

50376
1995
235

UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

THESE

présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE

EN SCIENCES DE LA VIE ET DE LA SANTE

PAR

MARIE-PIERRE LAGET

**CARACTERISATION D'UN MEMBRE DE LA FAMILLE ETS,
LE FACTEUR DE TRANSCRIPTION ERM :
PROPRIETES DE LIAISON À L'ADN
ET
POUVOIR TRANSACTIVATEUR.**

Présentée le 24 Octobre 1995 devant la commission d'examen:

Président : **Pr J-P BOHIN**
Rapporteurs : **Pr J. MARTIAL**
Dr J-P MORNON
Examineurs : **Dr A.D. SHARROCKS**
Pr D. STEHELIN
Dr Y. de LAUNOIT

SCD LILLE 1



D 030 230076 4

REMERCIEMENTS

Ce mémoire est pour moi l'occasion de remercier le Pr Dominique Stéhelin de m'avoir accordé l'hospitalité à un moment où, selon sa propre expression, il avait déjà fait le plein en matière d'étudiants.

J'exprime ma reconnaissance aux membres de mon jury qui ont tous accepté de gaieté de coeur de s'atteler à la tâche ingrate d'évaluer mon manuscrit.

Je remercie Isa, Jean-Paul ainsi que toute leur équipe de m'avoir accueillie comme un des leurs au sein de leur laboratoire et de m'avoir initiée aux plaisirs des "HCA foldingueries".

Le Dr J. R. Vandenneede qui m'a généreusement fait don de quelques uns des anticorps produits par son équipe et même si ces recherches n'ont pas abouti au résultat que nous espérons, je tiens à lui faire part de ma gratitude.

Yvan II m'a acceptée avec toutes mes casseroles, supportée avec toutes mes humeurs (et je sais que, malgré ses affirmations, il a dû me trouver désagréable à certains moments), adouci mes aigreurs avec quelques tonnes de chocolat et pensé à moi 28 heures sur 24 pendant la période pénible de la gestation de ce manuscrit. Je lui suis reconnaissante d'avoir fait preuve de tant de patience.

Je pense que je ne remercierai jamais assez Pierre-Antoine pour son soutien sans faille dans l'adversité et pour la confiance totale qui s'est établie entre nous et qui nous a permis de travailler ensemble.

Anne Chotte, BQ, Christelle, Darfik, Elizabeth, Gomez, Sweetie, ma Vache Marguerite et Zoukette (dans l'ordre alphabétique) ont écouté mes doléances, m'ont régulièrement offert des douceurs pour remonter ma glycémie et mon moral, et communiqué la technique de dissection des crocodiles Haribo; je ne suis pas encore partie (bientôt) mais je sais déjà qu'ils vont tous me manquer. Une mention particulière pour Anne Chotte que j'ai obligée à relire ma thèse en un temps record, mais si, chérie, tu sais que je t'aime.

Je remercie également Isabelle, Agnès, Emmanuel, Nini, Marie-Christine et Caïda qui m'ont tous aidée à leur façon à apprécier ce laboratoire.

Enfin un coup de chapeau admiratif à Daniel, Sophana et Thierry sans les talents desquels beaucoup d'illustrations de ce mémoire n'existeraient pas.

"Les gens qui se laissent abuser par le mensonge sont plus dangereux que ceux qui mentent; et ceux qui se laissent corrompre sont plus misérables que les corrupteurs. Car c'est une loi psychologique, que les imbéciles et les faibles recherchent -et nullement de façon inconsciente- ceux dont ils attendent mensonge et corruption, et ne s'accordent aucun repos avant de les avoir trouvés."

Arthur Schnitzler (La Transparence impossible)

"Et il y avait le même sourire triste et indéchiffrable sur son visage, cet air de la dernière minute que l'on charge d'exprimer tant de choses, mais qui n'est qu'un masque pincé par un sourire vide."

Henry Miller (Tropique du Cancer)

"Le véritable désert, donc le véritable enchantement, est à 10 000 mètres d'altitude, là où la terre apparaît dans sa lumière bleue et géologique, comme une essence inhumaine sans autre repère que la sinuosité des fleuves ou les ondulations minérales, et la fixité du temps est parfaite si vous avez la chance de voler dans le même sens que le soleil."

Jean Baudrillard (Cool Memories)

A Olivier qui avec tant d'amitié a suivi mes premiers pas scientifiques.

A Paulette et feu Ciboulette qui ont fait le meilleur usage
de ce que l'on nomme la "littérature scientifique".

Et à Xavier qui arrive toujours au moment de faire la vaisselle
et qui en dépit de toutes mes dénégations,
croit en mon prénom prédestiné.

PUBLICATIONS

- Predicted common structural features of DNA-binding domains from Ets, Myb and HMG transcription factors.

Laget, M-P, Callebaut, I., de Launoit, Y. Stéhelin, D. and Mornon, J-P
Nucleic Acid Research, 21, pp 5987-5996, 1993

- The PEA3 group of transcription factors in breast cancer.

Monte, D., Baert, J-L, **Laget, M-P**, Defossez, P-A, Coutte, L., Pelczar, H., Angeli, I., Dewitte, F., Stéhelin, D. and de Launoit, Y.
sous presse dans *Annales d'Endocrinologie*

- Characterization of the 3 β -hydroxysteroid dehydrogenase/ Δ^5 - Δ^4 -isomerase activity of *NOCARDIA* cholesterol dehydrogenase : peculiar substrate specificity.

Sanchez, R., Monte, D., **Laget, M-P**, Horinouchi, S., Beppu, T., Simard, J. and de Launoit, Y.
soumis à *Molecular and Cellular Endocrinology*

- Two functionally distinct domains cooperate for transactivation by the Ets family member ERM..

Laget, M-P, Defossez, P-A, Baert, J-L, Albagli, O., Dewitte, F., Stéhelin, D. and de Launoit, Y.
soumis à *Molecular and Cellular Biology*

COMMUNICATIONS

Aromatic residues within the ETS domain as well as a carboxy-terminal domain play a crucial role in the DNA-binding function of ERM.

Laget, M-P, Baert, J-L, Dewitte, F., Defossez, P-A, Monte, D., Stéhelin, D. and de Launoit, Y.
Annual meeting on Oncogenes, Frederick, 1994

Functional dissection of ERM, an Ets family member.

Laget, M-P, Defossez, P-A, Baert, J-L, Dewitte, F., Stéhelin, D. and de Launoit, Y.
Cold Spring Harbor Laboratory, cancer cells : mechanisms of eukaryotic transcription, 1995

SOMMAIRE

RESUME	1
Acte I : INTRODUCTION	3
<u>1. Portrait de famille.</u>	4
<u>2. Le domaine ETS.</u>	6
2.1. Définition.	6
2.2. A chaque protéine Ets sa vérité nucléotidique?	8
2.3. Modulation intramoléculaire de l'activité de fixation à l'ADN ou l'autocensure.	13
2.3.1. c-Ets-1 et v-Ets : les frères ennemis.	13
2.3.2. Et le reste de la famille.	17
2.4. Modulation intermoléculaire de la fixation à l'ADN	18
2.4.1. Les complexes GABP α -GABP β et c-Ets-1-CBF.	18
2.4.2. Coopération de Elf-1 et SPI-1 avec Rb.	20
2.5. Anatomie du domaine ETS (à défaut d'en obtenir les coordonnées atomiques).	22
2.5.1. Les prédictions classiques (une histoire d'ordinateur).	22
2.5.2. La longue histoire d'amour entre MYB et ETS.	24
<u>3. De l'importance du motif structural hélice-coude-hélice.</u>	26
3.1. Définition du motif hélice-coude-hélice (a star is born).	27
3.2. Les homéodomains.	30
3.2.1. Portrait-robot.	30
3.2.2. Ni tout à fait la même, ni tout à fait une autre...	33
3.2.3. Comparaison avec les protéines hélice-coude-hélice.	34
3.2.4. Les fauteurs de troubles.	35

3.3. Les protéines POU.	35
3.4. Les protéines à hélice-coude-hélice type papillon.	35
<u>4. La transactivation par les protéines Ets.</u>	37
4.1. Délimitation des domaines transactivateurs.	37
4.1.1. c-Ets-1 et c-Ets-2 (un effeuillage presque achevé).	37
4.1.2. Erg, Fli et les autres (J'enlève le haut mais pas le bas!).	39
4.2. Intervention des protéines Ets dans des complexes transactivateurs.	41
4.2.1. Un seul être vous manque et tout est dépeuplé.	41
4.2.2. SRF et les protéines Ets (le Dom Juan moléculaire).	43
4.2.3. Elf-1, HMG-I(Y) et NF-kB (des relations triangulaires).	47
4.2.4. c-Ets-1, LEF-1, CBF et ATF/CREB (les parties fines).	49
<u>5. A quoi peuvent donc bien servir les facteurs de transcription Ets?</u>	51
5.1. Les protéines Ets dans le développement.	52
5.1.1. L'incroyable histoire des récepteurs R7 et des protéines Ets.	52
5.1.2. E74.	55
5.2. Ets et l'oncogénèse.	57
5.2.1. Transformation.	57
5.2.2. Ets et la réaction stromale dans le processus de tumorigénèse.	60
5.2.3. Activation par insertion rétrovirale.	61
5.2.4. Activation par translocation.	62
<u>6. Le cas particulier du groupe PEA3.</u>	65
6.1. Les différents membres du groupe PEA3.	65
6.2. Expression des ARN messagers.	68
6.3. Localisation chromosomique.	68
6.4. Un rôle pour les membres du groupe PEA3?	69
6.5. Des gènes cibles pour les protéines du groupe PEA3?	70

Acte II : MATERIEL ET METHODES	73
<u>1. Prédiction de structure.</u>	73
1.1. Etablissement d'un diagramme HCA.	74
1.2. Validation statistique.	75
<u>2. Construction des vecteurs nécessaires.</u>	75
2.1. Construction des vecteurs d'expression contenant tout ou partie de ERM.	75
2.2. Construction nécessaire à l'étude du pouvoir transactivateur de ERM.	77
2.3. Construction des chimères GAL4 ERM.	77
2.4. Construction du domaine ETS de ERM sauvage ou muté.	80
2.5. Construction des mutants rêve.	83
<u>3. Traduction <i>in vitro</i> et analyse des protéines produites.</u>	83
<u>4. Gel-retard.</u>	84
<u>5. Transfection transitoire.</u>	85
5.1. Transfection des cellules.	85
5.2. Mesure de l'activité luciférase.	86
5.3. Mesure de l'activité CAT.	87
Acte III : RESULTATS	88
<u>1. Prédiction de la structure du domaine ETS par la méthode HCA.</u>	88
1.1. Comparaison des motifs structuraux Myb et HMG.	90
1.1.1 Comparaison des structures tridimensionnelles.	90
1.1.2. Alignement des diagrammes HCA.	92
1.1.3. Alignement de séquence.	92
1.2. Alignements Ets-Myb et Ets-HMG.	96
<u>2. Etude des capacités de liaison à l'ADN de ERM.</u>	99
2.1. Etude du domaine ETS.	99
2.1.1. Mutations ponctuelles.	99
2.1.2. Les mutants rêve.	105
2.1.3. Localisation de régions de ERM inhibitrices de la liaison à l'ADN.	106

<u>3. Etude des propriétés de transactivateur de ERM.</u>	111
3.1. Interférence du domaine ETS	
 dans l'étude de la transactivation par ERM.	111
3.2. Pouvoir de transactivation intrinsèque de α et de Ct.	115
3.3. Délimitation des domaines minimaux nécessaires	
 à la transactivation par ERM.	117
3.4. VP16 "sqelche" certains facteurs nécessaires	
 au fonctionnement de α mais non à celui de Ct.	119
3.5. Synergie entre α et Ct.	120
Acte IV : DISCUSSION	122
<u>1. Les relations délicates entre structure et prédiction de structure.</u>	123
1.1. Les charmes dévoilés du domaine ETS de Fli-1.	123
1.2. L'exercice subtil et périlleux de la prédiction de structure.	124
1.3. Le cas particulier de notre prédiction de structure.	126
1.4. Le mensonge des mutations moutonnes.	129
<u>2. Heurs et malheurs du domaine ETS.</u>	132
<u>3. Les qualités cachées de ERM.</u>	133
3.1. Le problème technique posé par l'utilisation de protéines de fusion.	133
3.2. ERM comme facteur de transcription bipartite.	135
Acte V : CONCLUSION	141
REFERENCES	145

RESUME

L'objectif de ce travail a été la caractérisation de la structure et de la fonction du facteur de transcription ERM, plus particulièrement de son domaine de liaison à l'ADN et de ses domaines responsables de la transactivation. ERM constitue avec PEA3 et ER81 le groupe PEA3 appartenant à la famille Ets. Une région d'environ 85 acides aminés, baptisée domaine ETS, constitue le signe de reconnaissance de cette famille. Le domaine ETS porte l'activité de liaison à l'ADN et toutes les protéines Ets se lient à un motif riche en purine GGA. Une relative spécificité de reconnaissance est assignée à chaque membre de la famille par la nature des nucléotides flanquant ce noyau. Le domaine ETS ne présente aucune identité de séquence avec les domaines de liaison à l'ADN déjà caractérisés.

Grâce à la méthode de "Hydrophobic Cluster Analysis", nous avons prédit que le domaine ETS était organisé en cinq hélices comprenant peut-être un motif hélice-coude-hélice. Nous nous sommes ensuite attachés à l'étude de l'effet de mutations ponctuelles dans le domaine ETS de ERM. Puis nous avons délimité deux domaines diminuant l'activité de liaison à l'ADN de ERM; une zone située dans la région carboxy-terminale de ERM (Ct) et l'autre localisée dans la partie centrale (domaine nommé ADID) inhibent de manière indépendante la fixation à l'ADN de ERM. L'activation transcriptionnelle par ERM requiert la présence d'un peptide correspondant aux 72 premiers acides aminés de ERM (α) qui recouvre le domaine riche en résidus acides conservé entre les trois membres actuellement connus du groupe PEA3. La transactivation par α est réprimée par l'expression du domaine transactivateur de VP16, ce qui indique que α et VP16 possèdent probablement des cofacteurs communs présents en quantité limitante. α présente une activité intrinsèque d'augmentation de la transcription au même titre que le second domaine transactivateur de ERM, Ct. Ct possède donc un rôle double puisqu'il exhibe une activité d'inhibition de la liaison à l'ADN. Contrairement à celle de α , l'activité de

Ct n'est pas réprimée par l'expression de VP16. α et Ct fonctionnent en synergie pour augmenter la transcription, ce qui confirme qu'ils n'agissent pas par les mêmes mécanismes. La séquence correspondant à α et à Ct ne semble pas conservée au sein de la famille Ets, indiquant que la spécificité de ERM repose sur des modes de fonctionnement distincts et sur des interactions avec des cofacteurs spécifiques.

INTRODUCTION

De la somme d'informations accumulées sur la régulation de la transcription chez les eucaryotes émergent deux idées essentielles : la régulation de la transcription est complexe et elle met en jeu l'action conjointe de nombreux facteurs qui modulent de concert la transcription. Ces facteurs sont des protéines qui se fixent de manière spécifique à l'ADN sur un ou plusieurs éléments de réponse (appelés également séquences cibles), en amont ou en aval du site d'initiation de la transcription, près de lui ou au contraire à une distance éloignée. Les facteurs de transcription peuvent interagir directement ou non avec la polymérase et exercer un effet activateur et/ou répresseur. Certains s'avèrent incapables de se fixer de manière spécifique sur l'ADN mais pourront réguler la transcription en agissant sur un autre facteur lié à sa séquence cible, en modulant par exemple la stabilité du complexe ADN-protéine formé. Dans ces conditions, il convient de considérer le terme de facteur de transcription comme polysémique. Liaison à l'ADN, modulation de la transcription, multimérisation, fixation d'un ligand ou d'un facteur modulant l'activité de cette protéine sont autant d'exemples de fonctions portées par le facteur de transcription.

Pour ma part, j'adopterai ce vocable pour désigner une protéine portant au minimum les polypeptides requis pour les activités de liaison à l'ADN et de modulation de la transcription. Ces deux fonctions sont traditionnellement attribuées à des domaines distincts et indépendants. La nomenclature des facteurs de transcription s'est établie sur la base de leur identité de séquence dans le domaine de liaison à l'ADN, étant donnée la grande conservation de celui-ci. Chaque famille de facteurs de transcription utilise la même structure de reconnaissance de l'ADN : le doigt de zinc, le domaine basique hélicoïdal des protéines à "leucine zipper" ou à motif hélice-boucle-hélice et le motif hélice-coude-hélice sont les motifs les plus répandus et

l'essentiel de leur structure spatiale et de leur mode d'interaction avec l'ADN est connu. En revanche, les domaines de modulation de la transcription sont beaucoup moins bien caractérisés : la richesse en certains acides aminés (proline, glutamine, résidus acides ou hydroxylés...), souvent adoptée comme signe distinctif, ne constitue pas un indicateur absolu de la présence d'un domaine responsable de la transactivation. En outre, l'élucidation du mode de fonctionnement de ces domaines tient toujours en haleine la communauté scientifique.

1. Portrait de famille.

A l'origine, le proto-oncogène *ets* a été distingué comme l'homologue aviaire de *v-ets*, partie intégrante de la fusion *gag-myb-ets* du rétrovirus E26 (Leprince *et al.*, 1983). *c-ets-1* qui, chez le poulet, en vertu d'un épissage alternatif, encode deux protéines, p54 (Duterque *et al.*, 1988) et p68 (Leprince *et al.*, 1988), constitue le membre fondateur d'une famille qui s'agrandit régulièrement. Avec la découverte d'autres gènes présentant des identités de séquence avec *c-ets-1*, il est rapidement apparu que le domaine ETS par sa grande conservation constituait les armoiries de la famille ETS (Karim *et al.*, 1990).

En utilisant la technique d'amplification en chaîne par polymérase (PCR) à l'aide d'oligonucléotides dégénérés recouvrant le domaine ETS, Degnan *et al.* (1993) ont montré que des membres de la famille *ets* étaient présents chez des invertébrés inférieurs tels que les éponges, les cnidaires, les nématodes, etc... confirmant que les gènes de la famille *ets* sont apparus avant la dichotomie diploblastes-triploblastes. En revanche, cette méthode n'a pas permis de détecter de gènes *ets* chez les plantes, les champignons ni les protozoaires, ce qui ne signifie pas obligatoirement qu'il n'existent pas : il est en effet toujours plus aisé de démontrer l'existence d'un gène que de prouver qu'il est absent.

L'analyse structurelle et fonctionnelle de l'ensemble des membres de la famille *ets* a permis dans un premier temps une classification grossière : par exemple SPI-1 qui ne

présente que 37% d'identité de séquence en acides aminés dans le domaine ETS avec *c-Ets-1* a été catalogué comme le membre le plus éloigné (Lautenberger *et al.*, 1992). L'analyse du taux de divergence montre que la famille *ets* constitue un groupe très ancien de gènes qui peuvent être répartis en deux sous-familles : ETS et SPI, la sous-famille ETS comprenant alors neuf groupes : ETS, ERG, ELG, ER71, PEA3, ELK, ELF, DETS4 et YAN/POK (Fig. 1) (Laudet *et al.*, 1993). Le groupe PEA3 comprend ERM (gène isolé dans notre groupe), ER81 et PEA3. Il est de plus probable que la diversité observée à ce jour au sein de la famille *ets* existait déjà avant la divergence arthropodes/vertébrés, soit il y a plus de 500 millions d'années. En effet, les gènes *D-ets-2*, *D-ets-3*, *elg* et *E74* sont les équivalents respectifs chez la drosophile des gènes de vertébrés *c-ets-1/c-ets-2*, *erg/fli-1*, *GABP α* et *elf-1*.

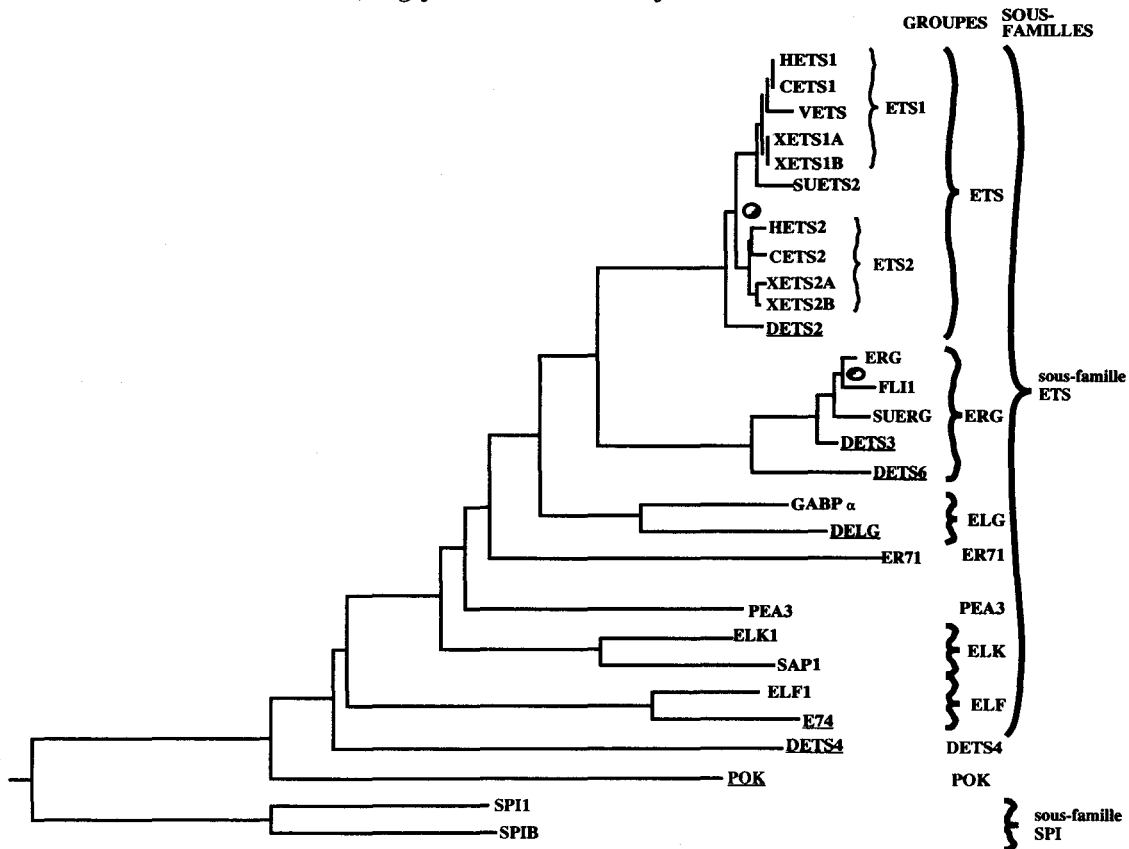


Figure 1 : Arbre phylogénétique des gènes de la famille *ets* établi en fonction de leur identité de séquence dans le domaine ETS selon Laudet *et al.* (1993). Les gènes de la drosophile sont soulignés. Les cercles ombrés représentent les duplications de gènes ancestraux. Certains membres de la famille dont la séquence a été publiée ultérieurement ne sont pas représentés. H, C, V, X, SU et D signifient respectivement Homme, Poulet, Virus, Xénope, Oursin et Drosophile.

Parmi les curiosités, on peut noter la disposition en tandem des gènes *c-ets-1/fli-1* et *c-ets-2/erg*. En effet, chez l'homme *c-ets-1* et *fli-1* sont tous deux localisés sur le chromosome 11 alors que *c-ets-2* et *erg* se situent sur le chromosome 21. La duplication du tandem aurait donc eu lieu avant l'apparition des amphibiens puisque le xénope possède à la fois *c-ets-1* et *c-ets-2*.

Le second point intéressant est celui de la position du domaine ETS : pour Elk-1, SAP-1 et ERP/NET (nommé aussi SAP-2), il est en position amino-terminale, pour Elf-1 et YAN/POK, il est central et pour le reste de la famille *ets* il se situe en carboxy-terminal. Un phénomène d' "exon shuffling" (qui correspond à un mouvement d'exons dans le gène) serait responsable de cette disposition.

2. Le domaine ETS.

2.1. Définition.



Figure 2 : Alignement des domaines ETS de 27 membres de la famille *ets* selon Laudet *et al.* (1993).

La ligne CONS représente les résidus strictement conservés. Les résidus conservés dans 75% des cas ont été notés par une étoile. Les tirets et les points d'interrogation représentent respectivement les discontinuités introduites pour optimiser l'alignement et les acides aminés non déterminés. Les flèches désignent les trois tryptophanes régulièrement espacés.

La compilation et l'alignement des séquences des différents gènes *ets* a mis en lumière l'existence d'un domaine de 85/90 acides aminés appelé domaine ETS qui est largement conservé au cours de l'évolution (Fig. 2). On a commencé à soupçonner la nature de la fonction du domaine ETS quand sa partie carboxy-terminale s'est avérée essentielle à la fixation non spécifique sur de l'ADN de thymus de veau (Boulukos *et al.*, 1989). L'attribution de la fonction de facteur de transcription aux produits des gènes *c-ets-1* et *c-ets-2* a exhorté à une étude plus poussée du domaine ETS (Wasylyk *et al.*, 1990)

Par la technique *in vitro* de gel-retard, un peptide c-Ets-1 de 116 résidus comprenant le domaine ETS est donné pour responsable de l'activité de fixation à l'ADN (Wang *et al.*, 1992). Par des délétions successives, la limite en amino-terminal du domaine de liaison à l'ADN a été fixée pour les protéines c-Ets-1, c-Ets-2 et SPI-1 (Wasylyk *et al.*, 1992; Wasylyk et Wasylyk, 1992). La longueur du domaine de liaison à l'ADN ainsi délimité est respectivement de 112, 116 et 106 résidus. Un fragment de 95 acides aminés contenant le domaine ETS de GABP α est responsable de la liaison spécifique à l'ADN (Thompson *et al.*, 1991). L'étude de Elk-1 a montré que si un fragment de 92 résidus recouvrant le domaine ETS est nécessaire et suffisant pour la liaison spécifique à l'ADN, la délétion des cinq derniers acides aminés du domaine ETS entraînait une diminution conséquente de la fixation (Janknecht et Nordheim, 1992). Finalement, un peptide c-Ets-1 restreint aux 85 acides aminés du domaine ETS a été montré comme nécessaire et suffisant pour la fixation à l'ADN et la délétion de ses dix résidus amino-terminaux abolit totalement l'activité de liaison (Nye *et al.*, 1992; Jonsen *et al.*, 1994). Tels qu'ils ont été définis, ces domaines de liaison à l'ADN recouvrent tous la totalité du domaine ETS. La seule exception et le record à battre en matière de plus petit peptide Ets permettant la fixation à l'ADN est ERM pour lequel 82 résidus suffisent à la liaison à l'ADN (Monté *et al.*, 1994). Signalons que par une technique plus sensible qui est celle de l'hybridation sur filtre,

il a été prouvé qu'un peptide de 37 résidus du domaine ETS était doué d'une capacité de liaison à l'ADN, capacité très faible cependant (Mavrothalassitis *et al.*, 1994).

Il faut néanmoins préciser que l'ensemble des données précitées a été obtenu en interaction avec un oligonucléotide qui est reconnu par la protéine mais qui ne correspond pas à la séquence optimisée (point détaillé plus loin). De plus, les conditions de manipulation sont variables d'une publication à l'autre, ce qui ne permet pas de comparer en toute rigueur les résultats obtenus puisque les facteurs de transcription Ets sont connus pour être affectés dans leur modalité de liaison à l'ADN par la salinité, par la présence de sérumalbumine bovine dans le tampon de réaction ainsi que par le potentiel d'oxydo-réduction (Wasylyk et Wasylyk, 1993).

2.2. A chaque protéine Ets sa vérité nucléotidique?

Plusieurs éléments ont de prime abord laissé penser que chaque protéine Ets afficherait une spécificité de reconnaissance de l'ADN qui lui est particulière. L'analyse des résultats d'interférence de méthylation ainsi que des mutations réalisées dans la séquence cible du domaine ETS ont mis en évidence que les nucléotides encadrant le noyau 5'-GGAA/T-3' reconnu par les protéines Ets étaient nécessaires pour la fixation à l'ADN et la transactivation. Par ailleurs, il a été montré que la présence d'une thymidine aux deux positions bornant le noyau 5'-GGAA-3' était incompatible avec la fixation à l'ADN de c-Ets-1 et de c-Ets-2, et que ces deux protéines se fixaient avec une affinité variable aux diverses séquences cibles examinées (Ascione *et al.*, 1992). Les protéines c-Ets-1, Elk-1 et Erg affichent également des spécificités de fixation légèrement différentes (Rao et Reddy, 1992). Enfin, au sein d'un groupe de quatre éléments de réponse naturels, c-Ets-1 et Elf-1 reconnaissent de manière préférentielle des cibles différentes (Wang *et al.*, 1992). Une telle observation illustre parfaitement le fait que ces deux protéines ne sont pas redondantes : c-Ets-1 et Elf-1, tous deux exprimés fortement dans les cellules T pourront néanmoins aider à la transcription de gènes placés sous le contrôle de

promoteurs différents. Enfin, SPI-1 et le produit p68 de l'épissage alternatif de c-Ets-1 de poulet affichent des préférences variables au sein d'un ensemble de huit cibles dérivées du motif PEA3 de l'élément de réponse aux oncogènes contenu dans l'amplificateur du virus du polyome (Wasylyk *et al.*, 1992). Ce faisceau de résultats laissait donc imaginer que le domaine de fixation, variable d'un membre de la famille Ets à l'autre, posséderait une spécificité de reconnaissance sans ambiguïté et qu'inversement les nucléotides entourant le quadruplet GGAA/T médieraient la préférence de fixation des protéines Ets.

Afin de mettre à l'épreuve ce modèle séduisant, la méthode de détection *in vitro* de la cible préférentielle a été employée. Cette technique de TDA (pour Target Detection Assay) consiste à incuber la protéine en présence d'une série d'oligonucléotides dégénérés marqués radioactivement, à les isoler par immunoprécipitation ou par gel-retard, à amplifier par PCR et à séquencer les produits obtenus. La comparaison des résultats mène à la conclusion décevante que tous les membres de la famille Ets testés reconnaissent des cibles similaires (Fig. 3).

	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
c-Ets-1 (homme)					G/A	C/G	C	G	G	A	A	G	T	C/T	
c-Ets-1 (souris)					G/A	C	C/A	G	G	A	A	G/A	T/C		
c-Ets-1/2 (poulet)					A	C	C/A	G	G	A	A	G/A	T	T	
E74 (drosophile)		A	A	C/T	C	C/A	G	G	A	A	G/A	T			
GABP α (souris)					G/A	C/G	C/A	G	G	A	A	G/A	T/C		
ER81 (souris)					G/A	C/G	C/A	G	G	A	A	G/A	T/C		
ER71 (souris)					G/C	C/G	C/A	G	G	A	A	G/A	T/C	C	
Erg-2 (homme)						C/G	C/A	G	G	A	A	G/A	T		
Fli-1 (souris)		T/G	N	G/A	A	C	C	G	G	A	A	G/A	T/C	G/A	G/A
Elk-1 (homme)						C/A	C/A	G	G	A	A	A/G			
SPI-1 (souris)	A	A	A	A	A/T	G/C	G/C/A	G	G	A	A	G/C	T	G/A	G/C
SPI-B (homme)	A	A	A	A	A/T	G/C	C/A	G	G	A	A	G/C	T	A/T	N

Figure 3 : Séquence de liaison à l'ADN préférentielle déterminée par TDA de quelques protéines Ets. N représente n'importe quel nucléotide.

C'est, assez logiquement, le cas pour c-Ets-1 et c-Ets-2 qui présentent 96% d'identité de séquence dans le domaine ETS (Fisher *et al.*, 1990; Woods *et al.*, 1992) mais aussi pour les protéines c-Ets-1 et Erg-2 humaines (70% d'identité de séquence). Cette tendance a été confirmée par les données obtenues sur ER81, ER71 et GABP α (Brown et McKnight, 1992). Un bémol est cependant apporté par la comparaison des éléments préférentiels reconnus par c-Ets-1 et E74A (50% d'identité de séquence) : E74A se fixe spécifiquement sur 5'-GGAA-3' et non sur 5'-GGAA/T-3'. De plus le nucléotide en position -3 est différent (Urness *et al.*, 1990). Un second argument allant dans le sens d'une spécificité de liaison à l'ADN est apparu dans le cas de SPI-1 et SPI-B : à l'instar de E74 et de Fli-1 (Mao *et al.*, 1994), ces membres de la famille Ets reconnaissent spécifiquement le motif 5'-GGAA-3' (Ray-Gallet *et al.*, 1995). Au cours de ces mêmes expériences de TDA, 7% et 25% des sites sélectionnés respectivement par SPI-1 et SPI-B contenaient le noyau 5'-AGAA-3' alors que les résultats de gel-retard stipulent que les oligonucléotides contenant le noyau 5'-AGAA-3' lient moins bien les protéines SPI que lorsque le noyau 5'-GGAA-3' est utilisé. On peut cependant imaginer que l'effet de cette mutation G en A soit contrebalancée par un nucléotide extérieur au noyau.

Dans le cas de c-Ets-1, les données obtenues *in vitro* correspondent à peu près au consensus établi après compilation des séquences des éléments promoteurs reconnus par les protéines Ets. Seule varie la fréquence de certains nucléotides : le consensus obtenu par TDA favorise la présence d'une thymidine en position +3 alors que parmi 11 éléments promoteurs reconnus par c-Ets-1, on ne la retrouve qu'au niveau de l'amplificateur du virus du polyome. Dans le cas de SPI-1 et SPI-B, la séquence optimale déterminée par TDA comporte des modifications par rapport au consensus établi après compilation des séquences reconnues *in vivo* (Fig. 4). Il est vrai qu'on ne prend en compte que les éléments promoteurs recensés...















amplificateur SV40	C T G A A A G A		C T T G G T T A
E1AV LTR	T A A A A A C A		G T T A A C G C
promoteur EBV LMP/TP2	T A G A A A G G		G T A G A A A G
amplificateur 3' Igκ	C C C T T T G A		C T G A A A A C
amplificateur Igλ 2-4	A A A T A A A A		G T G A A A C C
promoteur Ig Vκ 19	T C A A A T A A		G T A A A T T A
promoteur récepteur M-CSF	C T A A A A G G		G A A G A G G A
promoteur Fcγ RI	A G A A A A G A		G G A A A T T G
promoteur Fcγ RIIIA	C A A A A A G A		G G A A A T G G
amplificateur IL-4	T G A A A A C A		C T G A A A T G
promoteur chaîne J Ig	A A G A A A G C		G C A G C A T
promoteur récepteur nécrophage	A A A A A A A G		G T G A A A A G
promoteur récepteur CD11b	A G A A A A G G		G T A G G A G G
consensus séquence cible	A G A A A A G G		G T A G G A G G
	tc G		c G C
	A		

Figure 4 : Compilation de tous les sites de fixation *in vivo* identifiés pour SPI-1 selon Ray-Gallet *et al.* (1995). Les nucléotides inscrits en minuscules correspondent aux cas les moins fréquents.

On doit émettre quelques critiques vis-à-vis de la technique de TDA. La première (et la plus facile) est que les conditions *in vitro* ne sont pas représentatives de la réalité "cellulaire" au moins en ce qui concerne le rapport des quantités de protéine et d'oligonucléotides. L'illustration la plus frappante a été apportée par les travaux de Woods *et al.* (1992) dans lesquels l'un des oligonucléotides utilisés contient le site BamHI 5'-GGATCC-3' et est reconnu par c-Ets-1. Le résultat de ces expériences est donc à prendre avec grande précaution. La seconde est que rien ne prouve que la séquence définie comme préférentielle par la technique de TDA soit accessible dans le type cellulaire dans lequel la protéine s'exprime. De plus, on ne connaît pas la relation entre l'affinité pour l'ADN d'un facteur Ets et son efficacité en transactivation de sorte qu'il est peut-être compréhensible d'observer une telle disparité entre les éléments définis comme préférentiels par la technique de TDA et

les éléments sur lesquels la protéine se fixe et transactive de manière efficace. Enfin, il n'est pas exclu que la spécificité de fixation des protéines Ets relève également d'autres mécanismes : épissage alternatif ou régulation post-traductionnelle, sont autant de solutions à cette situation non désespérée.

L'interaction avec un cofacteur peut aussi modifier les capacités de liaison à l'ADN comme il a été élégamment montré dans le cas de la protéine X (pX) du virus de l'hépatite B (HBV). Ni pX, ni CREB, ni ATF-2 ne se lient de manière isolée à l'élément de réponse situé sur l'amplificateur de HBV. En revanche, les complexes pX-CREB et pX-ATF-2 ont une activité de liaison sur cette même séquence et l'analyse des données de protection contre la DNase I montre que la fixation sur l'élément de réponse se réalise par l'intermédiaire de CREB ou de ATF-2 et non par celui de pX. Le gain de l'activité de liaison à l'ADN par CREB et ATF-2 est donc attribuable au cofacteur pX. Cette association transitoire se révèle être à bénéfice réciproque puisqu'elle permet d'élargir le spectre d'action des deux facteurs de transcription CREB et ATF-2 en même temps qu'elle autorise pX à participer à la régulation transcriptionnelle (Maguire *et al.*, 1991).

En outre la spécificité de liaison, quel que soit le moyen par lequel elle est acquise, n'est pas le seul critère à prendre en compte, comme cela a été suggéré de manière convaincante par Hahn et Wasylyk (1994). Par la technique de TDA, ils ont montré que, parmi les sites à haute affinité, v-Ets fixait un nombre de séquences différentes plus important que c-Ets-1, en d'autres mots, v-Ets est moins sélectif que c-Ets-1. De plus, v-Ets se fixe toujours mieux à l'ADN que c-Ets-1, que l'on considère les sites de faible ou de haute affinité, mais dans le cas des sites à faible affinité, la différence entre l'intensité de la fixation de v-Ets et de c-Ets-1 est plus importante. Ces deux constatations : moindre sélectivité de v-Ets pour les sites à haute affinité et meilleure fixation de v-Ets sur des sites à faible affinité constituent deux explications probables au pouvoir oncogénique de v-Ets : il se fixe sur des sites non reconnus par c-Ets-1 et il est capable de fixer un plus large spectre de séquences; en conséquence,

l'activité transcriptionnelle de v-Ets peut être considérée comme un dérèglement de l'activité transcriptionnelle normale de c-Ets-1.

2.3. Modulation intramoléculaire de l'activité de fixation à l'ADN ou l'auto-censure.

2.3.1. c-Ets-1 et v-Ets : les frères ennemis.

Les séquences de c-Ets-1 et v-Ets ne divergent que par quatre points :

- 1) v-Ets est partie intégrante de la protéine de fusion Gag-Myb-Ets (p135),
- 2) les 16 résidus carboxy-terminaux de v-Ets sont remplacés par 13 acides aminés totalement différents chez c-Ets-1,
- 3) l'alanine 285 de c-Ets-1 (p68) est remplacée par une valine dans la p135,
- 4) l'isoleucine 445 (dans le domaine ETS de la p68) est remplacée par une valine.

La différence essentielle portant sur la nature de la queue carboxy-terminale, des expériences variées ont tenté d'en déterminer l'importance dans la liaison à l'ADN. Dans le cas de c-Ets-1 comme dans celui de v-Ets, elle n'est pas nécessaire à la fixation. Sa délétion dans c-Ets-1 (p54) augmente considérablement la liaison à l'ADN (Hagman et Grosschedl, 1992). D'autres arguments suggèrent un rôle inhibiteur dans la liaison à l'ADN de la queue carboxy-terminale de c-Ets-1 : la délétion de la queue carboxy-terminale de c-Ets-2 conduit à une augmentation de la fixation à l'ADN comme c'est le cas pour c-Ets-1 (Hagman et Grosschedl, 1992). De plus si l'on substitue la queue carboxy-terminale de c-Ets-2 par celle de c-Ets-1, on observe une inhibition de la liaison à l'ADN d'importance comparable. Ceci suggère que l'effet inhibiteur des 13 résidus carboxy-terminaux de c-Ets-1 peut s'exercer sur d'autres domaines ETS proches en séquence (c-Ets-1 et c-Ets-2 sont identiques à 96% dans le domaine ETS). En revanche, cette substitution est sans conséquence dans le cas de protéines Ets plus divergentes : l'ajout de la queue carboxy-terminale de c-Ets-1 à SPI-1 ne modifie pas son activité de liaison à l'ADN (Hagman et Grosschedl, 1992). Il est intéressant de s'attarder sur les effets de l'inhibition de la liaison à

l'ADN de c-Ets-1 par sa queue carboxy-terminale puisque cette modulation permettrait d'expliquer la différence d'activité entre c-Ets-1 et son équivalent viral v-Ets. Des expériences de protéolyse ménagée ainsi que l'utilisation d'un panel de constructions de délétion ont montré que le domaine ETS de c-Ets-1 (p68), isolé de son contexte protéique, avait une capacité de fixation accrue par rapport à la protéine de pleine longueur et que les régions de la protéine en dehors du domaine ETS inhibaient la liaison à l'ADN. Le mécanisme de répression de la fixation passe par une diminution de l'accessibilité du domaine ETS et nécessite l'intervention conjointe de la région centrale de 70 résidus ainsi que de la queue carboxy-terminale (Lim *et al.*, 1992). Même si cela n'a pas été montré dans le cas de p54, on peut supposer que les deux produits de l'épissage alternatif chez le poulet de *c-ets*, p54 et p68, sont gouvernés par le même mécanisme d'inhibition puisqu'ils ne diffèrent que par leur partie amino-terminale. Selon Lim *et al.* (1992), la plus forte fixation du domaine ETS par rapport à la protéine de pleine longueur n'est pas due à une augmentation de l'affinité mais à une augmentation du nombre de molécules actives. Un modèle a été proposé selon lequel la conformation inactive "fermée" due à l'interaction des deux domaines inhibiteurs de la protéine c-Ets-1 de pleine longueur serait modifiée (d'une manière non élucidée pour l'instant) en une conformation ouverte permettant la liaison à l'ADN (Fig. 5). En revanche, v-Ets qui diffère de c-Ets-1 entre autres par le remplacement de la queue carboxy-terminale de 13 résidus par un peptide de 16 acides aminés n'est pas soumis à une inhibition intramoléculaire similaire; il est donc constitutivement capable de lier l'ADN, ce qui pourrait constituer une explication à son pouvoir oncogène.

Dans le contexte de la p54 (Hagman et Grosschedl, 1992) ou de la p68 (Hahn et Wasylyk, 1994), la substitution des 13 résidus carboxy-terminaux par la queue carboxy-terminale de v-Ets conduit à une augmentation de la fixation. Certains résultats obtenus dans notre laboratoire vont cependant à l'encontre de ces données. En effet, lorsqu'on effectue une substitution comparable sur la p35, forme tronquée correspondant aux 311 derniers acides aminés de la p68, on note une diminution de

la liaison à l'ADN (Leprince *et al.*, 1992). On peut cependant imaginer que la queue carboxy-terminale de v-Ets ait un comportement variable selon qu'on étudie son effet sur une protéine de pleine longueur (p54 ou p68) ou sur un peptide tronqué (p35).

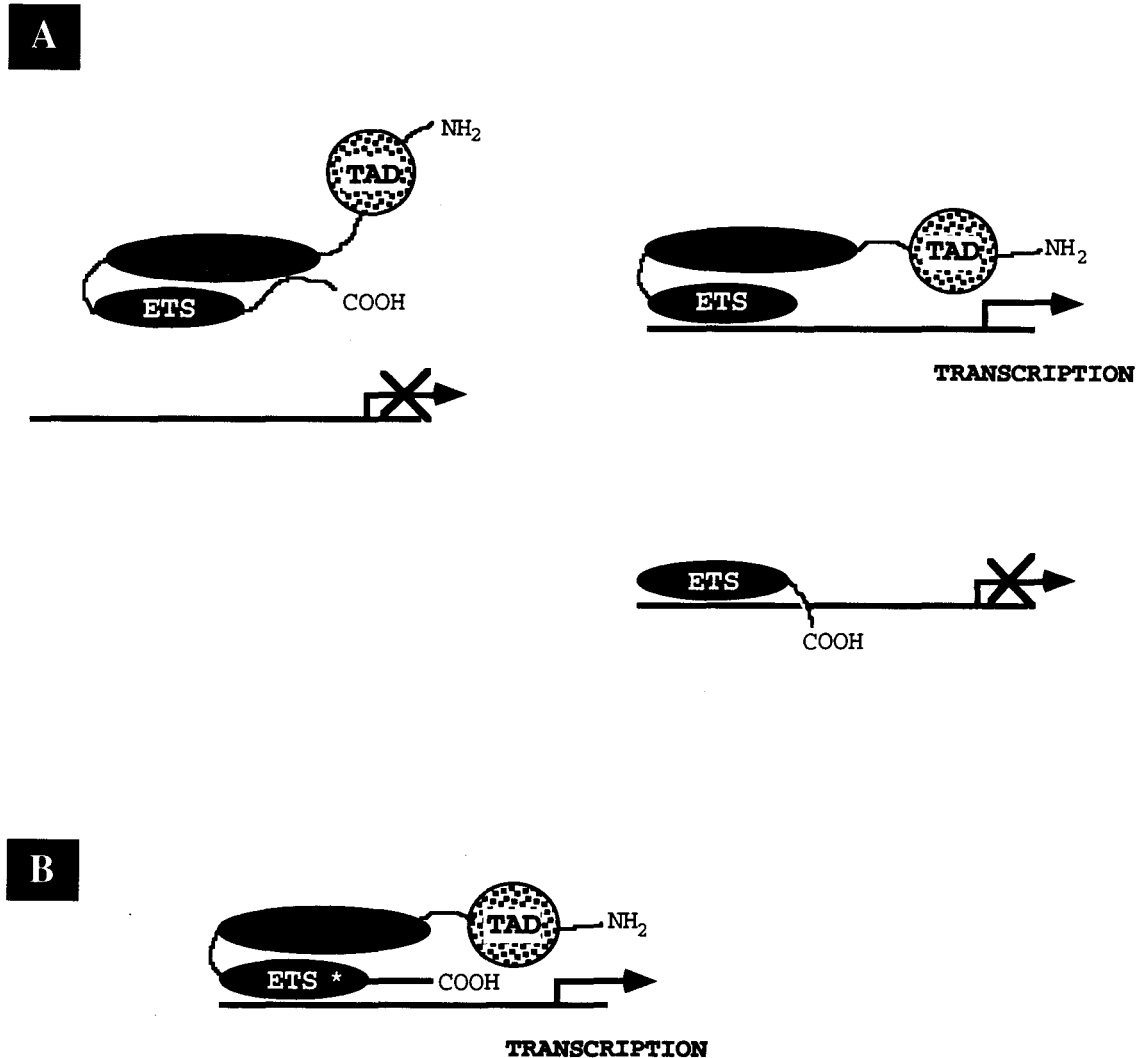


Figure 5 : Conformation et liaison à l'ADN de c-Ets-1 de pleine longueur (A, à gauche), ou tronqué (A, à droite) et de v-Ets (B) selon Lim *et al.* (1992).

ETS et TAD représentent respectivement le domaine ETS et le domaine transactivateur de c-Ets-1 (vue schématique). Le domaine de 70 résidus (noté 70) et la queue carboxy-terminale de c-Ets-1 (que ne possède pas v-Ets) inhibent sa liaison à l'ADN. Les étoiles représentent les mutations dans v-Ets.

Il est peut-être futile de se pencher sur l'explication possible de telles contradictions : sans doute est-il plus judicieux de considérer le comportement des 16 résidus de la queue carboxy-terminale dans le contexte de v-Ets, c'est-à-dire en prenant en compte les mutations ponctuelles et la fusion avec les séquences Gag et Myb. En effet,

chaque mutation ponctuelle (en position 285 ou 445), introduite indépendamment dans la p35 provoque une diminution de la liaison à l'ADN (Leprince *et al.*, 1992; Soudant *et al.*, 1994) mais la double mutation a un effet moins drastique (Leprince *et al.*, 1992), ce qui laisse penser qu'il existe un phénomène de compensation des mutations entre elles. De même, on peut supposer que l'effet négatif de la queue carboxy-terminale sur la fixation à l'ADN pourrait être compensé par les mutations ponctuelles et la fusion avec Gag et Myb. En effet, lorsqu'on introduit de manière indépendante, chacune des mutations ponctuelles présentes dans v-Ets au sein d'une construction Gag-Myb-p68, on n'observe pas d'augmentation du pouvoir transformant. De même, une construction Gag-Myb-p68 où les 13 résidus carboxy-terminaux sont remplacés par la queue carboxy-terminale de v-Ets est moins transformante que Gag-Myb-p68. En revanche, Gag-Myb-v-Ets est plus transformante que Gag-Myb-p68 (D. Leprince, communication personnelle), ce qui signifie que les effets des mutations ponctuelles et de la queue carboxy-terminale de v-Ets se contrebalancent respectivement. Il n'est donc pas certain que l'acquisition des 16 acides aminés de la queue de v-Ets seule soit un gain : l'activation oncogénique s'est réalisée par inversion de la queue carboxy-terminale de c-Ets-1 (Leprince *et al.*, 1993; Lautenberger et Papas, 1993) et la queue carboxy-terminale de v-Ets est codée par le brin non codant de l'extrémité carboxy-terminale de c-Ets-1. L'acquisition par v-Ets de ces 16 résidus peut donc être considérée plus comme un passage obligé dans la naissance du virus. Les mutations ponctuelles en position 285 et 445, et peut-être la fusion avec Gag et Myb, apparues ultérieurement, auraient corrigé l'effet négatif de la queue carboxy-terminale de v-Ets et accordé au virus son pouvoir transformant.

Il existe dans c-Ets-1 d'autres éléments de régulation de l'activité de liaison à l'ADN. Outre la région centrale de 70 acides aminés (Lim *et al.*, 1992; Nye *et al.*, 1992; Jonsen *et al.*, 1994) (résidus 207-280 de p68), l'exon VII (résidus 288 à 374), situé en amont du domaine ETS, s'est révélé responsable de l'inhibition de la fixation à l'ADN (Wasylyk *et al.*, 1992) (Fig. 6).

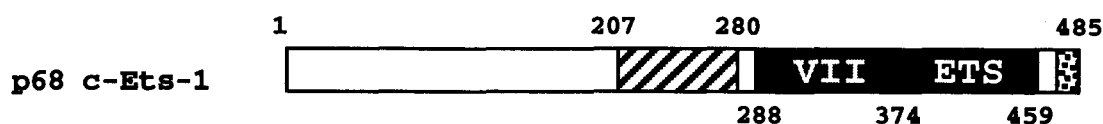


Figure 6 : Représentation schématique de tous les domaines d'inhibition de la liaison à l'ADN chez c-Ets-1. Le domaine ETS (en noir) s'étend des acides aminés 374 à 459. L'exon VII (en gris) comprend la région des résidus 288 à 374. Les zones hachurée et en gris clair représentent respectivement le domaine défini par Lim *et al.* (1992) et les 13 résidus inhibiteurs de la queue carboxy-terminale.

Certains indices laissent suggérer que la taille de ce domaine inhibiteur pourrait être réduite puisque certaines délétions aux extrémités de l'exon VII ou en son centre n'affectent pas l'inhibition de fixation à l'ADN de manière conséquente. Que l'épissage alternatif soit un artifice permettant la modulation de la liaison à l'ADN est une hypothèse qui a été vérifiée sur la protéine humaine c-Ets-1. HuEts-1 est sujette à différents épissages alternatifs l'amputant de tout ou partie de la région codée par l'exon VII (Koizumi *et al.*, 1990; Jorcyk *et al.*, 1991). La construction HuEts-1 Δ VII qui est dépourvue de l'exon VII, se fixe beaucoup mieux que HuEts-1 à l'ADN ce qui confirme que l'exon VII contient un domaine d'inhibition de la fixation à l'ADN (Hahn et Wasylyk, 1994). Les auteurs ont de plus démontré que l'effet de l'exon VII était supérieur à celui de la queue carboxy-terminale que l'on considère la protéine humaine (HuEts-1) ou celle de poulet (p68). C'est indirectement la preuve que contrairement au domaine 207-280 défini par Lim *et al.* (1992), l'exon VII n'interagit pas avec la queue carboxy-terminale. Si tel était le cas, l'augmentation de la fixation à l'ADN serait la même, que l'on délète l'exon VII ou la queue carboxy-terminale.

2.3.2. Et le reste de la famille.

c-Ets-2 est soumis à un mécanisme d'inhibition similaire impliquant une région en amont du domaine ETS qui présente 55% d'identité de séquence avec le domaine équivalent de c-Ets-1 et la queue carboxy-terminale (Wasylyk *et al.*,

1992; Hagman et Grosschedl, 1992). SPI-1, testé dans les mêmes conditions, ne présente pas de domaine d'inhibition de la fixation à l'ADN.

Les cas particuliers de Elk-1, SAP-1a et ERP/NET seront abordés plus loin.

