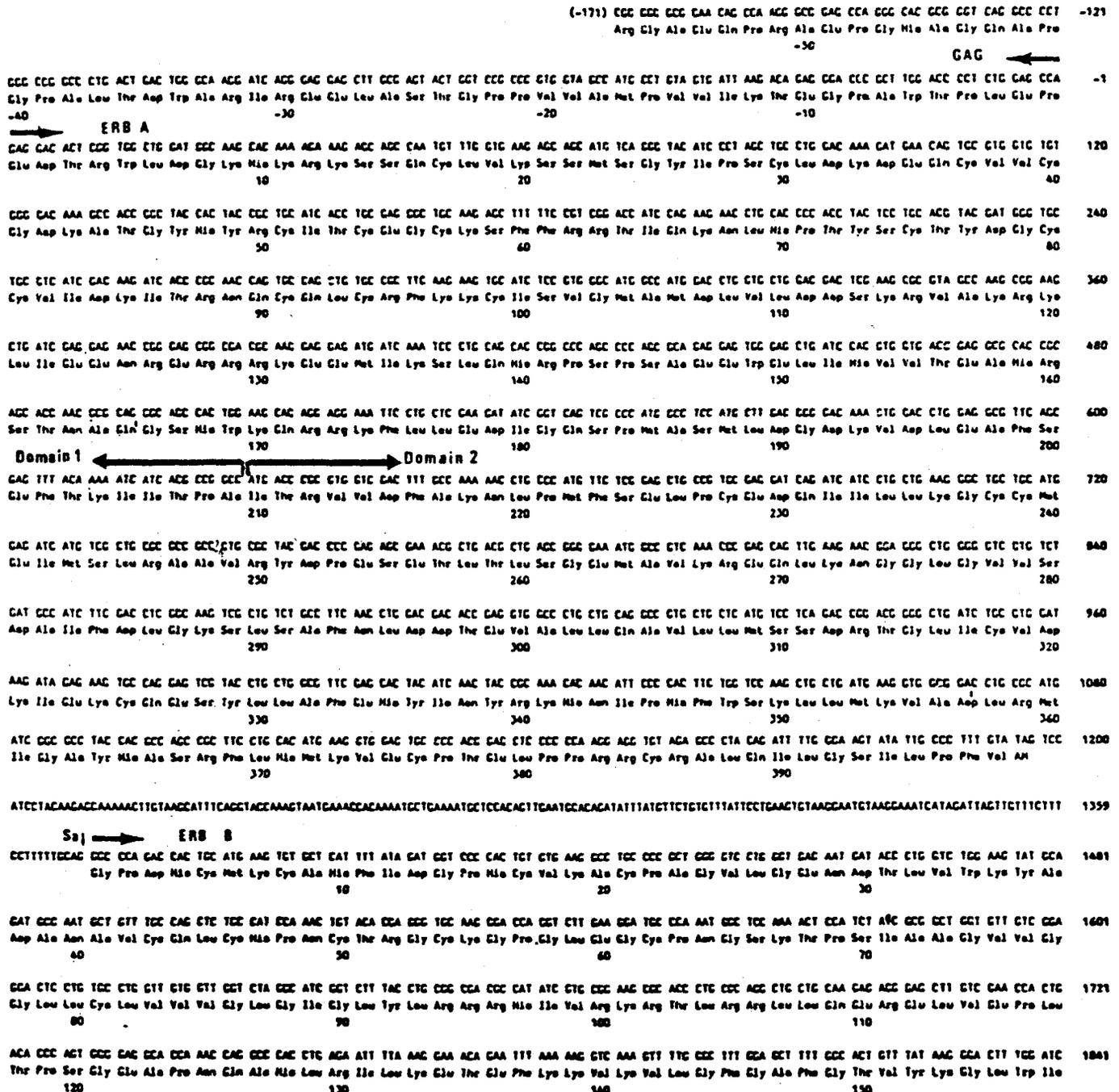


Fig. 2. Nucleotide sequence of the *erbA* gene and of the 5' end of *erbB*. The sequence of 1202 nucleotides encompassing the *erbA* gene is shown. The deduced amino acid sequence of the *erbA* domain of P75<sup>RAR-erbA</sup> is indicated from nucleotides 1 to 1194 and the deduced amino acid sequence of the 5' end of *erbB* is indicated from the putative splice acceptor site (Sa, nucleotide 1370) to nucleotide 1841. Nucleotides are numbered in the right column and every tenth amino acid is numbered.

...va I site at the extreme left of our sequence (Fig. 1) is within the coding domain for the virion protein p10<sup>gag</sup>. The nucleotide sequences of Pr-C and AEV are almost identical from this position rightward to residue 1 in the AEV sequence, where complete divergence of the two sequences marks the point of insertion of *erbA*. This insertion is located within the coding domain for the virion protein p27<sup>gag</sup> (32).

