UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE



N° d'ordre 283

# 50376 1988 261

# THESE

Présentée pour l'obtention du grade de DOCTEUR DE L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE option : Biochimie

par

Martine DUTERQUE-COQUILLAUD

# STRUCTURE ET EXPRESSION DU PROTO-ONCOGENE c-ets-1 (POULET)

soutenue le 9 décembre 1988 devant la commission d'examen

Président : Rapporteurs :

50376

1988 261

Examinateurs :

**Professeur J. MONTREUIL** 

Professeur J. KREMBEL Docteur P. TAMBOURIN

Professeur G. BISERTE Docteur B. DEBUIRE Docteur D. STEHELIN

#### DOYENS HONORAIRES DE L'ANCIENNE FACULTE DES SCIENCES

M. H. LEFEBVRE, M. PARREAU.

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

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

### PROFESSEUR EMERITE

M. A. LEBRUN

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

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

### PRESIDENT DE L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

M. A. DUBRULLE.

### **PROFESSEURS - CLASSE EXCEPTIONNELLE**

M. CONSTANT Eugène
M. FOURET René
M. GABILLARD Robert
M. MONTREUIL Jean
M. PARREAU Michel
M. TRIDOT Gabriel

Electronique Physique du solide Electronique Biochimie Analyse Chimie appliquée

### PROFESSEURS - lère CLASSE

M. BACCHUS Pierre M. BIAYS Pierre M. BILLARD Jean M. BOILLY Bénoni M. BONNELLE Jean Pierre M. BOSCQ Denis M. BOUGHON Pierre M. BOURIQUET Robert M. BREZINSKI Claude M. BRIDOUX Michel M. CARREZ Christian M. CELET Paul M. CHAMLEY Hervé M. COEURE Gérard M. CORDONNIER Vincent M. DEBOURSE Jean Pierre M. DHAINAUT André M. DOUKHAN Jean Claude M. DYMENT Arthur M. ESCAIG Bertrand M. FAURE Robert M. FOCT Jacques M. FRONTIER Serge M. GRANELLE Jean jacques M. GRUSON Laurent N. GUILLAUME Jean M. HECTOR Joseph M. LABLACHE-COMBIER Alain M. LACOSTE Louis M. LAVEINE Jean Pierre M. LEHMANN Daniel Mme LENOBLE Jacqueline M. LEROY Jean Marie M. LHOMME Jean M. LOMBARD Jacques M. LOUCHEUX Claude M. LUCQUIN Michel M. MACKE Bruno M. MIGEON Michel M. PAQUET Jacques M. PETIT Francis M. POUZET Pierre M. PROUVOST Jean M. RACZY Ladislas M. SALMER Georges M. SCHAMPS Joel M. SEGUIER Guy M. SIMON Michel Mle SPIK Geneviève M. STANKIEWICZ François M. TILLIEU Jacques M. TOULOTTE Jean Marc M. VIDAL Pierre M. ZEYTOUNIAN Radyadour

Astronomie Géographie Physique du solide Biologie Chimie-Physique Probabilités Algèbre Biologie végétale Analyse numérique Chimie-Physique Informatique Géologie générale Géotechnique Analyse Informatique Gestion des entreprises Biologie animale Physique du solide Mécanique Physique du solide Mécanique Métallurgie Ecologie numérique Sciences Economiques Algèbre Microbiologie Géométrie Chimie organique Biologie végétale Paléontologie Géométrie Physique atomique et moléculaire Spectrochimie Chimie organique biologique Sociologie Chimie physique Chimie physique Physique moléculaire et rayonnements atmosphériques E.U.D.I.L. Géologie générale Chimie organique Modélisation - Calcul scientifique Minéralogie Electronique Electronique Spectroscopie moléculaire Electrotechnique Sociologie Biochimie Sciences Economiques Physique théorique Automatique Automatique Mécanique

#### PROFESSEURS - 2ème CLASSE

M. ALLAMANDO Etienne M. ANDRIES Jean Claude M. ANTOINE Philippe M. BART André M. BASSERY Louis Mme BATTIAU Yvonne M. BEGUIN Paul N. BELLET Jean M. BERTRAND Hugues M. BERZIN Robert M. BKOUCHE Rudolphe N. BODARD Marcel M. BOIS Pierre M. BOISSIER Daniel M. BOIVIN Jean Claude M. BOUQUELET Stéphane M. BOUQUIN Henri M. BRASSELET Jean Paul M. BRUYELLE Pierre M. CAPURON Alfred M. CATTEAU Jean pierre M. CAYATTE Jean Louis M. CHAPOTON Alain M. CHARET Pierre M. CHIVE Maurice M. COMYN Gerard M. COQUERY Jean Marie M. CORIAT Benjamin Mme CORSIN Paule M. CORTOIS Jean M. COUTURIER Daniel M. CRAMPON Norbert M. CROSNIER Yves M. CURGY Jean jacques Mle DACHARRY Monique M. DAUCHET Max M. DEBRABANT Pierre M. DEGAUQUE Pierre M. DEJAEGER Roger M. DELORME Pierre M. DELORME Robert M. DEMUNTER Paul M. DENEL Jacques M. DE PARIS Jean Claude M. DEPREZ Gilbert M. DERIEUX Jean Claude Mle DESSAUX Odile M. DEVRAINNE Pierre Mme DHAINAUT Nicole M. DHAMELINCOURT Paul M. DORMARD Serge M. DUBOIS Henri M. DUBRULLE Alain M. DUBUS Jean Paul

Composants électroniques Biologie des organismes Analyse Biologie animale Génie des procédés et réactions chimiques Géographie Mécanique Physique atomique et moléculaire Sciences Economiques et Sociales Analyse Algèbre Biologie végétale Mécanique Génie civil Spectrochimie Biologie appliquée aux enzymes Gestion Géométrie et topologie Géographie Biologie animale Chimie organique Sciences Economiques Electronique Biochimie structurale Composants électroniques optiques Informatique théorique Psychophysiologie Sciences Economiques et Sociales Paléontologie Physique nucléaire et corpusculaire Chimie organique Tectolique Géodynamique Electronique Biologie Géographie Informatique Géologie appliquée Electronique Electrochimie et Cinétique Physiologie animale Sciences Economiques Sociologie Informatique Analyse Physique du solide - Cristallographie Microbiologie Spectroscopie de la réactivité chimique Chimie minérale Biologie animale Chimie physique Sciences Economiques Spectroscopie hertzienne Spectroscopie hertzienne Spectrométrie des solides

M. DUPONT Christophe Mme EVRARD Micheline M. FAKIR Sabah M. FAUQUEMBERGUE Renaud M. FONTAINE Hubert M. FOUOUART Yves M. FOURNET Bernard M. GAMBLIN André M. GLORIEUX Pierre M. GOBLOT Rémi M. GOSSELIN Gabriel M. GOUDMAND Pierre M. GOURIEROUX Christian M. CREGORY Pierre N. GRENY Jean Paul M. GREVET Patrice M. GRIMBLOT Jean M. GUILBAULT Pierre M. HENRY Jean Pierre M. HERMAN Maurice M. HOUDART René M. JACOB Gérard M. JACOB Pierre M. JEAN Raymond M. JOFFRE Patrick M. JOURNEL Gérard M. KREMBEL Jean M. LANGRAND Claude M. LATTEUX Michel Mme LECLERCQ Ginette M. LEFEBVRE Jacques M. LEFEVRE Christian Mle LEGRAND Denise Mle LEGRAND Solange M. LEGRAND Pierre Mme LEHMANN Josiane M. LEMAIRE Jean M. LE MAROIS Henri M. LEROY Yves M. LESENNE Jacques M. LHENAFF René M. LOCQUENEUX Robert M. LOSFELD Joseph M. LOUAGE Francis M. MAHIEU Jean Marie M. MAIZIERES Christian M. MAURISSON Patrick M. MESMACQUE Cérard M. MESSELYN Jean M. MONTEL Marc M. MORCELLET Michel M. MORTREUX André Mme MOUNIER Yvonne M. NICOLE Jacques M. NOTELET Francis M. PARSY Fernand M. PECQUE Marcel M. PERROT Pierre

Vie de la firme (I.A.E.) Génie des procédés et réactions chimiques Algèbre Composants électroniques Dynamique des cristaux Optique atmosphérique Biochimie structurale Géographie urbaine, industrielle et démographie Physique moléculaire et rayonnements atmosphériques Algèbre Sociologie Chimie physique Probabilités et statistiques I.A.E. Sociologie Sciences Economiques Chimie organique Physiologie animale Génie mécanique Physique spatiale Physique atomique Informatique Probabilités et statistiques Biologie des populations végétales Vie de la firme (I.A.E.) Spectroscopie hertzienne Biochimie Probabilités et statistiques Informatique Catalyse Physique Pétrologie Algèbre Algèbre Chimie Analyse Spectroscopie hertzienne Vie de la firme (I.A.E.) Composants électroniques Systèmes electroniques Géographie Physique théorique Informatique Electronique Optique - Physique atomique Automatique Sciences Economiques et Sociales Cénie Mécanique Physique atomique et moléculaire Physique du solide Chimie organique Chimie organique Physiologie des structures contractiles Spectrochimie Systèmes électroniques Mécanique Chimie organique Chimie appliquée

M. PERTUZON Emile M. PONSOLLE Louis M. PORCHET Maurice M. POSTAIRE Jack M. POVY Lucien M. RICHARD Alain M. RIETSCH François M. ROBINET Jean CLaude M. ROGALSKI Marc M. ROY Jean Claude Mme SCHWARZBACH Yvette M. SLIWA Henri M. SOMME Jean M. STAROSWIECKI Marcel M. STERBOUL François M. TAILLIEZ Roger M. THERY Pierre M. THIEBAULT François M. THUMERELLE Pierre Mme TJOTTA Jacqueline M. TOURSEL Bernard M. TREANTON Jean rené M. TURREL Georges M. VANDORPE Bernard M. VASSEUR Christian M. VAST Pierre M. VERBERT André M. VERNET Philippe M. WACRENIER Jean Marie M. WALLART Francis M. WARTEL Michel M. WATERLOT Michel M. WEINSTEIN Olivier M. WERNER Georges M. WOZNIAK Michel Mme ZINN JUSTIN Nicole

Physiologie animale Chimie physique Biologie animale Informatique industrielle Automatique Biologie animale Physique des polymères EUDIL Analyse Psychophysiologie Géométrie Chimie organique Géographie Informatique Informatique Génie alimentaire Systèmes électroniques Sciences de la terre Démographie - Géographie Humaine Mathématiques Informatique Sociologie du Travail Spectrochimie infrarouge et Raman Chimie minérale Automatique Chimie inorganique Biochimie Génétique Electronique Spectrochimie infrarouge et Raman Chimie inorganique Géologie générale Analyse économique de la recherche et développement Informatique théorique Spectrochimie Algèbre

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

Je remercie, en particulier le Docteur B. DEBUIRE, qui m'a accueillie dans son groupe. Ses conseils, son dynamisme et la confiance qu'elle m'a accordée m'ont été d'une aide constante. Je tiens à lui exprimer ma très vive reconnaissance.

Ce travail doit également beaucoup à Monsieur le Professeur G. BISERTE. Je tiens à le remercier pour la bienveillance qu'il m'a toujours témoignée.

J'exprime ici toute ma reconnaissance à Monsieur le Docteur D. STEHELIN, pour les conseils stimulants qu'il m'a prodigués au cours de ce travail et pour l'accueil qu'il m'a toujours réservé dans son laboratoire. Je remercie particulièrement D. LEPRINCE dont l'enthousiasme et la rigueur scientifique m'ont beaucoup aidée.

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

Madame le Docteur M.H. LOUCHEUX m'a intégrée au sein de l'Unité INSER M (U. 124) qu'elle dirige, qu'elle en soit remerciée.

Ce mémoire est le fruit d'un travail d'équipe avec laquelle fai eu plaisir à travailler. Je remercie tous les membres du laboratoire pour leur aide et leur sympathie.

Mademoiselle M. LESTIENNES a dactylographié ce mémoire, sa compétence et sa disponibilité m'ont été d'une aide permanente.

Je tiens à exprimer toute ma reconnaissance à mes parents qui m'ont permis de poursuivre mes études aussi longtemps que je le désirais.

Merci enfin, à Hervé pour sa compréhension et ses encouragements continus.

L'accomplissement dectravail n'aurait pas été possible sans l'aide financière de l'Institut de Recherches sur le Cancer de LILLE et l'Association pour la Recherche sur le Cancer.

# TABLE DES MATIERES

INTRODUCTION	
CHAPITRE I : RETROVIRUS ET ONCOGENES	
I. <u>LES RETROVIRUS</u> 4	
<ol> <li>Morphologie</li> <li>Cycle réplicatif et structure du génome viral</li> </ol>	
II. <u>LES ONCOGENES</u>	
III. MODES DE TRANSFORMATION PAR LES v-onc	0
<ul> <li>a) Les oncogènes viraux codant pour des facteurs de croissance</li> <li>b) Les oncogènes viraux codant pour des tyrosine-kinases</li> <li>c) Les oncogènes viraux codant pour les protéines ras</li> <li>d) Les oncogènes codant pour des sérine/thrénonine kinases</li> <li>e) Les oncogènes viraux dont les produits sont nucléaires</li> </ul>	
IV. LES MECANISMES D'ACTIVATION DES ONCOGENES CELLULAIRES AU COURS DE LA CARCINOGENESE	7
<ol> <li>La transfection d'ADN de cellules tumorales dans les cellules NIH 3T3</li> <li>Différents types de remaniement des oncogènes cellulaires conduisant à leur activation         <ul> <li>Amplification des c-onc ou ONC</li> <li>Les transfections characterisment</li> </ul> </li> </ol>	
<ul> <li>c) Les translocations chromosomiques</li> <li>c) Activation transcriptionnelle des oncogènes cellulaires</li> <li>d) Activation par modification de la séquence nucléotidique des oncogènes cellulaires</li> <li>e) Coopération d'oncogènes</li> </ul>	
<u>CHAPITRE II</u> : LA SYNTHESE ALTERNATIVE D'ARN MESSAGERS A L'ORIGINE DE L'EXPRESSION MULTIPLE D'UN MEME GENE	3
I. EXPRESSION D'UNE UNITE TRANSCRIPTIONNELLE SIMPLE 2	4
<ol> <li>Initiation de la transcription et signaux promoteurs</li> <li>Terminaison de la transcription et formation de l'extrémité 3'</li> <li>Epissage des ARN messagers</li> </ol>	

II.	LES UNITES TRANSCRIPTIONNELLES COMPLEXES
	<ol> <li>1) Gènes utilisant des sites de polyadénylation alternatifs et l'épissage alternatif des exons</li> <li>2) Gènes utilisant de multiples sites d'initiation et l'épissage alternatif des exons</li> <li>3) Gènes dont les ARNm présentent un seul site d'initiation, une extrémité 3' commune mais qui diffèrent par l'organisation de leurs exons</li> <li>4) Gènes produisant des ARNm hétérogènes quant à leurs extrémités 5' et/ou 3' mais dont l'épissage des exons est constant</li> </ol>
СНАР	ITRE III : LES RETROVIRUS AMV et E 26
Ι.	ACTIVITE BIOLOGIQUE
II.	STRUCTURE DU GENOME VIRAL
	a. L'oncogène <u>v-myb</u> b. L'oncogène <u>v-ets</u>
III.	LES PROTO-ONCOGENES c-myb ET c-ets
СНАР	ITRE IV : RESULTATS 41
Ι.	ORGANISATION DES SEQUENCES HOMOLOGUES À v-ets DANS LE GENOME DE POULET (ARTICLE 1 EN ANNEXE)
II.	ETUDE DE L'ADN COMPLEMENTAIRE DE L'ARNm c-ets-1 CODANT POUR LA PROTEINE P54 C-ets-1 (ADNC TYPE I) (ARTICLE 2 EN ANNEXE)
	1) Expression dans un système eucaryote 2) Analyse de la séquence nucléotidique
III.	ETUDE D'UN ADNC D'UN ARNM c-ets-1 CODANT POUR UNE PROTEINE P68 C-ets-1 (ADNC TYPE II) (ARTICLE 3 EN ANNEXE) 46
	1) Expression dans un système eucaryote 2) Analyse de la séquence nucléotidique
IV.	TRANSDUCTION DE c-ets-1 49
<u>C O N C</u>	<u>51 USION</u>
REFE	RENCES BIBLIOGRAPHIQUES 57
ANNE	X E S 88

# INTRODUCTION

--

La croissance cellulaire est un processus minutieusement contrôlé qui répond aux besoins spécifiques de l'organisme, mais parfois les mécanismes subtils de contrôle des multiplications cellulaires se dérèglent, une cellule commence à croître et à se diviser de façon anarchique et ce clone de cellules indésirables aboutit à la formation d'une tumeur. 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 de la structure et de l'expression du proto-oncogène **c-ets-1** (poulet), présent dans le rétrovirus E26. Celui-ci induit chez le poulet des leucémies de type mixte érythroīde/myéloīde à prédominance érythroīde. Outre l'oncogène **v-ets**, ce rétrovirus contient un second oncogène **v-myb**, il offre donc un modèle d'appréhension des mécanismes de la double oncogénèse. Notre travail a permis de montrer qu'un mécanisme d'épissage alternatif à l'intérieur du locus **c-ets-1** donnait naissance à deux protéines distinctes mais apparentées exprimées à des taux variables dans les organes lymphoīdes de poulet.

Dans ce texte, après un rappel sur les rétrovirus et les oncogènes, nous décrirons les mécanismes d'épissage alternatif déjà observés chez les eucaryotes puis nous présenterons notre travail en essayant d'en dégager les points essentiels et les perspectives qui s'y rattachent.

2

## CHAPITRE I

### RETROVIRUS ET ONCOGENES



Figure 1 : morphologie d'une particule virale (extrait de Martin P , 1986)

### I. LES RETROVIRUS :

Les travaux sur les rétrovirus, virus à ARN ayant comme hôtes des cellules eucaryotes, ont permis de mettre en évidence l'existence de gènes oncogènes ou **v-onc** volés ou "transduits" par ces rétrovirus à la cellule lors du cycle infectieux. Ces **v-onc** ont donc des équivalents cellulaires normaux, les proto-oncogènes ou **c-onc**. Ils sont phylogénétiquement conservés, ce qui prouve le rôle important qu'ils jouent probablement dans la physiologie cellulaire normale.

Ces rétrovirus oncogènes ont un patrimoine génétique relativement simple et ils constituent un modèle pour l'étude des mécanismes de cancérisation de la cellule. Pour infecter la cellule hôte, les rétrovirus ont la particularité de dépendre, d'un intermédiaire ADN synthétisé par un enzyme viral : l'ADN polymérase ARN dépendante ou rétropolymérase (Temin et Baltimore, 1972).

Les rétrovirus sont classés en trois sous-familles :

- <u>les spumavirus</u> induisent des infections chroniques sans pouvoir pathogène apparent (Ruckle, 1958; Jonhston, 1971; Stiles **et al.**, 1964).

- <u>les lentivirus</u> induisent des affections pulmonaires chroniques et une dégénérescence nerveuse après une longue période d'incubation. Le virus humain HIV responsable du syndrome d'immunodéficience acquise (SIDA) appartient vraisemblablement à cette catégorie (Rabson et Martin, 1985).

- <u>les oncovirus</u> possèdent un pouvoir transformant rapide. Ils induisent une grande variété de cancers chez l'animal. Ils seront les seuls décrits dans ce mémoire.

C'est en 1911 que le premier rétrovirus sarcomatogène a été isolé chez le poulet par Peyton Rous. Actuellement, de nombreux oncovirus ont été identifiés chez un grand nombre d'espèces animales. Les plus étudiés ont comme hôtes le poulet, la souris, le chat, le singe et l'homme.

Nous décrirons plus particulièrement dans ce texte les oncovirus aviaires.

### 1) Morphologie : (Figure 1)

Les virions, observés au microscope électronique, se présentent sous forme de particules sphériques de 80 à 110 nm de diamètre (Sharp et Bernhard, 1952 ; Bernhard, 1958). Ils sont entourés d'une enveloppe constituée d'un fragment de membrane cytoplasmique emprunté à la cellule hôte lors de leur extrusion par bourgeonnement. Cette enveloppe est hérissée de spicules de



Figure 2 : cycle réplicatif des rétrovirus

nature glycoprotéique dont la partie protéique est codée par le gène **env** du virus (Rifkin et Compans, 1971). Ces glycoprotéines interagissent avec des récepteurs cellulaires non encore identifiés permettant la reconnaissance virus-cellule (Vogt, 1977). La nucléocapside ou "core", icosaèdre constitué de protéines **gag** codées par le virus, renferme un dimère d'ARN monocaténaire (Mangel **et al.**, 1974 ; Weissman **et al.**, 1975) et quelques molécules de rétropolymérase (Baltimore, 1970; Temin et Mizutami, 1970).

### 2) Cycle réplicatif et structure du génome viral.

### a - Cycle réplicatif :

La propagation des rétrovirus se décompose en 3 étapes (Figure 2) :

- <u>l'infection</u>, qui débute par l'adsorption du virus par ses spicules aux récepteurs membranaires de la cellule et se termine par la pénétration du "core" viral dans le cytoplasme cellulaire.

- <u>la réplication</u> par la rétropolymérase virale, de l'ARN viral en ADN bicaténaire, qui est alors intégré dans le génome de la cellule hôte sous forme de provirus. Il est alors pris en charge au même titre qu'un gène cellulaire normal par la cellule qui assure l'expression des ARN viraux et la maturation des protéines virales puis

- <u>l'encapsidation des ARN viraux</u>, <u>l'acquisition de l'enveloppe</u> par bourgeonnement et l'extrusion des particules virales matures.

### b - Structure du génome viral :

Il existe deux types d'oncovirus selon leur aptitude à se répliquer :

- les ALV ("Avian Leukosis Viruses") peu ou non oncogènes in vivo et non transformants in vitro dont le génome ne contient que les séquences nécessaires au cycle infectieux.

- les DLV ("Defective Leukemia Viruses") induisant rapidement des cancers in vivo et capables de transformer des cellules en culture. Le génome des DLV est un ARN défectif pour la réplication en raison de la perte de séquences virales remplacées par des séquences spécifiques (v-onc) d'origine cellulaire (c-onc). Les DLV ont besoin pour se propager d'un virus auxiliaire ALV (virus "helper"). Toutefois, une exception existe à cette règle, il s'agit du virus du sarcome de Rous (RSV) qui, à la suite de l'acquisition d'un v-onc n'a pas perdu les séquences virales nécessaires au cycle infectieux. (Figure 3).

Le génome d'un ALV possède toutes les séquences codantes ou non,

compétence pour la réplication



Figure 3 : génome des rétrovirus

requises pour le bon déroulement du cycle infectieux. Ce génome est formé d'un ARN diploïde dont chaque sous-unité, est capuchonnée en 5' (cap = 7 mGppp) (Furuichi et al., 1975) et polyadénylée en 3' (Bender et Davidson, 1976). La taille de cet ARN est d'environ 8000 nucléotides divisés en différentes régions fonctionnelles réparties comme suit :

Les tailles respectives de chacune de ces régions varient en fonction des rétrovirus considérés.

### - LES SEQUENCES NON CODANTES :

- la séquence R: (16 à 80 nucléotides) est répétée aux deux extrémités et joue un rôle dans la rétrotranscription (Haseltine et al., 1977; Coffin et al., 1978) de l'ARN en ADN.

- la séquence U5 : (76 à 120 nucléotides) n'a pas de fonction connue.

- <u>la séquence AT</u> : (16 à 20 nucléotides) est complémentaire à la région 3' de l'ARN de transfert qui servira d'initiateur pour la rétrotranscription (Dahlberg **et al.**, 1974).

- <u>la séquence L</u> : (250 à 300 nucléotides) contient des séquences impliquées dans l'encapsidation des ARN viraux (Shank et Linial, 1980 ; Nishizawa **et al.**, 1985 ; Darlix, 1986). Elle participe à la formation de dimères d'ARN par l'intermédiaire de ponts hydrogène. (Haseltine **et al.**, 1977). Enfin, elle contient une séquence qui serait le site de fixation des ribosomes (Petersen et Hackett, 1985).

- <u>la séquence C</u> : (250 à 1200 nucléotides) contient probablement des séquences d'encapsidation ainsi qu'une séquence de 11 nucléotides où s'initierait le brin d'ADN (+) lors de la rétrotranscription. (Sorge **et al.**, 1983).

- <u>la séquence U3</u> : (250 à 1200 nucléotides) contient le signal de polyadénylation de l'ARN viral, AAUAAA ainsi que les séquences qui contrôlent la promotion de la transcription du provirus. Les régions U3 et U5 sont dupliquées et se trouvent répétées aux deux extrémités de l'ADN sous la forme U3-R-U5 alors appelée LTR ("Long Terminal Repeat"). L'ADN viral est ensuite circularisé, il migre dans le noyau de la cellule hôte pour s'intégrer dans le

# ARN génomique



ARN sous-génomique



SD : site donneur d'épissure

SA : site accepteur d'épissure

Figure 4 : structure et expression des ARN génomique et sous-génomique d'un ALV

gènome de celle-ci. La séquence des éléments du provirus intégré est alors la suivante :

La région U3 se trouve alors en 5' du provirus et en dirige la transcription effectuée par l'ARN polymérase II de la cellule hôte (Dinowitz, 1975; Shank et Varmus, 1978). De plus, la transcription est activée par la présence d'une séquence activatrice, ou "enhancer", située en 5' de U3 (Cullen et al., 1985).

### - LES GENES VIRAUX : (Figure 4).

1) <u>Le gène gag</u> (2100 nucléotides) code pour un polypeptide de 76 kd (Pr76 gag) précurseur de toutes les protéines de la capside interne. Pr76 gag est phosphorylé, non glycosylé, et associé à la membrane plasmique (Eisenman et Vogt, 1978). Ce précurseur est ensuite scindé en 5 protéines p19, p10, p27, p12, p15 (de 19, 10, 27, 12, 15 kd respectivement). La p15, préalablement libérée par une protéase cellulaire, semble responsable de la protéolyse du précurseur Pr76 gag (Von Der Helm, 1977) synthétisé à partir de l'ARN génomique. Le codon d'initiation utilisé pour la synthèse de celui-ci est situé en 3' de la séquence L, il est suivi (à 18 nucléotides) d'un site donneur d'épissure qui permettra la formation d'un ARN sous-génomique codant pour les protéines env (RNA Tumors Viruses, 1982).

2) <u>Le gène pol</u> (2600 nucléotides) code pour la rétropolymérase qui possède plusieurs activités spécifiques : elle synthétise l'ADN (-) complémentaire de l'ARN viral utilisé comme matrice (Temin et Baltimore, 1972), elle possède une activité RNAse-H qui lui permet de dégrader spécifiquement l'ARN de l'hybride ADN-ARN (Verma, 1977), elle synthétise le brin d'ADN (+) (Varmus **et al.**, 1979), elle semble participer, grâce à une activité endonucléasique, à l'intégration de l'ADN proviral dans le génome de la cellule infectée (Grandgenett **et al.**, 1978; Grandgenett et Ajaykumar, 1985).

Cette protéine, formée de deux sous-unités, A (58 kd) et B (92 kd), est obtenue par coupure d'un précurseur de 180 kd, Pr 180 **gag-pol** (Operman et al., 1977 ; Witte et Baltimore, 1978). Les cadres de lecture des gènes gag et pol sont ouverts dans des phases différentes ; un changement de phase de lecture est donc nécessaire au cours de la traduction et la synthèse du précurseur est rendue possible par un comportement inhabituel des ribosomes : le changement de phase de lecture en cours de traduction (Jacks et Varmus, 1985).

3) Le gène env (1800 nucléotides).

Les protéines d'enveloppe sont synthétisées à partir d'un ARN sousgénomique provenant d'un mécanisme d'épissage entre le site donneur d'épissure (SD) situé au début du gène **gag** et le site accepteur d'épissure (SA) du gène **env** (Lee **et al.**, 1979 ; Hayward, 1977 ; Weiss **et al.**, 1977). Cet ARN sous-génomique est à l'origine d'un précurseur de 57 kd (Pr57<sup>env</sup>) qui est ensuite glycosylé (gPr92 **env**) (Hayman, 1978) et coupé en deux protéines glycosylées gp85 et gp37. Ces protéines unies par des ponts disulfures constitueront les spicules virales situées à la surface externe de la membrane plasmique entrainée par le "core" viral lors de l'extrusion par bourgeonnement (Leamnson et Halpern, 1976).

Les rétrovirus défectifs dont le génome contient un (ou deux) oncogène(s) (**v-onc**) qui se sont substitués à une partie des gènes de structure se répliquent avec l'aide d'un ALV comme il a été décrit précédemment.

- EXPRESSION DES v-onc :

Le point d'insertion de l'oncogène cellulaire dans le génome viral a des répercussions sur la transcription et sur la structure du produit formé :

- si l'oncogène viral est exprimé à partir de l'ARN génomique, deux possibilités sont à considérer :

 synthèse d'une protéine de fusion Pgag-v-onc lorsque gag et le v-onc sont en phase, ou

• synthèse d'une protéine p<sup>v-onc</sup>.

- si l'oncogène viral est exprimé à partir d'un ARN sous-génomique provenant de l'épissage entre le SD du gène **gag** et le SA du **v-onc**:

• une protéine comportant les 6 amino-acides N-terminaux de **gag** puis les séquences **v-onc** est formée, ou

. le **v-onc** possède ses propres signaux d'initiation et de terminaison et il en dérive essentiellement une protéine  $p^{v-onc}$ .

### II. LES ONCOGENES :

### - Origine des v-onc : (Tableau 1)

Depuis le travail initial de Stéhelin **et al.** en 1976, concernant le gène **src** de RSV (Rous Sarcoma Virus), nous savons que les **v-onc** dérivent de gènes cellulaires normaux appelés **c-onc** ou proto-oncogènes, accidentellement transduits par les rétrovirus au cours de leur réplication. Ces proto-oncogènes sont phylogénétiquement conservés, ce qui prouve le rôle important qu'ils jouent probablement dans la physiologie cellulaire normale.

- Les mécanismes de la transduction sont peu connus et plusieurs modèles ont été proposés (Graf et Stéhelin, 1982 ; Bishop, 1983). Le premier fait état de recombinaisons au niveau de l'ADN entre le provirus intégré et un gène cellulaire, le second fait intervenir des "sauts" de la rétropolymérase entre l'ARN du virus auxiliaire ("helper") et l'ARN cellulaire encapsidés dans le même virion. Plus récemment, l'étude des mécanismes impliqués dans la maturation des ARN pré-méssagers a permis d'envisager la formation d'ARN messagers hybrides. Ces ARN résulteraient d'un mécanisme d'épissage intermoléculaire illégitime (en trans) entre deux molécules différentes (Konarska et al., 1985 ; Solnick, 1985).

Certains rétrovirus possèdent la particularité d'avoir transduit dans leur génome deux oncogènes cellulaires, ces virus sont peu nombreux et jusqu'à présent, trois seulement ont été identifiés : le virus de l'érythroblastose aviaire, AEV (v-erbA et v-erbB), l'un des virus de la myélocytomatose aviaire MH2 (v-myc et v-mil) et enfin le virus de l'érythroblastose aviaire E26 (v-myb et v-ets), objet de ce mémoire. Ces rétrovirus illustrent l'intérêt et l'importance de la coopération de certains oncogènes dans la transformation et la différenciation cellulaires. Au sein du virus AEV, v-erbA et v-erbB coopèrent pour transformer les cellules érythroīdes aviaires immatures (Gazzolo et al, 1980) et bloquer de façon très stricte leur différenciation. Les virus ne contenant que l'oncogène v-myc n'ont pas d'effet décelable sur les cellules de neurorétines alors que le virus MH2, dans lequel les oncogènes v-mil et v-myc sont présents est capable d'induire la prolifération des cellules de neurorétines aviaires. Enfin, les deux oncogènes du virus E26 (v-myb et v-ets) lui confèrent une action à la fois myéloblastosante et érythroblastosante.

ONCOGENE	VIRUS	MALADIE	HOTE	PROT.V-ONC	ACTIVITE	LOCALISATION	PROT.C-ONC	ACTIVITE	LOCALISATION	REFERENCES
sis	SSV	Sarcome	Singe	p28 <sup>env-sis</sup>	PDGF	Frarátáa (1)	-20	PDCF	Excrélée	(1) Huang et al., 1984
515	Pi-FeSV	Sarcome	Chat	p76 <sup>gag-sis</sup>	(Chaine <sup>B</sup> )	Excretee	pou	(Chaine <sup>\$</sup> )		
erb B	AEV-ES4	Sarcome Erythro.	Poulet	gp74 <sup>v-erbB</sup>	Tyr. Kin.	Membrane plasmique	gp170 (Homme)	Tyr. Kìn. Réc.EGF	Trans. Mbr.	
fms	SM-FeSV	Sarcome	Chat	gp180 <sup>gag-fms</sup>	Tyr. Kin.	Membrane plasmique	gp165 <sup>(2)</sup>	Tyr. Kin. Réc.CSF–1	Trans. Mbr.	(2) Sherr et al., 1985
ros	UR2	Sarcome	Poulet	p68 <sup>gag-ros</sup>	Tyr. Kin.	Membranes		Tyr. Kin. Réc. ?	Trans. Mbr.	
src	RSV S1 S2	Sarcome	Poulet	pp60 <sup>v-src</sup> (3), pp62 <sup>s-src</sup>	Tyr. Kin.	Face int. du Plasmaleme	рр60	Tyr. Kin.	Mbr. et Cytosquel.	(3) Ikawa et al., 1986
fps/src	Fu-ASV	Sarcome	Poulet	p130 <sup>gag-fps</sup>	Tyr. Kin.	Membrane plasmique	p98	Tyr. Kin.		
	ST-FeSV	Sarcome	Chat	p85 <sup>gag-fes</sup>	Tyr. Kin.	Membrane plasmique	p92			
yes	Y73-ASV	Sarcome	Poulet	p90 <sup>gag-yes</sup>	Tyr. Kin.	Membrane plasmique	р59-62 <sup>(4)</sup>	Tyr. Kin.		(4) Sudol et Hanafusa, 1986
fgr	GR-FeSV	Sarcome	Chat	p70 <sup>gag-actin-fgr</sup>	Tyr. Kin.					
abl	Ab-MLV	Leuc.préB	Souris	p160 <sup>gag-ab1</sup>	Tyr. Kin.	Membrane plasmique	p150		Cytoplasme <sup>(5)</sup>	(5) Hunter et Cooper, 1985
~~ .	HZ2-FeSV	Sarcome	Chat	p98 <sup>gag-kit</sup>	Tyr. Kin.	Membrane plasmique				
kit	HZ4-F'eSV	Sarcome	Chat	p80 <sup>gag-kit</sup>	Tyr. Kin. <sup>(6)</sup>		p145 (homme)	Tyr. Kin. ? Réc. ?	Trans. Mbr.	(6) Besmer et al., 1986
sea	S13	Sarcome Erythro.	Poulet	gp155 <sup>env-sea</sup> → gp85+gp70	Tyr. Kin. (gp70) <sup>(7)</sup>	Cytoplasme <sup>(7)</sup>				(7) Hayman et al., 1985
	TP1-FeSV	Sarcome	, Chat	gp83 <sup>gag-onc</sup>	Tyr. Kin.					

TABLEAU 1 : Oncogènes viraux et leurs homologues cellulaires. (RNA Tumor Viruses, 1985).

.

Ha-ras	Ha-MVS	Sarcome Erythro.	Rat	n21v-Ha-ras	Thr. Kin.	Mbr.	<sub>n21</sub> Ha-ras	Thr. Kin.	Mbr.	
	Ba1B-MSV	Hémangio– Sarcome	Souris	P	fixe GTP	face int.	per	fixe GTP	face int.	
Ki-ras	Ki-MSV	Sarcome Erythro.	Rat	p21 <sup>v-Ki-ras</sup>	GTPase		p21 <sup>Ki-ras</sup>	CTPase	Tucco mu.	
mos	Mo-MSV	Sarcome	Souris	p37 <sup>env-mos</sup>	Thr/Ser. Kin.	Cytoplasme				
mil-raf	MH2	Carcinome	Poulet	p100 <sup>gag-mil</sup>	Thr/Ser. Kin.	<b></b>	(2)			
	MSV-3611	Sarcome	Souris	gp90 <sup>gag-</sup> raf p75 <sup>gag-raf</sup>	Thr/Ser. Kin.	Cytoplasme	pplasme p71 <sup>(8)</sup>		(8)Patshinsky et al, 1986	
fos	FBJ-MSV	Ostéo- Sarcome	Souris	p55 <sup>v-fos</sup>		Noyau	p55 <sup>(9)</sup>	· · · · · · · · · · · · · · ·	Noyau	(9)Curran et al, 1984
myc	MC29	Nyélocyto. Carcinome	Poulet	p110 <sup>gag-myc</sup>	Affinité	Novau	n58	Affinité	Novau	
	MH2	Carcinome	Poulet	p61/63 <sup>v-myc</sup>	l'ADN		poo	I'ADN	noyau	
	AMV	Myélobl.	Poulet	p45 <sup>v-myb</sup>	Affinité			Affinité		/
тур	E26	Myélobl. Erythro.	Poulet	p135 <sup>gag-myb-ets</sup>	pour l'ADN	Noyau	p75 (10)	pour l'ADN	Noyau	(10)Klempnauer et al, 1984
ski	SKV-ASV	Carcinome	Poulet	p110 <sup>gag-ski-pol</sup>		Noyau				
ets	E26	Erythro.	poulet	p135 <sup>gag-myb-ets</sup>	Aff. ADN	Noyau	p54 (11)		Noyau Cytoplasme	(11)Ghysdael et al,1986
erb A	AEV-ES4	Sarcome Erythro.	Poulet	p75 <sup>gag-</sup> erbA	Aff. ADN	Noyau <sup>(12)</sup>	p40/46	Réc. Hormones Thyr.		(12)Sap et al,1986
rel	REV-T	Reticulo- endothél.	Dinde	p59 <sup>env-rel</sup>		Noyau <sup>(13)</sup> Cytoplasme	p68 (14)		Cytoplasme	(13)Gilmore et Temin,1986 (14)Simek et al,1988
jun <sup>(15)</sup>	ASV17	Sarcome	Poulet	p55/65 <sup>gag-jun</sup>	Aff. ADN	Noyau	p47	AP1 <sup>(16)</sup>	Noyau	(15)Maki et al., 1987 (16)Bohmann et al., 1987
crk	CT10	Sarcome	Poulet	p47gag-crk (17)	Phospholipase Cll <sup>(18)</sup>	)				(17)Mayer et al., 1988 (18)Stahl et al., 1988

TABLEAU1:(suite)

### - Les proto-oncogènes ou c-onc (Tableau 1) :

La conservation phylogénétique des c-onc, ONC pour l'espèce humaine ainsi que leur transduction dans des rétrovirus devenus ainsi oncogènes, soulignent leur importance dans la vie cellulaire normale. Ceci est largement vérifié, quelques exemples suffisent à le montrer : SIS est homologue au précurseur de la chaine du facteur de croissance des plaquettes (PDGF) (Waterfield et al., 1983), ERB A est un récepteur d'hormones thyroïdiennes (Weinberger et al, 1986 ; Sap et al., 1986), FMS est une version tronquée du récepteur d'un facteur de croissance spécifique des macrophages (M-CSF ou CSF-1) (Sherr et al., 1985).

### - Le cas particulier des rétrovirus humains :

Chez l'homme, des rétrovirus ont été associés à des leucémies de type T ("Human T-Lymphocyte Viruses") : HTLV I (Poiesz et al., 1980b ; Miyoshi et al., 1981) et HTLV II (Kalyanaraman et al., 1982b). Les virus à l'origine du syndrome d'immunodéficience acquise (SIDA), d'abord appelé LAV ("Lymphotropic Associated Viruses") (Barré-Sinoussi et al., 1983) ou HTLV III (Gallo et al., 1984) puis HIV 1 et 2 ("Human Immunodeficiency Viruses") (Barré-Sinoussi et al., 1983 ; Gallo et al., 1984 ; Guyader et al., 1987) sont apparentés aux lentivirus. Ces rétrovirus contiennent dans leur génome outre les séquences gag, pol, env, des séquences additionnelles qui ne sont pas d'origine cellulaire appelées chez HIV tat, (transactivator), rev (regulator of expression of virian proteins), vif (virion infectivity factor), vpr (R), nef (negative factor), vpx (X uniquement chez HIV2 et SIV) (Gallo et al., 1988). Les protéines qui en dérivent ont un effet activateur sur la transcription virale et pourraient être responsables de l'activation de la transcription de certains gènes des cellules infectées (Sodroski et al., 1984b, Chen et al., 1985).

### III. MODES DE TRANSFORMATION PAR LES v-onc :

L'acquisition du pouvoir oncogène des rétrovirus est liée à la présence dans leur génome d'un oncogène viral ou **v-onc.** Pour rendre compte du pouvoir transformant, de ces derniers, il faut considérer les différences majeures qui existent entre **v-onc** et **c-onc**. Ces différences ont permis d'émettre deux hypothèses qui ne sont pas mutuellement exclusives :

- la première est dite hypothèse quantitative, elle est liée au mode

d'expression différent des v-onc et des c-onc. Chez un animal infecté, les rétrovirus sont capables d'exprimer à un taux élevé et dans de nombreux tissus les oncogènes qu'ils contiennent. Au contraire, les c-onc sont généralement exprimés à un taux relativement faible dans un petit nombre de cellules et de tissus. Cet effet de dose responsable du pouvoir transformant semble être vérifié pour quelques c-onc ou ONC qui, par exemple, transforment les cellules lorsque leur expression est placée sous le contrôle d'un promoteur viral : MYC (Lee et al., 1985), FOS, (Miller et al., 1984), SIS (Clarke et al., 1984; Gazit et al., 1984).

- la seconde hypothèse dite qualitative suppose que les différences de structure primaire acquises par le v-onc (mutations, délétions ou insertions) lui confèrent un pouvoir transformant. Le cas le plus représentatif est celui de l'oncogène ras : l'allèle T24/EJ du proto-oncogène Ha-ras doit ses propriétés transformantes non seulement à la mutation bien connue responsable du changement de l'acide aminé en position 12, mais aussi à une seconde altération ponctuelle dans un intron qui entraîne une augmentation sévère (dix fois) de l'expression de ce gène (Cohen J.B. et al., 1988).

Par ailleurs, de nombreux **v-onc** sont exprimés sous forme de protéine de fusion avec des déterminants de gènes de structure viraux, le plus souvent des déterminants **gag** qui pourraient contribuer à l'oncogénicité de la protéine synthétisée.

La caractéristique essentielle de la cellule tumorale est sa capacité à proliférer. Les produits des **v-onc** semblent, pour transformer la cellule, agir de façon analogue aux facteurs impliqués dans la mitose et la différenciation cellulaire (Heldin **et al.**, 1984). De plus, les cellules transformées, in vitro, ont une dépendance réduite vis à vis des facteurs de croissance. Les produits des **v-onc** pourraient déréguler la chaîne d'événements qui va du signal mitotique à la synthèse de l'ADN ou induire la production de facteurs de croissance auxquels le cellule infectée serait sensible.

Il a été démontré que les produits des oncogènes viraux peuvent être des facteurs de croissance (v-sis) ou des récepteurs membranaires (v-erbB, v-fms, v-kit) ou des transducteurs membranaires (v-src, v-ras) ou des effecteurs cytosoliques (v-mil, v-mos) ou encore des protéines nucléaires (v-myb, v-myc, v-fos, v-jun).

# a - Les oncogènes viraux codant pour des facteurs de croissance :

L'oncogène **v-sis** qui a été isolé à partir du virus du sarcome de singe (SSV) est exprimé sous forme d'une protéine de 28 kd, p28 **v-sis**. Cette protéine présente une très forte homologie avec la chaine  $\beta$  du PDGF (Platelet Derived Growth Factor) (Doolittle **et al.**, 1983 ; Waterfield **Et al.**, 1983). Le PDGF est responsable de la croissance et de la réparation de divers types tissulaires et il est composé de deux sous-unités  $\alpha$  et  $\beta$  (Heldin **et al.**, 1979 ; Antoniades, 1981 ; Deuel **et al.**, 1981 ; Raines et Ross, 1982), présentant entre elles 40 % d'homologie.

Le produit de v-sis est formé de deux sous-unités identiques reliées par des ponts disulfures (Robbins et al., 1983, 1985) et son activité est comparable à celle du PDGF (Huang et al., 1984; Owen et al., 1984). La stimulation autocrine du récepteur du PDGF a été observée dans des cellules normales de rein de rat (NRK) transformées par v-sis. Mais à la différence de ce qui se déroule dans les cellules normales, le PDGF se fixe sur son récepteur dans le compartiment intracellulaire alors que celui-ci se trouve encore dans l'appareil de Golgi et n'a pas atteint la surface cellulaire, (Betsholtz et al., 1984 ; Keating et al., 1988). Ceci démontre que l'activation intracellulaire de récepteurs par un mécanisme autocrine peut jouer un rôle dans la transformation cellulaire.

### b - Les oncogènes viraux codant pour des tyrosine-kinases :

Le produit de l'oncogène **v-src**, présent dans le virus du Sarcome de Rous (RSV), est une phosphoprotéine de 60 kd, pp 60<sup>**v-src**</sup>, qui a été la première protéine transformante pour laquelle une activité enzymatique ait été mise en évidence (Brugge et Erikson, 1977). Elle est capable de phosphoryler certaines protéines cibles sur les tyrosines (Colett **et al.**, 1978).

Les produits de plusieurs oncogènes (erbB, fms, fps/fes, yes) partagent cette activité pourtant relativement rare dans la cellule (1 à 1,5 % des phosphorylations dans la cellule normale). Ces protéines se répartissent en deux classes :

- des récepteurs de facteur de croissance essentiellement membranaires.

- les protéines tyrosine kinases associées à la face interne de la membrane plasmique.

 Les oncogènes viraux codant pour des récepteurs de facteur de croissance membranaire :

Les récepteurs de facteur de croissance ont en général une structure transmembranaire, la partie externe fixant le facteur de croissance et la partie interne étant susceptible de transmettre le signal mitotique. Leur activité tyrosine-kinase pourrait jouer un rôle essentiel dans la transmission de ce signal.