2.4. Modulation intermoléculaire de la fixation à l'ADN.

Toutes les protéines Ets ne possèdent sans doute pas de domaine d'inhibition de la fixation ce qui permet d'imaginer que si leur activité de facteur de transcription est régulée au niveau de leur fixation à l'ADN, elle fait intervenir un élément extérieur.

2.4.1. Les complexes GABP α -GABP β et c-Ets-1-CBF.

L'étude de GABP a montré que la présence d'au moins deux polypeptides (GABP α et GABP β) était requise pour sa fixation spécifique optimale à l'ADN (LaMarco *et al.*, 1991; Thompson *et al.*, 1991). GABP α qui fait partie de la famille Ets en vertu de son identité de séquence dans le domaine ETS est capable de se fixer seul et avec une affinité modérée à l'ADN. GABP β 1, au contraire de GABP β 2 forme des homodimères stables en solution. Deux molécules GABP α recrutent un homodimère GABP β 1-GABP β 1 qui favorise la formation du complexe GABP α -ADN par l'établissement de contacts protéine-protéine (Fig. 7). Dans le complexe GABP α -GABP β 1-ADN, GABP β 1 ne lie pas l'ADN mais en est assez proche pour modifier le patron de méthylation du complexe GABP α -ADN. Les résidus de GABP α impliqués dans l'interaction avec GABP β 1 recouvrent le domaine ETS et les 37 résidus carboxy-terminaux consécutifs. Les quatre motifs "ankyrin-repeat" de 33 résidus de GABP β 1 sont nécessaires à son interaction avec GABP α . Il est à noter que l'interaction GABP β 1 avec GABP α est hautement spécifique puisque GABP β 1 ne se lie pas à d'autres protéines Ets comme ER81 et ER71 (Brown et McKnight, 1992). Cette discrimination de la part de GABP β au sein de la famille Ets

constitue donc un niveau supplémentaire de régulation de l'activité de liaison à l'ADN.

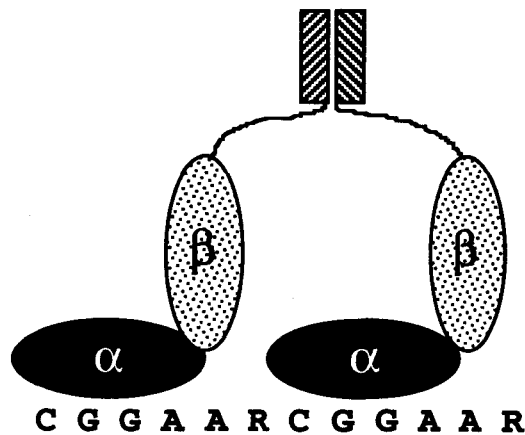


Figure 7 : Modèle d'interaction du complexe GABP avec la séquence dérivée de l'amplificateur du gène ICP4 du virus HSV-1 selon Thompson *et al.* (1991). R représente G ou A. Les rectangles correspondent aux séquences nécessaires à l'homodimérisation de GABPβ1.

Un second exemple de modulation de la fixation à l'ADN par interaction intermoléculaire est celui impliquant l'amplificateur du gène du récepteur β des cellules T. TCRβ présente deux éléments contenant chacun un site ETS sur lequel il a été prouvé que c-Ets-1 et c-Ets-2 pouvaient se fixer, et un site de liaison pour la protéine CBF. Si la fixation sur l'un de ces éléments de CBF et de c-Ets-1 (ou c-Ets-2) peut se réaliser de manière indépendante, il est intéressant de noter que la liaison d'un partenaire protéique stabilise le complexe ADN-protéine formé avec l'autre partenaire de sorte que CBF et c-Ets-1 coopèrent pour la liaison à l'ADN (Wotton *et al.*, 1994). Puisqu'on peut détecter une fixation coopérative de c-Ets-1 et CBF lorsque les sites reconnus sont séparés par 33 nucléotides, l'espacement ne constitue pas une contrainte forte mais il apparaît comme fonction combinatoire de l'affinité pour l'ADN de chaque protéine isolée : lorsque c-Ets-1 et CBF reconnaissent des cibles pour lesquelles elles ont une plus forte affinité, la distance maximale nécessaire à une fixation coopérative est plus importante que lorsque les sites sont faiblement reconnus par c-Ets-1 et CBF.

hétérodimère incapable d'activer la transcription. Puis la phosphorylation de Rb, suite à l'activation des cellules T, entraîne la dissociation de ce complexe, lève l'inhibition exercée par Rb sur le facteur de transcription Elf-1 permettant en conséquence l'activation du gène GM-CSF par Elf-1 (Fig. 8). Le fait que le site de fixation Rb soit adjacent au domaine de Elf-1 supposé responsable de la transactivation suggère, sans qu'on en ait la preuve, que l'inhibition de Elf-1 se joue plus au niveau de l'activité transcriptionnelle (au moins par encombrement stérique) qu'à celui de la fixation à l'ADN.

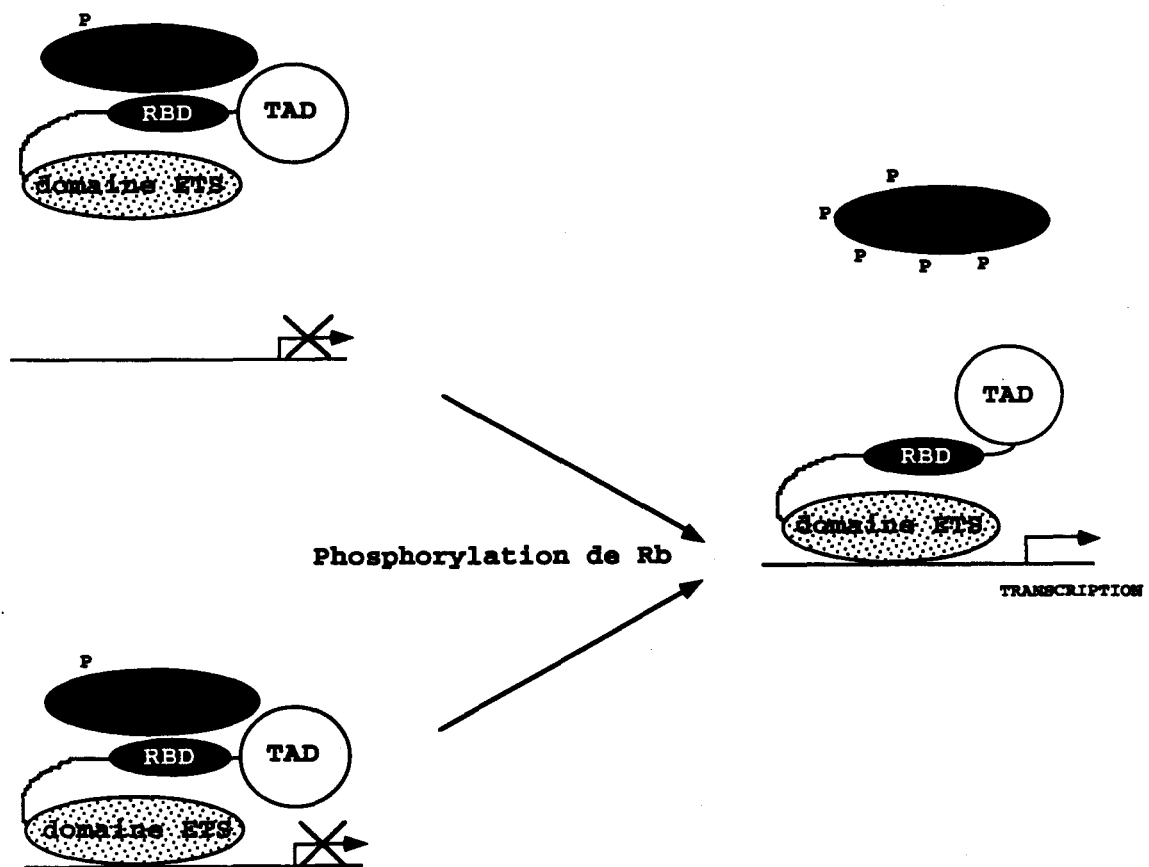


Figure 8 : Un modèle de coopération entre Rb et Elf-1 selon Wang *et al.* (1993). RBD et TAD représentent respectivement le domaine de Elf-1 qui permet la fixation de Rb et le domaine transactivateur. L'association avec la forme sous-phosphorylée de Rb empêche soit la fixation à l'ADN de Elf-1 (en haut à gauche) soit la transactivation (en bas à gauche). La phosphorylation de Rb permet la dissociation du complexe.

Notons qu'un autre membre de la famille dont l'expression est restreinte aux cellules lymphoïdes, SPI-1, se lie aussi directement à Rb. Les premiers 75 résidus du domaine de transactivation de SPI-1 médient la fixation directe sur Rb mais, contrairement à Elf-1, SPI-1 ne contient pas le peptide consensus de la liaison à Rb (Hagemeier *et al.*, 1993). On ignore également si SPI-1 est soumis au même mécanisme de régulation que Elf-1.

Les exemples de modulation intramoléculaire ou intermoléculaire des protéines Ets indiquent que, assez contradictoirement, on a accumulé moult données sur la régulation de l'activité de liaison à l'ADN en ignorant beaucoup de la structure du domaine ETS et de son mode de fixation à l'ADN.

2.5. Anatomie du domaine ETS (à défaut d'en obtenir les coordonnées atomiques).

2.5.1. Les prédictions classiques (une histoire d'ordinateur).

Dès qu'il a été reconnu comme une région nécessaire à la fixation à l'ADN, on a cherché à identifier le domaine ETS à l'une des classes de domaines de liaison à l'ADN déjà constituées. Le domaine ETS ne présente aucune identité de séquence avec les domaines en doigt de zinc ni avec les domaines basiques hélicoïdaux ou les motifs de reconnaissance organisés en feuillet β . Cependant, la prédiction de structure utilisant l'algorithme de Chou-Fasman a indiqué l'existence d'une zone organisée en un motif hélice-boucle-hélice (Reddy et Rao, 1990) (Fig. 9). L'extension progressive de la famille Ets a permis d'utiliser d'autres méthodes d'alignement de séquence et de prédiction de structure comme l'algorithme de Garnier-Robson-Kyte qui met en évidence une zone de conformation hélicoïdale précédant un domaine basique (Wang *et al.*, 1992) (Fig. 9).

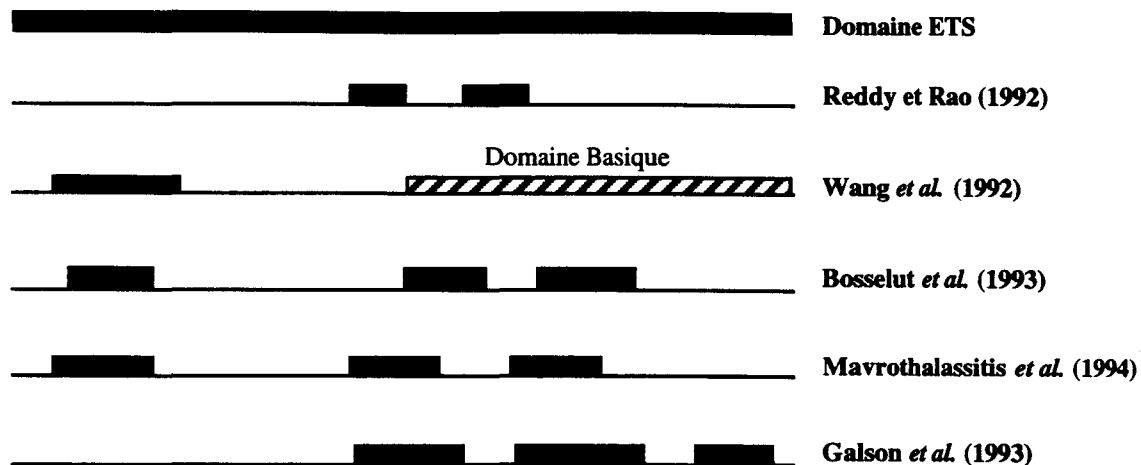


Figure 9 : Détail des différentes prédictions de structure effectuées sur le domaine ETS. Il est important de signaler que la prédiction de Reddy et Rao (1990) est partielle. De plus, la prédiction de Galson *et al.* (1993) n'a été réalisée que sur SPI-1.

Une potentielle ressemblance avec des domaines organisés en hélice-boucle-hélice ou en homéodomaine a été indirectement confirmée par des techniques de protection chimique et d'interférence d'alkylation (Nye *et al.*, 1992). En effet, il a été montré que c-Ets-1 reconnaissait, à l'intérieur d'un oligonucléotide correspondant à la séquence cible de c-Ets-1, le triplet GGA, situé dans le grand sillon, qui constitue le "noyau dur" du consensus établi à l'époque. Une des caractéristiques majeures du mode d'interaction des protéines à hélice-coude-hélice et des homéodomains est justement l'insertion d'une hélice (dont on soupçonnait le domaine ETS d'être pourvu) dans le grand sillon de l'ADN. Cependant la comparaison avec les résultats d'interférence d'éthylation et de protection contre l'attaque des radicaux hydroxyles n'a pas été convaincante, laissant une fois de plus la place à de nouvelles spéculations.

Plus récemment, Mavrothalassitis *et al.* (1994) par une méthode de prédiction de structure non précisée proposaient l'existence d'une hélice en amino-terminal du

domaine ETS, plus courte mais superposable à celle prédite par Wang *et al.* (1992) et d'un motif hélice-coude-hélice situé au début de la zone basique. Pour appuyer cette hypothèse, il a été montré qu'un peptide de 37 résidus contenant ce motif fixe la séquence reconnue par le domaine ETS entier mais avec une affinité cent fois moindre (Mavrothalassitis *et al.*, 1994). La méthode de prédiction de structure "mpred" utilisée par Bosselut *et al.* (1993) qui pour plus de précision a l'avantage de prendre en compte le domaine ETS de tous les membres de la famille suggérait une architecture en trois hélices. Selon les affirmations de l'auteur, ce résultat est comparable à celui obtenu en combinant les deux méthodes de prédiction Chou-Fasman et GOR. De plus, l'analyse du cas particulier de SPI-1 a montré qu'il existait une chance raisonnable que les deux dernières hélices correspondent au motif classique hélice-coude-hélice (Galson *et al.*, 1993).

En dehors des prédictions classiques, un alignement de séquences entre le domaine ETS et le domaine de fixation à l'ADN d'un facteur de transcription de la famille HMG a été proposé, basé essentiellement sur la conservation de trois résidus hydrophobes (Waterman *et al.*, 1991).

2.5.2. La longue histoire d'amour entre MYB et ETS.

Une importance a sporadiquement été accordée à la présence de trois tryptophanes très conservés au sein du domaine ETS (Fig. 2). Ces trois tryptophanes sont espacés d'environ 18 acides aminés, caractéristique déjà mise en lumière dans le domaine de fixation à l'ADN d'un autre facteur de transcription, Myb. En effet, cette région de Myb, longue de 120 acides aminés environ, est constituée de la répétition imparfaite de trois unités de 51/52 résidus. Chaque unité contient trois tryptophanes espacés régulièrement de 18/19 résidus (Frampton *et al.*, 1991) (Fig. 10).

Dans les deux cas, l'intervention d'une protéine extérieure (GABP β ou CBF) permet d'augmenter la capacité de liaison à l'ADN du membre de la famille Ets. Chaque système présente en outre l'avantage d'un niveau de régulation supplémentaire : dans le cas de GABP α -GABP β , c'est la spécificité du partenaire extérieur; dans celui de c-Ets-1-CBF, c'est la présence au sein du complexe de deux domaines de fixation à l'ADN, nécessitant la présence conjointe des deux séquences nucléotidiques reconnues. Les deux modèles exposés ci-dessus présentent néanmoins une différence : la capacité à lier l'ADN du partenaire de la protéine Ets. En effet, GABP β ne se lie pas à l'ADN au contraire de CBF. On ignore aussi pour l'instant si CBF se lie spécifiquement à c-Ets-1.

Les conséquences au niveau de l'activité de transcription de l'interaction c-Ets-1-CBF seront détaillées dans un chapitre ultérieur.

2.4.2. Coopération de Elf-1 et SPI-1 avec Rb.

Le produit du gène du rétinoblastome (Rb) est une protéine nucléaire dont la phosphorylation intervient dans la régulation de la progression du cycle cellulaire. L'identification de plusieurs protéines auxquelles Rb se lie a permis d'établir un consensus pour le motif peptidique nécessaire à la fixation à Rb. La partie amino-terminale de Elf-1 contient un site putatif de liaison à Rb et des expériences de "pull-down" (technique décrite plus loin) ont permis de démontrer une interaction directe entre Rb et la région des acides aminés 1 à 108 de Elf-1 *in vitro* et *in vivo* (Wang *et al.*, 1993). Elf-1 se fixe préférentiellement à la forme sous-phosphorylée de Rb. On a par ailleurs constaté que l'expression d'une forme constitutivement sous-phosphorylée de Rb inhibe la transactivation exercée par Elf-1 au niveau du promoteur GM-CSF. Un modèle simple propose que, dans les cellules T quiescentes où la présence de Elf-1 est requise pour l'induction de la transcription de certains gènes, Elf-1 et la forme sous-phosphorylée de Rb forment un

hétérodimère incapable d'activer la transcription. Puis la phosphorylation de Rb, suite à l'activation des cellules T, entraîne la dissociation de ce complexe, lève l'inhibition exercée par Rb sur le facteur de transcription Elf-1 permettant en conséquence l'activation du gène GM-CSF par Elf-1 (Fig. 8). Le fait que le site de fixation Rb soit adjacent au domaine de Elf-1 supposé responsable de la transactivation suggère, sans qu'on en ait la preuve, que l'inhibition de Elf-1 se joue plus au niveau de l'activité transcriptionnelle (au moins par encombrement stérique) qu'à celui de la fixation à l'ADN.

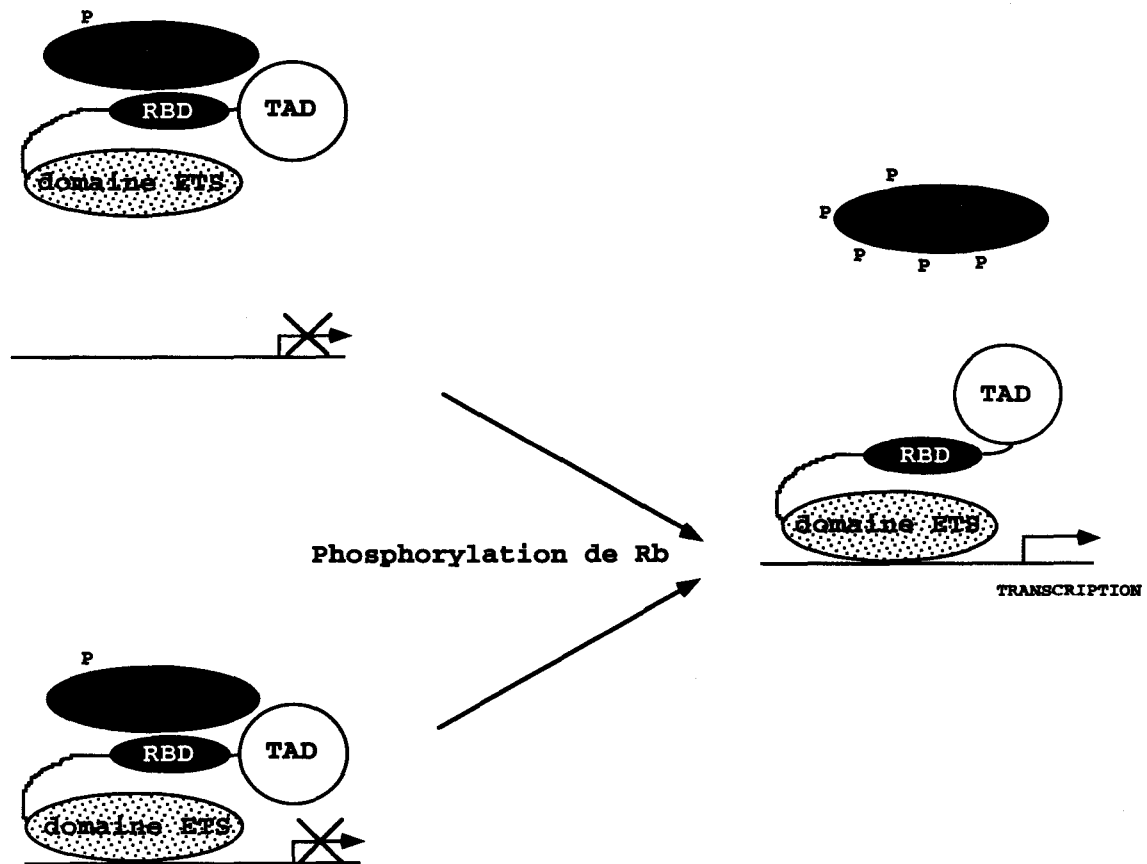


Figure 8 : Un modèle de coopération entre Rb et Elf-1 selon Wang *et al.* (1993).

RBD et TAD représentent respectivement le domaine de Elf-1 qui permet la fixation de Rb et le domaine transactivateur. L'association avec la forme sous-phosphorylée de Rb empêche soit la fixation à l'ADN de Elf-1 (en haut à gauche) soit la transactivation (en bas à gauche). La phosphorylation de Rb permet la dissociation du complexe.

Notons qu'un autre membre de la famille dont l'expression est restreinte aux cellules lymphoïdes, SPI-1, se lie aussi directement à Rb. Les premiers 75 résidus du domaine de transactivation de SPI-1 médient la fixation directe sur Rb mais, contrairement à Elf-1, SPI-1 ne contient pas le peptide consensus de la liaison à Rb (Hagemeier *et al.*, 1993). On ignore également si SPI-1 est soumis au même mécanisme de régulation que Elf-1.

Les exemples de modulation intramoléculaire ou intermoléculaire des protéines Ets indiquent que, assez contradictoirement, on a accumulé moult données sur la régulation de l'activité de liaison à l'ADN en ignorant beaucoup de la structure du domaine ETS et de son mode de fixation à l'ADN.

2.5. Anatomie du domaine ETS (à défaut d'en obtenir les coordonnées atomiques).

2.5.1. Les prédictions classiques (une histoire d'ordinateur).

Dès qu'il a été reconnu comme une région nécessaire à la fixation à l'ADN, on a cherché à identifier le domaine ETS à l'une des classes de domaines de liaison à l'ADN déjà constituées. Le domaine ETS ne présente aucune identité de séquence avec les domaines en doigt de zinc ni avec les domaines basiques hélicoïdaux ou les motifs de reconnaissance organisés en feuillet β . Cependant, la prédiction de structure utilisant l'algorithme de Chou-Fasman a indiqué l'existence d'une zone organisée en un motif hélice-boucle-hélice (Reddy et Rao, 1990) (Fig. 9). L'extension progressive de la famille Ets a permis d'utiliser d'autres méthodes d'alignement de séquence et de prédiction de structure comme l'algorithme de Garnier-Robson-Kyte qui met en évidence une zone de conformation hélicoïdale précédant un domaine basique (Wang *et al.*, 1992) (Fig. 9).

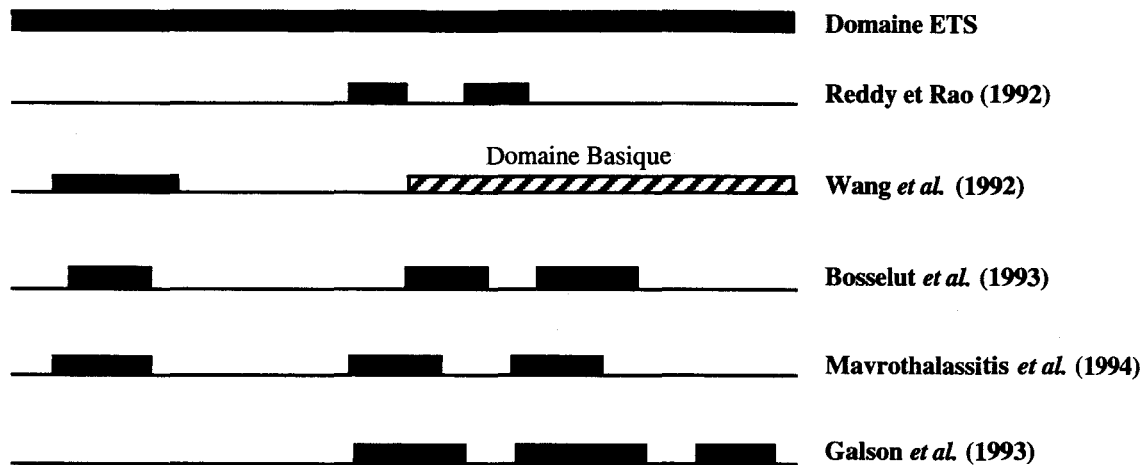


Figure 9 : Détail des différentes prédictions de structure effectuées sur le domaine ETS. Il est important de signaler que la prédiction de Reddy et Rao (1990) est partielle. De plus, la prédiction de Galson *et al.* (1993) n'a été réalisée que sur SPI-1.

Une potentielle ressemblance avec des domaines organisés en hélice-boucle-hélice ou en homéodomaine a été indirectement confirmée par des techniques de protection chimique et d'interférence d'alkylation (Nye *et al.*, 1992). En effet, il a été montré que c-Ets-1 reconnaissait, à l'intérieur d'un oligonucléotide correspondant à la séquence cible de c-Ets-1, le triplet GGA, situé dans le grand sillon, qui constitue le "noyau dur" du consensus établi à l'époque. Une des caractéristiques majeures du mode d'interaction des protéines à hélice-coude-hélice et des homéodomains est justement l'insertion d'une hélice (dont on soupçonnait le domaine ETS d'être pourvu) dans le grand sillon de l'ADN. Cependant la comparaison avec les résultats d'interférence d'éthylation et de protection contre l'attaque des radicaux hydroxyles n'a pas été convaincante, laissant une fois de plus la place à de nouvelles spéculations.

Plus récemment, Mavrothalassitis *et al.* (1994) par une méthode de prédiction de structure non précisée proposaient l'existence d'une hélice en amino-terminal du

domaine ETS, plus courte mais superposable à celle prédite par Wang *et al.* (1992) et d'un motif hélice-coude-hélice situé au début de la zone basique. Pour appuyer cette hypothèse, il a été montré qu'un peptide de 37 résidus contenant ce motif fixe la séquence reconnue par le domaine ETS entier mais avec une affinité cent fois moindre (Mavrothalassitis *et al.*, 1994). La méthode de prédiction de structure "mpred" utilisée par Bosselut *et al.* (1993) qui pour plus de précision a l'avantage de prendre en compte le domaine ETS de tous les membres de la famille suggérait une architecture en trois hélices. Selon les affirmations de l'auteur, ce résultat est comparable à celui obtenu en combinant les deux méthodes de prédiction Chou-Fasman et GOR. De plus, l'analyse du cas particulier de SPI-1 a montré qu'il existait une chance raisonnable que les deux dernières hélices correspondent au motif classique hélice-coude-hélice (Galson *et al.*, 1993).

En dehors des prédictions classiques, un alignement de séquences entre le domaine ETS et le domaine de fixation à l'ADN d'un facteur de transcription de la famille HMG a été proposé, basé essentiellement sur la conservation de trois résidus hydrophobes (Waterman *et al.*, 1991).

2.5.2. La longue histoire d'amour entre MYB et ETS.

Une importance a sporadiquement été accordée à la présence de trois tryptophanes très conservés au sein du domaine ETS (Fig. 2). Ces trois tryptophanes sont espacés d'environ 18 acides aminés, caractéristique déjà mise en lumière dans le domaine de fixation à l'ADN d'un autre facteur de transcription, Myb. En effet, cette région de Myb, longue de 120 acides aminés environ, est constituée de la répétition imparfaite de trois unités de 51/52 résidus. Chaque unité contient trois tryptophanes espacés régulièrement de 18/19 résidus (Frampton *et al.*, 1991) (Fig. 10).

A

R1			
	*	*	*
c-Myb souris	LGKTRWTR EDEKLK LVEQNGTDDWKVIANYLPNRTDVQCQHR		WQKVLNPE
A-Myb humain	WNRVKWTRDETDKLKKLVEQHGTDWTLIASHLQNRSDFCQHR		WQKVLNPE
B-Myb humain	KCKVKWTHEETEQLRALVRFQGDWKF LASHFPNRTDQQCQYR		WLRVLNPD
Myb drosophile	GRGKRWSKSEDLVLLKQLVETHG ENWEIIGPHFKDRLEQQVQQR		WAKVLNPE
BAS1 levure	LGYENGL EDVKTIQQSNHLSK CIA	WDVLATRFKHTVRTSKDVRKRWTGSLDPN	

R2			
	*	*	*
c-Myb souris	LIKGPWTR EEDQRVIKLVQKYGPKRWSVIAKHLK	GRIGKQCRERWHNHLNPE	
A-Myb humain	LIKGPWTK EEDQRVIELVQKYGPLRWSLIAKHLK	GRIGKQCRERWHNHLNPE	
B-Myb humain	LVKGPWTK EEDQKVIELVKKYGTQWTLIAKHLK	GRLGRQCRERWHNHLNPE	
Myb drosophile	LIKGPWTR EEDDMVIKLVRFNGPKKWTLIARYLN	GRIGKQCRERWHNHLNPN	
C1 Zea mays	VKRGAWTS KEDDALAAYVKAHGEGKWREVPQKAGLRRCGKSCRLRWLNLYLRPN		
BAS1 levure	LKKGKWTQE EDEQLLKAYEEH GPH	WLSISMDIP	GRTEDQCAKRYIEVLGPGSK

R3			
	*	*	*
c-Myb souris	VKKT SWTE EEDRIIYQAHKRLGNRWAEIAKLLPGRTDNAIKNHWNSTMRK V		
A-Myb humain	VKKSSWTE EEDRIIYEAHKRLGNRWAEIAKLLPGRTDN SIKNHWNSTMRK V		
B-Myb humain	VKKSCWTE EEDRIICEAHKVLGNRWAEIAKMLPGRTDN AVKNHWNSTIKR K V		
Myb drosophile	IKKTAWTE KEDEI IYQAHLELGNQWAKIAKRLPGRTDN AIKNHWNSTMRK Y		
C1 Zea mays	IRRGNISY DEEDLIIRLHRL LGNRWSLIAGRLPGRTDN ELKNYWNSTLGRRA		

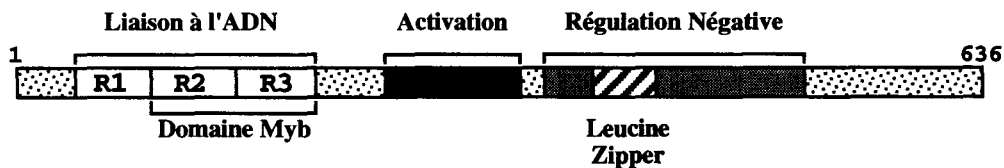
B

Figure 10 : (A) Alignement des séquences répétées de six membres de la famille Myb selon Frampton *et al.* (1991). Les étoiles signalent la présence des trois tryptophanes espacés régulièrement. Tous les membres de la famille ne contiennent pas les trois séquences répétées mais le domaine MYB (R2R3) est suffisant pour une fixation spécifique à l'ADN.

(B) Schéma de l'organisation fonctionnelle de c-Myb murin.

La possibilité d'une similarité de fonction entre les tryptophanes de Myb et ceux du domaine ETS est appuyée par la très bonne conservation au sein des familles Myb et Ets de ces tryptophanes (Anton et Frampton, 1988; Ascione *et al.*, 1992). L'interprétation des données de mutagenèse dirigée a mis en évidence l'importance

des tryptophanes de Myb et du domaine ETS dans la fixation à l'ADN et clarifié leur rôle. De plus, la résolution de la structure tridimensionnelle d'une unité de répétition Myb a montré l'importance des tryptophanes dans la constitution du noyau hydrophobe (Ogata *et al.*, 1992).

L'ensemble de ces résultats a relancé le débat sur la similarité de fonction entre les tryptophanes de Myb et Ets. En usant de l'argument que Myb et Ets coopèrent pour la transformation de cellules de la lignée érythrocytaire de poulet (Metz et Graf, 1991a), la communauté scientifique a longtemps été tentée d'imaginer une identité de structure entre les domaines de fixation à l'ADN de Myb et Ets, jusqu'aux récents travaux de RMN qui ont définitivement mis fin à cette idylle.

3. De l'importance du motif structural hélice-coude-hélice.

Comme nous l'avons détaillé dans un chapitre précédent, de nombreuses interactions entre le domaine ETS et d'autres régions du facteur de transcription ou une protéine extérieure modifient la spécificité de fixation à l'ADN. Néanmoins, on s'autorisera à définir le domaine de fixation à l'ADN d'un facteur de transcription comme le peptide nécessaire et suffisant à une fixation spécifique. Lorsqu'entre deux facteurs de transcription ces domaines sont interchangeables, ils sont considérés comme des unités de repliement protéiques autonomes. Pour clarifier la situation, on emploiera le mot "domaine" pour désigner une structure qui adopte une conformation tri-dimensionnelle indépendamment de reste de la protéine et le mot "motif" pour faire référence à une partie conservée de cette structure. A l'inverse d'un domaine, un motif ne se replie pas de manière indépendante, pas plus qu'il n'est suffisant à la fixation spécifique à l'ADN. Plusieurs types de domaines de fixation à l'ADN ont pu être identifiés sur la base de leur identité de séquence, autorisant ainsi

une taxonomie rendue chaque jour plus complexe par la résolution de structures atypiques ou de domaines présentant une identité structurale avec les représentants classiques de ce motif sans révéler aucune identité de séquence avec eux.

Ne seront détaillés ici que les domaines de fixation à l'ADN qui comprennent un motif hélice-coude-hélice dont la connaissance est indispensable à la compréhension de l'organisation du domaine ETS.

3.1. Définition du motif hélice-coude-hélice (a star is born).

Historiquement, c'est la comparaison de la structure tri-dimensionnelle des domaines responsables de la liaison à l'ADN des protéines Cro et λ , à première vue d'organisation très différente, qui a permis de définir le motif hélice-coude-hélice (Anderson *et al.*, 1981; Ohlendorf *et al.*, 1982; Pabo et Lewis, 1982; Pabo *et al.*, 1982; Ohlendorf *et al.*, 1983; Jordan *et al.*, 1988). En fait, les deuxième et troisième hélices de ces deux structures sont parfaitement superposables tant par l'orientation des hélices l'une par rapport à l'autre que par la longueur du peptide qui les relie. En outre, les seconde et troisième hélices jouent un rôle équivalent dans la liaison à l'ADN des deux protéines. Le motif hélice-coude-hélice était né.

Il est défini comme un peptide de 20 résidus comprenant deux hélices qui forment un angle de 120° environ (Fig. 11). La jonction entre les deux hélices se fait par une boucle de quatre acides aminés, le premier étant systématiquement une glycine. Certains résidus hydrophobes de la boucle et des deux hélices sont conservés (Fig. 11) et semblent jouer un double rôle dans la formation du noyau hydrophobe du domaine de fixation à l'ADN et dans les contacts qu'entretient ce domaine avec le reste de la protéine.

Des modèles d'interaction ADN-protéines ont été proposés puis la résolution par des méthodes directes de complexes ADN-protéine hélice-coude-hélice a permis de mettre en lumière le mode d'interaction de ce motif avec sa séquence nucléotidique cible. L'axe de la première hélice du motif n'est colinéaire ni à l'axe du grand sillon

ni à l'axe de la double hélice d'ADN, elle est donc "loin" de la séquence reconnue par le domaine de fixation. Cependant, par l'intermédiaire de ses résidus amino-terminaux, elle établit des contacts avec l'ADN. La seconde hélice se place dans le grand sillon, mais son axe n'est pas exactement colinéaire à celui du grand sillon puisque sa partie amino-terminale est plus proche de l'ADN que son extrémité carboxy-terminale.

A

R	Q	E	I	G	Q	I	V	G	C	S	R	E	T	V	G	R	I	L	K	CAP
Q	R	E	L	K	N	E	L	G	A	G	I	A	T	I	T	R	G	S	N	Trp
L	Y	D	V	A	E	Y	A	G	V	S	Y	Q	T	V	S	R	V	V	N	Lac
I	K	D	V	A	R	L	A	G	V	S	V	A	T	V	S	R	V	I	N	Gal
T	R	K	L	A	Q	K	L	G	V	E	Q	P	T	L	Y	W	H	F	K	Tet
Q	T	R	A	A	L	M	M	G	I	N	R	G	T	L	R	K	K	L	K	Tn10
Q	A	E	L	A	Q	K	V	G	T	T	Q	Q	S	I	E	Q	L	E	N	Fis
Q	T	E	L	A	T	K	A	G	V	K	Q	Q	S	I	Q	L	I	E	A	434 Rep
Q	T	K	T	A	K	D	L	G	V	Y	Q	S	A	I	N	K	A	I	H	434 Cro
Q	E	S	V	A	D	K	M	G	M	G	Q	S	G	V	G	A	L	F	N	λ Cro
																				λ Rep

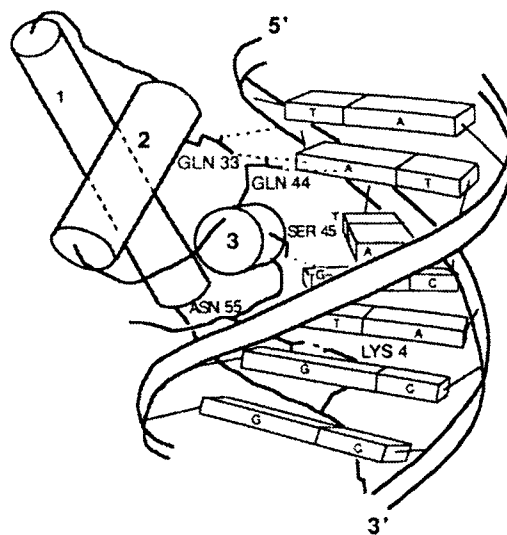
B

Figure 11 : (A) Alignement des séquences des motifs hélice-coude-hélice de protéines régulatrices de phage ou de bactérie selon Pabo et Sauer (1992). Les acides aminés apparaissant en caractères gras dans les cadres sont les résidus conservés dans la plupart des protéines hélice-coude-hélice.

(B) Vue schématique de l'organisation du motif hélice-coude-hélice du répresseur λ et de ses interactions avec l'ADN selon Jordan et Pabo (1988).

3.2. Les homéodomaines.

3.2.1. Portrait-robot.

L'homéodomaine est un peptide de 60 acides aminés nécessaire et suffisant à une fixation spécifique à l'ADN et que l'on retrouve dans une famille de protéines impliquées dans la régulation des gènes eucaryotes. Virtuellement présent dans toutes les espèces de la levure à l'homme, l'homéodomaine a originellement été mis en évidence chez les facteurs de transcription intervenant dans le développement de la drosophile.

La première structure élucidée a été celle de la protéine Antennapedia (Antp) (Fig. 12). Les données de RMN ont montré que l'homéodomaine de Antp était organisé en quatre hélices (Billeter *et al.*, 1989; Qian *et al.*, 1990). Les hélices HI et HII sont antiparallèles l'une à l'autre, les hélices HIII et HIV sont perpendiculaires aux deux premières et forment entre elles un coude d'environ 30° (Fig. 12). L'hélice HI est précédée d'un bras flexible qui établit des contacts avec le petit sillon. Le motif hélice-coude-hélice correspond aux seconde et troisième hélices qui sont reliées par quatre acides aminés. Les contacts entre HI, HII et HIII forment le noyau hydrophobe de l'homéodomaine.

Contrairement à Antp, Engrailed (en) et MAT α 2 qui sont également des protéines à homéodomaine possèdent un domaine de fixation à l'ADN organisé en seulement trois hélices (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991), HIII correspondant au couple d'hélices HIII/HIV présent dans Antp. Seuls le début de HIII (pour en et MAT α 2) et l'hélice HIII (pour Antp) interviennent dans la liaison à l'ADN. La dernière partie des homéodomaines ne semble pas correspondre à un élément de structure bien défini : fin d'une hélice simple longue (en, MAT α 2) ou couple d'hélices (Antp), ce segment correspond à un bras flexible dans le cas de fushi tarazu (ftz) (Qian *et al.*, 1994a). La variabilité que présente ce peptide semble donc indiquer qu'il ne joue qu'un rôle secondaire dans la fixation à l'ADN.

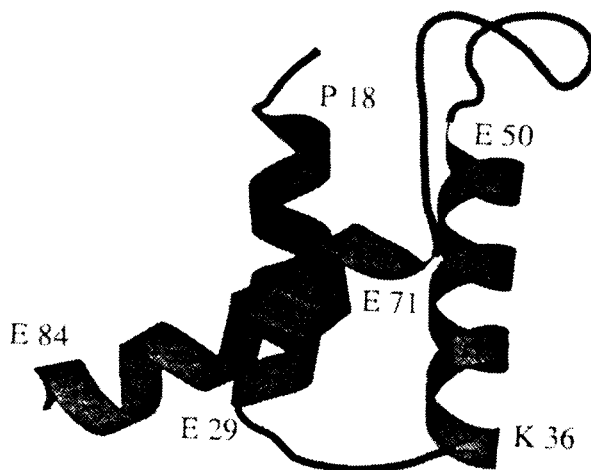
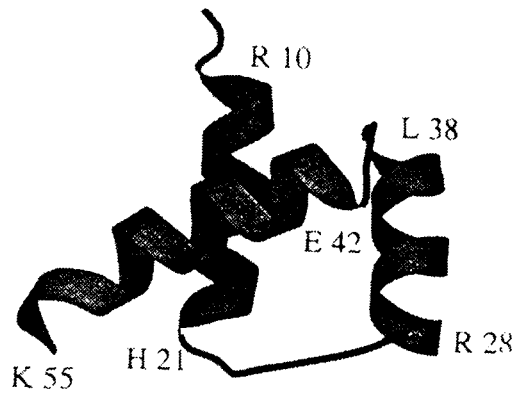
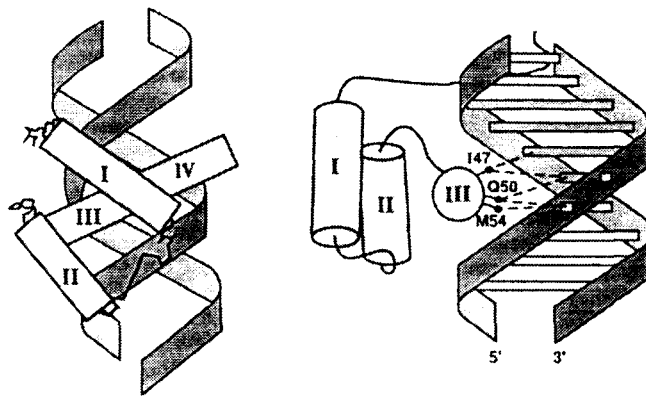


Figure 12 : Organisation spatiale de l'homéodomaine de *Antennapedia* (Antp) (en haut). Vue perpendiculaire à l'axe de l'hélice de reconnaissance à gauche et dans la direction de cette hélice à droite, selon Gehring *et al.* (1994). Comparaison des homéodomaines de bon goût (Antp) au milieu et de mauvais goût (LFB1) en bas, selon Leiting *et al.* (1993). L'insertion de 21 acides aminés dans le motif hélice-coude-hélice ne modifie pas l'architecture générale de l'homéodomaine.

La structure de Antp, en et MAT α 2 en complexe avec l'ADN (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991; Qian *et al.*, 1993) a clairement prouvé que les contacts avec l'ADN se réalisaient majoritairement dans le grand sillon par l'hélice de reconnaissance (HIII ou HIII/HIV) et minoritairement dans le petit sillon par le bras situé en amino-terminal de HI, équivalent du bras amino-terminal du domaine de fixation à l'ADN du répresseur λ . Il est intéressant de constater que l'absence de ce bras amino-terminal ne modifie en rien l'architecture de l'homéodomaine de Antp mais en revanche entraîne une importante diminution de l'affinité de fixation à l'ADN (Qian *et al.*, 1994b).

Le cas particulier du facteur de transcription de foie de rat (LFB1=HNF1) soulève un point intéressant puisque la structure de son homéodomaine a montré que si le nombre et l'orientation des hélices étaient conservés par rapport à en et à MAT α 2, l'homéodomaine de LFB1 présentait une insertion de 21 résidus entre HII et HIII, avec pour conséquence directe un allongement de HII d'un tour d'hélice et le remplacement du tour par une longue boucle qui n'est pas impliquée dans la fixation à l'ADN (Ceska *et al.*, 1993; Leiting *et al.*, 1993) (Fig. 12). La dimérisation de LFB1 pourrait s'effectuer par l'intermédiaire de cette boucle supplémentaire.

L'organisation générale de l'homéodomaine et son mode d'interaction avec l'ADN ne semblant pas modifiés dans le cas de LFB1, la communauté scientifique s'est donc vu contrainte d'élargir la notion de motif hélice-coude-hélice. La seconde implication de cette découverte est que la pression de sélection pour maintenir un motif de reconnaissance de l'ADN organisé en deux hélices est forte. Ceci est confirmé par le fait que l'homéodomaine de MAT α 2 présente une architecture très proche de celle de en et Antp (la seule variation se situe au niveau de la jonction entre HI et HII) tout en ne montrant qu'une faible identité de séquence (27% avec en et 28% avec Antp; MAT α 2 est considéré comme le membre le plus divergent de la famille des protéines à homéodomaine).

3.2.2. Ni tout à fait la même, ni tout à fait une autre...

Tous les protéines à homéodomaine reconnaissent toutes une séquence consensus 5'-ATTA-3' et ont des propriétés de fixation apparemment très similaires. De plus, la forte identité structurale entre les homéodomains pose le problème de la spécificité de fixation. Trois réponses peuvent être apportées à cette angoissante question :

1) Alors que certains contacts avec l'ADN sont identiques pour les différents homéodomains, les bases peuvent aussi établir des interactions avec les chaînes latérales non conservées, ce qui constituera une première manière d'augmenter la spécificité des homéodomains.

2) La combinaison possible avec un second domaine de fixation à l'ADN qui nécessitera donc plusieurs éléments de fixation est une seconde solution. Les protéines POU qui possèdent un second domaine dénommé domaine POU-spécifique en constituent un exemple frappant.

3) Enfin, d'autres mécanismes comme un effet de synergie pour la fixation à l'ADN par interaction avec un cofacteur modifieront les interactions avec les séquences cibles. L'hétérodimérisation des protéines MAT α 2 et MAT α 1 est essentielle au maintien du stade diploïde mais MAT α 2 peut également former avec la protéine MCM1 un hétérotétramère qui contribue au maintien des mini-chromosomes de la levure.

3.2.3. Comparaison avec les protéines hélice-coude-hélice.

Malgré l'existence d'un motif commun, les différences entre les protéines hélice-coude-hélice et les homéodomains sont multiples.

1) Même si les homéodomains comportent parfois une insertion au niveau du coude (cas de LFB1) ou une hélice de reconnaissance coudée (Antp), la topologie générale du domaine de fixation à l'ADN est très homogène comparée à celle des

protéines hélice-coude-hélice où le motif de reconnaissance est "décoré" d'un nombre très variable d'hélices ou de brins β .

2) Le motif hélice-coude-hélice de l'homéodomaine fait un angle de 90° quand il est de 120° environ pour les protéines hélice-coude-hélice.

3) Les hélices de l'homéodomaine sont plus longues que leur analogues chez les protéines hélice-coude-hélice.

4) Les résidus de l'hélice de reconnaissance qui sont impliqués dans l'interaction avec l'ADN n'occupent pas des positions équivalentes. Ils sont amino-terminaux pour les protéines hélice-tour-hélice et centraux pour les homéodomains.

5) Enfin, dans les protéines hélice-coude-hélice, le premier résidu du tour est une glycine. La nature de cet acide aminé très conservé ne semble paradoxalement pas jouer un rôle primordial puisque la mutation de ce résidu en acide glutamique dans le répresseur du phage λ donne un mutant tout à fait fonctionnel (Hochschild *et al.*, 1983). En revanche, les homéodomains présentent rarement une glycine à cette position, les seules exceptions connues étant en et Inverted.

3.2.4. Les fauteurs de troubles.

Pour perturber la classification bien établie entre les protéines à hélice-coude-hélice et les protéines à homéodomaine, Hin qui appartient à la famille des recombinases semble posséder des caractéristiques de ces deux familles structurales et s'impose comme une transition idéale (Feng *et al.*, 1994). A l'instar des homéodomains, Hin présente une organisation antiparallèle de HI et HII ainsi qu'un bras amino-terminal qui établit des contacts avec l'ADN sur le petit sillon. Mais la longueur des hélices et des jonctions entre hélices rappelle nettement l'architecture d'une protéine de type hélice-coude-hélice...

De plus, le domaine paired est organisé en deux sous-domaines chacun remarquablement superposable au domaine de fixation de la recombinase Hin (Xu *et al.*, 1995). Résultat troublant : l'hélice de reconnaissance de ces domaines (HIII et

HVI) entretient des rapports beaucoup plus étroits avec le grand sillon de l'ADN que n'importe quelle hélice de reconnaissance, protéines hélice-coude-hélice et homéodomains confondus.

3.3. Les protéines POU.

Les facteurs de transcription POU exercent une fonction critique dans la mise en route et la régulation de l'organogenèse. Le domaine spécifique des protéines POU (qui sera nommé POU_s) comporte environ 70 acides aminés et il forme une unité architecturalement indépendante de l'homéodomaine des protéines POU en amino-terminal duquel il est placé (Herr et Cleary, 1995). Sa structure indépendamment obtenue par deux équipes a montré qu'il consistait en quatre hélices (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993). Il présente deux particularités par rapport aux homéodomains qui sont un core hydrophobe très volumineux et une quatrième hélice distordue par la présence d'une proline. Le motif hélice-coude-hélice correspond aux deuxième et troisième hélices et diffère du motif classique par un point : la deuxième hélice ainsi que la boucle sont beaucoup plus longues. D'une manière surprenante le domaine POU_s de Oct-1 présente de nombreuses similarités avec les répresseurs λ et 434.

3.4. Les protéines à hélice-coude-hélice "type papillon".

Parmi les déclinaisons exotiques autour du motif hélice-boucle-hélice, on peut compter, outre les exemples précédemment cités, un ensemble de structures regroupées sous le terme de protéines à hélice-boucle-hélice de "type papillon" (traduction libre M.P.L. de "winged helix-turn-helix") (Fig. 13). En effet, avec un peu d'imagination, la forme de ces domaines se calque approximativement sur celle d'un papillon pour lequel l'hélice de reconnaissance ferait office de thorax, le reste du domaine constituant deux grandes ailes. Cette variante zoologique comprend des

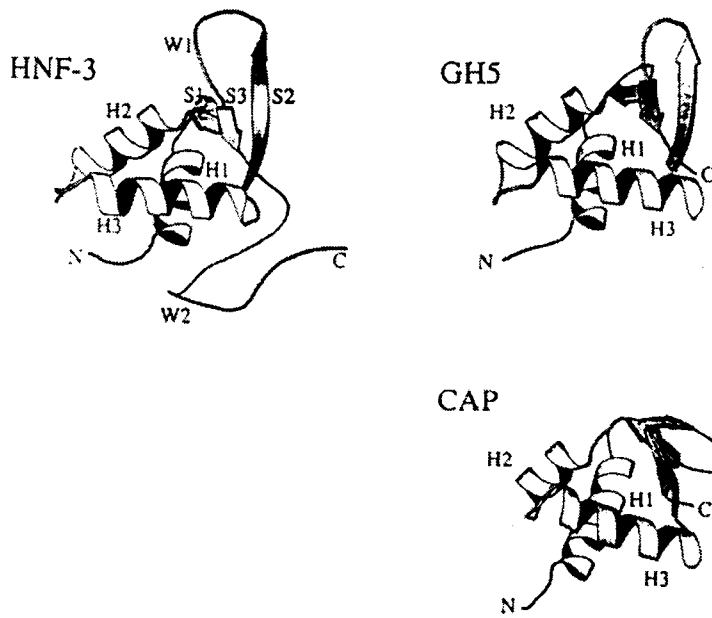


Figure 13 : Organisation tridimensionnelle de trois représentants de la famille des protéines à hélice-coude-hélice "type papillon" selon Brennan (1993).

protéines d'origine diverse. En font partie la transposase Mu, le répresseur de la biotine (Wilson *et al.*, 1992), la protéine CAP de pleine longueur, les facteurs de choc thermique (Harrison *et al.*, 1994; Vuister *et al.*, 1994a; Vuister *et al.*, 1994b), le facteur nucléaire hépatocytaire 3 (HNF3, nommé aussi forkhead) (Clark *et al.*, 1993), le domaine globulaire de l'histone H5 (Ramakrishnan *et al.*, 1993) et **au moins un membre de la famille Ets qu'on évoque ENFIN**. La structure précise du domaine ETS sera abordée dans la partie DISCUSSION.