The reading frame of *erbA* continues uninterrupted from the p27<sup>gag</sup> reading frame at position +1 until it is terminated by an amber codon at position 1195; this is followed by a stretch of noncoding sequences up to a putative splice acceptor (Sa in Fig. 2) site (33), TTTCC-TTTTGCAG.G (T, thymine; C, cytosine; G, guanine; A, adenine) for the *erbB* gene at nucleotide 1370. It is unlikely that the other two *erbA* reading frames are used because they are frequently closed by termination codons (vertical bars in Fig. 1B). The consensus splice acceptor site at position 1370 could generate the subgenomic *erbB* mRNA in an open reading frame different from that of *erbA*, and we present the deduced partial amino acid sequence of this reading frame (Fig. 2). Thus the *erbB* product may start at the AUG (U, uracil) codon of *gag* used to produce P75<sup>gag-erbA</sup> (assuming that AEV and Pr-C use the same splice donor site in *gag*); the two proteins then share a few common amino acids at their NH<sub>2</sub> terminus. Alternatively, the *erbB* product could initiate at the in-frame AUG codon at position 1386 in

our sequence (Fig. 2). The structure of *erbB* is virtually identical, in the region we studied, to the one recently described for another independent isolate of AEV, namely AEV-H, which lacks the *erbA* oncogene (34). Importantly, *erbB* is closely related to the *src* gene of avian sarcoma virus [(34) and our sequence data] and to a portion of the epidermal growth factor receptor (35).

The deduced amino acid sequence of *erbA* (398 amino acid residues) is shown under the nucleotide sequence in Fig. 2. The calculated molecular size of the *erbA* polypeptide is 45.4 kD, yielding for P75<sup>gag-erbA</sup> (36) a deduced molecular size of 72 kD. Cellular adenosine 3',5'-monophosphate-dependent protein kinases phosphorylate serine or threonine residues within sequences x-y-z-(Ser or Thr) (37), where x and y are basic residues; two such serine residues are found in the *erbA* stretch at positions 14 and 15. There are no potential glycosylation sites Asn-x-(Thr or Ser) (38) in this polypeptide. Thus *erbA* may be phosphorylated but not glycosylated. Two domains can be defined within the *erbA* protein; notable features of the amino acid composition are found in the first domain of *erbA* (amino acids 1 to 209 in Fig. 2) with elevated levels of cysteine and basic residues (10 percent and 21 percent, respectively, within amino acids 1 to 131), in strong contrast with the rest of the molecule. Hydrophilicity studies with the Hopp-Woods procedure (39) indicate that *erbA* encodes a relatively hydrophilic product (the value for the *erbA* poly-

peptide is 0.1, whereas the average protein has a net hydrophilicity of 0.07), as expected for a cytoplasmic protein (21).

A search for similarities between the *erbA* protein and the other proteins contained in the protein data bases (40, 41) revealed no obvious relationships with other oncogenes. Thus *erbA* may be a genuinely new member of this class of proteins. Domain 1 showed no salient homology with other known proteins, but a relatedness was found between the carboxyl terminal half of *erbA* (domain 2 in Fig. 2) and the carbonic anhydrase family (Fig. 3). The homology begins with *erbA* amino acid residue 219 (amino acid residue 10 for the carbonic anhydrases) and extends across 180 residues up to the carboxyl terminus of the molecule (residue 398 of *erbA* or residue 195 of the carbonic anhydrases, which continue for another 64 residues). When compared with the four mammalian carbonic anhydrases II available in the data banks we used, the overall homology was 27 percent for 180 amino acids, clustered in some portions of the molecule. For example, close to one of the known active sites of carbonic anhydrases (42), the homology increases to more than 51 percent (*erbA* residues 248 to 274), although the two residues reported to be important in the active site (His<sup>63</sup> and Asn<sup>66</sup>) (42) are not found in the *erbA* protein. This may imply a structural, but not a functional relationship. Another example is *erbA* residues 310 to 330, with more than 40 percent homology. This homology between the

