

**AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE**

(Article 3 de l'arrêté du 25 septembre 1985)

1) Titre de la thèse :

CONTRIBUTION A L'ETUDE DE LA FONCTION ET DE L'EVOLUTION DE  
DEUX FACTEURS DE TRANSCRIPTION : PAX6 ET COUP-TF

2) Nom de l'auteur et prénom : Mlle LANGLOIS MARIE CLAIRE

3) Président du Jury : *Mr CURGY*

4) Membres du jury :

LAUDET V.  
VERNIER P.  
DANAN J.L.  
STEHÉLIN D.  
CURGY J.J.



5) Date de la soutenance : 09-12-1998

6) Reproduction de la thèse soutenue : (barrer les mentions inutiles)

~~- thèse pouvant être reproduite en l'état~~

~~- thèse ne pouvant être reproduite~~

*α* - thèse pouvant être reproduite après corrections  
suggérées au cours de la soutenance

NON CORRIGÉE LE 30 MARS 2000  
DATE DE MAJEURITÉ POSITION  
DES LECTEURS - 1998  
SERV. DE LA BIBLIOTHÈQUE UNIVERSITAIRE  
59655 VILLENEUVE D'ASCQ CEDEX

Villeneuve d'Ascq le *09*.12.1998

Signature du Président de Jury

*Prof JJ CURGY*

## REMERCIEMENTS

Je tiens à remercier les membres du jury, Philippe Vernier, Jean-Louis Danan et Jean-Jacques Curgy, qui ont aimablement accepté de juger ce travail.

Je remercie également Dominique Stéhelin, de m'avoir acceptée dans son laboratoire. Un grand merci aussi à Simon Saule et à Vincent Laudet de m'avoir tous deux accueillie dans leur groupe respectif. J'ai appris beaucoup avec vous deux, et il m'a été infiniment profitable de m'adapter à deux façons de travailler différentes, j'ai "élargi mes horizons". C'est grâce à vous deux que j'ai eu l'envie de continuer l'étude des gènes exprimés dans le système nerveux central, auxquels appartiennent mes deux "chéris" COUP-TF et Pax-6, que je retrouverais très bientôt dans le thalamus... Vincent, je n'oublierai jamais ton enthousiasme et tes encouragements. En ces temps d'aigreurs et de désillusions, ça fait du bien de voir qu'il existe encore des chercheurs qui gardent leur foi pour la science ... et tu en fais partie!

Je tiens à remercier particulièrement Agnès Begue, l'encyclopédie universelle vivante des méthodes de clonages et des trucs pratiques qui permettent les raccourcis. Et puis, pour tout ce que tu es Agnès, je ne pourrais tout écrire, de peur de heurter ta modestie, mais tu as été pour moi le soleil qui manquait au ciel du Nord. Merci aussi à Manuella de m'avoir initiée au clonage dès mon arrivée à l'Institut Pasteur et pour ses sourires qui illuminent son regard bleu azur. A Nathalie Spruyt aussi, qui fait des sourires tout aussi jolis. A Brigitte, le rayon de gaieté qui m'a appris comment produire les anticorps chez les lapins. A Papy qui nous plonge dans les secrets des cellules, qui n'en n'ont plus pour lui. A Serge et Christine, qui m'ont initié aux techniques de fixation à l'ADN et de transactivation et avec qui j'ai pu avoir de bonnes discussions scientifiques. A Martin aussi, pour les discussions scientifiques, pour son soutien, sa générosité et sa disponibilité dans les coups durs. Merci aussi à Thierry Dupressoir avec qui j'ai partagé mon bureau ces trois derniers mois, pour sa patience et son soutien dans ma dernière ligne droite. Merci à Sophana, le roi des gags et de la photo.

Merci à tous les autres, qui m'ont souri un jour et plus particulièrement à Cateline, Fatima, Hélène, pour leur gentillesse. A propos de gentillesse, merci beaucoup à Catherine, Jean-Marc du groupe endocrin'os et à Nini et Marie-Christine, pour la relecture du manuscrit. Un merci particulier à Nini et Marie-Christine pour leur gentillesse tout au long de l'année.

Je remercie aussi l'Institut Pasteur, la région Nord Pas de Calais et L'Association pour la Recherche contre le Cancer qui ont financé mes années de thèse.

## Résumé du mémoire

La première partie de ce travail, a porté sur la régulation du gène Pax-6 qui code des protéines pourvues de deux domaines de fixation à l'ADN, le domaine paired et le domaine homéo. Les effets des gènes Pax sont dépendants de leur dose dans l'organisme ce qui suggère qu'il est important de mieux comprendre la régulation de leur expression. Cette régulation de l'expression peut s'effectuer à différents niveaux. Nous avons caractérisé un élément régulateur *cis*, capable d'activer la transcription de l'un des promoteurs du gène Pax-6, P0, dont l'activité est spécifique des cellules de neurorétine et du stade du développement. Nous avons vérifié que cette fonction enhancer était conservée chez la souris. Par ailleurs, nous avons mis en évidence que les produits du gène engrailed inhibaient la transcription de Pax-6. Nous avons montré enfin, que la régulation de la quantité des isoformes de Pax-6 s'exerçait à un niveau transcriptionnel, mais aussi post-transcriptionnel, avec l'exemple de l'ARNm B1. Cet ARNm produit en effet uniquement des protéines à domaine paired. Ces petites protéines ne sont pas capables de fixer des cibles de type homéo et n'ont pas encore de fonction connue chez les vertébrés.

Dans une deuxième approche nous avons tenté de mieux comprendre l'évolution d'un gène codant un facteur de transcription lui aussi important dans le développement du système nerveux : COUP-TF, un membre de la famille des récepteurs nucléaires d'hormones. Il nous a semblé intéressant d'isoler son homologue chez un céphalocordé, l'amphioxus, un procordé proche des vertébrés, possédant un cerveau simplifié. Amphi COUP-TF n'est pas exprimé dans la vésicule cérébrale de l'amphioxus, ce qui suggère que le rôle de COUP-TF dans la segmentation du diencephale est apparue après la séparation cordés/vertébrés. Par contre, la fonction biochimique (fixation à l'ADN, répression transcriptionnelle) de Amphi-COUP-TF est très conservée entre l'amphioxus et les vertébrés. L'isolement, de COUP-TF chez l'hydre, un organisme très précoce dans la classification animale suggère une conservation de la fonction de COUP-TF dans la détermination des cellules nerveuses. Il serait intéressant dans l'avenir d'établir les relations entre COUP-TF et Pax-6 et de comprendre les modifications apportées dans leur réseau de régulation au cours de l'évolution.

## Liste des publications

- **Identification and characterization of a neuroretina-specific enhancer element in the quail Pax-6 (Pax-QNR) gene.**

S.Plaza, C. Dozier, M-C Langlois et S. Saule.  
*Molecular and Cellular Biology* , 15, 892-903, 1995

- **The homeobox-containing Engrailed (EN-1) product down-regulates the expression of Pax-6 through a DNA-binding independent mechanism.**

S.Plaza, M-C Langlois, N. Turque, M. Bailly, A. Begue, B. Quatannens, C. Dozier et S. Saule.  
*Cell Growth and Differentiation* , 8, 1115-1125, 1997.

- **A quail particular pax-6 mRNA (pax-QNR) leads to the increase of paired less proteins**

M-C Langlois, P. Martin, C. Dozier et S. Saule.  
*En préparation.*

- **Ligand binding was acquired during the evolution of nuclear receptors**

Hector Escriva, Rachid Safi, Catherine Hänni, Marie-Claire Langlois, Pierre Saumitou-Laprade, Dominique Stéhelin, André Capron, Raymond Pierce, and Vincent Laudet.  
*Proc. Natl. Acad. Sci. USA* , 94, 6803-6808, 1997

- **A unique homologue for COUP-TF and EAR2 orphan nuclear receptors in Hydra.**

Marie-Claire Langlois, Dominique Gauchat, Hector Escriva, Vincent Laudet et Brigitte Galliot  
*En préparation pour "Genes, Development and Evolution"*

- **AmphiCOUP-TF and AmphiTR2/4, two nuclear orphan receptors of the lancelet *Branchiostoma floridae*, are implicated in retinoic acid signaling pathways.**

Marie-Claire Langlois, Nicolas D. Holland, Jean-Marc Vanacker, Hector Escriva, Christophe Quéva, Vincent Laudet et Linda Holland.  
*En préparation pour "Development"*.

- **Systématique et évolution des récepteurs nucléaires d'hormones.**

Hector Escriva, Ricardo Mendonca, Marie-Claire Langlois, Jean-Marc Vanacker, Raymond Pierce et Vincent Laudet.  
*Biosystema* , vol 15, pages 79-98, 1997

-- **Evolution and diversification of the nuclear hormone receptors superfamily.**

Hector Escriva, Marie-Claire Langlois, Ricardo Mendonca, Raymond Pierce and Vincent Laudet.

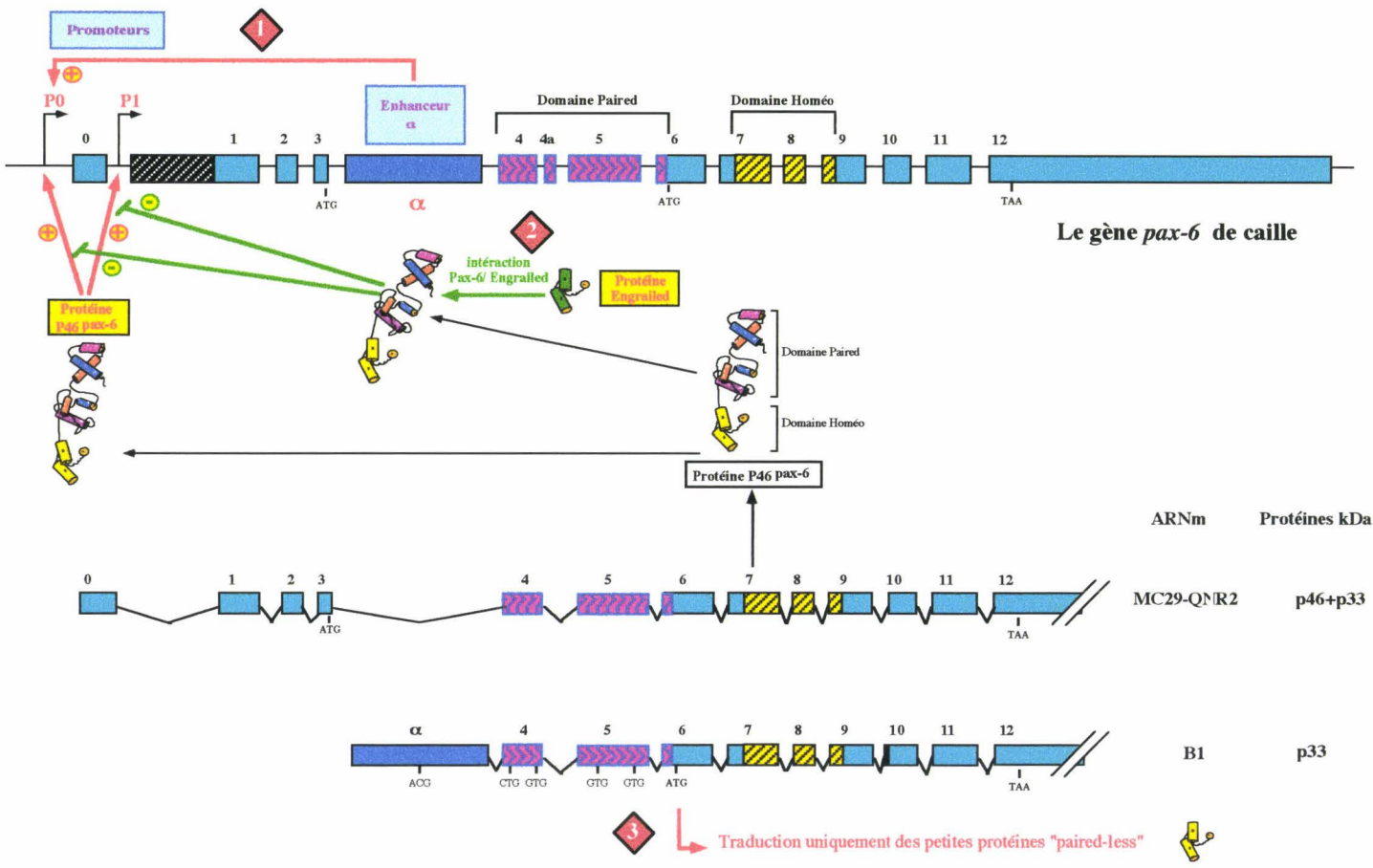
*Annals of the New York Academy of Sciences* , vol 839, May 15, 1998

## Présentation du mémoire

Les gènes qui participent à une régulation physiologique ou au développement embryonnaire sont inscrits dans des réseaux de régulation complexes faisant intervenir plusieurs gènes et souvent même plusieurs réseaux de régulation. Pour comprendre le rôle d'un gène dans un système donné, il faut pouvoir le replacer dans le contexte des régulations qu'il va induire et des régulations auxquelles il est lui-même soumis. L'analyse exhaustive de ces régulations n'est malheureusement pas possible dans le temps de la préparation d'une thèse et nécessitera les travaux de nombreux chercheurs et de nombreuses années d'études. Néanmoins, au cours de la préparation de ma thèse, j'ai choisi d'aborder plusieurs aspects de ces régulations qui participent au développement du système nerveux central et complètent le "puzzle d'informations" apporté par la "littérature" scientifique. Mon choix s'est porté sur l'étude de Pax-6 et COUP-TF, deux facteurs de transcription impliqués dans le développement du système nerveux central durant la vie embryonnaire.

Dans un premier temps, mon sujet d'étude a porté sur la régulation du gène *pax-6*. Ce gène appartient à la famille des gènes Pax composée de neuf membres chez les vertébrés qui s'expriment dans le système nerveux central, sauf Pax-1 et Pax-9 qui ne s'expriment que dans des dérivés mésodermiques. Ces gènes codent des facteurs de transcription dotés de deux domaines de fixation à l'ADN, un domaine paired et un domaine homéo. L'implication des gènes Pax dans le développement embryonnaire et leur mode de fixation à l'ADN sont résumés dans l'introduction de la première partie de ce manuscrit. *pax-6* s'exprime dans le système nerveux central, l'œil et le pancréas. Il a notamment été décrit comme le gène "maître" de l'œil, pour sa capacité à induire des yeux morphologiquement normaux dans des régions très distinctes de celles à l'origine de la formation de l'œil chez la Drosophile. Pax-6 est par ailleurs impliqué dans des phénomènes d'"effet-dose". Les mutations monoalléliques génèrent en effet, des malformations chez l'embryon hétérozygote (haploinsuffisance). De même, la surexpression de Pax-6 dans l'œil de souris induit des défauts dans cet organe. D'où l'importance de comprendre les mécanismes de régulation du gène *pax-6*.

# Etude de la régulation du gène *pax-6* de caille



- 1
 Caractérisation d'un enhancer du gène Pax-6 chez la caille et chez la souris  
 Plaza S., Dozier C., Langlois M.C. et al. , *M.C.B.* 1995 (article 1)
- 2
 La protéine Engrailed réprime l'expression du gène *pax-6* par un mécanisme indépendant de sa fixation à l'ADN  
 Plaza S., Langlois M.C. et al. *Cell Growth and Differentiation*, 1997 (article 2)
- 3
 Caractérisation d'un ARNm particulier de *pax-6* augmentant la quantité de protéines "paired-less"  
Langlois M.C. et al. , *en préparation* (article 3)

Figure A

Mon approche de travail a consisté à caractériser différents mécanismes de la régulation du gène *pax-6*.

D'une part, j'ai participé avec Serge Plaza, au sein du groupe de recherche dirigé par Simon Saule, à la caractérisation d'un enhancer chez la caille et chez la souris. Cet enhancer, noté  $\alpha$ , stimule la transcription initiée à partir de Po, l'un des deux promoteurs du gène *pax-6* de caille. Son activité est à la fois spécifique des cellules de neurorétine et du stade du développement embryonnaire et est conservée chez la souris. Cette étude fait l'objet du premier article du chapitre "Résultats" de la première partie de ce manuscrit (cf figure A).

Dans le deuxième article, toujours en collaboration avec Serge Plaza, nous montrons qu'un facteur de transcription à domaine homéo, la protéine Engrailed, peut réprimer l'expression du gène *pax-6*, et ce de manière indépendante de sa fixation à l'ADN. L'implication du gène *engrailed*, dans la régulation négative du gène *pax-6* nous avait été suggérée d'une part, par les territoires d'expression le plus souvent mutuellement exclusifs entre les deux gènes, d'autre part, par l'analyse des résultats obtenus sur des greffes poulet/caille. La greffe de régions mésodermiques de caille, - exprimant *engrailed*, mais pas *pax-6*, dans des régions diencéphaliques de poulet exprimant *pax-6* mais pas *engrailed* -, induit l'extinction de l'expression du gène *pax-6* chez l'hôte. La protéine Engrailed était par ailleurs connue pour réprimer l'expression de nombreux gènes. Nous avons donc cherché à savoir si cette protéine réprimait aussi l'expression de *pax-6* chez la caille. L'introduction de vecteurs exprimant la protéine Engrailed dans des cellules de neurorétine a en effet permis de montrer une extinction de l'expression du gène *pax-6*. Nous avons montré que la protéine Engrailed interagissait avec le domaine paired de la protéine P46 de *pax-6* et diminuait ainsi l'affinité de cette dernière pour sa séquence de reconnaissance sur l'ADN. Or la P46<sup>*pax-6*</sup> se fixe sur les promoteurs Po et P1 du gène *pax-6* et autoactive sa transcription. La protéine Engrailed perturbe par conséquent, cette boucle d'auto-régulation et l'expression du gène diminue (cf figure A).

Nous avons pu montrer par ailleurs, que la régulation de l'expression du gène *pax-6* ne s'exerçait pas seulement sur l'activité des promoteurs du gène, mais qu'elle pouvait également impliquer un mécanisme de régulation de la traduction des protéines Pax-6. Nous avons en effet caractérisé un ARNm

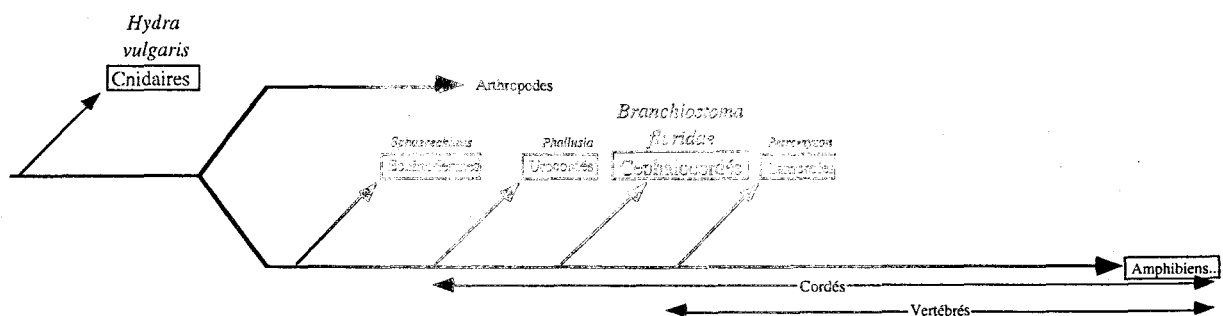




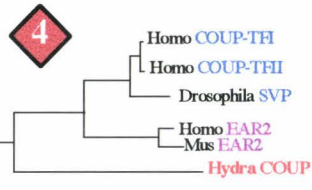
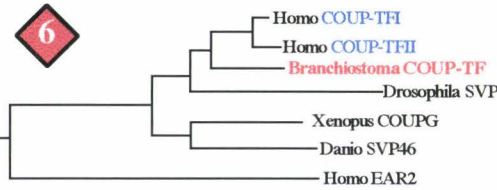
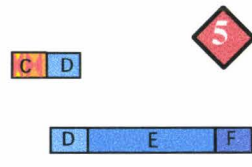



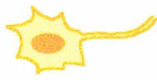

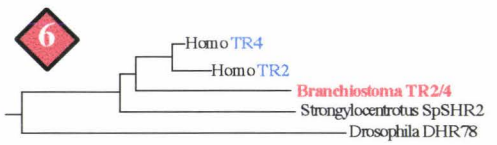




particulier du gène *pax-6* de caille, l'ARNm B1 dans le troisième article présenté dans ce mémoire (et cf. figure A). Cet ARNm, malgré une longue phase ouverte de lecture en 5', ne présente pas de codon d'initiation de traduction de type AUG avant la région située entre sa boîte *paired* et sa boîte *homéo*. Il possède néanmoins quatre codons alternatifs de traduction et produit en forte quantité seulement des petites protéines sans domaine *paired*. Ces observations suggèrent l'existence d'un mécanisme de "leaky scanning" qui consiste en l'accrochage des ribosomes par des codons alternatifs de traduction qui glissent ensuite jusqu'aux codons internes de type AUG et stimulent la traduction à partir des codons internes. Ce mécanisme de traduction offre un rendement meilleur pour la traduction des petites protéines initiées à partir des codons internes de traduction que celle observée avec d'autres ARNm de *pax-6* comportant un AUG en amont de leur boîte *paired*. L'ARNm B1 et les petites protéines sans domaine *paired* sont plus fortement exprimés dans la neurorétine et dans le pancréas, contrairement aux protéines Pax-6 à domaine *paired* et *homéo*, rencontrées en majorité dans le tube neural. Ces petites protéines ne reconnaissent pas les cibles classiques de fixation à l'ADN des protéines à domaine *homéo* et leur rôle reste encore à déterminer. Néanmoins, l'existence d'un mécanisme aussi fin pour la régulation de leur traduction, la proportion non négligeable d'ARNm B1 et de petites protéines rencontrées dans l'organisme souligne l'intérêt présenté par ces protéines. Chez le nématode, ces petites protéines existent, bien que produites à partir d'un promoteur interne et sont impliquées dans la détermination des organes sensoriels. Leur mécanisme d'action moléculaire reste encore à découvrir, elles pourraient interagir avec d'autres protéines à domaine *homéo* ou des cofacteurs transcriptionnels pour se fixer à l'ADN. L'étude de ces isoformes sans domaine *paired* constitue un axe de recherche intéressant et nouveau pour comprendre la régulation et la fonction du gène *pax-6*.

A la lumière de ces résultats, on ne peut perdre de vue que pour étudier la fonction d'un gène, il faut tenir compte de la quantité de chacune des isoformes produites par type de cellules étudié. De plus, l'expression de ces isoformes dépendra des facteurs *trans* présents dans la cellule et capables interagir avec les éléments de régulation *cis* du gène considéré. De même, les protéines exprimées dans le type cellulaire considéré peuvent aussi interagir avec les protéines Pax-6

traduites et interférer avec leur fonction. Dans la Discussion, nous replaçons les résultats obtenus sur Pax-6 dans le contexte des résultats publiés dans la littérature sur le développement embryonnaire.

L'évolution peut par ailleurs, constituer un excellent modèle pour l'étude de la fonction d'un gène dans un système donné. Les gènes sont en effet moins nombreux chez les organismes précoces, puisqu'ils n'ont pas encore subi toutes les étapes de duplication qui ont conduit aux nombreux gènes paralogues chez les vertébrés les plus récents. On peut alors déterminer la fonction commune à tous les métazoaires et replacer son rôle dans un "réseau de régulation", moins complexe chez des organismes précoces. Nous avons, par conséquent, choisi dans un deuxième temps, de caractériser deux autres facteurs de transcription COUP-TF et TR2/4 chez des organismes précoces dans l'échelle évolutive des métazoaires. COUP-TF et TR2/4 appartiennent à la famille des récepteurs nucléaires, une grande famille de gènes composée de plus de 60 membres dont l'activité transcriptionnelle est induite par un ligand. Cependant à côté de ces récepteurs qui fixent des petites molécules hydrophobes, comme les hormones, il existe de nombreux récepteurs pour lesquels aucun ligand n'a pu être identifié, appelés récepteurs orphelins. Dans l'introduction de la deuxième partie de ce manuscrit, nous présenterons cette grande famille des récepteurs, leur mode d'interaction à l'ADN, leur capacité à fixer un ligand et comment ils régulent la transcription. Nous aborderons ensuite, dans un dernier chapitre, l'évolution de ces récepteurs et les deux modèles choisis pour notre étude, une hydre, *Hydra vulgaris* et un amphioxus, *Branchiostoma floridae*, deux organismes situés à des positions clés dans l'échelle évolutive des métazoaires.



COUP-TF	 <p>Hydre</p>	 <p>Amphioxus</p>	Vertébrés
Position phylogénétique			
Isoformes			
Rôle dans le système nerveux	 <p>Exprimé dans des cellules de morphologie neuronale</p> <p style="text-align: right;">5</p>	 <p>Exprimé dans les neurones de la corde neurale mais pas dans la vésicule cérébrale.</p> <p style="text-align: right;">6</p>	 <p>COUP-TF I et II sont exprimés dans les neurones du tube neural et dans les vésicules cérébrales.</p>
Interférences avec la voie des rétinoïdes	<p style="text-align: center;">?</p>	<p>Réprime la transactivation médiée par les acides rétinoïques</p> <p style="text-align: right;">6</p>	<p>Répriment la transactivation médiée par les acides rétinoïques</p>
TR2/4	Position phylogénétique		
	Isoformes		
	Rôle dans le système nerveux	 <p>Exprimé dans la vésicule cérébrale.</p> <p style="text-align: right;">6</p>	 <p>TR4 est exprimé dans les vésicules cérébrales.</p>
	Interférences avec la voie des rétinoïdes	<p style="text-align: center;">?</p> <p style="text-align: right;">6</p> <p>L'acide rétinoïque réprime l'expression d'Amphi TR2/4</p>	<p>TR4 réprime la transactivation médiée par les acides rétinoïques</p> <p style="text-align: center;">?</p>

4

Ligand binding was acquired during the evolution of nuclear receptors  
Hector Escriva, Rachid Safi, Catherine Hanni, Marie-Claire Langlois, et al.  
*Proc. Natl. Acad. Sci. USA*, 1997 (article 4)

5

A unique homologue for COUP-TF and EAR2 orphan nuclear receptors in Hydra.  
Marie-Claire Langlois et al.  
*En préparation pour "Genes, Development and Evolution"* (article 5)

6

AmphiCOUP-TF and AmphiTR2/4, two nuclear orphan receptors of the lancelet *Branchiostoma floridae*, are implicated in retinoic acid signaling pathways.  
Marie-Claire Langlois et al.  
*En préparation pour "Development"*, (article 6)

Figure B

Dans la partie résultats, nous étudierons la fonction des deux récepteurs orphelins COUP-TF et TR2/4, chez l'hydre et l'amphioxus, deux récepteurs impliqués dans le développement du système nerveux central chez les vertébrés. Cette étude a été réalisée dans le groupe de Vincent Laudet. Pax-6 et COUP-TF sont en l'occurrence co-exprimés dans les vésicules antérieures du cerveau et dans les neurones moteurs de la moelle épinière au cours du développement embryonnaire. TR2/4 s'exprime aussi dans le cerveau chez l'embryon de souris. Etudier la fonction de COUP-TF chez l'hydre, un métazoaire très précoce, permettra d'identifier s'il existe une fonction commune de COUP-TF chez tous les métazoaires. De plus, l'étude de la fonction de COUP-TF et de TR2/4 chez l'amphioxus, un céphalocordé situé juste avant l'embranchement des vertébrés, permettra de comprendre le rôle de ces gènes dans le développement du cerveau au cours de l'évolution des vertébrés. Nous avons donc isolé les homologues de COUP-TF chez l'hydre et l'amphioxus et l'homologue de TR2/4 chez l'amphioxus.

Dans le quatrième article, nos résultats ont permis d'appuyer l'hypothèse des premières étapes de la duplication des récepteurs nucléaires avec l'émergence des métazoaires. La position phylogénétique de COUP-TF d'hydre située avant la duplication ayant conduit aux deux groupes de récepteurs EAR2 et COUP-TF, suggère en effet, une première étape de diversification des récepteurs nucléaires pouvant corrélérer avec l'émergence des métazoaires (figure B).

Dans le cinquième article, nous montrons que l'homologue de COUP-TF produit plusieurs isoformes chez l'hydre, contrairement aux COUP-TFs<sub>1</sub> de vertébrés. Ce gène est par ailleurs exprimé dans des cellules peu différenciées, comportant néanmoins des prolongements cytoplasmiques évocateurs d'une détermination de ces cellules vers la lignée neuronale (figure B).

Les résultats qui font l'objet de l'article 6 ont permis de montrer un rôle conservé de l'homologue des COUP-TF I et II de vertébrés, AmphiCOUP-TF, dans la détermination des cellules neuronales chez l'amphioxus. AmphiCOUP-TF est en effet exprimé dans la corde neurale chez la larve tardive. De plus, l'absence d'expression de AmphiCOUP-TF dans la vésicule cérébrale du céphalocordé suggère une acquisition plus récente de l'implication de COUP-TF dans la segmentation du diencéphale, comme observée chez les vertébrés. Cette

différence pourrait être une des raisons pour lesquelles le cerveau de l'amphioxus est si restreint comparé à celui des vertébrés. Amphi TR2/4, l'homologue des TR2 et TR4 de vertébrés est quant à lui, exprimé très tôt durant la neurulation et retrouvé dans la vésicule cérébrale de l'amphioxus et dans les fentes branchiales chez la larve (figure B). Ces observations suggèrent un rôle de AmphiTR2/4 dans la vésicule cérébrale et dans les fentes branchiales au cours du développement de l'amphioxus.

Nous avons par ailleurs, étudié l'interférence entre AmphiCOUP-TF et Amphi TR2/4 sur la voie de signalisation des acides rétinoïques. COUP-TFI et II répriment en effet, la voie des rétinoïdes chez les vertébrés. Les acides rétinoïques présentent notamment un gradient de concentration antéro-postérieur chez les vertébrés et engendrent des malformations dans le tube neural antérieur lorsqu'ils sont injectés dans des blastulas précoces. Ils empêchent par ailleurs la formation des fentes branchiales chez l'amphioxus et de ce fait, pourraient interférer avec AmphiTR2/4 dans cette région. Nous avons pu montrer que AmphiCOUP-TF, comme ses homologues de vertébrés, réprimait la transactivation médiée RAR et/ou RXR en présence de leurs ligands (par cotransfection transitoire dans des cellules Cos). D'autre part, les acides rétinoïques injectés au stade Blastula, répriment l'expression de AmphiTR2/4 à la fois dans les fentes branchiales mais aussi dans la vésicule cérébrale de l'amphioxus (figure B). En conséquence, nos résultats montrent que AmphiCOUP-TF et AmphiTR2/4 interfèrent avec la voie de signalisation des acides rétinoïques; AmphiCOUP-TF réprime la transactivation médiée par les acides rétinoïques, *in vitro* et les acides rétinoïques répriment *in vivo* l'expression de AmphiTR2/4.

La conservation de la fonction d'interférence du gène AmphiCOUP-TF sur la voie de signalisation des acides rétinoïques suggère que celle-ci pourrait être conservée chez les céphalocordés. Jusqu'à présent, l'existence de cette voie de régulation n'a été mise en évidence que chez les vertébrés.

Par ailleurs, la répression de l'expression des gènes TR2 ou TR4 par les acides rétinoïques n'a pas encore été mise en évidence. Néanmoins, peu de choses ont été publiées sur les gènes TR2 et TR4 à nos jours et il est fort probable que ce mode de régulation existe aussi chez les vertébrés.

Nos résultats suggèrent de plus, que la fonction commune des COUP-TFs chez tous les métazoaires est de participer à la détermination des cellules neuronales.

Curieusement, les patrons d'expression spatio-temporels de *pax-6* et COUP-TF présentent des analogies décrites dans la discussion générale de ce manuscrit. L'homologue de *pax-6* a été isolé par d'autres groupes de recherche chez l'hydre et l'amphioxus. Il serait intéressant dans un futur proche d'établir les liens directs ou indirects existant entre Pax-6 et COUP-TF. Par ailleurs, dans un futur plus lointain on peut imaginer que les réseaux de régulation impliqués chez l'hydre dans la régénération et la détermination neuronale pourraient être utilisés chez l'homme pour régénérer des neurones endommagés, qui malheureusement ont perdu leur capacité à se renouveler chez les mammifères et nombre de métazoaires.

Cette thèse représente une étape dans la compréhension des mécanismes de régulation qui conduisent à la différenciation neuronale. Ces mécanismes nécessiteront encore le travail de nombreux groupes de recherche durant plusieurs années et une interdisciplinarité entre l'embryologie, l'évolution et les neurosciences pour comprendre le rôle des gènes impliqués dans les fonctions neuronales.

# SOMMAIRE

## Introduction des gènes Pax

I	Présentation des gènes Pax	1
1	Comment l'étude des gènes du développement chez la drosophile ont conduit à la découverte des gènes Pax	1
1-1	<u>Les gènes homéotiques</u>	1
1-2	<u>Les gènes de segmentation</u>	2
1-3	<u>Les gènes Pax</u>	2
1-4	<u>Bilan des grands réseaux de régulation gouvernant le développement de <i>Drosophila melanogaster</i></u>	3
2	Importance des gènes Pax dans le développement embryonnaire	3
2-1	<u>La famille des gènes Pax</u>	3
2-2	<u>Expression des gènes Pax</u>	4
	a) Expression des gènes Pax dans le système nerveux central	4
	b) Expression des gènes Pax dans le tube neural	5
2-3	<u>Les mutants du développement induits par les gènes Pax</u>	6
	a) mutations dans le gène <i>pax-1</i> ( <i>undulated</i> )	6
	b) mutations dans le gène <i>pax-2</i> (Krd)	6
	c) mutations dans le gène <i>pax-3</i> ( <i>splotch;Sp</i> )	7
	d) mutations dans le gène <i>pax-4</i>	8
	e) mutations dans le gène <i>pax-5</i>	8
	f) mutations dans le gène <i>pax-6</i> , <i>small eye</i>	8
	g) mutations dans le gène <i>pax-8</i>	9
	h) <i>pax-9</i>	9
2-4	<u>Conclusion</u>	10
II	Paired & homéo, deux domaines de fixation à l'ADN	10
1	Le domaine homéo	10
1-1	<u>Le domaine homéo, généralités</u>	10
1-2	<u>Structure tridimensionnelle du domaine homéo</u>	11
	a) Structure tridimensionnelle générale	11
	b) acides aminés conservés indispensables à la reconnaissance du site TAAT et à la liaison du domaine homéo à l'ADN	11
	c) acides aminés liés à la spécificité de reconnaissance des domaines homéo	12
	- résidu 9 de l'hélice de reconnaissance	12
	- bras amino terminal	13
	- acides aminés non conservés/ liaison avec des cofacteurs	14

d) Conclusion sur la fixation du domaine homéo à l'ADN	15
1-3 <u>Le complexe de fixation : ADN - domaine homéo de type paired</u>	15
a) Séquence cible de fixation <i>in vitro</i>	15
b) Structure tridimensionnelle	16
- <i>structure tridimensionnelle générale : apparition d'un dimère</i>	16
- <i>liaisons impliquées dans l'interface du dimère</i>	17
c) Conclusion sur la fixation à l'ADN du domaine homéo de type paired	17
<b>2 Le domaine paired</b>	18
2-1 <u>Structure tridimensionnelle du domaine paired</u>	18
a) Structure tridimensionnelle générale	18
b) Acides aminés impliqués dans la liaison à l'ADN	18
c) Conclusion	19
2-2 <u>Etudes biochimiques</u>	19
a) Rôle de PAI	19
- <i>Les mutations de substitutions des gènes Pax sont situées dans PAI</i>	20
- <i>PAI est nécessaire pour la capacité de liaison à l'ADN du domaine paired</i>	21
- <i>conclusion : rôle de PAI dans la fixation à l'ADN</i>	22
b) Rôle de RED	22
- <i>Le cas des protéines Pax au domaine RED "actif"</i>	22
- <i>Le cas des protéines Pax "RED inhibé"</i>	23
c) Bilan sur la fixation du domaine paired : rôles respectifs de PAI et RED	24
<b>3 Coopérativité entre le domaine paired et le domaine homéo des protéines Pax pour la fixation à l'ADN</b>	24
3-1 <u>Séquence cible de fixation</u>	25
3-2 <u>Etudes biochimiques</u>	25
<b>4 Conclusions sur la fixation des protéines Pax</b>	25
<b>III Fonction des gènes PAX</b>	26
<b>1 Les gènes Pax sont des facteurs de transcription</b>	26
1-1 <u>Les gènes Pax ont une activité transactivatrice</u>	26
a) Le domaine transactivateur est situé dans la région C terminale	26
b) Modulation intramoléculaire de l'activité transactivatrice des protéines Pax	27
- <i>activité inhibitrice intramoléculaire</i>	27



- <i>synergie activatrice entre homodimères</i>	28
c) Conclusion sur l'activité transactivatrice des gènes Pax	28
<b>1-2 Effet-dose des gènes Pax</b>	28
a) Caractère haplo-insuffisant des gènes Pax	28
b) La surexpression d'un gène Pax cause des anomalies du développement	29
<b>2 Gènes cibles des protéines Pax</b>	29
- <i>les gènes qui interviennent dans la mise en place du système nerveux</i>	29
- <i>les gènes qui interviennent dans l'organogénèse</i>	30
<b>IV Le cas du gène <i>pax-6</i></b>	30
<b>1 <i>pax-6</i>, gène organisateur de l'oeil</b>	30
<b>2 Notre modèle d'étude, la neurorétine d'oiseau</b>	30
<b>3 Le gène Pax-QNR</b>	31
3-1 <u>Stratégie d'isolement</u>	31
3-2 <u>Structure génomique de Pax-QNR et protéines traduites</u>	32
3-3 <u>La régulation transcriptionnelle du gène Pax-QNR</u>	33
3-4 <u>Sujet de thèse</u>	33
<b>Résultats des gènes Pax</b>	34
<b>Article 1</b>	34
Discussion de l'article 1	36
<b>Article 2</b>	38
Discussion de l'article 2	40
<b>Article 3</b>	42
Discussion de l'article 3	44
<b>Discussion des gènes Pax</b>	45
<b>Références</b>	51

# Introduction des récepteurs nucléaires 56

## A Les récepteurs nucléaires, structure et fonction 56

### I Caractéristiques des récepteurs nucléaires 56

#### 1 Une famille de plus de 60 membres 56

1-1 Caractéristiques structurales d'un récepteur 56

1-2 Composition de la famille 56

#### 2 Mode de fixation à l'ADN 58

2-1 Vue d'ensemble 58

2-2 Structure du domaine de fixation à l'ADN 58

2-2-1) Les régions du domaine impliquées dans la fixation à l'ADN 58

2-2-2) Structure tridimensionnelle du domaine de liaison à l'ADN 59

2-3 Spécificité de reconnaissance à l'ADN 59

2-3-1) L'ADN : l'élément de réponse 59

a) Les palindromes et les palindromes inversés 59

b) Les répétitions directes 60

c) Les demi-sites 60

d) Flexibilité dans la spécificité de reconnaissance 60

2-3-2) Le récepteur : les boîtes impliquées 61

a) La boîte P 61

b) La boîte D 61

c) La boîte A 61

2-4 Fixation en monomère, homo- ou hétéro-dimère 62

2-4-1) Caractéristique des dimères 62

2-4-2) Exemples d'hétérodimères 62

2-4-3) Polarité des hétérodimères RAR et RXR 63

2-4-4) Interfaces de dimérisation 63

*homodimères* 63

*hétérodimères* 64

#### 3 La fixation du ligand 65

3-1 Une affinité variable du récepteur pour son ligand 65

3-2 Structure tridimensionnelle de l'interaction ligand-récepteur 65

3-3-1) Structure du domaine de fixation du ligand de l'apo RXR $\alpha$  65

3-3-2) Structure du domaine de fixation du ligand de l'holo RAR $\gamma$  65

3-3 Le coeur AF-2AD 66

<b>II</b>	<b>La modulation de la transcription</b>	<b>67</b>
<b>1</b>	<b><u>Principes généraux sur l'initiation de la transcription</u></b>	<b>67</b>
1-1	La machinerie transcriptionnelle de base	67
1-2	L'accessibilité de l'ADN	67
<b>2</b>	<b><u>Rôle des récepteurs dans l'activation de la transcription</u></b>	<b>68</b>
<b>2-1</b>	<b>Les domaines de transactivation sur le récepteur</b>	<b>68</b>
2-1-1)	Localisation des domaines AF-1 et AF-2	68
	<i>a) AF-1, situé dans le domaine A/B</i>	68
	<i>b) AF-2, situé dans le domaine E</i>	68
2-1-2)	Le coeur AF-2AD, une région d'interaction avec les acteurs de la transcription	69
2-1-3)	La régulation de l'activité transactivatrice : ligand et phosphorylation	69
	<i>a) régulation de l'activité de AF-2</i>	69
	<i>b) régulation de l'activité de AF-1</i>	70
<b>2-2</b>	<b>Recrutement de la machinerie transcriptionnelle de base</b>	<b>70</b>
2-2-1)	Recrutement direct	71
	<i>a) Recrutement de TFIIB</i>	71
	<i>b) Recrutement de TFIID</i>	71
2-2-2)	Recrutement indirect, association de co-facteurs	71
	<i>a) Quelques exemples de co-facteur</i>	72
	<i>b) motif d'interaction, co-facteur-récepteur</i>	73
<b>2-3</b>	<b>Association avec des co-facteurs modulant le degré d'acétylation des histones</b>	<b>73</b>
<b>3</b>	<b><u>Rôle des récepteurs dans la répression de la transcription</u></b>	<b>73</b>
<b>3-1</b>	<b>Les interférences transcriptionnelles</b>	<b>74</b>
<b>3-2</b>	<b>La répression directe</b>	<b>74</b>
3-2-1)	Domaine transrépresseur	74
3-2-2)	Les co-répresseurs associés aux récepteurs : N-CoR et SMRT	75
	<i>a) N-CoR</i>	75
	<i>b) SMRT</i>	76
	<i>c) Mode d'action des co-répresseurs</i>	76
<b>4</b>	<b><u>Conclusion</u></b>	<b>76</b>

<b>B</b>	<b>Evolution des récepteurs nucléaires</b>	<b>77</b>
<b>I</b>	<b>La diversification des récepteurs au cours de l'évolution</b>	<b>77</b>
1	<b><u>Etude phylogénétique des récepteurs</u></b>	77
2	<b><u>L'hypothèse de la duplication du génome</u></b>	78
2-1	La duplication comme moyen d'évolution	78
2-2	Les différents événements possibles de duplication	78
3	<b><u>Un autre exemple de choix pour l'étude des duplications géniques au cours de l'évolution : les gènes Hox</u></b>	79
3-1	Les gènes Hox au cours du développement embryonnaire	79
3-2	L'émergence du complexe Hox	80
3-3	La transition invertébrés/vertébrés	81
3-4	Le complexe des gènes ParaHox	81
4	<b><u>Les récepteurs orphelins, les ancêtres des récepteurs nucléaires classiques ?</u></b>	82
4-1	Arguments en faveur de cette hypothèse	82
4-1-1)	Pas de corrélation dans l'arbre entre récepteur et type de ligand considéré	82
4-1-2)	Un ligand peut être fixé par un récepteur de vertébré mais pas par son homologue chez un organisme plus précoce	83
4-1-3)	Seuls des récepteurs orphelins ont pu être isolés chez des organismes précoces	83
<b>II</b>	<b>Critères de choix des modèles d'études</b>	<b>84</b>
1	<b><u>Choix des modèles biologiques</u></b>	84
1-1	<b>L'invertébré le plus proche des vertébrés, l'amphioxus</b>	84
1-1-1)	Phylogénie moléculaire des cordés	84
1-1-2)	Caractéristiques physiques, développementales et évolutives de l'amphioxus	85
1-1-3)	Le système neveux central et les gènes du développement	86
	a) <i>La notion d'homologie</i>	86
	b) <i>Les gènes du développement, marqueurs des divisions du cerveau</i>	87
	c) <i>Les gènes Amphihox et la réponse aux acides rétinoïques</i>	90
1-2	<b>L'hydre, un métazoaire précoce</b>	90

1-2-1)	Caractéristiques structurales et morphologiques	90
1-2-2)	La régénération	91
	<i>a) le mécanisme de régénération</i>	91
	<i>b) Les voies de régulation impliquées dans la régénération</i>	91
	<i>c) Les gènes du développement sollicités lors de la régénération</i>	92
1-2-3)	Le système nerveux	92
	<i>a) Un réseau de cellules nerveuses</i>	92
	<i>b) L'Hydre possède-t-elle un système nerveux sensoriel?</i>	93
<b>2</b>	<b><u>Choix des récepteurs pour notre étude</u></b>	<b>93</b>
<b>2-1</b>	<b>Les COUP-TF : des récepteurs orphelins très conservés</b>	<b>93</b>
2-1-1)	Isolement	93
2-1-2)	Patron d'expression des gènes COUP-TF	94
	a) vésicules cérébrales	94
	b) Cellules de crêtes neurales et dérivés	95
	c) Expression dans la moelle épinière	95
	d) Expression pendant l'organogénèse	95
	e) Expression chez l'adulte	96
	f) Expression dans d'autres organismes	96
2-1-3)	Mutants des gènes COUP-TF	96
2-1-4)	COUP-TF interfère aussi avec d'autres récepteurs nucléaires	96
<b>2-2</b>	<b>Les récepteurs orphelins TR2 et TR4</b>	<b>97</b>
2-2-1)	Isolement	97
	<i>a) TR2</i>	97
	<i>b) TR4</i>	97
2-2-2)	Expression	97
	<i>a) Au cours de l'embryogénèse</i>	97
	<i>b) Chez l'adulte</i>	98
2-2-3)	Mécanisme d'action de TR2 et TR	99
	<b>Résultats des récepteurs nucléaires</b>	<b>100</b>
	Article 4	100
	Discussion de l'article 4	102
	Article 5	104
	Discussion de l'article 5	106
	Article 6	108
	Discussion de l'article 6	110
	<b>Bibliographie des récepteurs nucléaires</b>	<b>114</b>
	<b>Conclusion générale</b>	<b>122</b>

*Première partie*

*Introduction*  
*gènes Pax*

## I Présentation des gènes Pax

### 1 Comment l'étude des gènes du développement chez la drosophile

#### ont conduit à la découverte des gènes Pax

##### 1-1 Les gènes homéotiques

La Biologie du Développement a réalisé des progrès considérables ces dix dernières années grâce à la Biologie Moléculaire. Ce grand bond en avant a été favorisé par les précédentes études génétiques portées sur les mutants du développement chez *Drosophila melanogaster* (Poulsen, 1950). La mutation homéotique est une mutation fréquente chez la Drosophile. Chez les insectes mutants, certaines parties du corps de l'animal sont remplacées par d'autres parties normalement portées par des segments placés ailleurs sur l'axe antéro-postérieur. Soit le segment est transformé en un segment qui lui est antérieur (mutation perte de fonction) ou en un segment plus postérieur (mutation gain de fonction). Dans les deux cas, l'insecte se retrouve avec deux segments identiques, d'où l'appellation de mutations homéotiques, le terme *homéo* signifiant semblable. Par exemple, des pattes remplacent des antennes ou des éléments de la bouche comme dans le cas de la mutation *Antennapedia* (Gehring, 1966); (Lewis, 1978); (Garber *et al.*, 1983).

Les chercheurs ont tout d'abord repéré le locus ou les loci à l'origine de ces mutations (Nüsslein-Volhard et Wieschaus, 1980). De cette façon, ils ont aussi pu mettre en évidence l'agencement des gènes homéotiques. Ces gènes sont regroupés sur un même chromosome chez la drosophile et forment le complexe HOM-C (**H**OMeotic **C**omplex). Ces observations faites à partir des études génétiques (Lewis, 1978) seront confirmées par la suite par la biologie moléculaire (Kaufman *et al.*, 1990). Par la technique de marche sur le chromosome, ils ont ensuite isolé les gènes à l'origine des fonctions perdues (Bender *et al.*, 1983). En comparant les séquences des gènes obtenus, ils ont mis en évidence la présence de motifs conservés (appelés "boîte" sur une séquence nucléique ou "domaine" sur une séquence protéique. De cette manière, le domaine homéo des gènes homéotiques a été le premier domaine conservé, identifié chez *Drosophila melanogaster* (Scott et Weiner, 1984).


























### 1-2 Les gènes de segmentation

Les gènes homéotiques, de par le phénotype induit par leur mutation sont classés comme étant les gènes responsables de l'identité des segments chez l'insecte. En parallèle des gènes homéotiques, d'autres gènes impliqués dans la métamérisation de l'embryon de *Drosophila melanogaster* ont été isolés. Ces gènes, dénommés gènes de segmentation, se composent des gènes "gap", des gènes "pair-rule" et des gènes "polarity gene". Les mutations "gap" entraînent la perte d'un segment; les mutations "pair-rule", la disparition d'un segment sur deux; et les mutations "polarity gene", agissent sur la polarité antéro-postérieure des segments (Nüsslein-Volhard et Wieschaus, 1980; Nüsslein-Volhard *et al.*, 1985).

### 1-3 Les gènes Pax

Le gène *paired*, isolé appartient à la classe des gènes "pair-rule". Il comporte un domaine homéo, mais il est localisé sur un chromosome différent du complexe HOM-C, chez la Drosophile. Bopp et Frigerio *et al.* en 1986 ont découpé le gène *paired* nouvellement identifié, en plusieurs fragments avec lesquels ils ont criblé des banques chez *Drosophila*. Frigerio *et al.* ont sorti 11 loci en utilisant la partie 3' du gène (Frigerio *et al.*, 1986); (Bopp *et al.*, 1986). Dans cette région du gène, en aval du domaine homéo, se situe le domaine défini comme PRD (pour "Proline Repeat Domain"), constitué d'une simple répétition Histidine-Proline, sur une vingtaine d'acides aminés. Le domaine PRD a été identifié en même temps que le domaine homéo, domaine de 60 acides aminés (Scott et Weiner, 1984). Parmi les 11 loci isolés, Frigerio *et al.* ont retrouvé le gène *bicoid* (Carroll *et al.*, 1986), un gène maternel impliqué dans les étapes précoces du développement embryonnaire, ce qui validait le concept du réseau génétique. Ce concept supposait en effet, que des protéines à domaines fonctionnels similaires pouvaient être impliqués dans les mêmes réseaux de régulation de gènes. Bopp *et al.* en utilisant la partie 5' du gène *paired* ont pu mettre en évidence parmi les gènes isolés, l'existence d'un troisième motif conservé, le motif *paired*, un domaine constitué de 128 acides aminés sur la protéine. Les gènes qui comportent le motif *paired* sont appelés gènes Pax, "Pax" signifiant "paired box". *gooseberry* et *gooseberry neuro* étaient parmi les gènes isolés par Bopp *et al.* ; ils comportent aussi ce motif *paired*. Ils s'appelaient à l'époque, BSH9 pour *gooseberry* (*gsb*) et BSH11 pour *gooseberry neuro* (*gsbn*) (Bopp *et al.*, 1986) puis (Gutjahr *et al.*, 1993).

GENE	CHROMOSOME		STRUCTURE			CLASSE
	Souris	Humain	Prd	OCT	HD	
<b>Pax-1</b>	2	20p11				<b>I</b>
<b>Pax-2</b>	19	10q25				<b>III</b>
<b>Pax-3</b>	1	2q35				<b>II</b>
<b>Pax-4</b>	6	7q22-qter				<b>IV</b>
<b>Pax-5</b>	4	9p13				<b>III</b>
<b>Pax-6</b>	2	11p13				<b>IV</b>
<b>Pax-7</b>	4	11p36.2				<b>II</b>
<b>Pax-8</b>	2	2q12-q14				<b>III</b>
<b>Pax-9</b>	12	14q12-q13				<b>I</b>

**Figure 1: Division en 4 classes de la famille des gènes *pax*, en fonction de leur structure (boîtes Prd de même couleur par classe). Prd: domaine paired, OCT: octapeptide, HD: homéodomaine.**

## 1-4 Bilan des grands réseaux de régulation gouvernant le développement de *Drosophila melanogaster*.

Pour résumer, l'embryogenèse chez la Drosophile est gouvernée par des réseaux de gènes qui régulent d'une manière spatio-temporelle, les différentes étapes du développement. Les gènes maternels comme le gène *bicoid*, sont les premiers gènes activés. Ils s'expriment dans l'oeuf sous forme d'un gradient de concentration. Ce gradient est à l'origine de l'établissement de l'axe antéro-postérieur de l'embryon et de l'activation des gènes de segmentation. Ces derniers sont divisés en trois classes; les gènes "gap", les gènes "pair-rule" et les gènes de "polarité de segment". Chacune de ces classes établit une segmentation séquentielle et ordonnée l'embryon, débutant par une segmentation "grossière" pour les gènes "gap" à une segmentation en double segment pour les gènes "pair-rule". Les gènes de "polarité de segment" achèvent la métamérisation par la segmentation unitaire. L'identité des segments est ensuite définie par l'activation des gènes homéotiques. Les gènes Pax, auxquels appartient aussi le gène "pair-rule" *paired* de Drosophile, sont impliqués dans l'organisation du système nerveux chez les vertébrés. Le chapitre suivant traitera de l'importance des gènes Pax dans le développement embryonnaire.

## **2 Importance des gènes Pax dans le développement embryonnaire.**

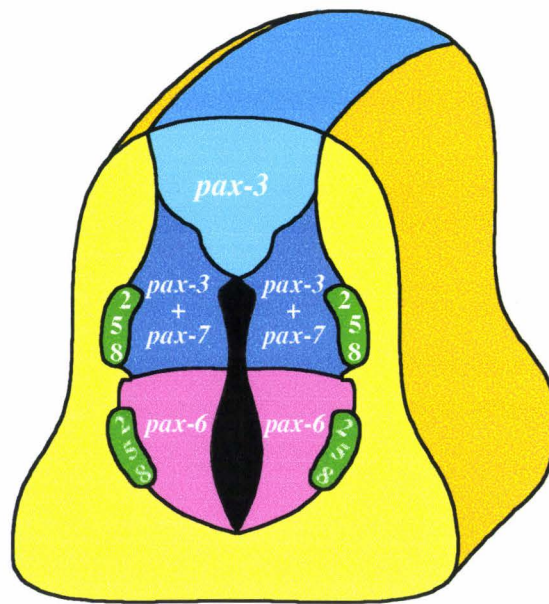
Chez les vertébrés, 9 gènes Pax ont été isolés, pour la plupart, par homologie avec le domaine *paired* du gène *paired* de Drosophile.

### 2-1 La famille des gènes Pax

Les gènes Pax sont divisés en quatre groupes de part leur identité de séquences et leur organisation moléculaire (Noll, 1993). Tous contiennent bien sûr, le motif *paired* de 128 acides aminés qui caractérise la famille des gènes Pax. Ce domaine est situé en N terminal de la protéine (cf. figure 1)

Chez les vertébrés, le groupe I est constitué des gènes *pax-1* et *pax-9*. Ils codent des protéines composées en plus du domaine *paired*, d'un motif octapeptide.

Les gènes du groupe II, *pax-3* et *pax-7*, contiennent un domaine *paired* et un domaine homéo entier et un domaine octapeptide.



**Figure 2 : Les gènes *pax* dans le tube neural de Mammifères**

Carte d'expression des gènes *pax* dans la moëlle épinière d'un embryon de Souris de 11jours.

*pax-3* est exprimé dans la région dorsale et *pax-6* dans la région ventrale.

Les protéines codées par les gènes *pax-2*, *pax-5* et *pax-8* contiennent en plus du domaine paired, un motif octapeptide et un domaine homéo partiel, tronqué en C terminal. Ces gènes constituent le groupe III de la famille des gènes Pax.

*pax-4* et *pax-6* sont dépourvus de cet octapeptide, mais ils contiennent, comme les gènes du groupe III, un domaine paired et un domaine homéo entiers; ils constituent le groupe IV.

L'octapeptide est un domaine qui participe à la dimérisation des protéines homéotiques chez la Drosophile. Les domaines paired et homéo sont des domaines de fixation à l'ADN, une caractéristique qui sera abordée dans le paragraphe suivant.

## 2-2 Expression des gènes Pax.

Tous les gènes Pax, excepté *pax-1* et *pax-9* (Deutsch *et al.*, 1988); (Stapleton *et al.*, 1993); (Wallin *et al.*, 1993; Wallin *et al.*, 1994), présentent une expression spatiale et temporaire dans le système nerveux au cours du développement embryonnaire (cf. figure 2)

### a) *Expression des gènes Pax dans le système nerveux central.*

Les gènes Pax s'expriment dans le tube neural en développement (Goulding *et al.*, 1991); (Walther *et al.*, 1991). *pax-6*, comme *pax-3* et *pax-7* s'expriment très tôt dans le développement embryonnaire: à partir de 8,5 jours *post-coïtum*. Ces trois gènes Pax s'expriment tout le long du tube neural, à l'exception de *pax-6* qui ne s'exprime pas dans le mésencéphale. *pax-6* continue de s'exprimer dans le cerveau antérieur jusqu'à des stades tardifs du développement et même après la naissance (Stoykova et Gruss, 1994). Par contre, l'expression des gènes *pax-3* et *pax-7* dans le cerveau se restreint dans le télencéphale au cours de l'embryogenèse, et se superpose ensuite à la région rostrale du diencéphale.

*pax-2*, *pax-5* et *pax-8* s'expriment plus tardivement; à partir de 10,5 jours post-coïtum et sont confinés au mésencéphale/métencéphale et à la moelle épinière. Leur expression est localisée dans les cellules post-mitotiques de la zone intermédiaire. La frontière entre le cerveau moyen et le cerveau postérieur correspond à la limite rostrale d'expression des gènes *pax-2*, *pax-5* et *pax-8*. Néanmoins *pax-5* est aussi exprimé dans *tegmentum* mésencéphalique postérieur.

L'analyse de l'activité des gènes Pax dans le cerveau en développement révèle une expression spatiale délimitée par les frontières anatomiques. Chez le zébrafish, des

anticorps dirigés contre la protéine Pax[zf-b] (l'homologue de Pax-2) induisent des défauts du développement de l'isthme qui relie le cerveau moyen au cerveau postérieur (Krauss *et al.*, 1992). Ces expériences suggèrent que les membres de la famille des gènes Pax pourraient jouer un rôle dans l'établissement de la ceinture moyenno-postérieure du cerveau. Par ailleurs, des modèles récents décrivent l'organisation du cerveau antérieur en quatre (D1-D4, (Figdor et Stern, 1993) ou six (p1-P6, (Puelles et Rubenstein, 1993) neuromères. Or, les limites d'expression des gènes *pax-3*, *pax-6* et *pax-7* sont consistantes avec l'existence de ces neuromères.

L'expression des gènes Pax est également détectée dans le cerveau adulte. La distribution des ARN messagers chez l'adulte est comparable à celle des embryons en cours de gestation le long de l'axe antéro-postérieur. L'expression des gènes Pax chez l'adulte suggère une implication des gènes Pax, non seulement dans le développement embryonnaire et la régionalisation du cerveau, mais aussi dans le maintien de l'identité régionale.

#### *b) Expression des gènes Pax dans le tube neural.*

Les gènes Pax pourvus d'un domaine homéo complet sont les premiers à s'exprimer dans le tube neural (Goulding *et al.*, 1991); (Jostes *et al.*, 1990) . Ils sont détectés à partir de 8,5 jours post-coïtum dans des régions spécifiques de la zone ventriculaire. Tandis que *pax-6* est présent dans les plaques basales et intermédiaires, *pax-3* et *pax-7* sont exprimés dans les plaques alaires. Les domaines d'expression de ces trois gènes Pax se superposent partiellement (figure 2) Ce type d'expression région spécifique suggère que les gènes *pax-3*, *pax-6* et *pax-7* pourraient jouer un rôle important durant l'établissement de la polarité dorso-ventrale de la moelle-épineuse.

Le second sous groupe de gènes Pax, sans domaine homéo, (*pax-2*, *pax-5* et *pax-8*) commence à s'exprimer à partir de 10,5 jours post-coïtum dans un sous type de neuroblastes post-mitotiques, agencés en deux colonnes dans la substance grise intermédiaire, de chaque côté du *sulcus limitans* de la moelle épinière (figure 2) (Nornes *et al.*, 1990); (Asano et Gruss, 1992). L'analyse de l'expression de *pax-2* comparée à celle de son homologue de zébrafish *pax-[zf-b]*, a identifié ces neurones comme étant des neurones intermédiaires (Krauss *et al.*, 1991) ; (Mikkola *et al.*, 1992) Cette observation suggère que les gènes Pax pourraient être impliquées dans la détermination des cellules vers un destin neuronal particulier dans la moelle épinière.

<b>GENE</b>	<b>PHENOTYPE PERTE DE FONCTION</b>	<b>PHENOTYPE GAIN DE FONCTION</b>
PAX1	Homme ? Souris : <i>undulated</i>	?
PAX9	?	?
PAX2	Homme : problèmes réiniens et rénaux  Souris : Krd	Homme : implication dans la tumeur de Wilms ?  Souris : anomalies du développement du rein
PAX5 (BSAP)	Souris : anomalies du cerveau et des cellules B	Homme : implication dans des glioblastomes et des médulloblastomes?
PAX8	?	Homme : implication dans la tumeur de Wilms ?
PAX3	Homme : WS1, WS3, syndrome craniofacial-surdité Souris : Splotch	Rhabdomyosarcome
PAX7	?	?
PAX4	absence de cellules matures productrices d'insuline et de somatostine	?
PAX6	Homme : aniridie, anomalie de Peter, cataracte Souris : small-eye	?

**Figure 3 : Syndromes malformatifs humains et défauts du développement murin associés à des mutations dans les gènes *pax* (complété, d'après Stuart et Gruss, 1995).**

L'importance des gènes Pax dans l'embryogenèse est illustrée par les défauts du développement induits par leurs mutations.

### 2-3 Les mutants du développement induits par les gènes Pax

Les mutations des gènes Pax, même sur un seul allèle, entraînent des malformations causées par des défauts du développement (figure 3).

#### *a) mutations dans le gène pax-1 (undulated)*

Les gènes *pax-1* et *pax-9* sont les seuls gènes Pax à ne pas être exprimés dans le système nerveux central (Deutsch *et al.*, 1988); (Wallin *et al.*, 1993). A 8,25 jours post-coïtum, l'expression de *pax-1* est tout d'abord détectée dans la région ventrale des somites (le futur sclérotome); puis dans la colonne vertébrale et le squelette facial.(Deutsch *et al.*, 1988); (Timmons *et al.*, 1994).

Le gène *pax-1* a été localisé sur le chromosome 2 (cf. figure 1) chez la souris, autour de la mutation *undulated* (Balling *et al.*, 1988). L'analyse des souris porteuses de cette mutation révèle qu'il existe trois allèles possibles présentant des mutations dans le gène *pax-1* responsables du phénotype *undulated*. Le phénotype responsable du phénotype *undulated* est caractérisé par une queue tordue et des défauts de la colonne vertébrale (Wallin *et al.*, 1994; Dietrich et Gruss, 1995). Les mutants présentent également une diminution de la taille du thymus et une altération dans la maturation des thymocytes (Spörle *et al.*, 1996). Outre, l'implication de *pax-1* dans le devenir des cellules mésenchymateuses, ce gène est aussi impliqué dans les processus de différenciation du thymus et la maturation des lymphocytes T.

#### *b) mutations dans le gène pax-2 (Krd)*

L'expression de *pax-2* est détectée à partir de 9,0 jours de gestation, dans le tube neural, le futur rein et l'oeil. La mutation Krd (*Krd, kidney and retinal effects*) induite par transgénèse, est provoquée par la délétion d'un large fragment sur le chromosome 19, chevauchant le locus *pax-2* (Keller *et al.*, 1994). Bien que la délétion s'étende sur plusieurs loci, il semble que le phénotype induit soit essentiellement causé par l'absence de *pax-2*. L'hétérozygote Krd est caractérisé par un retard de croissance et des anomalies dans le rein et la rétine. Cette dernière présente en effet un électrorétinogramme anormal et un nombre réduit de cellules dans la couche nucléaire interne et dans la couche ganglionnaire. La



mutation homozygote *Krd* est létale, les embryons de souris meurent *in utero*. le gène *pax-2* est, par conséquent, impliqué dans les processus de différenciation des cellules du rein et de la rétine (Keller *et al.*, 1994).

*c) mutations dans le gène pax-3 (splotch;Sp)*

La cartographie d'expression du gène *pax-3*, et l'analyse des défauts induits par des mutations dans ce gène (mutant *splotch*) (Epstein *et al.*, 1991) , prouvent que *pax-3* est un important facteur de régulation durant le développement du tube neural, la migration des cellules de crête neurale et le devenir des somites (Moase et Trasler, 1990) .

*pax-3* a été localisé à proximité du locus *splotch* sur le chromosome 1 chez la souris (Tassabehji *et al.*, 1993) (Goulding *et al.*, 1993). Six allèles *splotch* ont été décrits chez la souris. A l'état hétérozygote, les mutants *splotch* se distinguent par des anomalies de la pigmentation (présence de taches blanches sur le ventre, les membres et la queue), liées à des déficiences dans la migration des cellules de crête neurale (Epstein *et al.*, 1991) . Les souris *splotch* présentent aussi un retard de croissance. Les Souris homozygotes mutées dans *pax-3* (Souris *Splotch*) présentent entre autres défauts, une fermeture incomplète du tube neural (phénotype *spina bifida*) et une exencéphalie (Epstein *et al.*, 1991). La mutation *splotch* est létale à l'état homozygote, les embryons meurent entre 13 et 14 jours de gestation. Les anomalies observées chez les souris homozygotes sont comparables à celles observées chez les patients atteints des syndromes de Waardenburg de types I (SW1) (Tassabehji *et al.*, 1992) et III (SW3) (Baldwin *et al.*, 1992); ces deux syndromes découlent de mutations dans le domaine paired de PAX3 (Hoth *et al.*, 1993) . Le SW1 se caractérise par une dystopie des canthi (déplacement latéral des angles internes des yeux), des anomalies de pigmentation et une surdit  uni- ou bilatérale. Quant au SW3, il associe aux caractéristiques du SW1 des malformations des membres supérieurs (contractures et réduction de la taille des muscles). Il existe quelques variantes dans les phénotypes induits par les différentes mutations de PAX3, mais d'une façon générale, ils présentent de fortes homologues avec les mutations de type *Splotch*. Ces observations suggèrent un rôle de Pax-3 dans le contrôle très précoce du développement des cellules dérivées de la zone dorsale du tube neural.

d) mutations dans le gène *pax-4*

L'inactivation du gène *pax-4* par recombinaison homologue (Sosa-Pineda *et al.*, 1997) résulte en l'absence de cellules productrices d'insuline et de somatostatine matures dans le pancréas. L'activité du gène *pax-4* contribue au maintien des cellules  $\beta$ .

e) mutations dans le gène *pax-5*

Le gène *pax-5* a lui aussi été inactivé par recombinaison homologue dans les cellules souches embryonnaires de souris. Les mutants homozygotes naissent vivants, présentent un retard de croissance et meurent trois semaines plus tard (Urbanek *et al.*, 1994). La morphogenèse de la zone postérieure du cerveau moyen est affectée à partir de 16,5 jours embryonnaires. De plus, tous les mutants produisent des lymphocytes B immatures, stoppés à un stade précoce de leur différenciation. Le gène *pax-5* intervient par conséquent à la fois dans la lymphopoïèse des cellules B et dans la mise en place du cerveau moyen postérieur (Urbanek *et al.*, 1994). De plus, l'expression de *pax-5* a aussi été détectée dans les testicules chez la souris adulte, mais son rôle dans cet organe n'a pas encore été déterminé.

g) mutations dans le gène *pax-6*, *small eye*.

Un autre mutant du développement est associé au gène *pax-6*. Les mutants *small eye* (*Sey*) ont été identifiés chez la souris (Roberts, 1967) ; (Hill *et al.*, 1991) et chez le rat (Matsuo *et al.*, 1993). *pax-6* est situé sur le chromosome 2, sur le locus *Sey* (Walther et Gruss, 1991). L'expression précoce de *pax-6* dans l'oeil est confinée au trognon optique, à la vésicule optique, puis à la rétine et à l'ectoderme de surface qui donnera plus tard le cristallin.

L'importance de *pax-6* dans le développement de l'oeil est attestée par l'existence de mutations phénotypiques dans cet organe, induites par un gène *pax-6* altéré. Ainsi, les souris *Small eye*, à l'état hétérozygote ont des petits yeux et à l'état homozygote, sont anophtalmiques (Roberts, 1967); (Hill *et al.*, 1991). A l'état homozygote, les embryons *Small eye* de souris ne développent ni oeil, ni nez, ni nerf optique, ni bulbe olfactif. Finalement les souriceaux homozygotes meurent à la naissance parce-qu'ils n'ont pas de nez pour respirer (Hogan *et al.*, 1986),(Hogan *et al.*, 1988). Les défauts observés au niveau du nez correspondent aussi à des territoires d'expression de *pax-6*. Chez l'Homme, les mutations dans PAX6 sont à l'origine des cas d'Aniridie (malformations oculaires affectant plus ou moins gravement l'iris, le cristallin, le nerf optique, problèmes auxquels peuvent s'ajouter des glaucomes) et des anomalies de Peter (présence d'adhérences entre le cristallin et la cornée). Des mutations dans le gène *eyeless*, l'orthologue de *pax-6* chez la

Drosophile, exprimé dans le disque imaginal oeil/antenne, empêchent la formation correcte des yeux (Quiring *et al.*, 1994).

D'autre part, des souris mutantes, pour lesquelles le gène Pax-6 a été délété, sont déficientes en cellules productrices de glucagon dans le pancréas. Ces observations suggèrent que Pax-6 est aussi requis pour la différenciation des cellules  $\alpha$  productrices de glucagon (St Onge *et al.*, 1997). Par ailleurs, les souris *Small eye* présentent une diminution de la vitesse de prolifération des cellules, dans la région alaire du diencephale où s'exprime *pax-6*. Les thalamus ventraux et dorsaux, dérivés du diencephale alaire, n'offrent plus de différence de densité dans leur populations cellulaires respectives. De plus, la frontière diencephalo-mésencéphalique est moins bien déterminée. En conséquence, *pax-6* intervient aussi sur le contrôle de la prolifération et de la différenciation du diencephale, ainsi que dans l'établissement de la frontière morphologique existant entre le diencephale et le mésencéphale (Warren et Price, 1997).

Il ne fait aucun doute aujourd'hui, que le gène *pax-6* joue un rôle majeur dans le développement de l'oeil chez les vertébrés comme chez les insectes. Son rôle dans le système nerveux central est moins bien compris pour le moment, du fait que aucun des mutants induits par les allèles *Sey* analysées ne présentent d'anomalies de la moelle épinière.

#### *f) mutations dans le gène pax-8*

Aucune mutation n'a été reportée jusqu'à ce jour dans le gène *pax-8*. Néanmoins, l'expression de *pax-8* corrèle avec la différenciation de la thyroïde. Effectivement, plusieurs lignées cellulaires de la thyroïde indifférenciées ne présentent pas d'ARNm de *Pax-8* détectable (Zannini *et al.*, 1992).

#### *e) pax-9*

*pax-9*, gène pour lequel aucune mutation n'a été reportée, présente des territoires d'expression communs à ceux de *pax-1*, dans le mésenchyme, puis plus tard dans les disques intervertébraux, les tissus mésenchymateux bordant les régions crâniennes et la colonne vertébrale ainsi que dans le thymus embryonnaire et adulte. Par contre, *pax-9* est exprimé distinctement dans le tissu mésenchymateux du métatarsus médian en développement, dans les organes olfactifs et dans l'épithélium longeant l'oesophage embryonnaire et adulte (Peters *et al.*, 1995).

## 2-4 Conclusion

L'étude des effets induits par les mutations des gènes Pax est un excellent outil pour l'analyse de leur fonction. L'expression des gènes Pax et les conséquences induites par leurs mutations confirme leur implication dans le développement embryonnaire. En l'occurrence, *pax-3*, *pax-6* et *pax-7*, sont exprimés très tôt au cours du développement du système nerveux central (8,5j p.c.). *pax-2*, 5 et 8 s'expriment un peu plus tard, à 10,5j p.c. dans le cerveau postérieur et la moelle épinière. Ces gènes sont liés aux processus du développement embryonnaire à l'origine de la mise en place du système nerveux central. Par ailleurs, les gènes pax jouent aussi un rôle dans la différenciation et le maintien des cellules différenciées, i.e. *pax-1* et lymphocytes T, *pax-2* dans le rein et la rétine, *pax-3* dans les cellules musculaires, *pax-4* et les cellules  $\beta$  du pancréas, *pax-5* et les lymphocytes B, *pax-6* dans la neurorétine et les cellules  $\alpha$  du pancréas, *pax-7* dans les myocytes (Schafer *et al.*, 1994), *pax-8* dans la thyroïde, et *pax-9* dans les organes olfactifs et l'oesophage.

Le chapitre suivant aborde l'étude de la fonction des gènes Pax à un niveau moléculaire.

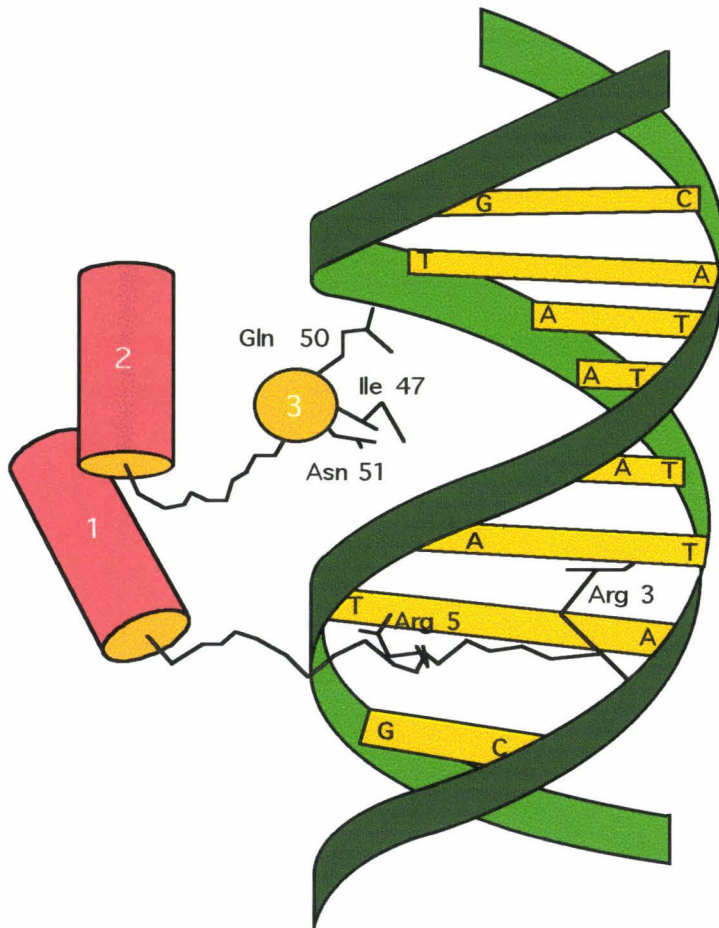
## II Paired & homéo, deux domaines de fixation à l'ADN.

### 1 Le domaine homéo

#### 1-1 Le domaine homéo. généralités

La boîte homéo est une séquence de 180 bp qui code un domaine homéo de 60 acides aminés. Ce domaine est capable de se fixer à l'ADN.

C'est par analogie avec les domaines homéo de levure MATa1 et MATa2, que les domaines homéo codés par les gènes homéotiques de Drosophile, puis de vertébrés, ont été supposés se fixer à l'ADN (Scott et Weiner, 1984). Les premières études ont été faites sur le domaine homéo du gène *antennapedia*. Ce dernier se fixe sur des séquences spécifiques, à partir desquelles un site consensus a pu être défini: le site ATTA (ou TAAT sur le brin complémentaire). Tous les domaines homéo connus peuvent reconnaître des séquences contenant ce motif, avec des affinités variables. Ces différences d'affinité ne sont pas suffisantes pour expliquer la spécificité d'action des protéines contenant un domaine homéo (Hoey *et al.*, 1988); revue de (Hayashi et Scott, 1990). La spécificité d'action des protéines pourrait s'expliquer par des mécanismes divers, tels que le contrôle spatio-



**Figure 4 : Structure du domaine *Homéo* de la protéine *Engrailed* fixé à l'ADN.**

Deux différences essentielles entre cet homéodomaine et celui de Pax-6. (i) la glutamine 50 est une sérine dans Pax-6 et cet acide aminé est essentiel dans la sélection de la séquence cible liée par l'homéodomaine. (ii) l'homéo-domaine de Pax-6 est capable de former des homodimères sur son ADN cible.

temporel de leur expression et/ou l'hétérodimérisation avec d'autres facteurs, ou bien encore l'association sur la même protéine d'un deuxième domaine de fixation à l'ADN, comme dans le cas des gènes Pax. Par ailleurs, il existe tout de même une spécificité relative au sein du domaine homéo lui-même, comme le suggèrent les études référencées dans ce chapitre.

## 1-2 Structure tridimensionnelle du domaine homéo

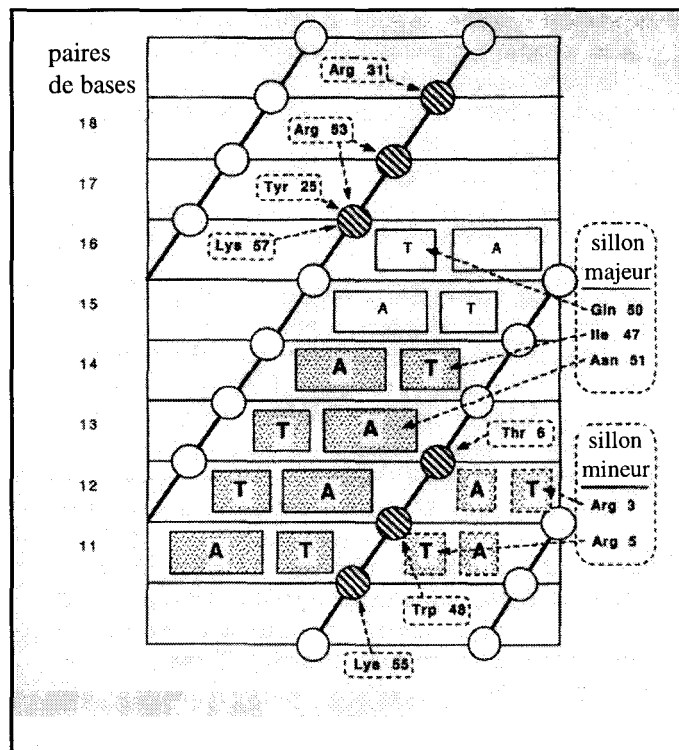
### a) *Structure tridimensionnelle générale.*

L'analyse tridimensionnelle par résonance magnétique nucléaire (RMN) du domaine homéo d'Antennapedia a mis en évidence une structure organisée en trois hélices  $\alpha$  précédées d'un bras N terminal de structure désorganisée. Le bras N terminal contient 9 acides aminés. Les acides aminés 10 à 22 du domaine homéo forment la première hélice  $\alpha$ , ceux de 28 à 37, la deuxième hélice, et les acides aminés 42 à 58 constituent la troisième hélice. Les hélices 1 et 2 sont séparées par une boucle, tandis qu'un tour raccorde les deux dernières hélices. L'analyse par cristallographie du domaine homéo de la protéine Engrailed fixée à l'ADN (Kissinger *et al.*, 1990), et du domaine homéo de la protéine MATA2 (Wolberger *et al.*, 1991), a révélé le même arrangement en trois hélices  $\alpha$ . De plus, le contact de l'ADN avec chacune de ces protéines se fait de façon similaire: la partie amino terminale du domaine contacte le petit sillon de l'ADN au niveau du site de fixation, tandis que la troisième hélice, appelée "hélice de reconnaissance" contacte le grand sillon adjacent (Figure 4).

### b) *acides aminés conservés indispensables à la reconnaissance du site TAAT et à la liaison du domaine homéo à l'ADN*

On peut définir trois groupes d'acides aminés conservés dans tous les domaines homéo et indispensables pour la fixation de ces peptides sur un site TAAT. On retrouve en effet, des acides aminés hydrophobes, tels que le Trp 48 et la Phé 49, qui pourraient permettre la cohésion des hélices 1 et 3 et maintenir la structure de la protéine (Kissinger *et al.*, 1990).

D'autre part, le deuxième groupe est composé d'acides aminés qui participent au maintien du complexe ADN-protéine. Ces acides aminés n'engagent pas de liaison directe avec les bases du site TAAT. Ils permettent l'"encrage" de la protéine sur l'ADN, en réalisant des contacts avec les désoxyriboses ou les phosphates du squelette de la double



**Figure 5 :** Acides aminés impliqués dans l'interaction de l'homéodomaine de la protéine Engrailed avec son site de reconnaissance sur l'ADN (d'après *Kissinger et al.*, 1990)

hélice, par des liens de natures diverses: liens hydrogènes, hydrophobes ou faisant intervenir une ou plusieurs molécules d'H<sub>2</sub>O, forces de Van Der Waals ou électrostatiques. Cette fonction est remplie par la Thr6 du bras N terminal, par la Tyr 25 de la boucle reliant les hélices 1 et 2, l'Arg 31 de l'hélice 2, le Trp 48, l'Arg 53 et les Lys 55 et 57 de l'hélice 3 (cf figure 5).

Le troisième groupe d'acides aminés établit des contacts directs avec les bases du site TAAT, par des liens variés, similaires à ceux du deuxième groupe. Dans le bras N terminal du domaine homéo, l'Arg 3, réalise un lien hydrogène avec l'O<sub>2</sub> de la thymine en position 2 (base complémentaire du A en position 2, en lisant du 5' vers le 3') du site TAAT. L'Arg 5, effectue un lien hydrogène avec l'O<sub>2</sub> de la thymine en position 1 du site TAAT. Ces deux liaisons se situent dans le sillon mineur de l'ADN.

Dans le sillon majeur, L'Ile 47 réalise des contacts hydrophobes avec le groupement méthyle de la thymine en position 4, TAAT.

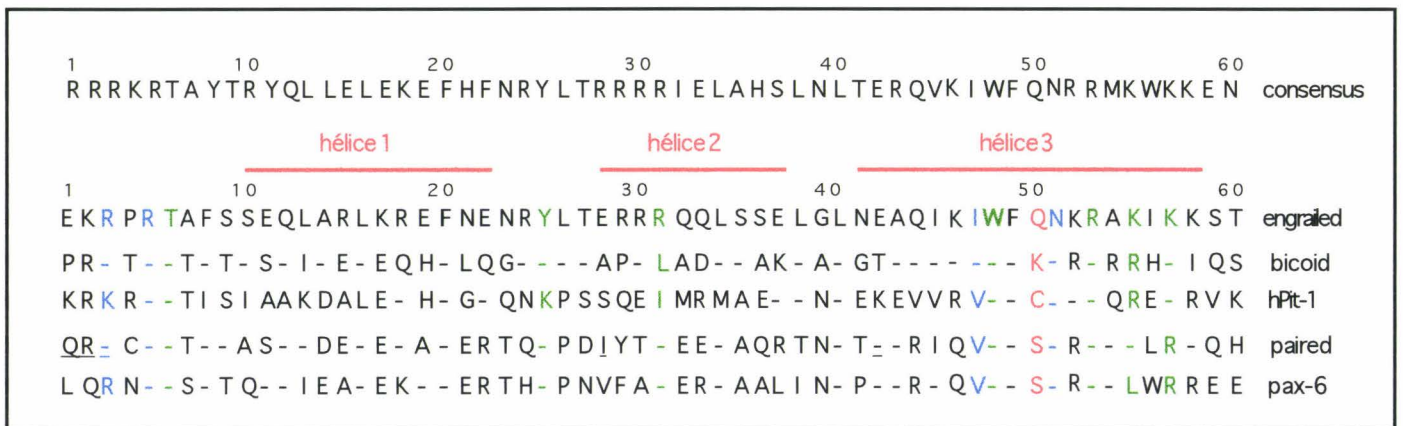
Nous venons de voir les acides aminés impliqués dans la reconnaissance du coeur TAAT du site de fixation des domaines homéo, par l'association d'acides aminés intervenant à la fois dans la cohésion du peptide, dans la reconnaissance du site TAAT et dans l'encrage de la protéine sur l'ADN. Un quatrième groupe d'acides aminés intervient dans la spécificité de reconnaissance et permet de distinguer les séquences cibles respectives des domaines homéo. Ces acides aminés interagissent soit avec les bases flanquantes du site TAAT, soit avec des cofacteurs susceptibles de modifier leurs propriétés de liaison à l'ADN. Les positions de ces acides aminés "de spécificité de reconnaissance" seront étudiés dans le paragraphe suivant (figure 6).

#### *c)acides aminés liés à la spécificité de reconnaissance des domaines homéo*

##### \_ résidu 9 de l'hélice de reconnaissance

La glutamine 50 (= résidu 9 de la 3e hélice) réalise des contacts de Van der Waals avec le groupement méthyle de la thymine (base située à deux nucléotides en 3' du site TAAT), dans le cas du domaine homéo de la protéine Engrailed (Kissinger *et al.*, 1990). Le résidu 50 permet la constitution des classes de domaine homéo. Il existe plus de 20 classes qui ne seront pas détaillées dans cet exposé (pour revue, cf (Duboule et Morata, 1994). A titre d'exemples, les domaines homéo de type Hox (gènes de vertébrés homologues aux gènes homéotiques du complexe HOM de Drosophile), ceux de type Caudal, TCL, Msh, NK1/2, Engrailed et certains apparentés au type paired possèdent une glutamine en position 50. Une lysine occupe cette position pour les domaines homéo de type Bcd et





**Figure 6:** Alignement de domaines homéo de différentes classes mettant en valeur les acides aminés impliqués dans la liaison à l'ADN dans la structure 3D.

- En bleu, acides aminés impliqués dans la reconnaissance du site TAAT
  - En vert, acides aminés qui contactent le squelette de la double hélice (interactions avec les phosphates ou les désoxyriboses)
  - En rouge, acide aminé n°50, impliqué dans la reconnaissance spécifique (interactions avec les bases flanquantes du site TAAT)
  - En gras, acides aminés conservés, pouvant permettre la cohésion des hélices et le maintien de la structure de la protéine.
  - En souligné, acides aminés impliqués dans l'interface du dimère des domaines homéo de type paired (cf Wilson *et al.*, 1995).
  - "Consensus" = séquence consensus du domaine homéo, établi d'après les acides aminés rencontrés le plus fréquemment aux positions considérées (selon Duboule *et al.*, 1994).
- Les tirets représentent les acides aminés conservés par rapport à la protéine Engrailed.

certaines apparentés au type paired, ou une **cystéine** pour ceux de type POU. Les produits des gènes de type Pax ou Paired se caractérisent par la présence d'une **sérine** en position 50 de leur domaine homéo (cf figure 6). Ce résidu 50 est lié à la spécificité de reconnaissance du domaine homéo et son identité peut influencer sur le type de bases qui entourent le coeur TAAT du site de reconnaissance. Ainsi, le domaine homéo de Fushi-tarazu (Ftz) possède une Gln en position 50 et reconnaît une séquence de type CCATTA. Le domaine homéo de Bicoïd (Bcd) porte une **lysine** en position 50 et reconnaît une séquence de type GGATTA. Si la séquence cible de Ftz est mutée en GGATTA, l'affinité de la protéine pour l'ADN diminue fortement. De façon complémentaire, la substitution de la Gln 50 de Ftz en Lys, accroît l'affinité de la protéine pour la séquence GGATTA (Schier et Gehring, 1992). L'acide aminé en position 50 du domaine homéo (ou 9 de l'hélice de reconnaissance) détermine le choix des bases situées en 3' du coeur TAAT sur l'ADN.

#### - bras amino terminal

Il arrive souvent que de nombreuses protéines possèdent le même résidu en position 50 de leur domaine homéo, comme c'est le cas du produit des gènes *Ubx*, *Dfd*, *Scr* et *Antp*. D'autres éléments de la protéine participent alors à la spécificité de reconnaissance de la séquence cible, tel le bras amino terminal du domaine homéo. Ainsi, les protéines des gènes *Antp* et *Sex comb reduced (Scr)* ont des spécificités d'action différentes *in vivo*, et leurs domaines homéo ne diffèrent que par 4 acides aminés localisés dans la partie amino terminale du domaine. Une protéine *Antp* pour laquelle, ces 4 acides aminés ont été remplacés par leurs équivalents dans *Scr*, reconnaît *in vivo*, les gènes normalement contrôlés par *Scr* (Zeng *et al.*, 1993). De même, le remplacement des six acides aminés qui distinguent le bras amino-terminal du domaine homéo de Deformed (*Dfd*) de celui d'Ultrabithorax (*Ubx*) suffit pour changer la spécificité *in vivo* du produit de *Dfd* en celle du produit de *Ubx*, bien qu'il y ait d'autres acides aminés qui diffèrent dans les structures en hélice  $\alpha$  (Lin et McGinnis, 1992). Cependant, dans d'autres cas, le bras amino-terminal, bien que nécessaire à la spécificité de reconnaissance, n'est pas toujours suffisant: l'échange de deux acides aminés différents dans cette région entre *Ubx* et *Antp* est nécessaire mais ne suffit pas au changement de spécificité entre les deux protéines. Il est indispensable de muter trois des 5 acides aminés qui diffèrent dans les régions en hélice  $\alpha$  (Chan et Mann, 1993).

#### - acides aminés non conservés / liaison avec des cofacteurs

Le bras amino-terminal, étudié dans le paragraphe précédent participe à la fois à la fixation à l'ADN et à la spécificité de reconnaissance du domaine homéo. Dans ce

paragraphe, le sujet d'intérêt portera sur les acides aminés qui ne contactent pas l'ADN, et qui sont pourtant impliqués dans la spécificité de reconnaissance de la séquence cible. Comme nous venons de le voir, il arrive parfois que deux domaines homéo possèdent le même résidu 50, et que leurs bras N terminaux, bien que différents et nécessaires à la spécificité de reconnaissance, ne soient néanmoins pas suffisants pour distinguer leurs cibles respectives, comme dans le cas des protéines Ubx et Antp. Par contre, la substitution de trois acides aminés d'Antp par ceux d'Ubx suffit à échanger les spécificités respectives de ces deux protéines: acides aminés 22 et 24, situés entre les deux premières hélices, et acide aminé 56, localisé à la fin de la troisième hélice (Chan et Mann, 1993). Les chaînes extérieures des acides aminés 22 et 24 sont exposées au solvant et peuvent interagir avec d'autres facteurs. C'est l'analyse du phénotype des mutants du gène *extradenticle* (*exd*) qui a permis d'identifier Exd comme le facteur coopérant avec Ubx et Antp. Les mutants du gène *exd* présentent en effet des transformations homéotiques dans les régions spécifiées par les gènes *Ubx* et *Antp*, en dépit de l'absence d'altération dans leur cartographie d'expression. Or, les produits du gène *exd* sont capables de se lier aux protéines Ubx et Antp par l'intermédiaire des 3 acides aminés qui diffèrent entre ces deux protéines (pour revue, cf. (Wilson *et al.*, 1995). Pourtant, bien qu' interagissant avec le même cofacteur, ces protéines ne régulent pas les mêmes gènes cibles. Soit, parce-qu'elles ne s'expriment pas dans les mêmes tissus, ou soit parce-que la structure globale du complexe Ubx-Exd diffère de celle du complexe Antp-Exd. Les acides aminés 22, 24 et 56 du domaine homéo pourraient jouer un rôle dans la conformation des deux types de complexes, et en conséquence leur octroyer une spécificité de reconnaissance propre à chacun. L'interaction avec des cofacteurs accroît l'affinité de la protéine pour l'ADN. Ainsi, le domaine homéo de Exd\* accroît l'affinité de la protéine Ubx pour son site de reconnaissance sur le promoteur du gène Dpp (Decapentaplégique) de 30 fois (Wilson *et al.*, 1995).

L'exemple d'Exd n'est pas un cas isolé; d'autres protéines à domaine homéo vont former des dimères avec d'autres protéines. Ces associations influent sur leur spécificité de reconnaissance à l'ADN. Ainsi, le produit d'un gène POU, oct-1, interagit avec la protéine VP16 du virus de l'herpès simplex grâce, là encore, à l'acide aminé en position 22 du domaine homo (Lai *et al.*, 1992). De même, le produit du gène *HOXD8* peut réprimer l'expression de *HOXD9*, non pas en se fixant sur le promoteur du gène *HOXD9*, mais en empêchant les produits des gènes *HOXD9* et *HOXD10* de s'y fixer. Seules la partie amino terminale et la première hélice  $\alpha$  de la protéine *HOXD8* sont nécessaires à cette répression

*in vivo*. *In vitro*, ces domaines sont suffisants pour permettre à la protéine HOXD8 d'interagir avec HOXD9 (Zappavigna *et al.*, 1994).

*d) Conclusion sur la fixation du domaine homéo à l'ADN.*

En résumé, le domaine homéo contient des acides aminés à fonctions différentes, mais complémentaires. Certains acides aminés hydrophobes que j'appellerai "acides aminés de structure" sont conservés sur l'ensemble des domaines homéo, et participent à l'organisation et à la cohésion des hélices qui constituent le peptide. D'autres, que j'appellerai "acides aminés de liaison à l'ADN", conservés eux aussi, sont placés à des positions clefs, et engagent des liaisons de natures variées avec l'ADN. Parmi ceux-ci, existent des "acides aminés cibleurs", qui vont repérer et fixer le site TAAT. D'autres, "acides aminés d'encrage à l'ADN", vont concourir à maintenir l'hélice de reconnaissance dans le sillon majeur, en réalisant des liaisons avec les sucres et les phosphates de l'ADN. Enfin, parmi ces "acides aminés de liaison à l'ADN", certains vont intervenir dans la spécificité de reconnaissance de la séquence cible. Ainsi, le résidu en position 50 dicte la nature des deux paires de bases qui prolongent en 3' le site TAAT. D'autre part, il existe aussi des "acides de spécificité de reconnaissance", qui n'interagissent pas avec l'ADN, mais s'associent avec des cofacteurs qui augmentent leur affinité pour l'ADN ou définissent une nouvelle séquence cible.

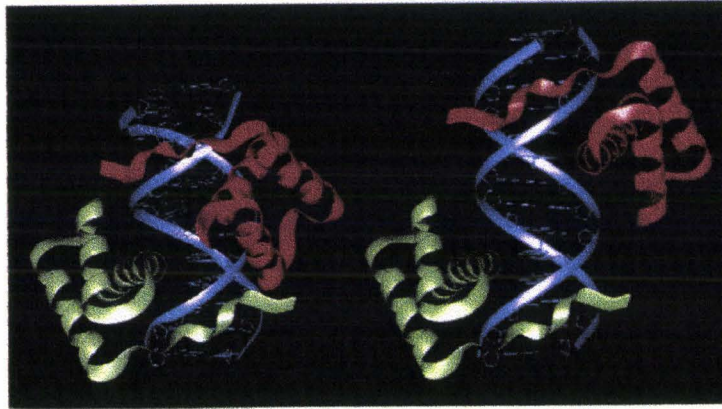
### 3 Le complexe de fixation: ADN - domaine homéo de type paired

*a) Séquence cible de fixation in vitro*

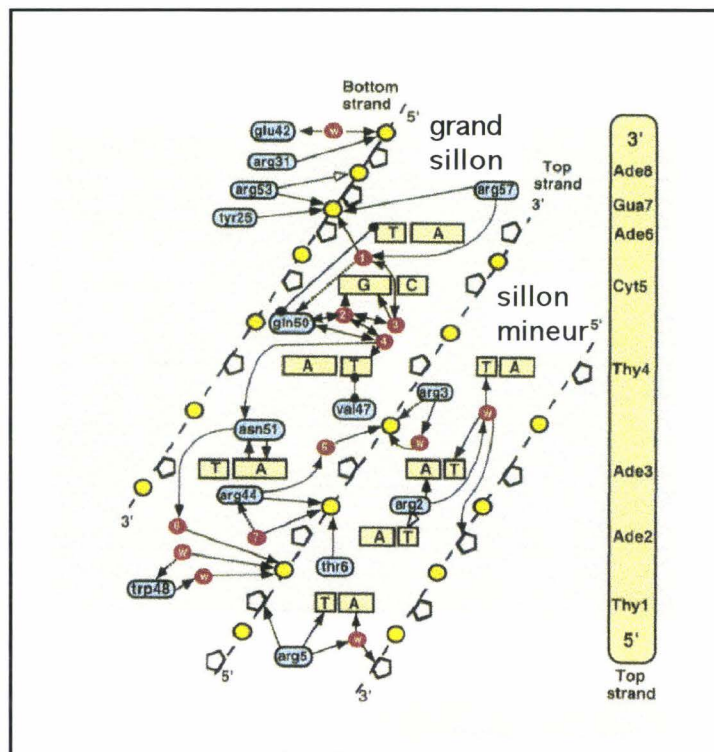
Les domaines homéo des protéines Pax sont caractérisés par la présence d'une sérine en position 50.

A partir d'oligonucléotides variant au hasard sur leur 18 bases centrales, Wilson a mis en évidence la fixation préférentielle des domaines homéo de type paired pour un oligonucléotide de type P2 (méthode SELEX réalisée avec le peptide correspondant aux 60 acides aminés du domaine homéo, flanqués de 4 acides aminés en N terminal dérivés de la protéine Paired), (Wilson *et al.*, 1993). Cette séquence de fixation est constituée d'un motif palindromique TAAT, séparé par deux paires de bases. **P2: TAATPyGATTA**

D'autres domaines homéo de la classe "paired like" possèdent une glutamine ou une lysine en position 9 de l'hélice de reconnaissance. Avec ce type de domaine homéo, un



a) Schéma représentatif de la structure 3D de l'interaction du complexe ADN/homéo de type paired



b) Contacts réalisés entre le dimère des domaines homéo de type paired et le site de reconnaissance sur l'ADN

**Figure 7** : Le domaine homéo de type paired se fixe en dimère sur l'ADN, de chaque côté du sillon majeur de l'ADN.

nouveau consensus a pu être établi, noté P3, il comporte trois paires de bases au centre d'un palindrome TAAT.P3: TAATPyNPuATTA

La fixation sur ces sondes palindromiques se présente sous forme d'une "dimérisation coopérative", c'est-à-dire que deux monomères sont capables de se fixer sur chacun des demi sites du palindrome, le premier fixé permet au second de se lier avec une meilleure affinité. Ainsi, la coopérativité de liaison pour un domaine homéo de type S9 sur P2 est de l'ordre de 50 fois, et celle d'un domaine homéo de type Q9 sur P3, de l'ordre de 180 fois . Par ailleurs, le domaine homéo de type paired se fixe aussi sur P3, avec cependant une fréquence plus faible, du moins d'après les résultats de recherche d'une séquence cible *in vitro* (Wilson *et al.*, 1993).

#### *b) Structure tridimensionnelle*

L'étude cristallographique du complexe "domaine homéo de type paired/ADN" a été réalisée avec la sonde P3. Le résidu sérine en position 50 du peptide homéo utilisé, a été muté en glutamine, car la coopérativité de liaison est plus forte pour un domaine homéo Q9 sur P3 que pour un domaine homéo S9 sur P2 (Wilson *et al.*, 1995).

#### - structure tridimensionnelle générale: apparition d'un dimère

Le complexe se présente comme un arrangement de tête à tête des deux domaines homéo sur le site palindromique. Chacun des domaines réalise des liens hydrogènes ou des contacts électrostatiques (contacts directs ou par l'intermédiaire de molécules d'H<sub>2</sub>O) avec la séquence ADN cible sur 11pb, et particulièrement, des liens spécifiques sur 6 pb. Les deux peptides du dimère sont suffisamment rapprochés, de façon à ce que le bras N-terminal d'un domaine homéo interagisse avec le début de l'hélice II de l'autre, et les régions N-terminales des hélices de reconnaissance s'approchent l'une de l'autre (Figure 7). La fixation du dimère sur l'ADN entraîne la courbure de la double hélice en forme B d'un angle de 20°.

Les liaisons engagées entre un demi site (TAATgt) de l'ADN et un domaine homéo sont identiques à celles réalisées entre un domaine homéo de type Engrailed sur son site de fixation (TAATat); se reporter au chapitre II-1-2.

#### - liaisons impliquées dans l'interface du dimère

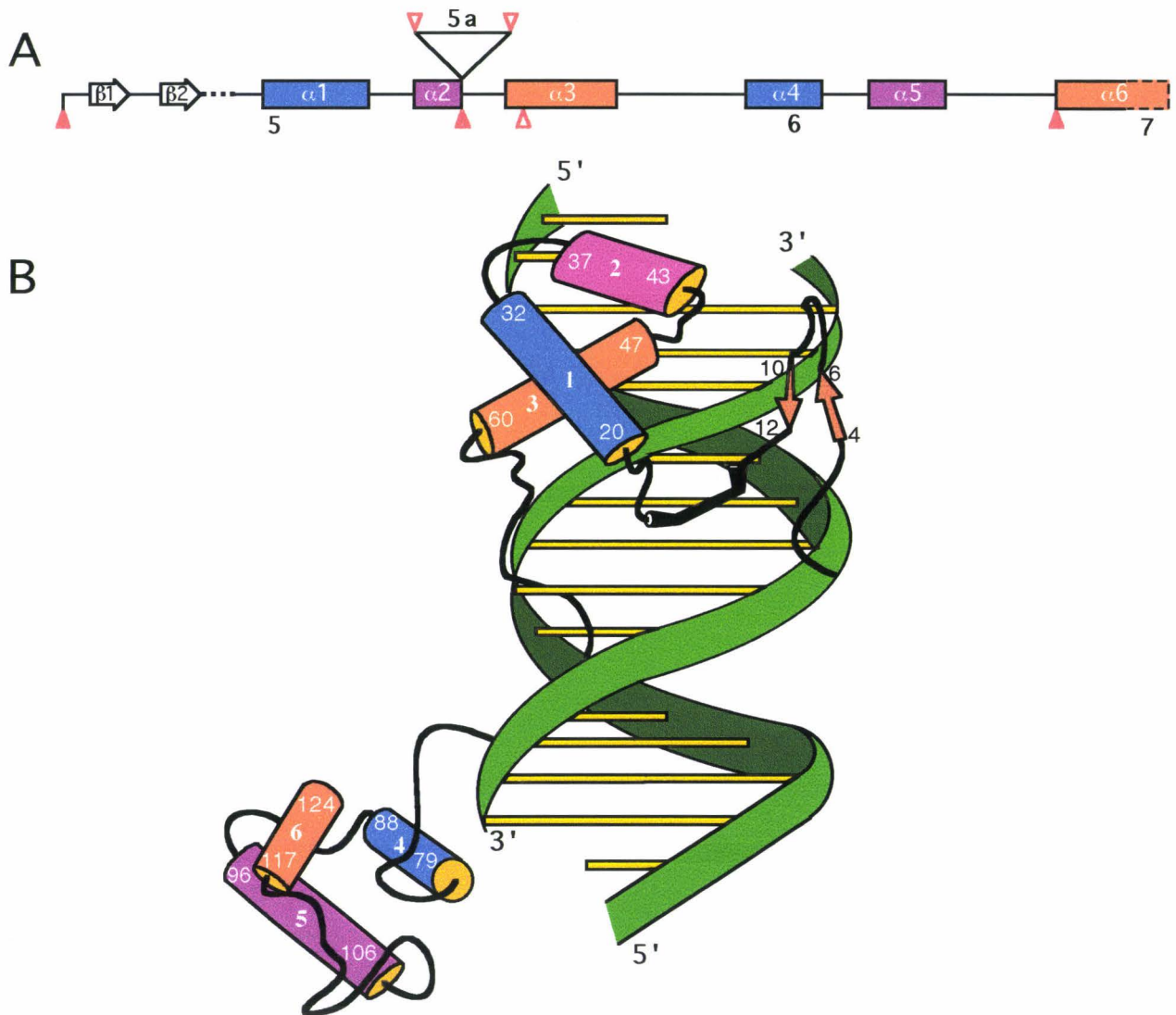
Le bras N terminal d'un domaine homéo s'approche de la partie N terminale de l'hélice 2 de l'autre domaine homéo. Ce rapprochement est permis par une complémentarité de forme et de charge entre les résidus 1-3 d'un domaine homéo et l'Ile

28 (le premier acide aminé de l'hélice 2) et la Glu 42 (le premier résidu de l'hélice de reconnaissance) de l'autre domaine. La chaîne latérale de l'Ile 28 se colle contre le squelette du bras N terminal et la chaîne latérale de l'Arg-3 de l'autre molécule. De plus, le résidu 1 accepte un lien hydrogène avec le résidu 28 de l'autre domaine homéo. Par ailleurs, un contact hydrophobe relie les groupements méthyle des Ala-43 de chacun des domaines homéo.

*c) Conclusion sur la fixation à l'ADN du domaine homéo de type paired*

Le domaine homéo de type paired se fixe sur un motif palindromique de coeur ATTA, séparé par deux à trois paires de bases. Les domaines homéo sont disposés de part et d'autre du grand sillon de l'ADN, de façon symétrique. Les contacts réalisés entre chacun des domaines homéo et chaque demi site de l'ADN sont similaires à ceux observés pour les autres domaines homéo étudiés par RMN ou cristallographie (cf. Qian *et al.*, 1989 pour l'étude par RMN du domaine d'Antp et (Kissinger *et al.*, 1990) pour l'étude du domaine d'Engrailed). De plus, les domaines homéo fixés en dimères établissent des contacts d'interface entre chacun des monomères. La dimérisation des domaines homéo de type paired est une dimérisation coopérative; c'est-à-dire que la fixation du premier monomère améliore l'affinité du second pour le deuxième demi site du palindrome. Cette dimérisation n'est réalisée qu'à partir du domaine homéo seul, contrairement aux autres protéines où elle nécessite un domaine supplémentaire, comme le domaine Pou spécifique des gènes Pou ou l'octapeptide des gènes Hox.

La structure secondaire des domaines homéo de type Pax fixés sous forme de monomère n'a pas été établie. Par SELEX (Wilson *et al.*, 1993), le consensus sortant est toujours un palindrome (contenant le motif ATTA), qui permet par conséquent, l'accrochage de deux domaines homéo sur chacun des deux sites définis par ce palindrome. Il semble que la fixation en monomère d'un domaine homéo sur un demi site sur l'ADN ne soit qu'un événement transitoire, permettant "d'apprêter" le terrain (courbure de l'ADN ou aide à la fixation) pour la fixation du deuxième domaine homéo. La fixation en monomère n'est pas observée *in vivo* à moins d'une intervention concomitante avec le domaine paired sur la même séquence cible. La liaison à l'ADN du domaine paired est étudiée dans le chapitre suivant.



**Figure 8 : Structure du domaine *Paired* du gène *paired* de *Drosophila*.**

**A,** Les hélices  $\alpha$  (h1-6), définies par l'analyse cristallographique [17] sont replacées sur la séquence codante de *PAX-6* (l'exon 5 code les hélices 1 et 2, l'exon 6 les hélices 3, 4 et 5, l'hélice 6 est codée par l'exon 7). Les limites des exons codant le domaine sont figurées par des triangles (pleins pour les exons codant la p46 et vide pour les exons alternatifs présentés dans la figure 3 ; la p43 perd ainsi les hélices 3, 4, et 5).

**B,** Schématisation des domaines de liaison à l'ADN du domaine *Paired*.



## 2 Le domaine paired

### 2-1 Structure tridimensionnelle du domaine paired

#### a) *Structure tridimensionnelle générale*

La structure du domaine paired a été déterminée par cristallographie en 1995 par Xu, W. et *al.*, à partir d'un oligonucléotide de 15 pb, optimisé *in vitro*, et du domaine paired de la protéine Paired de drosophile. La structure du cristal met en évidence deux structures globulaires définies dans le domaine paired (Figure 8). La région amino terminale, région la plus conservée du domaine paired établit le plus de contacts avec l'ADN, et plusieurs structures secondaires différentes participent à cette liaison. Deux feuillets  $\beta$  (résidus 4 à 6 et 10 à 12) s'agrippent au squelette phosphaté de l'ADN, et sont immédiatement suivis par un tour  $\beta$  qui réalise des contacts importants pour le maintien de la liaison dans le sillon mineur (résidus 13 à 16, appelé  $\tau 2$ ). La première région hélicoïdale contient un motif hélice-tour-hélice (HTH); l'hélice 2 effectue de nombreux contacts avec les phosphates, et l'hélice 3 se fixe dans le sillon majeur de l'ADN. La queue C terminale de ce domaine (résidues 65 à 72) réalise aussi des contacts dans le sillon mineur, à côté de ceux réalisés par le tour  $\beta$ . Un court "linker" (résidus 73 à 78) relie les deux structures globulaires amino et carboxy terminale du domaine paired. La structure C terminale n'effectue aucun contact avec l'ADN (Xu et *al.*, 1995).

#### b) *Acides aminés impliqués dans la liaison à l'ADN.*

Comme pour le complexe homéo/ADN, certains acides aminés du sous domaine N terminal vont se fixer sur les bases de la séquence cible, d'autres vont aider à la stabilité du complexe en réalisant des liaisons avec les sucres et les phosphates de l'ADN. La séquence consensus est constituée des 15 bp suivantes :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
A	A	C	G	T	C	A	C	G	G	T	T	G	A	C		
		G	C	A	G	T	G	C	C	A	A	C	T	G	T	T

Les bases en 5' du site, en particulier, les bases 4-8 du sillon majeur de l'ADN échangent des liaisons avec l'hélice de reconnaissance (hélice 3) de la région N terminale du domaine paired. La structure  $\beta$  turn dans le bras N terminal du domaine paired réalise

des liens avec les bases 9-11, et la queue C terminale du sous domaine N-ter de paired effectue des contacts avec les bases 11 à 14, dans le sillon mineur de l'ADN.

Les structures en hélices  $\alpha$  et le petit tour  $\tau_1$  participent aux points d'encrage du peptide sur le squelette désoxiribo-phosphaté dans le sillon majeur de l'ADN; tandis que la protéine est amarrée dans le sillon mineur par le tour  $\tau_2$  et la queue C terminale du sous domaine Nter de paired.

Toutes les acides aminés impliqués dans des liaisons avec l'ADN sont conservés dans tous les domaines paired, sauf l'histidine en position 47 qui est remplacée par une asparagine dans le cas des gènes *Pax-6* et *Pax-4*. Or, il s'avère que cette substitution d'acide aminé corrèle avec une différence de spécificité. L'His 47 permet la reconnaissance d'une guanine en position 4 du consensus ci-dessus, pour les gènes *Pax-2*, *Pax-5* et *Pax-8* (Czerny *et al.*, 1993), (Epstein *et al.*, 1994a; Epstein *et al.*, 1994b)); tandis que l'Asn 47 reconnaît une thymine dans la même position, dans le cas de la protéine *Pax-6* (Epstein *et al.*, 1994a).

### *c) conclusion*

En conclusion, seules les trois premières hélices du domaine paired participent à la liaison à l'ADN de la protéine Paired, et constituent le sous domaine N terminal, noté également PAI. Le sous domaine C ter ne participe pas à la fixation à l'ADN dans le cas de la protéine Paired. Il est pourtant constitué lui aussi de trois hélices, notées RED, agencées comme les domaines classiques de liaison à l'ADN de type hélice-tour-hélice ou hélice-boucle-hélice.. De plus, le sous domaine RED des protéines *Pax-6* comporte des motifs conservés avec le domaine de liaison à l'ADN de la Hin recombinase (Xu *et al.*, 1995). Par ailleurs, les études biochimiques ont montré que RED jouerait un rôle dans la fixation des protéines *Pax-6* à l'ADN. Le paragraphe suivant résume les résultats des études biochimiques parallèles ou antérieures à l'étude cristallographique du complexe ADN/ domaine paired.

## 2-2 Etudes biochimiques

Les études génétiques et biochimiques ont anticipé la cristallographie et ont conclu à l'existence de deux sous domaines dans le domaine paired. Par simplicité et en référence aux études cristallographiques, le sous domaine N terminal sera noté PAI et le sous domaine C terminal, RED.

### a) rôle de PAI

L'importance du sous domaine PAI dans la fonction des protéines Pax est illustrée d'une part, par les défauts du développement induits par la présence de nombreuses mutations de substitutions dans cette région. D'autre part, des études biochimiques plus ciblées ont permis d'identifier PAI comme la région responsable de la fixation à l'ADN du domaine paired.