4. La transactivation par les protéines Ets.

Le travail qui fait l'objet de ce présent manuscrit s'attachant essentiellement à la caractérisation de la protéine ERM, c'est-à-dire à la détermination des régions nécessaires aux activités de liaison à l'ADN et d'activation de la transcription ainsi qu'à l'élucidation du mode de fonctionnement de ce facteur de transcription, nous avons volontairement oblitéré la question des séquences cibles pour pouvoir nous focaliser sur l'architecture des domaines transactivateurs des protéines Ets et nous avons développé de manière extensive le cas des protéines c-Ets-1 et c-Ets-2.

4.1. Délimitation des domaines transactivateurs.

4.1.1. c-Ets-1 et c-Ets-2 (un effeuillage presque achevé).

c-Ets-1 et c-Ets-2 se sont vu coller l'étiquette de facteurs de transcription suite aux expériences établissant qu'ils étaient capables de transactiver sur une séquence dérivée de l'amplificateur du virus du polyome (Wasylyk *et al.*, 1990) et sur un élément de réponse contenu dans la longue répétition terminale du

virus HTLV-1 (Bosselut *et al.*, 1990; Gégonne *et al.*, 1992). En cherchant à délimiter les domaines responsables de cet effet, certains auteurs ont montré que la délétion des 98 premiers acides aminés de c-Ets-1 (p68) diminuait de cinq fois l'effet transactivateur par la protéine native, suggérant donc la présence d'un premier domaine transactivateur, et ont ensuite indiqué qu'il existait un second domaine (acides aminés 250 à 312) nécessaire à un effet transactivateur optimal. En réalité, la fonction d'activation de la transcription par c-Ets-1 (p68) est portée par les 71 premiers résidus amino-terminaux (région RI) ainsi que par la région RIII de 104 résidus qui est localisée juste en amont du domaine ETS (Schneikert *et al.*, 1992) et qui recouvre le domaine 250-312 (Fig. 14). La bien nommée région RII située entre RI et RIII, sans exercer aucune activité de transcription altèrerait la fonction de la région RI et stimulerait celle de RIII. Il est à noter que la région RI est spécifique du produit d'épissage alternatif p68 de c-Ets-1; c-Ets-1 (p54) pour laquelle les 71 premiers acides aminés (correspondant aux exons α et β de c-Ets-1) sont remplacés par un peptide totalement différent de 27 acides aminés possède donc la seule région RIII comme domaine transactivateur (Schneikert *et al.*, 1992).



Figure 14 : Organisation des domaines transactivateurs de c-Ets-1 (p68) selon Schneikert *et al.* (1992).

En outre, les travaux de Schneikert *et al.* (1992) ont suggéré une organisation fonctionnelle similaire pour c-Ets-1 (p68) et c-Ets-2, ce qui a été confirmé en particulier pour l'extrémité amino-terminale. Son organisation génomique est identique dans *c-ets-1* (p68) et *c-ets-2* (Albagli *et al.*, 1994) et elle présente une région de forte identité de séquence étendue sur un segment hydrophobe-acide de 13

résidus. Cette région recouvre l'exon β de c-Ets-1 (p68) et a été baptisée BEC pour Ets-1-Beta-Ets-2 Conserved region (Fig. 15).

Ets-2 xénope	Y	EEE QAVPTGLDSYPHD	SSSCELPLLTPCS
Ets-2 humain	N	EEQTLOEVPTGLDSISHD	SANCELPLLTPCS
Ets-2 poulet	N	EDQTLOEVPTGFDSTSYE	SNNCELPLLTPCS
Ets-1 (p68) poulet	W	DEARIQEVPTGLEHYSTD	MECADVOLLTPSS
Ets-1 (p68) tortue	W	DESTIQEVPTGLERYNSD	MECADV

Figure 15 : Définition du domaine BEC (encadré en gris) par alignement entre c-Ets-1 (p68) et c-Ets-2 selon Albagli *et al.* (1994).

La délétion du BEC dans les domaines RI de c-Ets-1 (p68) (correspondant aux exons α et β) et de c-Ets-2 compromet sérieusement la transactivation dans le cas de l'utilisation d'une chimère avec un domaine de liaison à l'ADN hétérologue mais diminue plus faiblement la transactivation dans le contexte de la protéine native. Les extrémités amino-terminales de c-Ets-1 (p68) et de c-Ets-2 sont dorénavant considérées comme portant des fonctions équivalentes. Pour corroborer cette identité de comportement, il faut signaler que c-Ets-1 (p68) et c-Ets-2 ont des rôles similaires dans la coopération avec le facteur de transcription AP-1 (Wasylyk *et al.*, 1990).

4.1.2. Erg, Fli et les autres (J'enlève le haut mais pas le bas!).

Un certain nombre de membres de la famille Ets ont dévoilé leurs charmes de facteurs de transcription sans que tous les domaines responsables de cette fonction aient été circonscrits de manière précise.

Les différentes formes résultant de l'épissage alternatif du gène *erg* codent des facteurs de transcription contenant plusieurs domaines transactivateurs dont une région de 46 acides aminés immédiatement en aval du domaine ETS. De plus, la

délétion des 208 premiers acides aminés entraîne une perte de 70% de l'effet transactivateur, suggérant l'existence d'un second domaine transactivateur (Reddy et Rao, 1991; Duterque-Coquillaud *et al.*, 1993; Prasad *et al.*, 1993; Siddique *et al.*, 1993). De manière intéressante cette région contient une zone d'identité de séquence que l'on retrouve dans divers membres de la famille Ets (Fli-1, c-Ets-1, c-Ets-2 et son homologue chez la Drosophile Pnt, YAN/POK et GABP α). Cette séquence commune est recouverte par le domaine responsable de la transactivation chez Erg et chez son proche parent Fli-1 (80% d'identité de séquence) mais, curieusement, elle ne joue pas un rôle équivalent pour c-Ets-1, c-Ets-2 ni Pnt. Fli-1 possède un second domaine transactivateur en aval du domaine ETS mais il semble plus étendu que l'équivalent présent chez Erg (Rao *et al.*, 1993; Zhang *et al.*, 1993). Erg et Fli-1 possèdent donc une organisation très similaire.

Le facteur de transcription SPI-1 contient une région riche en résidus acides (acides aminés 1 à 165), plus précisément la zone des résidus 75 à 93 comprend 31% de glutamine ce qui la plaçait *a priori* comme un domaine transactivateur. Les résultats obtenus par différentes équipes au sujet de la délimitation exacte du domaine transactivateur sont à première vue contradictoires. Selon Hagemeyer *et al.* (1993), les 75 premiers résidus en fusion avec un domaine de liaison à l'ADN hétérologue sont suffisants pour la transactivation et exercent le même effet qu'une chimère comportant les 134 résidus amino-terminaux. Selon Shin *et al.* (1993), la délétion des 70 premiers acides aminés n'a aucun effet sur la transactivation vis-à-vis d'un élément de réponse naturel et tout le potentiel transactivateur est porté par la région des résidus 75 à 93. La région amino-terminale, riche en résidus proline qui conférerait à la protéine une structure flexible apte aux interactions protéine-protéine. Cette seconde hypothèse peut être considérée comme la plus probable puisque les conditions utilisées sont nettement plus proche de la réalité.

Un des membres les plus éloignés de la famille, Elk-1, contient pour l'instant comme SPI-1 un domaine de transactivation unique (Rao et Reddy, 1992; Janknecht *et al.*, 1994). La région absolument indispensable à la transactivation est restreinte aux

résidus 376 à 404 tandis que les acides aminés encadrant cette zone, sans être essentiels, augmentent par leur présence l'effet de transactivation. Cet exemple a l'extrême avantage de montrer que la transactivation ne relève pas d'un effet tout ou rien mais d'un phénomène où l'importance des résidus est hiérarchisée.

Le cas de facteurs de transcription du groupe PEA3 sera évoqué dans la partie RESULTATS.

Toutes les protéines Ets n'ont pas été démontrées comme étant des facteurs de transcription, cependant, il apparaît d'ores et déjà que les différents domaines responsables de la transactivation des membres de la famille Ets ne partagent aucune caractéristique commune contrairement au domaine ETS. Modérons néanmoins cette assertion en soulignant que les critères en cours, basés sur la richesse en certains acides aminés (proline, glutamine, résidus acides ou hydroxylés...) seront sans doute affinés dans les années à venir puisque par exemple les résidus du domaine acide responsable de la transactivation par la protéine VP16 ne sont justement pas les résidus acides mais les résidus hydrophobes (Cress *et al.*, 1991).

4.2. Intervention des protéines Ets dans des complexes transactivateurs.

4.2.1. Un seul être vous manque et tout est dépeuplé.

Les premiers éléments d'une coopération entre une protéine Ets et un autre facteur de transcription sont apparus lorsque Wasylyk *et al.* (1989) ont montré entre autres qu'une protéine Ets et le facteur de transcription AP-1 se liaient tous deux à un élément de réponse de l'amplificateur du polyome (contenant un site Ets et un site AP-1) et que la somme des effets des mutations sur les sites ETS et AP-1 était moins importante que l'effet de la mutation simultanée des deux sites, suggérant donc qu'une protéine Ets et AP-1 agissaient de concert pour l'augmentation de la transactivation. L'existence d'une coopération de c-Ets-1 (ou c-Ets-2) avec AP-1 a été indirectement mise en évidence dans le cas de la transactivation sur l'élément de réponse de l'amplificateur du virus du polyome (Wasylyk *et al.*, 1990). L'intégrité

des sites ETS et AP-1 est requise pour cette synergie mais on ignore si cet effet est le simple écho d'une coopération de fixation ou implique une "collaboration" dans la transcription proprement dite.

Alors que c-Ets-2 ne transactive pas seul sur l'élément de réponse de l'amplificateur de *mim-1*, il coopère avec c-Myb pour l'augmentation de la transcription (Dudek *et al.*, 1992) et cet effet n'est pas dû à une coopération de fixation à l'ADN. La recherche d'hétérodimère c-Ets-2-Myb par co-immunoprécipitation s'est soldée par un échec, ce qui laisse supposer soit que l'association n'est pas directe et que c-Ets-2 et c-Myb s'intègrent au sein d'un complexe de transcription multimérique, soit que l'interaction est trop faible pour être détectée par la méthode utilisée.

L'exemple de c-Ets-1 et Sp1 montre que la coopération entre deux facteurs de transcription pour la transactivation peut être le reflet d'une simple coopération de fixation (Gégonne *et al.*, 1993). La formation du complexe ternaire comprenant c-Ets-1, Sp1 et l'élément de réponse contenu dans la longue répétition terminale du virus HTLV1 dépend strictement de l'intégrité des sites de fixation des deux protéines et des expériences de protection contre la DNase I ont clairement indiqué que c-Ets-1 se liait d'une manière préférentielle sur l'élément de réponse déjà lié à Sp1.

NF-EM5 nécessite la présence de SPI-1 lié au site ETS adjacent pour sa fixation à l'élément de réponse de l'amplificateur 3' de l'immunoglobuline κ (Pongubala *et al.*, 1992). Cependant des expériences de cofractionnement sur héparine-agarose n'ont amené aucun indice permettant de conclure à une interaction directe de ces deux protéines du moins en l'absence d'ADN. L'analyse de l'activité de divers mutants de délétion de SPI-1 a révélé que la présence d'une région de 43 acides aminés était nécessaire (mais pas forcément suffisante) pour l'établissement de contacts indirects entre les deux protéines. Cette région contient trois sérines cibles potentielles de phosphorylation par la caséine kinase II et l'état phosphorylé d'un de ces trois résidus (la sérine 148) est requis pour une interaction effective entre NF-EM5 et SPI-1 (Pongubala *et al.*, 1993). Cette région recouvre la zone d'identité de séquence de

SPI-1 avec les régions PEST (riches en proline (P), acide glutamique (E), sérine (S) et thréonine (T)) qui seraient impliquées dans la dégradation des protéines cellulaires. On peut donc facilement concevoir que l'interaction de NF-EM5 avec SPI-1 joue un double rôle d'aider à la fixation à l'ADN (pour le bien de NF-EM5) et de protéger SPI-1 de la dégradation. Un exemple éloquent d'association transitoire à bénéfice réciproque.

Récemment a été reportée l'existence de Pip (PU.1 interaction partner) qui est très probablement NF-EM5 (Eisenbeis *et al.*, 1995). Comme NF-EM5, Pip ne se lie pas seul à l'élément de réponse λ B de l'amplificateur B de l'immunoglobuline λ mais recrute SPI-1. De plus, la formation du complexe est hautement dépendante de l'intégrité des sites de liaison de SPI-1 et Pip ainsi que de la présence dans SPI-1 d'une sérine en position 148. Bien que cela n'ait pas été démontré explicitement, il semble que Pip possède des propriétés transactivatrices : Pip appartient à la famille de facteurs de transcription IRF et contient des régions riches en glutamine (15% des résidus sur la région 354-419) et en proline (20% des résidus sur la région 151-237). De plus, alors que dans un autre contexte il est doué de la capacité à transactiver (voir plus haut), SPI-1, ainsi que Pip, est incapable de transactiver seul à partir d'un élément λ B concatémérisé. Quand les deux protéines sont coexprimées, on observe néanmoins une augmentation de la transcription. Ceci ne prouve pas d'ailleurs que Pip soit un facteur de transcription. On pourrait le considérer comme une protéine douée de l'activité de liaison à l'ADN, qui recrute SPI-1 pour sa fixation sur l'élément λ B et qui, en contrepartie, aide à la transcription par SPI-1.

4.2.2. SRF et les protéines Ets (le Dom Juan moléculaire).

Je prendrai la liberté de m'attarder sur ce modèle d'interaction protéine-protéine et protéine-ADN dont l'étude a été très fouillée. Le SRE (pour Serum Response Element) médie l'expression des gènes précoces (comme *c-fos*) en réponse au stress, aux facteurs de croissance ou aux cytokines. C'est un complexe

ternaire composé du facteur de réponse au sérum (SRF pour Serum Response Factor) et de TCF (pour Ternary Complex Factor) médie la réponse à partir de l'élément SRE du gène *c-fos*. L'étude de ses caractéristiques de liaison à l'ADN a permis d'identifier TCF comme Elk-1 (Hipskind *et al.*, 1991), SAP-1a (Dalton et Treisman, 1992) ou ERP/NET (SAP-2) (Lopez *et al.*, 1994; Giovane *et al.*, 1994). Elk-1, SAP-1a et ERP/NET présentent des identités de séquence au niveau de trois régions ayant des fonctions distinctes :

- 1) la boîte A qui correspond au domaine ETS (située en amino-terminal de la protéine),
- 2) la boîte B qui médie les contacts directs SRF-TCF,
- 3) la boîte C, domaine de transactivation qui est régulé par phosphorylation (Marais *et al.*, 1993) (Fig. 16).

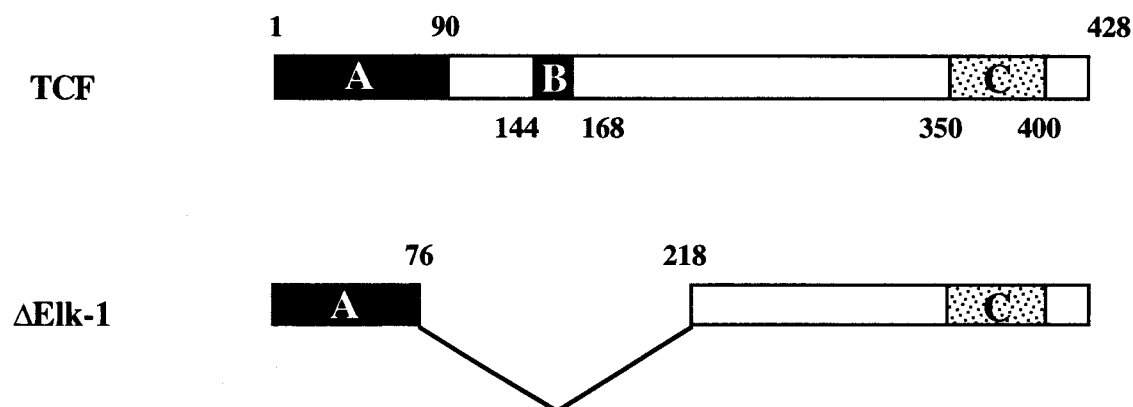


Figure 16 : Schéma récapitulatif de la structure des TCF (en haut) et du produit de l'épissage alternatif de Elk-1(en bas). Les numérotations correspondent à Elk-1.

Elk-1 est capable de se fixer de manière autonome (sans le SRF) sur les oligonucléotides PEA3, MSV LTR et, plus fortement, sur E74BS (Rao et Reddy, 1992) mais pas sur le SRE (Hipskind *et al.*, 1991). Pour la fixation autonome, l'intégrité de la partie carboxy-terminale du domaine ETS est requise. En revanche, les 14 premiers acides aminés du domaine ETS de Elk-1 peuvent être délétés sans que l'on observe une abolition de la fixation à l'ADN (Rao et Reddy, 1992b). La

troncature en carboxy-terminal de Elk-1 (Rao et Reddy, 1992b) ou de SAP-1 (Dalton et Treisman, 1992) augmente la fixation à l'ADN suggérant que ces protéines contiennent un domaine d'inhibition de la liaison autonome à l'ADN. C'est la boîte B qui semble responsable de cette activité (Janknecht *et al.*, 1994). Comme elle est également en partie responsable de l'interaction avec SRF (Janknecht et Nordheim, 1992; Treisman *et al.*, 1992), on peut lui attribuer une fonction ambivalente : lorsque Elk-1 intervient de manière autonome (*in vitro*), la boîte B impose des contraintes structurales qui répriment la fixation à l'ADN, tandis qu'elle se pose comme un élément nécessaire pour l'interaction *in vivo* avec SRF et la fixation sur le SRE, en d'autres termes la boîte B favorise le recrutement de Elk-1 au niveau du promoteur de *c-fos*. Notons que la formation du complexe ternaire Elk-1-SRF-SRE est dépendante non seulement de la boîte B mais également de l'intégrité du domaine ETS (Rao et Reddy, 1992b; Janknecht et Nordheim, 1992). A première vue, ces données sont contradictoires avec les résultats signalant que la boîte B est nécessaire et suffisante à l'interaction directe avec le SRF (Shore et Sharrocks, 1994). Néanmoins ces derniers essais ont été réalisés en l'absence du SRE, on peut donc supposer que si l'association SRF-Elk-1 dépend strictement de la boîte B, les contraintes structurales apportées par des délétions dans le domaine ETS perturbent cette interaction. Il faut préciser que la formation du complexe ternaire est indépendante de la capacité de fixation autonome de Elk-1 : en effet, une mutation ponctuelle introduite dans le domaine ETS inhibe la liaison de Elk-1 sur E74BS sans perturber la formation du complexe ternaire (Janknecht et Nordheim, 1992). A l'inverse, la délétion des 14 premiers acides aminés du domaine ETS, tout en préservant la fixation autonome, abolit la formation du complexe ternaire. D'autres informations suggèrent que les mécanismes gouvernant la formation du complexe ternaire sont relativement indépendants de la fixation à l'ADN de Elk-1 ou SAP-1, en particulier le fait que les sites TCF et SRF présents sur le promoteur de *c-fos* peuvent être placés en orientation inverse ou éloignés d'un ou de deux tours d'hélice sans dommage pour la formation du complexe ternaire (Treisman *et al.*, 1992). La

coopération pour la fixation de TCF sur le SRE repose donc surtout sur l'association des deux protéines.

L'étude comparée du mode de fixation à l'ADN *in vitro* et *in vivo* des trois TCF a montré que ERP/NET se fixe au *c-fos* SRE très faiblement contrairement à Elk-1 et SAP-1a, et que la formation de complexes ternaires engageant ERP/NET est négligeable (Price *et al.*, 1995). Les trois TCF, Elk-1, SAP-1 et SAP-2 sont des facteurs de transcription. L'activation transcriptionnelle par les TCF est augmentée par la phosphorylation de résidus hydroxylés conservés situés dans la partie carboxy-terminale des protéines. La substitution en alanine de quatre de ces résidus diminue fortement l'activité de Elk-1 et SAP-1a, la seule variation s'observant pour la mutation de la sérine 381 dont l'effet s'avère moins drastique dans le cas de SAP-1a (Price *et al.*, 1995). La régulation de la transcription de *c-fos* est exercée par les TCF en réponse à l'activation du signal de transduction impliquant les kinases MAP du groupe ERK (telle ERK2) mais aussi en réponse à d'autres signaux qui induisent une cascade différente comme l'exposition aux rayons UV ou le CSF (pour Colony-Stimulating Factor).

Les kinases MAP du groupe ERK phosphorylent Elk-1 (Gille *et al.*, 1992; Gille *et al.*, 1995) et les autres TCF (Price *et al.*, 1995). Elk-1 est phosphorylé sur les sérines et les thréonines par les kinases de type ERK et JNK (Whitmarsh *et al.*, 1995). La phosphorylation de Elk-1 par les ERK s'effectue autant sur les sérines que sur les thréonines alors que celle réalisée par les JNK est plus spécifique des thréonines. En dépit de cette différence, il semble que ces deux types de kinases MAP affectent les mêmes groupes de sites. La phosphorylation par les kinases de type ERK, médiant le signal induit par le TPA (Tumor Promoting Agent), ou de type JNK, médiant le signal induit par l'IL-1 (interleukine-1), résulte en une augmentation de la formation de complexes ternaires et de l'activité transcriptionnelle par Elk-1.

Malgré les similarités observées dans le fonctionnement de Elk-1, SAP-1a et ERP/NET, certaines différences sont notables : la liaison à l'ADN et la formation de complexes ternaires impliquant ERP/NET sont plus faibles; le patron de

phosphorylation de SAP-1a est légèrement modifié par rapport à celui de Elk-1 et SAP-1a est moins phosphorylé par les JNK que Elk-1 (Whitmarsh *et al.*, 1995). Il est donc peu vraisemblable que les TCF soient totalement redondants : on peut imaginer que SRF va interagir avec un partenaire différent, selon le niveau de transactivation requis et, sans doute, le promoteur considéré (seul le cas de *c-fos* a été examiné).

Outre les qualités intrinsèques du mécanisme décrit ci-dessus, qui placent plusieurs facteurs de transcription dans le contexte de la transduction du signal mitotique, ce modèle est également l'occasion d'apprécier les joies de la régulation de l'activité de transcription par épissage alternatif. Le locus *elk-1* code Elk-1 mais aussi une autre forme appelée Δ Elk-1 qui est amputée de 143 acides aminés (Rao et Reddy, 1993) (Fig. 16). Cette délétion prive Elk-1 des onze derniers résidus du domaine ETS et de toute une région couvrant la boîte B. Les conséquences sont importantes : la délétion des 11 derniers acides aminés du domaine ETS permet la fixation autonome de la construction mais uniquement sur le site le plus affiné, en l'occurrence E74BS; PEA3 et MSV LTR ne sont pas reconnus. De plus, Δ Elk-1, privé de la boîte B, n'interagit plus avec SRF.

4.2.3. Elf-1, HMG-I(Y) et NF- κ B (des relations triangulaires).

La régulation de la transcription de la chaîne β du récepteur de l'interleukine 2 passe par deux éléments, PRRI (nucléotides -299 à -228) et PRRII (nucléotides -137 à -64) qui mettent en jeu les interactions entre trois protéines au moins (John *et al.*, 1995). Sur PRRI se fixe une protéine de la famille NF- κ B tandis que PRRII recrute un membre de la famille Ets et de la famille HMG (Fig. 17). Interférence d'alkylation et protection contre la DNase I ont mis en évidence qu'il s'agissait respectivement de Elf-1 et de HMG-I(Y) (Thompson *et al.*, 1992). La fixation sur l'élément PRRII de Elf-1 se réalise indépendamment de HMG-I (Y) et n'est pas favorisée par la présence de HMG-I (Y) ni par celle de NF- κ B. Elf-1 est

capable de transactiver seul mais il agit de manière coopérative avec HMG-1(Y) ce qui laisse supposer que ces protéines sont en contact l'une avec l'autre. L'interaction directe entre deux protéines peut se mettre en évidence par la technique de "pull-down" : un partenaire est exprimé sous la forme d'une protéine de fusion avec la glutathion-S-transférase. La chimère ainsi obtenue sera retenue sur une colonne de glutathion-sépharose et, s'il y a interaction directe, fixera le second partenaire contenu dans un extrait protéique. Dans le cas présent, Elf-1 interagit directement avec HMG-1(Y) et avec un membre de la famille NF- κ B par l'intermédiaire du domaine ETS.

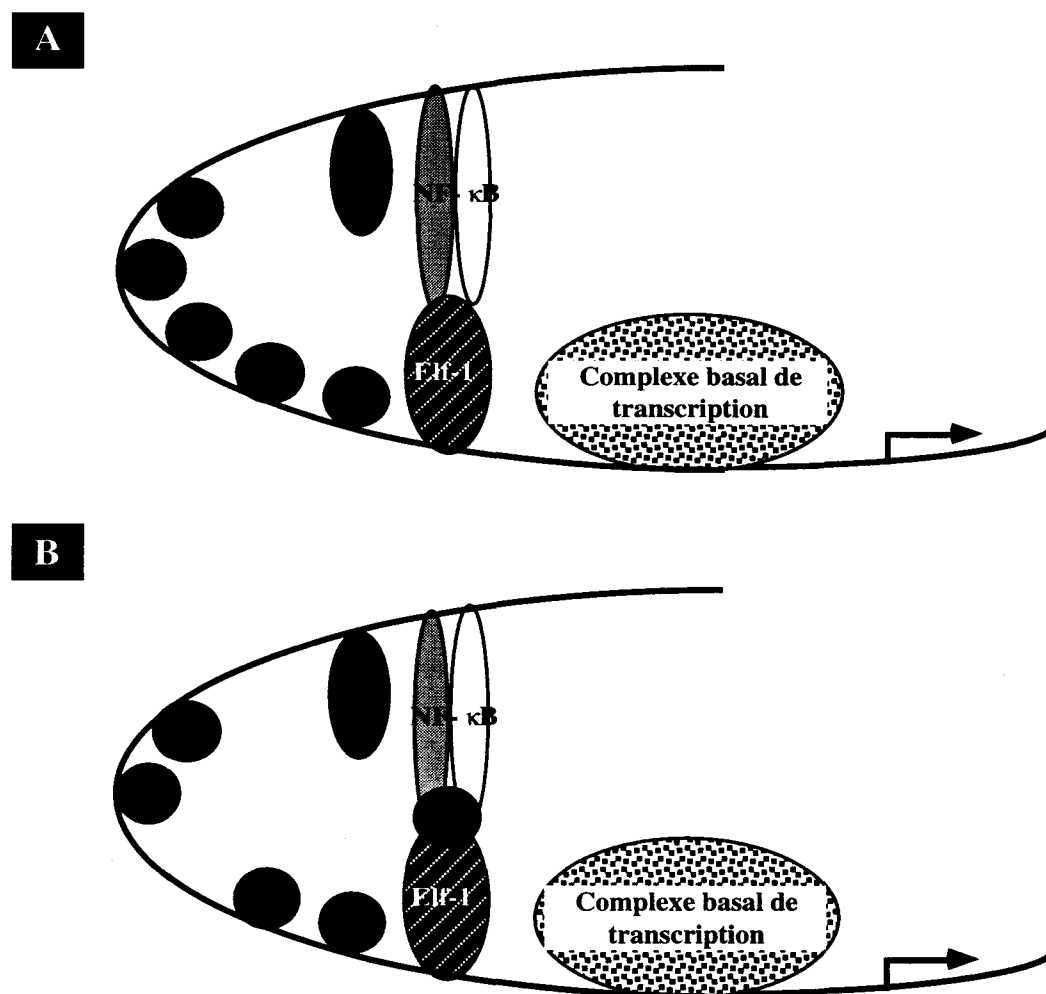


Figure 17 : Deux modèles possibles pour l'établissement du complexe protéique sur l'amplificateur du récepteur de l'interleukine 2 selon John *et al.* (1995). Les sphères noires représentent les molécules HMG-1(Y). (A) HMG-1(Y) n'intervient que

pour courber l'ADN. (B) HMG-1(Y) intervient également pour augmenter la transcription.

Le modèle proposé suggère que Elf-1 qui est spécifique des tissus lymphoïdes est exprimé suite à la stimulation des cellules T quiescentes et se fixe sur PRRII. Son interaction directe avec un membre de la famille NF- κ B (lié, lui, à PRRI) est facilitée par la présence de HMG-1(Y), véritable colle protéique, dont le rôle serait de déformer l'ADN, plaçant ainsi Elf-1 à proximité de la protéine de type NF- κ B, et/ou de servir de cofacteur "de pontage" entre Elf-1 et un autre facteur de transcription encore inconnu (Fig. 17). Cet exemple illustre parfaitement la dichotomie entre la fonction de liaison à l'ADN (Elf-1 est indépendant des autres protagonistes) et celle de transactivation (Elf-1 seul transactive de manière négligeable mais coopère avec NF- κ B en présence de HMG-1).

4.2.4. c-Ets-1, LEF-1, CBF et ATF/CREB (les parties fines).

L'amplificateur du gène du récepteur α des cellules T (TCR α) possède des sites de fixation pour les protéines c-Ets-1 (Ho *et al.*, 1990), LEF-1, ATF/CREB (Activating Transcription Factor/ cAMP Response Element-Binding) et CBF. CBF appelé aussi PEBP2 (pour Polyomavirus Enhancer Binding Protein 2) est constitué d'une sous-unité α spécifique des cellules T, et d'une sous-unité β exprimée de manière ubiquiste qui augmente faiblement l'activité de liaison à l'ADN de la sous-unité α sans le contacter. Des expériences de gel-retard et de "pull-down" ont démontré que c-Ets-1 seul est capable de se lier très faiblement à l'ADN mais que la présence de la protéine PEBP2 α qui interagit directement avec la région des acides aminés 123 à 240 de c-Ets-1 élève considérablement la capacité de liaison à l'ADN (Giese *et al.*, 1995). Réciproquement, c-Ets-1 augmente la fixation spécifique de PEBP2 α sur l'amplificateur du TCR α . Le détail de la coopération c-Ets-1-PEBP2 α pour la fixation sur l'élément de réponse du TCR β est exposé au chapitre 2.4.1., on

peut supposer qu'il est identique sur l'élément de réponse du TCR α . Le complexe établi n'est rendu stable que par la présence conjointe de deux autres polypeptides : ATF/CREB et LEF-1 qui est un membre de la famille HMG (Giese *et al.*, 1995) (Fig. 18). Il a été montré que c-Ets-1 et LEF-1 ainsi que c-Ets-1 et PEBP2 α coopéraient pour la transactivation à partir d'un élément de réponse TCR α .

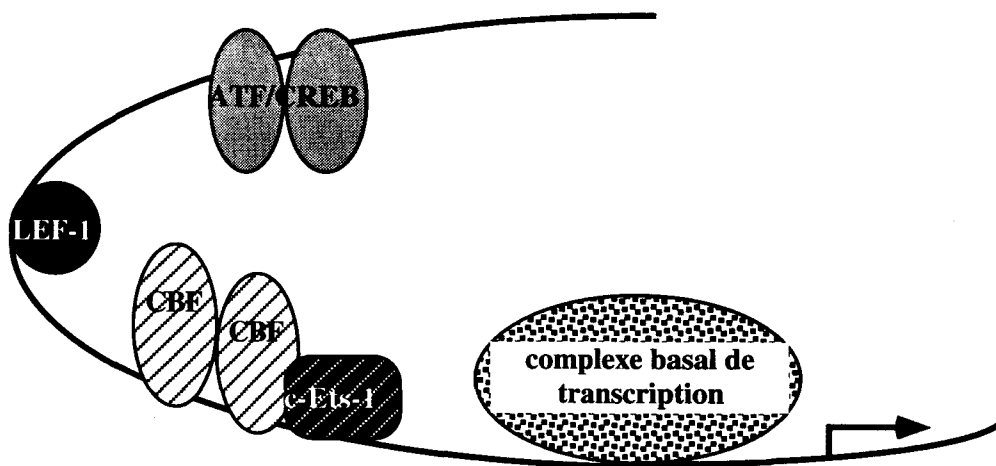


Figure 18 : Représentation schématique du complexe multiprotéique défini au niveau de l'amplificateur du TCR α selon Giese *et al.* (1995). La position de LEF-1 indique son rôle dans le "bending" de l'ADN.

Les différents membres de la famille Ets ne sont pas interchangeable dans ce contexte : même si on ne peut attribuer une spécificité de fixation à l'ADN propre à chaque protéine Ets, le remplacement de c-Ets-1 par Fli-1 abolit l'augmentation de la transcription du complexe, soit parce que le mode de transactivation de c-Ets-1 et de Fli-1 est différent, soit parce que Fli-1 n'interagit pas avec les mêmes partenaires. Il semble en particulier que Fli-1 contrairement à c-Ets-1 ne soit pas capable de s'associer à ATF/CREB. Hors domaine ETS, le contexte protéique octroie donc un niveau de discrimination supplémentaire aux facteurs de transcription Ets. En revanche, la participation de LEF-1 est moins spécifique que celle de c-Ets-1 et relève plus d'une activité de "bending" (LEF-1 courbe l'ADN sur lequel il se lie) que d'interactions spécifiques protéine-protéine, ainsi que le suggèrent les expériences

suivantes : premièrement, LEF-1 peut être substitué par son domaine de liaison à l'ADN sans que cela entraîne une diminution catastrophique du niveau de transactivation exercé par le complexe c-Ets-1-LEF-1-PEBP2 α . Deuxièmement, le domaine de liaison de LEF-1 peut être substitué par le domaine de liaison à l'ADN d'un autre membre de la famille HMG avec lequel il ne présente que 24% d'identité de séquence, celui de SRY, SRY possédant lui aussi la capacité de courber l'ADN. Le rôle de LEF-1 serait donc plus celui d'une mise en place de l'architecture nécessaire à la transactivation que celui d'une augmentation de la transcription. Cette constatation est à rapprocher de celle faite pour le complexe Elf-1-HMG-1(Y)-NF- κ B dans lequel HMG-1(Y) semble jouer le rôle de "colle moléculaire".

Un parallèle peut être établi entre le complexe SRF-TCF et celui détaillé dans ce chapitre. D'abord, la liaison à l'amplificateur de TRC α de c-Ets-1 et CBF est augmentée d'environ 20 fois, comme c'est le cas pour SRF avec Elk-1 ou SAP-1. Ensuite, la boîte B des TCF et la région 123-240 de c-Ets-1 possèdent toutes deux des fonctions similaires et ambivalentes. L'inhibition de la liaison à l'ADN exercée par la boîte B est levée par l'interaction du SRF sur la boîte B; de même, l'inhibition de la fixation exercée par la région 207-280 de c-Ets-1 est contrebalancée par la fixation de CBF sur une zone (acides aminés 123-240) qui recouvre largement la région d'inhibition de la liaison à l'ADN. Selon le contexte (absence ou présence de cofacteur), la boîte B des TCF et la région d'inhibition 207-280 de c-Ets-1 jouent donc un rôle ambivalent.

5. A quoi peuvent donc bien servir les facteurs de transcription Ets?

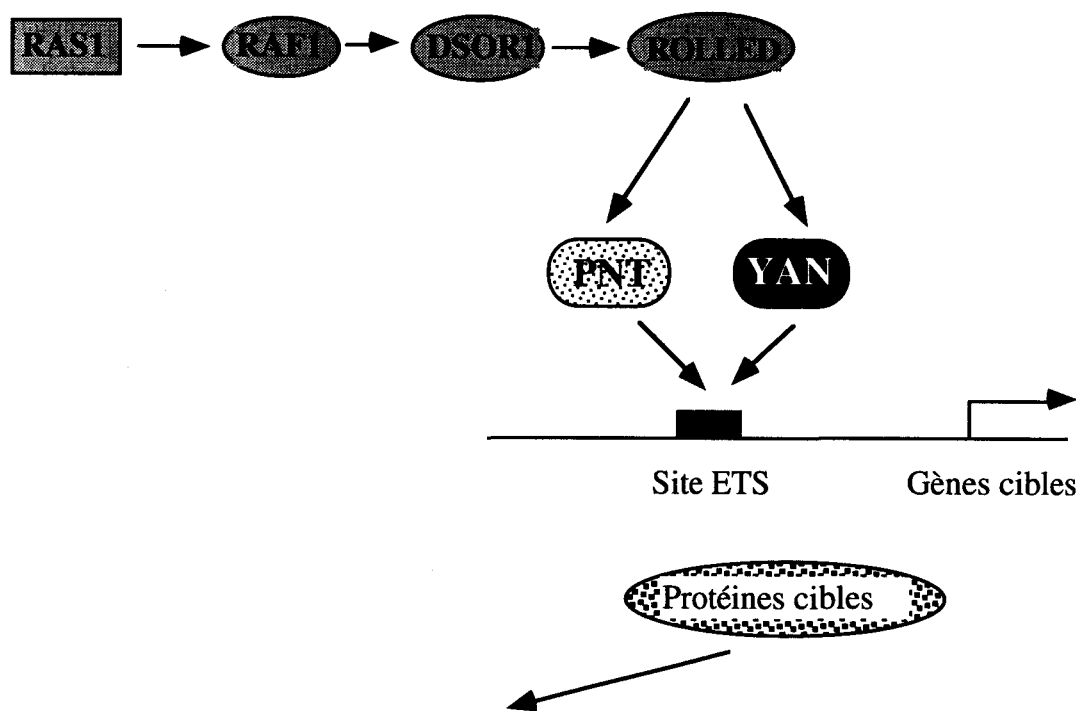
Nous avons choisi de ne développer que quelques exemples impliquant les protéines Ets et nous nous étendrons uniquement sur deux modèles de

développement mettant en jeu les facteurs de transcription Ets : la différenciation des photorécepteurs et le développement de la drosophile. Il semblerait que certains membres de la famille soient également engagés dans le développement embryonnaire d'autres espèces tels le xénope ou le poulet. Nous avons cependant pris le parti de passer sous silence ces études (sans que cela n'ôte rien de leur valeur) car la description de la répartition des ARN messagers et, chez le xénope, l'examen de l'impact après micro-injection d'ARN antisens ne constituent pas un matériel suffisant pour juger l'implication d'une protéine dans un processus de différenciation. La présence d'ARN messagers n'est pas le garant absolu de celle des protéines, encore moins d'une protéine ayant subi toutes les modifications post-traductionnelles nécessaires à son activité maximale. Enfin, il faut remarquer que les études d'hybridation *in situ* ne prennent pas toujours en compte la possibilité d'un épissage alternatif qui peut donner naissance à des protéines d'activité différente, comme nous l'avons signalé pour Elk-1 et Δ Elk-1.

5.1. Les protéines Ets dans le développement.

5.1.1. L'incroyable histoire des récepteurs R7 et des protéines Ets.

L'oeil de la drosophile est composé d'ommatidies, chacune de ces structures étant composée de huit photorécepteurs (R1 à R8). Le phénotype du mutant *sevenless* correspond à l'absence du photorécepteur R7. Le produit du gène *sevenless* est un récepteur à tyrosine kinase par lequel, en réponse au signal produit par R8, les récepteurs R7 se différencient. L'élucidation du mécanisme moléculaire de cette différenciation a été l'objet de nombreux efforts qui ont permis d'identifier plusieurs éléments de cette cascade dont Ras, Raf, MEK (kinase MAP codée par le gène *Dsor1*) et la kinase MAP codée par le gène *rolled* (Fig. 19). Les tentatives d'identification des cibles nucléaires de cette cascade ont mis à jour deux candidats sérieux : pntP2 et YAN/POK (Tei *et al.*, 1992) qui présentent un intérêt particulier pour nous, amoureux de la famille Ets.



Différenciation des récepteurs R7

Figure 19 : Modèle pour l'implication dans le développement de la drosophile de deux protéines Ets (Pnt et YAN/POK) selon O'Neill *et al.* (1994).

YAN/POK a été originellement identifié comme un potentiel facteur de transcription présent, au cours du développement de l'oeil, dans le noyau des cellules indifférenciées et qui serait réprimé après l'induction de la différenciation (Lai et Rubin, 1992). L'étude de différentes mutations sur *yan* a permis de suggérer que le produit du gène *yan* jouait le rôle de régulateur négatif dans le développement des cellules R7 en limitant l'engagement des cellules précurseurs dans la voie de différenciation des cellules R7 (Lai et Rubin, 1992; Tei *et al.*, 1992).

Les produits de l'épissage alternatif du gène *pointed*, *pntP1* et *pntP2*, codent les protéines qui ont en commun le domaine ETS, situé dans leur partie carboxy-terminale mais qui diffèrent par leur extrémité amino-terminale (Klämbt, 1993). Brunner *et al.* (1994) ont clairement montré que *pntP2* et YAN/POK étaient effectivement phosphorylés *in vitro* et *in vivo* par la kinase MAP encodée par *rolled*.

La phosphorylation de pntP2 dépend d'un seul résidu, unique site potentiel des kinases MAP; celle de YAN/POK dépend également d'un seul acide aminé parmi les dix sites potentiels. L'analyse de mutants de drosophile pour *yan* et *pointed* a fortement laissé penser que les produits de ces gènes jouaient un rôle important dans la différenciation des cellules R7. En effet, les mutants *yan* sont caractérisés par la présence des nombreux récepteurs R7 tandis que les doubles mutants *yan-pointed* ne présentent pas de récepteurs R7, ce qui indique que la différenciation des cellules R7 est sensible à la présence des protéines pntP2 et YAN/POK, même si elles n'interviennent pas forcément dans les mêmes voies de transduction du signal extracellulaire (Brunner *et al.*, 1994). La démonstration claire du rôle des protéines Ets comme cible de la phosphorylation par les kinases MAP a été faite par O'Neill *et al.* (1994). Sachant que, *in vivo*, pntP2 agissait en aval de Ras, des essais de transfection *in vitro* ont montré que l'activité de transcription des produits du gène *pointed* était différemment affectée par la coexpression d'une protéine Ras activée : l'activité de pntP2 est augmentée, au contraire de celle de pntP1. En outre, l'activité de transcription d'une version de pntP2 mutée en son site de phosphorylation n'est plus augmentée ni par Ras, ni par la kinase MAP codée par le gène *rolled*, connue pour agir en aval de Ras; ceci dénote l'importance de la phosphorylabilité de pntP2. Les mêmes techniques, confortées par l'analyse génétique, montre que YAN/POK est phosphorylé en réponse à Ras activé et à la kinase MAP, et que cette phosphorylation est sans doute le moyen par lequel Ras et la kinase MAP régulent négativement la répression de la transcription par YAN/POK. Le mécanisme de l'inhibition de la répression exercée par YAN/POK a récemment été élucidé : la phosphorylation par la kinase MAP entraîne une modification de la localisation cellulaire de YAN/POK qui quitte le noyau, puis une dégradation rapide de la protéine (Rebay et Rubin, 1995). Cet article propose également que YAN/POK, qui est exprimé dans l'embryon et dans les disques imaginaux, structures précurseurs des différents segments du corps de la drosophile, fonctionne comme un inhibiteur général de la différenciation. Les données obtenues pour l'instant sur le mode de

fonctionnement de YAN/POK ne permettent pas de trancher entre une répression passive (par la séquestration d'un cofacteur important, par exemple) et l'activité d'un vrai domaine répresseur de la transcription.

En résumé, la transduction du signal extracellulaire semble passer par Ras et une kinase MAP, deux éléments de la cascade qui vont permettre la phosphorylation de pntP2 (alors activé) et de YAN/POK (alors réprimé), deux facteurs de transcription dont l'activité est requise pour la différenciation normale des cellules R7. L'étude de l'implication de pntP2 et YAN/POK dans la transduction du signal extracellulaire pour la différenciation des photorécepteurs R7 est un petit bijou qui combine élégamment les techniques classiques de la génétique et de la biologie moléculaire. C'est aussi à l'heure actuelle le seul modèle sérieux qui permette d'attribuer définitivement un rôle aux protéines Ets.

5.2.2. E74.

Chez la drosophile, l'hormone stéroïde qui coordonne de multiples événements au cours du développement et de la croissance de certains tissus est la 20-hydroxy-ecdysone (ecdysone). A chacun des six stades du développement de la drosophile, on observe un pic du taux d'ecdysone. Ashburner a proposé un modèle pour la régulation génétique par l'ecdysone de la formation de "puffs" sur les chromosomes polyténiques (Fig. 20) (Ashburner *et al.*, 1974). On peut définir le "puff" comme un renflement qui est présent sur les chromosomes polyténiques géants des glandes salivaires et qui est la manifestation d'une activité transcriptionnelle intense. Selon le modèle suggéré, l'ecdysone s'associerait à son récepteur et agirait directement pour induire la formation précoce de "puffs" et pour réprimer la formation de "puffs" tardifs. Les protéines qui sont les produits des "puffs" précoces activeraient transitoirement les "puffs" tardifs puis, après une période déterminée, réprimeraient leur propre expression.

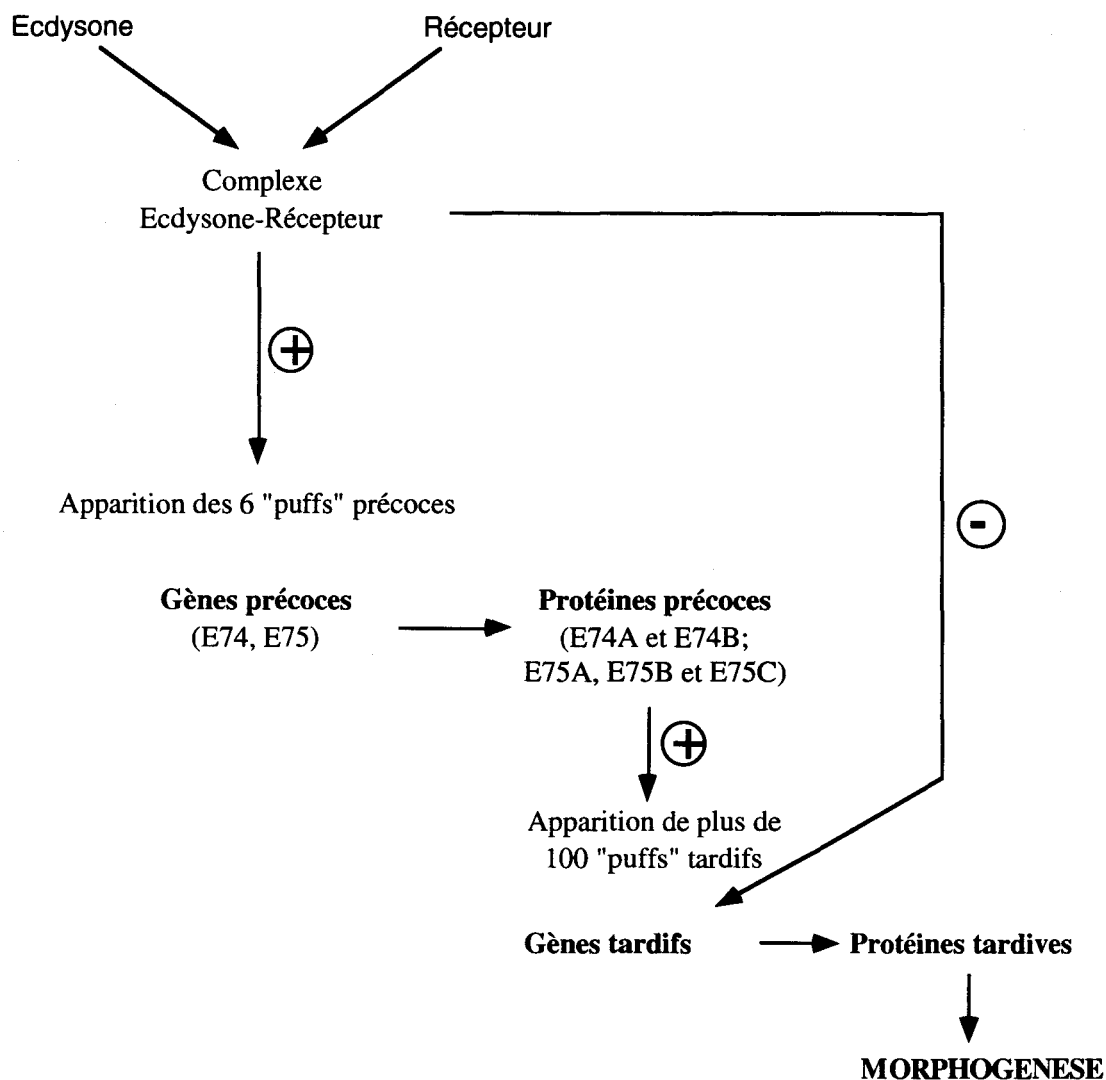


Figure 20 : Modèle d'Ashburner pour le contrôle génétique par l'ecdysone de "puffs" du chromosome polyténique selon Ashburner *et al.* (1974) et Burtis *et al.* (1990).

Un des gènes précoces inductibles par l'ecdysone, E74, est responsable de la formation de "puffs" dans la région 74EF des chromosomes polyténiques. L'épissage alternatif auquel il est soumis donne trois ARN messagers codant les deux protéines E74A et E74B qui possèdent une extrémité carboxy-terminale commune contenant un domaine ETS (Burtis *et al.*, 1990; Urness et Thummel, 1990). Leur extrémité

amino-terminale, spécifique de chaque protéine et riche en résidus acides, en fait des facteurs de transcription potentiels. L'analyse génétique de mutants E74 a récemment permis d'examiner le rôle joué par E74A et E74B non seulement pendant la métamorphose mais aussi au niveau de la formation de "puffs" dans les chromosomes polyténiques de la glande salivaire (Fletcher *et al.*, 1995). La plupart des mutations examinées sont létales, suggérant que E74A et E74B interviennent de manière fondamentale avant le stade de la chrysalide, en particulier à l'étape de l'émergence de la tête à partir du thorax. E74B préviendrait la dégénérescence induite par l'ecdysone des muscles abdominaux de la larve. La perte de la fonction de E74B affecte aussi les disques imaginaux. L'existence de E74A et E74B est également requise pour la formation correcte des "puffs" des chromosomes de la glande salivaire en réponse à l'ecdysone bien que les mutations sur E74B affectent peu ce processus. Par contre, les mutations sur E74A entraînent la formation de "puffs" tardifs surnuméraires, ce qui renforce le modèle d'Ashburner selon lequel les produits des gènes des "puffs" précoces (en l'occurrence E74A) réguleraient la formation des "puffs" tardifs.

5.2. Ets et l'oncogénèse.

5.2.1. Transformation.

Dans des conditions de culture en concentration de sérum faible ou nulle, les protéines c-Ets-1 aviaire (Seth et Papas, 1990) et c-Ets-2 murine (Seth *et al.*, 1989) sont capables de modifier des lignées de cellules comme les NIH3T3. Les cellules sont alors douées de capacité à pousser en milieu semi-solide et induisent l'apparition de tumeurs chez les souris chez lesquelles elles ont été injectées, indiquant par conséquent que c-Ets-1 et c-Ets-2 possèdent une activité oncogénique. Il est difficile de concilier ces résultats avec le fait que l'expression ectopique de c-Ets-1 dans les lignées humaines de cancer du colon DLD-1 et

HCT116 semble réverter leur phénotype (Suzuki *et al.*, 1995). L'expression ectopique de c-Ets-1 par transfection stable diminue leur capacité à croître en milieu semi-solide et à former des tumeurs après injection chez la souris; cet effet est proportionnel à la quantité de c-Ets-1 transfectée et il est inhibé par une mutation ponctuelle qui abolit l'activité d'augmentation de la transcription de c-Ets-1 sans modifier sa capacité de liaison à l'ADN. Peut-être est-il inutile de s'interroger plus avant sur le modèle expliquant ces effets contradictoires car on ne peut considérer la transfection transitoire comme biologiquement "relevante".

Les propriétés de transformation par des membres de la famille Ets ont été beaucoup plus approfondies dans le cas de v-Ets. Parmi les virus induisant une leucémie aiguë chez le poulet, le virus E26 possède un statut à part pour plusieurs raisons. Premièrement, il transduit deux oncogènes, *v-myb* et *v-ets* qui codent tous deux des facteurs de transcription. Deuxièmement, ces deux oncogènes sont exprimés dans une protéine chimère tripartite Gag-Myb-Ets de 135 kDa de poids moléculaire. Et troisièmement, il cause de manière prédominante des érythroleucémies mais possède la capacité de transformer les cellules myéloïdes *in vitro* et *in vivo*, touchant ainsi plusieurs lignées du système hématopoïétique du poulet.

Lorsque des cellules érythroïdes sont transformées *in vitro* par *v-myb* ou *v-ets*, elles poussent lentement et ont une durée de vie limitée. Lorsqu'elles sont transformées par la combinaison de *v-myb* et *v-ets* (produisant alors des protéines distinctes), elles poussent plus vite et ont une durée de vie plus importante (Metz et Graf, 1991a). Par contre, des virus contenant la séquence chimère *v-myb-v-ets*, transforment des cellules beaucoup plus immatures. Ces expériences montrent que les deux oncogènes sont capables de coopérer pour la transformation des cellules érythroïdes mais ont des cibles cellulaires préférentielles selon que les deux protéines sont exprimées sous forme de chimère ou indépendamment. Les spéculations portant sur cette différence de comportement ont mis en cause soit une nouvelle spécificité acquise par la protéine chimère, acquise par des modifications conformationnelles ou due à la présence de deux domaines de liaison à l'ADN, soit des nouvelles facultés de

transactivation acquises par la présence de plusieurs domaines responsables de la transactivation ou de l'inhibition de la transactivation.