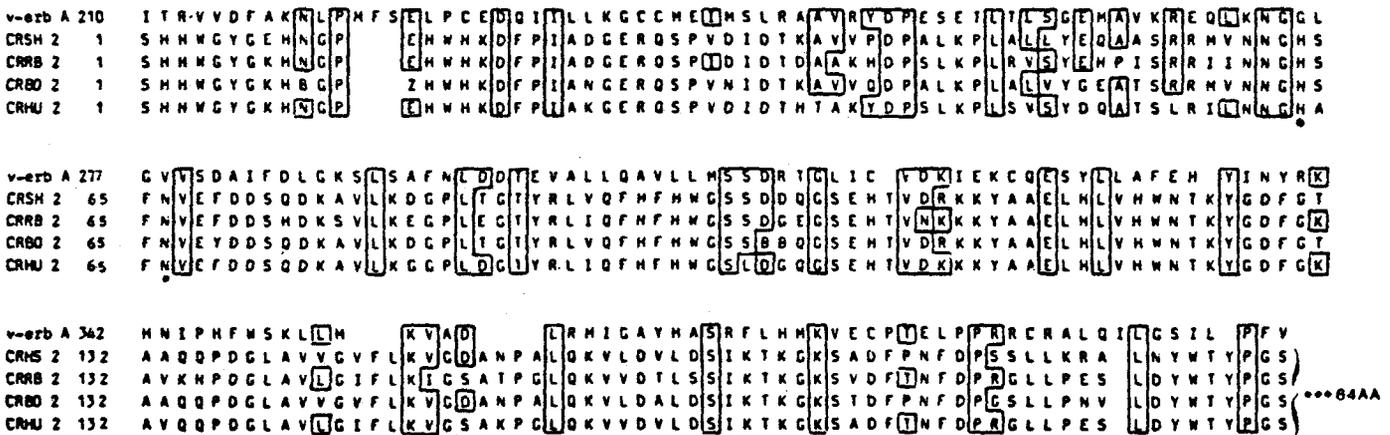


Fig. 3. Relatedness of *erbA* domain 2 to the carbonic anhydrase family. Using the sequence of the putative *erbA* polypeptide, we performed exhaustive homology searches on the NBRF (41) and NEWAT (40) protein data banks with our computer system (47). These searches revealed a homology encompassing 180 amino acid residues between *erbA* domain 2 and the carbonic anhydrase family. The significance of this homology was further assessed by an alignment program (48) and adapted by one of us (J.M.C.). The amino acid sequences of the predicted *erbA* polypeptide and of sheep (CRSH2) (49), rabbit (CRRB2) (50), bovine (CRBO2) (51), and human (CRHU2) (42) carbonic anhydrases II were compared. The optimal alignment was found when we used a deletion weight of 1 for each gap plus 0.2 times the number of residues in each gap. In this case, the Needleman-Wunsch similarity value (52) was found 3.8 standard deviation units above the mean of 30 shuffled sequences. Scores that are 3 or more standard deviation units above the mean can reasonably be expected to represent authentic relationships (53). Common residues are boxed. Closed circles indicate amino acids known to be important in the active site of carbonic anhydrases. The one-letter symbols for the amino acids are A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; Q, glutamine; E, glutamic acid; G, glycine; H, histidine; I, isoleucine; L, leucine; K, lysine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; V, valine; B, asparagine or aspartic acid; and Z, glutamine or glutamic acid.

*erbA* product and, for example, human carbonic anhydrase II (CRHU 2 in Fig. 3) is 3.8 standard deviation units above the mean of 30 shuffled sequences, considerably higher than the cutoff value of 3, which indicates, with statistical significance, an evolutionary relatedness (40) (for details of the computer program see Fig. 3). Moreover, in human carbonic anhydrase I, which has a 61 percent homology with human carbonic anhydrase II, most of the homologies depicted in Fig. 3 are conserved (not shown). Although *erbA* domain 2 is related to carbonic anhydrases, we showed that it is not the enzyme itself. Using a viral *erbA* DNA probe corresponding to domain 2, we screened both chicken and human genomic DNA libraries and found that normal DNA in both species contained a similar locus. The human cellular *erbA* domain 2 was cloned, and a nucleotide sequence was determined (corresponding to nucleotides 675 to 936 in the *erbA* domain 2 of Fig. 2). Homology at the nucleotide level was 83 percent, and the deduced amino acid sequence in the corresponding open reading frame showed 96 percent homology (the two other reading frames were closed by stop codons) (43). The homologies of the human *erbA* domain 2 with human carbonic anhydrase remained but were not augmented. Thus the *erbA* gene is related to but not identical to known carbonic anhydrases.

In conclusion, AEV is an unusual retrovirus in that it can specifically block the maturation of erythroid cells at an immature stage of differentiation. This is achieved by the synergistic action of two distinct oncogenes, *erbA* and *erbB*. The *erbB* product is a membrane glycoprotein (20, 21) that is homologous to a portion of the epidermal growth factor receptor (35) and shares extensive homology with the *src* oncogene family [(34) and our results]. This latter homology is also functional, since both *erbB* in the absence of *erbA* and other *src* family viral genes such as *src* or *fps* (44) can transform erythroblasts (28). Such cells require complex growth conditions and partially differentiate into mature erythrocytes in vitro (28). In contrast, *erbA* in combination with *erbB* is capable of ar-

resting erythroid leukemic cells at an early stage of differentiation where they are able to grow in simple tissue culture media unsuitable for normal erythroid precursors (45, 46).

The finding that the deduced *erbA* protein bears statistically significant relatedness, but not identity, with carbonic anhydrases is interesting because such enzymes play a fundamental role in the CO<sub>2</sub> transport by erythrocytes, the precursors of which are precisely the main targets of AEV.