Le produit de **v-erbB** est une version tronquée du récepteur de l'EGF (Epidermal Growth Factor) (Downward et al., 1984 ; Ullrich et al., 1984), seules subsistent dans la protéine gp76 **v-erbB**, les régions transmembranaires et cytoplasmiques du récepteur normal. De plus, l'extrémité C-terminale de la protéine **v-erbB** diffère de celle du récepteur normal. Ces deux différences, absence de région extracellulaire et extrêmité C-terminale différente, pourraient contribuer à l'activité néoplasique de la protéine gp76 **v-erbB**. De plus, alors que l'activité tyrosine kinase est déclenchée par la fixation de l'EGF sur le récepteur normal, l'activité tyrosine kinase est constitutive dans le cas du produit de **v-erbB** qui pourrait fonctionner comme relai mitotique même en l'absence de ligand.

Les produits des oncogènes v-fms, c-fms et v-kit présentent de fortes homologies de séquence avec le récepteur du PDGF en particulier au niveau de leur domaine tyrosine kinase (Sherr et al., 1985). Le mode d'action des oncogènes v-fms et v-kit est encore mal défini cependant le pouvoir transformant de v-fms pourrait être dû aux différences v-fms/c-fms siégant dans la région 3' : les 40 derniers acides aminés de la protéine c-fms sont remplacés par 11 amino acides d'origine rétrovirale dans la protéine v-fms.

La séquence nucléotidique de **v-ros** suggère qu'il pourrait être homologue à un récepteur de facteur de croissance proche de l'insuline (Hunter, 1985; Hunter et Cooper, 1985).

### Les oncogènes viraux codant pour les protéines tyrosine kinases

associées à la surface interne de la membrane plasmique :

Certains produits d'oncogènes viraux (v-abl, v-fes/fps, v-src) présentent une activité tyrosine kinase sans être toutefois apparentés à des récepteurs de facteurs de croissance connus. Ces protéines sont cytoplasmiques et souvent associées à la membrane plasmique (Bishop and Varmus, 1982). La plus étudiée est la protéine pp60 v-src. Elle est myristylée à son extrémité amino terminale (Sefton et al., 1982) et associée à la membrane plasmique par son groupement myristyl ; cette localisation membranaire est requise pour la transformation (Buss et al., 1984 ; Kamps et al., 1986). La pp60 v-src est également phosphorylée sur deux sites majeurs, un résidu tyrosine en position 416 (Tyr 416) et un résidu sérine en position 17 (Ser 17) (Collett et al., 1979). Elle est capable de s'autophosphoryler, régulant probablement ainsi son activité tyrosine kinase.

Le pouvoir transformant de pp60 **v-src** pourrait être dû à la phosphorylation d'autres constituants cellulaires que des protéines, les phosphoinositols par exemple qui ont des propriétés de second messager, relais entre les signaux extracellulaires et l'intérieur de la cellule. En effet, les cellules transformées par **v-src** présentent un métabolisme accru des phosphoinositols et la protéine pp60 **v-src** a, in vitro, une activité kinase s'exerçant sur les phosphoinositols (Sugimoto et al., 1984 ; Maccara et al., 1984). Les produits du métabolisme du phosphoinositol, diacylglycérol et inositoltriphosphate ont la capacité d'une part de mobiliser le calcium intracellulaire et d'autre part, d'activer la protéine kinase C (Berridge et Irvine 1984). Cette double action pourrait conduire à plusieurs réponses cellulaires dont la prolifération.

### c - Les oncogènes viraux codant pour les protéines ras :

Chez les mammifères, cette famille comprend au moins trois membres dont deux c-Ha-ras et c-Ki-ras ont été transduits dans des rétrovirus de sarcome murin (Harvey, 1964 ; Kirsten et al., 1967). Le troisième N-ras, isolé à partir d'une tumeur humaine (un neuroblastone d'où le terme N-ras), n'a pas d'équivalent viral connu. Existent également des gènes apparentés aux gènes ras, appelés ral, rap, rho et rab (Chardin P., 1988).

Les protéines codées par les gènes ras sont homologues et ont un poidds moléculaire de 21 kd, p21 ras. Elles fixent et hydrolysent le GTP et sont localisées à la face interne de la membrane plasmique. Les protéines v-ras ou c-ras mutées en position 12 hydrolysent beaucoup moins efficacement le GTP (Gibbs et al., 1984; Mc Grath et al., 1984, Sweet et al., 1984). Il semble que la superfamille de protéines ras joue un rôle critique à l'interface membranecytosquelette et régule quelques uns des échanges entre la cellule et son environnement. Cependant une implication directe de ces protéines dans ces évènements reste à démontrer. Les protéines ras possèdent des analogies structurales et biochimiques avec les sous-unités  $\alpha$  des protéines G présentes dans beaucoup de systèmes "récepteur/effecteur" ; il a été suggéré qu'elles pourraient être des médiateurs de signaux transmis d'un récepteur de facteur de croissance à une effecteur tel que la phospholipase C ou A2. Mais, des données récentes montrent que l'action de ras sur ces phospholipases est un effet indirect (Seuwen et al., 1987, Yu et al., 1988). Il est probable que les protéines ras soient activées de façon éphémère par des récepteurs de facteurs de croissance, mais les bases biochimiques de cette activation ne sont pas encore comprises (Chardin P., 1988).

### d - Les oncogènes codant pour des sérine/thréonine kinases.

Les produits des oncogènes v-mil, v-raf et v-mos ont une activité sérine/thréonine kinase in vitro (Moelling et al., 1984 ; Kloetzer et al., 1984), activité également rencontrée chez la protéine kinase C. Cela suggère que ces protéines oncogènes comme la protéine kinase C pourraient intervenir dans la transmission du signal mitotique. Cependant, les connaissances relatives aux propriétés des oncogènes cités ci-dessus sont encore trop limitées pour leur attribuer un rôle comparable à celui de la protéine kinase C. Il faut, en effet, souligner que les produits de v-mos, v-mil et v-raf sont localisés dans le cytosol alors que la protéine kinase C est associée à la membrane plasmique (Bunte et al., 1983 ; Papkoff et al., 1983).

### e - Les oncogènes viraux dont les produits sont nucléaires:

Les produits de v-myc (Donner et al., 1982 ; Alitalo et al., 1983), v-myb (Klempnauer et al., 1984), v-fos (Curran et al., 1984), v-ski (Stavnezer et al., 1981), v-erbA (Sap et al., 1986 ; Weinberger et al., 1986), v-jun (Bohmann et al., 1987) ont une localisation nucléaire. De par cette localisation et leur capacité à se lier à l'ADN, ces protéines oncogènes pourraient déréguler la transcription de certains gènes. Des expériences d'expression in vitro ont montré que la protéine myc pouvait réguler le promoteur de transcription d'un gène codant pour une protéine de choc thermique (Kingston et al., 1984). Il pourrait donc y avoir interaction de la protéine myc avec des séquences nucléotidiques du promoteur du gène de choc thermique.

Les produits de v-myc, v-myb et E1A (gène précoce de l'adénovirus) présentent des homologies de séquence (Ralston et Bishop, 1983), et des durées de vie courtes compatibles avec un rôle de régulation (Hann et Eisenman, 1984; Ramsay et al., 1984; Klempnauer et al., 1984; Spindler et Berk, 1984).

Il a été démontré en 1986 que la partie v-erbA de la protéine P75 gag-erbA synthétisée dans les cellules transformées par AEV présentait une homologie avec les récepteurs d'hormones stéroïdes, (Weinberger et al., 1986 ; Green et al., 1986 ; Loosfelt, 1986) et plus récemment que la protéine c-erbA était un récepteur d'hormones thyroïdiennes (Weinberger et al., 1986 ; Sap et al., 1986). La localisation de ces protéines est nucléaire, suggérant que leurs effets pourraient être associées à la régulation de la transcription de certains gènes.

La protéine v-jun présente, dans sa partie C-terminale, une homologie de 44 % avec le domaine de fixation à l'ADN d'un facteur de transcription de la levure, GCN4 (Vogt et al., 1987). D'autre part, le protooncogène humain c-jun code pour une protéine dont les propriétés structurales et fonctionnelles sont comparables à celles du facteur de transcription humain AP-1 (Bohmann et al., 1987; Bos et al., 1988). Celui-ci se lie spécifiquement aux séquences stimulatrices des régions de contrôle en cis de SV40 et du gène humain de la métallothionéine IIA (Angel et al., 1987; Lee et al., 1987ab). Récemment, il était démontré qu'une fraction significative du facteur de transcription c-jun/AP-1 cellulaire interagissait avec la protéine FOS (Chiu et al., 1988) et représenterait probablement la protéine de 39 kd qui lui est associée (Curran et al., 1985). Ceci fut confirmé par l'utilisation d'un antisérum (anti c-jun) (Angel et al., 1988 ; Sassone-Corsi et al., 1988). La protéine FOS serait donc un activateur de transcription qui n'agirait pas en se fixant directement sur l'ADN mais par l'intermédiaire de c-jun/AP-1.

# IV. LES MECANISMES D'ACTIVATION DES ONCOGENES CELLULAIRES AU COURS DE LA CARCINOGENESE :

Les mécanismes de transformation des cellules normales en cellules cancéreuses impliquent une cascade d'évènements dont la nature exacte n'est pas encore élucidée. L'une de ces étapes pourrait être l'activation d'un oncogène cellulaire. En effet, la transduction par les rétrovirus de séquences d'origine cellulaire leur conférant un pouvoir oncogène, a incité à rechercher le rôle éventuel de ces séquences cellulaires dans les cancers spontanés, non viroinduits.

L'étude de quelques c-onc ou ONC trouvés activés dans des tumeurs montre que les modifications qu'ils ont subies participent au processus transformant. D'autre part, il est observé que deux oncogènes sont activés dans certains cancers et des expériences de reconstitution in vitro ont montré que la coopération de ces deux oncogènes est nécessaire pour transformer des cellules normales en cellules tumorales.

Deux stratégies principales ont permis de mettre en évidence l'activation d'oncogènes cellulaires : la transfection de l'ADN de cellules transformées dans les cellules NIH 3T3 et la recherche des modifications de structure des oncogènes cellulaires provoquant leur activation ainsi que l'étude des effets de ces modifications sur la régulation ou les propriétés de ces gènes.

# 1) <u>La transfection d'ADN de cellules tumorales dans les cellules</u> NIH 3T3 :

L'ADN de cellules transformées par des cancérogènes chimiques contient des gènes capables de transformer les cellules NIH 3T3.

Après transfection au phosphate de calcium de l'ADN cellulaire étudié (Weinberg, 1981 ; Cooper, 1982) les foyers de NIH 3T3 morphologiquement

Oncogènes	Produits	Cellules d'origine	Références
Ha-Ras*	p21	EJ: carcinome de la vessie	Santos et al., 1982
Ki-Ras*	(Fixe le GTP	Carcinome pulmonaire	Der et al., 1982
N-Ras	et GTPase)	Sk-N-SH: Neuroblastome	Taparowsky et al., 1983
B-1ym	p 8 apparenlée aux lransférines	RP9: lymphome B aviaire	Goubin et al., 1983
T-1ym-1		S49T: Lymphome murin de type T	Lane et al., 1984
RAF*	10 10 10 10 10 10 10 10 10 10 10 10 10 1	Carcinome pulmonaire	Shimizu et al., 1985
neu / c-erb2	p 1 8 5 Tyrosine kinase	Neuroblastome	Bargmann et al., 1986
TRK	p 6 4 Tyrosine kinase	Carcinome du colon ascendant	Martin-Zanca et al., 1986
MET	Tyrosine kinase ?	MNNG-HOS: Lignée cellulaire humaine transformée par un carcinogène chimique	Cooper et al., 1984
mel	14 - T 1 3 B	NK14: mélanome	Padua et al.,1984
db1		Lymphome B diffus	Eva et Aaronson, 1985
ret		Lymphome de type T	Takahashi et al., 1985
mcf-2 ROS*	Tyrosine kinase ?	MCF-7: Carcinome mammaire	Fasano et al., 1984b Birchmeier et al., 1986
MAS	p 3 3 Transmembrane ?	Carcinome épidermoïde	Young et al., 1986
hst	2	Carcinome gastrique	Sakamoto et al., 1986
KS	Famille des FCF	Sarcome de Kaposi	Delli Bovi et al., 1987 a.b

\* Oncogène déjà trouvé au sein d'un rétrovirus.

TABLEAU 2 : ONCOGÈNES ISOLÉS PAR TRANSFECTION DE L'ADN DE CELLULES TRANSFORMÉES.

transformés sont détectés, et grâce aux techniques du génie génétique, les gènes responsables de cette transformation peuvent être isolés. 10 à 50 % des ADN de tumeurs ou de lignées cellulaires étudiées de cette façon contiennent un gène transformant NIH 3T3, gène qui, en général, est un membre de la famille **ras**.

Un autre gène NEU, apparenté à v-erbB, mais n'ayant pas d'équivalent viral connu, a été détecté de cette façon (Schechter et al., 1984). Par cette méthode, une dizaine de gènes ne présentant pas d'homologie avec des v-onc connus, ont été mis en évidence (Tableau 2).

## 2) <u>Différents types de remaniement des oncogènes cellulaires</u> conduisant à leur activation :

### a) Amplification des c-onc ou ONC :

L'amplification d'un gène se traduit par la présence de plusieurs copies de ce gène dans le génome cellulaire et a souvent pour conséquence la production accrue des ARN messagers correspondants. Ces modifications du génome sont mises en évidence soit par coloration des chromosomes, soit par étude du caryotype; dans le premier cas, on observe la présence de boucles anormalement larges appelées HSR (Homogeneously Staining Regions), et cette amplification est stable, dans le second, l'amplification entraîne l'apparition de chromosomes double-minutes relativement instables.

Une amplification de l'oncogène MYC a été observée dans des lignées cellulaires provenant de cancers humains, tels que la leucémie promyélocytaire HL 60 (Collins et al., 1982) ou un carcinome du colon COLO 320 (Alitalo et al., 1983).

Dans la leucémie myéloïde K 562, l'amplification de ABL en même temps que celle des gènes des chaînes lambda d'immunoglobulines a été décrite (Collins **et al.**, 1983).

### b) Les translocations chromosomiques :

La connaissance de la localisation chromosomique des c-onc sur les chromosomes humains (Tableau 3) a permis d'orienter les recherches vers une éventuelle activation de certains ONC par des translocations spécifiques de certains cancers. Dans plusieurs cas, les points de cassure sont localisés sur les mêmes bandes chromosomiques que les ONC (Pour revue : Rowley, 1983 ; Yunis, 1983).

Chromosome	Bande	Oncogène		
1	Cen p21 p32 p32 p36.1 - p36.2 q12 - q ter q31 - q32	N - RAS B - LYM L - Myc Fgr Ski Jun		
2	p23 - p ter.	N – MYC		
3	p25	MIL1/RAF1		
4		MIL2/RAF2		
5	q34	FMS		
6	p23 - q12 q16 - q22 q21 q22 - q24	Ki - RAS1 ROS FYN MYB		
7	p ter - q22 p11 - q ter	ERB B MET		
8	q11 q24	MOS MYC		
10 9	q34	ABL		
11	p15 q13 q13 q23 - q24	Ha - RAS1 INT2* BCL1* ETS1*		
	p12 - p ter q14 - q ter	Ki - RAS2 INT 1*		
14	q21 - q31	FOS		
16 > 15	q25 - q26	FES/FPS		
17	p13 q11 - q12 q11 - q22	p53 ERB A* NEU		
19 18	q21	BCL2*		
20	q12 - q13	SCR		
21	q22	ETS2*		
22	q11 q12 - q13	BCR* SIS		
X	p21 - q11 q12 - q13	A - RAF1 Ha - RAS2		

\* : pas d'évidence directe d'une activité transformante.

TABLEAU 3 : LOCALISATION CHROMOSOMIQUE DES "ONCOGENES"

Deux translocations ont été particulièrement étudiées, elles impliquent l'oncogène MYC dans les lymphomes de Burkitt et l'oncogène ABL dans la leucémie myéloïde chronique.

Dans le lymphome de Burkitt, la translocation implique un échange de matériel génétique entre le chromosome 8 et le chromosome 14 (85 % des cas) ou les chromosomes 2 ou 22 (15 % des cas). Le point de cassure sur le chromosome 8 est situé dans la bande q24 qui contient le gène MYC humain (Kaplan et Szafnert, 1985). Les points de cassure des trois autres chromosomes impliqués sont situés en 14q32, 2p13 et 22q11 correspondant aux localisations respectives des chaînes lourdes et légères kappa et lambda des immunoglobulines. Dans la plupart des lymphomes de Burkitt, il a été démontré que l'un des deux allèles MYC est transloqué du chromosome 8 sur le chromosome 14 dans le locus des chaînes lourdes d'immunoglobulines (IgH) (Dalla Favera et al., 1982). Dans ces translocations, la proximité des gènes codant pour les gènes d'immunoglobulines active la transcription de MYC (Hayday et al., 1984), stabilise ses ARNm (Adams et al., 1983).

Un autre type de translocation a été décrit dans une leucémie myéloïde chronique. Le caryotype des cellules leucémiques montre un raccourcissement du chromosome 22, appelé alors chromosome 22q<sup>-</sup> ou chromosome Philadelphie (Yunis, 1983). Il résulte d'une translocation (9;22) (q34,q11) dans laquelle l'oncogène ABL (9q34) et le gène BCR (22q11) (BCR : Breakpoint Cluster Region) sont impliqués (Groffen et al., 1984). Ces séquences sont alors réunies en une unité transcriptionnelle, dans laquelle l'oncogène ABL a perdu ses trois premiers exons, traduite en une protéine de fusion P210 BCR-ABL (Shtivelman et al., 1985). Cette protéine possède une activité tyrosine-kinase non observée pour la protéine ABL normale (Konopka et al., 1984).

Ainsi les translocations peuvent affecter l'organisation structurale (comme dans le cas de ABL) ou la régulation (comme dans le cas de MYC) d'un oncogène et contribuer à l'apparition du processus malin.

### c) Activation transcriptionnelle des oncogènes cellulaires :

L'activation transcriptionnelle d'un oncogène cellulaire, sans phénomène de translocation peut être induite par l'insertion d'un promoteur fort au voisinage du **c-onc.** Il s'agit le plus souvent de LTR rétroviraux, l'oncogène
pouvant être intact ou interrompu par l'insertion du LTR. C'est l'oncogène c-myc qui est le plus souvent activé dans les lymphomes T ; chez le chat, par exemple, c-myc est activé par le virus FeLV (Feline Leukemia Virus) (Neil et al., 1984).

Chez le poulet, une érythroblastose peut être induite par l'intégration d'un ALV (Avian Leukosis Virus) dans le locus c-erbB, dont seule la partie 3' est alors exprimée (Raines et al., 1985).

# d) Activation par modification de la séquence nucléotidique des oncogènes cellulaires :

Plusieurs types de modifications sont susceptibles d'activer un oncogène cellulaire :

- des mutations ponctuelles.

- l'addition ou la délétion de séquences codantes ou non.
- \* Activation par mutations ponctuelles :

Les gènes **ras** ont permis de mettre en évidence le rôle des mutations ponctuelles dans l'activation d'un oncogène. La mutation du gène **ras** la plus fréquemment observée dans les tumeurs ou certaines lignées cellulaires se trouve sur le douzième codon de la séquence codante (Reddy **et al.**, 1982 ; Tabin **et al.**, 1982 ; Taparowsky **et al.**, 1982 ; Capon **et al.**, 1983). L'importance de cette mutation sur le pouvoir transformant de **ras** a été confirmée par des expériences de mutagénèse dirigée où le 12éme acide aminé de la protéine **Ha ras-1** était remplacé par n'importe quel autre acide aminé à l'exception de la proline (Seeburg **et al.**, 1984). De plus, la mutagénèse chimique, à l'aide de bisulfate de sodium de ce même gène a montré que des mutations sur les 12e, 13e, 59e (site d'autophosphorylation), 63e acide aminé activaient son pouvoir transformant (Fasano **et al.**, 1984).

## \* Activation par addition ou délétion de séquences codantes ou non :

Les modifications structurales telles que la fusion de séquences exogènes ou la délétion partielle d'un gène jouent un rôle important dans l'acquisition des propriétés transformantes.

Ces modifications ont lieu essentiellement lors de translocations chromosomiques et surtout lors de la transduction des oncogènes cellulaires dans un rétrovirus.

> - Rôle des séquences exogènes dans l'activité transformante des c-onc :

L'oncogène v-abl illustre cet exemple. Le virus murin A-MulV est un virus défectif pour la réplication qui contient dans son génome des séquences c-abl. Ce virus code pour une protéine de fusion gag-abl de masse moléculaire variable selon les isolats mais contenant toujours 30 kd de déterminants gag. L'importance de la partie amino-terminale exogène dans l'activation du pouvoir transformant du produit c-abl peut s'expliquer comme suit :

- elle confère à la protéine une activité tyrosine-kinase dont est dépourvue la protéine **c-abl**. Cette même protéine remaniée par suite de la translocation du gène ABL dans le locus BCR au cours des leucémies myéloïdes chronique possède cette activité (Konopka et al., 1984). Ces séquences BCR pourraient alors jouer le même rôle que les séquences gag dans la protéine virale.

- elle pourrait également intervenir dans la localisation subcellulaire de la protéine.

Rôle des délétions dans l'activité transformante des c-onc :

Le mode d'activation particulier porte sur les séquences régulatrices dont l'action s'exerce au niveau transcriptionnel ou post-transcriptionnel.

L'exemple le plus significatif du mode d'activation transcriptionnel est celui de l'oncogène **c-mos.** Cet oncogène est contrôlé par une séquence inhibitrice de la transcription située en 5' des séquences codantes **c-mos.** Une délétion de cette séquence inhibitrice contribue à l'activation du gène (Wood **et al.**, 1984).

Le produit de **c-fos**, pour transformer les fibroblastes de souris NIH 3T3, doit être sous le contrôle transcriptionnel d'un promoteur rétroviral et doit être amputé d'une séquence de 67 nucléotides située en 3' du signal de polyadénylation du gène, donc en dehors de sa séquence codante. **c-fos** est donc un modèle d'activation par délétion de séquences impliquées dans la régulation



Figure 5 : structure et expression des rétrovirus contenant deux oncogènes dans leur génome

post-transcriptionnelle (Meijlink et al., 1985).

### e) Coopération d'oncogènes :

La tumorigénèse est un phénomène multi-étapes dans lequel deux oncogènes peuvent coopérer. Le premier modèle étudié est celui des oncogènes **myc** et **ras**. Des fibroblastes embryonnaires de rat (REF) (Land **et al.**, 1983) et des cellules épithéliales de reins de jeunes rats (BRK) (Ruley, 1984) peuvent être établis en lignées (immortalisés) et transformées par co-transfection d'un gène **myc** et d'un gène **ras** activés. Chacun de ces deux gènes, agissant seul, ne peut reproduire le phénomène. Ce premier modèle laisse supposer que des associations priviligiées d'oncogènes interviennent dans le cas d'évènements conduisant à la cancérisation de la cellule.

Certains virus défectifs des leucémies aviaires (DLV) possèdent la particularité d'avoir transduit dans leur génome deux oncogènes distincts et ils offrent un modèle de choix pour l'étude de la coopération d'oncogènes dans la tumorigénèse. Ces virus sont peu nombreux, jusqu'à présent trois seulement ont été identifiés : le virus de l'érythroblastose aviaire AEV, l'un des virus de la myélocytomatose aviaire MH2 et enfin le virus de l'érythroblastose aviaire E26. La figure 5 donne l'organisation génomique de ces rétrovirus. Chez AEV, v-erbB est responsable de la transformation des érythroblastes et v-erbA potentialise son action en bloquant la différenciation de ces cellules à un stade immature, le stade CFU-E (Graf et Beug, 1978 ; Gazzolo et al., 1980). Le virus MH2 induit des carcinomes du foie et du rein et transforme les fibroblastes et les macrophages in vitro. Ceci est également observé pour d'autres virus (MC 29, OK 10 et CM II) qui n'abritent dans leur génome que l'oncogène myc (Roussel et al., 1979). La présence de v-mil dans le rétrovirus MH2 lui confère les propriétés d'induire également la prolifération des cellules de neurorétines de poulet en culture (Béchade et al., 1985; Martin et al., 1986). Nous avons choisi d'étudier E26 dont nous décrirons les principales caractéristiques dans le Chapitre III. Il faut cependant dès à présent mentionner que dans les deux premiers cas (AEV et MH2) les deux oncogènes sont exprimés sous forme de deux protéines distinctes issues de deux ARN messagers différents (Anderson et al., 1980 ; Lai et al., 1980 ; Saule et al., 1983 ; Pachl et al., 1983) alors que chez E26, une seule protéine transformante est synthétisée, la P135 gag-myb-ets (Bister et al., 1982).

# CHAPITRE II

# LA SYNTHESE ALTERNATIVE D'ARN MESSAGERS A L'ORIGINE DE L'EXPRESSION MULTIPLE D'UN MEME GENE

L'expression des gènes eucaryotes requiert la mise en oeuvre d'activités biochimiques cellulaires complexes pour la transcription, la maturation et le transport des ARN messagers avant qu'ils ne soient traduits en protéines fonctionnelles. Ces activités sont impliquées dans la régulation du taux d'expression spécifique d'un gène au cours du développement.

Chez les eucaryotes existent, d'une part, des unités transcriptionnelles simples dont l'information code pour une seule protéine et, d'autre part, des unités transcriptionnelles complexes qui engendrent la synthèse de protéines distinctes. Les séquences non codantes de ces unités transcriptionnelles peuvent être directement impliquées dans la régulation de l'expression du gène.

Nous décrivons, ci-après, l'expression d'une unité transcriptionnelle simple puis le rôle de l'épissage alternatif dans l'expression des unités transcriptionnelles complexes en prenant comme exemple, le plus souvent, des oncogènes.

### I. EXPRESSION D'UNE UNITE TRANSCRIPTIONNELLE SIMPLE:

Chez les eucaryotes, la transcription par l'ARN polymérase II d'une unité transcriptionnelle simple implique l'existence de signaux spécifiques d'initiation et de terminaison, la coupure et polyadénylation de l'extrémité 3' et l'élimination des introns du transcrit primaire.

#### 1) Initiation de la transcription et signaux promoteurs:

L'étude des sites d'initiation spécifiques de l'ARN polymérase II a permis d'identifier des séquences promotrices telles que la "TATA box" (Nevins, 1983). Ces séquences bien qu'absentes dans un petit nombre de cas chez les eucaryotes et les gènes viraux, sont indispensables à la transcription **in vitro** par l'ARN polymérase II. Des mutations ou délétions dans ces séquences entraînent la perte de la spécificité d'initiation **in vivo**. C'est ainsi que l'initiation de la transcription se déroule en général à environ 25 nucléotides en aval de la séquence TATA, à un site qui engendre la formation du chapeau 5' (5' Cap) de l'ARN mature. De plus, d'autres éléments situés en amont tels que les séquences CCAAT et GGGCGG sont caractéristiques de nombreux promoteurs.

## 2) Terminaison de la transcription et formation de l'extrémité 3':

Bien que les signaux de terminaison de la transcription soient encore peu connus, il est clair que la transcription s'achève en aval du ou des sites de polyadénylation. Il est généralement observé (Nevins et al., 1978 ; Nevins et al., 1980; Ford et al., 1978) que la transcription prend fin à quelques centaines ou milliers de nucléotides en aval du site poly A où plusieurs autres signaux indiquent la terminaison (Lemeur et al., 1984; Hagenbuchle et al., 1984; Citron et al., 1984). Lorsque la transcription a franchi un site poly A potentiel, une coupure de l'ARN messager naissant, par une endonucléase, peut intervenir, créant l'extrémité 3' à laquelle une séquence poly A d'environ 200 nucléotides est ajoutée (Nevins, 1983; Birnstiel et al., 1985). Dans la plupart des transcrits, une séquence consensus AAUAAA est observée à 10-30 nucléotides en amont du site poly A (Proudfoot et al., 1976). Cette séquence consensus est très conservée chez les différentes espèces animales, sa délétion perturbe la stabilité des ARN messagers et sa mutation en AAGAAA réduit considérablement la coupure adéquate du transcrit (Fitzgerald et al., 1981; Montell et al., 1983). La séquence AAUAAA ne semble pas être seule impliquée dans la polyadénylation correcte des ARN messagers. En effet, des études de mutation des régions en aval de cette séquence laissent supposer l'importance de leur structure secondaire dans ce processus. Dans le cas des gènes cellulaires encaryotes, la polyadénylation peut précéder l'élimination des introns bien qu'il ne puisse être exclu qu'elle se déroule en même temps (Leff et al., 1986).

## 3) Epissage des ARN messagers:

Les unités transcriptionnelles des cellules eucaryotes se composent de régions codantes entre lesquelles s'intercalent des régions non codantes ou introns. Ces séquences sont éliminées du transcrit primaire par un mécanisme appelé "épissage" (en anglais : splicing) ; pour assurer un épissage correct et fidèle, de courtes séquences consensus sont présentes aux confins des introns (Sharp, 1981 ; Mount, 1982).

Exon 1  $5' \stackrel{A}{C} AG^{\dagger} GT^{A}_{G} AGT$  — intron — (Py)<sub>6</sub> X C<u>AG</u>  $G^{G}_{T}$  3' Site depreum Site donneur Site accepteur



Figure 6 : mécanisme d'épissage des ARNmessagers (extrait de m/s nº 7, 1986) L'élimination des introns se fait en deux étapes :

- la première concerne la coupure du site d'épissure donneur au niveau du résidu guanosine consensus (Figure 6). Cette extrémité 5' guanosine libre réagit avec un résidu, généralement l'adénosine, du site d'épissure accepteur en 3', pour former une liaison 5', 2' phosphodiester (Keller, 1984). La structure secondaire de l'intron est alors comparable à un cordage ou un lasso, ce qui explique l'utilisation des termes "épissage" et épissure.

- la deuxième étape implique un phénomène concomittant d'élimination de l'intron et ligation des exons. Ce mécanisme reste encore en bien des points obscur. Cependant, pour certains gènes, le choix du site d'épissure semble guidé par la structure secondaire du transcrit primaire (Munroe, 1984).

De plus, des ribonucléoprotéines nucléaires de petit poids moléculaire (small nuclear RiboNucleoProtein : snRNPs) s'associent aux ARN pré-messagers (Figure 7) (Lerner et al., 1980 ; Rogers et Wall, 1980) et sont supposées maintenir dans le même environnement les exons à venir ainsi que les enzymes participant aux réactions (Maniatis et al., 1987 ; Sharp, 1987a). La structure complexe formée de l'ARN pré-messager et des snRNPs est appelée "spliceosome". Elle est mise en place avant la première étape d'élimination des introns décrite précédemment.

Ce phénomène d'épissage a d'abord été décrit comme un mécanisme intramoléculaire, mais des observations récentes mettent en évidence un mécanisme d'épissage intermoléculaire à partir de deux ou plusieurs ARN messagers précurseurs (Sharp, 1987b).

#### **II. LES UNITES TRANSCRIPTIONNELLES COMPLEXES**

La synthèse de multiples ARN messagers à partir d'une seule unité transcriptionnelle implique l'utilisation de signaux alternatifs à chacune des étapes de production ou de maturation de l'ARNm décrites précédemment. Dans certains cas, les régions codantes n'en sont pas affectées et les conséquences de cette production alternative d'ARNm restent obscures. Dans d'autres cas, la protéine engendrée est différente.

Pour certains gènes dont les produits protéiques alternatifs sont supposés avoir un retentissement sur le phénotype cellulaire, un moyen de réguler la maturation de l'ARN est nécessaire. En retour, une telle régulation peut être soumise à un contrôle spécifique du tissu ou du stade de

26



# Figure 7 : formation du spliceosome (Sharp et al., 1987)

GENE	Synthèse de protéines multiples	Mécanisme spécifique du tissu ou du stade de développement				
1 SITES DE POLYADENYLATION MULTIPLES, EPISSAGE ALTERNATIF DES EXONS :						
* Chaines lourdes d'Ig M, Ig D, Ig E, Ig G	oui	oui				
* Calcitonine de rat	oui	oui				
* Fibrinogène humain	oui	oui				
2 SITES D'INITIATION MULTIPLES, EPISSAGE ALTERNATIF DES EXONS :						
* Chaine légère de myosine de poulet	oui	oui				
* Chaine légère de myosine de souris	oui	oui				
* a amylase de foie et de glande salivaire de souris	oui	oui				
* ABL	?	?				
* p53	oui	?				
3 EPISSAGE ALTERNATIF DES EXONS :						
* Récepteur de l'interleukine 2 humaine	oui	?				
* Fibronectine humaine	oui	oui				
* Troponine T de rat	oui	oui				
* Hormone de croissance humaine	oui	non				
* c-Ki-ras2	oui	?				
* c-myb	?	?				
* c-mil	oui	?				
4 SITES D'INITIATION ET DE POLYADENYLATION MULTIPLES SANS EPISSAGE ALTERNATIF DES EXONS :						
* Dihydrofolate réductase de souris (5', poly A)	non	?				
* Invertase de levure (5')	oui	oui				
* c-fos (poly A)	non	?				
* N-ras humain (poly A)	non	?				

# TABLEAU 4 : EXEMPLES DE GENES ENGENDRANT DE MULTIPLES ARN MESSAGERS.(Leff et al., 1986)

développement. Des exemples d'unités transcriptionnelles complexes peuvent être regroupés en fonction de leur structure génomique (Tableau 4).

La formation de multiples ARNm peut provenir de l'utilisation de sites multiples d'initiation ou polyadénylation, de l'utilisation d'un épissage alternatif des exons ou le plus souvent d'une combinaison de ces deux processus.

# Gènes utilisant des sites de polyadénylation alternatifs et l'épissage alternatif des exons:

Des gènes eucaryotes utilisent lors de leur transcription des sites de polyadénylation alternatifs ainsi qu'un épissage alternatif des exons (Tableau 4). Deux membres de cette classe, les gènes de la calcitonine de rat et de chaines lourdes d'immunoglobulines sont représentés dans la Figure 8.

## \* Chaines lourdes d'immunoglobulines:

Deux formes de régulation affectent l'expression des gènes d'immunoglobulines : l'une est le réarrangement qui permet la création des différentes classes d'immunoglobulines (Honjo **et al.**, 1978 ; Cory **et al.**, 1980 ; Coleclough **et al.**, 1980), l'autre est la maturation des ARNm par sélection de site de polyadénylation ou des sites d'épissage ou de terminaison de la transcription (Perry **et al.**, 1979 ; Alt **et al.**, 1980 ; Rogers **et al.**, 1980 ; Early **et al.**, 1980).

Comme on le voit dans la Figure 8, le passage de la forme membranaire de l'IgM à sa forme sécrétée résulte d'un épissage alternatif de l'ARNm entraînant une modification de la région C-terminale de la protéine codée.

## \* Calcitonine/CGRP:

Le gène de la calcitonine peut être transcrit différemment selon les tissus : la thyroïde contient surtout des ARNm calcitonine alors que dans l'hypothalamus et le cerveau est mis en évidence un transcrit donnant naissance à un "calcitonine gene related peptide" ou CGRP, Figure 8, (Rosenfeld **et al.**, 1981 ; Rosenfeld **et al.**, 1982 ; Amara **et al.**, 1982 ; Amara **et al.**, 1984).

# 1 Immunoglobuline $\mu$



A(n) : Site de polyadénylation

S : régions codantes

Figure 8 : gènes utilisant des sites de polyadénylation multiples et l'épissage alternatif des exons

# 1 $\alpha$ amylase de souris



2 Chaînes légères de myosine



- A(n) : site de polyadénylation
  - S:régions codantes
- Figure 9 : gènes utilisant de multiples sites d'initiation et l'épissage alternatif des exons

# 2) <u>Gènes utilisant de multiples sites d'initiation et l'épissage</u> alternatif des exons :

## \* α amylase et chaines légères de myosine :

Deux gènes eucaryotes, celui de l' $\alpha$  amylase et celui des chaînes légères de la myosine, sont connus pour utiliser plusieurs sites d'initiation et engendrer des ARN messagers différents (Figure 9).

Les ARNm des chaînes légères de myosine diffèrent en 5', ce qui affecte la séquence des protéines isoformes produites (Nabeshima **et al.**, 1984). Par contre, les ARNm des  $\alpha$  amylases, spécifiques du foie et des glandes salivaires diffèrent à leur extrémité 5' non traduites, les protéines codées sont donc identiques.

## \* ABL :

Dans les cellules normales humaines, le proto-oncogène ABL donne naissance à deux types de transcrits qui s'initient dans deux régions promotrices différentes et contiennent deux premiers exons distincts (Ben-Neriah et al., 1986; Shtivelman et al., 1986).

Cet épissage alternatif se produit aussi lors de la transcription du locus **BCR-ABL** spécifique de certaines leucémies myéloīdes chroniques (LMC). Il suggère que la capacité de l'exon II du gène **ABL** de participer à un épissage alternatif, pouvant s'effectuer sur de très longues distances, est indispensable à la formation de l'ARNM **BCR-ABL** important dans la pathogénèse.

## \* p53 :

La protéine **p53** est une protéine associée à la transformation cellulaire et un taux élevé de cette phosphoprotéine nucléaire est un point commun des cellules transformées.

Au moins deux types d'ARNm existent dans ces cellules, ils diffèrent à leurs extrémités 5' : le premier type d'ARNm correspond à la transcription de 11 exons et l'initiation de la traduction a lieu dans l'exon 2 ; le deuxième type d'ARNm est amputé des deux premiers exons et le site d'initiation est différent (Matlashewski et al., 1987) (Figure 10).

La signification biologique de ces différences n'est pas encore connue et l'étude des produits du gène **p53** dans plusieurs types cellulaires en présence ou en l'absence d'autres oncogènes devrait conduire à une meilleure compréhension du rôle de ces protéines dans la transformation.



p53

Figure 10 : gène utilisant de multiples sites d'initiation et l'épissage alternatif des exons

: intron 2

1 c-Ki-ras-2



exons myb présents dans un rétrovirus
exons myb non présents dans un rétrovirus

Figure 11: gènes dont les ARNm présentent un seul site d'initiation , une extrémité 3' commune mais qui diffèrent par l'organisation de leurs exons

# 3 - <u>Gènes dont les ARNm présentent un seul site d'initiation,</u> <u>une extrémité 3' commune mais qui diffèrent par l'organi-</u> sation de leurs exons.

\* c-Ki-ras-2 :

Plusieurs gènes utilisent l'épissage alternatif des exons pour engendrer une diversité des ARNm produits (Figure 11).

Cela peut se produire par inclusion ou exclusion d'un exon particulier comme dans le cas du proto-oncogène **c-Ki-ras - 2** (Mac Grath **et al.**, 1983 ; Shimizu **et al.**, 1983 ; Capon **et al.**, 1983a). Tous les gènes **ras**, y compris **Ha-ras et N-ras**, codent pour des protéines dont la masse moléculaire avoisine 21000 daltons, **p21<sup>ras</sup>** (Chang **et al.**, 1982 ; Capon **et al.**, 1983b ; Hall **et al.**, 1983 ; Taparowsky **et al.**, 1983) **c-Ki-ras-2** donne naissance à deux transcrits majeurs de 5,5 et 3,8 kb respectivement. Le gène comporte cinq exons codants et dans 98 % des transcrits le quatrième exon n'est pas utilisé, seul le cinquième exon participant à la synthèse de la région C-terminale de la protéine. (Capon **et al.**, 1983a). Le gène viral, **c-Ki-ras**, présent dans le virus du sarcome murin MSV, au contraire, utilise le quatrième exon et donc code pour une protéine **p21<sup>ras</sup>** différente de **c-Ki-ras** dans la région C-terminale.

Ce polymorphisme de l'ARNm a été observé chez un autre membre de la famille ras, N-ras, (Hall et al., 1983 ; Murray et al., 1983 ; Hall et al., 1985). Il s'agit là de l'utilisation d'un site de polyadénylation différent sans que la séquence de la protéine qui en dérive en soit altérée.

## \* c-myb :

Deux ARNm c-myb, un ARNm majoritaire de 3,8 kb et un ARNm minoritaire de 4,2 kb, ont été mis en évidence dans plusieurs types de cellules normales de souris (Figure 11).

Récemment, un exon additionnel (E6A) de 363 nucléotides a été observé chez des transcrits **gag-myb** exprimé dans des tumeurs murines de type ABPL (Abelson virus-induced plasmacytoid lymphosarcoma) (Shen-Ong, 1987). On pourrait croire que sa présence résulte d'un épissage alternatif aberrant dû à





# 2 Invertase de levure



- i: site d'initiation
- A(n) : site de polyadénylation
  - S: régions codantes

Figure 12 : gènes produisant des ARNm hétèrogènes quant à leurs extrémités 5' et/ou 3' mais dont l'épissage des exons est constant l'insertion provirale en amont du gène c-myb. Cependant, cet exon E6A a également été identifié dans l'ARNm minoritaire de 4,2 kb isolé à partir du thymus. Bien que la fonction de la protéine c-myb soit inconnue, nous savons que c-myb est exprimé dans les cellules hématopoïétiques immatures et qu'une délétion de la région amino-terminale de la protéine c-myb correspondante est associée à des tumeurs myéloïdes. Il est donc possible que c-myb contienne différents domaines fonctionnels requis pour différentes activités biologiques particulièrement pendant l'hématopoïése. Par conséquent, l'exon E6A trouvé dans une population mineure de transcrits c-myb normaux pourrait constituer un domaine additionnel indispensable à une fonction biologique particulière.

\* c-mil :

Le rétrovirus MH2 possède dans son génome deux oncogènes distincts mil et myc. L'oncogène mil représente une forme tronquée du gène cellulaire de poulet c-mil. En effet, 11 exons de la région 3' de ce dernier ont été transduits dans MH2. Dans les cellules normales de poulet, deux phosphoprotéines c-mil de 71 et 73 kd sont synthétisées, elles correspondent respectivement à deux ARN messagers qui diffèrent l'un de l'autre par l'absence ou la présence d'un exon de 60 bp (Dozier et al., 1988). Cet exon n'est pas transduit dans MH2 et son insertion dans celui-ci ne change pas les propriétés de la protéine de fusion P 100 gag-mil. Les rôles respectifs des deux protéines c-mil restent à définir.

# 4) Gènes produisant des ARNm hétérogènes quant à leurs extrémités 5' et/ou 3' mais dont l'épissage des exons est constant.

Plusieurs gènes utilisent uniquement des sites d'initiation ou de polyadénylation alternatifs pour engendrer des ARNm différents et dans la plupart des cas, ces ARNm codent pour des protéines identiques. Par exemple, 4 ARNm différents peuvent coder pour la dihydrofolate réductase dans les cellules de souris (Figure 12) (Setzer et al., 1982; Kaufman et al., 1983). Une exception à cette généralité est représentée par le gène de l'invertase de levure (Figure 12). Deux sites d'initiation peuvent être utilisés pour engendrer une forme cytoplasmique ou secrétée de l'enzyme. Le pouvoir transformant du proto-oncogène **c-fos** est inhibé par des séquences localisées en 3' du gène, non codantes, qui semblent interagir avec l'extrémité 3' des séquences codantes pour réduire l'efficacité de traduction et de transport des ARNm **c-fos** (Miller **et al.**, 1984).

#### CONCLUSION

Comme nous l'avons vu précédemment, un même gène viral ou eucaryote peut être transcrit en différents ARNm donnant naissance à des protéines distinctes mais apparentées. Ces protéines peuvent avoir des fonctions différentes ou, pour les protéines eucaryotes, être spécifiques de certains tissus. Les mécanismes d'épissage alternatif constituent un des modes de régulation de l'expression d'un gène. En ce qui concerne les virus, on peut émettre l'hypothèse que l'épissage alternatif donne une plus grande capacité fonctionnelle à un génome relativement limité.

Les exemples de gènes à l'origine de plusieurs types de protéines sont de plus en plus nombreux chez les eucaryotes et on peut penser que les mécanismes d'épissage alternatif sont très fréquents voire même systématiques. Cette observation bouscule l'ancien dogme qui associait un gène à une protéine. Peut-être faut-il plutôt associer un gène à une famille de protéines plus ou moins reliées entre elles.

Notre travail de thèse a été consacré au proto-oncogène **c-ets-1** de l'ADN cellulaire de poulet, gène hautement apparenté à l'oncogène viral **v-ets** présent dans le génome du rétrovirus E26. Ce proto-oncogène est à l'origine d'au moins deux protéines distinctes chez le poulet et nous avons mis en évidence l'existence d'un mécanisme d'épissage alternatif directement impliqué dans cette multiplicité.

Nous faisons ci-après un bref rappel des connaissances actuelles sur le rétrovirus E26 puis l'exposé de nos résultats.

# CHAPITRE III

LES RETROVIRUS AMV ET E 26



CFU : Colony Forming Unit	Hb : Hémoglobine
BFU : Burst Forming Unit	Br : Antigène de cerveau
	Im : Antigène Im (spécifique des érythrocytes).

Figure 13 : cellules cibles des rétrovirus E26 et AMV dans le système hématopoïétique

#### I. ACTIVITE BIOLOGIQUE

E26 est un rétrovirus défectif pour la réplication qui a été isolé à partir d'une érythroblastose spontanée du poulet par Ivanov et al., 1962. Ce rétrovirus induit chez cet animal des leucémies de type mixte érythroīde, myéloīde à prédominance érythroīde in vivo (Moscovici et al., 1981; Radke et al., 1982). In vitro, E26 transforme les myéloblastes et les érythroblastes (Graf et al., 1979; Moscovici et al., 1981; Radke et al., 1982). Des cellules érythroīdes ou méyloīdes peuvent être propagées sélectivement à partir de cellules leucémiques mises en culture en présence de milieu CFU-E (Erythroid Colony Forming Unit) ou CFU-M (Myeloid Colony Forming Unit) (Beug et al., 1982; Radke et al., 1982; Moscovici et al., 1983; Samarut et Bouabdelli, 1980). Par ailleurs, les cultures de moëlle infectées par E26 et ensemencées en milieu CFU-M semi-solide bourgeonnent en colonies myéloīdes tandis qu'en milieu CFU-E semi-solide, elles bourgeonnent en colonies érythroīdes (Radke et al., 1982).