#### - les mutations de substitutions des gènes Pax sont situées dans PAI

Les études génétiques ont montré que la majorité des mutations de substitutions des gènes Pax sont localisées dans le sous domaine PAI de paired. Les autres mutations des gènes Pax induisent des délétions dans la protéine ou des changements de cadre de lecture, qui toutes modifient complètement la conformation de la protéine. Par contre, les mutations de substitutions ne changent qu'un seul acide aminé. Retrouver des substitutions situées dans le domaine PAI à l'origine de mutants du développement, suggère une implication importante de ce domaine dans la fonction de la protéine, en particulier pour la fixation à l'ADN, comme l'illustre la mutation *undulated*. Ce mutant présente en effet une substitution de la Gly 15 en sérine (Balling *et al.*, 1988) et la protéine Pax-1 traduite présente une affinité moindre pour l'ADN (Chalepakakis *et al.*, 1991). D'autre part, cette mutation introduite dans la protéine Paired diminue aussi la fixation de celle-ci pour l'ADN (Treisman *et al.*, 1991). Dans la structure tridimensionnelle, la Gly 15 est située dans le turn  $\tau_2$  et se lie à la guanine en position 10 et à la thymine en position 11 dans le sillon mineur de l'ADN.

Par ailleurs, le syndrome de Waardenburg (qui se manifeste par des anomalies de la pigmentation) peut être causé par diverses substitutions dans le gène *Pax-3*: l'Asn 14 en His (Baldwin *et al.*, 1992) ou la Pro 17 en Leu (Hoth *et al.*, 1993). Ces deux positions sont aussi localisées dans le turn  $\tau_2$ . L'Asn 14 se lie à la guanine en position 9, et la Pro 17 forme une surface hydrophobe avec la Phe 12 qui se colle sur les désoxyriboses. De même, la Gly 48 substituée en Ala induit le syndrome de Waardenburg (Tassabehji *et al.*, 1993). Or, la glycine en position 48 réalise des contacts avec trois bases dans le sillon majeur de l'ADN. La substitution de l'arginine en position 23 a été identifiée à la fois dans le syndrome de Waardenburg pour le gène *Pax-3* (Hanson *et al.*, 1994), et dans l'anomalie de Peter pour le gène *Pax-6* (Hoth *et al.*, 1993). Cet acide aminé établit des contacts avec les sucres et les phosphates de l'ADN et intervient par ailleurs dans le maintien de la structure de la protéine, en effectuant une liaison avec la chaîne carbonée du résidu 63.

En somme, les mutations de substitutions causent des dommages considérables parce-qu'elles atteignent soit des acides aminés situés à des positions clefs pour l'établissement des liaisons avec l'ADN (avec les bases du consensus ou avec le squelette désoxiribo-phosphaté de l'ADN); soit elles touchent des acides aminés impliqués dans le maintien de la structure de la protéine durant sa fixation sur l'ADN. La plupart de ces mutations, localisées dans le sous domaine PAI de *paired* ont pour effet d'abolir ou de diminuer l'affinité de la protéine pour l'ADN. Ces observations renforcent l'idée du rôle primordial, ou tout au moins indispensable, de PAI dans la capacité de liaison à l'ADN du domaine *paired*.

- PAI est nécessaire pour la capacité de liaison à l'ADN du domaine *paired*

En 1991, Treisman et al. ont mis en évidence l'existence de deux sous domaines dans le domaine *paired* de la protéine *Paired* de *Drosophila*.

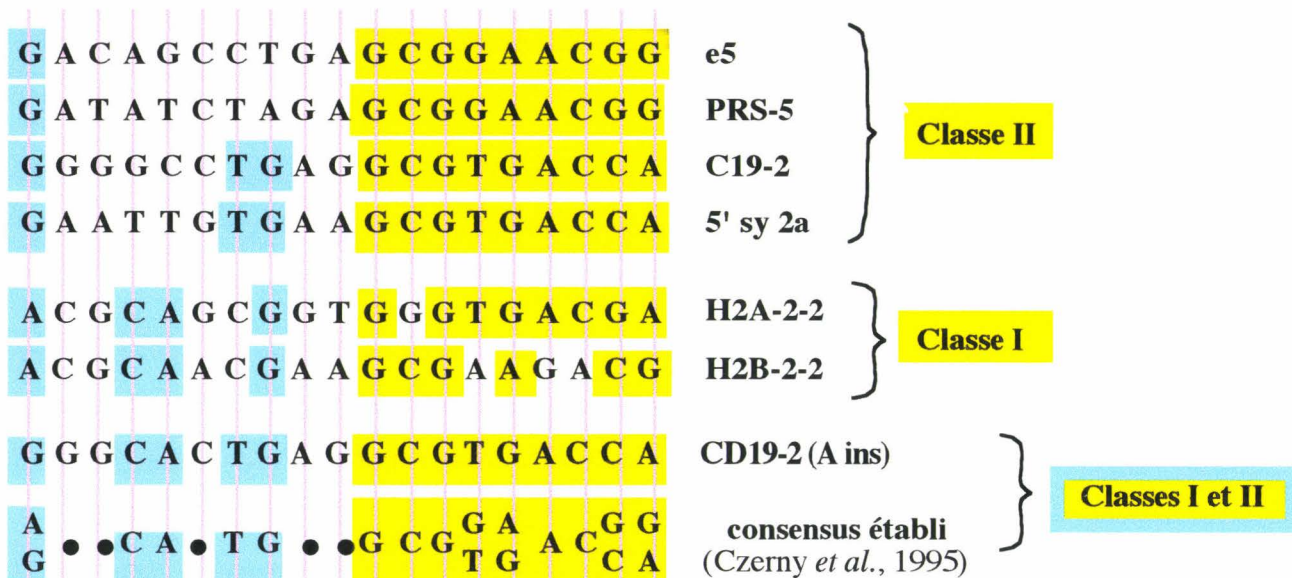
Le domaine déléte de ses 50 acides aminés C ter est toujours capable de reconnaître une séquence cible de fixation e5 (localisée sur le promoteur du gène *evenskipped* de *Drosophila*). Or, les 78 acides aminés Nter restant, comprennent le sous domaine PAI (Treisman *et al.*, 1991).

De même, le domaine PAI seul de Pax-5 (appelée aussi BSAP) est encore capable de reconnaître ses séquences cibles (dans le cas d'un consensus parfait seulement, cf. paragraphe suivant: "rôle de RED) (Czerny *et al.*, 1993).

De plus, des expériences de "sauvetage" sur des *Drosophiles* qui ont perdu la fonctionnalité du gène *paired*, par l'introduction de transgènes (contenant des portions différentes du gène *paired*), ont montré que le domaine RED n'est pas indispensable pour la fonction de la protéine *in vivo*. Du moins, pour les fonctions testées: le gène *paired* est indispensable pour l'expression correcte des gènes *wingless*, *engrailed*, et *gooseberry*. L'introduction d'un transgène codant la protéine *Paired* déléte de sa région RED permet encore de rétablir une expression normale de ses gènes cibles (Bertuccioli *et al.*, 1996).

- conclusion: rôle de PAI dans la fixation à l'ADN

En conclusion, le domaine PAI est effectivement indispensable à la fixation du domaine *paired* comme le démontrent les études biochimiques et les mutations de substitutions situées dans cette région, qui abolissent la fixation à l'ADN,. De plus, PAI est suffisant pour la fixation du domaine *paired*, dans le cas des protéines *Paired* et Pax-5 (s'il rencontre un consensus parfait). Mais RED est-il vraiment dispensable pour la fixation du



A A C G T C A C C G T G A C G T **oligo paired cristallo**  
(Xu *et al.*, 1995)

T T C A C G C G C T T G A G T T C A C A G C T C G A G T A **P6CON**  
(Xu *et al.*, 1995)

T T C A C G C <sup>A</sup> T G A T T C A C A C G C <sup>A</sup> T G A **5aCON**  
T T C A C G C T G A T T C A C A C G C T G A  
(Epstein *et al.*, 1994)

**Figure 9 :** Quelques séquences de fixation pour les domaines paired des protéines Pax.

Les séquences de classe I contiennent un consensus en 3' permettant la fixation de la portion N terminale du domaine paired (PAI).

Les séquences de classe II contiennent un consensus en 5' permettant la fixation de la portion C terminale du domaine paired (RED).

domaine paired chez toutes les protéines Pax? Le paragraphe suivant tente de répondre à cette question.

#### *b) rôle de RED*

Les études de Czerny en 1993 (Czerny *et al.*, 1993) ont permis, d'une part, de mettre en évidence la structure bipartite du domaine paired; d'autre part, de révéler la possibilité de liaison à l'ADN du sous domaine RED.

Czerny a établi une séquence consensus de fixation pour le domaine paired de BSAP (Pax-5), en alignant les séquences rencontrées dans ses gènes cibles. Par des expériences de retard sur gel, en utilisant des séquences cibles mutées et une protéine BSAP délétée ou non de son domaine RED, il a montré que la partie 5' de la sonde fixait la région amino terminale du domaine paired (PAI) et la partie 3', la région carboxy terminale (RED). Partant de cette observation, il a classé les séquences cibles en deux groupes. Celles du groupe II comportent une séquence en 5' proche du consensus de fixation pour PAI. La séquence consensus de cette classe est **TGGTCACGC**. On reconnaît le consensus de fixation pour l'hélice 3 de PAI (bases soulignées). Les séquences de groupe I possèdent des bases clefs en 3' pour la liaison de RED (figure 9).

#### le cas des protéines Pax au domaine RED "actif"

D'après les résultats de retard sur gel réalisés avec la protéine Pax-5, Czerny a défini le "principe de bases compensatoires". Celui-ci stipule que si une séquence s'éloigne quelque peu du consensus établi pour la fixation de PAI, alors le domaine paired peut tout de même se lier à l'ADN si le domaine RED rencontre des bases favorables à sa fixation. Ce principe s'applique aux protéines Pax-5 et Pax-1.

Toutes les protéines Pax sont capables de fixer les séquences de classe II, sauf Pax-6. Cette dernière ne fixe que deux séquences sur 5 de la classe II définie par Czerny (1993). Cependant, cette protéine fixe relativement bien les séquences de classe I consensuelles pour RED. De ces observations peuvent découler deux hypothèses. *i)* RED est une région importante pour la fixation de Pax-6. RED pourrait même surpasser la fonction de PAI dans sa capacité à lier l'ADN, contrairement aux autres protéines Pax. Dans ce sens iraient les résultats d'Epstein (Epstein *et al.*, 1994b) qui mettent en évidence la fixation de Pax-6 par RED, dans le cas où la conformation de l'hélice 3 du domaine PAI est bouleversée par l'insertion d'un exon supplémentaire (exon 5a). *ii)* Il est possible aussi que le domaine PAI

de Pax-6 soit capable de se fixer seul, ou de participer davantage à la liaison à l'ADN s'il rencontre une séquence proche du consensus établi pour sa fixation. Cette protéine nécessite en effet, la substitution du G le plus en 5' du site de liaison à l'ADN de Pax-5 GTCACGC, par un T, pour permettre à l'Asn, en position 47, spécifique de Pax-4 et Pax-6, de reconnaître sa séquence cible (Epstein *et al.*, 1994; Xu *et al.*, 1995). En réalité, les deux hypothèses i) et ii) sont possibles et dépendent du type de séquence cible que rencontre la protéine Pax-6. En effet, récemment, Yamaguchi *et al.*, (Yamaguchi *et al.*, 1997) ont montré que le domaine paired de Pax-6 (ou la protéine entière) se fixe bien par son domaine PAI sur une sonde de type P6CON (cf. figure 9). Tandis que cette même protéine se fixe par son domaine RED sur une sonde de type 5aCON. En conséquence, une isoforme de Pax-6, qui contient une insertion de 14 acides aminés dans l'hélice de reconnaissance de PAI (exon 5a non épissé), ne peut plus reconnaître les séquences consensus pour la fixation de PAI. Elle ne se lie alors à l'ADN que par son domaine RED et se fixe sur des séquences de type 5aCON.

#### le cas des protéines Pax "RED inhibé"

Les protéines Pax-3, Gsb, Prd et Poxn ne fixent que des séquences proches du consensus 5', région qui permet la liaison de PAI. Par contre, aucune d'entre elles ne se fixe sur les séquences pourvues des bases importantes pour la fixation de RED (bases déterminées à partir de Pax-5) (Czerny *et al.*, 1993). Deux hypothèses peuvent expliquer ces observations: ou le sous domaine RED de ces protéines se fixe sur un consensus très différent de celui obtenu pour Pax-5, ou bien il est absolument incapable de se fixer sur l'ADN. La seconde hypothèse semble la plus probable, étant donné que la séquence obtenue par la méthode classique de SELEX pour la protéine Paired dans l'étude cristallographique, a abouti à un consensus uniquement pour la partie 5'. Or, cette région correspond à la liaison de PAI, comme le confirme la structure tridimensionnelle du complexe ADN/domaine paired. Par conséquent, au moins pour la protéine Paired, il est démontré que le sous domaine RED est inactif pour la fixation directe de la protéine à l'ADN.

#### *c) bilan sur la fixation du domaine paired: rôles respectifs de PAI et RED.*

Tous les domaines paired des protéines Pax utilisent leur sous domaine PAI pour se fixer à l'ADN sur un consensus similaire pour toutes les protéines. (Seul le domaine paired

de Pax-6 préfère un T contre un G; cf. ci-dessus). Cette observation issue des études biochimiques et cristallographiques corrèlent avec la forte conservation du sous domaine PAI. En comparaison, le sous domaine RED diverge beaucoup plus que PAI. RED est incapable de se lier à l'ADN dans le cas des protéines Pax-3, Prd, Gsb et Poxn (Czerny *et al.*, 1993). Par contre, dans le cas d'un consensus incomplet pour PAI, si quelques bases du consensus de RED sont respectées (*bases compensatoires*), alors RED permet la fixation du domaine paired chez certaines protéines comme Pax-1 et Pax-5. Dans le cas de Pax-6, RED et PAI sont deux domaines de liaison à l'ADN indépendants. RED ou PAI se fixent sur leur cible respective. Lorsque le second sous domaine ne trouve pas de séquence consensus pour sa fixation, il n'établit que des liaisons non spécifiques avec l'ADN (Yamaguchi *et al.*, 1997) et solidifie la fixation à l'ADN de PAI dans la protéine Pax-6.

### 3 Coopérativité entre le domaine paired et le domaine homéo des protéines Pax pour la fixation à l'ADN.

Les protéines Pax contiennent donc deux grands domaines de fixation à l'ADN: le domaine paired et le domaine homéo. Le domaine homéo de type paired se fixe le plus fréquemment en dimères sur l'ADN, dans les études *in vitro*. Le domaine paired rencontré au chapitre précédent, est composé de deux sous domaines: PAI et RED. Selon les types de protéines Pax, seul le domaine PAI sera sollicité pour la fixation à l'ADN ou PAI et RED de façon synergique, ou seulement RED dans des cas isolés. Comment se comportent alors les protéines Pax dans leur entité, pour celles comme Pax-3/4/6 et 7 qui contiennent en plus d'un domaine paired, un domaine homéo? Pour répondre à cette question, la séquence consensus de fixation pour une protéine Pax contenant à fois un domaine paired et un domaine homéo a été établie *in vitro*.

#### 3-1 Séquence cible de fixation.

Par des expériences de SELEX, une séquence consensus pour la fixation de la protéine Paired, contenant un domaine paired et un domaine homéo entiers, a pu être établie. Cette séquence, notée **PHO** est constituée d'un site consensus pour la fixation de PAI, **GTCACGC** collé à un site **ATTA**, permettant la fixation d'un seul domaine homéo .

**PHO:** GATCTCTTCCAATTAGTCACGCTTGAGTG (Jun et Desplan, 1996)



### 3-2 Etudes biochimiques

Sur le promoteur du gène *even-skipped* de *Drosophila*, il existe un site PTE, superposable à **PHO**: PTE : GGCAATTAGCCACGGTGAGG. Le gène *even-skipped* est un gène "pair-rule" qui joue un rôle clef dans l'établissement antéro-postérieur de la segmentation chez la larve. Il comporte une phase d'expression précoce et une seconde plus tardive. Le site PTE appartient à un élément *cis* dit "tardif", car impliqué dans l'induction de l'expression tardive du gène *even-skipped*. La protéine Paired est capable de se fixer *in vitro* sur l'élément PTE. La mutation du site ATTA ou de la séquence en aval, requise pour la fixation du domaine paired, réduit l'activité de l'élément PTE. Ces résultats suggèrent que les deux domaines de fixation à l'ADN de la protéine Paired sont requis pour la régulation de l'expression tardive du gène *even-skipped*, via l'élément PTE (Fujioka *et al.*, 1996).

D'autre part, des expériences de "sauvetage" de la perte de fonction de la protéine Paired de *Drosophila* ne sont possibles seulement si les deux domaines PAI et homéo sont présents et situés sur la même molécule, suggérant l'existence probable d'interactions entre ces deux domaines (Bertuccioli *et al.*, 1996). De plus, la mutation Sp (d), caractérisée par la substitution de la glycine en position 9 du domaine paired par une arginine, induit non seulement une chute de l'affinité de la protéine Pax-3 mutée pour e5, mais aussi une diminution de sa fixation sur la séquence P2, séquence de reconnaissance pour le domaine homéo. Cette observation suggère que le domaine paired coopère avec le domaine homéo pour sa fixation sur P2.

## 4 Conclusions sur la fixation des protéines Pax.

Les protéines Pax comportent deux domaines de fixation à l'ADN: le domaine homéo et le domaine paired. Ce dernier est lui-même subdivisé en deux sous domaines PAI et RED. La fixation des protéines Pax est une combinaison complexe des différentes capacités de liaison de ces domaines, assortie d'une possibilité de dimérisation et/ou de coopérativité entre domaines identiques ou distincts. Ces aptitudes à "spécificités variables" des protéines Pax assurent un choix de séquences cibles varié mais spécifique et suggèrent la régulation de nombreux gènes cibles et l'implication des protéines Pax dans différentes fonctions.

### III Fonction des gènes PAX

#### 1 Les gènes Pax sont des facteurs de transcription

##### 1-1 Les gènes Pax ont une activité transactivatrice

Les protéines Pax agissent en facteurs de transcription et plus précisément en activateurs de la transcription. Cette fonction de modulation de la transcription est permise par les propriétés communes aux facteurs de transcription d'être localisés dans le noyau de la cellule (Bopp *et al.*, 1989); (Carrière *et al.*, 1993); (Dressler et Douglass, 1992), de lier l'ADN (cf. chapitre II), et d'interagir avec les facteurs de la machinerie transcriptionnelle basale (Chalepakis *et al.*, 1991); (Plaza *et al.*, 1993) ).

##### a) *Le domaine transactivateur est situé dans la région C terminale*

Le domaine transactivateur est situé derrière le domaine homéo, dans la région carboxy terminale riche en sérine, thréonine, proline (Carrière *et al.*, 1995); (Czerny et Busslinger, 1995) pour la protéine Pax-6 et pour la majorité des protéines Pax étudiées. L'importance de cette région dans la protéine Pax-6 est illustrée par la colocalisation de mutations responsables d'aniridie et du type mutant *small eye* chez la souris (Glaser *et al.*, 1994); (Hill *et al.*, 1991). La région transactivatrice de la protéine Pax-3 correspond à ses 78 acides aminés C terminaux (Chalepakis *et al.*, 1994). De même, les protéines Pax-2/5 et 8 possèdent une région transactivatrice de 55 acides aminés dans leur région C terminale, derrière leur domaine homéo tronqué. La protéine Pax-9, qui ne présente aucun vestige de domaine homéo, possède un domaine transactivateur localisé dans la partie C terminale de la protéine, c'est-à-dire derrière l'octapeptide (Nornes *et al.*, 1996). Cette région, fusionnée au domaine de liaison à l'ADN GAL-4, stimule en effet l'activité d'un promoteur minimal tk, par la fixation du domaine GAL-4 sur ses sites de reconnaissance placés en orientation enhancer ou promoteur.

De façon globale, les protéines Pax activent la transcription de gènes rapporteurs en se fixant sur des séquences cibles localisées dans leur promoteur ou en en situation enhancer. Cependant, cette modulation positive de la transcription est due à la somme d'une capacité transactivatrice liée au domaine C terminal des protéines Pax, associée à des activités inhibitrices réparties dans diverses régions de la molécule. Le paragraphe suivant résume les interactions connues entre ces différents domaines intramoléculaires aux effets antagonistes ou parfois synergiques.

## *b) Modulation intramoléculaire de l'activité transactivatrice des protéines Pax*

### *- activité inhibitrice intramoléculaire*

Le domaine paired de Pax-9, fusionné au domaine de liaison à l'ADN, GAL-4, réprime un promoteur minimal tk précédé de 5 sites de fixation pour GAL-4. De même, il est capable d'inhiber la fonction transactivatrice du domaine C terminal de Pax-9 (Nornes *et al.*, 1996). Par ailleurs, la protéine Pax-3 contient deux régions inhibitrices: ses 90 acides aminés N terminaux, dont 57 acides aminés du domaine paired, et son domaine homéo (Chalepakakis *et al.*, 1994). Pour les protéines Pax-2/5 et 8, la région inhibitrice est située derrière le domaine transactivateur, à l'extrémité C terminale (Dörfler et Busslinger, 1996). D'autre part, PAI et RED du domaine paired de Pax-6 agissent de façon antagoniste sur la régulation de la transcription. Ainsi, l'activation de la transcription d'un gène rapporteur par la fixation du domaine PAI sur sa séquence cible (P6CON) est meilleure lorsque le domaine voisin, RED est muté. De même, la réciproque est observée aussi: meilleure activation de la transcription par la fixation du domaine RED sur sa séquence cible (5aCON) lorsque le domaine PAI est muté. Ce phénomène s'explique par une liaison non spécifique du sous domaine paired voisin. La mutation de ce dernier l'empêche alors de se fixer à l'ADN et inhibe l'activation de la transactivation induite par le deuxième sous domaine paired fixé sur sa séquence cible (Yamaguchi *et al.*, 1997). Cette situation existe *in vivo*, dans le cas de l'isoforme Pax-6/ exon 5a, qui comporte un domaine PAI incapable de se fixer à l'ADN (Yamaguchi *et al.*, 1997); (Wilson *et al.*, 1993).

### *- synergie activatrice entre homodimères*

Le domaine homéo de type paired se fixe en dimères coopératifs sur l'ADN. Cette association améliore d'une part la spécificité de liaison à l'ADN et d'autre part, induit l'activation synergique de la transcription; dans le cas notamment de deux domaines homéo fixés en dimères coopératifs sur un site P2 (Czerny *et al.*, 1993).

## *c) conclusion sur l'activité transactivatrice des gènes Pax*

Les protéines Pax sont avant tout des régulateurs positifs de la transcription. Leur mécanisme de régulation est complexe, puisqu'elles possèdent un domaine transactivateur, dont l'activité est modulée par une ou plusieurs régions inhibitrices. A ce jour, rien n'est connu sur le fonctionnement moléculaire de ces régions. Les domaines inhibiteurs pourraient interagir directement avec la région transactivatrice, bloquant ainsi son activité.

La levée d'inhibition pourrait être déclenchée par une phosphorylation. Ces domaines pourraient aussi fixer des cofacteurs qui interagiraient avec les éléments de la machinerie transcriptionnelle et stopperaient l'activation de la transcription médiée par les domaines transactivateurs des protéines Pax. Des mécanismes de régulation aussi poussés suggèrent l'importance d'une intensité modérée de l'activité des protéines Pax. Ces protéines présentent en effet une activité relativement faible par rapport à nombre de facteurs de transcription. Les gènes cibles doivent être en conséquence extrêmement sensibles à l'activation par ces protéines. Ce qui suggère qu'une activation trop intense de ces gènes serait néfaste au bon fonctionnement de l'organisme. Cette "suractivation" des gènes cibles des protéines Pax pourrait être causée par une protéine hyperactive ou par une concentration intracellulaire trop élevée. Il a effectivement été montré qu'un nombre de copies surnuméraires des gènes Pax pouvait induire des malformations chez l'embryon. Réciproquement, la délétion ou la mutation d'un allèle entraîne également divers troubles selon le gène Pax étudié. On parle alors d'"effet-dose" des gènes Pax.

#### 1-2 Effet dose des gènes Pax

##### a) *Caractère haplo-insuffisant des gènes Pax.*

L'haploinsuffisance se définit par l'apparition d'un phénotype anormal causée par la mutation ou la délétion d'un seul allèle du gène considéré. Tous les membres de la famille des gènes Pax présentent un caractère haplo-insuffisant qui se manifeste par leurs mutations semi-dominantes. Les mutations *Undulated* de *pax-1*, *Spotch* de *pax-3*, *Small eye* de *pax-6*, chez la souris, etc. (se référer au paragraphe I-2-3) révèlent l'importance du nombre de copies des gènes Pax pour leur activité totale. Par ailleurs, récemment, il a été mis en évidence que la surexpression du gène Pax-6 pouvait aussi induire des malformations.

##### b) *La surexpression d'un gène Pax cause des anomalies du développement*

Le locus Pax-6 humain, assorti de ses séquences de régulation (ou du moins prolongé de séquences suffisamment longues en 3' et en 5' pour supposer que toutes ses séquences de régulation sont présentes: 200 kb de chaque côté) a été introduit dans un vecteur YAC et injecté dans des oocytes de souris. Les souris transgéniques obtenues ont permis d'une part de rétablir un phénotype normal chez des souris porteuses de la mutation *Small eye* homozygote ou hétérozygote. D'autre part, chez des souris normales, l'augmentation du

nombre de copies induite par la présence du transgène, a provoqué des défauts dans l'oeil au niveau du corps ciliaire, de l'iris et de la cornée (Schedl *et al.*, 1996). Ces résultats prouvent que les gènes cibles activés par les gènes Pax sont sensibles à une sous expression comme à une surexpression de leur transactivateur. Le mécanisme par lequel les gènes cibles réagissent de façon dose-dépendante aux protéines Pax n'a pas encore été mis en évidence.

## 2 Gènes cibles des protéines Pax

En réalité, le nombre de gènes cibles isolés jusqu'à présent n'est pas très élevé si l'on considère les multiples capacités de reconnaissance à l'ADN des protéines Pax (se référer au chapitre II). Les sites de fixation des protéines Pax retrouvés dans les promoteurs naturels se superposent assez bien avec les consensus déterminés préalablement au cours des études biochimiques, consensus pour le domaine paire (type PRD), pour le domaine homéo (type HD), ou pour le domaine PAI +Homéo (type PHO ou PHn),

Les gènes cibles des protéines Pax peuvent être classés en deux groupes:

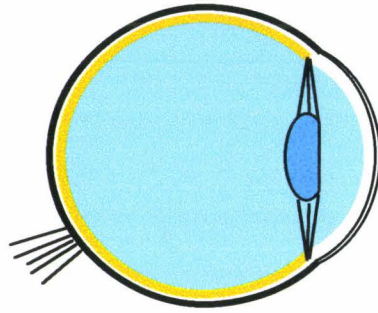
### -les gènes qui interviennent dans la mise en place du système nerveux

Y sont inclus, les gènes qui participent à l'établissement des frontières morphologiques, comme *engrailed*, *goosberry* et les gènes qui codent des protéines de structure, comme les molécules d'adhésion cellulaire neurales (N-CAM).

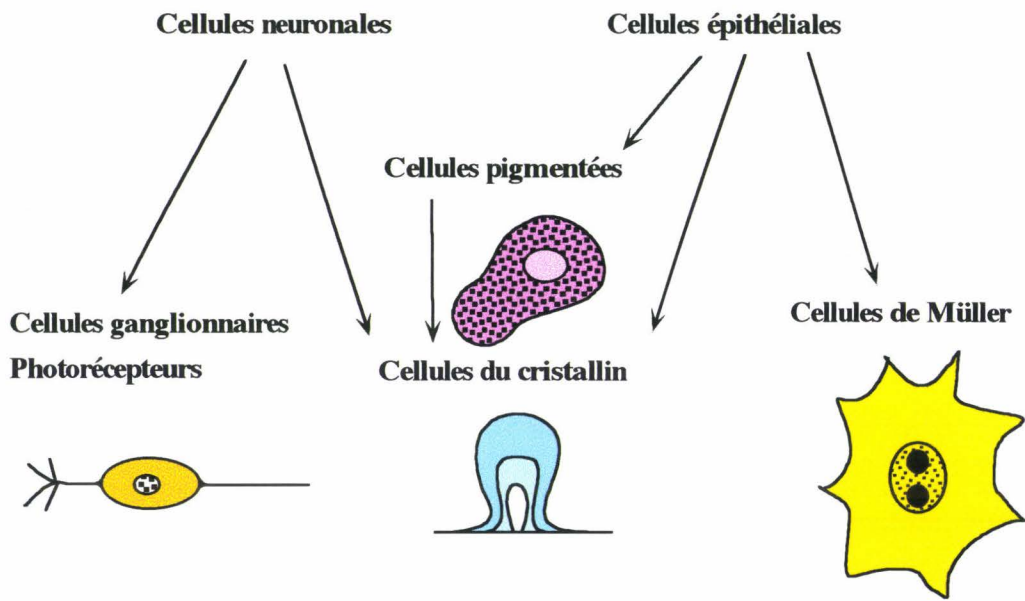
### - les gènes qui interviennent dans l'organogenèse

Parmi lesquels on trouve, des gènes qui codent des protéines de structure, comme les cristallines du cristallin, et des gènes qui traduisent des protéines intervenant dans le fonctionnement de l'organe, comme les gènes de la rhodopsine, du glucagon, de l'insuline, de la somatostatine, de la thyroglobuline, de la thyropéroxidase...

Le rôle des protéines codées par les gènes cibles des gènes Pax rejoint le rôle estimé des gènes Pax d'après l'analyse de leur cartographie d'expression et des défauts induits par leurs mutations au cours de l'embryogenèse (cf. chapitre I, paragraphe I-2). A savoir, les gènes Pax ont d'une part, un rôle dans la mise en place du système nerveux central au cours du développement. D'autre part, ils jouent aussi un rôle dans l'organogenèse, où ils interviennent à la fois, sur l'apparition des structures différenciées (en stimulant l'expression de gènes qui codent des protéines de structures), comme sur le fonctionnement de l'organe (en activant des gènes synthétisant des protéines régulatrices).



**Cellules souches multipotentes**



**Figure 10 : Potentialités de transdifférenciation de l'œil des Vertébrés.**

Le gène Pax-6 est le gène Pax pour lequel le plus grand nombre de gènes cibles ont été identifiés, qui sont pour la plupart, impliqués dans l'organogenèse de l'oeil. Le chapitre suivant est focalisé sur ce gène qui est aussi le centre d'étude de ma recherche de thèse.

#### IV Le cas du gène *pax-6*

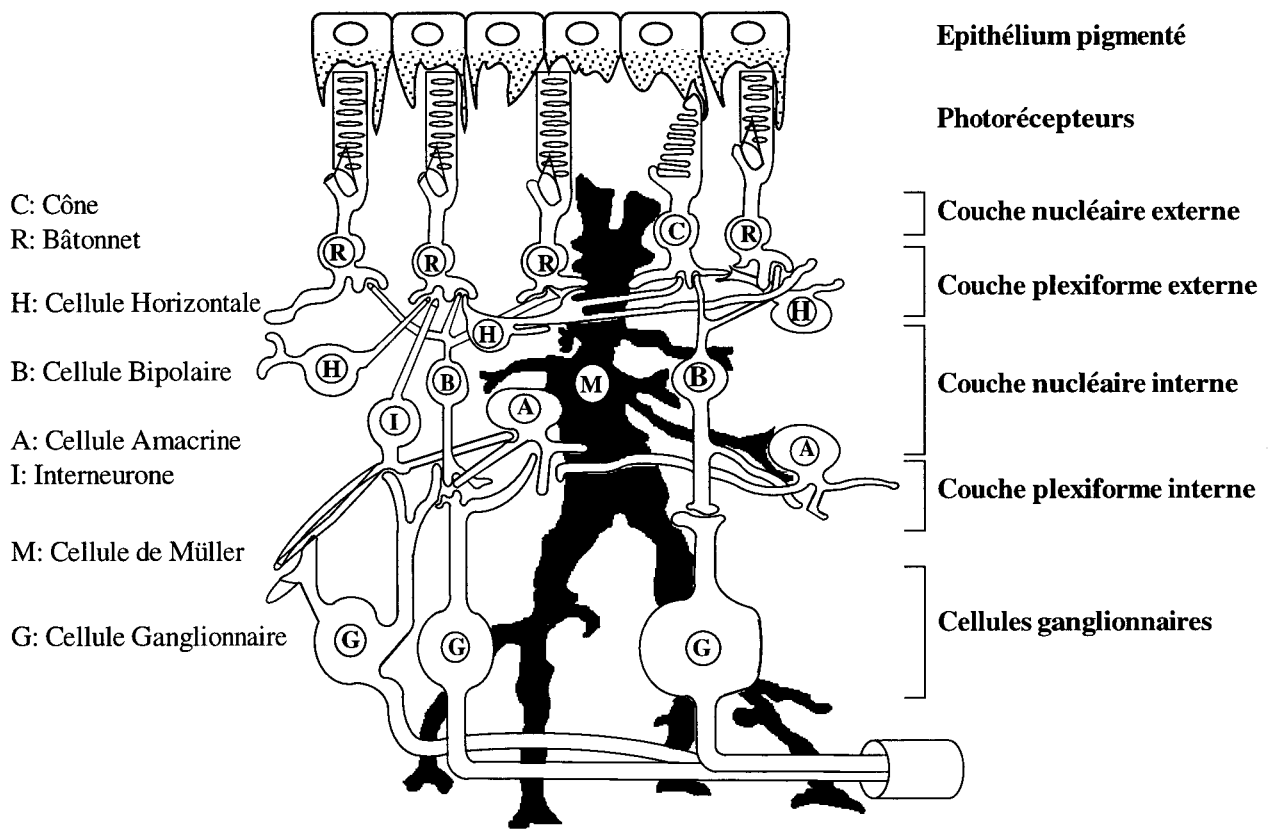
##### 1 *pax-6*, gène organisateur de l'oeil

L'introduction de *eyeless*, l'homologue de *pax-6* de *Drosophila*, dans les disques imaginaires des embryons de l'insecte induit l'apparition d'yeux ectopiques (localisés dans les ailes, les pattes ou les antennes de l'insecte) morphologiquement normaux et composés d'ommatidies totalement différenciées avec un nombre de cellules photoréceptrices normal (cf. (Halder *et al.*, 1995). Ces résultats suggèrent que *pax-6* est le gène maître pour l'organisation de l'oeil, puisqu'introduit dans un tissu non programmé pour générer un oeil, comme les disques imaginaux à l'origine des pattes, des ailes ou des antennes, il est capable à lui-seul de donner naissance à un oeil normalement constitué.

En outre, les résultats de nombreux groupes vont dans le sens d'une implication importante du gène *pax-6* dans l'oeil tant à un niveau structural que fonctionnel, comme le suggèrent ses gènes cibles identifiés à ce jour. En effet, *pax-6* régule aussi bien des gènes de structure de l'oeil comme les cristallines, et des gènes intervenant dans la transduction du signal lumineux, comme la rhodopsine 1. Par ailleurs, sa cartographie d'expression s'étend sur toutes les structures de l'oeil, de la couche pigmentaire à la couche nucléaire interne et laisse présager l'existence de nombreux autres gènes cibles dans cet organe, encore non identifiés.

##### 2 Notre modèle d'étude, la neurorétine d'oiseau

L'oeil des vertébrés au cours du développement embryonnaire se présente comme un modèle biologique fructueux pour l'étude des bases moléculaires de la différenciation, du fait de la potentialité que possèdent ses cellules à s'interconvertir en un type cellulaire ou un autre. En effet, les cellules de la neurorétine possèdent la capacité de se transdifférencier en cellules cristallines ou en cellules pigmentées (Okada, 1976). Cette transdifférenciation peut être induite *in vitro* sous l'effet de facteurs de croissance comme le FGF (Fibroblast Growth factor) basique (Pittack *et al.*, 1991) (figure 10).



**Figure 11:** Organisation structurale de la neurorétine d'oiseau



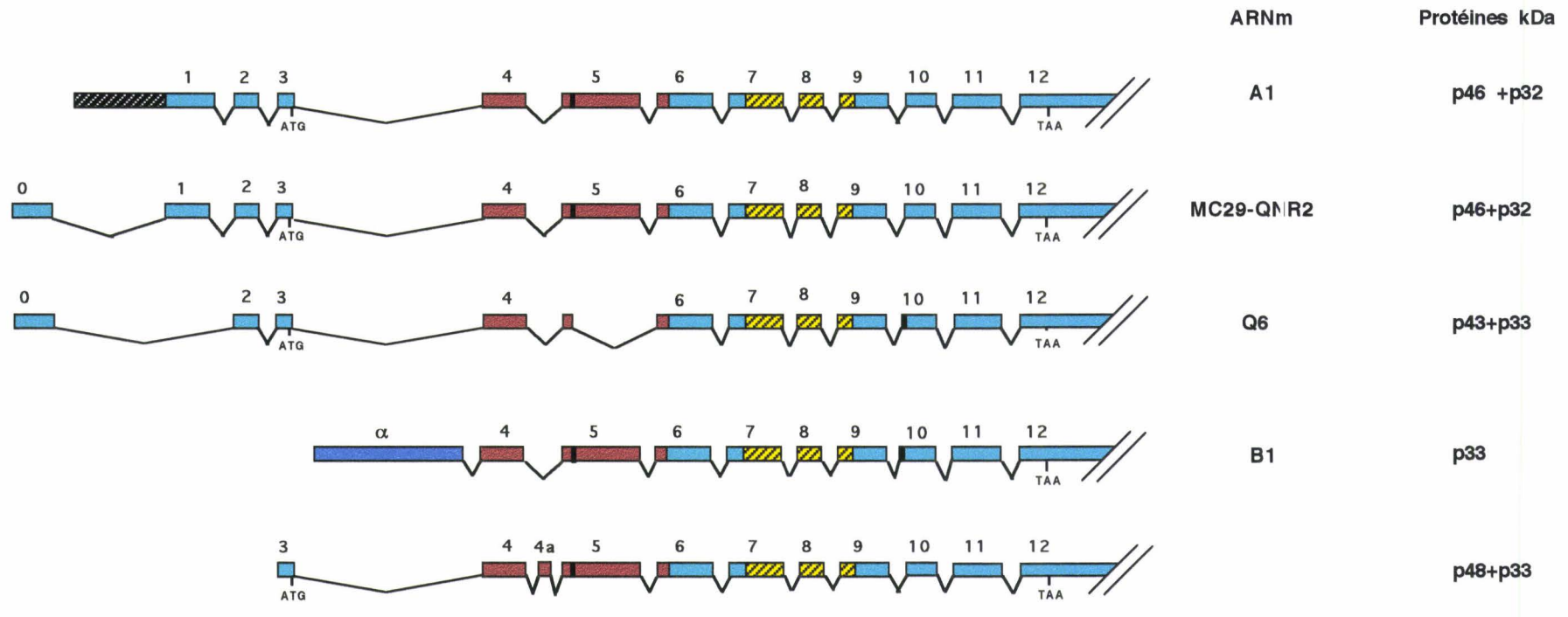
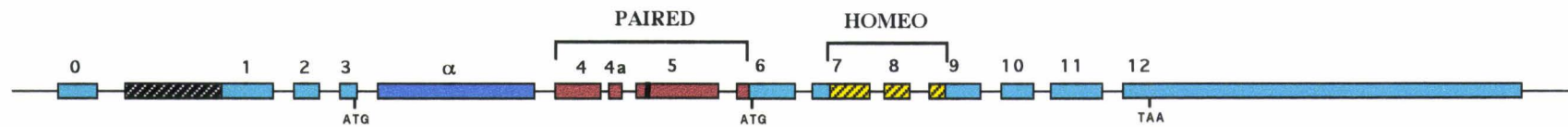
Notre groupe qui s'intéresse aux phénomènes de différenciation et de multiplication cellulaire a choisi de travailler sur la neurorétine pour mettre en évidence des gènes impliqués dans les processus de différenciation. La rétine d'oiseau (caille et poulet) a été retenue, d'une part, pour la facilité d'accès des embryons d'oiseaux, et d'autre part, pour la taille de l'oeil chez l'oiseau, qui offre ainsi un rendement avantageux .

Trois périodes sont définies dans le développement de la neurorétine aviaire: une période de prolifération des précurseurs neuroectodermiques en bordure de la cupule optique du deuxième au huitième jour d'incubation, une période de migration où les différentes couches cellulaires vont se mettre en place de E8 à E10 et enfin une période de différenciation terminale de E10 à l'éclosion (àE21). Le développement de la rétine suit un cours très similaire chez tous les vertébrés avec des différences temporelles liées à la durée du développement. La neurorétine différenciée est constituée de plusieurs types cellulaires: les neurones et les interneurones (photorécepteurs, cellules horizontales, bipolaires, amacrines et ganglionnaires) et les cellules gliales, les cellules de Müller (figure 11).

### 3 Le gène Pax-QNR

#### 3-1 Stratégie d'isolement

Des cellules de neurorétine disséquées au stade de la prolifération et mises en culture, cessent rapidement de se diviser et expriment plusieurs marqueurs de différenciation neuronale (Crisanti-Combes *et al.*, 1978). L'oncogène v-myc a été sélectionné pour ses propriétés relativement opposées puisqu'il peut aussi bien intervenir dans des processus de différenciation (il induit par exemple des caractères de différenciation dans les cellules de crête neurale comme il participe aux phénomènes de prolifération cellulaire (Fauquet *et al.*, 1990). Ainsi transformées par l'oncogène v-myc, les cellules de neurorétine se remettent à proliférer et d'autre part, les cellules se colorent en noir. Dans le but d'isoler les gènes potentiellement contrôlés par l'oncogène v-myc, un criblage différentiel a été réalisé entre une banque d'ADN complémentaire construite à partir de ces cellules de neurorétines infectées par le virus MC29 (appelées QNR MC29) et une banque d'ADN complémentaire construite à partir de cellules fibroblastiques infectées par ce même virus (QEC MC29). Ceci afin de n'isoler que des gènes spécifiques de la neurorétine infectée par le virus MC29. Deux ADNc ont été ainsi retenus, MC29QNR2 et QNR-71. Le gène QNR-71 est un gène marqueur de la mélanogénèse (l'homologue de Pmel17 humain), dont l'expression est



0.1 Kbp

Figure 12 : Structure génomique et isoformes du gène Pax-QNR.

induite par l'oncogène *v-myc* et par les produits du gène *microphthalmia*, gène clef du processus de mélanogénèse. Il est exprimé essentiellement dans la rétine pigmentaire et le derme de l'embryon d'oiseau au cours du développement embryonnaire, ainsi que dans les cellules de neurorétines transformées par *v-myc* en culture (cf. (Turque *et al.*, 1996). Mon sujet de thèse a porté sur le gène *Pax-QNR*, gène dont est issu l'ADNc MC29QNR2, l'autre ADNc isolé par criblage différentiel et totalement indépendant du gène *QNR-71*. Mon exposé sera donc concentré sur ce gène *Pax*.

### 3-2 Structure génomique de Pax-QNR et protéines traduites.

Le gène *Pax-QNR* s'étend sur 17 milles paires de bases. Il est constitué de 14 exons (figure 12). La boîte paired est composée de trois exons qui codent les 128 acides aminés du domaine paired: l'exon 4 code les deux feuillets  $\beta 1$  et  $\beta 2$ , les deux tours  $\tau 1$  et  $\tau 2$  et les deux premières hélices  $\alpha 1$  et  $\alpha 2$ , l'exon 5 code les hélices  $\alpha 3$ ,  $\alpha 4$  et  $\alpha 5$ , puis l'exon 6 qui code l'hélice 6. Le domaine homéo, composé de 60 acides aminés est lui codé par la fin de l'exon 7, l'exon 8 et le début de l'exon 9.

*Pax-QNR* code cinq protéines *in vivo*, notées P32, P33, P43, P46 et P48, en raison de leur masse moléculaire. Toutes ces protéines possèdent une région carboxy-terminale semblable, comprenant le domaine homéo, puis une queue C terminale. Les petites protéines de 32 et 33 KD, sont traduites à partir des codon(s) méthionine (s) interne(s), situés entre les domaines paired et homéo. Par contre, les grandes protéines varient dans leur domaine paired. Ainsi, la P48, possède dans son domaine paired 14 acides aminés de plus que la P46. Cette insertion provient d'un mécanisme d'épissage alternatif apportant l'exon 4a dans l'ARN messager codant la protéine P48. L'insertion de cet exon, noté exon 5a chez les autres espèces est conservé au cours de l'évolution, avec seulement un acide aminé substitué distinguant l'humain, la souris, la caille, l'axoloth et le zébrafish. Cet exon n'existe pas chez l'oursin et la drosophile. Sa présence modifie les capacités de fixation à l'ADN du domaine paired (cf. (Epstein *et al.*, 1994a) ; (Czerny et Busslinger, 1995). La P43 est traduite à partir d'un ARN messager chez lequel l'exon 5 est épissé; de ce fait, le domaine paired perd ses hélices 3, 4 et 5 et cette protéine est incapable de se fixer sur les cibles classiques de *pax-6*.

### 3-3 La régulation transcriptionnelle du gène Pax-QNR

Deux promoteurs ont été isolés, dénommés Po pour l'élément situé en 5' de l'exon 0 et P1 pour l'élément situé en 5' de l'exon 1 (Plaza *et al.*, 1993); (Plaza *et al.*, 1995a); ils produisent respectivement les ARNm MC29QNR2 et A1 qui divergent dans leur 5'UTR. Po et P1 sont tous deux activés par la p46 Pax-QNR. Po est également transactivé par la P75 c-myb (Plaza *et al.*, 1995a).

### 3-4 Sujet de thèse

*pax-QNR* génère aussi par un mécanisme d'épissage alternatif, un ARN messager noté B1, dont les exons 0, 1, 2 et 3 ont été épissés et remplacés par l'exon  $\alpha$ , situé entre les exons 3 et 4 sur l'ADN génomique. Cet ARNm ne contient pas de codon AUG en amont du domaine paired. Une partie de notre étude s'est portée sur l'ADNc B1 pour déterminer d'une part, si une protéine B1 pouvait être traduite par un codon alternatif de traduction dans le domaine paired, et d'autre part, étudier les capacités de fixation des protéines traduites à partir de ce nouveau messenger. Par ailleurs l'exon  $\alpha$  s'est avéré être un enhancer chez la caille. Une autre partie de mon étude a consisté à cloner l'homologue de l'exon  $\alpha$  chez la souris et vérifier si la fonction d'activation de la transcription était conservée. Une autre étude nous a permis de mettre en évidence une interaction entre la protéine P46Pax-QNR et la protéine Engrailed. Cette interaction inhibe la capacité à fixer l'ADN de la P46Pax-QNR et en conséquence diminue l'activité transcriptionnelle de P0 et P1 (eux mêmes activés par la protéine P46Pax-QNR et l'expression des ARNm de Pax-6 s'éteint).

*Résultats  
gènes Pax*

*Article 1*

## Article 1

### Identification and characterization of a neuroretina-specific enhancer element in the quail Pax-6 (Pax-QNR) gene.

S.Plaza, C. Dozier, M-C Langlois et S. Saule.

Molecular and Cellular Biology, 15, 892-903, 1995

Nous avons vu que Pax-6 s'exprimait essentiellement dans le système nerveux central et le pancréas au cours du développement embryonnaire. Dans l'oeil, les ARN messagers MC29QNR2 initiés au promoteur Po sont détectés exclusivement dans la neurorétine, tandis que l'ARNm A1 initié au promoteur P1 et l'ARNm B1 s'expriment aussi dans la rétine pigmentaire. Dans cet article, nous avons cherché à localiser les séquences *cis* sur le gène Pax-6 responsables de la restriction de l'expression des ARNm MC29QNR2 dans la neurorétine. Une région localisée à 7,5 kpb en aval du promoteur Po a été identifiée, appelée enhancer  $\alpha$ . Cet enhancer, placé en amont ou en aval du promoteur Po ou du promoteur tk, en orientation sens ou antisens, stimule la transcription d'un gène rapporteur dans les cellules de neurorétine mais pas dans les cellules de rétine pigmentaire ou dans les fibroblastes de poulet. L'activité de cet enhancer est non seulement spécifique à la neurorétine mais elle est également spécifique d'un moment précis du développement embryonnaire. La stimulation de la transcription par l'enhancer  $\alpha$  commence à être efficace à partir de 7 jours de développement dans les cellules de neurorétine de poulet. Parallèlement, on observe une augmentation d'expression graduelle des ARNm MC29QNR2 de E6 à E14 (stade à partir duquel le taux d'expression reste constant jusqu'à l'éclosion). Par ailleurs, l'activité de l'enhancer  $\alpha$  ne s'exerce pas sur P1, le deuxième promoteur du gène Pax-6.

Par des expériences d'empreintes sur l'ADN - l'enhancer  $\alpha$  incubé avec des extraits protéiques de neurorétine E7 puis digéré à la DNase I - 4 régions de fixation de protéines présentes dans la neurorétine aviaire ont pu être mises en évidence, dénommées DF1, DF2, DF3 et DF4. Ces boîtes fixent toutes une protéine de 110 kDa ainsi que des protéines de taille inférieure.

Afin de s'assurer que l'enhancer  $\alpha$  n'était pas spécifique de l'embryon de poulet, le fragment génomique correspondant a été isolé chez la souris. Les 4 boîtes DF1 à DF4 ont été retrouvées chez cette espèce et présentent de fortes identité de séquence avec leur homologue chez le poulet: 100 % pour DF3 et DF4 et 88,5% et 92,3 % respectivement pour

DF1 et DF2. Ce fragment de souris montre aussi une activité enhancer spécifique des cellules de neurorétine. Cloné en 3' d'un vecteur rapporteur comprenant un promoteur minimal (le promoteur tk) lié au gène CAT, le fragment a de souris est capable, après transfection transitoire dans des cellules de neurorétine de caille, à 7 jours de développement, de stimuler la transcription initiée à partir du promoteur minimum, comme son homologue de caille. La fonction enhancer du fragment  $\alpha$  est par conséquent conservé chez la souris.



## Identification and Characterization of a Neuroretina-Specific Enhancer Element in the Quail *Pax-6* (*Pax-QNR*) Gene

SERGE PLAZA, CHRISTINE DOZIER, MARIE-CLAIRE LANGLOIS, AND SIMON SAULE\*

Laboratoire de différenciation cellulaire et moléculaire, Centre National de la Recherche Scientifique EP 56, Institut Pasteur, 59019 Lille Cedex, France

Received 29 September 1994/Returned for modification 27 October 1994/Accepted 21 November 1994

Using nuclear run-on assays, we showed that the tissue-specific expression of quail *Pax-6* (*Pax-QNR*) P0-initiated mRNAs is due in part to regulation of the gene at the transcriptional level. Regulatory sequences governing neuroretina-specific expression of the P0-initiated mRNAs were investigated. By using reporter-based expression assays, we characterized a region within the *Pax-QNR* gene, located 7.5 kbp downstream from the P0 promoter, that functions as an enhancer in neuroretina cells but not in nonexpressing P0-initiated mRNA cells (quail embryo cells and quail retinal pigment epithelial cells). This enhancer element functioned in a position- and orientation-independent manner both on the *Pax-QNR* P0 promoter and the heterologous thymidine kinase promoter. Moreover, this enhancer element exhibited a developmental stage-specific activity during embryonic neuroretina development: in contrast to activity at day E7, the enhancer activity was very weak at day E5. This paralleled the level of expression of P0-initiated mRNAs observed at the same stages. Using footprinting, gel retardation, and Southwestern (DNA-protein) analysis, we demonstrated the existence of four neuroretina-specific nuclear protein-binding sites, involving multiple unknown factors. In addition we showed that the quail enhancer element is structurally and functionally conserved in mice. All of these results strongly suggest that this enhancer element may contribute to the neuroretina-specific transcriptional regulation of the *Pax-6* gene in vivo.

A large number of vertebrate developmental control genes have been identified by their homology to *Drosophila* genes that regulate pattern formation and contain conserved DNA binding domains, like the homeobox or the paired box (for a review, see reference 14). The *Pax* family, identified in the vertebrate genome by their homology to the paired box sequences of the *Drosophila* segmentation gene *paired*, consists of nine unlinked genes (36, 44) highly conserved during evolution, in various distantly related species such as nematodes, zebra fish, *Drosophila* species, mice, humans, quails, and chickens (25, 27, 33; for a review, see reference 8). The paired box is present alone (in the genes *Pax-1*, *Pax-2*, *Pax-5*, and *Pax-8*) or together with the homeobox domain (*Pax-3*, *Pax-4*, *Pax-6*, and *Pax-7*) (for a review, see reference 41). The predicted structure, nuclear localization, and in vitro binding strongly indicate that Pax proteins are transcriptional activators, and this has been directly demonstrated for the *Pax-1* (7), *Pax-2* (16), *Pax-5* (1), *Pax-6* (31), and *Pax-8* (46) proteins.

The *Pax* genes are expressed in a spatially and temporally restricted pattern during embryonic development (for a review, see reference 41), indicating that these genes may be involved in the control of morphogenesis and pattern formation. Indeed, the importance of these genes in development is demonstrated by the developmental alterations in mice associated with mutations in the *Pax* genes (for a review, see reference 41). Mutations in the *Pax-6* gene have been associated with the mouse mutation *small eye* (21), and the corresponding human gene (*AN*) has been found deleted or mutated in many cases of the human disorder aniridia (17; for a review, see reference 38). Moreover, this gene is the homolog of the *eyeless* gene of *Drosophila* species, which is also involved in eye morphogen-

esis (33). The *Pax-6* gene is expressed in the developing central nervous system, the optic cup, the lens, and the overlying epithelium prior to morphological differentiation, and later in the neuronal layer of the retina (25, 27, 43). All of this strongly suggests that *Pax-6* may be involved in the regulation of some of the inductive events that occur during the formation of the eye. In addition, *Pax-6* is also expressed in the endocrine pancreas, where the cells have some traits in common with neurons, despite their distinct embryological origin (42). Thus, the restricted expression of this gene suggests transcriptional regulation of expression, with sequence-specific DNA-binding proteins interacting with tissue-specific enhancer and promoter elements. However, the *cis*-acting elements and *trans*-acting factors for directing the tissue-specific expression of the *Pax-6* gene have not yet been identified.

Our research focuses on the regulatory elements involved in the tissue-specific expression of the quail *Pax-6* gene (termed *Pax-QNR*). This gene expresses at least two mRNAs, named MC29-QNR2 and B1, differing by their 5' untranslated region and resulting from alternative splicing (13). We have recently characterized the promoter region initiating the MC29-QNR2 mRNA (termed P0), and we have shown that the promoter sequences located between positions -1386 and +127 (with respect to the transcription start site) function in *Pax-QNR*-expressing cells (quail neuroretina cells [QNR]) and *Pax-QNR*-nonexpressing cells (quail embryo cells [QEC]) (31). This finding suggested that tissue-specific expression of *Pax-QNR* requires additional upstream or downstream unidentified sequences or is the result of posttranscriptional regulation, as for the *N-myc* gene (2).

In this report, we show that neuroretina-specific expression of P0-initiated mRNAs is, in part, the result of transcriptional control. We have analyzed further the molecular mechanisms responsible for this neuroretina-specific regulation. We have identified a neuroretina-specific 3' enhancer in the quail *Pax-6* gene, located 7.5 kbp downstream from the P0 promoter. In

\* Corresponding author. Mailing address: Laboratoire de différenciation cellulaire et moléculaire, CNRS EP 56, Institut Pasteur, 1 rue Calmette, 59019 Lille Cedex, France. Phone: 33 20 87 77 86. Fax: 33 20 87 79 08.

constructs containing this promoter, the 3' enhancer functions in a cell type-specific manner, increasing the activity of the reporter gene from the P0 or the minimal thymidine kinase (TK) promoter in QNR, but not in cells in which the P0-initiated mRNAs are not detected. We have shown that this neuroretina-specific enhancer functions in a developmental stage-dependent manner, and we have demonstrated the existence of nuclear proteins in neuroretina cells that can interact specifically with this enhancer region. In addition, this enhancer is also structurally and functionally conserved in the mouse *Pax-6* gene, strongly suggesting that this regulatory element plays an important role in *Pax-6* gene expression in the neuroretina.

## MATERIALS AND METHODS

**Oligonucleotides.** The following set of synthetic oligonucleotides was used (all sequences are written 5' to 3'): 1F, gacCTATTGTGCTGACGATGCAATTTCA; 1R, gatTGAAAATTGCATCGTACGACAAAATAG; 2F, gatCTGTCTATAAAGTGACA; 2R, gatIGTCACTTATGACAG; 3F, gatCTGATCTTTCAATTAGCCTTCCA; 3R, gatTGGAAGGCTAATTTGAAAAGATGAC; 4F, gatCCGGGGCGACTTCCGCTATTCCAGAAAATTAAGCTCAAACATGACGTGACGACTA; 4R, gatCTAGCTGCACGTCGAAGTTTGAGCTTAAATTTCTGGAATAGGCGGAAGTCGCCCC; 5R, CTAGCTGCACGTCAAATTTGAGC; 3X1F, CTATTGTGCTGACGATGCAATTTTCTTATTTGCTGACGATGCAATTTTCTTATTTGCTGACGATGCAATTTTCTTATTTGCTGACGATGCAATTTTCTGAGCT; 3X1R, CGAAAATGTCATCGTACGACAAAATAGGAAAATTTGCATCGTACGACAAAATAGGAAAATTTGCATCGTACGACAAAATAG; and Sp1 (containing a consensus site for the transcription factor Sp1, shown in boldface type), GATCGATCGGGGGCGGATC. The underlined sequences are 5' overhanging ends compatible with *Bam*HI or *Bgl*II restriction sites.

**Construction of CAT reporter plasmids.** Plasmids pHP and pHE were constructed by cloning the 9.5-kbp *Hind*III-partial *Pst*I and 9-kbp *Hind*III-partial *Eco*RV *Pax-QNR* genomic fragments, respectively, into the *Hind*III-*Pst*I and *Hind*III-blunt-ended *Xba*I sites of the promoterless vectors pBLCAT6 (described in reference 5) and pBLCAT5 (described in reference 37). Construct pPOCAT was obtained by cloning the 1.5-kbp *Hind*III-*Bam*HI *Pax-QNR* promoter fragment, excised from the pXA promoter construct (described in reference 31), into the *Hind*III-*Bam*HI sites of pBLCAT6. pTKCAT corresponds to the pBLCAT5 vector described in reference 5, which contains the chloramphenicol acetyltransferase (CAT) gene under the control of the TK promoter. pPOCAT EP and pTKCAT EP were constructed by cloning the 460-bp *Eco*RV-blunt-ended *Pst*I fragment, in its normal genomic orientation, downstream of the CAT gene, into the *Sma*I sites of pPOCAT and pTKCAT, respectively. Constructs pTKCAT ED and pTKCAT DP were obtained by blunt-end cloning of the 210-bp *Eco*RV-*Dra*I fragment (corresponding to the 5' part of the 460-bp *Eco*RV-*Pst*I fragment) and the 250-bp *Dra*I-*Pst*I fragment (corresponding to the 3' part of the 460-bp *Eco*RV-*Pst*I fragment) to the CAT gene, respectively, into the *Sma*I site of pTKCAT. Construct pTKCAT P, which contains the mouse *Pax-6* enhancer element, was obtained by cloning the blunt-ended 1-kbp *Pst*I mouse genomic fragment into the *Sma*I site of pTKCAT. DF deletion constructs of the 210-bp *Eco*RV-*Dra*I fragment were created by PCR amplification with the oligonucleotide primers listed above. The individual amplified fragments were inserted into the *Sma*I site of pTKCAT. The resulting plasmids are the following: pTKCAT 1234, a 163-bp fragment containing sites DF1 to DF4 (42 to 204 in Fig. 5); oligonucleotides 1F and 5R); pTKCAT 234, a 125-bp fragment containing sites DF2 to DF4 (80 to 204 in Fig. 5; oligonucleotides 2F and 5R); pTKCAT 34, an 88-bp fragment containing sites DF3 and DF4 (117 to 204 in Fig. 5; oligonucleotides 3F and 5R); pTKCAT 123, a 99-bp fragment containing sites DF1 to DF3 (42 to 140 in Fig. 5; oligonucleotides 1F and 3R); and pTKCAT 12, a 54-bp fragment containing sites DF1 and DF2 (42 to 95 in Fig. 5; oligonucleotides 1F and 2R). To obtain the constructs containing the individual DF cloned in one or multiple copies, the complementary oligonucleotides listed above were annealed, dimerized, blunt ended, and inserted 3' to the CAT gene into the blunt-ended *Kpn*I site of the pTKCAT vector. The resulting plasmids are the following: pTKCAT 1, one copy of the 27-bp DF1 fragment (42 to 68 in Fig. 5; oligonucleotides 1F and 1R); pTKCAT (DF2)4, four copies of the 16-bp DF2 fragment (80 to 95 in Fig. 5; oligonucleotides 2F and 2R); pTKCAT (DF3)4, four copies of the 24-bp DF3 fragment (117 to 140 in Fig. 5; oligonucleotides 3F and 3R); pTKCAT 4, pTKCAT (DF4)2, and pTKCAT (DF4)3, one, two, or three copies of the 58-bp DF4 fragment (146 to 204 in Fig. 5; oligonucleotides 4F and 4R). pTKCAT (DF1)3 contains three copies of the 26-bp DF1 fragment (42 to 67) and was constructed by annealing oligonucleotides 3X1F and 3X1R and cloning them into the *Sma*I-*Sac*I sites of pTKCAT. All constructs involving PCR and the cloning of synthetic oligonucleotides were verified by double-stranded dideoxynucleotide sequencing.

**Cell culture and transfection.** QNR, prepared from 5- or 7-day-old embryos, were seeded at one neuroretina per 35-mm dish in Dulbecco modified Eagle

medium-10% fetal calf serum, 24 h prior to transfection. Retinal pigment epithelial cells (RPE), prepared from 7-day-old quail embryos, and QEC were seeded at  $1.5 \times 10^5$  cells per 35-mm-diameter dish in the same conditions. Transfections were performed by the calcium phosphate method. Cells were transfected with 10  $\mu$ g of CAT constructs. Each experiment was performed in duplicate with two different DNA preparations, without or with 0.5  $\mu$ g of plasmid SV-Luc, as an internal control of transfection efficiency. As transfection efficiencies were very similar for all constructs, the results obtained with or without the internal control were identical.

**Luciferase and CAT assays.** After transfection (48 h), the cells were harvested and lysed in the reporter lysis buffer (Promega). One-tenth of the extract was tested for luciferase activity as described by Crepeux et al. (11). CAT assays were performed on the remaining cell extracts as described by Gorman et al. (19).

**RNAse protection analysis.** Total RNA was extracted from quail embryonic neuroretina at day 5 to day 14 of development by the RNazol method (9). A 653-bp *Xho*I-*Asp*718 genomic fragment encompassing exon 0 and cloned into the pGEM3 vector (described in reference 31) was used to synthesize the antisense RNA probe, with T7 polymerase in the presence of [ $\alpha$ - $^{32}$ P]CTP under conditions provided by the riboprobe system (Promega). RNAse protections were performed with 20  $\mu$ g of RNA as described previously (12), and the protected fragments were separated on a 6% sequencing gel. When the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) RNA probe was used, RNAse protections were performed with 2  $\mu$ g of RNA supplemented to 20  $\mu$ g with yeast tRNA.

**Nuclear run-on experiments.** Nuclei were prepared from embryonic day 7 (E7) QNR or RPE and 10-day-old QEC as described by Schibler et al. (34), except that 0.33% (vol/vol) Nonidet P-40 was added to the lysis buffer. The in vitro elongation reaction and labeled RNA extraction were performed as described by Spegele et al. (35), except that the nuclei were labeled for 30 min at 30°C. Labeled RNA ( $10^7$  cpm) from E7 QNR, QEC, or RPE nuclei was hybridized at 42°C for 48 h to prehybridized nitrocellulose filters containing 5  $\mu$ g of test or control denatured plasmid in a solution of 50% formamide, 3 $\times$  SSC (1 $\times$  SSC is 0.15 M NaCl and 0.015 M sodium citrate), 20 mM NaPO<sub>4</sub> (pH 6.5), 1 $\times$  Denhardt's solution, and 100  $\mu$ g of *Escherichia coli* tRNA per ml. After hybridization, the filters were washed for 10 min at 50°C in 0.1 $\times$  SSC, then in 0.3 M NaCl containing 10  $\mu$ g of RNase A per ml, and again in 0.1 $\times$  SSC, dried, and autoradiographed. The blots were quantified after exposure of the filters to a PhosphorImager screen. Three plasmids were used: the vector pGEM3; pExon 0, containing the 653-bp *Xho*I-*Asp*718 genomic fragment encompassing exon 0 cloned into the pGEM3 vector; and pGAPDH, containing a 450-bp fragment of the quail GAPDH cDNA that we isolated from a 15-day-old quail spinal cord cDNA library (kindly supplied by C. Dulac).

**Nuclear protein preparation.** Nuclei were prepared as described above, and nuclear extracts were prepared according to Lavery and Schibler (23) except that these were dialyzed twice for 2 h against DB buffer (20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES; pH 7.9], 60 mM KCl, 0.25 mM EDTA, 0.125 mM ethylene glycol-bis( $\beta$ -aminoethyl ether)-*N,N,N',N'*-tetraacetic acid [EGTA], 20% [vol/vol] glycerol, 1 mM dithiothreitol, and a cocktail of proteinase and phosphatase inhibitors described by Carrière et al. [6]) before being stored at -70°C. Final protein concentrations were determined by the Bradford method (Bio-Rad).

**Electrophoretic mobility shift assay (EMSA).** Double-stranded oligonucleotides were 5' end labeled with T4 polynucleotide kinase and [ $\gamma$ - $^{32}$ P]ATP, 5' overhangs were filled in with the Klenow fragment of DNA polymerase I (Boehringer) and excess deoxynucleotide triphosphates, and the oligonucleotides were purified by electrophoresis on a 20% polyacrylamide gel. Labeled oligonucleotide (1 ng) was incubated on ice for 1 h with 1  $\mu$ g of nuclear extracts in the presence of 5  $\mu$ g of poly(dI-dC) in a final volume of 16  $\mu$ l of 0.5 $\times$  DB buffer. For competition experiments, a 10- or 100-fold molar excess of unlabeled double-stranded oligonucleotides was added simultaneously with the probes. The DNA-protein complexes were analyzed by electrophoresis on a 6% polyacrylamide gel in 0.25 $\times$  Tris-borate-EDTA at room temperature for 2 h at 180 V, and the gel was dried for autoradiography.

**Southwestern analysis.** Southwestern (DNA-protein) blotting was performed with 100  $\mu$ g of E7 QNR, QEC, or RPE crude nuclear extracts electrophoresed through a sodium dodecyl sulfate (SDS)-polyacrylamide gel. Proteins were electrophoretically transferred to an Immobilon membrane (Millipore) for 5 h at room temperature in the transfer buffer (0.025 M Tris-HCl, 0.192 M glycine, 10% [vol/vol] methanol). After transfer, the membrane was treated as described by Leprince et al. (24), except that the double-stranded probes were obtained as described above and were used at  $10^5$  dpm per ml.

**DNase I footprinting.** The footprinting probes were prepared by end labeling of DNA restriction fragments with the Klenow fragment of DNA polymerase I (Boehringer) in the presence of [ $\alpha$ - $^{32}$ P]dCTP and [ $\alpha$ - $^{32}$ P]dATP (50  $\mu$ Ci each) and purified, after secondary restriction, by electrophoresis on a 5% polyacrylamide gel. End-labeled DNA fragment (5 ng) was incubated on ice for 1 h with 30 or 100  $\mu$ g of nuclear proteins in the presence of 5  $\mu$ g of poly(dI-dC) in 50  $\mu$ l of 0.5 $\times$  DB buffer supplemented with 5 mM MgCl<sub>2</sub> and 2.5 mM CaCl<sub>2</sub>. After incubation, digestion by 20 ng of DNase I (Bethesda Research Laboratories) was performed for 30 s at 20°C. The reaction was stopped by addition of 30  $\mu$ l of 50 mM EDTA and 0.5 mg of proteinase K per ml, phenol extracted, and ethanol precipitated. The recovered DNA was then loaded on an 8% sequencing gel.

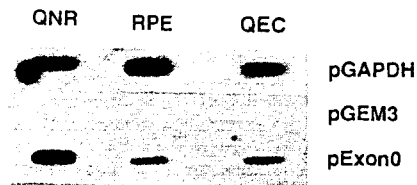


FIG. 1. Nuclear run-on experiments. Shown is an autoradiograph of the hybridization of in vitro transcripts from embryonic E7 QNR (left panel), RPE (middle panel), and QEC (right panel) nuclei. pExon0 contains the 653-bp *XhoI-Asp* 718 *Pax-QNR* gene fragment (from -526 to +127 [31]) subcloned into the pGEM3 vector. pGAPDH, containing part of the quail GAPDH cDNA, was included as a positive control, while pGEM3 was used as control for nonspecific hybridization signal.

**Nucleotide sequence accession number.** The nucleotide sequence data of the quail enhancer are included in the sequence appearing in the EMBL nucleotide sequence database under accession number X68168. The nucleotide sequence of the *Mus musculus* enhancer appears in the EMBL nucleotide sequence database under accession number X82150.

## RESULTS

**Nuclear run-on experiments.** In order to determine whether neuroretina-specific expression of *Pax-QNR* P0-initiated mRNAs is due to transcriptional control, in vitro nuclear run-on assays were performed. We compared embryonic day 7 QNR with QEC and embryonic RPE; cells of the latter two types do not express these mRNAs. Thus, nuclei isolated from these cells were incubated with nucleoside triphosphates in the presence of [<sup>32</sup>P]UTP, and the nuclear RNA was extracted. An equivalent amount of labeled nuclear RNA from each cell type was hybridized to filters containing the first exon of the *Pax-QNR* gene (exon 0, [13]) cloned in the plasmid vector pGEM3 (pExon0). The controls used included plasmid vector pGEM3 and a plasmid containing a fragment of the quail GAPDH cDNA. The results (Fig. 1) show a higher level of transcripts of the *Pax-QNR* gene in E7 QNR nuclei than in QEC or RPE nuclei, whereas transcription of the GAPDH gene was similar in all three cells. No significant hybridization to a probe upstream of the transcription initiation site (data not shown) was found. Therefore, these data demonstrate that transcriptional regulation takes place in the neuroretina-specific expression of MC29-QNR2 mRNA.

**Identification of a neuroretina-specific 3' enhancer within the *Pax-QNR* gene.** Previous studies have shown that the -1386 to +127 promoter sequences function relatively equally in both QNR and QEC (31). Thus, additional upstream and/or downstream regulatory elements might be responsible for up-regulating the transcription of the *Pax-QNR* gene in neuroretina cells. To locate potential *cis* control regions regulating the neuroretina-specific transcription of the P0-initiated mRNAs, we tested 5'- and 3'-flanking promoter DNAs for their ability to enhance transcription from the promoter in neuroretina cells. Among the 3'-flanking promoter DNA constructs tested (data not shown), two retained our attention: the pHP and pHE constructs (shown in Fig. 2A), which contain about 8 and 7.5 kbp, respectively, of 3'-flanking promoter sequences upstream of the CAT gene. After transient transfection of these constructs into E7 QNR cells, pHP showed CAT activity (about 8-fold) higher than that observed for the deletion construct pHE (Fig. 2A). This suggests that the 460-bp *EcoRV-PstI* fragment, located 7.5 to 8 kbp 3' to the *Pax-QNR* P0 promoter and deleted in the pHE construct, contains DNA sequences that may increase expression of the reporter gene.

To determine whether this region contains a neuroretina-specific transcriptional enhancer element, the 460-bp *EcoRV-*

*PstI* fragment was subcloned, in its normal genomic orientation, 3' to the CAT gene in the pP0CAT and pTKCAT vectors, in which the CAT reporter gene is driven by the *Pax-QNR* P0 promoter (31) or the minimal TK promoter (vector pBLCAT5 [5]), respectively (Fig. 2B). The resulting constructs (pP0CAT EP and pTKCAT EP; Fig. 2B) and their parental plasmids pP0CAT and pTKCAT were transiently transfected into E7 QNR, QEC, and RPE and assayed for CAT activity. In this case, the CAT activities of the chimeric CAT constructs were expressed relative to the values for the parental plasmids. The results show (Fig. 2B) that in E7 QNR, the 460-bp *EcoRV-PstI* fragment enhanced CAT activity by the *Pax-QNR* P0 and TK promoters by about 6-fold above the levels of the vectors without enhancer. However, in QEC and RPE, which do not express the P0-initiated mRNAs, the *EcoRV-PstI* fragment failed to enhance CAT activity by the *Pax-QNR* P0 or TK promoters above basal levels (Fig. 2B). The enhancer activity of the 460-bp *EcoRV-PstI* fragment was observed to be slightly lower (about twofold) when assayed in the opposite orientation (data not shown). No enhancer activity was detectable when this fragment was inserted 3' to the CAT gene in the promoterless pBLCAT6 vector (data not shown). These results demonstrate that the 460-bp *EcoRV-PstI* fragment contains a functional neuroretina-specific enhancer element. The activity of this 3' enhancer is not restricted to the *Pax-QNR* P0 promoter and is orientation independent.

In order to characterize more precisely the enhancer element, deletions from the 5' and 3' ends of the *EcoRV-PstI* enhancer fragment were generated and tested for enhancer activity after transient transfection into E7 QNR. The results show (Fig. 2C) that the 5' portion of the *EcoRV-PstI* fragment (construct pTKCAT ED containing the 210-bp *EcoRV-DraI* fragment) retained the full enhancer activity of the *EcoRV-PstI* fragment (compare constructs pTKCAT ED and pTKCAT EP). In contrast, the 3' portion of the *EcoRV-PstI* fragment (construct pTKCAT DP containing the 250-bp *DraI-PstI* fragment) had no enhancer activity (compare pTKCAT DP and pTKCAT). Thus, the enhancer element is located more precisely in a 210-bp *EcoRV-DraI* fragment.

**The neuroretina-specific 3' enhancer functions in a developmental stage-dependent manner.** We have previously shown that the level of *Pax-QNR* mRNAs, monitored by Northern (RNA) blot analysis, varied during the development of the embryonic avian neuroretina, increasing from day 5 to day 8, when it reaches a plateau (27). Since the *Pax-QNR* gene expresses multiple mRNAs differing by their 5' untranslated region (13), we examined the pattern of expression of the P0-initiated mRNAs during the embryonic quail neuroretina development by RNase protection assays. An antisense labeled RNA probe encompassing exon 0 (described in reference 31 and used to determine the P0 transcriptional start site) was hybridized to total RNAs extracted from embryonic quail neuroretina at day 5 (E5) until day 14 (E14) of development. Results show (Fig. 3A) that the level of P0-initiated mRNAs (indicated by a bracket) was very low in E5 neuroretina, increased markedly from E5 until E8, and remained constant until E14. As a control, the levels of GAPDH mRNA were monitored and were similar among these samples (Fig. 3B). These data are reminiscent of those obtained previously by Northern blot analysis (27).

The low level of P0-initiated mRNAs detected in E5 neuroretina, followed by a marked increase until E8, prompted us to investigate whether the enhancer activity varies with development. Thus, the enhancer constructs described in Fig. 2B and their parental plasmids were transiently transfected into E5 and E7 QNR and assayed for CAT activity. The CAT activities

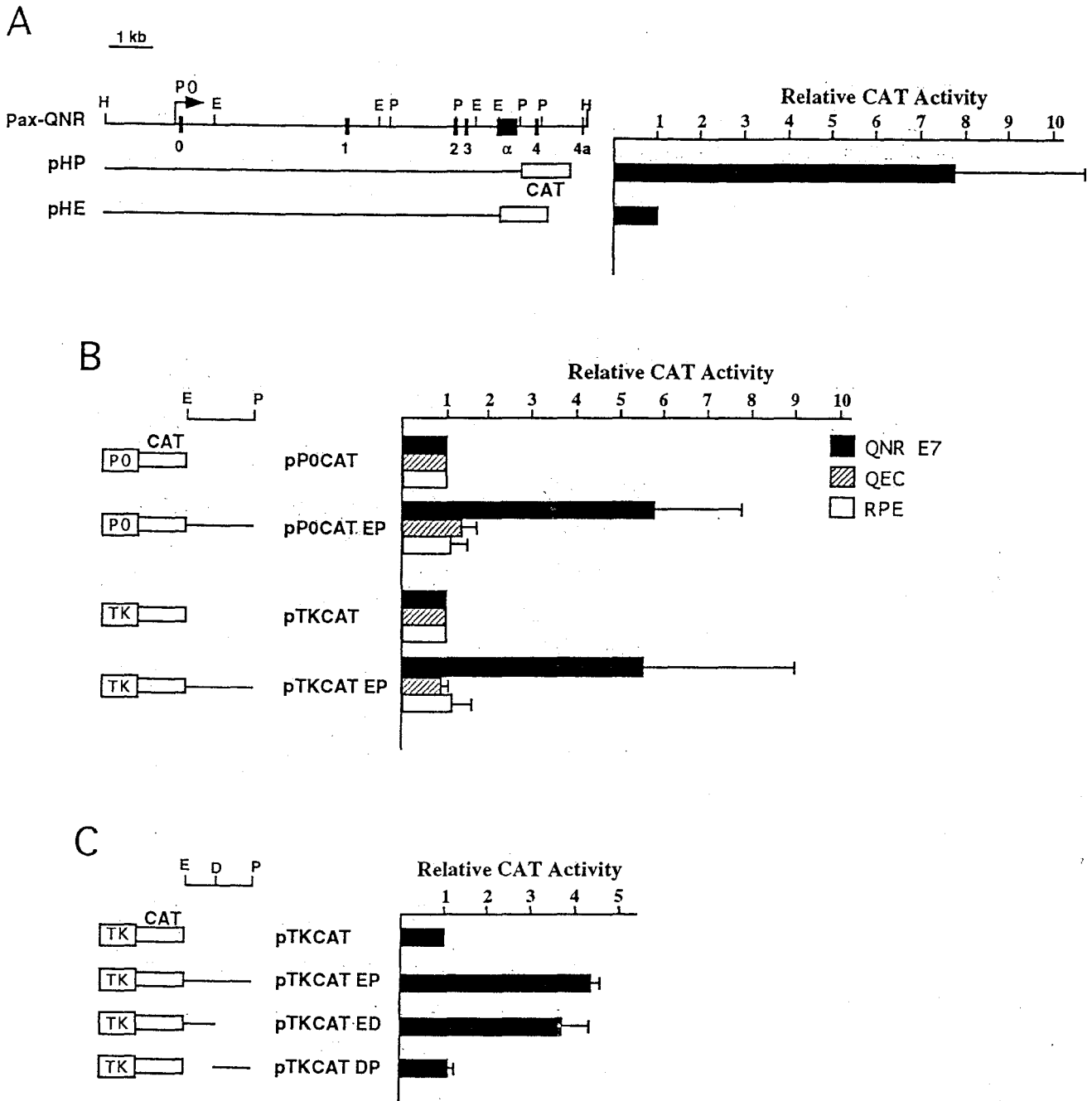


FIG. 2. Identification of a neuroretina-specific 3' enhancer element within the *Pax-QNR* gene. (A) Test of *Pax-QNR* 3' regions for the presence of *cis* regulatory regions. The top line shows a partial map of the *Pax-QNR* gene with the exons boxed. The positions of the *Hind*III (H), *Pst*I (P), and *Eco*RV (E) restriction enzyme sites used to generate the CAT constructs are marked. pHP and pHE CAT constructs contain the 9.5-kbp partial *Hind*III-*Pst*I and the 9-kbp partial *Hind*III-*Eco*RV fragments, respectively, cloned upstream of the CAT gene. These constructs were transiently transfected into E7 QNR cells, and levels of CAT activity were quantitated after exposure of the thin-layer chromatograms to a PhosphorImager screen. CAT activity is expressed relative to that of the pHE construct, chosen as a reference and arbitrarily set at 1. (B) The *Pax-QNR* 3' region contains a neuroretina-specific enhancer functional on the *Pax-QNR* P0 or TK promoter. pPOCAT EP and pTKCAT EP contain the 460-bp *Eco*RV-*Pst*I fragment inserted in the natural orientation, 3' to the CAT gene in plasmids pPOCAT and pTKCAT, respectively. These plasmids were transiently transfected into E7 QNR, QEC, and RPE, and CAT activities were measured with an Instant Imager (Packard). CAT activities of the enhancer constructs are expressed relative to those of the parental plasmids. (C) Delineation of the minimum enhancer element. The position of the *Dra*I (D) restriction enzyme site used to generate the deletion constructs is marked. pTKCAT ED and pTKCAT DP contain the 210-bp *Eco*RV-*Dra*I (5' portion of the *Eco*RV-*Pst*I sequence) and the 250-bp *Dra*I-*Pst*I (5' portion of the *Eco*RV-*Pst*I sequence) fragments, respectively, cloned in the natural orientation 3' to the CAT gene in plasmid pTKCAT. After transient transfection of these constructs into E7 QNR cells, CAT activities were quantified as for panel A. CAT activities of the constructs are expressed relative to that of the pTKCAT plasmid. The results shown in all three panels are the averages of at least two independent transfection experiments performed in duplicate with two different DNA preparations, with the standard deviations indicated by error bars.

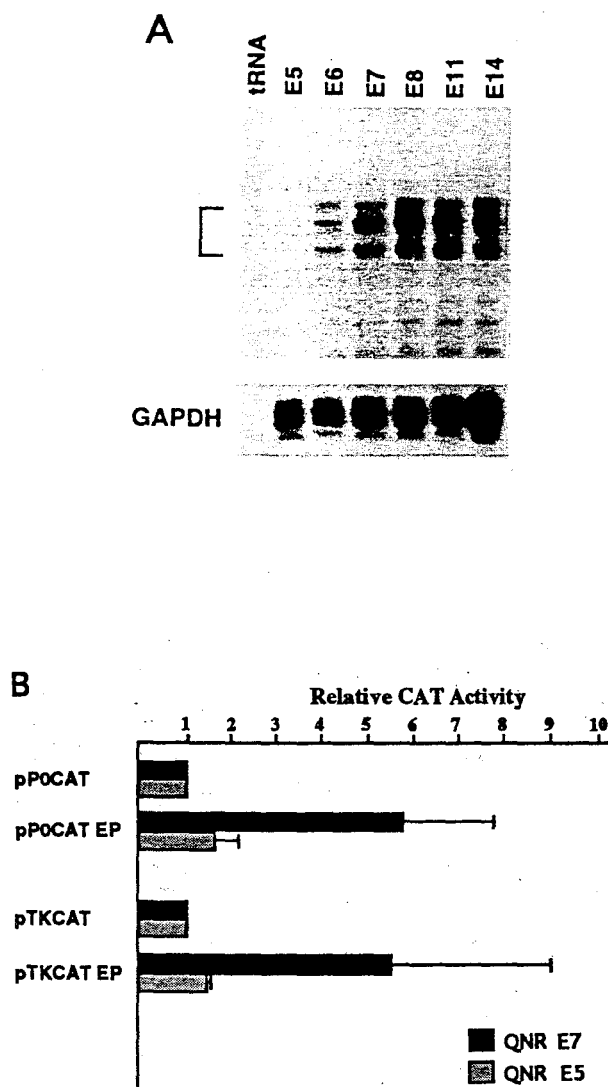


FIG. 3. Developmental analysis of the P0-initiated mRNAs and of the 3' enhancer activity. (A) RNase protection analysis. Total RNA (20  $\mu$ g) from QNR at stages E5 to E14 was hybridized with an antisense RNA probe encompassing exon 0 (described in reference 31). Following RNase digestion, the samples were analyzed on a sequencing gel. The *Pax-QNR* mRNAs initiated at the P0 promoter and protected by the probe are indicated by a bracket. The probe (top of the gel) is not shown. tRNA was included as a negative control. As a control, hybridization of the mRNAs with an antisense GAPDH RNA probe is shown (bottom panel). (B) Functional analysis of the 3' enhancer in neuroretina at different developmental stages. The CAT constructs were transiently transfected into E5 and E7 QNR, and CAT activities were measured with an Instant Imager (Packard). CAT activities of the enhancer constructs are expressed relative to those of the parental plasmids. The results shown are the averages of three independent transfection experiments performed in duplicate with two different DNA preparations, with the standard deviations indicated by error bars.

of the enhancer CAT constructs were expressed relative to the values for the parental plasmids. The results show (Fig. 3B) that, in contrast to E7 QNR in E5 QNR the enhancer element increased CAT activity by the *Pax-QNR* P0 or TK promoters very weakly above basal levels. These data indicate that the activity of the 3' enhancer in embryonic neuroretina is developmentally controlled and suggest that its increase between E5 and E7 could be responsible for the variation in the level of P0-initiated mRNAs observed between these two stages.

**Footprinting studies.** The neuroretina-specific 3' enhancer was located to a 210-bp *EcoRV-DraI* fragment (Fig. 2C). In

DNase I footprinting experiments, we investigated whether this sequence could be specifically recognized *in vitro* by *trans*-acting factors. The *EcoRV-PstI* fragment, end labeled on the noncoding strand, was incubated with either E7 QNR, QEC, or RPE nuclear extracts and subjected to moderate DNase I digestion. With E7 QNR nuclear extracts, four protected regions, named DNase I footprints (DF) DF1, DF2, DF3, and DF4, were observed in the *EcoRV-DraI* fragment (Fig. 4). The footprints were faint when 30  $\mu$ g of nuclear extracts was used (lanes 1) but were much more visible when 100  $\mu$ g of nuclear extracts was used (lanes 2). The 5' border of DF1 is not clearly visible in Fig. 4; however, this footprint was much more delineated in other experiments (data not shown). Footprint locations and sequences are shown in Fig. 5. With QEC and RPE nuclear extracts, the four footprints observed with E7 QNR nuclear extracts were not found (Fig. 4B). It is noteworthy that with QEC nuclear extracts a protected region extending 5' of and partially overlapping DF4 was observed (Fig. 4B), suggesting that a different factor binds this region. A footprinting experiment performed with a probe overlapping the *EcoRV* site failed to reveal any further protected region between DF1 and this site (data not shown). These results demonstrate the presence of four neuroretina-specific nuclear protein-binding sites in the 3' enhancer element.

**Deletional analysis of the neuroretina-specific enhancer.** To test directly the role of these protein-binding sequences in the neuroretina-specific enhancer activity, we constructed DF deletion mutants of the 210-bp *EcoRV-DraI* fragment by PCR amplification (summarized in Fig. 6A). The constructs were transiently transfected into E7 QNR cells and assayed for enhancer function. In this case, the enhancer activities of the constructs were expressed relative to the value for the full-size enhancer construct pTKCAT ED (set at a value of 100). Results show (Fig. 6A) that a 163-bp fragment extending through the four footprints, construct pTKCAT 1234, retained the full enhancer activity. Deletion of site DF1 (construct pTKCAT 234) or DF4 (construct pTKCAT 123) strongly reduced the enhancer activity. However, a single copy of either DF1 (construct pTKCAT 1) or DF4 (pTKCAT 4) mediated a very weak enhancer activity, if any. Additional deletion of DF2 (pTKCAT 34 and pTKCAT 1) or DF3 (pTKCAT 1) or DF3 (pTKCAT 12 and pTKCAT 4) also decreased the enhancer activity, suggesting that these binding sites contribute to the enhancer activity. We cloned multiple copies of the individual DF sites, 3' to the CAT gene in the pTKCAT vector (summarized in Fig. 6B). After transient transfection in E7 QNR, the CAT activities of the constructs were expressed relative to the value for the pTKCAT vector. Results show (Fig. 6B) that when multiple copies of either the DF1, DF2, or DF3 site were fused with the TKCAT [pTKCAT (DF1)3, pTKCAT (DF2)4, and pTKCAT (DF3)4], a weak enhancement, if any, was observed. In contrast, whereas a single copy of the DF4 site (pTKCAT 1) had no enhancer activity, multiple copies of the DF4 site [pTKCAT (DF4)2 and pTKCAT (DF4)3] mediated enhancer activity, which increased with the number of the copies cloned. Thus, multimerization of the DF4 site mimics the enhancer activity, suggesting that, in this context, DF4 binds *trans*-acting factors able to mediate the enhancer activity. Since, in the natural enhancer element, maximal enhancing occurs when the four binding sites are present, we conclude that the DF sites cooperate to mediate full enhancer activity.

**Characterization of nuclear protein binding to individual DF within the neuroretina enhancer.** To further characterize individual protein-binding sites within the neuroretina-specific enhancer, EMSA were performed with double-stranded labeled oligonucleotides corresponding to DF1, DF2, DF3, and

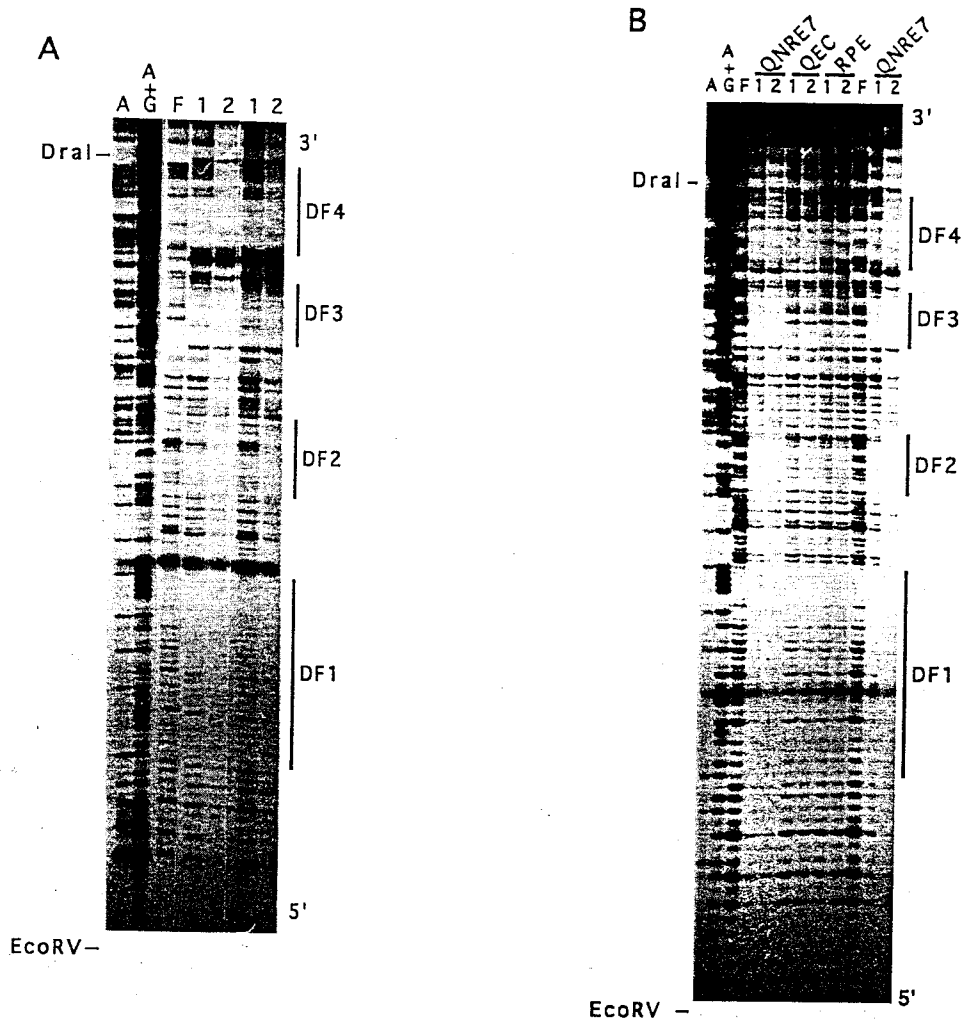


FIG. 4. Footprinting studies. (A) DNase I footprinting analysis of the *EcoRV-PstI* fragment in the absence (F) or presence of 30  $\mu$ g (lanes 1) or 100  $\mu$ g (lanes 2) of E7 QNR nuclear extracts. (B) DNase I footprinting analysis of the *EcoRV-PstI* fragment in the absence (F) or presence of 30  $\mu$ g (lanes 1) or 100  $\mu$ g (lanes 2) of nuclear extracts from E7 QNR, QEC, or RPE. For both panels, the positions of the *EcoRV* and *DraI* restriction sites, delineating the 5' and 3' borders, respectively, of the enhancer element, are indicated. The positions of the DF referred to in the text are also marked. Lanes A and A+G, sequencing reactions on the probe.

DF4 (defined in Fig. 5; also see Materials and Methods) and with nuclear extracts from E7 and E5 QNR, as well as from QEC and RPE (Fig. 7A). When the oligonucleotides were incubated with E7 QNR nuclear extracts, multiple complexes were formed (Fig. 7A, lanes 1, and Fig. 8, lanes 2), some of which were discrete (indicated by small arrows). The competition experiments presented in Fig. 8 show that these complexes were specific, since their formations were inhibited or strongly reduced by the presence of a 100-fold excess of the corresponding unlabeled oligonucleotide (for probe DF1, see Fig. 8A, lane 4; for probe DF2, see Fig. 8B, lane 6; for probe DF3, see Fig. 8C, lane 8; and for probe DF4, see Fig. 8D, lane 10), but not by the presence of a similar excess of a nonspecific competitor corresponding to the Sp1 consensus binding site (Fig. 8, lanes 12). These complexes were also detected, albeit at lower levels, with E5 QNR nuclear extracts (Fig. 7A, lanes 4). This might suggest that the factors binding to the oligonucleotides are less abundant in neuroretina at day E5 of development or that the affinity of these factors to their sequences is lower. These results might explain the differential enhancer activity observed between E5 QNR and E7 QNR (Fig. 3B). These complexes were not detected with the QEC or RPE

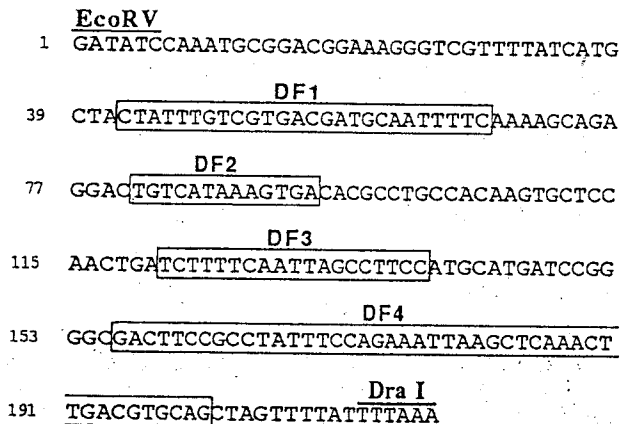


FIG. 5. Nucleotide sequence of the 3' enhancer element. The sequence of the 216-bp *EcoRV-DraI* fragment is given, with the four neuroretina-specific protein-binding sites (DF) boxed. Restriction sites used for subcloning are indicated. Nucleotide numbers are shown on the left.

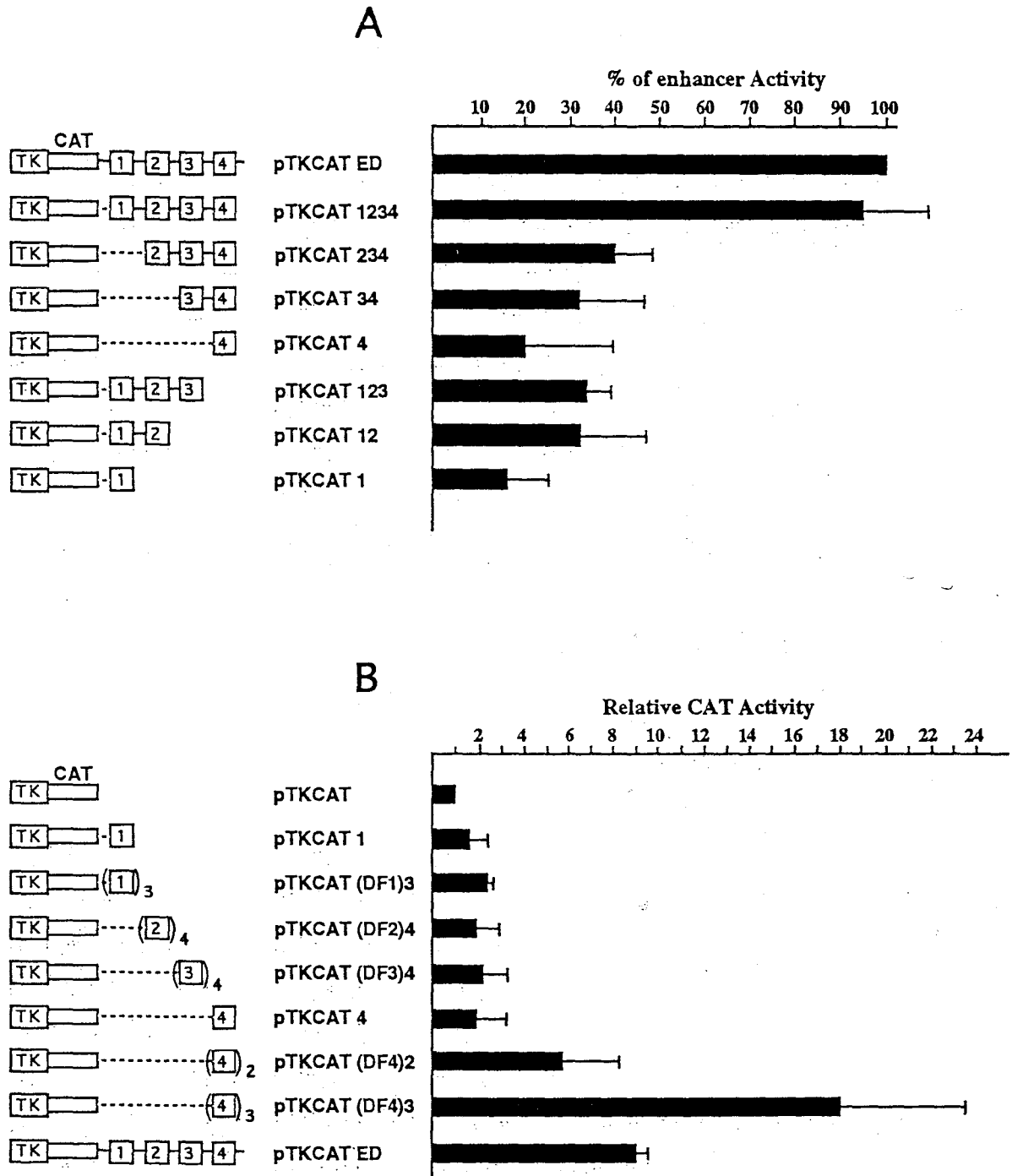


FIG. 6. Role of the four neuroretina-specific protein-binding sequences in the enhancer activity. (A) Deletional analysis of the neuroretina-specific enhancer. The DF deletion constructs, listed in the center of the panel, were transiently transfected into E7 QNR cells, and CAT activities were measured as for Fig. 2A. The CAT activities of the DF deletion constructs were expressed relative to that of the full-size enhancer construct pTKCAT ED (set at a value of 100). (B) Biological activity of each protein-binding site. Multiple copies of either DF1, DF2, DF3, or DF4 were cloned, 3' to the CAT gene, in the pTKCAT vector. The resulting constructs, listed in the center of the panel, were transiently transfected into E7 QNR cells, and CAT activities were measured as for Fig. 2A. CAT activities of the constructs are expressed relative to that of the parental plasmid pTKCAT. The results shown in both panels are the averages of at least three independent transfection experiments performed in triplicate with three different DNA preparations, with the standard deviations indicated by error bars.

nuclear extracts (except with the DF2 probe, in which faint complexes were observed with RPE nuclear extracts; see Fig. 7A). Incubation of QEC or RPE nuclear extracts with oligonucleotides containing consensus binding sites for Ap1 or upstream stimulatory factor indicated that these nuclear extracts

were fully functional (data not shown). These results are consistent with the footprinting experiments and confirm the binding of neuroretina-specific factors to the enhancer element. To identify the nuclear proteins that interact with the neuroretina-specific enhancer, Southwestern analyses were per-

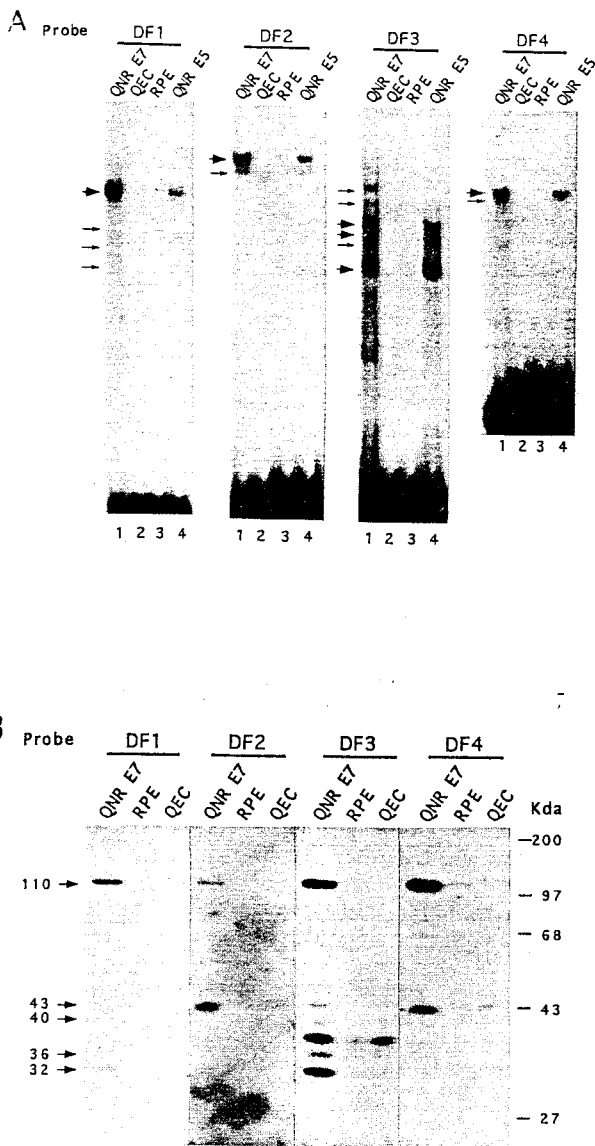


FIG. 7. Characterization of nuclear protein binding to individual DF within the neuroretina enhancer. (A) EMSA analysis of the four protein-binding sites. EMSAs were performed with the radiolabeled DF1, DF2, DF3, and DF4 oligonucleotides, and 1  $\mu$ g of nuclear extracts from E7 QNR (lanes 1), QEC (lanes 2), RPE (lanes 3), or E5 QNR (lanes 4). The arrows indicate bands of altered mobility representing specific protein-DNA complexes. Big arrows, major complexes; small arrows, minor complexes. (B) Identification of nuclear proteins interacting with the four DF by Southwestern analysis. Nuclear extracts (100  $\mu$ g) from E7 QNR, QEC, or RPE were subjected to SDS-PAGE. After protein transfer, the labeled DF1, DF2, DF3, and DF4 oligonucleotides were incubated with the renatured nuclear proteins. The potential proteins interacting with the probes are indicated by arrows.

formed. E7 QNR, QEC, or RPE nuclear proteins were fractionated by SDS-polyacrylamide gel electrophoresis (PAGE), immobilized onto nylon membrane, and probed with equal radiospecific activity probes of DF1, DF2, DF3, and DF4. Results show (Fig. 7B) that the four probes bound multiple neuroretina nuclear proteins, albeit with different affinities (if we compare the differences in the intensities of the multiple proteins detected with one probe considered). These results are consistent with the EMSA data showing multiple complexes formed when these probes were incubated with QNR nuclear extracts (Fig. 7A). Two proteins, of 110 and 43 kDa,

were bound by all four nucleotides, whereas DF1 and DF3 bound additional proteins of equal molecular masses. These proteins were essentially detected in neuroretina and very faintly in RPE (see Discussion for explanation) and could correspond to the proteins bound to the enhancer element. These results suggest that the four nuclear protein-binding sites detected in the neuroretina-specific enhancer bind a set of proteins, some of which are identical.

To investigate whether identical proteins are present in the complexes detected by EMSAs with probe DF1, DF2, DF3, or DF4 and E7 QNR nuclear extracts (Fig. 7A), we performed competition studies. Competition results are presented in Fig. 8. The major complex formed with the DF1 probe (Fig. 8A, lane 2) was unexpectedly increased by a 100-fold molar excess of either DF2 (lane 6) or DF3 (lane 8), whereas it was partially inhibited by a similar excess of DF4 (lane 10); however, the minor complexes formed with this probe were inhibited by an excess of either DF2, DF3, or DF4. The two complexes formed with the DF2 probe (Fig. 8B, lane 2) were inhibited totally by a 100-fold molar excess of DF4 (lane 10) but only partially by a similar excess of either DF1 (lane 4) or DF3 (lane 8). Likewise, the complexes obtained with the DF3 probe (Fig. 8C, lane 2) were partially inhibited by a 100-fold molar excess of either DF1 (lane 4), DF2 (lane 6), or DF4 (lane 10). Finally, only the minor complex formed with the DF4 probe (Fig. 8D, lane 2) was inhibited by a 100-fold molar excess of either DF1 (lane 4), DF2 (lane 6), or DF3 (lane 8). The specificity of these competition experiments was demonstrated by the fact that the Sp1 oligonucleotide failed to compete with DF1, DF2, DF3, or DF4 for binding (Fig. 8, lanes 11 and 12). The competition results demonstrate that the four DF oligonucleotides are able to modulate the binding among them. These data are consistent with the Southwestern blot results and indicate that some protein-DNA complexes visualized by EMSAs involved a set of identical neuroretina factors. However, these common factors are expected to bind to the four DF with different affinities, as suggested by the partial or total competition obtained.

Our gel retardation assays revealed a complex pattern of protein-DNA interactions involving different as well as common factors, prompting us to compare the sequences of the four neuroretina protein-binding sites. Results show (Fig. 9) that these AT-rich oligonucleotides exhibited some stretches of homologies, but these do not correspond to already described protein-binding sites. However, as pointed out by a referee, octamer-like and CREB-like proteins (binding sites ATGCAAAT and TGACGTA/CA/G, respectively) may be involved.

**Nucleotide sequence and enhancer activity of the 3'-flanking homologous region of the mouse *Pax-6* gene.** To help determine whether the 3' quail enhancer is phylogenetically conserved, we isolated a mouse *Pax-6* clone from a genomic library using the quail *EcoRV-PstI* fragment as a probe. The sequence of the mouse homologous region was determined and compared with the sequence of the quail *Pax-6* enhancer determined above (Fig. 5). The two noncoding regions exhibited very high nucleotide sequence identity (89.8%) (Fig. 10A), with the four protein-binding sites identified previously on the quail *Pax-6* enhancer being extremely well conserved: DF3 and DF4 exhibited 100% nucleotide identity, whereas DF1 and DF2 exhibited 88.5 and 92.3% nucleotide identities, respectively. It is striking that of the 22-bp differences between the quail and the mouse sequences, most of them (16) are outside the four binding sites described in the quail *Pax-6* enhancer, suggesting a functional role for these motifs. These results suggest that there could be an active enhancer within the mouse *Pax-6* gene.



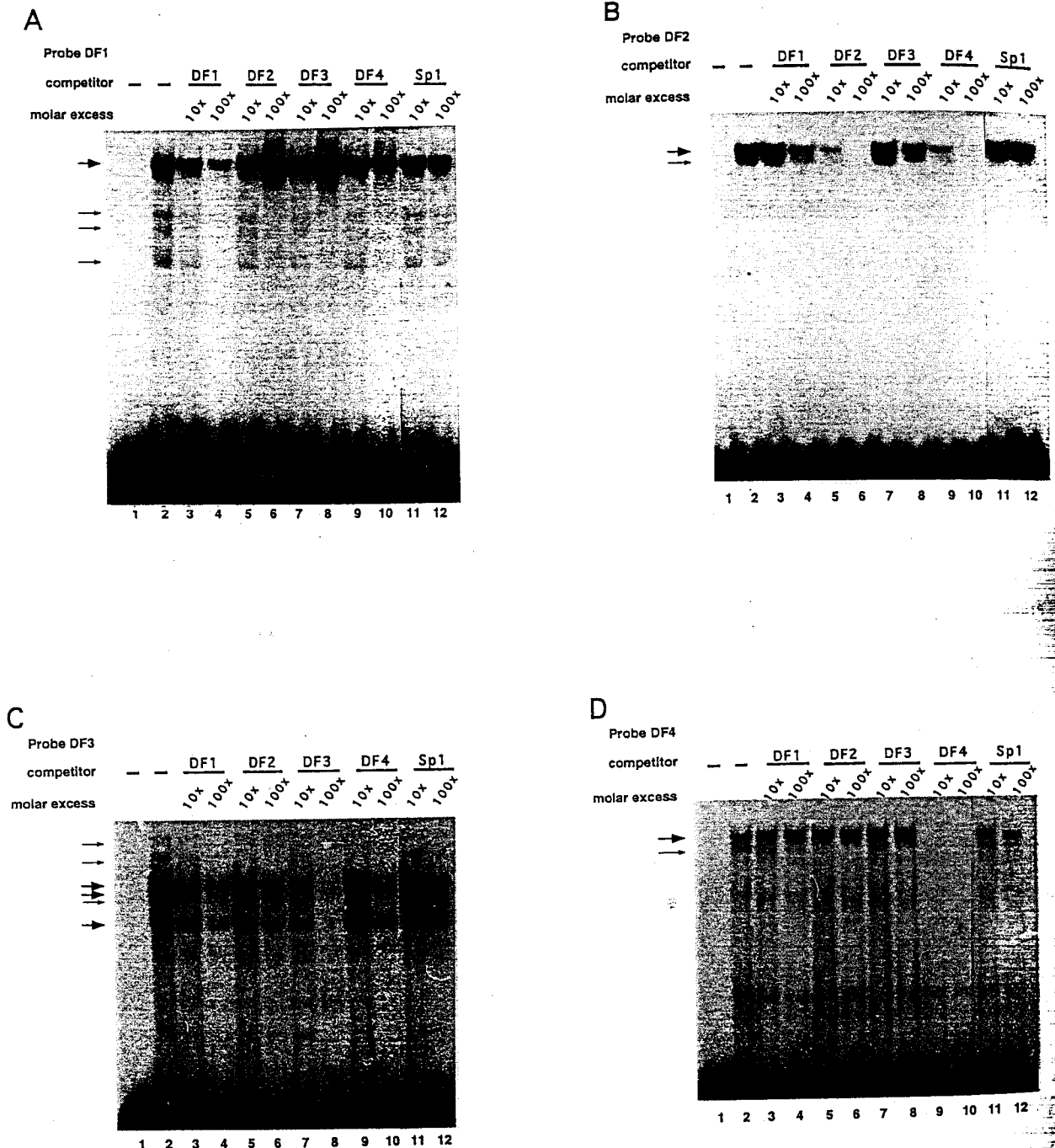


FIG. 8. EMSA analysis of the four neuroretina protein-binding sites. Labeled oligonucleotides DF1 (A), DF2 (B), DF3 (C), and DF4 (D) were incubated alone (lanes 1) or with 1  $\mu$ g of E7 QNR nuclear extracts (lanes 2 to 12) in the absence (lanes 2) or in the presence (lanes 3 to 12) of the indicated (fold) molar excess of various unlabeled competitors: DF1 (lanes 3 and 4), DF2 (lanes 5 and 6), DF3 (lanes 7 and 8), DF4 (lanes 9 and 10), and Sp1 (lanes 11 and 12).

To test this hypothesis, a 1-kbp *Pst*I mouse fragment containing the putative enhancer was inserted 3' to the CAT gene in the pTKCAT vector, yielding pTKCAT P. This construct was transiently transfected into E7 QNR and tested for enhancer activity. As a positive control, we used the quail homologous pTKCAT ED construct. Results show (Fig. 10B) that the 1-kbp *Pst*I mouse fragment (construct pTKCAT P) exhibited an enhancer activity, albeit 2-fold lower than that produced by the quail *Pax-6* enhancer (construct pTKCAT ED).

The ability of the mouse fragment to enhance CAT activity from the TK promoter in QNR demonstrates that this region contains a functional enhancer element.

#### DISCUSSION

In this study, we have demonstrated, by nuclear run-off experiments, that neuroretina-specific expression of *Pax-6* P0-initiated mRNAs is, in part, the result of transcript



FIG. 9. Nucleotide sequence comparison of the four protein-binding sites DF1, DF2, DF3, and DF4. The stretches of homologies are boxed. Gaps, indicated by dots, were created to maximize the alignment of conserved sequences.

control, and we have identified a neuroretina-specific 3' enhancer within the *Pax-QNR* gene, 7.5 kbp downstream from the P0 promoter. This neuroretina-specific enhancer exerts an enhancer activity on the homologous P0 *Pax-QNR* or heterologous TK promoters in a position- and orientation-independent manner. Moreover, the activity of the 3' enhancer in embryonic neuroretina is developmentally controlled and parallels the level of expression of P0-initiated mRNAs. Thus, the increase in the level of P0-initiated mRNAs during the embryonic neuroretina development could be attributed, in part, to the activity of this enhancer.