Pour la leucémogénèse, il semble que la coexpression des oncoprotéines *v-Myb* et *v-Ets* ne soit pas suffisante et que les cellules transformées sélectionnées ont "recrée" la protéine chimère *Myb-Ets*. La présence des domaines de liaison à l'ADN de *Ets* et *Myb* ainsi que celle du domaine de transactivation de *Myb* semble requise pour cet effet : la délétion d'une partie du domaine de transactivation de *Myb* est critique pour la leucémogénèse, mais aucune délétion n'a été réalisée sur les domaines de fixation qui permettrait de montrer leur implication directe (Metz et Graf, 1991b). En revanche, on sait que le virus E26 est nécessaire à la stimulation de la croissance des fibroblastes de poulet (Ravel-Chapuis *et al.*, 1991) et que dans ce cas, l'intégrité dans la protéine chimère *Gag-Myb-Ets* des domaines de liaison à l'ADN de *Myb* et de *Ets* est requise, ce qui laisse supposer non seulement que *Myb* et *Ets* coopèrent pour l'effet mitogène de E26 mais aussi que leur liaison à l'ADN est indispensable.

Notons que la coopération de *v-ets* avec un second oncogène ne passe pas dans tous les cas par la synthèse d'une chimère, loin de là. Le virus AEV exprime deux oncogènes : *v-erbA* et *v-erbB*, le produit du premier augmentant l'effet de la tyrosine-kinase encodée par le second en bloquant la différenciation des cellules érythroïdes. Un virus artificiel portant *v-erbA* et *v-ets* induit une érythroleucémie caractéristique de celles induite par AEV (Metz et Graf, 1992). En dépit du fait qu'ils ne jouent pas le même rôle, *v-Ets* et *v-ErbA* ne peuvent induire d'érythroleucémie l'un sans l'autre. Les deux oncoprotéines exprimées indépendamment coopèrent, *v-ErbA* se cantonnant à son activité de blocage de la différenciation pendant que *v-Ets* lève la dépendance vis-à-vis du sérum.

Il a aussi été décrit des situations où la protéine *Gag-Myb-Ets* et peut-être la partie *Ets* plus spécifiquement, est impliquée dans la coopération avec des protéines cellulaires. En réponse à la transformation par l'équivalent murin de E26, l'expression du facteur de transcription GATA-1 augmente dans les cellules érythroïdes (Aurigemma *et al.*, 1992). Le haut niveau d'expression du gène de

GATA-1 est "positivement corrélé" à l'activation du gène du récepteur de l'érythropoïétine (EpoR) supposant que GATA-1 active la transcription de EpoR. De plus, les auteurs ont supposé que le produit du rétrovirus E26 pouvait interagir avec GATA-1 en modifiant l'activité de liaison à l'ADN ou le pouvoir transactivateur de GATA-1. Une telle hypothèse, si elle était vérifiée, suggère que les protéines cellulaires c-Myb ou c-Ets seraient impliquées dans le processus normal de l'érythropoïèse et que la transformation interviendrait par des protéines dénaturées. Le mécanisme de la coopération de v-Ets avec une autre oncoprotéine ou avec une protéine cellulaire reste encore à analyser.

5.2.2. Ets et la réaction stromale dans le processus de tumorigénèse.

Dans les tumeurs malignes, une étape cruciale du phénomène de métastase est l'invasion au cours de laquelle l'adhésion des cellules est modifiée et le remodelage de la matrice extracellulaire induit. Certaines protéases telles la collagénase-1, l'u-PA (pour urokinase-Plasminogen Activator) ou la stromélysine-1 semblent impliquées dans la dégradation de la matrice extracellulaire (Blood et Zetter, 1990) tant au niveau post-traductionnel que transcriptionnel. Certains facteurs de transcription prennent part à la régulation des gènes de ces protéases : *in vitro* en transfection transitoire, c-Ets-1 augmente la transcription à partir du promoteur de la stromélysine-1 (Wasylyk *et al.*, 1991), de l'u-PA (Rorth *et al.*, 1990) et peut-être de la collagénase (Wasylyk et Wasylyk, 1992). Dans le modèle murin et dans de nombreux carcinomes humains, les transcrits c-Ets-1 sont détectés au niveau des fibroblastes du stroma alors qu'ils sont absents au niveau des cellules épithéliales de la tumeur (Wernert *et al.*, 1994). De plus, lorsque des fibroblastes en culture sont stimulés par le facteur de croissance basique des fibroblastes (bFGF) ou par le facteur α de nécrose des tumeurs (TNF α), l'expression de transcrits c-Ets-1, collagénase-1 et stromélysine-1 est corrélée ce qui laisse penser que c-Ets-1 pourrait

être impliqué dans l'invasion tumorale par le contrôle transcriptionnel qu'il pourrait exercer sur les protéases impliquées dans la dégradation de la matrice extracellulaire. Les transcrits c-Ets-1 sont détectés dans deux structures néoformées : les vaisseaux sanguins en voie de formation dans l'embryon de poulet (Vandenbunder *et al.*, 1989) et dans les vaisseaux sanguins qui irriguent la tumeur (Wernert *et al.*, 1992). Il est tentant d'imaginer que c-Ets-1 contrôle effectivement ces phénomènes en modulant l'expression des protéases impliquées dans la formation de nouveaux tissus, normaux ou tumoraux.

5.2.3. Activation par insertion rétrovirale.

La transformation par certains rétrovirus peut impliquer l'intégration de l'ADN proviral dans des gènes cellulaires spécifiques ou à leur proximité et la conséquence de la plupart de ces intégrations est une augmentation du niveau d'expression de ces gènes le plus souvent par l'insertion d'un promoteur ou par activation *via* un amplificateur.

Une leucémie peut être induite chez la souris après injection de l'un des deux variants du virus de Friend : FV-A ou FV-P qui sont un complexe composé d'un virus compétent pour la réplication (F-MuLV) et un virus déficient pour la réplication SFFVa ou SFFVp (Moreau-Gachelin, 1994). Le SFFV est l'agent causal direct de l'érythroleucémie aiguë et sa glycoprotéine d'enveloppe gp55, en interagissant directement avec le récepteur de l'érythropoïétine, perturbe la prolifération et la différenciation des cellules de la lignée érythroïde. L'expression de la gp55 ne suffit pas à induire des tumeurs malignes et des événements tardifs impliquent l'activation d'oncogènes de la famille Ets et l'inactivation de la p53 (que nous n'évoquons pas ici) (Ben-David *et al.*, 1991a).

Chez la souris, le gène *spi-1* (pour SFFV proviral integration-site 1) est activé par insertion rétrovirale dans 95% des érythroleucémies induites par injection des souches FLV-P et FLV-A du virus de Friend (Moreau-Gachelin *et al.*, 1988) et dans

les cellules où le locus *spi-1* est réarrangé, les transcrits de ce gène ne sont pas modifiés mais leur taux est élevé. Pour les tumeurs infectées par le F-MuLV, ce n'est pas *spi-1* mais *fli-1* (pour Friend leukemia integration-site 1) qui est réarrangé. Ce phénomène ne touche que 75% des érythroleucémies induites et n'affecte les leucémies myéloïdes ou lymphoïdes induites par le F-MuLV (Ben-David *et al.*, 1990). Dans les deux cas, l'insertion provirale correspond à un réarrangement en amont de la phase codante du gène *ets* et en orientation inverse par rapport au sens de transcription.

Chacun des deux membres de la famille *ets* répond donc à un type d'induction rétrovirale : pour *spi-1*, ce sont les leucémies induites par les complexes du virus de Friend, pour *fli-1* les érythroleucémies induites par le virus auxiliaire F-MuLV. Cette préférence affirmée pourrait être le reflet d'une spécificité de reconnaissance de l'ADN ou de transactivation des deux oncogènes impliqués (Ben-David *et al.*, 1991b). Néanmoins, les clones leucémiques induits par infection des différentes souches du virus de Friend et par F-MuLV présentent un phénotype similaire, ce qui laisse à penser que les gènes activés par le processus oncogénique sont identiques. Une explication plus probable est que SSFV et F-MuLV ont des cibles cellulaires différentes. En effet, F-MuLV n'induit des érythroleucémies que chez les souriceaux nouveau-nés.

5.2.4. Activation par translocation.

En outre, les membres de la famille *Ets* sont impliqués dans un mécanisme de tumorigenèse conduisant à la production de protéines chimères aberrantes. Chez l'homme, la translocation t(11;22) (q24;q12) du sarcome d'Ewing (tumeur d'origine neuroectodermique) entraîne la synthèse d'une protéine chimérique EWS-Fli-1 dans laquelle le domaine de liaison à l'ARN putatif de EWS est remplacé par le domaine de liaison à l'ADN de Fli-1 humain (Delattre *et al.*, 1992). Les conséquences sont de plusieurs ordres : le promoteur de EWS sous lequel est placée

la chimère EWS-Fli-1, n'est par définition pas celui de Fli-1. La protéine EWS-Fli possèdera donc un patron d'expression différent de celui de la protéine Fli-1 et altèrera le fonctionnement normal de Fli-1 : le domaine ETS de la chimère pourrait soit occuper des séquences reconnues normalement par Fli-1, agissant ainsi par compétition vis-à-vis des gènes cibles, soit canaliser des cofacteurs potentiels de Fli-1 (phénomène de "squelching"). En outre, la translocation met au voisinage du domaine ETS de Fli-1 la partie amino-terminale de EWS qui, riche en glutamine, thréonine, proline et en motifs sérine-tyrosine, présente toutes les caractéristiques d'un domaine responsable de la transactivation. En effet, alors que Fli-1 et EWS-Fli-1 affichent les mêmes capacités de liaison à l'ADN et la même affinité pour une séquence cible (Mao *et al.*, 1994), il a été démontré que la partie EWS de la chimère EWS-Fli-1 était un transactivateur plus puissant que la partie Fli-1 qu'elle remplace (May *et al.*, 1993a). Cette propriété a été vérifiée dans le cadre de la protéine de pleine longueur (Bailly *et al.*, 1994). De plus cette chimère, contrairement à Fli-1, possède la capacité de transformer des cellules NIH3T3 (May *et al.*, 1993a; May *et al.*, 1993b). Il semble donc qu'au moins dans le cas de Fli-1, l'existence concomitante d'une chimère engageant EWS et un membre de la famille Ets et d'une tumeur ne soit pas une simple coïncidence.

Le second membre de la famille Ets impliqué dans le sarcome d'Ewing est Erg. La translocation t(16;21) (p11;q22) est une anomalie qui touche la région codante de *erg* (Shimizu *et al.*, 1993). Les ARN messagers correspondant aux deux protéines de fusion possibles (extrémité amino-terminale de Erg et extrémité carboxy-terminale du gène du chromosome 16, ou inversement) sont exprimés. Il est intéressant de noter que la cassure s'effectue exactement au même endroit pour *fli-1*, par la translocation évoquée dans le paragraphe précédent, et *erg*, par la translocation t(16;21), deux gènes *ets* qui présentent environ 80% d'identité de séquence en amino-acides. On peut donc aisément supposer que la chimère contenant le domaine de liaison à l'ADN de Erg et la partie carboxy-terminale du gène situé sur le chromosome 16 est l'équivalent de la protéine EWS-Fli-1.

La translocation t(7;22) (p22;q12) du sarcome d'Ewing induit la production d'une protéine chimère EWS-ETV1, où ETV1 est l'équivalent humain de ER81. Comme dans les deux cas exposés ci-dessus, on obtient une fusion de la partie amino-terminale de EWS avec le domaine ETS de ETV1 qui se lie de manière spécifique à un oligonucléotide reconnu par les protéines Ets (Jeon *et al.*, 1995). Ce nouvel exemple confirme que les chimères de EWS avec le domaine de liaison à l'ADN de différents membres de la famille Ets mènent à l'expression de phénotypes comparables.

Sur un ensemble de treize protéines de fusion, produits des translocations impliquant la bande q12 du chromosome 22, toutes contenaient la partie amino-terminale de EWS et le domaine ETS de Fli-1 ou de Erg (Zucman *et al.*, 1994). Des résultats plus récents indiquent que les chimères impliquant ETV1 sont très minoritaires par rapport à celle impliquant Fli-1 ou Erg (O. Delattre, communication personnelle). A défaut d'un marqueur histologique ou pathologique, c'est la translocation impliquant EWS et un membre de la famille Ets qui peut pour l'instant être considéré comme marqueur du sarcome d'Ewing.

Afin de déterminer la contribution des chimères EWS impliquant des membres de la famille Ets, certaines cibles de EWS-Fli-1 ont été récemment mises en évidence par la technique de RDA (pour Representational Difference Analysis) (Braun *et al.*, 1995). Cette technique a permis d'identifier une dizaine de gènes exprimés différenciellement dans des NIH3T3 transfectées stablement avec Fli-1 ou avec EWS-Fli-1. Parmi ces gènes, on notera :

- 1) la stromélysine 1, métalloprotéinase qui digère diverses protéines de la matrice extracellulaire, et dont le promoteur contient deux sites ETS,
- 2) la cytokératine 15 dont un proche voisin, la cytokératine 8, est modulé par un amplificateur qui répond aux protéines Ets,
- 3) le cytochrome p450.

Il faut souligner qu'on ignore si ces gènes sont des cibles directes de EWS-Fli-1. Une version EWS-Fli-1 tronquée des 80 résidus amino-terminaux de EWS qui néanmoins

est capable, au même titre que EWS-Fli-1, de transformer les cellules NIH3T3, ainsi que EWS-Erg et TLS-Fli-1, où TLS est partie intégrante d'une chimère TLS-Erg exprimée dans certaines leucémies myéloïdes aiguës. Ces trois chimères régulent les mêmes gènes que EWS-Fli-1, au contraire de EWS-ETV1.

Les protéines aberrantes détectées dans les tumeurs du sarcome d'Ewing ne sont pas les seules à impliquer des membres de la famille Ets. La leucémie myélomonocytaire chronique est caractérisée par l'expression d'une protéine de fusion comprenant les 154 résidus amino-terminaux d'un dernier né Ets baptisé Tel (pour Translocation Ets Leucémie) avec les domaines transmembranaire et tyrosine-kinase du récepteur β du PDGF (Golub *et al.*, 1994). Cette chimère contient à la fois un domaine qui pourrait jouer un rôle dans la transactivation par Tel et le domaine ETS de Tel; elle a en revanche perdu le domaine responsable de la liaison au ligand et constituerait par conséquent une kinase constitutivement active. Cependant, un lien plus direct entre l'existence de cette chimère et la tumorigenèse reste à établir.

6. Le cas particulier du groupe PEA3.

6.1. Les différents membres du groupe PEA3.

PEA3, le membre fondateur du bien nommé groupe PEA3, est une protéine de 555 résidus présentant une région d'identité avec les autres membres de la famille Ets, le domaine ETS, qui est conservé à environ 60% avec la plupart des membres de la famille mais seulement 49% et 36% respectivement avec E74 et SPI-1. Le gène PEA3 (pour Polyoma Enhancer Activator 3) a été isolé d'extraits de cellules 3T6 de souris, hôte naturel du virus du polyome. L'ADN complémentaire encode une protéine qui se fixe de manière spécifique sur le motif PEA3 compris dans l'amplificateur du virus du polyome (Xin *et al.*, 1992) puisque des mutations

dans le motif PEA3 sont capables d'inhiber *in vitro* la fixation de la protéine PEA3. La séquence consensus établie à partir de ces expériences est 5'-A/CGGAAGT-3'. Outre le domaine ETS, PEA3 présente deux régions singularisées l'une (résidus 124 à 150) par sa richesse en acides aminés acides (33%), l'autre (résidus 217 à 315) par la présence de nombreuses glutamines (17%) (Fig. 21).

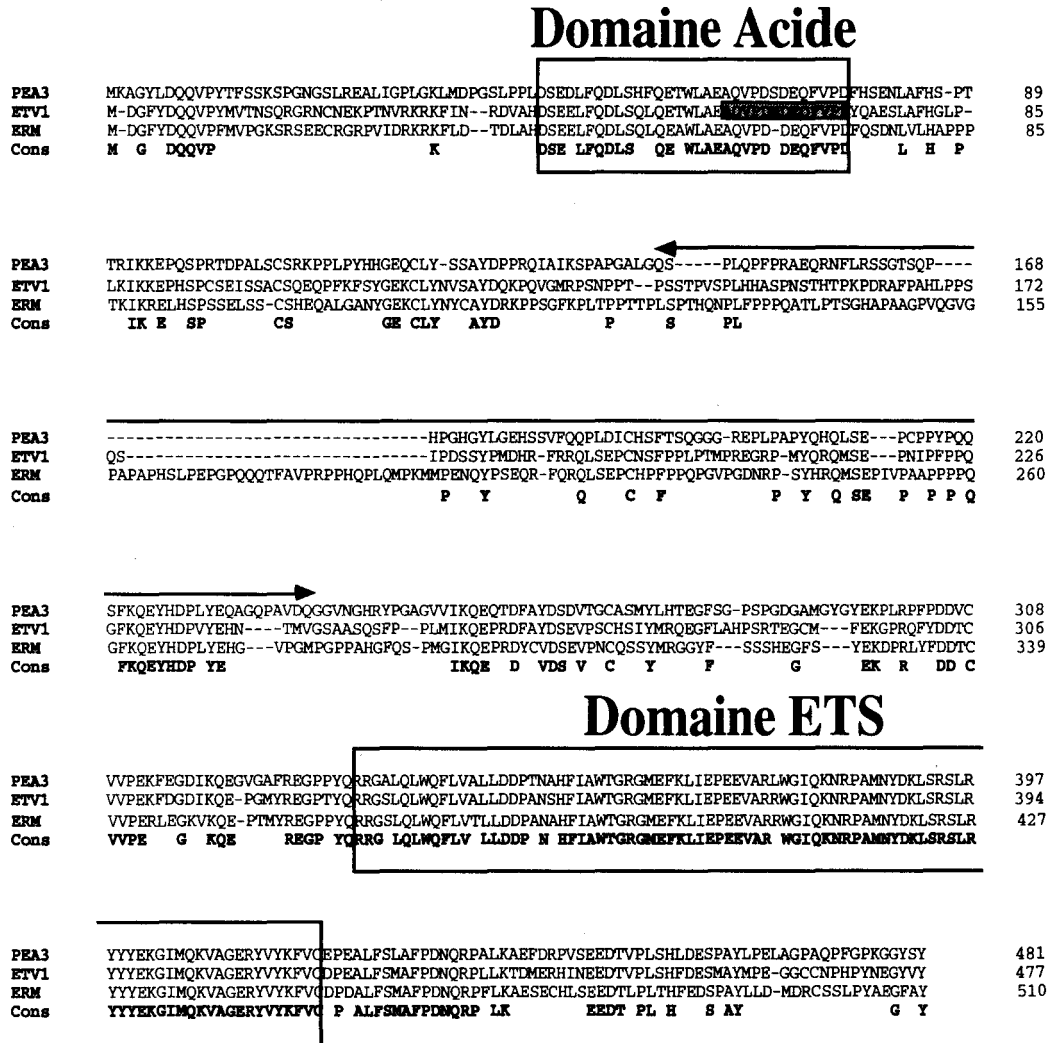


Figure 21 : Alignement des protéines humaines PEA3, ETV1 et ERM. Les domaines acide et ETS sont marqués par des rectangles. La région riche en glutamine est notée par une flèche à tête double et la zone du domaine acide absente dans le clone ER81 isolé dans notre équipe correspond au cadre gris. La protéine PEA3 humaine ne comprend pas 555 résidus comme la protéine PEA3 murine et l'on suspecte que les 79 premiers acides aminés de PEA3 murin soient un artéfact de clonage (J-H Chen, communication personnelle).

Son appartenance à la famille Ets ainsi que la présence de deux régions *a priori* candidates à porter la fonction de transactivation ont incité à une étude plus approfondie. Le domaine ETS de PEA3 a été confirmé dans ses fonctions de domaine de liaison à l'ADN puisque la délétion de ses trois premiers ou de ses 17 derniers acides aminés abolit totalement la fixation spécifique à l'ADN. De plus, PEA3 ne s'est pas dérobé à l'attente générale : c'est un facteur de transcription puisqu'il est capable de transactiver sur un élément de réponse contenant quatre sites PEA3 (Xin *et al.*, 1992).

Comme la plupart des membres de la famille Ets, PEA3 est conservé durant l'évolution : son homologue humain, isolé d'extraits de cellules HeLa et nommé pour l'occasion E1A-F (ou ETV4), présente 94% d'identité de séquence totale avec PEA3 murin (Higashino *et al.*, 1993) (Fig. 21). En revanche, toutes les tentatives de notre laboratoire pour isoler l'équivalent aviaire de PEA3 ont échoué pour l'instant. On peut supposer néanmoins qu'il existe puisque PEA3 est présent chez une espèce moins évoluée, le poisson zèbre (A. Sharrocks, communication personnelle).

Le criblage d'une banque d'embryon de souris de 8,5 jours a permis d'exhumer le second membre du groupe PEA3, ER81 (pour Ets Related 81) (Brown et McKnight, 1992). La protéine encodée, longue de 477 résidus, présente 95% d'identité de séquence dans le domaine ETS avec PEA3 mais seulement 54% sur la totalité de la phase de lecture de PEA3. La région riche en résidus acides (domaine acide) est néanmoins conservée entre les deux protéines (Fig. 21). Ainsi qu'il a été évoqué dans un chapitre précédent, l'équivalent humain de ER81, ETV1, a récemment été reconnu comme partie intégrante d'une chimère comprenant également un peptide de EWS (Jeon *et al.*, 1995). L'analyse de cette protéine de fusion suggère fortement que ETV1 présente des capacités de fixation à l'ADN semblables à celle de PEA3 : la liaison à l'ADN requiert l'intégrité du domaine ETS. Indépendamment de ces travaux, notre groupe a isolé ER81 humain (identique à 97% à ER81 murin) à partir d'une banque de rein humain et montré qu'il différait de ETV1 par l'absence d'une partie du domaine acide (Monté *et al.*, 1995) (Fig. 21). Par analyse de l'organisation

génomique, il a été démontré que cette disparité est le résultat d'un épissage alternatif. Les conséquences de ce processus seront examinées dans la partie DISCUSSION.

Le petit dernier du groupe PEA3, ERM (pour Ets Related Molecule) est né d'une banque foetale humaine (Monté *et al.*, 1994). La séquence en acides aminés déduite de la phase de lecture ouverte correspond à 510 résidus comprenant le domaine ETS ainsi que le domaine acide présent dans PEA3 et ER81. Les homologues murin (Hélène Pelczar, résultats non publiés) et aviaire (Pierre-Antoine Defossez, résultats non publiés) de ERM ont permis, par l'alignement de toutes les séquences des protéines du groupe PEA3, de mettre en évidence qu'outre ces deux domaines d'identité de séquence, la partie carboxy-terminale immédiatement adjacente au domaine ETS était également conservée (Fig. 21).

6.2. Expression des ARN messagers.

La distribution tissulaire des trois membres du groupe PEA3 n'est pas similaire : des expériences de transfert par la technique de Northern ont montré que la présence des ARN messagers de PEA3 est restreinte à l'épididyme et au cerveau ainsi qu'à la glande mammaire dans une moindre mesure (Xin *et al.*, 1992). Au contraire, il n'y a guère que dans le foie qu'on ne détecte pas ER81 (Brown et McKnight, 1992). Enfin ERM peut être considéré comme ubiquiste mais il est exprimé principalement dans le cerveau, le placenta et un peu moins dans le pancréas, le poumon et le coeur (Monté *et al.*, 1994).

6.3. Localisation chromosomique.

En collaboration avec R. Berger (Institut de Génétique Moléculaire-INSERM U301), notre équipe a montré que ERM se situait en position 3q27-29 (résultats non publiés). La localisation chromosomique de PEA3 (Brody *et al.*, 1995;

Isobe *et al.*, 1995; Osborne-Lawrence *et al.*, 1995) et de ER81 (Jeon *et al.*, 1995) est respectivement 17q21 et 7q22. Notons que PEA3 est situé à proximité du gène BRCA1 responsable du cancer héréditaire du sein et de l'ovaire chez l'homme.

6.4. Un rôle pour les membres du groupe PEA3?

Un faisceau d'arguments suggère que les membres du groupe PEA3 pourraient peut-être éventuellement hypothétiquement être impliqués dans la transformation. Premièrement, l'induction de la différenciation des cellules de carcinome embryonnaire murin F9 et P19 est contemporaine de l'augmentation du taux de transcrits PEA3 (Xin *et al.*, 1992). Deuxièmement, dans les tumeurs du sein exprimant l'oncogène *HER2/neu* et où l'augmentation du taux de transcrits *HER2* ne correspond pas à une élévation du nombre de copies de ce gène, les résultats obtenus indiquent de manière indirecte qu'une protéine Ets augmenterait la transcription du gène *HER2/neu* (Scott *et al.*, 1994). Parmi les candidats sérieux, ERM mais aussi $GABP\alpha$, Elk-1, Elf-1 ou n'importe quelle protéine Ets non encore caractérisée...

Des études plus approfondies ont montré que dans les adénocarcinomes mammaires induits chez des souris transgéniques par l'oncogène *neu* et dans les métastases apparaissant ensuite dans le poumon, l'ARN messager de PEA3 est fortement exprimé alors qu'il est virtuellement absent dans le tissu normal entourant la tumeur ou la métastase (Trimble *et al.*, 1993). Cette équipe a obtenu des résultats similaires pour des tumeurs induites par l'antigène moyen T du virus du polyome. De plus, notre groupe a montré que ERM et ER81 se comportaient de la même manière que PEA3 dans ces mêmes tumeurs (résultats non publiés). Cependant aucun mécanisme permettant d'expliquer l'augmentation du taux de transcrits PEA3 n'a pour l'instant été décrit.

En outre, une trentaine de lignées de tumeurs humaines mammaires ont été testées dans notre groupe en collaboration avec R. Sutherland (Garvan Institute de Sidney) et les résultats indiquent clairement que l'ARN messager de ER81 est exprimé dans

les tumeurs négatives pour les récepteurs aux oestrogènes et aux progestagènes et qu'inversement ER81 n'est pas exprimé (ou très faiblement) dans les lignées négatives (résultats non publiés). L'expression des protéines correspond à celle des ARN messagers (Jean-Luc Baert, résultats non publiés). Les données obtenues pour ERM et PEA3 suivent la même tendance bien qu'elle soit moins marquée. Notons de plus que ERM est exprimé non seulement dans des lignées de tumeurs mammaires mais également dans des lignées cancéreuses d'origine diverse.

L'ensemble de ces résultats peut être considéré comme encourageant mais il se limite pour le moment à une corrélation entre l'expression des membres du groupe PEA3 et le phénomène de transformation ou de métastase. Aucun système impliquant de manière directe et non ambiguë le rôle des protéines du groupe PEA3 n'a été décrit.

Une étude plus détaillée de l'expression des membres du groupe PEA3 ainsi que la recherche d'un modèle ont été entreprises dans notre groupe par Anne Lelièvre-Chotteau et Hélène Pelczar. Une analyse plus fine par la méthode d'hybridation *in situ* rendant possible la localisation cellulaire des ARN messagers, a permis d'avancer l'hypothèse d'une spécificité de fonction pour ERM, ER81 et PEA3. En effet, il apparaît clairement qu'au cours de l'organogenèse, par exemple dans le développement de la dent, de la glande salivaire, du poumon et de la glande mammaire de souris, l'expression de ces trois gènes n'est ni systématiquement redondante, ni toujours exclusive (résultats non publiés).

6.5. Des gènes cibles pour les protéines du groupe PEA3?

Le versant moléculaire de l'implication de PEA3 dans la tumorigenèse a aussi été exploré. Entre autres, il faut signaler que PEA3, de concert avec Oct-1 et Oct-2, est capable d'augmenter *in vitro* la transcription de c-Ets-1 (Chen et Wright, 1993), un exemple séduisant quand on songe que c-Ets-1, à basse concentration en sérum, cause la transformation des cellules NIH3T3 quand il est surexprimé (Seth *et*

al., 1990). Une modification de l'expression de c-Ets-1 pourrait dans ce cas être causée par la dérégulation de PEA3.

D'autre part, il a récemment été montré que l'expression de PEA3 corrélait celle de la vimentine, qui intervient dans la formation des filaments eucaryotes intermédiaires de classe III : dans les lignées de tumeurs mammaires humaines et murines, PEA3 et la vimentine sont exprimées alors qu'ils sont absents dans les lignées épithéliales mammaires normales (J-H Chen, communication personnelle). De plus, le promoteur de la vimentine est régulé par PEA3. D'autres membres de la famille Ets ont été testés en transfection pour leur aptitude à transactiver à partir de ce même promoteur mais ils semblent exercer un effet beaucoup moins important que PEA3. Signalons néanmoins que la quantification des différentes protéines Ets transfectées n'a pas été vérifiée, ce qui introduit peut-être un biais considérable. Cependant, si l'intervention spécifique de PEA3 dans la transcription du gène de la vimentine était confirmée, cela placerait PEA3 dans les facteurs pronostic du cancer du sein, au même titre que la vimentine.

A également été suggérée l'implication de PEA3 dans l'activation du promoteur des métalloprotéinases, dont la surexpression est associée au processus de la métastase : la transcription de la stromélysine et des collagénases de type I et IV seraient en particulier sous le contrôle de PEA3 (Higashino *et al.*, 1995). Cependant, il convient d'insister sur le fait que PEA3 n'a pas été déclaré comme candidat unique : aucun autre membre de la famille Ets n'a été testé en parallèle. Au regard du consensus établi pour leur site de reconnaissance nucléotidique, il est envisageable que la plupart des protéines Ets auraient eu le même effet.

Avec les mêmes précautions, on signalera que PEA3, ERM et ER81 activent la transcription des molécules d'adhérence ICAM-1 et intégrine $\alpha 4$ (Yvan de Launoit, résultats non publiés). Ces données préliminaires sont à mettre en relation avec celles obtenues sur l'activation par c-Ets-1 des gènes des protéases intervenant dans la dégradation de la matrice extracellulaire. Il est tentant d'imaginer (mais cela reste une supposition pour le moment) que divers membres de la famille Ets aient des

rôles complémentaires au niveau de la formation de nouveaux tissus, tant au niveau de la dégradation de la matrice extracellulaire (c-Ets-1) qu'à celui de la "solidification" de la structure nouvellement formée (c-Ets-1 et les protéines du groupe PEA3).

D'une manière générale, il sied de rappeler que ces expériences sont la plupart du temps effectuée *in vitro*, ce qui sous-entend que les promoteurs examinés sont sortis de leur contexte naturel; expression tissu-spécifique de chaque Ets, cofacteurs propres à chacun et régulation post-traductionnelle sont autant de critères supplémentaires dont il conviendrait de tenir compte.

MATERIEL ET METHODES

1. Prédiction de structure.

1.1. Etablissement d'un diagramme HCA.

A l'évidence, toute l'information nécessaire au repliement d'une protéine et à l'élaboration de sa structure tridimensionnelle est comprise dans sa séquence en acides aminés. Les méthodes de prédiction de structure, qui visent au décriptage de cette information pullulent et lorsque j'ai commencé ma thèse nous avons décidé d'appliquer la plus innovatrice d'entre elles à l'étude du domaine ETS. La méthode HCA, pour Hydrophobic Cluster Analysis, ce qui pourrait approximativement être traduit par méthode des amas hydrophobes, est un outil sensible qui permet de détecter des identités de structure secondaire et tertiaire entre des protéines révélant une identité de séquence très faible. Travailler avec deux protéines ne présentant que 10%-20% d'identité de séquence reste possible.

En vue d'une telle analyse, la séquence protéique est inscrite sur une hélice α classique (3,6 acides aminés par tour d'hélice) puis aplatie dans les limites d'un cylindre (Lemesle-Varloot *et al.*, 1990) (Fig. 22). La figure ainsi obtenue sera dupliquée de sorte que chaque acide aminé retrouve son plus proche voisin éventuellement "perdu" au cours de l'aplatissement du cylindre (par exemple, la phénylalanine 253 "perd" son voisin la leucine 254). Tous les acides aminés hydrophobes sont alors entourés et constituent donc les amas hydrophobes. La méthode HCA s'appuie sur les deux constatations fondamentales que l'établissement des interactions hydrophobes constitue l'événement essentiel du repliement des protéines et, par conséquent, que les résidus hydrophobes ne sont pas répartis au hasard au sein de la protéine mais sont au contraire représentatifs des faces internes des structures secondaires (Woodcock *et al.*, 1992). C'est la comparaison de la forme, de l'orientation et de la taille des amas hydrophobes qui permettra alors de détecter des identités de structure entre deux protéines. Retenons que dans le cas présent, le domaine ETS a été comparé à des protéines auparavant sélectionnées pour présenter un minimum d'identité de séquence avec le domaine ETS.

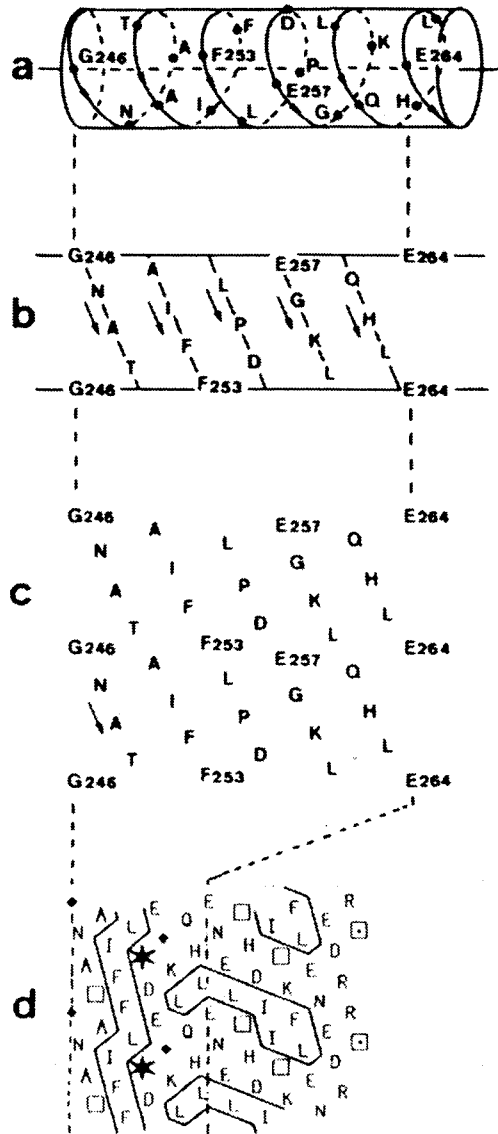


Figure 22 : Manuel du bon petit styliste HCA.

La séquence est enroulée sur une hélice α classique (A), puis aplatie en cylindre (B) et dupliquée (C). Les amas hydrophobes sont entourés (D). La proline est symbolisée par une étoile, la glycine par un losange, la sérine par un carré contenant un rond et la thréonine par un carré vide.

1.2. Validation statistique.

Le score statistique Z est établi en vue de déterminer si l'alignement HCA réalisé et les identités de structure relevées sont significatives ou reflètent une composition en acides aminés biaisée. Le score Z est calculé avec le programme RDF2 (Lipman et Pearson, 1985; Pearson, 1990) et il représente la différence entre le score de l'alignement considéré et la moyenne des scores établis entre la séquence 1 et la séquence 2 "randomisée" (au sein de laquelle les acides aminés sont distribués aléatoirement). Le score Z, calculé en SD (déviation standard), est considéré comme "peut-être significatif" au delà de 3 SD et comme significatif au delà de 6 SD.

2. Construction des vecteurs nécessaires.

2.1. Construction des vecteurs d'expression contenant tout ou partie de ERM.

L'ADN complémentaire correspondant aux versions tronquées en amino-terminal de ERM est produit par réaction de polymérisation en chaîne (PCR) à l'aide de, à l'extrémité 5', une amorce recouvrant une méthionine interne déjà présente dans la molécule ou introduite artificiellement et, à l'extrémité 3' un l'oligonucléotide suivant 5'-ACCAAACCACTGCCCTTGTTTGC-3'. Les constructions B, C et D ont été élaborées en partant des méthionines Met 185 (amorce 5'-CAGCCCCTGGCCATGCCAAAGATGA-3'), Met 272 (amorce 5'-TCACCAATGGGAATCAAGCAGGAGC-3') et Met 355 (amorce 5'-TACCATGTATCGAGAGGGGCCCCCT-3'), respectivement (Fig. 23). A correspond au produit de traduction de l'ADN complémentaire entier de ERM.

En utilisant le même oligonucléotide à l'extrémité 5' et un oligonucléotide situé à la fin du domaine ETS (5'-ATCTGGTTATCAGCAGACAAATTTGTAG-3'), nous avons obtenu les constructions ΔA , ΔB , ΔC and ΔD qui sont dépourvues des 61 derniers acides aminés de la protéine de pleine longueur (queue carboxy-terminale). ΔD correspond au domaine ETS. $\Delta D+$, qui recouvre la totalité du domaine ETS ainsi que

les vingt et un premiers acides aminés de la queue carboxy-terminale a aussi été produit par PCR, en utilisant l'armorce à l'extrémité 5' identique à celle du domaine ETS et à l'extrémité 3' l'amorce 5'-TAGGATCCTTACTTCAGGAACGGACGCTGG-3'.

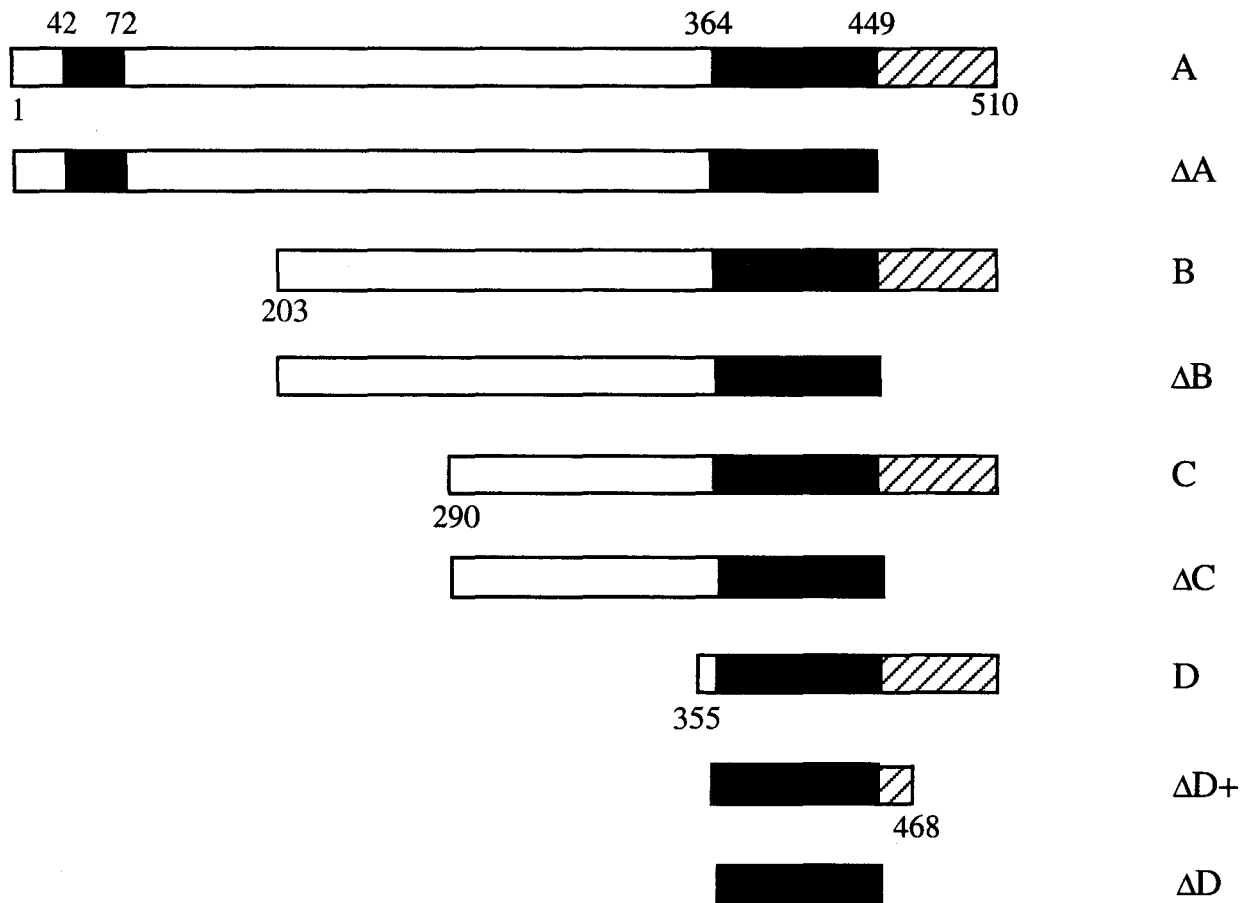


Figure 23 : Schéma récapitulatif des différentes constructions délétées en amino- et en carboxy-terminal de ERM. Les domaines ETS et acide et la queue carboxy-terminale sont respectivement en gris, noir et hachuré.

Tous les produits de PCR ont été clonés dans le vecteur PCRII (Kit TA cloning, Invitrogen) et séquencés dans leur totalité avant d'être reclésés dans le vecteur pSG5 (Green *et al.*, 1988) par l'intermédiaire des sites de restriction *EcoRI* du vecteur pSG5.

2.2. Construction nécessaire à l'étude du pouvoir transactivateur de ERM.

L'ADN complémentaire entier de ERM a aussi été introduit dans un vecteur dérivé de pSG5, le vecteur pSV (don de Guillaume Adelmant) en utilisant ses sites de restriction *EcoRI* et *XbaI*.

2.3. Construction des chimères GAL4 ERM.

Toutes les chimères GAL4 α , GAL4 β , GAL4 γ , GAL4 δ , GAL4 ϵ , GAL4 Db, GAL4 Db*, GAL4 Ct, GAL4 aD, GAL4 Ct 449-468, GAL4 Ct 449-489, GAL4 Ct 469-510, GAL4 $\alpha\beta\gamma$, GAL4 $\gamma\delta$ and GAL4 $\beta\gamma\delta\epsilon$ ont été construites en clonant divers produits de PCR (voir Table 1 pour les amorces correspondantes), en aval du domaine de fixation à l'ADN de la protéine GAL4 (acides aminés 1 à 147) et en phase avec ce domaine, en utilisant le vecteur pSG424 (Sadowski *et al.*, 1989) (don du Dr Achim Leutz) (Fig. 24). Les produits de PCR ont d'abord été clonés dans le vecteur PCRII, totalement séquencés puis sous-clonés dans le vecteur pSG424 en utilisant ses sites *Asp718* et *XbaI*.

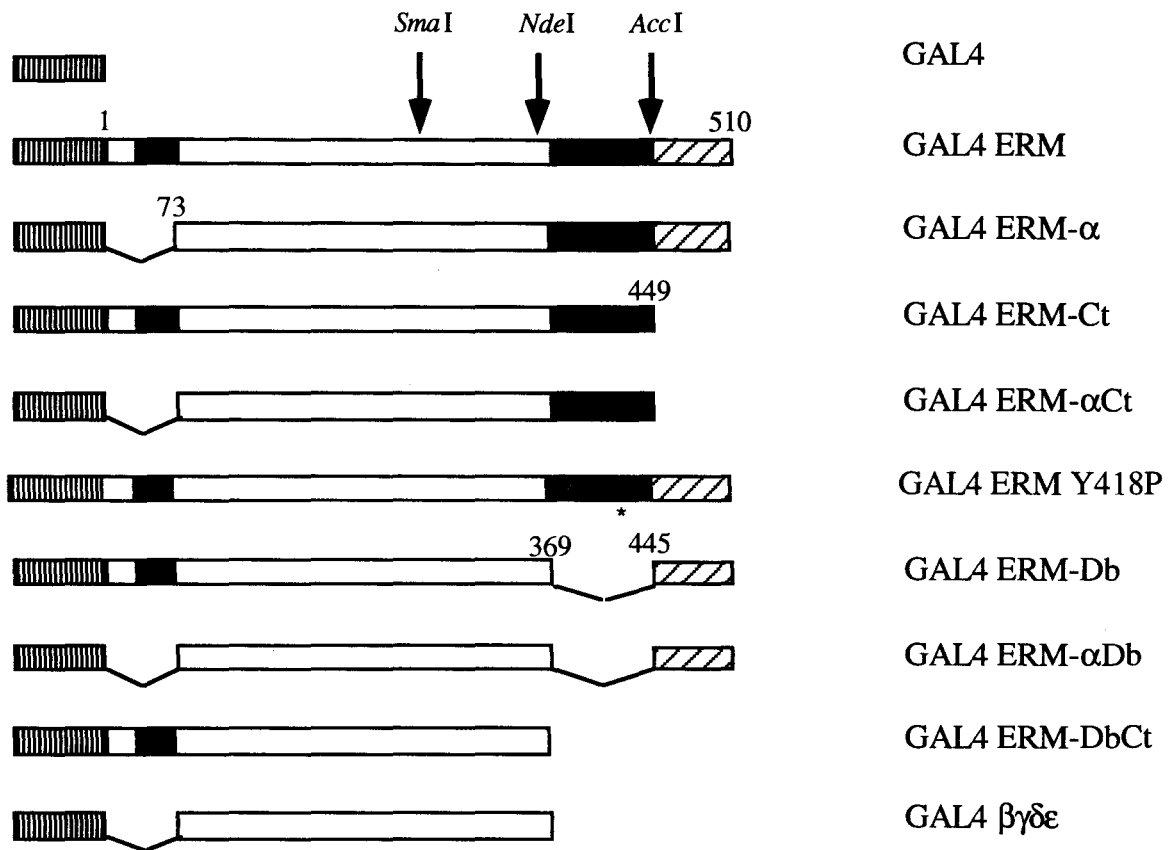


Figure 24 : Représentation schématique des différentes constructions GAL4 ERM délétées de un ou plusieurs domaines. Les sites enzymatiques utilisés (exceptés ceux de pSG424) sont indiqués par une flèche. Les domaines acide, ETS et la queue carboxy-terminale sont respectivement en noir, gris et hachuré.

<u>construction</u>	<u>amorce à l'extrémité 5' (5'->3')</u>	<u>amorce à l'extrémité 3' (5'->3')</u>
GAL4 α	TGGTACCCCATGGACGGGTTTTATGAT	TTCTAGATTAATCTGGGACAAACTGTTCA
GAL4 β	TGGTACCATTTTCAGTCTGATAACCTGGT	TTCTAGATTATGAGAGGGGGGTTGTAGGA
GAL4 γ	TGGTACCCACCCACCCATCAGAATCCCCT	TTCTAGATTAATGTGGTGGTCGGGGGACC
GAL4 δ	TGGTACCATCAGCCCCTGCAGATGCCAAA	TTCTAGATTATGCTGGGGGCCCTGGCATG
GAL4 ϵ	TGGTACCCACACGGGTTCCAGTCACCAAT	TTCTAGATTACTGGTAAGGGGGCCCCTCT
GAL4 Db	TGGTACCAGAGGCGAGGTTCCCTTCAGCT	TTCTAGATCAGACAAATTTGTAGACGTAT
GAL4 Db*	TGGTACCAGAGGCGAGGTTCCCTTCAGCT	TTCTAGATCAGACAAATTTGTAGACGTAT
GAL4 Ct	TGGTACCTCTGTGACCCAGATGCCCTCT	TTCTAGATTAGTAAGCAAAGCCTTCGGCA
GAL4 aD	TGGTACCACGATTCTGAAGAGCTATTTTCAG	TTCTAGATTAATCTGGGACAAACTGTTCA
GAL4 Ct 449-468	TGGTACCTCTGTGACCCAGATGCCCTCT	TTCTAGATTACTTCAGGAACGGACGCTGG
GAL4 Ct 469-510	TGGTACCACGCAGAGTCCGAGTGCCACCT	TTCTAGATTAGTAAGCAAAGCCTTCGGCA
GAL4 Ct 449-489	TGGTACCTCTGTGACCCAGATGCCCTCT	TTCTAGATTAGCTGTCTTCAAAGTGGGT
GAL4 $\alpha\beta\gamma$	TGGTACCCCATGGACGGGTTTTATGAT	TTCTAGATTAATGTGGTGGTCGGGGGACC
GAL4 $\gamma\delta$	TGGTACCCACCCACCCATCAGAATCCCCT	TTCTAGATTATGCTGGGGGCCCTGGCATG
GAL4 $\beta\gamma\delta\epsilon$	TGGTACCATTTTCAGTCTGATAACCTGGT	TTCTAGATTACTGGTAAGGGGGCCCCTCT
GAL4 ERM	TGGTACCCCATGGACGGGTTTTATGAT	ATTCCCATTTGGTGACTGGAAC
GAL4 ERM- α	TGGTACCATTTTCAGTCTGATAACCTGGT	ATTCCCATTTGGTGACTGGAAC
NdeI-AccI fragment	ATCATATGAAAAAGATCCCCCGA	TAGGATCCTTATTAATAATTTGTAGACGTATCG

Table 1 : Liste des oligonucléotides utilisés comme amorce pour les PCR.

Pour la chimère GAL4 ERM, afin de pouvoir cloner ERM en phase avec le domaine de fixation à l'ADN de GAL4, nous avons préalablement amplifié par PCR et séquencé un fragment correspondant aux 1175 premiers nucléotides codants de ERM et nous l'avons cloné dans le plasmide PCR^{II} ERM qui contient ERM total en utilisant les sites de restriction *Asp*718 et *Sma*I. La construction obtenue a ensuite été clonée dans le vecteur pSG424. La construction GAL4 ERM- α pour laquelle les 72 premiers acides aminés sont supprimés, a été obtenue en utilisant la même stratégie.

Afin d'obtenir les constructions GAL4 ERM-Ct (pour laquelle les 61 derniers résidus ont été supprimés) et GAL4 ERM- α Ct (dépourvue des 72 premiers et des 61 derniers acides aminés de ERM), nous avons digéré par *Acc*I (dans ERM) et *Xba*I (dans le pSG424) les constructions GAL4 ERM et GAL4 ERM- α , rendu franches et ligaturé les extrémités des produits de digestion.

La même technique employée sur GAL4 ERM, GAL4 ERM- α et GAL4 ERM-Ct avec les enzymes de restriction *Nde*I et *Acc*I a permis d'obtenir les constructions GAL4 ERM-Db, GAL4 ERM- α Db et GAL4 ERM-DbCt, pour lesquelles le domaine ETS a été supprimé.

Un fragment *Nde*I-*Acc*I comprenant le domaine ETS muté de manière ponctuelle (la tyrosine 418 est changée en proline) est produit par la méthode de double PCR décrite ci-dessous en utilisant une amorce interne 5'-GCCATGAACCCTGACAAGCTG-3' et son antisens complémentaire. Le fragment obtenu a été ensuite inséré dans la construction GAL4 ERM, donnant ainsi la chimère GAL4 ERM Y418P.

2.4. Construction du domaine ETS de ERM sauvage ou muté.

Le fragment codant les 82 acides aminés du domaine ETS de ERM a été amplifié par PCR en utilisant les amorces 5'-ATGAATTCAATGGGTTCCCTTCAGCTGTGG-3' à l'extrémité 5' et 5'-TAGGATCCTTATTA AAATTTGTAGACGTATCG-3' à l'extrémité 3'. Le produit de

PCR a été cloné dans le vecteur PCRII et utilisé comme matrice pour l'obtention des domaines ETS mutés sur la tyrosine 418 ou 442.

Les mutants sur la tyrosine 418 ont été produits par la méthode de double PCR décrite Figure 25: l'amorce 5'-GCCATGAACNNNGACAAGCTG-3' où NNN correspond à GCT, CCT, TTC and TGG et l'antisens complémentaire correspondant ont donné naissance à une mutation spécifique de la tyrosine 418 en alanine, proline, phénylalanine ou tryptophane.

Les mutants sur la tyrosine 442 ont été synthétisés en utilisant l'amorce à l'extrémité 5' identique à celle utilisée pour la construction du domaine ETS seul et une amorce à l'extrémité 3' qui porte la mutation : 5'-TAGGATCCTTATTAAAATTTGTAGACNNNTCG-3', NNN correspondant à AGC, GGG, GAA et CCA, ce qui permet la mutation spécifique de la tyrosine 442 en alanine, proline, phénylalanine ou tryptophane.

La construction portant la double mutation en tryptophane des tyrosines 418 et 442 a été produite par la méthode de double PCR en employant le mutant Y418W (tyrosine 418 en tryptophane) en guise de matrice, l'amorce 5' employée pour le domaine ETS et l'amorce 5'-TAGGATCCTTATTAAAATTTGTAGACCCATCG-3' à l'extrémité 3'.

Toutes ces constructions, obtenues par PCR ont été totalement séquencées et sous-clonées dans le vecteur pSG5 à l'aide des sites de restriction *EcoRI* et *BamHI*.