L'identification des cellules cibles de E26 a été rendue possible grâce à l'obtention de deux antiséra caractérisant respectivement les cellules progénitrices érythroīdes BFU-E et CFU-E (Gazzolo et al., 1980 ; Moscovici et al., 1983 ; Radke et al., 1982) ; c'est ainsi qu'il a été observé que les cellules cibles de E26 se situent à un stade de développement intermédiaire entre les cellules BFU-E et CFU-E (Figure 13). Les cellules myéloïdes transformées par E26 sont des cellules plus immatures que celles transformées par AMV (qui ne contient que l'oncogène v-myb), et leur différenciation est bloquée de façon beaucoup plus stricte (Beug et al., 1984). Au contraire, les colonies érythroïdes obtenues à partir de cellules de moëlle infectées par E26 contiennent des érythroblastes mais aussi des érythrocytes, ce qui prouve que le virus E26 ne bloque pas de façon très rigide la différenciation de ces cellules.

Les myéloblastes transformés par E26 nécessitent pour proliférer in vitro la présence d'un facteur de croissance de type cMGF (chicken Myelomonocytic Growth Factor) obtenu initialement à partir du surnageant de cellules de rate stimulées à la concanavaline A. Ce facteur semble requis pour la croissance des cellules de la lignée myéloïde (Beug et al., 1982; Leutz et al., 1984). Les oncogènes viraux dont les produits possèdent une activité tyrosinekinase sont capables d'induire ces myéloblastes à produire de façon autocrine un

Transformation par	E26 wt		E26 ts	
	myélob	érythr	myélob	érythro
température permissive 37°C	+	+	+	+
température non permissive 42°C	+	+		+

В



Figure 14 : A - Capacité des mutants E26ts à transformer les érythroblastes et les myéloblastes B - Localisation des mutations dans l'oncogène v-myb du rétrovirus E26ts143 facteur de croissance de type cMGF et à s'auto-stimuler (Adkins **et al.**, 1984). En effet, les myéloblastes transformés par E26 puis surinfectés par des rétrovirus ayant transduit des oncogènes de la famille **src**, poussent indépendamment de la présence de cMGF.

Le rétrovirus E26 est capable d'induire quelques critères de transformation des fibroblastes embryonnaires de caille (Graf **et al.**, 1979). Par contre, il ne transforme pas complètement ceux de poulet (Jurdic **et al.**, 1987), mais leur confère une relative indépendance vis-à-vis des conditions de culture.

Quatre mutants thermosensibles pour la transformation des myéloblastes (ts E26) ont été isolés récemment (Beug et al., 1984). Leur étude permet de mieux comprendre le rôle respectif de v-myb et de v-ets dans la transformation des myéloblastes et des érythroblastes.

Les mutants tsE26 possèdent la capacité de transformer les érythroblastes à la température non permissive de 42°C. Les érythroblastes transformés par les tsE26 isolés à 37°C puis cultivés à 42°C présentent les mêmes caractéristiques que les érythroblastes transformés par le virus E26 de type sauvage à 42°C. Le provirus biologiquement actif de l'un de ces mutants (E26 ts 143) a été cloné et de la séquence nucléotidique de l'oncogène v-myb correspondant a été déterminé. Elle montre la présence de trois mutations ponctuelles, l'une d'entre elles est silencieuse, l'autre conservative, enfin la troisième provoque le remplacement d'une thréonine par une arginine dans la protéine codée. Cette substitution intervient dans la région responsable de l'affinité de la protéine pour l'ADN. Cette unique mutation réintroduite dans le gène v-myb d'un rétrovirus E26 de type sauvage (Figure 14) est suffisante pour induire le phénotype thermosensible (Li et al., en préparation).

### II. STRUCTURE DU GENOME VIRAL

L'organisation structurale de AMV est la suivante :  $5^{\circ}$  LTR - gag -  $\triangle$  pol - myb -  $\triangle$  env - LTR 3'

Quant au rétrovirus E26, il a transduit dans son génome deux oncogènes distincts **v-myb** et **v-ets** (Leprince **et al.**, 1983 ; Nunn **et al.**, 1983). Son organisation structurale est du type :

5' LTR -  $\triangle$  gag - myb- ets -  $\triangle$  env - LTR 3'

ALV



Figure 15 : structure et expression des rétrovirus AMV et E26

### a. L'oncogène v-myb

**\* v-myb** est le seul oncogène présent dans le virus de la myéloblastose aviaire (AMV). Ce virus induit chez le poulet une leucémie monocytaire, les cellules cibles étant celles de la lignée macrophagique, qui sont bloquées à un stade de différenciation intermédiaire entre le monoblaste et le macrophage (Figure 13) (Gazzolo **et al.**, 1979 ; Beug **et al.**, 1979 ; Durban et Boettiger, 1981).

Une recombinaison entre le génome du virus auxiliaire ("helper") MAV1 et une séquence cellulaire **c-myb** sont à l'origine de la formation du génome de AMV. Cette recombinaison s'effectue en 5' de **v-myb** dans le gène **pol**, 111 nt en amont du codon de terminaison de ce gène et en 3' dans le gène **env**, 33 nt en amont du codon de terminaison de ce gène (Figure 15) (Roussel **et al.**, 1979; Klempnauer **et al.**, 1982).

Le provirus AMV est transcrit en un ARN génomique de 7,5 kb qui code pour une protéine de 180 kd, P180**gag-pol** et un ARN sous-génomique de 2,5 kb (Klempnauer **et al.**, 1982) traduit en une protéine transformante de 45 à 48 kd, p45**v-myb** (Figure 15) (Rushlow **et al.**, 1982 ; Boyle **et al.**, 1983). Elle présente à son extrémité N-terminale, six acides aminés du produit **gag** et à son extrémité C-terminale, onze acides aminés du produit **env**. Des expériences de délétion ont montré que les séquences rétrovirales **gag** et **env** ne sont pas nécessaires à l'activité transformante de AMV (Lipsick **et al.**, 1987). Cette protéine, p45**v-myb** est nucléaire dans les myéloblastes transformés mais acquiert une localisation cytoplasmique lorsque ces derniers sont induits à se différencier en macrophages (Klempnauer **et al.**, 1984 ; Ibanez **et al.**, 1988).

\* L'oncogène v-myb présent dans le virus E26 (v-myb E26) est tronqué en 5' et 3' par rapport à l'oncogène v-myb présent dans le virus AMV (v-myb AMV) (Klempnauer et al., 1982; Leprince et al., 1983; Nunn et al., 1983, 1984) Figure 15. D'autre part, l'oncogène v-myb E26 n'est pas traduit en une protéine v-myb mais en une protéine transformante de fusion gag-myb-ets de masse moléculaire 135 kd, P135 gag-myb-ets dont la localisation est nucléaire (Bister et al., 1982; Klempnauer et al., 1984; Boyle et al., 1984).

#### b. L'oncogène v-ets

Le deuxième oncogène transduit dans le rétrovirus E26 et spécifique



Figure 16 : structure et expression des rétrovirus contenant deux oncogènes dans leur génome de celui-ci est l'oncogène v-ets (E-twenty-six specific) dont la taille est de 1,6 kbp. (Leprince et al., 1983; Nunn et al., 1983). Alors que dans AEV et MH2, les deux oncogènes sont exprimés sous forme de deux protéines distinctes issues de deux ARNm différents, chez E26, une seule protéine transformante est synthétisée, il s'agit d'une protéine de fusion dont la masse moléculaire avoisine 135 kd : P135 gag-myb-ets (Figure 16) (Beug et al., 1982; Bister et al., 1982). Cette protéine est issue d'un ARN génomique de 5,7 kb et sa localisation est nucléaire (Boyle et al., 1984; Klempnauer et al., 1984). Elle possède la propriété de se lier à l'ADN in vitro (Moelling et al., 1985).

#### III. LES PROTO-ONCOGENES c-myb ET c-ets

#### a. c-myb

- Structure

Le gène **c-myb** chez le poulet, contient sept régions homologues à **v-myb** : El à E7 (Figure 17) (Klempnauer **et al.**, 1982). Les régions cellulaires E2, E3, E4, E5, E6 possèdent les caractéristiques des séquences exoniques : un cadre de lecture ouvert et des sites d'épissure donneur et accepteur nécessaires à l'élimination des introns.

La région E1 présente dans AMV correspond à un exon cellulaire flanqué en 5' d'une région intronique qui comporte le site d'épissure accepteur utilisé dans AMV lors de la formation de l'ARN sous-génomique  $\Delta$  gag-myb de 2.5 kb. Dans E26 la région E1 est amputée en 5' de sa partie intronique et du début de l'exon, la jonction gag-myb se fait de telle sorte que le cadre de lecture de c-myb soit conservé.

La région E7 correspond dans AMV à un exon cellulaire tronqué dans sa partie 3' et dont le cadre de lecture se poursuit dans **env.** Dans E26, l'homologie avec **c-myb** s'arrête dans l'exon cellulaire E6, lui aussi, tronqué dans sa partie 3' et dont le cadre de lecture se poursuit dans **v-ets** (Leprince **et al.**, 1983; Nunn **et al.**, 1983).

Le proto-oncogène **c-myb** est très conservé phylogénétiquement (Leprince **et al.**, 1983 ; Franchini **et al.**, 1983 ; Gonda **et al.**, 1985 ; Slamon **et al.**, 1986 ; Katzen **et al.**, 1985). Chez l'homme, **MYB** est localisé sur le chromosome 6q22-24 (Harper **et al.**, 1983).

L'étude des ADN complémentaires des ARNm **c-myb** chez le poulet (Rosson et Reddy, 1986), la souris (Gonda **et al.**, 1985) l'homme (Slamon **et al.**, 1 Kbp



Figure 17 : exons du locus c-myb transduits dans AMV et E26

1,2,3,4,5,6,7 : exons E1,E2,E3,E4,E5,E7

1986), la drosophile (Katzen et al., 1985) montre une très forte homologie entre les protéines c-myb des différentes espèces considérées. Les séquences protéigues correspondant aux exons E1 et E2 sont particulièrement bien conservées : 92 % d'homologie entre les exons humains et aviaires, et 99,5 % entre les exons humains et murins. Ces exons correspondent au domaine aminoterminal des protéines c-myb où trois séquences d'environ 50 résidus se trouvent répétées. La région de la protéine c-myb responsable de l'affinité pour l'ADN est localisée dans le domaine amino-terminal de la protéine virale p45v-myb. Ce domaine est particulièrement bien conservé chez les protéines virales et cellulaires (Klempnauer et al., 1986c; Klempnauer et Sippel, 1986a; Peters et al., 1987). Elles possèdent au moins deux sites antigéniques situés dans cette région ; lorsque l'un d'entre eux est bloqué par un anticorps monoclonal, la protéine v-myb exprimée dans un système bactérien ne fixe plus l'ADN. D'autre part, la délétion de ce domaine entraîne la perte de la capacité de cette protéine à se fixer sur l'ADN. La protéine p45<sup>v-myb</sup> ainsi tronquée n'a plus d'activité transformante (Ibanez et al., 1988). De plus, l'étude du produit des mutants E26, thermosensibles pour la transformation des myéloblastes a montré que le domaine myb de la protéine de fusion P135 gag-myb-ets mutée n'était plus capable de se lier à l'ADN (Moelling et al., 1985). Enfin, il existe 97 % d'homologie entre les séguences protéigues humaines et aviaires correspondant de l'exon E5. La séquence protéique des protéines p45v-myb et p75<sup>c-myb</sup> diffèrent par 11 substitutions d'acides aminés. Des expériences de mutagénèse dirigée, ont montré que ces divergences n'étaient pas nécessaires à l'activité transformante de la protéine p45<sup>v-myb</sup> (Stober-Grasser et al., 1988).

### - Expression

Le proto-oncogène **c-myb** est transcrit chez le poulet en un ARN messager de 4.0 kb qui code pour une protéine de 75 kd, p75 **c-myb**, dont la  $p45^{v-myb}$  représente une version tronquée aux extrémités N- et C-terminales.

Chez l'homme, le produit du proto-oncogène **MYB** est une protéine de 80 kd, p80**MYB** (Klempnauer **et al.**, 1986b).

Les protéines c-myb sont nucléaires, elles possèdent la capacité de se lier à l'ADN. Ainsi, la majeure partie des protéines c-myb est associée à la chromatine, une fraction d'environ 5 % reste associée à la matrice nucléaire (Klempnauer et al., 1986b ; Klempnauer et Sippel, 1986a). Les protéines c-myb ont une demi-vie d'environ une heure. Cela suggère un rôle régulateur éventuel, rôle qui semble être une caractéristique des protéines nucléaires (Hann et Eisenman, 1984; Reich et al., 1983).

Récemment, un deuxième transcrit du proto-oncogène **c-myb** a été mis en évidence dans les cellules normales et tumorales de souris (Rosson **et al.**, 1987 ; Shen-Ong **et al.**, 1987). Cet ARN messager comporte un exon complémentaire nommé E6A non transduit dans les virus AMV ou E26. Le rôle de la synthèse multiple de protéines **myb** à partir d'un seul gène **myb** reste à définir.

Le proto-oncogène c-myb est exprimé d'une part, dans les cellules hématopoïétiques et plus particulièrement dans les cellules immatures (Gonda et al., 1982; Coll et al., 1983; Sheiness et Gardiener, 1984; Gonda et Metcalf, 1984) et d'autre part, pendant la prolifération des fibroblastes embryonnaires de poulet (CEF) (Thompson et al., 1986). L'expression de c-myb ne semble pas spécifique d'un type cellulaire mais plutôt du niveau de prolifération de celui-ci. La différenciation cellulaire s'accompagnerait d'une diminution de l'expression de c-myb (Westin et al., 1982; Sheiness et Gardiener, 1984; Craig et Block, 1984).

La relation entre le taux d'expression de **c-myb** et les différentes étapes du cycle cellulaire semble dépendre du type de cellules que l'on considère et de leurs conditions de culture. Il semble que l'augmentation du taux de transcription de **c-myb** précède les changements dans la morphologie et le contenu en ARN de la cellule (Stern **et al.**, 1986).

Ce taux de transcription mesuré de **c-myb** ne varie pas ou peu dans les cellules CEF ou MSB1 quiescentes ou en phase exponentielle de croissance. Cependant, le traitement de ces cellules par la cycloheximide (inhibiteur de la synthèse peptidique) induit une augmentation de la transcription de **c-myb** (Thompson **et al.**, 1986).

#### b. c-ets

En 1985, la connaissance des équivalents cellulaires normaux de **v-ets**, était très limitée.

Chez l'homme, deux loci homologues à **v-ets** localisés sur deux chromosomes différents avaient été mis en évidence.

- Hu-c-ets-1 ou ETS-1 sur le chromosome 11q23-24 (De Taisne et al., 1984; Watson et al., 1985).



Figure 18 : homologie entre les ADNc ETS1 et ETS2 et l'oncogène <u>v-ets</u>

- Hu-c-ets-2 ou ETS-2 sur le chromosome 21q22-1-22-3.

Ces deux loci ont également été localisés sur des chromosomes différents dans le génome de la souris et du chat (Watson **et al.,** 1986).

La comparaison de la séquence nucléotidique partielle d'un fragment génomique du locus ETS-1 et d'un ADNc du locus ETS-2 avec le gène viral v-ets permettait à cette époque de définir trois régions d'homologie dont deux seraient spécifiques de ETS-1 et de ETS-2, la troisième, (14 codons), étant commune aux deux loci (Figure 18). Le cadre de lecture reste ouvert en 5' et en 3' de ces régions d'homologie et aucun site d'épissure donneur ou accepteur ne les délimite.

En 3', les 13 derniers codons de ETS-2 sont totalement différents de v-ets. Ainsi la partie v-ets de la protéine virale P135 gag-myb-ets et le produit de ETS-2 diffèrent à leur extrémité C-terminale.

Les séquences protéiques codées par ETS-1 et ETS-2 (Watson et al., 1985) présentent respectivement 98% et 95% d'homologie avec leurs équivalents viraux. Cette conservation au cours de l'évolution suggère l'importance du rôle probablement joué par ce gène dans la cellule normale.

ETS-1 est transcrit en un ARN messager unique de 6,8 kb. ETS-2 donne naissance à trois ARN messagers de 4,7 ; 3,2 ; 2,7 kb (Watson et al., 1985).

Chez le poulet, l'étude des produits du gène c-ets a été réalisée grâce à l'utilisation d'un antisérum spécifique de v-ets (Chen et al., 1985; Ghysdael et al., 1986a; Ghysdael et al., 1986b). Deux familles de protéines exprimées par deux loci différents c-ets-1 et c-ets-2, ont été identifiées :

- une protéine de masse moléculaire 54 kd, p54 **c-ets-1** exprimée à un taux élevé dans les lymphocytes du thymus et de la bourse de Fabricius, à un taux faible dans les lymphocytes circulants et dans les autres tissus (Ghysdael **et al.**, 1986a). La p54 **c-ets-1** est localisée dans le noyau cellulaire comme la protéine virale, P135 **gag-myb-ets** et la protéine p45 **v-myb** (J. Ghysdael, communication personnelle). Une protéine de 56 kd, p56 **c-ets-1** ainsi que d'autres en moins grande quantité sont exprimées à un taux plus faible dans les mêmes cellules. La p56 **c-ets-1** est apparentée à la protéine p54 **c-ets-1** et des travaux récents (Pognonec **et al.**, 1988) montrent qu'elle en dérive par modifications post-traductionnelles.

- Trois protéines de 60, 62, 64 kd, p60 **c-ets-2**, p62 **c-ets-2**, p64 **c-ets-2**, comportant un domaine limité d'homologie avec la protéine

p54 c-ets-1, sont exprimées à taux élevé dans les macrophages normaux ou transformés mais ne sont pas mises en évidence dans les cellules de moëlle osseuse ni dans les cellules lymphoïdes immatures (Ghysdael et al., 1986b). De plus, la synthèse des protéines  $p62^{c-ets-2}$  et  $p64^{c-ets-2}$  est augmentée dans les myéloblastes transformés par AMV dont la différenciation en macrophages est induite par le TPA (12-0-tétra-décanoyl-phorbol-1.3-acétate). Les protéines p60, p62, p64 c-ets-2 pourraient donc jouer un rôle dans la différenciation des macrophages.

Ainsi l'expression du proto-oncogène **c-ets-1** chez le poulet et chez l'homme se révèle très complexe. Cette complexité nous a incités à étudier chez le poulet l'organisation génomique et l'expression de ce gène (par le biais des ADN complémentaires (ADNc) des ARN messagers **c-ets-1** codant pour la protéine cellulaire p54 **c-ets-1**). Cette étude nous permettra d'une part, d'analyser les différences existant entre le gène viral et son homologue cellulaire normal ainsi que la responsabilité de ces divergences dans l'activité transformante du rétrovirus E26. D'autre part, nous essaierons de proposer un modèle de mécanisme à l'origine de la transduction de **c-ets-1**.

40
### CHAPITRE IV

### RESULTATS

Clonage moléculaire du gène Ck-c-ets-l







Figure 19 : organisation du locus <u>c-ets-1</u> chez le poulet (<u>Ck-c-ets-1</u>)

### I. ORGANISATION DES SEQUENCES HOMOLOGUES A v-ets DANS LE GENOME DE POULET (ARTICLE 1 EN ANNEXE)

Cette étude a été réalisée en collaboration avec D. Leprince et A. Gegonne en utilisant des banques d'ADN cellulaire de poulet. 80 kbp d'ADN cellulaire ont été clonés dans 11 phages (vecteur Charon 4A) et sélectionnés à l'aide de sondes virales (Figure 19). L'étude en microscopie électronique des "hétéroduplex" (hybrides d'ADN génomique et proviral) montre que les séquences homologues à **v-ets** se répartissent en deux domaines que séparent 40 kbp d'ADN cellulaire non apparenté au gène viral.

Le premier domaine comporte deux régions d'homologie  $\alpha$  et  $\beta$  qui représentent environ 230 nucléotides de l'extrémité 5' de **v-ets**, le deuxième domaine comporte 6 régions d'homologie, nommées **a** à **f**, qui rendent compte de la majeure partie du gène viral, i.e. 1200 nucléotides.

Afin de déterminer la nature exacte de  $\alpha$ ,  $\beta$ , **a** et **f**, ainsi que leur relation avec l'oncogène viral **v-ets**, nous en avons déterminé la séquence nucléotidique, (Figure 20).

Il apparaît que  $\alpha$  et  $\beta$  possèdent les caractéristiques de séquences exoniques : un cadre de lecture ouvert, colinéaire à **v-ets** flanqué de sites accepteur et donneur d'épissure. Il en est de même pour a qui fusionne avec l'exon  $\beta$  grâce à des sites d'épissure donneur et accepteur adéquats pour engendrer l'extrémité 5' de **v-ets** (Figure 20).

Quant à la région **f**, elle se présente comme un exon tronqué (Figure 21). Les résultats de séquence montrent que les 13 derniers codons du gène cellulaire (exon **f**) sont remplacés par 16 codons différents dans **v-ets**. Ainsi la protéine cellulaire p54 **c-ets-1** et la partie **v-ets** de la protéine virale P135 **gag-myb-ets** pourraient différer à leur extrémité C-terminale.

Par ailleurs, l'utilisation d'une sonde spécifique de  $\alpha$  et  $\beta$  a permis d'observer que ces séquences ne sont pas transcrites dans l'ARNm majoritaire de 7,5 kb codant probablement pour la protéine p54 C-ets-l exprimée à taux élevé dans les cellules MSB1 (lignée de cellules lymphoïdes transformées par le virus de Marek). Ainsi la protéine cellulaire p54 C-ets-l et la partie v-ets de la protéine P135 gag-myb-ets diffèrent également dans leur région N-terminale.

Cela nous a amenés à nous poser les questions suivantes : tout d'abord, quelle est la relation entre le domaine **v-ets** de P135 **gag-myb-ets** et

v-ets AT CAT GGC ACC TCA GAG ATG ATG AGT TAC TAC ATG GAC ACA ACC ATT GGC AGC ACG GGT Ck c-ets CT GCA GCC ACC TCA GAG ATG ATG AGT TAC TAC ATG GAC ACA ACC ATT GGC AGC ACG GGT TSa? v-ets CCT TAT CCT TTG GCT CGC CCT GGA GTG ATG CAA GGT GCT AGC AGC TGC TGT Ck c-ets CCT TAT CCT TTG GCT CGC CCT GGA GTG ATG GTA AGA AAG GGA TGG GGG TTA Sd

v-ets CTG GGA ATC CCC AAA GAT CCC CAG CAG TGG Ck c-ets CTG GGA ATC CCC AAA G<u>GT</u> AAG AGT AGG CAG

Ŝ₫

v-myb

X

# Β

Sd Ck c-etsB GAG CAC TAC AGC ACA GGT AGG AAC CCC TGT v-ets GAG CAC TAC AGC ACA GAC ATG GAA TGT GCA Ck c-etsa CTT CTG GTG TTT TC<u>A G</u>AT ATG GAA TGT GCA Sa

Figure 20: A-séquence nucléotidique des exons  $\alpha$ ,  $\beta$  et a B-jonction des exons  $\beta$  et a dans <u>v-ets</u>



Ck-c-ets-1 : locus c-ets-1 chez le poulet Hu-c-ets-2 : locus c-ets-2 chez l'homme

> Figure 21: comparaison des extrémités 3' de <u>v-ets</u> et de c-ets-1 chez le poulet et <u>c-ets-2</u> chez l'homme

son équivalent cellulaire présumé p54 c-ets-1? Ensuite, quelle est la nature de  $\alpha$  et  $\beta$  et comment furent-ils transduits dans E26? Enfin, quelle est l'origine des 16 derniers codons de **v-ets**?

Pour répondre à ces questions, nous avons étudié des ADN complémentaires des ARNm c-ets-1 issus de banques d'embryon de poulet ou de cellules de rate de poulet.

## II. <u>ETUDE DE L'ADN COMPLEMENTAIRE DE L'ARNM c-ets-1</u> CODANT POUR LA PROTEINE P54 c-ets-1 (ADNC TYPE I) (ARTICLE 2 EN ANNEXE)

L'utilisation d'une banque d'ADN complémentaires des ARNm c-ets-1 de cellules de rate de poulet, cellules qui expriment à taux élevé la protéine p54 c-ets-1 (Chen et al., 1985) a permis d'isoler dans un premier temps un ADNc de 1,4 kb correspondant à la majeure partie de v-ets (à l'exception de la région 5'). Il représente la partie codante de l'ARNm majoritaire c-ets-1 dont la taille est de 7,5 kb et ne possède pas en 3' la région non traduite et la queue de poly A.

En effet, il existe un site EcoRI localisé au niveau du codon de terminaison de l'exon f (Figure 21), ce site n'a pas été correctement protégé par la méthylation au moment de la construction de la banque d'ADNc et a été utilisé comme site de clonage. Afin d'isoler le fragment correspondant aux séquences 3' non codantes de l'ADNc, une deuxième banque d'ADNc de fibroblastes embryonnaires de poulet a été criblée avec une sonde virale. Un ADNc de 2.0 kbp environ a été isolé, il se compose de deux fragments EcoRI, l'un de 1,3 kbp correspondant à la majeure partie de **v-ets** et l'autre de 0,8 kbp non homologue au gène viral, s'hybride avec le clone génomique **c-ets-5** contenant l'exon f et correspond à la partie 3' non codante de l'ARNm **c-ets-1** (Figure 19).

#### 1) Expression dans un système eucaryote

Afin de comparer la protéine codée par l'ADNC décrit ci-dessus et la protéine p54<sup>c-ets-1</sup> normale, cet ADNC a été cloné dans un vecteur d'expression, pSVL, contenant un promoteur SV40 et les signaux d'épissage et de polyadénylation nécessaires à la transcription et à la traduction du fragment



1, 3, 5, 7 : immunoprécipitation par un sérum anti-ets-A.

2, 4, 6 : immunoprécipitation par un sérum anti-ets-A inhibé par un excès de protéines bactériennes immunogènes.

Figure 22 : expression de l'ADNc p54<sup>c-ets-1</sup>dans les cellules COS-1 comparaison avec les protéines <u>c-ets-1</u> normales exprimées dans le thymus de poulet sain inséré (Figure 22). Le vecteur pSVLets ainsi obtenu, est transfecté par la méthode au sulfate de dextran dans les cellules COS-1 de singe (Gluzman et al., 1981). Ces cellules expriment la protéine large T de SV 40 et sont permissives pour sa réplication. Elles permettront la transcription et donc la traduction dans un système cellulaire eucaryote du gène inséré dans un vecteur contenant une origine de réplication de SV40. Celles-ci s'accompagnent respectivement de la maturation des ARN messagers et des modifications posttraductionnelles des protéines, mimant ainsi l'expression d'un gène eucaryote normal. Les cellules sont marquées à la méthionine 35S puis lysées. Les lysats, immunoprécipités par un sérum anti-etsA, correspondant à une région située à une centaine de résidus en amont de l'extrémité C-terminale (Gégonne et al., 1987) révèlent que les cellules transfectées par pSVLets synthétisent deux polypeptides majeurs de 54 et 56 kd environ. Ces protéines malgré une masse moléculaire théorique de 48 kd migrent en gel d'acrylamide-SDS comme les protéines p54c-ets-1 et p56c-ets-1 normales exprimées dans le thymus de poulet. De plus, leurs cartes peptidiques sont identiques à celles des protéines normales (J. Ghysdael, communication personnelle).

Cet ADNc est donc capable de coder pour p54 c-ets-l et pour p56 c-ets-l qui en dérive par modification post-traductionnelle (Pognonec et al., 1988).

#### 2) Analyse de la séquence nucléotidique

La séquence nucléotidique de cet ADNc a été déterminée par la méthode des terminaisons de chaînes de Sanger **et al.**, 1977, (Figure 23). Elle commence en 5' par un codon d'initiation situé dans un contexte favorable selon la séquence consensus d'initiation des ARNm eucaryotes définie par Kozak (1986).

- Un large cadre de lecture ouvert (nt 1 à nt 1404) apparaît colinéaire à **v-ets** à 5 mutations ponctuelles près. Ce domaine s'étend de l'exon **a** à la première partie (100 nt, environ) de l'exon **F** de l'ADN cellulaire.

- En 3', les séquences de **v-ets** et de l'ADNc divergent et nous ne connaissons pas pour l'instant l'origine des 16 derniers codons de **v-ets**. En raison de cette divergence, nous avons appelé **f** la partie de l'exon cellulaire homologue à **v-ets** et **F** l'exon lui-même.

- En 5', les 79 premiers nucléotides codants de cet ADNc sont différents de  $\alpha$  et  $\beta$ . Une sonde spécifique (fragment EcoRI-ClaI Figure 24) a

ANG TTC TGC ATG ANC GGA GCT GCC CTG TGC GCC CTG GGC ANG GAG TGC TTC CTG GAG CTA CGG CCT GAC TTT Lys Phe Cys Met Asn Gly Ala Ala Leu Cys Ala Leu Gly Lys Glu Cys Phe Leu Glu Leu Ala Pro Amp Phe +00 GTG GGA GAT ATC CTT TGG GAA CAC CTG GAG ATC TTG CAG AAA GAA GAA GAA GCA AAA CCA TAC CCA GCA AAT GGA Val Giy Amp lie Leu Trp Giu His Leu Giu IIe Leu Gin Lys Giu Giu Ala Lys Pro Tyr Pro Ala Asn Giy 4 50 GTG AAT GCA GCG TAT CCA GAA TCC CGC TAT ACT TCA GAC TAC TTC ATT AGT TAT GGC ATC GAG CAC GCA CAG Yal Asn Ala Ala Tyr Pro Glu Ser Arg Tyr Thr Ser Asp Tyr Phe lle Ser Tyr Gly lle Glu His Ala Gin 550 TGC GTG CCT CCC TCC GAG TTC TCT GAG CCC AGC TTC ATC ACA GAG TCC TAC CAG ACC CTC CAT CCC ATC AGC Cys Val Pro Pro Ser Glu Phe Ser Glu Pro Ser Phe Ile Thr Glu Ser Tyr Gin Thr Leu His Pro Ile Ser 600 TCG GAA GAG CTT CTG TCC CTC AAG TAC GAG AAC GAC TAT CCC TCA GTC ATC CTT CGT GAC CCC GTC CAG ACG Ser Glu Glu Leu Leu Ser Leu Lys Tyr Glu Asn Amp Tyr Pro Ser Val Ile Leu Arg Amp Pro Val Gin Thr 650 GAC TCC CTG CAG ACA GAC TAC TTC ACA ATC AAG CAA GAA GTG GTA ACG CCA GAC AAC ATG TGC ATG GGA CGT Asp Ser Leu Gin Thr Asp Tyr Phe Thr Ile Lys Gin Glu Val Val Thr Pro Asp Asn Met Cys Met Gly Arg 750 GCC AGT CGA GGT AAA CTG GGT GGC CAG GAC TCC TTT GAG AGC ATA GAG AGC TAC GAC AGC TGT GAC CGC CTG Ala Ser Arg Gly Lys Leu Gly Gly Gla Asp Ser Phe Glu Ser Ile Glu Ser Tyr Asp Ser Cys Asp Arg Leu Val 850 100 ACA CAG TCC TC' AGC AGC CAG TCC TCC TTC CAG AGC CTG CAG CGC GTC CCC TCC TAC GAT AGC TTT GAC TCA The Gin Ser Trp Cer Ser Gin Ser Ser Phe Gin Ser Leu Gin Arg Val Pro Ser Tyr Asp Ser Pive Asp Ser 900 CAG GAC TAC CCC GCC GCC CTG CCC AAC CAC AAG CCC AAG GGC ACC TTC AAG GAC TAT GTT CGA GAT CGG GCT GNu Asp Tyr Pro Ala Ala Leu Pro Asn His Lys Pro Lys Gly Thr Phe Lys Asp Tyr Val Arg Asp Arg Ala . 950 1000 GAC ATG AAC AAG GAC AAG CCT GTC ATT CCT GCC GCT GCC CTC GGC TAC ACA GGC AGT GGA CCC ATC CAA Asp Met Asn Lys Asp Lys Pro Val lle Pro Ala Ala Ala Leu Ala Gly Tyr Thr Gly Ser Gly Pro llê Gln 1030 CTG TGG CAA TTC CTG CTG GAG CTG CTC ACT GAC AAG TCC TGT CAG TCC TTC ATA AGC TGG ACG GGT GAT GGC Lew Trp Gin Phe Lew Lew Giu Lew Thr Asp Lys Ser Cys Gin Ser Phe lile Hind HI 1100 Hind HI 1100 1150 TGG GAG TTC AAG CTT TCC GAT CCA GAT GAG GTG GCC AGG CGG TGG GGC AAG AGG AAA AAC AAG CCC AAG ATG Top Glu Phe Lys Leu Ser Am Pro Am Glu Val Ala Arg Arg Top Gly Lys Arg Lys Asn Lys Pro Lys Met And TAT GAG AAG CTG AGC CGT GGT CTG CGT TAC TAT TAC GAC AAG AACATC ATC CAC AAG ACG GCC GGC AAG Asm Tyr Glu Lys Leu Ser Arg Gly Leu Arg Tyr Tyr Tyr Asp Lys Asn Ile Ile His Lys Thr Ala Giy Lys Asp

- 50

C-213

ATG AAG GCG GCG GTG GAC CTG AAG CCC ACC CTG ACC ATC AAC AAG ACG GAG AAG GTG GAC ATC GAT CTC TTC Met Lys Ala Ala Val Aap Leu Lys Pro Thr Leu Thr Ile Ile Lys Thr Glu Lys Val Aap Ile Aap Leu Phe

TAC AGC ACA C 100 Meal CCG TCC CCCCATATG GAA TGT GCA GAT GTG CCT TTG TTA ACC CCC AGC AGG GAA ATG ATG TCT CAG GCA Pro Phe Pro Asp Met Glu Cys Ala Asp Val Pro Leu Leu Thr Pro Ser Ser Lys Glu Met Met Ser Gin Ala Tyr Ser Thr Asp 200

CTG AAA GCC ACC TTC AGT GGC TTC GCA AAG GAG CAG CAG CGG CTG GGA ATC CCC AAA GAT CCC CAG CAG TGG Leu Lys Ala The Phe Ser Cly Phe Ala Lys Clu Gin Gin Arg Leu Cly Ile Pro Lys Amp Pro Gin Gin Trp 230 ACA GAG ACG CAC GTG CGG GAC TGG GTG ATG TCG GCA GTG AAT GAG TTC AGC CTG AAG GGA GTG GAT TTC CAG The Glu The His Val Arg Amp Trp Val Met Trp Ala Val Asn Glu Phe Ser Leu Lys Gly Val Amp Phe G

Figure 23 : séquence nucléotidique de l'ADNc p54 c-ets-1 comparée à <u>v-ets</u>





permis de localiser leur présence sous forme d'un petit exon appelé I puisqu'il contient le codon d'initiation de la protéine p54<sup>c-ets-1</sup>, dans le domaine de 40 kbp non apparenté au gène viral. Cette région génomique d'homologie avec l'ADNc est limitée en 3' par un site d'épissure donneur. Par contre, en 5' du codon d'initiation, aucun site d'épissure accepteur n'a été mis en évidence et le cadre de lecture reste ouvert en 5' de cet ATG.

L'ADNC p54 **c-ets-1** permet de détecter dans la cellule normale trois ARN messagers de tailles différentes : 7,5 kb pour l'espèce majeure, 2,2 et 1,5 kb pour les espèces mineures. Actuellement, nous ne savons pas quel ARNm correspond effectivement à l'ADNc p54 **c-ets-1**. La taille de celui-ci (environ 2,0 kbp) laisse supposer qu'il correspond à l'ARN messager **c-ets-1** de 2,2 kb.

Deux hypothèses peuvent être émises quant à la nature et à la localisation des régions promotrices qui dirigent la transcription de ce locus :

- les deux ARNm (7,5 et 2,2 kb) possèdent la même origine de transcription et résultent d'un épissage alternatif d'exons traduits ou non.

les deux ARNm possèdent des régions promotrices différentes.

L'étude des séquences leaders des deux ARNm (7,5 et 2,2 kb) non traduites dans la protéine p54 C-ets-1 permettrait de le vérifier et de localiser dans l'ADN génomique les régions promotrices de ces ARNm. Pour ce faire, deux types d'expériences sont envisagés :

- l'isolement et la séquence d'ADNc de plus grande taille dont les séquences leaders en 5' sont plus longues,
- la localisation de ces séquences leaders dans l'ADN génomique.

Actuellement seule une centaine de nucléotides de la région 5' non traduite est connue. La séquence de son homologue cellulaire dont nous disposons, ne présente pas de signal caractéristique du début de la transcription tel que la séquence TATA. Elle comporte un cadre de lecture ouvert ainsi que des séquences répétitives GGA. De plus, cette région est riche en G et C, ce qui semble caractériser les promoteurs de certains gènes (Dynan **et al.**, 1986). Les exemples actuellement connus appartiennent à la famille des gènes codant pour des enzymes impliqués dans des réactions métaboliques essentielles de la cellule (housekeeping gene).

Ces seules informations ne nous permettent pas de choisir entre les deux hypothèses émises précédemment.

#### CONCLUSION

L'ADNC de l'ARNM p54 C-ets-1 isolé est colinéaire à v-ets, sauf aux extrémités 5' et 3'. L'origine de l'extrémité 5' se trouve dans les 40 kbp d'ADN non homologues à v-ets, alors que celle de l'extrémité 3' est encore inconnue.

Ainsi, l'ARNm codant pour la p54 c-ets-l ne contient pas les exons  $\alpha$  et  $\beta$  situés en amont dans l'ADN cellulaire. Ces exons présentent pourtant des structures exoniques typiques et sont accolés, par un épissage adéquat, à l'exon a dans v-ets. Cela pourrait suggérer l'existence d'un mécanisme d'épissage alternatif qui accolerait tantôt l'exon I, tantôt les exons  $\alpha$  et  $\beta$  à un groupe commun d'exons a à F pour engendrer au moins deux protéines distinctes mais apparentées.

Pour vérifier cette hypothèse, il fallait donc isoler un deuxième type d'ADNc correspondant à un ARNm c-ets-1 qui serait le progéniteur direct de l'oncogène viral v-ets puisque possédant les séquences  $\alpha$  et  $\beta$  en position 5'. C'est ce que nous relatons dans le paragraphe suivant.

## III. <u>ETUDE D'UN ADNC D'UN ARNM c-ets-1 CODANT POUR UNE</u> PROTEINE P68 c-ets-1 (ADNC TYPE II) (ARTICLE 3 EN ANNEXE).

Les séquences  $\alpha$  et  $\beta$  ne sont pas transcrites ou trop faiblement pour être détectées parmi les ARNm des cellules lymphoïdes exprimant la protéine p54C-ets-1 (Gégonne et al., 1987). L'utilisation d'une sonde spécifique de  $\alpha$  et  $\beta$  pour cribler de nombreux clones de deux banques d'ADN complémentaires des ARNm, l'une préparée à partir d'embryon de poulet, l'autre à partir de cellules de rate de poulet, a permis d'isoler deux ADNc de type II (respectivement ADNc 68-1 : 1,5 kbp et 68-2 : 1,4 kbp) d'un même ARNm c-ets-1 possédant les séquences  $\alpha$  et  $\beta$  en position 5' à la place des séquences de l'exon I.



2, 5, 8, 11 : immunoprécipitation par un sérum anti-ets-A inhibé par un excès de

protéines bactériennes immunogènes.

3, 6, 9, 12 : immunoprécipitation par le sérum anti-ets-B spécifique des domaines  $\alpha$  et  $^{B}$ 

Figure 25: expression des ADNc p68:1 et 2 dans les cellules COS-1 exprimées dans les cellules lymphoïdes T et la rate comparaison avec les protéines normales <u>c-ets-1</u>

de poulet sain

#### 1) Expression dans un système eucaryote

Ces ADNc ont été exprimés dans des cellules COS-1 après clonage dans un vecteur d'expression pSVL, comme il a été décrit précédemment pour l'ADNc p54<sup>c</sup>-ets-1. Ils dirigent la synthèse d'une protéine de 68 kd spécifiquement immunoprécipitée par deux séra anti-ets (Figure 25), le sérum anti-etsA (Ghysdael et al., 1986a ; Gégonne et al., 1987) qui reconnaît les séquences communes à la protéine virale P135gag-myb-ets et à la protéine p54 c-ets-1, le sérum anti-etsB (Gégonne et al., 1987) spécifique de la région codée par  $\alpha$  et  $\beta$ .

La protéine déduite de la séquence nucléotidique de ces ADNc de type II comporte 485 acides aminés, soit une masse moléculaire de 53 kd, alors que celle observée en gel d'acrylamide -SDS est de 68 kd. Cette différence entre les masses moléculaires prédite et observée peut s'expliquer en partie par la richesse en proline de l'extrémité amino-terminale de la protéine p68.

Cette nouvelle protéine p68 c-ets-1 est exprimée à un taux relativement faible comparativement à p54 c-ets-1 dans les cellules de rate de poulet, par contre on ne la détecte pas dans les thymocytes ou des lignées de cellules lymphoïdes. Afin de confirmer la parfaite correspondance entre la protéine exprimée dans les cellules COS-1 à partir de l'ADNc de type II et la protéine p68 exprimée dans les cellules de rate de poulet, une carte peptidique de ces protéines par hydrolyse partielle par la protéase V8 a été établie (Figure 26) (Cleveland et al., 1977). Les cartes peptidiques obtenues dans les deux cas sont les mêmes, les deux protéines sont donc identiques.

#### 2) Analyse de la séquence nucléotidique

Les ADNc 68-1 et 2 présentent un large cadre de lecture ouvert colinéaire à **v-ets** sauf dans la région 3' où les séquences divergent. La comparaison de **v-ets** et de ces ADNc révèle la présence de sept mutations ponctuelles dont cinq sont conservatives (Figure 27 et 28).