Our finding of this new type of oncogene leads to the idea that some *src*-related oncogenes found—with the exception of AEV—to be single transforming genes in retroviruses and to transform mainly fibroblasts can affect early hematopoietic differentiation when acting in synergy with a specific activated cell-derived gene such as *erbA*.

B. DEBUIRE\*  
C. HENRY  
M. BENAÏSSA  
G. BISERTE

U 124 INSERM, Institut de  
Recherches sur le Cancer de Lille,  
Faculté de Médecine, Place de Verdun,  
59045 Lille Cedex France

J. M. CLAVERIE  
Unité d'Informatique, Scientifique  
Institut Pasteur, 25 Rue du Docteur  
Roux, 75724 Paris Cedex 15 France

S. SAULE  
P. MARTIN  
D. STEHELIN

U 186 INSERM, Institut Pasteur,  
15 Rue C. Guérin, 59019  
Lille Cedex France

#### References and Notes

- H. Land, L. F. Parada, R. A. Weinberg. *Nature (London)* 304, 596 (1983).
- H. E. Ruley. *ibid.*, p. 602.
- R. F. Newbold and R. W. Overell. *ibid.*, p. 648.
- B. Vennstrom and J. M. Bishop. *Cell* 28, 135 (1982).
- D. LePrince, A. Gegonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagron, D. Stehelin. *Nature (London)* 306, 395 (1983).
- J. Coll *et al.*, *EMBO J.* 2, 2189 (1983).
- T. Graf, B. Royer-Pokora, G. E. Schubert, H. Beug. *Virology* 71, 423 (1976).
- T. Graf and H. Beug. *Biochim. Biophys. Acta* 516, 269 (1978).
- K. Bister and P. H. Duesberg. *Proc. Natl. Acad. Sci. U.S.A.* 76, 5023 (1979).
- M. M. C. Lai, S. S. F. Hu, P. K. Vogt. *Virology* 97, 366 (1979).
- M. Roussel *et al.*, *Nature (London)* 281, 452 (1979).
- B. Vennstrom, L. Fanshier, C. Moscovici, J. M. Bishop. *J. Virol.* 36, 575 (1980).
- S. Saule, M. Roussel, C. Lagrou, D. Stehelin. *ibid.* 38, 409 (1981).
- J. M. Coffin *et al.*, *ibid.* 40, 953 (1981).
- D. Sheinss, B. Vennstrom, J. M. Bishop. *Cell* 23, 291 (1981).
- S. M. Anderson, W. S. Hayward, B. G. Neel, H. Hanafusa. *J. Virol.* 36, 676 (1980).
- M. M. C. Lai, J. C. Neil, P. K. Vogt. *Virology* 100, 475 (1980).
- T. Pawson and G. S. Martin. *J. Virol.* 34, 280 (1980).
- M. L. Privalsky and J. M. Bishop. *Proc. Natl. Acad. Sci. U.S.A.* 79, 3958 (1982).
- M. L. Privalsky, L. Sealy, J. M. Bishop, J. P. McGrath, A. D. Levinson. *Cell* 32, 1257 (1983).
- M. J. Hayman *et al.*, *ibid.*, p. 579.
- A. Sergeant *et al.*, *EMBO J.* 1, 237 (1982).
- J. Coll *et al.*, *Exp. Cell. Res.* 149, 151 (1983).
- Y. Samarut and L. Gazzolo. *Cell* 28, 921 (1982).
- H. Beug, A. Von Kirchbach, G. Doderlein, J. F. Conscience, T. Graf. *ibid.* 18, 375 (1979).
- L. Gazzolo, J. Samarut, M. Bouadbelli, J. P. Blanchet. *ibid.* 22, 683 (1980).
- J. K. T. Fung, W. G. Lewis, L. B. Crittenden, H. J. Kung. *ibid.* 33, 357 (1983).
- L. Frykberg *et al.*, *ibid.* 32, 227 (1983).
- T. Graf and H. Beug. *ibid.* 34, 7 (1983).
- B. Vennstrom, L. Fanshier, C. Moscovici, J. M. Bishop. *J. Virol.* 36, 575 (1980).
- A. M. Maxam and W. Gilbert. *Methods Enzymol.* 65, 499 (1980).
- D. Schwartz, R. Tizard, W. Gilbert. *Cell* 32, 853 (1983).
- S. M. Mount. *Nucleic Acids Res.* 10, 459 (1982).
- T. Yamamoto *et al.*, *Cell* 35, 71 (1983).
- J. Downward *et al.*, *Nature (London)* 307, 521 (1984).
- M. J. Hayman, B. Royer-Pokora, T. Graf. *Virology* 92, 31 (1979).
- E. G. Krebs and J. A. Beavo. *Annu. Rev. Biochem.* 48, 923 (1979).
- R. D. Marshall. *Biochem. Soc. Symp.* 40, 17 (1974).
- T. P. Hopp and K. R. Woods. *Proc. Natl. Acad. Sci. U.S.A.* 78, 3824 (1981).
- R. F. Doolittle. *Science* 214, 149 (1981).
- M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt. *Protein Sequence Database Release 83* (National Biomedical Research Foundation, Washington, D.C., 1983), document PRSD-0183C.
- A. Liljas *et al.*, *Nature (London) New Biol.* 235, 131 (1972).
- S. Saule, C. Henry, P. Martin, in preparation.
- P. Kahn and T. Graf, personal communication.
- H. Beug *et al.*, *Cell* 28, 907 (1982).
- H. Beug and M. J. Hayman. *ibid.*, in press.
- J. M. Claverie. *Nucleic Acids Res.* 12, 397 (1984).
- T. F. Smith, M. S. Waterman, W. M. Fitch. *J. Mol. Evol.* 18, 385 (1981).
- R. J. Tanis, R. E. Ferrell, R. E. Tashian. *Biochim. Biophys. Acta* 371, 534 (1974).
- R. E. Ferrell, S. K. Stroup, R. J. Tanis, R. E. Tashian. *ibid.* 533, 1 (1978).
- M. Sciaky, N. Limozin, D. Filippi-Foveau, J. M. Gulian, G. Laurent-Tabusse. *Biochimie* 58, 1071 (1976).
- S. B. Needleman and C. D. Wunsch. *J. Mol. Biol.* 48, 443 (1970).
- W. C. Barker and M. O. Dayhoff, in *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, pp. 101-110.
- We thank B. Vennstrom and J. M. Bishop for providing the clone p-AEV 11, A. Begue and A. Schneeberger for helpful advice and discussion, and N. Devassine for manuscript preparation. Supported by the Institut National de la Santé et de la Recherche Médicale (U 124 CRL 822020 and U 186 CRL 822033); Institut de Recherches sur le Cancer de Lille, Institut Pasteur de Lille, and private foundations (Association pour la Recherche sur le Cancer and Fondation pour la Recherche médicale).