La production des protéines encodées par toutes les constructions chimères a été vérifiée par immunoprécipitation à l'aide d'un anticorps dirigé contre le domaine de fixation à l'ADN de GAL4 (don du Dr Yves Lutz).

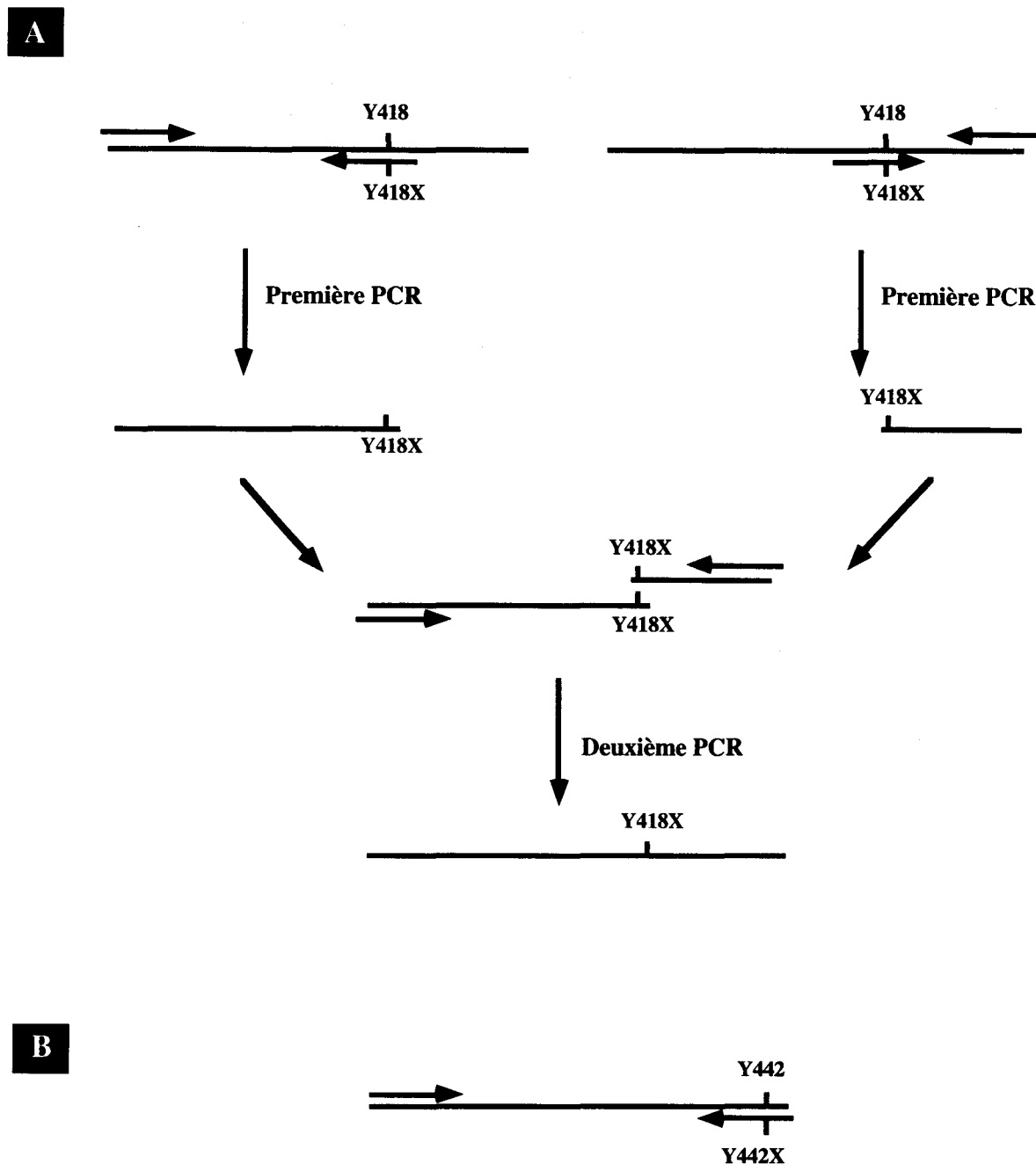


Figure 25 : Méthode de mutagenèse dirigée employée pour l'obtention de domaines ETS mutés. X représente l'alanine, la phénylalanine, le tryptophane ou la proline.

(A) Double PCR utilisée pour les mutants sur la tyrosine 418.

L'oligonucléotide recouvrant la tyrosine 418 porte la mutation désirée et sert d'amorce pour les deux premières PCR effectuées en parallèle. Les produits résultant des PCR contiendront la mutation désirée et serviront de matrice pour la seconde PCR. Le résultat final est un produit recouvrant le domaine ETS et portant la mutation désirée.

(B) Simple PCR utilisée pour l'obtention des mutants sur la tyrosine 442 et pour le double mutant.

La mutation est assez proche de l'extrémité du fragment pour être comprise dans l'oligonucléotide de l'extrémité 5'.

2.5. Construction des mutants rêve.

Le domaine de liaison à l'ADN de la protéine c-Myb humaine tel qu'il est défini dans Frampton *et al.* (1991) a été amplifié par PCR avec l'amorce 5'-ATGAATTCAATGCTCATCAAGGGTCCTT-3' à l'extrémité 5' et l'amorce 5'-TAGGATCCTTATTACTTCCGACGCATTGTAGA-3' à l'extrémité 3' puis inséré dans le vecteur pCRII (construction D Myb).

Le domaine ETS de ERM (construction Δ D) a été utilisé comme matrice pour l'obtention du mutant rêve ERM qui a été construit suivant la technique de double PCR mentionnée ci-dessus, l'oligonucléotide 5'-CGGCCAGCCATGCGCTGGGCCGAGATAGCCCGCTCTCTCCGC-3' et son antisens complémentaire. Nous avons ainsi introduit la mutation de cinq acides aminés du domaine ETS en les cinq acides aminés du domaine Myb occupant des positions équivalentes. Ce fragment a ensuite été cloné dans le vecteur pCRII.

3. Traduction *in vitro* et analyse des protéines produites.

Toutes les constructions en pSG5, puisqu'elles possèdent un promoteur T7 qui permet la transcription (puis la traduction) dans un système eucaryote, sont susceptibles de produire des protéines dans le système *in vitro* de lysat de réticulocytes de lapin. Pour les constructions rêve, selon qu'elles ont été clonées en sens ou en antisens dans le vecteur pCRII, on utilisera le promoteur SP6 ou T7 pour la traduction. Les conditions conseillées par le fournisseur (Promega) sont les suivantes:

- 1 μ g de plasmide pSG5
- 10 unités d'ARN polymérase T7
- 25 μ l de lysat de réticulocytes de lapin
- 40 μ Ci de méthionine marquée au 35 S

pendant 90 min à 30°C dans le tampon approprié fourni par Promega.

Les produits de traduction, qui ont incorporé la méthionine marquée peuvent être visualisés sur un gel de polyacrylamide en gradient 10%-30% en présence de SDS (Sambrook *et al.*, 1989). L'utilisation de l'analyseur d'images Phosphorimager (Molecular Dynamics) et du logiciel ImageQuant (Molecular Dynamics) a permis de quantifier les protéines marquées à la méthionine ³⁵S et d'en employer une égale quantité pour les expériences de gel-retard.

Les protéines de faible poids moléculaire (les domaines ETS de ERM sauvage et mutants comprennent 82 acides aminés) ont été visualisés sur un gel tricine en présence de SDS, conditions qui permettent une bonne résolution pour des poids moléculaires inférieurs à 20 kDa (Schägger *et al.*, 1987).

4. Gel-retard.

La technique du gel-retard rend possible la visualisation d'un complexe ADN-protéine. Dans notre cas, l'ADN est un oligonucléotide marqué au ³²P et la protéine est le produit de traduction du lysat de réticulocytes de lapin. Lorsque la protéine considérée possède une propriété de fixation à l'oligonucléotide, le complexe ADN-protéine formé après incubation, plus lourd, sera séparé de l'oligonucléotide libre par migration sur un gel de polyacrylamide.

L'oligonucléotide utilisé ici correspond à une partie de la séquence de l'amplificateur du virus du polyome (Martin *et al.*, 1988) sur lequel se fixent des protéines Ets (Wasylyk *et al.*, 1990) et dont la séquence est 5'-GATCTTCGAAAGGAAGTTTCGAG-3'.

L'oligonucléotide double brin est marqué au ³²P dans les conditions suivantes :

2 µl de Tris/HCl 1 M pH 7.6

4 µl de MgCl₂ 50 mM

2 µl de β-mercapto-éthanol 0.2 M

5 µl de γ³²P ATP à 3000 Ci/ml

4 unités de polynucléotide kinase T4
50 ng d'oligonucléotide double brin
dans un volume final de 20 μ l.

Après incubation à 37°C pendant 60 min, on arrête la réaction par l'ajout de 80 μ l d'EDTA 20 mM et la sonde est purifiée sur une colonne de Sephadex G50 (Pharmacia). L'incubation de la protéine avec l'oligonucléotide marqué, qui permettra la fixation de la protéine sur sa séquence cible se réalise dans les conditions suivantes :

20 mM Hepes pH 7.9
20% glycerol
0.1 mM EDTA
1 mM DTT
1 μ g/20 μ l polyIdC qui diminue les interactions non spécifiques
50 mM NaCl
en présence d'un excès d'oligonucléotide marqué
pour un volume final de 20 μ l

Après incubation 60 min dans la glace, les échantillons sont déposés sur un gel non dénaturant en gradient de polyacrylamide 5%-20%. La migration s'effectue sur la nuit à 150 V à 4°C.

5. Transfection transitoire.

5.1. Transfection des cellules.

Les cellules de carcinome du col de l'utérus humain (HeLa) et les cellules de carcinome de rein de lapin (RK13) sont entretenues en milieu Dulbecco à 10% de sérum de veau foetal à 37°C en atmosphère saturée en eau à 5%CO₂. Trois à 5.10⁵ cellules réparties dans des puits de 35 mm de diamètre sont transfectées dans les conditions suivantes :

10 μ l de lipofectamine (Gibco-BRL)

0.8 µg de plasmide d'expression (vecteur pSV ou pSG424)

0.1 µg de vecteur normalisateur (vecteur pSV2luc)

une quantité variable de vecteur rapporteur.

Le vecteur rapporteur contient plusieurs copies de la séquence reconnue par les protéines produits de l'expression des vecteurs pSV ou pSG424, copies placées en amont du gène de la Chloramphénicol Acétyl Transférase (CAT).

Dans le cas des constructions chimères avec le domaine de fixation à l'ADN de GAL4, le vecteur rapporteur, utilisé à 0.3 µg par puits, est 5xGAL4-tk-CAT.

Dans le cas de la construction en pSV (protéine ERM de pleine longueur), il s'agit de vecteur 3xTORU-tk-CAT (utilisé à 0.1 µg par puits). TORU (pour TPA Oncogene Responsive Unit) est une séquence dérivée de l'amplificateur du virus du polyome 5'-GATCTCGAGGAAAGTGACTAACTTGAGCA-3' (Wasylyk *et al.*, 1990).

Lorsque la protéine reconnaît sa séquence cible sur le vecteur 5xGAL4-tk-CAT ou 3xTORU-tk-CAT, elle active la transcription du gène CAT si elle possède un pouvoir transactivateur. De sorte que la mesure de l'activité CAT est alors l'exact reflet du pouvoir transactivateur de la construction utilisée.

Afin de prendre en compte les variations dues à des différences dans l'efficacité de transfection (nombre de cellules, qualité et transfectabilité des plasmides utilisés) les activités CAT sont corrigées (normalisées) à l'aide de l'activité luciférase du vecteur pSV2luc cotransfecté. Les cellules sont récupérées 24 heures après transfection et lysées dans le tampon Promega qui rend possible la mesure des activités luciférase et CAT sur le même extrait.

5.2. Mesure de l'activité luciférase.

Une fois les cellules récupérées dans 500 µl de tampon de lyse Promega et centrifugées, 85 µl du surnageant sont utilisés pour doser l'activité luciférase en

présence de 100 µl d'un tampon contenant les deux substrats nécessaires à la réaction (0.25 mM luciférine et 2 mM ATP).

5.3. Mesure de l'activité CAT.

En fonction de l'efficacité de transfection (c'est-à-dire de l'activité luciférase), une quantité variable d'extrait (de 50 à 200 µl) est incubée de 60 min à deux heures à 37°C en présence de :

20 µl d'acétyl-Coenzyme A à 4 mg/ml

0.5 µl de chloramphénicol ¹⁴C.

L'ajout de 500 µl d'acétate d'éthyle permet l'extraction du chloramphénicol acétylé ou non acétylé qui est soluble en phase organique. Celle-ci est évaporée et l'extrait sec est repris dans 10 µl d'acétate d'éthyle avant d'être déposé sur une plaque de silice. La migration dans un tampon 95% CHCl₃-5% méthanol permet de séparer les formes acétylée et non acétylée. Leur rapport est quantifié à l'aide de l'analyseur d'images Phosphorimager (Molecular Dynamics).

RESULTATS

1. Prédiction de la structure du domaine ETS par la méthode HCA.

Le domaine ETS ne présente aucune identité de séquence notable avec les domaines de liaison à l'ADN dont la structure et le mode d'association au site de reconnaissance spécifique est aujourd'hui connu. En utilisant une technique sensible, la méthode HCA, notre but a été de détecter une identité de séquence avec d'autres protéines, de déterminer les caractéristiques structurales du domaine ETS et d'en fournir une prédiction de structure. Par conséquent, nous nous sommes attachés, en compagnie d'Isabelle Callebaut et sous la direction de Jean-Paul Mornon (Jussieu), à la comparaison des séquences et des structures (quand elles étaient déterminées) des domaines de liaison à l'ADN des protéines Ets, Myb et HMG.

Les particularités des domaines de fixation des facteurs de transcription Ets et Myb ont déjà été décrites dans la partie INTRODUCTION mais il convient de s'attarder sur le cas des protéines HMG. Elles ont en commun un domaine d'environ 80 acides aminés appelé la boîte HMG. Le pourcentage moyen d'identité de séquence pour la boîte HMG entre les membres de cette famille est de 25%, ce qui est faible comparé à la famille Ets : au sein de la famille Ets, l'identité de séquence minimale est de 37% entre les domaines ETS de c-Ets -1 et de SPI-1 (le membre le plus éloigné) mais elle peut atteindre 96% (cas de c-Ets1 et c-Ets-2, par exemple). Certains membres peuvent de plus posséder plusieurs boîtes HMG tels HMG-1, HMG-2 et UBF, ce qui confirme que la famille HMG est moins homogène que la famille Ets. La diversité de la famille HMG est également perceptible au niveau de la fonction de la boîte HMG. Celle de LEF-1 semble nécessaire et suffisante à la liaison à l'ADN (Giese *et al.*, 1992) mais les associations ADN-protéine HMG ne correspondent pas toujours à une interaction spécifique (Parisi *et al.*, 1991; Diffley et Stillman, 1991; Ferrari *et al.*, 1992), ce qui remet en cause l'appartenance des HMG à la classe des facteurs de transcription. Pour appuyer cette assertion, les exemples de coopération entre des membres des familles Ets et HMG évoqués dans l'introduction montrent de plus que les protéines HMG

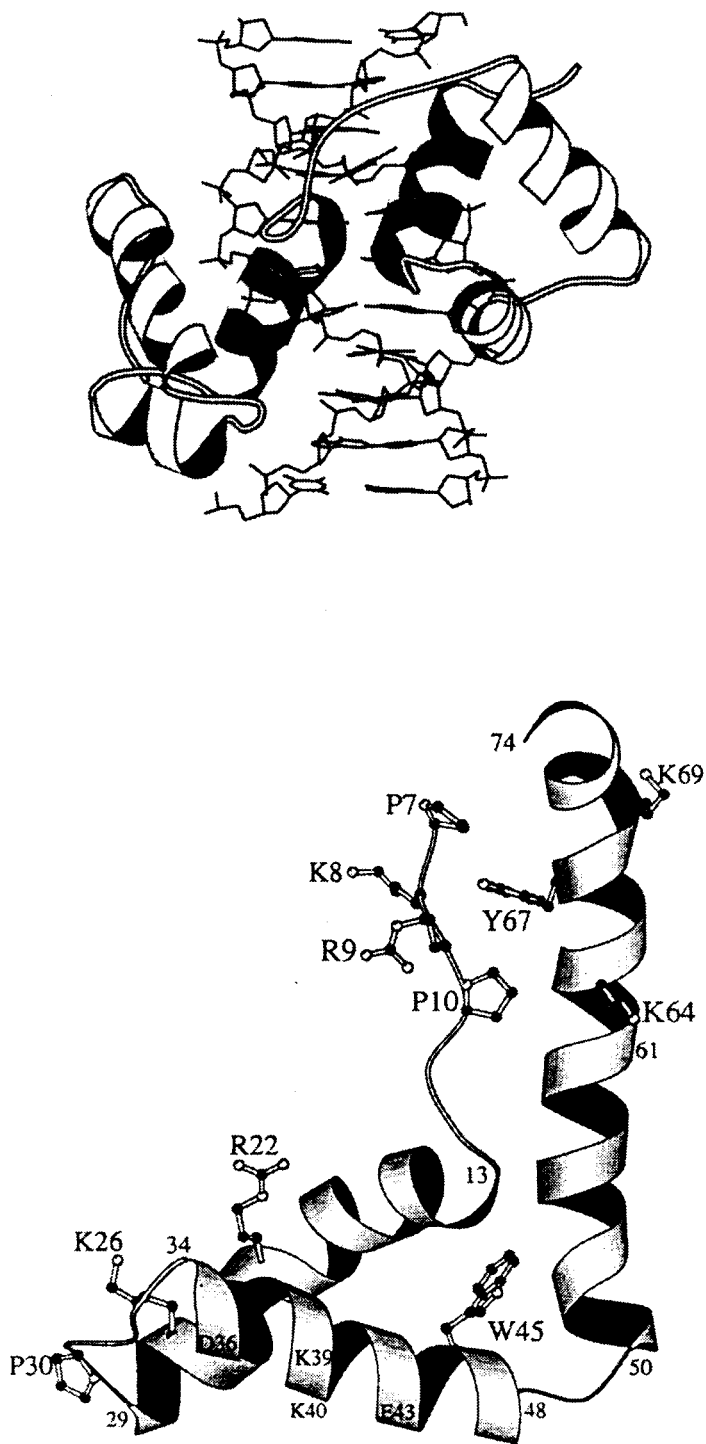


Figure 26 : Organisation tridimensionnelle du domaine MYB (R2R3) selon Ogata *et al.* (1994) (en haut, les hélices de reconnaissance de l'ADN sont en noir) et de HMG-1B selon Weir *et al.* (1993).

semblent dévolues plus au rôle de "colle moléculaire" qu'à celui de facteur de transcription.

1.1. Comparaison des motifs structuraux Myb et HMG.

De faibles identités de séquences ont été décrites entre le domaine ETS et la boîte HMG d'une part, entre le domaine ETS et le domaine de liaison à l'ADN des protéines Myb (appelé domaine MYB dans la suite de cet exposé) d'autre part. A l'époque où ce travail a été initié, on connaissait la structure tridimensionnelle de la boîte B de HMG-1 de rat (HMG-1B) (Weir *et al.*, 1993) ainsi que celle d'une partie du domaine MYB de souris (domaine MYB-R3) (Ogata *et al.*, 1992). Nous avons en premier lieu comparé les structures tri-dimensionnelles et les diagrammes HCA de ces deux domaines de liaison à l'ADN.

1.1.1 Comparaison des structures tridimensionnelles.

HMG-1B et MYB-R3 sont tous deux composés de trois hélices successives (Fig. 26). Si l'on superpose leurs structures tridimensionnelles, on constate que les hélices occupant des positions équivalentes n'ont pas une longueur égale. La variation dans la longueur des hélices est un phénomène courant, il peut même concerner des protéines appartenant à une famille identique, comme c'est le cas pour les protéines régulatrices de phage ou de bactérie (Pabo et Sauer, 1992). La superposition des structures spatiales de domaines très éloignés mais ayant en commun un même motif structural a également permis d'observer une variation dans la longueur des hélices (MYB-R3, 434 Cro et Engrailed) (Ogata *et al.*, 1992).

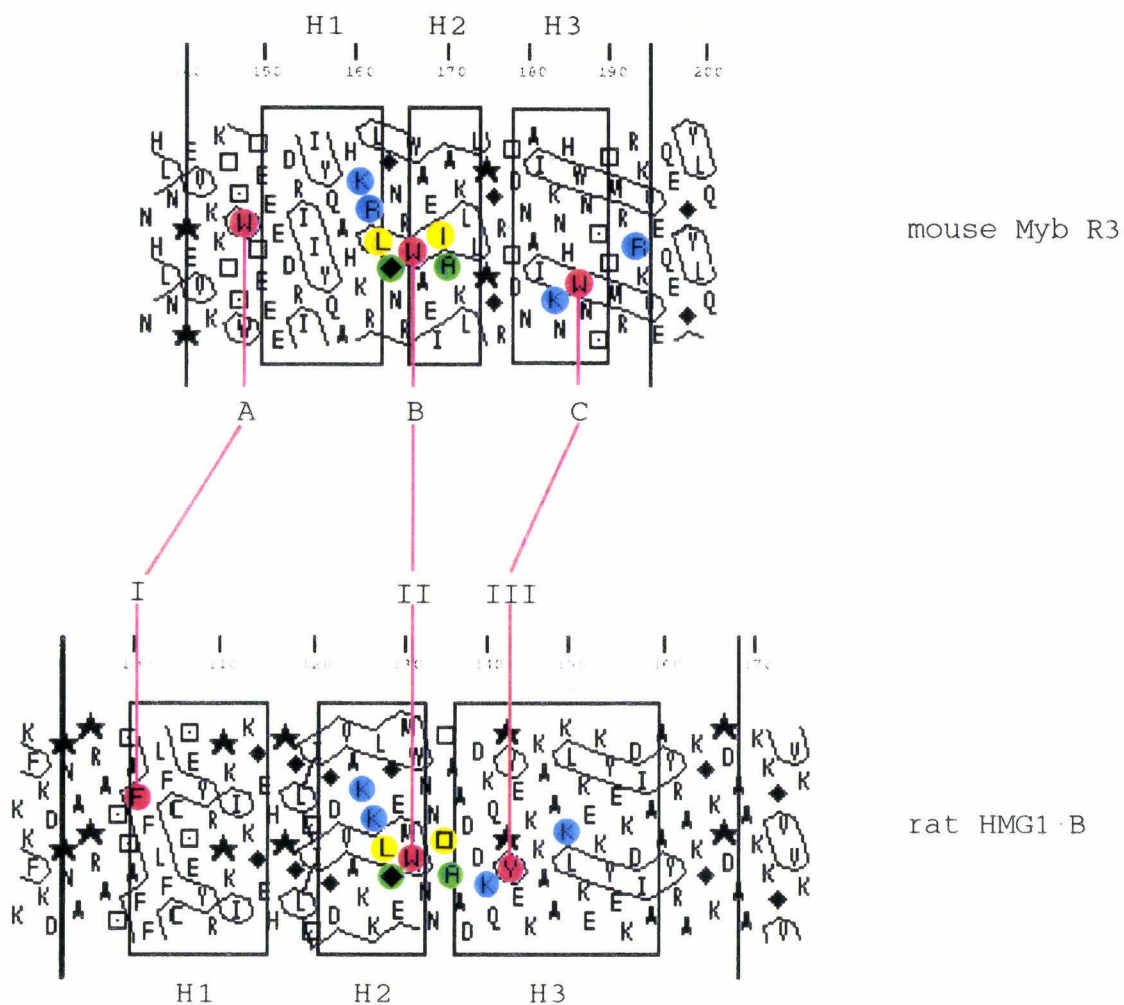


Figure 27 : Diagrammes HCA de R3-Myb et HMG-1B.

Les trois tryptophanes de Myb espacés régulièrement sont notés A, B et C. Les résidus aromatiques équivalents chez HMG-1B sont I, II et III. Les résidus aromatiques, hydrophobes, basiques et les acides aminés à chaîne latérale courte sont respectivement colorés en rouge, jaune, bleu et vert.

1.1.2. Alignement des diagrammes HCA.

La comparaison des diagrammes HCA de MYB-R3 et de HMG-1B est présentée Figure 27. Les résidus des hélices sont compris dans les rectangles et les formes circonscrites correspondent aux amas hydrophobes, dont la comparaison va permettre de détecter des identités de structure. La position respective des résidus hydrophobes (en jaune) et aromatique (en rouge) est similaire pour MYB-R3 et HMG-1B. De plus, l'emplacement des acides aminés basiques (en bleu) ou des résidus à chaîne latérale courte (glycine en losange et alanine, tous deux en vert) est aussi conservé. L'alignement proposé sur la base de la comparaison des amas hydrophobes est donc renforcé par la conservation, à l'extérieur des amas, des résidus hydrophiles. Outre une organisation structurale comparable, MYB-R3 et HMG-1B possèdent donc une identité de séquence détectable par la méthode HCA.

1.1.3. Alignement de séquence.

droso Myb	A	H	E	L	N	G	N	Q	W	A	K	I	A	MYB
BAS1	K	V	K	A	Y	G	T	K	W	R	K	I	S	
Zm C1 Myb	L	H	R	L	L	G	N	R	W	S	L	I	A	
Am 305 Myb	L	H	A	K	W	G	N	R	W	S	K	I	A	
Hv 33 Myb	L	H	K	I	L	G	N	R	W	S	Q	I	A	
v-Myb	A	H	K	R	L	G	N	R	W	A	E	I	A	
chicken Myb	A	H	K	R	L	G	N	R	W	A	E	I	A	
mouse Myb	A	H	K	R	L	G	N	R	W	A	E	I	A	
consensus Myb	a/l h k				l	G	n	r	W			I	a	
consensus HMG	v/i		k		l/i	g	e	r/k	w			l	s	
mouse SRY	I	S	K	Q	L	G	C	R	W	K	S	L	T	HMG
HMG1-A	F	S	K	K	C	S	E	R	W	K	T	M	S	
TCF-1	I	N	Q	I	L	G	R	R	W	H	A	L	S	
human UBF-1	L	T	K	I	L	S	K	K	Y	K	E	L	P	
ABF2-1	I	S	K	I	A	G	E	K	W	Q	N	L	E	
MC	V	S	K	L	V	G	E	M	W	R	N	E	S	
HMG1-B	V	A	K	K	L	G	E	M	W	N	N	T	A	
human UBF-2	V	K	D	S	L	G	K	Q	W	S	Q	L	S	
human UBF-5	R	M	V	E	I	G	S	R	W	Q	R	I	S	
tet HMGB	I	V	S	K	I	S	E	E	Y	K	A	L	P	
CCG1					R	I	Q	E	Q	L	R	R	L	K

Figure 28 : Alignement déduit de HCA, de diverses séquences des domaines MYB et HMG. Les résidus compris dans les rectangles correspondent aux acides aminés organisés en hélice. Lorsque le consensus est strict, le résidu est en majuscule.

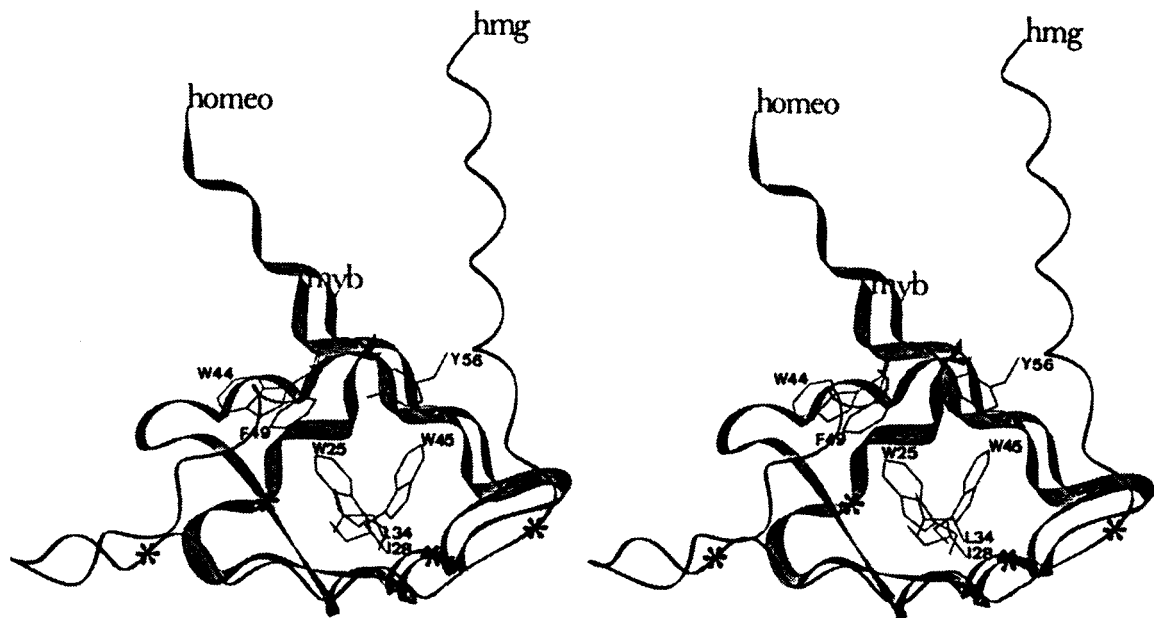


Figure 29 : Vue en stéréoscopie de Myb R3 (ruban large), de l'homéodomaine de Engrailed (ruban de largeur moyenne) et de HMG-1B (ruban fin).

La comparaison des consensus établis après compilation des séquences du domaine MYB et de la boîte HMG confirme l'alignement des diagrammes HCA (Fig. 28). Un alignement de séquences étendu sur 13 acides aminés est possible. Comme nous l'avons indiqué précédemment, les limites des hélices sont différentes d'une famille à l'autre. Ainsi, en dépit de la relative hétérogénéité de la famille HMG, l'alignement de séquence proposé pour MYB-R3 et HMG-1B est généralisable à l'ensemble des protéines Myb et HMG. La validité de cet alignement a été vérifiée par le calcul d'une valeur statistique appelée le score Z. Le score Z permet de déterminer si l'alignement de séquence réalisé est interprétable ou reflète simplement une composition en acides aminés biaisée. Lorsque Z est supérieur à 6 SD, on considère l'alignement comme significatif et lorsque Z est supérieur à 3 SD, on le juge comme peut-être significatif. Signalons également que la comparaison de séquences de petite taille donne toujours des scores Z plus faibles. Sur la portion de 13 résidus, le score Z entre MYB-R3 et HMG-1B est de 3.58 SD, suggérant donc que l'alignement que nous avons réalisé entre les deux familles était significatif.

En prenant en compte la similarité de séquence observée ci-dessus entre les hélices H2 et H3 de MYB-R3 et HMG-1B, nous avons superposé ces deux domaines de liaison à l'ADN (Fig. 29). Certains résidus occupent des positions équivalentes : ainsi les résidus aromatiques W25 et W44 de MYB-R3 correspondent à ceux de HMG-1B nommés W45 et Y56. L'angle formé entre les hélices H2 et H3 étant respectivement de 70° et 80° pour MYB-R3 et HMG-1B, les hélices H3 ne sont pas exactement superposables mais sont orientées dans une direction similaire.

On en conclut donc que ces deux domaines de liaison à l'ADN possèdent une identité de structure significative. MYB-R3 et HMG-1B se liant à des séquences spécifiques différentes, il semble logique d'observer de légères différences entre les structures spatiales de ces deux domaines de liaison à l'ADN.

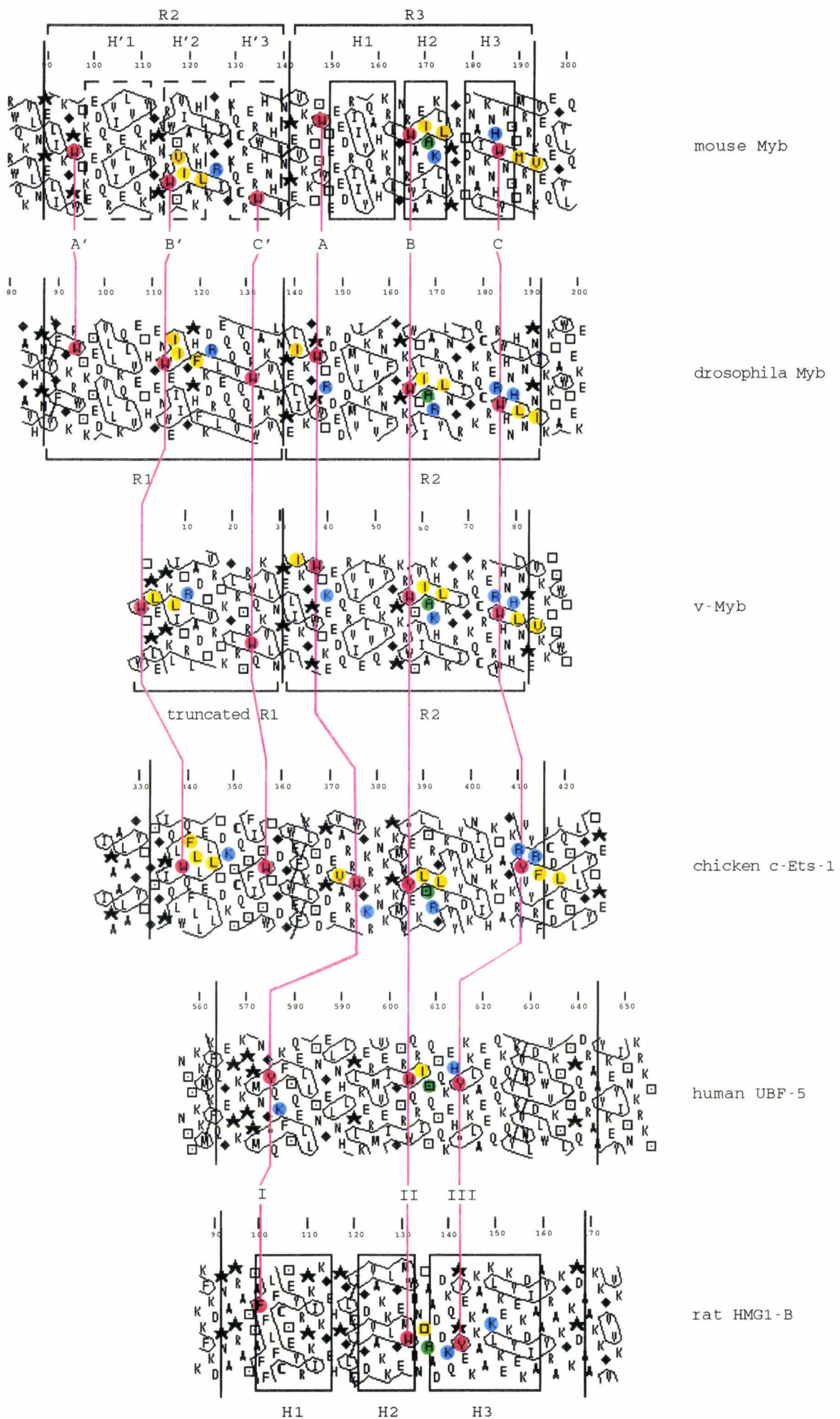


Fig. 30 : Diagramme HCA de trois membres de la famille Myb, de c-Ets-1 de poulet et de deux membres de la famille HMG.

Puisqu'il a été fortement supposé et démontré ultérieurement à notre étude (Ogata *et al.*, 1994) que l'hélice H3 de MYB-R3 était l'hélice de reconnaissance de l'ADN, une des implications de l'alignement que nous avons réalisé entre MYB-R3 et HMG-1B a été de désigner l'hélice H3 comme l'hélice de reconnaissance putative de HMG-1B. L'hélice H3 de HMG-1B n'a pas pour l'instant été désignée comme hélice de reconnaissance; les études structurales d'un complexe HMG-ADN sont en effet plus complexes car HMG-1 fixe l'ADN par le petit sillon, et que le petit sillon de l'ADN de forme B utilisé couramment dans les études de RMN n'est pas assez grand pour permettre l'insertion d'une hélice.

1.2. Alignements Ets-Myb et Ets-HMG.

Ayant validé la méthode HCA sur les domaines de liaison à l'ADN MYB-R3 et HMG-1B, dont la structure tridimensionnelle était connue et sachant que des identités de séquence, certes non significatives, entre les familles Myb et Ets d'une part et les familles HMG et Ets d'autre part, avaient été décrites, nous avons entrepris une prédiction de structure sur le domaine ETS.

La méthode HCA est surtout intéressante à utiliser dans le cas de protéines partageant des identités de séquence faibles. Si celle-ci est forte, comme c'est le cas pour la famille Ets, les diagrammes HCA construits sont très proches et leur utilisation redondante. Nous avons par conséquent choisi pour digne représentant de la famille le facteur de transcription c-Ets-1 (p54) de poulet mais nous avons vérifié que ce qui était valable pour c-Ets-1 l'était également pour toutes les autres protéines Ets.

La recherche sur banque de séquences protéiques de peptides similaires au domaine ETS par un programme de type FASTA a exhumé comme candidat potentiel Myb. Sur la figure 30, une partie du domaine MYB de drosophile est accompagné de celui de Myb de souris, dont on connaît la structure tridimensionnelle. La très forte identité de séquence non seulement entre R1, R2 et R3 (Fig. 31) mais aussi au sein de la famille Myb (Fig. 11) nous a permis d'une part d'aligner R1/R2 de Myb de drosophile avec

R2/R3 de Myb de souris (Fig. 30) et d'autre part d'extrapoler la structure spatiale obtenue pour R3 à R1 et R2.

```

R1  LGKTRWTREEDEKLKKLV EQNGTDDWKVIANYL PNRTDVQCQHRWQKVLNPE
R2  LIKGPWTREEDQRVIKLVQKYGPKRWSVIAKHLKGRIGKQCRERWHNHLNPE
R3  VKKTSWTEEEEDRIIYQAHKRLG NRWAEIAKLLPGRTDNAIKNHWNSTMRRKV
CONS      K  WT  EED                G  W  IA  L  R          W

```

Figure 31 : Alignement des séquences répétées R1, R2 et R3 de Myb souris. La ligne CONS représente le consensus.

La position des résidus aromatiques (marqués en rouge) est comparable pour les domaines ETS et MYB, au même titre que la forme des amas hydrophobes. Cette similarité est plus saisissante lorsqu'on examine les peptides correspondant aux hélices H'2, H2 et H3 pour lesquels non seulement les résidus aromatiques mais également les acides aminés hydrophobes et basiques ou les résidus possédant une chaîne latérale courte sont conservés. L'éventualité d'une relation de structure entre les domaines de liaison à l'ADN des protéines Ets et Myb est renforcée par la constatation que cet alignement est possible avec tous les membres de la famille Myb (non détaillé ici). Signalons que les scores statistiques Z calculés sont compris entre 3.65 SD et 7.94 SD pour l'alignement sur 67 résidus de différents domaines ETS avec le domaine MYB de drosophile, ce qui valide les résultats que nous avons obtenus.

En outre, la comparaison de c-Ets-1 avec v-Myb pour lequel R1 est tronqué (l'hélice H1 est absente) permet de confirmer les similarités observées et de mettre en avant le fait que le domaine ETS semble comporter trois hélices correspondant à une séquence répétée complète de Myb (en carboxy-terminal du domaine ETS) plus deux hélices en amino-terminal correspondant à une séquence répétée tronquée. MYB-R3 comprend trois hélices, H2 et H3 formant le motif hélice-coude-hélice. Un domaine ETS correspondant, selon la méthode HCA, aux deux tiers d'un domaine MYB (soit un

motif hélice-tour-hélice) suivi d'un domaine MYB entier (soit un second motif hélice-tour-hélice), on pouvait alors suggérer que le domaine de liaison à l'ADN des protéines Ets comprenait deux motifs hélice-coude-hélice séparés par une hélice.

Comme on pouvait s'y attendre au vu des similarités de structure et de séquence entre MYB-R3 et HMG-1B, il était possible d'aligner le domaine ETS avec les boîtes HMG (Fig. 30). Comme pour Ets-Myb, l'identité de séquence observée repose à la fois sur la conservation de résidus hydrophobes (intégrés dans les amas) et sur celle de résidus hydrophiles. Néanmoins le nombre de résidus conservés rend délicate l'interprétation. La faible longueur des séquences comparées ne permet pas non plus d'obtenir des scores statistiques valables.

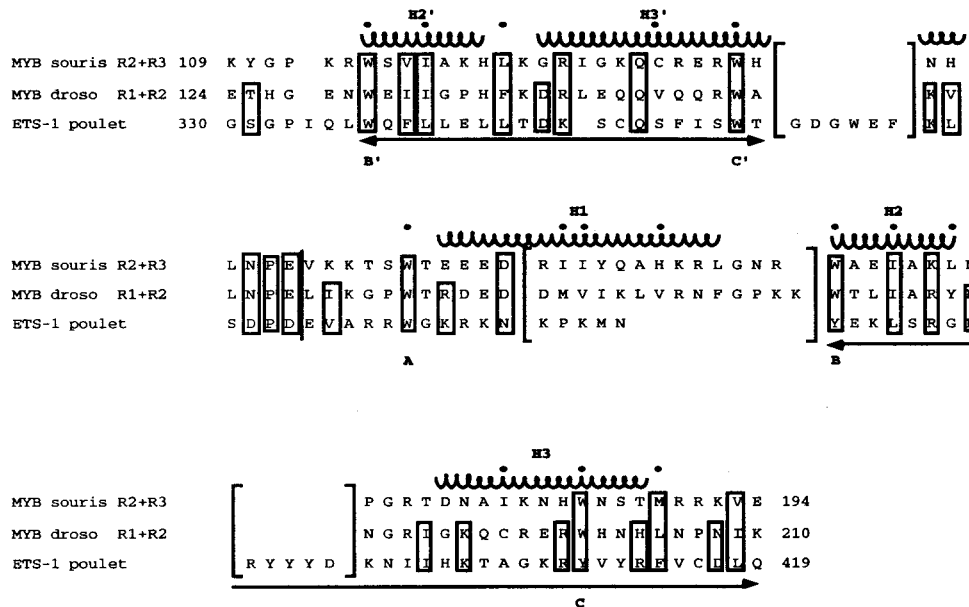


Figure 32 : Alignement déduit par la méthode HCA de deux domaines de la famille Myb (souris et drosophile) et du domaine ETS de c-Ets-1 de poulet. Les hélices H'2 et H'3 ont été positionnées d'après l'alignement des domaines R2 et R3. La barre verticale indique la limite entre deux domaines consécutifs de Myb. Les points correspondent aux résidus impliqués dans la formation du noyau hydrophobe. Les résidus encadrés sont conservés et les hélices correspondent à la structure obtenue sur le domaine MYB-R3.

En résumé, la méthode HCA nous a permis d'aligner une partie des domaines MYB et HMG. De plus, l'alignement des séquences des domaines MYB et ETS, déduit de la comparaison de leurs diagrammes HCA respectifs, a permis de proposer une

prédiction de structure pour le domaine ETS (Fig. 32). Nous avons suggéré une similarité dans l'organisation spatiale des domaines de liaison à l'ADN des familles de facteurs de transcription Ets, Myb et HMG. Selon cette proposition, le domaine ETS semble comprendre cinq hélices correspondant à une séquence répétée tronquée plus une séquence répétée entière du domaine de fixation MYB. Le domaine contiendrait au moins un motif hélice-coude-hélice.

2. Etude des capacités de liaison à l'ADN de ERM.

2.1. Etude du domaine ETS.

2.1.1. Mutations ponctuelles.

La prédiction de structure et par conséquent la similarité d'organisation tridimensionnelle entre les domaines ETS et MYB avait deux conséquences importantes : la première était que le caractère aromatique des tryptophanes de Myb régulièrement espacés de 18/19 acides aminés était conservé dans le domaine ETS. Sur les cinq résidus, trois étaient conservés de manière stricte et deux changés en tyrosine (Fig. 32). Les données de RMN ont montré que le noyau hydrophobe qui maintient le squelette de la structure était formé par une dizaine de résidus dont ces tryptophanes (Ogata *et al.*, 1992). On pouvait donc supposer qu'une organisation similaire était présente dans le domaine ETS. La seconde conséquence était que le rôle de reconnaissance de l'ADN endossé par l'hélice H3 de Myb devait sans doute être porté par l'hélice équivalente dans Ets.

Pour appuyer la prédiction de structure que nous avons réalisée, nous avons entrepris un travail de mutations ponctuelles sur des résidus du domaine ETS. Au regard des considérations exposées ci-dessus, les résidus aromatiques semblaient un morceau de choix pour ces expériences, d'autant plus qu'ils sont conservés au sein de la famille Ets et que, la prédiction de structure par HCA valable pour l'ensemble des domaines ETS, il importait de considérer des résidus présents dans la plupart des protéines.

ERM étant l'objet des soins attentionnés de l'équipe SPRITE, nous avons donc choisi de muter son domaine ETS afin de préciser sa fonction de liaison à l'ADN. Nous avons focalisé notre attention sur les tyrosines 418 et 442 pour trois raisons essentielles (Fig. 33). Elles sont situées dans la zone où notre prédiction de structure est la plus assurée. De plus, ces résidus sont à la fois conservés entre Myb et Ets et impliqués potentiellement dans le noyau hydrophobe équivalent à celui de Myb. Enfin, les tyrosines 418 et 442 sont chacune localisées dans une hélice différente d'un des potentiels motifs hélice-coude-hélice du domaine ETS.

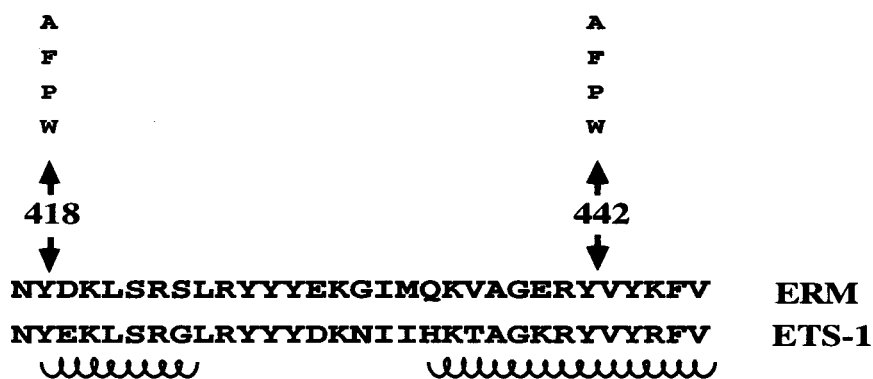


Figure 33 : Position des tyrosines 418 et 442.

Les hélices positionnées en dessous de la séquence de c-Ets-1 correspondent à notre prédiction de structure. Les tyrosines sont mutées en alanine (A), phénylalanine (F), proline (P) ou tryptophane (W).

La stratégie a consisté à muter en quatre acides aminés différents :

la proline qui introduit un coude dans la plupart des hélices où elle est présente, et donc devait affecter la structure du domaine ETS de manière drastique,

l'alanine, mutation certes non conservative mais qui est censée amener un minimum de perturbations tridimensionnelles,

le tryptophane,

la phénylalanine.

La mutation en ces deux derniers résidus est conservative et permet d'estimer le degré de sensibilité de la structure : la phénylalanine possède le noyau benzène de la tyrosine et ne diffère que par l'absence d'un groupement hydroxyle, tandis que la structure du

noyau indole du tryptophane est plus éloignée de celle de la tyrosine tout en conservant son caractère aromatique.

L'effet des mutations sur la structure tridimensionnelle du domaine ETS a été mesuré, en utilisant la technique de gel-retard, par la capacité à fixer une sonde reconnue par le domaine ETS de ERM non muté. Il s'agit d'une séquence dérivée de l'amplificateur du polyome. Ainsi que le montre la figure 34, les effets de la mutation sur les tyrosines 418 et 442 sont différents. La mutation de la tyrosine 418 en alanine (Y418A), en proline (Y418P), en tryptophane (Y418W) et en phénylalanine (Y418F) abolit totalement l'activité de liaison à l'ADN, par comparaison avec le domaine ETS de ERM non muté (ΔD). En revanche, la tyrosine 442 réagit de manière variable : la mutation en proline (Y442P) a un effet drastique sur la liaison à l'ADN; la mutation en alanine (Y442A) ou en tryptophane (Y442W) entraîne une diminution considérable tandis que la mutation en phénylalanine (Y442F) permet de préserver l'intégrité de la capacité de fixation à l'ADN. La quantification, réalisée à l'aide de l'analyseur d'image Phosphorimager (Molecular Dynamics), permet d'estimer la perte de la capacité à lier l'ADN à 90% suite à une mutation en alanine ou en tryptophane. Il est intéressant de remarquer que la double mutation des tyrosines Y418 et Y442 en tryptophane (Y418W/Y442W) conserve une faible activité de liaison à l'ADN alors que la mutation en tryptophane de la seule tyrosine 418 (Y418) suffit à abolir cette fonction.

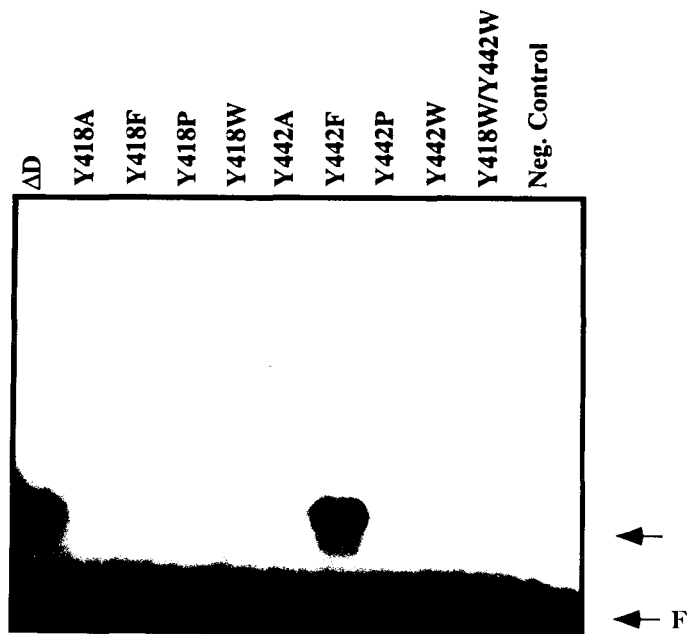


Figure 34 : Importance des tyrosines 418 et 442 dans la liaison à l'ADN du domaine ETS de ERM.

Gel-retard en présence de l'oligonucléotide dérivé de l'amplificateur du virus du polyome. La flèche accompagnée d'un F ou non indique respectivement l'oligonucléotide marqué libre ou complexé à la protéine.

Ces expériences ont été reproduites sur une autre séquence cible reconnue par le domaine ETS de ERM (Fig. 35). Elles ont conforté les données exposées ci-dessus pour deux raisons : premièrement, l'effet des mutations décrites n'est pas restreint à un seul type d'oligonucléotide, en l'occurrence l'oligonucléotide dérivé de l'amplificateur du virus du polyome; deuxièmement, l'effet des mutations est observable sur des oligonucléotides que le domaine ETS reconnaît différemment. En effet, ce second oligonucléotide, nommé EBS, est moins bien reconnu que l'oligonucléotide dérivé du virus du polyome (Fig. 36).

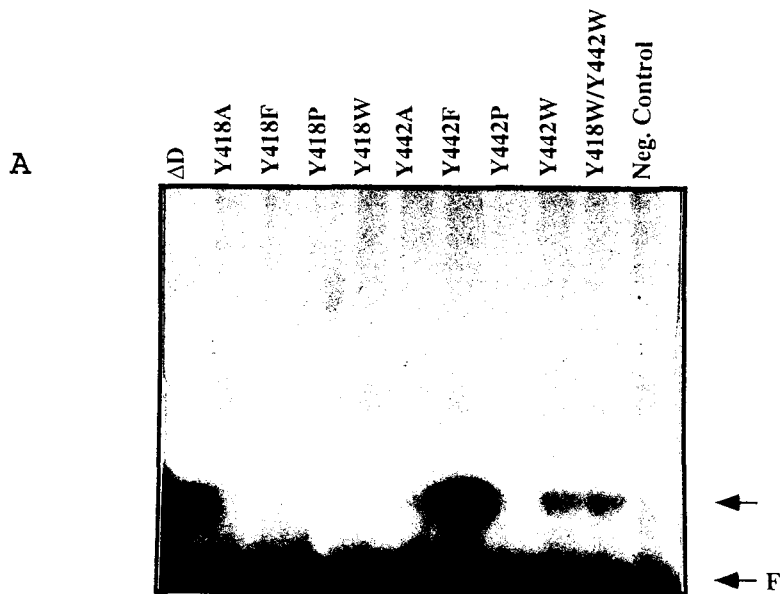


Figure 35 : Importance des tyrosines 418 et 442 dans la liaison à l'ADN du domaine ETS de ERM.
Gel-retard en présence de l'oligonucléotide EBS.