We have found, both in a DNase I footprinting assay and in an EMSA, that the enhancer element contains four neuroretina-specific nuclear protein-binding sites, and we also demonstrated a functional role for these binding sites in a reporter-based expression assay. The protein-DNA complexes visualized in EMSA are detected at lower levels with E5 QNR nuclear extracts, and this is correlated with a lower enhancer activity at this stage. This might suggest that the factors binding to the enhancer are less abundant in neuroretina at day E5 of development or that the affinity of these factors to their sequences is lower as a result of posttranslational modification, for example. On the basis of dissection of the enhancer region, we propose that the enhancer activity is mediated by a synergy between the four protein-binding sites. Thus, the factors bound to these sites must cooperate with each other in order to activate the neuroretina enhancer. The picture of a tissue-specific enhancer made of several interacting enhansons agrees with the modular nature of the enhancers described so far and reveals similarities with the control elements regulating the tissue-specific expression of other genes (20, 22, 28). Southwestern analysis and EMSA competition experiments demonstrate that the four sites bind a set of unknown identical proteins. Although these proteins are detected essentially in the neuroretina, they are also faintly detected in RPE. One explanation could be the fact that the retinal pigment epithelium possesses the capacity to transdifferentiate into neuroretina *in vivo* (10) and to generate cells with neuronal traits in certain culture conditions (29, 32; for a review, see reference 15).

However, several questions remain unanswered. Why do the multimerized DF1, DF2, and DF3 sites have weak or no enhancer activity, although they bind factors some of which are identical to those bound by the active multimerized DF4 site? We might suggest different hypotheses. Since EMSAs reveal different protein-DNA complexes between the four DF, perhaps additional proteins (not detectable by Southwestern blotting) are present in the complexes observed with the DF4 fragment and are important to *in vivo* enhancer activity. That additional proteins are present in the DF4 complexes is strongly suggested by the fact that in EMSA competition experiments only the minor complex is inhibited by unlabeled

DF1, DF2, or DF3 fragments. Alternatively, the geometric spacing of the *cis* elements is important for the enhancer activity and this is not respected in the multimerized DF1, DF2, or DF3 fragments. Characterization of the factors involved should provide an answer to these questions. The nucleotide sequences of the four binding sites suggest the possibility that the DNA-binding proteins involved recognize AT-rich boxes and therefore belong to the homeodomain-containing transcription factors (for a review, see reference 14). Such proteins

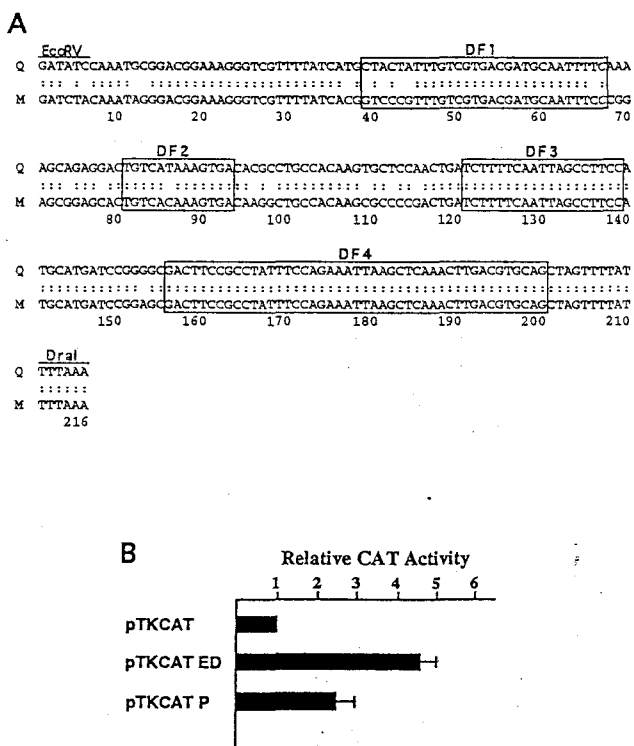


FIG. 10. Nucleotide sequence and enhancer activities of the 3' flanking homologous region of the mouse *Pax-6* gene. (A) Nucleotide sequence homology between the quail *Pax-6* enhancer (Q) and the homologous region in the mouse *Pax-6* gene (M). Double dots indicate identity of bases at the positions shown. The four protein-binding sequences observed previously on the quail *Pax-6* enhancer are boxed. (B) Enhancer activities of the 3' flanking homologous region of the mouse *Pax-6* gene. The pTKCAT P and pTKCAT ED constructs were transiently transfected into E7 QNR, and CAT activities were measured as for Fig. 2A. CAT activities of the constructs are expressed relative to that of the parental plasmid pTKCAT. The results shown are the averages of three independent transfection experiments performed in duplicate with two different DNA preparations, with the standard deviations indicated by error bars.

are effectively expressed in the eye (3, 26) and might be involved in *Pax-6* gene regulation.

Nucleotide sequence identity between the quail enhancer and the homologous murine region is approximately 90%, and the four protein-binding sites are also found in the murine sequence. Like the quail sequence, the murine homologous region also exhibits an enhancer activity when tested in transient transfection assays into neuroretina cells. However the mouse enhancer activity is 2-fold lower than that of the quail, despite the great degree of identity of the four protein-binding sites. One explanation could be point mutations in important sequences: the DF1 and DF2 sites differ by 5 and 1 bp from their quail homologs, respectively. Alternatively, there could be an inhibitory sequence, outside the enhancer region, within the 1-kbp mouse genomic fragment cloned in pTKCAT.

Taken together, our results indicate that the neuroretina enhancer described in this study is conserved in mammalian species and is a complex transcriptional regulatory element that potentially contributes to the neuroretina specificity of *Pax-6* gene expression. In this respect, it is interesting to note that the enhancer has no activity in pancreatic cell line  $\beta$ TC-3, which also expresses the *Pax-6* gene (42), and that Southwestern experiments performed with nuclear extracts from these cells fail to detect the proteins found in E7 QNR nuclear extracts (30).

We recently identified within the *Pax-QNR* gene a second promoter, termed P1, located 3 kbp downstream from the P0 promoter (32). However, the 3' neuroretina-specific enhancer described in this study has no activity on this promoter. Since the TATA box sequences of the two responsive promoters (P0 and TK) are identical (ATATTAAGG) and differ from those of the P1 promoter (ATAAAGCAATAT), it was tempting to test whether the specificity of the enhancer for the P0 promoter and its failure to interact with the P1 promoter were attributable to the divergent TATA box sequences of the two promoters, as has been described for the myoglobin gene (45). However, when the P0 TATA box sequence was changed to that of the P1 promoter, the responsiveness of the enhancer was not abolished (data not shown). Thus, the differential in enhancer activities between P0 and P1 promoters may reside in other *cis*-acting promoter sequences. Alternatively, since recent studies suggest that some enhancer factors are dependent upon auxiliary factors (coactivators) for activation functions (for a review, see reference 39), it remains possible that the enhancer is unable to activate the P1 promoter because this promoter lacks its full complement of required factors to interact with the active enhancer. Another explanation is the presence of an inhibiting factor that prevents the P1 promoter from interacting with the proteins of the active enhancer. These hypotheses about enhancer mechanisms could also be invoked in the differential in enhancer activities observed in the neuroretina between E5 and E7. In E5 neuroretina, the enhancer could be inactivated by a repressor or the enhancer activity needs the presence of an activator, absent at this stage. Cloning the gene(s) encoding the enhancer element-binding factor(s) and analyzing the molecular interactions between the enhancer element-binding factor(s) and other transcription factors will allow us to further define precise molecular mechanisms involved in the regulation of the *Pax-6* gene.

Preliminary results suggests that there may be another important *cis*-acting element(s) within the promoter-distal region that also contribute to the neuroretina-specific expression of the *Pax-QNR* gene. Defining additional regulatory elements that control the tissue-specific developmental expression of the *Pax-6* gene may be essential for in vivo studies of the role of this gene in normal eye development and for determining the

mechanisms of pathogenesis associated with *Pax-6*. Indeed, it remains possible that mutations in the *Pax-6* regulatory elements cause developmental defects associated with this gene, as has been suggested for some aniridia cases showing no mutations in the *Pax-6* coding region (18, 40). Although point mutations in the regulatory sequences associated with pathogenesis are rare, this has been demonstrated for the  $\beta$ -globin cluster (4).

#### ACKNOWLEDGMENTS

This work was supported by grants from the Centre National de la Recherche Scientifique, the Institut Pasteur de Lille, the Association de Recherche contre le Cancer, and the Association Française Retinitis Pigmentosa.

#### REFERENCES

- Adams, B., P. Dörfler, A. Aguzzi, Z. Kozmik, P. Urbanek, I. Maurer-Fogy, and M. Busslinger. 1992. *Pax-5* encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS and adult testis. *Genes Dev.* 6:1589-1607.
- Babiss, L. E., and J. M. Friedman. 1990. Regulation of *N-myc* gene expression: use of an adenovirus vector to demonstrate posttranscriptional control. *Mol. Cell. Biol.* 10:6700-6708.
- Beebe, D. 1994. Homeobox genes and vertebrate eye development. *Invest. Ophthalmol. Visual Sci.* 35:2897-2900.
- Berry, M., F. Grosfeld, and N. Dillon. 1992. A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature (London)* 358:499-502.
- Boshart, M., M. Klüppel, A. Schmidt, G. Schütz, and B. Luckow. 1992. Reporter constructs with low background activity utilizing the *cat* gene. *Gene* 110:129-130.
- Carriere, C., S. Plaza, P. Martin, B. Quatannens, M. Bailly, D. Stehelin, and S. Saule. 1993. Characterization of quail *Pax-6* (*Pax-QNR*) proteins expressed in the neuroretina. *Mol. Cell. Biol.* 13:7257-7266.
- Chalepakis, G., R. Fritsch, H. Fickenscher, U. Deutsch, M. Goulding, and P. Gruss. 1991. The molecular basis of the *undulated/Pax-1* mutation. *Cell* 66:873-884.
- Chalepakis, G., A. Stoykova, J. Wijnholds, P. Tremblay, and P. Gruss. 1993. *Pax*: gene regulators in the developing nervous system. *J. Neurobiol.* 24:1367-1384.
- Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156-159.
- Coulombre, J. L., and A. Coulombre. 1965. Regeneration of neural retina from the pigmented epithelium in the chick embryo. *Dev. Biol.* 12:79-92.
- Crepiex, P., D. Leprince, A. Flourens, O. Albagli, E. Ferreira, and D. Stehelin. 1993. The two functionally distinct amino termini of chicken *c-eis-1* products arise from alternative promoter usage. *Gene Expression* 3:215-225.
- Dozier, C., S. Ansseau, E. Ferreira, J. Coll, and D. Stehelin. 1991. An alternatively spliced *c-mil/raf* mRNA is predominantly expressed in chicken muscular tissues and conserved among vertebrate species. *Oncogene* 6:1307-1311.
- Dozier, C., C. Carriere, D. Grevin, P. Martin, B. Quatannens, D. Stehelin, and S. Saule. 1993. Structure and DNA-binding properties of *Pax-QNR*, a paired box- and homeobox-containing gene. *Cell Growth Differ.* 4:281-289.
- Duboule, D. 1994. Guidebook to the homeobox genes. Oxford University Press, Oxford.
- Eguchi, G., and R. Kodama. 1993. Transdifferentiation. *Curr. Opin. Cell Biol.* 5:1023-1028.
- Fickenscher, H. R., G. Chalepakis, and P. Gruss. 1993. Murine *Pax-2* protein is a sequence-specific trans-activator with expression in the genital system. *DNA Cell Biol.* 12:381-391.
- Glaser, T., L. Jepeal, J. G. Edwards, R. Young, J. Favor, and R. Maas. 1994. *PAX6* gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nature Genet.* 7:463-471.
- Glaser, T., D. S. Walton, and R. L. Maas. 1992. Genomic structure, evolutionary conservation and aniridia mutations in the human *Pax-6* gene. *Nature Genet.* 2:232-239.
- Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2:1044-1051.
- Hambor, J. E., J. Mennone, M. E. Coon, J. H. Hanke, and P. Kavathas. 1993. Identification and characterization of an *Alu*-containing, T-cell-specific enhancer located in the last intron of the human CD8 $\alpha$  gene. *Mol. Cell. Biol.* 13:7056-7070.
- Hill, R. E., J. Favor, B. Hogan, C. C. Ton, G. Saunders, I. Hanson, J. Prosser, T. Jordan, N. Hastie, and V. V. Heyningen. 1991. Mouse *small eye* results from mutations in a paired-like homeobox-containing gene. *Nature (London)* 354:522-525.

22. Jacquemin, P., C. Oury, B. Peers, A. Morin, A. Belayew, and J. A. Martial. 1994. Characterization of a single strong tissue-specific enhancer downstream from the three human genes encoding placental lactogen. *Mol. Cell. Biol.* 14:93-103.
23. Lavery, D. J., and U. Schibler. 1993. Circadian transcription of the cholesterol 7 $\alpha$  hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes Dev.* 7:1871-1884.
24. Leprince, D., P. Crepieux, and D. Stehelin. 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.
25. Li, H.-S., J.-M. Yang, R. Jacobson, D. Pasko, and O. Sundin. 1994. *Pax-6* is first expressed in a region of ectoderm anterior to the early neural plate: implications for stepwise determination of the lens. *Dev. Biol.* 162:181-194.
26. Liu, I. S., J.-D. Chen, L. Ploder, D. Vidgen, D. Van der Kooy, V. I. Kalnins, and R. R. McInnes. 1994. Developmental expression of a novel murine homeobox gene (*Chx10*): evidence for roles in determination of the neuroretina and inner nuclear layer. *Neuron* 13:377-393.
27. Martin, P., C. Carriere, C. Dozier, B. Quatannens, M. Mirabel, B. Vandebunder, D. Stehelin, and S. Saule. 1992. Characterization of a paired box and homeobox-containing quail gene (*pax-QNR*) expressed in the neuroretina. *Oncogene* 7:1721-1728.
28. O'Prey, J., S. Ramsay, I. Chambers, and P. R. Harrison. 1993. Transcriptional up-regulation of the mouse cytosolic glutathione peroxidase gene in erythroid cells is due to a tissue-specific 3' enhancer containing functionally important CACC/GT motifs and binding sites for GATA and Ets transcription factors. *Mol. Cell. Biol.* 13:6290-6303.
29. Pittack, C., M. Jones, and T. Reh. 1991. Basic fibroblast growth factor induces retinal pigment epithelium to generate neural retina in vitro. *Development* 113:577-588.
30. Plaza, S. Unpublished data.
31. Plaza, S., C. Dozier, and S. Saule. 1993. Quail PAX-6 (PAX-QNR) encodes a transcription factor able to bind and *trans*-activate its own promoter. *Cell Growth Differ.* 4:1041-1050.
32. Plaza, S., C. Dozier, N. Turque, and S. Saule. Submitted for publication.
33. Quiring, R., U. Walldorf, U. Kloter, and W. Gehring. 1994. Homology of the eyeless gene of *Drosophila* to the small eye gene in mice and aniridia in humans. *Science* 265:785-789.
34. Schibler, U., O. Hagenbüchle, P. K. Wellauer, and A. C. Pittet. 1983. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene *Amy-1a* in the parotid gland and the liver. *Cell* 33:501-508.
35. Spegeleere, P., B. van Hille, N. Spruyt, S. Faisst, J. J. Cornelis, and J. Rommelaere. 1991. Initiation of transcription from the minute virus of mice P4 promoter is stimulated in rat cells expressing a *c-Ha-ras* oncogene. *J. Virol.* 65:4919-4928.
36. Stapleton, P., A. Weith, P. Urbanek, Z. Kozmik, and M. Busslinger. 1993. Chromosomal localization of seven *PAX* genes and cloning of a novel family member, *PAX-9*. *Nature Genet.* 3:292-298.
37. Stein, B., H. J. Rahmsdorf, A. Steffen, M. Litfin, and P. Herrlich. 1989. UV-induced DNA damage is an intermediate step in UV-induced expression of human immunodeficiency virus type 1, collagenase, *c-fos*, and metallothionein. *Mol. Cell. Biol.* 9:5169-5181.
38. Strachan, T., and A. P. Read. 1994. *PAX* genes. *Curr. Opin. Genet. Dev.* 4:427-438.
39. Tijan, R., and T. Maniatis. 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77:5-8.
40. Ton, C. C., H. Hirvonen, H. Miwa, M. Weil, P. Monaghan, T. Jordan, V. Van Heyningen, N. Hastie, H. Meijers-Heijboer, M. Drechsler, B. Royer-Pokora, F. Collins, A. Swaroop, L. C. Strong, and G. F. Saunders. 1991. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* 67:1059-1074.
41. Tremblay, P., and P. Gruss. 1994. *Pax*: genes for mice and men. *Pharmacol. Ther.* 61:205-226.
42. Turque, N., S. Plaza, F. Radvanyi, C. Carrière, and S. Saule. 1994. *Pax-QNR/Pax-6*, a paired box- and homeobox-containing gene expressed in neurons, is also expressed in pancreatic endocrine cells. *Mol. Endocrinol.* 8:929-938.
43. Walther, C., and P. Gruss. 1991. *Pax-6*, a murine paired box gene, is expressed in the developing CNS. *Development* 113:1435-1449.
44. Walther, C., J.-L. Guenet, D. Simon, U. Deutch, B. Jostes, M. D. Goulding, D. Plachov, R. Balling, and P. Gruss. 1991. *Pax*: a murine gene family of paired box containing genes. *Genomics* 11:424-434.
45. Wefald, F. C., B. H. Devlin, and S. R. Williams. 1990. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature (London)* 344:260-262.
46. Zannini, M., H. Francis-Lang, D. Plachov, and R. Di Lauro. 1992. *Pax-8*, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol. Cell. Biol.* 12:4230-4241.

## Discussion de l'article 1

Dans cette étude, il a été montré l'existence d'un enhancer, dénommé  $\alpha$ , responsable de la stimulation dans la neurorétine à partir de 7 jours du développement embryonnaire du promoteur Po du gène *Pax-6* de caille. Cet enhancer n'agit pas sur le promoteur P1 de *Pax-6*. De plus, nous avons pu mettre en évidence la conservation de la fonction enhancer du fragment génomique  $\alpha$  chez la souris.

Les expériences de protection à la DNaseI, d'une part, et les expériences de retard sur gel, d'autre part, ont permis de mettre en évidence le fait que les éléments de l'enhancer contiennent quatre sites spécifiques de fixation pour des protéines nucléaires de neurorétine, DF1, DF2, DF3 et DF4. Par des tests d'expression avec des gènes rapporteurs, nous avons aussi démontré que ces sites de fixation jouent un rôle fonctionnel. Les complexes protéines-ADN mis en évidence par la technique de retard sur gel (EMSA) sont détectés à de faibles concentrations dans les extraits cellulaires de neurorétine de caille à 5 jours de développement et ceci est en corrélation avec une activité plus faible de l'enhancer à ce stade. Ces observations suggèrent que les facteurs qui se lient à l'enhancer sont en quantité limitée dans la neurorétine au jour E5 du développement ou que l'affinité de ces séquences pour leurs facteurs est moindre.

Nous proposons que l'activité enhancer est causée par une synergie entre les quatre sites de fixation des protéines DF1 à DF4. Ainsi les facteurs qui se fixent à ces sites doivent collaborer afin d'activer l'enhancer. L'idée d'un enhancer spécifique d'un tissu, constitué de plusieurs sous-unités agissant en synergie, est en accord avec la nature modulaire bien connue des enhancers (O'Prey et al., 1993, Jacquemain et al., 1994). Des analyses par la technique de "South-Western" et des expériences de compétitions en retard sur gel démontrent que les quatre sites de fixation lient un ensemble inconnu de protéines identiques. Même si ces protéines sont détectables principalement dans la neurorétine, elles sont aussi faiblement détectées dans la rétine pigmentaire embryonnaire. Ceci peut s'expliquer par le fait que l'épithélium pigmentaire rétinien conserve la capacité de se transdifférencier *in-vivo* en neurorétine (Coulombre et Coulombre, 1965).

L'enhancer  $\alpha$  chez la caille présente une identité de séquence de 90% avec la région homologue murine. De plus, les quatre sites de fixations des protéines sont présents dans la séquence de souris. Lors d'études menées en transfection transitoire nous avons observé que l'activité enhancer est aussi conservée chez la souris. En revanche, malgré la grande

identité de séquence qui existe entre la caille et la souris, l'activité de l'enhancer de souris est de moitié inférieure à celle de la caille.

Dans l'ensemble, nos résultats indiquent que l'enhancer de la neurorétine est conservé chez les mammifères et est un élément régulateur transcriptionnel complexe, qui peut contribuer à l'expression spécifique de Pax-6 dans la neurorétine. Un élément supplémentaire en faveur de ce mécanisme est que cet enhancer ne présente aucune activité dans la lignée cellulaire pancréatique  $\beta$ TC-3. Or, cette lignée exprime le gène Pax-6 et lors des expériences de "South-Western" réalisées avec des extraits cellulaires, il a été impossible de détecter les protéines qui se fixent sur l'enhancer et qui sont présentes dans les extraits nucléaires de neurorétine (résultats non publiés et Turque et al., 1994).

Dans l'avenir il sera bien évidemment important de réussir à cloner le(s) gène(s) qui codent le(s) facteur(s) de liaison aux éléments caractérisés sur l'enhancer et d'analyser les interactions moléculaires entre ce(s) facteur(s) et d'éventuels autres facteurs de transcription. Ainsi nous pourrions mieux comprendre les mécanismes moléculaires impliqués dans la régulation du gène Pax-6. La compréhension de ces mécanismes est essentielle pour les études *in-vivo* qui portent sur le rôle de l'expression du gène Pax-6 dans le développement normal de l'oeil ou pour les mécanismes pathogènes qui lui sont associés. Il est possible que des mutations dans les éléments régulateurs du gène Pax-6 soient importantes dans le développement de certaines des pathologies associées à ce gène. Un tel mécanisme a été suggéré pour certains cas d'aniridie qui ne présentaient pas de mutations au niveau de la région codante du gène Pax-6 (Ton et al., 1991 ; Glaser et al., 1992). Même si les pathologies associées avec des mutations ponctuelles de séquences régulatrices sont rares, cela a déjà été observé pour le cluster des gènes de  $\beta$ -globine (Berry et al., 1992).

## *Article 2*

## Article 2

### The homeobox-containing Engrailed (EN-1) product down-regulates the expression of Pax-6 through a DNA-binding independent mechanism.

S.Plaza, M-C Langlois, N. Turque, M. Bailly, A. Begue, B. Quatannens, C. Dozier et S. Saule.

Cell Growth and Differentiation, 8, 1115-1125, 1997.

La régulation de l'expression de Pax-6 doit être très fine car non seulement Pax-6 s'exprime dans des tissus spécifiques mais son taux d'ARNm fluctue au cours du développement embryonnaire et une quantité anormale d'un type de messager peut entraîner des mutations importantes chez l'animal concerné (cf. introduction). Les gènes Engrailed-1 et 2 s'expriment dans des territoires d'expression mutuellement exclusifs avec ceux du gène Pax-6 dans le système nerveux central au cours de l'embryogenèse précoce. En-1 et En-2 sont détectés dans le cerveau moyen, une région où les ARNm de Pax-6 ne sont pas retrouvés. De plus, les régions du cerveau moyen de caille greffées dans les régions diencephaliques de poulet induisent non seulement l'apparition de structures de type mét-mésencéphaliques chez l'hôte mais aussi l'expression du gène En-2 et l'extinction de l'expression du gène Pax-6. Par ailleurs, En-1 et En-2 sont fortement exprimés dans les couches ganglionnaires et amacrines de la rétine après l'éclosion, à une période où les taux d'expression des ARNm Pax-6 chutent considérablement. En conséquence, il semblerait qu'Engrailed intervienne dans la régulation négative du gène Pax-6.

Pour tester la véracité de cette hypothèse, nous avons étudié dans l'article présenté ci-dessous, l'effet de vecteurs exprimant En-1 ou En-2, dans des cellules de rétine, sur les promoteurs Po et P1 du gène Pax-6, transactivés ou non par la protéine P46 codée par le gène Pax-6. Les résultats obtenus montrent tout d'abord, que des cellules de rétine pigmentaire infectées par le virus SFCV-LE comportant la séquence codante d'En-1 n'expriment plus Pax-6. Une telle répression est également observée lorsque En-2 est utilisé. Elle s'exerce sur les promoteurs P1 et P0, à la fois sur leur transcription basale et sur leur transcription activée par la protéine P46. Par ailleurs, cette répression exercée par la protéine Engrailed est indépendante de son activité de fixation à l'ADN : elle s'exerce encore en l'absence du domaine homéo et des sites de fixation pour En-2 sur le promoteur Po. En revanche, elle nécessite une interaction entre la protéine Engrailed et le domaine paired de Pax-6, mise en évidence par la technique de co-rétention protéique ("pull-down"



assay). L'interaction entre ces deux protéines altère la capacité à fixer l'ADN de la protéine P46 et, en conséquence, diminue le taux d'expression des ARNm initiés à partir des promoteurs Po et P1, deux promoteurs activés par cette protéine P46. Les protéines Engrailed interfèrent donc avec l'autorégulation de Pax-6.

# The Homeobox-containing Engrailed (En-1) Product Down-Regulates the Expression of *Pax-6* through a DNA Binding-independent Mechanism<sup>1</sup>

Serge Plaza, Marie-Claire Langlois, Nathalie Turque,<sup>2</sup> Sébastien LeCornet, Manuella Bailly, Agnès Bègue, Brigitte Quatannens, Christine Dozier, and Simon Saule<sup>3</sup>

Centre National de la Recherche Scientifique EP560 [S. P., M-C. L., N. T., S. L., M. B., C. D., S. S.] and UMR319 [A. B., B. Q.], Institut Pasteur, 59019 Lille Cedex, France

## Abstract

By *in situ* hybridization of quail neuroretinas, we observed that *Engrailed (En-1)* is expressed both in the ganglionic and the amacrine cell layers, similarly to *Pax-6*. Because we observed a decrease of *Pax-6* expression in the neuroretina of hatched animals, we studied the effect of the chicken *En-1* and *En-2* proteins on *Pax-6* expression. *En-1* and to some extent *En-2* were able to repress the basal and the p46<sup>*Pax-6*</sup>-activated transcription from the two *Pax-6* promoters. Infection of retinal pigmented epithelium by a virus encoding the *En-1* protein repressed the endogenous *Pax-6*, and a similar effect was observed with a homeodomain-deleted *En-1*. *In vitro* interaction indicates that *En* proteins are able to interact with the p46<sup>*Pax-6*</sup> through the paired domain. This interaction negatively regulates the DNA-binding properties of the p46<sup>*Pax-6*</sup>. These results suggest an interplay between *En-1* and *Pax-6* during the central nervous system development and indicate that *En-1* may be a negative regulator of *Pax-6*.

## Introduction

Transcriptional regulation by the homeodomain proteins is thought to coordinate the precise spatial and temporal sequence of growth and differentiation (1). Through homology to the paired box, nine paired box-containing genes, the *Pax* gene family, have been isolated in vertebrates (2, 3), and mutations in these genes are associated with mouse developmental mutants (2, 4). Mutations in the *Pax-6* gene have been associated with the mouse mutant *small eye* (5, 6), and

the corresponding human gene (*AN*) has been found deleted or mutated in aniridia (7). The *Pax-6* gene is expressed in the developing central nervous system, the optic cup, lens, overlying ectoderm, neuronal layers of the retina, and endocrine pancreas (8-10). In the developing brain of young embryos, *Pax-6* is not expressed in the mesencephalon in the vertebrates species (3, 11, 12). The restricted expression pattern of this gene suggests a transcriptional regulation of expression by tissue-specific factors. Among several mechanisms used to restrain gene expression, active repression is often used, and several homeodomain proteins are known as active repressors (13, 14, 15). *Engrailed (En)* is a homeodomain-containing protein (16) that binds to DNA with a sequence specificity similar to that of several other homeodomain proteins (17). *En* can repress transcription *in vitro* by competing with general transcription factor IID for binding to the TATA box (18). It can also repress the transcription activated by different activators bound to distinct sites separated from the *En* binding sites, and this repression requires a protein sequence outside the homeodomain (19, 20). In vertebrates, two genes, *En-1* and *En-2*, have been identified (21), both expressed in a band of the early neural plate around the mid/hindbrain junction, and continue to be expressed in this region throughout development. Both *En-1* (22) and *En-2* (23) mouse mutants showed abnormal mid/hindbrain development, and the rescue of the *En-1* mutant phenotype by replacement of *En-1* with *En-2* suggests similar biochemical properties for the two proteins (24). When grafted ectopically into the diencephalon of a chick host embryo, portions of the quail mid/hindbrain domain are able to induce, in the contacted host tissue, the ectopic expression of *En-2* (25, 26) and the development of met-mesencephalic structures (26). Consequently, to the *En-2* expression in the diencephalic host tissue, *Pax-6* becomes undetectable (12). We have recently characterized two promoter regions in the quail *Pax-6* gene termed P0 (27) and P1 (28). Only the mRNAs initiated from the P1 promoter could be detected in the quail RPE.<sup>4</sup> P0- and P1-initiated mRNAs were accumulated differentially during neuroretina development, P0-initiated transcripts being detected after the neuronal differentiation, and these two *Pax-6* transcripts return close to the background level in the neuroretina of hatched animals (28).

In this report, we demonstrate that the full-length *En-1* and *En-2* are able to repress, in transient transfection assays, *Pax-6* P0 and P1 basal as well as protein-induced transactivation. We were able to footprint three *En-2* binding sites on

Received 6/23/97; accepted 8/19/97.

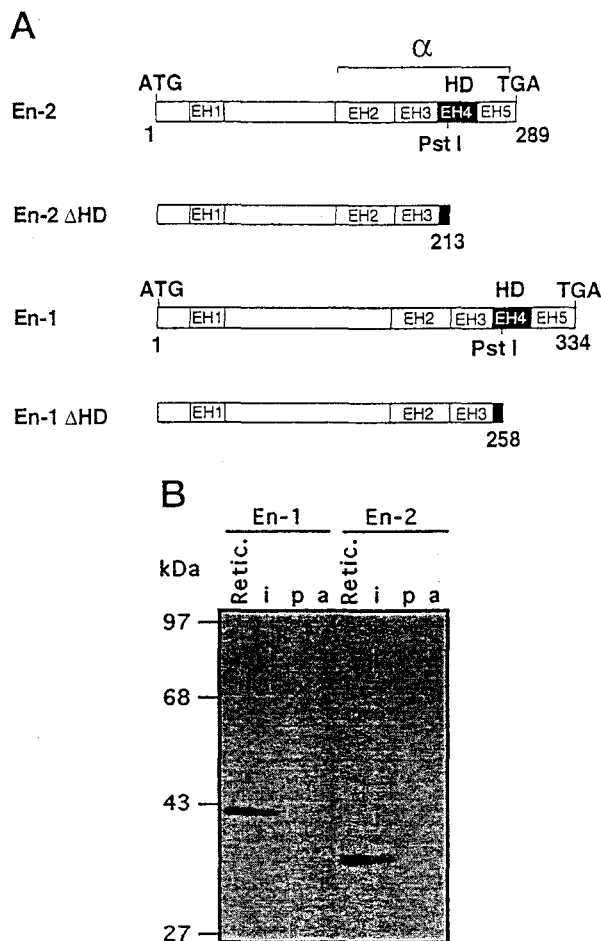
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> This work was supported by grants from the Centre National de la Recherche Scientifique, Institut Pasteur de Lille, Association pour la Recherche contre le Cancer, and the Association Française Retinitis Pigmentosa.

<sup>2</sup> Present address: Developmental Neurogenetics, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD 20892.

<sup>3</sup> To whom requests for reprints should be addressed. Phone: (33) 3-20-87-77-86; Fax: (33) 3-20-87-79-08.

<sup>4</sup> The abbreviations used are: RPE, retinal pigmented epithelium; CAT, chloramphenicol acetyltransferase; QNR, quail neuroretina; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GST, glutathione S-transferase.



**Fig. 1.** A, structure of the chicken *En* genes. Black boxes indicate the homeodomain. The bracket above *En-2* delineates the cDNA part used to produce the  $\alpha$ En serum. EH, engrailed homology domain, deduced from a sequence comparison of the different En proteins (29); numbers below the boxes refer to amino acids. B, the *En* cDNAs were transcribed *in vitro* in a sense orientation and translated in a rabbit reticulocyte lysate (Retic. lanes). Translated proteins were immunoprecipitated with  $\alpha$ En rabbit antiserum prepared against bacterially expressed peptides (Lanes i), the cognate preimmune serum (Lanes p), or the immune serum blocked by an excess of peptide (Lanes a).

the P0 promoter. However, the *En-1* protein deleted for the homeodomain was as efficient as the wild-type *En-1* in P0 and P1 repression, suggesting that a DNA binding-independent mechanism is operative in *En-1*-mediated repression on *Pax-6* promoters. In addition, the chicken *En-1* coding sequence as well as the *En-1* coding sequence deleted for the homeodomain inserted in a retrovirus and introduced in the RPE cells are both able to down-regulate the endogenous *Pax-6* expression. Finally, we were able to demonstrate by *in situ* hybridization that, at the time of *Pax-6* down-regulation to nearly background level after hatching, *En-1* is expressed in the same neuroretina layers as *Pax-6*.

## Results

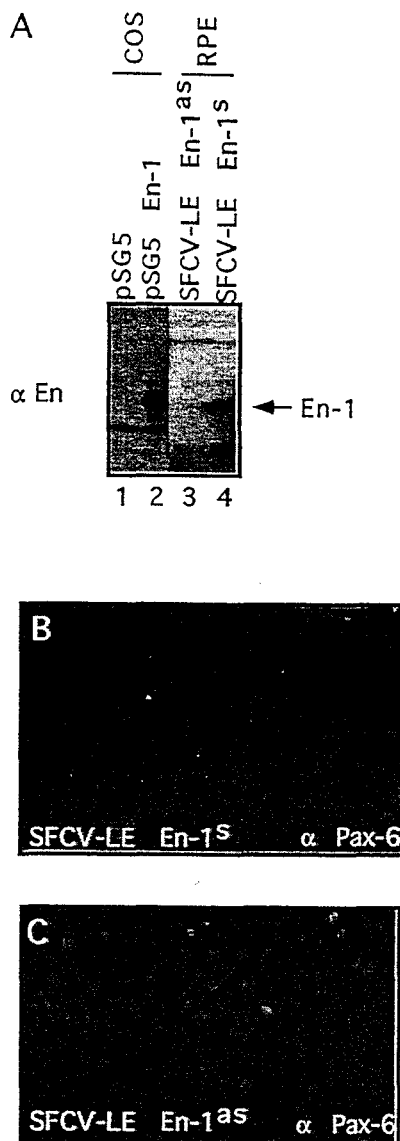
**Cloning of the Chicken *En* cDNAs.** We isolated *En* cDNAs from a chick embryo cDNA library (Fig. 1A). Nucleotide se-

quencing of part of these cDNAs revealed that they corresponded to the chicken *En-1* and *En-2*, respectively (29). These cDNAs were able to direct the synthesis of the full-length *M<sub>r</sub>* 44,000 and 39,000 proteins as described previously (21). From bacterially expressed *En-2* COOH-terminus peptide (amino acids 154–285, which include the homeodomain), we prepared rabbit antiserum able to specifically recognize both the *En-1* and *En-2* proteins (Fig. 1B).

***En-1* Expression Resulted in *Pax-6* Down-Regulation *in Vivo*.** To test the possibility that *En* was able to down-regulate *Pax-6* expression *in vivo*, we introduced the *En-1* coding sequence into a retroviral vector (30). We prepared two constructs containing the *En-1* coding sequence in the sense (SFCV-LE *En-1<sup>s</sup>*) or antisense (SFCV-LE *En-1<sup>as</sup>*) orientation. Transfection of RPE cells expressing *Pax-6* with the retroviral constructs, followed by a G418 treatment to select cells expressing the virus, produced cells expressing *Pax-6* only when the SFCV-LE *En-1<sup>as</sup>* virus was used. As shown in Fig. 2C, *Pax-6* could be observed by immunofluorescence staining in the cell nuclei after SFCV-LE *En-1<sup>as</sup>*, but not SFCV-LE *En-1<sup>s</sup>* expression (Fig. 2B). Western blot experiments performed on these RPE cell lysates showed that the p44<sup>En-1</sup> was detected in cells expressing the SFCV-LE *En-1<sup>s</sup>* but not SFCV-LE *En-1<sup>as</sup>* viruses (Fig. 2A). As shown by the RNase protection assay (Fig. 3), the decrease in *Pax-6* protein is due to a decrease in *Pax-6* RNA in SFCV-LE *En-1<sup>s</sup>*-transfected RPE cells. A PhosphorImager (Molecular Dynamics) quantification of the RNA signal (Fig. 3A) shows an 80% signal reduction in the SFCV-LE *En-1<sup>s</sup>*-transfected versus control cells (Fig. 3B). These results demonstrate that an overexpression of p44<sup>En-1</sup> negatively regulates *Pax-6* expression *in vivo*.

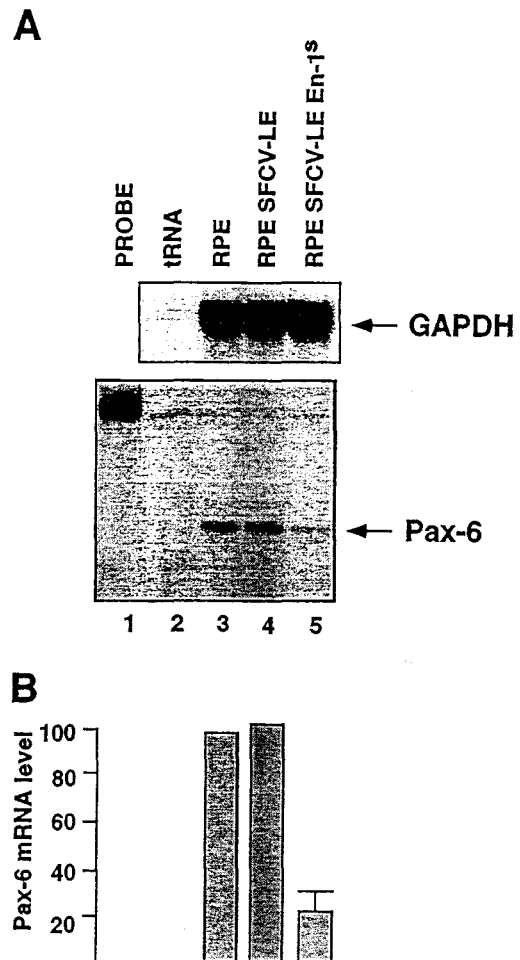
**Regulation of the *Pax-6* P0 and P1 Promoters by the *En* Proteins.** Two promoters were characterized in the quail *Pax-6* gene (Fig. 4A), termed P0 (27) and P1 (28). P0- and P1-initiated mRNAs were accumulated differentially during neuroretina development; P0-initiated transcripts are detected only after the neuronal differentiation, and both *Pax-6* transcripts return close to the background level in the neuroretina of hatched animals (28). To determine whether the chicken *En* proteins were able to regulate the *Pax-6* promoters, equal amounts of the CAT-expressing vectors [P0XA containing the 1.5-kb *Xba*I-Asp718 P0 promoter fragment (27) and pP1N<sup>S</sup> containing the 1.01kbp *Not*I P1 promoter fragment (28)] were cotransfected together with the pSG5*En-1* or pSG5*En-2* eucaryotic expression vectors, into primary QNR cells. To study the effect of *En* proteins on the activated promoters, we performed cotransfections of the P0 and P1 constructs with the vectors expressing *Pax-6* or *c-myb* proteins demonstrated previously as *Pax-6* activators (27, 31). Cell lysates were collected 2 days after transfection, and the level of CAT activity in the lysates was determined.

As shown in Fig. 4B, in QNR cells CAT activity driven by the P0 (Lane 1) and P1 (Lane 4) promoters are reduced in the presence of both *En* products (90% for P0 and 50% for P1). CAT activity was increased compared to control when the P0- and P1-*Pax-6* promoter CAT constructs were transfected in the presence of p46<sup>Pax-6</sup> activator (Fig. 4, C and D, Lanes 2). Again, coexpression of *En* products together with



**Fig. 2.** Characterization of En-1 protein in transfected cells. **A**, transfected cells are indicated in the top of the figure. Western blot was performed on cell extracts with  $\alpha$ En homeodomain serum. Bound antibodies were detected with an enhanced chemiluminescence detection kit. The p44<sup>En-1</sup> protein is indicated by an arrow. Pax-6 products were detected by immunofluorescence in SFCV-LE En-1<sup>S</sup> (**B**)- and SFCV-LE En-1 <sup>$\Delta$ S</sup> (**C**)-expressing RPE cells after G418 selection. Subcellular localization of Pax-6 products was assayed by indirect immunofluorescence on fixed cells with anti-Pax-6 serum, serum 11 (recognizing the paired domain). Anti-Pax-6 immunoreactive proteins were detected with FITC-labeled swine antirabbit immunoglobulin secondary reagent.

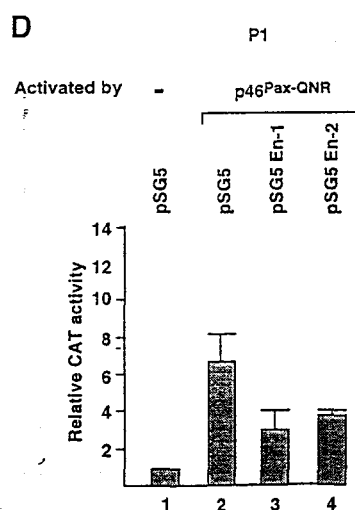
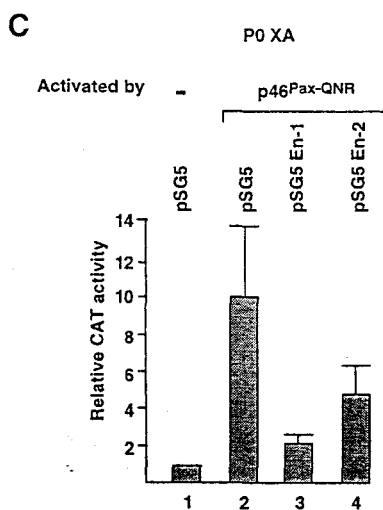
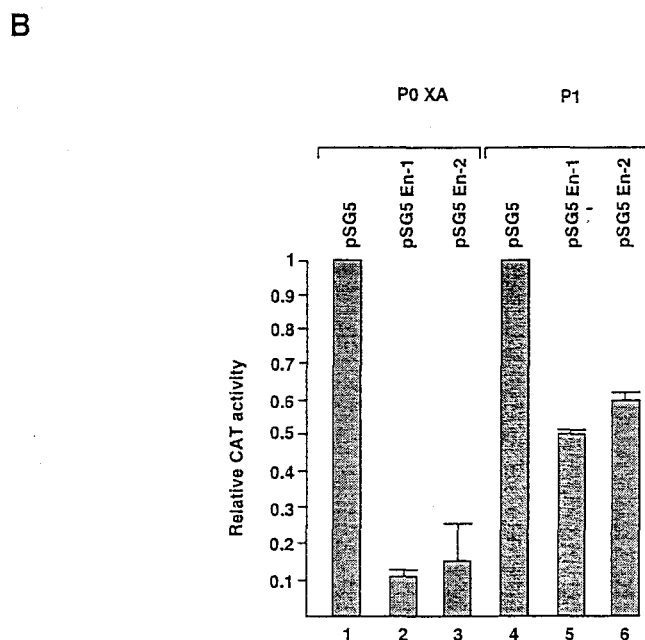
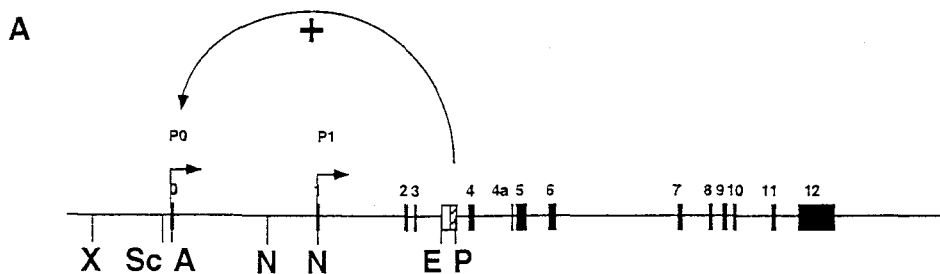
p46<sup>Pax-6</sup> activator resulted in marked reduction of the CAT activity driven by the P0 (Fig. 4C) and P1 (Fig. 4D) promoters. The activated transcription was inhibited by the two En proteins, En-1 being usually a more efficient repressor than En-2, with the pSG5-expressing vector used expressing a similar amount of En-1 and En-2 as judged by a Western blot experiment performed with transfected cell lysates (data not shown). Similar results were obtained with the c-myc-expressing vector (data not shown). These results indicate



**Fig. 3.** **A**, variations in the accumulation of P1-initiated Pax-6 mRNAs in RPE transfected with the SFCV-LE En-1<sup>S</sup> or control vector DNA and recovered after G418 treatment. **Top**, RNase protection experiment performed with 2  $\mu$ g of total RNA with a GAPDH probe (28). **Bottom**, the 323-nucleotide fragment (arrow) detect the P1-initiated messengers. Twenty  $\mu$ g of total RNA were used for each point. **B**, the signal generated by the 323-nucleotide fragment shown in **A** was measured with a PhosphorImager (Molecular Dynamics) and expressed as arbitrary units of RNA level. Data are means; bar, SD.

that En-1 and En-2 proteins are able to repress Pax-6 promoter expression.

**En DNA-binding Homeodomains Were Dispensable for En-induced Pax-6 Promoter Repression.** Three En-2 homeodomain binding regions were identified within the 214-bp Scal-Asp718 P0 fragment, the expression of which was inhibited by the En proteins (data not shown). However, because several transcription factors have been shown to act in a DNA binding independent way, we questioned whether or not En required the DNA-binding domain to inhibit the Pax-6 promoters. We prepared for both En-1 and En-2 an expression vector encoding a truncated protein, deleted from the COOH-terminus DNA-binding homeodomain (Fig. 1A, mutant En- $\Delta$ HD; these deletants retained only five amino acids of the homeodomain first helix, as deduced from the *Drosophila* En homeodomain; Ref. 32). All of these constructs produced proteins of the expected size in transfected



**Fig. 4.** Repression of the *Pax-6* promoters by the En proteins. **A**, schematic representation of the *Pax-6* locus. X, XbaI; Sc, ScaI; A, Asp718; E, EcoRV; N, NotI; P, PstI. The + refers to the positive activatory effect of the EP fragment enhancer included in exon  $\alpha$  on the P0 promoter into the neuroretina cells. **B**, QNR cells were transfected with 1  $\mu$ g of CAT construct (pXA containing the 1.5-kb Xba-Asp718 *Pax-6* genomic fragment for P0 or pP1N<sup>S</sup> containing the 1.01-kb NotI fragment *Pax-6* genomic fragment for P1) and 5  $\mu$ g of pSG5En-1 encoding the chicken p44<sup>En-1</sup> or pSG5En-2 encoding the chicken p39<sup>En-2</sup> as indicated (top). The control DNA in the transfection is the pSG5. **C** and **D**, QNR cells were transfected with 1  $\mu$ g of CAT construct plus 5  $\mu$ g of the pJ3 $\Omega$ 103 plasmid containing the MC29-QNR2 cDNA insert encoding the p46<sup>Pax-6</sup> and 5  $\mu$ g of the pSG5En-1 encoding the chicken p44<sup>En-1</sup> or pSG5En-2 encoding the chicken p39<sup>En-2</sup> as indicated (top). The control DNA in the transfection is the pSG5. The CAT activity of the promoter constructs in the absence of exogenous transactivator was designated arbitrarily as 1. Bars, SD, calculated from one transfection performed in duplicate with two distinct DNA preparations.

COS-1 as shown after immunoprecipitation of the COS cell lysate (the antiserum used is still able to recognize the deleted proteins), and we tested by electrophoretic mobility shift assay the ability of the homeodomain-deleted En proteins to bind an oligonucleotide bearing an Engrailed binding site (33). No specific complex could be observed in En- $\Delta$ HD COS cell extracts, although these proteins were found in the nucleus, suggesting that these deleted proteins are no longer able to bind the target DNA (data not shown).

We observed that the En- $\Delta$ HD products were as efficient as the wild-type En-1 product in P0 and P1 promoter repression. The repression observed was specific, because no inhibition was observed on the 4xGal4 thymidine kinase promoter activated by the Gal4VP16 protein cotransfected with the vector encoding the En- $\Delta$ HD proteins, indicating that Pax-6 promoter repression is not the result of a squelching of an element involved in the basal transcription machinery (data not shown).

To strengthen this observation, we introduced the En-1  $\Delta$ HD coding sequence into the pCRNCM vector and tested the *in vivo* effect of this protein on Pax-6 expression. Transfection of RPE cells expressing Pax-6 with the retroviral constructs, followed by a G418 treatment, produced cells expressing a reduced amount of Pax-6 when the CRNCM En-1  $\Delta$ HD virus was used (Fig. 5A, Lanes 3 and 4). A PhosphorImager quantification of the RNA signal shows an 80% signal reduction in the amount of Pax-6 mRNA when compared to GAPDH amount (Fig. 5B). These results demonstrate that, as for p44<sup>En-1</sup>, an overexpression of the En-1 protein devoid of homeodomain negatively regulates Pax-6 expression *in vivo*.

#### En-2 Recognizes the p46<sup>Pax-6</sup> through the Paired Domain and Can Inhibit the DNA-binding Activity of Pax-6.

The homeodomain-containing Pbx and En-2 proteins are capable of cooperative DNA binding (35), and this interaction involves the Engrailed homology domain 2 (EH2; Fig. 1A) of En-2 (36) also present in the En-1 protein (29). To investigate the possibility that the En products are able to repress the Pax-6 promoters in a DNA binding-independent way by binding directly to the Pax-6 protein, we prepared expression vectors encoding the En-1 protein fused to a Tag-His (Tag-His En-1) and part of En-2 (residues 154–285, domains EH2–EH5) fused to heterologous GST domain. Fusion proteins immobilized on Sepharose beads were incubated with L-[<sup>35</sup>S]methionine Pax-6 products (we used the five distinct Pax-6 proteins p48, p46, p43 and p32/33, produced by alternative splicing or internal AUG initiation described previously; Ref. 37) synthesized in reticulocyte lysates and subsequently purified using glutathione-Sepharose or nickel-Sepharose. The results obtained show that both Tag-His En-1 and GST-En-2 EH2-EH5 interacted with the paired containing protein p46 but not the paired less p32/33 (Fig. 6A). En-2EH2-EH5 interacted with the p48 and p46 Pax-6 proteins, very weakly with p43, and not at all with p32/33. To confirm that the paired domain was the target of En proteins, we prepared a GST-p46 paired domain (residues 2–134) and as a control a GST-p43 paired domain. The p43 protein exhibits, due to an alternative splicing event, a paired domain deleted from most of exon 5 (37). The paired domain of p46,

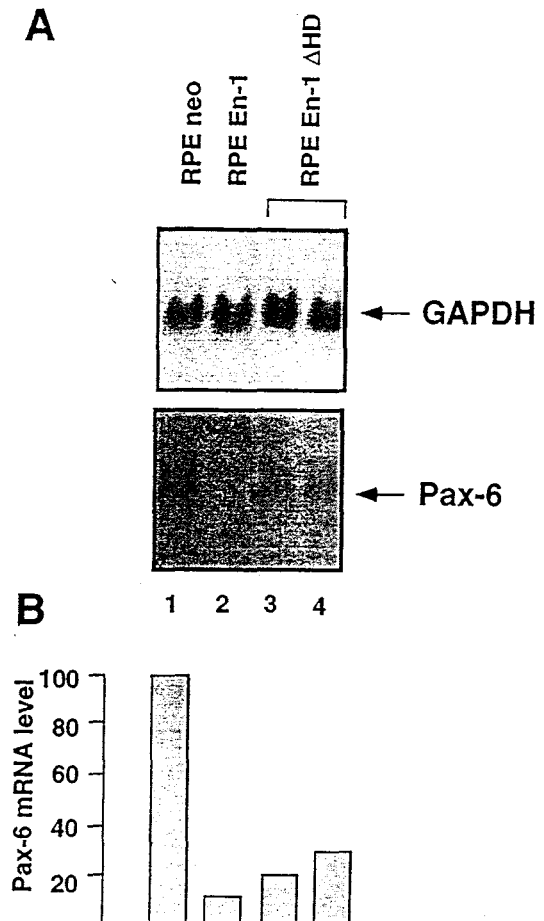
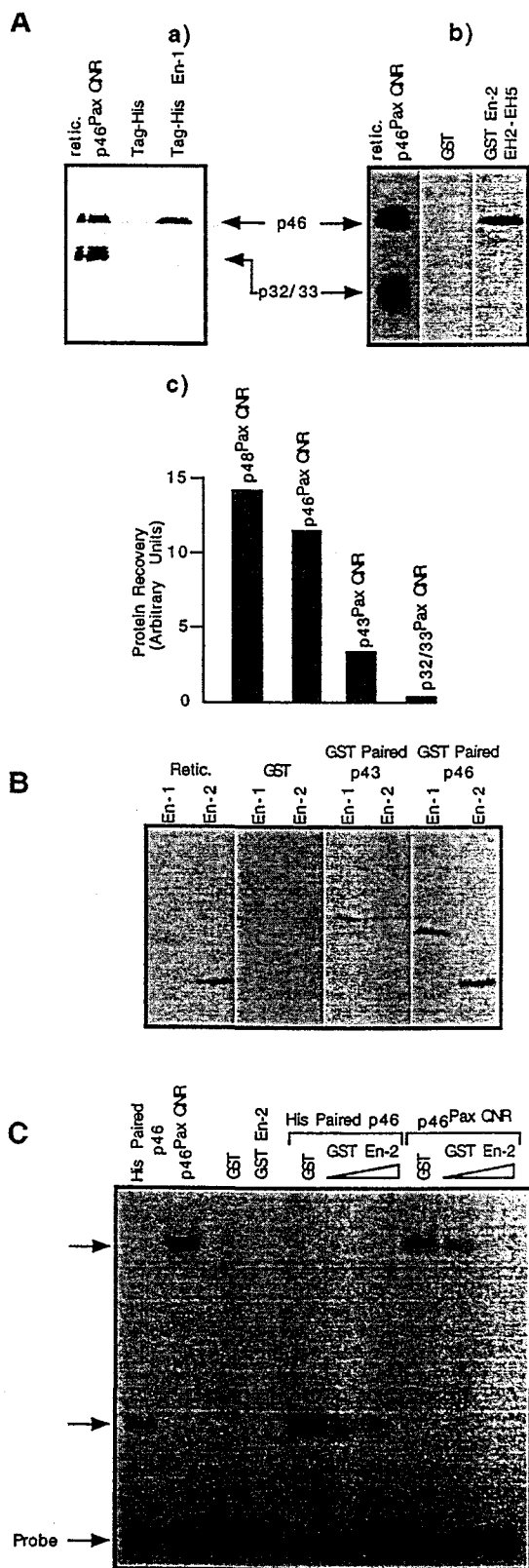


Fig. 5. Repression of Pax-6 mRNAs in RPE by the En-1  $\Delta$ HD. A, Northern blot analyses were performed with 20  $\mu$ g of total cellular RNA of cells transfected with either pCRNCM En-1  $\Delta$ HD or control DNA and selected by G418 treatment. Top, hybridization with the GAPDH probe (28); bottom, hybridization with Pax-6 probe (9). B, the signal generated by the hybridizations and indicated by an arrow (A) was measured with a PhosphorImager (Molecular Dynamics) and expressed as arbitrary units of RNA level after correction for the GAPDH signal.

but not the one of p43 (Fig. 6B), interacted with the En-1 and En-2 proteins synthesized in reticulocyte lysates, suggesting further that the critical domain of En/Pax-6 interaction lies in the paired exon 5 peptide.

Because coexpression of En products together with p46<sup>Pax-6</sup> activator resulted in marked reduction of the activity driven by the Pax-6 promoters (Fig. 4, C and D), we tested whether En binding to the p46 protein was able to modulate the DNA-binding activity of this Pax-6 product using the ability of p46 to bind an oligonucleotide bearing a high-affinity paired binding site, Pax6CON (38). As shown in Fig. 6C, when increasing amounts of GST-En-2 protein were added to either the reticulocyte-synthesized p46 or to the Tag-His-p46 paired domain, a marked decrease of binding to Pax6CON oligonucleotide was observed, and similar results were obtained with the P0 and P1 promoters fragments used as probe (data not shown). This DNA-binding inhibition of the paired domain resulted from protein interactions, because a



**Fig. 6.** A, p46<sup>Pax-6</sup>/En interaction *in vitro*. a), the Pax-6 cDNA encoding the p46 and p32/33 proteins (37) was translated in a rabbit reticulocyte lysate in the presence of L-[<sup>35</sup>S]methionine (left lane) and subjected to interaction with either Tag-His alone (middle lane) or Tag-His En-1 protein (right lane). Only the paired containing protein is recognized by En-1. b),

nonrelevant DNA-binding domain (such as ERM from the ets family of transcription factors or the c-myb R2R3 repeats of the DNA-binding domain) binding to their target DNA was not modified by the highest amount of GST-En-2EH2-EH5 protein used (data not shown).

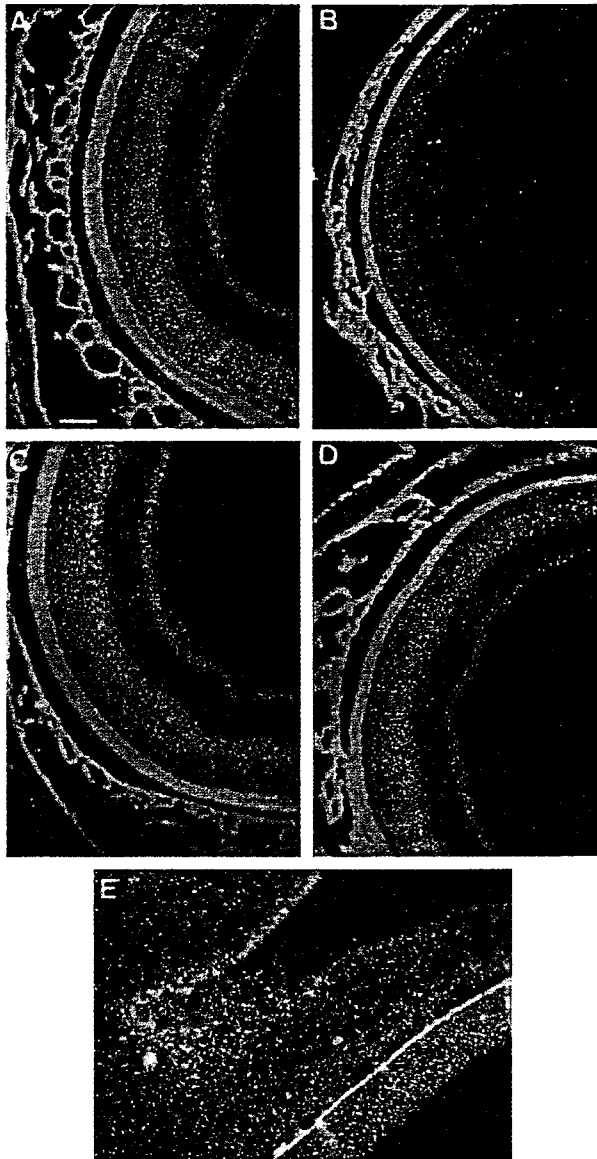
**En-1 Is Expressed in the Neuroretina of Hatched Animals.** The level of Pax-6 RNAs initiated from P0 and P1 declined when the neuroretina is dissected from hatched animals (28). Therefore, we performed *in situ* hybridization to study the possible En-1 expression in this tissue. Eyes from 6-day-old embryos or 3-day-old hatched quails were used. We performed serial sections of the retina hybridized with En-1 and Pax-6 as a control. As shown in Fig. 7E, En-1 was not expressed in the neuroretina of 6-day-old embryos. In contrast, in hatched animals, En-1 was expressed in both the ganglionic and inner nuclear layers of the retina (Fig. 7A), with the two cell layers expressing Pax-6; P0-initiated messengers (Fig. 7D); and P1-initiated transcripts (Fig. 7C). This En-1 signal was specific, because the sense probe used as a control gave no signal above the background (Fig. 7B). Taken together, all these results suggest that the Pax-6 down-regulation observed in the neuroretina after hatching is due to the inhibitory activity of the En-1 expression occurring in this tissue.

## Discussion

### Pax-6 Is Down-Regulated by the Chicken En Products.

We tested the possibility that Pax-6 was down-regulated by the En-1 product in the embryo central nervous system. Avian species express two genes (En-1 and En-2) related to the *Drosophila engrailed* (21, 29), and the *Drosophila En* is able to repress gene expression *in vitro* and *in vivo* (20, 39). During central nervous system development, Pax-6 and En-1 or En-2 are coexpressed in few neuronal populations, e.g., in the cerebellum, where the granular cell layer expressed En-1, En-2, and Pax-6 (40, 41). The possibility that Pax-6 could be under the control of En genes was suggested by transplantation experiments in avian embryos. Induction by isthmic transplant of an ectopic expression of En in the normal Pax-6

same experiment as in a), except that the reticulocyte lysate was subjected to interaction with either GST alone (middle lane) or GST-En-2EH2-EH5 protein (right lane). As with En-1, only the paired containing protein is recognized by En-2. c), interaction of Pax-6-translated proteins (p48, p46, p43, and p32/33) with GST-En-2EH2-EH5 protein. Equal amounts of proteins synthesized in reticulocyte lysates were used in GST precipitation assays. The amount of protein recovered with the GST alone was deduced from the amount of protein recovered with GST-En-2EH2-EH5. B, Pax-6 paired/En interaction *in vitro*. En-1 and En-2 proteins translated in a rabbit reticulocyte lysate in the presence of L-[<sup>35</sup>S]methionine (Retic. lane) were used in GST precipitation assays with GST alone (GST lane), GST-p43 paired domain (p43 lane), or GST-p46 paired domain (p46 lane). Note that the faint band observed with En-1 reticulocyte lysate in the GST-p43 paired domain did not correspond to En-1 protein. C, gel retardation assay of DNA oligonucleotide spanning the Pax-6 binding site (Pax6CON, described in Ref. 38) in the presence of increasing amounts of GST-En-2EH2-EH5 proteins. End-labeled DNA was incubated in the presence of histidine-tagged p46 paired domain, and p46 translated in reticulocyte lysate GST or GST-En-2EH2-EH5 proteins. The retarded DNA, only observed in the His paired and p46 lanes, was indicated by an arrow. Increasing amounts of GST-En-2EH2-EH5 protein (1 or 5  $\mu$ g) as indicated (top) were then added in the binding reaction. As control, 5  $\mu$ g of GST alone were also used without any effect (GST lanes).



**Fig. 7.** Detection of En-1 mRNAs by *in situ* hybridization in quail. **A**, hybridization of an eye transversal section from a 3-day-old hatched quail with a  $^{35}\text{S}$ -labeled En-1 antisense probe. Bar, 125  $\mu\text{m}$ . **B**, adjacent section hybridized with a  $^{35}\text{S}$ -labeled En-1 sense probe. **C**, hybridization of an eye transversal section from a 3-day-old hatched quail with a  $^{35}\text{S}$ -labeled Pax-6 P1 antisense probe. **D**, adjacent section hybridized with a Pax-6 P0 antisense probe (encompassing exon 0). **E**, hybridization of an embryo eye transversal section from a 6-day-old quail embryo with a  $^{35}\text{S}$ -labeled En-1 antisense probe. The retina is not labeled.

expression domain resulted in the concomitant repression of Pax-6 expression around the graft (12). Pax-6 is expressed from two promoters, P0 and P1, P0 being activated in the neuroretina at the onset of neuronal differentiation. The expression of the Pax-6 mRNAs declined strongly after hatching (28). The pigmented retina expresses Pax-6 through P1 only, and infection of RPE cells with a retroviral vector encoding the p44<sup>En-1</sup> resulted in a down-regulation of Pax-6 below a detectable level. Therefore, we conclude that Pax-6 is down-regulated *in vivo* by En-1.

#### Pax-6 Promoters Are Repressed by En through a DNA Binding-independent Mechanism.

The *Drosophila* En repression activity is mediated through the DNA binding of the homeodomain on target sequences; the target sequence could be the TATA box itself (18) or a sequence located some distance away from those bound by several activators (19, 20). We found three En-2 homeodomain binding sites in the P0 promoter. However, when these sites were mutagenized in a way suppressing the En-2 homeodomain binding, the p44<sup>En-1</sup> was still able to repress Pax-6-activated expression (data not shown), suggesting that a DNA binding-independent mechanism occurs in such a repressive event. An alternative explanation would be that the En-1 homeodomain recognizes a DNA sequence different from that of En-2. We do not favor this hypothesis, because (a) the En-1 and En-2 homeodomains are 92% identical; (b) two of the three foot-printed sites were also recognized by the divergent Pax-6 homeodomain; and (c) in mice, En-2 can rescue the En-1 mutant phenotype (24), suggesting that the two genes regulate an overlapping set of target genes. DNA binding-independent repressive activity was demonstrated further by the ability of a homeodomain-deleted En product to repress the activated Pax-6 promoters, both in transient transfection experiments, and infection of RPE cells with a retroviral vector encoding the En-1  $\Delta\text{HD}$  resulted in a down-regulation of Pax-6 mRNA. Pax-6 promoter repression is specific, because the thymidine kinase promoter bearing 4xGal4-binding sites is not repressed by En or homeodomain-deleted proteins upon activation by Gal4VP16. 4xGal4 thymidine kinase-activated promoter is instead repressed strongly by a Gal4En protein (data not shown), suggesting that both DNA binding-dependent and -independent mechanisms are involved in the En repressive activity. In the *Drosophila* system, the fushi tarazu (ftz) protein lacking the homeodomain DNA-binding region retains significant biological activity (42). In mice, through a DNA binding-independent mechanism, HOXD8 protein is able to inhibit the activating function of HOXD9 and HOXD10 on the HOX cross-talk region (43).

Although the mechanisms that promote transcriptional activation have been studied extensively, only few repressor proteins have been well characterized. These fall into two general classes: those that use a sequence-specific DNA-binding function to select target genes such as Krüppel (44), Engrailed (20), Even-skipped (45), or WT1 (46), and those that lack discernible DNA-binding activities and presumably use protein-protein interactions to direct target selection, such as Dr1 (47) NC1, NC2 (48, 49), Id (50), or I-POU (51). However, the distinction between DNA binding-dependent and -independent interactions may be artificial and linked to the experimental circumstances; in fact, Krüppel can repress transcription even in the absence of its cognate DNA-binding sites (52), and two homeodomain-bearing transcriptional repressors, Msx-1 (14) and SCIP (53), do not require homeodomain DNA-binding sites to repress. The relatively degenerate DNA recognition specificity displayed by homeodomains has led investigators to propose that homeodomains must act in conjunction with other proteins (54). Several examples of protein-protein interactions involving homeodomains have now been reported. The yeast  $\alpha 2$  homeodomain



interacts with Mcm 1, forming cooperative DNA-protein complexes (55). In vertebrates, the serum response factor bearing, like Mcm1, a MADS box, performs a cooperative interaction with the paired-type homeodomain of the Phox1 protein (56). Another theme is direct interaction between two homeodomains, as exemplified by the cooperative homodimer formation observed with the paired class homeodomains (57). Direct interactions have also been observed for the yeast homeodomain proteins  $\alpha 1$  and  $\alpha 2$  (58); the  $\alpha 1$ -related *Drosophila* extradenticle protein (*exd*); the homeotic gene products (33, 59); and the human HOXD8 and HOXD9 proteins (43). En has been shown to interact with *exd* (33), and the *Drosophila exd* has at least three vertebrates equivalent, *Pbx-1*, *Pbx-2*, and *Pbx-3*, *Pbx-1* being expressed in the central nervous system (60, 61). Our homeodomain-deleted En clones still contained five amino acids of the homeodomain helix 1, and this part of the homeodomain may be involved in the protein-protein contacts, as suggested for the HOXD8 homeodomain (43). Vertebrate En proteins interacted with the Pbx products through the two tryptophan residues in the EH2 domain (36). EH2 was required for cooperative DNA binding with Pbx (36), and we observed that part of the En-2 protein containing the EH2 domain was able to interact with the Pax-6 paired domain, most probably through amino acids encoded by the paired exon 5, resulting in a decreased DNA-binding ability of the p46<sup>Pax-6</sup> paired domain. Direct interaction between En and the p46<sup>Pax-6</sup> paired domain may explain the inhibitory effect of En on the p46-induced Pax-6 promoters activation, as well as the Pax-6 down-regulation observed in En-1-expressing RPE cells.

The mechanisms that promote repression are multiple. If some factors, such as Id or I-POU, repress transcription of specific genes by sequestering site-specific DNA-binding activators, other repressors affect transcription by directly recognizing the general transcription machinery, such as NC-1, NC-2 (48, 49) Dr1 (47), as well as the homeodomain-bearing Msx-1 (17) or Eve repressors (15). Repression by Msx-1 requires multiple protein regions that contain a high percentage of alanine, proline, and glycine residues (17) and resemble repressor domains of the *Drosophila* En (20). In addition, the amphipathic  $\alpha$ -helix located at the COOH terminus of Dr1, a TATA box binding protein-associated protein and inhibitor of class II gene transcription, is rich in glutamine and alanine residues and shares some homology with En repressor domains (47). Altogether, these data suggest that these regions may provide several unique surfaces for multiple protein-protein contact with the basal transcription machinery and may then negatively modulate transcription. The chicken En-1 and En-2 are only distantly related to the *Drosophila* En in their amino terminus (29) and in the domains involved in transcriptional repression (20). However, the chicken En-1 contains a glycine-, proline-, and alanine-rich NH<sub>2</sub>-terminal region, whereas the chicken En-2 exhibits a shorter alanine- and glycine-rich region (29). This difference may explain in part why En-2 is less efficient than En-1 in transcriptional repression of the Pax-6 promoters. Preliminary data suggest that at least two distinct negative domains exist in the chicken En-1 protein.

### The Biological Relevance of Pax-6 Repression by En.

During mouse brain development, *En-1* is first detected at the one-somite stage (62), whereas *Pax-6* is detected a few hours later, at the four-somite stage, as is *En-2* (8, 62). The transcriptional regulation of *Pax-6* is complex, and a fine tuning of its expression seems to be required for a correct embryo development, because *Pax-6* dosage is critical for normal eye and central nervous system development (6, 63). In the mouse embryo, *Pax-6* mRNA is expressed at E8.0 in a broad region of the head surface ectoderm covering the prosencephalon but not the hindbrain region (64), where *En-1* is expressed, and may prevent *Pax-6* expression in this part of the embryonic brain. The small eye (Sey) mice homozygous for mutation in the *Pax-6* gene exhibit a down-regulation of *Pax-6* expression in the tissue, giving rise to the lens placode, whereas at the same time, the *Pax-6* expression is maintained in the optic vesicle (64). This suggests that, whereas the initial induction of *Pax-6* may depend on tissue-specific activators (e.g., *c-myc* in the neuroretina; Ref. 31), its sustained expression in the ectoderm is dependent on its autoregulation (27, 28). Therefore, when the autoregulation is settled, to down-regulate the *Pax-6* expression, a specific inhibitor should be required. For example, in the neuroretina, a marked reduction in *Pax-6* expression is observed after hatching in quails (28), and at this time, we observed the expression of *En-1* in the cellular layers of the neuroretina where *Pax-6* is expressed. This suggests that *En-1* may be a negative regulator of *Pax-6*, specifying the patterning of its expression in the developing central nervous system before any segmentation and after the fully differentiated neuroretina has developed. In addition, because the *En-2* gene has been proposed to be controlled by Pax-2, Pax-5, and Pax-8 (65), our results suggest a more general regulatory loop between *En* and *Pax*, in which En may be involved in *Pax* gene regulation.

### Materials and Methods

**Cloning, DNA Sequencing, Construction of the En Chimeras, in Vitro Mutagenesis, and Vectors Used.** The En cDNAs were isolated from a chicken embryo cDNA library (supplied kindly by Jan Sap) using a fragment of the chicken En-2 cDNA as a probe (kindly provided by Diana Darnell). The En cDNAs were ligated into M13 vectors mp18 and Puc19. Partial nucleotide sequences were determined by the dideoxy sequencing method using an Applied Biosystems 370A automatic sequencer. The En-1 and En-2  $\Delta$ HDS were created by the deletion of the *Pst*I-*Eco*RI fragment. The GST-En-2EH2-EH5 was created by inserting the en-2 PCR fragment (from amino acids 154–285 and containing the EH2, EH3, EH4, and EH5 conserved boxes; see Ref. 29) generated to produce the bacterial expression vector (see below) inserted in the pGEX-4T3 (Pharmacia) vector. The GST-p46 paired and GST-p43 paired were created by inserting the relevant paired box PCR fragment (encoding from amino acids 2 to 134) using the oligonucleotides 5'-TCTGGATCCCAGAACAGTCACAGCGGAGTG-3' and 5'-TTCCTCGAGTCACTTTTCGTAGCCAGGTTG-3'. The amplified fragments were inserted in the pCRII vector DNA using an Invitrogen TA cloning kit. The recombinant DNA was then digested by *Bam*HI-*Xho*I, and the purified fragments were inserted into the pGEX-4T3 (GST fusion) or pQE30 (His-tag fusion). The Tag-His En-1 was produced by inserting the En-1 full-length coding sequence *Eco*RI-*Bam*HI fragment into the pQE30 expression vector.

**Cell Culture, Virus, and Transfection.** QNR cells and quail RPE cells, prepared from 7-day-old embryos, and quail embryo cells were seeded at 10<sup>6</sup> cells per six-well plate in DMEM-10% FCS 24 h prior to transfection. Transfections were performed in duplicate with two distinct DNA prepa-

rations by the calcium phosphate method, except for the COS-1 cells transfected by the lipofectamine reagent (Life Technologies, Inc.). For the experiments with the vector expressing the c-myc or Pax-6 proteins, cells were cotransfected with 1  $\mu$ g of the CAT construct and 1–5  $\mu$ g of expression vector. The total amount of transfected DNA was kept constant by the addition of empty vector DNA: pJ3 $\Omega$  for Pax-6 and pSG5 or N-Cla for c-myc.

The SFCV-LE En-1<sup>s</sup> and SFCV-LE En-1<sup>as</sup> viruses were created by the insertion of the 1.1-kb *EcoRI* fragment of the En-1 cDNA into the avian retroviral vector SFCV-LE (Ref. 30; provided kindly by Bjorn Vennström) digested by *EcoRI*. Because this vector was unable to allow expression of the En-1  $\Delta$ HD product, we instead used the pCRNCM vector (34) containing an internal cytomegalovirus promoter (provided kindly by Karl-Heinz Klempner). For stable transfection experiments, cells were transfected with 3  $\mu$ g of the molecular clone with the lipofectamine reagent, and the transfected cells selected in medium containing geneticin (400  $\mu$ g/ml) were passaged twice on gelatin-coated dishes before testing.

**Expression of En-2 Homeodomain Domains in Bacteria and Preparation of Rabbit Serum.** The appropriate fragment was obtained from En-2 plasmid as template in PCR experiments using oligonucleotides corresponding to sequences containing 5' (the *Bam*HI restriction site; ATCGCATTGGATCCAGATAGCTCCAGGCCGGCTCCAAAC) and 3' (the *Hind*III site; TGCATCTAAAGCTTACTACTTGGCGTCTTTGCGCGTGGT). The expected band subcloned in the pCRII (invitrogen) vector was sequenced, subjected to *Bam*HI-*Hind*III restriction enzyme digestion, and inserted between the *Bam*HI-*Hind*III sites of a pLC24-derived vector. The subcloning results in the inframe fusion of the desired sequence (from amino acids 154 up to 285) to the first 99 amino acids of the polymerase of phage MS2. For expression, plasmids grown in LE392 ( $\lambda$ ) were transferred into an *Escherichia coli* host (SG4044) that has a temperature-sensitive repressor of the PL promoter. Cultures of exponentially growing bacteria carrying the desired vector were induced at 42°C for 3 h. Bacterial pellets were washed twice, boiled in sample buffer, and electrophoresed in a 15% SDS-PAGE. Proteins were visualized by staining with Coomassie brilliant blue and cut from the gel. Rabbit antiserum ( $\alpha$ En) was prepared as published for Pax-6 (37).

**Cell Labeling, Immunoprecipitation, and Western Blotting.** Transfected COS-1 cells were incubated for 45 min in the presence of 100  $\mu$ Ci/ml L-[<sup>35</sup>S]methionine (Amersham Corp.; specific activity, 1000 Ci/mmol), lysed in radioimmunoprecipitation assay buffer, and immunoprecipitated with rabbit anti-En-2 homeodomain serum ( $\alpha$ En). Immunoprecipitated proteins were analyzed by SDS-PAGE followed by fluorography.

Western blots were performed with 10<sup>6</sup> cells boiled in loading buffer, subjected to SDS-PAGE, and transferred electrophoretically to Immobilon membranes (Millipore). Filters were treated for 60 min in blocking buffer (5% nonfat dry milk in PBS). Filters were then treated overnight in blocking buffer containing rabbit serum  $\alpha$ En diluted 1:200, followed by a 30-min washing in blocking buffer-0.1% Tween 20. Bound antibodies were revealed with an enhanced chemiluminescence detection kit (Amersham).

**RNA Extraction and RNase Protection Analysis.** Total RNA was extracted from transfected as well as normal RPE cells by the RNazol method (66). A 424-bp *EcoRI-HincII* fragment corresponding to the 5' part of the MC29-QNR2 cDNA and cloned into the pGEM4 vector was used to synthesize the antisense RNA probe used as described previously (67).

**Northern Blot Analysis.** Total RNA was extracted from cell cultures by using the guanidinium thiocyanate-cesium chloride method. RNAs were denatured at 68°C in a formamide-formaldehyde mixture, separated by electrophoresis in a 1% agarose-2.2 M formaldehyde gel, transferred to a nitrocellulose filter in 20 $\times$  SSC, and hybridized to DNA probes labeled by random priming. Pax-6 probe was the fragment containing the full open reading frame (9). The GAPDH probe was already described (28).

**Immunofluorescence Analysis.** Normal or transfected RPE cells cultured on collagen-coated 12-mm microscope coverslips were fixed for 20 min with 3.7% paraformaldehyde in PBS and then treated with rabbit anti-Pax-6 serum 11 or 14 (37). Bound antibodies were revealed using swine antirabbit FITC (DAKO S.A.).

**In Vitro Transcription.** To characterize the chicken En products, a fragment of cDNA En-1 (1.1-kb *EcoRI*) or En-2 (1.0-kb *EcoRI*) was cloned into the *EcoRI* site of the pSG5 vector. The sense RNA probe was transcribed *in vitro* from the plasmid using T7 polymerase according to the manufacturer's specifications (Promega).

**Protein Interaction in Vitro.** To construct GST fusions, En-2 and p46 or p43 paired boxes were amplified by PCR. These fragments were subcloned into pGEX (Pharmacia). GST fusions were expressed in *E. coli* and extracted according to Pharmacia recommendations. For interaction, 10  $\mu$ g of GST fusion proteins were incubated with 50  $\mu$ l of 50% slurry glutathione Sepharose 4B beads in incubation buffer [12 mM HEPES (pH 7.9), 4 mM Tris-HCl (pH 7.9), 50 mM NaCl, 10 mM KCl, 1 mM EDTA, 1 mM DTT, and 1 mM phenylmethylsulfonyl fluoride] for 20 min at room temperature, washed, and resuspended in a total volume of 30  $\mu$ l containing 10  $\mu$ l of L-[<sup>35</sup>S]methionine-labeled rabbit reticulocyte lysates in incubation buffer for 40 min in ice. Beads were then washed four times with 1 ml of washing buffer [0.5% NP40, 1 mM EDTA, 20 mM Tris-HCl (pH 8.0), and 0.1 M NaCl] at room temperature. Beads were then recovered in SDS-PAGE loading buffer, and proteins were analyzed by SDS-PAGE followed by fluorography.

**In Situ Hybridization.** A 918-bp *EcoRI-BglII* genomic fragment encompassing most of the En-1 coding sequence cloned into the pGEM3Z vector (Promega) was used to synthesize the RNA probe specific for the En-1 mRNAs. A 653-bp *XhoI-Asp718* genomic fragment encompassing exon 0 and cloned into the pGEM3 vector was used to synthesize the RNA probe specific for the P0-initiated mRNAs. To synthesize the RNA probe specific for the P1-initiated mRNAs, a 416-bp *SstII* genomic fragment (from -153 to +263) cloned into the pGEM3 vector was used (28). Sense and antisense RNA probes were prepared with <sup>35</sup>S-labeled CTP and used as described previously (10).

## Acknowledgments

We thank Patrick Martin, Karin Opdekamp, and Marion Wassef for critical reading of the manuscript.

## References

- Duboule, D. Guidebook to the Homeobox Genes. Sambrook and Tooze Publication. Oxford, United Kingdom: Oxford University Press, 1994.
- Tremblay, P., and Gruss, P. Pax: genes for mice and men. *Pharmacol. Ther.*, 61: 205–226, 1994.
- Walther, C., Guenet, J.-L., Simon, D., Deutch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R., and Gruss, P. Pax: a murine gene family of paired box containing genes. *Genomics*, 11: 424–434, 1991.
- Hastie, N. D. Pax in our time. *Curr. Biol.*, 1: 342–344, 1991.
- Hill, R. E., Favor, J., Hogan, B., Ton, C. C., Saunders, G., Hanson, I., Prosser, J., Jordan, T., Hastie, N., and Heyningen, V. V. Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature (Lond.)*, 354: 522–525, 1991.
- Schmal, W., Knoedseder, M., Favor, J., and Davidson, D. Defects of a neuronal migration and pathogenesis of cortical malformations are associated with small eye (sey) in the mouse, a point mutation at the *pax-6* locus. *Acta Neuropathol.*, 86: 126–135, 1993.
- Glaser, T., Waldon, D., and Maas, R. L. Genomic structure, evolutionary conservation and aniridia mutations in the human *pax-6* gene. *Nat. Gen.*, 2: 231–239, 1992.
- Walther, C., and Gruss, P. Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development (Camb.)*, 113: 1435–1449, 1991.
- Martin, P., Carrière, C., Dozier, C., Quatannens, B., Mirabel, M., Vandembunder, B., Stéhelin, D., and Saule, S. Characterization of a paired box- and homeobox-containing quail gene (*Pax-6*) expressed in the neuroretina. *Oncogene*, 7: 1721–1728, 1992.
- Turque, N., Plaza, S., Radvanyi, F., Carrière, C., and Saule, S. Pax-6, a paired box- and homeobox-containing gene expressed in neurons, is also expressed in pancreatic endocrine cells. *Mol. Endocrinol.*, 8: 929–938, 1994.
- Püschel, A., Gruss, P., and Westerfield, M. Sequence and expression pattern of *pax-6* are highly conserved between zebrafish and mice. *Development (Camb.)*, 114: 643–651, 1992.
- Bally-Cuif, L., and Wassef, M. Ectopic induction and reorganization of Wnt-1 expression in quail/chick chimeras. *Development (Camb.)*, 120: 3379–3394, 1994.

13. Han, K., Levine, M. S., and Manley, L. Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. *Cell*, **56**: 573-583, 1989.
14. Catron, K., Zhang, H., Marshall, S. C., Inostroza, J. A., Wilson, H., and Abate, C. Transcriptional repression by Msx-1 does not require homeo-domain DNA-binding sites. *Mol. Cell. Biol.*, **15**: 861-871, 1995.
15. Um, M., Li, C., and Manley, J. The transcriptional repressor Even-skipped interacts directly with TATA-binding protein. *Mol. Cell. Biol.*, **15**: 5007-5016, 1995.
16. Poole, S. J., Kauvar, L. M., Drees, B., and Kornberg, T. The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell*, **40**: 37-43, 1985.
17. Catron, K., Iler, N., and Abate, C. Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeo-domain proteins. *Mol. Cell. Biol.*, **13**: 2354-2365, 1993.
18. Ohkuma, Y., Horikoshi, M., Roeder, R. G., and Desplan, C. Engrailed, a homeodomain protein can repress *in vitro* transcription by competition with the TATA box-binding protein transcription factor IID. *Proc. Natl. Acad. Sci. USA*, **87**: 2289-2293, 1990.
19. Jaynes, J., and O'Farrell, P. Active repression of transcription by engrailed homeodomain protein. *EMBO J.*, **10**: 1427-1433, 1991.
20. Han, K., and Manley, J. Functional domains of the *Drosophila* Engrailed protein. *EMBO J.*, **12**: 2723-2733, 1993.
21. Davis, C., Holmyard, D., Millen, K., and Joyner, A. Examining pattern formation in mouse, chicken and frog embryos with an En-specific anti-serum. *Development (Camb.)*, **111**: 287-298, 1991.
22. Wurst, W., Auerbach, A. B., and Joyner, A. Multiple developmental defects in Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development (Camb.)*, **120**: 2065-2075, 1994.
23. Millen, J. K., Wurst, W., Herrup, K., and Joyner, A. Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development (Camb.)*, **120**: 695-706, 1994.
24. Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A. B., and Joyner, A. Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. *Science (Washington DC)*, **269**: 679-682, 1995.
25. Gardner, C. A., and Baral, K. F. The cellular environment controls the expression of engrailed-like protein in the cranial neuroepithelium of quail-chick chimeric embryos. *Development (Camb.)*, **113**: 1037-1048, 1991.
26. Martinez, S., Wassef, M., and Alvarado-Mallart, R-M. Induction of a mesencephalic phenotype in the 2 day-old chick prosencephalon is preceded by the early expression of the homeobox gene *en*. *Neuron*, **6**: 971-981, 1991.
27. Plaza, S., Dozier, C., and Saule, S. Quail Pax-6 (*Pax-QNR*) encodes a transcription factor able to bind and *trans*-activate its own promoter. *Cell Growth Differ.*, **4**: 1041-1050, 1993.
28. Plaza, S., Dozier, C., Turque, N., and Saule, S. The quail Pax-6(*Pax-6*) mRNAs are expressed from two promoters used differentially during retina development and neuronal differentiation. *Mol. Cell. Biol.*, **15**: 3344-3353, 1995.
29. Logan, C., Hanks, M. C., Noble-Topham, S., Nallainathan, D., Provart, N. J., and Joyner, A. L. Cloning and sequence comparison of the mouse, human, and chicken *engrailed* genes reveal potential functional domains and regulatory regions. *Dev. Genet.*, **13**: 345-358, 1992.
30. Fuerstenberg, S., Beug, H., Introna, M., Khazaie, K., Munoz, A., Ness, S., Nordström, K., Sap, J., Stanley, I., Zenke, M., and Vennström, B. Ectopic expression of the erythrocyte band 3 anion exchange protein, using a new avian retrovirus vector. *J. Virol.*, **64**: 5891-5902, 1990.
31. Plaza, S., Turque, N., Dozier, C., Bailly, M., and Saule, S. c-myb acts as transcriptional activator of the quail Pax-6(*Pax-6*) promoter through two different mechanisms. *Oncogene*, **10**: 329-340, 1995.
32. Kissinger, C., Liu, B., Martin-Blanco, E., Kornberg, T., and Pabo, C. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**: 579-590, 1990.
33. van Dijk, M. A., and Murre, C. extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell*, **78**: 617-624, 1994.
34. Foos, G., Grimm, S., and Klemppner, K-H. The chicken A-myb protein is a transcriptional activator. *Oncogene*, **9**: 2481-2488, 1994.
35. van Dijk, M. A., Peltenburg, L. T. C., and Murre, C. Hox gene products modulate the DNA binding activity of Pbx1 and Pbx2. *Mech. Dev.*, **52**: 99-108, 1995.
36. Peltenburg, L. T. C., and Murre, C. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.*, **15**: 3385-3393, 1996.
37. Carrière, C., Plaza, S., Martin, P., Quatannens, B., Bailly, M., Stéhelin, D., and Saule, S. Characterization of quail pax-6 (*Pax-6*) proteins expressed in the neuroretina. *Mol. Cell. Biol.*, **13**: 7257-7266, 1993.
38. Epstein, J., Cai, J., Glaser, T., Jepeal, L., and Maas, R. Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J. Biol. Chem.*, **269**: 8355-8361, 1994.
39. Eaton, S., and Kornberg, T. B. Repression of *ci-D* in posterior compartments of *Drosophila* by *engrailed*. *Genes Dev.*, **4**: 1068-1077, 1990.
40. Davis, C. A., and Joyner, A. L. Expression patterns of the homeo box-containing genes *En-1* and *En-2* and the proto-oncogene *Int-1* diverge during mouse development. *Genes Dev.*, **2**: 1736-1744, 1988.
41. Stoykova, A., and Gruss, P. Roles of Pax genes in developing and adult brain as suggested by expression patterns. *J. Neurosci.*, **14**: 1395-1412, 1994.
42. Fitzpatrick, V. D., Percival-Smith, A., Ingles, C. J., and Krause, H. M. Homeodomain-independent activity of the fushi tarazu polypeptide in *Drosophila* embryos. *Nature (Lond.)*, **356**: 610-612, 1992.
43. Zappavigna, V., Sartori, D., and Mavilio, F. Specificity of HOX protein function depends on a DNA-protein and protein-protein interactions both mediated by the homeo domain. *Genes Dev.*, **8**: 732-744, 1994.
44. Licht, J. D., Gossel, M. J., Figge, J., and Hansen, U. M. *Drosophila* Krüppel protein is a transcriptional repressor. *Nature (Lond.)*, **346**: 76-79, 1990.
45. Han, K., and Manley, J. Transcriptional repression by the *Drosophila* Even-skipped protein: definition of a minimal repression domain. *Genes Dev.*, **7**: 491-503, 1993.
46. Madden, S. L., Cook, D. M., and Rauscher, F. J., III. A structure-function analysis of transcriptional repression mediated by the WT1, Wilms' tumor suppressor protein. *Oncogene*, **8**: 1713-1720, 1993.
47. Inostroza, J. A., Mermelstein, F. H., Ha, I., Lane, W. S., and Reinberg, D. DR1: a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. *Cell*, **70**: 477-489, 1992.
48. Meisterernst, M., and Roeder, R. G. Family of proteins that interact with TFIID and regulate promoter activity. *Cell*, **67**: 557-567, 1991.
49. Meisterernst, M., Roy, A. L., Lieu, H. M., and Roeder, R. G. Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell*, **66**: 981-993, 1991.
50. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. The protein Id: a negative regulator of helix-loop-helix DNA-binding proteins. *Cell*, **61**: 49-59, 1990.
51. Treacy, M. N., Neilson, L. I., Turner, E. E., He, X., and Rosenfeld, M. G. Twin of I-POU: a two amino acid difference in the I-pou homeodomain distinguishes an activator from an inhibitor of transcription. *Cell*, **68**: 491-505, 1992.
52. Zuo, P., Stanojevic, D., Colgan, J., Han, K., Levine, M., and Manley, J. L. Activation and repression of transcription by the gap proteins *hunchback* and *Krüppel* in cultured *Drosophila* cells. *Genes Dev.*, **5**: 254-264, 1990.
53. Monuki, E., Kuhn, R., and Lemke, G. Repression of the myelin *Po* gene by the POU transcription factor SCIP. *Mech. Dev.*, **42**: 15-32, 1993.
54. Hayashi, S., and Scott, M. P. What determines the specificity of action of *Drosophila* homeodomain proteins? *Cell*, **63**: 883-894, 1990.
55. Smith, D. L., and Johnson, A. D. A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an  $\alpha 2$  dimer. *Cell*, **68**: 133-142, 1992.
56. Grueneberg, D., Simon, K. J., Brennan, K., and Gilman, M. Sequence-specific targeting of a nuclear signal transduction pathways by homeodomain proteins. *Mol. Cell. Biol.*, **15**: 3318-3326, 1995.

57. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N., and Desplan, C. Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.*, 7: 2120-2134, 1993.
58. Mak, A., and Johnson, A. D. The carboxy-terminus tail of the homeodomain protein *a2* is required for function with a second homeodomain protein. *Genes Dev.*, 7: 1862-1870, 1993.
59. Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. S. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with Extradenticle, another homeoprotein. *Cell*, 78: 603-615, 1994.
60. Monica, K., Galili, N., Nourse, J., Saltman, D., and Cleary, M. *PBX2* and *PBX3*, new homeobox genes with extensive homology to the human proto-oncogene *PBX1*. *Mol. Cell. Biol.*, 11: 6149-6157, 1991.
61. Rauskolb, C., Pfeifer, M., and Wesichaus, E. extradenticle, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene *pbx1*. *Cell*, 74: 1101-1112, 1993.
62. McMahon, A. P., Joyner, A. L., Bradley, A., and McMahon, A. The midbrain-hindbrain phenotype of *Wnt-1<sup>-/-</sup>/Wnt-1<sup>-/-</sup>* mice results from step-wise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell*, 69: 581-595, 1992.
63. Glaser, T., Jepeal, L., Edwards, J. G., Young, R., Favor, J., and Maas, R. L. *Pax-6* gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat. Genet.*, 7: 463-471, 1994.
64. Grindley, J. C., Davidson, D. R., and Hill, R. E. The role of *pax-6* in eye and nasal development. *Development (Camb.)*, 121: 1433-1442, 1995.
65. Song, D-L., Chalepakis, G., Gruss, P., and Joyner, A. L. Two Pax-binding sites are required for early embryonic brain expression of an *Engrailed-2* transgene. *Development (Camb.)*, 122: 627-635, 1996.
66. Chomczynski, P., and Sacchi, N. Single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, 162: 156-159, 1987.
67. Dozier, C., Carrière, C., Grévin, D., Martin, P., Quatannens, B., Stéhelin, D., and Saule, S. Structure and DNA-binding properties of *Pax-QNR*, a paired box- and homeobox-containing gene. *Cell Growth Differ.*, 4: 281-289, 1993.

***Le produit du gène En de poulet régule négativement Pax-6***

Nous avons vérifié l'hypothèse que le gène En-1 pouvait réguler négativement l'expression du gène Pax-6 dans le système nerveux central. Les deux gènes En-1 et En-2 sont les homologues aviaires du gène Engrailed de Drosophile (Davis *et al.*, 1991); (Logan *et al.*, 1992). Une caractéristique de ce gène est qu'il peut réprimer l'expression de gènes tant *in-vitro* que *in-vivo* (Han et Manley, 1993); (Eaton et Kornberg, 1990). Lors du développement du système nerveux central, le gène Pax-6 et En-1 ou En-2 sont co-exprimés dans des régions neuronales très spécifiques. Par exemple dans le cervelet les gènes En-1, En-2 et Pax-6 sont exprimés au niveau de l'assise granuleuse (Davis et Joyner, 1988); (Stoykova et Gruss, 1994). Pax-6 et les gènes Engrailed s'expriment dans des territoires mutuellement exclusifs pour le reste des vésicules cérébrales. La possibilité que le gène Pax-6 soit sous le contrôle des gènes En a été suggérée par des expériences de transplantation chez les oiseaux. Il a été observé qu'un transplant qui produisait une expression ectopique des gènes En induisait par ailleurs, une répression de l'expression de Pax-6 chez un embryon qui pourtant exprimait normalement les gènes Pax-6 à proximité du greffon (Bally et Wassef, 1994). L'expression du gène Pax-6 est gouvernée par les deux promoteurs P0 et P1. Le promoteur P0 est activé dans la neurorétine lors de la différenciation neuronale. La production des transcrits décroît fortement après l'éclosion (Plaza *et al.*, 1995a). Dans la rétine pigmentaire seul le promoteur P1 fonctionne, et l'infection de cellules de rétine pigmentaire par un vecteur rétroviral portant le gène En-1 produit une régulation négative du gène Pax-6 dont l'expression devient indécélable. Nous concluons donc que *in-vivo* le gène Pax-6 est régulé négativement par le gène En-1.

***Les gènes En répriment les promoteurs des gènes Pax-6 par un mécanisme indépendant de la fixation à l'ADN.***

Chez la drosophile la répression causée par En est réalisée par la liaison de l'homéodomaine à l'ADN. Les cibles pourraient être les séquences des boîtes TATA (Ohkuma *et al.*, 1990), 1990 #316] ou une séquence située à une certaine distance des séquences liant plusieurs activateurs (Jaynes et O'Farrell, 1991); (Han et Manley, 1993). Pour le promoteur P0 nous avons observé la présence de trois sites de fixation pour l'homéodomaine du produit du gène En-2. Cependant, les deux protéines En-1 et En-2 délétées de leur domaine homéo sont toujours capables de réprimer l'expression du gène

Pax-6 . Nous avons pu mettre en évidence le fait que les produits des gènes En répriment l'activité transcriptionnelle de Pax-6 par interaction avec le domaine paired de la protéine P46 qui active la transcription à partir des promoteurs P0 et P1. Cette interaction entre les protéines En et Pax-6, diminue l'affinité de la P46 pour ses séquences cibles.

Ces résultats suggèrent qu'il existe, pour les gènes En, deux types de mécanismes de répression de la transcription, soit un mécanisme dépendant de la fixation à l'ADN, comme un facteur de transcription classique, soit un mécanisme indépendant de la fixation à l'ADN. Le gène Fushi-tarazu (Ftz) de drosophile code une protéine qui ne possède pas d'homéodomaine et qui est donc incapable de se fixer à l'ADN. Cette protéine conserve néanmoins une activité biologique significative (Fitzpatrick *et al.*, 1992) ce qui montre bien que de tels mécanismes de régulation transcriptionnelle indépendants d'une fixation à l'ADN peuvent être importants au cours du développement embryonnaire.

### ***La régulation du gène Pax-6***

Nous avons mis en évidence un enhancer qui active la transcription à partir du promoteur P0 dans les cellules de neurorétine à partir de 6 à 7 jours de développement chez la caille (voir Article 1). Par ailleurs, il a été démontré que la protéine P46 de Pax-6 était capable de stimuler la transcription à partir des promoteurs P0 et P1 (Plaza *et al.*, 1993; Plaza *et al.*, 1995a). De plus, la protéine P75 c-myb active elle aussi la transcription à partir du promoteur P0 (Plaza *et al.*, 1995b). C-myb pourrait intervenir avant l'auto-induction du gène Pax-6 par sa protéine P46, comme un facteur d'initiation de la transcription de Pax-6. L'article 2 met l'accent sur les acteurs de la régulation négative de Pax-6, à l'origine de la chute d'expression des ARNm à l'éclosion dans la neurorétine et de leur répartition cloisonnée dans le système nerveux central. Les protéines Engrailed 1 et 2 interfèrent en effet, avec la protéine P46 et empêchent celle-ci de fixer sa séquence de reconnaissance et par conséquent, de transactiver les promoteurs P0 et P1.

*Article 3*

### Article 3

#### A quail particular pax-6 mRNA (pax-QNR)

leads to the increase  
of paired less proteins

M-C Langlois, P. Martin, C. Dozier et S. Saule.

En préparation.

Nous venons de voir qu'il existe une régulation complexe de la transcription du gène Pax-6 à partir des deux promoteurs P0 et P1. Néanmoins, la régulation des fonctions d'un gène passe aussi par d'autres étapes, comme la régulation post-transcriptionnelle et post-traductionnelle. Dans l'article qui suit, nous nous sommes intéressés à la régulation post-transcriptionnelle d'un ARNm particulier, dénommé ARNm B1, divergent des ARNm classiques du gène Pax-6 (comme ceux représentés par les ADNc 103 ou A1) par son extrémité 5'. L'ADNc B1 correspondant ne contient pas l'exon 3 retrouvé dans les ADNc 103 et A1 et en conséquence ne possède pas d'ATG avant le domaine paired. Le premier ATG de ce transcrit est localisé dans la région située entre le domaine paired et le domaine homéo. De plus, en amont de l'exon 4 dans la boîte paired, s'étend sur 500 bp un exon spécifique appelé exon  $\alpha$  qui possède un ACG en phase. Ce codon putatif de traduction est suivi d'une séquence en épingle à cheveux, l'ensemble formant une structure favorisant la fixation des ribosomes ainsi que celle du premier acide aminé lié à l'ARNt. Un tel ARNm pourrait engendrer une protéine Pax-6 pourvue en N-terminal d'une portion du domaine paired qui pourrait modifier la fonction de la protéine (voir la figure 1 de l'article). Les ARNm A1 et 103 traduisent tous deux des protéines identiques dont les plus grandes ne possèdent que trois acides aminés en amont du domaine paired. Par ailleurs, le premier acide aminé du domaine paired de la protéine B1 se trouverait être une glycine plutôt qu'une Sérine, ce qui pourrait également modifier les propriétés de la protéine.

Dans un premier temps, nous avons donc préparé des anticorps dirigés contre le peptide  $\alpha$  chez le lapin, pour chercher si l'ARNm B1 était capable d'initier une protéine à partir de l'exon  $\alpha$ . Cette démarche nous a amenés à la conclusion que l'ARNm B1 ne produit pas de protéine comportant d'exon  $\alpha$  *in vitro* ou dans la cellule et qu'il traduit essentiellement des protéines de 32 et 33 kDa. Ces petites protéines dépourvues de domaine paired sont également traduites à partir des ARNm A1 et 103 mais avec un



rendement moindre. En conséquence, la phase ouverte de lecture en 5' non traduite favorise la traduction à partir des AUG internes situés après la boîte paired. Ces protéines sans domaine paired sont incapables de fixer une sonde de type homéo, bien qu'elles possèdent le domaine correspondant de fixation à l'ADN. L'expression de ces protéines sans domaine paired est en corrélation avec les taux d'expression de l'ARNm B1. Par conséquent, le rôle de l'ARNm B1 pourrait bien être d'augmenter le taux d'expression des protéines sans domaine paired - une fonction non négligeable pour un gène tel que *Pax-6*, pour lequel une sous-expression ou au contraire une sur-expression entraîne des défauts du développement.

**Title :**

**A particular quail pax-6 mRNA (pax-QNR)  
leads to the increase  
of paired-less proteins**

Marie-Claire LANGLOIS, Patrick MARTIN, Christine Dozier and Simon SAULE  
EP 560 du CNRS, Institut de Biologie de Lille, Institut Pasteur de Lille, 1 rue  
Calmette, 59021 Lille Cedex, France.

**Keywords :** pax 6, alternative splicing, AUG codons, paired domain

Running title : Pax-6 isoform and pairedless proteins

## ABSTRACT

We report here the characterization of the product produced by a particular quail Pax-6 mRNA, called B1. The cDNA B1 presents an alternative exon in its 5' end when compared to the other transcripts generated from the Pax-6 gene. Surprisingly the B1 isoform mRNA does not possess an AUG initiation codon upstream of its paired box. We show here that this cDNA generate essentially paired-less proteins, with a better yield than other Pax-6 cDNAs. Moreover, the paired-less proteins show a similar distribution than the mRNA B1 in the embryo suggesting that the B1 isoform could be instrumental to produce specifically the paired-less protein concentration during development. The paired-less proteins have already been shown to be unable to bind on a e5 probe, which contained both homeo domain and paired domain binding sequences and which is recognized by normal Pax-6 proteins. More surprisingly, we show that the paired-less proteins are also unable to bind to a specific homeo domain probe although they possess all features of a transcription factor. These proteins might thus require a cofactor to recognize their target sequences or act on transcription through a DNA binding independent mechanism.

## INTRODUCTION

Homeobox proteins play a crucial role as transcriptional regulators to coordinate the precise temporal and spatial sequences of differentiation and growth (Duboule and Morata, 1994). Through homology with the paired box, 9 paired-box containing genes, the pax gene family, have been isolated in vertebrates (Walther and others, 1991).

Mutations in the Pax-6 gene have been associated with the mouse mutant small eye (Hill and others, 1991). The human Pax-6 gene is deleted or mutated in patients with aniridia (Glaser and others, 1992) and the *Drosophila* homologue of Pax-6 is responsible of the eyeless mutation (Quiring and others, 1994). The Pax-6 gene is expressed in the developing central nervous system, in the optic cup, lens and overlying epithelium, as well as in the neuronal layers of the retina (Walther and Gruss, 1991).

We have characterized *pax-QNR*, the quail homologue of the Pax-6 gene, defined its complete genomic organization and characterized two promoters

called P0 and P1 used differentially during retina development and neuronal differentiation (Martin and others, 1992, Dozier and others, 1993, Plaza and others, 1993; Plaza and others, 1995b). The quail Pax-6 (pax-QNR) gene produces several isoforms by alternative splicing events and differential use of the two different promoters (Fig. 1). Three proteins of 48, 46 and 43 kDa encoded by the quail Pax-6 gene and containing the paired domain were characterized both *in vitro* and *in vivo*. Furthermore, this gene also code for two proteins of 33 and 32 kDa which are devoid of the DNA-binding paired domain but who still contains the DNA binding homeo domain. All known isoforms contain exon 3 which includes the AUG initiation codon except the particular B1 isoform which presents an alternative exon in its 5' end that excludes the AUG-bearing exon 3. This exon is replaced by an AUG-less exon, called exon  $\alpha$ , as previously described. (Dozier and others, 1993). This raised the possibility that an alternative initiator codon located upstream the paired domain will be used to create an additional Pax-6 protein differing in its NH2 terminus. Of note, exon  $\alpha$  is part of a neuroretina specific enhancer element which governs the expression of transcripts from the P0 promoter and is structurally and functionally conserved in mouse (Fig. 1) (Plaza and others, 1995a).

In this paper, we show that the B1 isoform encodes mainly the small paired-less proteins of 32 and 33 kDa that are efficiently translated from this mRNA. We also provide evidences that the 5' exon,  $\alpha$ , does not code for a protein by using an alternative start codon since antibodies directed against the putative product of this exon do not recognize any product *in vitro* and *in vivo*. We observed that the paired-less proteins and the B1 mRNA show a similar tissue distribution, further suggesting that the B1 mRNA is instrumental for the synthesis of these proteins. Furthermore, we studied the DNA binding properties of these paired-less proteins and we show using various types of probes bearing homeo box and/or paired box binding sites that they are unable to bind alone to DNA. This suggest that these proteins may need a cofactor to bind to DNA or that they may act by a DNA binding independent mechanism. Finally, we show that a B1-like mRNA exists in mouse, further stressing its important role *in vivo*.

## MATERIAL AND METHODS

**Plasmids used.** All the pax 6 expression vectors and reporter plasmids, except the B1 isoform constructs were already described previously (Carrière and others, 1993 ;Dozier and others, 1993). The B1 AUG was constructed by PCR from the B1 (Dozier et al., 1993). The 5' AUG-containing oligonucleotide was as follow : CGAGAATTCCGTGCAGCTAGTT and the 3' oligonucleotide : CCCGTTGACGAACACCCCGCCG . The PCR fragment was cloned in the TA cloning vector PCR2.1™ (from Invitrogen) and subsequently cloned inside a B1 pSG5 construct predigested by the compatible enzymatic sites.

**Translation in reticulocytes lysates and immunoprecipitation.** 10 µg of plasmid DNA was transcribed and translated in a TNT rabbit reticulocyte lysate according to the manufacturer's specifications (Invitrogen). The immunoprecipitation procedure was carried out as described previously, using 1/10 of reticulocyte lysate (see Carriere and others, 1995).

**Immunofluorescence study.** Transfected cells (COS-1) cultured on 12-mm microscope cover slips were fixed for 20 min. with 4% paraformaldehyde in PBS, 78 hours after the transfection, and then treated with sera. (The immunofluorescence experiment was conducted as already described, cfCarriere and others, 1995)

**In situ hybridization.** We chose fragments in order to obtain a specific probe of each 5' quail alternative transcripts. Exon  $\alpha$  was cloned from a 452 bp EcoRV-PstI genomic fragment from the B1Nsi psG5 vector (containing a transcribed region of 392 bp) in the pGEM3 expression vector. Exon 0 was cloned from a 653 bp XhoI-Asp718 genomic fragment from the MC29QNR2 (containing a 108 bp transcribed region), and exon 1, from a 416 bp SstII genomic fragment (containing a 250 bp transcribed region), each one in the pGEM3 vector, as previously described (Queva and others, 1992) Sense and antisense RNA probes were prepared with <sup>35</sup>SCTP and used as described previously ( Queva and others, 1992 ).

**Cellular extracts preparation and gel mobility-shift assays.**  $5 \times 10^5$  COS-1 cells were transfected by 3 µg of DNA, using 18 µl lipofectamin (Gibco-BRL). Cells extracts were prepared essentially as described (Plaza and others, 1993). All buffers contains proteinase and phosphatase inhibitors. Binding reactions were performed for 20 min. on ice, using 1 ng radiolabelled DNA probe, 3 µg cellular extract in 20 µl 10% glycerol, 10 mM HEPES (pH 7.9), 30 mM KCl, 4 mM spermidine, 0.1 mM EDTA, 0.25 mM DTT, 1 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 7.2) and 3 µg poly dI-dC.

The probe was radiolabelled by kinasing the oligonucleotides with  $\gamma^{32}$ P ATP.

- probe P2 : 5' GATCCTGTACAGTAATCAATTAGACTCAG

## RESULTS

### *Characterization of B1-encoded proteins*

In order to check if the B1 cDNA was able to produce a protein from its exon  $\alpha$  which is devoid of AUG codon, we produced an antibody against the peptide  $\alpha$ . As shown in Figure 2A, the B1 cDNA is able to drive the synthesis of paired-containing and paired-less proteins (respectively at 43-45 kDa and 32-33 kDa) in a programmed reticulocyte lysate, as well as in transfected COS cells as shown in Figure 2B. All these proteins can be immunoprecipitated with the serum 14 which recognize the C-terminal part of the Pax-6 proteins whereas only the paired-containing proteins can be immunoprecipitated with the serum 11 which recognize the Pax-6 Paired domain (Carrière and others, 1993). These proteins were also efficiently produced in transfected COS-1 cells as shown by immunofluorescence experiments (Figure 2B, panel a and b). However, no protein produced from the B1 cDNA could be recognized with the B1 specific serum ( $\alpha$ 1 S3) by immunoprecipitation (Figure 2A) or by immunofluorescence experiments of transiently transfected COS-1 cells conducted with the same serum (Figure 2B, panel c). On line with these findings, these  $\alpha$ -specific proteins were not observed *in vivo* by immunoprecipitation of neuroretina lysate in which the B1 cDNA is expressed (data not shown). To ascertain the efficiency of the  $\alpha$ -specific serum we generate an  $\alpha$  B1 construct in which an ATG was introduced upstream of the  $\alpha$  open reading frame (B1 AUG construct, Figure 2A). Using this construct in reticulocyte lysate, the B1 specific immunoserum was able to recognize specifically a protein of ca. 60 kDa. This protein was also observed in COS-1 transfected cells (Figure 2B, panel f). These results show that in these conditions no protein could be translated from the exon  $\alpha$  of the cDNA B1 and the corresponding mRNA.

Interestingly, we could detect proteins translated from the B1 isoform, some 43-45 kDa proteins containing a Paired domain since they are recognized by the serum 11. These proteins might be generated by an alternative initiation codon present upstream the paired domain since there is no in frame AUG in this region. This unusual translation initiation could explain why these paired containing proteins of 43-45 kDa B1 are produced in such a small amount. Several authors have shown that, although this kind of translation initiation is

frequently used in eukaryotic cells, it is a relatively inefficient system ( and Kozak, 1996). This alternative initiation codon is likely to be present inside the paired domain since these proteins are just a little smaller than regular paired-containing proteins generated from a classical Pax-6 cDNA such as MC29QNR2 (Figure 2A, right part). These 43-45 kDa proteins are produced in smaller amount than the products of 32-33 kDa which are devoid of paired domain. The 32-33 kDa paired-less proteins have already been shown to be translated from three internal AUG located between the paired and the homeo domains (Carrière and others, 1993). These proteins are present in a much higher amount than the 43-45 kDa paired containing proteins, an observation that is thus not surprising since the former are initiated from a non AUG codon. Consequently, the B1 cDNA translate essentially Pax-6 paired-less proteins. Of note, the B1 AUG construct produced, although at a relatively low level, essentially proteins originated from the artificially introduced AUG. In that case, probably because of competition for ribosome binding the paired-less proteins are inefficiently translated. The MC29QNR2 cDNA produced paired domain containing proteins much more efficiently than the paired less proteins as already observed (Carrière et al., 1993). Taken together these data show that the B1 mRNA generates principally paired-less products and moreover that this cDNA may be *in vivo* a specific source for these proteins.

#### *Are the paired-less able to bind DNA?*

Previous studies have shown that the most frequent DNA target sequence of the paired type homeo domain proteins is the palindromic sequence P2 which contained two inverted TAAT motifs separated by two nucleotides. *In vitro*, the paired type homeo domain alone binds preferentially a P2 target sequence and less often a P3 target sequence, which differs from P2 in that it contains an additional nucleotide in the spacer located between the two TAAT motifs (Wilson and others, 1993). The paired-less proteins possess only their homeo domain as a DNA binding domain. We thus studied their ability to bind the P2 sequence in gel shift assay with as a source of proteins extracts from COS-1 cell transiently transfected with the various cDNAs. The 46 kDa protein which is the regular product of the MC29QNR2 cDNA is able bind the P2 probe and this

binding is specific as shown by supershift and competition experiments (Figure 3 and data not shown). In sharp contrast, no binding was observed when the B1 or the B1 AUG cDNA were used to transfect COS-1 cells showing that the paired-less proteins do not bind the P2 probe. Since this absence of binding could be due to the presence in the B1 cDNA of an insertion of 6 amino acids in the exon 10, due to an alternative splicing event we tested whether the inclusion of these 6 amino acid modify the binding of the 46 kDa protein on the P2 probe. As expected, the protein translated from the MC29QNR2\* cDNA (possessing these additional 6 amino acids) is able to bind the P2 sequence. The reverse experiment confirms this notion, since the  $\Delta$ Nhe MC29QNR2 construct, a mutant of the MC29QNR2 mutant that expressed only the paired-less protein is also unable to yield protein binding the P2 probe (Figure 3, Carrière and others, 1993). This confirms the specific inability of the paired-less proteins to recognize the P2 probe. Previous studies have shown that these paired-less proteins are not able to bind the e5 target sequence, a specific target of the 46 kDa protein derived from the *Drosophila* even-skipped gene (Dozier and others, 1993; Carrière and others, 1993; Plaza and others, 1993).