Il est difficile d'identifier avec certitude le codon d'initiation de p68 **c-ets-1** parmi les cinq ATG présents dans la région 5'. Cependant, deux d'entre eux (situés aux nucléotides 85 et 175) ne sont probablement pas utilisés car ils se situent dans un contexte peu favorable (Kozak, 1986). Les trois autres codons d'initiation se trouvent dans un bon contexte et sont tous susceptibles d'être



Figure 26: action de la V8 protéase de Staphyloccus aureus sur les protéines p68<sup>c-ets-1</sup>extraites de cellules de rate de poulet sain (A) ou exprimées dans les cellules COS-1(B)

TACAAGTGTGGGGAGCCGTGGAGGACATAAGCTTTTTCCTTGGTCTTCCAGATAAGTAGCACCTCAGAG v-ets 159 c-ets Met Met Ser Tyr Tyr Met Asp Thr Thr Ile Gly Ser Thr Gly Pro Tyr Pro Leu Ala Arg Pro Gly Val Met Gln Gly Ala Ser Ser Cys v-ets TOT GAG GAC CCC TGG ANG CCA TGC AGG CTG CAG TCT GCC TGC TGC CCG CCC AGG TCG TGT TGC CCG TGG GAT GAG GCG GCC ATC CAG Cys Glu Asp Pro Trp Met Pro Cys Arg Leu Gin Ser Ala Cys Cys Pro Pro Arg Ser Cys Cys Pro Pro Trp Asp Glu Ala Ala Ile Gin 40 60 GAA GTT CCC ACT GGC CTG GAG CAC TAC AGE ACA GAT ATG GAA TGT GCA GAT GTG CCT TTG TTA ACC CCC AGE AGE AAG GAA ATG ATG TCT 339 Glu Val Pro Thr Gly Leu Glu His Tyr Ser Trp Asp Met Glu Cys Ala Asp Val Pro Leu Leu Thr Pro Ser Ser Lys Glu Met Met Ser Ast AG GCA CTG AAA GCC ACC TTC AGT GGC TTC GCA AAG GAG CAG CAG CGG CTG GGA ATC CCC AAA GAT CCC CAG CAG TGG ACA GAG ACG CAC 429 Gin Ala Leu Lys Ala Thr Phe Ser Gly Phe Ala Lys Glu Gin Gin Arg Leu Gly Ile Pro Lys Asp Pro Gin Gin Trp Thr Glu Thr His 100 110 120 GTG CGG GAC TGG GTG ATG TGG GCA GTG AAT GAG TTC AGC CTG ANG GGA GTG GAT TTC CAG AAG TTC TGC ATG AAC GGA GCT GCC CTG TGC 519 Val Arg Asp Trp Val Met Trp Ala Val Asn Glu Phe Ser Leu Lys Gly Val Asp Phe Gln Lys Phe Cys Met Asn Gly Ala Ala Leu Cys 140 130 150 GCC CTG GGC AAG GAG TGC TTC CTG GAG CTA GCG CCT GAC TTT GTG GGA GAT ATC CTT TGG GAA CAC CTG GAG ATC TTG CAG AAA GAA GAA 609 Ala Leu Gly Lys Glu Cys Phe Leu Glu Leu Ala Pro Asp Phe Val Gly Asp Ile Leu Trp Glu His Leu Glu Ile Leu Gln Lys Glu Glu Glu 160 170 180 GCA AAA CCA TAC CCA GCA AAC GGA GTG AAT GCA GCG TAT CCA GAA TCC CGC TAT ACT TCA GAC TAC TTC ATT AGT TAT GGC ATC GAG CAC Ala Lys Pro Tyr Pro Ala Asn Gly Val Asn Ala Ala Tyr Pro Glu Ser Arg Tyr Thr Ser Asp Tyr Phe Ile Ser Tyr Gly Ile Glu His 699 Asn 190 200 210 GCA CAG TGC GTG CCT CCC TCT GAG TTC TCT GAG CCC AGC TTC ATC ACA GAG TCC TAC CAG ACC CTC CAT CAC AGC TCG GAA GAG CTT Ser Glu Phe Ser Glu Pro Ser Phe Ile Thr Glu Ser Tyr Gln Thr Leu His Pro Ile Ser Ser Glu Glu Leu Ala Gin Cys Val Pro Pro 220 230 Ser 240 CTG TCC CTC AAG TAC GAG AAC GAC TAT CCC TCA GTC ATC CTT CGT GAC CCC GTC CAG ACG GAC TCC CTG CAG ACA GAC TAC TTC ACA ATC 879 Leu Ser Leu Lys Tyr Glu Asn Asp Tyr Pro Ser Val Ile Leu Arg Asp Pro Val Gln Thr Asp Ser Leu Gln Thr Asp Tyr Phe Thr Ile 260 270 ANG CAN GAN GTG GTA ACG CCA GAC ANC ATG TGC ATG GGA CGT GCC AGT CGA GGT ANA CTG GGT GGC CAG GAC TCC TTT GAG AGC ATA GAG 969 Lys Gin Giu Val Val Thr Pro Asp Asn Met Cys Met Gly Arg Ala Ser Arg Gly Lys Leu Gly Gly Gln Asp Ser Phe Glu Ser Ile Glu 280 Val 290 300 AGE THE GAE AGE TGT GAE CGE CTG ACA CAG TEE TGG AGE AGE CAG TEE TTE CAG AGE CTG CAG CGE GTE CEE TEE TAE GAT AGE TTT 1059 Ser Tyr Asp Ser Cys Asp Arg Leu Thr Gin Ser Trp Ser Ser Gin Ser Ser Phe Gin Ser Leu Gin Arg Val Pro Ser Tyr Asp Ser Phe 310 120 330 GAC TEA CAG GAC TAC CEC GET GEC CTG CEC AND CAC ANG CEC ANG GEC ACC TTE ANG GAC TAT GTT CGA GAT CGG GET GAC ATG AAC ANG 1149 Asp Ser Glu Asp Tyr Pro Ala Ala Leu Pro Asn His Lys Pro Lys Gly Thr Phe Lys Asp Tyr Val Arg Asp Arg Ala Asp Met Asn Lys AL a 340 350 360 GAD ANG CET GTE ATT CET GEE GEE GEE CTE GEE GGE TAC ACA GGE AGT GGA CEE ATE CAA CTE TGE CAA TTE CTE GTE GAG CTE ACT ACT 1239 Asp Lys Pro Val Ile Pro Ala Ala Ala Leu Ala Gly Tyr Thr Gly Ser Gly Pro Ile Gln Leu Trp Gln Phe Leu Leu Glu Leu Thr 370 380 390 GAC AAG TCC TGT CAG TCC FTC ATC AGC TGG ACG GGT GAT GGC TGG GNG TTC AAG CTT TCC GAT CCA GAT GAG GTG GCC AGG CGG TGG GGC 1329 Asp Lys Ser Cys Gin Ser Phe Ile Ser Trp Thr Gly Asp Gly Trp Glu Phe Lys Leu Ser Asp Pro Asp Glu Val Ala Arg Arg Trp Gly 410 420 400 ANG AGG ANA ANC ANG CCC ANG ANG ANG TAT GAG ANG CTG AGC CGT GGT CTG CGT TAC TAT TAC GAC ANG ANC ATC ATC CAC ANG ACG GCC 1419 Lys Arg Lys Asn Lys Pro Lys Met Asn Tyr Glu Lys Leu Ser Arg Gly Leu Arg Tyr Tyr Tyr Asp Lys Asn Ile His Lys Thr Ala Asp 430 Val 450 CAC TCA TCA GCA TCT GGC TTG ACG GGC AND CGC TAC GTC TAC CGC TTC GTC TGC GAC CTG CAG AGC CTG CTG GGC TAC ACA CCA GAG GAG CTG CAC GCC ATG CTG GAC GTC AAG 1509 Gly Lys Arg Tyr Val Tyr Arg Phe Val Cys Asp Leu Gln Ser Leu Leu Gly Tyr Thr Pro Glu Glu Leu His Ala Met Leu Asp Val Lys His Ser Ser Ala Ser Gly Leu Thr 460 470 TCC AGC ATG GCG TGC AGC TCC TTT TGA CCA GAT GCT GAT GAG TGA Atte 1527 Pro Asp Ala Asp Glu \*\* Ser Ser Met Ala Cys Ser Ser Phe \*\*\*

Figure 27: séquence nucléotidique de l'ADNc p68 c-ets-1 comparée à v-ets







إيدا

]



Figure 29 : détection des ARNm <u>c-ets-1</u> dans les cellules de thymus, de lignées lymphoïdes et de rate de poulet sain: A-utilisation d'une sonde spécifique de l'ADNc p68<sup>c-ets-1</sup> B-sonde commune aux ADNc p54 et p68 c-ets-1

a márion a da motation d'ADNm (dátai) das sandes utilisáes et das tra

utilisés alternativement pour engendrer des protéines différentes par leur extrémité N-terminale comme cela a déjà été observé par Strubin **et al.,** 1986, pour un antigène (Ia) d'histocompatibilité. On observe d'ailleurs dans les cellules COS-1 une protéine de 60 kd qui pourrait résulter d'un tel mécanisme.

L'analyse des séquences nucléotidiques des ADNc correspondant aux protéines p54 **C-ets-1** et p68 **C-ets-1** a montré que l'homologie entre les deux ADNc débutait avec l'exon **a**, au niveau du site d'épissage accepteur de celui-ci. En amont de l'exon a se trouve tantôt l'exon I tantôt les exons  $\alpha$  et  $\beta$ , le point de jonction avec l'exon **a** correspond pour chacun d'eux à un site d'épissage donneur.

Ces résultats prouvent l'existence d'un mécanisme d'épissage alternatif à l'intérieur du locus **c-ets-1** poulet, épissage alternatif donnant naissance à deux protéines distinctes mais apparentées : p54 **c-ets-1** exprimée à taux élevé dans les organes lymphoïdes et p68 **c-ets-1** exprimée à taux plus faible dans les cellules de rate de poulet.

L'utilisation d'une sonde spécifique de l'ADNc p68 C-ets-1 a permis de détecter dans les cellules de rate de poulet et dans les embryons de poulet un ARN messager c-ets-1 de 7,5 kb. La même sonde ne le détecte pas dans le thymus ni dans les cellules lymphoïdes, ce qui confirme nos observations précédentes (Article 1). Les ARNm c-ets-1 de 2,2 et 1,5 kb ne sont pas détectés non plus.

Curieusement, la taille de l'ARNm détecté par l'ADNc p68 **c-ets-1** est de 7,5 kb, taille de l'espèce majeure des ARNm **c-ets-1**. Ces résultats suggèrent l'existence d'un doublet d'ARNm de 7,5 kb correspondant aux ADNc p54 et p68 **c-ets-1** et sont confirmés par des expériences de protection d'ARN (Figure 29) (Kruijer **et al.**, 1984).

Un fragment de l'ADNc p68-2, (EcoRI/Bgl II de 566 bp), correspondant aux exons  $\alpha$ ,  $\beta$ , **a** et **b** a été cloné puis répliqué par la polymerase T7 en un ARN complémentaire. Ce dernier est hybridé aux ARNm de différentes types cellulaires (cellules de thymus, cellules lymphoïdes et cellules de rate de poulet). L'hydrolyse obtenue par la RNase permet de visualiser la taille des hybrides ARNm/ARNc. Elle est de 305 bp (exon **a** et **b**) pour le fragment commun aux deux ADNc p54 **C-ets-1** et p68 **C-ets-1**, cet hybride est retrouvé dans tous les types cellulaires où sont exprimés la protéine p54 **C-ets-1** et/ou p68 **C-ets-1**. Le second hybride, de 566 bp, correspond aux

48

ADNc p68 18 TACAAGTGTGGGGGAGCCG 68-2 tctgcag<u>GC</u>ACC\*\*\*\*\* exon α TGGAGGACATAAGCTTTTTCCTTGGTCTTCCAGATAAGTAGCACCTCAGAG 69 ADNc p68 E26 exon a ACG: 108 ADNc p68 ATG GAC ACA ACC ATT GGC AGC ATG E26 d exon a GTG ATG aga gta ADNc p68 ATG CAA GGT 147 GCT CGC CCT GGA GTG GGT CCT TAT CCT TTG E26 \*\*\* \*\*\* \*\*\* \*\*\* exon ß CAA GGT cag aag exon  $\beta$ AGG 186 ADNc P68 ATG TGC AGC TGC TGT GAG GAC CCC TGG CCA GCT AGC E26 \*\*\* Pst I exon ß \*\*\* ADNc P68 CCC CCG 225 CTG CAG GCC TGC TGC CCG ACC TCG TGT TGC TCT ADNc p54 ATG AA\* GCG GC\* GT\* GAC CTG AA\* GCT \*C\* G\*C AC\* T\*A E26 exon a GGC 264 GCC ATC CAG GAA GTT CCC ACT ADNc p68 CCG TGG GAT GAG GCG ADNc p54 ATC ATC \*AG AC\* \*\*G AAG GTG GAC AT\* ACC CTG ACC E26 \*\*\* d exon a \*\*\* AGC ACA Ggt agg g.. 303 TAC AGC ACA GAT ADNc p68 CTG GAG CAC ATG GAA TGT GCA GAT GTG C\*C ADNc p54 CCG TC\* GAT CTC TT\* E26 \*\*\* \*\*C \*\*\* GAT ATG \*\*\* exon a . . . ttt tca Hpa Т **T**. AC 314 ADNc p68 TTG TTA CCT ADNc p54 E26 exon a

Figure 30: extrémités 5' des ADNc p54 et p $68^{c-ets-1}$ comparées aux exons  $\alpha$ ,  $\beta$ , a et à la jonction myb-ets dans E26 exons  $\alpha$ ,  $\beta$ , a et b où  $\alpha$  et $\beta$  sont spécifiques de l'ADNc p68 **c-ets-1**. Cet hybride est retrouvé uniquement dans les cellules de rate de poulet où est exprimé la protéine p68 **c-ets-1**. Il existe donc deux ARNm de 7,5 kb, distincts, transcrits à partir d'une même unité transcriptionnelle, codant pour deux protéines distinctes mais hautement apparentées.

#### IV. TRANSDUCTION DE c-ets-1

Les séquences de **v-ets** et de l'ADNc de l'ARNm codant pour p68 **c-ets-1** sont très homologues sauf dans leurs régions 3'.

Cet ARNm p68 **c-ets-1** peut être considéré comme le progéniteur direct de l'oncogène viral **v-ets**. La comparaison des extrémités 5' de **v-ets** et de l'exon  $\alpha$  montre que la séquence nucléotidique de l'ADNc p68 **c-ets-1** est colinéaire à **v-ets** à partir du nucléotide 59 qui correspond à la région 5' de l'exon  $\alpha$  limitée par un site d'épissure accepteur d'une part et à la jonction **myb-ets** dans le génome viral d'autre part. Cette observation appelle deux commentaires :

. Il existe au moins 11 nucléotides en amont de l'ATG en position 70, non codants dans l'ADNc p68 C-ets-1 qui sont convertis en séquences codantes dans v-ets (Figure 30).

. L'observation que la jonction myb-ets correspond exactement au site d'épissure accepteur de l'exon  $\alpha$  ne peut être fortuite et nous avons donc comparé la séquence myb-ets aux séquences génomiques c-myb et c-ets en cherchant un site d'épissure donneur potentiel dans c-myb. Il apparaît que le point de recombinaison dans c-myb se situe au milieu de l'exon 6 (E6) et présente une homologie certaine (6 nucléotides sur 10) avec une séquence consensus donneur d'épissure (CAAGGgtagagt) (Figure 31) (Kozak, 1986).

Un mécanisme d'épissage aberrant entre un site d'épissure donneur cryptique de l'exon 6 de **c-myb** et le site d'épissure accepteur normal de l'exon  $\alpha$  de **c-ets-1** a donc été à l'origine de la transduction de ces deux gènes dans le rétrovirus E26. Des mécanismes d'épissage aberrants ont déjà été décrits pour l'activation de **c-myb** dans certains lymphosarcomes impliquant le virus d'Abelson (Rosson **et al.**, 1987) ou pour l'activation de **c-erbB** dans une érythroblastose du poulet ALV-induite (Nilsen **et al.**, 1985).

Trois modèles de recombinaison peuvent être proposés :

un épissage illégitime entre deux ARN messagers d'origines



sd: site d'épissure donneur

В

jonction myb/ets

ADNc p68	• • •	•••	TA	<u>AGT</u>	<u>AG</u> C	ACC	TCA	• • •	• • •
<u>c-ets-1</u> exon α	• • •	•••	tct	gca	gGC	ACC	TCA	• • •	• • •
E26	• • •	• • •	GAT	CAT	GGC	ACC	TCA	• • •	• • •
<u>c-myb</u> exon E6	•••	••••	GAT	сат 	GGT 	ŢGC	TTA	6	• • •
séquence consensus d'un site d'épissure donneur				C <sub>AG</sub> A	Ggt	ag g	t	.10	

Figure 31 : A—schéma général du mécanisme d'épissage à l'origine de la jonction myb—ets B—séquences nucléotidiques des sites d'épissure impliqués dans ce mécanisme



Figure 32 : A-représentation schématique de l'activation de <u>c-erbB</u> par un ALV B-comparaison entre la protéine synthétisée, la protéine <u>v-erbB</u>, et le récepteur de l'EGF différentes conduisant à la formation d'un ARN messager hybride composé d'exons appartenant à deux gènes différents (Konarska **et al.,** 1985; Solnick, 1985).

- une translocation réunissant c-myb et c-ets et un ADN proviral présent dans le génome cellulaire.
- l'intégration d'un virus transportant déjà l'un des deux oncogènes au voisinage du second oncogène cellulaire.

Bien que nous n'ayons pas d'évidence expérimentale étayant ce troisième modèle, il semble plus plausible que les deux autres et a déjà été exploré dans le cas des virus contenant un oncogène **c-erbB** tronqué (Miles **et al.**, 1985; Nilsen **et al.**, 1985). En effet, l'étude d'une érythroblastose aviaire induite par un ALV a montré que le provirus s'intégrait à l'intérieur du locus **c-erbB**. L'analyse des ARNm **c-erbB** transcrits dans les cellules de moëlle osseuse leucémiques, a mis en évidence deux ARNm (7 kb et 3,6 kb) contenant des séquences **c-erbB**. L'un d'entre eux a été plus particulièrement étudié et montre que la transcription s'initie dans le LTR en 5' du provirus, le parcourt entièrement et se termine dans le gène **c-erbB** (Figure 32). L'épissage de ce transcrit primaire conserve une partie des gènes rétroviraux  $\Delta$  **gag et env** et y associe les exons de **c-erbB** correspondant à ceux retrouvés dans **v-erbB**.

Dans le cas de E26, on peut imaginer qu'un rétrovirus contenant l'un des deux oncogènes **v-myb** ou **v-ets** s'est intégré au voisinage (en amont ou en aval) du second oncogène cellulaire. De la même manière que précédemment, un ARN mixte a pu être transcrit puis épissé pour engendrer l'ARN E26.

Pour vérifier cette hypothèse, il faudrait analyser l'isolat primaire du rétrovirus E26 qui contiendrait éventuellement le virus dont le génome ne possède encore qu'un oncogène (**v-myb** ou **v-ets**). Malheureusement, cet isolat primaire n'est plus disponible. Le mécanisme à l'origine de E26 restera hypothétique.

### CONCLUSION

Il existe CHEZ LE POULET au moins deux loci apparentés à v-ets.

- Un locus hautement apparenté au gène viral que nous appellons Ck-c-ets-1, en raison de son homologie avec ETS-1 étudié par J.P. Kerckaert, (communication personnelle), et des travaux publiés par le groupe de J. Ghysdael sur le locus Ck-c-ets-2 exprimé dans les macrophages. Les séquences Ck-c-ets-1 se répartissent en deux domaines séparés par 40 kbp d'ADN non homologues à v-ets. Ce locus est transcrit en un doublet d'ARN messager de 7,5 kb, et en deux ARN mineurs de 2,2 kb et 1,5 kb, qui codent pour deux protéines distinctes mais apparentées, p54 c-ets-1 exprimée à un taux élevé dans les organes lymphoïdes de poulet et p68 c-ets-1 exprimée à taux plus faible dans la rate de poulet.

L'étude de la localisation subcellulaire de ces deux protéines p54 c-ets-l (P. Pognonec, communication personnelle) et p68 c-ets-l (D. Leprince, communication personnelle) a montré qu'elles étaient nucléaires. Cette localisation peut se justifier par plusieurs observations :

- la richesse en proline de l'extrémité N-terminale de chacune de ces protéines (11 % pour la région protéique correspondant à l'exon I dans la protéine p54 c-ets-1, 14 % pour celle qui correspond aux exons  $\alpha$  et  $\beta$  de la protéine p68 c-ets-1).

- l'existence d'un domaine protéique, constitué d'acides aminés basiques (résidus 420-426 de la protéine p68 **c-ets-1**), de composition voisine de la région protéique responsable de la localisation nucléaire de l'antigène T du virus SV40 (Dingwall et Laskey, 1986; Fujiwara **et al.**, 1988).

Afin de préciser le rôle de ces deux protéines nucléaires, il conviendrait dans un premier temps de connaître leur capacité à se lier à l'ADN ou à former un complexe avec une autre protéine cellulaire comme c'est le cas pour les protéines virales et cellulaires **v-fos** et **c-fos** (Curran **et al.**, 1985).

La synthèse de ces deux protéines résulte d'un mécanisme d'épissage alternatif au sein du locus **Ck-c-ets-1**. Comme nous l'avons décrit au Chapitre II, ce mécanisme est observé chez de nombreux gènes viraux ou eucaryotes qui synthétisent alors plusieurs protéines distinctes à partir d'un même locus. Ce mécanisme confère souvent une spécificité fonctionnelle ou une localisation particulière aux différentes protéines produites. En ce qui concerne le gène **Ck-c-ets-1**, l'épissage accole tantôt les exons  $\alpha$  et  $\beta$ , correspondant à l'extrémité 5' de **v-ets**, tantôt l'exon I (localisé dans les 40 kb d'ADN cellulaire qui séparent les séquences homologues à **v-ets**) à un groupe commun d'exons **a-F**.

Bien que la fonction des protéines c-ets-1 soit encore inconnue, il est tentant de spéculer sur le fait que les séquences communes (a-F) pourraient jouer un rôle dans la fonction même des protéines c-ets-1 alors que les exons alternatifs ( $\alpha$ ,  $\beta$  et I) pourraient diriger ou réguler cette fonction.

La protéine p54 C-ets-l est exprimée dans de nombreux tissus alors que la protéine p68 C-ets-l n'a été détectée, pour l'instant, que dans la rate de poulet. Afin de préciser la nature des cellules qui expriment p68 C-ets-l, des expériences d'hybridation "in situ" et l'analyse de populations cellulaires purifiées à partir de la rate sont envisagées, elles permettront de définir une éventuelle spécificité tissulaire de la protéine p68 C-ets-l.

La comparaison des séquences non codantes des ARNm p54 **c-ets-1** et p68 **c-ets-1** et la localisation de leur équivalent cellulaire permettraient une meilleure compréhension de l'expression du gène **Ck-c-ets-1**. Il conviendrait de connaître la nature et le nombre de promoteurs (un ou deux) de ce gène, de les comparer aux exemples connus. En effet, on connaît actuellement des gènes, comme celui codant pour une déaminase humaine, (Chrétien **et al.**, 1988), qui codent pour deux protéines isoformes dont l'expression est dirigée par deux promoteurs différents. L'un d'entre eux présente une spécificité tissulaire, l'autre qualifié de "housekeeping" est actif dans toutes les cellules (Dynan **et al.**, 1986).

L'étude des ADNc p54 **c-ets-1** et p68 **c-ets-1** et de leur expression a permis de préciser les différences qui existent entre le gène cellulaire **Ck-c-ets-1** et l'oncogène **v-ets.** Ces différences peuvent être à l'origine de l'activation du gène viral, et sont au nombre de trois :

- la synthèse d'une protéine virale de fusion la P135 gag-myb-ets.

- l'introduction de mutations ponctuelles dans v-ets.

- l'amputation des séquences C-terminales, remplacées dans le virus par une région d'origine différente.

Le rétrovirus E26 est le seul exemple de virus à deux oncogènes traduits sous forme d'une protéine transformante unique. On peut imaginer que la fusion de ces protéines, toutes deux nucléaires, codées par **v-myb** et **v-ets** 

renforce et conjugue leur effet.

L'activation par mutation ponctuelle semble peu probable car les mutations entre l'ADNc p68 **c-ets-1**, considéré comme progéniteur du gène viral, et **v-ets** sont pour la plupart conservatives.

Par contre, l'activation par modification de l'extrémité C-terminale de la protéine semble plus plausible. De tels exemples ont été décrits pour des oncogènes appartenant à la famille des protéines à activité tyrosine-kinase (src, Cooper et al., 1986 ; Takeya et al., 1983 ; fms, Coussens et al., 1986 ; ros, Neckamayer et al., 1986) ou pour des oncogènes dont les protéines sont nucléaires (v-fos et FOS, Van Beveren et al., 1983). Cette différence C-terminale a pu être, dans le cas de src, directement reliée à l'activation de cet oncogène (Cooper et al., 1986).

La construction de mutants E26 dans lesquels la région 3' de v-ets serait remplacée par l'exon F du locus Ck-c-ets-1 permettrait de vérifier si cette différence C-terminale est seule responsable de l'oncogénicité de E26. Si c'est le cas, le rôle précis de l'extrémité C-terminale des protéines c-ets-1 sera à définir (activité enzymatique, capacité de se lier à l'ADN, ou autre signal peptidique).

Chez le poulet existe également :

- un locus de parenté plus lointaine à v-ets, Ck-c-ets-2 (Ghysdael et al., 1986; Boulukos et al., 1988) exprimé dans les macrophages normaux et transformés, sous forme de protéines p58, 60, 62, 64 C-ets-2, localisées dans le noyau des cellules.

L'homologie de p58-p64 avec les autres protéines c-ets-1 (p54 c-ets-1 et p68 c-ets-1) et la partie v-ets de P135 gag-myb-ets est limitée à deux régions de 96 et 175 amino-acides situées respectivement du côté N- et Cterminal de ces molécules.

Ce locus est transcrit dans la plupart des tissus du poulet, en un ARNm de 4 kb.

- Les travaux récents de Pognonec et al., 1988 ont apporté des éclaircissements quant à la modification post-traductionnelle des protéines c-ets-1 et 2 ainsi que leur rôle physiologique éventuel. L'hétérogénéité de ces protéines en gel d'acrylamide-SDS est dûe à leur phosphorylation sur des résidus de sérine ou, dans une proportion moindre, de thréonine. Ces

phosphorylations peuvent être spécifiquement et rapidement stimulées par traitement par l'ionophore de calcium A23187 ou abolies en abaissant la concentration calcium extra-cellulaire. modifications posten Ces traductionnelles sont conservées pour les différents membres (c-ets-1 et c-ets-2) d'une même famille de gènes. Pognonec et al., ont observé que dans les thymocytes où l'expression de c-ets-1 est élevée comparativement aux autres types cellulaires, la phosphorylation de la protéine p54 c-ets-l peut être induite par stimulation par la concanavaline A et dépend d'une façon stricte de la concentration en Ca<sup>2+</sup> extracellulaire. L'étude de l'expression des protéines c-ets-1 chez la souris, Mo-p63 c-ets-1, a montré que la maturation des thymocytes ne s'accompagne pas de la synthèse de Mo-p63 c-ets-l. Par contre, le taux d'expression de cette protéine augmente dans une population de thymocytes enrichie en cellules immatures. De plus, la stimulation de la division cellulaire des lymphocytes T s'accompagne d'une augmentation du taux d'ARN messagers et des protéines c-ets-1. Ces résultats suggèrent que le produit de c-ets-1 pourrait jouer un rôle dans les évènements cellulaires précoces liés à l'activation des cellules T.

Le fait que les produits des différents gènes d'une même famille soient modifiés par phosphorylation en fonction de la concentration en  $Ca^{2+}$ indique le rôle important joué par ces modifications post-traductionnelles dans la fonction de ces protéines. Ces modifications s'accompagnent de changements de mobilité électrophorétique très significatifs et laissent supposer leur rôle important dans la modification conformationnelle de ces protéines. Cela pourrait intervenir dans la modulation de la fonction des protéines **c-ets-1** et **cets-2** ou d'autres membres de la famille **c-ets.** 

CHEZ L'HOMME, existent également deux loci apparentés à v-ets ; ils sont localisés sur deux chromosomes différents :

- ETS-1, sur le chromosome 11q23-24, transcrit en un ARNm de 6,8 kb codant pour une protéine de 51 kd, p51 ETS-1 de localisation cytoplasmique (Fujiwara et al., 1988).

- ETS-2, sur le chromosome 21 q22-1-22-3, transcrit en trois ARNm de 4,2 ; 3,2 ; 2,7 kb. Une seule protéine a été mise en évidence, la p56 ETS-2, sa localisation est nucléaire (Fujiwara et al., 1988).

55

Cette localisation différente des protéines ETS-1 et ETS-2 suggère des fonctions différentes pour chacune d'elles.

Récemment, un nouveau membre de la famille des oncogènes ETS a été identifié dans une lignée cellulaire humaine COLO 320, il s'agit du gène ERG très apparenté à ETS-2 et localisé sur le chromosome 21 (Reddy et al., 1987; Rao et al., 1987).

Deux transcrits ERG-1 et ERG-2 en dérivent par épissage alternatif. ERG-2 diffère de ERG-1 par un mécanisme d'épissage qui entraîne un changement de cadre de lecture dans la région 5', provoquant l'insertion de 99 amino-acides dans la partie N-terminale de la protéine. L'expression in vitro des ADNc des ARNM ERG-1 et ERG-2 donne naissance à deux polypeptides de 41 et 52 kd respectivement.

Hu-ETS-1 et Hu-ETS-2 sont impliqués dans des translocations chromosomiques spécifiques de certains types de leucémies aiguës (Sacchi et al., 1986).

Dans notre Institut, le groupe de J.P. Kerckaert s'intéresse plus particulièrement à un cas de leucémie aiguë monocytaire avec translocation réciproque t(9,11) dans lequel l'oncogène **ETS-1** pourrait être impliqué. Cet oncogène a été cloné sur environ 120 kbp. Pour l'instant un site de polymorphisme de restriction a été mis en évidence. Il est situé dans la partie 3' du gène et correspond à la disparition d'un site de restriction XbaI par simple mutation T (normal)  $\rightarrow$  C (leucémique). L'analyse détaillée de la séquence environnante suggère la possibilité d'un mécanisme d'épissage alternatif dans cette région qui représente une séquence consensus accepteur d'épissage. Si une étude statistique plus large apportait une certitude quant au rôle de cette mutation dans la pathogénèse et venait étayer l'hypothèse d'un épissage alternatif dans ce locus **ETS-1**, une remarquable filiation entre nos travaux de recherche fondamentale et la compréhension d'une pathologie humaine pourrait être établie.

## REFERENCES BIBLIOGRAPHIQUES

- ADAMS, J.M., GERONDAKIS, S., WEBB, E., CORCORAN, L.M. & CORY, S. (1983). Cellular myb oncogene is altered by chromosome translocation in an immunoglobulin locus in murine plasma-cytomas and is rearranged similarly in human Burkitt lymphomas. Proc. Natl. Acad. Sci. USA, 80: 1982-1986.
- ADKINS, B., LEUTZ, A. & GRAF, T. (1984). Autocrine growth induced by src-related oncogenes in transformed chicken myeloid cells. Cell, <u>39</u>: 439-445.
- ALITALO, K., SCHWAB, M., LIN, C.C., VARMUS, H.E. & BISHOP, J.M. (1983). Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. Proc. Natl. Acad. Sci. USA, <u>80</u>: 1707-1711.
- ALT, F.W., BOTHWELL, A.L.M., KNAPP, M., SIDEN, E., MATHER, E., KOSHLAND, M. & BALTIMORE, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin our heavy chains is directed by mRNAs that differ at their 3' ends. Cell, 20: 293-301.
- AMARA, S.G., JONAS, V., ROSENFELD, M.G., ONG, E.S. & EVANS, R.M. (1982). Alternative RNA processing in Calcitonin gene expression generates mRNAs encoding different polypeptides products. Nature, <u>298</u>: 240-244.
- AMARA, S.G., EVANS, R.M. & ROSENFELD, M.G. (1984). Calcitonin/ calcitonin Gene-related peptide transcription unit : tissue-specific expression involves selective use of alternative polyadenylation sites. **Mol. Cell. Biol.**, <u>4</u>: 2151-2160.
- ANDERSON, S.M., HAYWARD, W.S., NEIL, B.G. & HANAFUSA, H. (1980).
  Avian erythroblastosis virus produces two mRNAs.
  J. Virol., <u>36</u>: 676-683.
- ANGEL, P., ALLEGRETTO, E.A., OKINO, S.T., HATTORI, K., BOYLE, W.J., HUNTER, T., KARIN, M. (1988). Oncogene jun encodes a sequence-specific trans-activator similar to AP-1. Nature, <u>332</u>: 166-171.
- ANTONIADES, H.N. (1981). Human platelet-derived growth factor (PDGF): purification of PDGF-I and PDGF-II and separation of their reduced subunits. **Proc. Natl. Acad. Sci. USA**, 78:7314-7317.
- BALTIMORE, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature, <u>226</u>: 1209-1211.

- BARGMANN, C.I., HUNG, M.C. & WEINBERG, R.A. (1986a). Multiple independant activations of the **neu** oncogene by a point mutation altering the transmembrane domain of p185. **Cell**, 45: 649-657.
- BARGMANN, C.I., HUNG, M.C. & WEINBERG, R.A. (1986b). The neu oncogene encodes an epidermal growth factor receptor-related protein. Nature, <u>319</u>: 226-230.
- BARRE-SINOUSSI, F., CHERMANN, J.C., REY, F., NUGEYRE, M.T., CHAMARET, S., GRUEST, J., DANGUET, C., AXLER-BLIN, C., VEZINET-BRUN, F., ROUZIOUX, C., ROZENBAUM, W. & MONTAGNIER, L. (1983). Isolation of T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science, 220: 868-870.
- BECHADE, C., CALOTHY, G., PESSAC, B., MARTIN, P., COLL, J., DENHEZ, F., SAULE, S., GHYSDAEL, J. & STEHELIN, D. (1985). Induction of proliferation or transformation of neuroretina cells by the mil and myc viral oncogenes. Nature, 316: 559-562.
- BECKNER, S.K., HATTORI, S. & SHIH, T.Y. (1985). The ras oncogene product p21 is not a regulatory component of adenylate cyclase. Nature, 317:71-75.
- BENDER, W. & DAVIDSON, N. (1976). Mapping of polyA sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. Cell, 7: 595-607.
- BEN-NERIAH, Y., BERNARDS, A., PASKINA, M., DALEY, G.Q. & BALTIMORE, D. (1986). Alternative 5' exons in c-abl mRNA. Cell, <u>44</u>: 577-586.
- BERNHARD, W. (1958). Electron microscopy of tumor cells and tumor viruses. A review. Cancer Res., 20:712-727.
- BERRIDGE, M.J. & IRVINE, R.F. (1984). Inositol trisphosphate, a novel second messenger in cellular signal transduction. Nature, <u>312</u>: 315-321.
- BESMER, P., MURPHY, J.E., GEORGE, P.C., QIU, F., BERGOLD, P.J., LEDERMAN, L., SNYDER, H.W. Jr., BRODEUR, D., ZUCHERMAN, E.E.Z. & HARDY, W.D. (1986). A new acute transforming feline retrovirus and relationship of its oncogene v-kit with the protein kinase gene family. Nature, <u>320</u>, 415-421.

- BETSHOLTZ, C., WETERMARK, B., HELDIN, C.H. & Ek B. (1984). Coexpression of a PDGF like growth factor and PDGF receptors in a human osteosarcoma cell line : implications for autocrine receptor activation. Cell, <u>39</u>, 447-457.
- BEUG, H., VON KIRCHBACH, A., DODERLEIN, G., CONSCIENCE, J.F. & 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.
- BEUG, H., HAYMAN, M.J. & GRAF, T. (1982). Myeloblasts transformed by the avian acute leukemia virus E26 are hormone-dependent for growth and for the expression of a putative myb-containing protein P135E26. EMBO J., <u>1</u>: 1069-1073.
- BEUG, H., LEUTZ, A., KAHN, P. & GRAF, T. (1984). Ts mutants of E26 leukemia virus allow transformed myeloblasts, but not erythroblasts or fibroblasts, to differentiate at the non permissive temperature. Cell, 39: 579-588.
- BIRCHMEIER, C., BIRNBAUM, D., WAITCHES, G., FASANO, O. & WIGLER, M. (1986). Characterization of an activated human ros gene. Mol. Cell. Biol., <u>6</u>: 3109-3116.
- BIRNSTIEL, M.L., BUSSLINGER, M. & STRUB, K. (1985). Transcription termination and 3' processing : the end is in site ! Cell, <u>41</u>: 349-359.
- BISHOP, J.M. & VARMUS, H.E. (1982). Fonctions and origins of retroviral transforming gene. In molecular biology of tumor viruses. Part III
   RNA tumor viruses. R. Weiss, N. Teich, H. Varmus and J. Coffin, eds (New York).
   Cold Spring Harbor Press., 999-1108.
- BISHOP, J.M. (1983). Cellular oncogenes and retroviruses. Ann. Rev. Biochem., 52: 301-354.
- BISTER, K., NUNN, M., MOSCOVICI, C., PERBAL, B., BALUDA, M.A. & DUESBERG, P. (1982). Acute leukemia viruses E26 and avian myeloblastosis virus have related transformation-specific RNA sequences but different genetic structures, gene products, and oncogenic properties. **Proc. Natl. Acad. Sci.**, USA : 3677-3681.
- BOHMANN, D., BOS, T.J., ADMON, A., NISHIMURA, T., VOGT, P.K. & TJIAN, R. (1987). Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science, 238, 1386-1392.

60
- BOS, T.J., BOHMANN, D., TSUCHIE, H., TJIAN, R. & VOGT, P.K. (1988).
  v-jun encodes a nuclear protein with enhancer binding properties of AP-1.
  Cell, <u>52</u>: 705-712.
- BOULUKOS, K.E., POGNONEC, A., BEGUE, A., GALIBERT, J.C., GESQUIERE, D., STEHELIN, D. & GHYSDAEL, J. (1988). Identification in chickens of an evolutionary conserved cellular ets-2 gene (c-ets-2) encoding nuclear proteins related to the products of the c-ets proto-oncogene. EMBO J., 7: 697-705.
- BOYLE, W.J., LIPSICK, J.S., REDDY, E.P. & BALUDA, M.A. (1983). Identification of the leukemogenic protein of avian myeloblastosis virus and of its normal cellular homologue. **Proc. Natl. Acad. Sci.**, <u>80</u>: 2834-2838.
- BOYLE, W.J., LAMPERT, M.A., LIPSICK, J.S. & BALUDA, M.A. (1984). Avian myeloblastosis virus and E26 virus oncogene products are nuclear proteins. **Proc. Natl. Acad. Sci., USA,** <u>81</u>: 4265-4269.
- BRUGGE, J.S. & ERIKSON, R.L. (1977). Identification of a transformationspecific antigen induced by an avian sarcoma virus. Nature, <u>269</u>: 346-348.
- BUNTE, T., GREISER-WILKE, I. & MOELLING, K. (1983). The transforming protein of the MC29 related virus CMII is a nuclear DNA-binding protein whereas MH2 codes for a cytoplasmic RNA-DNA binding protein. EMBO J., 2: 1087-1092.
- BUSS, J.E., KAMPS, M.P. & SEFTON, B.M. (1984). Myristic acid is attached to the transforming of Rous sarcoma virus during or immediately after synthesis and is present in both soluble and membrane-bound forms of the protein.
  Mol. Cell. Biol., 4: 2697-2704.
- CAPON, D.J., CHEN, E.Y., LEVINSON, A.D., SEEBURG, P.H. & GOEDDEL, D.V. (1983). Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature, <u>302</u>: 33-37.
- CAPON, D.J., SEEBURG, P.H., Mc GRATH, J.P., HAYTLICK, J.S., EDMAN, V., LEVINSON, A.D. & GOEDDEL, D.V. (1983). Activation of Ki-ras-2 gene in human colon and lung carcinomas by two different point mutations. Nature, 304: 507-513.
- CAPON, D.J., SEEBURG, P.H., Mc GRATH, J.P., HAYTLICK, J.S., EDMAN, V., LEVINSON, A.D. & GOEDDEL, D.V. (1984). Activation of Ki-ras-2 gene in human colon and lung carcinomas by two different point mutations. Nature, <u>312</u>: 507-512.

- CHANG, E.H., GONDA, M.A., ELLIS, R.W., SCOLNICK, E.M. & LOWY, D.R (1982). Human genome contains four genes homologous to transforming genes of Harvey and Kirsten murine sarcoma viruses. **Proc. Natl. Acad. Sci. USA**, 79: 4848-4852.
- CHARDIN, P. (1988). The ras superfamily proteins. Biochimie, 70: 865-868.
- CHEN, J.H. (1985). The proto-oncogene c-ets is preferentially expressed in lymphoid cells. Mol. Cell. Biol., <u>5</u>: 2993-3000.
- CHIU, R., BOYLE, W.J., MEEK, J., SMEAL, T., HUNTER, T. & KARIN, M. (1988). The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. Cell, 54: 541-552.
- CHRETIEN, S., DUBART, A., BEAUPAIN, D., RAICH, N., GRANDCHAMP, B., ROSA, J., GOOSSENS, M. & ROMEO, P.H. (1988). Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or in housekeeping expression. **Proc. Natl. Acad. Sci., USA, 85:** 6-10.
- CITRON, B., FALCK-PEDERSEN, E., SALDITT-GEORGIEFF, M. & DARNELL, J.E. (1984). Transcription termination occurs within a 1000 bases pair region downstream from the poly (A) site of the mouse -globin (major) gene. Nucleic Acids Res., <u>12</u>: 8723-8731.
- CLARKE, M.F., WESTIN, E., SCHMIDT, D., JOSEPHS, S.F., RATNER, L., WONG-STAAL, F., GALLO, R.C. & REITZ, M.S. (1984). Transformation of NIH 3T3 cells by a human c-sis cDNA clone. Nature, <u>308</u>: 464-467.
- CLEVELAND, D.W., FISCHER, S.G., KIRSCHNER, M.W. & LAEMMLI, U.K (1977). Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analyses by gel electrophoresis.
  J. Biol. Chem., 252: 1102-1106.
- COFFIN, J.M., HAGEMAN, T.C., MAXAM, A.M. & HASELTINE, W.A. (1978). Structure of the genome of Moloney murine leukemia virus : A terminally redundant sequence. Cell, <u>13</u>: 761-773.
- COHEN, J.B., LEVINSON, A.D. (1988). A point mutation in the last intron responsible for increased expression and transforming activity of the c-Ha-ras oncogene. Nature, 334:119-124.
- COLECLOUGH, C., COOPER, D. & PERRY, R.P. (1980). Rearrangement of immunoglobulin heavy chain genes during -lymphocyte development as revealed by studies of mouse plasmacytoma cells. **Proc. Natl. Acad. Sci.**, <u>77</u>: 1422-1426.

- COLL, J., SAULE, S., MARTIN, P., RAES, M.B., GRAF, T., BEUG, B., SIMON, I.E. & STEHELIN D. (1983). The cellular oncogenes c-myc, c-myb and c-erb are transcribed in defined types of avian hematopoietic cells. Exp. Cell. Res., <u>149</u>: 151-162.
- COLLETT, M.S. & ERIKSON, R.L. (1978). Protein kinase activity associated with the avian sarcoma virus src gene product. Proc. Natl. Acad. Sci. USA, <u>75</u>: 2021-2024.
- COLLETT, M.S., ERIKSON, E., PURCHIO, A.F., BRUGGE, J.S. & ERIKSON, R.L. (1979). A normal cell protein similar in structure and function to the avian sarcoma virus transforming gene product. **Proc. Natl. Acad. Sci. USA**, <u>76</u>: 3159-3163.
- COLLINS, S. & GOUDINE, M. (1982). Amplification of endogenous myc related DNA sequences in a human myeloid leukemia cell line. Nature, <u>298</u>: 679-681.
- COLLINS, S. & GROUDINE, M. (1983). Rearrangement and amplification of c-abl sequences in the human chronic myelogenous leukemia cell line K562.
   Proc. Natl. Acad. Sci. USA, 80: 4813-4817.
- COOPER, G.M. (1982). Cellular transforming genes. Science, 218: 801-806.
- COOPER, C.S., PARK, M., BLAIR, D.G., TAINSKY, M.A., HUCHNER, K., CROCE, C.M. & VAN DE WOODE, G.F. (1984). Molecular cloning of a new transforming gene from a chemically transformed human cell line. Nature, 311: 29-33.
- COOPER, J.A., GOULD, K.L., CARTWRIGHT, C.A. & HUNTER, T. (1986).
  Tyr 527 is phosphorylated in pp60<sup>src</sup>. Implications for regulation.
  Science, 231: 1431-1434.
- CORY, S. & ADAMS, J.M. (1980). Deletions are associated with somatic rearrangement of immunoglobulin heavy chain genes. Cell., 19: 37-51.
- COUSSENS, L., VAN BEVEREN, C., SMITH, D., CHEN, E., MITCHELL, R.L., ISACKE, C.M., VERMA, I.M. & ULLRICH, A. (1986). Structural alteration of viral homologue of receptor proto-oncogene fms at carboxyl terminus. Nature, 326: 277-280.
- CRAIG, R.W. & BLOCK, A. (1984). Early decline in **c-myb** oncogene expression in the differentiation of human myeloblastic leukemia (ML-I) cells induced with 2-0-Tetradecanoylphorbol 13-acetate. **Cancer Res.**, <u>44</u>: 442-446.

- CURRAN, T., VAN BEVEREN, C., LING, N. & VERMA, I.M. (1985). Viral and cellular fos proteins are complexed with a 39,000 daltons cellular protein.
   Mol. Cell. Biol., 5: 167-172.
- CULLEN, B.R., RAYMOND, K. & JU, G. (1985). Transcriptional activity of avian retroviral long terminal repeats directly correlates with enhancer activity.
   J. Virol., 53: 515-521.
- CURRAN, T., MILLER, A.D., ZOKAS, L. & VERMA, J.M. (1984). Viral and cellular fos proteins : A comparative analysis. Cell, <u>36</u>: 259-268.
- DALHBERG, J.E., SAYWER, R.C., TAYLOR, J.M., FARAS, A.J., LEVINSON, W.E., GOODMAN, H.M. & BISHOP, J.P. (1974). Transcription of DNA from the 70s of Rous sarcoma virus. I. Identification of a specific 4s RNA which serve as primer. J. Virol., 13: 1126-1133.
- DALLA-FAVERA, R., BREGNI, M., ERIKSON, J., PATTERSON, D., GALLO, R.C. & CROCE, C.M. (1982b). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc. Natl. Acad. Sci. USA, <u>79</u>: 7824-7827.
- DARLIX, J.L. (1986). Control of Rous sarcoma virus RNA translation and packaging by the 5' and 3' untranslated sequences.
   J. Mol. Biol., <u>189</u>: 421-434.
- DELLI BOVI, P. & BASILICO, C. (1987a). Isolation of a rearranged human transforming gene following transfection of Kaposi's sarcoma DNA. **Proc. Natl. Acad. Sci. USA**, <u>84</u>: 5660-5664.
- DELLI BOVI, P., CURATOLA, A.M., KERN, F.G., GRECO, A., ITTMANN, M. & BASILICO, C. (1987b). An oncogene isolated by transfection of Kaposi's sarcoma DNA encodes a Growth Factor that is a member of the FGF Family. Cell, 50: 729-737.
- DER, J.C., KRONTIRIS, T.C. & COOPER, G.M. (1982). Transforming gene of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. **Proc. Natl. Acad. Sci. USA**, 9: 3637-3640.
- DE TAISNE, C.A., GEGONNE, A., STEHELIN, D., BERNHEIM, A. & BERGER, R. (1984). Chromosomal localization of the human proto-oncogene c-ets. Nature (London), 310: 581-583.

- DEUEL, T.F., HUANG, J.S., PROFFITT, R.T., BAENZIGER, J.U., CHANG, D. & KENNEDY, B.B. (1981). Human platelet-derived growth factor : purification and resolution into two active protein fractions. J. Biol. Chem., <u>256</u> : 8896-8899.
- DHAR, R., NIETO, A., KOLLER, R., DEFEO-JONES, R.D., ROBINSON, P., TEMELES, G. & SCOLNICK, E.M. (1984). Nucleotide sequence of two H-ras-related isolated from the yeast Saccharomyces cerevisiae. Nucleic Acids Res., 12: 3611.
- DINOWITH, M. (1975). Inhibition of Rous sarcoma virus by alpha-amanitin : Possible role of cell DNA-dependent RNA polymerase form II.
   Virology, <u>66</u>: 1-9.
- DINGWALL, C. & LASKEY, R.A. (1986). Protein import into the cell nucleus. Ann. Rev. Cell. Biol., 2: 367-390.
- DONNER, P., GREISER-WILKE I. & MOELLING, K. (1982). Nuclear localization and DNA binding of the transforming gene product of avian myelocytomatosis virus. Nature, 296: 262-265.
- DOOLITTLE, R.F., HUNKAPILLER, M.W., HOOD, L.E., DEVARE, S.G., ROBBINS, K.C., AARONSON, S.A. & ANTONIADES, H.N. (1983). Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. Science, <u>221</u>: 275-277.
- DOWNWARD, J., YARDEN, Y., MAYES, E., SCRACE, G., TOTTY, N., STOCKWELL, P., ULLRICH, A., SCHLESSINGER, J. & WATERFIELD, M.D. (1984). Close similarity of epidermal growth factor receptor and v-erbB oncogene protein sequences. Nature, 307: 521-527.
- DOZIER, C., DENHEZ, F., HENRY, C., COLL, J., BEGUE, A., QUATANNENS, B., SAULE, S. & STEHELIN, D. (1988). Alternative splicing of RNAs transcribed from the chicken c-mil gene. Mol. Cell. Biol., 8: 1835-1838.
- DURBAN, E.M. & BOETTIGER, D. (1981). Differential effects of transforming avian RNA tumor viruses on avian macrophages. **Proc. Natl. Acad. Sci. USA**, 78: 3600-3604.
- DYNAN, W.S. (1986). Promoters for housekeeping genes. Trends Genet., 8: 196-197.
- EARLY, P., ROGERS, J., DAVIS, M., CALAME, K., BOND, M., WALL, R. & HOOD, L. (1980). Two mRNAs can be produced from a single immunoglobulin gene by alternative RNA processing pathways. Cell, <u>20</u>: 313-319.