\* To whom requests for reprints should be addressed.

13 February 1984; accepted 8 May 1984

The Four C-Terminal Amino Acids of the *v-erbA* Polypeptide Are Encoded  
by an Intronic Sequence of the *v-erbB* Oncogene

C. HENRY,\* M. COQUILLAUD,\* S. SAULE,† D. STEHELIN,† AND B. DEBUIRE\*<sup>1</sup>

\*U 124 INSERM, Institut de Recherches sur le Cancer de Lille, Faculté de Médecine, Place de Verdun,  
59045 Lille Cedex, France and †U 186 INSERM, Institut Pasteur, 15 Rue Camille Guérin,  
59019 Lille Cedex, France

Received June 19, 1984; accepted September 10, 1984

The genome of avian erythroblastosis virus (AEV), a defective acute leukemia retrovirus, carries two distinct cell-derived oncogenes in the structure 5'Agag-*erbA-erbB*-Δenv3'. The nucleotide sequence of the *v-erbA* gene was recently reported. In order to determine the boundary between the two adjacent oncogenes, the sequence of the *v-erbA/v-erbB* junction of AEV was compared to that of a recombinant lambda phage containing a chicken cellular sequence representing the 5' part of *c-erbB*. The four C-terminal amino acids of *v-erbA* are in fact encoded by a *c-erbB* intron-derived sequence thus demonstrating that the virus acquired a truncated *c-erbA* gene. Furthermore the 7 to 10 amino acid residues upstream from the 4 C-terminal amino acids mentioned above appeared to be derived from *env*-related sequences. The splice acceptor site at the beginning of the only open reading frame for *v-erbB* is also present and functional in *c-erbB* when expressed to generate a truncated EGF (epidermal growth factor) receptor. Thus AEV joins a truncated *erbA* gene to a truncated *erbB* gene through *env*-derived sequences and intronic sequences from *c-erbB*. © 1985 Academic Press, Inc.