*The level of the B1 mRNA and of the paired-less proteins are correlated in vivo.*

The paired-less proteins are highly expressed in the neuroretina and pancreas whereas they are absent in the neural tube (Dozier and others, 1993 ; Turque and others, 1994). Interestingly, the amount of paired-less proteins follows the fluctuations of the mRNA B1 expression level suggesting that indeed the B1 mRNA is instrumental for the synthesis of the paired-less proteins. By in situ hybridization we observed that B1 transcripts are effectively highly expressed in the neuroretina whereas they can be weakly detected in the neural tube and in the brain (figure 4).

By studying the enhancer which governs the neuronal specific expression of transcripts derived from the P0 promoter we previously noticed that the sequence of this enhancer, and thus the sequence of exon a, is conserved between quail and mouse (Plaza and others, 1995a). This prompted us to test if the mRNA B1 itself can be also present in mouse. Using a mouse probe corresponding to the enhancer conserved between mouse and quail and thus putatively corresponding



to a mouse homologue of exon  $\alpha$  we detect by *in situ* hybridization a specific expression of an  $\alpha$ -containing transcript in mouse eyes. This signal was observed in lens, in cornea *i.e.* in the same territories than the B1 mRNA in quail as well as in the same territories than the Pax-6 mRNAs in mouse (Dozier and others, 1993 ; Walther and Gruss, 1991). Consistent with the expression of the B1 like mRNA in mouse we detected paired-less proteins in the mouse neuroretina by immunoblotting with the serum 11 (data not shown).

## DISCUSSION

### *The B1 mRNA and the translation of the paired-less proteins.*

In this paper, we have shown that the B1 mRNA B1 translates essentially paired-less proteins of apparent molecular weight between 32 and 33 kDa. These proteins are initiated from an internal AUG which is located downstream of the paired domain. By inspecting the sequence we found that there are three successive AUG lying on 36 pb, 219 pb upstream of the homeobox. Two other Pax-6 cDNA, MC29QNR2 and A1, containing a whole paired box but exhibiting divergent 5'-UTR sequences, translate both paired-containing and paired-less proteins (Carrière and others, 1993). However, the translation of paired-less proteins is more efficient from the B1 cDNA than from the other *Pax-6* cDNA suggesting that this cDNA is instrumental for the production of paired-less proteins *in vivo*.

By studying a B1 cDNA for which an AUG codon was inserted upstream of the 5' ORF of exon  $\alpha$ , we noticed that such an insertion result in a weaker translation efficiently of the paired-less proteins (compare in figure 2A, "retic B1" with "retic B1 AUG"). Consequently, the long 5'UTR containing the open reading frame of the cDNA B1 seems to be essential for a high translation yield of the paired-less proteins. The presence in the upstream coding region of an initiation codon weakly recognized by the transcriptional machinery, like an AUG surrounded by poor consensus sequences for translation initiation, or an alternative translation initiation codon, enhance a "leaky scanning" for correct ATG by the pre-initiation translation complex (reviewed by Geballe and Morris, 1994). The B1 cDNA shows a comparable profile with a long open reading frame

in its 5'-end, an exon  $\alpha$  with an ACG in frame but unable to translate the protein translation. In addition, the B1 cDNA bears in its paired box additional alternative translation initiation codons, such as several GTG and a CTG which may be able to initiate a weak translation of protein products at least *in vitro* (Kozak, 1990). Nevertheless, this B1-paired-containing proteins are not found *in vivo*. Thus, the long open reading frame of the cDNA B1 seems to enhance the slipping of the pre-initiation translation complexes on the alternative translation codon located upstream of the coding sequence as suggested in the "leaky scanning" model..

*The amount of B1 mRNA correlates in vivo with the amount of paired-less proteins.*

In quail, the paired-less proteins are not detected in the neural tube (Dozier and others, 1993). Similarly, the B1 transcripts are poorly expressed in the brain and the neural tube. Thus, the paired-less proteins level appears to follow the mRNA B1 expression level *in vivo*. Homologues of B1 mRNA have not be isolated in other organisms. Nevertheless, we demonstrate in this paper that its expression is not quail specific, since it could be detected, by a mouse specific exon  $\alpha$  probe, in the mouse eye. The signal detected by *in situ* hybridization, is localized in the same territories as those expressing the B1 and MC29QNR2 mRNAs in quail and MC29QNR2 mRNA in mouse (Walther and Gruss, 1991). Moreover, the paired-less proteins are also present in the mouse neuroretina consistent with the notion that B1 mRNA are mainly responsible for the production of these proteins *in vivo*.

The correlation between the B1 mRNA expression level and the level of the paired-less proteins *in vivo*, suggests a determinant role of this particular mRNA in the production and regulation of the level of paired-less proteins *in vivo*. It is important in that respect to remember that the amount of a given pax protein product may be of critical importance during embryonic development. This notion of haplo-insuffisance of Pax gene family products is well illustrated by the developmental defects induced by the mutation of only one allele observed for several pax genes including pax 6 (for review, see Walther and others, 1991). It has been shown that the Pax-6 surexpression could induce defects in eye development (Schedl and others, 1996). Thus, an abnormal expression level of a

Pax gene product can induce developmental defects. Consequently, the level of the Pax proteins have to be very well regulated into the cells *in vivo*. The mRNA B1 could participate to this regulation by translating a high concentration paired-less proteins in specific regions.

### *What can be the role of paired-less proteins?*

We show in this paper that in contrast to the regular 46 kDa protein, the paired less proteins are unable to bind the P2 probe nor they can bind the e5 sequence. The inability of the homeo domain in the context of the paired-less proteins, to bind its recognition sequence (whereas it is able to bind this P2 sequence as an homeo domain peptide alone, data not shown), could be due to a conformational change in these small proteins that would hide their binding domain.

Even if they are unable to bind known DNA sequences, it is likely that the paired-less proteins containing a nuclear localization signal, a DNA-binding domain of the homeo type, and a transactivation domain are functional *in vivo* (Carriere and others, 1995). This is also suggested by the fact that paired-less proteins exist for Pax-6 genes in other species. For example, paired-less proteins have been found in *mab-18*, the Pax-6 homologue gene in *Caenorhabditis elegans* (Zhang and Emmons, 1995). In this species, the paired-less proteins play a determinant role in the specification of sensory organs.

The paired-less proteins can work in a number of ways *in vivo*. For example it is possible to speculate that they need a specific co-factor in order to bind DNA. Such a situation is known for example in the case of the *Drosophila* gene *Dorsal*, a member of the Rel/NF $\kappa$ B family. *Dorsal* is a morphogen which specify various territories such as mesoderm and dorsal ectoderm, in the embryos by forming an expression gradient. The *Dorsal* target genes present variable affinities for the morphogen depending on the territory in which they are activated. Interestingly, the *Dorsal* protein can interact with a cofactor of the bHLH class to induce the transactivation of a target gene bearing a weak affinity site (Jiang and others, 1992). It is possible to speculate that the paired-less proteins could act in a same manner : (i) their titer is variable following the various Pax-6 expressing regions ; (ii) they are not able to bind a classic homeo domain DNA

target and need probably a cofactor to interact efficiently with their recognition sequence. Other homeo domain proteins such as Engrailed or Hox proteins interact with co-factors like Pbx, the homologue of the *Drosophila extradenticle (exd)* factor.

It is also possible to propose that paired-less proteins can act in transcription without any efficient DNA binding activity. Numerous examples of non-DNA binding effects of transcription factors were described in the literature. Recently the glucocorticoid receptor has been shown *in vivo*, by knock out experiments to have an important part of its biological function that as not need specific DNA binding (Reichardt and others, 1998). Many mechanisms such as heterodimerization or interaction with transcriptional co-factors or factors belonging to the transcriptional apparatus may explain such effects.

Whereas the precise mechanisms by which the paired-less proteins exert their action is not known our analysis have clearly emphasized the importance of these products which are conserved and the expression of which appears to be tightly regulated. More work will be needed to decipher their precise function and mode of action.

## ACKNOWLEDGMENTS

This work was supported by CNRS, Institut Pasteur de Lille, Région Nord Pas de Calais and the Association contre le Cancer (ARC).

We thanks Agnès Bégue for technical assistance, Serge Plaza for helpful discussions and Vincent Laudet for critical reading of the manuscript.

## REFERENCES

- Carriere, C., S. Plaza, J. Caboche, C. Dozier, M. Bailly, P. Martin, and S. Saule. "Nuclear localization signals, DNA binding, and transactivation properties of quail Pax-6 (Pax-QNR) isoforms." *Cell Growth Differ* **6** (12 1995): 1531-40.
- Carrière, C., S. Plaza, P. Martin, B. Quatannens, M. Bailly, D. Stehelin, and S. Saule. "Characterization of quail Pax-6 (Pax-QNR) proteins expressed in the neuroretina." *Mol Cell Biol* **13** (12 1993): 7257-66.
- Dozier, C., C. Carriere, D. Grevin, P. Martin, B. Quatannens, D. Stehelin, and S. Saule. "Structure and DNA-binding properties of Pax-QNR, a paired box- and homeobox-containing gene." *Cell Growth Differ* **4** (4 1993): 281-9.
- Duboule, D. and G. Morata. "Colinearity and functional hierarchy among genes of the homeotic complexes." *Trends Genet* **10** (10 1994): 358-64.
- Geballe, A. P. and D. R. Morris. "Initiation codons within 5'-leaders of mRNAs as regulators of translation." *Trends Biochem Sci* **19** (4 1994): 159-64.
- Glaser, T., D. S. Walton, and R. L. Maas. "Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene." *Nat Genet* **2** (3 1992): 232-9.
- Hill, R. E., J. Favor, B. L. Hogan, C. C. Ton, G. F. Saunders, I. M. Hanson, J. Prosser, T. Jordan, N. D. Hastie, and Heyning V. van. "Mouse small eye results from mutations in a paired-like homeobox- containing gene [published erratum appears in Nature 1992 Feb 20;355(6362):750]." *Nature* **354** (6354 1991): 522-5.
- Jiang, J., C. A. Rushlow, Q. Zhou, S. Small, and M. Levine. "Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo." *Embo J* **11** (8 1992): 3147-54.
- Kozak, M. "Evaluation of the fidelity of initiation of translation in reticulocyte lysates from commercial sources." *Nucleic Acids Res* **18** (9 1990): 2828.
- Kozak, M. "Interpreting cDNA sequences: some insights from studies on translation." *Mamm Genome* **7** (8 1996): 563-74.
- Martin, P., C. Carriere, C. Dozier, B. Quatannens, M. A. Mirabel, B. Vandenbunder, D. Stehelin, and S. Saule. "Characterization of a paired box- and homeobox-containing quail gene (Pax-QNR) expressed in the neuroretina." *Oncogene* **7** (9 1992): 1721-8.
- Plaza, S., C. Dozier, M. C. Langlois, and S. Saule. "Identification and characterization of a neuroretina-specific enhancer element in the quail Pax-6 (Pax-QNR) gene." *Mol Cell Biol* **15** (2 1995a): 892-903.
- Plaza, S., C. Dozier, and S. Saule. "Quail Pax-6 (Pax-QNR) encodes a transcription factor able to bind and trans-activate its own promoter." *Cell Growth Differ* **4** (12 1993): 1041-50.
- Plaza, S., C. Dozier, N. Turque, and S. Saule. "Quail Pax-6 (Pax-QNR) mRNAs are expressed from two promoters used differentially during retina development and neuronal differentiation." *Mol Cell Biol* **15** (6 1995b): 3344-53.
- Queva, C., S. A. Ness, F. A. Grasser, T. Graf, B. Vandenbunder, and D. Stehelin. "Expression patterns of c-myb and of v-myb induced myeloid-1 (mim-1) gene during the development of the chick embryo." *Development* **114** (1 1992): 125-33.

Quiring, R., U. Walldorf, U. Kloter, and W. J. Gehring. "Homology of the *eyeless* gene of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans." *Science* **265** (1994): 785-789.

Reichardt, H. M., K. H. Kaestner, J. Tuckermann, O. Kretz, O. Wessely, R. Bock, P. Gass, W. Schmid, P. Herrlich, P. Angel, and G. Schutz. "DNA binding of the glucocorticoid receptor is not essential for survival [see comments]." *Cell* **93** (4 1998): 531-41.

Schedl, A., A. Ross, M. Lee, D. Engelkamp, P. Rashbass, Heyningen V. van, and N. D. Hastie. "Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities." *Cell* **86** (1 1996): 71-82.

Turque, N., S. Plaza, F. Radvanyi, C. Carriere, and S. Saule. "Pax-QNR/Pax-6, a paired box- and homeobox-containing gene expressed in neurons, is also expressed in pancreatic endocrine cells." *Mol Endocrinol* **8** (7 1994): 929-38.

Walther, C. and P. Gruss. "Pax-6, a murine paired box gene, is expressed in the developing CNS." *Development* **113** (4 1991): 1435-49.

Walther, C., J. L. Guenet, D. Simon, U. Deutsch, B. Jostes, M. D. Goulding, D. Plachov, R. Balling, and P. Gruss. "Pax: a murine multigene family of paired box-containing genes." *Genomics* **11** (2 1991): 424-34.

Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan. "Cooperative dimerization of paired class homeo domains on DNA." *Genes Dev* **7** (11 1993): 2120-34.

Zhang, Y. and S. W. Emmons. "Specification of sense-organ identity by a *Caenorhabditis elegans* Pax-6 homologue." *Nature* **377** (6544 1995): 55-9.

## FIGURES LEGENDS

**Figure 1 :** Map of the Pax-6 locus showing the various isoforms including the B1 mRNA originating in exon  $\alpha$ .

**Figure 2 :** Proteins translated from the B1 cDNA. (A) Immunoprecipitation of proteins produced by the B1 cDNA introduced in rabbit reticulocytes lysate transcription/translation system followed by SDS-PAGE electrophoresis. Several constructs were used for protein production : B1 : the regular B1 isoform, B1 AUG : a B1 cDNA in which an AUG codon was added in order to generate high amount of proteins and MC29QNR2, the exon 0 containing mRNA Pax-6 (cf fig1). The proteins produced in the reticulocytes lysates labeled with  $^{35}\text{S}$  Met are visualized (retch). Proteins were immunoprecipitated with either the 14, 11 or a1 S3 antibodies which recognized respectively the C terminal region (behind the homeo domain), the paired domain and the B1 specific a domain. Pi a1 : immunoprecipitation with the pre-immune serum corresponding to the a1 S3 serum ; ad a1 : immunoprecipitation with the a1 S3 serum previously adsorbed on a large excess of a peptide. (B) Immunofluorescence of COS-1 transiently transfected cells. Cells were transfected with either the B1 (panels a-c) or the B1 AUG (panel d-f) cDNAs and the immunofluorescence was conducted with respectively serum 11 (a, d), 14 (b, e) and a1 S3 (c, f).

**Figure 3 :** Electrophoretic mobility shift assay showing the DNA binding ability of the paired-less proteins. COS-1 cells were transiently transfected with control vector (-), MC29QNR2 cDNA, MC29QNR2\* cDNA (a MC29QNR2 cDNA containing 6 additional amino acids in exon 10 as the B1 cDNA), B1, B1 AUG or the DNhe MC29QNR2 construct and total cell lysates were prepared. MESA were performed using a  $^{32}\text{P}$  labeled P2 probe, the sequence of which is indicated on the bottom of the figure. The retarded complexes were supershifted with serum 14 (Ab 14).

### **Figure 4 : The expression of the mRNA B1 fluctuates with the territories of Pax-6 expression**

The mRNA B1 is highly expressed in the neuroretina and shows a weaker expression in the neural tube and in the brain. In contrast MC29QNR2 mRNA is highly expressed in the neural tube and the brain and its signal is weaker in the neuroretina.

a) 3,5 days quail embryos mapped with A) Pax-6 MC29QNR2 mRNA antisense probe; B) Pax-6 MC29QNR2 mRNA sense probe; C) Pax-6 B1 mRNA antisense probe; D) Pax-6 B1 mRNA sense probe. b) 9 days quail embryo neuroretina mapped with E) Pax-6 B1 mRNA sense probe; F) Pax-6 B1 mRNA antisense probe; G) Pax-6 MC29QNR2 mRNA sense probe; H) Pax-6 MC29QNR2 mRNA antisense probe I) telencephal; 2) diencephal; 3) mesencephal;

4) metencephal; 5) myelencephal; 6) neural tube; 7) pigmentosa neuroretina; 8) nuclear extern layer; 9) nuclear inter layer; 10) multipolar neurons.

**Figure 5 : mRNA containing the exon a are detected in the mouse eye**

*in situ* hybridization with a mouse specific exon a probe

- a) transversal section of a mouse eye showing an expression localized in the lens, the cornea, the ciliary corpus and the neuroretina.
- b) higher magnification of mouse neuroretina showing a similar expression with mouse a probe than with B1 or MC29QNR2 quail Pax-6 signal on quail neuroretina.



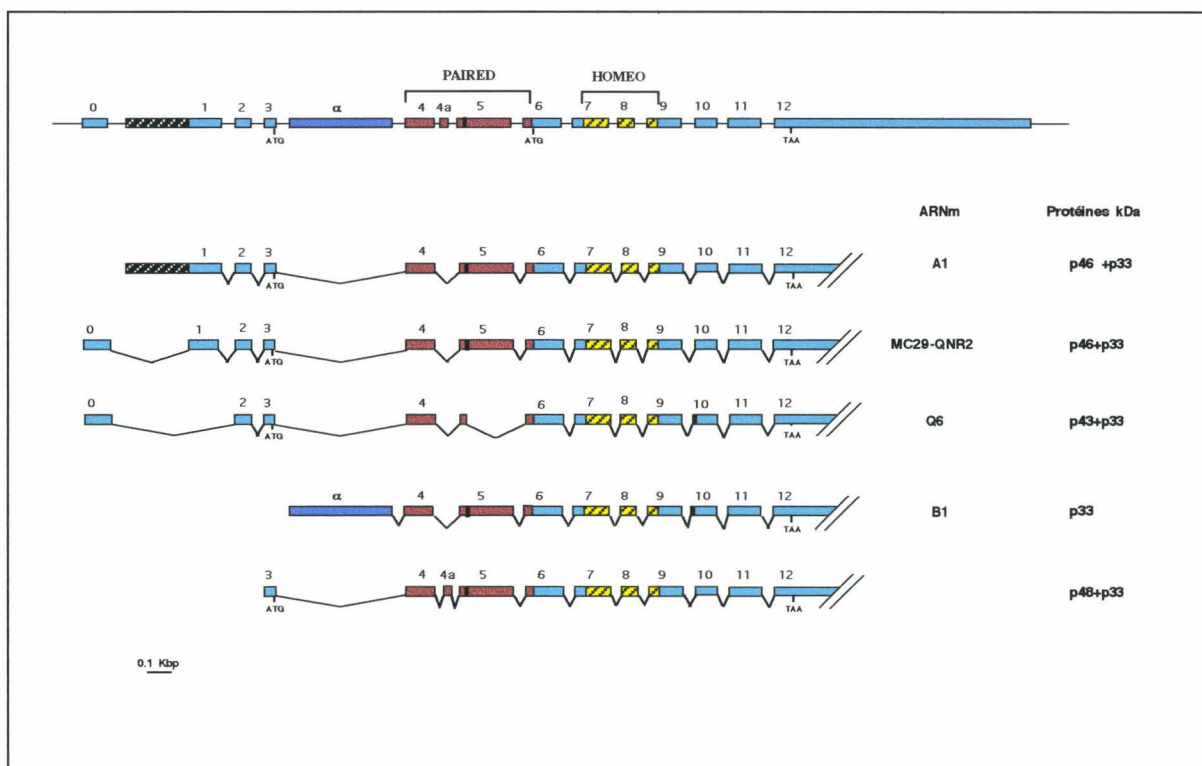
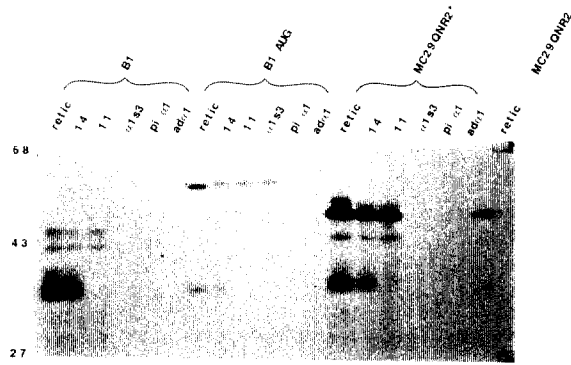
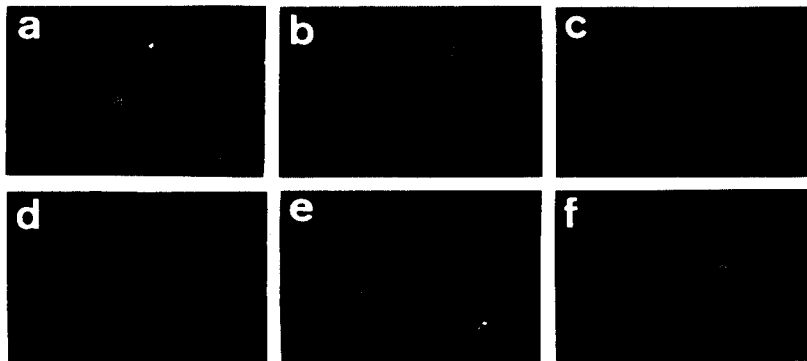


Figure 1: Map of the Pax-6 locus showing the various isoforms including the B1 mRNA originating in exon  $\alpha$ .

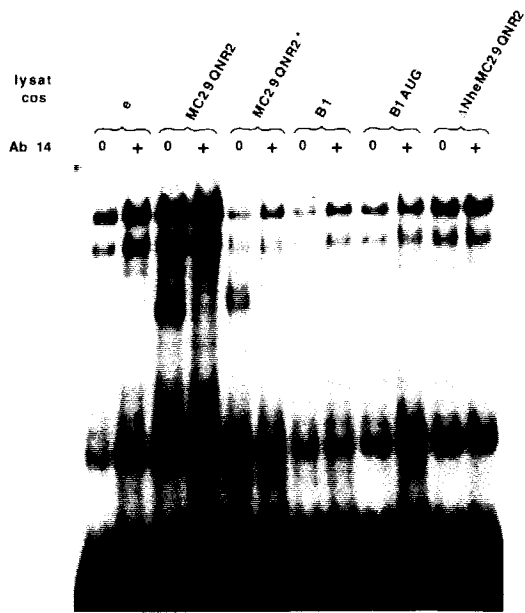




**Figure 2A**



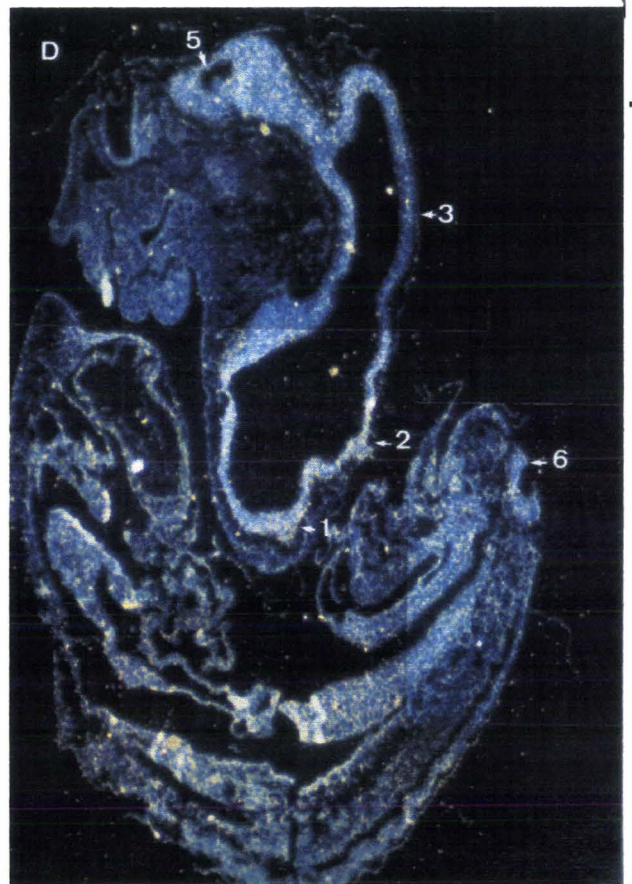
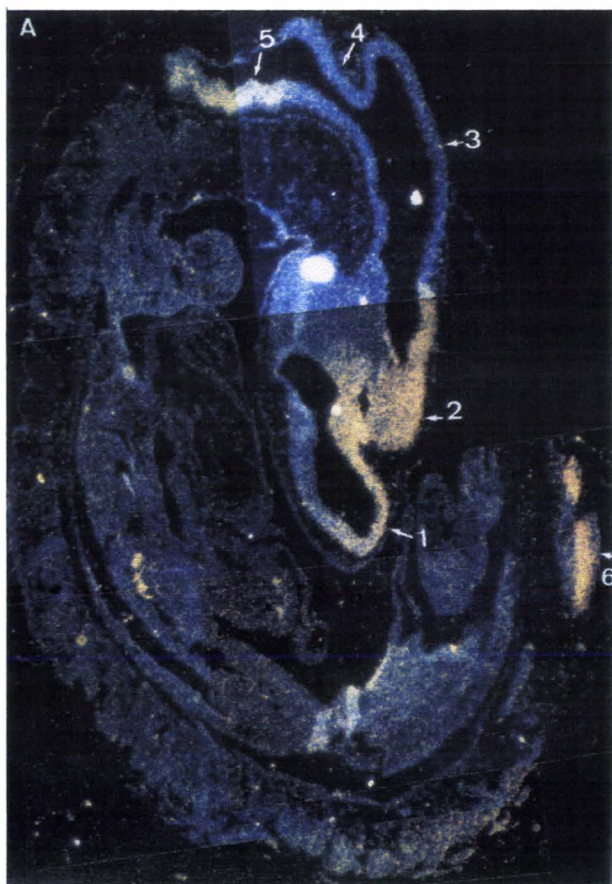
**Figure 2B**



probe P2: GATCCTGTACAGTAATCAATTAGACTCAG

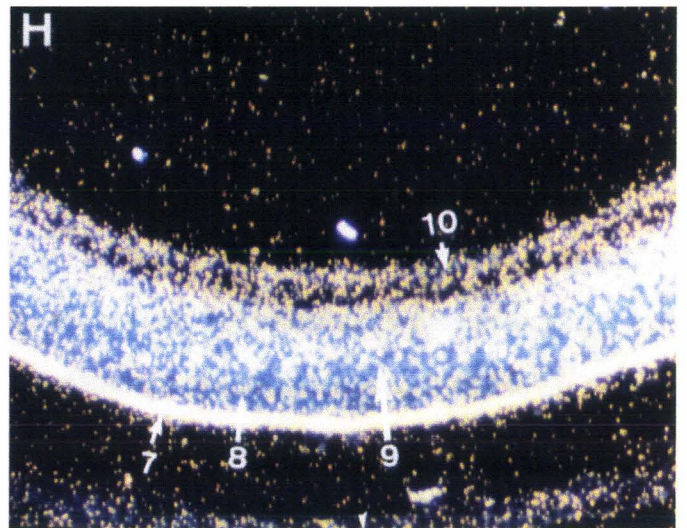
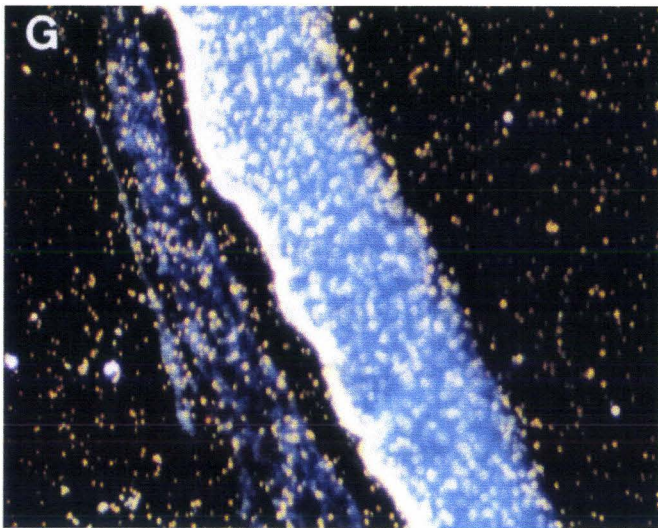
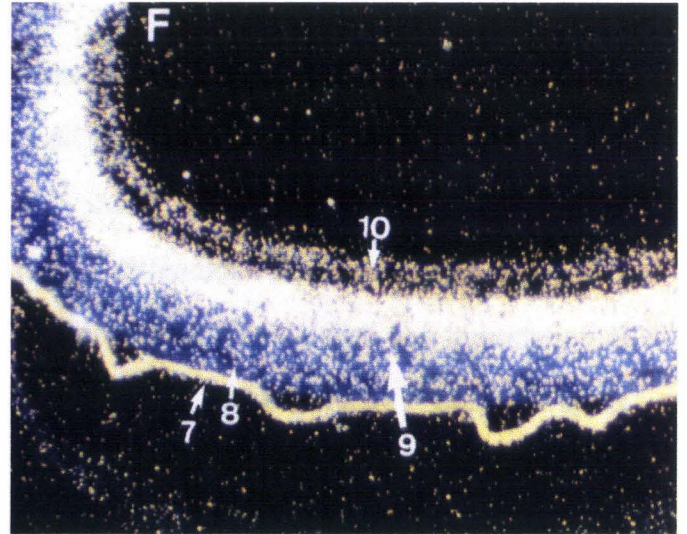
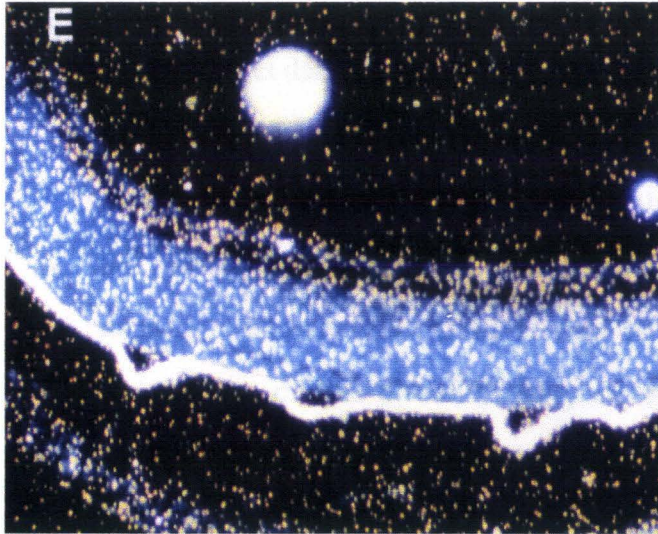
Figure 3



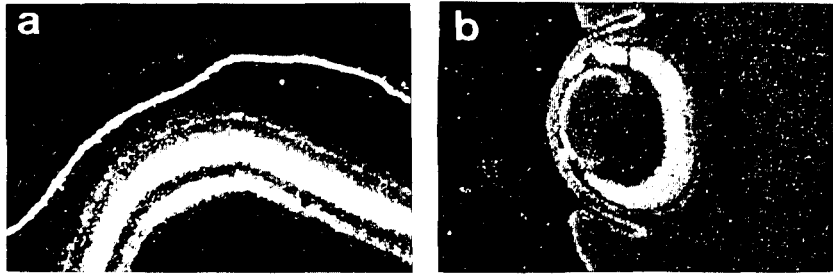


**Figure 4a: 3.5 days quail embryos mapped with**  
**A: antisense probe for Pax-6 MC29-QNR2 m RNA**  
**B: sense probe**  
**C: antisense probe for Pax-6 B1 m RNA**  
**D: sense probe**





**Figure 4b: 9 days quail embryo neuroretina mapped with**  
**E: sense probe for Pax-6 B1 mRNA**  
**F: antisense probe**  
**G: sense probe for Pax-6 MC29-QNR2 m RNA**  
**H: antisense probe**



**Figure 5**



### *Discussion de l'article 3*

Nous ne connaissons pas encore le rôle de ces protéines dépourvues de domaine paired chez les vertébrés. Néanmoins, nous savons qu'elles sont incapables de transactiver les promoteurs PO et P1 comme le fait la P46 et ne peuvent non plus se fixer sur une séquence de reconnaissance pour le domaine homéo bien qu'elles en possèdent un à part entière. Chez le nématode, des protéines de même type existent mais sont engendrées par un mécanisme différent de celui présent chez la caille puisqu'elles dérivent de l'utilisation d'un promoteur alternatif. Elles sont impliquées dans la mise en place des cils vibratiles, qui constituent un appareil sensoriel de l'animal. Pourquoi l'évolution a-t-elle introduit un mécanisme de régulation différent pour synthétiser ces protéines Pax-6 sans domaine paired ? Il est probable que l'intérêt de cet épissage alternatif repose sur un meilleur contrôle du titre des petites protéines. Comme nous l'avons vu au cours de l'introduction, le dosage des ARNm de Pax-6 est un facteur essentiel pour que les fonctions de ce gène soient assurées, un dérèglement du taux d'expression de Pax-6 pouvant entraîner des malformations embryonnaires.

Le rôle précis de ces petites protéines de 32-33 kDa dans le développement de l'embryon reste une énigme non résolue. Les cartographies d'expression de l'ARNm B1 concordent avec le taux des protéines de 32-33 kDa dans les différents organes qui expriment Pax-6. Les taux d'expression les plus élevés sont retrouvés dans l'oeil et le pancréas tandis qu'ils décroissent dans le tube neural et le cerveau. Leur mécanisme d'action en absence de toute fixation à l'ADN, dans l'état actuel de nos connaissances, reste à établir. Il est possible d'imaginer que ces protéines nécessitent l'aide d'un cofacteur pour se fixer à l'ADN ou bien, alternativement, qu'elle fonctionnent en l'absence de toute fixation à l'ADN, comme les protéines Engrailed qui régulent négativement l'expression de Pax-6 grâce à des interactions protéines-protéines.

*Discussion*  
*gènes Pax*



## DISCUSSION

L'étude des données de la littérature des gènes Pax nous indique que ce sont des "gènes tissu spécifique" qui s'expriment au cours du développement embryonnaire. Ils sont notamment impliqués dans la mise en place du système nerveux central (cerveau et moelle épinière, à l'exception de *Pax-1* et *Pax-9*), dans la formation des organes sensoriels et dans l'organogenèse d'autres organes tels que le pancréas (*Pax-4/6*), les reins (*Pax-2/8*), le thymus (*Pax-1/9*), la thyroïde (*Pax-8*) ou encore les testicules (*Pax-5*). De plus, leurs limites d'expression dans le cerveau coïncident souvent avec des limites morphologiques. Les gènes Pax semblent par conséquent participer à la segmentation du cerveau chez les vertébrés. Ces gènes s'expriment aussi chez les animaux adultes dans des territoires plus restreints, mais néanmoins comparables à ceux observés chez l'embryon. Chez l'adulte, ils maintiendraient l'identité cellulaire et régionale des tissus. (Cf. chapitre 1-2 de l'introduction pour la description de l'expression des gènes Pax).

Une étude phylogénétique (basée d'une part sur la méthode des distances et d'autre part sur la méthode de parcimonie du domaine paired des gènes Pax) (Balczarek *et al.*, 1997) a permis de rassembler les gènes Pax en quatre groupes: groupe I pour *Pax-1* et *Pax-9*, groupe II pour *Pax-2*, *Pax-5* et *Pax-8*, groupe III pour *Pax-3* et *Pax-7* et enfin le groupe 4, formé de *Pax-6* et *Pax-4*. Ces résultats, obtenus au moyen d'études phylogénétiques, coïncident avec des regroupements établis préalablement à partir de l'observation d'identités de séquence et de structure des protéines Pax (présence des domaines paired, octapeptide et homéodomaine) (Gruss et Walther, 1992). Or, il a été observé que les protéines qui appartiennent à un même groupe présentent des propriétés biochimiques similaires: spécificité de liaison à l'ADN commune. Ces observations sont conciliables avec la présence d'acides aminés conservés en des positions clés du domaine paired sur la protéine, telle l'asparagine 47, spécifique des protéines Pax-4 et Pax-6 qui confère leur spécificité de reconnaissance pour l'ADN (cf. p22, "le cas des protéines Pax au domaine RED actif"). Par ailleurs, on pourrait ajouter que les gènes d'un même groupe présentent également des territoires d'expression communs. Ainsi dans le tube neural, *Pax-3* et *Pax-7* sont localisés dans les plaques alaires, *Pax-2*, *Pax-5* et *Pax-8*, dans les neurones intermédiaires. De même, *Pax-2* et *Pax-5* sont colocalisés dans le met-mesencéphale et dans le tractus urogénital: *Pax-2* et *Pax-8* dans les reins, *Pax-5*, dans les testicules. *Pax-1* et *Pax-9* sont coexprimés dans le thymus et le mésenchyme qui longe l'oesophage, *Pax-4* et *Pax-6*

dans le pancréas, *Pax-3* et *Pax-7* dans les muscles squelettiques . Ces coexpressions entre gènes d'un même groupe appuient les résultats des études phylogénétiques sur l'idée d'un gène ancestral pour chacun de ces groupes. De plus, ces observations suggèrent l'existence de séquences *cis* régulatrices en amont des gènes, similaires entre gènes d'un même groupe, car provenant d'un gène ancestral et régulées par des "facteurs tissu-spécifiques" communs. Notre groupe a recherché les séquences régulatrices qui induisent l'expression de *Pax-6* dans la neurorétine. Un enhancer  $\alpha$  été localisé, dont la fonction activatrice de la transcription s'exerce seulement sur l'un des deux promoteurs préalablement identifiés, du gène *Pax-6* de caille (Po). Son activité est spécifique des cellules de neurorétine (l'enhancer est inactif dans les cellules Cos ou dans les cellules de rétine pigmentaire) et dépendante du stade du développement (elle s'exerce à partir de 7 jours de développement embryonnaire). La présence de cet enhancer  $\alpha$  pourrait expliquer la spécificité d'expression neuronale des ARNm de Pax-QNR initiés à Po. En effet, ces transcrits ne sont pas retrouvés dans les cellules de rétine pigmentaire contrairement aux ARNm produits à partir du promoteur P1. Néanmoins, d'autres éléments de régulation doivent intervenir pour réguler l'expression plus précoce et plus large du promoteur P1 dont les ARNm sont détectés à la fois dans la rétine pigmentaire et la neurorétine. Chez la drosophile, il existe un enhancer dans le locus *sparkling* (*spa*), l'homologue des gènes de vertébrés *Pax-2*, *Pax-5* et *Pax-8* , qui permet l'expression spécifique de *spa* dans les cônes et les cellules de pigments primaires. Par contre, la délétion de cet enhancer ne perturbe pas l'expression de *spa* dans les cils mécanosensoriels, ce qui signifie que cette expression est gouvernée par un enhancer distinct (Fu et Noll, 1997).

Pour l'enhancer  $\alpha$ , 4 boîtes fixant des protéines spécifiques de la rétine neuronale ont été mises en évidence par digestion à la DNase I. Ces boîtes sont conservées dans l'enhancer de souris. Chez la caille, elles fixent diverses protéines dont les masses s'échelonnent de 32 à 110 kD. Il serait intéressant de pouvoir identifier ces facteurs qui dirigent l'activité spécifique de Po dans les cellules de type neuronal de la rétine. L'identification de ces protéines permettra une avancée non négligeable dans l'étude des régulations du gène *Pax-6* . A l'heure actuelle, les résultats obtenus par différents groupes de recherche sur les gènes Pax portent davantage sur la détermination de leurs gènes cibles que sur leurs éléments de régulation *cis* ou encore moins *trans*. Pourtant, s'il est incontestable que les protéines Pax sont sensibles aux mutations pour accomplir leur fonction d'activateurs de la transcription de leurs gènes cibles, il reste non moins évident que des mutations dans les éléments de régulations *cis* ou dans les facteurs *trans*-

régulateurs pourraient être fatales pour l'expression "tissu-spécifique" des protéines Pax. Mais ces axes de recherche sont moins aisés à réaliser et plus longs à développer. Ainsi, notre groupe a mis en évidence la présence d'un enhancer responsable de l'expression "neurorétine spécifique", et celui-ci ne fonctionne pas dans les cellules de lignée pancréatique où s'exprime aussi *pax-6*. L'élément responsable de l'expression spécifique de *Pax-6* dans les cellules endocrines du pancréas reste encore inconnu.

Il existe de nombreux sites putatifs de fixation pour des facteurs *trans* sur les promoteurs Po et P1 du gène *Pax-6* de caille: Sp1, NF-kB, SIF, Ap2, myc/USF, Krox20/24, CRE, E-box, myb... Au sujet de Krox-20, il est intéressant de noter que ce facteur permet la spécificité d'expression du gène *Hoxb-2* (Vesque *et al.*, 1996) dans les rhombomères 3 et 5 du cerveau postérieur. De plus, *Pax-6* s'exprime aussi dans les rhombomères 3 et 5 et seulement ceux-ci. Krox-20 pourrait-être impliqué dans l'expression rhombencéphalique spécifique de *Pax-6*. Tous ces facteurs *trans*, régulateurs potentiels du gène *Pax-QNR*, n'ont pu être testés dans notre laboratoire. Néanmoins, il a été montré que c-Myb est capable de *trans*-activer les promoteurs Po et P1 de *Pax-QNR*, d'une part de façon directe en se fixant sur le promoteur et d'autre part, de manière indirecte, indépendamment de son domaine de fixation à l'ADN (Plaza *et al.*, 1995b). c-Myb et les gènes apparentés à c-myb sont des facteurs de transcription qui contrôlent les mécanismes de différenciation cellulaire au moyen d'une régulation directe de l'expression des gènes. En effet, les niveaux d'expression de c-myb décroissent rapidement sous la différenciation des myéloblastes, des cellules érythrocytaires ou encore des cellules de neuroblastome (Coll *et al.*, 1983); (Ramsay *et al.*, 1986). De plus, l'expression constitutive de c-myb empêche la différenciation des cellules érythrocytaires (Clarke *et al.*, 1988) et stimule la croissance et la réponse au bFGF des cellules de neurorétine de poulet (Garrido *et al.*, 1993). c-Myb pourrait-être le déclencheur de l'expression du gène *Pax-6* durant les phases précoces du développement embryonnaire, dans les cellules qui n'ont pas encore atteint leur stade de différenciation terminale. Puis, la protéine P46, traduite à partir du gène *Pax-QNR* (qui représente la protéine traduite de tous les gènes orthologues de *Pax-6*) pourrait prendre le relai en matière d'activation du gène *Pax-6*, puisqu'elle stimule la transcription du gène dont elle est issue, à la fois *via* Po et *via* P1 pour le gène *Pax-QNR* (Plaza *et al.*, 1993).

Les protéines c-Myb et la p46 de *Pax-QNR* sont capables d'activer la transcription du gène *Pax-6* de caille *via* les promoteurs Po et P1 mais elles ne sont pas suffisantes pour expliquer la cartographie d'expression des ARNm de *Pax-6*. Notamment dans le cerveau, *Pax-6* s'exprime dans le télencéphale, le diencéphale et dans les rhombomères 3 et 5 du

cerveau postérieur durant les stades précoces du développement embryonnaire. On le détecte plus tardivement dans le cervelet, mais on ne trouve aucune trace de son expression dans la région du met-mésencéphale ou cerveau moyen. Or cette région est couverte par les produits des gènes *En-1* et *En-2* et nous avons montré que la protéine *En-1* est capable de réprimer l'activation de la transcription médiée par la P46 de *Pax-QNR*. Cette régulation s'effectue par un mécanisme indépendant du domaine de fixation à l'ADN de la protéine *En-1*; *En-1* interagit avec la P46 et altère ainsi sa capacité à reconnaître sa séquence cible. De cette façon, la P46, bloquée par *En-1*, ne peut plus activer la transcription du gène *Pax-QNR* puisqu'elle n'est plus capable de se fixer ni sur *Po*, ni sur *P1*. Cette régulation s'effectue dans la région du cerveau moyen, mais aussi dans les cellules de neurorétine après l'éclosion, un stade où le taux des ARNm de *Pax-QNR* chute, tandis que l'expression d'*En-1* commence à apparaître.

Les mécanismes de la régulation transcriptionnelle du gène *Pax-QNR* s'exercent à différents niveaux dans les cellules de neurorétine: activation directe (P46 de *Pax-QNR* et *c-Myb*) et indirecte (*c-Myb*) et répression indirecte (*En-1*) des promoteurs *Po* et *P1*. Il est probable que l'activité de ces promoteurs soit aussi modulée par d'autres facteurs encore non identifiés. De plus, il existe au moins un élément *cis* situé à 7,5 kpb en aval du promoteur *Po*, appelé enhancer  $\alpha$ , qui stimule l'expression des ARNm MC29QNR2 dans la neurorétine. Cependant, les protéines qui interagissent avec cet élément n'ont pas encore été déterminées. D'autres éléments *cis* existent probablement pour induire notamment l'expression des ARNm *Pax-QNR* dans le pancréas. Par ailleurs, *Pax-6* génère de nombreuses isoformes par épissage alternatif (Carrière *et al.*, 1995) dont les proportions respectives doivent être impérativement respectées. En effet, l'altération de ces proportions peut entraîner des défauts des yeux chez l'embryon. De même qu'une diminution du nombre de copies des ARNm, causée par la perte ou la mutation d'un allèle provoque des malformations des yeux graves, on parle de mutations semi-dominantes (Mirzayans *et al.*, 1995). En conséquence, il apparaît évident qu'une régulation quantitative s'exerce sur le gène *Pax-6*. Cette régulation se réalise en premier lieu par l'intermédiaire d'épissages alternatifs à l'origine des diverses isoformes. En second lieu, l'ARNm lui-même peut participer à la régulation quantitative des protéines par un mécanisme post-transcriptionnel. En outre, l'ARNm B1 possède une phase de lecture ouverte dotée seulement de codons alternatifs de traduction dans son extrémité 5' et dans sa boîte paired qui ne sont pas ou peu reconnus par le complexe ribosomal. Cette constitution favorise le glissement du complexe d'initiation vers les codons internes de traduction de type AUG; ce

processus de fuite des ribosomes, nommé "leaky scanning" (Geballe et Morris, 1994) pour revue) permet un meilleur rendement de la traduction à partir des AUG internes. En conséquence, la présence de ARNm B1 augmente la proportion des protéines Pax-QNR de 32 et 33 kD. Le rôle de ces protéines n'a pas encore été clairement défini. Elles possèdent un domaine homo entier, encadré d'une séquence de localisation nucléaire et d'un domaine transactivateur. Néanmoins, elles ne sont pas capables d'interagir avec les cibles consensuelles classiques des domaines homéo. Peut-être nécessitent-elles des cofacteurs pour aider à leur fixation sur l'ADN? La fonction de ces protéines reste à déterminer dans l'avenir. Elles existent chez le nématode, bien qu'elles soient produites par un mécanisme différent (il existe un promoteur interne) et participent à la fonction sensorielle. Elles sont à l'origine de l'organisation et de la détermination des cils vibratiles situés à l'extrémité antérieure du ver et lui servent à détecter sa nourriture.

L'étude de la régulation du gène *Pax-6* et des gènes Pax en général est un vaste sujet de recherche. Leur régulation semble s'exercer de manière spatio-temporelle: les éléments *cis* ou *trans* intervenant varient en fonction du tissu et du stade du développement embryonnaire. Ainsi Pax-QNR s'exprime dans la rétine neuronale à partir de E7, sous l'action d'un enhancer stimulé par des protéines de neurorétine. Son expression précoce peut être liée aux propriétés stimulatrices de c-Myb sur ces deux promoteurs Po et P1. Et ses limites d'expression dans le cerveau font intervenir la protéine En-1 qui empêche la protéine P46 de Pax-6 d'autoactiver les promoteurs Po et P1. Par ailleurs, le respect des proportions de chacune des isoformes générées par le gène Pax-6 est indispensable pour maintenir la fonction correcte du gène. Les conclusions apportées par ces résultats amènent d'autres questions encore non résolues. D'une part c-Myb ne peut pas agir seul pour induire l'expression précoce de *Pax-6*, car il s'exprime aussi dans des tissus où *Pax-6* n'est pas retrouvé comme les cellules érythrocytaires et les cellules musculaires. Il pourrait agir en synergie avec d'autres facteurs comme des gènes induits par des facteurs de croissance précoce, tels les facteurs de type FGF. Nous avons vu par exemple que le bFGF associé à c-Myb augmente la croissance des cellules de neurorétine (Garrido *et al.*, 1993). Par ailleurs, un autre membre de la famille FGF, le FGF8 induit l'expression de BF1, un facteur essentiel pour le développement de l'oeil et du télencéphale, tissus où Pax-6 est détectée. D'autre part le NGF et le BDNF, des facteurs de croissance neuronaux stimulent aussi l'expression de *Pax-6* dans les cellules (Kioussi et Gruss, 1994). D'autres facteurs diffusibles peuvent au

contraire exercer un contrôle négatif sur l'expression de *Pax-6*. Ainsi, *Hedgehog* réprime *Pax-6*. Mais en ce qui concerne *Hedgehog*, rien n'est simple, car son action peut être activatrice ou inhibitrice selon sa concentration dans le milieu où il se trouve (cf. (Bumcrot et McMahon, 1996), pour revue). Par ailleurs, les récepteurs nucléaires, facteurs de transcription pour la plupart activés par des molécules diffusibles telles que les hormones ou les vitamines pourraient être de bons candidats pour la régulation du gène *Pax-6*. En effet, la carence en vitamine A (= précurseur des acides rétinoïques), de même que le "Knock out" des récepteurs aux acides rétinoïques ou au contraire, l'injection d'acide rétinoïque chez l'embryon précoce provoque des défauts du développement caractérisés par des anomalies des yeux allant jusqu'à leur disparition totale et une réduction du cerveau antérieur. Or, on retrouve des sites putatifs de fixation pour les récepteurs nucléaires dans l'enhancer  $\alpha$  et le promoteur P1. C'est une avenue de recherche qui pourrait-être empruntée...

L'ensemble de ces résultats suggère que la régulation du gène *Pax-6* est extrêmement complexe et pourrait faire intervenir de nombreux éléments *cis* et *trans* variés. De plus, ces réseaux de régulation seraient spécifiques d'un stade du développement et d'un tissu donné.

*Bibliographie*  
*gènes Pax*

## BIBLIOGRAPHIE Pax

- Asano, M. et Gruss, P. (1992). Pax-5 is expressed at the midbrain-hindbrain boundary during mouse development. *Mech Dev*, **39**, 29-39.
- Balczarek, K. A., Lai, Z. C. et Kumar, S. (1997). Evolution of functional diversification of the paired box (Pax) DNA-binding domains. *Mol Biol Evol*, **14**, 829-842.
- Baldwin, C. T., Hoth, C. F., Amos, J. A., da, S. E. et Milunsky, A. (1992). An exonic mutation in the HuP2 paired domain gene causes Waardenburg's syndrome [see comments]. *Nature*, **355**, 637-638.
- Balling, R., Deutsch, U. et Gruss, P. (1988). undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1. *Cell*, **55**, 531-535.
- Bally, C. L. et Wassef, M. (1994). Ectopic induction and reorganization of Wnt-1 expression in quail/chick chimeras. *Development*, **120**, 3379-3394.
- Bender, W., Spierer, P. et Hogness, D. S. (1983). Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster*. *J Mol Biol*, **168**, 17-33.
- Bertuccioli, C., Fasano, L., Jun, S., Wang, S., Sheng, G. et Desplan, C. (1996). In vivo requirement for the paired domain and homeodomain of the paired segmentation gene product. *Development*, **122**, 2673-2685.
- Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. et Noll, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*. *Cell*, **47**, 1033-1040.
- Bopp, D., Jamet, E., Baumgartner, S., Burri, M. et Noll, M. (1989). Isolation of two tissue-specific *Drosophila* paired box genes, Pox meso and Pox neuro. *Embo J*, **8**, 3447-3457.
- Bumcrot, D. A. et McMahon, A. P. (1996). Sonic hedgehog: making the gradient. *Chem Biol*, **3**, 13-16.
- Carrière, C., Plaza, S., Caboche, J., Dozier, C., Bailly, M., Martin, P. et Saule, S. (1995). Nuclear localization signals, DNA binding, and transactivation properties of quail Pax-6 (Pax-QNR) isoforms. *Cell Growth Differ*, **6**, 1531-1540.
- Carrière, C., Plaza, S., Martin, P., Quatannens, B., Bailly, M., Stehelin, D. et Saule, S. (1993). Characterization of quail Pax-6 (Pax-QNR) proteins expressed in the neuroretina. *Mol Cell Biol*, **13**, 7257-7266.
- Carroll, S. B., Laymon, R. A., McCutcheon, M. A., Riley, P. D. et Scott, M. P. (1986). The localization and regulation of Antennapedia protein expression in *Drosophila* embryos. *Cell*, **47**, 113-122.
- Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M. et Gruss, P. (1991). The molecular basis of the undulated/Pax-1 mutation. *Cell*, **66**, 873-884.
- Chalepakis, G., Jones, F. S., Edelman, G. M. et Gruss, P. (1994). Pax-3 contains domains for transcription activation and transcription inhibition. *Proc Natl Acad Sci U S A*, **91**, 12745-12749.
- Chan, S. K. et Mann, R. S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. *Genes Dev*, **7**, 796-811.
- Clarke, M. F., Kukowska, L. J., Westin, E., Smith, M. et Prochownik, E. V. (1988). Constitutive expression of a c-myc cDNA blocks Friend murine erythroleukemia cell differentiation. *Mol Cell Biol*, **8**, 884-892.
- Coll, J., Saule, S., Martin, P., Raes, M. B., Lagrou, C., Graf, T., Beug, H., Simon, I. E. et Stehelin, D. (1983). The cellular oncogenes c-myc, c-myb and c-erb are transcribed in defined types of avian hematopoietic cells. *Exp Cell Res*, **149**, 151-162.
- Crisanti-Combes, P., Pessac, B. et Calothy, G. (1978). Choline acetyl transferase activity in chick embryo neuroretinas during development in ovo and in monolayer cultures. *Dev Biol*, **65**, 228-232.
- Czerny, T. et Busslinger, M. (1995). DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Mol Cell Biol*, **15**, 2858-2871.
- Czerny, T., Schaffner, G. et Busslinger, M. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev*, **7**, 2048-2061.
- Davis, C. A., Holmyard, D. P., Millen, K. J. et Joyner, A. L. (1991). Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development*, **111**, 287-298.
- Davis, C. A. et Joyner, A. L. (1988). Expression patterns of the homeo box-containing genes En-1 and En-2 and the proto-oncogene int-1 diverge during mouse development. *Genes Dev*, **2**, 1736-1744.
- Deutsch, U., Dressler, G. R. et Gruss, P. (1988). Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell*, **53**, 617-625.
- Dietrich, S. et Gruss, P. (1995). undulated phenotypes suggest a role of Pax-1 for the development of vertebral and extravertebral structures. *Dev Biol*, **167**, 529-548.
- Dörfler, P. et Busslinger, M. (1996). C-terminal activating and inhibitory domains determine the transactivation potential of BSAP (Pax-5), Pax-2 and Pax-8. *Embo J*, **15**, 1971-1982.
- Dressler, G. R. et Douglass, E. C. (1992). Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor. *Proc Natl Acad Sci U S A*, **89**, 1179-1183.
- Duboule, D. et Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet*, **10**, 358-364.
- Eaton, S. et Kornberg, T. B. (1990). Repression of ci-D in posterior compartments of *Drosophila* by engrailed. *Genes Dev*, **4**, 1068-1077.



- Epstein, D. J., Vekemans, M. et Gros, P. (1991). Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell*, **67**, 767-774.
- Epstein, J., Cai, J., Glaser, T., Jepeal, L. et Maas, R. (1994a). Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem*, **269**, 8355-8361.
- Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S. et Maas, R. L. (1994b). Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev*, **8**, 2022-2034.
- Fauquet, M., Stehelin, D. et Saule, S. (1990). myc products induce the expression of catecholaminergic traits in quail neural crest-derived cells. *Proc Natl Acad Sci U S A*, **87**, 1546-1550.
- Figdor, M. C. et Stern, C. D. (1993). Segmental organization of embryonic diencephalon. *Nature*, **363**, 630-634.
- Fitzpatrick, V. D., Percival, S. A., Ingles, C. J. et Krause, H. M. (1992). Homeodomain-independent activity of the fushi tarazu polypeptide in Drosophila embryos. *Nature*, **356**, 610-612.
- Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. et Noll, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell*, **47**, 735-746.
- Fu, W. et Noll, M. (1997). The Pax2 homolog sparkling is required for development of cone and pigment cells in the Drosophila eye [see comments]. *Genes Dev*, **11**, 2066-2078.
- Fujioka, M., Miskiewicz, P., Raj, L., Gullledge, A. A., Weir, M. et Goto, T. (1996). Drosophila Paired regulates late even-skipped expression through a composite binding site for the paired domain and the homeodomain. *Development*, **122**, 2697-2707.
- Garber, J., Kuroiwa, A. et Gehring, W. (1983). Genomic end cDNA clones of the homeotic locus *Antennapedia* in *Drosophila*. *EMBO J*, **2**, 2027-2036.
- Garrido, C., Plaza, S., Gospodarowicz, D. et Saule, S. (1993). Fos expression induces growth arrest in tumorigenic neuroretina cells. *Oncogene*, **8**, 2713-2719.
- Geballe, A. P. et Morris, D. R. (1994). Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem Sci*, **19**, 159-164.
- Gehring, W. (1966). Bildung eines vollständigen Mittelbeines mit Sternopleura in der Antennenregion bei der Mutant *Nasobemia* (Ns) von *Drosophila melanogaster*. *Jul. Klaus. Arch*, **41**, 44-54.
- Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J. et Maas, R. L. (1994). PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects [published erratum appears in *Nat Genet* 1994 Oct;8(2):203]. *Nat Genet*, **7**, 463-471.
- Goulding, M., Sterrer, S., Fleming, J., Balling, R., Nadeau, J., Moore, K. J., Brown, S. D., Steel, K. P. et Gruss, P. (1993). Analysis of the Pax-3 gene in the mouse mutant splotch. *Genomics*, **17**, 355-363.
- Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. et Gruss, P. (1991). Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *Embo J*, **10**, 1135-1147.
- Gruss, P. et Walther, C. (1992). Pax in development. *Cell*, **69**, 719-722.
- Gutjahr, T., Frei, E. et Noll, M. (1993). Complex regulation of early paired expression: initial activation by gap genes and pattern modulation by pair-rule genes. *Development*, **117**, 609-623.
- Halder, Callaerts, P. et Gehring, W. T. (1995). New perspectives on eye evolution. *Curr. Opin. Genet. Dev.*, **5**, 602-609.
- Han, K. et Manley, J. L. (1993). Functional domains of the Drosophila Engrailed protein. *Embo J*, **12**, 2723-2733.
- Hanson, I. M., Fletcher, J. M., Jordan, T., Brown, A., Taylor, D., Adams, R. J., Punnett, H. H. et van, H. V. (1994). Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nat Genet*, **6**, 168-173.
- Hayashi, S. et Scott, M. P. (1990). What determines the specificity of action of Drosophila homeodomain proteins? [published erratum appears in *Cell* 1991 Mar 8;64(5):following 1046]. *Cell*, **63**, 883-894.
- Hill, R. E., Favor, J., Hogan, B. L., Ton, C. C., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D. et van, H. V. (1991). Mouse small eye results from mutations in a paired-like homeobox-containing gene [published erratum appears in *Nature* 1992 Feb 20;355(6362):750]. *Nature*, **354**, 522-525.
- Hoey, T., Warrior, R., Manak, J. et Levine, M. (1988). DNA-binding activities of the Drosophila melanogaster even-skipped protein are mediated by its homeo domain and influenced by protein context. *Mol Cell Biol*, **8**, 4598-4607.
- Hogan, B. L., Hirst, E. M., Horsburgh, G. et Hetherington, C. M. (1988). Small eye (Sey): a mouse model for the genetic analysis of craniofacial abnormalities. *Development*, **115**, 119.
- Hogan, B. L., Horsburgh, G., Cohen, J., Hetherington, C. M., Fisher, G. et Lyon, M. F. (1986). Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. *J Embryol Exp Morphol*, **97**, 95-110.
- Hoth, C. F., Milunsky, A., Lipsky, N., Sheffer, R., Clarren, S. K. et Baldwin, C. T. (1993). Mutations in the paired domain of the human PAX3 gene cause Klein- Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am J Hum Genet*, **52**, 455-462.
- Jaynes, J. B. et O'Farrell, P. H. (1991). Active repression of transcription by the engrailed homeodomain protein. *Embo J*, **10**, 1427-1433.
- Jostes, B., Walther, C. et Gruss, P. (1990). The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech Dev*, **33**, 27-37.
- Jun, S. et Desplan, C. (1996). Cooperative interactions between paired domain and homeodomain. *Development*, **122**, 2639-2650.

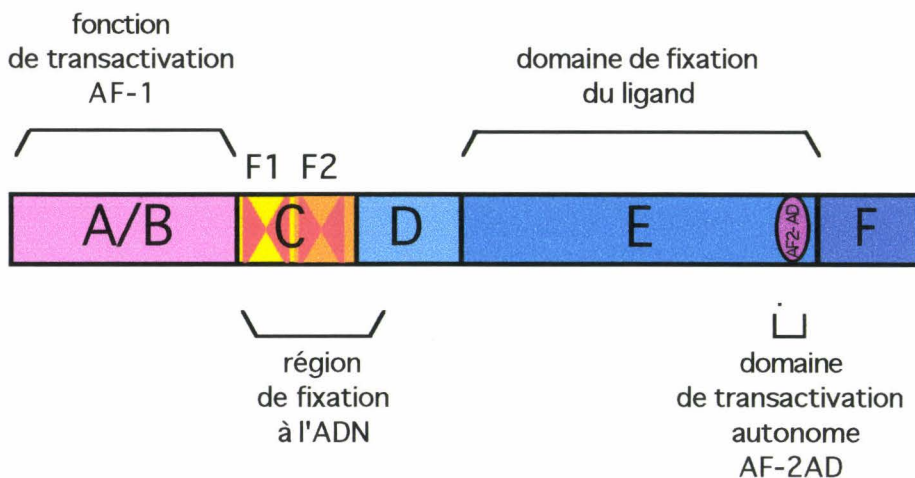
- Kaufman, T. C., Seeger, M. A. et Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet*, **27**, 309-362.
- Keller, S. A., Jones, J. M., Boyle, A., Barrow, L. L., Killen, P. D., Green, D. G., Kapousta, N. V., Hitchcock, P. F., Swank, R. T. et Meisler, M. H. (1994). Kidney and retinal defects (Krd), a transgene-induced mutation with a deletion of mouse chromosome 19 that includes the Pax2 locus. *Genomics*, **23**, 309-320.
- Kioussi, C. et Gruss, P. (1994). Differential induction of Pax genes by NGF and BDNF in cerebellar primary cultures. *J Cell Biol*, **125**, 417-425.
- Kissinger, C. R., Liu, B. S., Martin, B. E., Kornberg, T. B. et Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**, 579-590.
- Krauss, S., Johansen, T., Korzh, V. et Fjose, A. (1991). Expression of the zebrafish paired box gene pax[zf-b] during early neurogenesis. *Development*, **113**, 1193-1206.
- Krauss, S., Maden, M., Holder, N. et Wilson, S. W. (1992). Zebrafish pax[b] is involved in the formation of the midbrain-hindbrain boundary. *Nature*, **360**, 87-89.
- Lai, J. S., Cleary, M. A. et Herr, W. (1992). A single amino acid exchange transfers VP16-induced positive control from the Oct-1 to the Oct-2 homeo domain [published erratum appears in *Genes Dev* 1992 Dec;6(12B):2663]. *Genes Dev*, **6**, 2058-2065.
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature*, **276**, 565-570.
- Lin, L. et McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. *Genes Dev*, **6**, 1071-1081.
- Logan, C., Hanks, M. C., Noble, T. S., Nallainathan, D., Provart, N. J. et Joyner, A. L. (1992). Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions. *Dev Genet*, **13**, 345-358.
- Matsuo, T., Osumi, Y. N., Noji, S., Ohuchi, H., Koyama, E., Myokai, F., Matsuo, N., Taniguchi, S., Doi, H., Iseki, S. et al. (1993). A mutation in the Pax-6 gene in rat small eye is associated with impaired migration of midbrain crest cells. *Nat Genet*, **3**, 299-304.
- Mikkola, I., Fjose, A., Kuwada, J. Y., Wilson, S., Guddal, P. H. et Krauss, S. (1992). The paired domain-containing nuclear factor pax[b] is expressed in specific commissural interneurons in zebrafish embryos. *J Neurobiol*, **23**, 933-946.
- Mirzayans, F., Pearce, W. G., MacDonald, I. M. et Walter, M. A. (1995). Mutation of the PAX6 gene in patients with autosomal dominant keratitis. *Am J Hum Genet*, **57**, 539-548.
- Moase, C. E. et Trasler, D. G. (1990). Delayed neural crest cell emigration from Sp and Spd mouse neural tube explants. *Teratology*, **42**, 171-182.
- Noll, M. (1993). Evolution and role of Pax genes. *Curr Opin Genet Dev*, **3**, 595-605.
- Nornes, H. O., Dressler, G. R., Knapik, E. W., Deutsch, U. et Gruss, P. (1990). Spatially and temporally restricted expression of Pax2 during murine neurogenesis. *Development*, **109**, 797-809.
- Nornes, S., Mikkola, I., Krauss, S., Delghandi, M., Perander, M. et Johansen, T. (1996). Zebrafish Pax9 encodes two proteins with distinct C-terminal transactivating domains of different potency negatively regulated by adjacent N-terminal sequences. *J Biol Chem*, **271**, 26914-26923.
- Nüsslein-Volhard, C., Kluding, H. et Jurgens, G. (1985). Genes affecting the segmental subdivision of the *Drosophila* embryo. *Cold Spring Harb Symp Quant Biol*, **50**, 145-154.
- Nüsslein-Volhard, C. et Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, **287**, 795-801.
- Ohkuma, Y., Horikoshi, M., Roeder, R. G. et Desplan, C. (1990). Engrailed, a homeodomain protein, can repress *in vitro* transcription by competition with the TATA box-binding protein transcription factor IID. *Proc Natl Acad Sci U S A*, **87**, 2289-2293.
- Okada, T. S. (1976). "Transdifferentiation" of cells of specialized eye tissues in cell culture. In: *Ebert Jd, Marois M, ed. Tests of teratogenicity in vitro. Amsterdam, North Holland Publ.*, 91-105.
- Peters, H., Doll, U. et Niessing, J. (1995). Differential expression of the chicken Pax-1 and Pax-9 gene: *in situ* hybridization and immunohistochemical analysis. *Dev Dyn*, **203**, 1-16.
- Pittack, C., Jones, M. et Reh, T. A. (1991). Basic fibroblast growth factor induces retinal pigment epithelium to generate neural retina *in vitro*. *Development*, **113**, 577-588.
- Plaza, S., Dozier, C. et Saule, S. (1993). Quail Pax-6 (Pax-QNR) encodes a transcription factor able to bind and transactivate its own promoter. *Cell Growth Differ*, **4**, 1041-1050.
- Plaza, S., Dozier, C., Turque, N. et Saule, S. (1995a). Quail Pax-6 (Pax-QNR) mRNAs are expressed from two promoters used differentially during retina development and neuronal differentiation. *Mol Cell Biol*, **15**, 3344-3353.
- Plaza, S., Turque, N., Dozier, C., Bailly, M. et Saule, S. (1995b). C-Myb acts as transcriptional activator of the quail PAX6 (PAX-QNR) promoter through two different mechanisms. *Oncogene*, **10**, 329-340.
- Puelles, L. et Rubenstein, J. L. (1993). Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization. *Trends Neurosci*, **16**, 472-479.
- Quiring, R., Walldorf, U., Kloter, U. et Gehring, W. J. (1994). Homology of the *eyeless* gene of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans. *Science*, **265**, 785-789.

- Ramsay, R. G., Ikeda, K., Rifkind, R. A. et Marks, P. A. (1986). Changes in gene expression associated with induced differentiation of erythroleukemia: protooncogenes, globin genes, and cell division. *Proc Natl Acad Sci U S A*, **83**, 6849-6853.
- Roberts, R. C. (1967). *Small eyes*, a new dominant mutant in the mouse. *Genetic Research*, **9**, 121-122.
- Schafer, B. W., Czerny, T., Bernasconi, M., Genini, M. et Busslinger, M. (1994). Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes. *Nucleic Acids Res*, **22**, 4574-4582.
- Schedl, A., Ross, A., Lee, M., Engelkamp, D., Rashbass, P., van, H. V. et Hastie, N. D. (1996). Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell*, **86**, 71-82.
- Schier, A. F. et Gehring, W. J. (1992). Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature*, **356**, 804-807.
- Scott, M. P. et Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. *Proc Natl Acad Sci U S A*, **81**, 4115-4119.
- Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G. et Gruss, P. (1997). The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature*, **386**, 399-402.
- Spörle, R., Gunther, T., Struwe, M. et Schughart, K. (1996). Severe defects in the formation of epaxial musculature in open brain (opb) mutant mouse embryos. *Development*, **122**, 79-86.
- St Onge, L., Sosa, P. B., Chowdhury, K., Mansouri, A. et Gruss, P. (1997). Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. *Nature*, **387**, 406-409.
- Stapleton, P., Weith, A., Urbanek, P., Kozmik, Z. et Busslinger, M. (1993). Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX-9. *Nat Genet*, **3**, 292-298.
- Stoykova, A. et Gruss, P. (1994). Roles of Pax-genes in developing and adult brain as suggested by expression patterns. *J Neurosci*, 1395-1412.
- Tassabehji, M., Read, A. P., Newton, V. E., Harris, R., Balling, R., Gruss, P. et Strachan, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene [see comments]. *Nature*, **355**, 635-636.
- Tassabehji, M., Read, A. P., Newton, V. E., Patton, M., Gruss, P., Harris, R. et Strachan, T. (1993). Mutations in the PAX3 gene causing Waardenburg syndrome type 1 and type 2. *Nat Genet*, **3**, 26-30.
- Timmons, P. M., Wallin, J., Rigby, P. W. et Balling, R. (1994). Expression and function of Pax 1 during development of the pectoral girdle. *Development*, **120**, 2773-2785.
- Treisman, J., Harris, E. et Desplan, C. (1991). The paired box encodes a second DNA-binding domain in the paired homeo domain protein. *Genes Dev*, **5**, 594-604.
- Turque, N., Denhez, F., Martin, P., Planque, N., Bailly, M., Begue, A., Stehelin, D. et Saule, S. (1996). Characterization of a new melanocyte-specific gene (QNR-71) expressed in v-myc-transformed quail neuroretina. *Embo J*, **15**, 3338-3350.
- Urbanek, P., Wang, Z. Q., Fetka, I., Wagner, E. F. et Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP [see comments]. *Cell*, **79**, 901-912.
- Vesque, C., Maconochie, M., Nonchev, S., Ariza, M. L., Kuroiwa, A., Charnay, P. et Krumlauf, R. (1996). Hoxb-2 transcriptional activation in rhombomeres 3 and 5 requires an evolutionarily conserved cis-acting element in addition to the Krox-20 binding site [published erratum appears in EMBO J 1996 Dec 16;15(24):7188]. *Embo J*, **15**, 5383-5396.
- Wallin, J., Mizutani, Y., Imai, K., Miyashita, N., Moriwaki, K., Taniguchi, M., Koseki, H. et Balling, R. (1993). A new Pax gene, Pax-9, maps to mouse chromosome 12. *Mamm Genome*, **4**, 354-358.
- Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B. et Balling, R. (1994). The role of Pax-1 in axial skeleton development. *Development*, **120**, 1109-1121.
- Walther, C. et Gruss, P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development*, **113**, 1435-1449.
- Walther, C., Guenet, J. L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R. et Gruss, P. (1991). Pax: a murine multigene family of paired box-containing genes. *Genomics*, **11**, 424-434.
- Warren, N. et Price, D. J. (1997). Roles of Pax-6 in murine diencephalic development. *Development*, **124**, 1573-1582.
- Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. et Desplan, C. (1993). Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev*, **7**, 2120-2134.
- Wilson, D. S., Guenther, B., Desplan, C. et Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709-719.
- Wolberger, C., Vershon, A. K. et Liu, B. (1991). Crystal structure of a MATA2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, **67**, 517-528.
- Xu, W., Rould, M. A., Jun, S., Desplan, C. et Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell*, **80**, 639-650.
- Yamaguchi, Y., Sawada, J., Yamada, M., Handa, H. et Azuma, N. (1997). Autoregulation of Pax6 transcriptional activation by two distinct DNA-binding subdomains of the paired domain. *Genes Cells*, **2**, 255-261.

- Zannini, M., Francis, L. H., Plachov, D. et Di, L. R. (1992). Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol Cell Biol*, **12**, 4230-4241.
- Zappavigna, V., Sartori, D. et Mavilio, F. (1994). Specificity of HOX protein function depends on DNA-protein and protein-protein interactions, both mediated by the homeodomain. *Genes & Dev*, **8**, 732-744.
- Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A. et Scott, M. P. (1993). Ectopic expression and function of the Antp and Scr homeotic genes: the N terminus of the homeodomain is critical to functional specificity. *Development*, **118**, 339-352.

*Seconde partie*

*Introduction*  
*Récepteurs Nucléaires*



**Figure 13** : Schéma d'un récepteur nucléaire sous forme linéaire montrant les différents domaines fonctionnels :

**A/B** = domaines A/B qui contiennent la fonction de transactivation indépendante du ligand, notée AF-1.

**C** = domaine C, composé des deux doigts de zinc (en X rouge sur la figure) et des deux premières hélices situées dans les boîtes P et D des doigts F1 et F2, qui interagissent avec l'ADN. F2 participe aussi à la dimérisation des récepteurs. Des sites de localisation nucléaire ont aussi pu être localisés s'étendant de la fin du domaine C au début du domaine E.

**D** = domaine, dit charnière, contient les boîtes T/A qui participent à la fixation à l'ADN et à la dimérisation entre récepteurs.

**E** = domaine de fixation du ligand, participe aussi à la dimérisation des récepteurs. Il contient notamment la région AF2-AD autonome qui interagit avec les coactivateurs pour stimuler la transcription. De plus, différentes régions s'étendant sur les domaines E et D, selon les récepteurs, fixent les cofacteurs transcriptionnels.

**F** = domaine carboxy terminal facultatif, fonction inconnue.

## A Les récepteurs nucléaires, structure et fonction

### I Caractéristiques des récepteurs nucléaires

#### 1 Une famille de plus de 60 membres !

##### 1-1 Caractéristiques structurales d'un récepteur

Le clonage des premiers récepteurs nucléaires date du milieu des années 1980 (voir revue dans (Gronemeyer et Laudet, 1995)). La comparaison de leurs séquences en acides aminés et les analyses fonctionnelles qui ont suivi l'isolement de ces gènes ont mis en évidence l'existence d'une organisation structurale et fonctionnelle commune: six domaines constituent la protéine. Le domaine A/B, dans la partie amino terminale de la protéine, contient un domaine de transactivation noté AF-1 pour "Activation Function 1". Le domaine C, composé de 83 acides aminés, fixe l'ADN et permet également la dimérisation entre récepteurs. Derrière le domaine C, on trouve un domaine D dit "charnière", car on ne lui attribuait pas de fonction spécifique. Cependant, il est constitué d'une boîte T et A, qui nous le verrons, dans la partie 2 (Eléments de structure), participe également à la liaison à l'ADN. Le domaine E fixe le ligand et participe aussi à la dimérisation des récepteurs. Il contient de plus un domaine d'activation dépendant du ligand, appelé AF-2 (pour "Activation Function 2"). Enfin, en carboxy-terminal, le domaine F, facultatif, n'est pas retrouvé dans toutes les protéines (cf. figure 13).

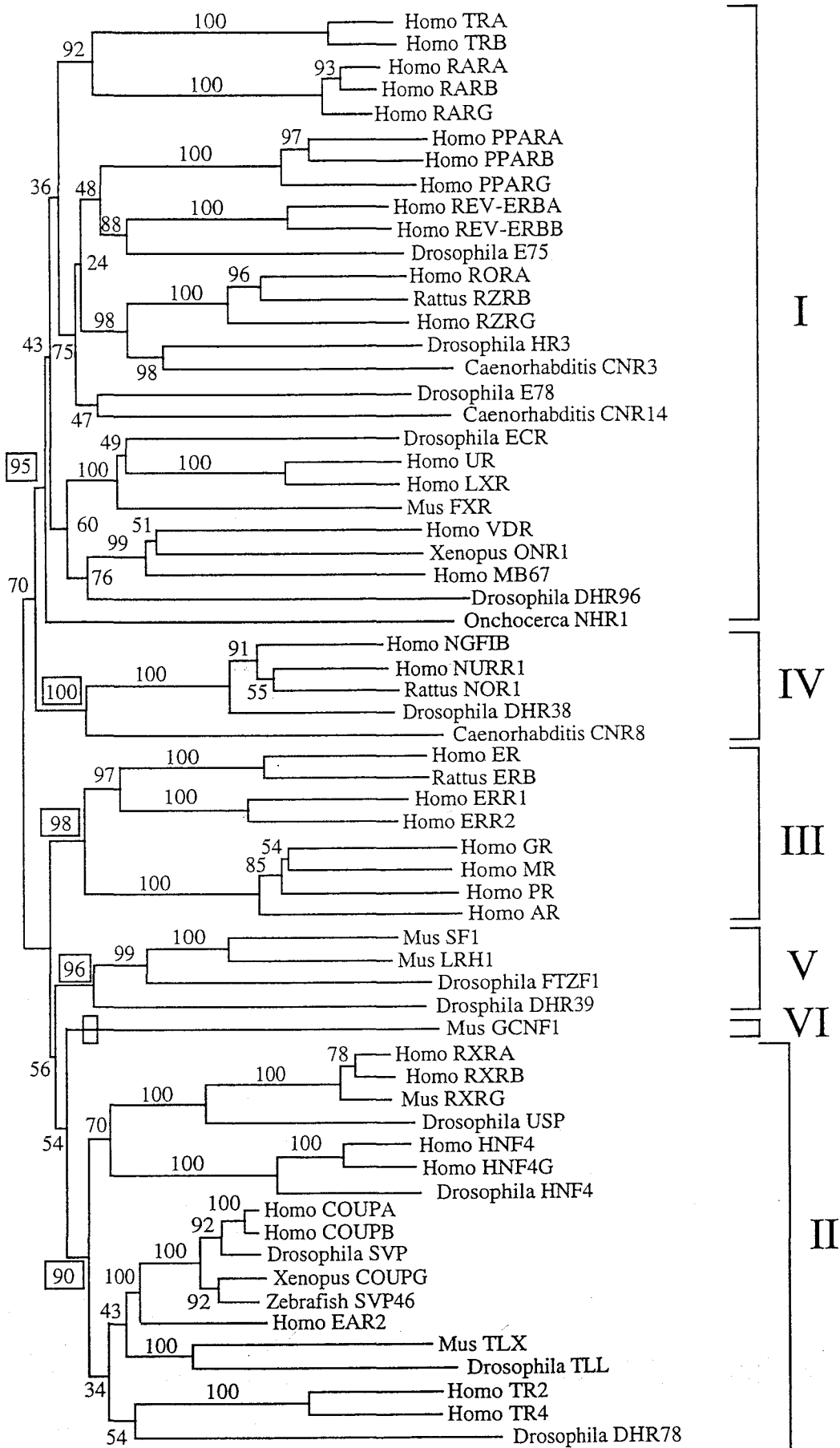
##### 1-2 Composition de la famille

Les récepteurs nucléaires constituent une grande famille de plus de 60 membres. L'appartenance à cette famille se définit par l'existence dans la protéine d'un domaine C et/ou d'un domaine E, domaines les plus conservés entre les récepteurs et au cours de l'évolution (Laudet *et al.*, 1992).

Un arbre phylogénétique établi à partir des domaines C et E (cf. figure 14) détermine l'existence de 6 sous familles (Escriva *et al.*, 1997a).

Parmi ces familles, on rencontre des récepteurs dit "classiques", qui comportent un ligand pour médier leur activité transcriptionnelle, tel que les hormones thyroïdiennes pour les récepteurs TR, les stéroïdes (progestérone oestrogènes, androgènes, minéralocorticoïdes et glucocorticoïdes), la vitamine D, les acides rétinoïques, dérivés de la vitamine A, etc... De part la nature de ces ligands, il





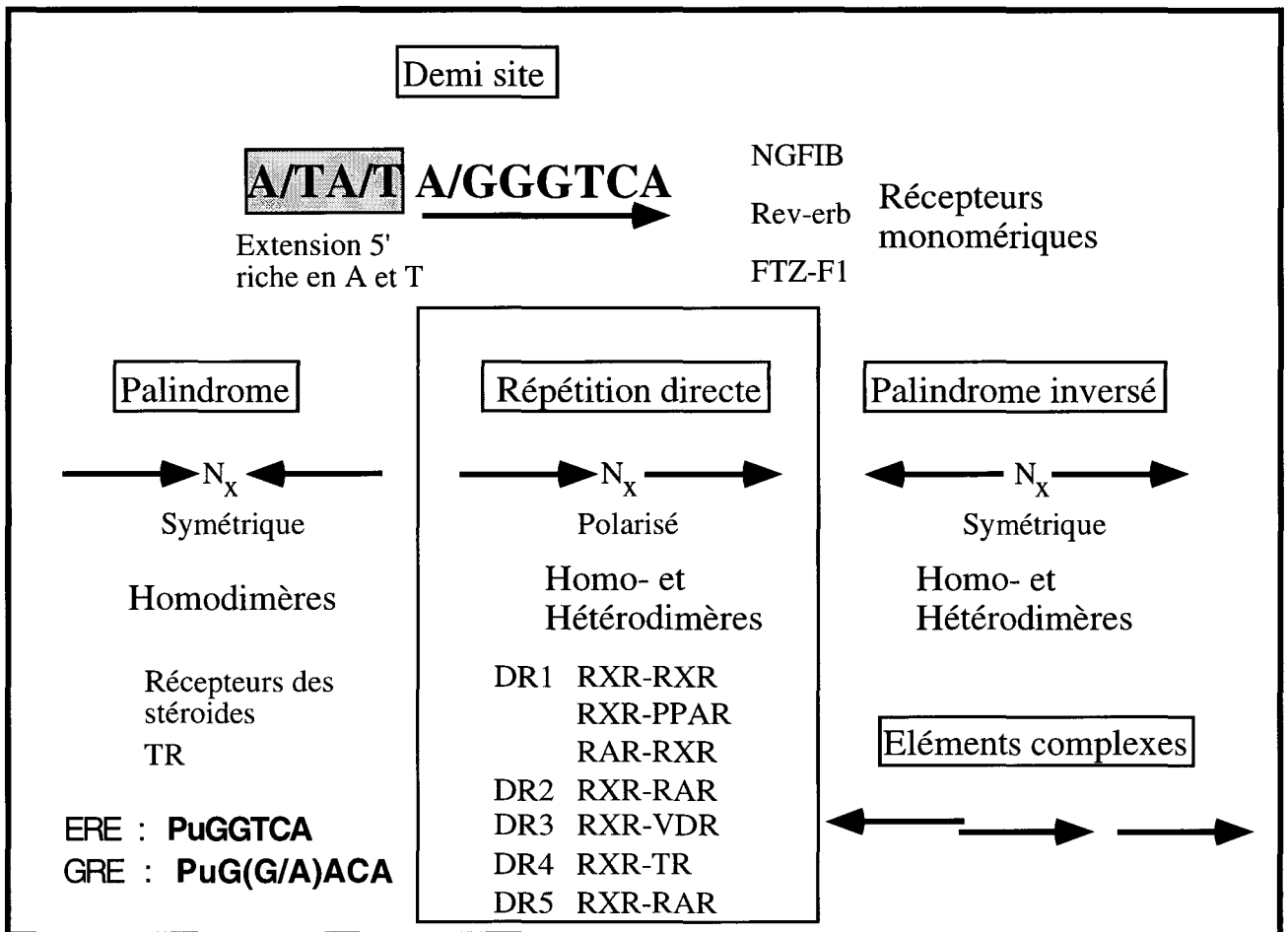
**Figure 14 :** Arbre phylogénétique (réalisé à partir des domaines C et E) montrant l'existence de 6 sous familles dans la superfamille des récepteurs nucléaires. (ref: Laudet *et al.*, 1997)

va de soit que les récepteurs nucléaires participent à l'homéostasie des organismes. Les acides rétinoïques jouent aussi un rôle dans l'organisation antéro-postérieure chez l'embryon et dans le développement du système nerveux (cf. (Sucov et Evans, 1995; Clagett et Plum, 1997), pour revues).

Il existe par ailleurs, des récepteurs pour lesquels aucun ligand n'a pu être identifié, on les appelle « récepteurs orphelins ». Ils sont distribués dans les 6 sous familles. Actuellement, rien ne permet de conclure sur la question: Sont-ils véritablement dépourvus de ligand ou bien leur ligand n'est pas encore connu? L'existence de ligands a pu être mise en évidence ces dernières années pour quelques récepteurs. Les prostaglandines et les leucotriènes ont ainsi été attribués aux PPAR (PPAR = Peroxisome Proliferator Activated Receptors) (Schoonjans *et al.*, 1997), pour revue), des récepteurs nucléaires dont l'activité est régulée par des molécules qui provoquent la prolifération des péroxisomes. Néanmoins, à l'heure actuelle, de nombreux récepteurs restent encore dépourvus de ligand, cas des gènes COUP-TF, TR4, NGFI-B qui sont notamment impliqués dans le développement du système nerveux (Pereira *et al.*, 1995; Maruyama *et al.*, 1998).

A quelques exceptions près, tous les récepteurs sont dotés de domaines C et E. Citons les exemples du récepteur orphelin DAX-1, chez l'humain, dont la mutation entraîne une déficience hormonale causée par l'hypoplasie congénitale de la glande adrénale (Zanaria *et al.*, 1994). DAX-1 code un récepteur particulier qui comporte un domaine E classique et un domaine de fixation composé de simples motifs répétés qui lui permettent cependant d'interagir avec des cibles communes aux autres récepteurs (Zanaria *et al.*, 1994). Knirps, un récepteur nucléaire impliqué dans la segmentation de la drosophile, appartient aussi à cette catégorie de récepteurs inhabituels. Celui-ci contient en effet uniquement un domaine C (Gerwin *et al.*, 1994).

Dans les paragraphes suivants, nous traiterons de l'apport des domaines C et E dans la fonction des récepteurs. (i) Comment le domaine C se fixe à l'ADN et quelles sont les cibles reconnues. (ii) Comment le domaine E fixe son ligand dans le cas des récepteurs classiques et quels sont les types de ligands rencontrés..



**Figure 15 : Eléments de réponse sur l'ADN des récepteurs nucléaires**

L'hexamère A/GGGTCA constitue l'élément de reconnaissance de base sur lequel se fixent les récepteurs nucléaires. Il peut-être agencé en palindrome, en répétition directe ou en palindrome inversé. N<sub>x</sub> représente le nombre de bases situées entre deux demi sites.

Certains récepteurs orphelins, NGFIB, FTZ-F1 reconnaissent des monosites précédés en 5' de séquences riches en A et T. RAR et RXR se fixent préférentiellement sur des répétitions directes, TR sur un palindrome de type Pal 0. Les récepteurs aux stéroïdes reconnaissent des sites palindromiques dont les bases de l'hexamère divergent quelque peu en 5' (cf texte).

(d'après Gronemeyer et Laudet *et al.*, 1995)

## 2 Mode de fixation à l'ADN

### 2-1 Vue d'ensemble

Les récepteurs nucléaires se fixent sur l'ADN le plus souvent en dimères, **homodimères** ou **hétérodimères** (essentiellement avec RXR) et parfois en **monomère**, cas notamment de certains récepteurs orphelins. Ils reconnaissent sur l'ADN un élément de réponse, noté HRE pour "Hormone Responsive Element" (revue dans (Gronemeyer et Laudet, 1995)). L'hexamère **A/GGT/GTCA** a été établi comme étant le consensus pour leurs demi sites de fixation. Les sites sont composés de deux hexamères orientés en répétition directe, **DR**, en palindrome, **Pal**, ou en répétition inversée, **IR**. Les demi sites sont le plus souvent précédés d'une séquence riche en A. De manière générale, les récepteurs nucléaires se fixent principalement sur des DR et des palindromes (Gronemeyer et Laudet, 1995) (figure 15).

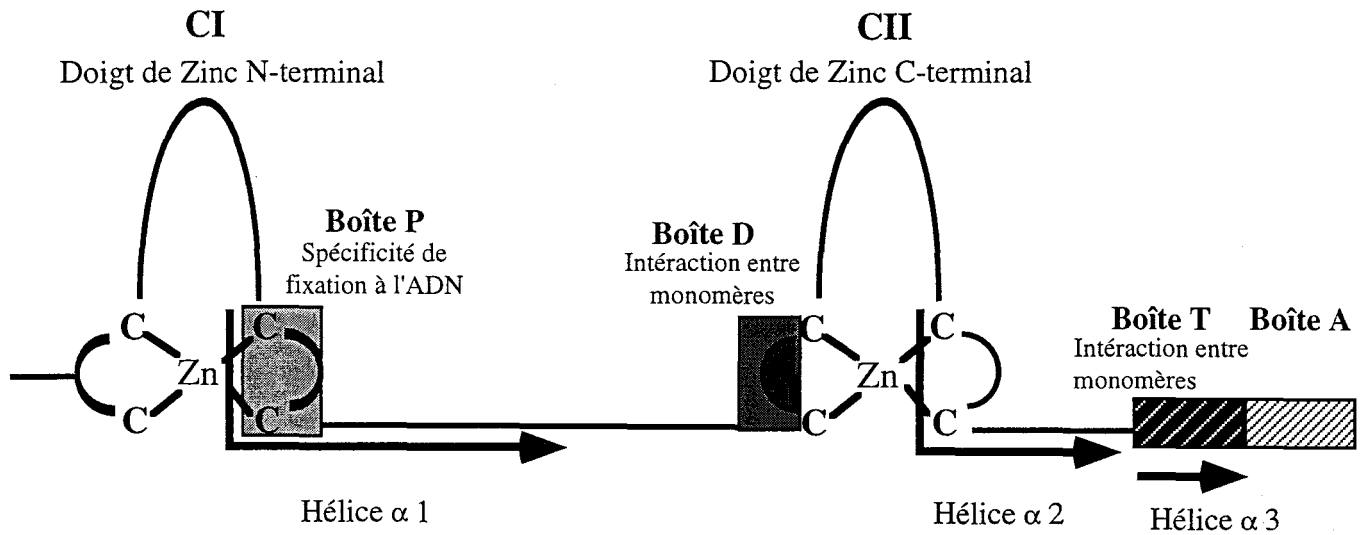
### 2-2 Structure du domaine de fixation à l'ADN

#### 2-2-1) Les régions du domaine impliquées dans la fixation à l'ADN

Des délétions réalisées, sur les récepteurs ER et GR ont permis d'identifier le segment peptidique impliqué dans la fixation à l'ADN (Danielsen *et al.*, 1986; Hollenberg *et al.*, 1987; Kumar *et al.*, 1987; Rusconi et Yamamoto, 1987) qui s'est avéré constituer la région la plus conservée des récepteurs, le domaine C, domaine long de 66 acides aminés (voir (Evans et Hollenberg, 1988) pour revue et (Gronemeyer et Laudet, 1995)) pour revue récente). Nous verrons que la séquence directement en aval du domaine C (portion amino terminale du domaine D) participe aussi à la fixation à l'ADN.

La région C est composée de motifs conservés sur l'ensemble des récepteurs. On distingue notamment deux groupes de 4 cystéines, séquence évocatrice de deux **doigts de zinc**. Le premier doigt, en amino terminal du domaine C, F1 exhibe le motif Cys-X2-Cys-X13-Cys-X2-Cys pour le récepteur à l'acide *9-cis* rétinolique, RXR. Le second, F2, en carboxy terminal du domaine C, contient le motif Cys-X5-Cys-X9-Cys-X2-Cys, pour le même récepteur. Dans chacun de ces motifs, deux paires de cystéines chélatent un ion Zn<sup>2+</sup>. La coordination du zinc est importante pour l'intégrité de la structure et la capacité de liaison à l'ADN de la protéine (Zilliaccus *et al.*, 1992). Deux boîtes intervenant dans la fixation à l'ADN ont été définies dans le domaine C. La première, la **boîte P**, a été identifiée pour son implication dans la spécificité de reconnaissance à l'ADN. La seconde, la **boîte D**, contribue aux interfaces de contacts

A



B

		Boîte P		
GR	419	KLCLVCSDEASGCHYGVLTCGSC	●	●
PR	565	KICLICGDEASGCHYGVLTCSSCKV	●	●
AR	555	KTCLICGDEASGCHYGALTCGSC	●	●
MR	601	KICLVCGDEASGCHYGVVTCGSC	●	●
ER	183	KYCAVCNEYASGYHYGVWSCEGCK	●	●
RARα	56	KPCFVCQDKSSGYHYGVSA	●	●
TRB	100	ELCVVCGDKATGYHYRCITCEGCK	●	●
VDR	22	RICGVCGDRATGFHFNAMTCEGCK	●	●
RXR	133	HICAI CGDRSSGKHYGVVSC	●	●
NGFI-B	264	GRCVAVCGDNASCQHYGVRTCEGCK	●	●
			○	○
		Boîte D	Boîte T	Boîte A
GR	457	CAGRNDCIIDKIRRKNCPACRYR	●	●
PR	603	CAGRNDCIIVDKIRRKNCPACRL	●	●
AR	593	CASRNDCTIDKFRRKNCPSCRL	●	●
MR	639	CAGRNDCIIDKIRRKNCPACRL	●	●
ER	221	CPATNQCTIDKNRRKSCQACRL	●	●
RARα	94	CHRDKNCIINKVTRNRRCQYCR	●	●
TRB	140	CKYEGKCVLDKVTRNQCECR	●	●
VDR	60	CPFNGDCRITKDNRRHCQACRL	●	●
RXR	171	CRDNKDCIIDKRQRNRRCQYCR	●	●
NGFI-B	302	CLANKDCPVDKRRRNRRCQFCR	●	●
			○	○
			○	○

**Figure 16 :** (A) Le domaine C des récepteurs nucléaires. Les deux atomes de zinc (Zn) sont maintenus par une structure tétraédrique de 4 cysteines (C). Les boîtes P et A importantes pour la fixation à l'ADN et les boîtes D et T intervenant dans la dimérisation sont indiquées. Les hélices  $\alpha$ , présentes dans le domaine C du récepteur RXR, sont indiquées. (B): Séquences en acides aminés du domaine de fixation à l'ADN de plusieurs récepteurs nucléaires. Les cysteines qui sont coordonnées avec un atome de zinc sont indiquées par cercle noir (●). Le cercle blanc (○) indique une cysteine conservée qui n'interagit pas avec un atome de zinc. Les trois résidus de la boîte P qui interviennent dans la spécificité de reconnaissance de la cible ADN sont encadrés, la boîte D qui intervient dans la mise en place d'une interface de dimérisation et les boîtes T et A qui contribuent à une meilleure fixation à l'ADN, sont aussi précisées (d'après Zilliaccus et al, 1995).

dans les dimères. Deux autres boîtes localisées en aval du domaine C, dans la région charnière du domaine D, les boîtes T et A participent au choix de l'espacement entre deux demi-sites ou aux bases en 5' d'un monosite, ainsi qu'à la surface d'interaction entre deux monomères (cf. figure 16).

### 2-2-2) Structure tridimensionnelle du domaine de liaison à l'ADN

Les structures tridimensionnelles des DBD de RXR (déterminée par RMN par (Lee *et al.*, 1993) et de GR (cristallographie par (Hard *et al.*, 1990b; Luisi *et al.*, 1991)) et ER (Schwabe *et al.*, 1990), sont similaires, à l'exception d'une hélice supplémentaire, chez RXR: l'hélice 3. Les acides aminés du premier doigt de zinc jusqu'à la troisième cystéine forment un feuillet  $\beta$  et le reste du domaine C est caractérisé par la présence de deux hélice  $\alpha$  pour GR et ER, et trois hélices  $\alpha$  pour RXR. La première, située en N terminal du premier doigt de zinc, juste derrière la troisième cystéine (appelée aussi boîte P sur la structure primaire), reconnaît l'HRE sur l'ADN et vient se coller dans le grand sillon. La seconde est localisée entre les deux premières cystéines du deuxième doigt de zinc. Elle se place de manière perpendiculaire à la première et vient stabiliser la complexe ADN/protéine (cf. figure 17) (Lee *et al.*, 1993). La troisième hélice en carboxy terminal du deuxième doigt de zinc, est spécifique à RXR; on ne la retrouve pas dans les protéines GR, ER ou RAR $\beta$  (Katahira *et al.*, 1992). Cette hélice se serre contre l'hélice 1 et l'extrémité N-terminale du premier doigt de zinc. Dans cette orientation, l'hélice 3 se projette vers le sillon mineur de l'ADN et établit des liens avec le squelette phosphaté de l'ADN renforçant ainsi la stabilité du complexe.

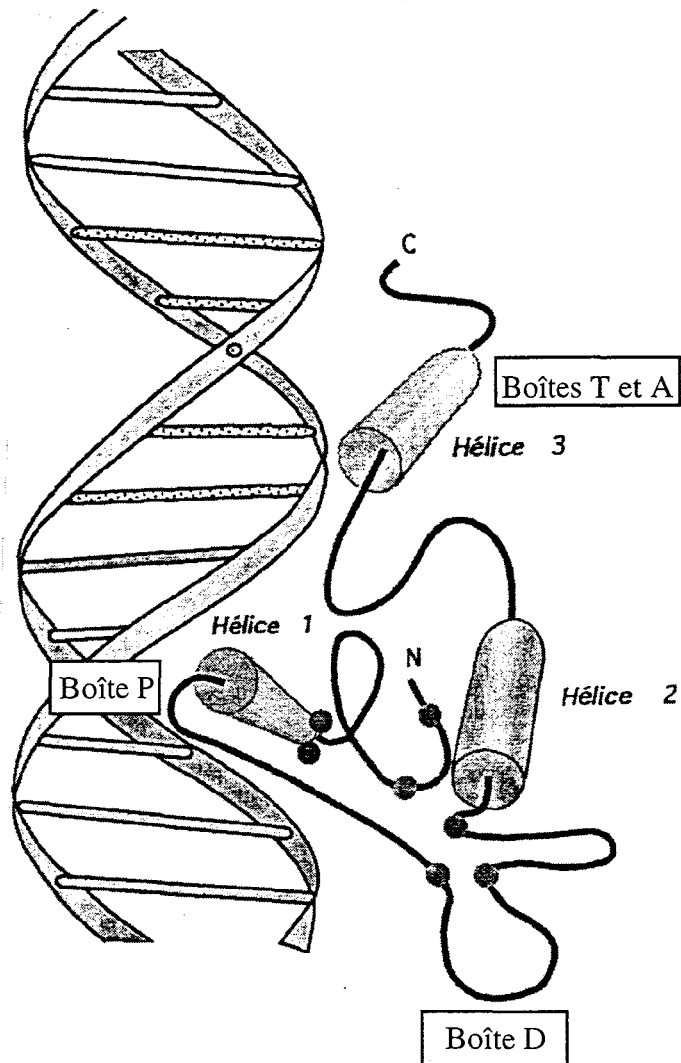
## 2-3 Spécificité de reconnaissance à l'ADN

### 2-3-1) L'ADN : l'élément de réponse

Dans le cas des dimères, la spécificité de liaison à l'ADN est conférée à la fois par la séquence du demi site et le nombre de nucléotides qui séparent ces deux demi sites (Näär *et al.*, 1991).

#### a) *Les palindromes et les palindromes inversés*

Les récepteurs aux hormones stéroïdiennes se fixent en homodimère sur des séquences agencées en répétition inversée (palindrome), de demi site AGAACA pour GR, AR et MR et AGGTCA pour ER, séparées par 3 nucléotides (Evans et Hollenberg, 1988; Beato, 1991; Glass, 1994) pour revue). De plus, TR, le récepteur aux



**Figure 17 :** Représentation schématique de la structure tridimensionnelle du domaine de fixation à l'ADN de RXR. La boîte P forme une hélice  $\alpha$  qui interagit avec le grand sillon de l'ADN. De plus dans la partie carboxy-terminale du domaine C les boîtes T et A forment une troisième hélice  $\alpha$  qui interagit directement avec le petit sillon de l'ADN et stabilise alors le contact récepteur/ADN (d'après Adelmant et Laudet, 1995).

hormones thyroïdiennes, peut aussi se fixer sur un palindrome inversé (appelé IR3 pour "Inverted Repeat" séparé par 3 paires de bases) (Näär *et al.*, 1991).

*b) Les répétitions directes*

La spécificité de l'élément de réponse pour un élément en répétition directe donné est basée sur l'espacement entre les deux demi sites. Pour les récepteurs couplés avec RXR, cet espace est de 5 nucléotides pour l'élément de réponse en répétition directe de RAR (appelé DR5 pour "Direct Repeat" séparé de 5 pb), de 4 pour l'élément de réponse de TR (DR4), de 3 nucléotides pour celui en répétition directe de VDR (DR3), tandis qu'il n'est que d'un nucléotide pour les récepteurs de type PPAR (DR1) (Brent *et al.*, 1989; de Thé *et al.*, 1990; Tugwood *et al.*, 1992).

Par ailleurs, RXR peut aussi former des homodimères qui interagissent de manière très affine avec des HRE de type DRs (Mangelsdorf *et al.*, 1991; Kliewer *et al.*, 1992) et activent la transcription en réponse à l'acide 9-cis rétinolique (Yang *et al.*, 1991; Heyman *et al.*, 1992; Kliewer *et al.*, 1992; Levin *et al.*, 1992; Zhang *et al.*, 1992). Dans ce dernier cas, le dimère des domaines C de RXR se fixe préférentiellement en homodimère, sur un DR1 de demi site: AGGTCA. Ce résultat est applicable à la protéine entière, car il a été montré que le domaine C isolé présentait la même spécificité pour l'ADN que la protéine entière (Kliewer *et al.*, 1992).

*c) Les demi-sites*

D'autres récepteurs nucléaires, et plus particulièrement certains récepteurs orphelins comme NGF-IB, SF-I, REV-ERB et ERRI, se fixent en monomère sur des demi-sites de type AGGTCA précédés en 5', de séquences riches en A, par exemple, AAA AGGTCA pour le site NBRE et TCA AGGTCA pour le site SFRE, reconnu par SF-1 et ERR-1 (Wilson *et al.*, 1993a; Bonnelye *et al.*, 1997).

*d) Flexibilité dans la spécificité de reconnaissance*

Les consensus définis ci-dessus par type de récepteur représentent les séquences pour lesquelles les récepteurs considérés possèdent la meilleure affinité. Néanmoins, il est bon de garder à l'esprit que l'espacement peut varier. Ainsi l'hétérodimère RXR/RAR est aussi capable de se fixer sur un DR4, puis de manière encore moins affine, sur un DR2 ou DR3. De plus, la séquence nucléotidique elle-même peut souffrir quelques écarts. Cette flexibilité de reconnaissance à l'ADN peut entraîner des processus d'interférences *in vivo* entre différents récepteurs.



## 2-3-2) Le récepteur : les boîtes impliquées

### a) La boîte P

La construction de protéines chimériques a permis d'illustrer le rôle de cette boîte dans la spécificité de reconnaissance à l'ADN. Elle est constituée de 5 acides aminés, incluant la dernière cystéine du doigt de zinc CI, de motif EGCKG pour RXR et la majorité des récepteurs (pour revue cf. (Parker, 1990; Mangelsdorf *et al.*, 1995)). Les boîte P des récepteurs aux oestrogènes (ER), (de séquence **EGCKA**) et aux glucocorticoïdes (GR) (de séquence **GSCKV**) diffèrent par 3 acides aminés (imprimés en caractères gras). ER se fixe sur l'hexamère AGGTCA, tandis que GR reconnaît le site AGAACA. Le remplacement des 3 acides aminés dans la boîte P de ER par les acides aminés équivalents de GR, confère au récepteur chimérique ER la spécificité de fixation au site AGAACA (=GRE) (Danielsen *et al.*, 1989; Umesono et Evans, 1989). De même, la substitution du résidu glycine (G) dans la boîte P du récepteur GR, par le résidu acide glutamique (E) présent dans les récepteurs TR, ER, RXR, RAR, suffit au récepteur muté à reconnaître un ERE tout en restant capable de se fixer sur un GRE (Umesono et Evans, 1989; Parker, 1990).

### b) La boîte D

Elle est responsable de la sélection entre un site palindromique sans espaceur Pal0 et un site palindromique comportant 3 bases dans l'espaceur, Pal3 (PP). ER se fixe sur un motif palindromique en orientation inverse dont les deux demi-sites sont espacés de 3 nucléotides (IR3). TR se fixe sur un Pal 0. Or, la substitution de la boîte D de ER par celle de TR, lui confère la capacité à se fixer sur la motif TRE pal (IR0). Ce récepteur ER muté n'est plus apte à reconnaître un motif de type IR3 (Umesono et Evans, 1989). La boîte D, nous l'avons vu, participe à l'interface entre deux dimères. D'après les résultats de Umesono (Umesono et Evans, 1989), les acides aminés de la boîte D, impliqués dans l'interface du dimère, vont déterminer la distance entre les deux monomères et conséquemment, l'espace entre les deux demi sites de la séquence de reconnaissance sur l'ADN.

### c) La boîte A

La boîte A fut décrite originellement dans le cas du récepteur orphelin NGFIB (Wilson *et al.*, 1993b). Ce récepteur, reconnaît sous forme de monomère, un demi site précédé en 5' par la séquence spécifique AA. L'ensemble de ce site est appelé un NBRE. Par des expériences de mutagenèse, il a été montré que la boîte A de NGFIB interagit spécifiquement avec 2 nucléotides dans le petit sillon de la molécule

d'ADN. On peut aussi obtenir un NGFIB capable de se fixer sur la séquence de SF1 (le SFRE : TCAAGGTCA), en plaçant la boîte A de SF1 à la place de celle de NGFIB (Wilson *et al.*, 1993b).

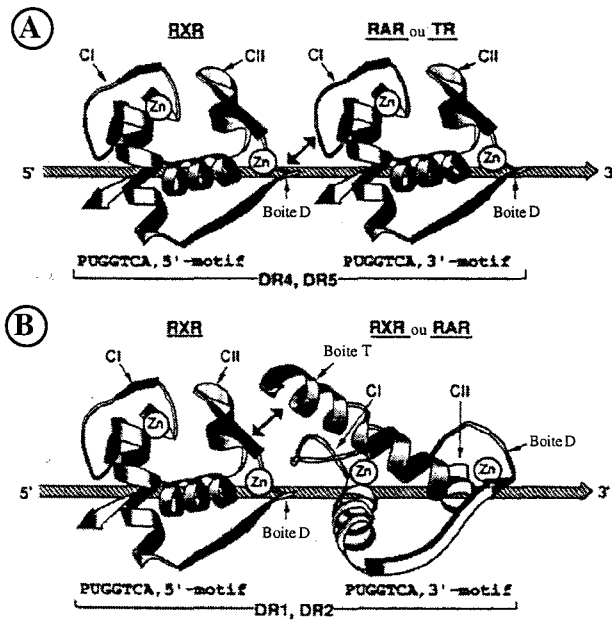
## 2-4 Fixation en monomère , homo- ou hétéro- dimère

### 2-4-1) Caractéristiques des dimères

Dans tous les complexes ADN/récepteurs fixés sous forme de dimères, les deux partenaires interagissent avec deux régions adjacentes du sillon majeur de l'ADN. C'est le cas par exemple, de l'homodimère GR (Luisi *et al.*, 1991), comme celui de ER, ainsi que des hétérodimères couplés avec RXR. De plus, ces dimères se fixent de manière coopérative sur l'ADN. Ceci signifie que la fixation du premier monomère sur l'élément de réponse augmente l'affinité du second monomère pour l'autre demi-site (Dahlman-Wright *et al.*, 1990; Hard *et al.*, 1990a). Par exemple, l'affinité du second monomère du dimère de GR pour un GRE est accrue de 10 fois en comparaison de l'affinité du premier monomère pour le demi-site (Lundback *et al.*, 1994). Par ailleurs, il existe une polarité dans l'ordre de placement des partenaires des hétérodimères en fonction de l'espacement situé entre les deux sites en répétition directe de leur élément de réponse (cf. ci-dessous). De plus, il est à noter que l'orientation des monomères est fonction de l'élément de réponse sur lequel ils se fixent. Sur un palindrome par exemple, les récepteurs sont placés tête-bêche, cas des récepteurs aux hormones stéroïdes, ou en tête à queue, sur les palindromes inversés. Dans le cas des répétitions directes au contraire, les récepteurs seront fixés dans la même orientation et on aura donc une préface d'interaction entre les deux monomères différente de celle présentée par les homodimères.

### 2-4-2) Exemples d'hétérodimères

Les hétérodimères les plus fréquemment rencontrés sont ceux formés avec RXR, tels que RXR/RAR, RAR/RXR, RXR/TR, RXR/VDR (Par convention, le récepteur situé en 5' sur l'élément de réponse est toujours noté à gauche de l'hétérodimère). L'existence d'hétérodimères entre RXR et COUP-TF (Kliwer *et al.*, 1992), entre NGFI-B et COUP-TF et entre NGFI-B et RXR (Wu *et al.*, 1997) a également été proposée. Cependant, l'existence d'hétérodimères formés entre RXR et COUP-TF a longtemps été discutée, entre autres par Butler *et al.*, (Butler et Parker, 1995), qui à l'aide d'un système double hybride en cellules intactes, ont pu détecter les hétérodimères RXR $\alpha$ /RAR $\alpha$ , les homodimères COUP-TFII, mais pas les



**Figure 18:** En fonction de la nature du partenaire de RXR, les domaines intervenant dans l'interface de dimérisation vont varier. (A) hétérodimère RXR-TR: l'interface de dimérisation sur un site DR4 comprend la boite D de RXR et le doigt CI de TR. (B) hétérodimère RXR-RAR: l'interface de dimérisation comprend le doigt CII de RXR et la boite T de RAR. (D'après Zechel *et al*, 1994b)

hétérodimères entre COUP-TFII et RXR $\alpha$ , ni en solution, ni sur un nombre varié d'éléments de réponse. Par ailleurs les hétérodimères RXR $\alpha$ /TR $\beta$  sont favorisés par rapport aux hétérodimères COUP-TFII/TR $\beta$ . Pourtant, les hétérodimères entre RXR et ARP-1 ont pu être détectés sur le promoteur de CRBP2 (Nakshatri et Chambon, 1994). Il est probable que la formation d'hétérodimères entre COUP-TF et RXR dépende de l'élément de réponse et du promoteur sur lequel ils se fixent.

#### 2-4-3) Polarité des hétérodimères RAR et RXR

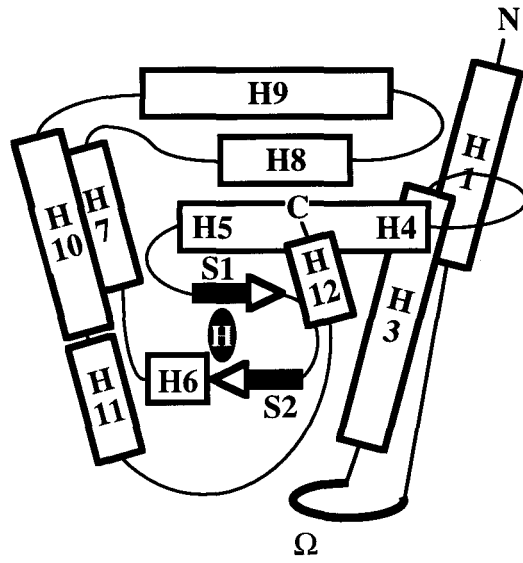
Perlmann *et al.* ont cherché à savoir si des deux partenaires RAR et RXR, l'un des deux avait une place préférentielle sur un DR donné, en amont ou en aval du site (Perlmann *et al.*, 1996). Pour répondre à cette question, ils ont remplacé la boîte P de RAR par celle de GR, afin de que le nouveau RAR<sup>GR</sup> ne puisse reconnaître que le demi site AGAACA. Ils ont ainsi mis en présence RAR<sup>GR</sup> et RXR et ont pu évaluer leur position respective sur AGGTCA(N)<sub>5</sub>AGAACA ou sur AGAACA(N)<sub>5</sub>AGGTCA (Kurokawa *et al.*, 1993; Perlmann *et al.*, 1993). C'est ainsi que sur un DR5, RAR a été montré capable de se fixer en 3' et RXR, en 5' de l'élément de réponse. De plus, les études de (Zechel *et al.*, 1994b), ont permis d'une part, de positionner RXR en 5' et RAR en 3' sur un DR2 et d'autre part, d'identifier les zones d'interface entre les deux monomères (cf. paragraphe suivant). L'hétérodimère RAR/RXR existe mais est transcriptionnellement inactif.