- EISENMAN, R.N. & VOGT, V.M. (1978). The biosynthesis of oncovirus proteins. Biochim. Biophys. Acta, 473: 187-239.
- EVA, A. & AARONSON, S.A. (1985). Isolation of a new human oncogene from a diffuse B-cell lymphoma. Nature, <u>316</u>: 273-275.
- FASANO, O., ALDRICH, T., TAMANOI, F., TAPAROWSKI, E., FURTH, M. & WIGLER, M. (1984). Analysis of the transforming potential of human H-ras by random mutagenesis. Proc. Natl. Acad. Sci. USA, 82:426-430.
- FITZGERALD, M. & SHENK, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for poly adenylation of late SV40 mRNAs. Cell, <u>24</u>: 251-260.
- FORD, J.P. & HSU, M.T. (1978). Transcription pattern of in vivo-labelled late simian virus 40 RNA : Equimolar transcription beyond the mRNA 3' terminus.
   J. Virol., <u>28</u>: 795-801.
- FOSTER, A.D., SHIBUYA, M. & HANAFUSA, M. (1985). Activation of the transforming potential of the cellular **fps** gene. **Cell**, <u>42</u>: 105-115.
- FRANCHINI, G., WONG-STAAL, F., BALUDA, M.A., LENGEL, C. & TRONICK, S.R. (1983). Structural organisation and expression of human DNA sequences related to the transforming gene of avian myeloblastosis virus.
   Proc. Natl. Acad. Sci. USA, 80: 7385-7389.
- FUJIWARA, S., FISHER, R.J., SETH, A. BHAT, N.K., SHOWALTER, S.D., ZWEIG, M. & PAPAS, T.S. (1988). Characterization and localization of the products of the human homology of the v-ets oncogene. Oncogene, 2:99-103.
- FURUICHY, Y., SHATKIN, A.J., STAVNEZER, E. & BISHOP, J.M. (1975). Blocked methylated 5 terminal sequence in avian sarcoma virus RNA. Nature, <u>257</u>: 618-620.
- GALLO, R.C., SALAHUDDIN, S.Z., POPOVIC, M., SHEARER, G.M., KAPLAN, M., HAYNES, B.F., PALKER, T.J., REDFIELD, R., OLESKE, J., SAFAI, B., WHITE, G., FOSTER, P. & MARKHAM, P.D. (1984). Human Tlymphotropic retrovirus, HTLV-III, isolated form AIDS patients and donors at risk for AIDS. Science, 224: 500-503.
- GALLO, R.C., WONG-STAAL, F., MONTAGNIER, L., HASELTINE, W.A., YOSHIDA, M. (1988). HIV/HTLV gene nomenclature. Nature, <u>333</u>: 504.

- GAZZIT, A., IGARASHI, H., CHIU, I.M., SRINIVASAN, A., YANIV, A., TRONICK, S.R., ROBBINS, K.C. & AARONSON, S.A. (1984). Expression of the normal human sis-PDGF-2 coding sequence induces cellular transformation. Cell, 39: 89-97.
- GAZZOLO, L., MOSCOVICI, C., MOSCOVICI, M.G. & SAMARUT, J. (1979). Response of hemopoietic cells to avian acute leukemia viruses : Effects on the differentiation of the target cells. Cell, <u>16</u>: 627-638.
- GAZZOLO, L., SAMARUT, J., BOUABDELLI, M. & BLANCHET, J.P. (1980). Early precursors in the erythroid lineage are the specific target cells of avian erythroblastosis virus in vitro. Cell, <u>22</u>: 683-691.
- GEGONNE, A., LEPRINCE, D., POGNONEC, P., DERNIS, D., RAES, M.B., STEHELIN, D. & GHYSDAEL, J. (1987). The 5' extremity of the v-ets oncogene of avian leukemia virus E26 encodes amino-acid sequences not derived from the major c-ets encoded cellular proteins. Virology, 156: 177-180.
- GHYSDAEL, J., GEGONNE, A., POGNONEC, P., DERNIS, D., LEPRINCE, D. & STEHELIN, D. (1986a). Identification of preferential expression in thymic and bursal lymphocytes of a c-ets oncogene encoded Mr-54000 cytoplasmic protein.
   Proc. Natl. Acad. Sci. USA, 83: 1714-1718.

\_\_\_\_

- GHYSDAEL, J., GEGONNE, A., POGNONEC, P., BOULUKOS, K., LEPRINCE, D., DERNIS, D., LAGROU, C. & STEHELIN, D. (1986b). Identification in chicken macrophages of a set of proteins related to but distinct from the chicken cellular c-ets encoded protein p54<sup>c-ets</sup>. EMBO J., <u>5</u>: 2251-2256.
- GONDA, T.J., GOUGH, N.M., DUNN, A.R. & DE BLAQUIERE, J. (1985). Nucleotide sequence of cDNA clones of the murine myb proto-oncogene. EMBO J., <u>4</u>: 2003-2008.
- GIBBS, J.B., SIGAL, I.S., POE, M. & SCOLNICK, E.M. (1984). Intrinsic GDPase activity distinguishes normal and oncogenic **ras** p21 molecules. **Proc. Natl. Acad. Sci. USA**, <u>81</u>: 5704-5708.
- GILMAN, A.G. (1984). G proteins and dual control of adenylate cyclase. Cell, <u>36</u>: 577-579.
- GILMORE, T.D. & TEMIN, H. (1986). Different localization of the product of the **v-rel** oncogene in chicken fibroblasts and spleen cells correlated with transformation by **REV-T**. Cell, <u>44</u>: 791-800.

- GLUZMANN, Y. (1981). SV 40-transformed simian cells support the replication of early SV 40-mutants. Cell, 23: 175-182.
- GONDA, T.J., SHEINESS, D.K. & BISHOP, J.M. (1982). Transcripts from the cellular homologs of retroviral oncogenes : distribution among chicken tissues.
   Mol. Cell. Biol., 2: 617-624.
- GONDA, T.J. & METCALF, D. (1984). Expression of myb, myc and fos proto-oncogenes during differentiation of a murine myeloid leukemia. Nature, <u>310</u>: 249-251.
- GOUBIN, G., GOLDMAN, D.S., LUCE, J., NEIMAN, P.E. & COOPER, G.M. (1983). Molecular cloning and nucleotide sequence of a transforming gene detected by transfection of Chicken B-cell lymphoma DNA. Nature, <u>302</u>: 114-119.
- GRAF, T. & BEUG, H. (1978). Avian leukemia viruses interaction with their target cells in vivo and in vitro. Biochem. Biophys. Acta., <u>516</u>: 269-299.
- GRAF, T., OKER-BLOM, N., TODOROV, T.G. & BEUG, H. (1979). Transforming capacities and defectiveness of avian leukemia viruses OK10 and E26. Virology, <u>99</u>: 431-436.
- GRAF, T. & STEHELIN, D. (1982). Avian leukemia viruses oncogenes and genome structure. Biochim Biophys. Acta., 651: 245-271.
- GRANDGENETT, D.P., VORA, A.C. & SCHIFF, R.D. (1978). A 32000dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. Virology, <u>89</u>: 119-132.
- GRANDGENETT, D.P. & AJAYKUMAR, C.V. (1985). Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. Nucleic Acids Res., <u>13</u>: 6205-6221.
- GREEN, S., WALTER, P., HUMAR, V., KRUST, A., BARNERT, J.M., ARGOS, P. & CHAMBON, P. (1986). Human receptor cDNA : sequence, expression and homology to v-erbA. Nature, 320: 134-139.
- GROFFEN, J., STEPHENSON, J.R., HEISTERKAMP, N., DE KLEIN, A., BARTRAM, C.R. & GROSVELD, D.G. (1984). Philadelphia chromosomal breakpoints are clustered within a limited region, bcr on chromosome 22. Cell, <u>36</u>: 93-99.

- GUYADER, M., EMERMAN, M., SONIGO, P., CLAVEL, F., MONTAGNIER, L. & ALIZON, M. (1987). Genome organization and transactivation of the human immunodeficiency virus type 2. Nature, 326: 662-669.
- HAGENBUCHLE, O., WELLAUER, P.K., CRIBBS, D.L. & SCHIBLER, U. (1984). Termination of transcription in the mouse α -amylase gene Amy-2a occurs at multiple sites downstream of the poly-adenylation site.
  Cell, <u>38</u>: 737-744.
- HALL, A., MARSHALL, C.J., SPURR, N. & WEISS, R.A. (1983). Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1. Nature, 303: 396-400.
- HALL, A. & BROWN, R. (1985). Human N-ras : cDNA cloning and gene structure. Nucleic Acids Res., 13: 5255-5268.

HANN, S.R. & EISENMAN, R.N. (1984). Proteins encoded by the human c-myc oncogene : differential expression in neoplastic cells. Mol. Cell. Biol., <u>4</u>: 2486-2497.

- HAPPER, M.E., FRANCHINI, G., LOVE, J., SIMON, M.I., GALLO, R.C. & WONG-STAAL, F. (1983). Chromosomal sublocalization of human c-myb and c-fes cellular oncogenes. Nature, 304: 169-171.
- HAPPER, J.R., ROOP, D.R. & YUSPA, S.H. (1986). Transfection of the Ej ras<sup>Ha</sup> gene into keratinocytes derived from carcinogen-induced mouse papillomas causes malignant progression. Mol. Cell. Biol., 6: 3144-3149.
- HARVEY, J.J. (1964). An unidentified virus which causes the rapid production of tumours in mice. Nature, <u>204</u>: 1104-1105.
- HASELTINE, W.A., MAXAM, A.M. & GILBERT, W. (1977). Rous sarcoma virus genome is terminally redundant : the 5' sequence. **Proc. Natl. Acad. Sci. USA**, 74 : 989-993.
- HAYDAY, A.C., GILLIES, S.D., SAITO, H., WOOD, C., WIMAN, K., HAYWARD, W.S. & TONEGAWA, S. (1984). Activation of a translocated human **c-myc** gene by an enhancer in the immunoglobulin heavy chain locus. **Nature**, <u>307</u>: 334-340.
- HAYMAN, M.J. (1978). Synthesis and processing of avian sarcoma virus glycoproteins. Virology, <u>85</u>: 475-486.

- HAYMAN, M.J., KITCHENER, G., VOGT, P.K. & BEUG, H. (1985). The putative transforming protein of S13 avian erythroblastosis virus is a transmembrane glycoprotein with an associated protein kinase activity. **Proc. Natl. Acad. Sci., USA**, <u>82</u>: 8237-8241.
- HAYWARD, W.S. (1977). Size and genetic content of viral RNAs in avian oncovirus-infected cells.
   J. Virol., <u>24</u>: 47-63.
- HELDIN, C.H., WESTERMARK, B. & WASTESON, A. (1979). Plateletderived growth factor : purification and partial characterization. **Proc. Natl. Acad. Sci. USA**, <u>76</u>: 3722-3726.
- HELDIN, C.H. & WESTERMAERK, B. (1984). Growth factors : mechanism of fonction and relation to oncogenes. Cell, <u>37</u>: 9-20.
- HONJO, T. & KATAOKA, T. (1978). Organisation of immunoglobulin heavy chain genes and allelic deletion model. Proc. Natl. Acad. Sci. USA, 75: 2140-2144.
- HUANG, J.S., HUANG, S.S. & DEUEL, T.F. (1984). Transforming protein of simian sarcoma virus stimulates autocrine cell growth of SSVtransformed cells through PDGF cell-surface receptors. Cell, <u>39</u>: 79-87.
- HUNTER, T. (1985). At last the insulin receptor. Nature, <u>313</u>: 740-741.
- HUNTER, T. & COOPER, J.A. (1985). Protein-tyrosine kinases. Ann. Rev. Biochem., 54: 897-930.
- IBANEZ, C.E. & LIPSICK, J.S (1988). Structural and functional domains of the **myb** oncogenes : requirements for nuclear transport, myeloid transformation, and colony formation. J. Virol., 62: 1981-1988.
- IKAWA, S., KAGINO-YAMAGISHI, K., KAWAI, S., YAMAMOTO, T. & TOYOCHIMA, K. (1986). Activation of the cellular src gene by transducing retrovirus.
   Mol. Cell. Biol., <u>6</u>: 2420-2428.
- IVANOW, X., MLADENOW, Z., NEDYALKOV, S. & TODOROV, T. (1962). Experimental investigation into avian leukosis. I. Transmission experiments of certain diseases of the avian leukosis complex found in Bulgaria.
   Bull. Inst. Pathol. Comp. Anim. Domest., <u>9</u>: 5-36.

- JACKS, T. & VARMUS, H.E., (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science, 230: 1237-1242.
- JOHNSTON, P. (1971). Taxonomic features of seven serotypes of simian and ape foamy viruses.
   Infect. Immun., 3: 793-799.
- JURDIC, P., BENCHAIBI, M., GANDRILLON, O. & SAMARUT, J. (1987). Transforming and mitogenic effects of avian leukemia Virus E26 on chicken hematopoietic cells and fibroblasts, respectively, correlated with level of expression of the provirus.
   J. Virol., 61: 3058-3065.
- KALYANARAMAN, V.S., SARNGAD-RAN, M.G., ROBERT-GUROFF, M., MIYOSHI, I., BLAYNEY, D., GOLDE, D. & GALLO, R.C. (1982). A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. Science, 218: 571-573.
- KAMPS, M.P., BUSS, J.E. & SEFTON, B.M. (1986). Mutation of NH<sub>2</sub>terminal glycine of p60<sup>Src</sup> prevents both myristoylation and morphological transformation. **Proc. Natl. Acad. Sci., USA**, <u>82</u>: 4625-4628.
- KAPLAN, J.C. & SZAJNERT, M.F. (1985). Le paradigme du lymphome de Burkitt. Médecine/Sciences, <u>1</u>: 17-23.
- KATZEN, A.M., KORNBERG, T.B. & BISHOP, J.M. (1985). Isolation of the proto-oncogene c-myb from D. Megalogaster. Cell, 41: 449-456.
- KAUFMAN, R.J. & SHARP, P.A. (1983). Growth-dependent expression of dihydrofolate reductase mRNA from modular cDNA genes.
   Mol. Cell. Biol., <u>3</u>: 1598-1608.
- KEATING, M.T., WILLIAMS, L.T. (1988). Autocrine stimulation of intracellular PDGF receptors v-sis-transformed cells. Science, 239: 914-916.
- KELLER, W. (1984). The RNA lariat : A new rung to the splicing of mRNA precursors. Cell, 39: 423-425.
- KINGSTON, R.E., BALDWIN, A.S., Jr. & SHARP, P.A. (1984). Regulation of heat shock protein 70 gene expression by c-myc. Nature, <u>312</u>: 280-282.

- KIRSTEN, W.H., MAYER, L.A., WOLLMANN, R.L. & PIERCE, M.I. (1967). Studies on a murine erythroblastosis virus. J. Natl. Cancer Inst., <u>38</u>: 117-139.
- KLEMPNAUER, K.H., GONDA, T.J. & BISHOP, J.M. (1982). Nucleotide sequence of the retrovirus leukemia gene v-myb and its cellular progenitor c-myb : the architecture of a transduced oncogene.
  Cell, <u>31</u>: 453-463.
- KLEMPNAUER, K.H., SYMONDS, G., EVAN, G.I. & BISHOP, J.M. (1984). Subcellular localization of proteins encoded by oncogenes of avian myeloblastosis virus and avian leukemia virus E26 and by the chicken c-myb gene. Cell, 37: 537-547.
- KLEMPNAUER, K.H. & SIPPEL, A.E. (1986a). Subnuclear localization of proteins encoded by the oncogene v-myb and its cellular homolog c-myb. Mol. Cell. Biol., <u>6</u>: 62-69.
- KLEMPNAUER, K.H., BONIFER, C. & SIPPEL, A.E. (1986b). Identi-fication and characterization of the protein encoded by the human c-myb protooncogene.
   EMBO J., 8: 1911-1986.
- KLEMPNAUER, K.H., BIEDENKAPP, H., BONIFER, C., PETERS, C. & SIPPEL, A.E. (1986c). Aminoacid sequences close to the aminoterminus of viral myb protein are involved in DNA-binding. Second Annual Meeting on Oncogenes, p. 244.
- KLOETZER, W.S., MAXWELL, S.A. & ARLINGHAUS, R.B. (1984). Further characterization of the P85gag-mos associated protein kinase activity. Virology, 138: 143-155.
- KONARSKA, M.M., PADGETT, R.A. & SHARP, P.A. (1985). Transplicing of mRNA precursors in vitro. Cell, <u>42</u>: 165-171.
- KONOPKA, J.B., WATANABE, S.M. & WITTE, O.N. (1984). An alteration of the human **c-abl** protein in K562 leukemia cells unmasks associated tyrosine kinase activity. **Cell**, 37: 1035-1042.
- KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulated translation by eukaryotic ribosomes. Cell, <u>44</u>: 283-292.
- KRUIJER, W., COOPER, J.A., HUNTER, T. & VERMA, I.M. (1984). Platelet derived growth factor induces rapid but transient expression of the c-fos gene and proteins. Nature (London), <u>312</u>: 711-716.

- LAI, M.M.C., NEIL, J.C. & VOGT, P.K. (1980). Cell-free translation of avian erythroblastosis virus RNA yields two specific and distinct proteins with molecular weights of 75 000 and 40 000.
   Virology, 100: 475-483.
- LAND, H., PARADA, L.F. & WEINBERG, R.A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes.
   Nature, 304: 596-602.
- LANE, M.A., SAINTEN, A., DOHERTY, K.M. & COOPER, M. (1984). Isolation and characterization of a stage specific transforming gene, Tlym-1 form T cell lymphomas.
   Proc. Natl. Acad. Sci., USA, <u>81</u>: 2227-2231.
- LEAMNSON, R.N. & HALPERN, M.S. (1976). Subunit structure of the glycoprotein complex of avian tumor virus. J. Virol., <u>18</u>: 956-968.
- LEE, J.S., VARMUS, H.E. & BISHOP, J.M. (1979). Virus-specific messenger RNAs is permissive cells infected by avian sarcoma virus.
   J. Biol. Chem., 245: 8015-8022.
- LEE, W.M.F., SCHWAB, M., WESTAWAY, D. & VARMUS, H. (1985). Augmented expression of normal c-myc is sufficient for cotransformation of rat embryo cells with a mutant ras gene. Mol. Cell. Biol., 5: 3345-3356.
- LEE, W., HASLINGER, A., KARIN, M. & TJIAN, R. (1987a). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40.
   Nature, 325: 368-372.
- LEFF, S.E. & ROSENFELD, M.G. (1986). Complex transcriptional units: diversity in gene expression by alternative RNA processing.
   Ann. Rev. Biochem., <u>55</u>: 1091-1117.
- LEMEUR, M.A., GALLIOT, B. & GERLINGER, P. (1984). Termination of the ovalbumin gene transcription.
   EMBO J., 3: 2779-2786.
- LEPRINCE, D., GEGONNE, A., COLL, J., DE TAISNE, C., SCHNEEBERGER, A., LAGROU, C. & STEHELIN, D. (1983). A putative second cell-derived oncogene of the avian leukemia retrovirus E26. Nature, London, <u>306</u>: 395-397.
- LERNER, M., BOYLE, J.A., MOUNT, S.M., WOLIN, S.L. & STEITZ, J. (1980). Are snRNPs involved in splicing? Nature, 283: 220-224.

- LEUTZ, A., BEUG, H. & GRAF, T. (1984). Purification and characterization of cMGF, a novel chicken myelomonocytic growth factor. **EMBO J.** <u>3</u>: 3191-3197.
- LIPSICK, J.S. & IBANEZ, C.E. (1987). env-encoded residues are not required for transformation by p48v-myb. J. Virol., <u>61</u>: 933-936.
- LOOSFELT, H., ATGER, M., MISHARI, M., GUIOCHON-MANTEL, A., MERIEL, C., LOGEAT, F., BENAROUS, R. & MIGROM, E. (1986). Cloning and sequence analysis of rabbit progesterone-receptor complementary DNA. Proc. Natl. Acad. Sci. USA, <u>83</u>: 9045-9049.
- MACARA, I.G., MARINETTI, G.V. & BALDUZZI, P.C. (1984). Transforming protein of avian sarcoma virus UR2 is associated with phosphatidylinositol kinase activity: possible role in tumorigenesis. **Proc. Natl. Acad. Sci. USA**, <u>81</u>: 2728-2732.
- MANGEL, W.F., DELIUS, H. & DUESBERG, P.H. (1974). Structure and molecular weight of the 60-70 S RNA and the 30-40 S RNA of the Rous sarcoma virus. **Proc. Natl. Acad. Sci. USA,** 71: 4541-4545.
- MAKI, Y., BOS, T.J., DAVIS, C., STARBUCK, M. & VOGT, P.K. (1987). Avian sarcoma virus 17 carries the jun oncogene. Proc. Natl. Acad. Sci. USA, 84 : 2848-2852.
- MANIATIS, T. & REED, R. (1987). The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature, <u>325</u>: 673-678.
- MARTIN, P., HENRY, C., DENHEZ, F., AMOUYEL, P., BECHADE, C., CALOTHY, G., DEBUIRE, B., STEHELIN, D. & SAULE, S. (1986). Characterisation of a MH2 mutant lacking the **v-myc** oncogene. **Virology**, <u>153</u>: 272-279.
- MARTIN-ZANCA, D., HUGHES, S.H. & BARBACID, M. (1986). A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. nature, 319: 743-748.
- MATLASHEWSKI, G., PIM, D., BANKS, L. & GRAWFORD, L. (1987). Alternative splicing of human p53 transcripts. Oncogene Res., 1, 77-85.
- MAYER, B.J., HAMAGUHAI, M. & HANAFUSA, H. (1988). A novel viral oncogene with structural similarity to phospholipase C.
   Nature, <u>332</u>: 272-275.

- Mc GRATH, J.P., CAPON, D.J., SMITH, D.H., CHEN, E.Y., SEEBURG, P.H., GOEDDEL, D.V. & LEVINSON, A.D. (1983). Structure and organisation of the human Ki-ras proto-oncogene and a related processed pseudogene. Nature, <u>304</u>: 501-507.
- Mc GRATH, J.P., CAPON, D.J., GOEDDEL, D.V. & LEVINSON, A.D. (1984). Comparative biochemical properties of normal and activated human ras p21 protein. Nature, 310: 644-649.
- MEIJLINK, F., CURRAN, T., MILLER, A.D. & VERMA, I. (1985). Removal of a 67 bases pair sequence in the non coding region of proto-oncogene fos converts it to a transforming gene. Proc. Natl. Acad. Sci. USA, 82:644-649.
- MESSING, J., CREA, R. & SEEBURG, P.H. (1981). A system for shotgun DNA sequencing. Nucleic Acids Res., 9: 309-321.
- MILES, B.D. & ROBINSON, H.L. (1985). High-frequency transduction of c-erbB in avian leukosis-induced erythroblastosis.
  J. Virol., <u>54</u>: 295-303.
- MILLER, A.D., CURRAN, T. & VERMA, T.M. (1984). c-fos protein can induce cellular transformation : A novel mechanism of activation of a cellular oncogene. Cell, 36: 51-60.
- MIYOSHI, I., KUBONISHI, I., YOSHIMOTO, S., AKAGI, T., OHTSUKI, Y., SHIRAISHI, Y., NAGATA, K. & HINUMA, Y. (1981). Type C virus particles in a cord T-cell line derived by cocultivating normal human cord leukocytes and human leukaemic T cells. Nature, 294: 770-771.
- MOELLING, K., HEIMANN, B., BEIMLING, P., RAPP, U.R. & SANDER, T. (1984). Serine-and threonine-specific protein kinase activities of purified gag-mil and gag-raf proteins. Nature, <u>312</u>: 558.
- MOELLING, K., PFAFF, E., BEUG, H., BEIMLING, P., BUNTE, T., SCHALLER, M.E. & GRAF, T. (1985). DNA-binding activity is associated with purified **myb** proteins from AMV and E26 viruses and is temperaturesensitive for E26 ts mutants. **Cell**, 40: 983-990.
- MONTELL, C., FISCHER, E.F., CARUTHERS, M.H. & BERK, A.J. (1983). Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' concensus sequence AAUAAA. Nature, <u>305</u>: 600-605.

- MOSCOVICI, C., SAMARUT, J., GAZZOLO, L. & MOSCOVICI, M.G. (1981). Myeloid and erythroid neoplastic responses to avian defective leukemia viruses in chickens and in quail.
   Virology, <u>113</u>: 765-768.
- MOSCOVICI, M.G., JURDIC, P., SAMARUT, J., GAZZOLO, L., MURA, C.V. & MOSCOVICI, C. (1983). Characterization of the hemopoietic target cells for the avian leukemia virus E26.
   Virology, 129: 65-78.
- MOUNT, S.M. (1982). A catalogue of splice jonction sequence. Nucleic Acids Res., 10: 459-472.
- MUNROE, S.H., (1984). Secondary structure of splice sites in adenovirus mRNA precursors. Nucleic Acids Res., <u>12</u>: 8437-8456.
- MURRAY, M.J., CUNNINGHAM, J.M., PARADA, L.F., PAUTRY, F. & LEBOWITZ, P. (1983). The HL-60 transforming sequence : a ras oncogene coexisting with altered myc genes in hematopoietic. Cell, 33: 749-757.
- NABESHIMA, Y., FUJII-KURIYAMA, Y., MURAMATSU, M. & OGATA, K. (1984). Alternative transcription and two modes of splicing result in two myosin light chains from one gene.
  Nature, <u>308</u>: 333-338.
- NECKAMAYER, W.S., SHIBUYA, M., HSU, M. & WANG, L. (1986). Protooncogene c-ros codes for a molecule with structural features common to those of growth factor receptors and displays tissue-specific and developmentally regulated expression.
   Mol. Cell. Biol., 6: 1478-1486.
- NEIL, J.C., HUGHES, D., Mc FARLANE, R., WILKIE, N.M., ONIONS, D.E., LEES, G. & JAREET, O. (1984). Transduction and rearrangement of the myc gene by feline leukemia virus in naturally occuring T-cell leukemias. Nature, <u>308</u>: 814-820.
- NEVINS, J.R. & DARNELL, J.E. (1978). Steps in the processing of Ad2 mRNA : poly (A)<sup>+</sup> nuclear sequences are conserved and poly (A) addition procedes splicing. Cell, <u>15</u>: 1477-1478.
- NEVINS, J.R., BLANCHARD, J.M. & DARNELL, J.E. (1980). Transcription units of adenovirus type 2 (termination of transcription beyond the poly A addition site in early regions 2 and 4).
   J. Mol. Biol., <u>144</u>: 377-386.
- NEVINS, J.R. (1983). The pathway of eukaryotic mRNA formation. Ann. Rev. Biochem., <u>52</u>: 441-466.

- NILSEN, T.W., MARONEY, P.A., GOODWIN, R.G., ROTTMAN, F.M., CRITTENDEN, L.B., RAINES, M.A. & KUNG, H.J. (1985). c-erbB activation in ALV-induced erythroblastosis : novel RNA processing and promotor insertion result in expression of an amino-truncated EGF receptor. Cell, <u>41</u>: 719-726.
- NISHIZAWA, M., KOYAMA, T. & KAWAI, S. (1985). Unusual features of the leader sequence of Rous sarcoma virus packaging mutant TK 15.
   J. Virol., <u>55</u>: 881-885.
- NUNN, M.F., SEEBURG, P.H., MOSCOVICI, C. & DUESBERG, P.H. (1983). Tripartite structure of the avian erythroblastosis virus E26 transforming gene. Nature, London, 306: 391-395.
- NUNN, M., WEIHER, H., BULLOCK, P. & DUESBERG, P. (1984). Avian erythroblastosis virus E26: nucleotide sequence of the tripartite onc gene and of the LTR, and analysis of the cellular prototype of the viral ets sequence.
  Virology, 139: 330-339.
- OPPERMAN, H., BISHOP, J.M., VARMUS, H.E. & LEVINSON, L. (1977). A joint product of the genes gag and pol of avian sarcoma virus : a possible precursor of reverse transcriptase. Cell, <u>12</u>: 993-1005.
- OWEN, A.J., PANTAZIS, P. & ANTONIADES, H.N. (1984). Simian sarcoma virus-transformed cells secrete a mitogen identical to plateletderived growth factor. Science, 225: 54-56.
- PACHL, C., BIEGALKE, B. & LINIAL, M. (1983). RNA and protein encoded by MH2 virus: evidence for subgenomic expression of v-myc.
   J. Virol., <u>45</u>: 133-139.
- PADUA, R.A., BARRAS, N. & GURRIE, G.A. (1984). A novel transforming gene in a human malignant melanoma cell line. Nature, <u>311</u>: 671-673.
- PAPAGEORGE, A.G., De FEO-JONES, D., ROBINSON, P., TEMELES, G. & SCOLNICK, E.M. (1984). Saccharomyces cerevisiae synthesizes proteins related to the p21 gene product of ras genes found in mammals. Mol. Cell. Biol., <u>4</u>: 23-29.
- PAPKOFF, J., NIGG, E.A. & HUNTER, T. (1983). The transforming protein of moloney murine sarcoma virus is a soluble cytoplasmic protein. Cell, <u>33</u>: 161-172.
- PATSCHINSKY, T., SCHROEER, B. & BISTER, K. (1986). Protein product of proto-oncogene c-mil. Mol. Cell. Biol., <u>6</u>: 739-744.

- PERRY, R.P. & KELLEY, D.E. (1979). Immunoglobulin messenger RNAs in Murine cell lines that have characteristics of immature B lymphocytes. Cell, 18: 1333-1339.
- PETERS, C.W.B., SIPPEL, A.E., VINGRON, M. & KLEMPNAUER, K.H., (1987). Drosophila and vertebrate myb proteins share two conserved regions, one of which functions as a DNA-binding domain. EMBO J., <u>6</u>: 3085-3090.
- PETERSEN, R.B. & HACKETT, P.B. (1985). Characterisation of ribosome binding a Rous sarcoma virus RNA in vitro.
   J. Virol., <u>56</u>: 683-690.
- POGNONEC, P., BOULUKOS, K.E., GESQUIERE, J.C., STEHELIN, D. & GHYSDAEL, J. (1988). Mitogenic stimulation of thymocytes results in the calcium-dependent phosphorylation of c-ets-1 proteins. EMBO. J., <u>7</u>: 977-983.
- POIESZ, B.J., RUSCETTI, F.W., GAZDAR, A.F., BUNN, P.A., MINNA, J.D. & GALLO, R.C. (1980b). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient cutaneous T-cell lymphoma.
  Proc. Natl. Acad. Sci. USA, 71: 7415-7419.
- POWERS, S., KATAOKA, T., FASANO, O., GOLDFARB, M., STRATHERN, J.S., BROACH, J. & WIGLER, M. (1984). Genes in S. cerevisiae encoding proteins with domains homologous to the mammalian ras proteins. Cell, <u>36</u>: 607-612.
- PROUDFOOT, N.J. & BROWNLEE, G.G. (1976). 3' Non-coding region sequences in eukaryotic messenger RNA. Nature, <u>263</u>: 211-214.
- PRYWES, R., FOULKES, J.G., ROSENBERG, N. & BALTIMORE, D. (1983). Sequences of the A-MuLv protein needed for fibroblast and lymphoid cell transformation. Cell, 34: 569-579.
- RABBITS, T.H., HAMLYN, P.H. & BAER, R. (1983). Altered nucleotide sequences of a translocated **c-myc** gene in Burkitt lymphoma. Nature, <u>306</u>: 760-765.
- RABSON, A.B. & MARTIN, M.A. (1985). Molecular organisation of the AIDS retrovirus. Cell, 40: 477-480.
- RADKE, K., BEUG, H., KORNFELD, S. & GRAF, T. (1982). Transformation of both erythroid and myeloid cells by E26, an avian leukemia virus that contains the **myb** gene. **Cell**, <u>31</u>: 643-653.

- RAINES, E.W. & ROSS, R. (1982). Platelet-derived growth factor. I. High yield purification and evidence for multiple forms.
  J. Biol. Chem., 257: 5154-5160.
- RAINES, M.A., LEWIS, W.G., CRITTENDE, L.B. & 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.
- RALSTON, R. & BISHOP, J.M. (1983). The protein products of the myc and myb oncogenes and adenovirus Ela are structurally related. Nature, <u>306</u>: 803-806.
- RAMSAY, G., EVAN, G.F. & BISHOP, J.M. (1984). The protein encoded by the human proto-oncogene c-myc. Proc. Natl. Acad. Sci. USA, <u>81</u>: 7742-7746.
- RAO, V.N., PAPAS, T.S. & REDDY, E.S.P. (1987). erg, a human etsrelated gene on chromosome 21 : alternative splicing, polyadenylation, and translation. Science, 237: 635-639.
- REDDY, E.P., REYNOLDS, R.K., SANTOS, E. & BARBACID, M. (1982). A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature, <u>300</u>: 149-152.
- REDDY, E.S., RAO, V.N. & PAPAS, T.S. (1987). The erg gene : a human gene related to the ets oncogene. Proc. Natl. Acad. Sci., USA,
- REICH, N., OREN, M. & LEVINE, A. (1983). Two distinct mechanisms regulate the levels of a cellular tumor antigen, p53. Mol. and Cell. Biol., 12: 2143-2150.
- RIFKIN, D.B. & COMPANS, R.W. (1971). Identification of the spike proteins of Rous sarcoma virus. Virology, 46: 485-489.
- ROBBINS, K.C., ANTONIADES, H.N., DEVARE, S.G., HUNKAPILLER, M.W.
  & AARONSON, S.A. (1983). Structural and immunological similarities between simian sarcoma virus gene product(s) and human platelet-derived growth factor.
  Nature, 305: 605-608.
- ROBBINS, K.C., LEAL, F., PIERCE, J.H., AARONSON, S.A. (1985). The v-sis/PDGF-2 transforming gene product localizes to cell membranes but is not a secretary protein.
  EMBO J., 4: 1783-1792.

- ROGERS, J., EARLY, P., CARTER, C., CALAME, K. & BOND, M. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell, 20: 303-312.
- ROGERS, J. & WALL, R. (1980). A mechanism for RNA splicing.
  Proc. Natl. Acad. Sci. USA, <u>77</u>: 1877-1879.
- ROSENFELD, M.G., AMARA, S.G., ROOS, B.A., ONG, E.S. & EVANS, R.M. (1981). Altered expression of the calcitonin gene associated with RNA polymorphism. Nature, 290: 63-65.
- ROSENFELD, M.G., LIN, C.R., AMARA, S.G., STOLARSKY, L., ROOS, B.A., ONG, E.S. & EVANS, M. (1982). Calcitonin mRNA polymorphism : peptide switching associated with alternative splicing events. **Proc. Natl. Acad. Sci. USA**, <u>79</u>: 1717-1721.
- ROSSON, D. & REDDY, P. (1986). Nucleotide sequence of chicken c-myb complementary DNA and implications for myb oncogene activation. Nature, <u>319</u>: 604-606.
- ROSSON, D., DUGAN, D. & REDDY, E.P. (1987). Aberrant splicing events that are induced by proviral integration : implications of myb oncogene activation. **Proc. Natl. Acad. Sci., USA**, <u>84</u>: 3171-3175.
- ROUSSEL, M., SAULE, S., LAGROU, C., ROMMENS, C., BEUG, H., GRAF, T. & STEHELIN, D. (1979). Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. Nature, <u>281</u>: 452-455.
- ROWLEY, J. (1983). Human oncogene locations and chromosome aberrations. Nature, 301: 290-291.
- RUCKLE, G. (1958). Studies with the monkey intranuclear agent (MINIA) and foamy-agent derived from spontaneously degenerating monkey kidney cultures. I isolation and tissue culture behavior of the agents and identification of MINIA as closely related to measles virus. Arch. Gesamte Virusforsch., 8:139-166.
- RULEY, H.E., MOOMAW, J.F. & MARUYAMA, K. (1984). Avian myelocytomatosis virus myc and adenovirus early reigon 1A promote the in vitro establishment of cultured primary cells. Cancer Cells, 2: 481-486.
- RUSHLOW, K.E., LAUTENBERGER, I.A., PAPAS, T.S., BALUDA, M.A., PERBAL, B., GHIRIKJIAN, L.G. & REDDY, E.P. (1982). Nucleotide sequence of the transforming gene of avian myeloblastosis virus. Science, <u>216</u>: 1421-1423.

- SACCHI, N., WATSON, D.K., VAN KESSEL, A.H.M.G., HAGEMEIJER, A., KERSEY, J., DRABKIN, H.D., PATTERSON, D. & PAPAS, T.S. (1986). Hu-ets-1 and Hu-ets-2 genes are transposed in acute leukemias with (4; 11) and (8; 21) translocations. Science, 231: 379-382.
- SAKAMOTO, H., MORI, M., TAIRA, M., YOSHIDA, T., MATSUKAWA, S., SHIMIZU, K., SEKIGUCHI, M., TERADA, M. & SUGIMURA, T. (1986). Transforming gene from human stomach cancers and a non cancerous portion of stomach mucrosa. **Proc. Natl. Acad. Sci., USA**, 83: 3997-4001.
- SAMARUT, J. & BOUABDELLI, M. (1980). In vitro development of CFU-E and BFU-E in cultures of embryonic and post-embryonic chicken hematopoietic cells. J. Cell. Physiol., 105: 553-565.
- SANGER, F.S., NICKLEN, S. & COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors.
   Proc. Natl. Acad. Sci., USA, <u>74</u>: 5463-5467.
- SANTOS, E., TRONICK, S.R., AARONSON, S.A., PULCIANI, S. & BARBACID, M. (1982). T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB and Harvey-MSV transforming genes. Nature, 298: 343-347.
- SASSONE-CORSI, P., LAMPH, W.W., KAMPS, M. & VERMA, M. (1988). fos-associated cellular p39 is related to nuclear transcription factor AP-1. Cell, <u>54</u>: 553-560.
- SAP, J., MUNOZ, A., DAMM, K., GOLBERG, Y., GHYSDAEL, J., LEUTZ, A., BEUG, H. & VEENSTROM, B. (1986). The c-erbA protein is a high-affinity recepteur for thyroid hormone. Nature, 324: 635-640.
- SAULE, S., COLL, J., RIGHI, M., LAGROU, C., RAES, M.B. & STEHELIN, D. (1983). Two different types of transcription for the myelocytomatosis viruses MH2 and CMII. EMBO J., <u>2</u>: 805-809.
- SCHECHTER, A.L., STERN, D.F., VAIDYANATHAN, L., DECKER, S.J., DREBIN, J.A., GREENE, M.I. & WEINBERG, R.A. (1984). The neu oncogene : An erb-B related gene encoding a 185 000-M tumor antigen. Nature, <u>312</u>: 513-516.
- SEEBURG, P.H., COLBY, W.W., CAPON, D.J., GOEDDEL, D.V. & LEVINSON, A.D. (1984). Biological properties of human c-Ha-ras genes mutated at codon 12. Nature, <u>312</u>: 71-75.

- SEFTON, B.M., TROWBRIDGE, I.S., COOPER, J. & SCOLNICK, E.M. (1982). The transforming proteins of Rous sarcoma virus, Harvey sarcoma virus and Abelson virus contain tightly bound lipid. Cell, <u>31</u>: 465-474.
- SETZER, D.R., Mc GROGAN, M. & SCHIMKE, R.T. (1982). Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductive gene.
   J. Biol. Chem., 257: 5143-5147.
- SEUWEN, K., LAGARDE, A. & POUYSSEGUR, J. (1988). Deregulation of hamster fibroblast proliferation by mutated **ras** oncogenes is not mediated by constitutive activation of phosphoinositide-specific phospholipase C. **EMBO J.**, <u>7</u>: 161-168.
- SHANK, P.R. & VARMUS, H.E., (1978). Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closely circular viral DNA in the nucleus. J. Virol., 25: 104-114.
- SHANK, P.R., & LINIAL, M. (1980). Avian oncornavirus mutant (SE21Q1b) deficient in genomic RNA : characterization of a deletion in the provirus. J. Virol., <u>36</u>: 450-456.
- SHARP, D.G. & BERNHARD, J.W. (1952). Counts of virus particles by sedimentation on agar and electron micrography. **Proc. Natl. Acad. Sci. USA**, <u>81</u>: 75-79.
- SHARP, P.A. (1987a). Splicing of messenger RNA precursors. Science, 235: 766-771.
- SHARP, P.A. (1987b). Speculations on RNA splicing. Cell, 23:643-646.
- SHAW, J., HAYMAN, M.J. & ENRIETTO, P.J. (1985). Analysis of a deleted MC29 provirus : gag sequences are not required for fibroblast transformation.
  J. Virol., 56: 943-950.
- SHEINESS, D. & GARDINIER, M. (1984). Expression of a proto-oncogene (proto-myb) in hemapoietic tissues of mice. Mol. Cell. Biol., <u>4</u>: 1206-1212.
- SHEN-ONG, G.L.C. (1987). Alternative internal splicing in c-myb RNAs occurs commonly in normal and tumor cells.
  EMBO J., 6: 4035-4039.

- SHERR, C.J., RETTENMIER, C.W., SACCA, R., ROUSSEL, M.F., LOOK, A.T. & STANLEY, E.R. (1985). The c-fms proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF-1. Cell, 41: 665-676.
- SHIMIZU, K., BIRNBAUM, D., RULEY, M.A., FASANO, O., SUARD, Y., EDLUND, L., TAPAROWSKY, E., GOLDFARD, M. & WIGLER, M. (1983). Structure of the **Ki-ras** gene of the human lung carcinoma cell line Calu-1. **Nature**, <u>304</u>: 497-500.
- SHIMIZU, K., NAKATSU, Y., SEKIGUCHI, M., HOKAMURA, K., TANAKA, K., TERADA, M. & SUGIMURA, T. (1985). Molecular cloning of an activated human oncogene, homologous to v-raf, from primary stomach cancer. Proc. Natl. Acad. Sci., USA, <u>82</u>: 5641-5645.
- SHTIVELMAN, E., LIFSHITZ, B., GALE, R.P. & CANAANI, E. (1985). Fused transcript of **abl** and **bcr** gene in chronic myelogenous leukemia. **Nature**, <u>315</u>: 550-554.
- SHTIVELMAN, E., LIFSHITZ, B., GALE, R.P., ROE, R.A. & CANAANI, E. (1986). Alternative splicing of RNAs transcribed from the human abl gene and from the bcr-abl fused gene.
  Cell, <u>47</u>: 277-284.
- SIMEK, S. & RICE, N.R. (1988). Detection and characterization of the protein encoded by the chicken c-rel proto-oncogene. Oncogene Res., 2:103-119.
- SLAMON, D.J., BOONE, T.C., MURDOCK, D.C., KEITH, D.E., PRESS, M.F., LARSON, R.A. & SOUZA, L.M. (1986). Studies of the human c-myb gene and its product in human acute leukemias. Science, 233: 347-351.
- SODROSKI, J.G., ROSEN, C.A. & HASELTINE, W.A. (1984b). Trans-acting transcriptional activation of the long terminal repeat of human T lymphotropic viruses in infected cells. Science, 225: 381-385.
- SOLNICK, D. (1985). Trans splicing of mRNA precursors. Cell, <u>42</u>: 157-164.
- SOLNICK, D. (1985). Alternative splicing caused by RNA secondary structure. Cell., <u>43</u>: 667-676.

- STAHL, M.L., FERENZ, C.R., KELLEHER, K.L., KRIZ, R.W. & KNOPK, J.L., (1988). Sequence similarity of phospholipase C with the noncatalytic region of src. Nature, 332: 269-272.
- SOMPAYRAC, L.M. & DANNA, K.J. (1981). Efficient infection of monkey cells with DNA of simian virus 40.
   Proc. Natl. Acad. Sci., USA, 78: 7575-7578.
- SORGE, J., RICCI, W. & HUGHES, S.H. (1983). cis-acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus.
   J. Virol., <u>48</u>: 667-675.
- SPINDLER, K.R. & BERK, A.J. (1984). Rapid intracellular turnover of adenovirus 5 early region 1A proteins.
   J. Virol., <u>52</u>: 706-710.
- STAVNEZER, E., GERHARD, D.S., BINARI, R.C. & BALAYES, I. (1981). Generation of transforming viruses in cultures of chicken fibroblasts infected with an avian leukosis virus.
   J. Virol., 39: 920-934.
- STERN, J.B. & SMITH, K.A. (1986). Interleukin-2 induction of T-cell G1 progression and c-myb expression. Science, 233: 203-206.
- STILES, G.E., BITTLE, J.L. & CABASSO, V.J. (1964). Comparison of simian foamy virus strains including a new serological type. Nature, <u>201</u>: 1350-1351.
- STOBER-GRASSER, U. & LIPSICK, J.S. (1988). Specific amino acid susbtitutions are not required for transformation by v-myb of avian myeloblastosis virus.
   J. Virol., 62: 1093-1096.
- STRUBIN, M., LONG, E.O. & MACH, B. (1986). Two forms of the Ia antigen-associated invariant chain result from alternative initiations at two in phase AUGs. Cell., 47: 619-625.
- SUDOL, M. & HANAFUSA, H. (1986). Cellular proteins homologous to the viral yes gene product.
  Mol. Cell. Biol., <u>6</u>: 2839-2846.
- SUGIMOTO, Y., WHITMAN, M., CANTLEY, L.C. & ERIKSON, R.L. (1984). Evidence that the Rous sarcoma virus transforming gene product phosphorylates phosphotidylinositol and diacylglycerol. **Proc. Natl. Acad. Sci.USA**, <u>81</u>: 2117-2121.
- SWEET, R.W., YOKOYAMA, S., KAMATA, T., FERAMISCO, J.R., ROSENBERG, M. & CROSS, M. (1984). The product of ras is a GTPase and the T24 oncogenic mutant is deficient in this activity. Nature, <u>311</u>: 273-275.