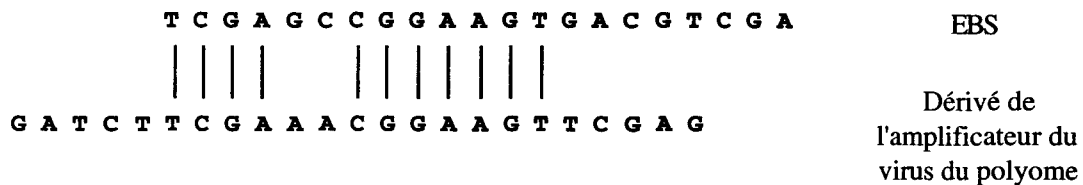


Figure 36 : Comparaison des séquences des oligonucléotides utilisés en gel-retard.

Signalons que les résultats que nous avons obtenus ne sont pas dus à un artéfact de production des protéines *in vitro* par lysat de réticulocytes (Fig. 37) : après séparation sur un gel d'acrylamide, les protéines correspondant aux différents mutants ont été quantifiées.

Ces données montrent que les tyrosines 418 et 442 du domaine ETS de ERM ne jouent pas un rôle équivalent dans la fonction de liaison à l'ADN : alors que la tyrosine 418 est très sensible aux mutations, le remplacement à la position 442 d'une tyrosine en phénylalanine préserve l'activité de liaison à l'ADN, ce qui semble indiquer qu'elle joue un rôle moins important.

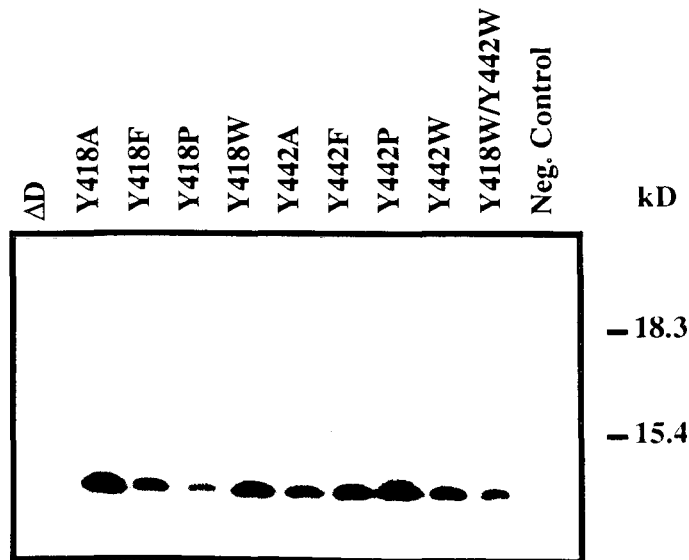


Figure 37 : Visualisation des différentes constructions du domaine ETS, native ou mutées sur la tyrosine 418 ou 442.

Les protéines ont été produites *in vitro* en lysat de réticulocytes et séparées sur gel d'acrylamide-tricine.

Deux éléments indiquent que les résultats obtenus dans le contexte du domaine ETS sont valables dans celui de la protéine de pleine longueur. Un peptide que nous avons amplifié par PCR, correspondant aux 221 derniers acides aminés de ERM et comprenant donc le domaine ETS, montrait une faible capacité à lier l'ADN. Nous l'avons donc séquencé et observé la mutation de la tyrosine 442 en cystéine. Ceci indique que le changement de la tyrosine en cystéine induit une diminution de l'activité de fixation à l'ADN, diminution comparable à celle causée par l'introduction d'une alanine ou d'un tryptophane. Cela signifie aussi que, en ce qui concerne les mutations affectant la fonction du domaine ETS, ce qui est valable dans le contexte du domaine ETS peut aussi l'être pour des versions tronquées de ERM de plus grande taille.

Le second élément est que l'introduction d'une proline en position 418 (construction qui nous a été utile pour l'étude du pouvoir transactivateur de ERM) dans la protéine ERM de pleine longueur inhibe la liaison à l'ADN (résultats non montrés). Ces deux

indices permettent donc de penser que l'effet des mutations réalisées dans le domaine ETS en vue d'abolir sa fonction n'est pas compensé par une autre partie de ERM.

2.1.2. Les mutants rêve.

Pour les récepteurs stéroïdes dont le domaine de liaison à l'ADN, organisé en doigts de zinc, est très conservé, la mutation de deux acides aminés suffit à modifier la spécificité de liaison à l'ADN. Le changement de la glycine 458 et la sérine 459 du récepteur aux glucocorticoïdes respectivement en glutamine et glycine, résidus qui occupent la place équivalente dans le récepteur aux oestrogènes, suffit à changer la spécificité de fixation du récepteur des glucocorticoïdes en celle du récepteur aux oestrogènes. L'opération réciproque conduit à une modification du récepteur aux oestrogènes qui acquiert alors la spécificité de reconnaissance du récepteur aux glucocorticoïdes (Berg *et al.*, 1989). En conséquence, afin de vérifier si l'hypothèse d'une identité de structure entre les domaines de liaison à l'ADN de Myb et Ets était vraie, nous avons essayé de réaliser le même genre d'échange entre le domaine MYB et le domaine ETS de ERM.

Un pentapeptide WAEIA couvrant la presque totalité de l'hélice H2 du domaine MYB et qui est suspecté de porter la spécificité de fixation à l'ADN du domaine MYB (D. Leprince, communication personnelle), a été introduit à la place du pentapeptide YEKLS du domaine ETS de ERM, donnant naissance à la chimère appelée rêve ERM (Fig. 38). Après production en lysat de réticulocytes, cette construction a été testée en gel-retard et ne se fixe pas sur une séquence reconnue par le domaine MYB (don du Dr S. Plaza) (Fig. 38), au contraire du domaine MYB, témoin positif de notre expérience.

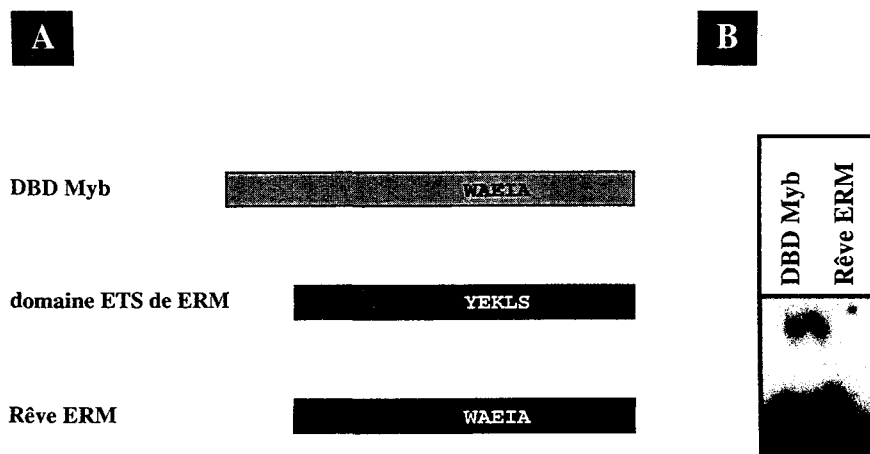


Figure 38 : (A) Les six résidus constituant la majeure partie de H2 dans le domaine de liaison à l'ADN de Myb ont été introduits dans le domaine ETS de ERM à la place des résidus équivalents (selon l'alignement présenté dans la figure 32) donnant ainsi naissance à la construction Rêve ERM.

(B) Gel-retard réalisé en présence d'un oligonucléotide reconnu par le domaine MYB.

Ce résultat négatif peut être interprété autant comme une réfutation du modèle que nous avons proposé que comme une perturbation trop importante du système. Il est en effet possible que la modification de structure consécutive à la mutation de cinq acides aminés soit trop drastique pour permettre un repliement correct de la chimère et le recouvrement de ses capacités de fixation à l'ADN. L'identité de structure que nous supposons entre le domaine MYB et le domaine ETS ne s'accompagne pas, comme c'est le cas pour les récepteurs stéroïdes, d'une identité de séquence élevée.

2.1.3. Localisation de régions de ERM inhibitrices de la liaison à l'ADN.

Dans le but de délimiter, si elles existent, des régions modulant l'activité du domaine ETS, nous avons construit une série de mutants de délétion à partir des extrémités amino-terminale et carboxy-terminale (Fig. 23 page 76). Leur capacité à

lier l'ADN a été examinée dans les mêmes conditions que pour les mutants du domaine ETS. Les gel-retards ont été réalisés avec la sonde dérivée de l'amplificateur du virus du polyome sur une même quantité de protéines produites en lysat de réticulocytes (Fig. 39).

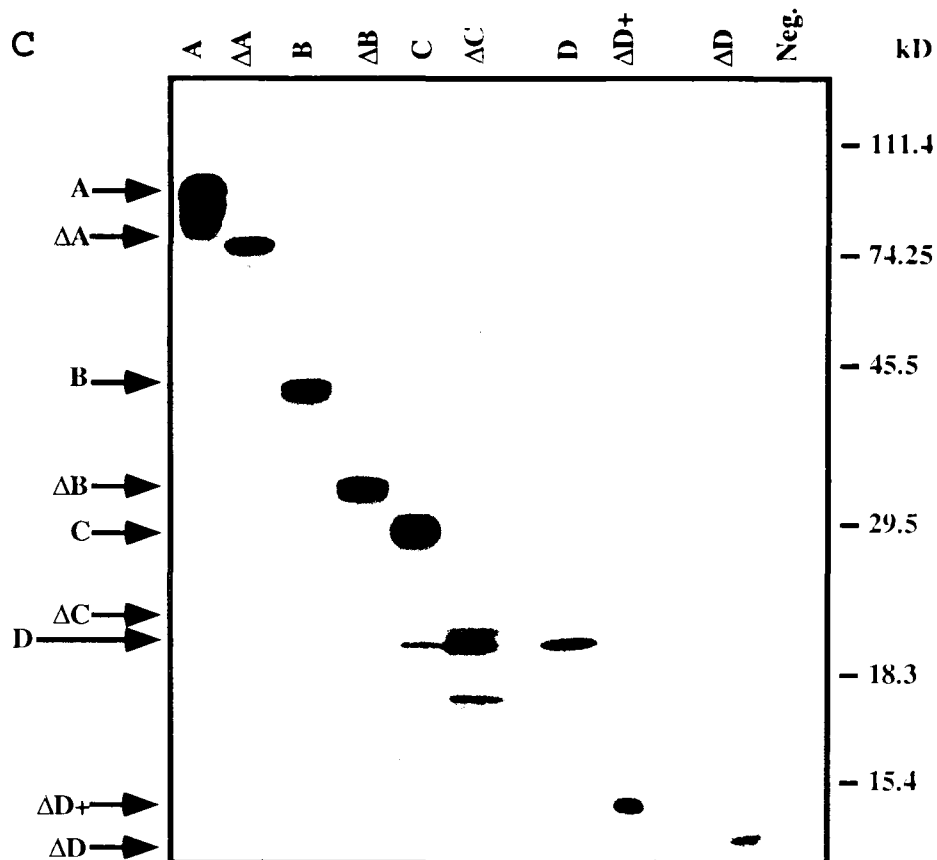


Figure 39 : Visualisation des différentes constructions délétées en amino- et en carboxy-terminal de ERM.

Les protéines ont été produites *in vitro* en lysat de réticulocytes et séparées en gradient d'acrylamide 5%-20%. Les flèches indiquent la taille du produit attendu.

On peut constater sur la Fig. 40 que la traduction en lysat de réticulocytes de nos constructions donne non seulement le produit attendu mais aussi des produits de poids moléculaire plus faible. Ceux-ci correspondent à des initiations de la traduction à partir de méthionines internes et sont considérés comme des artefacts de la traduction *in vitro* des protéines. Ces sous-produits contiennent vraisemblablement le domaine ETS et la plupart du temps, des bandes correspondant à leur association avec la sonde seront donc visibles en gel-retard (Fig. 40).

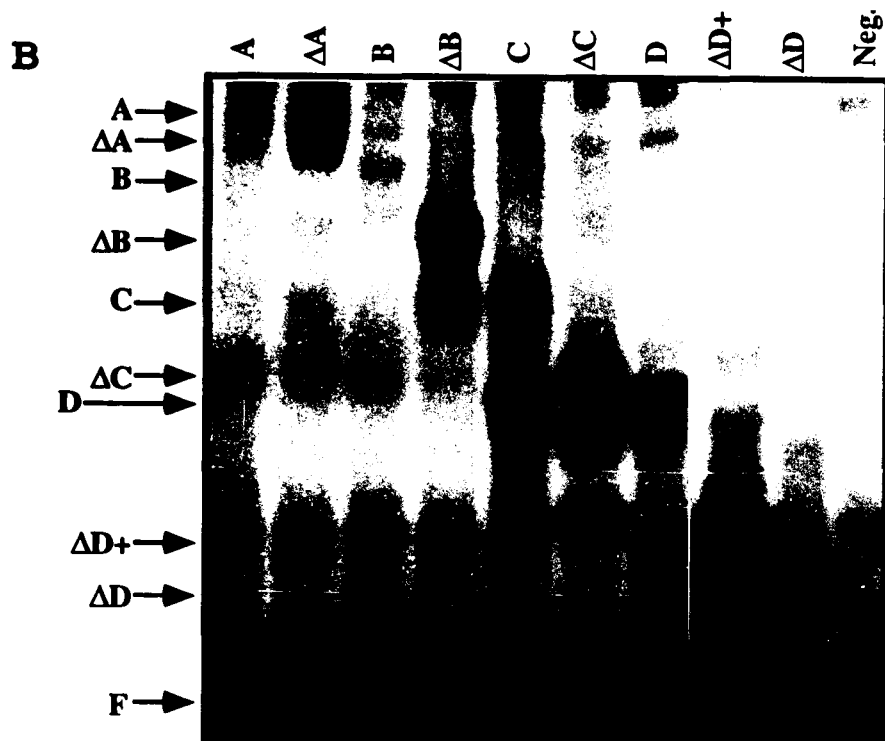


Figure 40 : Localisation des domaines d'inhibition de la liaison à l'ADN de ERM.

Gel-retard en présence de l'oligonucléotide dérivé du virus du polyome. Les flèches indiquent la taille du complexe attendu. F représente l'oligonucléotide libre.

La protéine ERM totale (A) se fixe beaucoup moins bien que le domaine ETS de ERM (ΔD) (Fig. 40). Une analyse plus détaillée montre que plus la protéine est tronquée en amino-terminal, plus sa liaison à l'ADN est forte (comparer A, B et C avec D). Cette constatation est valable que les constructions soient pourvues (A, B, C et D) ou tronquées de la queue carboxy-terminale (ΔA , ΔB , ΔC et ΔD).

De plus, sans qu'une quantification ait été faite, il semble que A et B se lient avec la même intensité à l'ADN alors que C se fixe plus fort (pistes A et B comparées à la piste C). Nous suggérons qu'il existe une région inhibitrice de la liaison à l'ADN entre les acides aminés 203 et 290, que nous avons nommée ADID pour Amino-terminal DNA-binding Inhibitory Domain. Le fait que le comportement des différentes versions tronquées soit indépendant de la présence de la queue carboxy-terminale conforte cette assertion.

Simultanément nous avons constaté que les mutants de délétion A, B, C et D se liaient à l'ADN beaucoup moins fortement que les constructions correspondantes tronquées de la queue carboxy-terminale (Fig. 40). Ceci suggérait que la queue carboxy-terminale était douée d'une capacité à inhiber la fonction du domaine ETS. Dans les conditions dans lesquelles nous avons effectué les gels-retard, cette propriété est indépendante des régions situées en amont du domaine ETS et de ADID en particulier. En revanche, cet effet est spécifique du domaine ETS de ERM puisque l'inhibition exercée par la queue carboxy-terminale est la même que celle exercée par les 72 premiers résidus de ERM (qui ne contiennent aucun domaine inhibiteur de la liaison à l'ADN) lorsqu'elle est fusionnée à un domaine hétérologue de liaison à l'ADN comme celui du proto-oncogène *c-erbA* (Fig. 41). Cet effet d'inhibition pourrait correspondre à une gêne stérique.

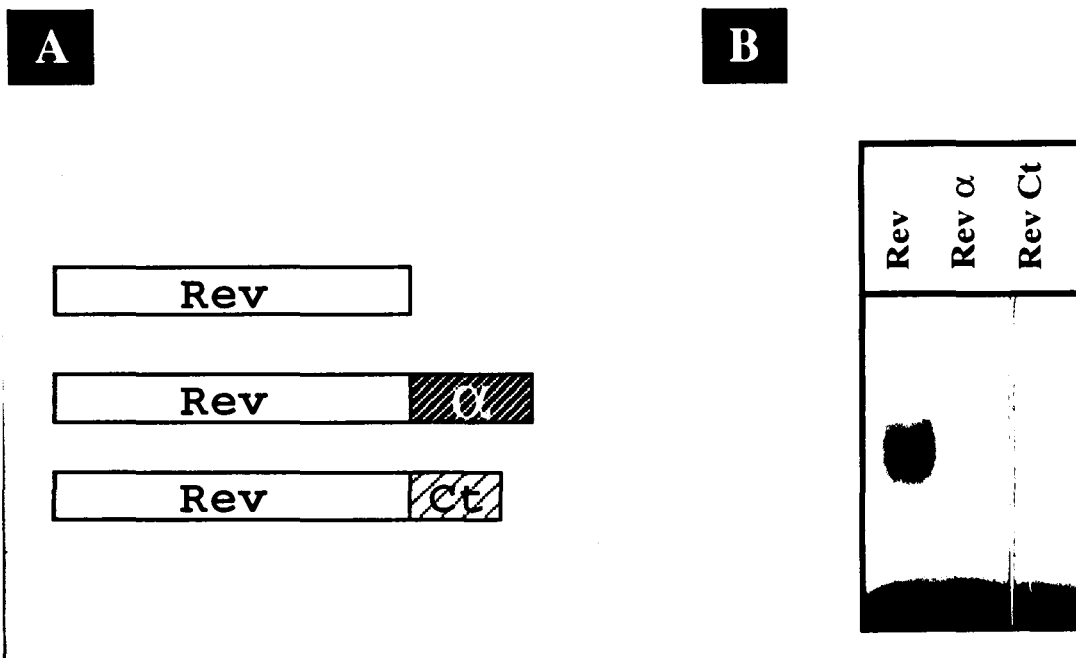


Figure 41 : L'effet inhibiteur de la queue carboxy-terminale est spécifique de ERM.

(A) La construction RevCt comprend le domaine hétérologue de liaison à l'ADN Rev et l'intégralité de la queue carboxy-terminale de ERM (Ct). De même a été construite une chimère comprenant Rev et les 72 premiers résidus de ERM (α)

(B) Gel-retard avec l'oligonucléotide reconnu par le domaine Rev.

L'alignement des séquences en acides aminés des trois membres connus du groupe PEA3 (PEA3, ER81 et ERM) indique que le peptide de 20 résidus contigu au domaine ETS et compris dans la queue carboxy-terminale était mieux conservé que l'extrémité carboxy-terminale (Fig. 42). Nous avons produit une construction $\Delta D+$ qui correspond au domaine ETS flanqué de cette région de 20 résidus et montré qu'elle se fixait à l'ADN comme le domaine ETS seul (comparer $\Delta D+$ à ΔD dans la Fig. 40). Cette donnée signifie que la fonction d'inhibition portée par l'extrémité carboxy-terminale n'est pas médiée par les 20 premiers acides aminés mais par la partie la moins conservée.

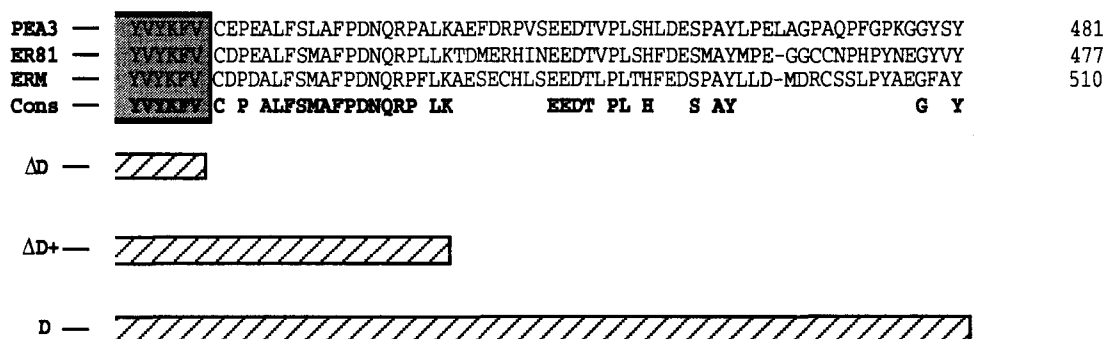


Figure 42 : Alignement des séquences des protéines humaines PEA3, ER81 et ERM sur la partie carboxy-terminale. CONS désigne le consensus. La limite carboxy-terminale de chaque construction est représentée en hachuré.

L'ensemble de nos résultats montre donc que ERM contient deux régions qui inhibent de manière indépendante l'activité de liaison à l'ADN : un domaine situé en amont du domaine ETS appelé ADID ainsi qu'un peptide correspondant à la région la moins conservée de la queue carboxy-terminale.

3. Etude des propriétés de transactivateur de ERM.

L'appartenance de ERM à la famille des facteurs de transcription de Ets ainsi que l'existence chez ERM d'une région (le domaine acide) qui comprend 32% de résidus acides sur 31 acides aminés et qui constitue donc un domaine transactivateur potentiel, nous ont fortement incités à examiner avec attention les propriétés d'augmentation de la transcription de ERM.

3.1. Interférence du domaine ETS dans l'étude de la transactivation par ERM.

Ainsi que le montre la figure 43, la protéine ERM de pleine longueur (pSV ERM) active la transcription à partir d'un plasmide rapporteur contenant trois copies d'une séquence reconnue par les protéines Ets. Dans les deux lignées que nous avons examinées, cellules de carcinome humain du col de l'utérus (HeLa) et cellules du carcinome du rein de lapin (RK13), l'effet transactivateur de ERM est respectivement de 10 fois et 25 fois celui du contrôle, le vecteur pSV.

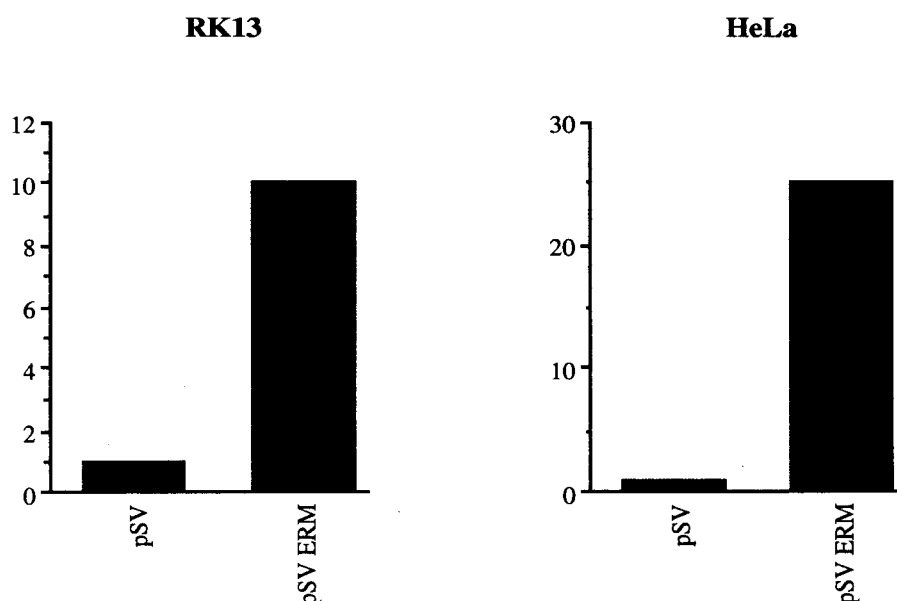


Figure 43 : Transactivation par ERM humain exprimée par rapport au témoin pSV. Le vecteur rapporteur 3xTORU-tk-CAT contient la séquence cible TORU 5'-GATCTCGAGGAAGTGACTAACTGAGCAGATCCTCGACTGTGCTCAGGTAGTCAC-3'.

Il nous a semblé délicat d'utiliser les différentes constructions tronquées de ERM pour délimiter les domaines transactivateurs de ERM pour trois raisons. La première est l'incapacité à discerner l'effet dû à une modulation de liaison à l'ADN de celui dû à une modulation du pouvoir transactivateur. De plus, on ignore pour l'instant les relations entre fonction de liaison à l'ADN et fonction de transactivation. Certains indices laissent néanmoins penser que, pour les protéines de la famille Ets et dans les conditions utilisées par notre laboratoire, l'effet transactivateur observé soit fonction inverse de la capacité à fixer l'ADN. La seconde raison est que l'emploi des versions tronquées de ERM sous-entend l'utilisation d'un vecteur rapporteur contenant des sites reconnus par les protéines Ets. Dans les lignées HeLa et RK13, en la seule présence de ce vecteur rapporteur (en l'absence de vecteur exprimant tout ou partie de ERM), l'activité de transactivation est assez élevée par rapport au bruit de fond, ce qui suggère que ces lignées expriment de manière endogène des membres de la famille Ets. A cause de cette activité basale importante, une variation fine dans le pouvoir transactivateur risque de passer inaperçue. La troisième raison est la nécessité de pouvoir quantifier les protéines par immunoprécipitation dans nos essais de transfection transitoire. En effet, l'anticorps dirigé contre le domaine ETS des membres du groupe PEA3 n'est pas assez performant pour les immunoprécipitations. D'ailleurs, dans l'hypothèse où il l'aurait été, rien ne permet de penser qu'il reconnaîtrait avec la même affinité les différentes formes tronquées. Nous nous sommes affranchis de ces contraintes par la construction de protéines de fusion avec un domaine de liaison à l'ADN hétérologue, les 147 résidus de la protéine de levure GAL4.

La protéine ERM de pleine longueur est introduite en phase dans le vecteur pSG424 (Fig. 44). Dans un premier temps, nous avons décidé de déléter les régions les plus conservées au sein du groupe PEA3 c'est-à-dire les 72 premiers résidus (région α) qui recouvre le domaine acide (résidus 42 à 72) et la queue carboxy-terminale (nommée Ct), soit de manière indépendante (GAL4 ERM- α , GAL4 ERM-Ct) soit concertée (GAL4 ERM- α Ct). Ces quatre constructions sont incapables de transactiver (Fig. 44). Ce défaut peut être causé par la présence simultanée de deux domaines de liaison à

l'ADN : le domaine ETS et le domaine hétérologue GAL4. Il est possible que le domaine ETS de la chimère lie des sites endogènes reconnus par les protéines Ets, séquestrant ainsi une partie des molécules chimères destinées à transactiver à partir du vecteur rapporteur GAL4.

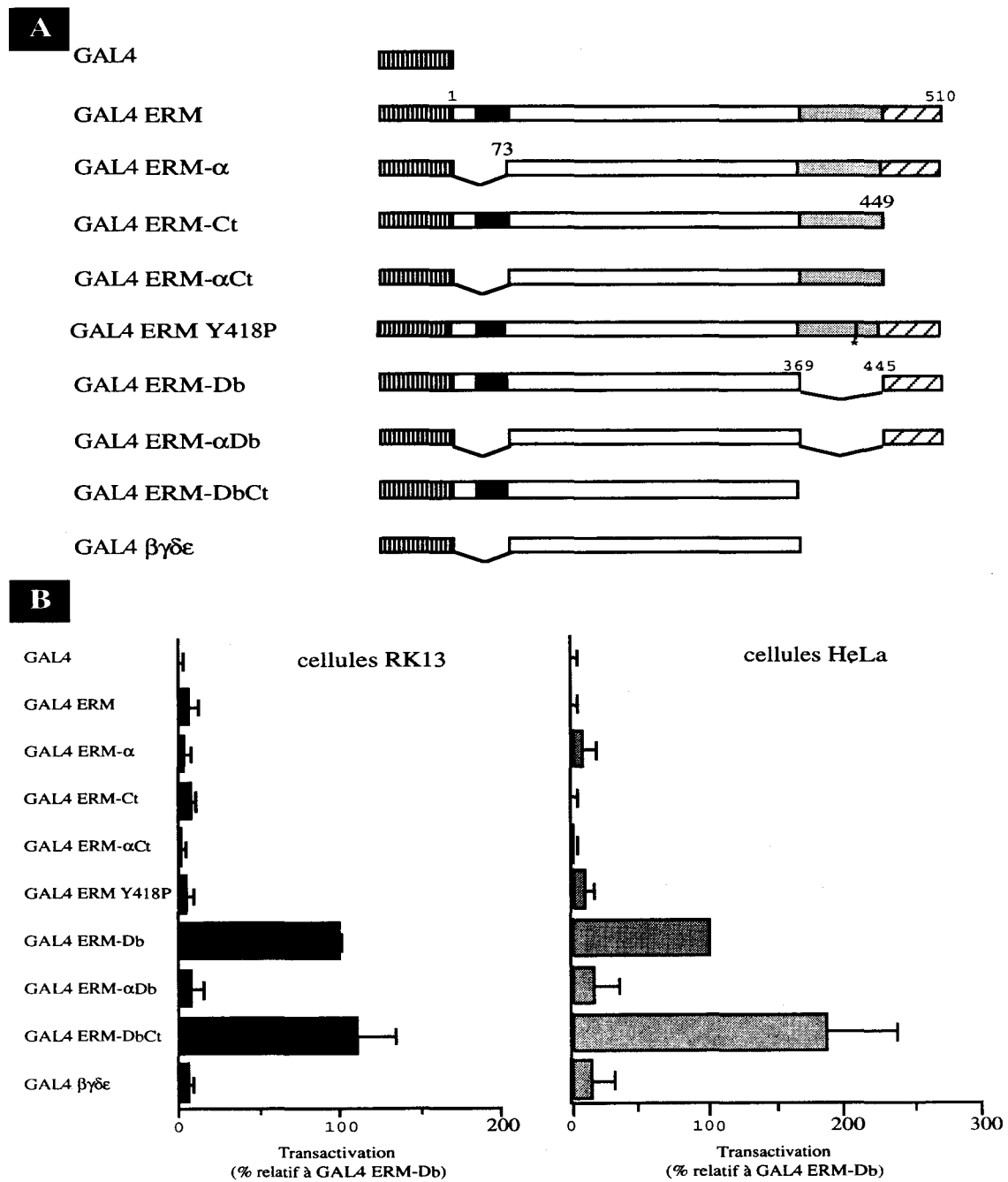


Figure 44 : Domaines essentiels à la transactivation par ERM.

Afin de mettre à l'épreuve cette hypothèse, nous avons introduit la mutation Y418P dans la construction GAL4 ERM (construction alors nommée GAL4 ERM Y418P). Cette mutation inhibe totalement l'activité de liaison à l'ADN comme il a été exposé ci-dessus. Il est surprenant de constater que l'abolition de la fonction du domaine ETS ne conduit pas à un effet transactivateur plus important (Fig. 44). Ceci signifie en premier lieu que l'hypothèse émise est fautive, et en second lieu, que la seule présence du domaine ETS, qu'il soit fonctionnel ou non, suffit à perturber la transactivation exercée par les chimères GAL4 ERM. On peut alors supposer que lorsqu'il n'est pas engagé dans une liaison ADN-protéine avec son site de reconnaissance, le domaine ETS adopte une conformation qui gêne considérablement la transactivation ou qu'il est impliqué dans des interactions intermoléculaires ou intramoléculaires. A ce propos, il convient de signaler que toutes les chimères GAL4 ERM utilisées contiennent le domaine d'inhibition de la liaison à l'ADN nommé ADID. Des contacts intramoléculaires entre ADID et le domaine ETS pourraient éventuellement expliquer la perturbation observée.

Afin d'examiner le pouvoir transactivateur de α et de Ct, nous avons donc repris les quatre constructions GAL4 ERM, GAL4 ERM- α , GAL4 ERM-Ct et GAL4 ERM- α Ct et nous les avons amputées du domaine ETS : GAL4 ERM-Db, GAL4 ERM- α Db, GAL4 ERM-DbCt et $\beta\gamma\delta\epsilon$. A notre grand soulagement, la chimère GAL4 ERM-Db est capable de transactiver, de même que la construction GAL4 ERM-DbCt qui est dépourvue de la queue carboxy-terminale. Les résultats sont similaires dans les HeLa et dans les RK13 et les deux constructions stimulent la transcription avec la même intensité. En revanche, les deux protéines de fusion pour lesquelles les 72 premiers résidus (α) ont été délétés ne sont pas dotées d'un pouvoir transactivateur supérieur à celui du vecteur seul.

En résumé, nous avons montré que le peptide de 72 résidus recouvrant le domaine acide était indispensable à la transactivation tandis que la queue carboxy-terminale ne l'est pas. Nous avons accessoirement prouvé que l'absence de pouvoir transactivateur observée pour des facteurs de transcription fusionnés à un domaine hétérologue de

reconnaissance de l'ADN n'était pas dû, dans le cas de ERM, à une séquestration des chimères par les sites endogènes.

3.2. Pouvoir de transactivation intrinsèque de α et de Ct.

Afin de déterminer si le fragment α que nous avons désigné comme responsable de l'augmentation transcriptionnelle par ERM exhibait un pouvoir transactivateur, nous avons construit des chimères en fusionnant GAL4 à toute une série de peptides recouvrant l'intégralité de la protéine ERM (Fig. 45). En utilisant les mêmes lignées de cellules que précédemment, il apparaît que les fragments α (les 72 premiers acides aminés) et Ct (correspondant à la queue carboxy-terminale) sont doués de la capacité à transactiver quand ils sont fusionnés avec le domaine GAL4. En revanche, les fragments β , γ , δ , ϵ et Db n'augmentent pas la transcription. Les deux lignées ne répondent pas de manière identique : en effet, le rapport GAL4 Ct/GAL4 α est respectivement de 37,5% et de 300% dans les RK13 et les HeLa (Fig. 45).

Par ailleurs, si le fragment Db ne transactive pas, on peut imaginer une fois de plus que c'est sa fonction de liaison à l'ADN qui perturbe le système. Nous avons donc testé la même construction en y introduisant la mutation Y418P. La construction GAL4 Db* n'est pas capable d'augmenter la transcription, suggérant ainsi que le domaine ETS ne contient aucun fragment transactivateur.

Nous avons voulu vérifier si la région comprise entre la fin du domaine acide et le début du domaine ETS ne comprenait pas un domaine transactivateur qui aurait été coupé en deux (ou plus) par le découpage en différents fragments. Pour tester cette éventualité, nous avons examiné les chimères GAL4 $\gamma\delta$ et GAL4 $\beta\gamma\delta\epsilon$... qui ne transactivent pas (Fig. 45). Le cas de la protéine de fusion GAL4 $\alpha\beta\gamma$ suggère de nouveau que le type cellulaire utilisé est important : dans les RK13, elle a la même activité que GAL4 α tandis que dans les HeLa, son effet est plus important.

L'ensemble de ces résultats indique que α et Ct sont les deux régions transactivatrices de ERM.

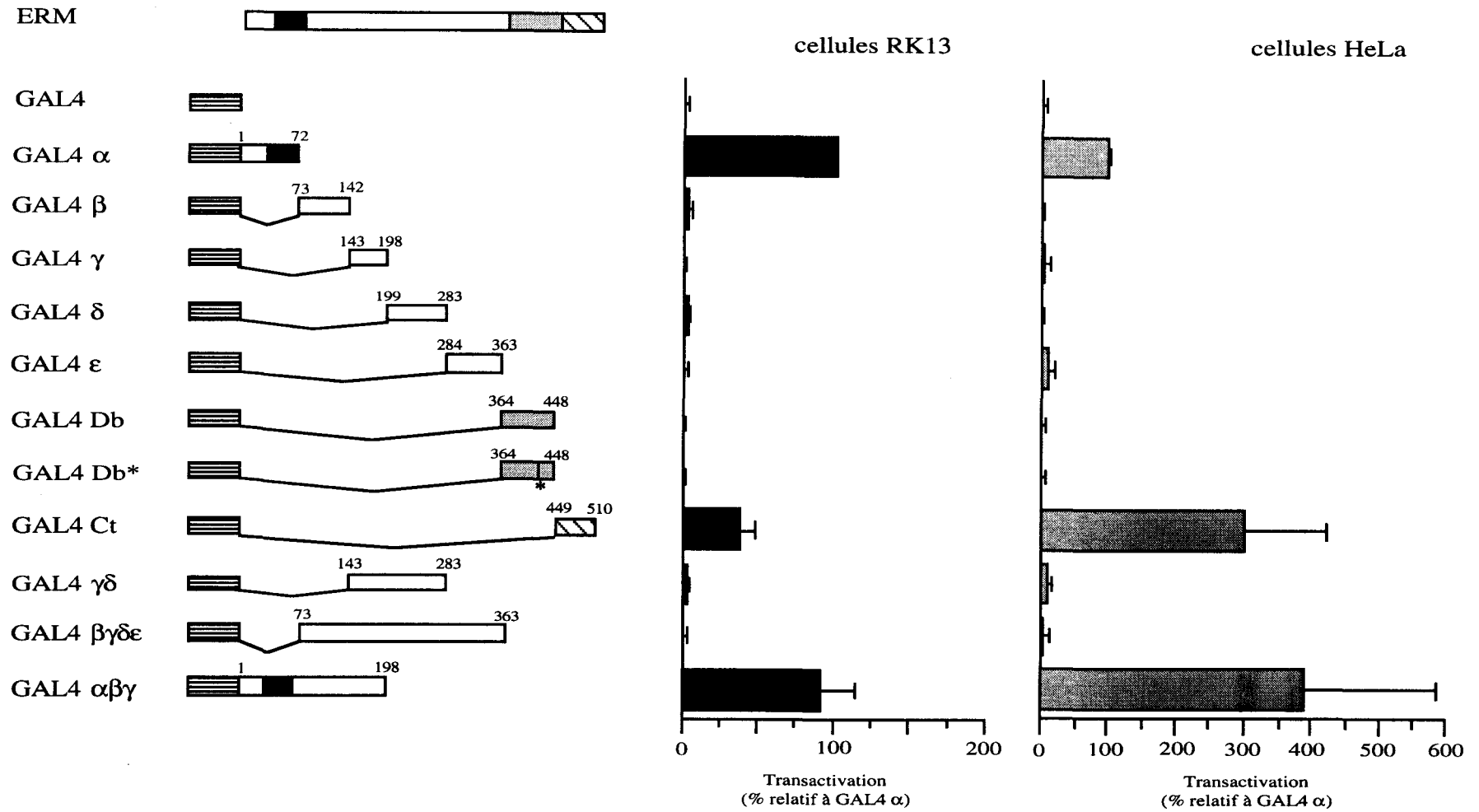


Figure 45 : Identification des régions de ERM portant une capacité de transactivation.
La transactivation est calculée par rapport à celle de GAL4 α .

3.3. Délimitation des domaines minimaux nécessaires à la transactivation par ERM.

Nous avons désiré déterminer les domaines transactivateurs avec plus de détail. Il était en particulier intéressant de préciser l'importance de la zone conservée présente dans α (le domaine acide, nommé aD) ou dans Ct (le peptide de 20 résidus situé juste en aval du domaine ETS) (Fig. 21). En vertu de sa richesse en résidus acides, aD (acides aminés 42 à 72) s'imposait comme un domaine transactivateur de choix. La chimère GAL4 aD exerce respectivement 10% et 30% de l'activité de GAL4 α dans les RK13 et dans les HeLa. Bien que ces valeurs soient notablement plus élevées que celles obtenues pour des fragments ne transactivant pas du tout, on peut considérer que aD, bien qu'actif, n'est pas responsable de la totalité de l'effet transactivateur dû à α (Fig. 46).

Pour l'étude de la queue carboxy-terminale, nous avons utilisé GAL4 Ct 449-468, la région 449 à 468 correspondant strictement à la zone des 20 résidus très conservés au sein du groupe PEA3 (Fig. 42); GAL4 Ct 469-510 englobe la région la moins conservée de Ct alors que GAL4 Ct 449-489 recouvre les 20 résidus conservés ainsi que la première partie de l'extrémité de Ct (Fig. 46). Alors que GAL4 449-468 et GAL4 449-489 n'ont aucun effet, la transactivation par GAL4 469-510 correspond respectivement à 20% et 40% de l'effet de GAL4 Ct dans les RK13 et les HeLa. La conclusion à tirer est par conséquent la même que dans le cas de GAL4 aD, en l'occurrence que GAL4 Ct 469-510 constitue la partie transactivatrice de Ct qui, pour un fonctionnement optimal requiert la présence de résidus compris dans le fragment 449-468.

Cependant on ne peut pas, pour GAL4 aD comme pour GAL4 Ct 469-510, exclure l'hypothèse que les fragments de ERM, tout en possédant un pouvoir transactivateur, sont de trop petite taille pour avoir la possibilité d'exercer correctement leur fonction.

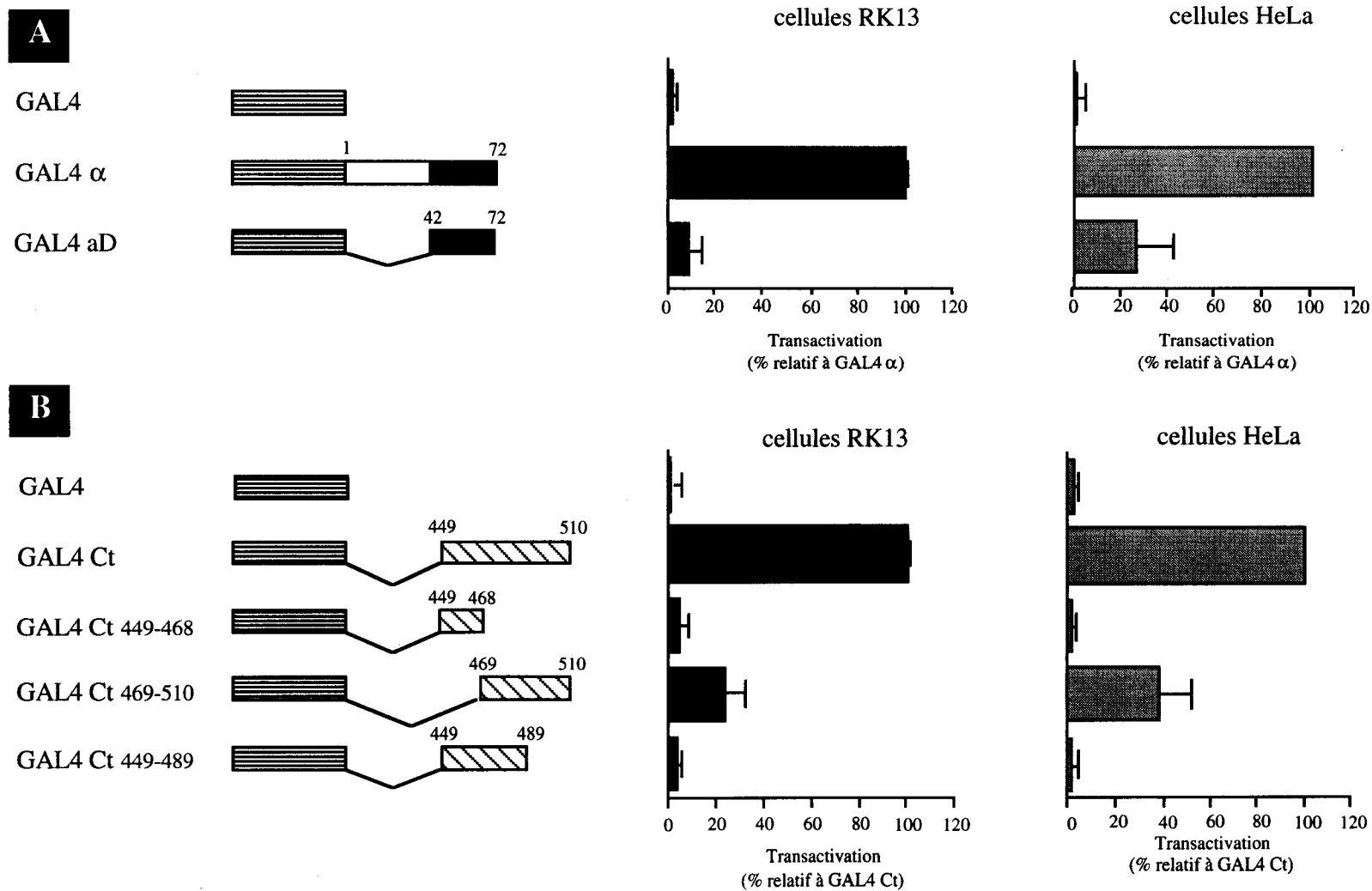


Fig. 46 : Délimitation des domaines transactivateurs minimaux.

3.4. VP16 "sqelche" certains facteurs nécessaires au fonctionnement de α mais non à celui de Ct.

Le fait que α contienne une région acide et soit impliqué dans la transactivation par ERM nous a encouragés à rechercher une similarité avec d'autres domaines transactivateurs. La figure 47 montre une identité de séquence entre une région de α qui recouvre en partie le domaine acide aD et la région d'activation de la transcription de la protéine VP16 du virus de l'herpes (Cress *et al.*, 1991). Cette identité s'étend sur un peptide court mais s'appuie sur la conservation de résidus acides et de résidus hydrophobes. La possibilité s'est alors offerte que cette conservation reflète une similarité des mécanismes de transactivation par α et par VP16.

Un des moyens envisagés pour valider cette supposition est l'expérience du "sqelching". Si deux protéines X et Y transactivent selon le même processus, elles sont censées utiliser des cofacteurs identiques. Si ces derniers sont présents en quantité limitée, alors l'ajout de Y dans un essai destiné à tester les capacités transactivatrices de X (en présence d'un vecteur rapporteur contenant les sites reconnus par X) diminue l'effet observé, puisque Y recrute les cofacteurs sans pouvoir exercer de transactivation. Nous avons coexprimé GAL4 α avec des quantités croissantes de domaine transactivateur de VP16 (Fig. 47). Le "sqelching" par VP16 est drastique : lorsqu'on cotransfecte 200 ng de pSV VP16 et 400 ng de GAL4 α , on note une diminution de l'effet de transactivation par GAL4 α de plus de 90%. Le contrôle négatif de l'expérience est GAL4 Ct, Ct ne partageant aucune identité de séquence avec VP16. L'effet de "sqelching" par VP16 est donc spécifique de la transactivation exercée par GAL4 α et suggère que α et VP16 interagissent avec un ou plusieurs cofacteurs communs et présents en quantité limitée dans le système que nous avons employé.

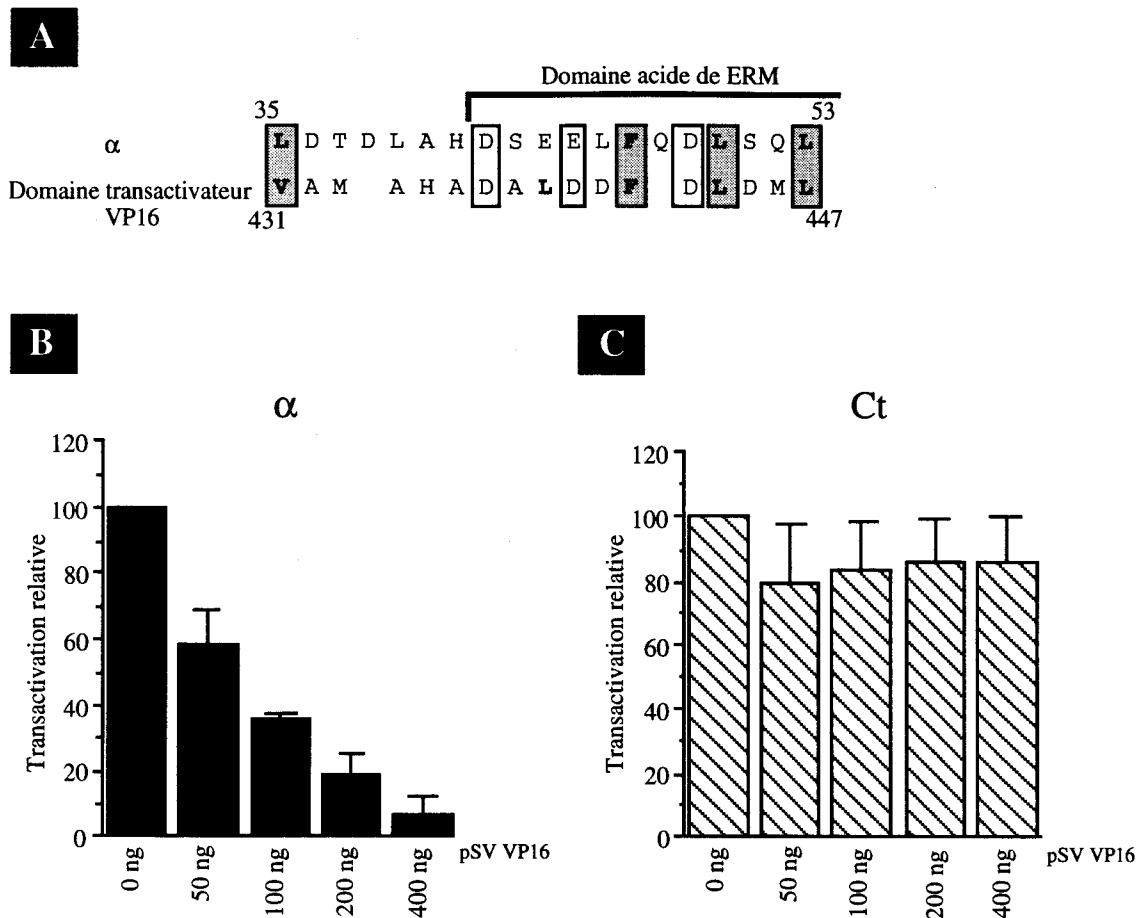


Figure 47 : Le domaine transactivateur de VP16 "squelche" l'effet transactivateur de α mais pas celui de Ct.

(A) Alignement entre le domaine transactivateur de VP16 et une partie de α . Les acides aminés entourés correspondent aux résidus conservés. Les acides aminés importants pour la fonction de VP16 sont dans les cadres, en caractères gras.

(B) et (C) Les transfections ont été faites en RK13. La construction pSV VP16 a été cotransfectée en quantité croissante avec GAL4 α (B) ou GAL4 Ct (C). Les taux de transactivation sont exprimés par rapport à celui de GAL4 Ct (C) ou en (B) GAL4 α .

3.5. Synergie entre α et Ct.

Puisqu'ils constituent les deux domaines transactivateurs d'un même facteur de transcription, il nous a semblé logique de nous intéresser à une synergie possible entre α et Ct. Lorsqu'on coexprime GAL4 α et GAL4 Ct en quantités égales (deux fois 400 ng), on constate que l'effet de transactivation observé est environ cinq fois plus élevé que pour GAL4 α seul (Fig. 48). Ceci laisse penser qu'il y a synergie entre les deux domaines activateurs de la transcription de ERM mais qu'elle est faible.

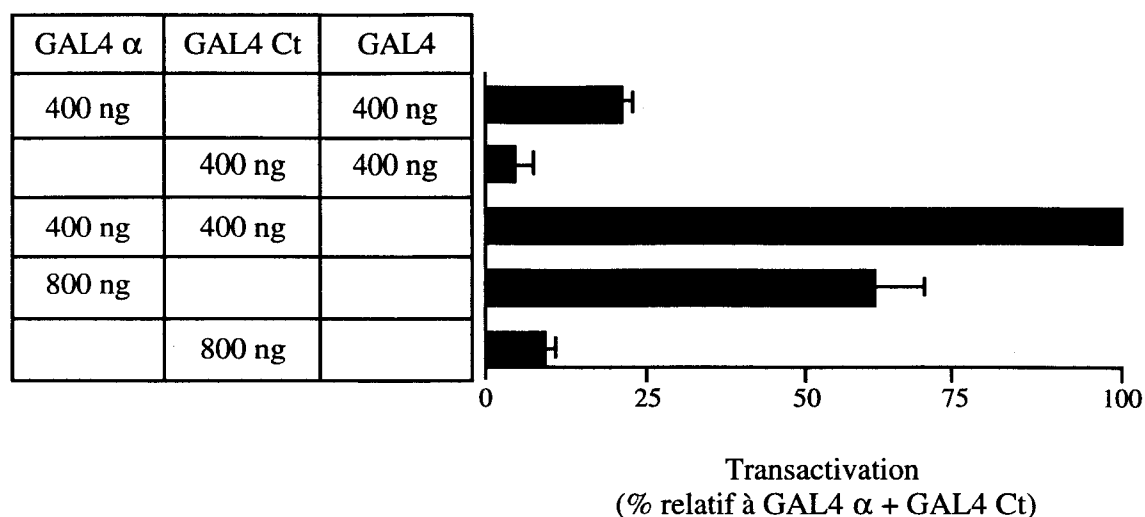


Figure 48 : Effet de synergie entre les deux domaines transactivateurs de ERM. Les transfections ont été réalisées en RK13 avec 100 ng de vecteur rapporteur.