#### 2-4-4) Interfaces de dimérisation

Dans tous les cas les contacts entre deux monomères d'un dimère s'établissent dans le sillon mineur de l'ADN. Dans les exemples cités ci-dessous, les zones de contact entre deux récepteurs ont été soit observées à partir d'une étude de leur structure tridimensionnelle, soit déduites suite à des expériences de mutagenèse. Les mutants obtenus ayant été testés par des expériences de retard sur gel, pour leur capacité à se lier à l'ADN sous forme de monomère ou de dimère. Voici quelques résultats obtenus par différents auteurs (figure 18) :

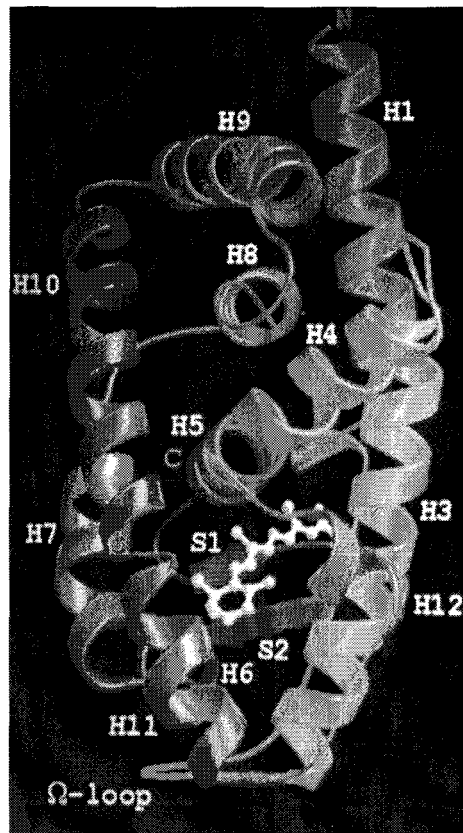
##### - *homodimères* :

Les deux partenaires sont positionnés tête-bêche sur un élément de réponse palindromique. Le dimère formé présente donc une symétrie axiale qui passe par le centre de la séquence de reconnaissance de l'ADN. L'interface de dimérisation s'étend des boîtes D de chacun des deux récepteurs jusqu'au sommet du doigt CII (Luisi *et al.*, 1991); étude cristallographique). Les récepteurs stéroïdiens comme GR ou ER forment ce type de complexe.



**a) Motifs structuraux du domaine E de holo-RAR $\gamma$ .**

Les hélices sont indiquées par H suivi de leur n°, S1 et S2 sont les feuillets bêta.



**Figure 19: b) Structure tridimensionnelle du domaine E de RAR $\gamma$ .**

(D. Moras)

- *hétérodimères* :

Les hétérodimères sont asymétriques. En conséquence, la surface de contact entre les deux récepteurs du complexe entre hétérodimères n'implique pas les mêmes acides aminés que ceux réalisant l'interface d'un complexe homodimérique. De plus la distance entre deux demi sites ajoute aux variations observées. Quelques exemples sont détaillés ci dessous.

Sur un **DR5**, pour l'hétérodimère RXR/RAR, l'interface de dimérisation est réalisée entre la boîte D du doigt de zinc CII de RXR, placé à gauche du complexe (en 5' sur l'ADN), et le sommet du doigt CI de RAR placé en dernier sur le site (Zechel *et al.*, 1994b).

Sur un **DR4**, le contact entre les deux partenaires de l'hétérodimère (RXR et TR) s'établit, d'une part avec la boîte D de RXR, comme dans l'exemple précédent, et cette fois, non plus avec le doigt CI, comme pour RAR, mais avec les 7 acides aminés qui précèdent CI. (Zechel *et al.*, 1994b). La structure tridimensionnelle du complexe RXR/TR lié à l'ADN *via* un élément DR4, obtenue par cristallographie (Rastinejad *et al.*, 1995), a apporté d'autres éléments d'information. Il s'avère qu'il existe effectivement un résidu dans la partie C terminale du domaine A/B qui participe à l'interface de dimérisation, inclus dans la région des 7 acides aminés en amont de CI de RXR, dans l'étude de Zechel (Zechel *et al.*, 1994b). En outre, d'autres résidus sont impliqués dans la surface de contact entre les deux récepteurs : deux acides aminés au sommet du doigt CI de TR, ainsi qu'un résidu Asn dans sa boîte T (Kurokawa *et al.*, 1993; Zechel *et al.*, 1994a).

Par ailleurs, l'homodimère RXR se fixe sur un élément de réponse de type **DR1**. Zechel *et al.* ont montré, à l'aide de mutants de délétions et d'expériences de retard sur gel, que les contacts entre les deux monomères s'établissaient par des liaisons entre les résidus de l'hélice 3 du partenaire en 3' sur l'ADN, et des résidus du second doigt de zinc C2 du récepteur RXR placé en 5' (Zechel *et al.*, 1994a).

Le choix préférentiel de l'espacement du DR, nous l'avons vu précédemment, dépend du type de récepteur impliqué dans l'hétérodimère avec son partenaire RXR. Ainsi, l'hétérodimère RXR/RAR présente une affinité plus forte pour DR5 que pour d'autres DR. Néanmoins ce dimère est capable, avec une affinité moindre, de se fixer sur d'autres DR. Dans ce cas, l'interface de dimérisation entre les deux récepteurs varie. Par exemple, le complexe RXR/RAR fixé sur un **DR2**, fait intervenir dans la

surface de dimérisation, d'une part la boîte T de RAR (situé toujours en 5') et d'autre part, le doigt CII de RXR, à l'exclusion de sa boîte D (Zechel *et al.*, 1994a).

### 3 La fixation du ligand

#### 3-1 Une affinité variable du récepteur pour son ligand

Les récepteurs nucléaires présentent une affinité variable pour leur ligand (quand celui-ci existe). Ainsi le Kd (constante de dissociation) du récepteur aux oestrogènes ER pour les oestrogènes est de 0,1nM, le Kd de RXR (l'un des plus faible de la famille) pour l'acide 9 cis rétinolique est de 14 à 18 nM (pour revue cf. (Keller *et al.*, 1993).

#### 3-2 Structure tridimensionnelle de l'interaction ligand-récepteur

La structure tridimensionnelle du domaine de fixation du ligand (LBD) a été déterminée par des études cristallographiques pour l'apo-récepteur hRXR $\alpha$  (c'est à dire le récepteur dépourvu de ligand) et l'holo-récepteur hRAR $\gamma$  (c'est à dire le récepteur fixé à son ligand).

##### 3-3-1) Structure du domaine de fixation du ligand de l'apo-RXR $\alpha$

Le domaine de fixation du ligand de l'apo-RXR $\alpha$  est constitué de 12 hélices  $\alpha$  (H1 à H12) qui forment trois types de couches empilées en "sandwich", de manière parallèle et antiparallèle (Bourguet *et al.*, 1995). Ainsi, les hélices H4, H5, H6, H8 et H9 sont coincées entre les hélices H1, H2 et H3 d'une part, et les hélices H7, H10 et H11 d'autre part (cf. figure 19). L'hélice H12 ferme l'entrée de la cavité pour le ligand.

Par ailleurs, les LBD des récepteurs RXR $\alpha$  forment des dimères observables en cristallographie (mais pas pour RAR $\gamma$ ), dont l'interface fait intervenir les hélices H10, H9 et la boucle entre H7 et H8.

##### 3-3-2) Structure du domaine de fixation du ligand de l'holo-RAR $\gamma$

hRAR $\gamma$  est constitué de 11 hélices contre 12 pour l'apo-RXR $\alpha$ . En effet, l'hélice H2 est absente dans cette structure (Renaud *et al.*, 1995). D'après le modèle proposé par ces auteurs, le ligand serait attiré dans la cavité du LBD par des forces électrostatiques. En effet, des lignes de champs électrostatiques de potentiel négatif à positif recouvrent la surface de la cavité et forment un faisceau de charges entre les résidus R278 et K236. Le ligand est ensuite maintenu dans la cavité principalement par des interactions hydrophobes, réalisées avec les résidus situés sur de très

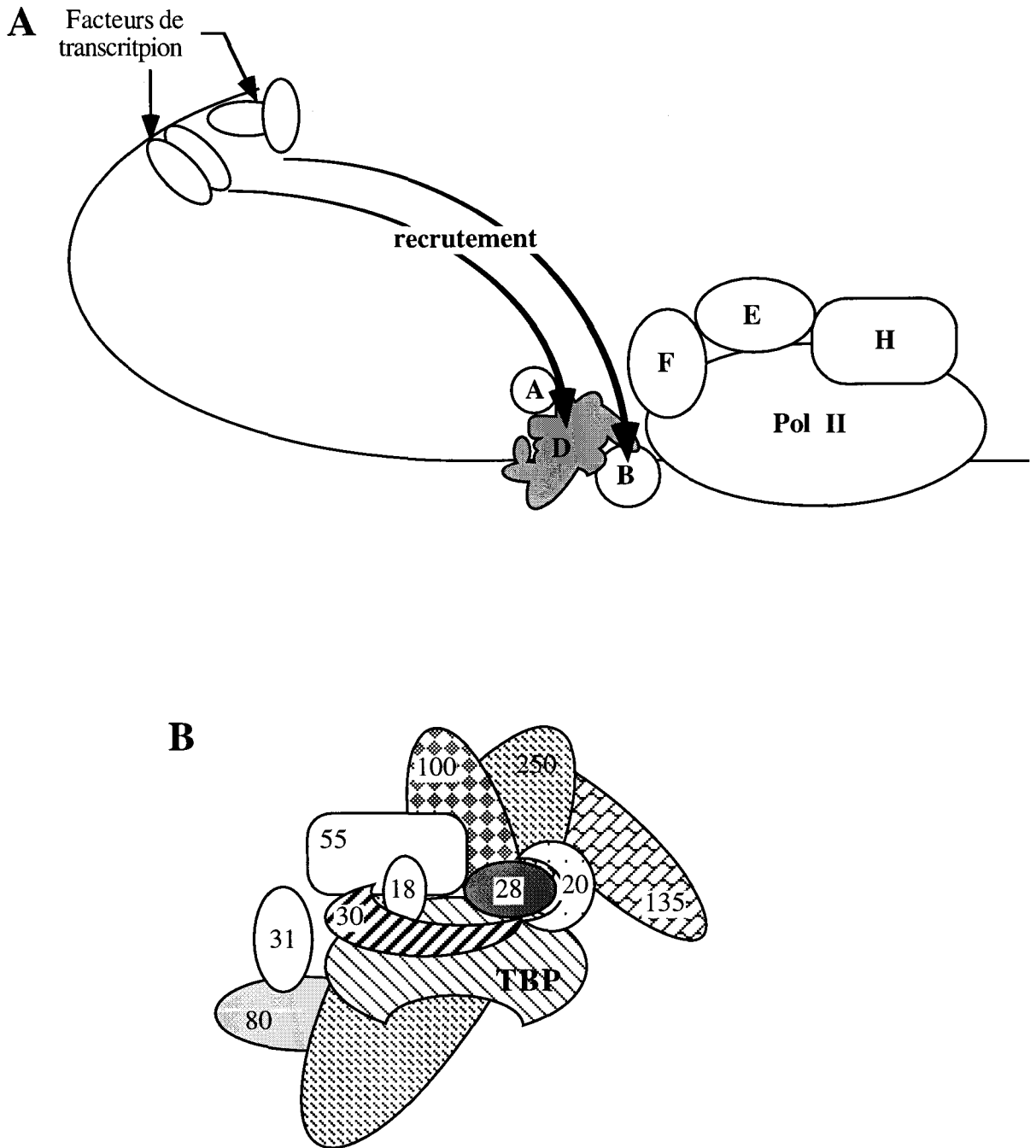
nombreuses hélices dans la structure. Des liens hydrogènes et des forces de Van der Waals stabilisent le ligand dans son compartiment. En outre, une fois le ligand enclavé dans sa cavité, l'hélice H11 est repositionnée dans la continuité de l'hélice H10 ce qui induit probablement un mouvement de H12 et une rotation d'une boucle appelée "boucle oméga" qui se positionne juste au dessous de H6, alors qu'elle était dans le voisinage de H2 et H3 en l'absence de ligand. Cette boucle entraîne une partie de H3 qui se recourbe comme un arc. H12 a subit un mouvement de rotation qui la colle contre le ligand et forme ainsi une sorte de "couvercle" sur la poche d'entrée du ligand. Ce mécanisme de liaison peut permettre d'expliquer la forte affinité des ligands pour leurs récepteurs nucléaires bien que l'on sache que le ligand puisse sortir librement de la poche.

### 3-3 Le coeur AF-2AD

Les LBD des récepteurs, après avoir fixé son ligand, se présente sous une forme plus compacte et moins accessible aux protéases. Les médiateurs transcriptionnels tels que TIF1 et SUG1, fixent les holo- mais pas les apo- LBD de RAR et/ou RXR (Lee *et al.*, 1995b; Le Douarin *et al.*, 1996b; vom Baur *et al.*, 1996). Ces interactions requièrent la présence du coeur AF-2AD, situé sur H12 (cf. II-3-2)

L'équipe de Pierre Chambon, après avoir résolu la structure en trois dimension des domaines de fixation du ligand de l'apo RXR et de l'holo RAR et avoir vérifié par un alignement de séquence la conservation des acides aminés cruciaux pour le maintien de la structure, propose un modèle conformationnel général très probablement valable pour toute la famille (Wurtz *et al.*, 1996). De plus, la résolution de la structure tridimensionnelle du domaine de fixation du ligand du récepteur TRb, en présence d'un analogue de la T3 (holo TR), réalisée par les équipes de John Baxter et Robert Fletterick, permet également de confirmer la modèle d'activation de ces récepteurs (Wagner *et al.*, 1995).





**Figure 20: A** L'initiation de la transcription par le recrutement de l'ARN polymérase II  
 Le recrutement de la machinerie de transcription peut se faire, entre autre, par l'intermédiaire de TFIIID (TBP et TAFIs), ou deTFIIB.

**B** Le complexe TFIIID est constitué de TBP et de différents TAFs, ici identifiés par leur masse moléculaire.

## II La modulation de la transcription

### 1 Principes généraux sur l'initiation de la transcription

#### 1-1 La machinerie transcriptionnelle de base

L'initiation de la transcription par la RNA polymérase II requiert la coopération de deux sortes de protéines: celles qui appartiennent à la machinerie transcriptionnelle de base et les facteurs de transcription qui reconnaissent une séquence spécifique sur l'ADN (dont font partie les récepteurs nucléaires, les protéines Pax, Hox, etc.).

La machinerie transcriptionnelle de base est constituée de l'ARN polymérase II et d'au minimum 12 facteurs associés, parmi lesquels se distinguent des unités mieux connues telles que: TFIIA, TFIIB, TFIIC, TFIID, TFIIE et TFIIF (cf. figure 20). TFIIA reconnaît l'ADN de façon non spécifique (cf. (Brunel *et al.*, 1991) pour revue). TFIID est composé de la protéine TBP ("TATA-box Binding Protein") et de diverses sous unités associées nommées TAF ("TBP associated factors") (Tansey et Herr, 1997). TBP reconnaît une séquence riche en A et en T, présente sur de nombreux promoteurs eucaryotes et procaryotes et désignée par son consensus "TATA". TBP, fixée sur cette boîte TATA, est située à proximité du site d'initiation (en moyenne à 10 pb en 5'), elle interagit avec des TAFs, qui à leur tour, vont se complexer aux différents facteurs TFIIB à TFIIF pour constituer la machinerie transcriptionnelle de base (Roeder, 1996) (cf. figure 20)

#### 1-2 L'accessibilité de l'ADN

L'ADN s'enroule sur un complexe d'histones. Le degré d'acétylation de la chromatine est corrélé à son activité transcriptionnelle: il est apparu que les régions les plus acétylées sont transcriptionnellement plus actives (ou activables) et réciproquement (Pazin et Kadonaga, 1997a). L'acétylation de la queue amino-terminale des histones rendrait l'ADN plus accessible à la fixation de différents facteurs et plus "perméable" à la progression de l'ARN polymérase, probablement par un "relâchement" des interactions ADN/histones et une mobilité de l'octamère d'histones accrue).

Nous verrons dans le chapitre suivant comment les récepteurs interagissent avec les éléments de la machinerie transcriptionnelle de base ou modulent le degré d'acétylation des histones pour activer la transcription.

## 2 Rôle des récepteurs dans l'activation de la transcription

### 2-1 Les domaines de transactivation sur le récepteur

L'utilisation de récepteurs tronqués et de constructions chimériques ont révélé l'existence de deux fonctions d'activation transcriptionnelle (=AF pour "Activation Function"). Celles-ci agissent de façon synergique et présentent des propriétés distinctes. La première, AF-1, possède une fonction constitutive, indépendante du ligand, localisée dans la région A/B de la protéine, en amino-terminal. La seconde, AF-2, dépendante du ligand, est située dans le LBD.

#### 2-1-1 Localisation des domaines AF-1 et AF-2

##### a) AF-1, situé dans le domaine A/B

Ce domaine a tout d'abord été mis en évidence, dans les récepteurs aux hormones stéroïdiennes (GR, PR) (pour revue cf. Beato *et al.*, 1995), puis, identifié très rapidement dans le récepteur aux hormones thyroïdiennes (TR) (pour revue (Thompson et Evans, 1989)). L'activité transactivatrice de GR chute ainsi de 95% après la délétion d'une région de son domaine A/B, dénommée AF-1. Parallèlement, la duplication de ce domaine dans le récepteur entier, triple sa capacité transactivatrice (Hollenberg et Evans, 1988). Le cas de RAR $\beta$ 2 illustre bien l'indépendance de AF-1 par rapport au ligand; l'activité transactivatrice du domaine AF-1 est en effet constante, en présence ou en l'absence d'acide rétinoïque (Folkers *et al.*, 1993).

##### b) AF-2, situé dans le domaine E

Des constructions, caractérisées soit par la délétion du domaine E, ou composées de la fusion du domaine E avec un domaine de fixation à l'ADN, comme Gal 4, co-transfectées dans des cellules en culture avec un gène rapporteur doté d'un élément de reconnaissance à l'ADN, ont permis de révéler la capacité transactivatrice de cette région, en présence du ligand approprié (pour exemples, cf. (Danielsen *et al.*, 1987) pour GR et (Webster *et al.*, 1988) pour GR et ER et aussi (Tora *et al.*, 1989), (Lees *et al.*, 1989), et (Nagpal *et al.*, 1993) pour RARs et RXRs). Des exemples *in vivo* illustrent aussi l'importance du domaine C-terminal dans la transactivation. La fonction AF-2 ligand-dépendante, a ainsi pu être soulignée par l'analyse de l'oncogène v-erbA, qui code un récepteur TR délété de 9 acides aminés dans son domaine AF-2. v-erbA se comporte en répresseur constitutif *via* des séquences cibles normalement transactivées par TR en présence de son ligand. (Zenke *et al.*, 1990; Danielian *et al.*, 1992; Baretino *et al.*, 1994; Durand *et al.*, 1994).

Les 9 acides aminés absents forment le "coeur AF-2AD" de TR, retrouvé dans l'hélice 12 de la structure 3D. La présence de ce coeur AF-2AD servirait donc de commutateur pour permettre au récepteur de passer de l'état de répresseur à l'état d'activateur transcriptionnel.

### 2-1-2) Le coeur AF-2AD , une région d'interaction avec les acteurs de la transcription

De nombreux domaines de transactivation comportent des séquences peptidiques riches en acides aminés acides (Lin *et al.*, 1991). D'autres domaines ont aussi été caractérisés incluant des séquences riches en glutamine (comme dans le facteur Sp1) ou riches en proline (comme dans le facteur CTF, (Mitchell et Tjian, 1989), pour revue). Les domaines d'activation de nombreux récepteurs ne cadrent avec aucun de ces domaines. AF-1 et AF-2 ne comportent en effet pas d'acides aminés particuliers (Tora *et al.*, 1989; Bradshaw *et al.*, 1991; Shemshedini *et al.*, 1992). Cependant, le domaine AF-2 contient en outre une courte région AF-2AD (AD ="Autonomous Domain"), dont l'activité est indépendante du ligand et qui isolée, reste efficace bien que plus faiblement que dans la molécule entière (cf. par exemple, (Barettino *et al.*, 1994) pour TR, (Durand *et al.*, 1994) pour RARs et RXRs). AF-2AD est, en l'occurrence, caractérisée par un motif très conservé parmi les récepteurs, de consensus ffXEff où f correspond à un résidu hydrophobe (Danielian *et al.*, 1992). La mutation de l'acide glutamique, E401 pour TR $\alpha$ 1 et E412 pour RXR $\alpha$ , engendre un récepteur dominant négatif, incapable de fixer le ligand et de transactiver (Barettino *et al.*, 1994)(Durand *et al.*, 1994). L'étude cristallographique du LBD de TR $\alpha$ 1 et RAR $\gamma$  a suggéré que ce résidu participerait d'une part, à la stabilité de l'hélice H12, une fois le ligand fixé, et d'autre part, à l'interaction avec des co-facteurs (Renaud *et al.*, 1995)(Wagner *et al.*, 1995). De plus, les résidus hydrophobes F399 et F403 qui encadrent l'acide glutamique dans le coeur AF-2AD du récepteur TR $\alpha$ 1, interagissent directement avec le ligand et participent ainsi au maintien de la conformation de l'holo récepteur (Barettino *et al.*, 1994) (Renaud *et al.*, 1995).

### 2-1-3) La régulation de l'activité transactivatrice : ligand et phosphorylation

#### *a) régulation de l'activité de AF-2*

La fixation du ligand sur le récepteur entraîne, en plus d'un changement de conformation (cf. paragraphe structure 3D), d'autres événements régulateurs, comme celui relié à la tyrosine Y541 et à sa phosphorylation putative en l'absence ou en présence de ligand. La mutation de Y541 par un résidu non phosphorylable,

induit la formation d'un récepteur capable de transactiver de manière constitutive, en l'absence ou en présence du 17 $\beta$  oestradiol. De plus, ce résidu a été montré jouer un rôle fondamental dans l'interaction de ER avec les co-activateurs RIP-140 et SRC-1 (White *et al.*, 1997). En parallèle, dans ce contexte, l'interaction entre ER et les co-activateurs RIP-140 et SRC-1, est observée en présence ou en absence du 17 $\beta$  oestradiol. Il serait donc possible que le ligand entraîne la déphosphorylation de Y541, ce qui entraînerait la fixation des cofacteurs transcriptionnels.

#### *b) régulation de l'activité de AF-1*

AF-1, domaine toujours présenté comme indépendant du ligand peut néanmoins être régulé par la présence de ce dernier. Son efficacité est en effet augmentée en présence du ligand. Cette variation d'activité est corrélée avec l'état de phosphorylation du domaine AF-1. Par exemple, la présence du ligand induit la phosphorylation de la sérine 118 du domaine AF-1 de ER. Parallèlement, l'inhibition de cette phosphorylation abolit l'activité transactivatrice de AF-1 (Ali *et al.*, 1993; Kato *et al.*, 1995). Cette phosphorylation peut aussi être réalisée *via* le facteur de croissance EGF, en l'absence de ligand (Bunone *et al.*, 1996). Par ailleurs, le domaine A/B de TR $\alpha$  présente une affinité accrue pour TFIIB en présence de T3 (Hadzic *et al.*, 1995). Par conséquent, la présence du ligand ou de facteurs de croissance peuvent induire directement ou indirectement la phosphorylation du domaine AF-1. La phosphorylation de ce domaine le rend plus efficace car il est dans un état tel qu'il interagit mieux avec les acteurs de la transcription (éléments de la machinerie transcriptionnelle de base ou cofacteurs).

Par spectroscopie RMN, le domaine AF-1 du récepteur aux glucocorticoïdes, GR, apparaît sous forme de trois hélices  $\alpha$  (Dahlman-Wright *et al.*, 1995). La déformation de cette structure diminuerait fortement la capacité d'AF-1 à transactiver. Or, l'arrivée du ligand, nous l'avons vu précédemment, entraîne un changement conformationnel (cf. II-2-2 b) du LBD. Ce mouvement pourrait se propager sur l'ensemble de la molécule et induire un agencement des hélices  $\alpha$  du domaine A/B, plus propice à l'interaction avec des cofacteurs. Néanmoins, le mode d'action d'AF-1 est à l'heure actuelle beaucoup moins bien connu que celui d'AF-2.

## **2-2 Recrutement de la machinerie transcriptionnelle de base**

Les éléments de réponse aux hormones peuvent être situés à une distance relativement grande des sites d'initiation de la transcription. Par conséquent, les

récepteurs ont du développer des mécanismes leur permettant de contacter physiquement directement ou indirectement les composés du complexe de pré-initiation.

#### 2-2-1) Recrutement direct

Le recrutement direct de la machinerie transcriptionnelle de base passe soit par TFIIB, soit par le complexe TFIID.

##### a) *Recrutement de TFIIB*

Les récepteurs ER, PR et COUP-TFI radiomarqués interagissent avec TFIIB, fusionnée à la GST, en test de co-rétention protéique ou "pull-down" (Ing *et al.*, 1992). Le domaine E de ER participe à cette liaison. Pour PR et COUP-TFI, les protéines entières furent utilisées; il est par conséquent impossible de déterminer lequel des deux domaines AF-1 ou AF-2 est incriminé. Le récepteur à la vitamine D se complexe lui aussi à TFIIB (Blanco *et al.*, 1995).

##### b) *Recrutement de TFIID*

**TBP** : L'hélice 12 de RXR, utilisée en tests "pull-down" et dans le système double hybride chez la levure, interagit directement avec TBP. Parallèlement, des mutations qui affectent la formation du complexe AF2/TBP entravent la capacité de RXR à activer la transcription (Schulman *et al.*, 1995). Par conséquent, TBP participe à la fonction transactivatrice de RXR.

**TAFII30** : (TAF pour TBP-associated factors) TAFII30 s'associe au récepteur des oestrogènes, ER (Jacq *et al.*, 1994). Cette interaction, démontrée par des expériences de "pull-down", implique le domaine AF2 et n'est pas modulée par la présence du ligand (Tanese et Tjian, 1993). De plus, des anticorps dirigés contre TAFII30 inhibent l'activité transcriptionnelle d'une protéine chimère dotée du domaine E de ER en test de transactivation *in vitro*, ce qui confirme le rôle de TAFII30 dans le relais de l'activité transcriptionnelle des récepteurs aux oestrogènes (Jacq *et al.*, 1994).

#### 2-2-2) Recrutement indirect, association de co-facteurs

Les cofacteurs réalisent un "pont" entre le domaine transactivateur et les facteurs de la machinerie transcriptionnelle de base. La plupart des criblages réalisés pour aller chercher ces cofacteurs impliqués dans l'activité transcriptionnelle des récepteurs, ont utilisé le domaine E comme appât. Dans la majorité des cas, le motif AF-2AD a été suffisant pour les criblages effectués par les techniques de pull-down ou par le système double hybride en levures.

<b>Facteur</b>	<b>Interaction avec les récepteurs</b>	<b>Effet du ligand</b>
<b>co-activateurs</b>		
hbrm	GR	Association
ERAP140	hERAF2	Association
GRIP1	GR, AR, ER	Association
p160	CBP et ER	Association
p300/CBP	T3R, RAR, RXR	Association
RAP46	AR, ER, PR, GR	dépend du récepteur
RIP 110	TR, RAR, RXR	dépend du RIP et du récepteur
RIP140	ER	Association
hSNF2a	ER, RAR	?
SRC1	ER, GR, PR, RXR, TR	Association
TRIP	RXR, TR	Différent selon les TRIP
TIF1	ER, TR, VDR, RAR, RXR	Association
TIF2	AR, ER, GR, PR, RAR, RXR, TR	Association
<b>Co-répresseurs</b>		
N-coR	RAR, TR	Dissociation
SMRT	TR, RAR, RXR	Dissociation

**Figure 21** : Les différents co-régulateurs des récepteurs nucléaires (d'après Gelman et al, 1997). Tableau récapitulatif des co-régulateurs connus, qui interagissent avec les récepteurs nucléaires. L'association des récepteurs avec leurs ligands est nécessaire pour l'interaction avec des molécules co-activatrices. Inversement, l'absence du ligand est indispensable pour l'interaction des récepteurs avec les co-répresseurs.

a) Quelques exemples de co-facteurs

(cf. figure 21)

- TIF-1

L'un des premiers isolé par la technique de double hybride, en utilisant les domaines E et D de RXR, en présence de son ligand. *In vitro*, TIF-1 interagit avec RXR $\alpha$ , RAR $\alpha$  et ER (Le Douarin *et al.*, 1995).

- SRC-1

La molécule SRC-1 (Steroid Receptor Co-activator) a été identifiée, dans le système double-hybride, sur la base de son interaction avec le LBD du récepteur humain de la progestérone (PR), en présence de l'hormone (Onate *et al.*, 1995). Les protéines p160, homologues de SRC-1, s'associent à RAR ainsi qu'aux protéines CBP (CREB-binding protein) et p300 (Kamei *et al.*, 1996).

- CBP et p300

CBP (= CREB binding protein)

CBP interagit avec une grande variété de protéines capables de fixer l'ADN, comme CREB ou AP-1, ou encore les protéines du complexe d'initiation de la transcription comme TBP ou TFIIB (cf. (Janknecht et Hunter, 1996), pour revue).

- Trip 1/SUG1

Trip-1 (TR interacting protein) est une des 15 protéines isolées au cours d'un criblage dans le système double-hybride en utilisant comme "appât", le domaine E de TR $\beta$ 1 en présence de son ligand ou celui de RAR $\alpha$  (vom Baur *et al.*, 1996). Trip-1 est l'homologue du gène de levure SUG1 (Swaffield *et al.*, 1992). La région carboxy-terminale est constituée d'un grand domaine doté d'une activité ATPase. De plus, SUG1 possède aussi une activité hélicase (Fraser *et al.*, 1997) et elle interagit avec TBP et TAFII30 (Swaffield *et al.*, 1995; vom Baur *et al.*, 1996)

- RIP140

RIP140 est une protéine isolée elle aussi par le système double-hybride, sur la base de son interaction avec les domaines DEF de RXR $\alpha$  humain (Seol *et al.*, 1995). En cellule, la transfection transitoire de vecteur d'expression pour RIP140 augmente la transcription ligand-dépendante médiée par ER (Cavailles *et al.*, 1995).

- SPT16

Des expériences de précipitation par la GST ont montré que SPT6 (isolée chez la levure) interagissait *in vitro* avec le domaine E de ER. De plus, dans une souche de levure déficiente pour le gène SPT6, la transactivation ligand-dépendante médiée



par ER est diminuée par rapport à celle observée dans une souche sauvage (Baniahmad *et al.*, 1995).

#### *b) Motif d'interaction, co-facteur-récepteur*

Le cofacteur possède un motif de reconnaissance pour le récepteur, appelé "boîte NR" et constitué en particulier de 4 résidus Leucine cruciaux de consensus: LXXLL. Par exemple, cette boîte NR est retrouvée dans la région de RIP140 impliquée dans l'interaction avec ER (Le Douarin *et al.*, 1996a; Heery *et al.*, 1997).

En résumé, les facteurs qui médient l'activité transcriptionnelle des récepteurs nucléaires, qu'ils appartiennent ou non à la machinerie transcriptionnelle de base, s'associent aux récepteurs *via* le coeur AF-2AD de ces derniers, en présence du ligand. Celui-ci en effet, une fois ancré dans la poche hydrophobe du récepteur, entraîne un changement conformationnel du LBD, rendant l'hélice constituée par le coeur AF-2AD accessible aux cofacteurs. Ces co-facteurs stimuleraient la transcription, soit en favorisant le recrutement des éléments de la machinerie transcriptionnelle, soit en rendant la structure de l'ADN plus accessible par leur activité ADN hélicase ou d'acétylation intrinsèques, ou portées par d'autres protéines auxquelles ils sont capables de s'associer.

### **2-3 Association avec des cofacteurs modulant le degré d'acétylation des histones**

Les cofacteurs aident au recrutement du complexe de transcription, mais beaucoup d'entre eux sont beaucoup plus que des "protéines de liaison". Ils possèdent souvent une activité "histone acétylase" intrinsèque. CBP par exemple, possède une activité "histone acétylase" (Ogryzko *et al.*, 1996), qui lui permettrait de préparer le terrain pour le facteur de transcription auquel elle est associée (cf. (Wolffe et Pruss, 1996; Pazin et Kadonaga, 1997b) pour revue). De même SPT6, fait partie d'une classe de protéines, identifiées génétiquement chez le levure, qui contrôlent la transcription en affectant la structure chromatinienne (Winston et Carlson, 1992).

### **3 Rôle des récepteurs dans la répression de la transcription**

De nombreux récepteurs, notamment les récepteurs orphelins, se fixent sur leur séquence de reconnaissance à l'ADN et répriment la transcription de leurs gènes cibles. Par commodité, on qualifie ce type de répression de répression directe. Par opposition, on parle de répression indirecte, pour les récepteurs capables de

réprimer la transcription de gènes cibles, sans se fixer sur leurs séquences de régulation. Ces récepteurs agissent par capture des co-activateurs indispensables à l'activation d'autres récepteurs ou d'autres facteurs de transcription, transactivateurs du gène cible considéré.

### 3-1 Les interférences transcriptionnelles

Les interférences entre co-activateurs communs peuvent dépasser le réseau des récepteurs nucléaires. Par exemple, GR, en présence de son ligand, peut entraver la capacité transactivatrice de l'hétérodimère AP-1 (composé des produits des proto-oncogènes c-fos et c-jun) (Jonat *et al.*, 1990; Schule *et al.*, 1990). Réciproquement, AP-1 réprime la fonction de transactivation de GR. Ces interférences s'élargissent aux récepteurs RAR et RXR, ER, TR, mais pas au MR et passent notamment par la capture de CBP (Kamei *et al.*, 1996). Le plus bel exemple de l'importance de CBP et p300 dans la régulation de nombreux gènes du développement est représentée par les effets de leur inactivation dans l'embryon de souris. Les animaux pour lesquels l'inactivation du gène codant CBP est homozygote meurent entre 9 à 11,5 jours de gestation. De plus, la mortalité est élevée aussi pour les animaux dont l'inactivation de CBP ou/et p300 est hétérozygote (Yao *et al.*, 1998). Ces résultats suggèrent que le développement de la souris est sensible à la concentration de p300 et CBP.

### 3-2 La répression directe

#### 3-2-1) Domaine transrépresseur

Certains récepteurs activateurs répriment la transcription en l'absence de ligand. C'est le cas par exemple de TR, de RAR ou de RXR. Il a été montré que cette répression s'effectuait *via* la fixation au récepteur libre d'un corépresseur, dont les deux principaux sont NCoR (pour Nuclear Receptor Co-Repressor) et SMRT (pour Silencing Mediator for Retinoid and Thyroid hormone receptors). Ceux-ci interagissent avec la partie amino terminale du domaine E ou la région C-terminale du domaine D, et sont libérés lors de la fixation du ligand, ce qui abolit la répression (Chen et Evans, 1995; Hörlein *et al.*, 1995; Kurokawa *et al.*, 1995; Perlmann et Jansson, 1995). Des mutations dans TR $\beta$ 1, ont permis de cartographier le site d'interaction avec N-CoR, appelé "CoRBox", situé entre les acides aminés 210 et 230 de TR $\beta$ 1, inclus dans l'hélice 1 du LBD (Kurokawa *et al.*, 1995; Wagner *et al.*, 1995).

Il a été montré récemment que le récepteur orphelin COUP-TF, lequel réprime l'activité transcriptionnelle de nombreux gènes (cf. § 2-3 du chapitre II) était lui-même capable d'interagir avec les co-répresseurs N-CoR et SMRT, interaction mise en évidence par la technique de double hybride en levure et par "pull-down" *in vitro* (Shibata *et al.*, 1997). Contrairement à TR et RAR, ce n'est pas l'hélice 1 qui, pour COUP-TF, s'associe aux co-répresseurs, mais sa région C-terminale: 35 acides aminés, donc le coeur AF-2AD. Or, ce motif chez COUP-TF, diverge du motif AF-2AD des autres récepteurs. Achatz *et al.* (Achatz *et al.*, 1997), ont étudié les effets d'un vecteur contenant le domaine E de COUP-TFI, fusionné au domaine de liaison à l'ADN de Gal4, sur la répression d'un gène rapporteur, doté de sites de fixation pour GAL4. La co-transfection d'un vecteur codant COUP-TFI en cellule, inhibe cette répression. Par contre, la co-transfection d'un vecteur d'expression pour COUP-TFI, délété de ces 35 acides aminés C-terminaux, n'interfère plus avec la répression médiée par le domaine E de COUP-TFI. En conséquence, ces 35 acides aminés C-terminaux de COUP-TFI semblent participer à la fixation de molécules spécifiques comme des co-répresseurs (cf. ci-dessus et (Shibata *et al.*, 1997)). En parallèle, Les auteurs qualifient ce domaine de "domaine de répression active", par opposition à d'autres régions de la protéine, impliquées dans la "transrépression" car capturant co-activateurs.

### 3-2-2) Les co-répresseurs associés aux récepteurs : N-CoR et SMRT

#### a) N-CoR

N-CoR (Nuclear Receptor Co-Repressor) a été isolé sur la base de son interaction avec le complexe TR/RXR/ADN en absence de ligand (Hörlein *et al.*, 1995). N-CoR interagit aussi avec RAR, mais pas VDR, RXR, ER et GR.

La fixation de N-CoR sur les récepteurs, serait notamment responsable de la répression transcriptionnelle de l'hétérodimère RAR/RXR fixé sur un DR2 ou sur un DR5. N-CoR interagit en effet avec les deux types d'hétérodimères formés entre RAR et RXR, quelque soit leur polarité: RAR/RXR fixé sur un DR2 ou RXR/RAR fixé sur un DR5. En l'absence de N-CoR, les deux types d'hétérodimères sont capables de fixer l'acide rétinoïque et les cofacteurs RIP140 et RIP160. En revanche, si N-CoR est complexé aux hétérodimères RAR/RXR ou RXR/RAR, l'arrivée du ligand dans le milieu dissociera le co-répresseur de RXR/RAR mais pas de RAR/RXR. Il en résulte que ce dernier, fixé sur un DR2 reste transcriptionnellement inactif (Kurokawa *et al.*, 1994; Hörlein *et al.*, 1995).

#### b) SMRT

(Silencing Mediator for Retinoid- and -Thyroid-hormone receptors)

SMRT interagit avec RAR et TR, en l'absence de ligand, mais pas avec d'autres récepteurs comme GR ou RXR (Chen et Evans, 1995). SMRT reste fixé à RAR ou TR, même quand ceux-ci sont complexés à leur éléments de réponse, sous forme d'hétérodimère. Lors de la fixation du ligand, les complexes SMRT/RAR ou SMRT/TR se dissocient, ce qui constitue vraisemblablement une étape nécessaire pour que s'exerce l'activité transcriptionnelle.

#### c) Mode d'action des co-répresseurs

Le co-répresseur, complexé au récepteur, lui-même fixé à l'ADN, empêche celui-ci d'interagir avec les éléments de la machinerie transcriptionnelle de base ou les co-activateurs, inhibant ainsi la stimulation potentielle d'activation par les récepteurs. De plus, ces co-répresseurs peuvent agir sur la structure de la chromatine. Par exemple, ils ont un effet massif sur la structure de la chromatine en promouvant la dé-acétylation des histones. N-CoR interagit avec le facteur Sin3A, qui lui-même s'associe à l'histone deacétylase HDAC1 (Li *et al.*, 1997). Cette association pourrait rapprocher une déacétylase des gènes cibles sur lesquels sont fixés les récepteurs associés à N-CoR et maintenir l'ADN sous sa forme non acétylée, et ainsi prévenir ou diminuer, l'accessibilité de l'ARN polymérase II pour le site d'initiation.

### 4 Conclusion

Nous venons de voir que les récepteurs nucléaires orphelins sont capables de d'entrer en compétition pour des co-facteurs utilisés par des récepteurs "classiques", comme le cofacteur CBP, par exemple (cf. § 3-1 ci-dessus). De cette façon, la disponibilité de CBP chute et l'activation des récepteurs à ligand fixés sur leur HRE, est réduite. Par ailleurs, ces récepteurs orphelins sont capables de reconnaître et d'entrer en compétition avec les mêmes cibles sur l'ADN que les récepteurs classiques (cf. § 2-3 du I). Ainsi fixés, ils peuvent interagir avec des co-répresseurs qui vont par exemple déacétyler l'ADN, et inhiber l'activation de l'expression du gène.

En conclusion, nous pouvons retenir que le contrôle de l'expression d'un gène est dépendante de la concentration des facteurs du milieu qui participent à sa régulation positive d'une part, (concentrations des récepteurs, de leurs ligands et de leurs cofacteurs) mais aussi de la concentration des facteurs qui interfèrent avec cette régulation positive. Une répression ou une activation, même directe sans cet

Première période

Deuxième période

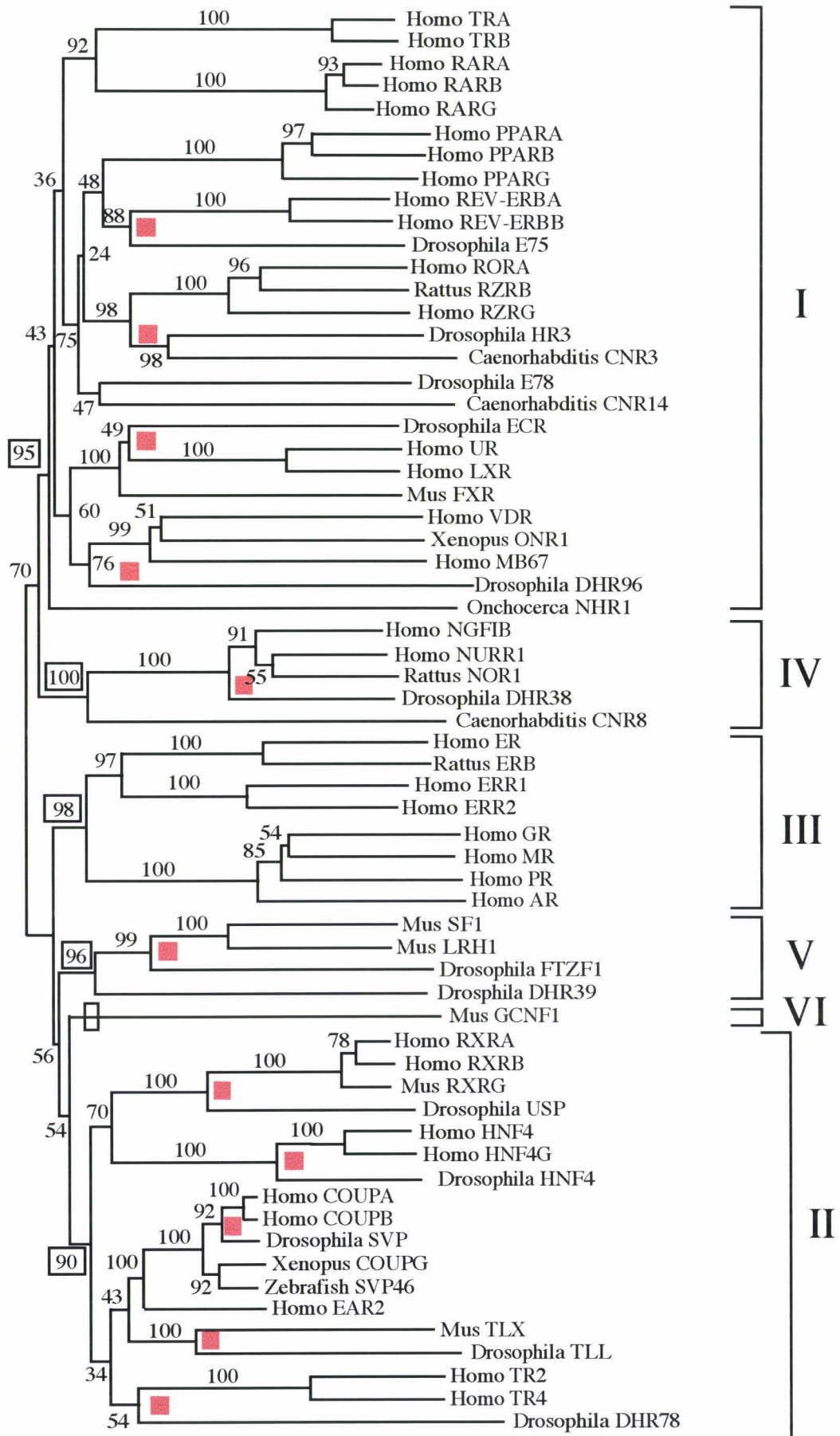


Figure 22 : Arbre phylogénique des récepteurs, montrant les deux grandes étapes de diversification, au cours de l'évolution.

■ indique la dichotomie arthropodes/vertébrés.

ensemble d'interférences entre plusieurs récepteurs, doit toujours être remise dans le contexte de la séquence régulatrice du gène considéré. La séquence régulatrice, promoteur ou enhancer, est en effet, un ensemble de sites de fixation pour de nombreux facteurs de transcription. Par conséquent, vont s'exercer *via* cette séquence, des synergies ou au contraire, des réductions d'activation entre plusieurs facteurs.

Dans ce contexte, les récepteurs orphelins jouent un rôle important en interférant avec les régulations médiées par les autres récepteurs.

## B Evolution des récepteurs nucléaires

### I La diversification des récepteurs au cours de l'évolution

L'étude de la phylogénie moléculaire des récepteurs nucléaires a permis de définir l'existence de 6 sous familles parmi les 60 membres des récepteurs de la superfamille (Laudet *et al.*, 1992; Escriva *et al.*, 1997a; Laudet, 1997) (cf. aussi figure 14). Il serait intéressant par ailleurs, de comprendre comment ces sous familles ont émergé et quels ont pu être leurs rôles au cours de l'évolution. Nous nous intéresserons notamment, aux différentes étapes de diversification qui ont engendrer les récepteurs actuels.

#### 1 Etude phylogénétique des récepteurs

L'arbre de la figure 22 permet de déduire deux grandes périodes dans la diversification des récepteurs nucléaires.

La première, correspond à l'émergence des récepteurs ayant donné naissance aux six grands groupes de récepteurs. Cette première vague de diversification est située avant la dichotomie arthropodes/vertébrés et pourrait corrélérer avec l'apparition des métazoaires. La seconde vague de duplication, ayant donné naissance à deux, voir trois gènes paralogues à partir d'un même gène. Par exemple, le gène ancestral TR va donner deux gènes TR $\alpha$  et TR $\beta$ , un RAR, un RAR $\alpha$ ,  $\beta$  et  $\gamma$ , un

RXR, les trois récepteurs RXR $\alpha$ , RXR $\beta$  et RXR $\gamma$ . Cette seconde étape de diversification est spécifique des vertébrés.

Ces deux grandes périodes de diversification des récepteurs correspondent-elles à des événements précis dans l'histoire de l'évolution des métazoaires ? L'hypothèse du zootype qui propose l'apparition de grands groupes de gènes avec l'émergence des métazoaires, pourrait rendre compte de la première période de diversification des récepteurs (Slack *et al.*, 1993). L'hypothèse d'une duplication du génome qui se serait produite spécifiquement dans la lignée des vertébrés et qui serait contemporaine de la mise en place de cet embranchement, permettrait d'expliquer la seconde étape de duplication génique observée dans le cas des récepteurs nucléaires.

## **2 L'hypothèse de la duplication du génome**

### **2-1 La duplication comme moyen d'évolution**

Ohno, en 1968, avait émis l'hypothèse d'une duplication du génome pour expliquer les innovations morphologico-anatomiques et développementales spécifiques des vertébrés, telles que les systèmes sensoriels et le cerveau particulièrement développés, ou encore la présence des arcs branchiaux, les crêtes neurales migratoires etc... (Ohno *et al.*, 1968). Ohno suppose que les gènes du développement sont intégrés dans des voies de régulation hiérarchiques et interdépendantes et qu'ils ne peuvent par conséquent, être mutés ou diverger vers d'autres fonctions sans bouleverser totalement toute l'organisation de l'animal. Il argumente en faveur de la polyploidie qui crée la redondance de copies de gènes, capables de diverger vers d'autres fonctions. Il propose également que certains de ces gènes n'étant soumis à aucune pression de sélection peuvent devenir des "pseudo-gènes" non fonctionnels ou être tout simplement perdus.

### **2-2 Les différents événements possibles de duplication**

Le génome des invertébrés est estimé contenir 15 000 à 20 000 gènes contre 60 000 à 80 000 pour le génome des vertébrés (Bird et Tweedie, 1995). La duplication des gènes est donc un événement fréquent dans l'histoire de l'évolution, cependant, elle diffère d'un point de vue qualitatif chez les vertébrés. Dans de nombreux cas, les duplications s'effectuent en effet "en tandem", ce qui signifie que la duplication d'un gène engendre deux gènes frères adjacents.

Chez les vertébrés, il n'y a que très peu de duplications en tandem. Les gènes dupliqués sont dispersés et non pas liés. Par exemple, les trois gènes paralogues  $RAR\alpha$ ,  $\beta$  et  $\gamma$  sont situés sur des chromosomes différents;  $RAR\alpha$  en 17q11;  $RAR\beta$  en 3p25;  $RAR\gamma$  en 12q21 ((Gronemeyer et Laudet, 1995) pour une revue). Cependant, la cartographie des gènes humains montre que nombreux d'entre eux sont regroupés en "régions de paralogie". Il s'agit de régions où les membres de plusieurs familles sont retrouvés dans les mêmes positions relatives les uns par rapport aux autres, sur des chromosomes distincts. Par exemple, les gènes  $RAR\alpha$ ,  $TR\alpha$  et  $Rev-erb\alpha$  sont colocalisés sur le chromosome 17, en position q11 et les gènes paralogues  $RAR\beta$ ,  $Tr\beta$ ,  $Rev-erb\beta$  sont colocalisés sur le chromosome 3, en position 3p25. L'explication la plus simple sur l'origine de ces régions de paralogies, est qu'elles dérivent d'un ancêtre commun, par la duplication d'une large région sur l'ADN génomique (Lundin, 1993).

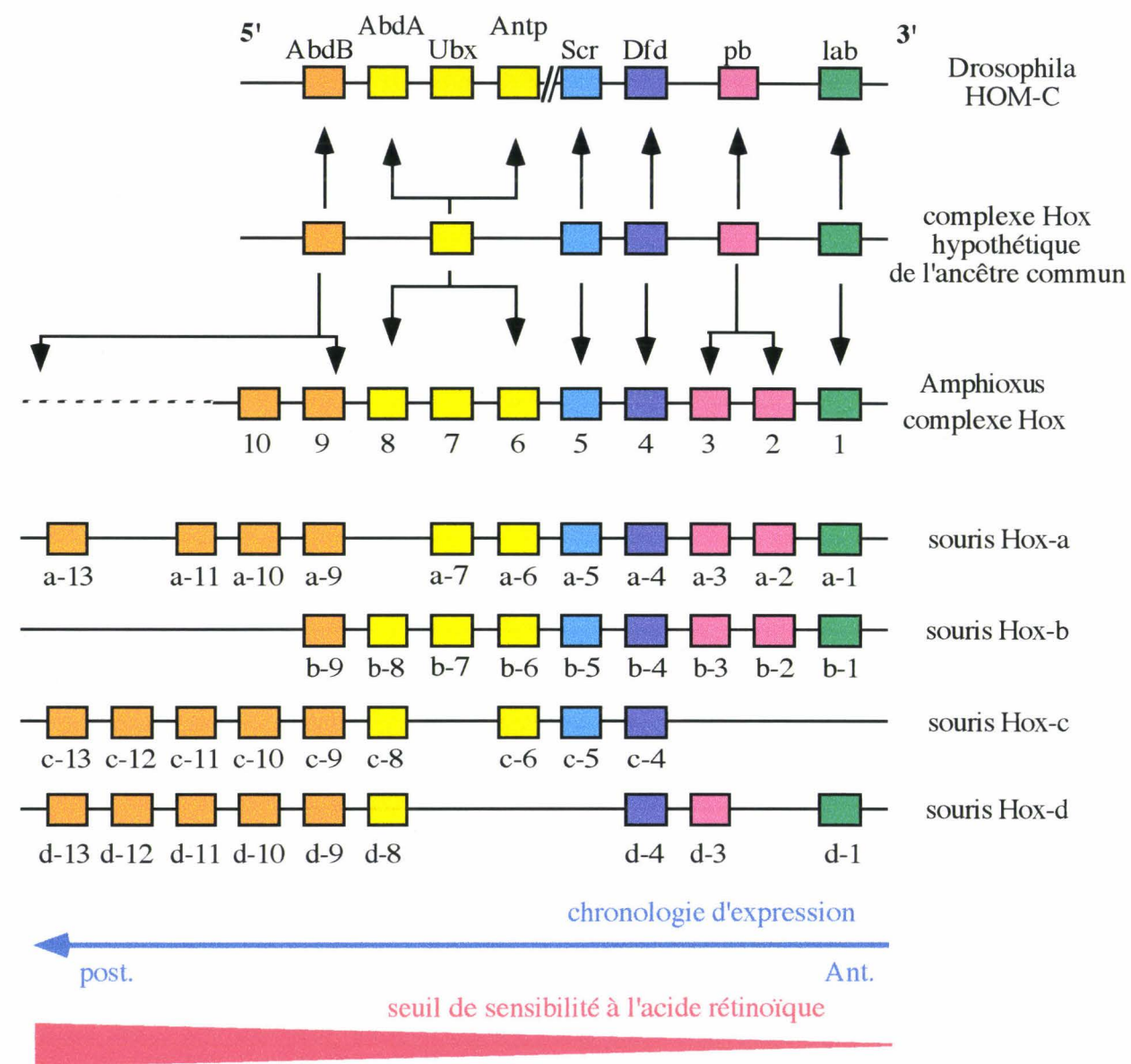
### **3 Un autre exemple de choix pour l'étude des duplications géniques au cours de l'évolution : les gènes Hox**

#### **3-1 Les gènes Hox au cours du développement embryonnaire**

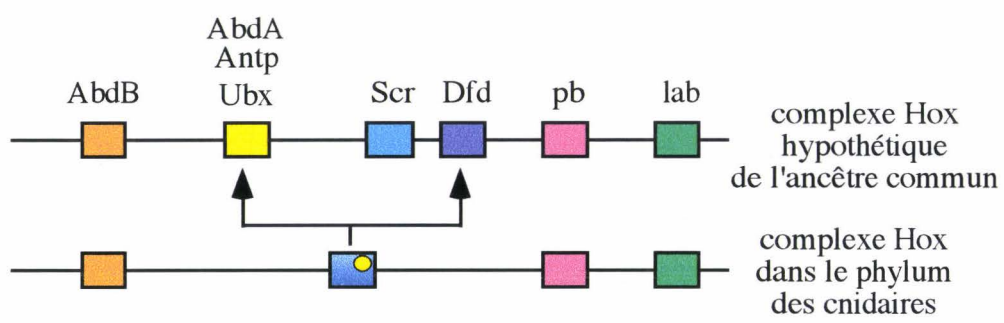
Les gènes Hox participent à l'organisation antéro-postérieure du plan du corps de l'embryon en cours de développement (cf. (Duboule et Morata, 1994) pour revue). Ils constituent par conséquent de formidables outils pour l'étude des caractères à l'origine des différences, notamment morphologiques, qui ont provoqué l'évolution des organismes.

Les gènes Hox forment une classe de gènes à boîte homéo. La boîte homéo représente le motif de fixation à l'ADN des protéines codées par les gènes Hox, qui sont des facteurs de transcription. L'une des principales caractéristiques de ces gènes est leur arrangement en regroupement sur les chromosomes, conséquence probable de la duplication en tandem d'un gène Hox primordial, très tôt dans l'évolution (Kappen *et al.*, 1989; Scott *et al.*, 1992). De plus, l'ordre des gènes sur le chromosome est corrélé avec l'ordre spatio-temporelle le long de l'axe embryonnaire antéro-postérieur. Les gènes situés en 5' d'un complexe s'expriment ainsi les premiers au cours du développement, et sont par ailleurs détectés en position antérieure par rapport aux gènes placés plus en 3', dans l'axe antéro-postérieur de l'animal.





(a)



(b)

**Figure 23 :** Comparaison de l'organisation chromosomique du complexe Hox chez la souris, la drosophile et l'amphioxus (a), et l'hydre (b). La règle d'expression spatio-temporelle est conservée chez l'hydre

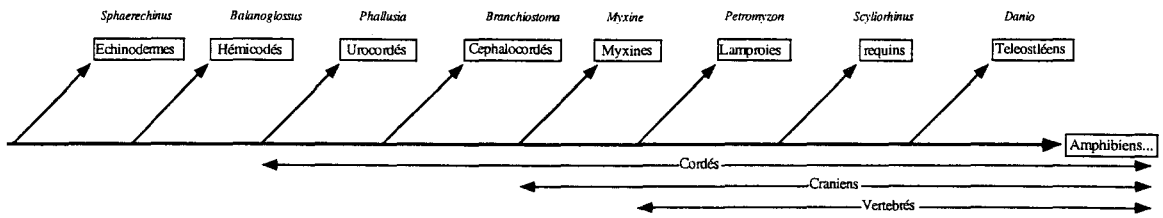
Par ailleurs, le nombre de complexes formés par les gènes Hox varie selon l'organisme étudié. Il en existe ainsi quatre chez la souris, un chez la drosophile. Les paragraphes qui suivent font le point sur les événements de duplication qui ont engendré l'agencement des gènes Hox sur les chromosomes des organismes actuels.

### 3-2 L'émergence du complexe Hox

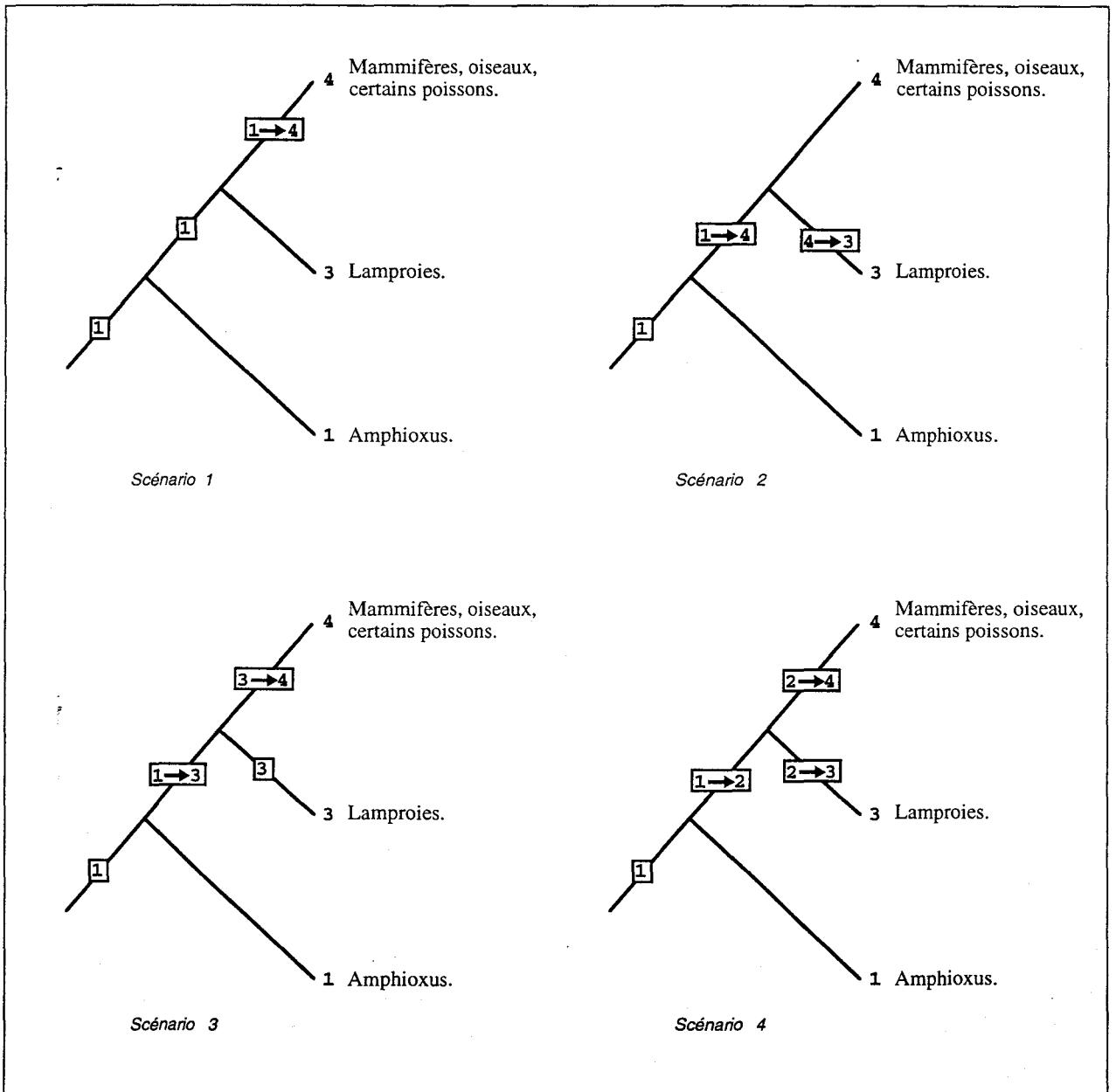
Les gènes Hox sont retrouvés uniquement dans les génomes animaux. Il existe cependant des gènes à boîte homéo chez les plantes et les champignons, mais ils ne sont pas regroupés en complexes. Les cnidaires représentent l'embranchement le plus précoce des métazoaires pour lequel un complexe Hox a pu être identifié (Martinez *et al.*, 1997). Quatre types de gènes Hox ont été retrouvés chez l'hydre, deux gènes de type antérieur, les ancêtres de *labial* et *proboscipedia* de drosophile, l'ancêtre de *Deformed*, un gène de type médian et *Abdominal-B*, un ancêtre des gènes postérieurs. Aucun ancêtre des gènes Hox médians de type *Sex Combs Reduced* n'a pu être isolé chez les cnidaires (Martinez *et al.*, 1997). Ce résultat est en contradiction avec les premières hypothèses concernant l'évolution des complexes Hox. En effet, il avait été proposé l'existence d'un complexe Hox complet chez l'ancêtre de tous les métazoaires, et un tel complexe ancestral aurait dû être retrouvé chez le métazoaire le plus précoce de la classification animale (Slack *et al.*, 1993). Selon cette hypothèse, 6 gènes Hox correspondant aux ancêtres de i) *Abdominal A*, ii) *Abdominal B*, *Ultrabithorax* et *Antenapedia*, iii) *Sex Combs Reduced*, iv) *Deformed*, v) *Proboscipedia* et vi) *Labial* auraient donc dû être retrouvés chez l'hydre, (cf. figure 23). Pourtant l'Hydre est bel et bien un animal (les critères morphologiques et structuraux qui la définissent comme tel, sont exposés dans le chapitre II). Par conséquent, l'explication la plus plausible est l'apparition tardive des gènes Hox médians, de type *Scr* ou *Ubx*, seulement après l'émergence du groupe des cnidaires.

Les gènes Hox chez l'hydre, suivent de plus la règle de co-linéarité d'expression spatio-temporelle en fonction de leur position sur le complexe. Ainsi, chez l'hydre *Chlorohydra viridissima*, *cnox1*, l'homologue de *labial* de drosophile, s'exprime au cours de la régénération, avant *cnox2*, l'homologue de *Deformed* (Schummer *et al.*, 1992). De plus, l'expression des homologues des gènes Hox chez l'hydre semblent être corrélée avec l'organisation de son axe antéro-postérieur. Les ARNm *cnox1*

a)



b)



**Figure 24 :** a) position de quelques phyla sur l'arbre évolutif  
 b) Les scénarios possibles de duplication des complexes Hox chez les cordés et les vertébrés.  
 (selon Sharman et Holland, 1998)

sont ainsi détectés en plus forte quantité dans la tête de l'hydre en cours de régénération (Shenk *et al.*, 1993a; Shenk *et al.*, 1993b).

### 3-3 La transition invertébrés/vertébrés

Les comparaisons de séquences suggèrent que l'évolution des gènes Hox s'est produite d'abord par une duplication en tandem, sur un même chromosome chez les invertébrés, formant ainsi un complexe. Puis, par la duplication d'un complexe donnant jusqu'à 4 clusters de 13 gènes Hox paralogues chez la souris (Duboule et Dolle, 1989). La recherche des gènes Hox chez les deutérostomiens, permet de situer les duplications ayant conduit aux 4 clusters. L'amphioxus, un céphalocordé, possède un seul complexe de gènes Hox (Garcia Fernandez et Holland, 1994). La lamproie, le vertébré le plus précoce actuellement vivant avec la myxine, contient, semble-t-il, 3 complexes Hox (Sharman et Holland, 1998). On s'attendait pourtant à en trouver 2, par duplication du complexe ancestral de l'amphioxus, selon l'hypothèse de la duplication du génome d'après Ohno (cf. ci-dessus) (cf. figure 24). Divers scénarios peuvent expliquer cette situation : (i) une triplification du cluster Hox chez les lamproies et une deux duplications indépendantes chez les vertébrés; (ii) une double duplication se produisant après la divergence des céphalocordés et une perte secondaire d'un des complexe chez les lamproies ; (iii) une triplification après les amphioxus et la duplication d'un seul cluster dans le groupe des vertébrés et enfin (iv) la duplication du complexe après la divergence de l'amphioxus , puis la duplication d'un seul complexe chez les lamproies et des deux complexe chez les mammifères, les oiseaux et les poissons téléostéens (Sharman et Holland, 1998). Pour éclaircir ce mystère il serait intéressant d'aller chercher les gènes Hox chez la myxine une espèce située qui a semble t-il divergée entre l'amphioxus et la lamproie. Si le nombre de complexes Hox est supérieur à 1, on aurait un idée plus précise des événements de duplication qui se sont produits après les céphalocordés.

### 3-4 Le complexe des gènes ParaHox

Il existe chez les vertébrés des gènes à boîte homéo, n'appartenant pas à un complexe, ils sont dispersés dans le génome. Il s'agit des familles Cdx, Xlox et Gsx. Or, très récemment, les ancêtres de ces gènes ont été isolés chez l'amphioxus et, sont situés sur un même chromosome. Ces gènes regroupés en un complexe appelé "ParaHox", présentent le même ordre d'expression spatial le long de l'axe du corps

de l'embryon que les gènes Hox (Brooke *et al.*, 1998). Ce complexe ParaHox, dont aucun orthologue n'a été recherché chez les organismes plus précoces comme les cnidaires, aurait pu apparaître à l'époque de l'explosion du Cambrien, dans la lignée des triploblastes et participer à la mise en place des tissus.

Parallèlement aux études faites sur les gènes Hox, il serait intéressant de replacer les événements de duplication qui ont donné naissance aux nombreux gènes de récepteurs nucléaires, et à leurs copies paralogues existant chez les vertébrés actuels. D'autre part, il ressort des études phylogénétiques des récepteurs nucléaires, une origine possible à partir d'un seul récepteur orphelin (Escriva *et al.*, 1997b). D'une part, on n'observe pas de corrélation dans l'arbre entre le récepteur et le type de ligand considéré. D'autre part, un ligand peut être fixé par un récepteur de vertébré mais pas par son homologue chez un organisme plus précoce, comme si la capacité à fixer un ligand était apparue plus tard dans l'évolution. Enfin, une recherche systématique des homologues de toutes les sous familles de récepteurs n'a permis d'isoler que des homologues de récepteurs orphelins chez les organismes précoces.

#### **4 Les récepteurs orphelins, les ancêtres des récepteurs nucléaires classiques?**

##### **4-1 Arguments en faveur de cette hypothèse**

##### **4-1-1) Pas de corrélation dans l'arbre entre récepteur et type de ligand considéré**

L'étude phylogénétique met en évidence qu'il n'existe aucun lien entre la position d'un récepteur dans l'arbre et la nature de son ligand. Par exemple, dans la famille I, les récepteurs TR, RAR et PPAR issus d'un ancêtre commun fixent des ligands totalement différents, respectivement, les hormones thyroïdiennes, les acides rétinoïques et les acides gras et leurs dérivés (dont les prostaglandines). De plus, RXR, qui a pour ligand l'acide 9-*cis* rétinoïque, fixé aussi par RAR, est placé très loin dans l'arbre phylogénétique de RAR, c'est-à-dire dans une autre sous famille de récepteurs. De même, les récepteurs orphelins sont dispersés dans tout l'arbre. La même observation peut-être faite pour les récepteurs fixant les stéroïdes. Ces observations suggèrent une indépendance entre l'histoire évolutive des récepteurs et leur capacité à fixer des ligands. Il semblerait par conséquent, que la capacité à fixer

un ligand ait évolué de façon indépendante des autres activités fonctionnelles du récepteur.

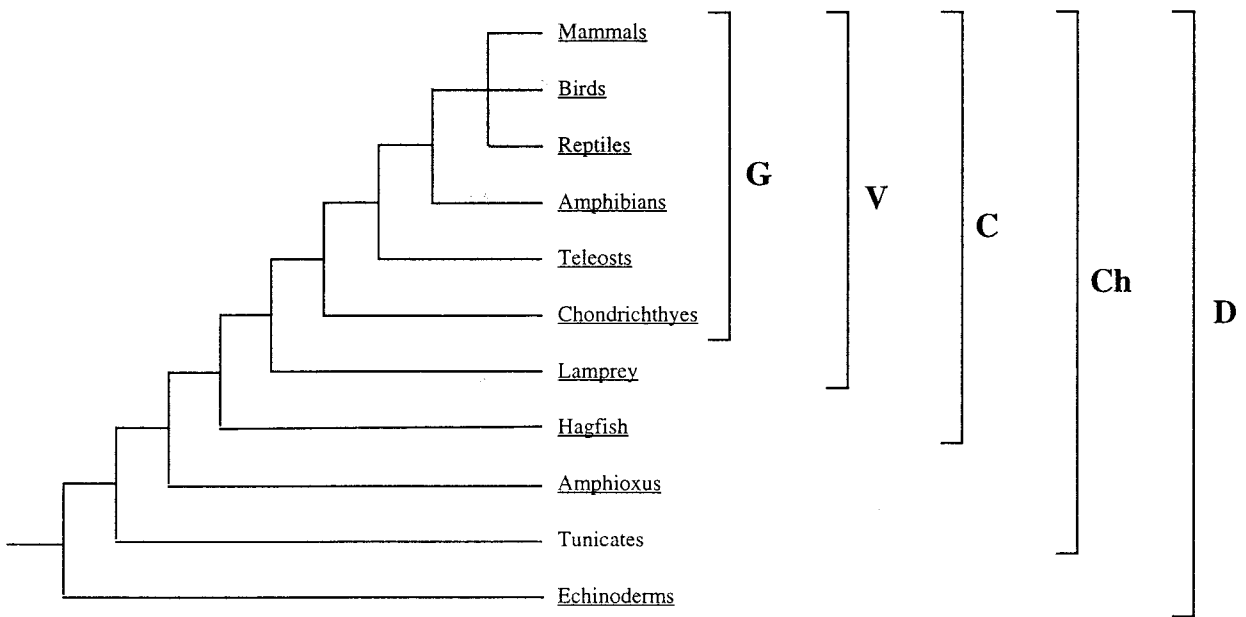
4-1-2) Un ligand peut être fixé par un récepteur de vertébré mais pas par son homologue chez un organisme plus précoce.

Des exemples de récepteurs à ligand (ou récepteurs "classiques") chez des vertébrés, auxquels correspondent des homologues invertébrés incapables de fixer ce ligand, sont en faveur de cette hypothèse d'acquisition récente de la capacité à fixer un ligand. Ainsi, le récepteur usp, l'homologue de RXR chez la drosophile est incapable de fixer les acides rétinoïques (Yao *et al.*, 1992). De même, récemment, deux gènes homologues de RXR et usp ont pu être isolés chez la tique *Amblyomma americanum*. Ceux-ci, bien que possédant un domaine E plus proche de celui des vertébrés que des insectes (71% d'identité de séquence contre 52%), sont encore incapables, comme usp, de fixer les acides rétinoïques (Guo *et al.*, 1998). Par ailleurs, très récemment, CiNR1, l'homologue de TR, isolé chez l'ascidie *Ciona intestinalis*, un urocordé, situé donc dans la lignée des vertébrés, est lui aussi incapable de fixer la triiodothyronine, le ligand de TR (Carosa *et al.*, 1998). Ces exemples de récepteurs à ligand connus chez les vertébrés, orphelins chez des organismes plus précoces, étayaient l'hypothèse de l'origine ancestrale des récepteurs orphelins.

4-1-3) Seuls des récepteurs orphelins ont pu être isolés chez des organismes précoces

Par ailleurs, au cours d'une recherche systématique des récepteurs chez tous les métazoaires (par la technique de "touch-down" avec des oligonucléotides dégénérés), nous n'avons pu amplifier que des fragments de récepteurs orphelins chez des organismes précoces dans l'échelle évolutive animale (Escriva *et al.*, 1997b). Pourtant, des oligonucléotides dégénérés pour des récepteurs à ligand, tels que ER, ont été employés selon la même méthode. Les organismes précoces ne possèderaient-ils que des récepteurs orphelins? Les résultats que nous venons d'exposer dans ce paragraphe et dans les deux paragraphes précédents vont dans ce sens. Cependant, l'absence de récepteur "classiques" est difficile à prouver formellement. Il est cependant tentant de croire avec les éléments dont nous disposons que seuls les récepteurs orphelins existaient au début de l'évolution.

Pour les raisons évoquées ci-dessous, à partir des fragments PCR obtenus par Escriva *et al.*, 1997, nous avons choisi d'isoler des récepteurs orphelins chez des



**Figure 25:** Phylogénie moléculaire des cordés d'après la séquence de l'ADN mitochondrial de l'amphioxu (Spruyt *et al.*, 1998)

organismes, tels que l'hydre et l'amphioxus, situés dans des positions clefs de la classification animale.

## III Critères de choix des modèles d'études

### 1 Choix des modèles biologiques

Afin d'analyser les deux grandes étapes de la diversification des récepteurs nucléaires, notre objectif visait à isoler, pour un récepteur considéré, l'homologue des copies des gènes paralogues, le plus proche des vertébrés, d'une part, et d'autre part, l'homologue chez un métazoaire précoce.

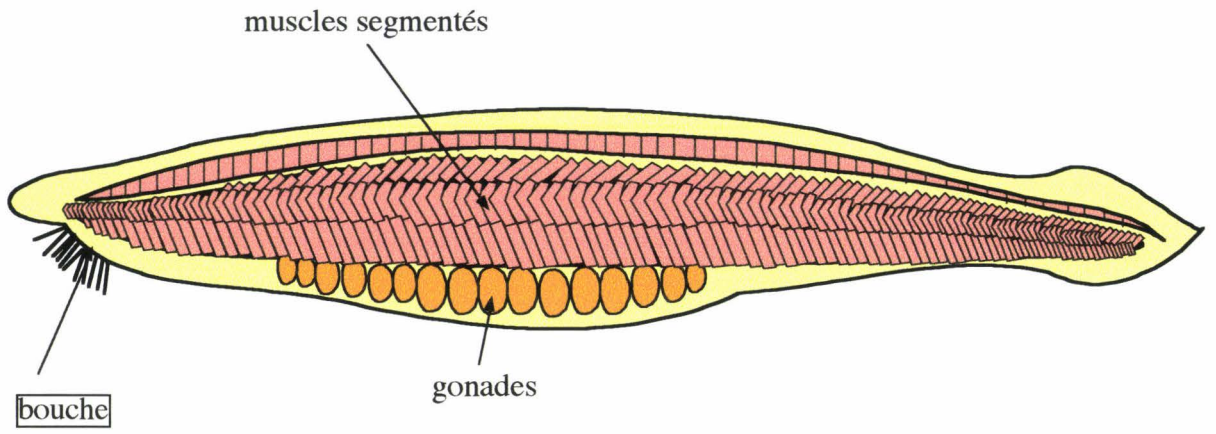
#### 1-1 **L'invertébré le plus proche des vertébrés, l'amphioxus.**

##### 1-1-1) Phylogénie moléculaire des cordés

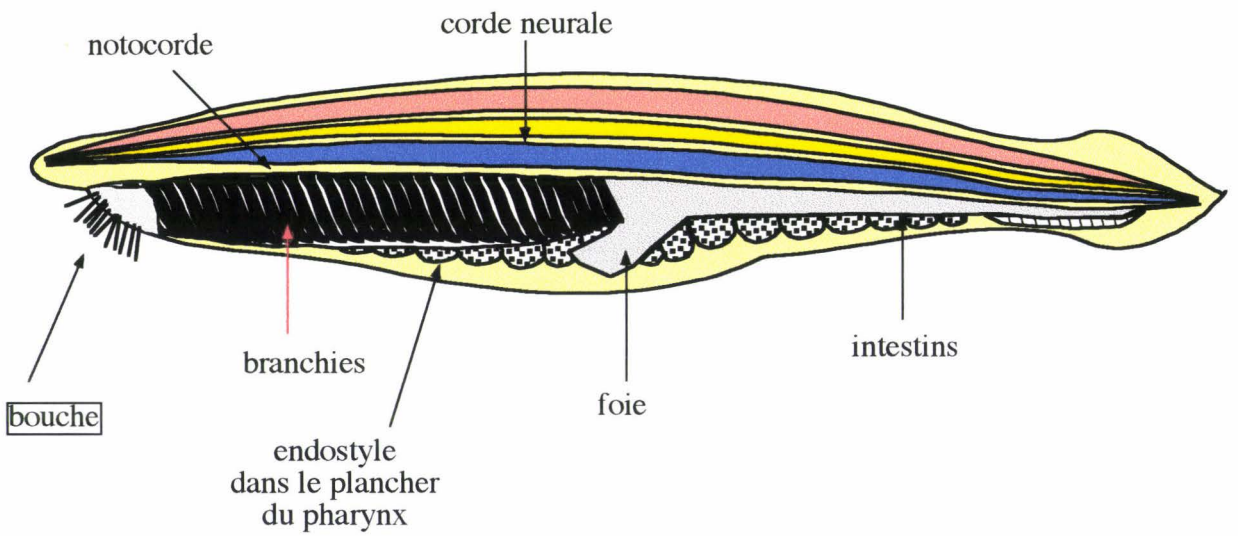
Le groupe des cordés comprend trois embranchements, les urocordés (tuniciers), les céphalocordés (amphioxus) et les vertébrés. Tous présentent des caractéristiques structurales communes qui définissent leur appartenance au groupe des cordés, notamment un tube neural dorsal, une notocorde et des fentes branchiales dans le pharynx. Une polémique a longtemps existé sur la phylogénie des procordés (urocordés et céphalocordés), et dans la recherche du groupe frère des vertébrés. Jefferies en 1986 plaçait les urocordés comme le groupe frère des crâniates, sur la base de la comparaison de caractères fossiles (cité par (Chen *et al.*, 1995)). L'hypothèse la plus commune, au contraire, supportait les céphalocordés comme le groupe frère des crâniates (Ohno *et al.*, 1968). La phylogénie moléculaire donne aujourd'hui raison à la deuxième hypothèse, par des études indépendantes de comparaison de séquences, ARN 18S (Stock et Whitt, 1992), de la séquence complète de l'ADN mitochondrial (Spruyt *et al.*, 1998) ou encore, de gènes impliqués dans le développement embryonnaire (cf. figure 25).

L'étude phylogénétique des gènes du développement des cordés, nous permet de replacer, par ailleurs, les événements de duplication qui ont permis la mise en place les nombreux gènes paralogues des vertébrés. Il est communément admis que deux étapes de duplications successives auraient conduit aux gènes actuels des vertébrés. Janvier en 1995 situe la première phase de duplication après le groupe des amphioxus et la seconde phase, juste après le groupe des lamproies, par conséquent, au cours de l'évolution précoce des vertébrés (Janvier et Grimont, 1995). Les études





a)



b)

**Figure 26 :** Représentation classique de l'amphioxus.  
 a) Vue à travers la peau transparente.  
 b) Coupe saggitale.

phylogénétiques actuelles vont dans ce sens. Un seul complexe Hox a été identifié chez l'amphioxus, contre trois chez la lamproie et quatre chez les mammifères et les oiseaux (cf. chapitre précédent, I-3). Les gènes Pax fournissent un autre exemple dans le même ordre d'idée. Ils forment une famille de neuf membres, divisée en quatre groupes, i) Pax1 et Pax-9, ii) Pax-3 et Pax-7, iii) Pax-2, Pax-5 et Pax-8 pour le troisième groupe, et iv) Pax-4 et Pax-6 pour le quatrième. AmphiPax-1 est l'ancêtre chez l'amphioxus, de Pax-1 et Pax-9 de mammifères et d'oiseau (Holland et Holland, 1996). De même, un seul gène a été identifié chez l'amphioxus, pour l'orthologue des gènes Pax2, Pax5 et Pax8 de souris (Holland L.Z., communication personnelle). Par conséquent, pour un groupe de gènes Pax chez les vertébrés correspond un seul gène Pax chez l'amphioxus. Parallèlement, un seul gène Otx a été isolé chez l'amphioxus. Il représente l'ancêtre des gènes paralogues otx1 et otx2 chez la souris et l'homme (Williams et Holland, 1998). Le génome de la lamproie, *Lampetra japonica* contient deux gènes Otx, ljotxA et ljotxB (Ueki *et al.*, 1998). Ces résultats obtenus sur les gènes Hox, les gènes Pax et otx, vont dans le sens d'une duplication du génome après la divergence des procordés. Le génome s'est-il dupliqué dans sa totalité après l'amphioxus ? Qu'en est-il des récepteurs nucléaires ? L'amphioxus représente un organisme de choix pour rechercher l'homologue des récepteurs avant leur duplication.

#### 1-1-2) Caractéristiques physiques, développementales et évolutives de l'amphioxus

*Branchiostoma floridae* est l'espèce d'amphioxus la plus étudiée en Europe et en Amérique, du fait de sa distribution dans l'Océan Atlantique. Elle est récoltée sur les côtes américaines. Les techniques de Biologie Moléculaire ont été appliquées à l'amphioxus seulement depuis le début des années 1990. Avant cela, la proximité de cet animal aux vertébrés était définie par ses caractéristiques anatomiques et structurales, sa notocorde, sa corde neurale tubulaire dorsale et ses paires de somites (Gee, 1994) (figure 26). Ces structures sont abordées ci-dessous d'un point de vue structural et définissent les caractères spécifiques de l'amphioxus.

Les somites de l'amphioxus n'ont pas de limite antérieure, elles sont distribuées tout le long du corps de l'animal. A l'opposé, chez les vertébrés, la limite antérieure des somites est située dans le rhombencéphale.

La notocorde, au lieu de s'arrêter au niveau du cerveau moyen comme chez les vertébrés, se prolonge jusqu'à l'extrémité antérieure.

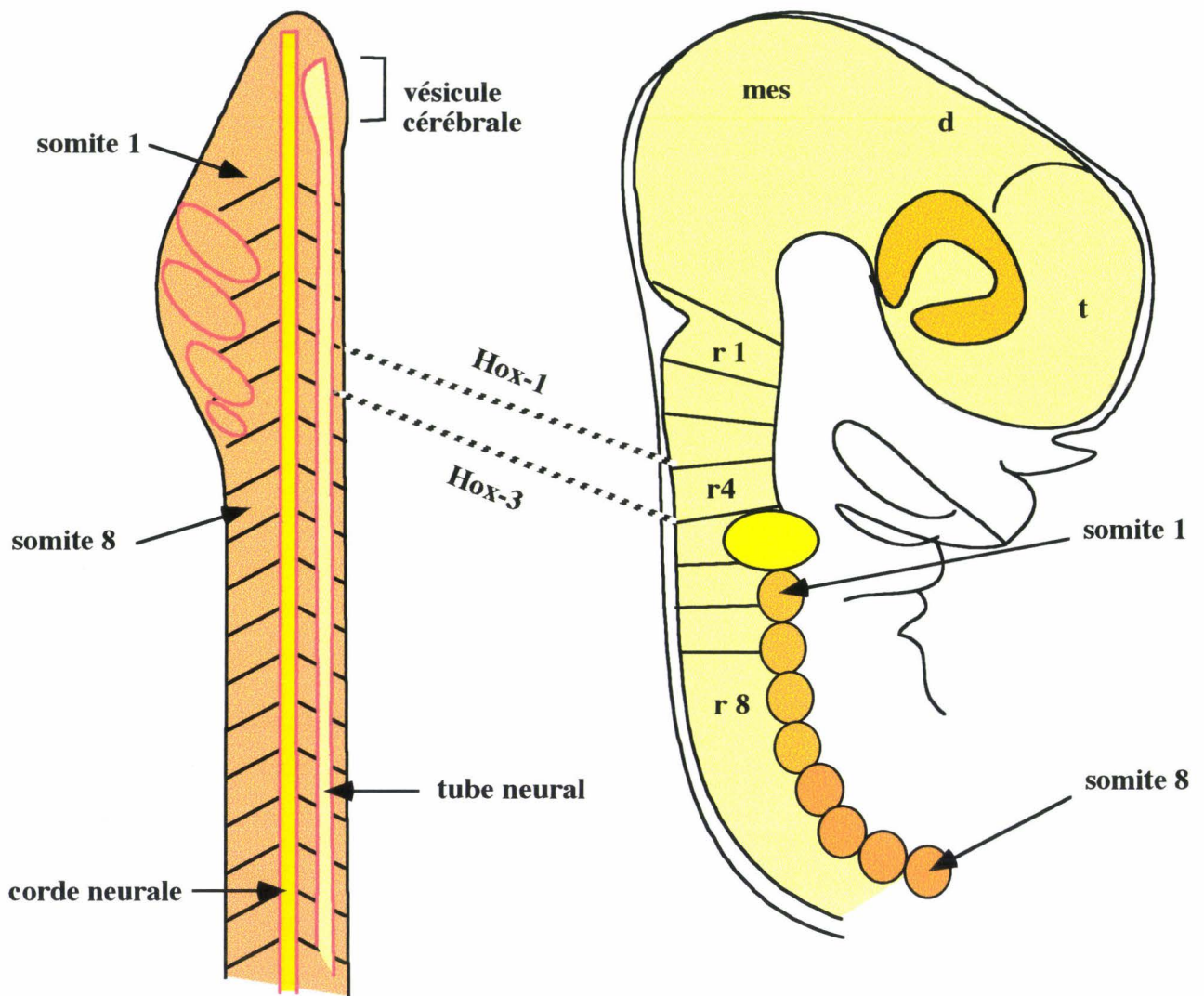
La vésicule cérébrale représente un renflement négligeable de la corde neurale qui se termine par une modeste et solitaire tache pigmentée. Lacalli et al. (Lacalli *et al.*, 1994) ont reconstruit l'architecture neurale de la vésicule cérébrale à partir de micrographies électroniques de 5000 coupes transverses sériées. Digitalisées et rentrées dans un ordinateur, elles ont permis une reconstruction en trois dimensions. Le résultat suggère qu'au moins une partie du tube neural peut être considérée comme un cerveau postérieur allongé, additionné d'un oeil ressemblant à une tache pigmentée ("eyespot"), qui, par son anatomie, semble être analogue à la paire d'yeux des vertébrés qui prend naissance dans le diencéphale. L'existence du télencéphale ou du cerveau moyen est difficile à déterminer selon les conclusions de Lacalli. L'existence des différentes divisions du cerveau chez l'amphioxus, a toujours été source de controverse. L'existence même d'un cerveau avait été remise en doute au siècle dernier.

Nous verrons dans le paragraphe suivant comment l'analyse de l'expression et de la fonction des gènes du développement chez l'amphioxus, peuvent compléter les données des études évolutives, pour déterminer l'homologie entre deux structures.

### 1-1-3) Le système neveux central et les gènes du développement

#### a) *La notion d'homologie*

Les cartographies d'expression de nombreux gènes du développement donnent des indications non seulement sur le développement de l'amphioxus, mais aussi sur l'évolution. Dans ce contexte, l'homologie entre deux structures peut-être évaluée. Deux structures chez deux espèces sont considérées homologues si elles ont dérivé de la même structure chez l'ancêtre de ces deux espèces. Cette définition de l'homologie est souvent étendue en incluant les origines développementales, sans lesquelles il est difficile de formuler le concept de "même" ou "similaire". Ainsi, Roth définit en 1984 la base de l'homologie comme le partage des voies de régulation du développement. Ce raisonnement implique que les limites antérieures de l'expression des gènes Hox chez l'amphioxus, par exemple, sont homologues aux mêmes limites chez les vertébrés. L'amphioxus possède par conséquent une partie de son cerveau homologue au cerveau postérieur des vertébrés (cf. ci-dessous).



**Figure 27** : Régions d'homologies entre le tube neural d'une larve d'amphioxus (à gauche) et le cerveau d'un embryon de vertébré supérieur (à droite), déterminées à partir d'une analyse comparative de l'expression des gènes Hox.

La limite rostrale d'expression du gène *AmphiHox-1* est comparée aux limites d'expression des gènes Hox du groupe paralogue 1, de la souris et du poulet.

Une comparaison analogue a été effectuée entre *AmphiHox-3* et le groupe paralogue 3 de la souris et du poulet.

d, diencéphale, mes, mésencéphale, r, rhombomère, t, télencéphale  
(d'après Holland et Garcia-Fernández, 1996)

*b) Les gènes du développement, marqueurs des divisions du cerveau*

Le cerveau postérieur : Selon, la notion d'homologie définie ci-dessus, on peut considérer que les limites antérieures de l'expression des gènes *AmphiHox-3* et *AmphiHox-1*, chez l'amphioxus, sont homologues aux limites d'expression des groupes paralogues 3 et 1 chez les embryons de vertébrés. Pour les gènes du groupe 1, cette limite est située entre les rhombomères 3 et 4. Pour les gènes du groupe 3, elle se situe à la frontière des rhombomères 4 et 5. Le tube neural de l'amphioxus en développement ne présente pas de segmentation externe claire, mais peut être décrit en se référant aux paires de somites adjacentes. *AmphiHox-3* a une limite rostrale dans le tube neural, adjacent à la séparation entre les somites 4 et 5. La limite pour *AmphiHox-1* s'étend sur une somite au dessus (Holland et Garcia-Fernandez, 1996) (figure 27). Le cerveau de l'amphioxus possède donc l'équivalent d'un cerveau postérieur qui présente, comme le rhombencéphale des vertébrés, un semblant de segmentation délimitée par l'expression des gènes Hox.

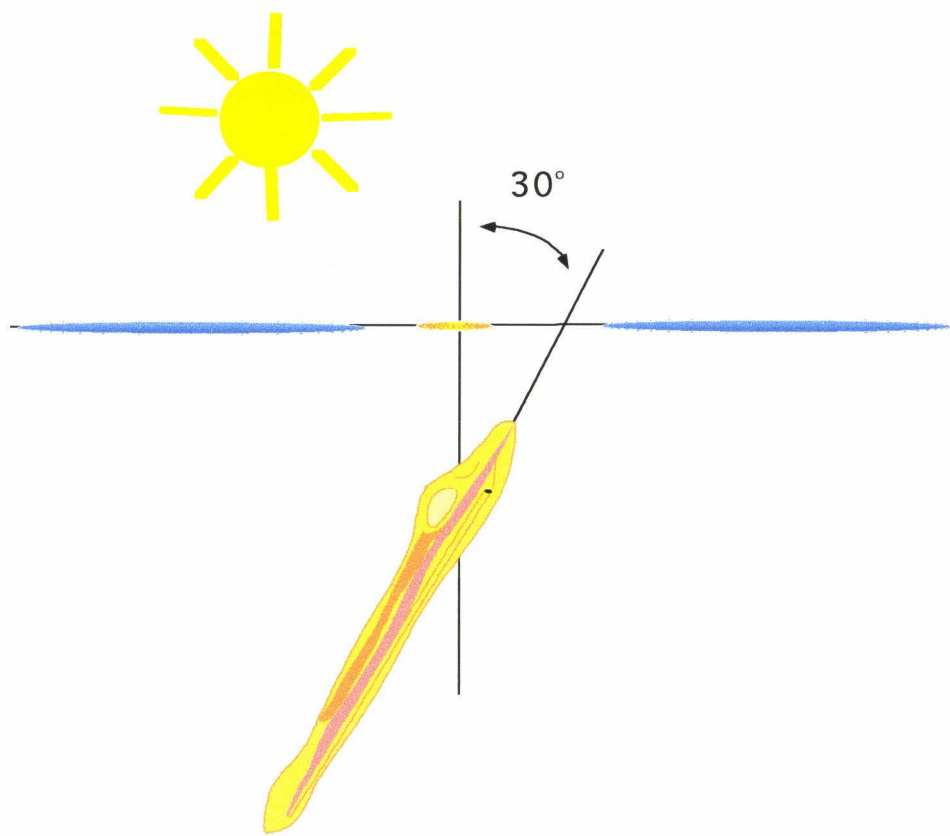
Dans la région antérieure au cerveau postérieur de l'amphioxus, existe-t-il des structures homologues aux différentes divisions du cerveau des vertébrés: télencéphale, diencéphale, cerveau moyen?

le cerveau moyen

*AmphiEn*, l'homologue des gènes *engrailed 1* et *2* de vertébrés (exprimés dans le cerveau moyen), isolé chez l'amphioxus, s'exprime dans une fine bordure en amont du cerveau postérieure de l'amphioxus. De plus, l'homologue de *Pax-2/5/8*, des gènes exprimés aussi dans le cerveau moyen chez les vertébrés, vient d'être isolé chez l'amphioxus (Holland L.Z., communication personnelle). L'existence du cerveau moyen n'est pas encore déterminée de façon claire chez l'amphioxus, mais ceci ne saurait tarder...

le cerveau antérieur (télencéphale et diencéphale)

*Otx* est un gène déterminant pour l'élaboration du cerveau antérieur chez la souris, il s'exprime dans le futur télencéphale, le diencéphale et plus tard dans l'oeil en cours de formation. L'expression de *Amphiotx* chez l'amphioxus est corrélée avec celle de ses orthologues chez les vertébrés. Chez ces deux phyla, les ARNm *otx* sont détectés à des stades précoces du développement dans le neuroectoderme antérieur et le mésoderme sous-jacent (Williams et Holland, 1998). Ces résultats d'expression suggèrent un maintien des mécanismes d'induction neuronaux, notamment les inductions mésoderme-ectoderme qui initient la formation de l'oeil



**Figure 28 :** Posture stationnaire de la larve d'amphioxus maintenue par des battements de cils. En présence de lumière, la larve s'oriente lentement de manière à placer son oeil dans l'ombre.  
(Selon Stokes at Holland, 1995)

à partir du diencephale des vertébrés. *AmphiPax-6* s'exprime dans les mêmes territoires que *otx*, chez les vertébrés comme chez l'amphioxus (Glardon *et al.*, 1998). De même *NK2-2*, un gène contrôlé par *Pax-6* chez les vertébrés est détecté dans les mêmes territoires. Ces résultats suggèrent l'existence d'un cerveau antérieur chez l'amphioxus, cependant rien ne permet de distinguer d'après les résultats d'expression le télencéphale du diencephale. De plus, Lacalli a observé des structures similaires au diencephale, mais pas au télencéphale.

#### Structures dérivées du cerveau antérieur, l'oeil et l'hypophyse

Quatre différents groupes de cellules sont supposées être des photorécepteurs (Ruiz et Anadon, 1991; Lacalli *et al.*, 1994; Lacalli, 1996). (i) Tout d'abord, l'oeil frontal qui comprend la tache pigmentaire antérieure et les cellules réceptrices et neuronales associées, situés en position antéro-ventrale de la vésicule cérébrale. Cet oeil semble être constitué de cellules photoréceptrices. En effet, un amphioxus, dont l'oeil frontal est placé en face d'une source lumineuse, se repositionnera de manière à ce que le rayon de lumière ne soit plus qu'à 30° des cellules de sa tache pigmentée, de manière à garder son "oeil" dans l'ombre (Stokes et Holland, 1995) (cf. figure 28). (ii) L'organe lamellaire, localisé dorsalement dans la moitié postérieure de la vésicule cérébrale, possède des neurones ciliés similaires à ceux observés dans l'oeil frontal. Cependant, cet organe semble plutôt être l'homologue de l'épiphyse des vertébrés, comme nous le verrons ci-dessous. (iii) Les cellules de Joseph, localisées dorsalement, s'étendent à partir de l'organe lamellaire sur quelques centaines de micromètres en aval. (iv) Les organes de Hesse, localisés ventralement, s'étendent à partir d'une région chevauchant les cellules de Joseph les plus postérieures, jusqu'à l'extrémité caudale de la corde nerveuse. Aucune de ces cellules pigmentées n'a véritablement pu être identifiée comme cellule photo-réceptrices, avant d'isoler des gènes homologues de vertébrés chez l'amphioxus.

Les cellules photo-réceptrices et les neurones de l'oeil frontal expriment *AmphiPax-6*, mais pas les autres photorécepteurs (Glardon *et al.*, 1998). De part ce résultat et les observations exposées précédemment, l'oeil frontal de l'amphioxus semble représenter l'homologue de l'oeil de vertébré, de drosophile et de nématode. Et de façon générale, il est fort probable que l'oeil ait eu une origine monophylétique dans différents phylas à travers l'évolution.

*AmphiPax-6* s'exprime aussi dans le puits de Hatschek qui était supposé être l'homologue de l'hypophyse antérieure selon des critères cytologiques et

endocriniens (Nozaki et Gobman, 1992; Welsch, 1978; Zhang, 1982). Pax-6 de vertébrés s'exprime aussi dans l'hypophyse. De plus, très récemment, l'expression de Pit-1 a pu être mise en évidence par immunofluorescence dans cette région, chez l'amphioxus (Candiani et Pestarino, 1998)

### La notocorde

La notocorde est un des éléments qui caractérisent les cordés. Brachyury (T) est en quelque sorte, le "gène maître" de la notocorde; il s'exprime dans la notocorde présomptive chez les vertébrés et induit sa formation. Amphi Brachyury s'exprime dans le mésoderme présomptif, autour du blastopore chez un embryon au stade gastrula (Holland *et al.*, 1995). Puis son expression est localisée dans toute la notocorde, jusque dans la région la plus rostrale, ce qui représente une caractéristique de l'amphioxus. C'est en effet le seul animal à posséder une notocorde aussi rostrale, puisque même son ancêtre probable *Pikaia* du Cambrien, présentait une notocorde se terminant en arrière de l'extrémité rostrale (Chen *et al.*, 1995). De plus, l'expression d'Amphi Brachyury est détectée également dans le mésoderme dorsal, caractéristique commune à celle des vertébrés. L'homologue de Brachyury chez l'ascidie, As-T, par exemple, est restreinte à la notocorde (Yasuo et Satoh, 1993). De plus, la conservation de la fonction de Brachyury a pu être vérifiée chez l'ascidie. L'injection d'ARNm As-T dans les blastomères de tuniciers, induit en effet, l'expression de marqueurs spécifiques de la notocorde, tels que Not-1, détecté par immunofluorescence (Yasuo et Satoh, 1993).

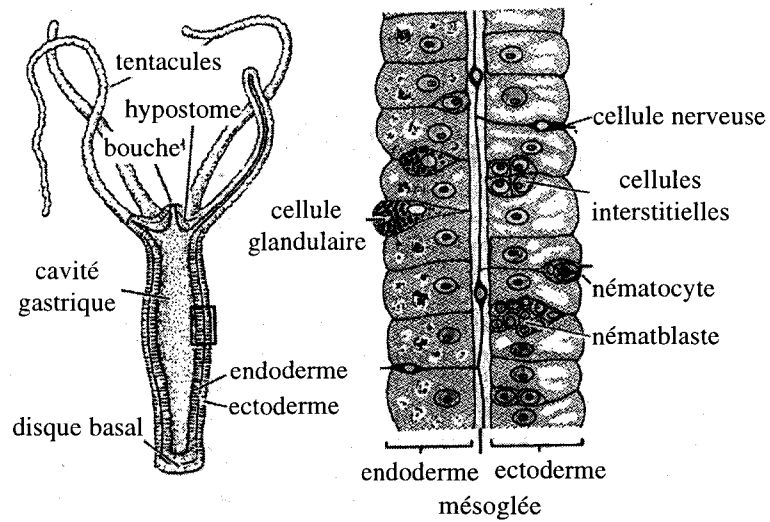
### Le tube neural

Comme chez les vertébrés, le facteur diffusible ventralisant HNF3, est exprimé dans le plancher du tube neural (dans la notocorde aussi) (Shimeld, 1997), ce qui représente un élément de plus pour assimiler la corde neurale de l'amphioxus comme le tube neural des vertébrés.

### Les cellules de crêtes neurales

On a longtemps cru l'amphioxus était dépourvu de cellules de crêtes neurales. Pourtant, ces dernières années, l'expression de *Dll*, un gène marqueur des cellules de crêtes neurales chez les vertébrés, a pu être mis en évidence dans des cellules dorsales du tube neural ([Holland LZ, 1995]). Il reste à déterminer si ces cellules possèdent des propriétés migratoires comme leurs homologues de vertébrés.





**Figure 29:** Schéma d'une hydre. L'hydre est un polype constitué de deux couches de cellules épithéliales, l'endoderme et l'ectoderme, séparées par la mésoglée (cf à droite du schéma). Ce diploblaste présente une organisation antéro-postérieure, avec une bouche à l'extrémité apicale et un pied ou disquebasal à son extrémité basale.

### c) Les gènes *AmphiHox* et la réponse aux acides rétinoïques

Les acides rétinoïques provoquent l'antériorisation de l'expression du gène *AmphiHox-1*, de la même façon que les gènes *Hox*, orthologues du groupe 1 des vertébrés (Shimeld *et al.*, 1996). De plus, cette antériorisation est marquée par l'absence de fentes branchiales et l'expression extra-temporelle de *AmphiPax-1* dans le pharynx de l'amphioxus (Holland *et al.*, 1995). Cependant l'existence d'acides rétinoïque n'a pas encore été mise en évidence chez des organismes autres que les vertébrés.

Tous les résultats convergent vers la proximité de l'amphioxus et des vertébrés. Cet animal possède en effet, au niveau de son système nerveux central, un tube neural, une notocorde, un cerveau restreint, mais présentant néanmoins la majorité des structures présentes chez les vertébrés et des cellules ressemblant aux cellules de crêtes neurales des vertébrés. Par ailleurs, les effets observés par l'adjonction d'acide rétinoïque sont similaires à ceux observés chez les vertébrés.

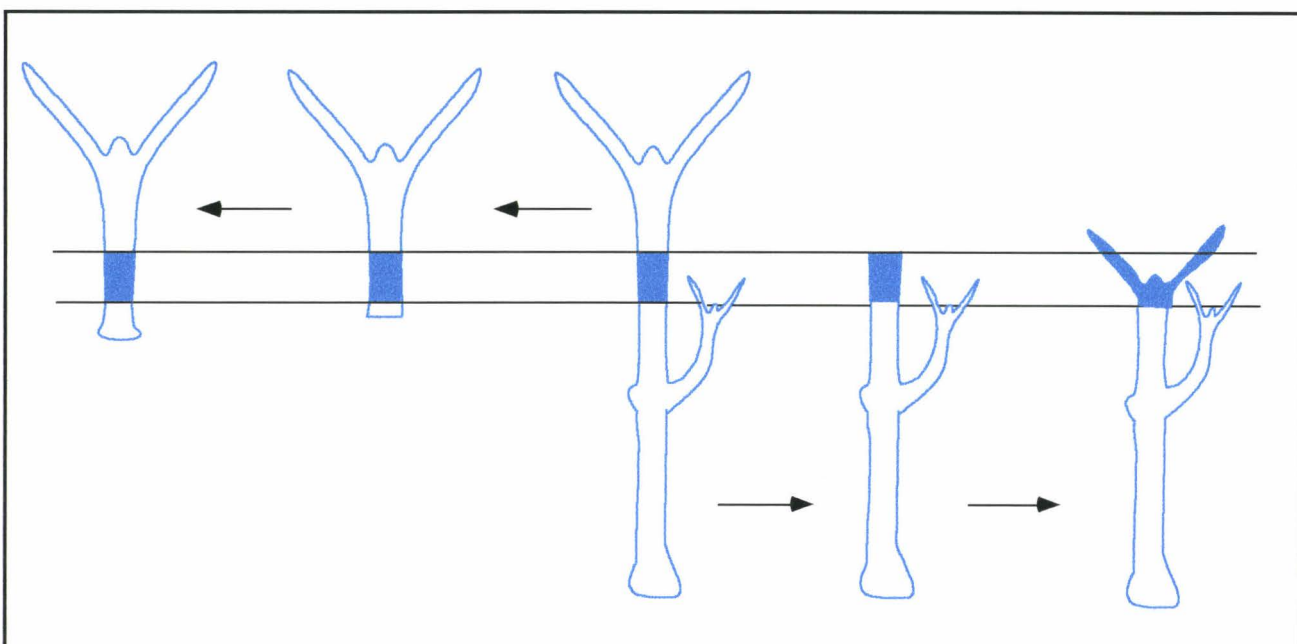
## 1-2 L'hydre, un métazoaire précoce

L'hydre est le métazoaire le plus précoce de la classification animale après les éponges. L'appartenance des éponges au groupe des métazoaires est encore très controversée, c'est pourquoi nous avons choisi de rechercher la fonction ancestrale de récepteurs orphelins des métazoaires chez l'hydre, qui offre par ailleurs de nombreux avantages méthodologiques.

### 1-2-1) Caractéristiques structurales et morphologiques

L'hydre est un animal diploblaste, car issu des deux feuilletts embryonnaires, l'ectoderme et l'endoderme. C'est en outre le premier métazoaire à posséder des cellules sensorielles nerveuses. Ce métazoaire, présente des caractéristiques surprenantes, telle que la capacité à générer une nouvelle hydre par bourgeonnement ou à régénérer n'importe quelle partie de son corps.

Le polype est un simple tube de 0,5 cm de longueur, avec une région apicale, appelée tête, et un pied à son extrémité opposée, qui lui permet d'adhérer au support sur lequel il est fixé (cf. figure 29). Le tube est composé de deux couches de cellules épithéliales, l'endoderme et l'ectoderme, séparées par une structure extracellulaire de collagène, la mésoglée. La tête possède une ouverture, la bouche, entourée de tentacules. Celles-ci sont surmontées de cellules contractiles, les nématocystes, utilisées pour la capture des proies. Seulement 20 types de cellules sont répertoriées



**Figure 30:**

Schéma représentatif de l'expérience classique de régénération de l'hydre.  
Les régions sectionnées à des niveaux identiques montrent que les mêmes cellules peuvent former soit la région gastrique, soit le pied ou la tête selon leur position le long de l'axe du polype.  
(selon Müller, 1996)

et on comptabilise 100 000 cellules pour l'animal entier (cf. (Müller, 1996) pour une revue récente sur l'hydre). En cas de nourriture abondante, un surplus de cellules est engendré et exporté par un processus de bourgeonnement, donnant naissance à une nouvelle hydre. La reproduction sexuée existe aussi chez l'hydre, mais elle est moins fréquente que le bourgeonnement. L'hydre est immortelle, car les cellules sont sans cesse renouvelées par de nouvelles cellules engendrées par les cellules souches. Ces dernières, cellules multipotentes, à capacité migrative, sont situées dans la cavité gastrique, au centre du polype. La propriété de l'hydre à régénérer une partie de son corps sectionné, présente un intérêt tant dans l'étude du fonctionnement de la mise en route de ce phénomène, que dans l'avantage qu'offre un tel modèle pour l'étude des gènes du développement sollicités pour régénérer de nouvelles structures.

### 1-2-2) La régénération

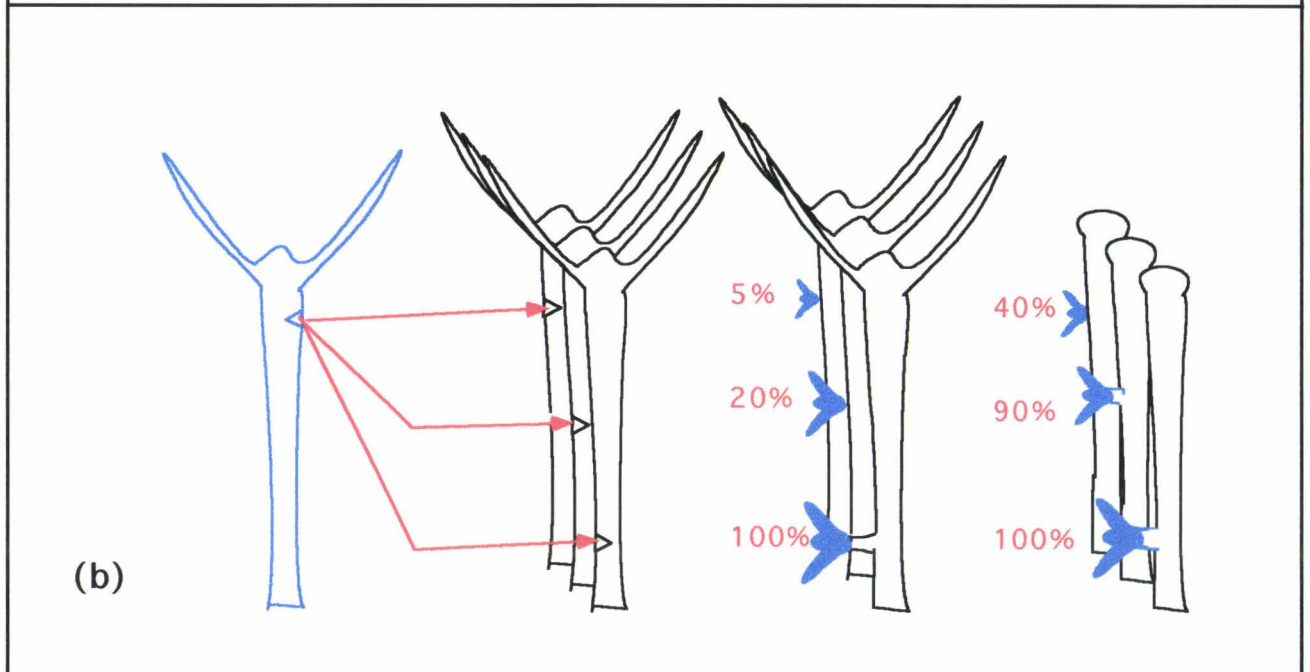
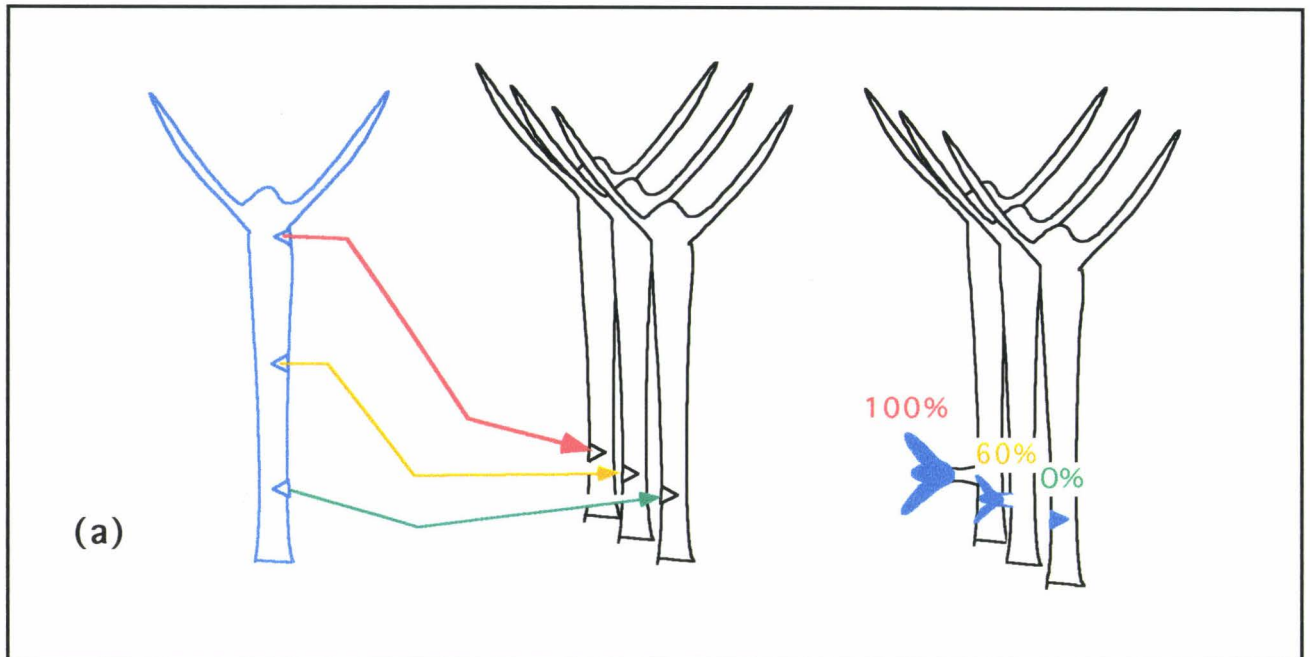
#### *a) le mécanisme de régénération*

Une hydre sectionnée en dessous de sa tête, est capable de régénérer un pied et vice versa (cf. figure 30). De plus, cette régénération obéit à une loi de position, c'est-à-dire que les cellules de l'hydre ont une "valeur positionnelle" selon l'endroit auquel elles sont situées. Par exemple, des cellules issues d'un tête greffée sur un pied redonneront un tête dans 100% des cas. Si par contre, ces mêmes cellules de tête sont greffées sur des régions plus proches de la tête, le pourcentage de donner un tête diminue (cf. figure 31). Il ressort de ces expériences qu'il existe un facteur capable d'engendrer une tête ou un pied et un autre capable d'inhiber la formation d'une telle structure. En fait, ce sont des molécules de type neuropeptides qui jouent ces rôles. On trouve un activateur de la tête (HA pour "Head Activator"), colocalisé avec un inhibiteur HI et parallèlement, un activateur du pied (FA pour "Foot Activator"), co-localisé avec un inhibiteur (FI pour "Foot Inhibitor") (Schaller *et al.*, 1989).