- TABIN, C.J., BRADLEY, S.M., BARGMANN, C.I., WEINBERG, R.A., PAPAGEORGE, A.G., SCOLNICK, E.M., DHAR, R., LOWY, D.R. & CHANG, E.H. (1982). Mechanism of activation of a human oncogene. Nature, <u>300</u>: 143-149.
- TAKAHASHI, M., RITZ, J. & COOPER, G.M. (1985). Activation of a novel human transforming gene, ret, by DNA rearrangement. Cell., <u>42</u>: 581-588.
- TAKEYA, T. & HANAFUSA, H. (1983). DNA sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. Cell, 32:881-890.
- TAPAROWSKY, E., SUARD, Y., FASANO, O., SHIMIZU, J., GOLDFARB, M. & WIGLER, M. (1982). Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. Nature, 300: 762-765.
- TAPAROWSKY, E., SHIMIZU, K., GOLDFARB, M. & WIGLER, M. (1983). Structure and activation of the human **N-ras** gene. **Cell**, <u>34</u>: 581-586.
- TEMELES, G.L., GIBBS, J.B., D'ALONZO, J.S., SIGAL, I.S. & SCOLNICK, E.M. (1985). Yeast and mammalian ras proteins have conserved biochemical properties. Nature, 313: 700-703.
- TEMIN, H.M. & MIZUTAMI, S. (1970). RNA directed DNA polymerase in virions of Rous sarcoma virus. Nature, 226: 1211-1213.
- TEMIN, H.M. & BALTIMORE, D. (1972). RNA directed DNA synthesis and RNA tumor viruses. Adv. Virus. Res., <u>17</u>: 129-186.
- THOMPSON, C.B., CHALLONER, P.B., NEIMAN, P.E. & GROUDINE, M. (1986). Expression of the c-myb proto-oncogene during cellular proliferation. Nature, 319: 374-380.
- TODA, T., UNO, I., ISHIKAWA, T., POWERS, S., KATAOKA, T., BROEK, D., CAMERON, S., BORACH, J., MATSUMOTO, K. & WIGLER, M. (1985). In yeast, <u>RAS</u>, proteins are controlling elements of adenylate cyclase. Cell, <u>40</u>: 27-36.

- ULLRICH, A., COUSSENS, L., HAYFLICK, J.S., DULL, T.J., GRAY, A., TAM. A.W., LEE, J., YARDEN, Y., LIVERMANN, T.A., SCHLESSINGER, J.S., DOWNVARD, J., MAYES, E.L.V., WHITTLE, N., WATERFIELD, M.D. & SEEBURG, P.H. (1984). Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature, 309: 418-425.
- VAN BEVEREN, C., VAN STRAATEN, F., CURRAN, T., MULLER, R. & 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.
- VARMUS, H.E., SHANK, P.R., HUGHES, S.E., KUNG, H.J., HEASLY, S., MAJORS, J., VOGT, P.K. & BISHOP, J.M. (1979). Synthesis, structure and integration of the DNA of RNA tumor viruses. Cold Spring Harbor - Symp. Quant. Biol., <u>43</u>: 851-864.
- VERMA, I.M. (1977). The reverse transcriptase. Biochim. Biophys. Acta., <u>473</u>: 1-38.
- VOGT, P.K. (1977). Genetics of RNA viruses.
  In comprehensive virology (ed. H. Fraenkel-Conrat and R. Wagner), <u>9</u>: 341-455. Plenum Press. New York.
- VOGT, P.K., BOS, T.J. & DOOLITTLE, R.F. (1987). Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein coded for by the oncogene jun. Proc. Natl. Acad. Sci., USA, 84: 3316-3319.
- VON DER HELM, K. (1977). Cleavage of Rous sarcoma viral poly protein precursor into internal structural proteins in vitro by viral protein p15.
   Proc. Natl. Acad. Sci., USA, <u>74</u>: 911-915.
- WATERFIELD, M.D., SCRACE, G.T., WHITTLE, N., STROOBANT, P., JOHNSON, A., WASTESON, A., WESTERMARK, B., HELDIN, C.H., HUANG, J.S. & DEUEL, T.F. (1983). Platelet derived growth factor is structurally related to the putative transforming protein p28<sup>sis</sup> of simian sarcoma virus. Nature, 304: 35-39.
- WATSON, D.K., Mc WILLIAMS-SMITH, M.J., NUNN, M.F., DUESBERG, P.H., O'BRIEN, S.J. & PAPAS, T.S. (1985). The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. **Proc. Natl. Acad. Sci., USA,** <u>82</u>: 7294-7298.

- WATSON, D.K., Mc WILLIAMS-SMITH, M.S., KOZAK, C., REEVES, R., GEARHART, J., NUNN, M.F., NASH, W.W., FORWLEIII, J.R., DUESBERG, P., PAPAS, T.S. & O'BRIEN, S.J. (1986). Conserved chromosomal positions of dual domains of the ets proto-oncogene in cats, mice and humans. Proc. Natl. Acad. Sci., USA, 79: 2194-2198.
- WEINBERG, R.A. (1981). Use of transfection to analyze genetic information and malignant transformation. Biochim. Biophys. Acta., 651:25-35.
- WEINBERGER, C., THOMPSON, C.C., ONG, E.S., LEBO, R., GRUOL, D.J. & EVANS, M. (1986). The c-erb, a gene encodes a thyroid hormon receptor. Nature, <u>324</u>: 641-646.
- WEISS, S.R., VARMUS, H.E. & BISHOP, J.M. (1977). The size and genetic composition of virus-specific RNAs in the cytoplasm of cells producing avian sarcoma-leukosis viruses. Cell, 12:983-992.
- WESTIN, E.H., GALLO, R.C., ARYA, S.K., EVA, A., SOUZA, L.M., BALUDA, M.A., AARONSON, S.A. & WONG-STAAL, F. (1982). Differential expression of the **amv** gene in human hematopoietic cells. **Proc. Natl. Acad. Sci., USA,** <u>79</u>: 2194-2198.
- WEISS, R., TEICH, N., VARMUS, M. & COFFIN, J. (1982, 1985). RNA tumor viruses. Eds Cold Spring Harbor Laboratory.
- WEISSMANN, C., PARSONS, J.T., COFFIN, J.W., RYMO, L., BILLETER, M.A. & HOFSTETTER, H. (1975). Studies on the structure and synthesis of Rous sarcoma virus RNA. Cold Spring Harbor Symp., Quant. Bio., <u>39</u>: 1043-1056.
- WITTE, O.N. & BALTIMORE, D. (1978). Relationship of retrovirus polyprotein cleavages to virion maturation studied with temperaturesensitive murine leukemia virus mutants. J. Virol., 26: 750-761.
- WOOD, T.G., Mc GEADY, M.L., BAROUDY, B.M., BLAIR, D.G. VAN DE WOOD, G.F. (1984). Mouse c-mos oncogene activation is prevented by upstream sequences. Proc. Natl. Acad. Sci., USA, <u>81</u>: 7817-7821.
- YOUNG, D., WAITCHES, G., BIRCHMEIER, C., FASANO, O. & WIGLER, M. (1986). Isolation and characterization of a new oncogene encoding a protein with multiple potential cellular transmembrane domains. Cell., 45: 711-719.
- YU, C., TSAI, M., STACEY, D.W. (1988). Cellular ras activity and phospholipid metabolism. Cell., <u>52</u>: 63-71.
- YUNIS, J.J. (1983). The chromosomal basis of human neoplasia. Science, <u>221</u>: 227-236.

# ANNEXES

# Multiple Domains for the Chicken Cellular Sequences Homologous to the v-*ets* Oncogene of the E26 Retrovirus

ANNE GEGONNE,<sup>1</sup> DOMINIQUE LEPRINCE,<sup>1</sup> MARTINE DUTERQUE-COQUILLAUD,<sup>2</sup> BERNARD VANDENBUNDER,<sup>1</sup> ANNE FLOURENS,<sup>1</sup> JACQUES GHYSDAEL,<sup>1</sup> BRIGITTE DEBUIRE,<sup>2</sup> AND DOMINIQUE STEHELIN<sup>1</sup>\*

Laboratoire d'Oncologie Moléculaire, Unité 186 Institut National de la Santé et de la Recherche Médicale-Centre National de la Recherche Scientifique 1160, Institut Pasteur, 59019 Lille Cédex,<sup>1</sup> and Unité 124 Institut National de la Santé et de la Recherche Médicale, Institut de Recherches sur le Cancer, 59045 Lille,<sup>2</sup> France

Received 11 August 1986/Accepted 28 October 1986

We have investigated the structure of chicken genomic DNA homologous to v-ets, the second cell-derived oncogene of avian retrovirus E26. We isolated a c-ets locus spanning ca. 30.0 kilobase pairs (kbp) in the chicken genome with homologies to 1,202 nucleotides (nt) of v-ets (total length, 1,508 nt) distributed in six clusters along 18.0 kbp of the cloned DNA. The 5'-distal part of v-ets (224 nt) was homologous to chicken cellular sequences contained upstream within a single 16.0-kbp *Eco*RI fragment as two typical exons but not found transcribed into the major 7.5-kb c-ets (or 4.0-kb c-myb) RNA species. Between these two v-ets-related cellular sequences we found ca 40.0 kbp of v-ets-unrelated DNA. Finally, the most 3' region of homology to v-ets in the cloned DNA was shown to consist of a truncated exon lacking the nucleotides coding for the 16 carboxy-terminal amino acids of the viral protein but colinear to one of the two human c-ets loci, c-ets-2.

Avian retrovirus E26 induces a mixed erythroid-myeloid leukemia predominantly of the erythroid type (13). In addition to v-myb, the genome of E26 contains a second stretch of cell-derived specific sequences, v-ets (1,508 nucleotides [nt]) 3, 16, 19). The cellular counterpart of v-ets, c-ets, is clearly distinct from c-myb since it is transcribed in some normal cell types as a major 7.5-kilobase (kb) polyadenylated RNA (as well as two minor 2.2- and 1.5 (2.0)-kp species) in contrast to the 4.0-kb c-myb mRNA (16, 27). Furthermore, the human cellular genes c-myb and c-ets have been assigned to different chromosomes: 6q22-24 for c-myb (14), 11q23-24 for c-ets-1 (8, 26), and 21q22.1-22.3 for c-ets-2 (26). The E26 viral transforming protein is a 135,000-Mr fusion protein, P135gag-myb-ets translated from the 5.7-kb genomic viral RNA and located in the nucleus of E26-transformed cells (2, 4, 5, 15). v-ets is not related to any other known viral oncogenes but its deduced amino acid sequence displays homology with the products of yeast cell cycle genes CDC4 and CDC36 (20).

In this report we demonstrate that the v-ets sequences originated mainly from two stretches of DNA 40.0 kbp distant. We isolated lambda recombinant phages corresponding to a c-ets locus transcribed as a major 7.5-kb mRNA and containing six exons homologous to 1,202 nt of a central portion of v-ets. In addition, the 5' part of v-ets was acquired from two exons located in a 16.0-kbp EcoRI fragment of chicken DNA which lies within the c-ets locus described here. These two exons do not appear to be transcribed into the major 7.5-kb c-ets RNA species seen in several cell types. Finally, by sequencing analysis of the most 3' cellular sequence homologous to v-ets, we identified a putative translation stop codon to this c-ets locus, colinear with that of the human c-ets-2 locus but distinct from the one used for the termination of the E26 viral protein. Thus, v-ets appears to represent a mosaic of probably at least three stretches of DNA.

## MATERIALS AND METHODS

**Preparation of DNAs.** High-molecular-weight chicken DNA was prepared from total embryos as previously described (23). Plasmid DNA and recombinant phage DNAs were prepared as described previously (22).

Hybridization probes. Lambda E26Q1 is a molecularly cloned E26 provirus. A 2.5-kbp EcoRI fragment containing the v-ets domain flanked 5' and 3' by E26 v-myb and env sequences, respectively, was subcloned in plasmid pKH47 to yield the plasmid clone pVE2.5 (Fig. 1) (16). The different v-ets and v-myb probes used in this study were prepared from the pVE2.5 clone.

Fragments representative of the viral *env* gene were derived from the Prague A strain of Rous sarcoma virus (Pr-RSVA) genome as illustrated by Saule et al. (22). These fragments were labeled by nick-translation with New England Nuclear Corp. nick translation kits.

Isolation of recombinant phages containing c-ets chicken sequences. A library of chicken erythrocyte DNA fragments in Charon 4A vectors was kindly provided by J. Dodgson and J. Engel (9). A total of  $2.5 \times 10^5$  plaques were screened with a v-ets probe by the in situ plaque purification procedure of Benton and Davis (1).

A recombinant DNA library in vector EMBL4 constructed with a partial Sau3A digest of total chicken embryo DNA by A. Begue in our lab was used to isolate lambda clones containing sequences homologous to the 5' part of v-ets.

Restriction mapping and gel electrophoresis of poly(A)containing cellular RNAs. Restriction endonucleases were obtained from Boehringer Mannheim and Bethesda Research Laboratories. Digested DNA fragments were size separated by electrophoresis in horizontal agarose gels and transferred to nitrocellulose filters by the method of Southern (24). Total cellular RNA was fractionated on oligo(dT)cellulose (T3; Collaborative Research). Polyadenylated RNAs were denatured, size separated, and transferred to nitrocellulose by the method of Thomas (25).

<sup>\*</sup> Corresponding author.



FIG. 1. (Top) Schematic drawing of the E26 provirus deduced from restriction endonuclease mapping and partial sequencing analysis of recombinant phages isolated from DNA libraries of two distinct E26-infected cells (22, 26, 28). Abbreviations: B, BamHI; E, EcoRI; H, HindIII; S, SstI; X, XhoI. gag, env, and LTR (long terminal repeat) refer to helper virus-related sequences (13). whereby myb and ets refer to the two cell-derived sequences present in the E26 genome (16, 19). The arrow indicates the 3.6-kbp of coding capacity required for the synthesis of the fusion protein P135Rag ers (4, 5, 15). (Middle) Restriction endonuclease map of the 2.5-kbp EcoRI fragment subcloned into plasmid pKH47 (pVE2.5). This fragment contains all the v-ets sequences flanked 5' and 3' by E26 v-myb and env sequences, respectively. Abbreviations: E, EcoRI; H, HindIII; P, PstI; N, NdeI; R, RsaI. S\* represents a convenient Sau3A restriction site located at the E26 myb-ets junction. The stippled area represents the regions of homology found between v-ets and the yeast cell cycle genes CDC4 and CDC36 (20). (Bottom) The different v-ets probes (A, B, C, and D) derived from the pVE2.5 subclone are indicated. After digestion of chicken chromosomal DNA by BamHI (lanes 1, 4, and 8), HindIII (lanes 2, 5, and 9), or EcoRI (lanes 3, 6, 7, 10, and 11) the DNA fragments were separated by electrophoresis in the same 0.8% agarose gel and transferred onto nitrocellulose sheets (24). These Southern blots were hybridized with the v-ets probes indicated above: probe A (lanes 1 to 3), probe B (lanes 4 to 6), probe C (lanes 8 to 10), and probe D (lane 11). Similarly, the chicken DNA was hybridized with an env probe (lane 7), since probes B and D were shown to contain such sequences (19).

Hybridization of blots to  ${}^{32}$ P-labeled DNA, washing, and autoradiography at  $-70^{\circ}$ C with Kodak X-ray films and Du Pont Lightning-Plus X-ray intensifying screens were performed as described previously (22).

Heteroduplex mapping. The lambda chicken c-ets DNAs were hybridized with either E26 recombinant phages containing the v-ets sequence in the same orientation versus the lambda vector arms: lambda E26-1, which has been described (16), or lambda E26-2 containing the 9.0-kbp BamHI fragment from lambda E26-1 but cloned in the opposite orientation into the lambda 47 vector. The heteroduplex molecules spread onto the hypophase were picked up on the parlodion-coated grids (300 mesh) and stained with uranyl acetate. The grids were subsequently dried, rotary shadowed, and viewed in a Hitachi HU12 electron microscope.

Sequencing analysis. The sequence of the EcoRI-PstI78-base-pair (bp) fragment was obtained by the dideoxy chain termination method (21) after cloning into the polylinker restriction sites of the M13 phage derivatives mp8 and mp9 and by the incorporation of [<sup>32</sup>P]deoxynucleotide triphosphates (dNTPs) onto a synthetic primer complementary to the terminus of the M13 vector in the presence of dideoxy dNTPs as described by Messing et al. (17).

# RESULTS

Restriction analysis of c-ets in chicken DNA. The probes used to analyze the restriction pattern of chicken c-ets were obtained by digestion with the restriction endonuclease RsaI of the plasmid subclone pVE2.5 containing all the v-ets domain (Fig. 1). This yielded two fragments containing roughly the 5' half of v-ets (768 bp, probe A) and the 3' half of v-ets (765 bp) together with 105 bp related to helper virus env sequences (870 bp, probe B) (19). These two probes hybridized to several DNA fragments in Southern blots of chicken DNA digested with either EcoRI, BamHI, or HindIII. After EcoRI digestion, 16.0-, 6.0- and 1.8-kbp EcoRI fragments hybridized with the 5' v-ets probe (Fig. 1, lane 3), whereas the 3' v-ets probe detected 8.0-, 7.5-, 4.6-, 3.9- (faint), and 1.8-kbp EcoRI fragments (Fig. 1, lane 6). However, since normal chicken DNA is known to contain endogenous retroviral information, we analyzed the same DNA sample with an env probe to differentiate the c-ets bands from the endogenous env sequences. Only the 3.9-kbp EcoRI fragment (hybridizing faintly with probe B) corresponded to env sequences (Fig. 1, lanes 6 and 7). Southern blots of BamHI- and HindIII-digested chicken DNA also hybridized with probe A (Fig. 1, lanes 1 and 2), probe B (lanes 4 and 5), or an env probe (data not shown). Finally, with a convenient Sau3A restriction site located at the v-myb-v-ets junction (19), we prepared probe C (Sau3A-PstI fragment, 135 bp) specific to the 5' end of v-ets. This probe detected the 16.0-kbp EcoRI fragment (Fig. 1, lane 10) as well as the 7.5-kbp BamHI and the 5.4- and 1.9-kbp HindIII fragments (Fig. 1, lanes 8 and 9) already visualized by probe A (Fig. 1, lanes 1 and 2). Similarly, a <sup>32</sup>P-labeled PstI-NdeI 194-bp fragment (probe D) containing the 3' end of v-ets and some env (78 nt) sequences detected in total DNA the 4.6and 3.9-kbp EcoRI fragments, the latter probably representing env-related endogenous sequences (Fig. 1, lanes 7 and 11).

In conclusion, the v-ets probes detected a small number of discrete bands in *Eco*RI-digested chicken DNA.

Cloning of c-ets from a library of chicken DNAbacteriophage recombinants. To gain information about the arrangement of the chicken DNA fragments homologous to v-ets, molecular clones of c-ets were obtained from a chicken erythrocyte DNA library (9). With probes A and B (Fig. 1), three such recombinant clones (lambda c-ets 3, 4, and 5 in Fig 2A) were isolated and studied by restriction enzyme digestion and Southern blot analysis. Results (Fig. 2A) showed the following: (i) the three recombinant clones, lambda c-ets 3, c-ets 4, and c-ets 5, contained different overlapping cellular inserts; (ii) none of these recombinant phages hybridized with the 5'-most v-ets probe (probe C) or



B



ets

myb

0.5 Kbp

env

contained part of the 16.0-kbp EcoRI fragment detected in cellular DNA by this probe (Fig. 1, lane 10); and (iii) lambda c-ets 4 and c-ets 5 hybridized with the 3' v-ets probe (probe D in Fig. 1) and contained the 4.6-kbp EcoRI fragment (Fig. 1, lane 11). These 3' v-ets-homologous sequences were flanked downstream by 8.0 kbp of v-ets-unrelated chicken DNA in lambda c-ets 5. Moreover, the regions of homology to v-ets in lambda c-ets 3 and c-ets 4 DNAs were precisely localized by heteroduplex analysis; six small double-stranded regions separated by five internal loops could be defined (Fig. 2B).

Incidentally, we also isolated from the same DNA library a recombinant phage named lambda c-ets 4\* which seemed to overlap with both lambda c-ets 3 and lambda c-ets 4 since each  $^{32}P$ -labeled *Eco*RI fragment of lambda c-ets 4\* crosshybridized with a corresponding fragment in one of the two clones (Fig. 2A). However, lambda c-ets 4\* and c-ets 4 contained two *Eco*RI fragments different in size (7.5 and 8.0 kbp) which were both detected by a v-ets probe in the chicken DNA used in the experiments shown in Fig. 1. This is probably due to allelic variation in this region of c-ets, since analysis of DNA obtained from seven individual chicken embryos revealed either both *Eco*RI fragments or only the 8.0-kbp *Eco*RI fragment (Fig. 2A).

In conclusion, cellular sequences corresponding to the major part of v-ets (1,202 nt; see below) are split into six regions of homology (presumably exons), interrupted by five regions of nonhomology (introns) within ca. 18.0 kbp of chicken DNA (Fig. 2) cloned in lambda c-ets 3, 4, and 5. It should be noted that both the ets-1 and ets-2 sequences found split on different chromosomes in human, cat, and mouse DNA (26) can be found contiguous in chicken cellular DNA.

Cloning of the cellular sequences homologous to the 5' part of v-ets. From a newly constructed library of Sau3A-digested chicken embryo DNA, we isolated two recombinant phages (named lambda c-ets 1 and lambda c-ets 2) hybridizing with the 5'-specific v-ets probe C. Lambda c-ets 1, although containing a single 11.0-kbp artificial EcoRI fragment instead of the 16.0-kbp EcoRI fragment (Fig. 1, lane 10), contained the 7.5-kbp BamHI as well as the 5.4- and 1.9-kbp HindIII fragments detected in total chicken DNA (Fig. 1, lanes 8 and 9, and Fig. 2C). These HindIII fragments contained cellular sequences, named  $\alpha$  and  $\beta$ , accounting for the 5' part of v-ets. Nucleotide sequence analysis revealed that  $\alpha$  and  $\beta$ exhibited the typical properties of eucaryotic exons, with open reading frames (82 and 142 nt, respectively) identical to that used by E26 virus and flanked by consensus splice signals (Fig. 3) (18). Nevertheless, we found no splice acceptor sequence at the leftward boundary of  $\alpha$  or upstream sequence homology to v-myb. Thus,  $\alpha$  represents an exon probably truncated during an illegitimate recombination with myb sequences. Appropriate splicing between  $\alpha$  and  $\beta$  would yield the most 5' sequences of v-ets. To determine more precisely the relationship between the two genomic regions homologous to v-ets, we also sequenced exon a, the first v-ets-related exon of lambda c-ets 3 (Fig. 2). From the results shown in Fig. 3, it appears that the most obvious mechanism to generate the v-ets sequences involves a consensus splicing of  $\beta$  to a.

These results as well as the fact that lambda c-ets 1 and lambda c-ets 3 did not contain overlapping DNA inserts raised the question whether they represented two distinct regions of the chicken genome or distant parts of a same locus. We thus extensively screened chicken DNA libraries with <sup>32</sup>P-labeled fragments of lambda c-ets 1 and lambda c-ets 3. Using in addition the "walking gene" method, we isolated five recombinant clones (lambda c-ets A, B, C, D, and E) containing v-ets-unrelated inserts totaling ca. 40 kbp which represented the junction between the two v-ets homologous cellular stretches (Fig. 2). The transcription of c-ets in chicken cells as a major 7.5-kb RNA species (16, 27) most likely correlates with a large cellular gene. E26 could therefore have transduced and joined in its genome two sequences, namely those found in lambda c-ets 1 and c-ets 3, that may belong to a same locus but are separated by a large region (40.0 kbp) of cellular DNA. To test this hypothesis, polyadenylated RNAs of MSB1 cells (a chicken line of T lymphoblasts transformed by Marek disease virus, expressing a high level of 7.5-kb c-ets RNA [16]) was hybridized with probe C. No signal was observed (Fig. 4, lane C1). The lack of hybridization was significant, since the same blot hybridized with the <sup>32</sup>P-labeled 870-bp Rsal fragment (probe B) revealed the typical c-ets RNA species (Fig. 4, lane B1); furthermore, probe C readily detected the 5.7-kb E26 genomic RNA as efficiently as did probe B (Fig. 4, lanes 2). To bolster this demonstration we analyzed MSB1 mRNAs with probe E (PstI-PstI. fragment, 420 bp) which encompasses myb sequences and probe C; only the 4.0-kb c-myb mRNA species was detected (Fig. 4, lane E1). In addition, when we used probe A, which represents the 5' end of v-ets (including  $\alpha$  and  $\beta$ ), only the c-ets mRNA species described above with probe B were detected (Fig. 4, lane A1). Therefore, the 5' part of v-ets is homologous to chicken cellular sequences which are not transcribed into the major 7.5-kb c-ets RNA species in MSB1 cells. However, we cannot rule out the possibility that the 5' part of v-ets may be expressed by an alternate splicing mechanism as part of a minor c-ets

FIG. 2. Structure of the c-ets locus. (A) Molecular clones of c-ets were obtained from different chicken (CK) DNA libraries: Alul-HaellI-digested erythrocyte DNA, EcoRI-digested fibroblast DNA, and Sau3A-digested total embryonic DNA (see the text). Vertical bars represent true EcoRI fragments, and arrows represent artificial EcoRI fragments due to the construction of some of the libraries used. Fragments containing sequences homologous to v-ets (diagonal lines) in clones lambda c-ets 1, 2, 3, 4, 4\*, and 5 were identified by Southern blot analysis with v-ets probes (data not shown). Recombinant clones A, B, C, D, and E which did not hybridize with v-ets probes were isolated by the walking gene technique. Orientation is 5' to 3'. Chromosomal chicken DNAs obtained from different embryos were restricted with EcoRI. The DNA fragments were separated by agarose gel electrophoresis, transferred to nitrocellulose sheets, and hybridized with the <sup>32</sup>P-labeled 8.0-kbp EcoRI fragment of lambda c-ets 4\*. (B) Heteroduplex analysis of chicken c-ets recombinant clones. Top to bottom: Representative heteroduplex, interpretative sketch (thick lines represent double strands and thin lines represent single strands), and schematic drawing of the contour lengths (features 1 to 5 refer to single strands [loop 3 was 3200 = 300] and features a to f refer to double strands, whose lengths in base pairs are summarized in the table). (C) Restriction map of a c-ets chicken locus. The sites of cleavage of the following restriction endonucleases into lambda c-ets 1, 3, and 4 are shown: BamHI (B), EcoRI (E), and HindIII (H). In addition the position of one HpaI site (Hp\*) is shown as it is also found in the viral v-ets sequence. The 5' and 3' limits of the homology found between v-ets and the second c-ets domain have been roughly determined by heteroduplex analyses. These studies revealed six regions of homology (a to f) separated by five regions of nonhomology (1 to 5) (see the table). The hatched box represents the most 3' sequences of v-ets which are not derived from the c-ets exon f and remain of unknown origin. The stippled box represents the homology with CDC4 and CDC36 (20).



v-ets GAG CAC TAC AGC ACA GAC ATG GAA TGT GCA IIIII IIIIII Ck c-etsa CTT CTG GTG TTT TC<u>A G</u>AT ATG GAA TGT GCA **T** Sa

FIG. 3. Nucleotide sequence of  $\alpha$ ,  $\beta$ , and a. Nucleotide sequence of relevant portions from lambda c-ets 1 and c-ets 3 clones is aligned with the published v-ets sequence (19). Three open reading frames (82, 142, and 120 nt) colinear to v-ets (with a single base change at the beginning of exon a) and flanked by typical splice signals (18) with the exception of the leftward boundary of  $\alpha$  (see the text) were detected. The v-ets sequence probably results from splicing of  $\alpha$  and  $\beta$  to a. Ck, Chicken; Sd, splice donor; Sa, splice acceptor.

RNA species present at a very low level in this cell line or expressed in so far unidentified cell types.

v-ets and c-ets differ at their 3' end. Sequencing analysis of v-ets and the topography of the E26 provirus suggested that the same translation stop codon was used for termination of synthesis of the viral protein  $P135^{gag-myb-ets}$  and the cellular c-ets protein  $P54^{c-ets}$  (6, 11, 19). Therefore, we investigated in more detail the cellular sequences homologous to the 3' end of v-ets. From hybridization experiments with total chicken DNA and with chicken DNA recombinant clones (Fig. 1, lanes 6, 7 and 11, and Fig. 2), these sequences should be localized within the 4.6-kbp *Eco*RI fragment of clones lambda c-ets 4 or lambda c-ets 5. A *PstI-Eco*RI 78-bp fragment hybridizing with the 3'-distal v-ets probe D (Fig. 1) was isolated from the 4.6-kbp *Eco*RI fragment of lambda c-ets 4, subcloned into phage M13mp8, and sequenced by the chain termination method (21). This sequence was com-

pared with the previously published v-ets sequence (19). A perfect homology beween the v-ets and the c-ets fragment was found (Fig. 5), but only for the first 34 nt following the PstI site. No significant homology was found with the last 82 nt of v-ets. However, the chicken c-ets sequence described here is colinear, with some minor base changes probably due to phylogenetic divergence, with that of a human c-ets cDNA clone corresponding to human c-ets-2 (Fig. 5) (27). Therefore, the viral v-ets sequence on one hand and the sequences of human c-ets-2 cDNA and chicken c-ets DNA on the other hand appear to specify different carboxy termini for their encoded proteins. It should also be noted that from sequence analysis (Fig. 5), probe D, although containing 78 env-related nt, 34 nt derived from the 4.6-kbp EcoRI fragment of lambda c-ets 4 but also 82 nt of unknown origin (see sequence data in Fig. 5), hybridized only with two cellular EcoRI fragments (4.6 and 3.9 kbp in size) with similar

intensities (Fig. 1, lane 11). A possible explanation could be that the 4.6-kbp band (lane 6) represented a doublet of a first fragment located within the lambda c-ets 4 recombinant clone and derived from the c-ets locus and a second fragment which could contain the termination codon used by E26 for its transforming protein. The nature and origin of this latter fragment remain open questions.

### DISCUSSION

In this paper, we investigate the nature and structure of the cellular sequences homologous to the E26-specific cellderived domain v-ets (1,508 nt). Surprisingly, the cellular chicken sequences accounting for the v-ets domain appeared to be split into two stretches (6.0 kbp and 18.0 kbp, respectively) separated by about 40.0 kbp of v-ets-unrelated DNA. The downstream stretch related to v-ets contained six clusters of homology, named a to f (Fig. 2), scattered along 18.0 kbp of DNA. Therefore, the ets-1 and ets-2 loci which appeared to be located on different chromosomes in several mammalian species, including humans, are found contiguous in chickens, the species from which the E26 virus arose. However, at least two sets of ets-related proteins can be observed in chicken tissues: P54<sup>c-ets</sup>, which is highly related to the v-ets domain of P135<sup>gag-myb-ets</sup> (6, 11), and the P60, P62, and P64 proteins, which have only limited homology with P54<sup>c-cts</sup> and P135<sup>gag-myb-cts</sup> (12). We do not know whether these two sets of ets-related proteins are translated from differentially spliced mRNA transcripts derived from the locus described here or from two distinct loci, one being highly related to v-ets, the other displaying only a small region of homology. The characterization of the nucleotide



FIG. 4. The 5' part of v-ets was not found transcribed into the c-ets or c-myb RNAs. Polyadenylated RNAs of MSB1 (a T-lymphoid cell line) (lanes 1) or of E26-transformed myeloblasts (lanes 2) were denatured with glyoxal-dimethyl sulfoxide, separated by agarose gel electrophoresis, bound to nitrocellulose sheets (25), and hybridized with the indicated probes (C, B, E, and A). See the Legend to Fig. 1 for symbols and abbreviations.



# V-ets AGGGAACATCCTTTCTTTGATAGGGCCTCCAA -

FIG. 5. P135 translational stop codon not derived from c-ets exon f. A 78-nt c-ets fragment that includes the 3' stretch of homology between v-ets and the 4.6-kbp EcoRI fragment lambda c-ets 4 recombinant clone (top) was subjected to nucleotide sequence analysis. Numbering starts at the C residue of a PstI restriction site (nt 1927 in reference 27). Sequencing was done as described by Sanger et al. (21) after cloning of a PstI-EcoRI fragment into phage M13 derivatives mp8 and mp9. \*\*\*, Translation stop codon used for synthesis of the E26 transforming protein. The border with the env gene is also shown. The human c-ets-2 cDNA sequence with its translation stop codon is also indicated (27). Vertical bars refer to homologous nucleotides found within the three sequences. Ck, Chicken; Hu, human.

sequence of the corresponding cDNAs should resolve the issue.

The 5' c-ets stretch contains sequences homologous to the first 224 nt of v-ets. This 5' c-ets stretch, although displaying a split structure typical of a eucaryotic gene (Fig. 3), is not transcribed into the 7.5-kb major RNA species (Fig. 4). This suggests that these sequences could be expressed in minor RNA species present at low levels (just above the limit of sensitivity of the method we used) in the MSB1 cellular RNAs tested or in so far unidentified cell types. However, antisera directed against these sequences corresponding to the amino-terminal v-ets-encoded domain, although identifying the P135 viral protein, failed to precipitate any of the described cellular c-ets proteins in many different cell types (10). Finally, in support of these results, the beginning of homology found between v-ets and the yeast cell cycle genes CDC4 and CDC36 (20) seems to correlate with the upstream part of the second cellular stretch of homology to v-ets defined in this study.

Turning now to the 3' end of v-ets, the topography of E26 proviral DNA as well as sequencing analysis strongly suggested that the translation stop codon used for the synthesis of the viral P135 protein was of cellular origin and derived from c-ets (16, 19). However, sequencing analysis of the cellular sequence homologous to the most 3' v-ets sequence in the 4.6-kbp *Eco*RI fragment of lambda c-ets 4 demonstrated that (i) the last 16 amino acids of the P135<sup>Kag-myb-ets</sup> protein are not derived from this part of c-ets and (ii) the region of homology (named f in Fig. 2C) found between c-ets and v-ets displays features typical of a truncated c-ets exon

#### 812 GEGONNE ET AL.

and is colinear (with some single base changes) to a human c-ets-2 cDNA which contains a translation stop codon. Therefore, this f probably provides the termination codon of at least one of the chicken c-ets proteins, namely P54 (11).

Using specific oligonucleotide probes, we are currently cloning the cellular sequences containing the translation stop codon of the E26 transforming protein and isolating *c-ets* cDNAs. These studies will shed light on the mechanism by which three probably distinct cellular domains have been transduced in E26 virus and fused to encode a single viral protein.

#### ACKNOWLEDGMENTS

We thank J. Coll and S. Saule for stimulating discussions and critical reading of this manuscript; A. Fritsch, G. Torpier, and A. Jacob for help with the heteroduplex analysis; J. Dogson for the chicken erythrocyte DNA library; A. Begue for the Sau3A chicken embryo DNA library; C. Lagrou for expert technical assistance in cell culture; and N. Devassine for patient typing.

This work was supported by the Institut National de la Santé et de la Recherche Médicale (U186-U-124), Centre National de la Recherche Scientifique (UA 04 1160), Association pour la Recherche sur le Cancer, and the Pasteur Institute of Lille.

#### LITERATURE CITED

- 1. Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones to hybridization to single plaques in situ. Science 196:180-182.
- Beug, H., M. S. Hayman, and T. Graf. 1982. Myeloblasts transformed by the avian acute leukemia virus E26 are hormone-dependent for growth and for expression of a putative myb-containing protein, P135 E26. EMBO J. 1:1069-1073.
- 3. Bishop, J. M. 1983. Cellular oncogenes and retrovirus. Annu. Rev. Biochem. 52:301-334.
- 4. Bister, K., M. Nunn, C. Moscovici, B. Perbal, M. A. Baluda, and P. Duesberg. 1982. Acute leukemia viruses E26 and avian myeloblastosis virus have related transformation specific RNA sequences but different genetic structures, gene products and oncogenic properties. Proc. Natl. Acad. Sci. USA 79:3677-3681.
- Boyle, W. S., M. A. Lampert, J. S. Lipsick, and M. A. Baluda. 1984. Avian myeloblastosis virus and E26 virus oncogene products are nuclear proteins. Proc. Natl. Acad. Sci. USA 81:4265– 4269.
- Chen, J. H. 1986. The proto-oncogene c-ets is preferentially expressed in lymphoid cells. Mol. Cell. Biol. 5:2993-3000.
- Coll, J., M. Righi, C. de Taisne, C. Dissous, A. Gegonne, and D. Stehelin. 1983. Molecular cloning of the avian acute transforming retrovirus MH2 reveals a novel cell-derived sequence (v-mil) in addition to the myc oncogene. EMBO J. 2:2189-2194.
- de Taisne, C., A. Gegonne, D. Stehelin, A. Bernheim, and R. Berger. 1984. Chromosomal localization of the human protooncogene *c-ets*. Nature (London) 310:581-583.
- 9. Dodgson, J. B., J. Strommer, and D. Engel. 1979. Isolation of the chicken B globin gene and a linked embryonic B-like globin gene from a chicken DNA recombinant library. Cell 17:879–887.
- Gegonne, A., P. Pognonec, D. Leprince, D. Dernis, E. Remaut, D. Stehelin, and J. Ghysdael. 1986. Preparation and characterization of specific antisera directed against different polypeptidic domains encoded by the v-ets oncogene of the avian acute leukemia virus E26. C. R. Acad. Sci. (Paris) 303:253-256.

- 11. Ghysdael, J., A. Gegonne, P. Pognonec, D. Dernis, D. Leprince, and D. Stehelin. 1986. Identification and preferential expression in thymic and bursal lymphocytes of a *c-ets* protooncogene encoded Mr 54 000 cytoplasmic protein. Proc. Natl. Acad. Sci. USA 83:1714-1718.
- Ghysdael, J., A. Gegonne, P. Pognonec, K. Boulukos, D. Leprince, D. Dernis, C. Lagrou, and D. Stehelin. 1986. Identification in chicken macrophages of a set of proteins related to, but distinct from, the chicken cellular *c-ets* encoded protein P54<sup>c-ets</sup>. EMBO J. 5:2251–2256.
- Graf, T., and D. Stehelin. 1982. Avian leukaemia viruses oncogenes and genome structure. Biochem. Biophys. Acta 651:245– 271.
- Harper, M. E., G. Franchini, J. Love, M. I. Simon, R. C. Gallo, and F. Wong-Staal. 1983. Chromosomal sublocalization of human *c-myb* and *c-fes* cellular onc genes. Nature (London) 304:169-171.
- 15. Klempnauer, K. H., G. Symonds, G. I. Evan, and J. M. Bishop. 1984. Subcellular localization of proteins encoded by oncogenes of avian myeloblastosis virus and avian leukemia virus E26 and by the chicken c-myb gene. Cell 37:537-547.
- Leprince, D., A. Gegonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagrou, and D. Stehelin. 1983. A putative second cell-derived oncogene of the avian leukaemia retrovirus E26. Nature (London) 306:395-397.
- Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309-321.
- Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10:459–472.
- Nunn, M. F., P. H. Seeburg, C. Moscovici, and P. H. Duesberg. 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. Nature (London) 306:391-395.
- Peterson, T. A., J. Yochem, B. Byers, M. F. Nunn, P. H. Duesberg, R. F. Doolittle, and S. I. Reed. 1984. A relationship between the yeast cell cycle genes CDC4 and CDC36 and the *ets* sequence of oncogenic virus E26. Nature (London) 309:556-558.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminator inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- Saule, S., J. Coll, M. Righi, C. Lagrou, M. B. Raes, and D. Stehelin. 1983. Two different types of transcription for the myelocytomatosis viruses MH2 and CMII. EMBO J. 2:805–809.
- Saule, S., A. Sergeant, G. Torpier, M. B. Raes, S. Pfeifer, and D. Stehelin. 1982. Subgenomic mRNA in OK10 defective leukemia virus transformed cells. J. Virol. 42:71–82.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by get electrophoresis. J. Mol. Biol. 98:503-517.
- Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA 77:5201-5205.
- Watson, D. K., M. S. McWilliams-Smith, C. Kozak, R. Reeves, J. Gearhart, M. F. Nunn, W. Nash, J. R. Fowle III, P. Duesberg, T. S. Papas, and S. J. O'Brien. 1986. Conserved chromosomal positions of dual domains of the *ets* protooncogene in cats, mice and humans. Proc. Natl. Acad. Sci. USA 83:1792-1794.
- 27. Watson, D. K., M. J. McWilliams, M. F. Nunn, P. H. Duesberg, S. S. O'Brien, and T. S. Papas. 1985. The *ets* sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. Proc. Natl. Acad. Sci. USA 82:7294– 7298.

Oncogene Research, 1988, Vol. 2, pp. 335-344 Photocopying permitted by license only © 1988 Harwood Academic Publishers GmbH Printed in the United States of America

# Cloning and Expression of Chicken $p54^{c-ets}$ cDNA<sub>s</sub>: the first $p54^{c-ets}$ coding exon is located into the 40.0 kbp genomic domain unrelated to v-ets

# MARTINE DUTERQUE-COQUILLAUD<sup>+</sup>, DOMINIQUE LEPRINCE, ANNE FLOURENS, CATHERINE HENRY<sup>+</sup>, JACQUES GHYSDAEL, BRIGITTE DEBUIRE<sup>+</sup>, and DOMINIQUE STEHELIN<sup>\*</sup>

†INSERM Unité 124-Institut de Recherches sur le Cancer-Place de Verdun-59045 LILLE-FRANCE. Laboratoire d'Oncologie Moléculaire, INSERM Unité 186-Institut Pasteur-l, rue Calmette-59019 LILLE CEDEX-FRANCE

We have isolated cDNA clones of chicken *c-ets* mRNA the longest of which, designated pCk E54A, contained  $\sim 2.0$  kb of a *c-ets* mRNA species. Nucleotide sequencing of this clone revealed a single long open reading frame, extending from the first ATG codon (nucleotide + 1) to a TGA termination codon at nucleotide 1324. The predicted translation product contains 441 amino acid residues and its molecular weight is 48 kd. Expression in COS-1 cells of this clone resulted in the synthesis of polypeptides immunologically indistinguishable from the authentic p54<sup>c-ets</sup> after one-dimensional gel electrophoresis.

Comparison of the nucleotide sequence of this cDNA to that of v-ets of avian acute leukemia virus E26 showed that both sequences are almost colinear with the exception of five point mutations but present striking differences in their 5' and 3' parts. 79 nucleotides downstream of the first ATG codon in c-ets cDNA are not found in the 5' part of v-ets where they are replaced by 223 different nucleotides. The 3' parts of v-ets and the coding region of the chicken c-ets cDNAs are also different : the last 13 codons of the cDNA are replaced by 16 different codons in v-ets.

Thus our results precisely define the structural differences between the *ets* encoded domain of E26 viral transforming protein (P135 gag-myb-ets) and the normal cellular protein p54<sup>c-ets</sup> expressed at high levels in chicken thymocytes and bursal lymphocytes. They also suggest the possibility of alternative splicing of different 5' exons to a common set of 3' exons.

KEYWORDS: E26, v and c-ets, oncogenes, alternative splicing

# INTRODUCTION

In chicken, the c-*ets* protooncogene, which appears to encode a protein of Mr 54000,  $p54^{c-ets}$ , is expressed at low levels in most cell lines and tissues and at high levels in thymocytes, bursal lymphocytes and normal spleen cells (Chen, 1985; Ghysdael et al., 1986a).  $p54^{c-ets}$  contains unique [<sup>35</sup>S] methionine containing tryptic peptides in addition to those (7 out of 10) shared with the v-*ets* encoded domain of the Mr 135000 transforming polyprotein (P135<sup>gag-myb-ets</sup>) (Bister et al., 1982; Ghysdael et al., 1986a; Leprince et al., 1983; Nunn et al., 1983) localized in the nucleus of E26 transformed cells (Klempnauer et al., 1984). We previously reported the structure of the chicken cellular DNA homologous to v-*ets* (Gegonne et al., 1987a). *c-ets* appears split into two domains

<sup>\*</sup>Corresponding author.


FIGURE 1 Structure of the c-ets locus and sequencing strategy for the c-ets cDNA clones lambda c-etsA and lambda c-etsB.

A diagrammatic figure of the previously described chicken *c-ets* locus is provided as well as a partial structure of the E26 virus with derivation of the probes used in this study and relevant features for comparison with the cDNA clones. Shown in bold line is the  $p54^{c-ets}$  protein coding region, translation initiation and termination codons and internal EcoRI restriction sites. EcoRI sites in parenthesis represent EcoRI sites generated during the cDNA libraries construction. Abbreviations used : B, Bam HI; Bg, Bg II; C, Cla I; E, EcoRI; Hd, Hind III; Hp<sup>+</sup>, Hpa I (only a site in exon a is represented); P, Pvu II; S, Sau 3A; SS, SS II.

separated by 40 kbp of v-ets unrelated DNA. (Fig. 1). The first domain accounts for 223 nucleotides of the 5' part of v-ets and is composed of two sets of sequences named  $\alpha$  and  $\beta$  which present the structural properties of typical eucaryotic exons. This domain however is neither transcribed into the major 7.5 kb c-ets mRNA species (Gegonne et al., 1987a; Leprince et al., 1983) nor translated into p54<sup>c-ets</sup> or more distantly related proteins (p60-64) expressed predominantly in macrophages (Gegonne et al., 1987b; Ghysdael et al., 1986b). The second domain contains 6 regions, named a through f, homologous to ca 1200 nucleotides of v-ets (Gegonne et al., 1987a). The most 3' region of homology to v-ets in the cloned cellular DNA ("exon f") was shown to be a truncated exon highly homologous to one of the two human c-ets loci (human c-ets 2) (De Taisne et al., 1984; Watson et al., 1985; Watson et al., 1986) lacking the nucleotides coding for the 16 carboxy-terminal amino acids of the viral protein.