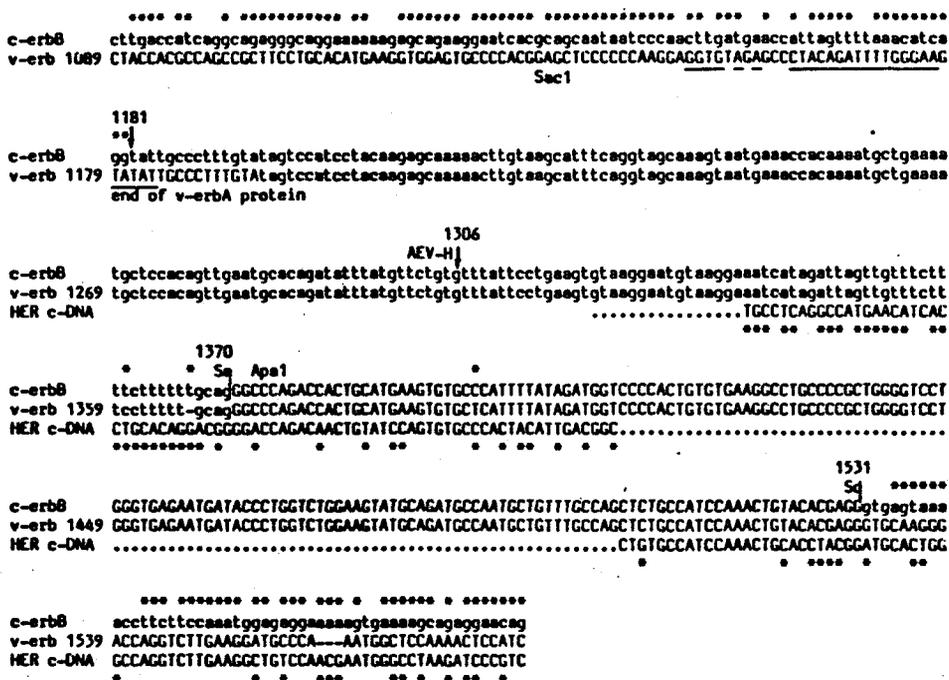
Avian erythroblastosis virus is a defective leukemia virus (DLV) which induces both erythroblastosis and fibrosarcomas in chickens (1). The genome of this virus contains two independent cell-derived oncogenes *v-erbA* and *v-erbB* (2, 3) expressed in transformed cells as two distinct mRNAs (4-6). A genomic 5.4-kilobase (kb) mRNA codes for a cytoplasmic 75,000-molecular weight (MW) fusion protein (P75<sup>gag-erbA</sup>) (7) while a subgenomic spliced 3.5-kb mRNA produces a *v-erbB* membrane glycoprotein of 74,000 MW (8, 9). Recently, the *v-erbB* sequence (10) was identified as a portion of the EGF receptor (11). We have sequenced the *v-erbA* oncogene (12) and shown that the carboxy-terminal half of its product exhibited a significant homology with the carbonic anhydrase family. This homology begins with *v-erbA* amino acid residue 219 (amino acid residue 10 for the carbonic anhydrases) and spans ca. 180 residues toward

the carboxyterminus of the molecule (residue 398 of *v-erbA* corresponding to residue 195 of the carbonic anhydrases, which continue for another 64 residues (12)). We also compared the nucleotide sequence of *v-erbA* to the sequences of avian viral structural genes. We found a significant homology (82%) between the end of *v-erbA* (nucleotides 1151 to 1183) and a portion of the *env* gene of Rous sarcoma virus (nucleotides 5804 to 5836) (Ref. (13) and Figs. 1A and B). However, the stop codon for the P75<sup>gag-erbA</sup> fusion protein (TAG at position 1195) did not belong to this stretch of homology.

In order to determine the exact boundary between the two oncogenes, we sequenced part of the 2.5-kbp *EcoRI* fragment of a previously described recombinant lambda phage (NO4 *c-erbB* in Ref. (2)) containing the cellular chicken sequences corresponding to the upstream region of *v-erbB* and we compared this sequence to the homologous region of *v-erbB*. The relevant nucleotide sequences are shown in Fig. 1A and allow the fol-

<sup>1</sup> Author to whom requests for reprints should be addressed.

## A



## B

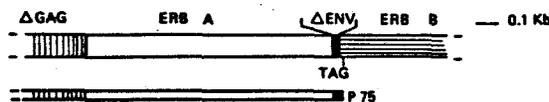


FIG. 1. (A) Nucleotide sequence of the 5' part of the chicken *c-erbB* locus: comparison to the *v-erb* (12) and human EGF receptor (HER) cDNA sequences (11). The *c-erbB* sequence was determined by the Maxam and Gilbert procedure (17). Coding sequences are written in capital letters and noncoding sequences in small letters. Numbers refer to *v-erb* nucleotides as in Ref. (12). Sa and Sd demarcate the splice acceptor and splice donor sites at the extremities of the *c-erbB* exon. Asterisks mark the differences between the cellular and the viral *erb* sequences (upper line) and between the *v-erb* and human EGF receptor cDNA nucleotide sequences (11) (lower line). AEV-H indicates the recombination site between *c-erbB* and an erythroblastosis virus described in Ref (10). The arrow shows the recombination point between *erbA* and *erbB* in AEV. Endonuclease restriction sites used to produce the probes described in Fig. 2 are indicated. Underlined nucleotides are homologous to *env* nucleotides of RSV (13). (B) Schematic representation showing the presence of *env*-related sequences between *v-erbA* and *v-erbB*.