#### *b) Les voies de régulation impliquées dans la régénération*

##### **La voie de la PKC / DAG**

Deux gènes de la PKC (protéine kinase C) ont été identifiés chez l'hydre, HvPKC1 et HvPKC2 (Hassel, 1998; Hassel *et al.*, 1998). HvPKC1 est exprimé dans toutes les régions du polype adulte, sauf le pied. HvPKC2 est exprimé très tôt dans les processus de régénération de la tête et du pied. En outre, l'augmentation de l'activité PKC est spécifiquement associée à la régénération de la tête. L'activité PKC participe



**Figure 31:** Schéma représentatif des expériences de transplantation conduisant à la formation de tête ou de pied ectopique.

a) Il existe un facteur activateur de la formation de la tête, HA (head activator), plus fortement concentré dans la tête. Plus le transplant est prélevé près de la tête et plus son potentiel de régénérer une tête est élevé.

b) Il existe par ailleurs, un facteur inhibiteur de la régénération de la tête, HI, qui est également plus concentré dans la tête. Par conséquent, plus un transplant, prélevé près d'une tête d'un polype est greffée loin de la tête de l'hôte et plus les chances d'obtenir une tête sont élevées. Corrolèremment, l'ablation de la tête chez le polype hôte augmente les chances de régénération d'une tête pour les transplants greffés dans les régions sous jacentes à l'ancienne tête.

(Selon Müller, 1996).

donc aux processus de régénération, et HvPKC1 plus particulièrement dans l'élaboration de la tête.

### **La voie de l'AMPc / PKA**

La voie de la PKC participerait au processus médié par l'AMPc, dans la détermination et la différenciation des cellules nerveuses et des cellules épithéliales de la tête. L'AMPc augmente en effet les concentrations de PKC dans les endroits où s'effectuent les processus de différenciation terminale de l'hydre, la tête et le pied (Galliot *et al.*, 1995). Chez les mammifères, l'AMPc passe par la voie de la PKA, qui active les facteurs de transcription CREB (cAMP responsive element, binding factors). Or, l'homologue de CREB a été isolé chez l'hydre verte *Chlorohydra viridissima*, et participe à un complexe de fixation à l'ADN, *via* un site CRE (cAMP responsive element). Ce complexe de fixation augmente aussi sous l'effet de HA (Galliot *et al.*, 1995). La voie de l'AMPc représente donc un médiateur de l'activateur de la tête et sans doute de l'activateur du pied, pour la détermination des cellules.

#### *c) Les gènes du développement sollicités lors de la régénération*

Les gènes Hox, nous l'avons vu précédemment, sont stimulés au cours de la régénération, ce qui est corrélé avec leur rôle dans l'organisation axiale des organismes. Par ailleurs, le gène *ks1* est exprimé dans les cellules épithéliales des tentacules et dans la région gastrique sous-jacente, après stimulation au TPA (activateur de la PKC) (Weinziger *et al.*, 1994). Il semble par conséquent participer à la formation de la tête. Récemment, un autre gène impliqué dans le développement de la tête *prdl-a* a été isolé (Gauchat *et al.*, 1998). *prdl-a*, apparenté au gène *aristalless* de drosophile, est restreint aux cellules nerveuses de la tête chez l'hydre. Par ailleurs, *msh*, l'homologue de *msx-2* des vertébrés impliqué dans le bourgeonnement des membres au cours de la vie embryonnaire, est exprimé faiblement dans les régions de l'hydre compétentes pour la formation de bourgeons. Puis l'expression de *msh* forme une tache circulaire intense à l'endroit où le prochain bourgeon va se former (Müller, 1996).

### **1-2-3) Le système nerveux**

#### *a) Un réseau de cellules nerveuses*

L'hydre possède le système nerveux le plus simple des métazoaires. Il consiste en un réseau de cellules nerveuses qui s'étendent dans tout le corps de l'animal, avec des densités plus fortes dans la tête et le pied (Bode *et al.*, 1990; Bode, 1992). Les

cellules nerveuses individuelles peuvent être caractérisées en tant que cellules sensorielles ou ganglionnaires, de part leur morphologie (Davis *et al.*, 1968; David et Gierer, 1974). Des sous populations de cellules contenant des neurotransmetteurs spécifiques (cellules aux récepteurs GABA ou à la sérotonine etc... ) sont localisés dans différentes régions de l'hydre (McCauley, 1997; Concas *et al.*, 1998). De plus, l'existence de monoamines, précurseurs de la dopamine, a été mise en évidence chez un autre cnidaire (Pani et Anctil, 1994).

b) *L'Hydre possède-t-elle un système nerveux sensoriel?*

Le polype est capable de répondre à des stimuli tactiles, chimiques ou lumineux (Bode, 1992). Sa couche épithéliale de cellules musculaires est longée de cellules ganglionnaires qui peuvent répondre à cette fonction. Cet organisme exprime-t-il des gènes impliqués dans le développement du système sensoriel chez d'autres espèces? Le gène *Achaete-scute* a été isolé chez *Hydra vulgaris*, CnASH (Grens *et al.*, 1995). *Achaete-scute* est impliqué dans l'induction des cils vibratoires tactiles localisés sur l'abdomen ventral des Drosophiles. De plus, sa fonction est conservée chez l'hydre puisque l'expression ectopique de CnASH dans des larves de drosophile, conduit à la formation d'organes sensoriels ectopiques (Grens *et al.*, 1995). L'expression de CnASH est localisée dans toute la région abdominale de l'hydre et dans les cellules du nématocyste (organe de capture des proies).

Les quelques gènes homologues aux gènes du développement embryonnaire de vertébrés, isolés chez l'hydre, semblent avoir conservé leur fonction principale au cours de l'évolution : gènes *Cnox/Hox* et organisation antéro-postérieure, *msh/msx-2* et bourgeonnement des membres, *prdl-a/prdl* et formation de la tête, *Achaete-scute* et détermination des cellules sensorielles. Qu'en est-il des récepteurs nucléaires?

## 2 Choix des récepteurs pour notre étude

### 2-1 Les COUP-TF : des récepteurs orphelins très conservés

#### 2-1-1) Isolement

COUP-TFI et COUP-TFII ont été isolés chez l'homme, la souris, le xénope, le poisson zèbre et l'oursin (Miyajima *et al.*, 1988; Wang *et al.*, 1989; Qiu *et al.*, 1995; Tsai et Tsai, 1997). Ce sont des gènes extrêmement bien conservés dans toutes les espèces.

Deux gènes existent chez tous les vertébrés connus, tandis que chez le xénope et le poisson-zèbre un gène supplémentaires (COUPG ou SVP46) a été mis en évidence.

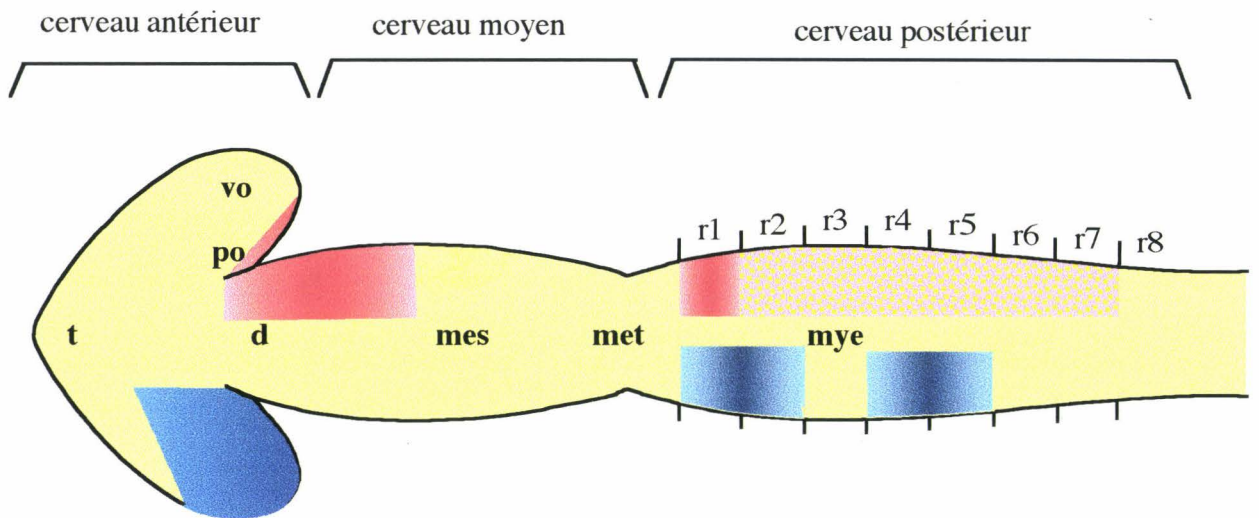
### 2-1-2) Patron d'expression des gènes COUP-TF.

COUP-TFI et COUP-TFII présentent des cartographies d'expression distinctes bien que chevauchantes au cours du développement embryonnaire (Tsai et Tsai, 1997). Chez la souris, ils sont initialement exprimés dans les trois couches germinatives et plus tard, dans certains de leurs dérivés (Pereira *et al.*, 1995). Leur expression est d'abord détectée à 8,5 jours post-coïtum, avec un pic à 14-15 jours, puis un déclin jusqu'à la naissance (Qiu *et al.*, 1995). Ils sont notamment exprimés dans le système nerveux central et périphérique, chez la souris, le poisson zèbre, la drosophile. Leur homologue de drosophile SVP joue un rôle dans la détermination des cellules photoréceptrices chez la drosophile (Fjose *et al.*, 1993). Ils s'expriment par ailleurs dans certains organes et dans les membres au cours du développement embryonnaire.

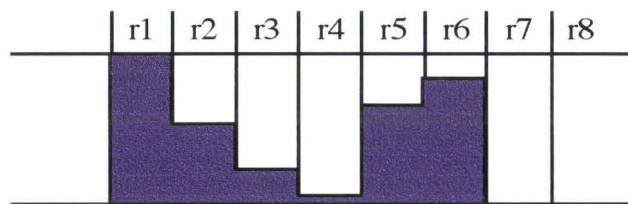
#### a) *vésicules cérébrales*

COUP-TFI et II s'expriment dans toutes les vésicules du cerveau chez la souris (Tsai et Tsai, 1997). On détecte en effet, leurs ARN messagers dans le cerveau antérieur : dans le tronçon optique, dans la portion caudale du télencéphale et dans le thalamus et dans le cerveau moyen. Dans le cerveau postérieur, leur expression démarre dans le rhombomère r1 et s'étend caudalement (Qiu *et al.*, 1994). A 14,5 jours post-coïtum, l'expression de COUP-TFI et II est aussi observée dans le cervelet (Jonk *et al.*, 1994a). COUP-TFI et II s'expriment différemment dans les neuromères du diencéphale (structures globulaires qui segmentent cette région du cerveau). Ils pourraient de ce fait, participer au maintien des neuromères D1 à D4 du diencéphale (qui représentent les structures primordiales des futures divisions représentées par le thalamus ventral et l'hypothalamus, D1, le thalamus dorsal, D2 et la région prétectale D3 et D4) ou intervenir dans la détermination d'un type de cellule neuronale (Qiu *et al.*, 1994). Chez la souris adulte, l'expression de COUP-TFI et COUP-TFII est aussi détectée dans le cerveau. Les ARN messagers de COUP-TFI sont retrouvés de façon prédominante dans les régions rostrales et caudales du cerveau, tandis que les ARN messagers de mCOUP-TFII sont majoritaires dans les régions médianes du cerveau (Lopes da Silva *et al.*, 1995).





**32a :** Expression de svp[44] (=COUP-TFI) ■ et svp[46] (=COUP-TFG) ■ chez un embryon de zébrafish de à 14 heures de développement embryonnaire  
 d, diencéphale, mes, mesencéphale, met, métencéphale, mye, myélencéphale, po, pédoncule optique, vo, vésicule optique, r, rhombomères



**32b :** Gradient d' expression de svp[40] (=COUP-TFII) dans les rhombomères d'un embryon de zébrafish

**Figure 32 :** Expression des COUP-TFs dans le système nerveux central de l'embryon de zébrafish

Chez le zébrafish, les homologues des COUP-TFs s'expriment aussi dans les vésicules cérébrales (figure 32).

*b) Cellules de crêtes neurales et dérivés*

COUP-TFI et II sont détectés sous forme de deux raies dans le futur cerveau postérieur à 8,5 jours post-coïtum, ce qui laisse penser à une expression dans les cellules de crête neurale en migration (Qiu et al., résultats non publiés). Dans des stades ultérieurs, les ARNm de COUP-TFI sont détectés en forte quantité dans les rhombomères r1 à r4 et à des faibles niveaux dans r5 et r6. L'expression de COUP-TFI est localisée très nettement dans les cellules de crête neurale avant et après leur migration au niveau des rhombomères. Ces cellules participent à la formation du 9<sup>e</sup> nerf cranien. L'inactivation du gène COUP-TFI, par recombinaison homologue, entraîne d'ailleurs, un défaut de morphogénèse de ce nerf et de son ganglion (Qiu *et al.*, 1997). Les souris homozygotes meurent à la naissance de déshydratation et de mal nutrition, l'absence de ce nerf entraînant des troubles de la déglutition.

*c) Expression dans la moelle épinière*

Chez le poulet, l'expression de COUP-TFII est localisée dans les neurones moteurs de la moelle épinière. De plus, la transplantation d'une notocorde supplémentaire à proximité de la région dorsale du tube neural induit l'expression ectopique de COUP-TFII. Dans cette région du tube neural normalement occupée par les neurones sensoriels, se différencient de plus, des neurones de type moteur, identifiés par l'expression ectopique des marqueurs de ce type de neurone, SC1 et Islet 1 (Lutz *et al.*, 1994). Ces résultats suggèrent que COUP-TFII participe à la détermination des cellules neuronales en neurones moteurs. Chez la souris, COUP-TFI présente une expression étendue à tout le tube neural, tandis que COUP-TFII est restreint aux neurones moteurs, comme chez le poulet.

*d) Expression pendant l'organogénèse*

mCOUP-TFI est exprimé dans le stroma du septum nasal, dans la langue, dans les follicules de vibrassae et dans la cochlée. mCOUP-TFII est exprimé dans les mêmes régions mais à des niveaux moindres. Par contre, mCOUP-TFII est fortement exprimé dans les glandes salivaires, les poumons, l'oesophage, l'estomac, le pancréas primaire, les mésonéphros, les ovaires, la rétine, le bourgeon de membre, la peau, l'oreille interne. Tandis que mCOUP-TFI s'exprime aussi dans ces tissus, mais plus faiblement, à l'exception des glandes salivaires.

e) Expression chez l'adulte

(81)

L'expression des COUP-TFs est très réduite chez la souris adulte. A la fois mCOUP-TFI et II présentent des cartographies d'expression distinctes et restreintes dans le cerveau adulte. mCOUP-TFI est exprimé dans la partie rostrale et caudale du cerveau de souris adulte, tandis que mCOUP-TFII est exprimé dans la partie médiane du cerveau. De plus forts niveaux de mCOUP-TFI sont détectés dans le noyau olfactif, le néocortex et le gyrus dentate et dans des régions variées de l'hippocampe et du cervelet. mCOUP-TFII est détecté dans les noyaux thalamiques et réticulés ainsi que dans la membrane mydaloïde et arachnoïdienne. Ces deux facteurs sont exprimés à des niveaux élevés dans de nombreuses autres régions (Lopes da Silva *et al.*, 1995).

f) Expression dans d'autres organismes.

*Seven-up* est exprimé dans les neuroblastes chez la *Drosophile*, qui donnent naissance au système nerveux central chez cet arthropode (Broadus et Doe, 1995). Chez l'oursin, l'expression de l'homologue de COUP-TF, spCOUP-TF est détecté dans l'ovaire et au cours de l'embryogenèse précoce (Chan *et al.*, 1992).

2-1-3) Mutants des gènes COUP-TF

Des mutants "nuls" (ou "knock-out" en anglais) de COUP-TFI ont été générés par la technique de recombinaison homologue (Qiu *et al.*, 1997). Les mutants homozygotes meurent après la naissance entre 8 et 36 h et 95 % d'entre eux présentent une fusion asymétrique des nerfs crâniens glossopharyngien (IX) et vague (X). De plus, l'arborescence des axones est sévèrement affectée dans les régions des plexus cervicaux. Les os exoccipitaux droit ou gauche sont fusionnés à l'os basioccipital dans 98 % des mutants "nuls". Des fusions similaires sont observées chez le double mutant nul RAR a1 et RARb. Comme COUP-TFI est régulé par les acides rétinoïques durant la différenciation des cellules de carcinome embryonnaire, il est considéré comme une des cibles en aval de la voie de signalisation des rétinoïdes (Jonk *et al.*, 1994b). Par ailleurs, comme nous le verrons dans le chapitre III, il est aussi capable d'antagoniser cette voie (Neuman *et al.*, 1995).

2-1-4) COUP-TF interfère aussi avec d'autres récepteurs nucléaires

Les protéines COUP-TF ont été décrites comme des répresseurs transcriptionnels. De plus, la protéine COUP-TFI interfère avec les réponses cellulaires aux rétinoïdes (tout-*trans*-RA, au 9-*cis*-RA) à la vitamine D, à l'hormone

thyroïdienne et aux (Cooney *et al.*, 1992; Kliewer *et al.*, 1992; Tran *et al.*, 1992). Pour ces raisons il nous a semblé particulièrement intéressant de caractériser COUP-TF, un récepteur orphelin très conservé et qui est fonctionnellement "connecté" aux voies de signalisation par les ligands des récepteurs nucléaires "classiques" comme RAR ou TR, dans une vaste gamme de métazoaires.

## 2-2 Les récepteurs orphelins TR2 et TR4

### 2-2-1) Isolement

#### a) TR2

TR2 a été isolé pour la première fois en 1988 par C. Chang, à partir d'une banque d'ADNc de testicules humains à l'aide d'une sonde oligonucléotides de 41 bases spécifique du domaine DNA-binding des récepteurs aux stéroïdes. Au total, 4 ADNc ont été identifiés, issus du même gène, probablement par un mécanisme d'épissage alternatif et différants dans leur région 3'. Ainsi, les protéines hTR2-11, hTR2-9 et hTR2-5 varient dans leur région C terminale à partir de l'acide aminé en position 465, tandis que hTR2-7 est totalement délété de son domaine E (Chang et Kokontis, 1988; Chang *et al.*, 1989).

Plus tard, hTR2-11 a été localisé sur le chromosome 10 chez la souris, ce qui laisse supposer une position en 12 q chez l'homme (Lee *et al.*, 1996a).

#### b) TR4

Trois groupes différents ont isolé des ADNc humain et/ou murin de TR4 en 1994, dénommés alors TR2R1 (Law *et al.*, 1994), TAK1 (Hirose *et al.*, 1994) et TR4 (Yoshikawa *et al.*, 1996a) en criblant des banques d'ADNc de testicules, de cerveau ou de prostate. Par ailleurs, hTR4 a été localisé par hybridation sur le chromosome, dans la région 3p25.

### 2-2-2) Expression

#### a) *Au cours de l'embryogénèse*

Les ARNm de TR2-11 fluctuent au cours du développement embryonnaire chez la souris. Ils apparaissent à 8 jours p.c.. En 24 h, leur proportion par rapport à la quantité des ARNm totaux, a doublé. Ils diminuent ensuite de 20% durant les 48 h suivantes. Puis, ils augmentent de 150% pour atteindre un sommet à E12. A partir de E13, l'expression décroît régulièrement pour atteindre un niveau basal en fin de gestation qui représente 10% de l'expression relevée à E12 (Lee *et al.*, 1996a). Ces

résultats suggèrent qu'il existerait différents mécanismes d'activation du gène TR2-11, au cours de l'embryogénèse. Premièrement, l'initiation de l'expression à 8/8,5 j p.c., deuxièmement, une stimulation additionnelle à E9, plus ou moins entretenue, puisqu'on observe une chute à E11. Puis, de nouveau une activation transcriptionnelle, très forte cette fois, à E12, qui ne sera pas maintenue car le taux d'ARNm de TR2-11 diminue jusqu'à la naissance. Néanmoins, il est difficile d'interpréter des résultats obtenus à partir d'ARNm totaux, car nous avons vu qu'après la naissance, TR2-11 ne s'exprime plus que dans des tissus spécifiques et par conséquent, le signal obtenu par Lee CH est dilué sur l'ensemble des ARNm.

*b) Chez l'adulte*

Chez l'adulte, l'expression de TR2 est détectée dans les testicules, débutant chez des souris âgées de 18 jours et plus, un stade où la méiose active commence à se produire. Le signal est localisé dans le compartiment des tubes séminifères où résident les cellules germinales avancées. De plus, des souris ayant subi une diète en vitamine A, présentent des testicules déplétés en cellules germinatives avancées et l'expression de la protéine TR2 n'est plus visualisée par immunocytochimie. Ces observations suggèrent une relation biologique de ce récepteur orphelin avec la différenciation des cellules germinatives (Lee *et al.*, 1995a; Lee *et al.*, 1996a) et potentiellement une régulation directe de l'expression de TR2 par l'acide rétinoïque, comme c'est le cas pour COUP-TF. TR2 s'exprime aussi chez l'adulte, dans d'autres organes que les testicules, comme les ovaires, la glande adrénale, le coeur, le thymus et le cerveau, mais à des niveaux moindres (Lee *et al.*, 1996a).

Les ARNmTR4 s'expriment plus fortement dans le système nerveux central. Par expérience de Northern sur des tissus de rat, en utilisant le signal obtenu pour l'hypothalamus comme base, soit 100 %, on observe: 35 % d'expression dans la glande adrénale, 34 % dans la rate, 33 % dans les testicules, 27% dans la glande thyroïde, 17% dans la prostate et 8% dans la glande pituitaire. Par ailleurs, bien qu'aucune expression n'a pu être détectée par Northern blot dans l'épididyme, la vésicule séminale, la glande sous maxillaire, le coeur, le rein, les poumons et le foie, une amplification a néanmoins pu être réalisée par PCR (Chang *et al.*, 1994). De plus, (Yoshikawa *et al.*, 1996b), détectent par Northern, sur des ARN poly A humains, une forte expression de TR4 dans le coeur, le cerveau, le placenta, les poumons, le muscle squelettique, le pancréas, la rate et le foie pour lequel l'expression est moins intense. Pour TAK1, Hirose *et al.* (Hirose *et al.*, 1994) obtiennent le même profil

d'expression, à cela près, qu'ils ont testé en plus, le sang périphérique, le petit intestin et les ovaires qui expriment aussi cet ARN messager. D'autre part, il détectent un ARNm plus petit, présent seulement dans les testicules (2,8 Kb pour l'ARNm le plus petit, et 9,4 Kb pour le plus grand qui est quasi ubiquitaire). TR2R1 code 2 ARNm de taille similaire à TAK1. Le plus grand est lui aussi quasi ubiquitaire et le plus petit n'est plus cette fois spécifique des testicules, mais du cerveau (Law *et al.*, 1994).

L'expression de TR4 dans le cerveaux est particulièrement intéressante. Les ARNm de TR4 sont localisés dans les aires hypothalamiques, thalamiques (cellules granulaires) et corticales. Une expression intense est par ailleurs observée, dans les cellules granulaires du cervelet (Chang *et al.*, 1994). L'expression dans le cerveau est notamment retrouvée dans le thalamus, le noyau sous thalamique, la substance nigria, l'hypothalamus, l'hippocampe, le corps calleux, le noyau caudal et les amygdales ((Yoshikawa *et al.*, 1996b), par Northern blot).

#### 2-2-3) Mécanisme d'action de TR2 et TR4

Comme les récepteurs COUP-TF, TR2 et TR4 sont des récepteurs orphelins qui ont une action de répression de la transcription. Le parallèle avec les COUP-TF peut même aller plus loin puisque ces récepteurs, comme les COUP-TF semblent être capables de jouer un rôle antagoniste de celui des récepteurs de l'acide rétinoïque. Il était donc intéressant de caractériser un homologue de TR2/4 chez l'amphioxus, en parallèle de la caractérisation d'un homologue de COUP-TF.



*Résultats*  
*Récepteurs Nucléaires*

*Article 4*



#### Article 4

### Ligand binding was acquired during the evolution of nuclear receptors

Hector Escriva, Rachid Safi, Catherine Hänni, Marie-Claire Langlois, Pierre Saumitou-Laprade, Dominique Stéhelin, André Capron, Raymond Pierce, and Vincent Laudet.

Proc. Natl. Acad. Sci. USA, 94, 6803-6808, 1997

Afin de retracer les différentes étapes de l'évolution des récepteurs nucléaires, nous avons recherché l'existence de ces récepteurs dans différents phyla des métazoaires ainsi que dans des organismes non métazoaires. Cette recherche a été menée en utilisant la méthode de PCR. Dix séries d'amorces ont été choisies pour amplifier le second doigt de zinc du domaine de fixation à l'ADN (domaine C) qui est le domaine le plus conservé des récepteurs nucléaires. Afin de vérifier si les récepteurs nucléaires existent en dehors des métazoaires, nous avons recherché ces molécules chez les plantes, les champignons et les algues. Nous avons également recherché ces récepteurs dans une vaste gamme de métazoaires depuis les éponges et les cnidaires jusqu'aux vertébrés en passant par les acoelomates (les vers plats tels que le schistosome), les protostomiens comme les arthropodes, les mollusques ou les annélides et différents deutérostomiens tels que des échinodermes et les différents organismes importants pour la transition cordés-vertébrés : hémicordés, urocordés et céphalocordés. C'est au moyen de deux amplifications successives PCR en utilisant toutes les combinaisons possibles d'amorces internes compatibles selon la technique de "PCR nesting" et par la méthode de PCR "touch-down" au cours de laquelle la température d'hybridation des amorces décroît tous les 5 cycles que nous sommes parvenus à isoler ces fragments de gènes de récepteurs nucléaires.

Nous avons obtenus 84 fragments de d'ADN pour tout l'éventail de métazoaires testés. Les récepteurs nucléaires sont spécifiques des métazoaires, puisque nous n'avons trouvé des fragments de récepteurs nucléaires que chez ces espèces. Aucun fragment n'a été obtenu chez les plantes, les champignons ou les algues.

Nous montrons de plus dans cet article, qu'il n'existe pas de lien entre la position d'un récepteur donné dans l'arbre et la nature chimique de son ligand. Les récepteurs orphelins sont d'ailleurs dispersés dans tout l'arbre évolutif. Par ailleurs,

les résultats de PCR montrent que chez les métazoaires les plus précoces comme l'hydre ou l'anémone de mer, nous n'avons trouvé que des homologues de récepteurs orphelins. Ces données suggèrent donc que les premiers récepteurs nucléaires étaient des récepteurs orphelins qui ont acquis la capacité à fixer un ligand plus tardivement, au cours de l'évolution.

# Ligand binding was acquired during evolution of nuclear receptors

JEFFREY A. ESCRIVA\*, RACHID SAFI†, CATHERINE HÄNNI†, MARIE-CLAIRE LANGLOIS†, PIERRE SAUMITOU-LAPRADE‡, DOMINIQUE STEHELIN†, ANDRÉ CAPRON\*, RAYMOND PIERCE\*, AND VINCENT LAUDET†§

Centre National de la Recherche Scientifique Unite Mixte de Recherche 319, Mécanismes du Développement et de la Cancérisation, and \*Institut National de la Santé et de la Recherche Médicale U167 Relations Hôte-Parasite et Stratégies Vaccinales, Institut Pasteur de Lille, 1 rue Calmette, 59019, Lille Cedex, France; †Centre National de la Recherche Scientifique Unite de Recherche Associée 1185 Génétique et Evolution des Populations Végétales, Université de Lille 1, S.M. 1185, Villeneuve d'Ascq, France

Communicated by Roy J. Britten, California Institute of Technology, Corona Del Mar, CA, April 25, 1997 (received for review December 5, 1996)

**ABSTRACT** The nuclear receptor (NR) superfamily comprises, in addition to ligand-activated transcription factors, members for which no ligand has been identified to date. We demonstrate that orphan receptors are randomly distributed on the evolutionary tree and that there is no relationship between the position of a given liganded receptor in the tree and the chemical nature of its ligand. NRs are specific to metazoans, as revealed by a screen of NR-related sequences in early- and non-metazoan organisms. The analysis of the NR gene duplication pattern during the evolution of metazoans shows that the present NR diversity arose from two waves of gene duplications. Strikingly, our results suggest that the ancestral NR was an orphan receptor that acquired ligand-binding ability during subsequent evolution.

The nuclear receptor (NR) superfamily contains ligand-activated transcription factors that exert widely different biological responses by regulating target gene expression positively and/or negatively (1, 2). Apart from receptors that bind steroid hormones, retinoic acid, or thyroid hormones, this superfamily contains so-called orphan receptors for which no ligand is known. Recently, ligands have been proposed for some orphan receptors such as PPAR, LXR, or FXR. Interestingly most of these newly discovered ligands act in an intracrine fashion. However, in most cases, and despite intensive efforts, ligands have not been found for most orphan receptors (e.g., COUP-TF, HNF4, SF1, etc.), the biological role of which remains enigmatic (1, 2).

The modular organization of NRs, the various degrees of conservation between their respective domains and the importance of NRs for many physiological processes in both invertebrates and vertebrates, have led several authors to study these molecules from an evolutionary point of view (3–5). We showed previously that the NR superfamily is ancient in origin and was well diversified before the arthropod/vertebrate split, because most of the subfamilies of NRs that we defined, and most of the groups of receptors, were already present before their divergence (3). However, the precise origin of the family is still unclear. To elucidate this crucial point, the origin and diversification of NRs was studied by amplifying fragments of NR genes in various organisms that are located at critical positions in the evolutionary tree. Our analysis demonstrates that NRs appear specific to metazoans and that two waves of gene duplications led to their present diversity. Furthermore, our results suggest that orphan receptors are ancient and that liganded receptors independently acquired the ability to bind ligand later in evolution.

Publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/946803-6\$2.00/0

Table 1. Authenticity of the PCR products

Clone	%AT	Southern blot	Genomic/cDNA
<i>Hydra</i> COUP	+		C
<i>Schistosoma</i> TR4	+	+	G
<i>Schistosoma</i> RXR	+	+	GC
<i>Schistosoma</i> FTZ-F1	+	+	C
<i>Schistosoma</i> COUP	+	+	
<i>Schistosoma</i> COUPII	+	+	
<i>Phallusia</i> FTZ-F1			G
<i>Branchiostoma</i> COUP			C
<i>Branchiostoma</i> TR2			C
<i>Danio</i> REVERB			C
<i>Danio</i> COUP			C
<i>Danio</i> PPAR			C
<i>Danio</i> RAR			C
<i>Danio</i> NGFIB			C

%A/T +: Sequence of the PCR fragment is rich in bases A and T, as is the genome of the corresponding organism. Southern blot +: Positive signal with the corresponding probe under high stringency conditions. G: Isolation of a genomic clone. C: Isolation of cDNA clones.

## MATERIALS AND METHODS

**DNA Extraction.** DNA was extracted according to standard procedures (6) from zebrafish (*Danio rerio*), amphioxus (*Branchiostoma lanceolatum*), acorn worm (*Balanoglossus* sp.), sea urchin (*Sphaerechinus granularis*), tunicate (*Phallusia mammillata*), mollusk (*Biomphalaria glabrata*), annelid (*Perinereis cultrifera*), brine shrimp (*Artemia salina*), flatworm (*Schistosoma mansoni*), cnidarians (*Hydra vulgaris* and *Anemonia sulcata*) fresh tissues; and from shark (*Scyliorhinus canicula*), lamprey (*Petromyzon marinus*), hagfish (*Myxine glutinosa*) frozen tissues. Sponge (*Pseudocorticium jarre*) DNA was a gift of R. Garrone (Lyon, France). Non-metazoan samples were from plants (*Beta vulgaris*, *Thymus vulgaris*, *Daucus carota*, and *Sparganium erectum*), a fungus (*Glomus intraradices*), a red alga (*Gracilaria verrucosa*), and a protist (*Trypanosoma cruzi*).

**PCR Analysis.** Three or four primers were designed for each studied group of receptors (TR, RAR, PPAR, REV-ERB, ROR, EcR, LXR, NGFIB, ER, steroid receptors, FTZ-F1, RXR, COUP-TF) and all possible combinations of primers were used in semi-nested PCR amplifications. Primer sequences are available from V.L. upon request. PCR experiments were done according to a modified "touch-down" method (7) using the *Taq* polymerase and buffer from Eurogentec (Brussels). DNA extraction, PCR amplification, and cloning/sequencing were performed under strict conditions to avoid contamination. All the PCR experiments were carried out in the laboratory of P.S.-L., where no animal DNA is

Abbreviation: NR, nuclear receptor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U93406–U93489).

§To whom reprint requests should be addressed. e-mail: vlaudet@infobiogen.fr.

**RAR (subfamily I)**

CKGFFRRSIQKNMVTCHRDKNCLINKVTRNRCQYCRLLQKCFEVMGMSKE	Homo RAR $\alpha$	
-----I-----V-----	Homo RAR $\beta$	
-----	Homo RAR $\gamma$	
-----E-S-Q-----	Danio RAR $\alpha$	(2)
-----R-E-S-----S-----	Petromyzon RAR	(7)
-----K-----Q-V-----V-----F--K--D-----	Branchiostoma RAR	(19)

**PPAR (subfamily I)**

GFFRRITIRLKLVDKCDRSCKIQKGNRNKCCYCRFHKLCSVGM	Homo PPAR $\alpha$	
-----M--E--E--E-----Q--AL-----	Homo PPAR $\beta$	
-----R-Q-H--N-----N--S-----	Xenopus PPAR $\beta$	
-----I--R--LN-R-H--S-----	Homo PPAR $\gamma$	
-----E--E--R-----Q--FPL-----	Danio PPAR $\alpha$	(8)
-----E--Q--E--R-----I-----	Danio PPAR $\alpha$ (II)	(5)
-----V--E--E--N-----K-----	Danio PPAR $\alpha$ (III)	(1)
-----V--ER-E-----S--F-----	Danio PPAR $\beta$	(6)
-----V--ER-E-A--V--S-----	Danio PPAR $\beta$ (II)	(1)
-----H--LH-R-H--S-----FP---	Danio PPAR $\gamma$	(8)
-----M--K--E--N-----	Scyliorhinus PPAR $\alpha$	(2)
-----V--D--Q--E--N-----S--Q--FP---	Scyliorhinus PPAR $\gamma$	(22)
-----M--E--E--QR-----S--H-----	Petromyzon PPAR $\alpha$	(2)
-----M--E--GN--GLN-----F-----	Petromyzon PPAR $\beta$	(1)
-----MSTQVQ--ET--QMN--V--S--R--F-----	Petromyzon PPAR $\gamma$	(7)
-----V--E--EA--A--H--RVH--S--F-----	Myxine PPAR	(3)
-----KS--T--KP--G--TVHRS-----Q--P-----	Branchiostoma PPAR	(11)

**NURR1/NGFIB (subfamily IV)**

VQKNAKYICLANKDCPVDKRRRNRCQFCRFQKCLAVGMVKE	Homo NGFIB	
-----V--N-----Y-----	Homo NURR1	
-----N-----S-----	Rattus NOR-1	
-----D--V--N-----Y--V-----	Danio NGFIB	(1)
-----GS--V--N-----Y--V-----	Danio NURR1	(1)
-----GS--V--N--I-----Y--V-----	Petromyzon NURR1	(1)

**FTZ-F1 (subfamily V)**

CECCKGFFKRTVQNNKHVYCT***ESQCKIDKTRKRCPPFCRFQKC	Mus SF1	
-----Q--R--I***--N--N--Q-----Y--K--	Mus LRH1	
-----K--V--V***--AER--H-----Y-----	Drosophila FTZ-F1	
-----R--I***--N--Q-----P--	Danio SF1	(3)
-----R--***--QN--D--G-----QP---	Danio LRH1	(3)
-----R--Q--I***--N--N--V-----S--P--	Branchiostoma FTZ-F1	(2)
-----G-----G--K--S--V***--HTSR--L-----G--QY--SY--	Phallusia FTZ-F1	(9)
-----KQLK--S--K***--K--D--EVT--FS--NN--QP---	Balanoglossus FTZ-F1	(2)
-----K--V--I***--NRN--A-----HP---	Sphaerechinus FTZ-F1	(4)
-----K--V--M***--ADR--H--S-----P--	Artemia FTZ-F1	(4)
-----KRI--V***--ADGK--E--T--A--N--QP---	Artemia FTZ-F1(II)	(1)
-----A--R--A--HRPNA--SR--E--NVAS--K--P--S--	Schistosoma FTZ-F1	(9)
-----A--KQLD--K***--ADGN--EVNQMT--N--HP--S--	Anemonia FTZ-F1	(3)
-----KQLQ--A--V***--NK--Q--NN--I--HP--S--	Anemonia FTZ-F1(II)	(1)

**RXR (subfamily II)**

CEGCKGFFKRTVVKDLTYTCRD*NKDCLIDKQRNRCQYCRYQK	Mus RXR $\alpha$	
-----I-----S--*-----TV-----	Mus RXR $\beta$	
-----I--I-----*	Mus RXR $\gamma$	
-----N-----*-----V-----S	Danio RXR $\alpha$	(1)
-----Q--S--Q--I--K--*--SQM--F--N--FS--S--PP--F	Biomphalaria RXR	(2)
-----E--I--*--SQE--Q-----L-----S--	Schistosoma RXR	(5)
-----S--QNKK-----NLT--PM--H-----S--	Anemonia RXR	(1)
-----R--SIK--R--Y--*--AGC--P--S--Y--S--H--SLPS	Anemonia RXR(II)	(1)

**COUP-TF (subfamily II)**

CEGCKSFFKRSVRNRLTYTCRANRNCPIDQHHRNQCYCRLKCKLVG	Homo COUP-TFI	
-----S-----	Homo COUP-TFII	
-----TI--S--S--D--Q-----	Homo EAR2	
-----S--GS-----	Danio SVP46	
-----S--G--D-----S	Drosophila SVP	
-----N--S--D--Q-----	Danio COUP	(5)
-----C-----S--V-----	Danio COUP(II)	(1)
-----GS--D-----S	Danio COUP(III)	(3)
-----G-----FA--	Danio COUP(IV)	(1)
-----C-----S	Danio COUP(V)	(12)
-----C-----	Danio COUP(VI)	(1)
-----C-----	Scyliorhinus COUP	(20)
-----C-----I	Scyliorhinus COUP(II)	(2)
-----GS-----I	Petromyzon COUP	(6)
-----S--G--T-----S	Myxine COUP	(10)
-----S--G-----S	Branchiostoma COUP	(3)
-----GS-----	Phallusia COUP	(3)
-----C-----GS-----	Balanoglossus COUP	(8)
-----G--T-----S	Sphaerechinus COUP	(7)
-----I--A--N--GTKQ--V--V-----	Biomphalaria COUP	(3)
-----G--Q--N--N--KQ--IN-----	Perinereis COUP	(4)
-----I--A--FQ--YS--LN-----	Schistosoma COUP	(6)
	Schistosoma COUP(II)	(5)
	Hydra COUP	(4)

FIG. 1. Deduced amino acid sequence alignment of the NR PCR products corresponding to 6 of the 13 studied groups. The sequences are compared with their mammalian or arthropod homologues, which are shaded. Only amino acids different to those in the first sequence are indicated.

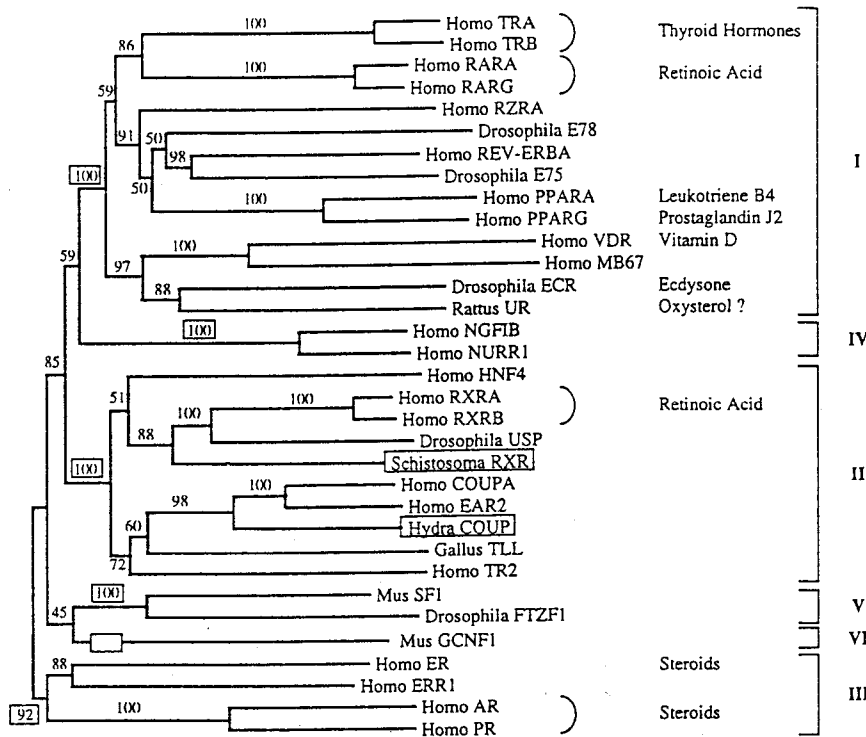


FIG. 2. Phylogenetical tree of 33 selected NRs. The six subfamilies are shown. Subfamilies are defined by clusters supported by high robustness "bootstrap" values above 90 starting from the base of the tree, which was rooted at mid-length of its longest path. Ancestral segments for subfamilies 1 to 5 have their bootstrap value boxed. Subfamily VI defined by only one member is shown by a small box along its branch. Ligands are indicated when applicable. The two boxed sequences correspond to cDNA clones of early metazoan NRs isolated in this study and corresponding to the PCR products of Fig. 1.

handled. To avoid crosscontamination, the DNA from chordates and vertebrates was extracted and submitted to PCR after completion of the PCR done on non- and early-metazoan samples.

**Cloning and Sequencing.** PCR fragments were cloned in the TA cloning vector (Invitrogen) and sequenced using an Applied Biosystems 373A automatic sequencer and methods and reagents of the supplier.

**Sequence Analysis.** Assignment of each sequence to a particular group was done after removal of the primer sequence and translation by alignment and construction of phylogenetic trees. The assignment of the sequence of the PCR products corresponding to genes, which were then isolated as cDNA (see Table 1), was proven to be correct, suggesting that no major assignment errors were made using short sequences. Nevertheless, in some cases such as *Danio* (a partially tetraploid organism) the precise assignment to a given paralogous gene could be problematic.

Sequences of all known NR from databases (see ref. 1) were aligned using the MUST package (8). Only the C and E domains were analyzed. Distance trees were calculated using the neighbor-joining method with 1,000 bootstrap replicates. To decrease calculation times three independent sets of 33 receptors that gave identical results were treated separately. The clustering into subfamilies and groups was confirmed by parsimony analysis using the PAUP program (9). The complete phylogenetic reconstruction will be published elsewhere (V.L., unpublished work).

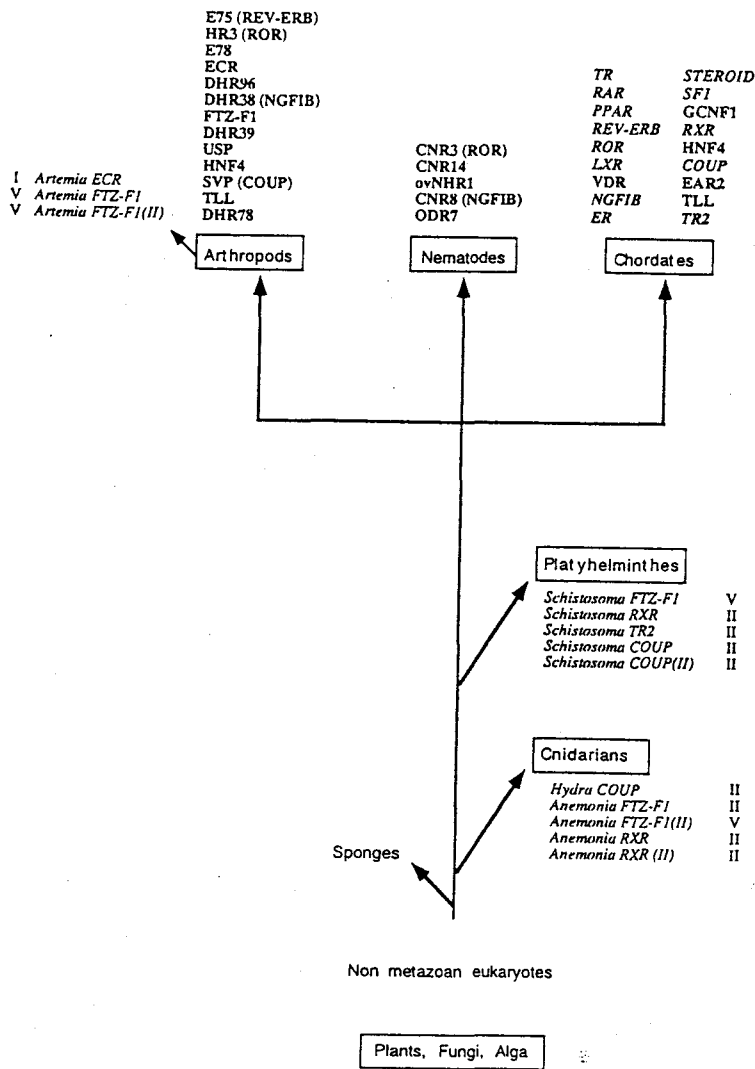
## RESULTS AND DISCUSSION

Ten sets of primers were designed to amplify the second exon encoding the highly conserved DNA-binding domain of NRs outside the animal kingdom (in plants, algae, and fungi) as well as in sponges, cnidarians, acoelomate, and coelomate triploblasts, and in key organisms of the transition from chordates to vertebrates. By two successive rounds of PCR cycles using all possible combinations of nested primers and a "touch-down" PCR method (see *Materials and Methods*), we amplified 84 similar genomic DNA fragments from a wide range of metazoans. We report here the first, to our knowledge, NR clones in early metazoans, such as cnidarians and acoelomates. Sponge DNA yielded no PCR fragments harboring a clear NR signature. The question of the presence of NR in this phylum, the phylogenetic position of which within metazoans is unclear, thus remains open (10, 11). Interestingly, we did not obtain any positive amplification outside metazoans. Although we cannot formally exclude the possibility that our primers may have artifactually missed NR sequences in these organisms, our observation confirms the apparent absence of NR observed in yeast (12). Thus, NRs appear to be confined to metazoans.

The NR gene fragments were cloned and sequenced (Fig. 1). Because the PCR method is particularly prone to contamination artifacts, we performed DNA extraction, PCR experiments, and cloning/sequencing in three different laboratories. Furthermore, the authenticity of the sequenced PCR products was ascertained by the following criteria (Table 1): (i) the A/T content of the PCR products was in accordance with the A/T

For each PCR product the number of sequenced clones is indicated. Gaps in the sequence alignment are indicated by stars. Interestingly, in *Anemonia* and *Schistosoma* we found several PCR products assigned to a given NR group (e.g., *Schistosoma* COUP and COUP II). In these cases the Roman number in parentheses is an arbitrary number needed to discriminate different gene versions. This suggests that, at least in these organisms, there is more diversity in NR than previously expected. The precise position of these extra copies was not determined, and it is not known whether these are secondary duplications specific to a given species or rather new receptor genes. In all cases these duplications are independent of the vertebrate gene duplication discussed in Fig. 3. GenBank codes of the sequences are U93406 to U93489.

A



B

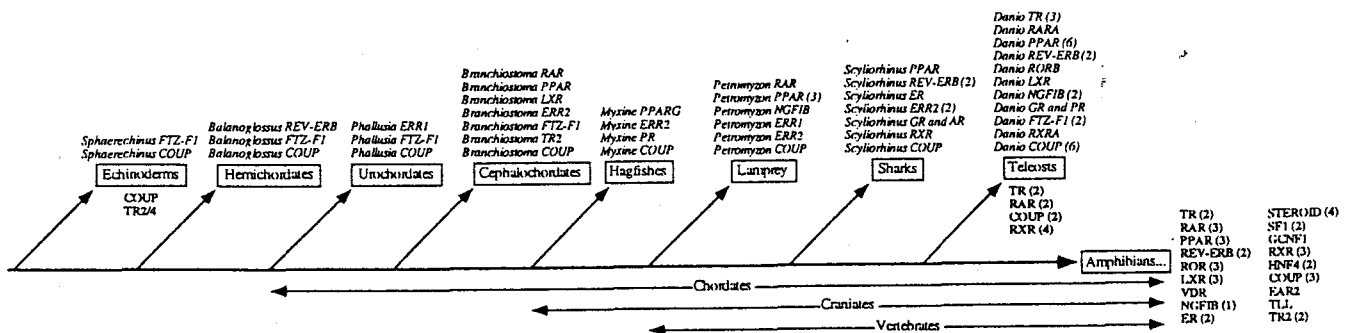


FIG. 3. Simplified phylogeny of metazoans (A) and chordates (B) showing the organisms where NR genes were found. In A, the subfamilies are indicated for each sequence. For B the number of different sequences found for a given group (e.g., PPAR) is indicated in parentheses. Zoological groups are boxed, previously described receptors from the literature are in bold, and receptors identified in this study are italicized.

content of the genomes in the case of *Hydra* and *Schistosoma* (13); (ii) Southern blots using the PCR products as probes were done when sufficient amounts of DNA were available and revealed specific bands under high-stringency conditions; (iii) cDNA and genomic libraries were screened using the PCR products and revealed clones containing a region 100% iden-

tical with the probe; (iv) several independent PCR amplifications in the same organisms produced identical sequences; and (v) the position of the fragments in phylogenetic trees was consistent with their origin (14). Taken together these criteria provide strong evidence that the amplified DNAs correspond to bona fide NR gene sequences in these various organisms.

The NR sequences were then compared with other NRs (Fig. 1). Interestingly, in cnidarians and *Schistosoma* we identified only members of COUP-TF, RXR, and FTZ-F1 groups of receptors. Despite an extensive search, no members of the other groups (such as RAR, TR, etc.) were found in these animals. Importantly, these primers have been successfully tested in a wide range of organisms that appeared more recently in evolution in which they give positive results. In addition, the NRs we found to be missing in cnidarians and *Schistosoma* are not evolving more rapidly in vertebrates than the sequences we effectively found. To position COUP-TF, RXR, and FTZ-F1 sequences, we constructed a phylogenetic tree connecting a representative selection of full-length NR sequences (Fig. 2). By observing the confidence "bootstrap" values of the most internal branches of the tree, the superfamily can be separated into six subfamilies and 26 groups of receptors. The relationships among the subfamilies remain unresolved. Nevertheless, the fact that the groups in which we found homologues in early metazoans only correspond to two subfamilies (II and V, Figs. 1 and 2) suggests that the origin of the superfamily can be positioned between these two subfamilies. Furthermore, the placement of the full-length sequence of the *Hydra* COUP cDNA that we recently isolated shows that *Hydra* COUP diverged before the split that led to COUP-TF and EAR2 and after the appearance of TLL and TR2 (Fig. 2). This indicates that a first wave of gene duplication took place before cnidarians diverged from bilaterians.

A second wave of gene duplications can be observed on the tree that led to the diversification of receptors inside each group, i.e., the appearance of the two to four paralogous copies of each receptor type (RARA, RARB, and RARG, etc.). Within the Deuterostomians, this wave of gene duplication is specific to vertebrates because in all cases we observed only one homologue in tunicates or amphioxus (Figs. 1 and 3B). This homologue (e.g., amphioxus COUP, RAR, or PPAR) is located as a nonduplicated version of the paralogous vertebrate receptors (data not shown). In contrast, in early vertebrates such as lamprey, shark, or zebrafish we found several paralogous genes. This means that orthologues of each gene (e.g., RARA, RARB, and RARG) in higher vertebrates are found in early vertebrates, hence the paralogous groups evolved before the latest common ancestor of living vertebrates (Fig. 1 and data not shown). In this sense, the four steroid receptors (PR, GR, MR, and AR) should be viewed as vertebrate-specific paralogous copies of a unique ancestral steroid receptor. Because NR genes are dispersed throughout the genome, our observation is clearly in favor of the Ohno model of vertebrate specific duplication of the entire genome (15, 16). A large number of new genes that were able to play a crucial role in the appearance of the vertebrate-specific developmental innovations were created by this event (15).

In contrast to orphan receptors (such as REV-ERBA, NGFIB, or SF1), vertebrate receptors with ligands have, in most cases, no arthropod homologues (Fig. 2), suggesting that orphan receptors are ancient (i.e., primitive), whereas receptors to known ligands are modern (i.e., derived). Indeed, receptors to known ligands appear to be present only in coelomates (Figs. 2 and 3A). In this context, the case of RXR is particularly interesting, because it has a homologue, called USP, in arthropods. Interestingly, the function of this homologue appears to be different in terms of ligand binding. In fact, RXR binds a ligand (9-*cis*-retinoic acid) in vertebrates but does not do so outside vertebrates, as exemplified by the *Drosophila* USP gene product (17). It is thus conceivable that there was a gain of 9-*cis*-retinoic acid binding to RXR specifically in vertebrates. Thus, the RXR/USP homologues that we found in early metazoans (*Anemonia* and *Schistosoma*) should be orphan receptors. In accordance with the above hypothesis of

acquisition of ligand binding function, RAR and RXR, which both bind retinoids, are located on different subfamilies in our phylogenetic tree. Strikingly, we noticed that, in general, there is no relationship between the position of a liganded NR in the tree and the chemical nature of its ligand. Thus, because the liganded receptors appear to be predominantly members of recent subfamilies of receptors, we propose that they have gained the ability to bind their ligands independently and that the ancestral NR was an orphan receptor. This model has important functional implications because it suggests that the conformational change that governs the activity of the receptor can be achieved in the absence of a ligand binding. Recently, it has been proposed that NRs do not exist in static off or on conformations but that ligand alters an equilibrium between the inactive and active states (18). Acquisition of ligand binding during the course of evolution could be an easy way to lock the conformation of the receptor into an active state. Indeed, the recent demonstration that the activation domain AF-2 of NRs such as RXR or PR may be active (i.e., providing binding surfaces for coactivators) in the absence of ligand (e.g., by phosphorylation) is in perfect accordance with our model (19–21).

Acquisition of ligand binding function within a transcription factor superfamily is not unique to NRs. For example, a member of the basic helix-loop-helix family has gained the ability to bind dioxin (22). Similarly, an independent gain of ligand binding ability also was suggested for the G-protein-coupled bioamine receptors in which divergent groups of receptors bind related ligands (exactly like RAR and RXR in NRs) whereas highly related groups bind structurally different molecules (as do TR and RAR in NRs) (23). Future structural and functional studies addressing the three-dimensional organization, pattern of expression, ligand-dependent structural and functional alterations, and dimerization abilities of early-metazoan NR homologues will cast a new light on how the multiple functions present in these molecules evolved and will allow the definition of the role of these genes in the evolutionary process itself.

We thank Laurent Meijer for access to laboratory facilities; Blandine Baratte, Murielle Fikry-Vennin, and Agnès Bègue for technical help; Brigitte Galliot, Jean-Guy Catford, François Fontaine, Robert Garrone, and Guillaume Lecointre for gifts of samples; and Hinrich Gronemeyer, François Catzeflis, Philippe Berta, Ricardo Mendonça, Jean Coll, and Jean Marc Vanacker for critical reading of the manuscript. We thank the anonymous reviewers for their comments. H.E. received support from the European Community (Human Capital and Mobility Program) and from Fondation Les Treilles. M.C.L. was funded by Association pour la Recherche Contre le Cancer. This work was supported by grants from Groupement de Recherche et D'étude Desgenomes, Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, and Institut Pasteur de Lille.

- Gronemeyer, H. & Laudet, V. (1995) *Protein Profile* 2, 1173–1308.
- Mangelsdorf, D. J. & Evans, R. M. (1995) *Cell* 83, 841–850.
- Laudet, V., Hänni, C., Coll, J., Catzeflis, F. & Stéhelin, D. (1992) *EMBO J.* 11, 1003–1013.
- Amero, S. A., Kretsinger, R. H., Moncrief, N. D., Yamamoto, K. R. & Pearson, W. R. (1992) *Mol. Endocrinol.* 6, 3–7.
- Detera-Wadleigh, S. D. & Fanning, T. G. (1995) *Mol. Phylogenet. Evol.* 3, 192–205.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, New York), 2nd Ed.
- Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991) *Nucleic Acids Res.* 19, 4008.
- Philippe, H. (1993) *Nucleic Acids Res.* 21, 5264–5272.
- Swofford, D. L. & Olsen, G. J. (1990) in *Molecular Systematics*, eds. Hillis, D. M. & Moritz, D. (Sinauer, Sunderland, MA), pp. 411–501.

10. Christen, R., Ratto, A., Baroin, A., Perasso, R., Grell, K. G. & Adoutte, A. (1991) *EMBO J.* **10**, 499–503.
11. Cavalier-Smith, T., Allsopp, M. T. E. P., Chao, E. E., Boury-Esnault, N. & Vacelet, J. (1996) *Can. J. Zool.* **74**, 2031–2045.
12. Dujon, B. (1996) *Trends Genet.* **12**, 263–267.
13. Musto, H., Alvarez, F., Tort, J. & Rodriguez-Maseda, H. (1994) *Int. J. Parasitol.* **24**, 277.
14. Philippe, H., Chenuil, A. & Adoutte, A. (1994) *Development (Cambridge, U.K.)*, Suppl., 15–25.
15. Holland, P. W. H., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. (1994) *Development (Cambridge, U.K.)*, Suppl., 125–133.
16. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, Heidelberg).
17. Mangelsdorf, D. J., Borgmeyer, U., Heyman, R. A., Zhou, J. Y., Ong, E. S., Oro, A. E., Kakizuka, A. & Evans, R. M. (1992) *Genes Dev.* **6**, 329–341.
18. Schulman, I. G., Juguilon, H. & Evans, R. M. (1996) *Mol. Cell. Biol.* **16**, 3807–3813.
19. Matkovits, T. & Christakos, S. (1995) *Mol. Endocrinol.* **9**, 232–242.
20. Perlmann, T. & Vennström, B. (1995) *Nature (London)* **377**, 387–388.
21. Power, R. F., Mani, S. K., Codina, J., Connelly, O. M. & O'Malley, B. W. (1991) *Science* **254**, 1636–1639.
22. Burbach, K. M., Poland, A. & Bradfield, C. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8185–8190.
23. Vernier, P., Cardinaud, B., Valdenaire, O., Philippe, H. & Vincent, J. D. (1995) *Trends Pharmacol. Sci.* **16**, 375–376.



#### Discussion de l'article 4

L'un des premiers enseignements tiré de cette étude est que nous n'avons pas obtenu d'amplification en dehors des métazoaires alors que nous avons trouvé des fragments de récepteurs nucléaires dans tous les métazoaires testés. Le résultat obtenu avec les éponges reste incertain. Nous n'avons en effet pas obtenu de fragments d'ADN caractéristiques des gènes de récepteurs nucléaires mais ce résultat reste à confirmer. Ces résultats suggèrent que les récepteurs nucléaires sont bien spécifiques des métazoaires. Nous ne pouvons toutefois pas exclure que nos amorces aient été incapables d'amplifier des séquences de récepteurs nucléaires dans les organismes non métazoaires. Cependant nos observations sont confirmées par l'absence de récepteurs nucléaires chez la levure *Saccharomyces cerevisiae* dont le génome a été entièrement séquencé (Dujon, 1996). Ainsi les récepteurs nucléaires se limiteraient aux métazoaires. De plus, nous avons isolé la séquence complète d'un ADNc correspondant au gène COUP-TF d'hydre (voir article 5) et lorsque nous avons replacé cette séquence dans un arbre évolutif nous avons constaté qu'elle se plaçait avant la divergence entre les groupes COUP-TF et EAR2 ce qui suggère bien que lorsque les diploblastes se sont séparés des autres métazoaires les récepteurs nucléaires étaient en pleine diversification.

Il est intéressant de constater que chez les cnidaires et le schistosome nous avons obtenus seulement des récepteurs des groupes COUP-TF, RXR et FTZ-F1. Malgré un effort pour rechercher d'autres types de récepteurs, aucune séquence appartenant à un autre groupe, comme RAR ou TR, n'a été trouvée. Il est important de noter que nos amorces ont été utilisées pour une grande variété d'organismes qui sont d'origine plus récente sur le plan évolutif et qui ont donné une réponse positive. Ce résultat négatif chez les métazoaires précoces ne peut donc pas s'expliquer par un artefact expérimental. Les groupes pour lesquels nous avons trouvé des homologues chez les métazoaires précoces se regroupent en deux sous-familles de récepteurs nucléaires seulement : les sous familles II (RXR, COUP-TF) et V (FTZ-F1).

L'étude de l'arbre évolutif nous permet de faire une autre observation intéressante. En effet, on peut noter qu'au contraire des récepteurs orphelins (REV-ERBA, NGFIB ou SF1) les récepteurs de vertébrés qui sont reconnus par des ligands

n'ont pas d'homologues chez les arthropodes. Ceci permet de suggérer que les récepteurs orphelins sont anciens (primitifs), alors que les récepteurs avec ligands sont modernes (dérivés). De plus, de manière marquante, nous avons constaté qu'il n'existe pas de relations entre la position d'un récepteur avec ligand sur l'arbre et la nature chimique de ce ligand. Ainsi, RAR et TR qui sont étroitement apparentés fixent des ligands totalement différents. A l'opposé, les RAR et les RXR qui reconnaissent tous des rétinoïdes sont situés dans des sous familles différentes.

L'ensemble de ces éléments nous a amenés à proposer l'hypothèse que les premiers récepteurs nucléaires étaient des récepteurs orphelins qui ont acquis secondairement la capacité à fixer un ligand au cours de l'évolution.

*Article 5*

## Article 5

### A unique homologue for COUP-TF and EAR2 orphan nuclear receptors in Hydra.

Marie-Claire Langlois, Dominique Gauchat, Hector Escriva, Vincent Laudet et Brigitte Galliot

En préparation pour "Genes, Development and Evolution"

Après avoir démontré que les récepteurs nucléaires étaient présents chez tous les métazoaires, il nous a semblé important de caractériser certains d'entre eux chez les premiers métazoaires. Ce travail permettrait en effet d'obtenir des informations sur la mise en place des fonctions de ces gènes au cours de l'évolution. De plus, une telle étude des récepteurs nucléaires chez les métazoaires précoces pourra nous permettre d'amener des informations permettant de tester le modèle de gain de capacité à fixer des ligands au cours de l'évolution.

Au cours d'une première approche, nous avons isolé un ADNc codant un récepteur nucléaire appartenant au groupe COUP-TF chez l'hydre d'eau douce *Hydra vulgaris*. Les membres du groupe COUP-TF sont impliqués, chez la drosophile et chez de nombreux vertébrés (souris, poulet, xénope et poisson-zèbre), dans la différenciation des neurones. L'hydre étant le premier métazoaire à présenter un réseau de cellules nerveuses, il nous a semblé intéressant de caractériser un homologue de COUP-TF chez cette espèce. Nous avons isolé un clone d'ADNc incomplet codant une partie du domaine C et le domaine D du gène COUP-TF d'hydre. En criblant des banques d'ADNc d'hydre à l'aide de ce clone, nous avons isolé 3 types différents d'ADNc divergents dans leur extrémité 5'. Ces isoformes montrent qu'un épissage alternatif complexe se produit en amont du domaine D de COUP-TF d'hydre. Aucun clone complet portant à la fois le domaine C et les domaines D et E n'a pu être isolé. Il est intéressant de noter que COUP-TF d'hydre semble posséder un intron dans son domaine C, intron qui est absent chez les COUP-TF de vertébrés.

L'hydre est un organisme en constant renouvellement. Cet animal possède un axe antéro-postérieur avec une tête et un pied. Au cours de la vie de l'animal les cellules migrent le long de l'axe antéro-postérieur, soit vers la tête, soit vers la queue et se différencient en migrant. Le patron d'expression d'un gène donné et

notamment la répartition de son expression le long de cet axe permet donc d'avoir une idée sur le degré de différenciation des cellules qui l'expriment. En collaboration avec Brigitte Galliot de l'université de Genève, nous avons étudié l'expression du gène COUP-TF d'hydre par hybridation *in situ in toto*. Les résultats obtenus montrent une expression dans le milieu du polype mais pas à ses extrémités ce qui suggère une expression de COUP dans des cellules relativement indifférenciées. L'étude de l'expression dans des cellules isolées après dilacération des polypes suggère que ces cellules pourraient être des précurseurs pluripotents, notamment impliqués dans la voie de différenciation neuronale.

**THE HYDRA COUP-TF, A UNIQUE HOMOLOGUE FOR COUP-TF AND EAR2 ORPHAN NUCLEAR RECEPTORS, IS A VERY EARLY MARKER FOR MORPHOGENETIC PROCESSES IN HYDRA**

Marie-Claire LANGLOIS<sup>1</sup>, Dominique GAUCHAT<sup>3</sup>, Hector ESCRIVA<sup>2</sup>, Agnès BEGUE<sup>1</sup>, Vincent LAUDET<sup>1, 2\*</sup> and Brigitte GALLIOT<sup>3</sup>

<sup>1</sup> UMR 319 du CNRS, Institut de Biologie de Lille, Institut Pasteur, 1 rue Calmette, 59019 Lille Cedex, France.

<sup>2</sup> UMR 49 du CNRS, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France.

<sup>3</sup> Department of Zoology and Animal Biology, University of Geneva, Sciences III, 30 Quai Ernest Ansermet, CH-1211 Geneve 4, Switzerland.

Keywords : Hydra, Cnidaria, Nuclear hormone receptors, COUP-TF, EAR2, Transcription factors, Orphan Receptors

Running title : Hydra COUP-TF

## ABSTRACT

The COUP-TF orphan receptors are implicated in the neuronal differentiation in a wide variety of organisms. Since hydra is an early metazoan belonging to the cnidarians, one of the first metazoan phyla possessing differentiated neuronal cells, we characterized the COUP-TF homologue in this organism. We have isolated several cDNAs types corresponding to a unique COUP-TF homologue which show that the COUP-TF gene product undergoes extensive alternative splicing events, in sharp contrast to its vertebrate homologues. The longest cDNA is still incomplete in its 5' region since it lacks the first 19 amino acids of the DNA binding domain and all the putative A/B region. Interestingly, we show by phylogenetical analysis that the Hydra COUP-TF gene derives from an ancestor common to the EAR2 and COUP-TFs groups of orphan receptors. This suggests that when cnidarians splitted from the other metazoan the first wave of nuclear receptor gene duplication was still ongoing. Expression studies revealed that the hydra COUP-TF gene is expressed along the body column as patches corresponding to groups of small interstitial cells which are the common precursor cells to nerve cells and nematocytes. During regeneration, COUP-TF transcripts are detected as early as 1 hour following bisection in endodermal cells of the stump. Similarly, COUP-TF is expressed as spot at the place where a new bud will emerge but preceding bud formation. These results suggest that COUP-TF is involved in early processes leading to morphogenesis.

## INTRODUCTION

Hydra belongs to the Cnidaria, the first metazoan phylum of the animal kingdom where the presence of highly specialized cell types such as muscle cells and neurons was demonstrated. For this reason hydra represents an excellent model to study the emergence of genes implicated in the differentiation of nerve cells. Moreover, hydra is also an excellent model system for developmental biology, since the animal displays an oral aboral axis and pattern formation processes can be easily studied in regeneration, transplantation and reaggregation experiments.

Some developmental genes have already been shown to be conserved both at the structural and expression levels in cnidarians. Several Hox-related genes have been isolated in hydra (Schummer *et al.*, 1992; Miller and Miles, 1992; Naito *et al.*, 1993; Aerne *et al.*, 1995; Kuhn *et al.*, 1996). Interestingly, Hox genes have been also shown to be differentially expressed during hydra regeneration, both at temporal and spatial levels: For instance, the hydra *cnox-1* gene, more related to the anterior

*Drosophila labial* gene is expressed earlier and more anteriorly than *cnox-2*, mostly similar to more posterior Hox genes (Schummer *et al.*, 1992 ; Shenk *et al.*, 1993a; BG and coll., unpublished). Thus, the function of Hox genes in the determination of the antero-posterior axis might be partially conserved between early metazoans such as hydra and triploblastic metazoans.

More recently, homologues of regulatory genes involved in the vertebrate head development, have been characterized in cnidarians. Noteworthy, these genes also display head-specific expression in cnidarians suggesting an extensive conservation of their developmental functions. *Budhead* , a hydra transcription factor related to the winged-helix transcription factor HNF-3 is expressed during head regeneration and budding (Martinez *et al.*, 1997). In the gastrozoid polyp *Hydractinia symbiolongicarpus*, *Cn-ems* , the homologue of the mouse homeobox *Emx* gene, is expressed in the most apical part of the animal, the hypostome, a tissue dome localized between the mouth opening and the tentacle ring (Mokady *et al.*, 1998). In adult hydra, *prdl-a*, a paired-like homeobox gene shows an expression that is restricted to the nerve cell lineage of the head region (Gauchat *et al.*, 1998). Moreover, at the time head organizer activity is induced during budding or apical regeneration, *prdl-a* is transiently expressed in endodermal cells of the stump. The current thought defines a "minimal head region" in hydra which uses homologues of vertebrate genes to induce its organization (Gauchat *et al.*, 1998).

However, little is known about the genes implicated in the differentiation of the nerve cell lineage in cnidarians. In hydra, the homolog of *Achaete-scute* , a gene involved in the induction of sensorial organ formation in *Drosophila*, is not expressed in the nerve cell lineage (Grens *et al.*, 1995). However, the function of this gene appears well conserved since partial rescue was observed in an *Achaete-scute* *Drosophila* mutant upon hydra's *achaete-scute* ectopic expression (Grens *et al.*, 1995).

The emergence of the nervous system is an important hallmark of metazoan organization. However, the underlying genetic mechanisms responsible for the appearance of the neuronal fact during evolution are still poorly understood. Characterizing *Hydra* version of genes known to be important for neuronal differentiation in other metazoan species is thus of utmost interest.

COUP-TFs are members of the nuclear hormone receptor superfamily (reviewed in Tsai and Tsai, 1997, Gronemeyer and Laudet, 1995). These proteins are ligand-dependent transcription factors. As all nuclear receptors (NRs), COUP-TFs genes underwent two waves of gene duplication during metazoan evolution : one



very early corresponding to the appearance of the COUP-TF group itself and a second one later on, specific of the vertebrate lineage which corresponds to a diversification inside the COUP-TF group with the appearance of the COUP-TFI and COUP-TFII paralogous genes (Laudet *et al.*, 1992 ; Laudet, 1997, M.C.L. and V.L. in preparation). All COUP-TFs from *Drosophila* to vertebrates share an extensive degree of sequence identity: 95-100 % in the DNA binding domain (also called C domain) and 89-96% in their ligand binding domain (or E domain). More recently, in a PCR screen of NR family members in metazoans, we had identified COUP-TF homologues in a wide range of metazoans from hydra to several protostomian and deuterostomian animals (Escriva *et al.*, 1997). Together with other published results, these observations demonstrated that the COUP-TF genes are the most conserved members of the NR superfamily, suggesting that they support essential functions among all metazoans (Laudet, 1997).

Although much is known about the *in vitro* transcriptional effects and the target genes of COUP-TF genes products, especially in vertebrates, there is still no cognate ligand identified. COUP-TFs are therefore still considered as "orphan" members of the NR superfamily, as is EAR2, its most closely related receptor.

In vertebrates, the COUP-TFs related genes are differentially expressed in the developing cerebral vesicles (Jonk *et al.*, 1994; van der Wees *et al.*, 1996; Schuhand Kimelman, 1995; Fjose *et al.*, 1993). In mouse, COUP-TFI contributes to the segmentation of the diencephalon (Qiu *et al.*, 1994) and participates to the regulation of contact formation between neurites (Connor *et al.*, 1995). COUP-TFII is expressed in spinal motor neurons and can be induced ectopically in the alar plate by adjacent notocord grafts (Lutz *et al.*, 1994). In addition, COUP-TFs mRNAs are also found in adult mouse brain (Lopes *et al.*, 1995). Null mutation of mCOUP-TFI induces defects in glossopharyngeal ganglion and abnormal axonal projections in several regions of mutant mouse embryos are observed (Qiu *et al.*, 1997). The counterpart of COUP-TF genes in *Drosophila*, Seven-up (*svp*) is implicated in photoreceptor determination. Altogether these data show that COUP-TF related genes are implicated both in developmental nervous processes during the embryonic life and also have a role in the nervous system of adult organisms.

In this paper, we describe the characterization of a COUP-TF homologue in hydra. In contrast to its vertebrate homologue, the products of this gene are alternatively spliced. Interestingly, the hydra COUP-TF gene is located on a phylogenetical tree before the split leading to the COUP-TF group and to the EAR2 orphan receptor. This suggests that at the time cnidarians diverged from other

metazoans, the NR family had not yet diversified. Finally we demonstrate that the hydra COUP-TF related gene is expressed in adult body, in cells which are the precursor cells for the nerve cell lineage and during pattern formation, at a place, endodermal cells of the stump, and a time where organizer activity was mapped. These data suggest that COUP-TF is involved in morphogenetic activity in hydra.

## **MATERIAL AND METHODS**

### **Isolation of Hydra COUP-TF cDNAs**

Using a 117bp PCR fragment corresponding to *Hydra vulgaris* COUP-TF (Escriva *et al.* 1997) as a probe, we screened a cDNA library constructed in lambda gt11 vectors with *Hydra vulgaris* cDNA. A single clone (A) was isolated out of  $10^6$  phages. The region of clone A encompassing the C and D domains was used to screen another cDNA library constructed with adult Hydra polyps RNAs (Bode *et al.*, 1994). Hybridization were performed over night at 37°C in 40% formamide, 7X sodium citrate (SSC), 5X Denhardt's solution, 0.5 % SSC, 0.1 mg/ml denatured salmon sperm DNA. Membranes were washed twice in 2XSSC, then 0.1XSSC at 37°C for 15 min.

The inserts were subcloned in BlueScript vectors and sequenced on both strands, using an Applied Biosystems 373A automatic sequencer and methods and reagents recommended by the supplier.

### **Sequence analysis**

The sequences were analyzed using the programs available at the Infobiogen network ([www://infobiogen.fr](http://www://infobiogen.fr)). Phylogenetical analysis were performed using the Must package (Philippe, 1993) with method previously described (Laudet, 1997).

### ***In situ* hybridization**

Whole-mount *in situ* hybridization and *in situ* hybridization performed on macerated animals, were carried out according to Gauchat *et al.*, 1998.

## RESULTS AND DISCUSSION

### Isolation of Hydra COUP-TF cDNAs.

We screened a *Hydra vulgaris* (adult polyps) cDNA library (Schummer *et al.*, 1992) with a probe corresponding to a 117nt COUP-TF fragment obtained during a PCR screen for nuclear hormone receptors in metazoans (Escriva *et al.*, 1997). A single clone (clone A, Figs. 1 and 2A) harboring a region 100% sequence identical to the probe was isolated. In order to obtain a full length cDNA, a part of clone A (see below) was used to screened the same library again as well as a second one (from Bode *et al.*, 1994). 10 distinct new clones were isolated (Fig.1), all encoding ORFs showing strong similarity with the human COUP-TF sequence (Fig. 4).

Among the different clones isolated, only clone A contained a portion of the putative DNA binding domain (DBD). However, this DBD sequence was not complete since it lacked the 19 N-terminal amino acids when compared to the human COUP-TF including the three first cysteines of the first zinc finger. Homology to NRs in clone A actually started inside the P-box, a highly conserved region (CEGCKS), which mediates the direct interaction of the receptor to the major groove of the DNA double helix. Instead, a 686 bp DNA stretch containing a stop codon at position 663 (Fig. 2A) and bearing no sequence identity to any sequence was present upstream of the putative DBD part. The first available initiation codon was located at the end of the C domain and is conserved in all NRs known to date. The reading frame of this clone is thus only open in a region corresponding to the D domain (Fig. 2A). The position from which the sequence diverged inside the C domain (inside the P-box) may correspond to an intron and clone A may therefore be viewed as resulting from either an alternative splice event or an uncompletely processed primary transcript in which an intron remained. Supporting the latter view the nucleotides located 5' to partial DBD (AG and GT, positions 680-683) are reminiscent of a splice site consensus. Since no intron in the P-box of nuclear hormone receptors has ever been observed, this would represent a particularity of the hydra COUP-TF (Figure 3). Alternatively, clone A could be also viewed as a chimera generated during the construction of the library.

All various strategies (inverted PCR, genomic library screening, PCR with conserved oligonucleotides, 5'RACE) designed to isolate the 5' end of the COUP-TF gene encoding the missing part of the DBD remained unsuccessful. Furthermore, we did not obtain any clone containing both the DBD and the ligand binding domain (LBD). Clone A deleted from the divergent 5'UTR was used to obtain the 10

other clones. All these cDNAs contained only the D domain in common with clone A. Clones 6 and 1/27 appeared full-size in the 3' part since they encompassed a complete E domain linked to an F domain, and harbored a stop codon (position 1228 of clone 6, Fig. 2B) and a 356 bp long 3'UTR sequence (clone 6). Surprisingly, the clones obtained during the second screening exhibited variable 5' sequences, diverging at the same point (position 831 in clone A). Three different clone types, bearing different 5' sequences, were identified (Fig. 1). Type I contained the longest 5' sequence and an ATG located 13 codons before the diverging point with clone A. The reading frame upstream of this ATG is closed. Type II and III did not exhibit any open reading frame. The first ATG that can be used to initiate translation from these clones lied in the D domain (position 876 of clone A or 64 of clone 6, Fig. 2). The precise boundary between these divergent 5' sequences and the common D and E domains is not strict. For each clone type we found two possibilities (Fig. 1B) : the boundary could be at the level of the GTT codon (position 831 of clone A and 22 of clone 6, Fig. 2) or 3 bp upstream (GCA codon; position 828 of clone A and 19 of clone 6). Interestingly, this region corresponds to a conserved splice site in the COUP-TF genes of vertebrates (Ritchie *et al.*, 1990) and is in good agreement with splice site consensus (Mount *et al.*, 1982). So far, there are two examples of perfect conservation of intron position within regulatory genes from hydra to vertebrates : both were noted within regions encoding DNA-binding domains, one within the CREB gene (Galliot *et al.*, 1995), the second within the *prdl-a* homeobox gene (Gauchat *et al.*, 1998). It is thus tempting to interpret these clones as resulting from alternative splicing or alternate promoter usage that generated several isoforms. Nevertheless, such a situation is unusual in that these isoforms will be devoid of DNA binding domain and will only encompass the D, E and F domains.

The complexity of the hydra COUP-TF gene organization is in sharp contrast to the simplicity of that observed in its vertebrate homologues. Indeed, in all vertebrates, COUP-TFI and II genes are splitted in only three exons: one encompassing the A/B and C domains, one covering the D domain and the N-terminal part of the E domain, and the last one encoding the remaining of the E domain. Interestingly the two introns present in the vertebrate COUP-TF genes are the two most positionally conserved accross the whole NR superfamily. The *Drosophila* homologue of COUP-TFs, SVP, is transcribed as two isoforms differing in their C-terminal part (Mlodzik *et al.*, 1990). In contrast, the hydra COUP-TF exhibits a complex structure with at least three possible sequences upstream of the exon coding for the D and E domains and, although this remains to be firmly

established, an intron inside the DBD which is an unusual location. Furthermore, the relationships between the various domains, and particularly C and DE, remain unclear since we have not found a cDNA harboring both parts. It is nevertheless highly probable that such a transcript should exist in order to generate a classical DNA binding COUP-TF protein. The function of COUP-TF isoforms containing only the D and E domains remains elusive. It can be hypothesized that such a protein could either regulate by heterodimerization the action of a « classical » DNA binding isoform or even the action of another NR present in hydra. Interestingly, two orphan nuclear receptors devoid of C domain Dax-1 and SHP have already been isolated. Both of them downregulate the action of other members of the nuclear hormone superfamily such as ER or SF-1 (Zanaria *et al.*, 1995, Seol *et al.*, 1998). For example, it has been shown recently that DAX-1 interacts with SF-1 and recruits the nuclear corepressor N-coR, maintaining SF-1 in a repressing state (Crawford *et al.*, 1998). In this respect it is important to note that, in our screen for NR genes in metazoans (Escriva *et al.*, 1997) we found four other types of NR fragments (two RXRs and two FTZ-F1) in *Anemonia sulcata*, another cnidarian. Since COUP-TF and RXR have been shown to interact together (although this remains controversial; reviewed in Tsai and Tsai, 1997) it is conceivable that the hydra COUP-TF isoform containing only the D and E domains interacts with a hydra homologue to RXR. Finally it is also possible that, since the E domain of COUP-TF interacts with transcriptional co-repressors such as N-CoR, the D and E domains of hydra COUP-TF may regulate gene transcription even in the absence of DNA binding simply by titrating out a hydra N-CoR like molecule (Shibata *et al.*, 1997).

#### **Comparison of COUP-TF functional domains in *Hydra* and vertebrates**

Figure 4 shows a comparison of the C, D and E domains of the hydra COUP-TF with the other known COUP-TF factors. Overall, the hydra COUP-TF sequence bears 47 % amino acid identity with the human COUP-TFI, 45.7 % with human COUP-TFII, 42.9 % with human EAR2 orphan receptor and 44.5% with the *Drosophila* SVP protein. By comparison, we observed that this level is less than 30 % with human RXR for example. This clearly shows that we have indeed isolated a member of the COUP-TF group in hydra. The C domain of the hydra COUP-TF is the most conserved one when compared to other COUP-TFs (80-83%). In particular, the part of the P-box present in hydra COUP-TF as well as the end of the helix which contacts the DNA (positions 19-27 in Fig. 4) are strongly conserved between hydra

and other species. Furthermore, Lys 7 and Arg 12 are in similar position to the Rev-erba Lys 22 and Arg 27 (respectively), implicated by 3D structure analysis in half-site recognition by this orphan receptor (Zhao *et al.*, 1998). This also true for residues 31, 32 and 35, identical to Rev-erba Arg 50, Asn 51 and Gln 54, involved in contacts with the DNA phosphate skeleton. A very strong selective pressure thus maintains, in this critical region, residues that are critical to the DNA binding function. These observations strongly argue in favor of the fact that a regular COUP-TF protein: i.e. containing a complete C domain and being able to bind to DNA should exist in hydra.

The high identity between the hydra COUP-TF E domain and the human COUP-TFI E domain is scattered over nine regions that roughly correspond to the  $\alpha$ -helices determined by the 3D structure resolution of the RAR- and RXR LBD (Bourguet *et al.*, Renaud *et al.*). This suggests that hydra COUP-TF, as its homologues in other metazoans, will fold in a common tertiary structure (Wurtz *et al.*) Since all the known COUP-TF are mainly transcriptional repressors and since the domain responsible for the repressing activity is the E domain (Achatz *et al.* 1997) it is tempting to speculate that the hydra COUP-TF will also be a transcriptional repressor. This remains to be experimentally addressed.

The last remarkable feature of hydra COUP-TF is the presence of a long F domain, C-terminal to the E domain. This domain does not harbour sequence identity to any other proteins. F domains are not uncommon in NRs although their functions often remain elusive (reviewed in Gronemeyer and Laudet, 1995).

Phylogenetic analyses done on the C domain only, or on the C and E domains together, confirm that hydra COUP-TF is a member of the COUP-TF group (Fig; 5). Since hydra COUP-TF is significantly more closely related to COUP-TF and EAR2 than with other receptors such as RXR or HNF4 the tree was rooted with a selection of divergent receptors belonging to different nuclear receptor sub-families (Laudet, 1997) Interestingly, on such an analysis, the hydra COUP-TF sequence appears as a divergent member of this group, consistent with the fact that cnidarians diverged early on during metazoan evolution. Indeed, the hydra COUP-TF splitted before the dichotomy that led to the COUP-TF group itself and to the EAR2 orphan receptor. This suggests that hydra COUP-TF could be viewed as a homologue of the common ancestor of COUP-TFs and EAR2 orphan receptors.

Our previous phylogenetical analysis (Laudet, 1997) has shown that the diversification of the NR superfamily occurred in two waves of gene duplication : an early one, corresponding to the emergence of the various groups of receptors and

the second one later, specific to the vertebrate lineage and giving rise to the various paralogous copies of receptors such as COUP-TFI and COUP-TFII. The precise datation of the first wave of gene duplication was difficult due to the fact that NR sequences were known only in few organisms such as *Drosophila* and vertebrates. Our PCR screen for homologues of NR genes in and outside metazoans drove us to conclude that the NR superfamily is specific to metazoans, suggesting that the first wave of gene duplication was active at the very beginning of metazoan evolution (Escriva *et al.*, 1997). This is consistent with the fact that only members of the NR superfamily belonging to two out of the six described sub-families could be found in early metazoans. The present result confirms this view. Indeed, when cnidarians diverged from the other metazoans the first wave of gene duplication was not completed since the COUP-TF/EAR2 dichotomy was still not done.

#### **Expression of the hydra COUP-TF gene.**

In order to determine the expression pattern of Hydra COUP-TF, we performed in situ hybridizations on whole animal mount with a probe containing the C and D domains (common to the clone A described in Figs 1 and 2A but without the divergent 5'UTR).

Hydra is composed of two cell layers, ectodermal and endodermal, which are separated by an extracellular matrix, the mesoglea. Hydra cells fall into two lineages, the myoepithelial lineage whose derivatives are the most abundant, and the interstitial cell lineage which provide precursor cells for the nerve cells, the gland cells, the nematocytes and the gametes when the animals follow the sexual cycle. Within the body column, hydra polyps maintain a continual renewal of their stock of stem cells, either epithelial or interstitial, while determination and differentiation of new mature head or foot-specific cells will occur concomitantly with their migration towards the extremities of the animals. Simultaneously, mature cells are permanently sloughed off at the tip of the tentacles or at the foot basis, preserving thus a balance in the animal cell number. Thus, for a given gene, the level of expression along the body column within a given cell type provides information about the differentiation stage of the cell where this gene is expressed.

In adult polyps, the hydra COUP-TF gene is strongly expressed in groups of ectodermal cells located in the body column while a diffuse and moderate staining is observed in endodermal cells of the head region (Fig. 6A). This pattern was observed in budless polyps as well as in late-stage buds. In order to determine the cell type where COUP-TF is expressed, in situ hybridizations were carried out on

isolated cells obtained after dissociation of polyps. Indeed, hydra can be incubated with pronase in order to obtain a cell suspension that can be labelled by in situ hybridization. Because the morphology of the various cell types is recognizable (Holstein, 1995), the cell types expressing a given gene may be identified. COUP-TF expression was observed in isolated small rounded-shape cells identified as the small cells of the interstitial lineage (Fig. 6B). These cells correspond to the precursors of either the nerve cells or the nematoblasts. In hydra, precursor cells to nerve cells are blocked in G2 before, upon appropriate signaling, undergoing a final mitosis which will provide two newly mature nerve cells (see Schaller *et al.*, 1989). In contrast, precursors to nematocytes will undergo several runs of synchronous divisions before terminally differentiated nematocytes reach the tentacles. A similar "patchy" ectodermal pattern was observed for the paired-like *prdl-b* homeobox gene and found to correspond to synchronously dividing nematoblasts (Gauchat *et al.*, 1998). Thus, the size, the shape and the location in the body column of the COUP-TF expressing cells suggest they belong to the sub-set of the interstitial cell lineage which will give rise to mature nematocytes. The moderate expression observed in endodermal cells of the head region might correspond to a different subset of the interstitial lineage, that one which will give rise to gland cells. (This result will require further confirmation).

During budding and regeneration, COUP-TF expression is dramatically modified and interestingly at very early stages. During budding, a row of endodermal COUP-TF expressing cells is observed, before any bud is observed and for this reason named bud spot (Fig. 6B). During regeneration, COUP-TF is expressed very early, as soon as one hour after mid-gastric section, again in endodermal cells of both head- and foot-regenerating stumps. In the head-regenerating stump, this expression is still detectable after 18 hours but hardly visible after 24 hours, about 12 hours before tentacle buds will appear. Transplantation experiments have demonstrated that head-organizer activity could be mapped within the regenerating stump as soon as 3 to 4 hours after mid-gastric section reaching a plateau level after 9 hours (MacWilliams, 1984). A similar early endodermal expression restricted to the regenerating stump was observed for both CREB gene (Kaloulis and BG, unpublished) and the paired-like homeobox gene, *prdl-a* (Gauchat *et al.*, 1998) about 2 and 4 hours after bisection respectively. Thus, the hydra COUP-TF gene is one of the earliest marker expressed during regeneration. Expression of these regulatory genes at the place where, and at the time when head-organizer activity was mapped, strongly suggest that COUP-TF,



CREB and *prdl-a* play a key role in the establishment of the organizer activity. Another main conclusion drawn from these expression studies relies on the fact that the primary organizer activity is likely located in the endoderm and not in the ectoderm. Finally, one should note that at the cellular level, within first hours of regeneration at the time these regulatory genes are expressed, proliferation of all cell types is repressed during at least 24 hours except the pool of nerve precursor cells which will terminally differentiate after a final mitosis (see for review Schaller *et al.*, 1989; Galliot 1997). It is thus tempting to speculate that the products of these genes are involved in repression of proliferation and / or terminal differentiation of new nerve cells.

### **Conclusion**

Despite the simplicity of its cell organisation and body plan, hydra expresses a range of highly conserved regulatory genes involved in developmental processes which provide a unique window whereby to trace back the ancestral functions of their vertebrate counterparts. The hydra COUP-TF gene, the first NR isolated in cnidarians, belongs to this category. A better understanding of the COUP function in hydra should help deciphering the ancestral role of COUP in cell differentiation and pattern formation.

### **ACKNOWLEDGEMENTS**

This work was supported by CNRS, IPL, ENS de Lyon, ARC, the Swiss National Foundation and the Canton de Genève. We thank Jean-Marc Vanacker for critical reading of the manuscript. M.C.L. was granted by ARC, Institut Pasteur de Lille and Région Nord Pas de Calais.

## REFERENCES

- Achatz, G., Holzl, B., Speckmayer, R., Hauser, C., Sandhofer, F., and Paulweber, B. (1997). Functional domains of the human orphan receptor ARP-1/COUP-TFII involved in active repression and transrepression. *Mol Cell Biol*, **17**, 4914-4932.
- Aerne, B. L., Baader, C. D. and Schmid, V. 1995. Life stage and tissue-specific expression of the homeobox gene *cnox1-Pc* of the hydrozoan *Podocoryne carnea*. *Dev Biol*, **169**, 547-566.
- Begemann, G., Paricio, N., Artero, R., Kiss, I., Perez, A. M., and Mlodzik, M. (1997). *muscleblind*, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development*, **124**, 4321-4331.
- Connor, H., Nornes, H., and Neuman, T. (1995). Expression screening reveals an orphan receptor chick ovalbumin upstream promoter transcription factor I as a regulator of neurite/substrate-cell contacts and cell aggregation. *J Biol Chem*, **270**, 15066-15070.
- Crawford, P.A., Dorn, C. Sadovsky, Y., Milbrandt, J. (1998). Nuclear receptor Dax-1 recruits nuclear corepressor N-CoR to steroidogenic factor 1. *Moll. Cell. Biol.* **18**, 2949-56.
- Escriva, H., Safi, R., Hanni, C., Langlois, M. C., Saumitou, L. P., Stehelin, D., Capron, A., Pierce, R., and Laudet, V. (1997). Ligand binding was acquired during evolution of nuclear receptors. *Proc Natl Acad Sci U S A*, **94**, 6803-6808.
- Fjose, A., Nornes, S., Weber, U., and Mlodzik, M. (1993). Functional conservation of vertebrate seven-up related genes in neurogenesis and eye development. *Embo J*, **12**, 1403-1414.
- Galliot, B. 1997. Signaling molecules in regenerating hydra. *BioEssays*, **19**, 37-46.
- Galliot, B., de Vargas C. and Miller D. 1998. Evolution of Homeobox Genes: Q50 Paired-like Genes Founded the Paired-Class. *Dev Genes Ev*, in press.
- Gauchat, D., Kreger, S., Holstein, T., and Galliot, B. (1998). *prdl-a*, a gene marker for hydra apical differentiation related to triploblastic paired-like head-specific genes. *Development*, **125**, 1637-1645.
- Grens, A., Mason, E., Marsh, J. L., and Bode, H. R. (1995). Evolutionary conservation of a cell fate specification gene: the Hydra *achaete-scute* homolog has proneural activity in *Drosophila*. *Development*, **121**, 4027-4035.
- Gronemeyer, H. and Laudet, V. (1995). Transcription factors 3: nuclear receptors. *Protein Profile*, **2**, 1173-1308.
- Hall, R. K., Sladek, F. M., and Granner, D. K. (1995). The orphan receptors COUP-TF and HNF-4 serve as accessory factors required for induction of phosphoenolpyruvate carboxykinase gene transcription by glucocorticoids. *Proc Natl Acad Sci U S A*, **92**, 412-416.
- Holstein, TW. 1995. *Zytologie. in Cnidaria: Hydrozoa, Kamptozoa. Süßwasserfauna von Mitteleuropa.* ed. J. Schwoerbel and P. Zwick, Gustav Fisher Verlag, Berlin.
- Huet, F., Ruiz, C., and Richards, G. (1995). Sequential gene activation by ecdysone in *Drosophila melanogaster*: the hierarchical equivalence of early and early late genes. *Development*, **121**, 1195-1204.

- Jonk, L. J., de, J. M., Pals, C. E., Wissink, S., Vervaart, J. M., Schoorlemmer, J., and Kruijer, W. (1994). Cloning and expression during development of three murine members of the COUP family of nuclear orphan receptors. *Mech Dev*, **47**, 81-97.
- Kuhn, K., Streit, B. and Schierwater, B. (1996). Homeobox genes in the cnidarian *Eleutheria dichotoma*: evolutionary implications for the origin of Antennapedia-class (HOM/Hox) genes. *Mol Phylogenet Evol* **6**, 30-38.
- Lopes, da Silva, S, Cox, J. J., Jonk, L. J., Kruijer, W., and Burbach, J. P. (1995). Localization of transcripts of the related nuclear orphan receptors COUP-TF I and ARP-1 in the adult mouse brain. *Brain Res Mol Brain Res*, **30**, 131-136.
- Lutz, B., Kuratani, S., Cooney, A. J., Wawersik, S., Tsai, S. Y., Eichele, G., and Tsai, M. J. (1994). Developmental regulation of the orphan receptor COUP-TF II gene in spinal motor neurons. *Development*, **120**, 25-36.
- MacWilliams, H.K. (1983). Hydra transplantation phenomena and the mechanism of Hydra head regeneration. II. Properties of the head activation. *Dev. Biol.* **96**, 239-257.
- Malik, S. and Karathanasis, S. (1995). Transcriptional activation by the orphan nuclear receptor ARP-1. *Nucleic Acids Res*, **23**, 1536-1543.
- Martinez, D. E., Dirksen, M. L., Bode, P. M., Jamrich, M., Steele, R. E., and Bode, H. R. (1997). Budhead, a fork head/HNF-3 homologue, is expressed during axis formation and head specification in hydra. *Dev Biol*, **192**, 523-536.
- Miller, D.J. and Miles, A. (1993). Homeobox genes and the zootype. *Nature* **365**, 215-6.
- Mokady, O., Dick, M. H., Lackschewitz, D., Schierwater, B., and Buss, L. W. (1998). Over one-half billion years of head conservation? Expression of an ems class gene in *Hydractinia symbiolongicarpus* (Cnidaria: Hydrozoa). *Proc Natl Acad Sci U S A*, **95**, 3673-3678.
- Mount, Stephen M., (1982) A catalogue of splice junction sequences. *N.A.R.* **10** (2) 459-472
- Naito, M., Ishiguro, H., Fujisawa, T. and Kurosawa, Y. 1993. Presence of eight distinct homeobox-containing genes in cnidarians. *Febs Letters*, **333**, 271-274.
- Power, R. F., Lydon, J. P., Conneely, O. M., and O'Malley, B. W. (1991). Dopamine activation of an orphan of the steroid receptor superfamily. *Science*, **252**, 1546-1548.
- Power, S. C. and Cereghini, S. (1996). Positive regulation of the vHNF1 promoter by the orphan receptors COUP- TF1/Ear3 and COUP-TFII/Arp1. *Mol Cell Biol*, **16**, 778-791.
- Qiu, Y., Cooney, A. J., Kuratani, S., DeMayo, F. J., Tsai, S. Y., and Tsai, M. J. (1994). Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon. *Proc Natl Acad Sci U S A*, **91**, 4451-4455.
- Qiu, Y., Pereira, F. A., DeMayo, F. J., Lydon, J. P., Tsai, S. Y., and Tsai, M. J. (1997). Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. *Genes Dev*, **11**, 1925-1937.

- Ritchie, H. H., Wang, L. H., Tsai, S., O'Malley, B. W., and Tsai, M. J. (1990). COUP-TF gene: a structure unique for the steroid/thyroid receptor superfamily. *Nucleic Acids Res*, **18**, 6857-6862.
- Schaller, H.C., S.A. Hoffmeister, and S. Dubel (1989). Role of the neuropeptide head activator for growth and development in hydra and mammals. *Development* **107**, 99-107
- Schuh, T. J. and Kimelman, D. (1995). COUP-TFI is a potential regulator of retinoic acid-modulated development in *Xenopus* embryos. *Mech Dev*, **51**, 39-49.
- Schummer, M., Scheurlen, I., Schaller, C., and Galliot, B. (1992). HOM/HOX homeobox genes are present in hydra (*Chlorohydra viridissima*) and are differentially expressed during regeneration. *Embo J*, **11**, 1815-1823.
- Shenk, M. A., Bode, H. R., and Steele, R. E. (1993a). Expression of Cnox-2, a HOM/HOX homeobox gene in hydra, is correlated with axial pattern formation. *Development*, **117**, 657-667.
- Shenk, M. A., Gee, L., Steele, R. E., and Bode, H. R. (1993b). Expression of Cnox-2, a HOM/HOX gene, is suppressed during head formation in hydra. *Dev Biol*, **160**, 108-118.
- Shibata, H., Nawaz, Z., Tsai, S. Y., O'Malley, B. W., and Tsai, M. J. (1997). Gene silencing by chicken ovalbumin upstream promoter-transcription factor I (COUP-TFI) is mediated by transcriptional corepressors, nuclear receptor-corepressor (N-CoR) and silencing mediator for retinoic acid receptor and thyroid hormone receptor (SMRT). *Mol Endocrinol*, **11**, 714-724.
- Tsai, S. Y. and Tsai, M. J. (1997). Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): coming of age. *Endocr Rev*, **18**, 229-240.
- van der Wees, J., Matharu, P. J., de, R. K., Destree, O. H., Godsave, S. F., Durston, A. J., and Sweeney, G. E. (1996). Developmental expression and differential regulation by retinoic acid of *Xenopus* COUP-TF-A and COUP-TF-B. *Mech Dev*, **54**, 173-184.
- Vlahou, A., Gonzalez, R. M., and Flytzanis, C. N. (1996). Maternal mRNA encoding the orphan steroid receptor SpCOUP-TF is localized in sea urchin eggs. *Development*, **122**, 521-526.
- Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M. A., and Rastinejad, F. (1998). Structural elements of an orphan nuclear receptor-DNA complex. *Mol Cell*, **1**, 849-861.

## FIGURE LEGENDS:

FIGURE 1 : (A) Structure of the various COUP-TF clones isolated in *Hydra* compared with the human COUP-TFI gene. Clone A comprises a part of the C domain with the D domain and linked in 5' an unrelated 5'UTR. Type I-III clones comprise the D and E domains linked to three different types of 5'UTR. Each type of clone is linked to the D domain at the same points at position 831 of clone A (vertical arrows). The clones depicted with a  $\Delta$  sign contains a deletion of 3 bp when compared to the others. ATG codons are shown by horizontal arrows whereas stop codons are shown by X. For each clone the position of ATG and stop codons are indicated. (B) Sequences of the various 5'UTR for types I-III clones. The two points of junction with the D-E regions are shown by vertical arrows. Stop and start codons are indicated.

FIGURE 2 : (A) Sequence of clone A. The region of junction with type I-III clones are indicated by stars. The DNA binding C domain is underlined together with the upstream stop codon. Region upstream of the stop codon is in lower case letters ; (B) Sequence of a type II clone, clone 6. The junction with clone A is indicated by stars. The E domain is underlined. Note the long F domain. For both clones the numbering of amino acid and nucleotides is indicated.

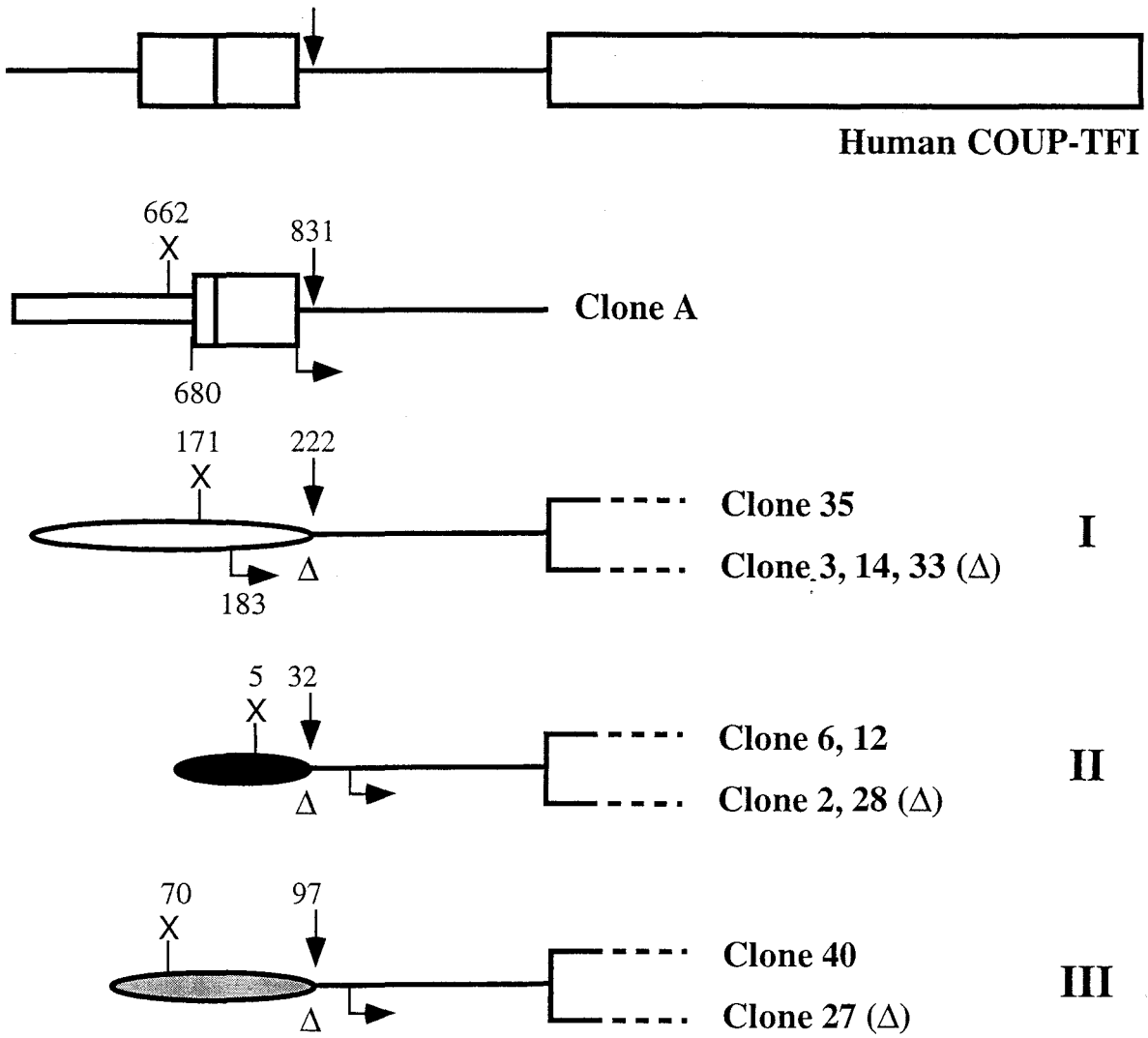
FIGURE 3 : Genomic organization of nuclear receptors for the C domain coding region. The putative introns of *Hydra* COUP-TF are indicated. The 5' intron of *Hydra* COUP-TF located in the P-box region is indicated with a question mark since it is located at a very unusual location. The 3' intron is located at the same position of a well conserved intron for all superfamily members. Two arrows are indicated for the 3' intron of *Hydra* COUP-TF corresponding to the two slightly different locations of the junctions between the different 5'UTR and the D-E domains.

FIGURE 4 : Alignement of several COUP-TF and EAR2 group members with the Hydra COUP-TF genes. Amino acid positions identical with the ones of Human COUP-TFI are indicated by dots, gaps are dashes. The A/B region was not included since two variable. The C domain is underlined in Human COUP-TFI. The E domain starts at position 127 of the alignment. Sequences are as follow with their genbank codes : COUPAH : human COUP-TFI (X12795), COUPAZF : zebrafish COUP-TFI (X70299), COUPBH : human COUP-TFII (M64497), COUBZF : zebrafish COUP-TFII (S80986), SVPD : drosophila SVP (M28863), COUPGX : xenopus COUP-TFG (X63092), COUPGZF : zebrafish COUP-TFG (X70300), EAR2H : human EAR2 (X12794), COUPHV : Hydra COUP-TF (this study).

FIGURE 5 : Phylogenetical Neighbor-Joining tree showing the placement of *Hydra* COUP-TF (arrows). Clearly this sequence split before the COUP-TF/EAR2 split. Two trees constructed with C+E domains (A) or C domain only (B) and showing similar topologies are shown. For the C domain tree *Hydra* COUP-TF is more closely connected with the COUP-TF/EAR2 clade. Bootstrap values (1000 replicates) are indicated on each branch.

FIGURE 6 : In situ Hybridization on *Hydra vulgaris* polyps.

The signal is localized in the gastric region of the polyps.