To gain more insight into the structure of c-ets and to precisely define the relationship between the v-ets encoded domain of P135<sup>gag-myb-ets</sup> and its putative normal cellular counterpart p54<sup>c-ets</sup>, we sequenced c-ets cDNAs isolated from chicken embryo or spleen cell cDNA library. In this report we present the nucleotide sequence of the entire coding domain of a c-ets mRNA encoding p54<sup>c-ets</sup>. We have detected the first coding exon of p54<sup>c-ets</sup> and localized it into the 40.0 kbp domain unrelated to v-ets. We also provide evidence suggesting alternative splicing within the previously described c-ets locus.

## RESULTS

#### Cloning of c-ets cDNAs

A cDNA library prepared using mRNA from chicken spleen cells was screened by hybridization with v-ets probes since this organ has been reported to express high levels of *c-ets* mRNA<sub>s</sub> and p54<sup>c-ets</sup> protein (Chen et al., 1985). The first round of screening with a v-ets probe C (Bgl II-Hind III 750 bp fragment, see Fig. 1), yielded a total of 15 clones.

These clones contained a single EcoRI insert hybridizing with v-ets probe B or v-ets probes A and B (see Fig. 1) (Gegonne et al., 1987a) and ranged in size from 0.6 to 1.4 kbp. Restriction mapping analysis of the longest insert, ca. 1.4 kbp in clone A, demonstrated that it corresponded to most of the v-ets sequence (1.5 kbp) (Leprince et al., 1983; Nunn et al., 1983) except in the 5' part ( $\alpha$  and  $\beta$ ). It represented only a small part of the major 7.5 kb c-ets mRNA and lacked the 3' untranslated sequences of the cDNAs as well as the poly A tail. We previously reported the presence of an EcoRI site in a coding region of c-ets located downstream of the recombination point between E26 and the cellular gene (Gegonne et al., 1987a). We therefore concluded that clone A terminates at the previously mapped EcoRI site which escapes to the EcoRI methylase treatment during the construction of the library. We screened a chicken embryonic fibroblast cDNA library (Sap et al., 1986) with a probe corresponding to the 5' part of clone A (EcoRI-Cla I 157 bp fragment, see Fig. 1) in order to isolate a clone containing both the 5' side of clone A together with the 3' untranslated sequences. Such a clone, lambda c-ets B (Fig. 1) was obtained and appeared to contain two EcoRI fragments of ca. 1.3 and 0.8 kbp respectively. This latter fragment failed to hybridize to v-ets probes but hybridized with a single 8.5 kbp EcoRI fragment in total chicken DNA (data not shown). Most of this genomic DNA is present in the previously described chicken c-ets recombinant phage, lambda c-ets 5 ((Gegonne et al., 1987a); see Fig. 5).

## Nucleotide sequence of p54<sup>c-ets</sup> cDNAs

The composite *c-ets* mRNA sequence deduced from the analysis of these cDNA<sub>s</sub> represents ca. 2.0 kb of *c-ets* mRNA and is shown in Fig. 2. The longest open reading frame extends on 1404 nucleotides from the beginning of the sequence and contains at position + 1 a potential initiation codon within a sequence CAACCATGA in good agreement with the "consensus" sequence at eukaryotic initiation codons CCAC-CATGG described by Kozak (Kozak, 1986). However, we cannot presently rule out the possibility that the AUG codon at position + 1 encodes an internal methionine residue since the upstream reading frame is not closed. Translation initiated at position + 1 would result in a protein of 441 amino acids highly homologous to the *v-ets* encoded domain of P135<sup>gag-myb-ets</sup> except for its amino and carboxy-terminal parts (see below). This protein would have a predicted molecular weight of 48 kd, a value shorter by 6 kd of the apparent molecular weight in SDS polyacrylamide gel of the *c-ets* protein (54 kd) (Chen, 1985; Ghysdael et al., 1986a). Such a discrepancy could be explained by the high level of Proline residues (29 out of 441) in p54<sup>c-ets</sup>.

To prove that clone A was able to encode p54<sup>c-ets</sup> and thus contained the true initiation codon for this protein, we subcloned it into the expression vector pSVL to yield pSVL etsA. pSVL etsA was transfected into COS-1 cells and the resulting polypeptides were examined by [<sup>35</sup>S] methionine labelling and SDS-polyacrylamide gel electrophoresis (Ghysdael et al., 1986a). Two predominant polypeptides of approximatively 54 and 56 kd apparent molecular weight as well as fainter bands of higher molecular weight were

GAAGGEGETEGGECEGECEGECEGECEGECAGECAGECEGECECAGTTECEGGETEGECETEGGECETEAACE

ATG AAG GCG GCG GTG GAC CTG AAG CCC ACC CTG ACC ATC ATC AAG ACG GAG AAG GTG GAC ATC GAT CTC TTC Met Lys Alla Alla Val Aap Leu Lys Pro Thr Leu Thr He He Lys Thr Giu Lys Val Aap Leu Phe TAC AGC ACA CC CCC GCC GAT ATG GAA TGT GCA GAT GTG CCT TTG TTA ACC CCC AGC AAG GAA ATG ATG TCT CAG GCA Pro Phe Pro Aap Mate Giu Cys Alla Aap Val Pro Leu Leu Thr Pro Ser Ser Lys Glu Met Met Ser Gin Alla Tyr Ser Thr Aap

C-815

CTG AAA GCC ACC TTC AGT GGC TTC GCA AAG GAG CAG CAG CGG CTG GGA ATC CCC AAA GAT CCC CAG CAG TGG Lew Lys Ala Thr Phe Ser Gly Phe Ala Lys Glu Gin Arg Leu Gly lie Pro Lys Amp Pro Gin Gin Trp 250 ACA GAG ACG CAC GTG CGG GAC TGG GTG ATG TGG GCA GTG AAT GAG TTC AGC CTG AAG GGA GTG GAT TTC CAG Thr Glu Thr His Val Arg Am Trp Val Met Trp Ala Val Am Glu Phe Ser Leu Lys Gly Val Am Phe G 300 . . . 350 AAG TTC TGC ATG AAC GGA GCT GCC CTG TGC GCC CTG GGC AAG GAG TGC TTC CTG GAG CTA CGG CCT GAC TTT Lys Phe Cys Met Asn Giy Ala Ala Leu Cys Ala Leu Gly Lys Glu Cys Phe Leu Glu Leu Ala Pro Amp Phe BOIN 400 GTG GGA GAT ATC CTT TGG GAA CAC CTG GAG ATC TTG CAG AAA GAA GAA GAA GCA AAA CCA TAC CCA GCA AAT GGA Val Gly Asp lie Leu Trp Giu His Leu Giu lie Leu Gin Lys Glu Giu Ala Lys Pro Tyr Pro Ala Asn Gly . 450 . 500 GTG AAT GCA GCG TAT CCA GAA TCC CGC TAT ACT TCA GAC TAC TTC ATT AGT TAT GGC ATC GAG CAC GCA Yal Asn Ala Ala Tyr Pro Giu Ser Arg Tyr The Ser Asp Tyr Phe IIe Ser Tyr Giy Ile Giu His Ala • . • 550 TGC GTG CCT CCC TCC GAG TTC TCT GAG CCC AGC TTC ATC ACA GAG TCC TAC CAG ACC CTC CAT CCC ATC AGC Cys Val Pro Pro Ser Glu Phe Ser Glu Pro Ser Phe Lie Thr Glu Ser Tyr Gin Thr Leu His Pro Lie Ser • 600 TCG GAA GAG CTT CTG TCC CTC AAG TAC GAG AAC GAC TAT CCC TCA GTC ATC CTT CGT GAC CCC GTC CAG ACG Ser Glu Glu Leu Leu Ser Leu Lys Tyr Glu Asn Amp Tyr Pro Ser Val IIe Leu Arg Amp Pro Val Gln Thr . 650 700 GAC TCC CTG CAG ACA GAC TAC TTC ACA ATC AAG CAA GAA GTG GTA ACG CCA GAC AAC ATG TGC ATG GGA CGT Asp Ser Leu Gin Thr Asp Tyr Phe Thr lie Lys Gin Glu Val Val Thr Pro Asp Asn Met Cys Me: Gly Arg TO THE TRANSPORT OF THE CASE OF CASE O . ACA CAG TCC TC'S AGC AGC CAG TCC TCC TTC CAG AGC CTG CAG CGC GTC CCC TCC TAC GAT AGC TTT GAC TCA Thr Gin Ser Trp Cer Ser Gin Ser Ser Phe Gin Ser Leu Gin Arg Val Pro Ser Tyr Amp Ser Phe Amp Ser 900 GAG GAC TAC CCC GCC GCC CTG CCC AAC CAC AAG CCC AAG GGC ACC TTC AAG GAC TAT GTT CGA GAT CGG GGT Glu Asp Tyr Pro Ala Ala Leu Pro Asn His Lys Pro Lys Gly Thr Phe Lys Asp Tyr Val Arg Asp Arg Ala 930 1000 GAC ATG AAC AAG GAC AAG CCT GTC ATT CCT GCC GCT GCC CTC GCC GGC TAC ACA GGC AGT GGA CCC ATC CAA Amp Met Asn Lys Asp Lys Pro Val Ile Pro Ala Ala Ala Leu Ala Gly Tyr Thr Gly Ser Gly Pro Ilé Gin Leu Trp Gin Phe Leu Leu Giu Leu Leu Int nap Lys de Gy 

FIGURE 2 Nucleotide and predicted amino-acid sequences of c-ets clones A and B and comparison with v-ets. Only those nucleotides and amino-acids in v-ets that differ from c-ets are boxed with the exception of  $\alpha$  and  $\beta$  which are not represented. The principal restriction sites are indicated.

specifically immuno precipitated by the previously described anti-*ets*A serum (Gegonne et al., 1987b; Ghysdael et al., 1986a) (Fig. 3, lanes 3 and 4). The same polypeptides were present in normal chicken thymus lysates immunoprecipitated and run on the same gel (Fig. 3, lane 7). They were absent in mock transfected COS-1 cells (Fig. 3, lanes 1 and 2)

#### NUCLEOTIDE SEQUENCE OF p54<sup>c-ets</sup> cDNA



FIGURE 3 Transient expression of p54<sup>c-ets</sup> in COS cells.

COS cells were transfected with  $10 \ \mu g$  of plasmid DNA and 72 h later the cells were metabolically labelled with [<sup>35</sup>S]methionine and analyzed by immunoprecipitation with anti *ets*A serum (8) (lanes 1, 3, 5 and 7) or anti *ets*A serum adsorbed with an excess of bp 17 <sup>MS2-etsA</sup> (8) (lanes 2, 4 and 6). Mock-transfected cells (lanes 1 and 2); COS cells transfected with the *c-ets* cDNA clone A into the pSVL COS expression vector in the sense orientation (lanes 3 and 4) and in the anti-sense orientation versus the SV 40 late promotor of transcription (lanes 5 and 6); cell suspension obtained from thymuses of 2-week-old chickens (lane 7).

or in cells transfected with a *ets*A cDNA cloned in pSVL in the opposite orientation versus the SV40 promotor of transcription (Fig. 3, lanes 5 and 6). In addition the peptide map of the cDNA encoded p54 from COS-1 cells appeared similar to that of p54<sup>c-ets</sup> from normal chicken thymuses (data not shown).

Our cDNA clone, now referred to as pCk E54A, directed the synthesis of several p54<sup>c-ets</sup> and p56<sup>c-ets</sup> polypeptides also detected in vivo in normal cells. The multiple products encoded by pCk E54A could represent either post-translational modifications of a single protein and/or alternative initiations at two in phase ATG<sub>s</sub> (Strubin et al., 1986). Initiation at the ATG codon at position 85 would yield a protein 28 amino acids shorter, consistent with a difference of 2 kd apparent molecular weight. However, we do not favor this latter hypothesis since this ATG codon lies within an unfavorable context according to the consensus sequence defined by Kozak (Kozak, 1986). In conclusion, the protein encoded by this cDNA (pCk E54 A) is p54<sup>c-ets</sup> which, by peptide mapping analysis (Ghysdael et al., 1986a) and by nucleotide sequencing, exhibits a strong homology with the v-ets domain of P135<sup>gag-myb-ets</sup> in contrast to the p60-64 proteins which display only a limited domain of homology with P135<sup>gag-myb-ets</sup> (Gegonne et al., 1987b; Ghysdael et al., 1986b).

## Relationship to v-ets and c-ets genomic DNA

The nucleotide sequence of the chicken c-ets cDNA was compared to the nucleotide sequence of its viral oncogenic counterpart v-ets (Nunn et al., 1983). The alignment shown in Fig. 2 reveals the following striking features.

First, a large region (1203 nucleotides from nucleotides 83 to 1285 of the cDNA and

339

FIGURE 4 Exon a is fused to exon I in pCK E54 A or to exon B in v-ets.

A perfect homology was observed between the cDNA pCk E54A and the genomic exon I until nucleotide 82 in the cDNA corresponding to a consensus splice donor site (Sd). In the genomic sequence, the putative 5' untranslated region of the cDNA contains a GA rich region and short sequence repeats such as GGAGG. The cellular exon I is fused to the previously described exon a to yield the c-ets coding sequence. By contrast exon  $\beta$  is fused to the same exon a to yield the v-ets sequence.

nucleotides 758 to 1960 of v-ets according to numbering in Nunn et al., 1983) corresponding to most of "exons" a to f was found homologous between the cellular and viral sequences with only 5 point mutations among which two are silent mutations and 3 others are chemically conservative. Moreover, the conservative mutation at position 1062 which abolishes a PVu II site observed in v-ets (AAGCTG/CAGCTG) could reflect an allelic variation since we did not find it in some other c-ets cDNAs by restriction mapping analysis (data not shown).

Second, 79 nucleotides downstream of the first ATG codon of the c-ets cDNA were not found in v-ets. This could account for some of the [35S] methionine labelled tryptic peptides which appeared unique either to p54<sup>c-ets</sup> or to the v-ets encoded portion of P135gag-myb-ets (Gegonne et al., 1987a; Ghysdael et al., 1986a). In E26 these 79 nucleotides are replaced by 223 nucleotides originating from two regions named  $\alpha$  and  $\beta$  found in the 5' genomic domain homologous to v-ets. The point where the viral and cDNA sequences diverge corresponds to the leftward boundary of c-ets exon a (Fig. 4 and 5) (Gegonne et al., 1987a). We previously reported that  $\alpha$  and  $\beta$  although not found in the major c-ets 7.5 kb mRNA species (Leprince et al., 1983) nor translated into c-ets proteins (Gegonne et al., 1987a), exhibited typical properties of true eukaryotic exons with an open reading frame identical to that used by E26 and consensus splice signals (Gegonne et al., 1987a). Therefore, the most obvious mechanism to generate the viral v-ets sequence involves the splicing of  $\alpha$  and  $\beta$  to a through f (Fig. 5). These observations, together with the nucleotide sequence of pCk E54A cDNA suggested an alternative splicing of 5' exons to a common set of 3' exons. To confirm this hypothesis we cloned and sequenced the genomic equivalent of the 5' part of the c-ets cDNA. This sequence was precisely mapped within a 2.7 kpb Hind III fragment in the previously described lambda c-etsB recombinant phage (Gegonne et al., 1987a). This cellular sequence is localized within the 40.0 kbp genomic DNA stretch unrelated to v-ets and separating the two v-ets related domains ( $\alpha$ ,  $\beta$  and a-f). Nucleotide sequence analysis showed that the position where v-ets and pCK E54 A cDNA diverge (nucleotide 82 in Fig. 4) corresponds to a splice



FIGURE 5 Schematic representation of the chicken *c-ets* locus as compared to the *v-ets* sequence of E26 and the cDNA encoding p54<sup>c-ets</sup>.

In the cDNA, the stippled box represents the putative end of the 3' untranslated region and the poly A tail which have not been investigated in detail. Abbreviations for restriction cleavage sites are as follows : B : BamHI; C : Cla I; H : Hind III;  $Hp^* : Hpa I$ . EcoRI sites are represented as vertical bars.

donor site in the genomic DNA. This genomic sequence appeared to be a single DNA stretch colinear both to the 5' untranslated region and the beginning of the coding region of our cDNA (Fig. 4). We named this exon I since it contains the Initiation codon for p54<sup>c-ets</sup>. We noted that the putative 5' untranslated regions contains a GA rich region (nucleotides -122 to -77) in which repeated sequences such as GGAGG can be observed.

Third, we confirmed and extended our previous observation that v-ets and c-ets encode proteins which differ in their carboxy-terminal parts since the last 13 codons of the cDNA are replaced by 16 other codons in the viral sequence (Gegonne et al., 1987a). Therefore we propose to name F the exon encoding the carboxy-terminal part of the cellular protein p54<sup>c-ets</sup> whereas f would refer to the last region of homology between v-ets and c-ets. In the virus, f is followed by 82 nucleotides of unknown origin which encode the different carboxy-terminus of the viral protein.

At last, we isolated several cDNA clones overlapping pCK E54 A and containing sequences located downstream of the open reading frame (for example 0.8 kb in clone lambda *c-etsB*). These sequences located downstream of the p54<sup>*c-ets*</sup> coding region were not investigated in detail as a preliminary analysis revealed no long open reading frames, thus indicating that they probably represented 3' untranslated sequences.

#### DISCUSSION

Several avian retroviruses (AEV,  $MH_2$  and E26) harbor in their genomes two distinct cell-derived oncogenes (Vennstrom et al., 1982; Coll et al., 1983; Leprince et al., 1983). In one of them, E26, these two oncogenes are expressed in a single fusion protein of Mr 135000 (P135<sup>gag-myb-ets</sup>). We have previously reported the genomic organization of the chicken c-ets protooncogene (Gegonne et al., 1987a) and identified the major c-ets encoded protein (Ghysdael et al., 1986a; Ghysdael et al., 1986b). In this study we have characterized several c-ets cDNAs in order to understand how v-ets is related to the normal c-ets gene and to gain some insight into the genesis of E26.

Our results show that c-ets is colinear to v-ets except in the 5' and 3' extremities. The amino-terminus of  $p54^{c-ets}$  is encoded by sequences located in the 40 kbp domain

341

separating the two v-ets related domains. Therefore this long DNA stretch does not represent a large intron but contains at least one cellular exon not transduced by the virus. This exon has to be spliced to exons a through F (Gegonne et al., 1987a) to yield the p54 coding region. The cellular sequences homologous to the 5' part of v-ets ( $\alpha$  and  $\beta$ ) although neither detectable in the major 7.5 kb c-ets mRNA species nor translated into p54<sup>c-ets</sup> nor the ets related p60-64 proteins display a typical exon structure (Gegonne et al., 1987a; Gegonne et al., 1987b) These sequences should be fused to exons a through an altered F to generate the v-ets part of P135gag-myb-ets. Taken together these data suggest that alternative splicing events could be operative in chicken c-ets to generate either the mRNA species coding for p54<sup>c-ets</sup> or a so far unindentified mRNA containing  $\alpha$  and  $\beta$ v-ets sequences. Consistent with that hypothesis we have recently isolated from a chicken embryo cDNA library a clone hybridizing with a specific probe for  $\alpha$  and  $\beta$ . Preliminary results indicate that this cDNA sensibly differs from v-ets in its 3' part but corresponds to the expected splicing of  $\alpha$  and  $\beta$  to a common set of 3' exons (a through F). (Leprince et al., in preparation). Therefore it could/or might exist a protein highly related to p54<sup>c-ets</sup> but having a distinct amino-terminus.

The  $p54^{c-ets}$  cellular protein and the v-ets part of the viral protein also differ in their carboxy-terminal part. Similar examples of cellular and viral proteins differing at their carboxy-termini have been reported for oncogenes of the tyrosine kinase family (*src*, Cooper et al., 1986; Takeya et al., 1983; *fms*, Coussens et al., 1986; *ros*, Neckamayer et al., 1986) and for the nuclear v- and c-fos proteins (Van Beveren et al., 1982). In some cases, the biological relevance in terms of oncogenic activation of this difference has been defined. The best documented example is the replacement of a negatively regulated Tyr (Tyr<sup>527</sup> in c-src) in the *src* family oncogenes (Cooper et al., 1986).

Although the function of the cellular *ets* proteins is unknown, evidences based on the sequence of the viral and cellular genes have enabled us to detect their structural differences. These changes, together with the association to the *v*-*myb* sequence, probably contribute to the oncogenic potential of *v*-*ets*.

### MATERIALS AND METHODS

#### cDNA cloning and isolation of c-ets cDNA clones

A SPF chicken (Spafas; 2 animals) spleen cell cDNA library in lambda gt 10 prepared by modification of the Hoffman-Gubler procedure was kindly provided by Drs K. Strebhardt and J. Mullins. A SPF chicken fibroblasts cDNA library in lambda gt 11 was a kind gift of Drs J. Sap and B. Vennstrom (Sap et al., 1986). Both libraries were screened by hybridization to v-ets specific probes (Gegonne et al., 1987a). Hybridization was at 41°C in 50% (W/V) formamide,  $3 \times SSC$  ( $1 \times SSC = 0.15M$  NaCl, 0.015M Sodium citrate, pH 7.0)  $5 \times$  Denhardt's solution, 100 µg/ml denatured Salmon sperm DNA and 0.1% NaDodSO<sub>4</sub>, and filters were washed at 50°C in 0.1  $\times$  SSC/0.1% NaDodSO<sub>4</sub>. Approximately 400,000 recombinant plaques were screened from each library.

#### Southern hybridization

The orientation of and overlap between the different cDNA clones were determined by restriction mapping and confirmed by hybridization to a v-ets subclone (Leprince et al., 1983) and to chicken genomic c-ets clones using the classical procedure (Gegonne et al., 1987a).

#### NUCLEOTIDE SEQUENCE OF p54<sup>c-ets</sup> cDNA

#### Nucleotide sequence analysis

The DNA sequence was determined by the dideoxy chain termination method (Sanger et al., 1977) after subcloning of specific restriction fragments into the polylinker restriction sites of M13 phage derivatives mp 10, mp 11, mp 18 and mp 19 (Messing et al., 1981).

## **COS-cells** transfection

The EcoRI sites of the *c-ets* cDNA clone pCk E54A were converted to blunt ends using the Klenow fragment of DNA polymerase I before ligation to Sal I linkers. After addition of linkers the mixture was digested with Sal I and the resulting 1.5 kb Sal I *c-ets* fragment was subcloned into the Xho I digested pSVL COS expression vector (Pharmacia) in both orientations versus the SV 40 late promotor of transcription.

Plasmid DNA, 10 µg, was transfected into  $1 \times 10^6$  COS-1 cells (11) in a 100 mm dish by the DEAE dextran protocol (Sompayrac et al., 1981) with the addition of a chloroquine treatment (Luthmann et al., 1983).

Labelling of cells with [<sup>35</sup>S] methionine 72 hours after the addition of the DNA, preparation of cell lysates, immunoprecipitation and NaDodSO<sub>4</sub>/polyacrylamide gel electrophoresis have been previously described (Ghysdael et al., 1986a).

#### Acknowledgments

We thank S. Saule for stimulating discussions, D. Hetuin and A. Begue for excellent technical assistance, M. Lestiennes for patient typing and M.B. Raes for help in manuscript preparation. This work was supported by the Institut National de la Santé et de la Recherche Médicale (U. 124-U. 186), Centre National de la Recherche Scientifique (ATP n° 98548 and UA 04-1160), Association pour la Recherche sur le Cancer and the Pasteur Institute of LILLE.

(Received September 9, 1987) (Accepted November 4, 1987)

#### REFERENCES

- Bister, K., Nunn, M., Moscovici, C., Perbal, B., Baluda, M.A. & Duesberg, P.H. (1982). Acute leukemia viruses E 26 and avian myeloblastosis virus have related transformation specific RNA sequences but different genetic structures, gene products and oncogenic properties. Proc. Natl. Acad. Sci., USA 79, 3677–3681.
- Chen, J.H. (1985). The protooncogene c-ets is preferentially expressed in lymphoid cells. Mol. & Cell. Biol., 5, 2993-3000.
- Coll, J., Righi, M., De Taisne C., Dissous, C., Gegonne, A. and Stehelin, D. (1983). Molecular cloning of the avian acute transforming retrovirus MH2 reveals a novel cell derived sequence (v-mil) in addition to the myc oncogene. EMBO J., 2, 2189-2194.
- Cooper, J.A., Gould, K.L., Cartwright, C.A. & Hunter, T. (1986). Tyr 527 is phosphorylated in pp 60<sup>c-src</sup>. Implications for regulation. Science, 231, 1431–1434.
- Coussens, L., Van Beveren, C., Smith, D., Chen, E., Mitchell, R.L., Isacke, C.M., Verma, I.M. & Ullrich, A. (1986). Structural alteration of viral homologue of receptor protooncogene *fms* at carboxyl-terminus. *Nature*, **320**, 277–280.
- De Taisne, C., Gegonne, A., Stehelin, D., Bernheim, A. & Berger, R. (1984). Chromosomal localization of the human protooncogene c-ets. Nature, 310, 581-583.
- Gegonne, A., Leprince, D., Duterque-Coquillaud, M., Vanderdunder, B., Flourens, A., Ghysdael, J., Debuire, B. & Stehelin, D. (1987a). Multiple domains for the chicken cellular sequences homologous to the v-ets oncogene of the E26 retrovirus. *Mol. Cell. Biol.*, 7, 806-812.
- Gegonne, A., Leprince, D., Pognonec, P., Dernis, D., Raes, M.B., Stehelin, D. & Ghysdael, J. (1987a). The 5' extremity of the v-ets oncogene of avian leukemia virus E26 encodes amino-acid sequences not derived from the major c-ets encoded cellular proteins. Virol., 156, 177–180.

Ghysdael, J., Gegonne, A., Pognonec, P., Dernis, D., Leprince, D. & Stehelin, D. (1986a). Identification and preferential expression in thymic and bursal lymphocytes of a c-ets oncogene encoded Mr 54,000 cytoplasmic protein. Proc. Natl. Acad. Sci. USA 83, 1714-1718.

Ghysdael, J., Gegonne, A., Pognonec, P., Boulukos, K., Leprince, D., Dernis, D., Lagrou, C. & Stehelin, D. (1986b). Identification in chicken macrophages of a set of proteins related to, but distinct from, the chicken cellular c-ets encoded protein p54<sup>c-ets</sup>. EMBO J., 5, 2251-2256.

Gluzmann, Y. (1981). SV 40-transformed simian cells support the replication of early SV 40-mutants. Cell, 23, 175-182

- Klempnauer, K., Symonds, G., Evan, C. & Bishop, J.M. (1984). Subcellular localization of proteins encoded by oncogenes of avian myeloblastosis virus and avian leukemia virus E26 and by the chicken c-myb gene. Cell 37, 537-547.
- Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell, 44, 283-292.
- Leprince, D., Gegonne, A., Coll, J., De Taisne, C., Schneeberger, A., Lagrou, C. & Stehelin, D. (1983). A putative second cell-derived oncogene of the avian leukemia retrovirus E26. Nature (London) 306, 395-397.
- Luthmann, H. and Magnusson, G. (1983). High efficiency polyoma DNA transfection of chloroquine treated cells. Nucleic. Acid. Res., 11, 1295-1308.
- Messing, J., Crea, R. & Seeburg, P.H. (1981). A system for shotgun DNA sequencing. Nucleic Acids Res., 9, 309-321
- Neckamayer, W.S., Shibuya, M., Hsu, M. & Wang, L. (1986). Protooncogene c-ros codes for a molecule with structural features common to those of growth factor receptors and displays tissue specific and developmentally regulated expression. *Mol. & Cell. Biol.* 6, 1478–1486.
- Nunn, M.F., Seeburg, P.H., Moscovici, C. & Duesberg, P.H. (1983). Tripartite structure of the avian erythroblastosis virus E26 transforming gene. Nature (London) 306, 391-395. Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain terminator inhibitor. Proc. Natl.
- Acad. Sci., USA, 74, 5463-5467.

Sap, J., Munoz, A., Damm, K., Goldberg, Y., Ghysdael, J., Leutz, A., Beug, H. and Vennstrom, B. (1986).The c-erbA protein is a high affinity receptor for thyroid hormone. Nature, 324, 635-639.

Sompayrac, L.M. and Danna K.J. (1981). Efficient infection of monkey cells with DNA of simian virus 40. Proc. Natl. Acad. Sci., USA, 78, 7575-7578.

Strubin, M., Long, E.O. and Mach, B. (1986). Two forms of the Ia antigen-associates invariant chain result from alternative initiations at two in phase AUGs. Cell, 47, 619-625

- Takeya, T. & Hanafusa, H. (1983). Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. Cell, 32, 881-890.
- Van Beveren, C., Van Straaten, F., Curran, T., Muller, R. and Verma, I.M. (1982). 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
- 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.
- Watson, D.K., Mc Williams-Smith, M.J., Nunn, M.F., Duesberg, P.H., O'Brien, S.J. & Papas, T.S. (1985). The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21 : both loci are transcriptionnaly active. Proc. Natl. Acad. Sci., USA 82, 7294-7298.
- Watson, D.K., Mc Williams-Smith, M.J., Kozak, C., Reeves, R., Gearhart, J., Nunn, M.F., Nash, W., Fowle, J.R., III, Duesberg, P., Papas, T.S. & O'Brien, S.J. (1986). Conserved chromosomal positions of dual domains of the ets protooncogene in cats, mice and humans. Proc. Natl. Acad. Sci., USA 83, 1792-1796.

# Alternative Splicing within the Chicken c-ets-1 Locus: Implications for Transduction within the E26 Retrovirus of the c-ets Proto-Oncogene

## DOMINIQUE LEPRINCE,<sup>1\*</sup> MARTINE DUTERQUE-COQUILLAUD,<sup>2</sup> RUO-PING LI,<sup>1</sup> CATHERINE HENRY,<sup>2</sup> ANNE FLOURENS,<sup>1</sup> BRIGITTE DEBUIRE,<sup>2</sup> AND DOMINIQUE STEHELIN<sup>1</sup>

Laboratoire d'Oncologie Moléculaire, Institut National de la Santé et de la Recherche Médicale Unité 186/Centre National de la Recherche Scientifique UA 1160, Institut Pasteur, 1 Rue Calmette, 59019 Lille Cedex,<sup>1</sup> and Institut National de la Santé et de la Recherche Médicale Unité 124, Institut de Recherches sur le Cancer, 59045 Lille,<sup>2</sup> France

#### Received 19 January 1988/Accepted 20 May 1988

Two overlapping c-ets-1 cDNA clones were isolated which contained the  $\alpha$  and  $\beta$  genomic sequences homologous to the 5' end of v-ets not detected in the previously described c-ets RNA species or proteins. Nucleotide sequencing demonstrated that these cDNAs corresponded to the splicing of  $\alpha$  and  $\beta$  to a common set of 3' exons (a through F) already found in the p54<sup>c-ets-1</sup> mRNA. They contained an open reading frame of 1,455 nucleotides which could encode a polypeptide of 485 amino acids with a predicted molecular mass of 53 kilodaltons. However, when expressed in COS-1 cells, the cDNAs directed the synthesis of a protein with an apparent molecular mass in sodium dodecyl sulfate-polyacrylamide gel electrophoresis of 68 kilodaltons,  $p68^{c-ets-1}$ , comigrating with a protein expressed at low levels in normal chicken spleen cells. These two proteins were shown to be identical by partial digestion with protease V8. Northern (RNA) blot hybridization analysis with the  $p68^{c-ets-1}$ -specific sequence and RNase protection experiments showed that the corresponding mRNA was expressed in normal chicken spleen and not in normal chicken thymus or in various T lymphoid cell lines. Thus, two closely related proteins, having distinct amino-terminal parts, are generated within the same locus by alternative addition of different 5' exons,  $\alpha$  and  $\beta$  or  $I^{54}$ , respectively, onto a common set of 3' exons (a to F). Finally, we demonstrate that an aberrant splicing event between a cryptic splice donor site in c-myb exon E6 and the normal splice acceptor site of c-ets-1 exon  $\alpha$  is involved in the genesis of the E26 myb-ets sequence.

The avian acute leukemia virus E26 induces a mixed erythroid-myeloid leukemia and transforms cells of both erythroid and myeloid lineages in vivo and in vitro (2, 22, 35, 41). The 5.7-kilobase (kb) E26 genomic RNA contains two distinct cell-derived oncogenes, v-myb and v-ets, expressed, together with a partial viral gag gene, as a 135-kilodalton (kDa) nuclear fusion protein, P135<sup>gag-myb-ets</sup> (6, 23, 29, 40).

We previously reported that the chicken cellular sequence from which v-ets presumably arose, c-ets, is a contiguous locus split into two domains (6.0 and 18.0 kilobase pair [kbp], respectively) separated by ca 40.0 kbp of v-etsunrelated DNA (15). These studies were further consolidated by the characterization and nucleotide sequencing of a c-ets cDNA clone able to encode the major chicken ets protein p54<sup>c-ets</sup> (8, 19) by transient expression in COS-1 cells (M. Duterque-Coquillaud, D. Leprince, A. Flourens, C. Henry, J. Ghysdael, B. Debuire, and D. Stehelin, Oncogene Res., in press). All these data can be summarized as follows.

(i) The 3' domain (18.0 kbp) contains six regions of homology to v-ets, named a to f. Regions a to e were homologous between the viral and cellular sequences, except for five point mutations, two silent and three chemically conservative (Duterque-Coquillaud et al., in press). By contrast, the last region, f, is a truncated and altered form of a c-ets cellular exon named F. The v-ets- and c-ets-encoded proteins differ in their carboxy-terminal parts, since the last 13 codons in the cellular exon are replaced by 16 codons of unknown origin in the viral sequence (15; Duterque-Coquillaud et al., in press).

(ii) The 40.0-kbp domain unrelated to v-ets contains the

first coding exon of the p54 cDNA clone. This exon (named I) which, for clarity, is referred to as  $I^{54}$ , should be fused to exons a to F to yield the p54<sup>c-ers</sup>-encoding sequence (Duterque-Coquillaud et al., in press).

(iii) The 5' domain accounts for 223 nucleotides at the 5' part of v-ets and is composed of two sets of sequences, named  $\alpha$  and  $\beta$ , presenting the structural properties of typical eucaryotic exons (15). This domain is not transcribed into the major 7.5-kb c-ets RNA species identified in lymphoid cells such as MSB1, translated into p54<sup>c-ets</sup> (15, 16; Duterque-Coquillaud et al., in press), or translated into the more distantly related p58 to p63 proteins found in macrophages (18) and originating from another locus (7). However, to yield the viral v-ets sequence,  $\alpha$  and  $\beta$  should be fused by splicing to exon a (15). Similarly, the  $I^{54}$  exon should be fused to exon a to yield the p54<sup>c-ets</sup>-encoding sequence (Duterque-Coquillaud et al., in press).

(iv) The human homologs of the *ets* oncogene consist of at least two distinct domains located on different chromosomes: ETS-1 on chromosome 11 (13, 60, 61) and ETS-2 on chromosome 21 (60, 61). Recently a new member of the *ets* oncogene family, *erg*, has been identified (42). The chicken genome also harbors two loci related to v-*ets*: a DNA locus highly related to v-*ets* and encoding  $p54^{c-ets-1}$ , known as chicken c-*ets-1* (15, 19; Duterque-Coquillaud et al., in press), and a recently characterized locus only distantly related to v-*ets*, encoding the p58 to p64 proteins expressed in macrophages, named chicken c-*ets-2* (7, 18).

In this report, we demonstrate that alternative splicing events are operative in the chicken c-*ets-1* locus to generate either the mRNA species coding for  $p54^{c-ets-1}$  or a previously undetected mRNA species containing  $\alpha$  and  $\beta$ . We cloned

<sup>\*</sup> Corresponding author.

two distinct overlapping cDNA clones differing from v-ets only in their 3' end. These cDNAs correspond to the expected splicing of  $\alpha$  and  $\beta$  sequences to the common set of 3' exons (a to F). These cDNAs, when expressed in COS-1 cells, directed by synthesis of a novel p68<sup>c-ets-1</sup> protein closely related to p54<sup>c-ets-1</sup>, but having a distinct aminoterminal part and detected at low levels in normal chicken spleen cells.

Finally, we demonstrate that an aberrant splicing between a cryptic splice donor site located in c-myb exon E6 and the normal splice acceptor site of c-ets-1 exon  $\alpha$  is involved in the genesis of the E26 myb-ets sequence.

#### MATERIALS AND METHODS

cDNA cloning and isolation of c-ets-I cDNA clones. An SPF chicken (two animals; SPAFAS) spleen cell cDNA library in lambda gt10 prepared by modification of the Hoffman-Gubler procedure was kindly provided by K. Strebhardt and J. Mullins. An SPF 11-day-old chicken embryo cDNA library in lambda gt11 was a kind gift of J. Sap and B. Vennstrom (47). Both libraries were screened by hybridization to v-ets-specific probes. Hybridization was carried out at 41°C in 50% (wt/vol) formamide-3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate [pH 7.0])-5× Denhardt solution-100  $\mu$ g of denatured salmon sperm DNA per ml-0.1% sodium dodecyl sulfate (SDS), and filters were washed at 50°C in 0.1× SSC-0.1% SDS. Approximately 400,000 recombinant plaques were screened from each library as previously described (Duterque-Coquillaud et al., in press).

Nucleotide sequence analysis. The DNA sequence was determined by the dideoxy chain termination method (46) after subcloning of specific restriction fragments into the polylinker restriction sites of M13 bacteriophage derivatives mp10, mp11, mp18, and mp19 (33).

COS cell transfection. The EcoRI sites of the c-*ets-1* cDNA clones were converted to blunt ends by using the Klenow fragment of DNA polymerase I and subcloned into the *SmaI*-digested pSVL COS expression vector (Pharmacia) in the correct orientation with respect to the simian virus 40 late promoter of transcription.

Plasmid DNA (10  $\mu$ g) was transfected into 10<sup>6</sup> COS-1 cells (20) in a 100-mm dish by the DEAE-dextran method (52) with the addition of a chloroquine treatment (30) as previously described (Duterque-Coquillaud et al., in press).

Radiolabeling of cells, immunoprecipitation, and peptidemapping analyses. Cells were labeled with L-[ $^{35}$ S]methionine in methionine-free Eagle medium supplemented with 5% dialyzed calf serum and lysed by boiling for 5 min in a 1% solution of SDS. After dilution with 10 volumes of 150 mM NaCl-1 mM EDTA-10 mM Tris hydrochloride (pH 7.4)-1% Triton X-100-0.5% sodium deoxycholate-1% Trasylol, the lysates were centrifuged at 100,000 × g for 60 min and immunoprecipitations were carried out as described previously with anti-*ets* sera (16, 19). Before immunoprecipitation, the spleen cell lysates were incubated with 50  $\mu$ l of 50% protein A-Sepharose beads (suspended in radioimmunoprecipitation buffer) for 30 min at 4°C with gentle shaking to eliminate the heavy chain of avian immunoglobulin (ca. 68 kDa) present in B lymphocytes of chicken spleens.

Partial digestion by *Staphylococcus aureus* V8 protease in gel slices containing [<sup>35</sup>S]methionine-labeled proteins was done essentially as described by Cleveland et al. (9).

**RNA blot analysis.** Total cellular RNAs were prepared by the guanidine thiocyanate procedure, fractionated by electrophoresis on a 1% agarose-formaldehyde gel, and trans-

ferred to Hybond membranes (Amersham Corp.). Transfer, hybridization, and washing were done as recommended by the manufacturer.

Nuclease protection experiment. The EcoRI-Bg/II 566-bp fragment from the p68-2 cDNA clone was subcloned in pGEM4. After linearization by EcoRI, an RNA probe complementary to the mRNA (cRNA) was synthesized by using T7 phage polymerase (Promega Biotec). Total RNA (20 μg) or 5  $\mu$ g of *Escherichia coli* tRNA was mixed with 5 × 10<sup>5</sup> cpm of [<sup>32</sup>P]cRNA (specific activity, 10<sup>9</sup> dpm/ $\mu$ g) in 30  $\mu$ l of 80% formamide-400 mM NaCl, 1 mM EDTA-40 mM piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES; pH 6.5). The mixture was heated at 85°C for 5 min and incubated for 4 h at 37°C and then for 10 h at 30°C (27). After hybridization, 300 µl of RNase digestion buffer (5 mM EDTA, 300 mM NaCl, 10 mM Tris hydrochloride [pH 7.5]) containing 30 µg of RNase A per ml and 2 µg of RNase T1 per ml was added, and the mixture was incubated for 30 min at 30°C. The nucleases were removed by digestion with proteinase K. extracted with phenol-chloroform, and ethanol precipitated with 5 µg of tRNA as carrier. Protected fragments were analyzed on 6% sequencing gels. The exact sizes of the protected fragments were determined by comparison with adjacent sequencing reactions.

RESULTS

Isolation and sequencing of cDNA clones containing  $\alpha$  and  $\beta$  v-ets sequences. We previously reported the cloning of p54<sup>c-ets-1</sup> cDNAs from chicken embryo or spleen cell cDNA libraries (Duterque-Coquillaud et al., in press). These libraries were screened at high density with a probe corresponding to  $\alpha$  and  $\beta$  v-ets sequences (Sau3a-HpaI 250-bp fragment) and containing 25 nucleotides from exon a also present in the p54<sup>c-ers-1</sup> cDNA clone. We focused our attention on two clones (known as lambda ets 68-1 and lambda ets 68-2) that were negative with an EcoRI-ClaI 150-bp probe specific for p54<sup>c-ets-1</sup> cDNA clones (Fig. 1) (Duterque-Coquillaud et al., in press). Lambda ets 68-1 (obtained from the chicken embryo cell cDNA library) contained two EcoRI fragments: a ca. 1.5-kbp fragment hybridizing with v-ets probes, and a 0.8-kbp fragment negative with these probes and presumably representing 3' untranslated sequences (see below). Lambda ets 68-2 (obtained from the spleen cell cDNA library) contained a single ca. 1.4-kb EcoRI fragment hybridizing with v-ets probes (Fig. 1). These data are in agreement with our previous reports that a true EcoRI site is present in the c-ets-1 cellular exon F and that some EcoRI sites escaped methylase treatment during the construction of the library (15; Duterque-Coquillaud et al., in press). The DNA of these two clones was sequenced by using the strategy outlined in Fig. 1, and the results are illustrated in Fig. 2.

Identification of a novel p68<sup>c-ets-1</sup> protein. The composite c-ets-1 mRNA sequence of the ets-68 cDNA clones was determined by the dideoxy chain-terminating method of Sanger et al. (46). The longest open reading frame starting with a methionine codon at position 70 in the nucleotide sequence is expected to encode a 485-amino-acid polypeptide with a predicted molecular mass of 53 kDa (Fig. 2). We transiently expressed these cDNA clones in COS-1 cells as previously described (Duterque-Coquillaud et al., in press). The v-ets-related EcoRI fragment of each cDNA clone was converted to blunt ends and cloned into the SmaI site of the simian virus 40-derived expression vector pSVL. After transfection into COS-1 cells, the resulting polypeptides were characterized by immunoprecipitation with anti-ets Vol. 62, 1988



FIG. 1. Structure of the chicken  $p68^{c-ets-1}$  cDNA clones compared with the chicken genomic c-ets-1 locus. (A) A schematic organization of the chicken c-ets-1 locus (A) A schematic organization of the chicken c-ets-1 locus (A) A schematic top. Nomenclature is as in reference 15 and Duterque-Coquillaud et al., in press. Only some relevant restriction sites, ClaI, EcoRI (E), HindIII (HdIII), and Hpal are shown. The restriction map of the two c-ets cDNA clones is represented below:  $\mathbb{ZZ}$ , 5' and 3' untranslated regions, the open portion, the c-ets-1 coding sequence, "E, EcoRI sites generated by the cloning. (B) Detailed restriction map of the lambda p68-2 cDNA clone showing cleavages for Bg/II (Bg), EcoRI (E), HindIII (Hd), PstI (P), PvuII (PII), and Sau3A (S). The arrows beneath this restriction map denote the sequencing strategy used. The EcoRI-HpaI and HindIII-EcoRI fragments of lambda p68-1 have also been sequenced.

sera and electrophoresis in an SDS-polyacrylamide gel (19). A predominant polypeptide of approximately 68 kDa, as well as bands of lower molecular mass (in particular a ca. 60-kDa polypeptide), not detectable in mock-transfected COS-1 cells (data not shown; Duterque-Coquillaud et al., in press), was specifically immunoprecipitated by the previously described anti-*etsA* serum (Fig. 3A, lanes 1 and 4), as well as by the anti-*etsB* serum (Fig. 3A, lanes 3 and 6) directed against the  $\alpha$ - $\beta$ -encoded domain of P135<sup>gag-myb-ets</sup> (16). This immunoprecipitation is specific, since it is abolished by anti-*ets* sera blocked with an excess of the bacterial immunoprecipital immunoprecipitation.