Afin de conforter ce résultat, nous avons fusionné α et Ct à un domaine hétérologue de liaison à l'ADN différent, en l'occurrence celui, nommé Rev, du proto-oncogène *c-erbA*. Les couples Rev α -GAL4 Ct et GAL4 α -Rev Ct ont été testés sur un rapporteur contenant des sites GAL4 et des sites reconnus par les chimères contenant le domaine de liaison à l'ADN Rev. La coexpression de GAL4 α et Rev Ct conduit à un effet de transactivation environ cinq fois plus important que celui induit par GAL4 α seul. Ces résultats confirment que l'effet de synergie observé n'est pas dépendant de la nature du domaine de liaison à l'ADN qui est fusionné aux différents fragments ERM.

DISCUSSION

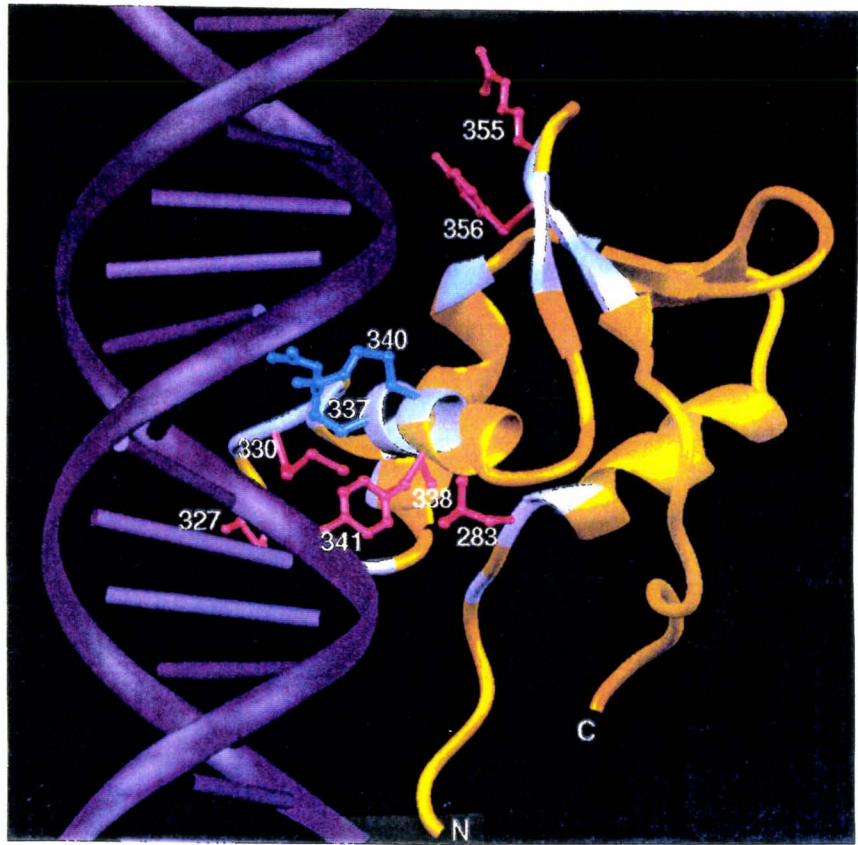
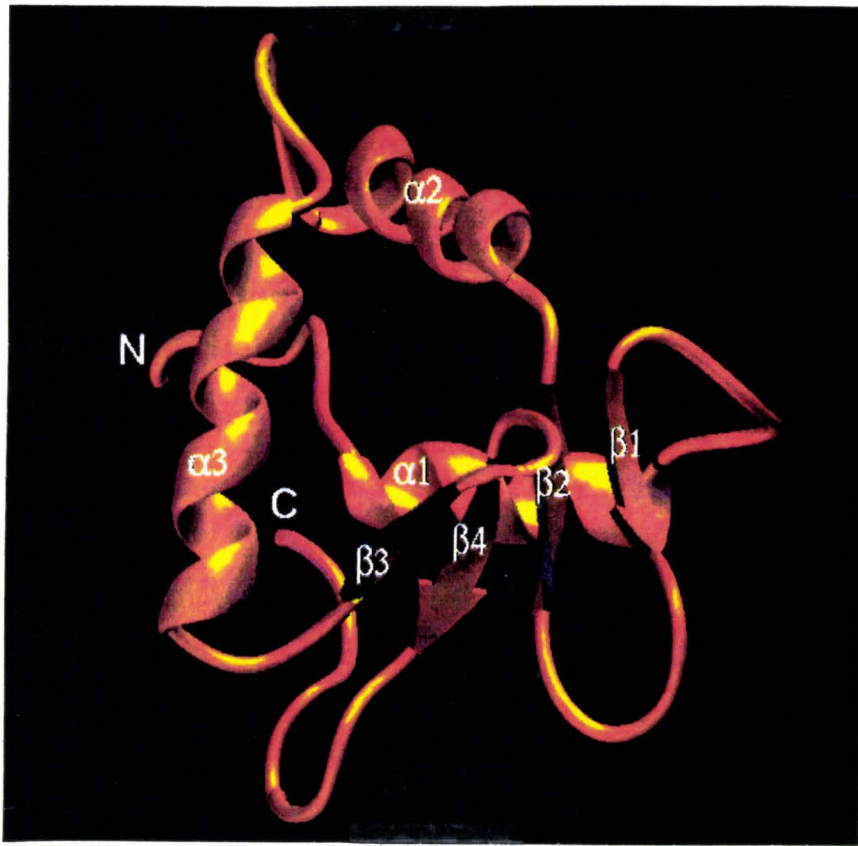


Figure 49 : Représentation schématique du domaine ETS de Fli-1 en solution (en haut) ou en complexe avec un oligonucléotide contenant sa séquence cible (en bas) selon Liang *et al.* (1994b).

1. Les relations délicates entre structure et prédiction de structure.

1.1. Les charmes dévoilés du domaine ETS de Fli-1.

Fin 1994, plus d'un an après la parution de l'article dans lequel nous proposons un modèle d'organisation spatiale du domaine ETS, des données sur la structure secondaire du domaine ETS ont été publiées. La RMN a permis de montrer que le domaine ETS de Fli-1 (Liang *et al.*, 1994a) et de c-Ets-1 (Donaldson *et al.*, 1994) étaient tous deux composés de trois hélices α et de quatre brins β organisés séquentiellement en $\alpha 1-\beta 1-\beta 2-\alpha 2-\alpha 3-\beta 3-\beta 4$. Le domaine ETS de Elk-1 est aussi susceptible de partager ces caractéristiques (Shore *et al.*, 1995). En effet, les techniques combinées de dichroïsme circulaire, de RMN et de spectroscopie de fluorescence montrent que le domaine ETS de Elk-1 comprend à la fois des brins β et des hélices α et que les trois tryptophanes conservés participent à la formation du noyau hydrophobe. L'obtention de la structure en solution du domaine ETS de Fli-1 en complexe avec sa séquence oligonucléotidique cible a montré que les quatre brins s'organisaient en un feuillet anti-parallèle et a désigné le couple d'hélices $\alpha 2$ et $\alpha 3$ comme le motif hélice-coude-hélice, $\alpha 3$ jouant le rôle d'hélice de reconnaissance (Liang *et al.*, 1994b) (Fig. 49). L'organisation générale de la structure tridimensionnelle du domaine ETS place celui-ci parmi les protéines à hélice-coude-hélice type papillon : l'hélice de reconnaissance $\alpha 3$ constitue le thorax et le reste du domaine ETS présente l'aspect général de deux ailes de papillon, la structure la plus proche de celle du domaine ETS étant celle de la protéine CAP (Fig. 13). Cependant, ses caractéristiques tridimensionnelles ne permettent pas d'identifier le domaine ETS comme appartenant à une famille structurale plus précise. En effet, il se rapproche du motif canonique hélice-coude-hélice par l'angle formé entre $\alpha 2$ et $\alpha 3$ mais l'hélice $\alpha 2$ est ornée d'un coude supplémentaire à son extrémité carboxy-terminale ce qui constitue une particularité des protéines POU.

Le fait que le domaine ETS partage des caractéristiques de différentes familles structurales a sans doute participé à la difficulté d'obtenir une prédiction de structure correcte.

1.2. L'exercice subtil et périlleux de la prédiction de structure.

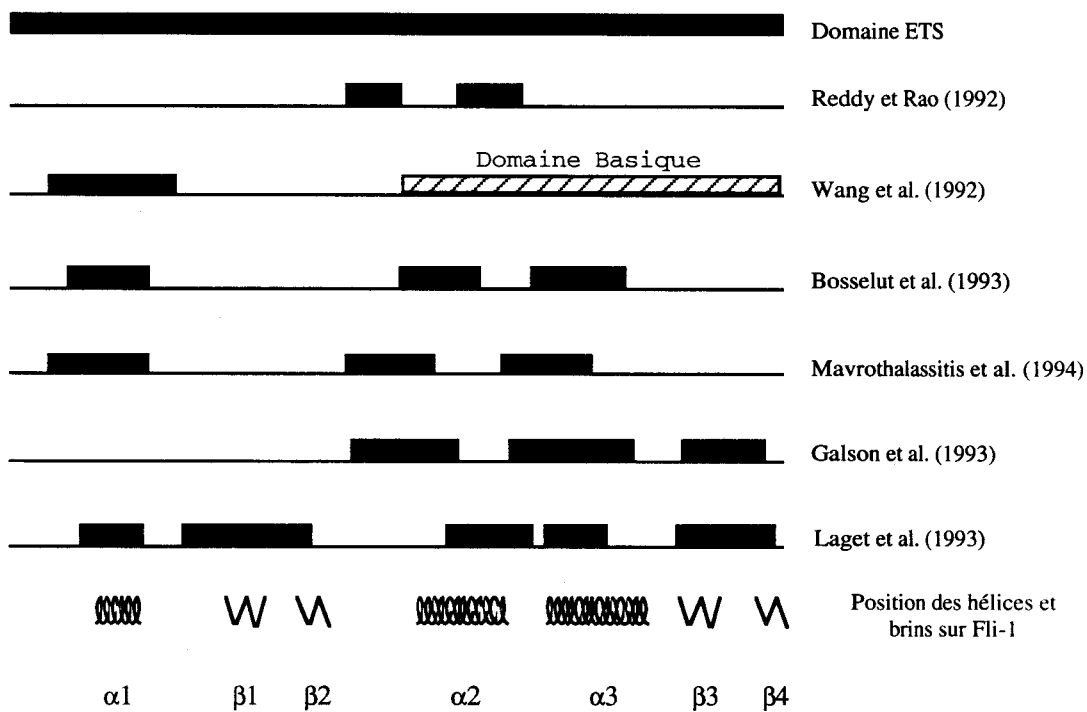


Figure 50 : Comparaison des différentes prédictions de structure proposées pour le domaine ETS et des données de RMN (Liang *et al.*, 1994b).

La comparaison des prédictions réalisées par notre équipe et dans d'autres laboratoires avec les résultats obtenus par la RMN met en avant le fait qu'aucun modèle d'organisation du domaine ETS n'est correct, la prédiction la plus proche restant celle de Bosselut *et al.* (1993) (Fig. 50). Cependant, des différentes méthodes utilisées, un consensus semble se dégager qui identifie correctement l'hélice $\alpha 1$ tandis que les deux dernières hélices sont détectées avec moins de précision en ce qui concerne tant leur position que leur longueur.

La constatation qu'aucune des méthodes utilisées n'a permis de détecter la présence des brins mérite réflexion. L'idée de base de la prédiction de structure est que toute l'information nécessaire au repliement correct d'une protéine est contenue dans sa formule développée : sa séquence en acides aminés. Tout le défi réside dans le décriptage de ce code qui permettra d'attribuer une organisation tridimensionnelle à une séquence en acides aminés particulière. La structure tridimensionnelle d'une protéine peut être prédite assez précisément lorsqu'il existe une protéine homologue dont l'organisation tri-dimensionnelle est connue (c'est le principe de la modélisation par homologie). Malheureusement, la structure tridimensionnelle de moins de 1% des séquences recensées dans les banques a été établie, et pour 80% des séquences, il n'existe pas d'homologue dont la structure soit connue (Rost *et al.*, 1994). Pour ces protéines ou peptides, tel le domaine ETS, toute tentative de prédiction de structure cause des difficultés pratiquement insurmontables ainsi que l'illustre le nombre (une bonne trentaine) et la variété des méthodes de prédictions de structure développées au cours de ces vingt dernières années. L'approche consiste généralement à simplifier le problème en fournissant non une prédiction de l'organisation tridimensionnelle mais une suite ordonnée de structures secondaires, pour laquelle chaque résidu se voit attribuer le caractère "brin", "hélice" ou "boucle". Or deux structures tridimensionnelles identiques n'ont pas toujours des structures secondaires identiques : la variation moyenne tourne autour de 12% (Rost *et al.*, 1994), la précision d'une prédiction de structure se voit donc "naturellement" restreinte. L'origine de cette variation réside dans le fait que la formation d'une structure secondaire (c'est-à-dire locale) est dépendante de deux composantes, l'une déterminée par des interactions locales et l'autre relevant d'interactions globales. L'importance de chaque composante varie selon la structure secondaire considérée. Les brins n'existent que comme partie d'un feuillet β , correspondent à une structure qui fait intervenir de manière plus conséquente des interactions globales (c'est-à-dire entre des parties éloignées d'une même protéine) et sont par conséquent moins faciles à prédire que les hélices. Ceci explique sans doute

que les quatre brins présents dans le domaine ETS n'aient été mis en évidence par aucune méthode.

1.3. Le cas particulier de notre prédiction de structure.

Les techniques de prédiction de structure telles que celle de Chou et Fasman s'appuient sur l'attribution du caractère brin, hélice ou boucle à chaque acide aminé considéré dans une fenêtre peptidique d'environ six résidus consécutifs. Un biais commun aux méthodes de prédiction de structure est donc qu'elles assimilent une protéine à une suite de peptides se recouvrant les uns les autres alors que la structure de chaque peptide peut varier selon qu'on le considère isolé ou dans le contexte de la protéine de pleine longueur. La méthode HCA utilisée pour la prédiction de structure se base sur la détection d'une identité de séquence faible par la comparaison de la taille, l'orientation et la forme des amas hydrophobes du domaine ETS et d'une protéine de structure connue. Logiquement, aucun biais dû à la sous-estimation des contraintes imposées sur la structure secondaire par la structure tridimensionnelle n'est introduit. Il est pourtant tentant de faire le lien avec les résultats récemment obtenus à l'Institut Pasteur de Lille par B. Odart et G. Lippens. Un peptide recouvrant les 37 acides aminés correspondant au fragment du domaine ETS qui, selon Mavrothalassitis *et al.* (1994), est doué d'une activité de liaison à l'ADN a été synthétisé. En présence de 30% de trifluoroéthanol, ce peptide présente deux centres d'hélices correspondant aux hélices $\alpha 2$ et $\alpha 3$ (Fig. 51).

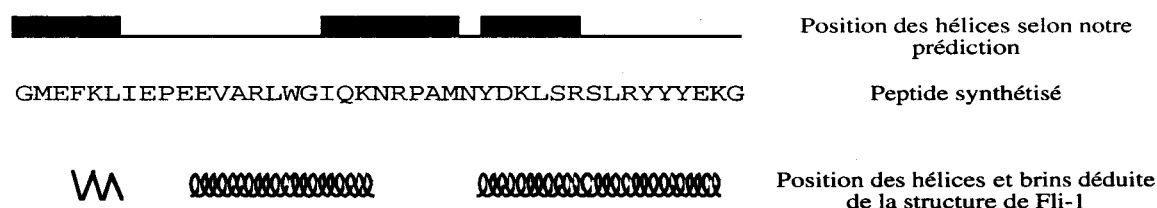


Figure 51 : Comparaison de la structure secondaire d'un peptide du domaine ETS avec la structure tertiaire obtenue sur le domaine ETS entier et avec notre prédiction de structure. Le peptide synthétisé (un fragment de PEA3) est l'équivalent du peptide défini par Mavrothalassitis *et al.* (1994) sur c-Ets-1.

En revanche, les sept premiers résidus, GMEFKLI, qui, selon les données de RMN obtenues sur le domaine ETS de Fli-1, recouvrent le brin $\beta 2$ situé en amino-terminal de l'hélice $\alpha 2$, ne sont pas, dans le peptide, organisés en brin mais en hélice, comme le suggérait notre prédiction de structure. Cette information laisse penser que la prédiction de structure réalisée avec l'aide d'HCA "voit" le domaine ETS plus comme une suite de peptides qui se recouvrent que comme une partie de la protéine, sans que l'on puisse proposer une explication.

Tenter de prédire la structure d'un domaine de liaison à l'ADN impose une contrainte supplémentaire : l'association du domaine de fixation avec son élément de reconnaissance nucléotidique peut avoir des conséquences sur la structure de chaque partenaire. Cependant, il a récemment été montré que le domaine ETS ne possédait pas la propriété de courber l'ADN (Shore et Sharrocks, 1995). De plus, la formation du complexe domaine ETS-séquence cible ne semble pas avoir de répercussions sur la structure du peptide engagé, ni stabiliser une zone flexible, ainsi que l'ont montré les études de RMN (Liang *et al.*, 1994b).

D'autres aléas inhérents à la prédiction de structure sur un domaine de liaison à l'ADN ont gêné notre travail. D'une part, les domaines de liaison à l'ADN ne dépassent généralement pas la taille de 100 résidus, ce qui rend les tests de validation statistique des alignements de séquence plus sensibles au biais introduit par une identité de séquence aléatoire. Ce point est détaillé plus loin. D'autre part, les domaines de liaison à l'ADN, par essence voués à la fixation d'un polynucléotide chargé électriquement, sont plus hydrophiles que la plupart des protéines. En conséquence, la place occupée par des acides aminés hydrophobes est moins importante. Dans le cas du domaine ETS, les amas hydrophobes sont nombreux et de taille réduite, ce qui ne facilite pas leur comparaison.

En revanche, la méthode HCA a été utilisée avec succès sur des protéines de taille et d'hydrophobicité plus importantes (Mornon *et al.*, 1989; Schoentgen *et al.*, 1992; Callebaut *et al.*, 1992; Callebaut *et al.*, 1994). Elle a par exemple permis la classification des cellulases en six familles et la localisation de leur site enzymatique

(Henrissat *et al.*, 1989; Henrissat *et al.*, 1991). La classification des cellulases en fonction de leur spécificité de substrat a confirmé la première prédiction et des mutations ponctuelles dans le site enzymatique putatif de l'endoglucanase EGZ a suggéré que l'histidine 98, sans intervenir dans la catalyse proprement dite était importante pour le repliement de l'enzyme tandis que la fonction carboxylique de l'acide glutamique 133 était impliquée dans la réaction (Py *et al.*, 1991). La méthode HCA a également permis de proposer un modèle selon lequel le domaine extracellulaire de liaison aux hormones des récepteurs aux cytokines/prolactine/hormone de croissance est organisé en deux sous-unités (résultant probablement d'une duplication) qui exhiberaient une similarité de structure avec les cytokines (Thoreau *et al.*, 1991). Ces données ont été pleinement confirmées par les résultats de RMN obtenus sur le domaine extracellulaire du récepteur à l'hormone de croissance humaine (de Vos *et al.*, 1992).

Une de nos erreurs a sans doute été d'accorder trop d'importance à la conservation des résidus hydrophobes et aromatiques en particulier. Pourtant, l'étude des structures tridimensionnelles de protéines appartenant à une même famille a mis en évidence un trait commun aux résidus impliqués dans la formation du noyau hydrophobe de la protéine : ils semblent plus conservés que les acides aminés hydrophobes périphériques (Lemesle-Varloot *et al.*, 1990). Cette caractéristique se reflète sur les diagrammes HCA qui sont topologiquement invariants alors que les acides aminés hydrophobes périphériques sont topologiquement moins stables. La comparaison des domaines de liaison des facteurs de transcription appartenant aux familles Ets, Myb et HMG a constitué une illustration parfaite de ce principe (Fig. 27 et 30). Alliée à la constatation que le domaine ETS présentait, à l'instar du domaine MYB, des résidus aromatiques régulièrement espacés, cette notion a servi de base à l'alignement des diagrammes HCA des deux domaines.

Pour aligner les domaines ETS, MYB et HMG, nous nous sommes basés sur la conservation des résidus aromatiques et sur celle des acides aminés les entourant, qu'ils soient hydrophobes ou non (Fig. 27 et 30). Les tests statistiques réalisés sur les alignements de séquence déduits de la méthode HCA se sont révélés significatifs. En

particulier, l'alignement Myb-HMG a été réalisé avec l'aide des données de RMN par la comparaison des structures tridimensionnelles des deux domaines, et le score obtenu, supérieur à 3 SD, nous a encouragés à considérer la présence des résidus aromatiques régulièrement espacés comme représentative d'une identité de structure entre les deux familles. En revanche, pour les alignements Ets-Myb et Ets-HMG, pour lesquels des scores statistiques élevés (supérieur à 6 SD) ont été obtenus, la présence de résidus régulièrement espacés relève plus d'une coïncidence que d'une caractéristique structurale commune. En effet, l'épreuve décisive de la comparaison des structures tridimensionnelles des domaines MYB (Fig. 26) et ETS (Fig. 49) invalide notre alignement.

C'est pour toutes les raisons énumérées dans ces deux derniers chapitres que la prédiction de structure que nous avons proposée après utilisation de la méthode HCA est partiellement fautive, même si nous avons détecté la présence des trois hélices.

1.3. Le mensonge des mutations moutonnes.

Afin d'asseoir notre modèle d'organisation spatiale du domaine ETS, nous avons décidé de muter certains acides aminés très conservés au sein de la famille Ets. Notre choix s'est porté sur les résidus aromatiques en vertu de leur conservation et de leur position par rapport aux hélices prédites par la méthode HCA, les tyrosines 418 et 442 étant respectivement censées se localiser sur les hélices H2 et H3. Les données obtenues par les expériences de gel-retard mettent en évidence que la tyrosine 418 est très sensible à la mutation, ce qui semble indiquer qu'elle joue un rôle plus important que la tyrosine 442 dont la mutation en phénylalanine, et dans une moindre mesure, en tryptophane ou en alanine est sans dommage pour la capacité de liaison à l'ADN du domaine ETS.

A défaut de connaître l'organisation tridimensionnelle du domaine ETS, ces résultats avaient conforté l'hypothèse de son identité structurale avec le domaine MYB, puisque les résidus équivalents de Myb réagissaient de manière identique (Fig. 52). En effet, le

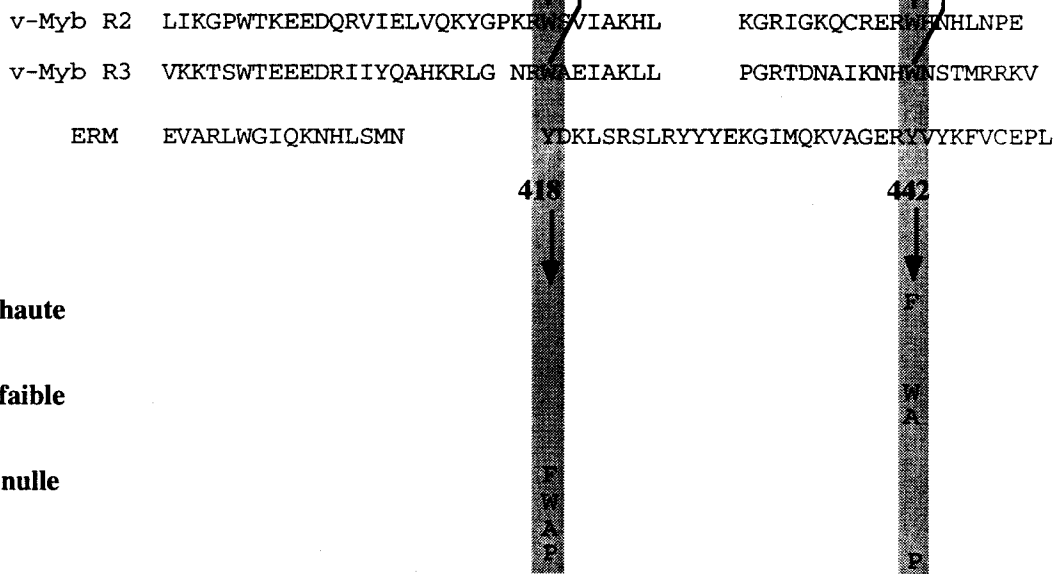
tryptophane occupant, selon notre alignement, une position équivalente à celle de la tyrosine 418 dans les séquences répétées R2 et R3, est sensible à la mutation en un autre résidu aromatique : la capacité de liaison à l'ADN est faible pour R2 ou totalement abolie pour R3. De plus, comme dans le cas du domaine ETS de ERM, la mutation du tryptophane équivalent de la tyrosine 442 touche de manière moins drastique la fonction de liaison à l'ADN.

**capacité de liaison
à l'ADN**

nulle

faible

haute



haute

faible

nulle

Figure 52 : Comparaison des mutations réalisées sur le domaine de liaison à l'ADN de v-Myb selon Frampton *et al.* (1991) et de celles effectuées dans notre groupe sur le domaine ETS de ERM. L'alignement des séquences est celui obtenu après comparaison des diagrammes HCA.

Compte-tenu de l'identité de séquence au sein du domaine ETS, il est raisonnable de penser que l'organisation tridimensionnelle du domaine ETS de ERM est très proche de celle du domaine ETS de Fli-1. La tyrosine 418, équivalent de la tyrosine 332 de Fli-1,

s'intégrerait donc au début de l'hélice α_3 qui est l'hélice de reconnaissance de l'ADN. Sa position à un emplacement stratégique lui confère la grande sensibilité vis-à-vis de la mutation que nous avons observée. Le cas de la tyrosine 442 est plus épineux : la tyrosine 356 qui est son équivalent dans Fli-1 se localise au début du brin β_4 et est en contact avec l'ADN. Ceci suggère qu'une mutation en alanine, proline, tryptophane ou phénylalanine, affectant dans tous les cas la charge du résidu considéré et probablement l'établissement de ponts hydrogènes avec le polynucléotide, devrait modifier la capacité du domaine ETS muté à fixer l'ADN. Manifestement, la tyrosine 442 est peu sensible à cette perturbation puisque les capacités de liaison à l'ADN du domaine ETS sont préservées quand la tyrosine 442 est mutée en alanine, tryptophane et phénylalanine. On peut donc supposer que la structure du domaine ETS de ERM diffère de celle de Fli-1 dans la partie carboxy-terminale et qu'elle varie de manière assez conséquente pour que la mutation de la tyrosine 442 n'affecte que peu sa capacité de liaison à l'ADN. Une alternative est que, sur ERM, la modification de structure imposée par la mutation est compensée par le reste du peptide et préserve ainsi les aptitudes du domaine ETS à lier l'ADN. Quant au double mutant Y418W/Y442W qui garde la capacité de lier l'ADN alors que la mutation Y418W l'abolit totalement, on peut supposer que l'effet de la mutation Y442W contrebalance celui de la mutation Y418W.

Les données de la RMN ont permis d'éclaircir certains effets de mutations qui pouvaient sembler surprenants en partant de l'hypothèse de l'importance des résidus aromatiques conservés. Par exemple, le second tryptophane du domaine ETS, parfaitement conservé (Fig. 2, page 6), peut être muté en leucine sans que cela affecte les capacités de liaison à l'ADN du domaine ETS de c-Ets-1 (Wang *et al.*, 1992). La structure tridimensionnelle montre que ce tryptophane est situé dans le brin β_1 , loin de l'ADN. En revanche, le troisième tryptophane situé dans l'hélice α_2 , plus proche de l'ADN ne peut être muté en arginine sans dommage (Wang *et al.*, 1992).

En résumé et en guise de morale, il faut garder à l'esprit que le gel-retard permet d'évaluer l'importance d'un acide aminé mais que le rôle exact joué par un résidu dans l'organisation tridimensionnelle d'une protéine ne peut être déterminé que par la RMN

ou la cristallographie. L'exemple est édifiant dans notre cas : les tyrosines 418 et 442 se comportent en gel-retard de manière sensiblement identique à leurs équivalents respectifs dans le domaine MYB (selon notre étude). Cependant, les causes de ces effets sont différentes et ces acides aminés n'occupent pas des positions similaires dans la structure spatiale des domaines ETS et MYB.

2. Heurs et malheurs du domaine ETS.

L'utilisation d'un éventail de constructions de délétion à partir des extrémités amino- et carboxy-terminale a rendu possible la délimitation de deux régions de ERM qui entravent l'activité de liaison à l'ADN. La fixation à l'ADN est diminuée par un domaine situé en amino-terminal du domaine ETS, entre les acides aminés 203 et 290 (ADID), et de la queue carboxy-terminale. De plus, l'effet inhibiteur exercé par la queue carboxy-terminale (comprenant 60 résidus) peut être levé par la suppression des 42 derniers acides aminés. Il est intéressant de constater que cette délétion recouvre entièrement une zone qui est peu conservée au sein du groupe PEA3. Des résultats préliminaires indiquent qu'un autre membre du groupe PEA3, PEA3 humain, est également soumis à un mécanisme d'inhibition de la liaison à l'ADN par la queue carboxy-terminale. Il reste à déterminer s'il dépend de la région la moins conservée de la queue carboxy-terminale. Si tel était le cas, on peut imaginer un scénario selon lequel cette région n'exercerait pas une fonction partagée par l'ensemble des membres du groupe PEA3 mais au contraire permettrait l'établissement d'une capacité de liaison variable selon le représentant du groupe PEA3.

Il convient enfin de signaler que nous n'avons pas détecté d'identité de séquence entre les différentes régions responsables du blocage de la liaison à l'ADN des protéines Ets et leurs analogues dans ERM. Cependant, la présence de plusieurs domaines d'inhibition de la liaison à l'ADN est un phénomène qui a déjà été observé pour c-Ets-1. Le domaine défini par Lim *et al.* (1992) (acides aminés 207-280 pour le produit p68 de c-Ets-1 de poulet) et la queue carboxy-terminale semblent interagir pour maintenir la

protéine dans une conformation fermée (Fig. 5 page 15). Un troisième domaine codé par l'exon VII possède apparemment un effet inhibiteur plus important mais ne semble pas interagir directement avec la queue carboxy-terminale (Hahn *et al.*, 1994). Le modèle proposé par les auteurs accorderait une part considérable dans l'inhibition de la fixation à l'exon VII qui est supposé interagir directement avec le domaine ETS et en altérer la conformation (ce point reste à confirmer) tandis que l'interaction de la queue carboxy-terminale avec le domaine 207-280 constituerait un verrou supplémentaire mais accessoire dans le maintien de la conformation fermée. Malgré la similarité d'organisation entre les domaines entravant la fixation à l'ADN chez c-Ets-1 et ERM, le parallèle entre ces deux membres de la famille Ets s'arrête là. En effet, pour ERM, l'activité inhibitrice de chaque région ne dépend pas de la présence de l'autre région : l'effet répresseur de la queue carboxy-terminale est observé, que les constructions testées contiennent la seconde région (ADID, résidus 203-290) ou non. Le mode d'action des deux domaines transiterait alors par une modification de la conformation du domaine ETS ou par une interaction avec des cofacteurs qui contacteraient chacune des régions délimitées, bloquant ainsi la liaison à l'ADN par encombrement stérique.

3. Les qualités cachées de ERM.

3.1. Le problème technique posé par l'utilisation de protéines de fusion.

Utiliser différentes formes délétées de ERM en vue d'identifier les domaines transactivateurs était délicat puisque, dans les conditions utilisées, il est impossible de discerner la part due à une modulation de la liaison à l'ADN de celle due à une variation du pouvoir transactivateur. Afin de délimiter les domaines transactivateurs, nous avons donc reprogrammé la spécificité de fixation à l'ADN des constructions ERM en les fusionnant au domaine de liaison à l'ADN de la protéine de levure GAL4.

De manière surprenante, alors que ERM a été défini comme facteur de transcription, la protéine de fusion GAL4 ERM s'est révélée inapte à exercer un effet transactivateur et

seule la délétion du domaine ETS de ERM a permis de lever cette incapacité. Les exemples répertoriés de fusion d'un membre de la famille Ets avec un domaine hétérologue ne permettent pas de fournir une explication générale. Dans le cas de c-Ets-1, la chimère GAL4 c-Ets-1 est douée d'activité de transcription, quelle soit dotée ou dépourvue du domaine ETS (Gégonne *et al.*, 1992). Ce comportement est spécifique de c-Ets-1 puisque des chimères GAL4 c-Ets-2 transactivent plus lorsqu'elles sont privées du domaine ETS (Chumakov *et al.*, 1993). En revanche, on note une différence de comportement de la même protéine lorsqu'elle est fusionnée à un autre domaine hétérologue de liaison à l'ADN, celui de LexA : privée de son domaine ETS, cette chimère transactive trois fois mieux (Schneikert *et al.*, 1992). Dans le cas de ERM comme dans celui de Elk-1 (Janknecht *et al.*, 1994) ou de Spi-1 (Hagemeier *et al.*, 1993), la chimère contenant la protéine de pleine longueur en fusion avec le domaine de liaison à l'ADN de GAL4 est incapable de transactiver mais l'activité de transcription est restaurée par la délétion du domaine ETS. Le comportement en transactivation des protéines Ets et l'effet de la suppression du domaine ETS ne sont donc pas prévisibles lorsque leur spécificité de fixation à l'ADN est reprogrammée par la fusion avec un domaine hétérologue de liaison à l'ADN.

La littérature fournit de nombreux exemples de facteurs de transcription qui, lorsqu'ils sont examinés sous forme de chimère avec le domaine de liaison à l'ADN GAL4, transactivent faiblement et dont l'activité d'augmentation de la transcription est restaurée par la délétion de leur domaine de liaison à l'ADN. C'est le cas de CRP2 (Williams *et al.*, 1995), de SRF (Johansen et Prywes, 1993) ou de TEF-1 (Hwang *et al.*, 1993). L'explication la plus fréquemment évoquée pour justifier les perturbations observées au niveau de la transactivation est la présence de deux domaines de liaison à l'ADN qui sont fonctionnels : le domaine de liaison "naturel" de la protéine de fusion va fixer des sites endogènes, diminuant ainsi la quantité de protéines de fusion qui se lient aux sites présents en amont du gène rapporteur et reconnus par le domaine de liaison artificiellement introduit dans la chimère. Cependant, puisqu'une mutation ponctuelle inhibant la capacité de liaison à l'ADN ne permet pas d'abolir l'effet perturbateur du

domaine ETS, cette hypothèse n'est pas valable dans le cas de ERM; la seule présence du domaine de fixation à l'ADN, qu'il soit muté ou non, est suffisante pour exercer cet effet. Il apparaît donc que lorsqu'il n'est pas lié à son site de reconnaissance, c'est-à-dire en l'absence de la cible naturelle ou lorsqu'il est muté, le domaine ETS exerce un effet inhibiteur sur la transactivation. Des résultats similaires ont été obtenus pour SRF : lorsque la protéine de pleine longueur est fusionnée à GAL4, elle ne transactive pas et l'activité d'augmentation de la transcription est restaurée non pas par des mutations dans le domaine basique qui abolissent la fonction de liaison à l'ADN mais par une délétion extensive dans le domaine de fixation à l'ADN (Johansen et Prywes, 1993; Hill *et al.*, 1994). Pour tous les facteurs de transcription qui, sous forme de protéine hybride avec GAL4, réagissent comme ERM et SRF, on peut imaginer un mécanisme général selon lequel le domaine de liaison à l'ADN, lorsqu'il n'est pas fixé sur son élément de réponse, bloque la transition structurale contemporaine de la reconnaissance de la séquence cible qui permet aux domaines transactivateurs d'exercer pleinement leurs fonctions (Weintraub *et al.*, 1991; Hill *et al.*, 1994; Williams *et al.*, 1995).

En résumé, il semble que la transactivation requière deux conditions : l'interaction entre le domaine transactivateur et le facteur potentiel qu'il reconnaît d'une part, et l'association entre le domaine de liaison à l'ADN et sa séquence cible d'autre part. Si l'une de ces conditions n'est pas remplie, comme c'est le cas pour les chimères GAL4 ERM pourvues du domaine ETS, la construction ne transactivera pas : toute perturbation du système qui modifie les interactions entre les trois partenaires abolit la transactivation.

3.2. ERM comme facteur de transcription bipartite.

En quête de domaines de transactivation au sein de la protéine ERM, nous avons examiné l'effet de la délétion de zones recouvrant les régions les mieux conservées. Celles-ci, susceptibles de porter une fonction importante, correspondent aux 72 premiers acides aminés de ERM (région nommée α), contenant le domaine

acide aD, très conservé, et à la queue carboxy-terminale (nommée Ct). Les tests de transfection en présence d'un vecteur rapporteur contenant des sites GAL4 ont permis de montrer que, contrairement à Ct, α est indispensable à la transactivation par la protéine de pleine longueur fusionnée à GAL4.

Des résultats préliminaires obtenus dans notre groupe infirmeraient à première vue ce résultat. En effet, Jean-Luc Baert a montré que le clone ER81 isolé dans notre groupe, dépourvu des 13 résidus carboxy-terminaux du domaine acide aD, exerce le même pouvoir transactivateur que la construction ETV1, clone "ER81" isolé par une autre équipe (Jeon *et al.*, 1995), qui contient la totalité du domaine acide (résultats non publiés). Il existe plusieurs explications possibles : premièrement, cette disparité dans les résultats obtenus peut relever de considérations techniques. Comme supposé dans le cas de C/EBP β (Williams *et al.*, 1995), des plasmides rapporteurs différents ont été utilisés. Ils contiennent soit des sites GAL4 (notre étude) soit des sites ETS (travail de Jean-Luc Baert) et, selon le système employé, la délétion d'une partie du domaine acide aurait des conséquences variables. La localisation du domaine transactivateur de SPI-1 est une illustration de ce phénomène (point détaillé dans l'introduction). Deuxièmement, la part de l'effet transactivateur exercé par aD est attribuable à la région de aD ne recouvrant pas les 13 derniers acides aminés et la délétion de cette région est sans conséquence sur le pouvoir transactivateur de ERM. Troisièmement, la zone responsable de la transactivation par α n'est pas le domaine acide aD mais la région située en amino-terminal de celui-ci; une construction dépourvue des 13 résidus carboxy-terminaux du domaine aD (telle ETV1) exercera le même effet transactivateur que ERM. Le fait que GAL4 aD ne transactive qu'à 10% (en RK13) ou 30% (en HeLa) de la valeur observée pour GAL4 α corrobore cette hypothèse. Une construction chimère recouvrant les 42 premiers acides aminés de α répondrait à cette interrogation. Enfin il ne faut pas écarter l'éventualité que chaque fragment (aD ou les 42 premiers acides aminés de ERM) peut être doué d'activité transcriptionnelle mais que la fusion avec GAL4 les place dans une conformation impropre à la transactivation.

La construction sous forme de protéine hybride avec GAL4, d'une série de fragments couvrant la totalité de ERM nous a permis de révéler, outre α , un autre domaine transactivateur : la queue carboxy-terminale (Ct). Signalons qu'en revanche la délétion de cette région est sans effet dans le contexte de la protéine de pleine longueur en fusion avec GAL4. En effet, GAL4 ERM-DbCt transactive à un niveau comparable à celui de GAL4 ERM-Db. Ces deux résultats sont néanmoins conciliables si l'on imagine que dans le contexte de GAL4 ERM, Ct joue le rôle d'un domaine transactivateur cryptique. L'interprétation du comportement de Ct lors d'études de transactivation sur une séquence reconnue par ERM (et non sur une cible GAL4) avec une construction ERM non fusionnée à GAL4, est sujette à discussion. Il semble que la délétion de Ct diminue l'effet transactivateur observé de deux fois (Yvan de Launoit, résultats non publiés). Ce résultat est à prendre en compte avec précaution puisqu'il est impossible dans ce système de faire la part des choses entre la diminution de l'effet transactivateur et l'augmentation de la fixation à l'ADN (rappelons que Ct constitue un domaine inhibiteur de la liaison à l'ADN). La présence d'un domaine transactivateur cryptique tel que Ct a déjà été mis en évidence pour le produit de l'épissage alternatif p68 de c-Ets-1. En effet, la délétion du domaine BEC a peu de répercussions sur l'effet transactivateur exercé par p68 sur un plasmide rapporteur contenant des sites Ets de type TORU ou HTLV-1 LTR. Cependant, deux éléments suggèrent que le domaine BEC joue un rôle important : premièrement, il est conservé dans au moins trois membres de la famille Ets (p68 c-Ets-1, c-Ets-2 et Pnt). Deuxièmement, lorsque le second domaine transactivateur de c-Ets-1 (RIII) est délété, la délétion du domaine BEC, qui est partie intégrante du domaine transactivateur RI, abolit l'effet transactivateur de RI (Albagli *et al.*, 1994).

Les capacités transactivatrices de Ct au sein du groupe PEA3 font l'objet de controverses. Il semble que chez PEA3 murin, un fragment de 164 résidus comprenant le domaine ETS soit suffisant pour l'activation de la transcription et que cet effet soit en particulier porté par la zone jouxtant la fin du domaine ETS, c'est-à-dire la région la plus conservée de la queue carboxy-terminale (Desjardins *et al.*, 1994). En opposition

avec ces résultats, il a été montré dans notre laboratoire, en utilisant le système GAL4, que la queue carboxy-terminale de PEA3 murin était dépourvue de toute activité de transcription (J-H Chen, communication personnelle). Ces données indiquent de nouveau que les données obtenues diffèrent selon le système utilisé et suggèrent que la position des domaines transactivateurs est variable d'un membre du groupe PEA3 à l'autre.

En dehors de α et Ct, aucune région de ERM ne semble douée de pouvoir transactivateur. Ni β (acides aminés 73 à 142) ni γ (acides aminés 143 à 198) n'augmentent la transcription, cependant nous avons noté que le fragment $\alpha\beta\gamma$ (acides aminés 1 à 198) produisait un effet beaucoup plus important que α dans les cellules HeLa. Plusieurs hypothèses peuvent être émises. D'une part, ceci laisse suggérer que $\beta\gamma$ contient un domaine transactivateur cryptique qui se replie plus facilement dans les cellules HeLa que dans les RK13, et que le découpage en différents fragments α , β , γ etc... l'a amputé d'une région essentielle. D'autre part, la région $\beta\gamma$ pourrait jouer un rôle identique au domaine RII de c-Ets-1 qui, démunie de capacité de transactivation, est néanmoins capable d'affecter positivement le domaine RI et négativement le domaine RIII (Schneikert *et al.*, 1992). Sans posséder de propriété transactivatrice, $\beta\gamma$ modulerait positivement le domaine α en recrutant une protéine exprimée dans les cellules HeLa et absente dans les cellules RK13. Un tel cofacteur serait également produit dans les cellules de rat FR3T3 pour lesquelles $\alpha\beta\gamma$ transactive aussi bien que α (résultats non exposés).

La présence au sein d'un même facteur de transcription de plusieurs domaines d'activation de la transcription est une propriété commune à de nombreux représentants de la famille Ets. C'est le cas de c-Ets-1 (Albagli *et al.*, 1994; Schneikert *et al.*, 1992), c-Ets-2 (Schneikert *et al.*, 1992; Chumakov *et al.*, 1993), Erg (Siddique *et al.*, 1993) et Fli-1 (Rao *et al.*, 1993). Nous n'avons noté aucune identité de séquence d'un quelconque de ces domaines avec α ni Ct. Cependant, aD avec 32% d'acides aspartiques ou glutamiques est un domaine acide au même titre que le domaine BEC de c-Ets-1 (4 résidus sur 13). Ct contient 21% d'acides aminés hydrophobes, 13% de

leucine et 10% de proline, alanine et sérine, on pourrait donc le qualifier de domaine à caractère hydrophobe-acide.

Notre étude a montré que α et Ct constituent deux domaines indépendants et ne correspondent pas à deux parties séparées d'un même domaine transactivateur. Ils sont capables d'agir en synergie, ce qui suggère qu'ils transactivent par des mécanismes différents, en impliquant sans doute soit des cofacteurs différents, soit différentes parties d'un même cofacteur. D'autres arguments permettent de penser qu' α et Ct n'interviennent pas de la même manière, en particulier le fait que le rapport des pouvoirs de transactivation Ct/ α est variable selon le type cellulaire : il est de 37,5% dans les cellules RK13, de 300% dans les cellules HeLa et de 75% dans les cellules FR3T3 (résultat non exposé). Enfin, tandis que VP16 peut "squelcher" la transactivation par GAL4 α , il est sans effet sur la transactivation exercée par GAL4 Ct.

Ces expériences de "squelching" permettent de proposer que α et VP16 possèdent un cofacteur commun présent en quantité limitée. Cependant, on ne peut éliminer l'hypothèse selon laquelle α , Ct et VP16 possèderaient un cofacteur commun présent en quantité non limitante. Dans ce cas, le domaine transactivateur de VP16, tout en séquestrant le cofacteur commun, n'aurait aucun effet visible puisque ce cofacteur est présent en quantité suffisante pour satisfaire les exigences de α , Ct et VP16. L'existence d'un cofacteur commun à ERM et à VP16, alliée à l'identité de séquence que nous avons révélée entre leurs domaines transactivateurs, suggère fortement que le domaine α de ERM et VP16 augmentent la transactivation par des mécanismes similaires. Cette hypothèse semble renforcée par les récents résultats que nous avons obtenus : le changement en proline du résidu équivalant à la phénylalanine 442 de VP16 (Fig. 47 abolit complètement la transactivation par α . Puisque la mutation de la phénylalanine 442 de VP16 produit le même effet, on peut supposer que l'acide aminé 47 de ERM et le résidu 442 de VP16 jouent des rôles similaires.

Très logiquement, on peut envisager que les cofacteurs de VP16 soient également ceux de ERM. ADA2 semble être le candidat de choix à ce poste pour plusieurs raisons. D'abord c'est un coactivateur de VP16 (Berger *et al.*, 1992; Silverman *et al.*, 1994).

Ensuite, ADA2, qui n'est pas un facteur essentiel à la croissance cellulaire, n'est sans doute pas un facteur de la machinerie transcriptionnelle (Berger *et al.*, 1992) et pourrait donc constituer un cofacteur présent en quantité limitée.

VP16 s'associe également avec TFIIB (Lin *et al.*, 1991), deux sous-unités de TFIID : TBP (Stringer *et al.*, 1990) et TAF_{II}40 (Goodrich *et al.*, 1993), TFIIF (Xiao *et al.*, 1994). Ces facteurs sont des candidats d'autant plus intéressants que certains membres de la famille Ets se sont révélés interagir avec des protéines de la machinerie transcriptionnelle de base. En effet des expériences de "pull-down" ont montré sans ambiguïté que SPI-1 était capable de s'associer directement avec TBP et que cette interaction requérait une région située dans la partie la plus conservée de TBP (résidus 202 à 272) et les acides aminés 1-75 du domaine transactivateur de SPI-1 (Hagemeyer *et al.*, 1993). De plus, GABP α , autre membre de la famille Ets, se lie à TBP sans qu'on ait déterminé si l'association était directe (T. Yoganathan, communication personnelle). Enfin, il a récemment été suggéré que PEA3 fixe TBP mais aussi l'ARN polymérase de type II par l'intermédiaire de deux peptides identiques localisés aux extrémités du domaine ETS (Desjardins *et al.*, 1994). La mise en évidence d'une interaction entre ERM et TFIIB, TFIID ou TAF_{II}40 placerait alors ERM parmi les facteurs de transcription qui, reconnaissant une séquence spécifique, interagissent avec le complexe d'initiation de la transcription et facilitent son association pour l'augmentation de la transcription d'un gène particulier (Roeder *et al.*, 1991).

CONCLUSION

Le but du présent travail a consisté en la caractérisation de la structure et de la fonction du facteur de transcription ERM, et plus particulièrement son domaine de liaison à l'ADN et ses domaines responsables de la transactivation. La définition des caractéristiques structurales du domaine ETS s'est réalisée à l'aide d'une méthode qui, bien que déjà appliquée avec succès dans d'autres cas, est susceptible d'engendrer des données éloignées de la "réalité biologique". En effet, la méthode de prédiction de structure HCA s'est révélée défailante au même titre que les autres méthodes de prédiction précédemment utilisées, et la structure du domaine ETS que nous avons proposée ne correspond que partiellement à la structure obtenue très récemment par RMN. Notons au passage que ces données de RMN doivent être considérées avec précaution puisque la caractérisation du domaine ETS n'a été réalisée que sur un fragment d'une centaine de résidus et ne prend donc pas en compte le contexte général de la protéine. En effet, si l'ensemble des membres de la famille Ets possèdent en commun le domaine ETS, d'autres domaines spécifiques de chaque protéine peuvent moduler la capacité de liaison à l'ADN de la protéine de pleine longueur et donc affecter de manière différentielle sa structure.

L'influence des mutations réalisées sur différents résidus du domaine ETS est variable en fonction de la nature de l'acide aminé introduit et diffère parfois du résultat escompté, ce qui souligne, si besoin était, que les lois qui régissent les interactions ADN-protéine sont multiparamétriques et que chaque acide aminé doit être examiné en fonction de son volume, de son hydrophobicité, de sa charge... A posteriori les effets engendrés par ces mutations sont difficilement interprétables d'autant plus qu'ils sont jugés par le biais d'une méthode indirecte : le gel-retard. Quoiqu'il en soit, les mutations testées fournissent aux investigateurs travaillant sur ces facteurs de

transcription des constructions d'intérêt dépourvues d'activité de liaison à l'ADN. De plus, toujours grâce aux méthodes de gel-retard, nous sommes en mesure de suggérer que dans nos conditions d'expérimentation la fonction de liaison à l'ADN du facteur de transcription ERM est influencée par deux domaines inhibiteurs : le domaine ADID (Amino-terminal DNA binding Inhibitory Domain, résidus 203 à 290) et la queue carboxy-terminale (Ct). Ces résultats obtenus par des techniques *in vitro* sont indicatifs mais n'ont pas la prétention d'être le reflet exact de la "réalité biologique".

En revanche, la caractérisation d'un domaine de transactivation au sein d'un facteur de transcription fait intervenir des méthodes qui se rapprochent plus de la "réalité biologique". En effet, ces essais sont réalisés dans un système cellulaire eucaryote produisant les différents intervenants de la machinerie basale de la transcription ainsi que les potentiels cofacteurs de ERM. Nous avons pu mettre en évidence la présence de deux domaines transactivateurs dans ERM : la partie amino-terminale (α) dont la présence est requise pour la fonction de transactivation et le domaine Ct qui, quoique non essentiel, joue un rôle significatif. En se basant sur l'ensemble des résultats acquis, nous sommes en mesure de présenter un modèle pour la fixation à l'ADN et la transactivation par ERM (Fig. 53). L'inhibition de la liaison à l'ADN exercée par le domaine ADID et par la queue carboxy-terminale est levée en réponse à un signal extracellulaire, sans doute par modification post-traductionnelle de ERM ou d'une protéine qui lui est associée (Fig. 53 A). Lorsque ERM est lié à son site de reconnaissance sur le promoteur du gène cible, la conformation qu'il a adoptée est propice à la transactivation et au recrutement des cofacteurs nécessaires à l'augmentation de la transcription par ERM (Fig. 53 B). α et Ct agissent en synergie et le (ou les) cofacteur(s) requis pour la transactivation peut (peuvent) être identique(s) à celui potentiellement nécessaire à l'inhibition de la liaison à l'ADN.

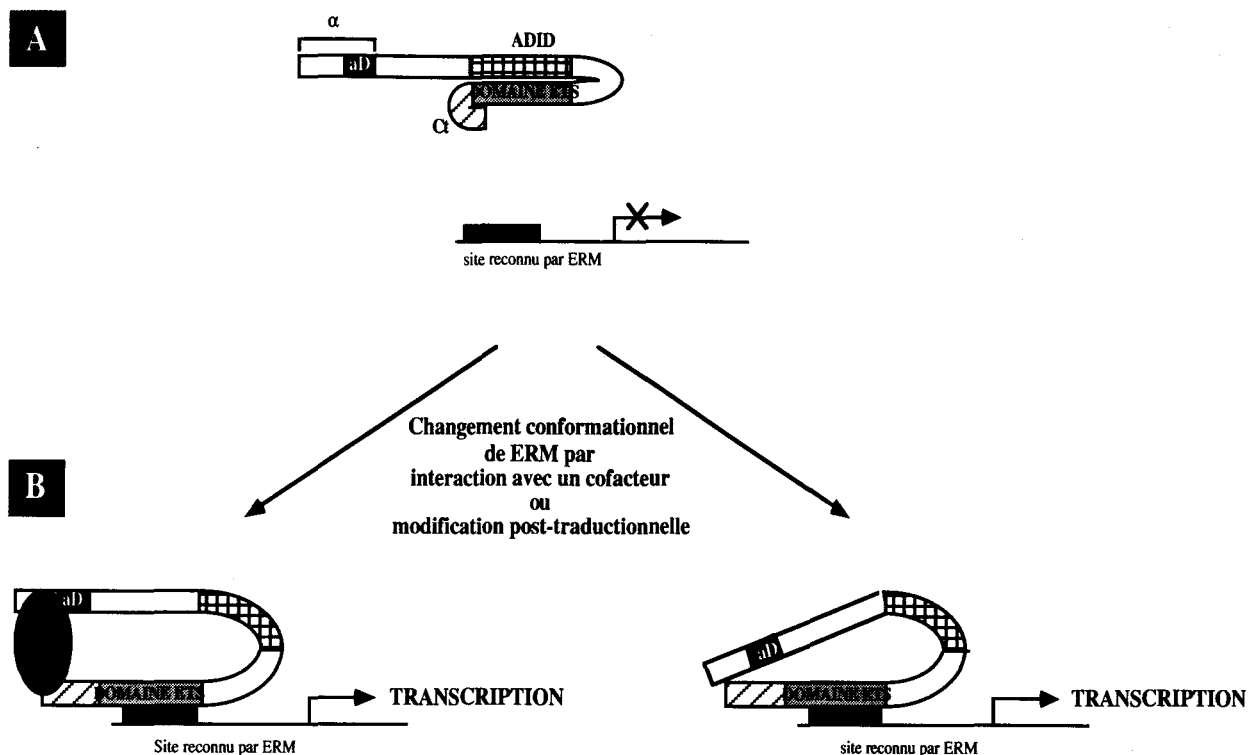


Figure 53 : Fonctionnement du facteur de transcription ERM : hypothèse.