**A**

Langlois et al., Figure 1A

# B

## Type I :

... taa agaaaagga ATG TAT AAT GAA GAT GAA GAT GAG ATT CAT AAC TAT  
M Y N E D E D E I H N Y

## Type II :

... tgag caaacttctattttcttaataaa

## Type III :

... tgag taatagatgatatctcaaac

↓ ↓  
GCA GIT CAG AAG GGA AGA ...  
A V Q K G R



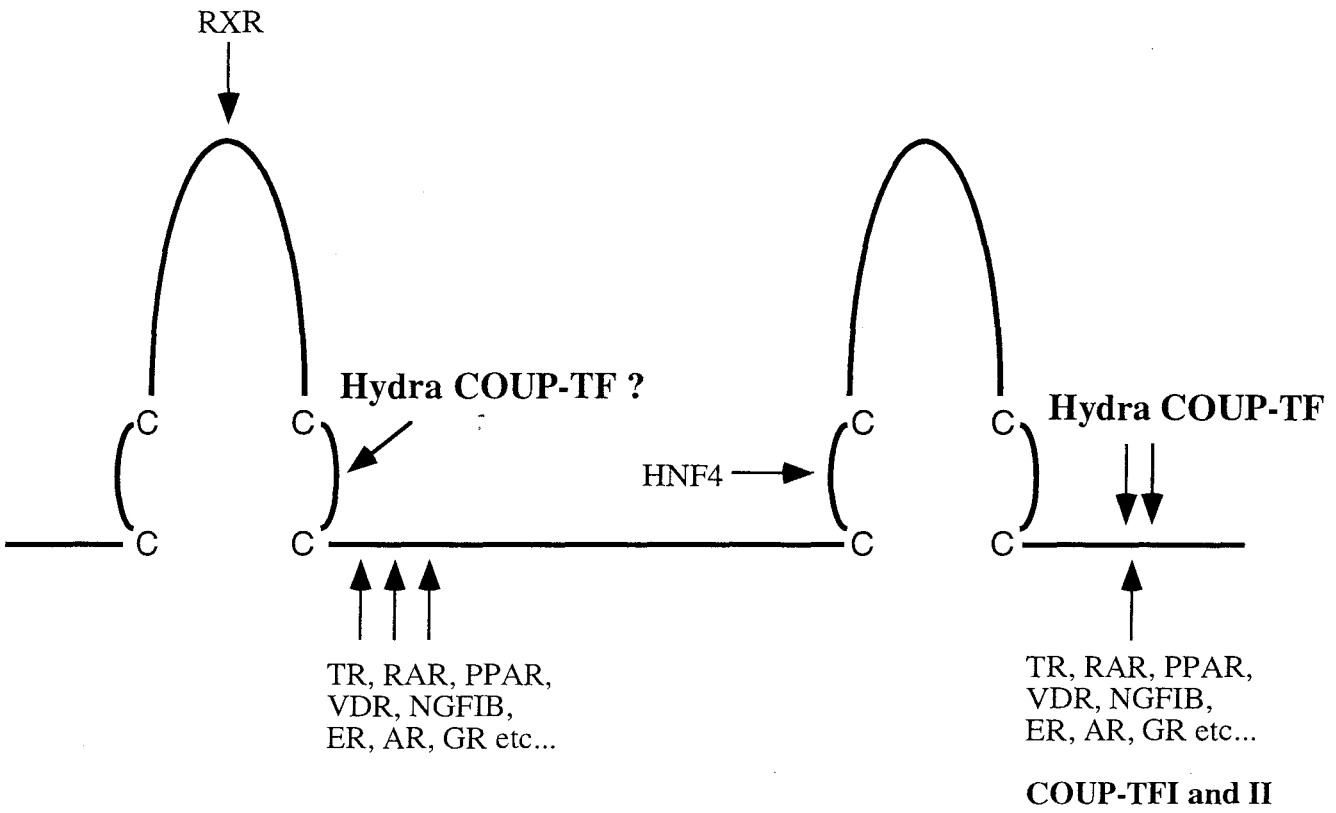
```

ggcacgagccttaattaacagcgatcagcaataagtcctttaatttacgattaaaagcagttagtttag 69
ataattagaagcaaaaaataaatacactaatttctctgtttttattatgagtcactctctatcctggattac 141
gtattaactaatggatagaaactggtaaaagaagcttatcttacattatgtcatcctaatgcaagcgtgagt 213
cataacaatcaaaacaataattcgtcttttgatcatgctaataatggatataatgtcaataattcaagagggttaa 285
tcaaaaccaatagattgtggagtttgcggtgataaaagctcaggaaaacattatggagtttacacttgcgaa 357
ggtagattactttcatcaattctacttattaatgtgttatttttacactttaatcatttataacataaaagt 429
attacgttaacctaataaacacattataatcacctaaaataaaacctaagcacattataaagtaatctacgat 501
attagataaatatgtttttttccatttaatcctttatgggtgatatagtggttaataactagaaaatttgcacag 573
gactacgtgctaggtttgggattattgtgtatacaaaaaacaaaaacaaaagcaaaaacctttgaagttgtt 645
          C I N V A G C K S F F K R 13
attaanaataatgtattaa TGT ATA AAT GTT GCA GGT TGC AAA AGT TTT TTC AAA CGA 704
S I R R N L A Y T C R A F Q N C S I 31
AGC ATT AGA CGA AAC TTA GCT TAT ACT TGC AGA GCC TTT CAA AAC TGT TCC ATT 758
D L N H R N O C O Y C R L K K C V K 49
GAC TTA AAT CAC AGA AAC CAA TGT CAA TAT TGC CGA TTG AAA AAA TGC GTC AAA 812
V G M R K D A * V * Q K G R L P K S P L 67
GTT GGA ATG AGA AAA GAC GCA*GTT*CAG AAG GGA AGA CTA CCC AAG TCA CCA CTA 866
E G S M V D Q F I A M N N N V S H Y 85
GAA GGT TCT ATG GTT GAT CAG TTT ATT GCT ATG AAT AAT AAT GTC AGC CAT TAT 920
N G L H Q I A N Y S N V L S A R L Q 103
AAT GGA TTA CAT CAG ATT GCA AAT TAT TCT AAC GTT TTA TCT GCA AGA TTG CAG 975
Q Q 105
CAA CAA 981

```

Langlois et al., Figure 2A

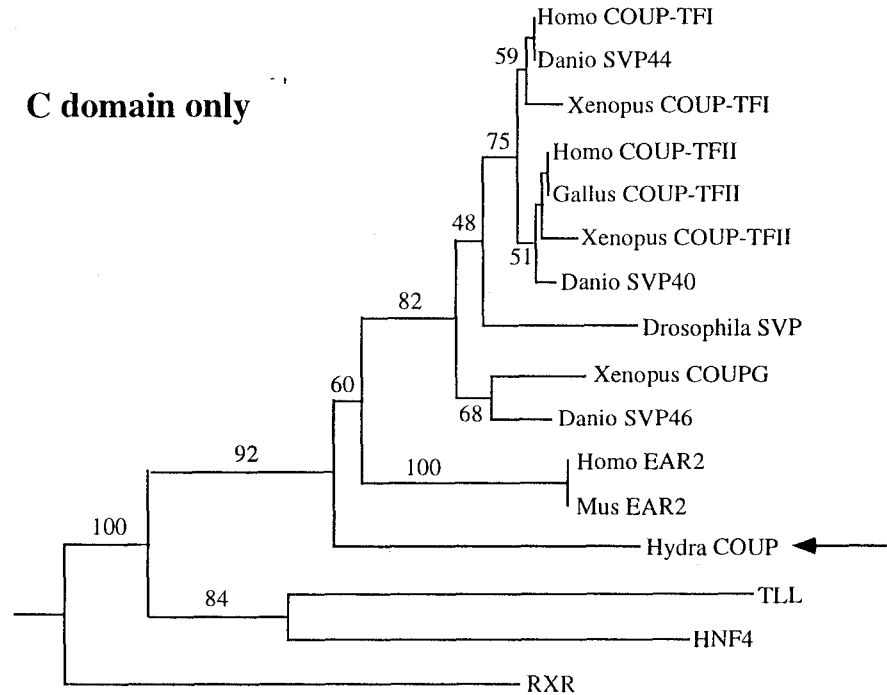
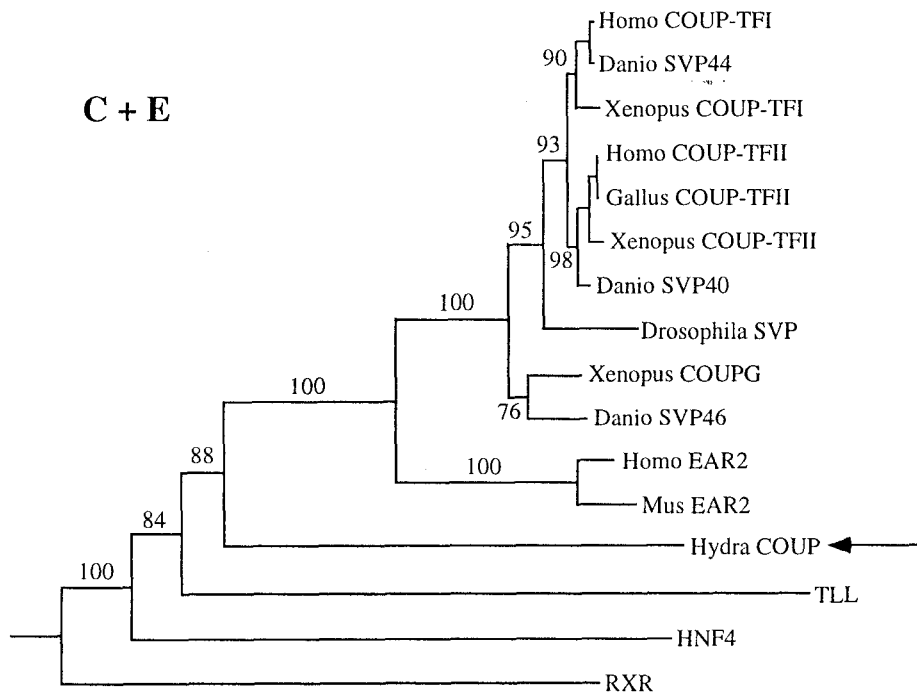
F	Y	F	L	N	K	*	A	*	V	Q	K	G	R	L	P	K	S	P	L	18	
TTC	TAT	TTT	CTT	AAT	AAA	*	GCA	*	GTT	CAG	AAG	GGA	AGA	CTA	CCC	AAG	TCA	CCA	CTA	54	
E	G	S	M	V	D	Q	F	I	A	M	N	N	N	N	V	S	H	Y		36	
GAA	GGT	TCT	ATG	GTT	GAT	CAG	TTT	ATT	GCT	ATG	AAT	AAT	AAT	GTC	AGC	CAT	TAT			108	
N	G	L	H	Q	I	A	N	Y	S	N	V	L	S	A	R	L	Q			54	
AAT	GGA	TTA	CAT	CAG	ATT	GCA	AAT	TAT	TCT	AAC	GTT	TTA	TCT	GCA	AGA	TTG	CAG			162	
Q	Q	Q	Q	N	O	Y	L	S	O	N	N	L	F	G	G	D	S			72	
CAA	CAA	CAG	CAA	AAT	CAA	TAC	TTA	AGT	CAA	AAC	AAC	TTG	TTT	GGT	GGG	GAT	AGT			216	
I	Y	E	M	A	T	R	L	L	F	N	A	V	E	W	A	R	N			90	
ATC	TAT	GAG	ATG	GCT	ACT	AGG	TTA	CTT	TTT	AAT	GCT	GTT	GAG	TGG	GCA	AGG	AAT			270	
V	P	F	F	S	A	L	P	T	S	D	O	I	A	L	L	K	S			108	
GTA	CCT	TTC	TTT	TCA	GCT	CTT	CCA	ACA	TCT	GAT	CAA	ATT	GCA	CTA	TTA	AAG	TCT			324	
S	W	S	E	L	Y	I	L	S	T	S	O	H	C	I	A	F	Q			126	
TCA	TGG	AGT	GAG	CTT	TAT	ATT	CTC	AGC	ACA	TCT	CAA	CAT	TGC	ATT	GCA	TTT	CAA			378	
I	N	A	R	A	L	T	S	E	O	N	L	S	E	I	K	K	R			144	
ATT	AAT	GCT	AGA	GCT	TTA	ACT	TCA	GAA	CAA	AAT	CTA	AGT	GAA	ATA	AAA	AAA	CGA			432	
S	E	G	A	N	E	A	S	V	K	M	F	E	E	L	V	E	R			162	
AGT	GAA	GGA	GCA	AAT	GAA	GCA	TCA	GTG	AAA	ATG	TTT	GAA	GAA	CTT	GTT	GAA	AGA			486	
F	K	N	L	O	T	D	A	A	E	F	S	C	L	K	A	L	V			180	
TTC	AAA	AAT	CTG	CAA	ACT	GAT	GCT	GCT	GAA	TTT	TCA	TGT	CTG	AAA	GCT	TTG	GTA			540	
L	F	N	P	D	S	P	G	L	G	N	P	S	L	I	E	N	L			198	
CTG	TTT	AAC	CCA	GAT	TCT	CCT	GGA	TTA	GGT	AAT	CCA	TCA	CTT	ATA	GAA	AAT	CTC			594	
Q	E	K	A	O	S	A	L	E	D	Y	L	R	O	O	N	A	T			216	
CAG	GAG	AAA	GCG	CAA	TCA	GCT	CTT	GAA	GAT	TAT	TTG	CGT	CAA	CAA	AAT	GCA	ACA			648	
Q	S	Y	H	N	R	F	G	K	L	L	L	R	L	P	A	L	S			234	
CAG	AGT	TAT	CAT	AAT	AGA	TTT	GGA	AAA	CTT	TTA	CTT	CGT	TTA	CCT	GCG	TTA	AGT			702	
L	I	R	P	A	T	I	E	A	L	F	F	P	R	H	H	G	S			252	
TTA	ATT	AGA	CCA	GCT	ACT	ATT	GAA	GCA	TTA	TTT	TTT	CCT	CGT	CAC	CAT	GGA	AGT			756	
Q	N	I	D	S	L	V	G	S	M	L	L	Y	G	L	G	S	N			270	
CAA	AAT	ATT	GAT	AGT	CTT	GTT	GGA	TCA	ATG	TTA	TTG	TAT	GGT	TTA	GGT	TCC	AAT			810	
N	N	T	S	F	V	N	N	G	I	P	A	V	G	L	N	G	G			288	
AAT	AAT	ACA	TCT	TTT	GTT	AAC	AAT	GGA	ATT	CCT	GCT	GTT	GGT	CTT	AAT	GGA	GGT			864	
S	S	G	S	I	L	N	M	T	G	L	A	G	H	V	T	S	G			306	
TCA	AGT	GGA	TCA	ATA	TTA	AAC	ATG	ACC	GGA	TTA	GCT	GGA	CAT	GTT	ACT	TCA	GGG			918	
F	I	N	N	A	N	L	A	F	S	N	A	N	N	L	S	S	L			324	
TTT	ATA	AAT	AAT	GCA	AAT	TTA	GCT	TTT	TCA	AAT	GCT	AAC	AAT	TTA	AGT	TCA	TTA			972	
Q	M	P	I	Q	D	N	I	Q	N	L	C	N	P	N	M	F	S			342	
CAA	ATG	CCA	ATA	CAA	GAT	AAT	ATA	CAG	AAT	CTT	TGT	AAT	CCT	AAT	ATG	TTC	TCT			1026	
A	T	G	I	N	P	V	N	G	L	N	A	S	N	G	Y	G	H			360	
GCA	ACA	GGT	ATC	AAT	CCT	GTC	AAT	GGC	TTA	AAT	GCA	TCA	AAT	GGG	TAT	GGT	CAT			1080	
H	V	I	N	N	M	M	N	Q	H	S	V	A	N	R	N	S	Q			378	
CAT	GTG	ATA	AAT	AAT	ATG	ATG	AAT	CAG	CAT	TCT	GTA	GCC	AAT	CGA	AAC	TCT	CAA			1134	
N	N	S	C	V	Q	S	M	Q	Q	N	V	Q	M	H	L	S	P			396	
AAT	AAC	TCA	TGT	GTG	CAA	AGC	ATG	CAA	CAA	AAT	GTC	CAA	ATG	CAT	TTA	AGT	CCA			1188	
T	T	R	H	H	I	K	I	E	P	P	L	R	*							409	
ACC	ACA	AGA	CAC	CAT	ATT	AAG	ATT	GAG	CCT	CCT	CTT	AGG	TGA	tacaatgatgtgcc						1245	
t	a	a	t	t	t	a	a	a	t	a	t	a	t	a	t	a	t	a	t	a	1316
t	g	c	a	t	g	t	t	t	t	a	a	a	a	a	a	a	a	a	a	t	1387
g	t	t	t	g	t	a	t	a	a	a	a	a	a	a	a	a	a	a	t	1458	
t	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	t	1529	
t	a	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	1586	



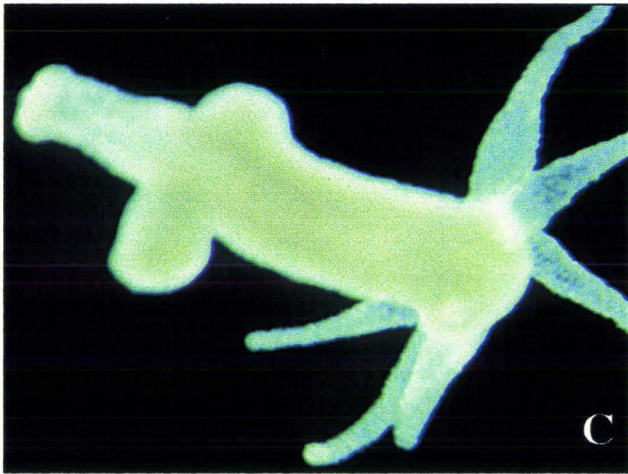
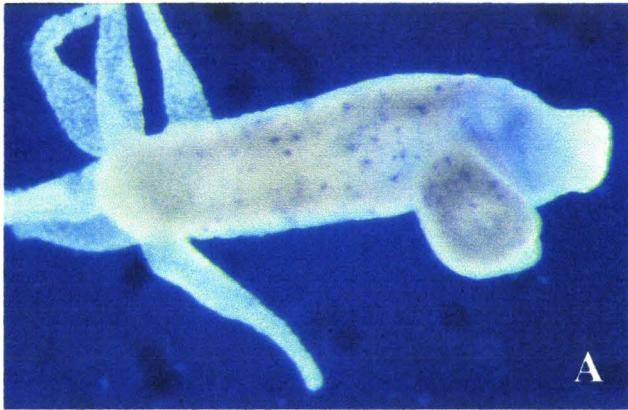
Langlois et al., Figure 3

	10	20	30	40	50	60	70	80	90	100	110
COUPAH	<u>IECVVCGDKSSGKHYGOFTCEGCKSFFKRSVRRNLTYTCRANRNCPIDOHHRNOCOYCRLLKCKLVGMRRREAVQRGRMPPTQP</u> ----NPGQYAL/INGD-----PLNGHCYLSG-----										
COUPAZF	.....										
COUPBH	.....S.....										
COUPBZF	.....TH..F.....C.S.....										
SVPD	.....S.GS.....M.....V.....GLAGMH...QIA...PMG---IAGF...S...S-----										
COUPGX	VD.L.....S..D.....SHP.T---S...T.N.V.-----Y..S..T.-----										
COUPGZF	VD.M.....S..G..D.....SNS.S---S...LSNGS.-----Y..QP.....										
EAR2H	VD.....V.....TI...S...S..D.Q.....FR...K.....I.HSL.-----AV.ASS.SPPGSALAAVAS.GDLFP.QPVSE										
COUPHV	-----I...A...FQ..S..LN.....V...KD...K..L.KSPLEG-SMVD.FIAM.NN-----VSHY..LHQIAN-----										
	120	130	140	150	160	170	180	190	200	210	220
COUPAH	YISLLLRAEPYP--TSRYGSQCMQPNNIMGIENICELAAARLLFSAVEWARNI PFFPDLQITDQVSLRLRLTWSELVFLNAAQCSMPLHVAPLLAAAGLHASPMSADRVVAFMDHIRIFQEQV										
COUPAZF	--.....										
COUPBH	--..F.....M.....A.....										
COUPBZF	--.....M.....A.....										
SVPD	--..-..D.....K.....E.V...A...V.....S.....A.....										
COUPGX	F.....A..L.....I..K.....F.LS...M.....R.....										
COUPGZF	F.....A...S..L.....K.....LM..A..MS.....E.....V.....										
EAR2H	L.AQ.....AAAG.F.AGGAAGAVL..D.V.....T...HG-..E.PVA..A..MS.....AAL...T.....A..A.E.A....QV.A....										
COUPHV	-Y.NV.S.RLQQ---QQQN.YLSQ..LF.GDS.Y.M.T...N.....V...SA.PTS..IA..KSS...YI.STS.HCIAFQINARALTSEQNL.EIKKRSEG.NEASVKM.E.L.										
	230	240	250	260	270	280	290	300	310	320	327
COUPAH	EKLKALHVDSA EYSCLKATVLF*PSDACGLSDAAHIESIQEKSQCALEEYVRSQYPNQ--PSRFGKLLRLPSLRTVSSSVIEQLFFVRLVGKTPIETLIRDMLLSGSSFNWPYMSIQ										
COUPAZF	.....I.....										
COUPBH	.....V..V.....T.....A..										
COUPBZF	.....V..V.....T.....										
SVPD	.....T.....VT.....C.T.....T.....Q.....N..S...LPSM										
COUPGX	.....A..P..V...IG.V..I.....N.....T..R.....I..AP.....PM.										
COUPGZF	.....Q..T.....S...M...V..V..I.....N.....N..R.....I..P.....L.....Y...PV.										
EAR2H	D..GR.Q.....G...A..P.....P..V.....A.V..T...A..S--Q..R.....A..A.PA.L.S...M.....T...G.G.										
COUPHV	.RF.N.QT.A..F...L...NP.SP..GNPSL..N...A.S...D.L.Q.NAT.SYHN.....A.SLIRPAT..A..P.HH.SQN.DS.VGS...Y.LGS.NNTSFWN										

Langlois et al., Figure 4



Langlois et al., Figure 5



**Figure 6 :**  
Whole mount *in situ* hybridization  
a) D domain antisense DIG probe,  
alkaline phosphatase revelation.  
b) overexposure  
c) negative control, sense probe.



## Discussion de l'article 5

Nous avons isolé l'homologue des gènes COUP-TFs chez un cnidaire, *Hydra vulgaris*. Nous avons obtenus par criblages de banques plusieurs types de clones : i) Le clone A présente un domaine C partiel et un domaine D, avec un cadre de lecture ouvert à partir du début du domaine D. ii) Les autres clones chevauchent le clone A sur le domaine D, comportent un domaine E et un domaine F, pour les clones complets en 3' (présence d'un codon stop). Ces clones dépourvus de domaine C, divergent dans leur extrémité 5'. Trois types de clones ont pu être obtenus qui se racordent tous au GTT en position 831 du clone A, additionné parfois d'un triplet GCA supplémentaire. Cette position est située derrière le deuxième doigt de zinc du domaine C et correspond à un site d'épissage commun à tous les récepteurs nucléaires. Par ailleurs, le clone A diverge en 5' en amont et dans les deux premiers acides aminés de sa boîte P, sur les positions AG et GT en 680 et 683 de la séquence du clone A. Ces éléments suggèrent la position d'un site d'épissage à cet endroit. Le clone A serait soit un produit alternatif, soit un pré-messager partiellement épissé. Néanmoins, on ne peut pas exclure la possibilité d'une chimère durant la rétrotranscription pour la construction de la banque (aucun site EcoRI ou linker n'a été retrouvé à cet endroit, ce qui tend à exclure la chimère de clonage). Cependant, si la présence de cet intron était vérifiée dans le génomique, les gènes COUP-TF représenteraient une catégorie particulière de récepteurs, qui auraient perdu un intron de leur domaine C, au cours de l'évolution.

Il est fort probable qu'il existe une isoforme de *Hydra* COUP-TF comportant un domaine C dans un cadre de lecture ouvert. Les acides aminés participant aux interactions avec les bases du site de reconnaissance dans le grand sillon de l'ADN, ou même ceux qui contactent les phosphates du squelette de la double hélice (si on se réfère à la structure 3D du récepteur orphelin, rev-erbA: (Zhao *et al.*, 1998)), sont, en effet, conservés chez l'hydre. Par ailleurs, les clones clones 6 et 1/27, dépourvus de domaine C, sont entiers (cadre de lecture fermé en 5' et codon stop en 3'). Quelle peut donc être la fonction de telles protéines déléetées de domaine C? Un récepteur de ce type, DAX-1, a déjà été isolé chez l'homme. Celui-ci réprime les transactivations médiées par les récepteurs aux stéroïdes. Il a été montré récemment, que DAX-1 interagissait avec SF-1, un récepteur orphelin usuel (doté d'un domaine A/B, C, D et E) impliqué dans de nombreux processus de régulation liés à la reproduction. DAX-1,

fixé sur SF-1, capture le corépresseur N-coR et maintient SF-1 dans un état de répresseur (Crawford *et al.*, 1998). Le domaine E de COUP-TF de souris interagit également avec N-coR (Shibata *et al.* 1997). Les isoformes d'hydre dépourvues de domaine C pourraient agir de la sorte, en s'hétérodimérisant avec un autre récepteur ou un autre facteur de transcription et de cette façon l'inactiver ou le maintenir dans un état de répresseur, comme dans le cas de DAX-1 et SF-1.

L'étude phylogénétique de *Hydra* COUP-TF situe ce gène avant la dichotomie EAR-2/COUP-TFs. *Hydra* COUP-TF serait donc l'homologue de l'ancêtre de la famille des gènes COUP-TFs des triploblastes.



*Article 6*

## Article 6

**AmphiCOUP-TF and AmphiTR2/4, two nuclear orphan receptors of the lancelet *Branchiostoma floridae*, are implicated in retinoic acid signaling pathways.**

**Marie-Claire Langlois**, Nicolas D. Holland, Jean-Marc Vanacker, Hector Escriva, Christophe Quéva, Vincent Laudet et Linda Holland.

En préparation pour "Development".

Chez plusieurs vertébrés, les récepteurs nucléaires orphelins des groupes COUP-TF et TR2/4 sont décrits comme des régulateurs négatifs de l'activité transcriptionnelle des récepteurs classiques, comme les récepteurs des acides rétinoïques (RAR) ou les récepteurs des hormones thyroïdiennes (TR). Ces observations suggèrent que l'une des fonctions conservées au cours de l'évolution de ces gènes, pourrait être liée à cette régulation des voies de réponses aux rétinoïdes ou aux hormones thyroïdiennes.

Nous décrivons dans cet article, la caractérisation de AmphiCOUP-TF et AmphiTR2/4, les homologues chez l'amphioxus *Branchiostoma floridae* de COUP-TFI, COUP-TFII et TR2, TR4, respectivement. Nous démontrons par des expériences de retard sur gel, que AmphiCOUP-TF se fixe sur un large éventail d'éléments de réponse, de la même façon que ses homologues de vertébrés. Par ailleurs, AmphiCOUP-TF est un répresseur transcriptionnel qui inhibe fortement la transactivation médiée par les acides rétinoïques via les récepteurs RAR et RXR. AmphiTR2/4 contient une délétion de 11 acides aminés dans ses boîtes T/A, impliquées dans la fixation à l'ADN. En accord avec cette observation, AmphiTR2/4 fixe faiblement l'ADN.

L'étude de l'expression de ces récepteurs menée en collaboration avec Linda Holland montre que leur patron d'expression rappelle celui de leurs homologues de vertébrés. De façon intéressante, l'expression de AmphiTR2/4 est régulée négativement par l'acide rétinoïque. L'ensemble de ces résultats suggère que (i) les voies de réponse aux rétinoïdes existent chez l'amphioxus et (ii) que l'une des fonctions commune aux récepteurs des groupes COUP-TF et TR2/4 chez les vertébrés est d'inhiber (donc de réguler finement) les voies de réponse aux ligands

classiques des récepteurs nucléaires comme les acides rétinoïques et les hormones thyroïdiennes.

Title page :

**AMPHICOUP-TF and AMPHITR2/4, TWO NUCLEAR ORPHAN RECEPTORS OF THE LANCELET *Branchiostoma floridae*, ARE IMPLICATED IN RETINOIC ACID SIGNALING PATHWAYS.**

Marie-Claire LANGLOIS<sup>1</sup>, Nicholas D. HOLLAND<sup>3</sup>, Jean-Marc VANACKER, Hector ESCRIVA<sup>2</sup>, Christophe QUEVA<sup>4</sup>, Vincent LAUDET<sup>2\*</sup> and Linda Z. HOLLAND<sup>3</sup>

<sup>1</sup> UMR 319 du CNRS, Institut de Biologie de Lille, Institut Pasteur, 1 rue Calmette, 59019 Lille Cedex, France.

<sup>2</sup> UMR 49 du CNRS, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France.

<sup>3</sup> Marine Biology Research Division, Scripps Institution of Oceanography, La Jolla, CA 92093-0202, USA.

<sup>4</sup> Fred Hutchinson Cancer Research Center, Seattle, USA.

Keywords : Amphioxus, Cephalochordate, Nuclear hormone receptors, COUP-TF, TR2, TR4, Transcription factors, Retinoic acid.

Running title : Amphioxus COUP-TF and TR2/4

## SUMMARY

In vertebrates, the orphan nuclear receptors of the COUP-TF and TR2/4 groups function as negative transcriptional regulators that down-regulate the hormonal induction of target genes mediated by classical members of the nuclear hormone superfamily such as the retinoic acid receptors (RARs) or the thyroid hormone receptors (TRs). To investigate the evolutionary conservation of the roles of COUP-TF and TR2/4 in negative cross-talk with the retinoid and thyroid hormone pathways, we have characterized AmphiCOUP-TF and AmphiTR2/4, the homologues of COUP-TFI, COUP-TFII and TR2, TR4 respectively, in the lower chordate amphioxus (*Branchiostoma floridae*), which is the closest living invertebrate relative of the vertebrates. Electrophoretic mobility shift assays (EMSA) show that AmphiCOUP-TF binds to a wide variety of response elements as do its vertebrate homologues. Furthermore, AmphiCOUP-TF is a transcriptional repressor that strongly inhibits retinoic acid mediated transactivation. AmphiTR2/4 contains an 11 amino-acid deletion in the T/A boxes which have been implicated in DNA binding. Consistent with this observation AmphiTR2/4 poorly binds to DNA. *In situ* hybridizations reveal expression of AmphiCOUP-TF in the nerve cord of late larvae in a pattern of dorso-ventral stripes, in a region corresponding to hindbrain and probably anterior spinal cord. Although the amphioxus nerve cord appears unsegmented at the gross anatomical level, this pattern evidently reflects segmentation at the cellular level with stripes of expressing cells occurring adjacent to the ends and the centers of each myotomal segment locations which may include respectively visceral motor neurons and somatic motor neurons among other cells. In contrast, AmphiTR2/4 is expressed in the neurula and early larva. Expression is strongest throughout the cerebral vesicle, the anterior-most region of the nerve cord, corresponding to the vertebrate diencephalon and perhaps anterior midbrain, Hatschek's anterior left diverticulum, part of which develops into Hatschek's pit, homologous to the adenohypophysis, and in the pharynx in the forming and newly-formed gill slits. Treatment at the gastrula stage with  $1 \times 10^{-6}$  all-*trans* retinoic acid (RA), which eliminates the gill slits and causes an anterior shift in Hox gene expression in the nerve cord, downregulates expression of

AmphiTR2/4. The only remaining expression in larvae is at the extreme anterior end of the pharynx. Thus, AmphiTR2/4 may function in RA-signalling pathways involved in formation of the gill slits and patterning of the nerve cord. A comparison of expression patterns of AmphiCOUP-TF and AmphiTR2/4 with those of their vertebrate homologues, suggests that their respective roles in patterning of the nerve cord and in RA-signalling evolved prior to the split between the amphioxus and vertebrate lineages.

## INTRODUCTION

The nuclear hormone receptor superfamily contains transcription factors that are ligand-dependent as well as so-called "orphan receptors" for which no ligand has yet been described (Enmark and Gustafsson, 1996; Gronemeyer and Laudet, 1995; Mangelsdorf and others, 1995). This superfamily of structurally-related members includes receptors for unrelated ligands such as thyroid hormone, retinoids, steroid hormones, and derivatives of fatty acids such as prostaglandins J2 or leukotriene B4. Whether orphan receptors are constitutive transcription factors or their ligands remain to be identified remains an open question.

The COUP-TF and TR2/4 transcription factors define two distinct groups of orphan receptors (Graupner and others, 1989; Gronemeyer and Laudet, 1995). The first member of the COUP-TF group, COUP-TFI was isolated as a regulator of the chicken ovalbumine gene and also, under the name of ear3, as being distantly related to the thyroid hormone receptor (Miyajima and others, 1988; Wang and others, 1989). A second gene, ARP-1, or COUP-TFII, was identified as a negative regulator of the apolipoprotein A1 gene transcription (Ladiaz and Karathanasis, 1991). COUP-TF proteins from various species have a high degree of sequence identity: 95-100% in the DNA binding domain (C domain) and 89-96% in the ligand binding domain (E domain) (Chan and others, 1992; Fjose and others, 1993; Fjose and others, 1995; Matharu and Sweeney, 1992; Mlodzik and others, 1990; van der Wees and others, 1996). By comparison little is known about the TR2 and TR4 orphan receptors. Both genes were isolated because of high sequence identity to the DNA binding domain of steroid receptors ([Chacon, 1986 #1415; Chang, 1994 #3166; Hirose and others, 1994; Chang and Kokontis, 1988) . Homologues in drosophila (DHR78) and sea urchin (SpSHR2) of TR2 and TR4 have been identified (Fisk and Thummel, 1998; Kontogianni and others, 1996). The TR2/4 receptor group is more variable than the COUP-TF group having 73-80% and 50-75% identities in the C and E domains respectively.

Much is known about the *in vitro* transcriptional effect and the target genes of COUP-TF and TR2/4 proteins, particularly in vertebrates. Although clearly different at the sequence and expression levels, both factors exhibit strikingly

similar functional characteristics. In all species examined, both bind to a wide range of Hormone Response Elements (HRE, reviewed in [Tsai, 1997 #3897; Hirose and others, 1994), including direct repeats of the AGGTCA core element (Cooney and others, 1991; Cooney and others, 1992; Hirose and others, 1995; Hwung and others, 1988; Ladas and Karathanasis, 1991; Lee and others, 1995), and both behave as potent negative transcriptional regulators on their response elements. Furthermore, they also down-regulate the hormonal induction of target genes by retinoic acid receptors (RARs), thyroid hormone receptors (TRs), Vitamin D receptor (VDR) and peroxisome proliferator activated receptors (PPARs) (reviewed in Lee and others, 1997; Lin and others, 1995a; Tsai and Tsai, 1997; Yan and others, 1998). However, both COUP-TF and TR4 may behave as transcriptional activators under some *in vitro* conditions (Chang and Wei, 1997; Hall and others, 1995; Lee and others, 1997; Malik and Karathanasis, 1995; Power and Cereghini, 1996).

However, spatio-temporal expression of COUP-TF and TR2/4 genes is strikingly different, suggesting that they have different functions in embryogenesis. Embryonic expression of COUP-TF genes in both vertebrates and drosophila is largely restricted to the central and peripheral nervous system (reviewed in Tsai and Tsai, 1997). In particular they are expressed in the diencephalon of the mouse (Qiu and others, 1994), in a segmental pattern in the hindbrain of the zebrafish (Fjose and others, 1993; Fjose and others, 1995), in embryonic motor neurons in mouse and chick (Lutz and others, 1994; Qiu and others, 1994), in the eye of zebrafish and xenopus (Fjose and others, 1993; van der Wees and others, 1996; C. Quéva and B. Eisenman, pers; com.), and in the brain and eye of drosophila, being implicated in cell fate determination of photoreceptor neurons (Mlodzik and others, 1990). Knockout experiments of mouse COUP-TFI have confirmed that it is critical for nervous system development (Qiu et al., 1997). In vertebrates, COUP-TF is also expressed in mesenchymal tissue of many developing organs, and also in many adult organs including the brain (Lopes da Silva and others, 1995).

In contrast, in embryos TR2/4 genes are generally expressed during early stages predominantly in non-nerual tissues. For example, mouse TR2 is expressed during early development of many tissues including kidney and intestine (Lin et al., 1995), while TR4 is expressed as two different transcripts - one in testes, the



other ubiquitously (Hirose and others, 1994). TR2/4 genes are typically downregulated in late embryogenesis but are subsequently upregulated in post embryonic stages in a variety of tissues including the brain, testis, pituitary, thyroid and spleen (Chang et al., 1994). In early sea urchin embryos, the homologue of TR2 and TR4, is stored as maternal RNA in the egg and not detected beyond the blastula stage (Kontrogianni and others, 1996). In drosophila DHR78 is expressed throughout development, with peak of expression in third instar larvae and prepupa that correlates with the known ecdysone pulses. drosophila mutant in the DHR78 gene die during the third larval instar with an inability to pupariate in response to ecdysone (Fisk and Thummel, 1995; Fisk and Thummel, 1998).

A major difference between drosophila and vertebrates, is that exogenously applied retinoic acid (RA) strongly affects embryogenesis of vertebrates, but is without effect on development of drosophila. In vertebrates, RA alters the expression patterns of both COUP-TF and TR2/4 (Fjose and others, 1995; van der Wees and others, 1996). In the xenopus neurula RA enhances COUP-TFG expression but decreases COUP-TFII (van der Wees and others, 1996), while overexpression of COUP-TFI alters the transcriptional response to RA in *ex vivo* explants (Schuh and Kimelman, 1995) and blocks RA-induced neuronal differentiation *in vitro* (Neuman and others, 1995). The expression patterns of TR2 and TR4 largely overlap the ones of RARs and RXRs and interestingly, TR2 levels decrease in the testis of vitamin-A depleted animals suggesting that the expression of this gene may be regulated by retinoids (Lee and others, 1996). In addition, in *in vitro* experiments, TR2 can suppress transcription of reporter constructs containing retinoic acid response elements such as the one found in the genes coding the cellular retinoic acid binding protein I and cellular retinol-binding protein II (Chinpaisal and others, 1997; Lin and others, 1995b). Taken together these data suggest that COUP-TF and TR2/4 generally function in vertebrates as negative controllers of retinoid signaling. In contrast, the absence of effect of exogenous RA on development in drosophila suggests that in drosophila, at least, COUP-TF and TR2/4 homologs do not function in retinoid signalling. Indeed, while drosophila and ticks have RXRs (but not RARs, which are known only from vertebrates), the drosophila RXR, unlike the vertebrate RXR,

does not bind 9-*cis* RA (Escriva and others, 1997; Guo and others, 1997; Guo and others, 1998; Mangelsdorf and others, 1992).

These differences between drosophila and vertebrates raise the questions as to when in evolution retinoid-signalling pathways arose and when did COUP-TF and TR2/4 genes acquire a function in repressing retinoid signalling. A role for retinoid-signalling in embryogenesis evidently arose at or before the evolution of lower chordates, which include tunicates and the cephalochordate amphioxus. In both groups exogenous RA is a teratogen and alters the expression of anterior Hox genes as it does in vertebrates (De Bernardi et al., 1994 ; Holland and Garcia, 1996; Katsuyama and others, 1995). In addition, in amphioxus, which is the closest living invertebrate relative of the vertebrates, exogenous RA affects the morphogenesis of the gill slits as it does in vertebrates, and alters the expression of Pax1/9 in the pharynx associated with gill slit formation (Holland and Garcia, 1996). However, nothing is known about the receptors involved in mediating retinoid signalling in lower chordates. In order to study the evolution of retinoic acid signalling pathways, we have begun a study of receptors in the steroid hormone superfamily in amphioxus (*Branchiostoma floridae*). Amphioxus is vertebrate-like, but much simpler, not having undergone the two rounds of genome duplication which occurred during vertebrate evolution (Garcia Fernandez and Holland, 1994). In recent years, developmental genetic studies of amphioxus have been providing insights into the evolution of patterning in vertebrate embryos (Holland and others, 1998; Holland and Garcia, 1996). In the present study, we report the cloning, expression and binding and activation characteristics of AmphiCOUP-TF and AmphiTR2/4 from amphioxus, the single amphioxus homologs of vertebrate COUP-TFI and COUP-FII and of vertebrate TR2 and TR4 genes. Phylogenetic analysis shows that the two amphioxus genes are located at the base of their respective vertebrate families. Like their vertebrate homologues, AmphiCOUP-TF is expressed in the nerve cord of late larvae whereas AmphiTR2/4 is expressed in embryos in a wide range of tissues, including forming and newly formed gill slits. Effects of exogenous RA on AmphiTR2/4 expression suggest that this gene plays a role in retinoid-signalling. Binding assays and transactivation assays support the conclusion that these genes, like their vertebrate homologs, may act as repressors in retinoid-signalling

pathways. We conclude that vertebrate-like retinoid-signalling pathways probably evolved before the separation of the vertebrate and amphioxus lineages.

## **MATERIALS AND METHODS**

### **Isolation of AmphiCOUP-TF and AmphiTR2/4 cDNA clones**

Using a 117bp PCR fragment corresponding to AmphiCOUP-TF (*Branchiostoma floridae*) (Escriva et al. 1997) as a probe, we screened a cDNA library constructed in lambda ZAP II (Stratagene, Inc. La Jolla, CA, USA) from 2-4 day larval amphioxus RNA (Holland et al., 1995). Three clones were isolated out of  $1 \times 10^6$  pfu. These clones were shown to be identical by digestion pattern and partial sequencing and contain a 1,9 kb insert. Only one AmphiTR2/4 3Kb sequence containing clone was obtained using a 134 pb genomic PCR fragment (Escriva et al., 1997), by screening the same cDNA library. Hybridization was done over night at 42°C in 50% formamide, 5X sodium citrate (SSC), 5X Denhardt's solution, 0,5 % SSC, 0,1 mg/ml denatured salmon sperm DNA. Membranes were washed twice in 2XSSC, 0,1XSSC at 37°C for 15 min. The inserts were sequenced on the two strands, using an Applied Biosystems 373A automatic sequencer and methods and reagents of the supplier.

### **Expression vectors and clones**

pSG5-mRXRa is a generous gift of P.Kastner and C.Reibel ; pSG5-mRARa of Zelent et al. (Zelent et al., 1989). mCOUP-TFI and COUP-TFII cloned in the pSG5 vector are generous gifts from M.J Tsai (Wang et al., 1989). pSG5-mTR4 is a generous gift of P. Maire.

### **Electrophoretic mobility shift assays**

Sequences of probes used:

- DRO: 5'CGCGATTTGAGGTCAAGGTCACACAGTTAG
- DR1 : 5'CGCGATTTGAGGTCAAAGGTCACACAGTTA
- DR2 : 5'GCGATTTGAGGTCAAGAGGTCACACAGTTA
- DR3 : 5'GCGATTTGAGGTCAAGGAGGTCACACAGTT
- DR4 : 5'CGATTTGAGGTCACAGGAGGTCACACAGTT

- DR5 : 5'CGATTTGAGGTCACCAGGAGGTCACACAGT
- HRE pal : 5'CGCGATTTGAGGTCATGACCTCACAGTTAG
- DR5 5'mut: 5'CGATTTGCCTAGCCCAGGAGGTCACACAGT
- DR5 3'mut: 5'CGATTTGAGGTCACCAGGCCTAGCCACAGT

DNA probes were [<sup>32</sup>P] labeled with the T4 polynucleotide Kinase (BRL) prior to annealing. Proteins were in vitro synthesized using the TNT kit (Promega, Madison, WI, USA). Protein production was evaluated by electrophoresis on acrylamide gels. Approximately equal amounts of proteins were added to each binding-assay.

Binding-assays were performed in 18 ml, with 5 ml reticulocyte lysate (unprogrammed reticulocyte lysate was added when necessary) with 1 ng of the DNA probe (50000 cpm) in a binding buffer containing 10% glycerol, 10 mM HEPES, 30 mM KCL, 0.5 mM DTT [dithiothreitol], 1.5 mg polydIdC, 0.25 mM PMSF (protease and phosphate inhibitors), 0.25 mM EDTA and 0.125 mM EGTA. Reactions were incubated for 10 min on ice, loaded on a 6% polyacrylamide gel and run at room temperature for 2h.

### **Cell transfections**

Cells were maintained in DMEM medium supplemented with 10% fetal calf serum. For transfections, 10<sup>5</sup> COS1 cells were seeded on 35 mm diameter plates. Transfections were performed using lipofectamine, with 0.4 mg reporter plasmid (DR1- or DR5CAT), 0.2 mg expression vectors. pSG5 vector was added as a carrier up to 1 mg. Cells were lysed 48 hours after transfection and analysed for CAT activity using standard methods.

### **Rearing of embryos and larvae and treatment with retinoic acid**

Adults of *Branchiostoma floridae* were obtained during the summer breeding season in Tampa, Florida, USA. Adults were spawned, and embryos and larvae raised as described in Holland and Holland (1993). For retinoic acid (RA) treatment, embryos were exposed to 1x10<sup>-6</sup> M all-*trans* RA from mid-blastula until hatching at the early neurula stage (Holland and Holland, 1996).

### ***In situ* hybridization**

*In situ* hybridization was according to the methods of Holland et al., (Holland and others, 1996). Since an antisense-riboprobe transcribed from a full-length cDNA of AmphiCOUP-TF which was about 1200 bp long and corresponded largely to 3' UTR gave a weak label, a 5' 1050 bp PstI fragment was subcloned into pBluescript SK and used as a template for a second riboprobe. To obtain a satisfactory signal, both riboprobes were combined. Similarly, for AmphiTR2/4 antisense riboprobes from transcription of a full-length cDNA and a 2030 5' AcclI subclone were combined to give an acceptable signal.

## RESULTS

### Isolation of AmphiCOUP-TF

We used a short PCR fragment obtained by "touch-down" PCR method on *B. floridae* DNA (Escriva et al., 1997) to screen a larval cDNA library for COUP-TF products. This screen yielded three 1.9 kbp clones that were identical as judged by partial sequencing. One clone was therefore completely sequenced (Fig.1A). It included a 1329 bp-long open reading frame (ORF) coding for a predicted 443 amino acids product. Although the 5' UTR of this clone is short (28 bp) and does not contain any in-frame stop codon upstream of the first methionine, the sequence around the first ATG includes the most conserved bases in the Kozak sequence, an A at -3 and a G at +4 (Kozak, 1996). In addition, the 15 first amino acids of the *B. floridae* COUP-TF exhibit a high level of identity (60%) with the amino-terminal amino acids of human COUP-TFI and COUP-TFII genes (Fig. 1B). A typical polyadenylation signal is lacking, although the long stretch of "A" residues at the 3' end suggests that the sequence AAATA (Fig. 1A, underlined) may be a non-standard polyadenylation signal.

The product encoded by the cDNA has all the distinctive features of a member of the nuclear receptor superfamily with a divergent N-terminal A/B domain, a strongly conserved C domain, a variable hinge (D domain) and a C-terminal well conserved E domain. When excluding the variable A/B domain, the percentage of amino acid identity between our clone and the human COUP-TFI and COUP-TFII is 93.5 % and 92 %, respectively, compared to 87.9 % to the drosophila Seven-up (SVP) gene product and 86.3 % to the the sea urchin *Strongylocentrotus purpuratus* COUP-TF. There is considerably less identity to the EAR2 gene (67.8 %). We conclude, therefore, that we have isolated the amphioxus homologue of the COUP-TF gene which will be hereafter referred to as AmphiCOUP-TF.

Fig. 1B compares the AmphiCOUP-TF amino acid sequence with COUP proteins from other species. We have included in this comparison and the subsequent phylogenetical analysis two new *xenopus* COUP-TF sequences recently isolated by one of us (C.Q.). The DNA binding domain of AmphiCOUP-TF is strongly conserved with only 4/67 amino acid replacements (94 % identity).

The P-box implicated in the selection of the DNA target sequence (CEGCKS, positions 158-163 in Fig 1B) is strictly conserved, whereas the D-box (positions 178-182) which chiefly functions in dimerization contains 2 substitutions (RANRN in human COUP-TFs, RGNRT in AmphiCOUP-TF). Interestingly, the product of the zebrafish SVP46 gene (Fjose et al., 1993) and the previously described *xenopus* COUP-TF gene (Matharu and Sweeney, 1992) which are divergent members of the COUP-TF group also contain two substitutions at identical positions in the D-box, among other divergences. The T/A boxes, located near the C-terminus of the AmphiCOUP-TF protein, which are implicated in DNA binding and dimerization of nuclear receptors, are also well conserved. The E domain of AmphiCOUP-TF (Fig. 1B) is 95.5 % and 94.1% identical with those of COUP-TFI and COUP-TFII, respectively, 92.3 % identical to that of drosophila SVP, 86 % identical to that of zebrafish SVP46, and 86.5% identical to that of *xenopus* COUPG. Some regions of the E domain, such its C-terminal part (positions 444-498) are completely conserved between AmphiCOUP-TF and the two mammalian COUP-TFs. Interestingly, this region corresponds to the CID domain that mediates the transcriptional repressive activity of COUP-TFs via co-repressor binding (Shibata and others, 1997). This suggests that the interaction surface between AmphiCOUP-TF and co-repressors is crucial and is strictly conserved from chordates to vertebrates.

### Isolation of AmphiTR2/4

Our PCR screen for amphioxus NR genes (Escriva and others, 1997) also yielded a 134 bp fragment corresponding to a member of the TR2/TR4 group of nuclear orphan receptors. In all vertebrates studied to date, this group comprises two paralogous genes (TR2 and TR4). Screening of the *B. floridae* cDNA library with our PCR fragment yielded a unique 3069bp long cDNA which contained a 1497bp long ORF predictably translated as a 499 amino acids long product (Fig. 2A). The 5' UTR of this clone is short (27 bp) and does not contain any in frame stop codon before the first methionine codon. Although, there is no match with the conserved -3A and +4G in the Kozak sequence, 4/7 bases outside the ATG start codon do match. Thus, this clone may represent a full-length cDNA. It has been pointed out that mRNAs lacking -3A and +4G are able to initiate translation from

the second ATG, which would seem a distinct possibility since the second ATG immediately follows the first (reviewed in Kozak, 1996). The product encoded by the cDNA shares 82.3 % amino acid identity with TR2 and TR4 in the DNA binding domain whereas the ligand binding domain is 65.3 and 73 % identical to the E domains of TR2 and TR4 respectively. The percentages of identity in the C and E domains compared to the sea urchin homologue SpSHR2 are 83.9 and 72.9 % respectively, whereas these values are 79.4 and 29.4% with the DHR78 gene in *drosophila*. This clearly indicates that we have isolated the amphioxus homologue of both TR2 and TR4 gene. This gene will thus be referred to as AmphiTR2/4.

AmphiTR2/4 has several interesting features, the most prominent of which is an 11 amino acid deletion in the region immediately adjacent the DNA binding domain, *i.e.* in the so-called T and A boxes (Fig. 2B). This region is functionally important since it is implicated in dimerization and DNA binding in a number of nuclear receptors. In TR2, as in a number of other nuclear receptor genes, an intron is present in this region and separates the DNA binding domain from the T/A boxes (Lee et al., 1995). To determine whether this short deletion resulted from alternative splicing or a mis-splicing, we performed RT-PCR experiments on RNA from 72 hr larvae and adult amphioxus using primers located on both sides of the deletion. Surprisingly, we obtained a single PCR fragment identical to the cDNA as determined by sequencing (data not shown). We thus concluded that this unusual feature of AmphiTR2/4 is genuine. Downstream of the T/A boxes, AmphiTR2/4 includes a region that is strongly conserved among all deuterostome TR2/4 proteins (22/28 amino acids) but apparently not in *drosophila* DHR78 or other insect homologues. The E domain of AmphiTR2/4 is similar to those of TR2/4 proteins in vertebrates, and, as noted above, shares greater identities with TR2 (65.3%) than TR4 (73%).

### **Phylogenetic analysis of the AmphiCOUP-TF and AmphiTR2/4 sequences**

We constructed a phylogenetic tree of COUP-TF proteins from *drosophila*, amphioxus and vertebrates, including a previously undescribed *xenopus laevis* COUP-TF sequence identified by one of us (C.Q.). The sea-urchin sequence Chan and others, 1992) was excluded from the analysis since its C-terminal part is



incomplete. A distance tree was first constructed using the Neighbor-Joining method and the C and E conserved domains (Fig. 3A). From this tree it is clear that the duplication leading to COUP-TFI and COUP-TFII genes occurred after the split between the cephalochordate and the vertebrate lineages, and that AmphiCOUP-TF is the homologue of the ancestral COUP-TF gene that led by duplication to COUP-TFI and COUP-TFII. Although the bootstrap value of chordate sequences *vs drosophila* SVP is low, the position of the latter sequence fits perfectly with the known phylogeny of metazoans. Trees based on the analysis of the C or E domain alone gave rise to similar topologies except that sometimes AmphiCOUP-TF clustered with low bootstrap values with *drosophila* SVP (data not shown). Identical conclusions could be drawn from a parsimony analysis using the PAUP program. The positions of *xenopus* and zebrafish COUP sequences are not entirely in agreement with known vertebrate phylogeny. For each species, there are three different COUP genes, which include homologs of COUP-TFI and COUP-TFII (respectively SVP44 and SVP40 in zebrafish) as well as a highly divergent COUP (*xenopus* COUPG and *Danio* SVP46). For COUP-TFI, the *xenopus* and zebrafish sequences are reversed compared to the known phylogeny.

A phylogenetic tree of TR2/4 sequences was also constructed (Fig. 3B). AmphiTR2/4 clustered as a unique homologue of both the vertebrate TR2 and TR4 sequences, being more closely related to the vertebrate TR2/4 sequences than to the *drosophila* homologue DHR78. Thus, it appears that AmphiTR2/4 is the unique TR2/4 homologue present in amphioxus.

#### **Developmental expression of AmphiCOUP-TF and AmphiTR2/4**

AmphiCOUP-TF is not expressed during development in *Branchiostoma floridae* until the 4-5 gill-slit stage, which occurs at about 3-4 weeks of development at 24°C (Fig. 4). Expression occurs in the nerve cord in the homolog of the hindbrain and spinal cord (Fig. 4B). The anteriormost portion of the nerve cord, the cerebral vesicle (the homolog of the diencephalon plus possibly anterior midbrain), does not express AmphiCOUP-TF (Fig. 4C). Expression is in dorsal and lateral groups of cells, which form dorso-ventral stripes in the nerve cord (Fig. 4). The floorplate does not express AmphiCOUP-TF. In amphioxus, because the nerve cord appears outwardly unsegmented, the axial muscles serve as positional

markers along the antero-posterior axis. The muscles on the right and left sides are out of register with one another such that the ends of muscles on one side are opposite the midpoints of muscles on the other (arrowheads Fig. 4E). These stripes of expression are adjacent to the ends of the muscle blocks, such that there are two stripes per muscle block.

In contrast, *AmphiTR2/4* is expressed only from the neurula through the early larva. In amphioxus, gastrulation results in an outer epiblast and inner hypoblast. In the early neurula the neural tube forms from the dorsal epiblast and the somites and notochord form from grooves in the dorsolateral and dorsal portions respectively of the hypoblast. The remainder of the epiblast is the ectoderm, and the remainder of the hypoblast forms the endoderm or larval gut. The mouth and gill slits form respectively on the left and right sides in the late neurula by fusion of the ectoderm and endoderm. There is no epithelial/mesenchymal interconversion in amphioxus. In the early neurula, expression of *AmphiTR2/4* is strongest in the cerebral vesicle, the posterior endoderm and in Hatschek's anterior left gut diverticulum, part of which develops in the adult into Hatschek's pit, the probable homolog of the adenohypophysis. There is weaker expression throughout the endoderm and moderate expression in the pharynx in the approximate region where the gill slits and mouth will later penetrate. As the neurula elongates and the mouth begins to form, expression is upregulated around the forming mouth and where the first two gill slits will soon penetrate. Expression remains fairly strong throughout the cerebral vesicle and in Hatschek's left anterior diverticulum. Prior to penetration of the third gill slit, *AmphiTR2/4* is upregulated where this gill slit will penetrate. Expression remains strong around the first three gill slits after they have formed.

In embryos treated with  $1 \times 10^{-6}$  all-trans RA mouth and gill slits do not form (Holland and Holland, 1996). Early larvae treated with exogenous RA may also exhibit slightly shorter anterior ends, particularly at concentration of RA in excess of  $1 \times 10^{-6}$  M RA. Expression of *AmphiTR2/4* is reduced by RA treatment. In early larvae, expression in the anterior nerve cord is eliminated and in the pharynx, expression occurs only at the extreme anterior end.

### Functional analysis of AmphiCOUP-TF and AmphiTR2/4.

The COUP-TF group members bind efficiently as homodimers to DR1 (AGGTCAnAGGTCA where n is any nucleotide) elements in the regulatory regions of target genes. Therefore, we performed electrophoretic gel mobility shift assays (EMSAs) to test binding of AmphiCOUP-TF to DR1 elements. Positive controls were mouse COUP-TFI and COUP-TFII (Fig.5A). The AmphiCOUP-TF protein bound strongly to the DR1 probe as did the mouse COUP-TFI or COUP-TF-II (Fig. 5A and data not shown). Binding was specific since it was out-competed by an excess of unlabelled oligonucleotide encompassing the DR1 sequence (lanes 3, 6 and 9) but not by an unrelated oligonucleotide (lanes 4, 7 and 10). Since the shift induced with AmphiCOUP-TF was approximately comparable to the one generated with the mouse COUP-TFI which binds this element as a homodimer, we suggest that AmphiCOUP-TF also bound the DR1 element as an homodimer. COUP-TFs are known to bind to a number of different response elements, we determined by competition EMSA whether AmphiCOUP-TF can bind to other known nuclear hormone response elements. Like its mouse homologs AmphiCOUP-TF can bind to all the DR elements (DR0 to DR5) as well as to the palindromic response element (HREpal) and the monomeric half site RevRE (Fig. 5B and data not shown). These competition experiments were confirmed by direct binding experiments on <sup>32</sup>P labelled probes encompassing the various DR elements (data not shown). However, it should be noted that the DR1 element was clearly the preferential AmphiCOUP-TF binding target. Taken together, these results suggest that the DNA binding activities of AmphiCOUP-TF are indistinguishable from those of its vertebrate homologues.

We also tested the *in vitro* synthesized AmphiTR2/4 protein in EMSA using mouse TR4 as a control (Fig 5C). Both proteins bound to the DR1 element, as expected for the mouse product (Hirose and others, 1995), although the level of AmphiTR2/4-DNA complex was lower than the one obtained with mouse TR4. Both complexes were specific since they could be out-competed by the homologous oligonucleotide (lanes 3 and 6) but not by an unrelated one (lanes 4 and 10). Neither factors bound to a DR5 element (data not shown).

Transcriptional activities of the AmphiCOUP-TF were next analyzed. To this end, COS-1 cells were cotransfected with a reporter construct consisting of a

trimer of the DR1 element cloned in front of the minimal tk promoter together with an AmphiCOUP-TF-expressing clone. AmphiCOUP-TF reduced expression of the reporter gene by about 20% possibly due to weak basal reporter expression. More interesting was the effect of COUP-TF on the transactivation mediated by RAR and RXR on a DR5 element. COS-1 cells were co-transfected with a reporter construct comprising a trimer of the DR5 element cloned in front of the minimal tk promoter together with either COUP-TFs, RXRa and/or RARa expression vectors (Fig.6). After transfection, retinoic acids ( $10^{-6}$  M *all-trans*- or *9-cis* retinoic acid) were added to the culture medium. As expected, these compounds generated a strong reporter activity when the cells were transfected solely with RAR, RXR or both. The fact that RXR, which binds only *9-cis* RA, could be stimulated by *all-trans* RA could be due to a high level of endogenous RAR that could heterodimerize with the exogenous RXR product. Alternatively, a significant amount of *all-trans* RA could be isomerized to *9-cis* RA within the cells. When the AmphiCOUP-TF expression vector was added to the co-transfection mixture, a strong reduction of the reporter activity was observed suggesting that AmphiCOUP-TF efficiently repressed the transactivation driven by the RAR/RXR heterodimer. This repressive effect of Amphi-COUP-TF was dose-dependant as shown in Fig 6B. Interestingly AmphiCOUP-TF also efficiently repressed transactivation driven by RXR in the presence of *9-cis* RA on a DR1-tk-CAT construct (Fig 6C). Taken together these results strongly suggest that AmphiCOUP-TF is a strong repressor of the retinoid signalling pathway in amphioxus, a phenomenon which is also conserved in all vertebrates.

## DISCUSSION

### Sequences of AmphiCOUP-TF and AmphiTR2/4

The amphioxus COUP-TF gene encodes a protein that is extremely similar to other COUP-TFs from vertebrates or arthropods. This high level of sequence conservation (in the range of 90 %) was already pointed out by numerous authors. This is also true outside protostomes and deuterostomes since this gene is also extremely well conserved in the cnidarian *Hydra vulgaris* (M.C.L., V.L. and B. Galliot, in preparation). There is no specific functional features of COUP-TFs that may explain such a conservation but this nevertheless suggests that the functions of the COUP-TF genes are ancient and vital for metazoans.

The phylogenetical position of COUP-TF allow us to suggest that the AmphiCOUP-TF gene should be viewed as a unique *bona fide* COUP-TF gene in amphioxus except if there was, specifically in the cephalochordates, an independant gene duplication giving rise to two closely related genes. The same reasoning led us to conclude that AmphiTR2/4 should be viewed as a unique homologue of TR and TR4 genes of vertebrates.

The study of the COUP-TF family tree reveal an striking placement of xenopus COUPG and *Danio* SVP46 which clustered together and splitted before the divergence leading to drosophila SVP and the deuterostomians COUP-TF. This allowed us to suggest that a homologue of this gene should be found in drosophila as well as in mammals. The fact that such a homologue remains undescribed despite an apparent strong conservation suggests that it may have been lost by a secondary event. We cannot exclude the existence of such a distantly related member of the COUP-TF group in amphioxus. Furthermore, and in contrast to a recent proposal (Tsai and Tsai, 1997) we suggest that xenopus COUPG and *Danio* SVP46 should be viewed as orthologues respectively in xenopus and zebrafish. In fact if these genes were not true orthologues this may imply that the genome of xenopus and *Danio* should each contain a COUPG and a SVP46 genes and that these two genes have been lost in all other species. It seems more parsimonious to conclude that, even if the sequences of xenopus COUPG and *Danio* SVP46 have diverged more rapidly than the sequences of the other COUP genes for an unknown reason, these two genes are orthologous.

The study of the sequence of AmphiTR2/4 reveal an interesting feature. This sequence exhibit a deletion of 11 amino acids in the T/A box which is responsible of the dimerisation of some nuclear receptors (such as RXR) and of the DNA binding to half-site sequences of monomer-binding orphan receptors such as Rev-erb. Since this box is apparently lacking in AmphiTR2/4 and despite a lack of functional characterization of this region in the TR2 and TR4 genes of vertebrates it is tempting to link the presence of this deletion with the poor DNA binding activity of the AmphiTR2/4 gene product. This remain to be experimentally adressed. This deletion lie in the precise position in which an intron is present in the mouse TR2 gene (Lee et al., 1995). This observation allow us to suggest that it may result from an mis-splicing event that result in the skipping of the 33 first base pairs of the 3' exon. This suggest that despite our negative RT-PCR result a form containing a complete TA box region should exist in AmphiTR2/4.

#### **AmphiCOUP-TF as a negative regulator of retinoid signalling.**

Our functional data clearly show that AmphiCOUP-TF is a negative regulator of retinoid signalling. The results presented in this paper demonstrate that the expression of AmphiCOUP-TF repress the signalling mediated by the RXR/RAR heterodimer acting on synthetic DR5 element in transient transfection analysis. Similar results were obtained on the RXR mediated signalling through a DR1 element. We have nevertheless to recall that these results were obtained using mouse versions of RXR and RAR and not the amphioxus homologues. This will wait the characterization of homologues of AmphiRAR and AmphiRXR cDNA currently underway in our laboratory (H. Escriva and V.L., in preparation).

Amphi-COUP-TF is expressed late in development, in fact only at the 4-5 gill slit stage. Nevertheless an early effect of retinoids was already observed during amphioxus development (Holland and Holland, 1996). Since early RA treatment result in mortality of embryos we cannot tested the effects of RA on COUP-TF expression during late development and thus we cannot exclude that RA regulates AmphiCOUP-TF expression. It is nevertheless clear that RA signalling occuring during early embryogenesis is not inhibited by AmphiCOUP-

TF. The *in vivo* significance of the observed inhibition of RXR-RAR-mediated transactivation by AmphiCOUP-TF thus remain to be firmly established. It is envisageable that during early embryogenesis other orphan members of the nuclear receptor superfamily play this role. AmphiTR2/4 is a good candidate for such a function.

In vertebrates COUP-TFs were not described only as negative regulators of retinoid signalling. In fact COUP-TF was shown to be antagonist of a large number of other nuclear receptor mediated pathways such as RXR-TR, RXR-PPAR, RXR-VDR, HNF4 etc... The lack of co-expression of COUP-TF and RAR may thus also indicate that *in vivo* the function of COUP-TF may be to antagonize another signalling pathway. For example, it is possible that thyroid hormones regulates amphioxus metamorphosis as in vertebrates. Since AmphiCOUP-TF is expressed late in embryogenesis it may be that it is required to antagonise thyroid hormone signalling. Thus, we believe that AmphiCOUP-TF should not be viewed only as a regulator of retinoid signalling but rather as a general inhibitor of developmental pathways governed by other nuclear receptors. Such a model has important assumptions since we recently proposed that classical liganded nuclear receptors were evolutionary novelties since they are known only in vertebrates and apparently not conserved in protostomes (Escriva and others, 1997; Laudet, 1997). The fact that the basic function of COUP-TF even before the emergence of vertebrates, was to inhibit the action of liganded receptors suggest either (i) that other homologues liganded receptors such as TR or PPAR should be found in amphioxus and/or (ii) that liganded receptors evolved from orphan receptors that were negatively regulated by COUP-TFs and that this negative regulation of their activity was conserved. The fact that indeed a PPAR gene exist in amphioxus (H. Escriva and V.L. unpublished result) is in favor of this model.

#### ACKNOWLEDGMENTS

We thanks CNRS, ENS de Lyon, MENRT, ARC and NSF (grant NSF IBN96-309938 to L.Z.H. and N.D.H) for financial support. M.C.L. was supported by ARC, Institut Pasteur de Lille and Région Nord Pas de Calais. H.E. is supported by an EMBO fellowship. We thank Jean-Marc Vanacker for critical reading of the manuscript and Agnes Bègue for excellent technical assistance.

## References

- Chan, S. M., N. Xu, C. C. Niemeyer, J. R. Bone, and C. N. Flytzanis. "SpCOUP-TF: a sea urchin member of the steroid/thyroid hormone receptor family." *Proc Natl Acad Sci U S A* **89** (22 1992): 10568-72.
- Chang, C. and J. Kokontis. "Identification of a new member of the steroid receptor super-family by cloning and sequence analysis." *Biochem Biophys Res Commun* **155** (2 1988): 971-7.
- Chang, L. and L. N. Wei. "Characterization of a negative response DNA element in the upstream region of the cellular retinoic acid-binding protein-I gene of the mouse." *J Biol Chem* **272** (15 1997): 10144-50.
- Chinpaisal, C., L. Chang, X. Hu, C. H. Lee, W. N. Wen, and L. N. Wei. "The orphan nuclear receptor TR2 suppresses a DR4 hormone response element of the mouse CRABP-I gene promoter." *Biochemistry* **36** (46 1997): 14088-95.
- Cooney, A. J., S. Y. Tsai, B. W. O'Malley, and M. J. Tsai. "Chicken ovalbumin upstream promoter transcription factor binds to a negative regulatory region in the human immunodeficiency virus type 1 long terminal repeat." *J Virol* **65** (6 1991): 2853-60.
- Cooney, A. J., S. Y. Tsai, B. W. O'Malley, and M. J. Tsai. "Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTC A response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors." *Mol Cell Biol* **12** (9 1992): 4153-63.
- Enmark, E. and J. A. Gustafsson. "Orphan nuclear receptors--the first eight years." *Mol Endocrinol* **10** (11 1996): 1293-307.
- Escriva, H., R. Safi, C. Hanni, M. C. Langlois, Laprade P. Saumitou, D. Stehelin, A. Capron, R. Pierce, and V. Laudet. "Ligand binding was acquired during evolution of nuclear receptors." *Proc Natl Acad Sci U S A* **94** (13 1997): 6803-8.
- Fisk, G. J. and C. S. Thummel. "Isolation, regulation, and DNA-binding properties of three Drosophila nuclear hormone receptor superfamily members." *Proc Natl Acad Sci U S A* **92** (23 1995): 10604-8.
- Fisk, G. J. and C. S. Thummel. "The DHR78 nuclear receptor is required for ecdysteroid signaling during the onset of Drosophila metamorphosis." *Cell* **93** (4 1998): 543-55.
- Fjose, A., S. Nornes, U. Weber, and M. Mlodzik. "Functional conservation of vertebrate seven-up related genes in neurogenesis and eye development." *Embo J* **12** (4 1993): 1403-14.
- Fjose, A., U. Weber, and M. Mlodzik. "A novel vertebrate svp-related nuclear receptor is expressed as a step gradient in developing rhombomeres and is affected by retinoic acid." *Mech Dev* **52** (2-3 1995): 233-46.
- Garcia Fernandez, J. and P. W. Holland. "Archetypal organization of the amphioxus Hox gene cluster." *Nature* **370** (6490 1994): 563-6.
- Graupner, G., K. N. Wills, M. Tzukerman, X. K. Zhang, and M. Pfahl. "Dual regulatory role for thyroid-hormone receptors allows control of retinoic-acid receptor activity." *Nature* **340** (6235 1989): 653-6.
- Gronemeyer, H. and V. Laudet. "Transcription factors 3: nuclear receptors." *Protein Profile* **2** (11 1995): 1173-308.



- Guo, B., F. Aslam, Wijnen Aj van, S. G. Roberts, B. Frenkel, M. R. Green, H. DeLuca, J. B. Lian, G. S. Stein, and J. L. Stein. "YY1 regulates vitamin D receptor/retinoid X receptor mediated transactivation of the vitamin D responsive osteocalcin gene." *Proc Natl Acad Sci U S A* **94** (1 1997): 121-6.
- Guo, X., Q. Xu, M.A. Harmon, X. Jin, V. Laudet, D.J. Mangelsdorf, and M.J. Palmer. "Isolation of two functional retinoid X receptor subtypes from the ixodid tick, *Amblyomma americanum* (L.)." *Mol Cell Endocrinol* **139** (1998): 45-60.
- Hall, R. K., F. M. Sladek, and D. K. Granner. "The orphan receptors COUP-TF and HNF-4 serve as accessory factors required for induction of phosphoenolpyruvate carboxykinase gene transcription by glucocorticoids." *Proc Natl Acad Sci U S A* **92** (2 1995): 412-6.
- Hirose, T., W. Fujimoto, T. Tamaai, K. H. Kim, H. Matsuura, and A. M. Jetten. "TAK1: molecular cloning and characterization of a new member of the nuclear receptor superfamily." *Mol Endocrinol* **8** (12 1994): 1667-80.
- Hirose, T., D. A. O'Brien, and A. M. Jetten. "Cloning of the gene encoding the murine orphan receptor TAK1 and cell- type-specific expression in testis." *Gene* **163** (2 1995): 239-42.
- Holland, L. Z. and N. D. Holland. "Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx." *Development* **122** (6 1996): 1829-38.
- Holland, L. Z., T. V. Venkatesh, A. Gorlin, R. Bodmer, and N. D. Holland. "Characterization and developmental expression of *AmphiNk2-2*, an NK2 class homeobox gene from Amphioxus. (Phylum Chordata; Subphylum Cephalochordata)." *Dev Genes Evol* **208** (2 1998): 100-5.
- Holland, N. D., G. Panganiban, E. L. Henyey, and L. Z. Holland. "Sequence and developmental expression of *AmphiDll*, an amphioxus *Distal-less* gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest." *Development* **122** (9 1996): 2911-20.
- Holland, P. W. and Fernandez J. Garcia. "Hox genes and chordate evolution." *Dev Biol* **173** (2 1996): 382-95.
- Hwung, Y. P., L. H. Wang, S. Y. Tsai, and M. J. Tsai. "Differential binding of the chicken ovalbumin upstream promoter (COUP) transcription factor to two different promoters." *J Biol Chem* **263** (26 1988): 13470-4.
- Katsuyama, Y., S. Wada, S. Yasugi, and H. Saiga. "Expression of the labial group Hox gene *HrHox-1* and its alteration induced by retinoic acid in development of the ascidian *Halocynthia roretzi*." *Development* **121** (10 1995): 3197-205.
- Kontrogianni, Konstantopoulos A., A. Vlahou, D. Vu, and C. N. Flytzanis. "A novel sea urchin nuclear receptor encoded by alternatively spliced maternal RNAs." *Dev Biol* **177** (2 1996): 371-82.
- Kozak, M. "Interpreting cDNA sequences: some insights from studies on translation." *Mamm Genome* **7** (8 1996): 563-74.
- Ladias, J. A. and S. K. Karathanasis. "Regulation of the apolipoprotein AI gene by ARP-1, a novel member of the steroid receptor superfamily." *Science* **251** (4993 1991): 561-5.

- Laudet, V. "Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor." *J Mol Endocrinol* **19** (3 1997): 207-26.
- Lee, C. H., L. Chang, and L. N. Wei. "Distinct expression patterns and biological activities of two isoforms of the mouse orphan receptor TR2." *J Endocrinol* **152** (2 1997): 245-55.
- Lee, C. H., N. G. Copeland, D. J. Gilbert, N. A. Jenkins, and L. N. Wei. "Genomic structure, promoter identification, and chromosomal mapping of a mouse nuclear orphan receptor expressed in embryos and adult testes." *Genomics* **30** (1 1995): 46-52.
- Lee, H. Y., M. I. Dawson, G. L. Walsh, J. C. Nesbitt, R. L. Eckert, E. Fuchs, W. K. Hong, R. Lotan, and J. M. Kurie. "Retinoic acid receptor- and retinoid X receptor-selective retinoids activate signaling pathways that converge on AP-1 and inhibit squamous differentiation in human bronchial epithelial cells." *Cell Growth Differ* **7** (8 1996): 997-1004.
- Lin, K. H., Y. W. Lin, H. F. Lee, W. L. Liu, S. T. Chen, K. S. Chang, and S. Y. Cheng. "Increased invasive activity of human hepatocellular carcinoma cells is associated with an overexpression of thyroid hormone beta 1 nuclear receptor and low expression of the anti-metastatic nm23 gene." *Cancer Lett* **98** (1 1995a): 89-95.
- Lin, T. M., W. J. Young, and C. Chang. "Multiple functions of the TR2-11 orphan receptor in modulating activation of two key cis-acting elements involved in the retinoic acid signal transduction system." *J Biol Chem* **270** (50 1995b): 30121-8.
- Lopes da Silva, S., J. J. Cox, L. J. Jonk, W. Kruijer, and J. P. Burbach. "Localization of transcripts of the related nuclear orphan receptors COUP-TF I and ARP-1 in the adult mouse brain." *Brain Res Mol Brain Res* **30** (1 1995): 131-6.
- Lutz, B., S. Kuratani, A. J. Cooney, S. Wawersik, S. Y. Tsai, G. Eichele, and M. J. Tsai. "Developmental regulation of the orphan receptor COUP-TF II gene in spinal motor neurons." *Development* **120** (1 1994): 25-36.
- Malik, S. and S. Karathanasis. "Transcriptional activation by the orphan nuclear receptor ARP-1." *Nucleic Acids Res* **23** (9 1995): 1536-43.
- Mangelsdorf, D. J., U. Borgmeyer, R. A. Heyman, J. Y. Zhou, E. S. Ong, A. E. Oro, A. Kakizuka, and R. M. Evans. "Characterization of three RXR genes that mediate the action of 9-cis retinoic acid." *Genes Dev* **6** (3 1992): 329-44.
- Mangelsdorf, D. J., C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, and a. l. et. "The nuclear receptor superfamily: the second decade." *Cell* **83** (6 1995): 835-9.
- Matharu, P. J. and G. E. Sweeney. "Cloning and sequencing of a COUP transcription factor gene expressed in *Xenopus* embryos." *Biochim Biophys Acta* **1129** (3 1992): 331-4.
- Miyajima, N., Y. Kadowaki, S. Fukushima, S. Shimizu, K. Semba, Y. Yamanashi, K. Matsubara, K. Toyoshima, and T. Yamamoto. "Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other." *Nucleic Acids Res* **16** (23 1988): 11057-74.

- Mlodzik, M., Y. Hiromi, U. Weber, C. S. Goodman, and G. M. Rubin. "The Drosophila seven-up gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fates." *Cell* **60** (2 1990): 211-24.
- Neuman, K., A. Soosaar, H. O. Nornes, and T. Neuman. "Orphan receptor COUP-TF I antagonizes retinoic acid-induced neuronal differentiation." *J Neurosci Res* **41** (1 1995): 39-48.
- Power, S. C. and S. Cereghini. "Positive regulation of the vHNF1 promoter by the orphan receptors COUP-TF1/Ear3 and COUP-TFII/Arp1." *Mol Cell Biol* **16** (3 1996): 778-91.
- Qiu, Y., A. J. Cooney, S. Kuratani, F. J. DeMayo, S. Y. Tsai, and M. J. Tsai. "Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon." *Proc Natl Acad Sci U S A* **91** (10 1994): 4451-5.
- Schuh, T. J. and D. Kimelman. "COUP-TFI is a potential regulator of retinoic acid-modulated development in Xenopus embryos." *Mech Dev* **51** (1 1995): 39-49.
- Shibata, H., Z. Nawaz, S. Y. Tsai, B. W. O'Malley, and M. J. Tsai. "Gene silencing by chicken ovalbumin upstream promoter-transcription factor I (COUP-TFI) is mediated by transcriptional corepressors, nuclear receptor-corepressor (N-CoR) and silencing mediator for retinoic acid receptor and thyroid hormone receptor (SMRT)." *Mol Endocrinol* **11** (6 1997): 714-24.
- Tsai, S. Y. and M. J. Tsai. "Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): coming of age." *Endocr Rev* **18** (2 1997): 229-40.
- van der Wees, J., P. J. Matharu, Roos K. de, O. H. Destree, S. F. Godsave, A. J. Durston, and G. E. Sweeney. "Developmental expression and differential regulation by retinoic acid of Xenopus COUP-TF-A and COUP-TF-B." *Mech Dev* **54** (2 1996): 173-84.
- Wang, L. H., S. Y. Tsai, R. G. Cook, W. G. Beattie, M. J. Tsai, and B. W. O'Malley. "COUP transcription factor is a member of the steroid receptor superfamily." *Nature* **340** (6229 1989): 163-6.
- Yan, C., M. Ghaffari, J. A. Whitsett, X. Zeng, Z. Sever, and S. Lin. "Retinoic acid-receptor activation of SP-B gene transcription in respiratory epithelial cells." *Am J Physiol* (1998): L239-46.

## FIGURE LEGENDS

FIGURE 1 : (A) Sequence of the AmphiCOUP-TF cDNA. Numbering of nucleotides and amino acids are indicated. Sequence of the C domain is between amino acid 106 to 171 (underlined) whereas the E domain starts at 205 and finishes at the end of the protein at position 443 (italicized). The putative polyadenylation site is underlined ; (B) Alignment of known members of the COUP-TF group including AmphiCOUP-TF (COUPBF). Human orphan receptor EAR2, which is the closest relative to the COUP group, was added for comparison. Alignment of the first 15 amino acids is shown to illustrate the conservation of the most N-terminal amino acids. Sequence of the A/B domain is not indicated because too weakly conserved. Sequence of the C domain of human COUP-TFI is underlined (positions 141-206). Sequence of the E domain of COUP-TFs starts at position 261 of the alignment. Sequences are as follow with their genbank codes : COUPAH : human COUP-TFI (X12795), COUPAM : mouse COUP-TFI (U07625), COUPAX : xenopus COUP-TFI (this study), SVP44 : zebrafish COUP-TFI (X70299), COUPBH : human COUP-TFII (M64497), COUPBM : mouse COUP-TFII (U07635), COUPBCK : chicken COUP-TFII (U00697), COUPBX : xenopus COUP-TFII (this study), SVP40 : zebrafish COUP-TFII (S80986), COUPGX : xenopus COUP-TFG (X63092), SVP46 : zebrafish COUP-TFG (X70300), COUPBF : amphioxus COUP-TF (this study), COUPSU : sea urchin COUP-TFII (L01104 and L01105), this sequence is incomplete in C-terminal, SVPD : drosophila SVP (M28863), the N-terminal sequence is not conserved in this sequence, EAR2H : human EAR2 (X12794).

FIGURE 2 : (A) Sequence of the AmphiTR2/4 cDNA. Numbering of nucleotides and amino acids are indicated. Sequence of the C domain is between amino acid 32 to 87 (underlined) whereas the E domain starts at 256 and finishes at the end of the protein at position 499 (italicized) ; (B) Alignment of known members of the TR2/4 group including AmphiCOUP-TF (TR2/4BF). Sequence of the A/B domain is not indicated because it is too weakly conserved. Sequence of the C domain of human TR2 is underlined (positions 3-74). Gaps are indicated with dashes. The gap inside the T/A box of AmphiTR2/4 (position 80-91) is visible. The alignment of the E domain reveals the good conservation of this domain in Amphi TR2/4 and its strong divergence in drosophila DHR78. Sequences are as follow with their genbank codes : TR2H : human TR2 (M29959), TR2M : mouse TR2 (U28265), TR4H : human TR4 (L27586), TR4M : mouse TR4 (S75970), TR4R : rat TR4 (L27513), TR2/4BF : amphioxus TR2/4 (this study), TR2/4SU : sea urchin SpSHR2 (U38528), DHR78 : drosophila DHR78 (U31517).

FIGURE 3 : Phylogenetical trees showing the placement of AmphiCOUP-TF (A) and AmphiTR2/4 (B) in their respective groups. The two amphioxus sequences are indicated by arrows. Trees were constructed from the alignments presented Fig. 1 and 2 by the distance Neighbor-Joining method with 1000 bootstrap replicates in order to test the robustness of the branches. The bootstrap values are indicated on each branch of the trees. For each tree the root was placed by using an outgroup which

contains at least two nuclear receptor sequences of the same sub-family as well as two sequences from other subfamilies (Laudet, 1997).

FIGURE 4 : Whole amount *in situ* Hybridization with Amphi-COUP-TF probe.

A. 2,5 j post-metamorphosis larvae. B. The expression of Amphi-COUP-TF is localized in the neural cord in a segmented manner. C-E. Different views along the antero-posterior axis of the neural cord. F. Transversal slide of the amphioxus larvae body showing the Amphi-COUP-TF expression restricted to the neural cord.

FIGURE 5 : DNA binding properties of AmphiCOUP-TF (A and B) and AmphiTR2/4 (C) gene products studied by EMSA. Protein produced in rabbit reticulocytes lysates were incubated with labelled oligonucleotide probes and submitted to electrophoresis. The same quantity of protein was used by experiment, previously evaluated by phosphorimager analysis. (A) Binding of AmphiCOUP-TF on the DR1 probe (lines 2 to 4). Mouse COUP-TFI (COUP-I) and COUP-TFII (COUP-II) were added to the experiment as references, see lines 8-10 and lines 5-7 respectively. In each case the probe was added alone (-) or a 100 fold molar excess of either cold DR1 specific (S) or cold unrelated non specific (NS) competitors were added. (B) Competition experiments showing the wide specificity of AmphiCOUP-TF binding. After binding of the COUP-TF protein to the DR1 probe (lane 2), this complex was competed out by 1, 10 or 100 fold excess of cold oligonucleotides encompassing the various direct repeat elements from DR0 to DR5. (C) Binding of AmphiTR2/4 to the DR1 probe compared to the complex formed with mouse TR4. In each case the probe was added alone (-) or a 100 fold molar excess of either cold DR1 specific (S) or cold unrelated non specific (NS) competitors were added. R : unprogrammed reticulocyte lysate.

FIGURE 6 : AmphiCOUP-TF represses RA mediated signalling. (A) COS-1 cells were transiently transfected with 100 ng of either pSG5 vector alone (pSG5), pSG5-RAR (RAR), pSG5-RXR (RXR) or a combination of both (RAR + RXR) and a reporter DR5tkCAT vector. In each case cells were cultivated in stripped medium (empty boxes) or stripped medium supplemented with  $10^{-6}$  M all-*trans* RA (stippled boxes) or  $10^{-6}$  M 9-*cis* RA (filled boxes). pSG5-Amphi-COUP-TF (200 ng) was added where indicated (+ Amphi-COUP-TF). (B) Dose-dependant effect of Amphi-COUP-TF on the repression of RAR + RXR signalling. COS-1 cells transfected with 200 ng of pSG5-RXR, pSG5-RAR and the DR1tkCAT vector were treated by  $10^{-6}$  M all-*trans* RA. increasing amounts of pSG5-Amphi-COUP-TF expression vector were added : 0.4 ng (lane 2), 16 ng (lane 3), 40 ng (lane 4), 100 ng (lane 5), 250 ng (lane 6), 625 ng (lane 7) and 1.5 mg (lane 8).

**A**

	M	A	M	A	V	S	T	W	R	E	10
cgggttcagtgtagccccgcgcgtgtaat	ATG	GCG	ATG	GCA	GTC	AGT	ACA	TGG	CGA	GAA	58
P G E D L S S P D K Q Q Q Q P V P 28											
CCG GGT GAA GAC CTC TCG AGC CCG GAC AAG CAG CAG CAG CAG CAG CCG GTA CCC 112											
A P V P G P G P Q Q P P P S Q P P G 46											
GCA CCC GTC CCG GGA CCA GGG CCG CAG CAG CCG CCG CCG TCG CAG CCC CCA GGG 166											
Q T Q Q S T P G P Q Q P G P P T P N 64											
CAG ACG CAG CAG TCG ACC CCG GGT CCG CAG CAG CCA GGC CCG CCG ACC CCG AAC 220											
Q Q T P N G G S M P Q Q T P T A P T 82											
CAG CAG ACG CCG AAC GGC GGG TCG ATG CCG CAG CAG ACG CCG ACG GCC CCG ACG 274											
Q P S S Q Q G Q Q Q S T P T S G G Q 100											
CAG CCG TCC TCG CAG CAG GGC CAA CAA CAG TCC ACC CCG ACT TCG GGA GGC CAG 328											
S Q H I E C V V C G D K S S G K H Y 118											
AGC CAG CAC ATC GAG TGC GTC GTG TGC GGG GAC AAG AGT TCC GGT AAG CAC TAT 382											
G Q F T C E G C K S F F K R S V R R 136											
GGC CAG TTC ACA TGC GAG GGT TGC AAG AGT TTC TTC AAG AGA AGC GGT CGT CCG 436											
N L T Y S C R G N R T C P I D O H H 154											
AAC CTG ACC TAC TCC TGC CCG GGC AAC CGC ACC TGC CCC ATC GAC CAA CAC CAC 490											
R N Q C Q Y C R L K K C L K M G M R 172											
CGG AAC CAG TGT CAG TAC TGC CCG CTC AAG AAG TGT CTG AAG ATG GGC ATG AGA 544											
R E A V Q R G R I P P T Q H P A G A 190											
CGA GAA GCG GGT CAA CCG GGC CCG ATC CCC CCG ACC CAG CAC CCA GCC GGT GCC 598											
Q Y A L T N G V D P I N G H S Y L S 208											
CAG TAC GCC CTG ACC AAC GGA GTA GAC CCG ATC AAC GGC CAC TCG TAC TTG TCC 652											
G Y I S L L L R A E P Y P T S R Y G 226											
GGG TAC ATC TCT CTC CTG CTA CCG GCT GAA CCC TAC CCA ACC TCG CGC TAT GGC 706											
T Q C M Q P N N I M G I D N I C E L 244											
ACA CAG TGT ATG CAG CCC AAC AAC ATT ATG GGC ATT GAC AAC ATC TGC GAG CTC 760											
A A R L L F S A V E W A R N I P Y F 262											
GCG GCT CGC CTG CTC TTC AGT GCC GTG GAA TGG GCC AGG AAT ATT CCC TAC TTC 814											
P E L Q V T D Q V A L L R L V W S E 280											
CCC GAG TTA CAA GTC ACC GAC CAA GTG GCG CTG TTG CGC CTG GTG TGG AGT GAG 868											
L F V L N A S Q C S M P L H V A P L 298											
CTA TTC GTG CTG AAC GCC TCA CAG TGC TCC ATG CCC CTT CAC GTC GCC CCA TTG 922											
L A A A G L H A S P M S A D R V V A 316											
CTT GCC GCC GCT GGC CTG CAT GCC AGC CCC ATG TCA GCC GAC CGA GTC GTG GCG 976											
F M D H I R I F Q E Q V E K L K A L 334											
TTC ATG GAT CAC ATA CGC ATA TTC CAG GAG CAA GTA GAG AAA CTG AAA GCA TTG 1030											
H V D S A E Y S C L K A I V L F T S 352											
CAC GTG GAT TCT GCA GAG TAC AGC TGT CTG AAG GCG ATC GTG CTC TTC ACC TCA 1084											
D A C G L S D A A H I D S L Q E K S 370											
GAT GCG TGC GGC CTG TCG GAC GCG GCT CAG AAT GAC TCC CTG CAG GAG AAG TCG 1138											
Q C A L E E Y V R S Q Y P N Q P T R 388											
CAG TGC GCG CTC GAG GAG TAC GTC CGC TCC CAG TAC CCC AAC CAA CCC ACG CCG 1192											
F G K L L L R L P S L R T V S S S V 406											
TTC GGC AAG CTG CTG CTG CGC CTG CCG TCG CTC CGC ACC GTC TCC TCT TCC GTC 1246											
I E Q L F F V R L V G K T P I E T L 424											
ATA GAG CAG TTG TTT TTC GTC CGC CTG GTA GGT AAG ACT CCC ATC GAA ACC TTA 1300											
I R D M L L S G S S F N W P Y M S I 442											
ATC CGA GAT ATG CTC TTG TCT GGA AGC TCC TTT AAC TGG CCA TAC ATG TCC ATC 1354											
Q *											443
CAG TGA gtgacttattgaaacacagtcctgcgacatagcgtcgaacggacatcttctgaacacacac 1423											
ttcagatggatttgcaggactctaaccgacgaagtaccgtagtagataaacgacttgccttgagcaca 1494											
gcgttagtcctgtactactctactccttttgtaaaaaggatcaaaaaagaaaagtgttattttctctgttt 1565											
ttcttttggtttcggattttattgataccagatgaaatgttccacttttgaaaaagctaagcctttgtcaatt 1636											
ggtaaacctataacagactaggggaattttaacaatttcgacaaaatgccaattatgtgtctatgtatct 1707											
tgctgttttatatacatggttccaaaccaaacttaagtgttactaataataattttgatataaccttg 1778											
ctctaagtgtatgctatcttgcgtatgtaaaaacttgtaaaggaaattttttgaaggaaaagaaatacgga 1849											
cggaaagaagctcaaaaaaaaaaaaaaaaaaaaaa 1882											



A

```

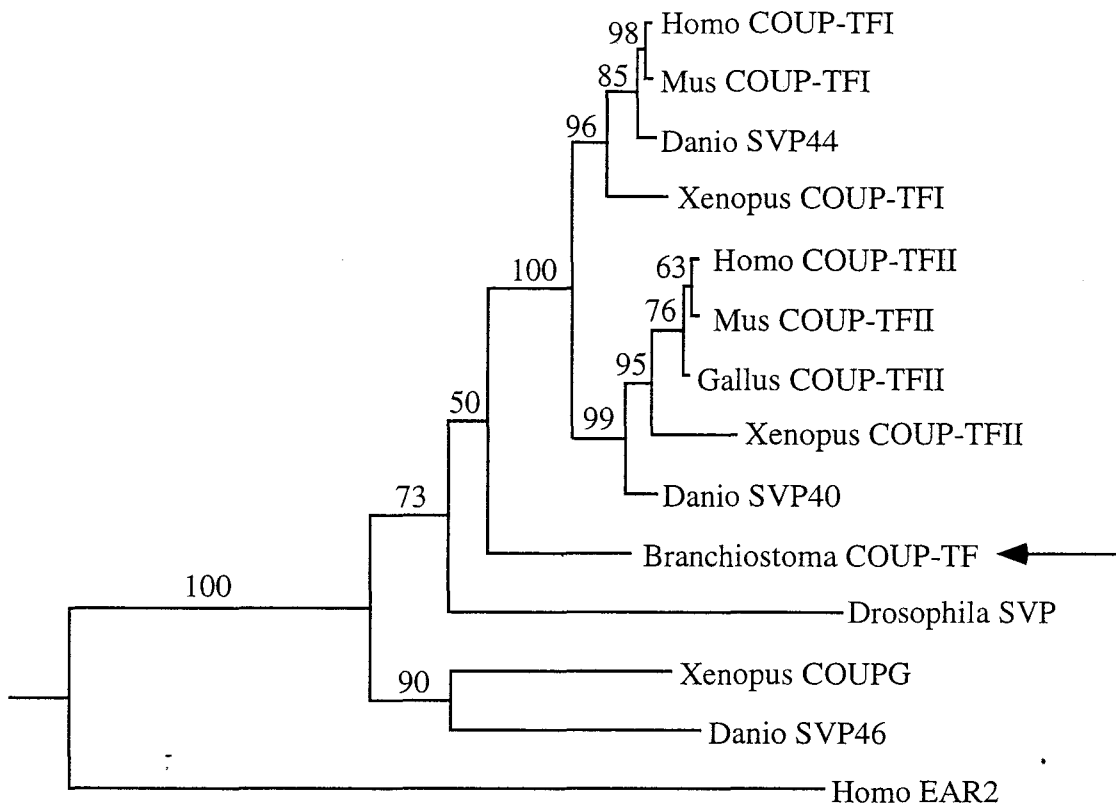
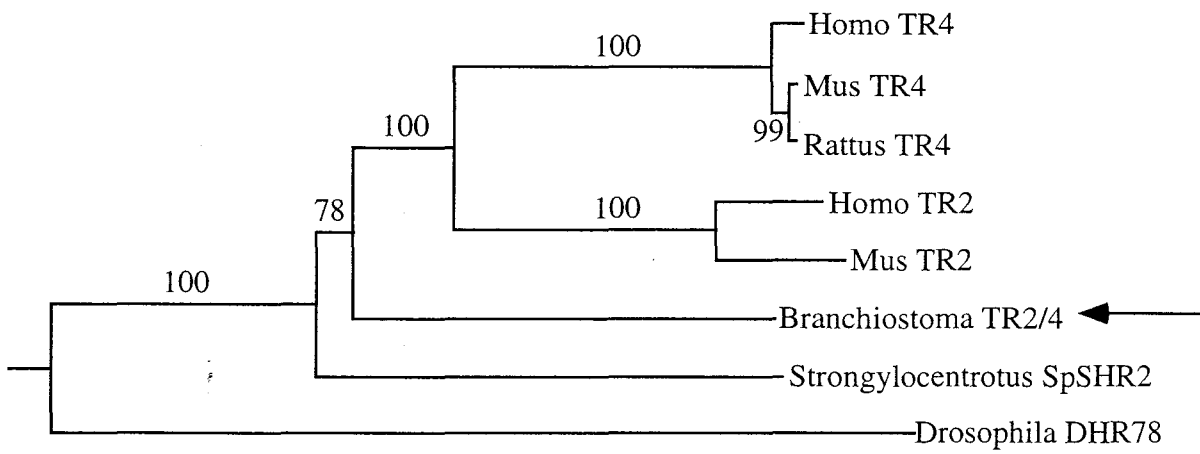
cggcccatagggcaccgcccgccgctc ATG ATG TCC CCC AAC AGT GTG CTG TTC ACA CCG GGC TCC GGC GAC ACC AAA CCA CAG GAG CTA 90
M M S P N S V L F T P G S G D T K P Q E L 21
TGT GTG GTG TGT GGG GAC AAG GCA TCA GGG CGG CAC TAT GGC GCA ATA AGT TGC GAG GGC TGT AAG GGC TTC TTC AAG CGC AGC ATT CGG 180
C V V C G D K A S G R H Y G A I S C E G C K G F F K R S I R 51
AAG AAC CTG GGC TAT GTG TGC CGG AGC AGC AAG GAG TGC CCC ATC AAC AAG CAC CAC CGG AAC CGC TGC CAG TAC TGC CGG CTG CAG AAG 270
K N L G Y V C R S S K E C P I N K H H R N R C O Y C R L O K 81
TGC CTG TCT GTC GGC ATG AGG TCG GAG TCT CGC GAG CGC CCG AAC TGC GCC ACC TCC ACC GAG AAA ATC TAC ATC CGC AAG GAC CTA 360
C L S V G M R S E S R E R P P N C A T S T E K I Y I R K D L 111
CGC AGC CCG CTG GCC ACC ACA CCC ACG TTC GTG ACG GAG AGA GAA CCC AAG TCG GGA GGG CTG TTC GAT GCC GGC ATG CTG GCA AAC ATA 450
R S P L A A T P T F V T E R E P K S G G G L F D A G M L A N I 141
CAG CAA GGT GAC AGC AAT GCA CCC ATA CTG CTT GCC GTG GAC CCT ACC AAG TCA GAG GGG AAG GAG GGG GAC GGT AAC ACG GAC CTG AGC 540
Q Q G D S N A P I L L A V D P T K S D G K E G D G N T D L S 171
ACG CTG GCG AAC GTG GTG ACG ACC CTG GCG AGC ATG AGT AAG AAC AAG GAA GGC CAG GCG CAG AAC GGA GGG GAG GGG GAC GCG GCG 630
T L A N V V T P L A S M S K N K E G Q A Q Q N G G E G D A A 201
ATC AGC GGT GAC GGG ACC GTG ATG GAG GTT CAG ACG TCC ACG GAC ATC TCC AAG GCC TTT GAT AAC CTC ACC AAG GCC CTC AAC ACA TCT 720
I S G D G T V M E V Q T S T D I S K A F D N L T K A L N T S 231
CAA ACA GGT GAA CAA GGC ATC ACA TCC GAC GGC TCC CAG GGG ATG GAG GGG ACG TCA GAC CAG GAC CAG ACC ACC GTG ATC GAG CTG GAG 810
Q T G E Q G I T S D G S Q G M E G T S D Q D Q T T V I E L E 261
GGT CCG CTG CTC ACG GAC GCG CAC ACG GCC TTC AAG CTC ACC ACG CCC TCG CCC ATG CCG GAG TTC CTC AAC GTC CAC TAC ATC TGC GAG 900
G P L L L T D A H T A F K L T T P S P M P E F L N V H Y I C E 291
TCT GCG TCA CGA CTG CTG TTC CTC ACC ATG CAC TGG GCA CGC TCC ATC CCG GCC TTC CAG GCC CTC GGA CAA GAG TGC CAT ACC GGG CTG 990
S A S R L L L F L T M H W A R S I P A F Q A L G Q E C H T G L 321
GTG CGG GTG TGT TGG AGT GAG CTG TTC ACC CTG GGC CTG AGC CAG TGT TCG GCT ATC TCC CTC CCA ACC ACC GCT ACC CTC ACC GCT ATC GTC 1080
V R V C W S E L F T L G L S Q C S A V M S L P T I L T A I V 351
AAC CAC CTG GAC TCC AGC GTA CAG GCA GAC AAG TIG AGT GCA GAC AGA GTG AAG GTA GTG ATG GAG CAC ATC TGG AAG CTA CAG GAG TTC 1170
N H L Q S S V Q A D K L S A D R V K V V M E H I W K L Q E F 381
ATC AAC AAG ATG GCC GCC CTG ACC GTG GAC CCT ACA GAG TAT GCC TAC ATG AAG GCC ATA GTT CTC TTC AGC ACA GAC CAC CCT GGC CTG 1260
I N K M A A L T V D P T E Y A Y M K A I V L F S T D H P G L 411
CTG AAC CCT CGG CAG ATC GAG AAG TTC CAG GAG AAG GCC CAC GAA GAG CTG AAG GAG TAC GAA GCT CAA GCG CAT CCT GAT GAC ACA GAA 1350
L N P R Q I E K F Q E K A H E E L K E Y E A Q A H P D D T E 441
AGG TTT GCG AAG CTT CTC CTA CGG CTA CCT GCC CTG CGC CTG CTA TCC CCG TCC ATC ATG GAG GAG CTG TTC TTC GCG GGA CTC ATC GGC 1440
R F A K L L L R L P A L R L L S P S I M E E L F F A G L I G 471
AAC GTC CAG ATC GAC AGC ATC ATT CCG TAC ATC CTT ATG GAG ACG GAC TAC AAC AGC CAG ATC GGG GGG ACC TCG CAG TAG agc 1530
N V Q I D S I I P Y I L R M E T A D Y N S Q I G G T S Q * 499
gcccagtcgtagcagcgtccattcgtaaaacaacacagcacaacatgttttctcccctgttttaaatagtagcacacattctgtacattgttgtgtgagcactccactcatagaacagaacatg 1648
tctggaaaagggtagacatttgaagagacctcccagagatcagctcgtgtgtaaaattttcccctcgtgtgtgtgtagtcatctggtttgggtgtgtcctcctcgtcgtggttggacatt 1766
aatgtgtcttctaatgtgtgtagctgtgacggtaggagttcattttagggcacaagtcacacctatcagcaaatatgaagaatgaatgtaaagtaggacaaaaagattcatttgg 1884
agatgaccacgactattttagcagcagcacaataacctataactctgatttgaagacgattgtctgtgtctacaaataagaatgtgaagggatctaggttgaagttttgacatttgg 2002
tacttacaatcgacaacctcacaagtagctgtatcagtggtgctttcctgtgtgatttaatagcagagatcccgcgaagcggcgcctcagcccgacgctggtctactcctgatt 2120
atccgccatgtaaatagagcagcactacaattgaccaagtcacaaatcttctgtgggatgtagggtgaagattctgtacatgtatctatccaccagtggtattccaatgggacctcatg 2238
tgatgtagttagacctcagagggcttgtggtaagagaatcagtcggactgttctagagtggtttccgatcagtagacgataatgtttcaagcaacgataatgtgacctgtaccata 2356
gtgaagaccagcgggaaggaacgggggtccctggaagacaagactgttgtgataaaccatacacatttgcacgtttctttccaatagattattgtgaaggaattgtagttcgaagg 2474
tagccgtgtgtagattgtccatcaaacacgctgtgtcgttggtagccttagaacaagtgaaaccattaacacaaaaatgaaatttgcacccagcgtttgcccaggaatattgcac 2592
tcacatgatgtattttaatcogaagccagataaagcgataatgctagtggttttcaatgtattgtaatttttaagaattgtttaaactgttatgagataataatagaggcgggttgg 2710
aagaggtgtctcggcttagaggcatgttgggacaacatttatgggggttttaagttgactatggatgccaggttgttaagaacttaagcttctgtaatagccgctactaaagagta 2828
gacaaaaacgatgtaaatgtgttggcgaatacgtagaggttgtttagcgtgtgactatggaaagatgtaataaacacggggttctatatgtgagtgtaagctatggc 2946
tagagctgttgggaacactgtgtgtggtatatacacctgggtgccaatgttcattagtagcttaacaacggtagcttctttaatggactgatctcgaacacctcaccctggctgaaga 3064
aatcc 3069

```

Langlois et al., Figure 2A





**A****B**

Langlois et al., Figure 3

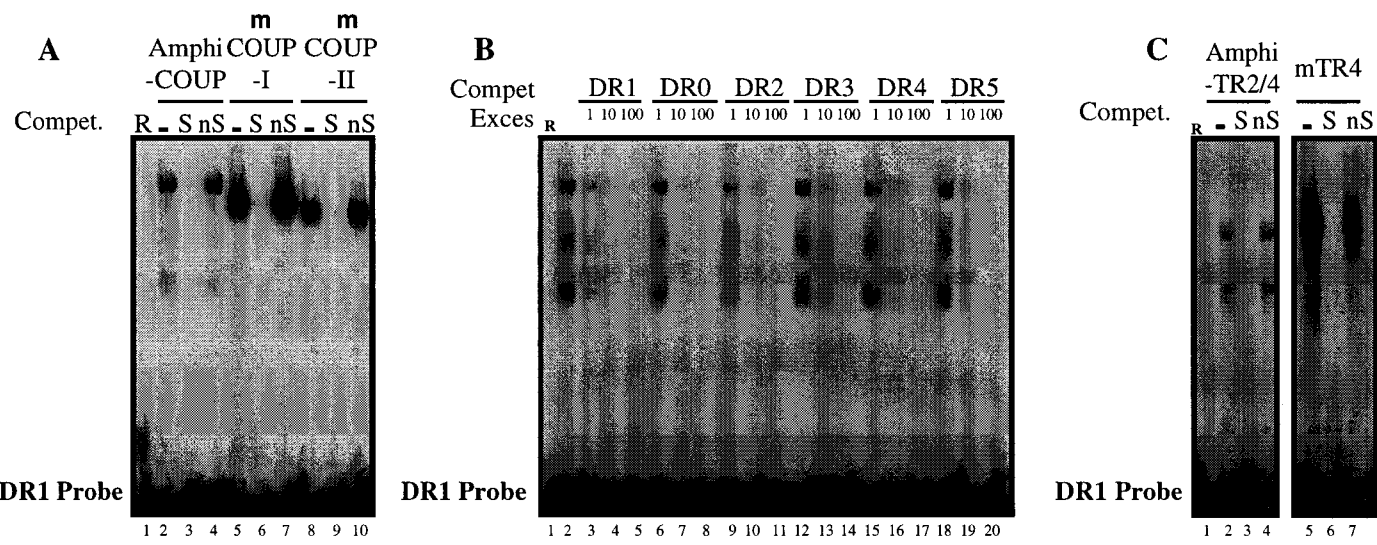
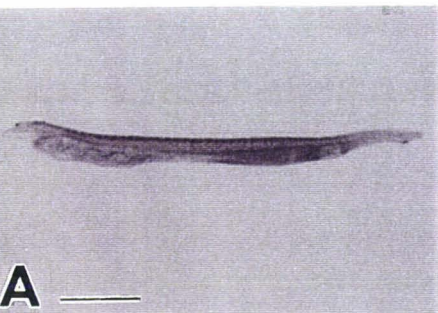
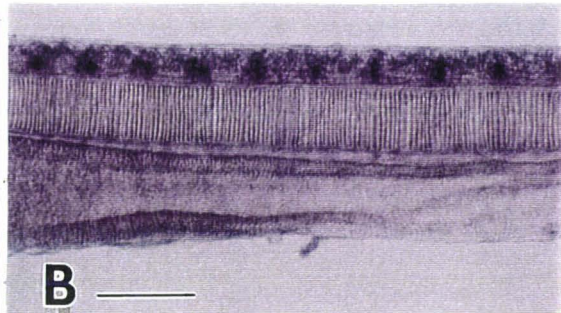


Figure 5: EMSA

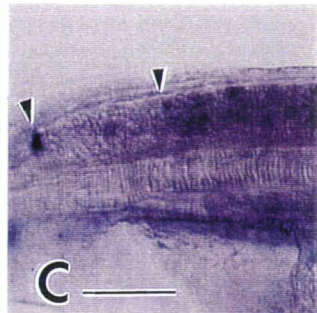




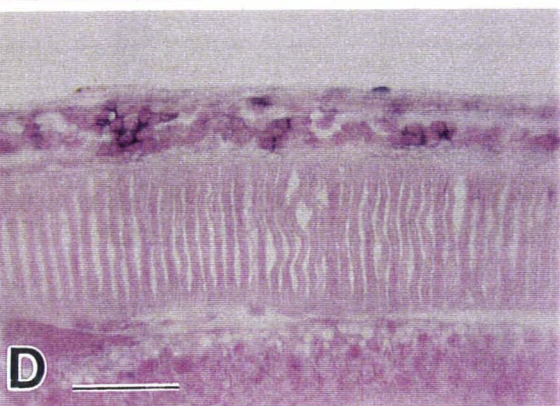
**A** —



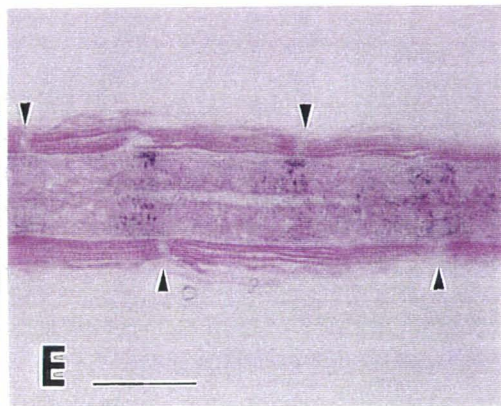
**B** —



**C** —



**D** —



**E** —



**F** —



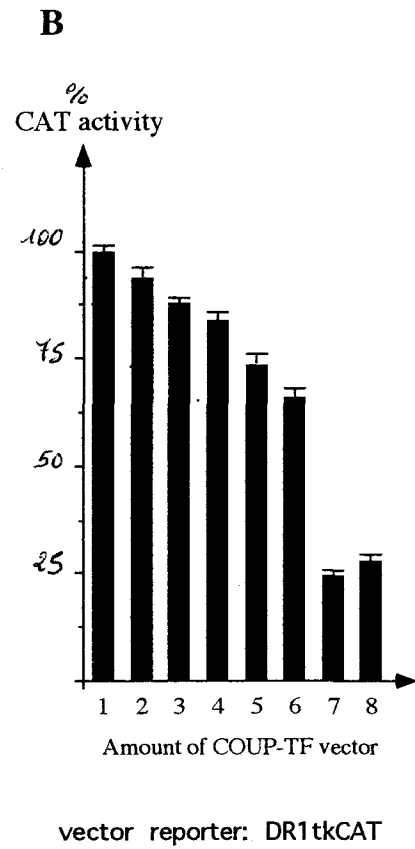
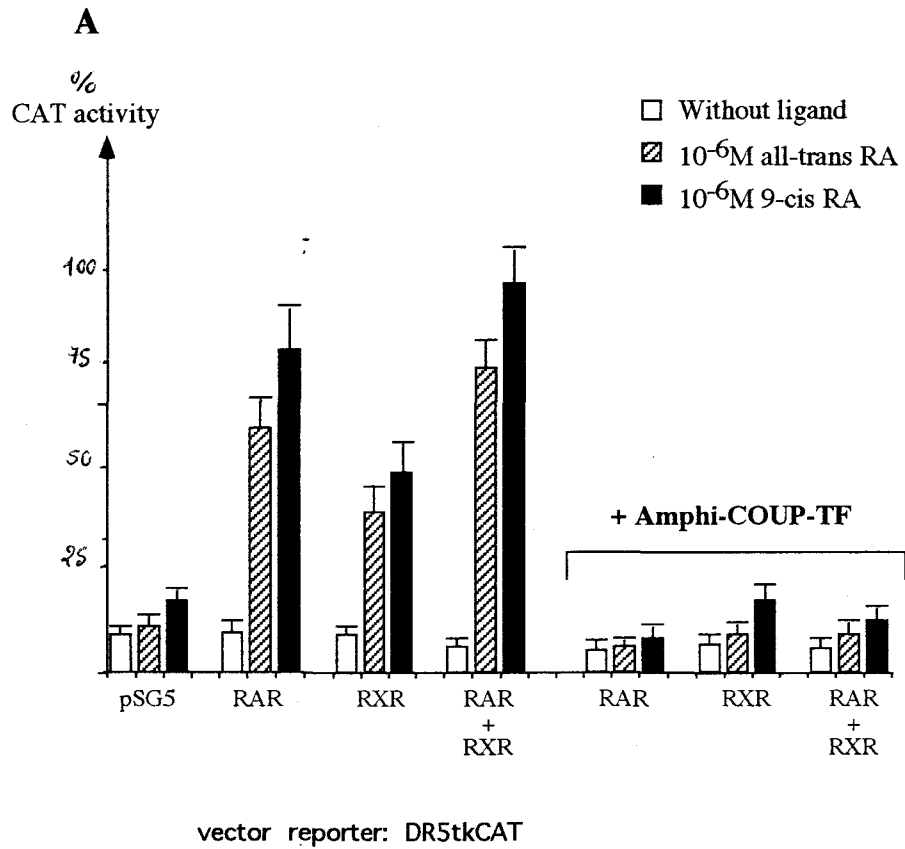


Figure 6: Transactivation tests



## Discussion de l'article 6

### **AmphiCOUP-TF et Amphi TR2/4 sont les homologues des ancêtres de COUP-TFI et II et de TR2 et TR4 de vertébrés.**

Nous avons montré dans cet article que les deux clones COUP-TF et TR2/4 isolés chez l'amphioxus sont bien les homologues de COUP-TFI et II et de TR2/4 de vertébrés, respectivement. D'une part, AmphiCOUP-TF est fortement conservé avec ses homologues de vertébrés : il présente respectivement, 93,5% et 92% d'identité de séquence avec les séquences COUP-TFI et COUP-TFII humaines, sur la totalité de la protéine, incluant 94% d'identité de séquence dans le domaine de fixation à l'ADN de COUP-TFI humain, et 95,5% dans son domaine E. De même, Amphi TR2/4, présente 82,3% d'identité de séquence avec TR2 et TR4 humains et respectivement 65,3% avec le domaine E de TR2 et 73% avec le domaine E de TR4. De plus, une analyse phylogénétique montre que ces deux récepteurs d'amphioxus représentent les homologues des ancêtres de vertébrés, de COUP-TFI et II d'une part et de TR2/4, d'autre part. Ils sont en effet, placés avant la dichotomie de COUP-TFI et II de vertébrés pour AmphiCOUP-TF et avant la dichotomie de TR2 et TR4 de vertébrés pour Amphi TR2/4.

### **Expression de AmphiCOUP-TF au cours du développement.**

AmphiCOUP-TF est exprimé de façon segmentée dans le cerveau postérieur de la larve tardive de l'amphioxus. Ces observations nous apportent plusieurs éléments nouveaux : (i) le cerveau postérieur de l'amphioxus qui apparaissait non segmenté par des études anatomiques, montrent une certaine segmentation par le profil d'expression d'AmphiCOUP-TF sous forme de raies répétées dans cette région. Des homologues de COUP-TF chez le poisson-zèbre s'expriment aussi de façon segmentée dans les rhombomères (Fjose *et al.*, 1995). (ii) Contrairement à ses homologues de vertébrés, AmphiCOUP-TF ne s'exprime que tardivement, seulement à partir du stade larvaire, et de plus, n'est pas retrouvé dans les vésicules antérieures de la corde neurale. En comparaison, les expressions des COUP-TFs ont été mises en évidence dans le cerveau antérieur et les rhombomères de tous les embryons de vertébrés étudiés, le poisson-zèbre (Fjose *et al.*, 1993; Fjose *et al.*, 1995), le xénope (van der Wees *et al.*, 1996) et la souris (Qiu *et al.*, 1994). A la lumière de ces

résultats, il semblerait que *AmphiCOUP-TF*, qui n'est pas exprimé dans le cerveau antérieur de l'amphioxus, n'ait pas de fonction précoce dans la segmentation de cet organe, contrairement à *COUP-TFI* et *II* dans le diencéphale de l'embryon de souris (Qiu *et al.*, 1994). Par contre, *AmphiCOUP-TF* s'exprime de façon segmentée dans le cerveau postérieur et partage probablement un rôle avec son homologue *COUP-TFI* de souris dans l'organisation des nerfs de cette région. L'inactivation du gène *COUP-TFI* chez l'embryon de souris, empêche en effet, la formation du nerf glossopharyngien (numéro X), (Qiu *et al.*, 1997). Il est à noter que les embryons dont le gène *COUP-TFI* est inactivé ne présentent pas d'altérations de la segmentation du diencéphale, probablement par un effet de redondance du gène *COUP-TFII* dans cette région. La fonction ancestrale de *COUP-TF* aurait pu être simplement d'intervenir dans la formation des nerfs et la segmentation du rhombencéphale ou simplement, l'organisation segmentée des nerfs de cette région. Les gènes *COUP-TF* ont d'autre part, été montrés jouer un rôle important dans la régulation de la formation des prolongements neuronaux (Connor *et al.*, 1995). La duplication des gènes *COUP-TFs* auraient permis aux gènes *COUP-TF* de jouer un rôle dans la segmentation du cerveau antérieur, seulement chez les vertébrés. Alternativement on peut également penser que la fonction ancestrale de *COUP-TF* était importante pour tous le système nerveux, antérieur et postérieure et qu'il y a eu perte secondaire de la fonction dans le cerveau antérieur chez l'amphioxus.

### **Interférences *COUP-TF* / acides rétinoïques**

L'acide rétinoïque réprime l'expression des *COUP-TFs* chez les embryons de vertébrés (van der Wees *et al.*, 1996). On ignore si l'acide rétinoïque modifie l'expression de *Amphi-COUP-TF* car les embryons traités à l'acide rétinoïque meurent avant la période d'expression de *COUP-TF*. De plus, il n'a pas encore été possible de disposer de jeunes larves d'amphioxus (qui expriment *AmphiCOUP-TF*) en quantité suffisante pour les traiter à l'acide rétinoïque. Cependant, les études fonctionnelles montrent que la fonction de répression des *COUP-TFs* sur la transactivation médiée par les acides rétinoïques est conservée chez l'amphioxus. Cette répression peut se faire par compétition avec les récepteurs *RAR* et *RXR* *via* les mêmes sites de reconnaissance sur l'ADN. Dans ce contexte, *AmphiCOUP-TF* reconnaît aussi les différents sites de fixation pour *RAR* et *RXR*, avec une meilleure affinité pour *DR1*, comme ses homologues de vertébrés. Par ailleurs, en

cotransfection dans des cellules COS, AmphiCOUP-TF est capable, comme les COUP-TFs de vertébré de réprimer la transactivation médiée par RAR/RXR, néanmoins, de manière moins efficace que ses homologues murins et humains.