The apparent discrepancy between the predicted molecular mass (53 kDa) and the apparent molecular mass observed in the SDS-polyacrylamide gel (68 kDa) could reflect some intrinsic properties of *ets* proteins such as their high content of proline residues. Such a discrepancy was previously observed, although to a lesser extent (48 instead of 54 kDa), with the cDNA encoding  $p54^{e-ets-1}$  (Duterque-Coquillaud et al., in press). We therefore compared the cDNA-encoded proteins in COS-1 cells with the cellular *ets* proteins encoded in normal chicken cells. Although only the  $p54^{e-ets-1}$  protein can be observed in thymus (data not shown) or in MSB-1 T lymphoid cells (Fig. 3A, lane 10), a novel  $p68^{e-ets-1}$  protein similar in an SDS-polyacrylamide gel electrophoresis onedimensional gel to the one observed in transfected COS-1 cells can be detected at low levels in normal spleen cells (Fig. 3A, lanes 7 to 9). To further substantiate the relationship between these two proteins, we compared the onedimensional cleavage patterns obtained from these polypeptides after partial digestion with S. aureus V8 protease, a procedure particularly suited to demonstrating the extensive sequence homology between polypeptides (9). The results (Fig. 3B) demonstrate that the COS-1 cDNA-encoded and the splenic cellular proteins p68<sup>c-ets-1</sup> are identical. In addition, the spleen, the organ from which one of the cDNA libraries was constructed, expressed considerably higher levels of p54<sup>c-ets-1</sup> (Fig. 3A, lanes 7 and 8) as previously described by Chen (8). We thus conclude that the cDNAs we sequenced encompass an intact coding element for c-ets-1. This was further confirmed by nucleotide sequencing, demonstrating that the three reading frames upstream of the ATG codon 70 are closed by termination codons (Fig. 2). However, from our data it is impossible to definitively authenticate the initiation codon for p68<sup>c-ets-1</sup>. There are actually five in-frame methionine codons close to each other at the 5' end of the cDNA clone (Fig. 2). Although none of the flanking sequences of these ATG codons shows a perfect match with the Kozak consensus sequence [CC(A/G)C CATGG] (26), two of them are presumably not used (positions 85 and 175 in Fig. 2), since their context appears very unfavorable, with a T instead of a purine at position -3 of the Kozak consensus sequence, a feature believed to be important for initiation of translation (26). Thus, we do not know which of the three remaining ATG codons (positions 70, 73, and 139 in Fig. 2) in a more favorable context, if any, is the true initiation codon. The multiple products encoded by the p68<sup>c-ets-1</sup> cDNA could result from alternative initiations at other internal in-frame ATG codons (53). Relation of the p54<sup>e-ets-1</sup> cellular protein. We aligned the

nucleotide sequences of the cDNAs coding for p68<sup>c-eis-1</sup> and p54<sup>c-ets-1</sup> with that of the previously described genomic exons  $\alpha$ ,  $\beta$ , and *a* (15; Duterque-Coquillaud et al., in press). These data (summarized in Fig. 4) were used to bolster our previous suggestions regarding both the nature and origin of the  $\alpha$  and  $\beta$  sequences and alternative splicing events within the c-ets-1 locus (Duterque-Coquillaud et al., in press). In addition to some point mutations, the sequence encoding p54<sup>c-ets-1</sup> and the sequence reported here diverge upstream of nucleotide 283 (Fig. 2 and 4). This nucleotide corresponds exactly to the leftward boundary of exon a. Thus, this cellular exon is a common acceptor exon, which can be fused by splicing either to the first p54<sup>c-eis-1</sup>-encoding exon,  $I^{54}$ , or to the  $\alpha$  and  $\beta$  p68<sup>c-ets-1</sup>-encoding exons as shown by the nucleotide sequence (Fig. 4). Taken together, these data clearly demonstrate that an alternative splicing mechanism within the c-ets-1 locus generates two highly related proteins having distinct amino-terminal parts: p54<sup>c-ets-1</sup> expressed at high levels in chicken thymus, and p68<sup>c-ets-1</sup>, detected so far only in chicken spleens.

Expression and tissue specificity of the  $p68^{c-ets-1}$  mRNA. We next determined the size of the mRNA species encoding  $p68^{c-ets-1}$ . We first prepared a probe specific for  $p68^{c-ets-1}$  and devoid of common sequences. The *Eco*RI-*Pst*I 190-bp fragment of lambda p68-2 (Fig. 1B and 4) was subcloned into pGEM4. Using this fragment as a probe, we analyzed total RNA prepared from spleens obtained from 14-day-old embryos or 2-day-old chickens, from normal chicken thymus, from MSB-1 (a T lymphoid cell line transformed by Marek disease virus), and from different clones of lymphoid cells transformed by reticuloendotheliosis virus and blocked at various immature stages of differentiation (3). High levels of

TACANGTGTGGGGAGCCGTGGAGGACATAAGCTTTTTCCTTGGTCTTCCAGATAAGTAGCACCTCAGAG 63 ATG ATG AGT TAC TAC ATG GAC ACA ACC ATT GGC AGC AGG GGT CCT TAT CCT TTG GCT CGC CCT GGA GTG ATG CAA GGT GCT AGC AGC TGC 159 Met Met Ser Tyr Tyr Met Asp Thr Thr Ile Gly Ser Thr Gly Pro Tyr Pro Leu Ala Arg Pro Gly Val Met Gln Gly Ala Ser Ser Cys 30 10 TET GAS GAC CCC TEG ATE CCA TOC AGE CTE CAG TCT GCC TEC TEC CCE CCE AGE TET TEC CCE CCE TEG GAT GAG GCE GCC ATC CAG 249 Cys Glu Asp Pro Trp Met Pro Cys Arg Leu Gin Ser Ala Cys Cys Pro Pro Arg Ser Cys Cys Pro Pro Trp Asp Glu Ala Ala Ile Gin GAA GTT CCC ACT GGC CTG GAG CAC TAC AGC ACA GAT ATG GAA TGT GCA GAT GTG CCT TTG TTA ACC CCC AGC AGC AAG GAA ATG ATG TCT 339 Glu Val Pro Thr Gly Leu Glu His Tyr Ser Thr Asp Het Glu Cys Ala Asp Val Pro Leu Leu Thr Pro Ser Ser Lys Glu Met Met Ser 70 80 429 cas ger ets and see are the ast see the sea and eng eng eng egg atte sea are end the eng tes aca eng aca eng ac Gin Ala Leu Lys Ala Thr Phe Ser Gly Phe Ala Lys Glu Gin Gin Arg Leu Gly Ile Pro Lys Asp Pro Gin Gin Trp Thr Glu Thr His 100 110 120 STE COS GAC TOS GTG ATE TOS GCA GTG AAT GAG TTC AGE CTG AAG GGA GTG GAT TTC CAG AAG TTC TOC ATG AAC GGA GCT GCC CTG TOC 519 Val Arg Asp Trp Val Met Trp Ala Val Asn Glu Phe Ser Leu Lys Gly Val Asp Phe Gln Lys Phe Cya Met Asn Gly Ala Ala Leu Cys 130 140 GCC CTG GGC ANG GAG TGC TTC CTG GAG CTA GCG CCT GAC. TTT GTG GGA GAT ATC CTT TGG GAA CAC CTG GAG ATC TTG CAG AAA GAA GAA 609 Ala Leu Gly Lys Glu Cys Phe Leu Glu Leu Ala Pro Asp Phe Val Gly Asp Ile Leu Trp Glu His Leu Glu Ile Leu Gln Lys Glu Glu 160 170 180 GEA ANA CEA TAE CEA GEA ANE CEA GEG ATE GEA GEG TAT CEA GAA TEE CEG TAT AET TEA GAE TAE TTE AET TAT GEE ATE GAG AET GAG CAE Ala Lys Pro Tyr Pro Ala Ash Gly Val Ash Ala Ala Tyr Pro Glu Ser Arg Tyr Thr Ser Asp Tyr Phe Ile Ser Tyr Gly Ile Glu His 699 λsn 190 200 210 SCA CAG TGE GTG CCT CCC TCT GAG TTC TCT GAG CCC AGE TTC ATC ACA GAG TCC TAC CAG ACE CTC CAT CCC ATC AGE TCG GAA GAG CTT 789 Ser Glu Phe Ser Glu Pro Ser Phe Ile Thr Glu Ser Tyr Gln Thr Leu His Pro Ile Ser Ser Glu Glu Leu Ala Gin Cys Val Pro Pro 220 230 240 CTO TEC CTC ANG TAC GAG AAC GAC TAT CCC TCA GTC ATC CTT CGT GAC CCC GTC CAG ACG GAC TEC CTG CAG ACA GAC TAC TTC ACA ATC 879 Leu Ser Leu Lys Tyr Glu Asn Asp Tyr Pro Ser Val Ile Leu Arg Asp Pro Val Gln Thr Asp Ser Leu Gln Thr Asp Tyr Phe Thr Ile 260 250 270 Ŧ ANS CAN GAN GTO GTA ACG CCA GAC ANC ATG TOC ATG GGA CGT GCC AGT CGA GGT ANA CTG GGT GGC CAG GAC TCC TTT GAG AGC ATA GAG 969 Lys Gin Glu Val Val Thr Pro Asp Asn Met Cys Met Gly Arg Ale Ser Arg Giy Lys Leu Gly Gly Gin Asp Ser Phe Glu Ser Ile Glu Val 290 300 1059 Ser Tyr Asp Ser Cys Asp Arg Leu Thr Gln Ser Trp Ser Ser Gln Ser Ser Phe Gln Ser Leu Gln Arg Val Pro Ser Tyr Asp Ser Phe 320 310 330 GAC TCA GAG GAC TAC CCC GCT GCC CTG CCC AAC CAC AAG CCC AAG GGC ACC TTC AAG GAC TAT GTT CGA GAT CGG GCT GAC ATG AAC AAG Asp Ser Glu Asp Tyr Pro Als Als Als Leu Pro Asm His Lys Pro Lys Gly Thr Phe Lys Asp Tyr Val Arg Asp Arg Als Asp Met Asm Lys 1149 ALa 350 360 GAC AAG CCT GTC ATT CCT GCC GCT GCC GCC GCC TAC ACA GGC AGT GGA CCC ATC CAA CTG TGG CAA TTC CTG CTG GAG CTG CTC ACT Asp Lys Pro Val Ile Pro Ala Ala Ala Ala Lau Ala Gly Tyr Thr Gly Ser Gly Pro Ile Gln Lau Trp Gln Phe Lau Lau Glu Lau Lau Thr 1239 370 380 390 GAC ANS TEE TOT ENG TEE THE ATE AGE TES AGE GET GAT GEE TES GAS ITE ANS ETT TEE GAT ECA GAT GAS STE SEE AGE EGE TES GEE 1329 Asp Lys Ser Cys Gin Ser Phe Ile Ser Trp Thr Gly Asp Gly Trp Glu Phe Lys Leu Ser Asp Pro Asp Glu Val Ala Arg Arg Trp Gly 400 410 420 ANG AGG ANA ANC ANG CCC ANG ATG ANC TAT GAG ANG CTG AGC CGT GGT CTG CGT TAC TAT TAC GAC ANG ANC ATC ATC CAC ANG ACG GCC Lys Arg Lys Asn Lys Pro Lys Het Asn Tyr Glu Lys Leu Ser Arg Gly Leu Arg Tyr Tyr Tyr Asp Lys Asn Ile Ile His Lys Thr Ala Asp 430 440 Val 450 1419 450 CAC TCA TCA GCA TCT GGC TTG ACG GEC AND CEC THE GTE THE CEC TTE GTE TEE GAE CTE CAS AGE CTE CTE GEC TAE ACA CEA GAE GAE GTE ANG CTE CAE GEC ATE CTE GAE GTE ANG 1509 Gly Lys Arg Tyr Val Tyr Arg Phe Val Cys Asp Leu Gin Ser Leu Leu Gly Tyr Thr Pro Glu Glu Leu His Ala Met Leu Asp Val Lys 460 470 His Ser Ser Ala Ser Gly Leu Thr TCC AGC ATG GCG TGC AGC TCC TTT TGA CCA GAT GCT GAT GAG TGA atte 152 . Pro Asp Ala Asp Glu \*\*\* Ser Ser Met Ala Cys Ser Ser Phe \*\*\*

FIG. 2. Nucleotide sequence of 1.5 kbp of the c-ets-1 cDNA clone deduced from that of cDNA clones p68-1 and p68-2. The translation of this sequence in the long open reading frame, which extends from the first in-frame ATG codon at position 70 to the termination codon TGA at position 1524, is shown. Termination codons upstream of the first ATG are underlined. Only the nucleotides and amino acids in v-ets that differ from c-ets-1 are indicated and boxed.

 $p68^{c-ets-1}$  mRNA were detected in spleen cells (Fig. 5A, lanes 6 and 7) and in poly(A)<sup>+</sup> RNA prepared from 11-day-old chicken embryos (data not shown), from which the p68 cDNAs have been isolated. The lack of hybridization signal in thymus and lymphoid cell lines (Fig. 5A, lanes 1 and 5)

was in agreement with our previous results (15, 16) and was significant since the same blot hybridized with a common *ets* probe encompassing exons a to F (1.2-kbp *HpaI-Eco*RI fragment) revealed the major 7.5-kb c-*ets-l* RNA species (Fig. 5B, lanes 1 to 7). The 2.2- and 1.5-kb RNA species,



FIG. 3.  $p68^{c-ets-l}$  transiently expressed in COS-1 cells transfected by  $p68^{c-ets-l}$  cDNA is similar to a cellular protein expressed in spleen cells. (A) COS-1 cells were transfected with 10 µg of plasmid DNA, labeled 72 h later with [<sup>35</sup>S]methionine, and analyzed by immunoprecipitation with anti-etsA serum (lanes 1 and 4), anti-etsA adsorbed with an excess of bp17<sup>MS2-etsA</sup> (lanes 2 and 5), and anti-etsB serum directed against the  $\alpha$ - $\beta$ -encoded domain of P135<sup>gag-myb-ets</sup> (lanes 3 and 6) (16). Cell suspensions obtained from the spleen of a 2-day-old chicken (lanes 7 through 9) were treated in the same way after preincubation with protein A-Sepharose beads (see Materials and Methods), as well as MSB-1 T lymphoid cells. The immunoprecipitation of  $p68^{c-ets-l}$  by anti-etsB sera, which is difficult to see in the figure (lane 9), was clearly visible on the original autoradiogram. Prestained protein molecular weight standards (Bethesda Research Laboratories, Inc.) were run on the SDS-polyacrylamide gel and used to determine the sizes of the immunoprecipitated proteins (not shown). (B) One-dimensional V8 protease mapping of COS-1 cDNA-encoded and normal  $p68^{c-ets-l}$  proteins. Spleen cells were labeled with 2 mCi of L-[<sup>35</sup>S]methionine per ml and treated as described in Materials and Methods. COS-1 cells were transfected with p68-2 cDNA clone and labeled 72 h later with 0.5 mCi of L-[<sup>35</sup>S]methionine. Labeled proteins were excised from gels similar to those in panel A and run on a 15% polyacrylamide gel after partial digestion for 30 min with either 20 or 200 ng of S. aureus V8 protease. Lanes: Sp, proteins obtained from spleen cells; Cos-1, proteins obtained from transfected COS-1 cells. The positions of prestained protein molecular weight standards are shown on the right (in thousands).

difficult to see on this autoradiogram, were clearly visible on a longer exposure (data not shown).

Interestingly enough, the major p68<sup>c-ers-1</sup> RNA was also 7.5 kb in size (Fig. 5A, lanes 6 and 7) and could not be resolved from the p54<sup>c-ets-1</sup> mRNA species even by extensive electrophoresis in low-concentration agarose gels (data not shown). The existence of two distinct proteins encoded by two distinct cDNAs containing alternative 5' exons upstream of the common exon a strongly supports the idea that the 7.5-kb RNA observed in spleen cells was in fact a doublet. To demonstrate this, we performed RNase protection experiments (27). We subcloned in pGEM4 an EcoRI-BglII 566-bp fragment from the p68-2 cDNA clone. The probe (608 nucleotides) complementary to the mRNA (cRNA) was transcribed from the T7 polymerase promoter and encompassed 42 nucleotides from the plasmid promoter and polylinker, 305 nucleotides from the common exons (a and part of b) and 261 nucleotides corresponding to the p68<sup>c-ets-1</sup>-specific 5' exons (Fig. 5C, bottom). RNA species specific for p54<sup>c-ets-1</sup> should protect 305 nucleotides of the cRNA (common fragment) from RNase digestion. Such RNase-resistant hybrids were formed with RNAs extracted from thymus or lymphoid cell lines (Fig. 6C, lanes 3 to 5). In contrast, with RNAs prepared from spleen cells, fully protected fragments were observed in addition to the 305nucleotide hybrids (Fig. 5C, lane 6).

The apparent ratio (ca. 1:20) in Fig. 5C, lane 6, between the fully protected fragments ( $p68^{c-ets-1}$  specific) and the common fragments ( $p68^{c-ets-1}$  and  $p54^{c-ets-1}$ ) seemed to be in good correlation with the relative ratio of  $p54^{c-ets-1}$  and  $p68^{c-ets-1}$  proteins synthesized in spleen cells (Fig. 3A, lane 7).

Thus, a single transcriptional unit directs the synthesis of two similar-sized mRNAs (7.5 kb) with differential expression and encodes two highly related proteins (Fig. 7, top row).

Transduction of c-ets. Sequences of v-ets and of the p68<sup>c-e/s-1</sup> cDNA are highly homologous except for their carboxy-terminal parts. The cellular protein p68<sup>c-ets-1</sup>, as well as p54<sup>c-ets-1</sup>, diverges from the viral protein in the last 13 codons. Seven point mutations, five of which are conservative, are found between p68<sup>c-ets-1</sup> and v-ets (Fig. 2). These reflect the differences observed between the common set of cellular exons (a to F) and their viral homologous sequences. Otherwise,  $p68^{c-ets-1}$  and the v-ets sequence appear colinear, especially in their 5' parts (Fig. 4). We conclude that p68<sup>c-ets-1</sup>, rather than p54<sup>c-ets-1</sup>, should be considered the cellular progenitor of the viral ets oncogene. Focusing now on the 5' end of the p68 cDNA, its nucleotide sequence appears colinear with that of v-ets until nucleotide 59, which corresponds to the leftward boundary of exon  $\alpha$  on one hand and to the myb-ets boundary in the E26 genome on the other. This observation deserves two comments. First, if we assume that the ATG codon 70 is not the true initiation codon, 11 nucleotides or more which are noncoding in p68<sup>c-ers-1</sup> since they are located upstream of the first ATG codon, are converted into coding sequences in the viral protein (Fig. 4). Second, the observation that the myb-ets boundary corresponds exactly to the boundary of c-ets-l exon  $\alpha$  might not be coincidental and suggests a region of splicing. Therefore, we compared the *myb-ets* sequence with c-ets-1 and c-myb genomic sequences and searched for a putative splice donor consensus sequence in this region of c-myb (Fig. 6). In fact, the rightward recombination point in E26 v-myb, although located in the middle of a cellular c-myb coding exon named E6 (24), displays salient homology (6 of 10 nucleotides conserved) with a splice donor site consensus sequence (CAAGGgtagagt) (36) (Fig. 6).

In conclusion, the transduction of the c-ets proto-oncogene presumably occurred via an abnormal splicing event between a cryptic splice donor site in c-myb exon E6 and the normal splice acceptor site of c-ets-1 exon  $\alpha$  as diagrammat-

#### 3238 LEPRINCE ET AL.

c-DNA p68	TACAAGTGTGGGGAGCCG
	68-2
" genomic	
C-DNA P68	TGGAGGACA <u>RAN</u> GCTTTTTCCTTGGTCTTCCAG <u>XTAAQTAQ</u> CACCTCAGAG ()
E26	······································
	v-myb <del>()</del> v-ets
" genomic	*** *** *** *** *** *** *** *** *** *** *** ***
C-DNA P68	ATG ATG AGT TAC TAC ATG GAC ACA ACC ATT GGC AGC ACG 108
E26	*** *** *** *** *** *** *** *** *** *** ***
	*.a 1
u genomic	*** *** *** *** *** *** *** *** *** GTG <u>ATG</u> gta aga
c-DNA p68	GUT CCT TAT CCT TTG GCT CGC CCT GGA GTG ATG CAA GGT 147
E26	*** *** *** *** *** *** *** *** *** *** ***
8 genomic	CAR ANT CAA GGT
	_ <b>T</b> .
β genomic	*** *** *** *** *** *** *** *** *** *** ***
C-DNA p68	GCT AGE AGE TGE TGT GAG GAE CEE TGG ATG CEA TGE AGG 186
E26	*** *** *** *** *** *** *** *** *** *** *** *** ***
	Pat I
ßgenomic	
C-DNA D68	CTG CMG TCT GCC TGC TGC CCG CCC MGG TCG TGT TGC CCG 225
C-DNA p54	GCT "C" G"C T"A AC" ATG AA" GCG GC" GT" GAC CTG AA"
E26	*** *** *** *** *** *** *** *** *** *** ***
β genomic	*** *** *** *** *** *** *** *** *** *** *** *** ***
C-DNA p68	CCG TGG GAT GAG GCG GCC ATC CAG GAA GTT CCC ACT GGC 264
c-DNA p54	**C ACC CTG ACC ATC ATC *AG AC* **G AAG GTG GAC AT*
E26	*** *** *** *** *** *** *** *** *** *** *** ***
	*jd
βgenomic	*** *** *** AGC ACA Ggt agg g
C-DNA p68	CTG GAG CAC TAC AGE ACA GAT ATG GAA TGT GCA GAT GTG 303
c-DNA p54	GAT CTC TT* CCG TC* C*C *** *** *** *** *** ***
226	*** *** *** *** *** *** *** *** *** *** *** *** ***
a genemic	··· ··· ttt tca gAT ATG *** *** *** ***
	<u>- Eps I</u>
C-DNA p68	CCT TTG TTA AC T. S 314
c-DNA p54	*** *** *** **
126	*** *** *** **
a genomic	*** *** *** **

FIG. 4. Comparison of the two-c-ets-1 cDNA types with v-ets and their corresponding genomic clones. The sequence of the 5' end of the  $p68^{c-ets}$  cDNA clones is aligned with that of v-ets (39, 40),  $p54^{c-ets-1}$  cDNA (Duterque-Coquillaud et al., in press), and the relevant parts of genomic  $\alpha$ ,  $\beta$ , and a exons (15; Duterque-Coquillaud et al., in press). Asterisks (\*) indicate identical sequences shared between the sequences. The putative ATG codons are underlined, whereas the termination codons are boxed. The intron consensus splice junctions (36) are indicated by an arrow with s. a. or s. d. The conserved GT and AG nucleotides are written in lower-case letters and underlined. Numbering in the right-hand column is similar to that used in Fig. 2.

ically outlined at the bottom of Fig. 7. In addition, the *myb-ets* boundary is not located at the second G or the GGC codon as previously suggested (39, 40), but, rather, at the first G of this codon (Fig. 6).

#### DISCUSSION

To understand how a cellular proto-oncogene, c-ets-1, has been converted into an activated viral oncogene, v-ets, we have compared the cellular ets coding sequences isolated from both genomic DNA (15) and cDNA cloning (Duterque-Coquillaud et al., in press) to the viral ets sequence. These studies confirmed and extended the relationship between the major c-ets-1-encoded protein, p54<sup>c-ets-1</sup>, and the etsencoded domain of the viral transforming protein P135<sup>gag-myb-ets</sup> (19). However, the nature and origin of the  $\alpha$ and  $\beta$  sequences remained obscure, since they were homologous to the 5' part of v-ets and displayed structural features of eucaryotic exons, but were never detected in c-ets-1 RNA species or proteins. These sequences deserved special attention, since they represented the leftward recombination point of c-ets-1 with c-myb in E26. In this communication we report the cloning and sequencing of a new c-ets-l cDNA which allowed us to draw several conclusions regarding both the transcription of the cellular c-*ets-l* locus and the transduction of c-*ets*.

First, these results clearly demonstrate that an alternative splicing mechanism is operative in the chicken c-ets-1 locus. Alternative splicing has been found in a number of DNA viruses and in cellular genes, including several proto-oncogenes (reviewed in reference 28). In the proto-oncogenes, alternative splicing (32, 48, 59) could affect the level of gene expression by selection of different promoters or could generate different proteins (i) by insertion of small amino acid sequences as in the examples of the brain-specific form of pp60<sup>c-src</sup> (31) and chicken c-mil (14) or (ii) by addition of different 5' exons onto a common set of 3' exons as in the examples of the murine (1) or human (49) c-abl genes and the chicken c-myb gene (17, 45). The alternative splicing in chicken c-ets-1 that we report here is an example of the latter case and yields two highly related proteins having different amino-terminal parts. Although the function of the c-ets-1 proteins is not known, it is tempting to speculate that they could have a common function, encoded by their common region (a to F), but directed towards different targets according to their specific N-terminal part encoded by the alternate 5' exons ( $\alpha$ ,  $\beta$ , or  $I^{54}$ ). In that view, it is important to note the high level of proline residues of the  $\alpha$  and  $\beta$  exons (10 residues of 71 [14%]), a criterion which has often been associated with nuclear proteins (58).

Although  $p54^{c-ets-1}$  has been described as a cytoplasmic protein (19), the chicken and mouse c-ets-2 products have been localized in the nucleus (4, 7). Preliminary results indicate that in cell fractionation experiments of COS-1 cells transfected by the  $p68^{c-ets-1}$  cDNA clones, a small but significant proportion of  $p68^{c-ets-1}$  could be located in the nucleus (D. Leprince, unpublished results).

Splicing of c-ets-1 exon a to upstream sequences can occur over long distances (ca. 40 kbp). In addition, splicing to exons  $\alpha$  and  $\beta$  jumps exon  $I^{54}$ . This observation supports previous experiments in vivo (1) and in vitro (51) which argued against the progressive scanning mechanism for splicing. The mechanism by which the complete  $p54^{c-ets-1}$ and  $p68^{c-ets-1}$  mRNA species are generated is not clear. In particular, it remains to be determined how the 5' and 3' noncoding sequences of these RNAs relate to each other. Cloning and sequencing of longer cDNAs are required to resolve this point.

These experiments also address the differential expression of the c-ets-1 proteins. The ubiquitous expression of p54<sup>c-ets-1</sup> predominantly in thymic and bursal lymphocytes versus the expression of p68<sup>c-ets-1</sup> found so far only in the spleen could result from a tissue-specific splicing of the same transcriptional unit, from the use of different promoters, from different 5' and 3' noncoding exons yielding two distinct but similar-sized mRNAs, or from a combination of these mechanisms. The restricted expression of p68<sup>c-e/s-1</sup> to the spleen, which is a peripheral lymphoid tissue, and the wide expression of  $p54^{c-ers-1}$  in the thymus and bursa of Fabricius, which are central lymphoid tissues, are characteristic of a protein family involved in various stages of the lymphoid maturation pathway. However, the ontogeny of the spleen is quite complex. Although its chief functions are the manufacture of leukocytes and the destruction of erythrocytes, during embryonic life the spleen is hemopoietic, with erythropoiesis and granulopoiesis both being very ac-tive (43). Thus,  $p68^{e-ets-1}$  could be expressed in a nonlymphoid cell type. The nature of the cell type(s) expressing

VOL. 62, 1988



FIG. 5. Tissue-specific expression of  $p68^{c-ets-l}$  mRNA. (A) Hybridization to a cDNA probe specific for  $p68^{c-ets-l}$  (EcoRI-PstI 190-bp fragment) to the indicated chicken tissues and lymphoid cells lines. Autoradiography was for 3 days. (B) The same Northern (RNA) blot was stripped and rehybridized with a probe corresponding to the common exons (a to F) (EcoRI-HpaI; 1.2-kbp fragment). Autoradiography was for 1 day. (C) RNase protection experiments. The structure of the [<sup>32</sup>P]cRNA probe and predicted fragments is shown diagrammatically at the bottom of the figure. Boxed portions of the restriction enzyme cleavage maps indicate cDNA sequences; single lines indicate plasmid sequences. Transcription orientation with the T7 promoter is also indicated. Restriction sites; B, BamHI-BgIII; E, EcoRI; H, HpaI. Lanes: 1, probe only; 2, control tRNA coli (5 µg); 3, total RNA from 2-day-old chicken thymus (20 µg); 4, total RNA from 2-week-old chicken thymus 20 µg; 5, total RNA from the T lymphoid cell line MSB1 20 µg; 6, total RNA from 2-day-old chicken spleen (20 µg). Exposure was for 12 h for lanes 1 to 4 and 2 days for lanes 5 and 6. The extra bands in lane 6 represent full protection of degradation products of the probe (compared with lane 1; the bands are trimmed by 42 bp corresponding to the polylinker).

p68<sup>c-ets-1</sup> is currently being investigated by in situ hybridization and analyses of purified spleen cell populations.

Second, our results shed light on the mechanism of transduction of *c-ets*. In striking contrast with the other avian retroviruses carrying two onc genes (avian erythroblastosis virus and MH2) (10, 57), in E26 the two oncogenes are expressed in a single fusion protein of  $M_r$  135,000 (P135<sup>gag-myb-ets</sup>) (6, 23). We demonstrate that the two cell-derived oncogenes were apparently fused by an abnormal splicing mechanism between a cryptic splice donor site in the coding *myb* sequence (E6) and the normal splice acceptor site of the *c-ets-l* exon  $\alpha$  (Fig. 6 and 7). Such aberrant splicing events have been previously reported for the acti-



FIG. 6. Genesis of the *myb-ets* sequence of E26 by aberrant splicing between a cryptic splice donor site in c-myb exon E6 and the normal splice acceptor site of c-ets-1 exon  $\alpha$ . The 5' point of c-ets transduction corresponds to the splice acceptor (s. a.) site of c-ets-1 exon  $\alpha$ . The 3' point of c-myb transduction, although located in a c-myb-coding exon, namely E6, (24), presents salient homology with a splice donor consensus sequence (36). The arrow in E26 indicates the accurate myb-ets boundary (the present study), which was previously assigned to the following G based on homology with the c-myb sequence (39, 40).

vation of the mouse c-myb oncogene in Abelson virusinduced plasmacytoid lymphosarcomas (44) and for the chicken c-erbB activation in avian leukosis virus-induced erythroblastosis (38). Since a recombination between processed mRNA species occurring exactly at splice junctions seems unlikely, we suggest instead a mechanism based on one of the following models: (i) intermolecular splicing between RNA molecules (25, 50); (ii) a DNA translocation



FIG. 7. Alternative splicing within the c-ets-l locus and genesis of the E26 myb-ets sequence. From top to bottom: schematic drawing of the c-ets-l locus organization and of the two c-ets cDNA types generated by alternative splicing within this locus. The E26 myb-ets sequence is generated through an abnormal splicing mechanism.

3239

event between c-myb, c-ets-1, and a proviral DNA; and (iii) integration of a virus carrying only one of the oncogenes in the vicinity of the second cellular proto-oncogene.

Although we have no experimental evidence for the third model, it seems more favorable than the other models. Furthermore, an abnormal splicing event has been proven to be efficient in avian leukosis virus-induced erythroblastosis. In that particular situation, an avian leukosis virus integrates into the c-erbB gene and, via an aberrant splicing between env and erbB sequences, generates with a high frequency (about 50%) retroviruses carrying truncated erbB oncogenes (34, 38).

A second recombination between the hybrid RNA transcribed from this first recombinational event and the RNA genome of a helper virus could occur via "copy choice" during reverse transcription (5, 21, 56). During this last step, the reverse transcriptase could also generate the differences observed in the carboxy-terminal part of v-ets as compared with c-ets-1 (15; Duterque-Coquillaud et al., in press) (Fig. 2). This alteration in its carboxy-terminal part represents, together with its linkage to the nuclear oncogene myb, the major differences observed between the coding regions of the cellular protein p68<sup>c-ets-1</sup> and the v-ets-encoded part of the viral transforming protein P135gag-myb-ets. Similar examples of cellular and viral proteins differing at their carboxyterminal ends and the biological relevance of this alteration have been well documented (11, 12, 37, 54, 55). With the availability of our cDNA clones, we are currently constructing E26 mutants in which different parts of the v-ets oncogene, especially the 3' end, are replaced by corresponding parts of the cDNA to analyze the relevance of these mutations in the oncogenic spectrum of E26.

#### ACKNOWLEDGMENTS

We are indebted to K. Strebhardt and J. Mullins, Harvard University, and to J. Sap and B. Vennstrom, European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany, for the gift of their cDNAs libraries; A. Gegonne and J. Ghysdael for the generous and constant supply of anti-ets sera; B. Vandenbunder for synthesizing the cRNA probe; and T. Graf and H. Beug for providing the reticuloendotheliosis virus-transformed clones. We thank many colleagues for stimulating discussions and helpful advice; K. E. Boulukos for a critical reading of the manuscript; D. Hetuin, C. Denis, B. Quatannens, and C. Lagrou for expert technical help; P. Martin for his patience and invaluable help with the computer-assisted graphics; and Nicole Devassine and Valérie Dufresnoy for patient typing.

This work was supported by the Institut National de la Santé et de la Recherche Médicale (U.124-U.186), Centre National de la Recherche Scientifique (UA 04-1160), Association pour la Recherche sur le Cancer, and the Pasteur Institute of Lille.

#### LITERATURE CITED

- Ben-Neriah, Y., A. Bernard, M. Paskind, G. O. Daley, and D. Baltimore. 1986. Alternative 5' exons in *c-abl* mRNA. Cell 44: 577-586.
- 2. Beug, H., A. Leutz, P. Kahn, and T. Graf. 1984. Ts mutants of E26 leukemia virus allow transformed myeloblasts, but not erythroblasts or fibroblasts to differentiate at the non-permissive temperature. Cell 39:579-588.
- Beug, H., H. Muller, S. Grieser, G. Doderlein, and T. Graf. 1981. Hematopoietic cells transformed in vitro by REV-T avian reticuloendotheliosis virus express characteristics of very immature lymphoid cells. Virology 115:295-309.
- Bhat, N. K., R. J. Fisher, S. Fujiwara, R. Ascione, and T. S. Papas. 1987. Temporal and tissue specific expression of mouse ets genes. Proc. Natl. Acad. Sci. USA 84:3161-3165.
- 5. Bishop, J. M. 1983. Cellular oncogenes and retrovirus. Annu. Rev. Biochem. 52:301-354.

- Bister, K., M. Nunn, C. Moscovici, B. Perbal, M. A. Baluda, and P. H. Duesberg. 1982. Acute leukemia viruses E26 and avian myeloblastosis virus have related transformation specific RNA sequences but different genetic structures, gene products and oncogenic properties. Proc. Natl. Acad. Sci. USA 79:3677– 3681.
- Boulukos, K. E., P. Pognonec, A. Begue, F. Galibert, J. C. Gesquiere, D. Stehelin, and J. Ghysdael. 1988. Identification in chickens of an evolutionary conserved cellular ets-2 gene (c-ets-2) encoding nuclear proteins related to the products of the c-ets proto oncogene. EMBO J. 7:697-705.
- 8. Chen, J. H. 1985. The proto-oncogene c-ets is preferentially expressed in lymphoid cells. Mol. Cell. Biol. 5:2993-3000.
- Cleveland, D. W., S. G. Fischer, M. W. Kirschner, and U. K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analyses by gel electrophoresis. J. Biol. Chem. 252:1102-1106.
- Coll, J., M. Righi, C. De Taisne, C. Dissous, A. Gegonne, and D. Stehelin. 1983. Molecular cloning of the avian acute transforming retrovirus MH2 reveals a novel cell derived sequence (v-mil) in addition to the myc oncogene. EMBO J. 2:2189-2194.
- Cooper, J. A., K. L. Gould, C. A. Cartwright, and T. Hunter. 1986. Tyr 527 is phosphorylated in pp60<sup>c-src</sup>. Implications for regulation. Science 231:1431-1434.
- Coussens, L., C. Van Beveren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma, and A. Ullrich. 1986. Structural alteration of viral homologue of receptor protooncogene *fins* at carboxyl-terminus. Nature (London) 310:277-280.
- De Taisne, C., A. Gegonne, D. Stehelin, A. Bernheim, and R. Berger. 1984. Chromosomal localization of the human protooncogene *c-ets*. Nature (London) 310:581-583.
- Dozier, C., F. Denhez, C. Henry, J. Coll, A. Begue, B. Quatannens, S. Saule, and D. Stehelin. 1988. Alternative splicing of RNAs transcribed from the chicken c-mil gene. Mol. Cell. Biol. 8:1835-1838.
- Gegonne, A., D. Leprince, M. Duterque-Coquillaud, B. Vanderbunder, A. Flourens, J. Ghysdael, B. Debuire, and D. Stehelin. 1987. Multiple domains for the chicken cellular sequences homologous to the v-ets oncogene of the E26 retrovirus. Mol. Cell. Biol. 7:806-812.
- Gegonne, A., D. Leprince, P. Pognonec, D. Dernis, M. B. Raes, D. Stehelin, and J. Ghysdael. 1987. The 5' extremity of the v-ets oncogene of avian leukemia virus E26 encodes amino-acid sequences not derived from the major c-ets encoded cellular proteins. Virology 156:177-180.
- Gerondakis, S., and J. M. Bishop. 1986. Structure of the protein encoded by the chicken proto-oncogene c-myb. Mol. Cell. Biol. 6:3677-3684.
- Ghysdael, J., A. Gegonne, P. Pognonec, K. Boulukos, D. Leprince, D. Dernis, C. Lagrou, and D. Stehelin. 1986. Identification in chicken macrophages of a set of proteins related to, but distinct from, the chicken cellular *c-ets* encoded protein p54<sup>c-ets</sup>. EMBO J. 5:2251-2256.
- Ghysdael, J., A. Gegonne, P. Pognonec, D. Dernis, D. Leprince, and D. Stehelin. 1986. Identification and preferential expression in thymic and bursal lymphocytes of a *c-ets* oncogene encoded Mr 54,000 cytoplasmic protein. Proc. Natl. Acad. Sci. USA 83: 1714–1718.
- Gluzmann, Y. 1981. SV 40-transformed simian cells support the replication of early SV 40-mutants. Cell 23:175-182.
- Huang, C. C., N. Hay, and J. M. Bishop. 1986. The role of RNA molecules in transduction of the proto-oncogene *c-fps*. Cell 44: 935-940.
- Ivanov, X., Z. Mladenov, S. Nedyalkov, and T. Todorov. 1962. Experimental investigation into avian leukosis. I. Transmission experiments of certain diseases of the avian leukosis complex found in Bulgaria. Bull. Inst. Pathol. Comp. Anim. Domest. 9: 5-36.
- Klempnauer, K., G. Symonds, G. Evan, and J. M. Bishop. 1984. Subcellular localization of proteins encoded by oncogenes of avian myeloblastosis virus and avian leukemia virus E26 and by the chicken *c-myb* gene. Cell 37:537-547.

VOL. 62, 1988

- 24. Klempnauer, K. H., T. J. Gonda, and J. M. Bishop. 1982. Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb the architecture of a transduced oncogene. Cell 31:453-463.
- Konarska, M. M., R. A. Padgett, and P. A. Sharp. 1985. Trans splicing of mRNA precursors in vitro. Cell 42:165-171.
- 26. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283-292.
- Kruiser, W., J. A. Cooper, T. Hunter, and I. M. Verma. 1984. Platelet derived growth factor induces rapid but transient expression of the *c-fos* gene and protein. Nature (London) 312: 711-716.
- Leff, S. E., M. G. Rosenberg, and R. M. Evans. 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annu. Rev. Biochem. 55:1091-1117.
- Leprince, D., A. Gegonne, J. Coll, C. De Taisne, A. Schneeberger, C. Lagrou, and D. Stehelin. 1983. A putative second cell-derived oncogene of the avian leukemia retrovirus E26. Nature (London) 306:395-397.
- Luthmann, H., and G. Magnusson. 1983. High efficiency polyoma DNA transfection of chloroquine treated cells. Nucleic Acids Res. 11:1295-1308.
- Martinez, R., B. Mathey-Prevot, A. Bernards, and D. Baltimore. 1987. Neuronal pp60<sup>c-src</sup> contains a six-amino acid insertion relative to its non-neuronal counterpart. Science 237:411-415.
- Matlashewski, G., D. Pim, L. Bank, and L. Crawford. 1987. Alternative splicing of human p53 transcripts. Oncogene Res. 1: 77-85.
- Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309-321.
- Miles, B. D., and H. L. Robinson. 1985. High-frequency transduction of c-*erbB* in avian leukosis-induced erythroblastosis. J. Virot. 54:295-303.
- Moscovici, C., J. Samarut, L. Gazzolo, and M. G. Moscovici. 1981. Myeloid and erythroid neoplastic response to avian defective leukemia viruses in chicken and quails. Virology 111:765-768.
- Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. 10:459-472.
- 37. Neckamayer, W. S., M. Shibuya, M. Hsu, and L. Wang. 1986. Proto-oncogene *c-ros* codes for a molecule with structural features common to those of growth factor receptors and displays tissue-specific and developmentally regulated expression. Mol. Cell. Biol. 6:1478-1486.
- Nilsen, T. W., P. A. Maroney, R. G. Goodwin, F. M. Rottman, L. B. Crittenden, M. A. Raines, and H. J. Kung. 1985. *c-erbB* activation in ALV-induced erythroblastosis: novel RNA processing and promotor insertion result in expression of an aminotruncated EGF receptor. Cell 41:719–726.
- 39. Nunn, M., H. Weiher, P. Bullock, and P. Duesberg. 1984. Avian erythroblastosis virus E26: nucleotide sequence of the tripartite onc gene and of the LTR and analysis of the cellular prototype of the viral ets gene. Virology 139:330-339.
- Nunn, M. F., P. H. Seeburg, C. Moscovici, and P. H. Duesberg. 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. Nature (London) 306:391-395.
- 41. Radke, K., H. Beug, S. Kornfeld, and T. Graf. 1982. Transformation of both erythroid and myeloid cells by E26, an avian leukemia virus that contains the *myb* gene. Cell 31:643-653.
- 42. Reddy, E. S., V. N. Rao, and T. S. Papas. 1987. The erg gene: a human gene related to the ets oncogene. Proc. Natl. Acad. Sci.

USA 84:6131-6135.

- 43. Romanoff, A. L. 1960. The avian embryo, p. 580-583. MacMillan, New York.
- Rosson, D., D. Dugan, and E. P. Reddy. 1987. Aberrant splicing events that are induced by proviral integration: implications of myb oncogene activation. Proc. Natl. Acad. Sci. USA 84:3171-3175.
- Rosson, D., and E. P. Reddy. 1986. Nucleotide sequence of chicken *c-myb* complementary DNA and implications for *myb* oncogene activation. Nature (London) 319:604-606.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 47. Sap, J., A. Munoz, K. Damm, Y. Goldberg, J. Ghysdael, A. Leutz, H. Beug, and B. Vennstrom. 1986. The *c-erbA* protein is a high affinity receptor for thyroid hormone. Nature (London) 324:635-639.
- Shimizu, K., D. Birnbaum, M. A. Ruley, O. Fasano, Y. Suard, L. Edlund, E. Taparowsky, M. Goldfarb, and M. Wigler. 1983. Structure of the *Ki-ras* gene of the human lung carcinoma cell line Calu-1. Nature (London) 304:497-500.
- 49. Shtivelman, E., B. Lifshitz, R. P. Gale, B. A. Roe, and E. Canaani. 1986. Alternative splicing of RNAs transcribed from the human *abl* gene and from the *bcr-abl* fused gene. Cell 47: 277-284.
- 50. Solnick, D. 1985. Trans splicing of mRNA precursors. Cell 42: 157-164.
- Solnick, D. 1985. Alternative splicing caused by RNA secondary structure. Cell 43:667-676.
- Sompayrac, L. M., and K. J. Danna. 1981. Efficient infection of monkey cells with DNA of simian virus 40. Proc. Natl. Acad. Sci. USA 78:7575-7578.
- 53. Strubin, M., E. O. Long, and B. Mach. 1986. Two forms of the Ia antigen-associates invariant chain result from alternative initiations at two in phase AUGs. Cell 47:619-625.
- 54. Takeya, T., and H. Hanafusa. 1983. Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. Cell 32:881– 890.
- 55. Van Beveren, C., F. Van Straaten, T. Curran, R. Muller, and I. M. Verma. 1982. 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.
- Varmus, H. E. 1982. Form and function of retroviral proviruses. Science 216:812-820.
- Vennstrom, B., and J. M. Bishop. 1982. Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. Cell 28:135-143.
- 58. Verma, I. 1984. From c-fos to v-fos. Nature (London) 308:317.
- Wang, L., S. Lijima, T. Dorai, and B. Lin. 1987. Regulation of the expression of proto-oncogene *c-src* by alternative RNA splicing in chicken skeletal muscle. Oncogene Res. 1:43-59.
- 60. Watson, D. K., M. J. McWilliams-Smith, C. Kozak, R. Reeves, J. Gearhart, M. F. Nunn, W. Nash, J. R. Fowle III, P. Duesberg, T. S. Papas, and S. J. O'Brien. 1986. Conserved chromosomal positions of dual domains of the *ets* proto-oncogene in cats, mice and humans. Proc. Natl. Acad. Sci. USA 83:1792-1796.
- 61. Watson, D. K., M. J. McWilliams-Smith, M. F. Munn, P. H. Duesberg, S. J. O'Brien, and T. S. Papas. 1985. The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. Proc. Natl. Acad. Sci. USA 82:7294-7298.

## RESUME

Le rétrovirus E26 induit chez le poulet des leucémies de type mixte érythoïde/ myéloïde à prédominance érythroïde. Son génome contient 2 oncogènes v-myb et v-ets traduits en une protéine de fusion P135gag-myb-ets,

Nous avons étudié la structure et l'expression du locus c-ets-1, équivalent cellulaire de v-ets chez le poulet.

Les séquences homologues au gène viral se répartissent en 2 domaines séparés par 40 kbp d'ADN cellulaire non apparenté à v-ets. Le premier domaine contient 2 exons alpha et bêta, correspondant à l'extrémité 5' de v-ets, ces exons ne sont pas transcrits dans l'ARNm c-ets codant pour la protéine P54<sup>c-ets-1</sup>. Le second domaine contient 6 exons (a-f) qui rendent compte de la majeure partie de v-ets, à l'exception de la partie 3' de l'exon f qui en diverge. P54<sup>c-ets-1</sup> et la partie v-ets de P135<sup>gag-myb-ets</sup> diffèrent donc à leurs extrémités N- et C- terminales. L'étude des ADNc des ARNm c-ets-1 de cellules d'embryon et de rate de poulet a montré que le locus c-ets-1, par un mécanisme d'épissage alternatif, est à l'origine de 2 protéines distinctes mais apparentées : P54<sup>c-ets-1</sup> exprimée à un taux élevé dans les organes lymphoïdes et P68<sup>c-ets-1</sup> (possédant les séquences alpha et bêta en position Nterminale) exprimée à un taux relativement faible dans les cellules de rate de poulet. L'ARNm P68<sup>c-ets-1</sup> serait donc le progéniteur direct de v-ets.

Il apparait d'autre part que la jonction **myb-ets** dans le virus provient d'un épissage aberrant entre un site d'épissure donneur cryptique de l'exon 6 de **c-myb** et le site d'épissure accepteur normal de l'exon alpha de **c-ets-1**.

## **MOTS CLEFS:**

- Leucémie
- Rétrovirus
- Oncogène
- Epissage alternatif
- Transduction