lowing observations to be made. First, the homology between *v-erbB* and *c-erbB* begins at nucleotide 1181 in the viral sequence (Ref. (12) and Fig. 1A); this defines the upstream boundary of the *c-erbB*-related sequences in AEV. Sur-

prisingly the four C-terminal amino acids of P75<sup>gag-erbA</sup> are encoded by *c-erbB*-derived nucleotides which also provide the termination codon of this protein (nucleotides 1195-1197) indicating that AEV acquired a truncated *c-erbA* locus. Second, the viral

sequences between nucleotides 1181 and 1370 correspond to part of an intron of *c-erbB*, closed in all three reading frames by numerous stop codons and ending with the conserved splice acceptor site TTTTCTTTTTGCAG/G (Sa in Fig. 1A). This splice acceptor site is used to generate the subgenomic *v-erbB* mRNA since the latter did not hybridize on Northern blots (Fig. 2, lane 1) with a *SacI*-*Apal* probe (restriction sites indicated in Fig. 1A) derived from AEV DNA (nucleotides 1139 to 1374 of *v-erbB*) (12). As shown as a control in Fig. 2 lane 2, this mRNA hybridized with a *SacI*-*Bam*HI probe (nucleotides 1139 to 1841 of *v-erbB*) (12). The splice acceptor site described above is also likely to be used during the maturation of the *c-erbB* transcript. Nucleotide 1371 of *v-erbB* corresponds to the beginning of a strong homology with the human EGF receptor cDNA (nucleotide 1918 in Ref. (11)). The *c-erbB* exon beginning with the splice acceptor site at position 1370 ends with the splice donor site G/GTGAG (14) at position 1528, since downstream from this site the *c-erbB* sequence diverges

from that of both *v-erbB* (downstream from position 1531 in Fig. 1A) and the EGF receptor cDNA (downstream from nucleotide 2067 in Ref. (11)). We concluded that the *v-erbB* gene is a truncated version of *c-erbB*, starting within an intron at the 5' end. These observations favor the idea that the recombination events involved to generate *v-erbB* occurred at the DNA level. How and when AEV acquired the *erbA* gene remains to be determined: *erbB* before *erbA*, *erbA* before *erbB*, or both together from a cell having undergone a translocation joining the two genes prior to their transduction by the virus? The homology detected between the end of *v-erbA* and the *env* gene of RSV suggests that two successive recombination events led to the capture of the *erb* genes by AEV. In fact, a possible mechanism leading finally to the *c-erbB* capture by an ALV (avian leukosis virus) could result from the upstream insertion of a LTR (long terminal repeat) as reported by Fung *et al.* (15). Such an activation would have generated an *erbB*-containing retrovirus as recently described by Yamamoto *et al.* (10) for AEV-H. A second recombination event at the DNA or RNA level would occur leading to the capture of *erbA*.

Finally a salient feature of *c-erbB* activation linked to transformation appears to involve the *c-erbB* exon described here, that encodes the carboxyterminal half of the EGF receptor (16). Indeed the three independent *erbB*-activated oncogenes in AEV (this paper), in AEV H (10), and in an LTR activated *c-erbB* chicken leukemia (15) all concern insertions (although at different sites, see Fig. 1) clustered in the intron preceding the *c-erbB* exon sequences presented in this paper. Ullrich *et al.* (11) proposed that the truncated EGF receptor may escape cell regulation by losing its EGF binding site and thus lead to cell transformation. Whether such recombinations are facilitated by a privileged domain contained within this intron remains to be documented. In any case, our observations together with the ones just quoted here, allow to predict that most *c-erbB* activations (by viral genomes, translocations, etc.) might involve recom-

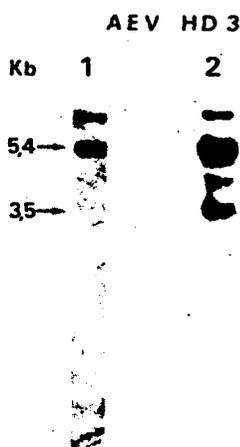


FIG. 2. Sizes of AEV-coded RNAs in AEV-transformed chicken erythroblasts HD3 (described in Ref. (18)). Poly(A)-containing RNA was denatured (19), separated on agarose gels, and transferred to nitrocellulose. Blots were hybridized with a  $^{32}$ P-nick-translated fragment isolated from AEV proviral DNA (20). Lane 1, *SacI*-*Apal* fragment; lane 2, *SacI*-*Bam*HI fragment, extending largely into *v-erbB*.

ination events in the intron mentioned above.

#### ACKNOWLEDGMENTS

We express gratitude to Professor G. Biserte for constant encouragement and support during the course of this work. We thank M. Benaissa and A. Begue for excellent technical assistance, P. Martin and N. Devassine for manuscript preparation. This work was supported by the Institut National de la Santé et de la Recherche Médicale (U 124 CRL 822020 and U 186 CRL 822033); Institut de Recherches sur le Cancer de Lille, Institut Pasteur de Lille; and private foundations, Association pour la Recherche sur le Cancer and Fondation pour la Recherche Médicale.