Certains points de ce schéma sont pure spéculation. Plus précisément, aucune interaction, directe ou par l'intermédiaire d'une protéine annexe, n'a été mise en évidence entre α et Ct et nos résultats préliminaires ne permettent pas de trancher sur ce point. Quoique les cofacteurs de VP16 semblent des candidats potentiels, il nous faudra définir les protéines susceptibles d'interagir avec ERM. Cette démarche est déjà entreprise dans notre équipe par Pierre-Antoine Defossez et elle nous fournira de précieux renseignements sur l'implication de ERM dans le mécanisme de la transactivation. En effet la formation d'un complexe multiprotéique peut avoir des conséquences sur la stabilité des interactions ADN-protéine ainsi que sur la spécificité des protéines engagées dans un tel assemblage. La sélection d'une protéine donnée au sein d'une famille de facteurs de transcription qui reconnaissent des sites de fixation proches peut être augmentée par des contacts protéine-protéine multiples et par la nature coopérative de l'assemblage d'un complexe multiprotéique lié à l'ADN.

Il apparaît cependant que notre hypothèse reste un modèle qui a le mérite d'être proposé et par conséquent de faire l'objet de critiques et d'études complémentaires qui le confirmeront ou l'infirmont. Pour ce faire, d'autres techniques d'investigation, si possible plus proches de la "réalité biologique" seront utilisées. Les récentes découvertes mettant en avant le rôle fonctionnel similaire de facteurs de transcription chez les vertébrés et de leurs homologues chez la *Drosophile* voire chez le nématode *Caenorhabditis elegans* encouragent la communauté scientifique des "transcriptionnistes" à utiliser ces modèles moins complexes et surtout mieux définis génétiquement. C'est dans cette optique que j'ai choisi de travailler au cours de mon stage post-doctoral sur l'implication du facteur de transcription CREB dans le signal de transduction de Hedgehog chez la drosophile.

REFERENCES

1. **Anderson, W. F., D. H. Ohlendorf, Y. Takeda, and B. W. Matthews.** 1981. Structure of the *cro* repressor from bacteriophage λ and its interaction with DNA. *Nature* **290**:754-758.
2. **Anton, I. A., and J. Frampton.** 1988. Tryptophans in *myb* proteins. *Nature* **336**:719.
3. **Ascione, R., D. M. Thompson, R. Thomas, A. Panayiotakis, R. Ramsay, M. Tymms, I. Kola, and A. Seth.** 1992. Influence of the nucleotides flanking the -GGAA- core sequence on ETS1 and ETS2 DNA-binding and the mechanism of ETS1 autoregulation. *Int. Journ. Onc.* **1**:631-637.
4. **Ashburner, M., C. Chihara, P. Meltzer, and G. Richards.** 1974. Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **38**:655-662.
5. **Assa-Munt, N., R. J. Mortishire-Smith, R. Aurora, W. Herr, and P. E. Wright.** 1993. The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA-binding domain. *Cell* **73**:193-205.
6. **Aurigemma, R. E., D. G. Blair, and S. K. Ruscetti.** 1992. Transactivation of erythroid transcription factor GATA-1 by a *myb-ets*-containing retrovirus. *J. Virol.* **66**:3056-3061.
7. **Bailly, R. A., R. Bosselut, J. Zucman, F. Cormier, O. Delattre, M. Roussel, G. Thomas, and J. Ghysdael.** 1994. DNA-binding and transcriptional activation properties of the EWS-FLI-1 fusion protein resulting from the t (11;22) translocation in Ewing sarcoma. *Mol. Cell. Biol.* **14**:3230-3241.
8. **Ben-David, Y., and A. Bernstein.** 1991a. Friend virus-induced erythroleukemia and the multistage nature of cancer. *Cell* **66**:831-834.
9. **Ben-David, Y., E. B. Giddens, and A. Bernstein.** 1990. Identification and mapping of a common proviral insertion site *Fli-1* in erythroleukemia cells induced by Friend murine leukemia virus. *Proc. Natl. Acad. Sci. USA* **87**:1332-1336.

10. **Ben-David, Y., E. B. Giddens, K. Letwin, and A. Bernstein.** 1991b. Erythroleukemia induction by Friend murine leukemia virus : insertional activation of a new member of the ets gene family, *Fli-1*, closely linked to *c-ets-1*. *Genes Dev.* **5**:908-918.
11. **Berg, J. M.** 1989. DNA binding specificity of steroid receptors. *Cell* **57**:1065-1068.
12. **Berger, S. L., B. Pina, N. Silverman, G. A. Marcus, J. Agapite, J. L. Regier, S. J. Triezenberg, and L. Guarente.** 1992. Genetic isolation of ADA2 : a potential transcriptional adaptator required for function of certain acidic activation domains. *Cell* **70**:251-265.
13. **Bhattacharya, G., L. Lee, E. P. S. Reddy, and V. N. Rao.** 1993. Transcriptional activation domains of *elk-1*, Δ *elk-1* and SAP-1 proteins. *Oncogene* **8**:3459-3464.
14. **Billeter, M., Y. Qian, G. Otting, M. Müller, W. J. Gehring, and K. Wüthrich.** 1990. Determination of the Three-dimensional Structure of the *Antennapedia* Homeodomain from *Drosophila* in solution by ¹H Nuclear Magnetic Resonance Spectroscopy. *J. Mol. Biol.* **214**:183-197.
15. **Blood, C. H., and B. R. Zetter.** 1990. Tumor interactions with the vasculature : angiogenesis and tumor metastasis. *Biochim. Biophys. Acta* **1032**:89-118.
16. **Bosselut, R., J. F. Duvall, A. Gégonne, M. Bailly, A. Hémar, J. Brady, and J. Ghysdael.** 1990. The product of the *c-ets-1* proto-oncogene and the related Ets2 protein act as transcriptional activators of the long terminal repeat of human T cell leukemia virus HTLV-1. *EMBO J.* **9**:3137-3144.
17. **Bosselut, R., J. Levin, E. Adjadj, and J. Ghysdael.** 1993. A single amino-acid substitution of Ets1 to that of the related related transcription factors Elf1 and E74. *Nucl. Acids Res.* **21**:5184-5191.
18. **Braun, B. S., R. Frieden, S. L. Lessnick, W. A. May, and C. T. Denny.** 1995. Identification of target genes for the Ewing's sarcoma EWS/FLI fusion protein by representational difference analysis. *Mol. Cell. Biol.* **15**:4623-4630.

19. **Brennan, R. G.** 1993. The winged-helix DNA-binding motif : another helix)turn-helix takeoff. *Cell* **74**:773-776.
20. **Brody, L. C., K. J. Abel, L. H. Castilla, F. J. Couch, D. R. McKinley, G. Y. Yin, P. P. Ho, S. Merajver, S. C. Chandrasekharappa, J. Xu, J. L. Cole, J. P. Struewing, J. M. Valdes, F. S. Collins, and B. L. Weber.** 1995. Construction of a transcription map surrounding the BRCA1 locus of human chromosome 17. *Genomics*. **25**:238-247.
21. **Brown, T. A., and S. L. McKnight.** 1992. Specificities of protein-protein and protein-DNA interaction of GABP α and two newly defined *ets*-related proteins. *Genes Dev.* **6**:
22. **Brunner, D., K. Dücker, N. Oellers, E. Hafen, H. Scholtz, and C. Klämbt.** 1994. The ETS domain protein Pointed-P2 is a target of MAP kinase in the Sevenless signal transduction pathway. *Nature* **370**:386-389.
23. **Burtis, K. C., C. S. Thummel, C. W. Jones, F. D. Karim, and D. S. Hogness.** 1990. The *Drosophila* 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two *ets*-related proteins. *Cell* **61**:85-99.
24. **Callebaut, I., M. G. Catelli, D. Portetelle, X. Meng, F. Cadepond, A. Burny, E. E. Baulieu, and J. P. Mornon.** 1994. Redox mechanism for the chaperone activity of heat shock proteins HSPs 60, 70 and 90 as suggested by hydrophobic cluster analysis : hypothesis. *C. R. Acad. Sci. Paris* **317**:721-729.
25. **Callebaut, I., J. M. Renoir, M. C. Lebeau, N. Massol, A. Burny, E. E. Baulieu, and J. P. Mornon.** 1992. An immunophilin that binds Mr 90,000 heat shock protein: main structural features of a mammalian p59 protein. *Proc. Natl. Acad. Sci. USA* **89**:6270-6274.
26. **Ceska, T. A., M. Lamers, P. Monaci, A. Nicosia, R. Cortese, and D. Suck.** 1993. The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for the DNA binding. *EMBO J.* **12**:1805-1810.

27. **Chen, J. H., C. Vercamer, Z. Li, D. Paulin, B. Vandebunder, and D. Stehelin.** Submitted. PEA3 transactivates vimentin promoter in mammary epithelial and tumor cells.
28. **Chen, J. H., and C. D. Wright.** 1993. PEA3, Oct 1 and Oct 2 positively regulate ETS1 promoter. *Oncogene* **8**:3375-3383.
29. **Chumakov, A. M., D. L. Chen, E. A. Chumakova, and H. P. Koeffler.** 1993. Localization of the c-Ets-2 transactivation domain. *J. Virol.* **67**:2421-2425.
30. **Clark, K. L., E. D. Halay, E. Lai, and S. K. Burley.** 1993. Co-crystal structure of the HNF3/*forkhead* DNA-recognition motif resembles histone H5. *Nature* **364**:412-420.
31. **Coleman, K. G., S. J. Poole, M. P. Weir, W. C. Soeller, and T. B. Kornberg.** 1987. The *invected* gene of *Drosophila* : sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes Dev.* **1**:19-28.
32. **Cress, W. D., and S. J. Triezenberg.** 1991. Critical structural elements of the VP16 transcriptional activation domain. *Science* **251**:87-90.
33. **Dalton, S., and R. Treisman.** 1992. Characterization of SAP-1, a protein recruited by serum response factor to the *c-fos* serum response element. *Cell* **68**:597-612.
34. **de Vos, A., M. Ultsch, and A. A. Kossiaakoff.** 1992. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**:306-312.
35. **Degnan, B. M., S. M. Degnan, T. Nagagnuma, and D. E. Morse.** 1993. The *ets* multigene family is conserved throughout the Metazoa. *Nucleic. Acids. Res.* **21**:3479-3484.
36. **Dekker, N., M. Cox, R. Boelens, C. P. Verrijzer, P. C. van der Vliet, and R. Kaptein.** 1993. Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature* **362**:852-855.

37. **Delattre, O., J. Zucman, B. Plougastel, C. Desmaze, T. Melot, M. Peter, H. Kovar, I. Joubert, P. de Jong, G. Rouleau, A. Aurias, and G. Thomas.** 1992. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**:162-165.
38. **Desjardins, P., B. Brinton, J. H. Xin, and J. A. Hassel.** 1994. PEA3, an activator of transcription and DNA replication, interacts with TBP and RPA through common domains. Tenth annual meeting on oncogenes, Fredericks.
39. **Diffley, J. F. X., and B. Stillman.** 1991. A close relative of the nuclear, chromosomal high-mobility group protein HMG1 in yeast mitochondria. *Proc. Natl. Acad. Sci. USA* **88**:7864-7868.
40. **Donaldson, L. W., J. M. Petersen, B. J. Graves, and L. P. McIntosh.** 1994. Secondary structure of the ETS domain places murine Ets-1 in the superfamily of winged helix-turn-helix DNA-binding proteins. *Biochemistry* **33**:13509-13516.
41. **Dudek, H., R. V. Tantravahi, V. N. Rao, E. S. P. Reddy, and F. P. Reddy.** 1992. Myb and Ets proteins cooperate in transcriptional activation of the *mim-1* promoter. *Proc. Natl. Acad. Sci. USA* **89**:1291-1295.
42. **Duterque-Coquillaud, M., D. Leprince, A. Flourens, C. Henry, J. Ghysdael, B. Debuire, and D. Stéhelin.** 1988. Cloning and expression of chicken p54^{c-ets} cDNAs : the first p54^{c-ets} coding exon is located into the 40.0 kbp genomic domain unrelated to *v-ets*. *Oncogene Res.* **2**:335-344.
43. **Duterque-Coquillaud, M., C. Niel, S. Plaza, and D. Stéhelin.** 1993. New human *erg* isoforms generated by alternative splicing are transcriptional activators. *Oncogene* **8**:1865-1873.
44. **Eisenbeis, C. F., H. Singh, and U. Storb.** 1995. Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator. *Genes Dev.* **9**:1377-1387.
45. **Feng, J. A., R. C. Johnson, and R. E. Dickerson.** 1994. Hin recombinase bound to DNA : The origin of specificity in major and minor groove interactions. *Science* **263**:348-355.

46. **Ferrari, S., V. R. Harley, A. Pontiggia, P. N. Goodfellow, R. Lovell-Badge, and M. E. Bianchi.** 1992. SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.* **11**:4497-4506.
47. **Fisher, R. J., G. Mavrothalassitis, A. Kondoch, and T. S. Papas.** 1991. High-affinity DNA-protein interactions of the cellular ETS1 protein : the determination of the ETS binding motif. *Oncogene* **6**:2249-2254.
48. **Fletcher, J. C., K. C. Burtis, D. S. Hogness, and C. S. Thummel.** 1995. The *drosophila E74* gene is required for metamorphosis and plays a role in the polytene chromosome puffing response to ecdysone. *Dev.* **121**:1455-1465.
49. **Frampton, J., T. J. Gibson, S. A. Ness, G. Doderlein, and T. Graf.** 1991. Proposed structure for the DNA-binding domain of the MYB oncoprotein based on model building and mutational analysis. *Prot. Eng.* **4**:891-901.
50. **Frampton, J., T. J. Gibson, S. A. Ness, G. Doderlein, and T. Graf.** 1991. Proposed structure for the DNA-binding domain of the MYB oncoprotein based on model building and mutational analysis. *Prot. Eng.* **4**:891-901.
51. **Friedman, A. M., T. O. Fishmann, and T. A. Steitz.** 1995. Crystal structure of *lac* repressor core tetramer and its implications for DNA looping. *Science* **268**:1721-1727.
52. **Galson, D. L., J. O. Hensold, T. R. Bishop, M. Schalling, A. D. D'Andrea, C. Jones, P. E. Auron, and D. E. Housman.** 1993. Mouse β -globin DNA-binding protein B1 is identical to a proto-oncogene, the transcription factor *Spi-1/PU.1*, and is restricted in expression to hematopoietic cells and the testis. *Mol. Cell. Biol.* **13**:2929-2941.
53. **Gégonne, A., R. Bosselut, R. A. Bailly, and J. Ghysdael.** 1993. Synergistic activation of the HTLV1 LTR Ets-responsive region by transcription factors Ets1 and Sp1. *EMBO J.* **12**:1169-1172.

54. **Gégonne, A., B. Punyammalee, B. Rabault, R. Bosselut, S. Seneca, M. Crabeel, and J. Ghysdael.** 1992. Analysis of the DNA binding and transcriptional activation properties of the Ets1 oncoprotein. *New Biol.* **4**:512-519.
55. **Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, and K. Wüthrich.** 1994. Homeodomain-DNA recognition. *Cell* **78**:211-223.
56. **Giese, K., J. Cox, and R. Grosschedl.** 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structure. *Cell* **69**:185-195.
57. **Giese, K., C. Kingsley, J. R. Kirshner, and R. Grosschedl.** 1995. Assembly and function of a TCRA enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev.* **9**:995-1008.
58. **Gille, H., M. Kortenjann, O. Thomae, C. Moomaw, C. Slaughter, M. H. Cobb, and P. E. Shaw.** 1995. ERK phosphorylation potentiates Elk-1-mediated ternary complex formation and transactivation. *EMBO J.* **14**:951-962.
59. **Gille, H., A. D. Sharrocks, and P. E. Shaw.** 1992. Phosphorylation of transcription factor p62^{TCF} by MAP kinase stimulates ternary complex formation at *c-fos* promoter. *Nature* **358**:414-417.
60. **Giovane, A., A. Pintzas, S. M. Maira, P. Sobieszczuk, and B. Waslyk.** 1994. Net, a new *ets* transcription factor that is activated by Ras. *Genes Dev.* **8**:1502-1513.
61. **Golub, T., G. F. Barker, M. Lovett, and D. G. Gilliland.** 1994. Fusion of PDGF receptor β to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* **77**:307-316.
62. **Goodrich, J. A., T. Hoey, C. J. Thut, A. Admon, and R. Tjian.** 1993. Drosophila TAF_{II}40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* **75**:519-530.

63. **Green, S., I. Issemann, and E. Sheer.** 1988. A versatile *in vivo* and *in vitro* eukaryotic expression vector for protein engineering. *Nucleic Acids Res.* **16**:369.
64. **Hagemier, C., A. J. Bannister, A. Cook, and T. Kouzarides.** 1993. The activation domain of transcription factor PU1 binds the retinoblastoma (RB) protein and the transcription factor TFIID *in vitro* : RB shows sequence similarity to TFIID and TFIIB. *Proc. Natl. Acad. Sci. USA* **90**:1580-1584.
65. **Hagman, J., and R. Grosschedl.** 1992. An inhibitory carboxyl-terminal domain in Ets-1 and Ets-2 mediates differential binding of ETS family factors to promoter sequences of the *mb-1* gene. *Proc. Natl. Acad. Sci. USA* **89**:8889-8893.
66. **Hagman, J., and R. Grosschedl.** 1992. An inhibitory carboxyl-terminal domain in Ets-1 and Ets-2 mediates differential binding of ETS family factors to promoters sequences of the *mb-1* gene. *Proc. Natl. Acad. Sci. USA* **89**:8889-8893.
67. **Hahn, S. L., and B. Wasyluk.** 1994. The oncoprotein v-Ets is less selective in DNA binding than c-Ets-1 due to the C-terminal sequence change. *Oncogene* **9**:2499-2512.
68. **Harrison, C. J., A. A. Bohm, and H. C. M. Nelson.** 1994. Crystal structure of the DNA binding domain of the Heat Shock Transcription Factor. *Science* **263**:224-227.
69. **Henrissat, B.** 1991. Classification of glycosyl hydrolases based on amino acids sequence identities. *Biochem. J.* **280**:309-316.
70. **Henrissat, B., M. Claeysens, P. Tomme, L. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by Hydrophobic Cluster Analysis. *Gene* **81**:83-95.
71. **Herr, W., and M. A. Cleary.** 1995. The POU domain : versatility in transcription regulation by a flexible two-in -one DNA-binding domain. *Genes Dev.* **9**:1679-1693.

72. **Higashino, F., K. Yoshida, Y. Fujinaga, K. Kamio, and K. Fujinaga.** 1993. Isolation of a cDNA encoding the adenovirus E1A enhancer binding protein : a new human member of the *ets* oncogene family. *Nucleic Acids Res.* **21**:547-553.
73. **Higashino, F., K. Yoshida, T. Noumi, M. Seiki, and K. Fujinaga.** 1995. Ets-related protein E1A-F can activate three different matrix metalloproteinase gene promoters. *Oncogene* **10**:1461-1463.
74. **Hill, C. S., J. Wynne, and R. Treisman.** 1994. Serum-regulated transcription by Serum Response Factor (SRF) : a novel role for the DNA binding domain. *EMBO J.* **13**:5421-5432.
75. **Hipskind, R. A., V. N. Rao, C. G. F. Mueller, E. S. P. Reddy, and A. Nordheim.** 1991. Ets-related protein Elk-1 is homologous to the *c-fos* regulatory factor p62^{TCF}. *Nature* **354**:531-534.
76. **Ho, I. C., N. K. Bhat, L. R. Gottschalk, T. Lindsten, C. B. Thompson, T. S. Papas, and J. M. Leiden.** 1990. Sequence-specific binding of human Ets-1 to the T cell receptor α gene enhancer. *Science* **250**:814-8818.
77. **Hochschild, A., N. Irwin, and M. Ptashne.** 1983. Repressor structure and the mechanism of positive control. *Cell* **32**:319-325.
78. **Hwang, J. J., P. Chambon, and I. Davidson.** 1993. Characterization of the transcription activation function and the DNA binding domain of transcriptional enhancer factor-1. *EMBO J.* **12**:2337-2348.
79. **Isobe, M., F. Yamagishi, K. Yoshida, F. Higashino, and K. Fujinaga.** 1995. Assignment of the *ets*-related transcription factor E1A-F gene (ETV4) to human chromosome region 17q21. *Genomics* **28**:357-359.
80. **Janknecht, R., W. H. Ernst, V. Pingoud, and A. Nordheim.** 1993. Activation of ternary complex factor Elk-1 by MAP kinases. *EMBO J.* **12**:5097-5104.
81. **Janknecht, R., and A. Nordheim.** 1992. Elk-1 protein domains required for direct and SRF-assisted DNA-binding. *Nucleic Acids Res.* **20**:3317-3324.
82. **Janknecht, R., R. Zinck, W. H. Ernst, and A. Nordheim.** 1994. Functional dissection of the transcription factor Elk-1. *Oncogene* **9**:1273-1278.

83. **Jeon, I.-S., J. N. Davis, B. S. Braun, J. E. Sublett, M. F. Roussel, C. T. Denny, and D. N. Shapiro.** 1995. A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. *Oncogene* **10**:1229-1234.
84. **Johansen, F. E., and R. Prywes.** 1993. Identification of transcriptional activation and inhibitory domains in serum response factor (SRF) by using GAL4-SRF constructs. *Mol. Cell. Biol.* **13**:4640-4647.
85. **John, S., R. B. Reeves, J. X. Lin, R. Child, J. M. Leiden, C. B. Thompson, and W. J. Leonard.** 1995. Regulation of cell-type interleukin -2 receptor α -chain gene expression : potential role of physical interactions between Elf-1, HMG-I(Y), and NF- κ B family proteins. *Mol. Cell. Biol.* **15**:1786-1796.
86. **Jonsen, M. D., J. M. Petersen, and B. J. Graves.** 1994. Intramolecular regulation of Ets-1 DNA binding. Tenth annual meeting on oncogenes, Fredericks.
87. **Jorcyk, C. L., D. K. Watson, G. J. Mavrothalassitis, and T. S. Papas.** 1991. The human *ETS1* gene : genomic structure, promoter characterization and alternative splicing. *Oncogene* **6**:523-532.
88. **Jordan, S. R., and C. O. Pabo.** 1988. Structure of the lambda complex at 2.5Å^o resolution : details of the repressor-operator interactions. *Science* **242**:893-899.
89. **Karim, F. D., and C. S. Thummel.** 1991. Ecdysone coordinates the timing and amounts of *E74A* and *E74B* transcription in *Drosophila*. *Genes Dev.* **5**:1067-1079.
90. **Karim, F. D., L. D. Urness, C. S. Thummel, M. J. Klemsz, S. R. McKercher, A. Celada, C. Van Beveren, R. A. Maki, K. Gunther, J. A. Nye, and B. J. Graves.** 1990. The ETS domain: A new DNA binding motif that recognizes a purine-rich core sequence. *Genes Dev.* **4**:1451-1453.
91. **Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å^o resolution : a framework for understanding homeodomain-DNA interactions. *Cell* **63**:579-590.

92. **Klämbt, C.** 1993. The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells. *Development* **117**:163-176.
93. **Klämbt, C.** 1993. The *drosophila* gene *pointed* encodes two-ETS-like proteins which are involved in the development of the midline glial cells. *Development* **117**:163-176.
94. **Koizumi, S., R. J. Fisher, S. Fujiwara, C. L. Jorcyk, N. K. Bhat, A. Seth, and T. S. Papas.** 1990. Isoforms of the human *ets-1* protein : generation by alternative splicing and differential phosphorylation. *Oncogene* **5**:675-681.
95. **Kortenjann, M., O. Thoma, and P. E. Shaw.** 1994. Inhibition of *v-raf*-dependent *c-fos* expression and transformation by a kinase-defective mutant of the mitogen-activated protein kinase Erk-2. *Mol. Cell. Biol.* **14**:4815-4824.
96. **Lai, Z. C., and G. M. Rubin.** 1992. Negative control of photoreceptor development in *Drosophila* by the product of the *yan* gene, an ETS domain protein. *Cell* **70**:609-620.
97. **LaMarco, K., C. C. Thompson, B. P. Byers, E. M. Walton, and S. L. McKnight.** 1991. Identification of Ets- and Notch-related subunits in GA binding Protein. *Science* **253**:789-792.
98. **Laudet, V., C. Niel, M. Duterque, D. Leprince, and D. Stéhelin.** 1993. Evolution of the *ets* gene family. *B. B. R. C.* **190**:8-14.
99. **Lautenberger, J. A., L. A. Burdett, M. A. Gunnell, Q. Shengmei, D. K. Watson, S. J. O'Brien, and T. S. Papas.** 1992. Genomic dispersal of the *ets* gene family during metazoan evolution. *Oncogene* **7**:1713-1719.
100. **Lautenberger, J. A., and T. S. Papas.** 1993. Inversion of a chicken *ets-1* proto-oncogene segment in avian leukemia virus E26. *J. Virol.* **67**:610-612.
101. **Leiting, B., R. De Francesco, L. Tomei, R. Cortese, G. Otting, and K. Wüthrich.** 1993. The three-dimensional NMR-solution structure of the polypeptide fragment 195-286 of the LFB1/HNF1 transcription factor from rat liver comprises a non-classical homeodomain. *EMBO J.* **12**:1797-1803.

102. **Lemesle-Varloot, L., B. Henrissat, C. Gaboriaud, A. Morgat, and J. P. Mornon.** 1990. Hydrophobic cluster analysis: procedures to derive structural and functional information from 2-D-representation of protein sequences. *Biochimie* **72**:555-574.
103. **Leprince, D., P. Crépieux, V. Laudet, A. Flourens, and D. Stéhelin.** 1993. a new mechanism of ncogenic activation : E26 retroviral *v-ets* oncogene has inverted the C-terminal end of the transcription factor *c-ets-1*. *Virology* **194**:855-857.
104. **Leprince, D., P. Crépieux, and D. Stéhelin.** 1992. *c-ets-1* DNA binding to the PEA3 motif is differentially inhibited by all the mutations found in *v-ets*. *Oncogene* **7**:9-17.
105. **Leprince, D., M. Duterque-Coquillaud, R. P. Li, C. Henry, A. Flourens, B. Debuire, and D. Stéhelin.** 1988. Alternative splicing within the chicken *c-ets-1* locus : implications for transduction within the E26 retrovirus of the *c-ets* proto-oncogene. *J. Virol.* **62**:3233-3241.
106. **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 virus E26. *Nature* **306**:395-397.
107. **Liang, H., X. Mao, E. T. Olejniczak, D. G. Nettesheim, L. Yu, R. P. Meadows, C. B. Thompson, and S. W. Fesik.** 1994b. Solution structure of the ets domain of Fli-1 when bound to DNA. *Nature Struct. Biol.* **1**:871-876.
108. **Liang, H., E. T. Olejniczak, X. Mao, D. G. Nettesheim, L. Yu, C. B. Thompson, and S. W. Fesik.** 1994a. The secondary structure of the ets domain of human Fli-1 resembles that of the helix-turn-helix DNA-binding motif of the *Escherichia coli* catabolite activator protein. *Proc. Natl. Acad. Sci. USA* **91**:11655-11659.
109. **Lin, Y. S., and M. R. Green.** 1991. Mechanism of action of an acidic transcriptional activator in vitro. *Cell* **64**:971-981.

110. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid sensitive protein similarity searches. *Science* **227**:1435-1441.
111. **Lopez, M., P. Oettgen, Y. Akbarali, U. Dendorfer, and T. A. Libermann.** 1994. ERP, a new member of the *ets* transcription factor/oncoprotein family : cloning, characterization, and differential expression during B-lymphocyte development. *Mol. Cell. Biol.* **14**:3292-3309.
112. **Maguire, H. F., J. P. Hoeffler, and A. Siddiqui.** 1991. HBV X protein alters the DNA-binding specificity of CREB and ATF-2 by protein-protein interactions. *Science* **252**:842-844.
113. **Mao, X., S. Miesfeldt, H. Yang, J. M. Leiden, and C. B. Thompson.** 1994. The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J. Biol. Chem.* **269**:18216-18222.
114. **Marais, R., J. Wynne, and R. Treisman.** 1993. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell* **73**:381-393.
115. **Martin, M. E., J. Piette, M. Yaniv, W. J. Tang, and W. R. Folk.** 1988. Activation of the polyoma virus enhancer by a murine activator protein 1 (AP1) homolog and two contiguous proteins. *Proc. Natl. Acad. Sci. USA* **85**:5839-5843.
116. **Mavrothalassitis, G., R. J. Fisher, F. Smyth, D. K. Watson, and T. S. Papas.** 1994. Structural interference of the ETS1 DNA-binding domain determined by mutational analysis. *Oncogene* **9**:425-435.
117. **May, W. A., M. L. Gishizky, S. L. Lessnick, L. B. Lunsford, B. C. Lewis, O. Delattre, J. Zucman, G. Thomas, and C. T. Denny.** 1993b. Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain by *FLI-1* for transformation. *Proc. Natl. Acad. Sci. USA* **90**:5752-5756.

118. **May, W. A., S. L. Lessnick, B. S. Braun, M. Klemsz, B. C. Lewis, L. B. Lunsford, R. Hromas, and C. T. Denny.** 1993a. The Ewing's sarcoma *EWS/FLI-1* fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than *FLI-1*. *Mol. Cell. Biol.* **13**:7393-7398.
119. **Metz, T., and T. Graf.** 1991a. *v-myb* and *v-ets* transform chicken erythroid cells and cooperate both in *trans* and in *cis* to induce distinct differentiation phenotypes. *Genes Dev.* **5**:369-380.
120. **Metz, T., and T. Graf.** 1991b. Fusion of the nuclear oncoproteins v-Myb and v-Ets is required for the leukemogenicity of E26 virus. *Cell* **66**:95-105.
121. **Metz, T., and T. Graf.** 1992. the nuclear oncogenes v-erbA and v-ets cooperate in the induction of avian erythroleukemia. *Oncogene* **7**:597-605.
122. **Monté, D., J. L. Baert, P. A. Defossez, Y. de Launoit, and D. Stéhelin.** 1994. Molecular cloning and characterization of human ERMp, a new member of the ETS family closely related to mouse PEA3 and ER81 transcription factors. *Oncogene* **9**:1397-1406.
123. **Monté, D., L. Coutte, J. L. Baert, I. Angeli, D. Stéhelin, and Y. de Launoit.** 1995. Molecular characterization of the ets-related human transcription factor ER81. *Oncogene* **11**:771-780.
124. **Moreau-Gachelin, F.** 1994. Spi1/PU1 : an oncogene of the Ets family. *Bioch. Biophys. Acta* **1198**:149-163.
125. **Moreau-Gachelin, F., A. Tavitian, and P. Tambourin.** 1988. *Spi-1* is a putative oncogene in virally induced murine erythroleukaemias. *Nature* **331**:277-280.
126. **Mornon, J. P., V. Bissery, C. Gaboriaud, A. Thomas, T. Ojasoo, and J. P. Raynaud.** 1989. Hydrophobic cluster analysis (HCA) of the hormone-binding domain of receptor proteins. *J. Steroid Biochem.* **34**:355-361.
127. **Murakami, K., G. Mavrothalassitis, N. K. Bhat, R. J. Fisher, and T. S. Papas.** 1993. Human ERG-2 protein is a phosphorylated DNA-binding protein - a distinct member of the ets family. *Oncogene* **8**:1559-1566.

128. **Nye, J. A., J. M. Petersen, C. V. Gunther, M. D. Jonsen, and B. J. Graves.** 1992. Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes Dev.* **6**:975-990.
129. **O'Neill, E. M., I. Rebay, R. Tjian, and G. M. Rubin.** 1994. The activities of two Ets-related transcription factors required for drosophila eye development are modulated by the Ras/MAPK pathway. *Cell* **78**:137-147.
130. **Ogata, K., H. Hojo, S. Aimoto, T. Nakai, H. Nakamura, A. Sarai, S. Ishii, and Y. Nishimura.** 1992. Solution structure of a DNA-binding unit of MYB: A helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl. Acad. Sci. USA* **89**:6428-6432.
131. **Ogata, K., S. Morikawa, H. Nakamura, A. Sekikawa, T. Inoue, H. Kanai, A. Sarai, S. Ishii, and Y. Nishimura.** 1994. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79**:639-648.
132. **Ohlendorf, D. H., W. F. Anderson, R. G. Fisher, Y. Takeda, and B. W. Matthews.** 1982. The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. *Nature* **298**:718-723.
133. **Ohlendorf, D. H., W. F. Anderson, M. Lewis, C. O. Pabo, and B. W. Matthews.** 1983. Comparison of the structures of Cro and λ repressor proteins from bacteriophage λ . *J. Mol. Biol.* **169**:757-7669.
134. **Osborne-Lawrence, S., P. L. Welsh, M. Spillman, S. C. Chandrasekharappa, T. D. Gallardo, M. Lovett, and A. M. Bowcock.** 1995. Direct selection of expressed sequences within a 1-Mb region flanking BRCA1 on human chromosome 17q21. *Genomics* **25**:248-255.
135. **Otting, G., Y. Q. Qian, M. Billeter, M. Müller, M. Affolter, W. J. Gehring, and K. Wüthrich.** 1990. Protein-DNA contacts in the structure of a homeodomain_DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**:3085-3092.

136. **Pabo, C. O., W. Krovatin, A. Jeffrey, and R. T. Sauer.** 1982. The N-terminal arms of λ repressor wrap around the operator DNA. *Nature* **298**:441-443.
137. **Pabo, C. O., and M. Lewis.** 1982. The operator-binding domain of λ repressor : structure and DNA recognition. *Nature* **298**:443-447.
138. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**:1053-95.
139. **Parisi, M. A., and D. A. Clayton.** 1991. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* **252**:965-969.
140. **Pearson, W. R.** 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enz.* **183**:63-98.
141. **Pongubala, J. M. R., S. Nagulapalli, M. J. Klemsz, S. R. McKercher, R. A. Maki, and M. L. Atchinson.** 1992. PU.1 recruits a second nuclear factor to a site important for immunoglobulin κ 3' enhancer activity. *Mol. Cell. Biol.* **12**:368-378.
142. **Pongubala, J. M. R., C. Van Beveren, S. Nagulapalli, M. J. Klemsz, S. R. McKercher, R. A. Maki, and M. L. Atchinson.** 1993. Effect of PU.1 phosphorylation on interaction with NF-EM5 and transcriptional activation. *Science* **259**:1622-1625.
143. **Prasad, D. D. K., V. N. Rao, L. Lee, and E. S. P. Reddy.** 1994. Differentially spliced *erg-3* product functions as a transcriptional activator. *Oncogene* **8**:669-673.
144. **Price, M. A., A. E. Rogers, and R. Treisman.** 1995. Comparative analysis of the ternary complex factors Elk-1, SAP-1a et SAP-2 (ERP/NET). *EMBO J.* **14**:2589-2601.
145. **Py, B., I. Bortoli-German, J. Haiech, M. Chippaux, and F. Barras.** 1991. Cellulase EGZ of *Erwinia chrysanthemi* : structural organization and importance of His 98 and Glu 133 residues for catalysis. *Prot. Eng.* **4**:325-333.

146. **Qian, Y. Q., M. Billeter, G. Otting, M. Müller, W. J. Gehring, and K. Wüthrich.** 1990. The Structure of the *Antennapedia* Homeodomain Determined by NMR Spectroscopy in Solution : Comparison with Prokaryotic Repressors. *Cell* **59**:573-580.
147. **Qian, Y. Q., K. Furukubo-Tokunaga, M. Müller, D. Resendez-Perez, W. J. Gehring, and K. Wüthrich.** 1994b. Nuclear Magnetic Resonance solution structure of the *fushi tarazu* homeodomain from *Drosophila* and comparison with the *Antennapedia* homeodomain. *J. Mol. Biol.* **238**:333-345.
148. **Qian, Y. Q., G. Otting, M. Billeter, M. Müller, W. J. Gehring, and K. Wüthrich.** 1993. NMR spectroscopy of a DNA complex with the uniformly ¹³C-labeled Antennapedia homeodomain and structure determination of the DNA-bound homeodomain. *J. Mol. Biol.* **234**:1070-1083.
149. **Qian, Y. Q., D. Resendez-Perez, W. J. Gehring, and K. Wüthrich.** 1994a. The des(1-6) Antennapedia homeodomain : comparison of the NMR solution structure and the DNA-binding affinity with the intact Antennapedia homeodomain. *Proc. Natl. Acad. Sci. USA* **91**:4091-4095.
150. **Ramakrishnan, V., J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet.** 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**:219-223.
151. **Rao, V. N., T. Ohno, D. D. K. Prasad, G. Bhattacharya, and E. S. P. Reddy.** 1993. Analysis of the DNA-binding and transcriptional activation functions of human *Fli-1* protein. *Oncogene* **8**:2167-2173.
152. **Rao, V. N., and E. S. P. Reddy.** 1992. A divergent *ets*-related protein, Elk-1, recognizes similar *c-ets-1* proto-oncogene target sequences and acts as a transcriptional activator. *Oncogene* **7**:65-70.
153. **Rao, V. N., and E. S. P. Reddy.** 1992b. Elk-1 domains responsible for autonomous DNA binding, SRE : SRF interaction and negative regulation of DNA-binding. *Oncogene* **7**:2335-2340.

154. **Rao, V. N., and E. S. P. Reddy.** 1993. Δ elk-1, a variant of *Elk-1*, fails to interact with the serum response factor and binds to DNA with modulated specificity. *Cancer Res.* **53**:215-220.
155. **Ravel-Chapuis, P., D. Leprince, B. Pain, R. Li, C. Domenget, D. Stéhelin, J. Samarut, and P. Jurdic.** 1991. *v-myb* and *v-ets* cooperate for the mitogenic stimulation of primary fibroblasts by avian E26 retrovirus. *J. Virol.* **65**:3928-3931.
156. **Ray-Gallet, D., C. Mao, A. Tavitian, and F. Moreau-Gachelin.** 1995. DNA binding specificities of Spi-1/PU.1 and Spi-B transcription factors and identification of a Spi-1/Spi-B binding site in the *c-fes/c-fps* promoter. *Oncogene* **11**:303-313.
157. **Rebay, I., and G. M. Rubin.** 1995. Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell* **81**:857-866.
158. **Reddy, E. S. P., and V. N. Rao.** 1990. Localization and modulation of the DNA-binding activity of the human *c-ets-1* protooncogene. *Cancer Res.* **50**:5013-5016.
159. **Reddy, E. S. P., and V. N. Rao.** 1991. *erg*, an *ets*-related gene, codes for sequence-specific transcriptional activators. *Oncogene* **6**:2285-2289.
160. **Roeder, R. G.** 1991. The complexities of eukaryotic transcription initiation : regulation of preinitiation complex assembly. *Trends Biochem. Sci.* **16**:402-408.
161. **Rorth, P., C. Nerlov, F. Blasi, and M. Johnson.** 1990. Transcription factor PEA3 participates in the induction of urokinase plasminogen activator transcription in murine keratinocytes stimulated with epidermal growth factor or phorbol-ester. *Nucleic Acids Res.* **18**:5009-5017.
162. **Rost, B., C. Sander, and R. Schneider.** 1994. Redefining the goals of protein secondary structure prediction. *J. Mol. Biol.* **235**:13-26.
163. **Sadowski, I., and M. Ptashne.** 1989. A vector for expressing GAL4(1-147) fusions in mammalian cells. *Nucleic Acids Res.* **17**:7539.

164. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press
165. **Schägger, H., and G. Von Jagow.** 1987. Tricine-sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**:368-379.
166. **Schneikert, J., Y. Lutz, and B. Wasylyk.** 1992. Two independent activation domains in c-Ets-1 and c-Ets-2 located in non-conserved sequences of the *ets* gene family. *Oncogene* **7**:249-256.
167. **Schoentgen, F., N. Seqqidi, S. Bucquoy, P. Jolles, L. Lemesle-Varloot, K. Provost, and J. P. Mornon.** 1992. Main structural and functional features of the basic cytosolic bovine 21 kDa protein delineated through Hydrophobic Cluster Analysis and molecular modelling. *Prot. Eng.* **5**:295-303.
168. **Scott, G. K., J. C. Daniel, X. Xiong, R. A. Maki, D. Kabat, and C. C. Benz.** 1994. Binding of an ETS-related protein within DNase I hypersensitive site of the *HER2/neu* promoter in human breast cancer cells. *J. Biol. Chem.* **269**:19848-19858.
169. **Scott, M. P., J. W. Tamkun, and G. W. Hartzell.** 1989. The structure and function of the homeodomain. *Biochim. Biophys. Acta* **989**:25-48.
170. **Seth, A., and T. S. Papas.** 1990. The *c-ets-1* proto-oncogene has oncogenic activity and is positively autoregulated. *Oncogene* **5**:1761-1767.
171. **Seth, A., D. K. Watson, D. G. Blair, and T. K. Papas.** 1989. *c-ets-2* protooncogene has mitogenic and oncogenic activity. *Proc. Natl. Acad. Sci. USA* **86**:7833-7837.
172. **Shimizu, K., H. Ichikawa, A. Tojo, Y. Kaneko, N. Maseki, Y. Hayashi, M. Ohira, S. Asano, and M. Ohki.** 1993. An *ets*-related gene, *ERG*, is rearranged in human myeloid leukemia with t(16;21) chromosomal translocation. *Proc. Natl. Acad. Sci. USA* **90**:10280-10284.

173. **Shin, M. K., and M. E. Koshland.** 1993. Ets-related protein PU1 regulates expression of the immunoglobulin J-chain through a novel Ets-binding element. *Genes Dev.* **7**:2006-2015.
174. **Shore, P., L. Bisset, J. Lakey, J. P. Waltho, R. Virden, and A. D. Sharrocks.** 1995. Characterization of the Elk-1 DNA-binding domain. *J. Biol. Chem.* **270**:5805-5811.
175. **Shore, P., and A. D. Sharrocks.** 1994. The transcription factors Elk-1 and Serum Response Factor interact by direct protein-protein contacts mediated by a short region of Elk-1. *Mol. Cell. Biol.* **14**:3283-3291.
176. **Siddique, H. R., V. N. Rao, L. Lee, and E. S. P. Reddy.** 1993. Characterization of the DNA binding and transcriptional activation domains of the erg protein. *Oncogene* **8**:1751-1755.
177. **Silverman, N., J. Agapite, and L. Guarente.** 1994. Yeast ADA2 protein binds to the VP16 protein activation domain and activates transcription. *Proc. Natl. Acad. Sci. USA* **91**:11665-11668.
178. **Soudant, N., O. Albagli, P. Dhordain, A. Flourens, D. Stéhelin, and D. Leprince.** 1994. A residue of the ETS domain mutated in the *v-ets* oncogene is essential for the DNA-binding and transactivating properties of the ETS-1 and ETS-2 proteins. *Nucleic Acids Res.* **22**:3871-3879.
179. **Stringer, K. F., C. J. Ingles, and J. Greenblatt.** 1990. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345**:783-786.
180. **Suzuki, H., V. Romano-Spica, T. S. Papas, and N. K. Bhat.** 1995. ETS1 suppresses tumorigenicity of human colon cancer cells. *Proc. Natl. Acad. Sci. USA* **92**:4442-4446.
181. **Tei, H., I. Nihonmatsu, T. Yokokura, R. Ueda, Y. Sano, T. Okuda, K. Sato, K. Hirata, S. C. Fujita, and D. Yamamoto.** 1992. Pokkuri, a *Drosophila* gene encoding an E26-specific (Ets) domain protein, prevents overproduction of the R7 photoreceptor. *Proc. Natl. Acad. Sci. USA* **89**:6856-6860.

182. **Thompson, C. B., C. Y. Wang, I. C. Ho, P. R. Bohjanen, B. Petryniak, C. H. June, S. Miesfeldt, L. Zhang, G. J. Nabel, B. Karpinski, and J. M. Leiden.** 1992. *cis*-acting sequences required for inducible interleukin-2 enhancer function bind a novel Ets-related protein, Elf-1. *Mol. Cell. Biol.* **12**:1043-1053.
183. **Thompson, C. C., T. A. Brown, and S. L. McKnight.** 1991. Convergence of Ets- and Notch-related structural motifs in a heteromeric DNA binding complex. *Science* **253**:762-768.
184. **Thoreau, E., B. Petridou, P. A. Kelly, J. Djiane, and J. P. Mornon.** 1991. Structural symmetry of the extracellular domain of the cytokine/growth hormone/prolactin receptor family and interferon receptors revealed by hydrophobic cluster analysis. *FEBS Lett.* **282**:26-31.
185. **Treisman, R., R. Maras, and J. Wynne.** 1992. Spatial flexibility in ternary complexes between SRF and its accessory proteins. *EMBO J.* **11**:4631-4640.
186. **Trimble, M. S., J. H. Xin, C. T. Guy, W. J. Muller, and J. A. Hassel.** 1993. PEA3 is overexpressed in mouse metastatic mammary adenocarcinomas. *Oncogene* **8**:3037-3042.
187. **Urness, L. D., and C. S. Thummel.** 1990. Molecular interactions within the ecdysone regulatory hierarchy : DNA binding properties of the *Drosophila* Ecdysone-Inducible E74A protein. *Cell* **63**:47-61.
188. **Vandenbunder, B., L. Pardanaud, T. Jaffredo, M. A. Mirabel, and D. Stéhelin.** 1989. Complementary patterns of expression *c-ets-1*, *c-myb* and *c-myc* in the blood forming system of chick embryo. *Development* **106**:265-274.
189. **Vuister, G. W., S. J. Kim, A. Orosz, J. Marquardt, C. Wu, and A. Bax.** 1994b. Solution structure of the DNA-binding domain of *Drosophila* heat shock transcription factor. *Str. Biol.* **1**:605-614.
190. **Vuister, G. W., S. J. Kim, C. Wu, and A. Bax.** 1994a. NMR evidence for similarities between the DNA-binding regions of *Drosophila melanogaster* Heat Shock Factor and the Helix-Turn-Helix and HNF-3/*forkhead* families of transcription factors. *Biochemistry* **33**:10-16.

191. **Wang, C. Y., B. Petryniak, I. C. Ho, C. B. Thompson, and J. M. Leiden.** 1992. Evolutionarily conserved Ets family members display distinct DNA binding specificities. *J. Exp. Med.* **175**:1391-1399.
192. **Wang, C. Y., B. Petryniak, C. B. Thompson, W. G. Kaelin, and J. M. Leiden.** 1993. Regulation of the Ets-related transcription factor Elf-1 by binding to the retinoblastoma protein. *Science* **260**:1330-1335.
193. **Wasylyk, B., C. Wasylyk, P. Flores, A. Bègue, D. Leprince, and D. Stéhelin.** 1990. The c-ets proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. *Nature* **346**:191-193.
194. **Wasylyk, C., P. Flores, A. Gutman, and B. Wasylyk.** 1989. PEA3 is a nuclear target for transcription activation by non-nuclear oncogenes. *EMBO J.* **8**:3371-3378.
195. **Wasylyk, C., A. Gutman, R. Nicholson, and B. Wasylyk.** 1991. The c-Ets oncoprotein activates the stromelysin promoter through the same elements as several non-nuclear oncoproteins. *EMBO J.* **10**:1127-1134.
196. **Wasylyk, C., J. P. Kerckaert, and B. Wasylyk.** 1992. A novel modulator domain of ETS transcription factors. *Genes Dev.* **6**:965-974.
197. **Wasylyk, C., and B. Wasylyk.** 1992. Oncogenic conversion alters the transcriptional properties of ets. *Cell Growth Diff.* **3**:617-625.
198. **Wasylyk, C., and B. Wasylyk.** 1993. Oncogenic conversion of Ets affects redox regulation *in-vivo* and *in-vitro*. *Nucleic Acids Res.* **21**:523-529.
199. **Waterman, M. L., W. H. Fischer, and K. A. Jones.** 1991. A thymus-specific member of the HMG protein family regulates the human T cell receptor C alpha enhancer. *Genes Dev.* **5**:656-669.
200. **Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott.** 1991. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* **5**:1377-1386.

201. **Weir, H. M., P. J. Kraulis, C. S. Hill, A. R. C. Raine, E. D. Laue, and J. O. Thomas.** 1993. Structure of the HMG box motif in the B-domain of HMG1. *EMBO J.* **12**:1311-1319.
202. **Wernert, N., F. Gilles, V. Fafeur, F. Bouali, M. B. Raes, C. Pyke, T. Dupressoir, G. Seitz, B. Vandebunder, and D. Stéhelin.** 1994. Stromal expression of c-Ets-1 transcription factor correlates with tumor invasion. *Cancer Res.* **54**:5683-5688.
203. **Wernert, N., M. B. Raes, P. Lasalle, M. P. Dehouck, B. Gosselin, B. Vandebunder, and D. Stéhelin.** 1992. *c-ets-1* proto-oncogene is a transcription factor expressed in endothelial cells during tumor vascularization and other forms of angiogenesis in humans. *Am. J. Pathol.* **140**:119-127.
204. **Whitmarsh, A. J., P. Shore, A. D. Sharrocks, and R. Davis.** 1995. Integration of MAP kinase signal transduction pathways at the serum response element. *Science* **269**:403-407.
205. **Williams, S. C., M. Baer, A. J. Dillner, and P. F. Johnson.** 1995. CRP2 (C/EBP β) contains a bipartite regulatory domain that controls transcriptional activation, DNA binding and cell specificity. *EMBO J.* **14**:3170-3183.
206. **Wilson, K. P., L. M. Shewchuk, R. G. Brennan, A. J. Otsuka, and B. W. Matthews.** 1992. *Escherichia coli* biotin holoenzyme synthetase/*bio* repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. Natl. Acad. Sci. USA* **89**:9257-9261.
207. **Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo.** 1991. Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**:
208. **Woodcock, S., J. P. Mornon, and B. Henrissat.** 1992. Detection of secondary structure elements in proteins by hydrophobic cluster analysis. *Prot. Eng.* **5**:629-636.

209. **Woods, D. B., J. Ghysdael, and M. J. Owen.** 1992. Identification of nucleotide preferences in DNA sequences recognised specifically by c-Ets-1 protein. *Nucl. Acids Res.* **20**:699-704.
210. **Wotton, D., J. Ghysdael, S. Wang, N. A. Speck, and M. J. Owen.** 1994. Cooperative binding of Ets-1 and Core Binding Factor to DNA. *Mol. Cell. Biol.* **14**:840-850.
211. **Xiao, H., A. Pearson, B. Coulombe, R. Truant, S. Zhang, J. L. Regier, S. J. Triezenberg, D. Reinberg, O. Flores, C. J. Ingles, and J. Greenblatt.** 1994. Binding of basal transcription factor TFIID to the acidic activation domains of VP16 and p53. *Mol. Cell. Biol.* **14**:7013-7024.
212. **Xin, J. H., A. Cowie, P. Lachance, and J. A. Hassel.** 1992. Molecular cloning and characterization of PEA3, a new member of the *Ets* oncogene family that is differentially expressed in mouse embryonic cells. *Genes Dev.* **6**:481-496.
213. **Xu, W., M. A. Rould, S. Jun, C. Desplan, and C. O. Pabo.** 1995. Crystal structure of a paired domain-DNA complex at 2.5Å resolution reveals structural basis for Pax developmental mutations. *Cell* **80**:639-650.
214. **Zhang, L., V. Lemarchandel, P. H. Romeo, Y. Ben-David, P. Greer, and A. Bernstein.** 1993. The *Fli-1* proto-oncogene, involved in erythroleukemia and Ewing's sarcoma, encodes a transcriptional activator with DNA-binding specificities distinct from other Ets family members. *Oncogene* **8**:1621-1630.
215. **Zucman, J., T. Melot, C. Desmaze, J. Ghysdael, B. Plougastel, M. Peter, J. M. Zucker, T. J. Triche, D. Sheer, C. Turc-Carel, P. Ambros, V. Combaret, G. Lenoir, A. Aurias, G. Thomas, and O. Delattre.** 1993. Combinatorial generation of variable fusion proteins in the Ewing family of tumours. *EMBO J.* **12**:4481-4487.