#### REFERENCES

1. GRAF, T., and BEUG, H., *Cell* **34**, 7-9 (1983).
2. SERGEANT, A., SAULE, S., LEPRINCE, D., BEGUE, A., ROMMENS, C., and STEHELIN, D., *EMBO J.* **1**, 237-242 (1982).
3. SPURR, N. K., SOLOMON, E., JANSSON, M., SHEER, D., GOODFELLOW, P. N., BODMER, W. F., and VENNSTROM, B., *EMBO J.* **3**, 159-163 (1983).
4. ANDERSON, R. W., HAYWARD, W. S., NEEL, B. G., and HANAFUSA, H., *J. Virol.* **36**, 676-683 (1980).
5. SAULE, S., ROUSSEL, M., LAGROU, C., and STEHELIN, D., *J. Virol.* **38**, 409-419 (1981).
6. SHEINESS, D., VENNSTROM, B., and BISHOP, J. M., *Cell* **23**, 291-300 (1981).
7. HAYMAN, M. J., ROYER-POKORA, B., and GRAF, T., *Virology* **92**, 31-45 (1979).
8. PRIVALSKY, M. L., SEALY, L., BISHOP, J. M., McGRATH, J. P., and LEVINSON, A. D., *Cell* **32**, 1257-1267 (1983).
9. HAYMAN, M. J., and BEUG, H., *Nature (London)* **309**, 460-462 (1984).
10. YAMAMOTO, T., NISHIDA, T., MIYAJIMA, N., KAWAI, S., OOI, T., and TOYOSHIMA, K., *Cell* **35**, 71-78 (1983).
11. ULLRICH, A., COURSENS, L., HAYFLICK, J. S., DULL, J. T., GRAY, A., TAM, A. W., LEE, J., YORDEN, Y., LIBERMANN, T. A., SCHLESSINGER, J., DOWNWARD, J., MAYES, E. L. V., WHITTLE, N., WATERFIELD, M. D., and SEEBURG, P. H., *Nature (London)* **309**, 418-425 (1984).
12. DEBUIRE, B., HENRY, C., BENAISSA, M., BISERTE, G., CLAVERIE, J. M., SAULE, S., MARTIN, P., and STEHELIN, D., *Science (Washington, D. C.)* **224**, 1456-1459 (1984).
13. SCHWARTZ, D. E., TIZARD, R., and GILBERT, W., *Cell* **32**, 853-869 (1983).
14. MOUNT, S. M., *Nucl. Acids Res.* **10**, 459-472 (1982).
15. FUNG, J. K. T., LEWIS, W. G., CRITTENDEN, L. B., and KUNG, H. J., *Cell* **33**, 357-368 (1983).
16. DOWNWARD, J., YORDEN, Y., MAYES, E., SCRACE, G., TOTTY, N., STOCKWELL, P., ULLRICH, A., SCHLESSINGER, J., and WATERFIELD, M. D., *Nature (London)* **307**, 521-527 (1984).
17. MAXAM, A. M., and GILBERT, W., In "Methods in Enzymology" (L. Grossman and K. Moldave, eds.), Vol. 65, pp. 499-560. Academic Press, New York.
18. COLL, J., SAULE, S., MARTIN, M., RAES, M. B., LAGROU, C., GRAF, T., BEUG, H., SIMON, I. E., and STEHELIN, D., *Exp. Cell Res.* **149**, 151-162 (1983).
19. McMASTER, G. K., and CARMICHAEL, G. M., *Proc. Natl. Acad. Sci. USA* **74**, 4835-4838 (1977).
20. VENNSTROM, B., MOSCOVICI, C., GOODMAN, H. M., and BISHOP, J. M., *J. Virol.* **39**, 625-631 (1981).

## RESUME

Le virus de l'érythroblastose aviaire (AEV) est un virus défectif des leucémies qui transforme principalement les érythroblastes et contient deux oncogènes indépendants v-erbA et v-erbB. Ces oncogènes ont des équivalents cellulaires c-erbA et c-erbB qui ont été conservés au cours de l'évolution.

Nous avons déterminé la séquence nucléotidique du gène v-erbA et montré que la protéine codée par ce gène présente une homologie avec la famille des anhydrases carboniques. Cette homologie est particulièrement intéressante puisque l'anhydrase carbonique joue un rôle essentiel dans le métabolisme des cellules rouges, les érythroblastes étant la principale cible de AEV.

En séquençant un fragment du locus c-erbB de l'ADN cellulaire de Poulet homologue à la partie 5' de v-erbB, nous avons mis en évidence la présence d'un intron de c-erbB dans le virus. Il apparaît également que les quatre amino-acides C-terminaux du polypeptide erbA sont codés par des nucléotides erbB.

Enfin, nous avons déterminé la structure primaire de la partie du locus c-erbA homologue à l'anhydrase carbonique dans l'ADN cellulaire du Poulet et de l'Homme.

**MOTS-CLES :** CANCEROGENESE VIRALE - ONCOGENE - SEQUENCE NUCLEOTIDIQUE - ANHYDRASE CARBONIQUE.