### **Conclusion sur COUP-TF**

Nous avons isolé AmphiCOUP-TF, l'homologue chez l'amphioxus, de l'ancêtre de COUP-TFI et II de vertébrés. L'expression de gène diffère avec celle de ces homologues de vertébrés et semble signifier que le rôle des COUP-TFs dans la segmentation du cerveau antérieur est apparue après la dichotomie de COUP-TI et COUP-TFII. Cependant l'expression segmentée d'AmphiCOUP-TF dans le cerveau postérieur indique que ce gène partage un rôle similaire avec ses homologues de vertébrés dans cette région du cerveau (au moins, mCOUP-TFI, d'après les résultats de "Knock-out"). Par ailleurs, nous avons noté une conservation de la répression d'AmphiCOUP-TF sur la transactivation médiée par RAR/RXR.

### **AmphiTR2/4, son expression au cours du développement de l'amphioxus**

Le profil d'expression d'AmphiTR2/4 suggère un rôle de ce gène dans la mise en place des tissus nerveux au cours des stades précoces du développement embryonnaire. Son expression la plus précoce est en effet détectée au stade neurula. De plus, l'expression d'AmphiTR2/4 dans la vésicule cérébrale chez la jeune larve de l'amphioxus, indique que ce gène joue probablement un rôle dans la région du cerveau correspondant au diencephale des vertébrés. L'expression de TR2 ou TR4 dans le cerveau a déjà été mise en évidence par Northern, chez l'ambystome, pour TR2, (Wirtanen *et al.*, 1997), chez le lapin, pour TR4 (Harada *et al.*, 1998) et par hybridation *in situ* sur un fœtus de souris à 16,5 jours de gestation (Lee *et al.*, 1996b). Cependant, l'expression de TR2 ou TR4 dans le cerveau en cours de développement, n'a pas encore été décrite de façon détaillée. De plus, l'expression d'AmphiTR2/4 est intense dans la région homologue de l'adénohypophyse, région de laquelle ont pu être isolés les ADNc de TR4 de rat et humain (Chang *et al.*, 1994).

### **Répression d'AmphiTR2/4 par les acides rétinoïques**

L'effet de AR sur l'expression de TR2 ou TR4 au cours du développement embryonnaire n'a été testée chez les vertébrés, que chez TR2 d'ambystome. L'acide rétinoïque injecté au stade jeune gastrula n'affecte pas l'expression de TR2 chez cet



animal. Cependant les stades ultérieurs n'ont pas été examinés. Nous montrons ici que l'expression de TR4 peut être affectée par l'acide rétinoïque chez l'amphioxus, puisqu'elle disparaît totalement dans la vésicule cérébrale et n'est plus localisée que dans la partie antérieure des fentes branchiales du céphalocordé.

En conséquence, AmphiTR2/4 jouerait un rôle dans la mise en place des structures de la vésicule cérébrale et des fentes branchiales de l'amphioxus et son expression est modulable par les variations de concentration d'acide rétinoïque. Pour l'instant l'absence d'un ADNc comprenant une boîte T/A complète et capable de fixer l'ADN correctement nous a empêché de réaliser des expériences de transactivation *in vivo* de façon à étudier si, fonctionnellement, AmphiTR2/4 pouvait inhiber l'action trans-activatrice des rétinoïdes.

# *Annexes*

# SYSTÉMATIQUE ET ÉVOLUTION DES RÉCEPTEURS NUCLÉAIRES D'HORMONES

Hector ESCRIVA<sup>1</sup>, Ricardo MENDONCA<sup>1</sup>, Marie-Claire LANGLOIS<sup>2</sup>, Jean-Marc VANACKER<sup>2</sup>,  
Ray PIERCE<sup>1</sup> & Vincent LAUDET<sup>2\*</sup>

<sup>1</sup> INSERM U 167 - Institut Pasteur de Lille, 1 rue Calmette, 59019 Lille Cedex

<sup>2</sup> CNRS UMR 319 - Institut Pasteur de Lille - Institut de Biologie de Lille,  
1 rue Calmette, 59021 Lille Cedex

## *Résumé*

Les récepteurs nucléaires forment une vaste famille de gènes tous construits sur le même modèle. Ces récepteurs sont des facteurs de transcription qui dépendent d'un ligand pour leur action. Cette famille groupe les récepteurs des hormones stéroïdes, des hormones thyroïdiennes et de l'acide rétinoïque. En plus de ces récepteurs « classiques » cette famille comprend des récepteurs pour lesquels aucun ligand n'a été mis en évidence et que l'on regroupe dans la famille sur la base de leur identité de séquence. Ces molécules sont appelées des « récepteurs orphelins ». Par une analyse de phylogénie moléculaire, nous démontrons que la famille des récepteurs nucléaire peut être séparée en six sous familles. de façon intéressante nos résultats montrent qu'il n'y a pas de relation entre la position d'un récepteur dans l'arbre et la nature chimique de son ligand. Nous démontrons de plus que les récepteurs nucléaires retrouvés chez les métazoaires précoces comme les cnidaires sont tous des homologues de récepteurs orphelins. Ceci suggère donc que les premiers récepteurs nucléaires étaient des orphelins qui auraient acquis la capacité à fixer des ligands tardivement au cours de l'évolution. Nos résultats suggèrent de plus que la diversité actuelle de ces récepteurs a été acquise par

deux vagues de duplications géniques au cours de l'évolution.

## *Summary*

The nuclear receptor (NR) superfamily comprises, in addition to ligand-activated transcription factors, members for which no ligand has been identified to date. We demonstrate that the nuclear receptor superfamily can be split into six subfamilies. Interestingly, our results show that orphan receptors are randomly distributed in the evolutionary tree and that there is no relationship between the position of a given liganded receptor in the tree and the chemical nature of its ligand. NR are specific to metazoans, as revealed by a screen of NR-related sequences in early- and non-metazoans organisms. The analysis of the NR gene duplication pattern during the evolution of metazoans shows that the present NR diversity arose from two waves of gene duplications. Strikingly, our results suggest that the ancestral NR was an orphan receptor which acquired ligand-ability during subsequent evolution.

---

\* Auteur correspondant

## INTRODUCTION

Tous les organismes, même les plus simples comme les procaryotes ou les eucaryotes unicellulaires, doivent assurer une communication harmonieuse avec l'extérieur, c'est-à-dire « interpréter » les conditions de l'environnement, de façon à s'y adapter le mieux possible. Ici, « l'extérieur » a une définition large : il peut s'agir des conditions physico-chimiques (température, salinité, présence de nutriments ou de substances toxiques) mais également de la simple présence d'autres organismes, semblables ou non.

L'obligation de communiquer est encore renforcée chez les organismes multicellulaires chez qui la réunion de plusieurs cellules pour former un organisme complexe unique rend vitale une parfaite cohésion du tout. Cette nécessité absolue a conduit les métazoaires à développer dans leurs différentes cellules des systèmes génétiques complexes pour « écouter », « voir », « sentir » l'environnement, interpréter ces « informations » et y répondre de la meilleure façon possible. Dans la plupart des cas, ces systèmes sont basés sur la collecte d'informations à l'extérieur de la cellule *via* des récepteurs extracellulaires, sur l'intégration de ces messages *via* un système complexe de relais, les voies de signalisation intracellulaire, et enfin sur la réponse à ces informations *via* des facteurs capables de modifier l'expression des gènes de la cellule et que l'on appelle facteurs de transcription. Le terme de "récepteur" extracellulaire doit s'entendre dans un sens très large. Il peut bien sûr s'agir d'un récepteur à une substance chimique (nutriment, hormone, facteur de croissance, toxique etc...) mais aussi à un stimulus physique (un photon, un son, une pression etc...). Les capacités de réponse développées par les organismes multicellulaires sont immenses et il existe une grande variabilité des voies de signalisation qui se connectent, se ramifient les unes dans les autres, se neutralisent ou au contraire se renforcent en un immense réseau de relais en tous sens... On peut même se demander si le

concept de voie lui-même n'est pas une façon trop simpliste d'envisager ce labyrinthe d'interactions entre protéines... Il en est bien sûr de même pour les facteurs de transcription dont l'action est basée sur un principe simple et conservé (dans la majorité des cas l'interaction d'une protéine avec l'ADN et avec la machinerie transcriptionnelle) qui, dans le détail, a subi une série d'innombrables modifications pour aboutir à des systèmes d'une méticulosité touchant à la maniaquerie...

Dans ce contexte, les récepteurs nucléaires d'hormones font figure d'exception. Il s'agit en effet de molécules qui combinent en une seule entité la capacité de fixer une molécule chimique (réception d'un message), celle d'interpréter cette fixation (transduction du message) et celle enfin de réguler l'expression des gènes (mise en route de la réponse cellulaire). Les récepteurs nucléaires fournissent donc un exemple de système de communication intracellulaire apparemment très simple (bien que cette simplicité cache en fait une grande complexité), et distinct des systèmes plus traditionnels que nous venons d'évoquer. Pour ces raisons, mais aussi parce que ces récepteurs reconnaissent des molécules extrêmement importantes pour la santé de l'homme, elles ont fait l'objet d'une intense activité de recherche. Il est apparu au cours de ces dernières années qu'en plus de leurs rôles importants en clinique, les récepteurs nucléaires forment un excellent modèle d'étude de l'évolution des facteurs de transcription.

### LA FAMILLE DES RÉCEPTEURS NUCLÉAIRES

Les récepteurs nucléaires sont des facteurs de transcription qui dépendent d'un ligand pour leur action (revue dans GRONEMEYER & LAUDET, 1995 ; MANGELSDORF *et al.*, 1995). Ils forment une famille homogène de molécules qui comportent un certain nombre de domaines structuraux conservés présentant une identité minimale de séquence de 30 à 40%. Un membre de la famille des récepteurs nucléaires comprend un domaine appelé C

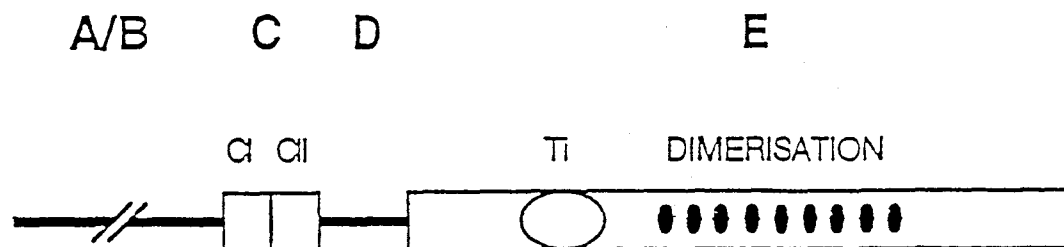


Figure 1 – Organisation en domaines fonctionnels d'un récepteur nucléaire type. Le domaine C est responsable de la fixation à l'ADN et le domaine E responsable de la fixation du ligand (lorsqu'un ligand est connu), de la dimérisation et de la régulation de la transcription.

responsable de la fixation à l'ADN et présentant deux structures dites « en doigt de zinc » qui comportent des cystéines répétées et conservées et un domaine, appelé E, toujours situé en C terminal du précédent lorsque les deux sont présents, responsable de la fixation du ligand (Fig. 1). Le domaine E est également impliqué dans la dimérisation des récepteurs ainsi que dans la régulation de la transcription. Dans la très grande majorité des cas, ces deux domaines sont présents dans la même molécule ; cependant il existe des exceptions (dont le nombre va croissant) où un seul de ces domaines (soit C, soit E) est présent.

#### Présentation d'une super(be)-famille

La famille des récepteurs nucléaires contient environ 80 membres connus dont une soixantaine chez les vertébrés et une vingtaine chez les insectes (GRONEMEYER & LAUDET, 1995). Le très grand déséquilibre dans les espèces au sein desquelles des récepteurs nucléaires ont été identifiés s'explique par le fait que ces molécules ont surtout été étudiées sous l'angle de la médecine et de la génétique du développement. On est donc très peu sorti des modèles classiques tels que l'homme, la souris, le poulet ou la drosophile.

Les récepteurs nucléaires regroupent tout d'abord des récepteurs « vrais » (Fig. 2)

comme les récepteurs des hormones stéroïdes (oestrogènes, glucocorticoïdes, minéralocorticoïdes, androgènes, progestérone ; respectivement ER, GR, MR, AR et PR), de la vitamine D (VDR), de l'ecdysone (ECR), des hormones thyroïdiennes (TR), des acides rétinoïques (isoformes tout-trans et 9-cis ; RAR et RXR). Ces molécules sont donc des récepteurs classiques à haute affinité qui relaient la réponse cellulaire à ces différents ligands. Cependant, la famille des récepteurs nucléaires comporte un grand nombre (plus de la moitié) de molécules pour lesquels le terme de récepteur est à vrai dire impropre puisqu'on ne leur connaît pas de ligand. Ces molécules collectivement appelées « récepteurs orphelins » ont été isolées sur la base leur degré d'identité de séquence avec les récepteurs classiques (ENMARK & GUSTAFSSON, 1996).

Dire que le ligand n'est pas connu ne signifie évidemment pas qu'il n'existe pas de ligand du tout et effectivement dans un certain nombre de cas on a, après de longues recherches et de nombreux rebondissements, une idée de ce que pourraient être certains ligands des récepteurs orphelins. Ainsi les PPAR (PPAR = Peroxisome Proliferator Activated Receptors), des récepteurs dont l'activité est régulée par des molécules qui provoquent la prolifération des péroxysomes (d'où leur nom), sont des récepteurs de certaines prostaglandines ou encore de leucotriènes (DEVCHAND *et al.*,

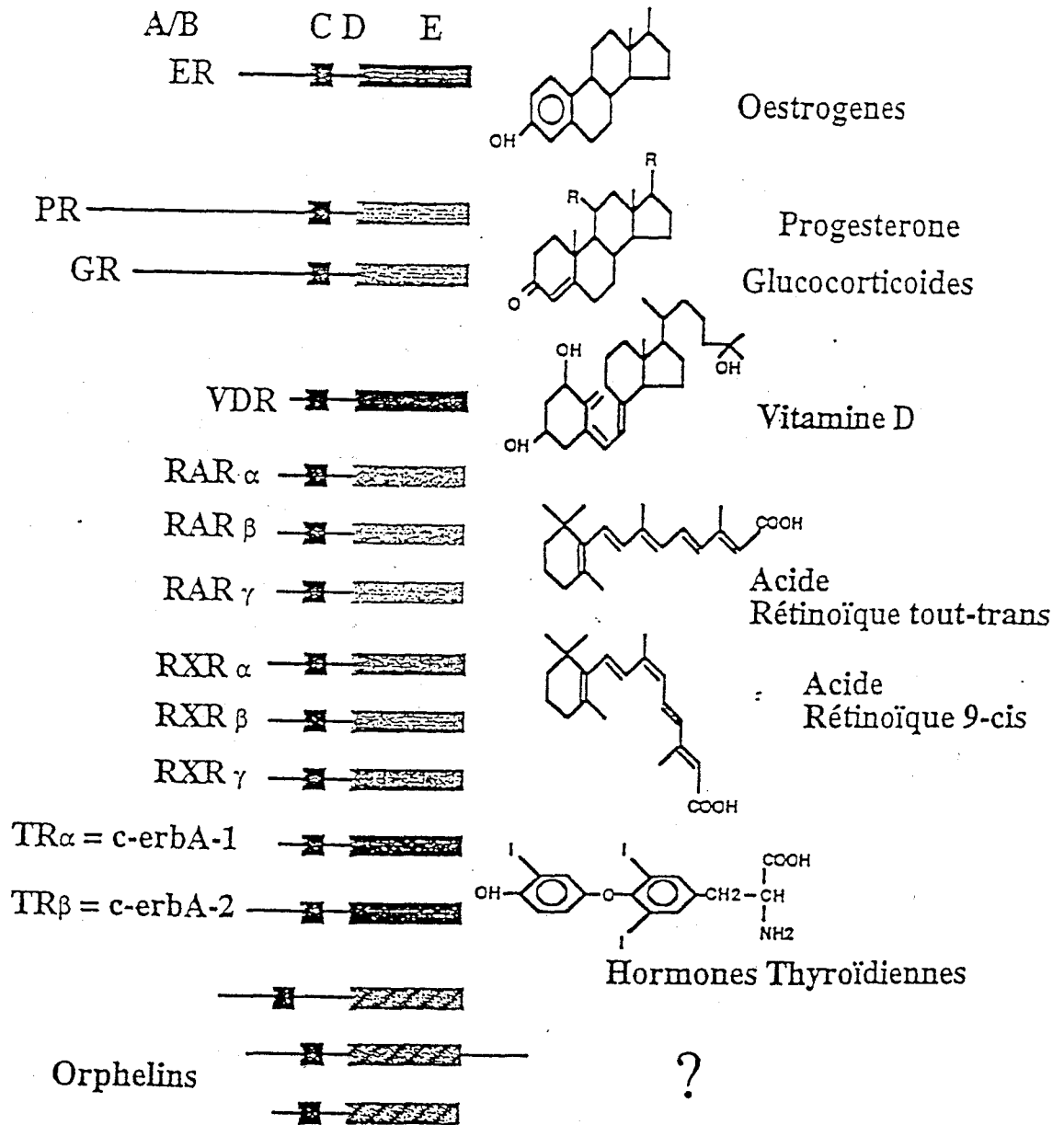


Figure 2 – Les grands types de récepteurs nucléaires et leurs ligands. Les récepteurs orphelins représentent en fait plus de la moitié des membres de la famille.

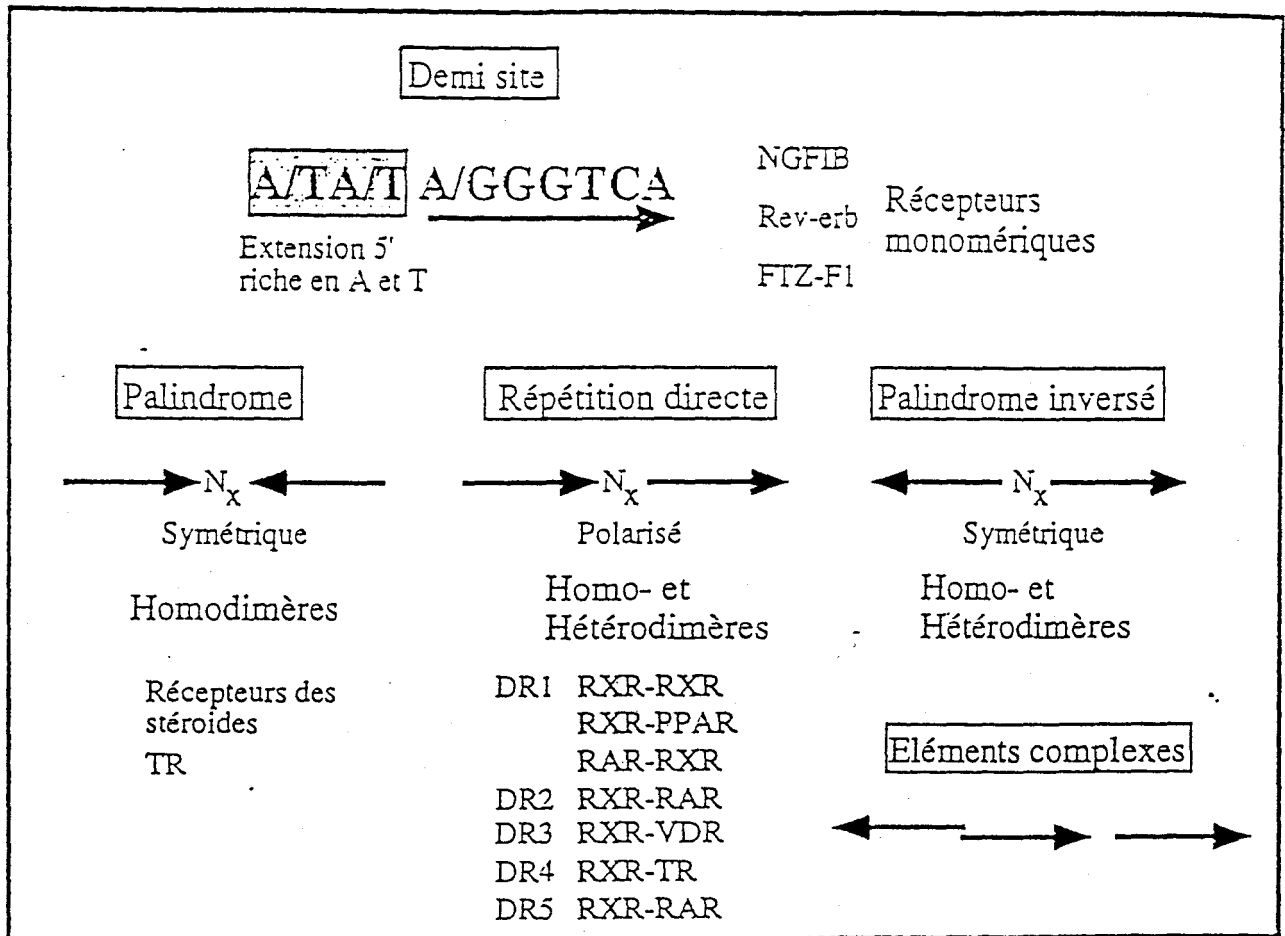


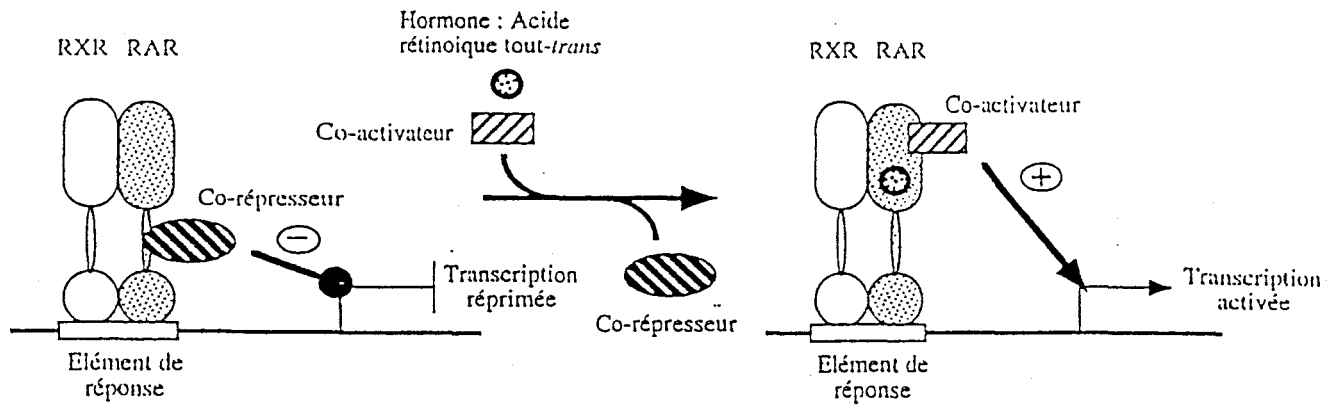
Figure 3 – Les éléments de réponse des récepteurs nucléaires. Les grands types d'éléments de réponse ont été encadrés. Les répétitions du motif A/GGGTCA sont représentées par des flèches. Certains des récepteurs capables de se fixer sur ces éléments sont indiqués. Pour les répétitions, Nx indique un nombre de nucléotides variable (de 1 à 5 pour les répétitions directes, DR1 à DR5).

1996). De même, le récepteur orphelin LXR récemment cloné, pourrait être un récepteur des oxysterols. Il est important de noter que ces nouveaux ligands sont en majorité des molécules synthétisées dans les cellules, dérivées de notre alimentation et qui pourraient avoir un rôle de signalisation intracrine (JANOWSKI *et al.*, 1996).

Cependant, dans la plupart des cas, les récepteurs orphelins continuent à nous poser un véritable défi, tant du point de vue fonctionnel que du point de vue évolutif. La chasse au ligand continue mais il n'est pas impossible que certains de ces orphelins le restent définitivement.

#### Mécanismes d'action des récepteurs nucléaires

Les récepteurs nucléaires se fixent à l'ADN *via* une séquence cible appelée HRE (pour « Hormone Responsive Element ») dont la structure est, dans la plupart des cas dérivée de AGGTCA (Fig. 3). Cette unité de base est le plus souvent présente sous forme de répétitions qui peuvent être organisées de différentes façons. On trouve par exemple des palindromes (deux unités tête-bêche), des palindromes inversés (deux unités « queue contre queue ») ou des répétitions directes (DR pour « Direct Repeat »). C'est ce dernier type d'élément qui est le plus fréquemment



**Figure 4** – Mode de fonctionnement des récepteurs nucléaires. Selon le modèle simplifié présenté sur cette figure, en absence d'hormone un récepteur nucléaire est reconnu par un co-répresseur qui empêche l'activation de la transcription. En présence d'hormone le co-répresseur est relâché, il y a changement de conformation du récepteur et fixation d'un co-activateur qui va permettre la mise en route de la transcription.

rencontré. La séquence précise de chaque motif ainsi que leur espacement détermine le type de récepteur capable de se fixer sur un site donné (GRONEMEYER & LAUDET, 1995).

Les récepteurs nucléaires se fixent souvent sous forme de dimères sur l'ADN. Tous les types de dimères possibles peuvent être observés, celui le plus fréquemment rencontré est l'hétérodimère formé entre le récepteur à l'acide rétinoïque 9-cis RXR et un autre type de récepteur, le tout fixé à une répétition directe. Là encore ce cas, le type d'hétérodimère qui se fixe à une répétition directe est au moins partiellement déterminé par la distance séparant les deux motifs. Ainsi, une répétition directe avec un espacement de 3 bases (appelée en jargon un élément DR3) sera reconnue par un hétérodimère formé entre RXR et le récepteur de la vitamine D (RXR-VDR), tandis qu'un DR4 sera reconnu par un hétérodimère RXR-TR, et un DR5 par un hétérodimère RAR-TR. Cette règle doit cependant être vue simplement comme un guide à l'expérimentation, certainement pas comme une règle déterministe stricte car elle comporte de très nombreuses exceptions.

Enfin, on s'est aperçu ces dernières années que les récepteurs nucléaires peuvent également se fixer à l'ADN sous forme de monomère. Dans ce cas, la séquence reconnue

est un site unique de type AGGTCA qui est toujours précédé en 5' d'une zone riche en base A et T. C'est la séquence de cette zone riche en A et T qui détermine le type de récepteur qui va se fixer. Ainsi la séquence TCA AGGTCA va être reconnue par les récepteurs orphelins SF1 et ERR1 tandis que la séquence AA AGGTCA sera reconnue par le récepteur orphelin NGFIB. Curieusement, pour l'instant tous les récepteurs qui se fixent à l'ADN sous forme de monomère sont des orphelins.

En absence de ligand, les récepteurs nucléaires sont fixés à l'ADN, et dans la majorité des cas ils répriment la transcription de leurs gènes cibles (HORVITZ *et al.*, 1996). Cette répression transcriptionnelle est obtenue par l'interaction des récepteurs avec des molécules spécifiques, appelées co-répresseurs qui les empêchent d'influencer la machinerie basale de transcription et bloquent donc toute transcription du gène cible (Fig. 4). En présence de ligand, le domaine E du récepteur (c'est à dire le domaine qui fixe le ligand) change de conformation et les zones du récepteur qui fixaient les co-répresseurs ne sont plus accessibles à ces dernières. Les co-répresseurs quittent donc le récepteur alors que d'autres molécules, appelées co-activateurs reconnaissent spécifiquement la conformation



active du récepteur. Comme les co-répresseurs, les co-activateurs connectent le récepteur à la machinerie transcriptionnelle, mais ils activent cette dernière et on obtient ainsi une régulation positive de la transcription. Le récepteur passe donc d'un état répresseur (et pas seulement inactif) à un état activateur de la transcription.

Les gènes cibles et les processus physiologiques régulés par les récepteurs nucléaires sont très nombreux. Citons par exemple le cas des hormones thyroïdiennes qui déclenchent la métamorphose du têtard d'amphibien en adulte ou, celui de l'ecdysone qui contrôle les mues des insectes (GILBERT *et al.*, 1996). Citons également les oestrogènes qui régulent le cycle sexuel des mammifères et de nombreux autres vertébrés ou l'acide rétinoïque qui est très important à de nombreux titres au cours du développement précoce des vertébrés. Parmi les gènes cibles on peut citer par exemple celui de l'hormone de croissance pour le récepteur des hormones thyroïdiennes (des enfants privés d'hormones thyroïdiennes sont ainsi nains) ou celui de la vitellogénine dans de très nombreuses espèces de vertébrés pour le récepteur des oestrogènes. Le rôle physiologique des récepteurs orphelins est moins clair mais on peut par exemple citer le rôle de SF1 dans la détermination sexuelle mâle des mammifères ainsi que dans le contrôle de l'expression des enzymes de la stéroïdogenèse (revue dans GRONEMEYER & LAUDET, 1995).

#### Phylogénie moléculaire des récepteurs nucléaires

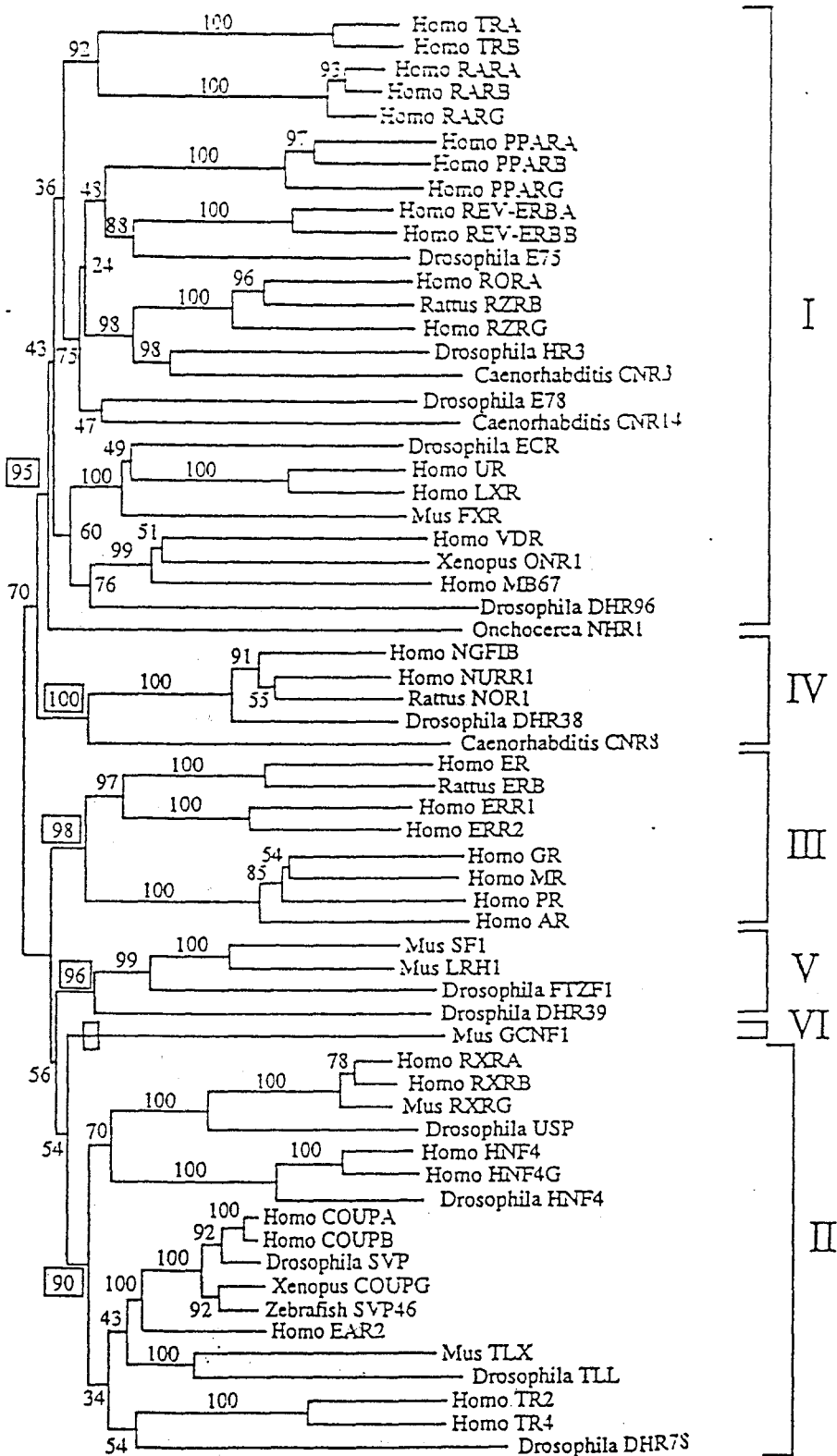
On sait maintenant que les gènes sont très souvent regroupés en familles, comme c'est le cas des récepteurs nucléaires. On connaît par exemple la famille des gènes d'actines, celle des gènes codant les protéines à homéodomains ou celle des gènes codant les protéines à activité kinase. Cette caractéristique commune amène à nous poser plusieurs questions appliquées aux récepteurs nucléaires : (i) comment de telles familles sont-elles apparues au cours de l'évolution ? ;

(ii) peut-on retracer des événements de mise en place de processus physiologiques (apparition d'une glande thyroïde endocrine) et les lier à l'apparition de certains gènes (les récepteurs des hormones thyroïdiennes) ; (iii) quelles sont les relations fonctionnelles qui unissent les membres d'une même famille ou d'un même groupe de gènes ? ; (iv) peut-on lier des caractéristiques structurales de certains gènes à des événements précis de leur histoire évolutive ? A plus long terme, la question qui guide cette approche est de déterminer s'il existe un lien entre diversification d'une famille de gènes et des événements précis de l'histoire évolutive des organismes. En fait, nous cherchons à déterminer quel a pu être le rôle des récepteurs nucléaires d'hormones dans l'évolution des métazoaires.

Pour tenter d'apporter un début de réponse à ces questions, nous avons retracé l'évolution de la super famille des récepteurs nucléaires en déterminant la phylogénie moléculaire de ces molécules en collaboration avec François CATZEFLIS, à l'Institut des Sciences de l'Evolution de Montpellier. Nous avons bâti un premier modèle d'évolution des récepteurs nucléaires qui a été affiné plus récemment (LAUDET *et al.*, 1992 ; LAUDET, soumis ; Fig 5A, B) Ce modèle a pour l'essentiel été confirmé par d'autres équipes (AMERO *et al.*, 1992 ; DETERA-WADLEIGH *et al.*, 1994).

Les séquences en acides aminés des récepteurs nucléaires ont été alignées à l'oeil à l'aide du logiciel MUST développé par Hervé Philippe à Orsay (PHILIPPE, 1993). Le choix d'utiliser les séquences protéiques a été déterminé par la faible conservation des séquences nucléotidiques, surtout dans le domaine E. Cette faible conservation est encore plus évidente lorsque l'on compare des espèces divergentes dont les génomes ont des compositions en bases très différentes. Du fait de l'extrême diversité en taille et en séquence des domaines A/B, D et F nous avons limité notre analyse aux domaines C et E bien conservés et qui sont en fait les seules régions alignables.

Les séquences alignées ont été traitées en distance (par la programme Neighbor-Joining)



**Figure 5A** – Arbre phylogénétique non raciné obtenu par la méthode Neighbour-Joining de 63 récepteurs nucléaires. Un représentant de chaque locus a été inclus dans l'arbre, les différentes séquences orthologues de vertébrés n'ont pas été incluses. L'arbre a été soumis au test du bootstrap avec 1000 répliquions (exprimées en %). Les sous familles sont indiquées en chiffre romain par des crochets. Les branches les plus internes de l'arbre ayant des valeurs de bootstrap supérieures à 90% sont encadrées et servent à la définition de ces sous familles.

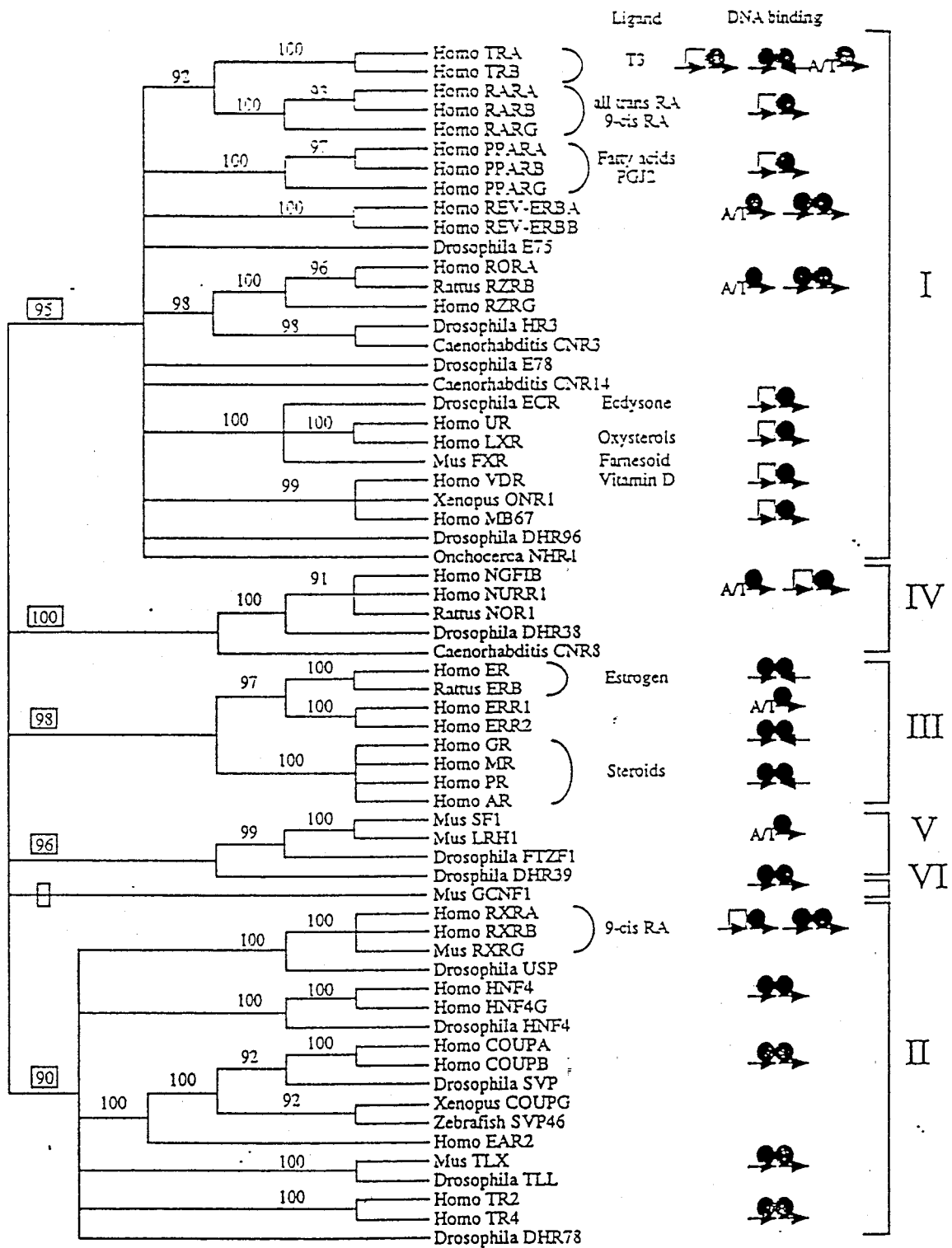


Figure 5B – Arbre consensus. Les branches de l'arbre de la partie A de la figure soutenues par des valeurs de bootstrap inférieures à 90% ont été fusionées pour ainsi former des polytomies indiquant une non résolution. Les capacités de fixation d'un ligand et de dimerisation sont schématisées sur la figure. Le flèche représente le motif AGGTCA sur lequel se fixent les récepteurs nucléaires. Les sens des flèches indiquent la polarité de l'élément de réponse : palindrome ou répétition directe. Le site monomérique est précédé d'une région riche en bases A et T (A/T). les récepteurs sont représentés par des ronds noirs. le récepteur RXR qui est la partenaire d'hétérodimerisation de nombreux autres récepteurs est indiqué par un carré blanc.

ou en parcimonie (à l'aide du logiciel PAUP). La robustesse des branches a été mesurée à l'aide de la méthode de « bootstrap ». Seules les branches présentant des valeurs de robustesse d'au moins 90% ont été jugées valides. Du fait du grand nombre de séquences de récepteurs nucléaires actuellement connues, plusieurs ensembles de données ont été traités :

1) Un ensemble exhaustif de 154 séquences de récepteurs nucléaires complets. Ces 154 séquences correspondent à environ 70 locii différents puisque de nombreuses séquences de récepteurs orthologues sont connus. Ainsi le récepteur de type alpha des hormones thyroïdiennes (TR $\alpha$ ) est connu chez l'homme, le rat, la souris, le poulet, le xénope, une grenouille, le poisson-zèbre et le flet japonais. Cet ensemble a été traité par une simple analyse en distance.

2) Le grand nombre de séquences est tout à fait préjudiciable à une analyse soignée. Nous avons donc effectué 3 choix indépendants de 30 récepteurs représentatifs des grandes subdivisions observées dans l'arbre complet (voir par exemple Fig. 8). Les trois jeux de données ont été traités de façon indépendante en distance et en parcimonie de manière à limiter autant que possible les artefacts liés à l'échantillonnage. Toutes ces séquences ont été traitées par la méthode de bootstrap avec 1000 répliques pour les analyses en distance et seulement 100 pour l'analyse en parcimonie. Au cours de ces analyses l'influence des événements d'insertion/délétion a été étudiée en comparant des arbres où toutes les positions contenant au moins un « indel » sont supprimées avec ceux où nous avons conservé un nombre variable de positions contenant des « indels ». Dans tous les cas nous avons obtenu des arbres présentant une topologie identique. En revanche les valeurs de robustesse des branches, surtout des branches les plus internes des arbres sont significativement réduites lorsque tous les « indels » sont supprimés. Nous avons ainsi conclu que ces « indels » portent une information phylogénétique et nous les avons conservés,

avec une pondération appropriée, dans les différentes analyses.

3) Enfin, un arbre dit « arbre locus » a été calculé en incluant un représentant et un seul de chaque type de récepteur (Fig 5A, B). Ainsi nous avons complètement supprimé les différentes séquences orthologues dans ce jeu de données. Les 61 séquences ainsi traitées contiennent seulement un représentant de chaque récepteur de vertébré ou d'arthropodes.

Ce travail a permis de préciser l'organisation de la famille des récepteurs nucléaires. En recherchant les branches les plus internes de l'arbre présentant des valeurs de robustesse au moins égales à 90 %, nous avons défini 5 sous-familles de récepteurs nucléaires (Fig 5A, B):

La sous-famille I, la plus grande, regroupe les récepteurs des hormones thyroïdiennes (TR), de l'acide rétinoïque (RAR), de la vitamine D (VDR) et de l'ecdysone (ECR) ainsi que de nombreux récepteurs orphelins comme les PPAR, Rev-erb, ROR, FXR, LXR etc... dont, nous l'avons vu, certains deviennent de vrais récepteurs avec ligand.

La sous-famille II, regroupe les récepteurs de l'acide rétinoïque 9-cis (RXR) ainsi que les groupes de récepteurs orphelins HNF4, COUP-TF, EAR2, TR2, TLL etc...

La sous-famille III, contient les récepteurs des stéroïdes et les orphelins ERR1 et ERR2. Cette sous famille est la seule à ne pas s'être agrandie notablement depuis notre première analyse.

La sous-famille IV, comprend les récepteurs orphelins du groupe NGFIB (NGFIB, NURR1 et NOR ainsi que leur homologue chez la drosophile et le nématode *Caenorhabditis*).

La sous-famille V, comprend les récepteurs orphelins du groupe FTZ-F1-F1 (SF1, LRH1 chez les vertébrés et FTZ-F1-F1 et DHR39 chez la drosophile).

Il n'est pas impossible qu'il y ait une sixième sous-famille puisque le récepteur orphelin GCNF, récemment découvert, ne se place clairement dans aucune des 5 sous-familles décrites.

Du fait de l'absence de groupe externe, c'est à dire d'une famille de molécules proches mais différentes des récepteurs nucléaires il nous a été impossible d'enraciner les arbres obtenus. Il semble logique de placer la racine de l'arbre entre les sous familles de récepteurs que nous venons de décrire mais cela reste encore du domaine de l'hypothèse. Il est clair en tout cas que les relations entre ces sous familles ne sont pas résolues pour le moment.

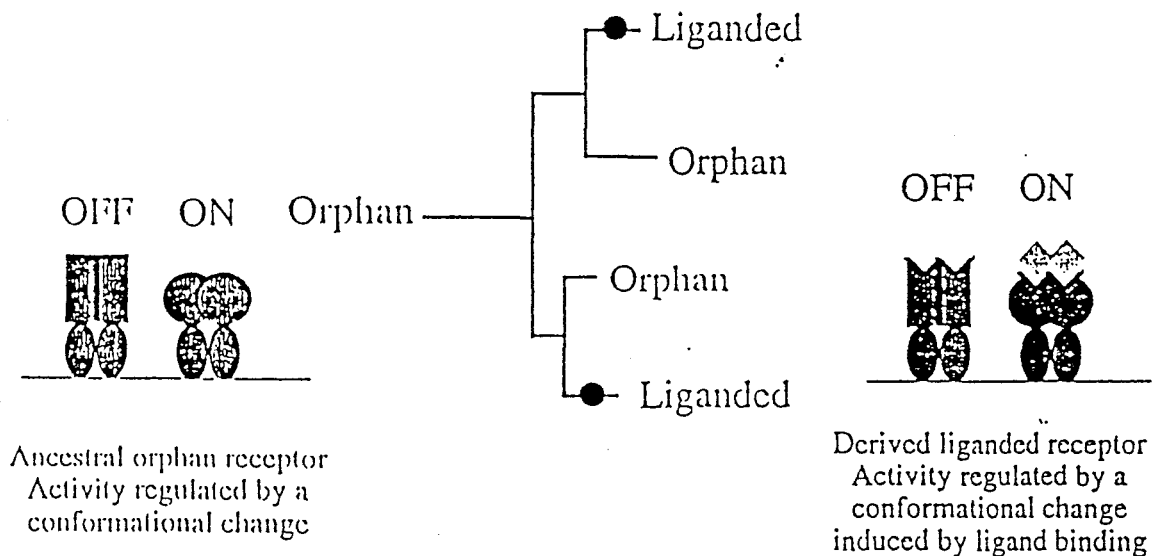
L'analyse de l'arbre obtenu permet de discuter deux points intéressants : premièrement, la concordance existant entre la fonction d'un récepteur et sa position dans l'arbre, et deuxièmement, la diversification de la famille au cours du temps.

### Évolution des fonctions des récepteurs

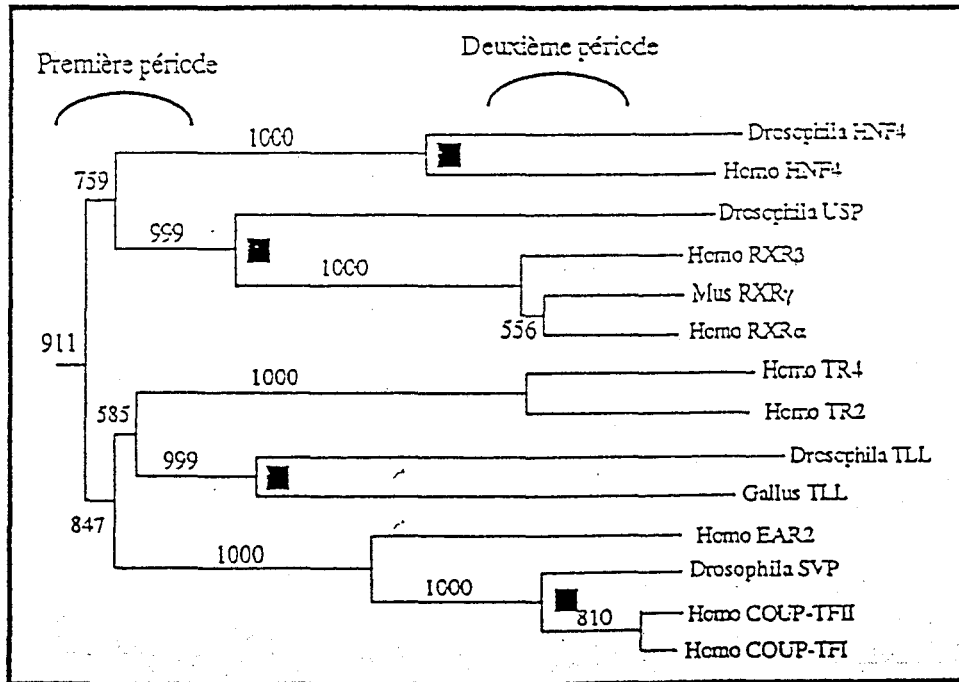
Nous avons cherché à savoir s'il existe une signification fonctionnelle aux différentes sous-familles que nous avons décrites. Première observation: il existe un lien clair entre position dans l'arbre et capacité à reconnaître l'ADN des récepteurs (Fig. 5B). On se souvient que les récepteurs nucléaires peuvent reconnaître la séquence AGGTCA organisée d'au moins trois façons différentes : sous forme d'un palindrome (normal ou inversé), d'une répétition directe ou d'un demi site AGGTCA précédé d'une zone riche en bases A et T (voir Introduction et figures 4 et 5). Si on considère l'arbre que nous avons obtenu et que l'on place près de chaque récepteur, un symbole représentant ses capacités à se fixer et à interagir avec RXR ou avec lui-même, un certain nombre de conclusions apparaissent immédiatement, qui renforcent la validité de l'arbre évolutif : on observe que la sous-famille III regroupe des récepteurs capables uniquement de se fixer à l'ADN sous forme d'homodimères. Pour qu'il y ait formation d'hétérodimère avec RXR, un récepteur doit appartenir aux sous-familles I et IV qui sont d'ailleurs proches l'une de l'autre sur l'arbre etc... Le mode de fixation à l'ADN qui semble à la fois le plus répandu et le plus dispersé dans la famille est la fixation d'un homodimère sur un élément palindromique.

Ceci nous pousse à proposer que ce mode de fixation à l'ADN est le mode ancestral de reconnaissance de l'ADN. A partir de cet homodimère simple, car symétrique, auraient dérivés plus tard deux comportements finalement plus complexes : l'hétérodimérisation avec RXR et la capacité à reconnaître de façon stable l'ADN sous forme d'un monomère. Ce type de modèle nous semble très intéressant car applicable à d'autres familles de facteurs de transcription. Bien sûr, seul l'identification de récepteurs provenant d'organismes « précoces » (*i.e.* ayant divergé très tôt dans l'arbre des métazoaires) pourra éventuellement permettre de répondre à ces questions (voir plus loin). En résumé et d'un point de vue systématique le caractère « fixation à l'ADN » peut présenter trois états différents : fixation homodimérique, fixation hétérodimérique et fixation monomérique. L'état de fixation homodimérique serait plésiomorphe alors que les deux autres (hétérodimères et monomères) seraient apomorphes, c'est à dire dérivés.

Deuxième observation, opposée à la première et très étonnante : il n'existe pas de lien entre le type de ligand reconnu par un récepteur et la position dans l'arbre de ce dernier (Fig. 5B). Ainsi, RAR et TR qui fixent des ligands complètement différents (respectivement l'acide rétinoïque et les hormones thyroïdiennes) sont toujours groupes frères. Au contraire RAR et RXR, qui fixent tous deux des rétinoïdes sont localisés dans des sous-familles différentes, de même que, par exemple, le récepteur de l'ecdysone et le récepteurs des oestrogènes qui tout deux fixent des dérivés des stéroïdes. De plus, les récepteurs orphelins sont dispersés dans tout l'arbre puisque chaque sous famille contient au moins un récepteur orphelin. Ceci suggère fortement que les récepteurs auraient acquis la capacité à fixer des ligands de façon indépendante les uns des autres sans tenir compte de leur proximité évolutive (ESCRIVA *et al.*, 1997). Donc, les récepteurs "vrais" auraient acquis la capacité à fixer des ligands relativement tard et les premiers récepteurs auraient été des orphelins, c'est-à-dire des facteurs de transcription classiques. L'état



**Figure 6** – Modèle de gain de capacité à fixer des ligands par les récepteurs nucléaires au cours de l'évolution. Les premiers récepteurs auraient été des orphelins dont l'activité aurait été régulable (passage de l'état « OFF » à l'état « ON ») par un changement de conformation induit par un événement tel que la phosphorylation. Au cours de la diversification des récepteurs, certains d'entre eux auraient acquis la capacité à fixer des ligands et ceux-ci auraient alors induit le changement de conformation qui passe le récepteur dans l'état activé « ON ».



**Figure 7** – Arbre phylogénétique (NJBOOT) groupant des récepteurs nucléaires de la sous-famille II. Sur cet arbre figurent un certain nombre de récepteurs homologues chez les vertébrés et les arthropodes. Les carrés noirs indiquent sur l'arbre la dichotomie entre des récepteurs d'arthropodes et de vertébrés, ces duplications correspondent donc à la séparation des protostomiens et des deutérostomiens. Les deux périodes de diversification des récepteurs sont indiquées.

orphelin d'un récepteur serait donc plésiomorphe et les récepteurs possédant des ligands seraient donc dérivés (Fig. 6). Ceci est en accord avec l'observation selon laquelle les récepteurs ayant des ligands apparaissent comme des nouveautés évolutives dans l'arbre. En effet, ces récepteurs ne sont présents que chez les vertébrés (TR, RAR, PPAR, les récepteurs stéroïdiens, VDR) et n'ont pas d'homologues connus chez les arthropodes. Ils apparaissent donc comme spécifiques des vertébrés.

Ce modèle est tout à fait l'inverse de ce que la plupart des spécialistes des récepteurs nucléaires imaginaient jusqu'à présent ce qui illustre bien tout l'intérêt de la démarche évolutive appliquée à cette famille. En effet chacun pensait que le premier récepteur nucléaire aurait pu être un récepteur d'un ligand quelconque qui aurait, suite à des duplications géniques successives, donné la super-famille que l'on connaît actuellement. Ce modèle simple ne permet cependant pas de comprendre les changements de spécificité des récepteurs car leurs ligands ne sont pas des peptides qui peuvent accumuler des mutations. Les ligands des récepteurs nucléaires sont des petites molécules hydrophobes produites par des cascades métaboliques complexes et la pression de sélection qui gouverne leur synthèse (donc leur structure finale) doit s'exercer sur une toute une série de gènes différents. Un exemple : l'hormone T3 et l'acide rétinoïque n'ont rien en commun, ni voie métabolique de synthèse, ni structure chimique. Dans un tel contexte comment imaginer le passage du récepteur ancestral non dupliqué qui précédait les deux groupes frères TR et RAR ? Alors comment imaginer quel ligand pouvait fixer le TR/RAR ancestral ? La T3 ou l'acide rétinoïque, ou bien encore les deux ? Il est en fait plus simple et plus parcimonieux d'imaginer que l'ancêtre commun de TR et RAR était un récepteur orphelin et qu'il y a eu un gain indépendant de la capacité à fixer des ligands après la duplication qui a créé TR d'une part et RAR d'autre part. Un tel modèle a des conséquences fonctionnelles importantes : ainsi puisque les

différents récepteurs auraient acquis la capacité à fixer des ligands indépendamment de façon convergente, on s'attend à ce que les poches hydrophobes de fixation du ligand soient différentes d'un récepteur à l'autre.

### Diversification des récepteurs

L'analyse de l'arbre phylogénique des récepteurs nucléaires nous a récemment permis de préciser la manière dont ces gènes se sont diversifiés au cours de l'évolution. En effet, sur un arbre on distingue nettement deux types d'événements distincts (Fig. 7). Tout d'abord, l'apparition des différents groupes de récepteurs, qui a eu lieu avant la divergence arthropodes/vertébrés. Cette diversification précoce est intéressante car laisse à penser qu'il peut y avoir une relation entre l'apparition des premiers organismes multicellulaires et la mise en place des grands groupes de récepteurs, comme c'est le cas également pour les gènes du complexe homéotique Hox/HOM (SLACK *et al.*, 1993). Il n'est pas inutile de rappeler que les récepteurs nucléaires sont des molécules intimement impliquées dans les systèmes de communication entre cellules, un processus vital pour un métazoaire... On peut donc émettre l'hypothèse que la première période d'évolution des récepteurs soit contemporaine de l'apparition des premiers métazoaires.

La deuxième période de diversification des récepteurs est plus tardive et correspond à une diversification au sein de chacun des groupes de récepteurs (Fig 7). C'est au cours de cette période que le gène ancestral codant le récepteur des hormones thyroïdiennes se duplique pour donner les deux copies paralogues actuellement connus : TRa et TRb. Le même phénomène se produit pour les groupes de récepteurs RAR, RXR, PPAR, COUP ou FTZ-F1-F1. Il y a en fait apparition des différentes copies de gènes paralogues de récepteurs capables d'exercer des actions similaires dans l'organisme. Il est tentant de penser que cet événement est contemporain à l'apparition des premiers vertébrés.

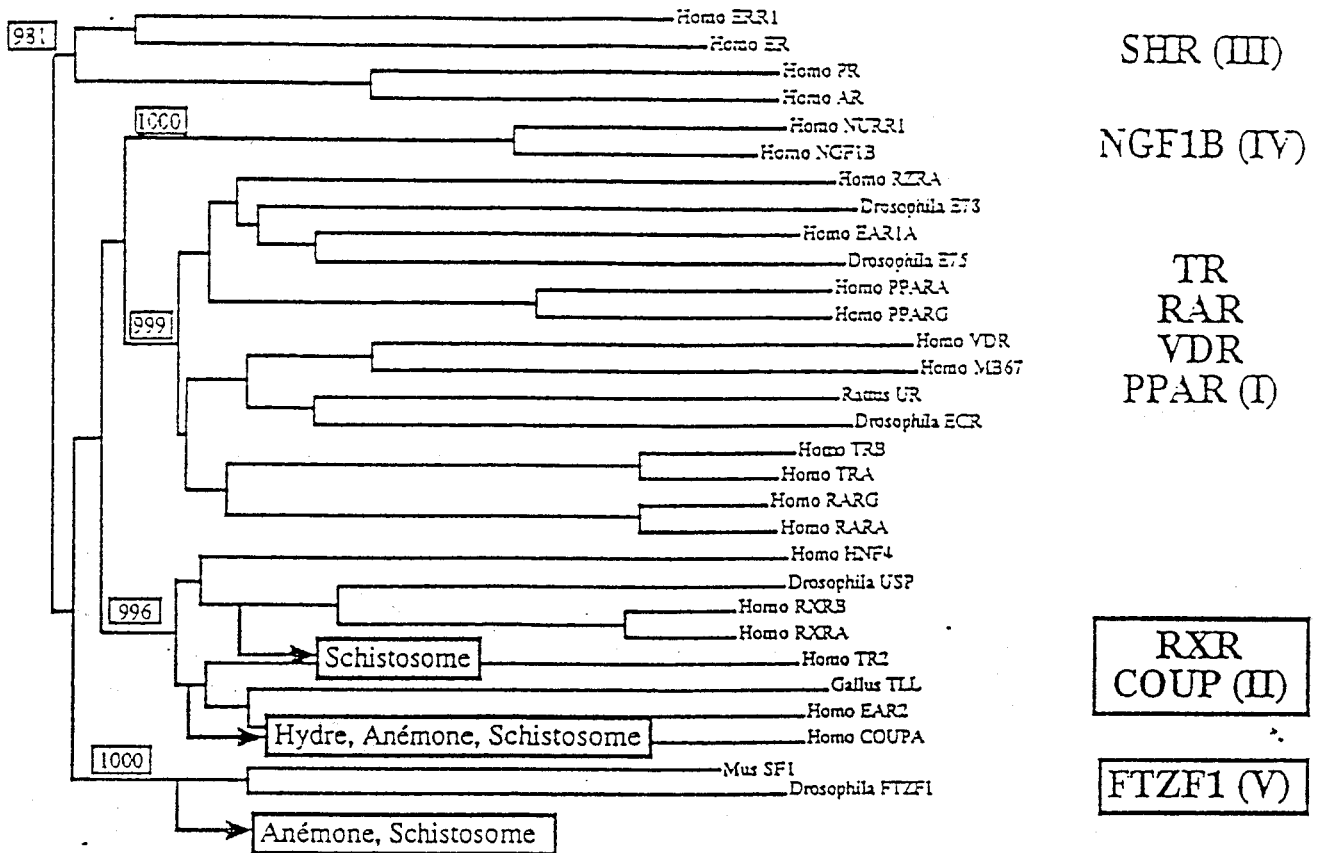


Figure 8 – Arbre phylogénétique non raciné obtenu par la méthode Neighbour-Joining de 30 récepteurs nucléaires représentatifs. La position des fragments de PCR obtenu chez les métazoaires précoces est indiquée. On voit bien que ces récepteurs ne représentent que deux des sous familles décrites et sont des homologues de récepteurs orphelins. Seules les valeurs de bootstrap correspondant aux sous familles sont indiquées.

Ce modèle d'évolution en deux étapes est valable pour d'autres familles de gènes. C'est le cas par exemple des gènes homéotiques Hox qui ont été très étudiés de ce point de vue (HOLLAND & GARCIA-FERNANDEZ, 1996). C'est également le cas de deux autres familles de facteurs de transcription que nous avons étudiées : celle des gènes ETS et celle des facteurs à boîte HMG (LAUDET *et al.*, 1993a, 1993b).

La mise en évidence d'une série de duplications de gènes très variés spécifiquement dans la lignée des vertébrés est à mettre en rapport avec le modèle proposé dans les années 1970 par Susumu Ohno (OHNO, 1970). En posant la question de savoir

quels types d'événements génétiques pouvaient entraîner des modifications majeures de la morphologie des organismes, comme par exemple l'apparition des structures propres aux vertébrés, cet auteur s'est penché sur le rôle des duplications de gènes. Il a proposé que la duplication de grandes fractions du génome (voire même l'apparition de polyploïdie), pouvait créer des gènes redondants qui seraient alors capables de diverger en restant relativement peu touchés par la sélection naturelle, jusqu'à ce qu'ils soient co-optés pour une nouvelle fonction. Selon ses propres termes (cités dans HOLLAND *et al.*, 1994) : « la sélection naturelle modifie, alors que la redondance crée ». En bref, la



duplication des gènes serait une force créative majeure pour engendrer des innovations en matière de développement. Le modèle que nous venons de proposer est évidemment en parfaite concordance avec cette vue et notre but (à long terme !) est donc maintenant de montrer comment les récepteurs nucléaires et leur diversification ont permis, en même temps que d'autres gènes (ceux du complexe Hox, par exemple) la mise en place d'innovations majeures. Dans une telle démarche, le premier objectif est bien sûr de rechercher des gènes de récepteurs nucléaires dans des organismes situés à des étapes clés de l'évolution, comme par exemple la transition chordés-vertébrés ou les toutes premières étapes de l'apparition des métazoaires. Nous sommes donc partis à la pêche aux récepteurs...

A la recherche des récepteurs nucléaires chez les métazoaires précoces

#### But du jeu

Notre objectif était d'étudier expérimentalement le modèle que nous avons élaboré d'évolution « en deux étapes » de la famille des récepteurs nucléaires. Nous souhaitons ainsi aborder les deux grandes époques de l'évolution des récepteurs nucléaires pour tenter de déterminer si elles correspondent à des étapes importantes de l'évolution des métazoaires. Bien sûr, ce type d'étude ne peut donner qu'une corrélation entre un événement génétique et un événement morphologique et ne permettra pour autant d'établir un lien causal entre les deux. Nous avons divisé cette recherche en deux grandes parties : (i) la recherche de récepteurs dans des métazoaires ayant divergé très tôt dans l'arbre évolutif et que nous appellerons par commodité des métazoaires « précoces » et (ii) la recherche de récepteurs au cours de la transition chordés-vertébrés.

#### Première période : les « métazoaires précoces »

En recherchant dans quels groupes actuels, autres que les vertébrés et les arthropodes, existent des récepteurs nucléaires, on peut

obtenir une première idée de l'ancienneté de cette super famille. Pour l'instant, aucune séquence de récepteur nucléaire convaincante n'a été décrite en dehors des vertébrés, des échinodermes, des insectes et des nématodes et le génome de la levure de boulanger qui a été entièrement séquencé en semble dépourvu.

Nous avons donc recherché des récepteurs nucléaires dans des métazoaires « précoces » comme les éponges (*Pseudocortidium jarre*), les coelentérés (une anémone de mer : *Anemonia sulcata*, une hydre : *Hydra vulgaris*) ou les plathelminthes (le schistosome : *Schistosoma mansoni* ; WRAY *et al.*, 1996 ; PHILIPPE *et al.*, 1994). Nous avons également recherché des récepteurs nucléaires dans des phyla plus récents comme les annélides (*Perinereis cultrifera*), les mollusques (*Biomphalaria glabrata*), les crustacés (*Artemia salina*). Nous avons donc réalisé des expériences de PCR avec une trentaine de couples d'oligonucléotides dégénérés et des conditions extrêmes pour éviter les contaminations. Cette recherche s'est avérée fructueuse puisque nous avons trouvé une vingtaine de fragments de PCR correspondant à des récepteurs nucléaires à partir de ces organismes dont 2 chez l'hydre, 4 chez l'anémone de mer et 5 chez le schistosome (ESCRIVA *et al.*, 1997). Les récepteurs nucléaires étant connus chez les coelentérés, il en découle qu'ils doivent logiquement exister chez tous les métazoaires, sauf événement de perte secondaire. En revanche, nous n'avons pas retrouvé de récepteurs nucléaires chez l'éponge. Ceci peut être dû au fait que les séquences recherchées sont trop divergentes ou que les récepteurs nucléaires n'existent effectivement pas chez l'espèce considérée.

Les courtes séquences dont nous disposons maintenant nous permettent de les assigner à un groupe de récepteurs donné (Fig. 8). De cette analyse phylogénétique encore très primaire, il ressort que chez les coelentérés étudiés on ne retrouve que des membres de deux des cinq sous-familles décrites : la sous-famille II (RXR, COUP, HNF4, TR2) et la sous-famille V (FTZ-F1-F1). Ce résultat peut, là encore, être dû à une incapacité technique d'amplifier des séquences des membres des

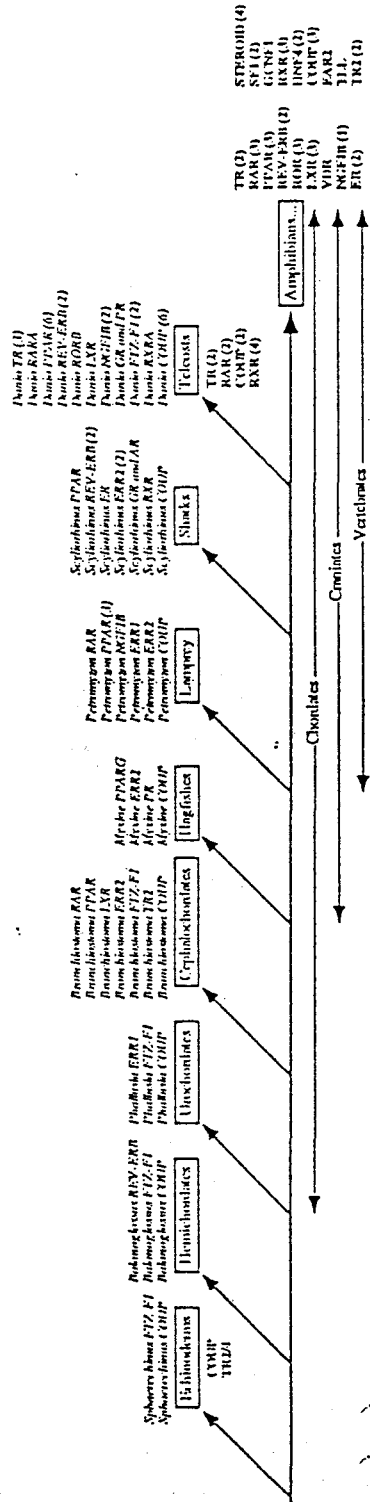


Figure 9 – Duplication des gènes de récepteurs nucléaires et transition cordés-vertébrés. Les différents groupes zoologiques impliqués dans cette transition sont indiqués. Au dessus de chaque groupe les fragments de PCR obtenus sont indiqués en italique avec entre parenthèse le nombre de copies paralogues (c'est à dire de gène sdifférents) trouvées. On voit bien que chez l'amphioxus on ne dispose que d'une copie de chaque groupe paralogue. Dans un arbre phylogénique ces copies sont situées avant la dichotomie menant au différents paralogues alors que chez la lamproie les différentes séquences trouvées correspondent a des copies déjà dupliquées (donnée non montrée). Les groupes indiqués en gras pour les téléostéens et les amphibiens (et les autres tétrapodes) sont les groupes paralogues déjà décrits dans la littérature avec le nombre de copies connus.

autres sous-familles. Mais il faut cependant noter que les couples d'oligonucléotides avec lesquels nous n'avons pas trouvé de membres des sous-familles I, III ou IV chez les coelentérés ont bien fonctionné dans d'autres organismes comme le schistosome ou les chordés. On peut donc émettre l'hypothèse que si l'on ne trouve que des membres des sous-familles II et V chez les coelentérés, cela reflète une réalité biologique. Ainsi, les sous-familles II et V seraient donc les premières apparues et les sous-familles I, III et IV seraient apparues plus tardivement.

De façon intéressante, les récepteurs que nous avons trouvés chez les coelentérés sont (sauf en ce qui concerne les homologues de RXR) des homologues de récepteurs orphelins. Ceci, joint aux observations tirées de l'arbre phylogénétique, nous permet de proposer que les premiers récepteurs nucléaires étaient des orphelins et que certains de ceux-ci ont acquis tardivement et indépendamment les uns des autres la capacité de fixer des ligands. Ce modèle, a priori étonnant, n'est pas si invraisemblable qu'il n'y paraisse. En effet on sait maintenant que l'activité transcriptionnelle des récepteurs nucléaires peut être enclenchée par autre chose que le ligand comme par exemple un événement de phosphorylation (WHITE *et al.*, 1997). Le mode de régulation par changement de conformation des récepteurs serait donc ancien, et la fixation d'un ligand serait venue, plus tard, affiner un mode de régulation allostérique basé sur d'autres déclencheurs.

Bien sur, ce modèle est encore spéculatif, mais il y a là une voie de réflexion et de tests passionnants. En outre, nous avons maintenant une piste directe pour caractériser ces récepteurs nucléaires « précoces ». C'est ce que nous sommes en train de réaliser, en clonant des cDNAs correspondant à ces différents récepteurs. Nous nous focalisons ainsi sur un récepteur de type RXR/USP trouvé chez le schistosome dont nous avons isolé un ADNc. L'étude de ce récepteur de schistosome devrait amener des données intéressantes pour tester notre modèle de gain de capacité à fixer des ligands. En effet, les RXR de vertébrés ont comme ligand l'acide

rétinoïque 9-cis alors que le produit du gène USP de *Drosophile* (qui est l'homologue des 3 RXR de vertébrés) en est incapable (MANGELSDORF *et al.*, 1995). Perte de la capacité à fixer l'acide rétinolique 9-cis chez USP ou gain chez les RXRs de vertébrés ? L'étude du clone RXR/USP de schistosome pourra amener une réponse à cette question.

#### *Deuxième période : la transition chordés /vertébrés*

L'étude de la transition chordés-vertébrés est un excellent exemple de la manière dont les interactions entre zoologie, biologie du développement et génétique moléculaire peuvent se croiser pour apporter des résultats nouveaux, tout ceci à la lumière de la systématique... En effet, un lien se précise de plus en plus entre duplication de gènes (en fait probablement des événements de polyploïdisation affectant tout le génome, comme proposé par Susumu OHNO, 1970) et apparition des caractéristiques morphologiques et développementales nouvelles, propres aux vertébrés (la crête neurale, les arcs branchiaux cartilagineux, etc...). Les gènes Hox ou Wnt fournissent un exemple de duplication de gènes propres aux vertébrés qui semble se produire entre les procordés (comme l'amphioxus) et les premiers vertébrés comme la lamproie. Sur ces bases, l'équipe de Peter HOLLAND a proposé un modèle : il y aurait eu deux duplications totales ou partielles du génome au cours de l'histoire des vertébrés : l'une entre l'amphioxus et la lamproie et l'autre juste avant l'apparition des chondrichthyens (HOLLAND *et al.*, 1994 ; HOLLAND & GARCIA-FERNANDEZ, 1996). La recherche du nombre de gènes dans les différents organismes qui se placent à des stades intéressants de cette transition nous a donc paru importante. Nous avons étudié un échinoderme (l'oursin, *Sphaerechinus granularis*), un hémicordé (*Balanoglossus sp.*), un tunicier (l'ascidie, *Phallusia mamillata*), un céphalocordé (l'amphioxus, *Branchiostoma lanceolatum*), deux agnates – bien que ce soit un groupe paraphylétique ! – (la lamproie *Petromyzon marinus* et la myxine *Myxine glutinosa*), un chondrichthyen (la roussette *Scyliorhinus*

*caniculata*) et un « poisson » téléostéen (le danio plus connu sous son nom anglais de « zebrafish », *Danio rerio*).

Les résultats obtenus vont dans le sens du modèle présenté ci-dessus. Nous pouvons par exemple discuter les résultats obtenus sur deux groupes de récepteurs, pour lesquels trois gènes paralogues sont connus chez les vertébrés supérieurs : les récepteurs de l'acide rétinoïque RAR et les récepteurs PPAR. Dans chacun de ces cas, nous n'avons trouvé chez l'amphioxus qu'une seule séquence dans une analyse phylogénétique (Fig. 9) qui sort en racine des 3 gènes connus chez les vertébrés. Chez la lamproie et la roussette il y a plusieurs gènes (2 ou 3) et il est donc clair qu'un événement de duplication au moins s'est produit entre les divergences de l'amphioxus et de la lamproie (HOLLAND, 1996).

#### *Duplications de gènes et évolution morphologique.*

Nous pensons donc qu'il y a une corrélation entre des événements de duplication affectant des familles de régulateurs du développement et des innovations majeures du développement embryonnaire, qui ont conduit à l'apparition de groupes zoologiques nouveaux comme les vertébrés. *A priori*, des nouveautés développementales peuvent être dues à différentes causes, telles que l'accumulation de mutations dans les séquences codantes ou régulatrices de ces gènes-clefs du développement, ou la duplication de gènes. Cependant, il paraît difficile d'expliquer des événements de macro-évolution par la lente accumulation de mutations dans des gènes. Les vagues de duplication dont nous tentons de retracer l'histoire peuvent-elles jouer un rôle dans un tel processus ?

La duplication des gènes offre une opportunité unique aux deux copies filles : celle « d'explorer » des opportunités nouvelles. Ainsi, la duplication d'un gène ne serait pas directement causale dans l'apparition d'un nouveau mécanisme développemental (il n'y a pas de « monstre prometteur »...) mais permettrait, par la divergence (souvent rapide après duplication) des copies filles, « d'accepter » plus facilement des

changements phénotypiques importants. Dans cette hypothèse, après duplication, on permet à un gène donné d'être modifié avec un coût phénotypique faible, voire nul, puisqu'il y a une copie de secours. Il faut d'ailleurs noter que, très souvent, les deux copies filles ne se partagent pas simplement les mutations qui apparaissent après duplication. Ainsi en étudiant les séquences de récepteurs des hormones thyroïdiennes de différents vertébrés que nous avons identifiées nous avons constaté que le récepteur de type  $\alpha$  accumulait beaucoup plus de mutations que le récepteur  $\beta$ . Dans ce cas on s'attend à ce que le récepteur unique non dupliqué qui serait retrouvé chez l'amphioxus soit plutôt proche du type  $\beta$  des vertébrés. Le récepteur aurait alors « exploré » des voies nouvelles alors que le récepteur  $\beta$  serait resté proche de sa fonction ancestrale et essentielle. L'analyse plus approfondie des récepteurs obtenus par notre équipe notamment chez l'amphioxus, nous fournira une magnifique opportunité de tester ce modèle.

Cette discussion sur le rôle moteur des duplications dans la mise en place d'innovations au cours de l'évolution est à rapprocher du débat qui agite les généticiens moléculaires à propos des expériences de recombinaison homologue où l'on mute sélectivement un gène donné pour étudier sa fonction. L'un des résultats majeurs de cette approche réalisée depuis une dizaine d'années chez la souris est... son absence de résultat ! En effet, très souvent, un gène considéré comme étant essentiel, par exemple un récepteur de l'acide rétinoïque, peut être muté avec un effet phénotypique très faible ! Une des explications de ce phénomène étrange, qui n'est pas observé chez des organismes plus simples comme la drosophile ou le nématode, est que l'on est probablement dans un cas de redondance génétique. Plusieurs gènes proches, paralogues, existent et si l'on mute l'un d'entre eux, les autres prennent le relais. Cette explication simple n'est pas toujours facile à démontrer car, comme l'a observé l'équipe de Pierre CHAMBON, si l'on mute un RAR donné on n'observe pas pour autant

# Evolution and Diversification of the Nuclear Receptor Superfamily<sup>a</sup>

HECTOR ESCRIVA,<sup>b</sup> MARIE-CLAIRE LANGLOIS,<sup>c</sup>  
RICARDO LUIS MENDONÇA,<sup>b</sup> RAY PIERCE,<sup>b</sup> AND VINCENT LAUDET<sup>c</sup>

<sup>b</sup>INSERM U167 and <sup>c</sup>CNRS URA1160, Institut Pasteur de Lille, 1 rue Calmette 59019, Lille Cedex, France

Nuclear receptors play an important part in the regulation of cell growth and differentiation by providing a direct link between signaling molecules and the transcriptional response.<sup>1</sup> These receptors are grouped into a large superfamily which includes receptors for steroid hormones, vitamin D, ecdysone, retinoic acids (all-*trans* or 9-*cis* isoforms), or thyroid hormones. It was recently discovered that fatty acids, farnesol, or prostaglandin J<sub>2</sub> metabolites may act through binding to nuclear receptors. In addition to these receptors for known ligands, an increasing number of "orphan" receptors have been described. It is not yet known if all of these orphan receptors indeed have a ligand still to be identified, if they act in a constitutive manner, or if they have an alternative transcriptional regulation mechanism. To date, the family of nuclear receptors contains more than 60 members.<sup>1</sup>

The modular organization of nuclear receptors, the various degrees of conservation between their respective domains, and the importance of nuclear receptors for many physiological processes in both arthropods and vertebrates have led several authors to study these molecules from an evolutionary point of view. Molecular phylogeny studies have led to a classification of the family into three subfamilies: (1) a large subfamily containing TR, RAR, PPAR, VDR, and EcR; (2) a subfamily containing RXR, HNF4, and COUP-TF; and finally (3) a subfamily containing the steroid receptors with the orphan receptors ERR1 and ERR2.<sup>2</sup> This classification was confirmed by other studies using only the C domain.<sup>3,4</sup>

Nevertheless, the number of characterized receptors has greatly increased in the last few years, and the power of resolution of molecular phylogeny methods has been improved. In this paper we present briefly a new molecular phylogeny of the nuclear receptor superfamily. Furthermore, we tested the validity of this phylogenetic model by identifying homologs of known nuclear receptors in early metazoans.

## MOLECULAR PHYLOGENY OF NUCLEAR RECEPTORS

A phylogenetic tree (FIG. 1) produced from alignment of the conserved domains of nuclear receptors is in agreement with previous analyses<sup>2-4</sup> done with fewer sequences, but extends the number of subfamilies into which the superfamily can be divided. In our previous analysis we defined three subfamilies of nuclear receptors. In addition to these three subfamilies the newly generated tree harbors three smaller

<sup>a</sup>This work was supported by IPL, CNRS, INSERM, ARC, and GREG.

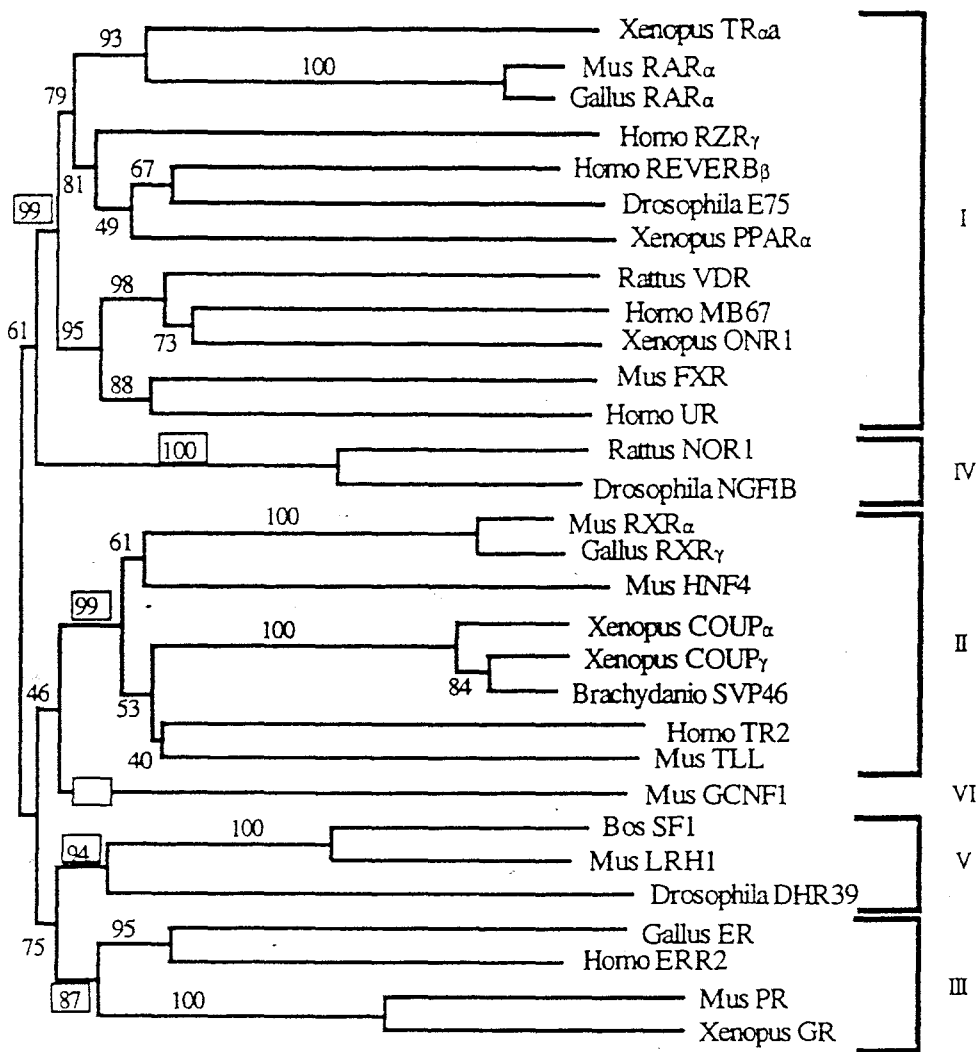
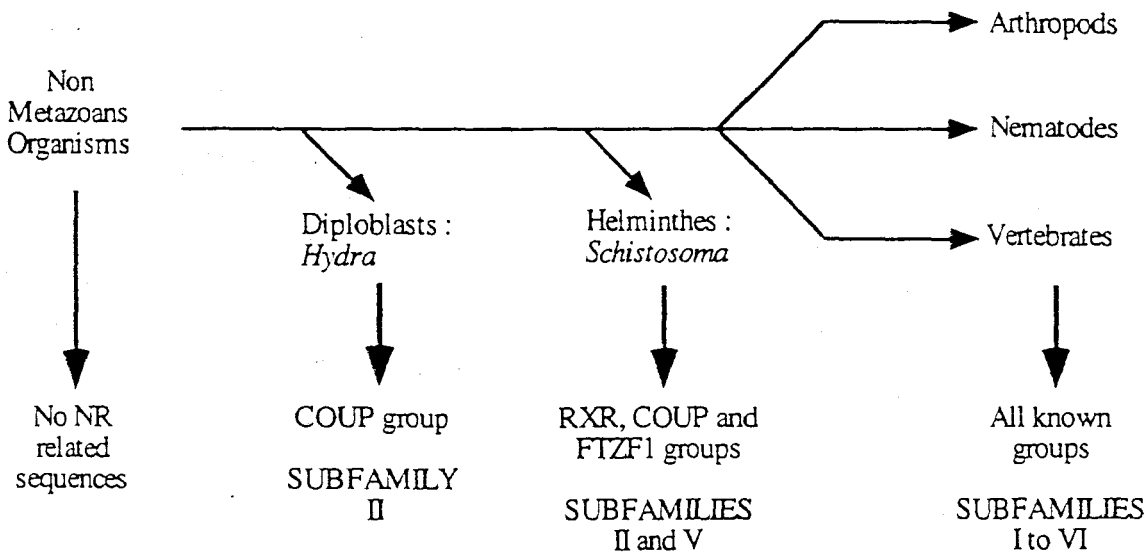


FIGURE 1. Unrooted phylogenetic tree connecting 30 representative nuclear receptors. The tree was calculated by the distance matrix analysis using the Neighbour-Joining method. One thousand bootstrap replicates were performed on each set and bootstrap values are indicated along each branch, expressed as a percentage. Bootstrap values of the branches that define subfamilies are *boxed*. Sequences were taken from Gronemeyer and Laudet.<sup>1</sup>

ones which have in common the characteristic that they cluster orphan receptors together: subfamily IV contains NGFIB receptors, subfamily V members of the FTZ-F1 group, and subfamily VI the unique GCNF1 orphan receptor. These six major branches are clearly distinguished on trees constructed with different phylogenetic methods such as distance analysis or parsimony. With each of these methods, bootstrap analysis to test the validity of each branch clearly shows that the most robust internal branches are those corresponding to these six subfamilies.

Interestingly, no clear relationship exists between the ligand-specificity of a receptor and its position in the phylogenetic tree. For example, TRs and RARs which recognize two structurally unrelated ligands, thyroid hormones and retinoic acid, are stably clustered together in all phylogenetic analyses. This is in contrast to recent claims of a possible prediction of ligand relatedness from the position in the tree.<sup>4</sup> In line with the previous point, orphan receptors are not clustered in a given subfamily but are randomly dispersed inside all the subfamilies. Taken together these ele-



**FIGURE 2.** Nuclear receptors are specific to metazoans. Interestingly, in our PCR-based strategy we identified only members of subfamilies II and V in early metazoans, suggesting that these subfamilies may be ancestral.

ments support the notion that nuclear receptors may have gained ligand binding ability during subsequent evolution.

From the phylogenetic tree we cannot directly infer a time scale from a given branch length because the length is proportional not to time but to sequence divergence. However, by comparing the splits between homologous receptors in different taxa (such as arthropods and vertebrates, or *Xenopus* or *Danio* compared to other vertebrates), we have some reference points in the tree. By doing this we can clearly observe two periods of gene duplication. The first which occurred very early on, before the arthropod/vertebrate split, corresponds to the appearance of the six subfamilies and within these to the appearance of the various groups of receptors (RXR, ECR, etc.). It should be noted that the superfamily is already highly diversified at the level of the arthropod/vertebrate split. The second wave of duplications took place specifically in the vertebrate lineage, after the arthropod/vertebrate split and corresponds to a diversification within each group of receptor, that is, to the appearance of the paralogous versions of a given receptor (RAR $\alpha$ , RAR $\beta$ , and RAR $\gamma$  or TR $\alpha$ , and TR $\beta$ ).

### IDENTIFICATION OF NUCLEAR RECEPTORS IN EARLY METAZOANS

By analogy with Hox genes, another gene superfamily that has been extensively studied from an evolutionary point of view,<sup>5</sup> it is tempting to suggest that the first nuclear receptors emerged when the first multicellular organisms evolved. This hypothesis has been tested by the polymerase chain reaction (PCR) identification of nuclear receptor homologs in early metazoans such as the diploblast *Hydra vulgaris* and the helminth *Schistosoma mansoni*. Using degenerate oligonucleotides directed towards a conserved exon of the DNA binding C domain and a "touch down" PCR

method, we were able to identify PCR fragments corresponding to the COUP-TF orphan receptor in *Hydra* and *Schistosoma* and the RXR and FTZ-F1 receptors in *Schistosoma*. The authenticity of these PCR products was confirmed by the isolation of full-size cDNA clones encoding these receptors. All our tentatives to identify nuclear receptor-related sequences in nonmetazoan organisms such as plants, algae, or fungi were unsuccessful. Thus, as for Hox/HOM genes, nuclear receptors appear to be specific for metazoans (FIG. 2).

Our results demonstrate that the molecular synapomorphy of a metazoan, the existence of a Hox/HOM gene complex, may be extended to the existence of a given set of gene families that could eventually form a regulatory network critical for embryonic development. Interestingly, in vertebrates, Hox gene transcription is regulated by retinoic acid receptors (RARs and RXRs). It is not known if this regulation was established earlier in evolution, but it is tempting to suggest a link between the homeotic genes which dictate cell identity in the developing embryo and nuclear receptors which are important regulators of cell-to-cell communication; both features are extremely important for a primitive multicellular organism. This network may have played a direct role in the emergence and diversification of metazoans. We thus now have to consider nuclear receptors, like Hox genes, as actors in the evolution of animals.

#### ACKNOWLEDGMENTS

We thank Dominique Stéhelin and André Capron for support, Catherine Hänni, Laurent Meijer, Brigitte Galliot, Murielle Fykry, and Pierre Saumitou-Laprade for invaluable help, and Rachid Safi for help in cloning/sequencing.

#### REFERENCES

1. GRONEMEYER H. & LAUDET V. 1995. Transcription factors. 3. Nuclear receptors. *Protein Profile* **2**: 1173–1308.
2. LAUDET V., HÄNNI C., COLL J., CATZEFLIS F. & STÉHELIN D. 1992. Evolution of the nuclear receptor gene superfamily. *EMBO J.* **11**: 1003–1013.
3. AMERO S.A., KRETSINGER R.H., MONCRIEF N.D., YAMAMOTO K.R. & PEARSON W.R. 1992. The origin of nuclear receptor proteins: A single precursor distinct from other transcription factors. *Mol. Endocrinol.* **6**: 3–7.
4. DETERA-WADLEIGH S.D. & FANNING T.G. 1994. Phylogeny of the steroid receptor superfamily. *Mol. Phyl. Evol.* **3**: 192–205.
5. HOLLAND P.W.H., GARCIA-FERNANDEZ J., WILLIAMS N.A. & SIDOW A. 1994. Gene duplications and the origins of vertebrate development. *In* The Evolution of Developmental Mechanisms. Akam M., Holland P., Ingham P. & Wray G., Eds. Development Supplement: 125–133.



*Bibliographie*  
*Récepteurs Nucléaires*

## Références

- Achatz, G., Holzl, B., Speckmayer, R., Hauser, C., Sandhofer, F. et Paulweber, B. (1997). Functional domains of the human orphan receptor ARP-1/COUP-TFII involved in active repression and transrepression. *Mol Cell Biol*, **17**, 4914-4932.
- Ali, S., Metzger, D., Bornert, J. M. et Chambon, P. (1993). Modulation of transcriptional activation by ligand-dependent phosphorylation of the human oestrogen receptor A/B region. *Embo J*, **12**, 1153-1160.
- Baniahmad, A., Leng, X., Burris, T. P., Tsai, S. Y., Tsai, M. J. et O'Malley, B. W. (1995). The tau 4 activation domain of the thyroid hormone receptor is required for release of a putative corepressor(s) necessary for transcriptional silencing. *Mol Cell Biol*, **15**, 76-86.
- Baretino, D., Vivanco, R. M. et Stunnenberg, H. G. (1994). Characterization of the ligand-dependent transactivation domain of thyroid hormone receptor. *Embo J*, **13**, 3039-3049.
- Beato, M. (1991). Transcriptional control by nuclear receptors. *Faseb J*, **5**, 2044-2051.
- Beato, M., Herrlich, P. et Schutz, G. (1995). Steroid hormone receptors: many actors in search of a plot. *Cell*, **83**, 851-857.
- Bird, A. et Tweedie, S. (1995). Transcriptional noise and the evolution of gene number. *Philos Trans R Soc Lond B Biol Sci*, **349**, 249-253.
- Blanco, J. C., Wang, I. M., Tsai, S. Y., Tsai, M. J., O'Malley, B. W., Jurutka, P. W., Haussler, M. R. et Ozato, K. (1995). Transcription factor TFIIB and the vitamin D receptor cooperatively activate ligand-dependent transcription. *Proc Natl Acad Sci U S A*, **92**, 1535-1539.
- Bode, H. R. (1992). Continuous conversion of neuron phenotype in hydra. *Trends Genet*, **8**, 279-284.
- Bode, H. R., Gee, L. W. et Chow, M. A. (1990). Neuron differentiation in hydra involves dividing intermediates. *Dev Biol*, **139**, 231-243.
- Bonnelye, E., Vanacker, J.-M., Dittmar, T., Begue, A., Desbiens, X., Denhardt, D. T., Aubin, J. E., Laudet, V. et Fournier, B. (1997). The ERR-1 orphan receptor is a transcriptional activator expressed during bone development. *Mol. Endocrinology*, **11**, 905-916.
- Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H. et Moras, D. (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha [see comments]. *Nature*, **375**, 377-382.
- Bradshaw, M. S., Tsai, S. Y., Leng, X. H., Dobson, A. D., Conneely, O. M., O'Malley, B. W. et Tsai, M. J. (1991). Studies on the mechanism of functional cooperativity between progesterone and estrogen receptors. *J Biol Chem*, **266**, 16684-16690.
- Brent, G. A., Dunn, M. K., Harney, J. W., Gulick, T., Larsen, P. R. et Moore, D. D. (1989). Thyroid hormone aporeceptor represses T3-inducible promoters and blocks activity of the retinoic acid receptor. *New Biol*, **1**, 329-336.
- Broadus, J. et Doe, C. Q. (1995). Evolution of neuroblast identity: seven-up and prospero expression reveal homologous and divergent neuroblast fates in Drosophila and Schistocerca. *Development*, **121**, 3989-3996.
- Brooke, N. M., Garcia, F. J. et Holland, P. W. (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature*, **392**, 920-922.
- Brunel, C., Romby, P., Westhof, E., Ehresmann, C. et Ehresmann, B. (1991). Three-dimensional model of Escherichia coli ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J Mol Biol*, **221**, 293-308.
- Bunone, G., Briand, P. A., Miksicek, R. J. et Picard, D. (1996). Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *Embo J*, **15**, 2174-2183.
- Butler, A. J. et Parker, M. G. (1995). COUP-TF II homodimers are formed in preference to heterodimers with RXR alpha or TR beta in intact cells. *Nucleic Acids Res*, **23**, 4143-4150.
- Candiani, S. et Pestarino, M. (1998). Expression of the tissue-specific transcription factor Pit-1 in the lancelet, Branchiostoma lanceolatum. *J Comp Neurol*, **392**, 343-351.
- Carosa, E., Fanelli, A., Ulisse, S., Di, L. R., Rall, J. E. et Jannini, E. A. (1998). Ciona intestinalis nuclear receptor 1: A member of steroid/thyroid hormone receptor family [In Process Citation]. *Proc Natl Acad Sci U S A*, **95**, 11152-11157.
- Cavailles, V., Dauvois, S., L'Horsset, F., Lopez, G., Hoare, S., Kushner, P. J. et Parker, M. G. (1995). Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *Embo J*, **14**, 3741-3751.
- Chan, S. M., Xu, N., Niemeyer, C. C., Bone, J. R. et Flytzanis, C. N. (1992). SpCOUP-TF: a sea urchin member of the steroid/thyroid hormone receptor family. *Proc Natl Acad Sci U S A*, **89**, 10568-10572.
- Chang, C., Da, S. S., Ideta, R., Lee, Y., Yeh, S. et Burbach, J. P. (1994). Human and rat TR4 orphan receptors specify a subclass of the steroid receptor superfamily. *Proc Natl Acad Sci U S A*, **91**, 6040-6044.
- Chang, C. et Kokontis, J. (1988). Identification of a new member of the steroid receptor super-family by cloning and sequence analysis. *Biochem Biophys Res Commun*, **155**, 971-977.
- Chang, C., Kokontis, J., Acakpo, S. L., Liao, S., Takeda, H. et Chang, Y. (1989). Molecular cloning of new human TR2 receptors: a class of steroid receptor with multiple ligand-binding domains. *Biochem Biophys Res Commun*, **165**, 735-741.
- Chen, J.-Y., Dzik, J., Edgecombe, G. D., Ramsköld, L. et Zhou, G.-Q. (1995). A possible early Cambrian chordate. *Nature*, **377**, 720-722.
- Chen, J. D. et Evans, R. M. (1995). A transcriptional co-repressor that interacts with nuclear hormone receptors [see comments]. *Nature*, **377**, 454-457.

- Clagett, D. M. et Plum, L. A. (1997). Retinoid-regulated gene expression in neural development. *Crit Rev Eukaryot Gene Expr*, **7**, 299-342.
- Concas, A., Pierobon, P., Mostallino, M. C., Porcu, P., Marino, G., Minei, R. et Biggio, G. (1998). Modulation of gamma-aminobutyric acid (GABA) receptors and the feeding response by neurosteroids in *Hydra vulgaris*. *Neuroscience*, **85**, 979-988.
- Connor, H., Nornes, H. et Neuman, T. (1995). Expression screening reveals an orphan receptor chick ovalbumin upstream promoter transcription factor I as a regulator of neurite/substrate-cell contacts and cell aggregation. *J Biol Chem*, **270**, 15066-15070.
- Cooney, A. J., Tsai, S. Y., O'Malley, B. W. et Tsai, M. J. (1992). Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTC A response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors. *Mol Cell Biol*, **12**, 4153-4163.
- Crawford, P. A., Dorn, C., Sadovsky, Y. et Milbrandt, J. (1998). Nuclear receptor DAX-1 recruits nuclear receptor corepressor N-CoR to steroidogenic factor 1. *Mol Cell Biol*, **18**, 2949-2956.
- Dahlman-Wright, K., Baumann, H., McEwan, I. J., Almlof, T., Wright, A. P., Gustafsson, J. A. et Hard, T. (1995). Structural characterization of a minimal functional transactivation domain from the human glucocorticoid receptor. *Proc Natl Acad Sci U S A*, **92**, 1699-1703.
- Dahlman-Wright, K., Siltala, R. H., Carlstedt, D. J. et Gustafsson, J. A. (1990). Protein-protein interactions facilitate DNA binding by the glucocorticoid receptor DNA-binding domain. *J Biol Chem*, **265**, 14030-14035.
- Danielian, P. S., White, R., Lees, J. A. et Parker, M. G. (1992). Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors [published erratum appears in *EMBO J* 1992 Jun;11(6):2366]. *Embo J*, **11**, 1025-1033.
- Danielsen, M., Hinck, L. et Ringold, G. M. (1989). Mutational analysis of the mouse glucocorticoid receptor. *Cancer Res.*, 2286s-2291s.
- Danielsen, M., Northrop, J. P., Jonklaas, J. et Ringold, G. M. (1987). Domains of the glucocorticoid receptor involved in specific and nonspecific deoxyribonucleic acid binding, hormone activation, and transcriptional enhancement. *Mol Endocrinol*, **1**, 816-822.
- Danielsen, M., Northrop, J. P. et Ringold, G. M. (1986). The mouse glucocorticoid receptor: mapping of functional domains by cloning, sequencing and expression of wild-type and mutant receptor proteins. *Embo J*, **5**, 2513-2522.
- David, C. N. et Gierer, A. (1974). Cell cycle kinetics and development of *Hydra attenuata*. III. Nerve and nematocyte differentiation. *J Cell Sci*, **16**, 359-375.
- Davis, L. E., Burnett, A. L. et Haynes, J. F. (1968). Histological and ultrastructural study of the muscular and nervous systems in *Hydra*. II. Nervous system. *J Exp Zool*, **167**, 295-331.
- de Thé, H., Vivanco Ruiz, M., Tiollais, P., Stunnenberg, H. et Dejean, A. (1990). Identification of a retinoic acid responsive element in the retinoic acid receptor beta gene. *Nature*, **343**, 177-180.
- Duboule, D. et Dolle, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. *Embo J*, **8**, 1497-1505.
- Duboule, D. et Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet*, **10**, 358-364.
- Dujon, B. (1996). The yeast genome project: what did we learn? *Trends Genet*, **12**, 263-270.
- Durand, B., Saunders, M., Gaudon, C., Roy, B., Losson, R. et Chambon, P. (1994). Activation function 2 (AF-2) of retinoic acid receptor and 9-cis retinoic acid receptor: presence of a conserved autonomous constitutive activating domain and influence of the nature of the response element on AF-2 activity. *Embo J*, **13**, 5370-5382.
- Escriva, H., Mendonca, R., Langlois, M.-C., Vanacker, J.-M., Pierce, R. et Laudet, V. (1997a). Systematique et evolution des récepteurs nucléaires d'hormones. *Biosystema*, **15**, 79-98.
- Escriva, H., Safi, R., Hanni, C., Langlois, M. C., Saumitou, L. P., Stehelin, D., Capron, A., Pierce, R. et Laudet, V. (1997b). Ligand binding was acquired during evolution of nuclear receptors. *Proc Natl Acad Sci U S A*, **94**, 6803-6808.
- Evans, R. M. et Hollenberg, S. M. (1988). Cooperative and positional independent trans-activation domains of the human glucocorticoid receptor. *Cold Spring Harb Symp Quant Biol*, **2**, 813-818.
- Fjose, A., Nornes, S., Weber, U. et Mlodzik, M. (1993). Functional conservation of vertebrate seven-up related genes in neurogenesis and eye development. *Embo J*, **12**, 1403-1414.
- Fjose, A., Weber, U. et Mlodzik, M. (1995). A novel vertebrate svp-related nuclear receptor is expressed as a step gradient in developing rhombomeres and is affected by retinoic acid. *Mech Dev*, **52**, 233-246.
- Folkers, G. E., van, der, Leede, B., van, der, Saag et Pt (1993). The retinoic acid receptor-beta 2 contains two separate cell-specific transactivation domains, at the N-terminus and in the ligand-binding domain. *Mol Endocrinol*, **7**, 616-627.
- Fraser, R. A., Rossignol, M., Heard, D. J., Egly, J. M. et Chambon, P. (1997). SUG1, a putative transcriptional mediator and subunit of the PA700 proteasome regulatory complex, is a DNA helicase. *J Biol Chem*, **272**, 7122-7126.
- Galliot, B., Welschof, M., Schuckert, O., Hoffmeister, S. et Schaller, H. C. (1995). The cAMP response element binding protein is involved in hydra regeneration. *Development*, **121**, 1205-1216.
- Garcia Fernandez, J. et Holland, P. W. (1994). Archetypal organization of the amphioxus Hox gene cluster. *Nature*, **370**, 563-566.

- Gauchat, D., Kreger, S., Holstein, T. et Galliot, B. (1998). *prdl-a*, a gene marker for hydra apical differentiation related to triploblastic paired-like head-specific genes. *Development*, **125**, 1637-1645.
- Gee, H. (1994). Vertebrate morphology. Return of the amphioxus [news; comment]. *Nature*, **370**, 504-505.
- Gerwin, N., La, R. A., Sauer, F., Halbritter, H. P., Neumann, M., Jackle, H. et Nauber, U. (1994). Functional and conserved domains of the *Drosophila* transcription factor encoded by the segmentation gene *knirps*. *Mol Cell Biol*, **14**, 7899-7908.
- Glardon, S., Holland, L. Z., Gehring, W. J. et Holland, N. D. (1998). Isolation and developmental expression of the amphioxus Pax-6 gene (*AmphiPax-6*): insights into eye and photoreceptor evolution. *Development*, **125**, 2701-2710.
- Glass, C. K. (1994). Differential recognition of target genes by nuclear receptor monomers, dimers, and heterodimers. *Endocr Rev*, **15**, 391-407.
- Grens, A., Mason, E., Marsh, J. L. et Bode, H. R. (1995). Evolutionary conservation of a cell fate specification gene: the Hydra *achaete-scute* homolog has proneural activity in *Drosophila*. *Development*, **121**, 4027-4035.
- Gronemeyer, H. et Laudet, V. (1995). Transcription factors 3: nuclear receptors. *Protein Profile*, **2**, 1173-1308.
- Guo, X., Xu, Q., Harmon, M. A., Jin, X., Laudet, V., Mangelsdorf, D. J. et Palmer, M. J. (1998). Isolation of two functional retinoid X receptor subtypes from the ixodid tick, *Amblyomma americanum* (L.). *Mol Cell Endocrinol*, **139**, 45-60.
- Hadzic, E., Desai, Y. V., Helmer, E., Guo, S., Wu, S., Koudinova, N., Casanova, J., Raaka, B. M. et Samuels, H. H. (1995). A 10-amino-acid sequence in the N-terminal A/B domain of thyroid hormone receptor alpha is essential for transcriptional activation and interaction with the general transcription factor TFIIB. *Mol Cell Biol*, **15**, 4507-4517.
- Harada, H., Kuboi, Y., Miki, R., Honda, C., Masushige, S., Nakatsuka, M., Koga, Y. et Kato, S. (1998). Cloning of rabbit TR4 and its bone cell-specific activity to suppress estrogen receptor-mediated transactivation. *Endocrinology*, **139**, 204-212.
- Hard, T., Dahlman, K., Carlstedt, D. J., Gustafsson, J. A. et Rigler, R. (1990a). Cooperativity and specificity in the interactions between DNA and the glucocorticoid receptor DNA-binding domain. *Biochemistry*, **29**, 5358-5364.
- Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt, D. J., Yamamoto, K. R., Gustafsson, J. A. et Kaptein, R. (1990b). Solution structure of the glucocorticoid receptor DNA-binding domain. *Science*, **249**, 157-160.
- Hassel, M. (1998). Upregulation of a *Hydra vulgaris* cPKC gene is tightly coupled to the differentiation of head structures. *Dev Genes Evol*, **207**, 489-501.
- Hassel, M., Bridge, D. M., Stover, N. A., Kleinholz, H. et Steele, R. E. (1998). The level of expression of a protein kinase C gene may be an important component of the patterning process in *Hydra*. *Dev Genes Evol*, **207**, 502-514.
- Heery, D. M., Kalkhoven, E., Hoare, S. et Parker, M. G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors [see comments]. *Nature*, **387**, 733-736.
- Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A., Stein, R. B., Eichele, G., Evans, R. M. et Thaller, C. (1992). 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor. *Cell*, **68**, 397-406.
- Hirose, T., Fujimoto, W., Tamaai, T., Kim, K. H., Matsuura, H. et Jetten, A. M. (1994). TAK1: molecular cloning and characterization of a new member of the nuclear receptor superfamily. *Mol Endocrinol*, **8**, 1667-1680.
- Holland, L. Z. et Holland, N. D. (1996). Expression of *AmphiHox-1* and *AmphiPax-1* in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. *Development*, **122**, 1829-1838.
- Holland, P. W. et Garcia-Fernandez, J. (1996). Hox genes and chordate evolution. *Dev Biol*, **173**, 382-395.
- Holland, P. W., Koschorz, B., Holland, L. Z. et Herrmann, B. G. (1995). Conservation of *Brachyury* (T) genes in amphioxus and vertebrates: developmental and evolutionary implications. *Development*, **121**, 4283-4291.
- Hollenberg, S. M. et Evans, R. M. (1988). Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. *Cell*, **55**, 899-906.
- Hollenberg, S. M., Giguere, V., Segui, P. et Evans, R. M. (1987). Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. *Cell*, **49**, 39-46.
- Hörlein, A. J., Naar, A. M., Heinzl, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Soderstrom, M., Glass, C. K. et al. (1995). Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor [see comments]. *Nature*, **377**, 397-404.
- Ing, N. H., Beekman, J. M., Tsai, S. Y., Tsai, M. J. et O'Malley, B. W. (1992). Members of the steroid hormone receptor superfamily interact with TFIIB (S300-II). *J Biol Chem*, **267**, 17617-17623.
- Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P. et Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell*, **79**, 107-117.
- Janknecht, R. et Hunter, T. (1996). Transcription. A growing coactivator network [news; comment]. *Nature*, **383**, 22-23.
- Janvier, M. et Grimont, P. A. (1995). The genus *Methylophaga*, a new line of descent within phylogenetic branch gamma of Proteobacteria. *Res Microbiol*, **146**, 543-550.
- Jonat, C., Rahmsdorf, H. J., Park, K. K., Cato, A. C., Gebel, S., Ponta, H. et Herrlich, P. (1990). Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone. *Cell*, **62**, 1189-1204.

- Jonk, L. J., de, J. M., Pals, C. E., Wissink, S., Vervaart, J. M., Schoorlemmer, J. et Kruijer, W. (1994a). Cloning and expression during development of three murine members of the COUP family of nuclear orphan receptors. *Mech Dev*, **47**, 81-97.
- Jonk, L. J., de, J. M., Vervaart, J. M., Wissink, S. et Kruijer, W. (1994b). Isolation and developmental expression of retinoic-acid-induced genes. *Dev Biol*, **161**, 604-614.
- Kamei, Y., Xu, L., Heinzl, T., Torchia, J., Kurokawa, R., Gloss, B., Lin, S. C., Heyman, R. A., Rose, D. W., Glass, C. K. et Rosenfeld, M. G. (1996). A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell*, **85**, 403-414.
- Kappen, C., Schughart, K. et Ruddle, F. H. (1989). Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. *Proc Natl Acad Sci U S A*, **86**, 5459-5463.
- Katahira, M., Knegt, R., Schilthuis, J., Boelens, R., Eib, D., van, der, Saag, P et Kaptein, R. (1992). The structure of the human retinoic acid receptor-beta DNA-binding domain determined by NMR. *Nucleic Acids Symp Ser*, **1992**, 65-66.
- Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., Uchiyama, S., Sasaki, H., Masushige, S., Gotoh, Y., Nishida, E., Kawashima, H. et et, a. l. (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science*, **270**, 1491-1494.
- Keller, H., Mahfoudi, A., Dreyer, C., Hihi, A. K., Medin, J., Ozato, K. et Wahli, W. (1993). Peroxisome proliferator-activated receptors and lipid metabolism. *Ann N Y Acad Sci*, **684**, 157-173.
- Kliewer, S. A., Umesono, K., Heyman, R. A., Mangelsdorf, D. J., Dyck, J. A. et Evans, R. M. (1992). Retinoid X receptor-COUP-TF interactions modulate retinoic acid signaling. *Proc Natl Acad Sci U S A*, **89**, 1448-1452.
- Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R. et Chambon, P. (1987). Functional domains of the human estrogen receptor. *Cell*, **51**, 941-951.
- Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M. G., Heyman, R. A. et Glass, C. K. (1994). Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature*, **371**, 528-531.
- Kurokawa, R., Soderstrom, M., Horlein, A., Halachmi, S., Brown, M., Rosenfeld, M. G. et Glass, C. K. (1995). Polarity-specific activities of retinoic acid receptors determined by a co-repressor [see comments]. *Nature*, **377**, 451-454.
- Kurokawa, R., Yu, V. C., Naar, A., Kyakumoto, S., Han, Z., Silverman, S., Rosenfeld, M. G. et Glass, C. K. (1993). Differential orientations of the DNA-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. *Genes Dev*, **7**, 1423-1435.
- Lacalli, T. C. (1996). Frontal eye circuitry, rostral sensory pathways and brain organization in amphioxus larvae: evidence from 3D reconstructions. *Phil. trans. R. Soc. Lond.*, **351**, 243-263.
- Lacalli, T. C., Holland, N. D. et West, J. E. (1994). Landmarks in the anterior central nervous system of amphioxus larvae. *Phil. Trans. R. Soc. Lond.*, **B344**, 165-185.
- Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J Mol Endocrinol*, **19**, 207-226.
- Laudet, V., Hanni, C., Coll, J., Catzeflis, F. et Stehelin, D. (1992). Evolution of the nuclear receptor gene superfamily. *Embo J*, **11**, 1003-1013.
- Law, S. W., Conneely, O. M. et O'Malley, B. W. (1994). Molecular cloning of a novel member of the nuclear receptor superfamily related to the orphan receptor, TR2. *Gene Expr*, **4**, 77-84.
- Le Douarin, B., Pierrat, B., vom, B. E., Chambon, P. et Losson, R. (1995). A new version of the two-hybrid assay for detection of protein-protein interactions. *Nucleic Acids Res*, **23**, 876-878.
- Le Douarin, B., vom Baur, E., Zechel, C., Heery, D., Heine, M., Vivat, V., Gronemeyer, H., Losson, R. et Chambon, P. (1996a). Ligand-dependent interaction of nuclear receptors with potential transcriptional intermediary factors (mediators). *Philos Trans R Soc Lond B Biol Sci*, **351**, 569-578.
- Le Douarin, B., vom, B. E., Zechel, C., Heery, D., Heine, M., Vivat, V., Gronemeyer, H., Losson, R. et Chambon, P. (1996b). Ligand-dependent interaction of nuclear receptors with potential transcriptional intermediary factors (mediators). *Philos Trans R Soc Lond B Biol Sci*, **351**, 569-578.
- Lee, C. H., Chang, L. et Wei, L. N. (1996a). Molecular cloning and characterization of a mouse nuclear orphan receptor expressed in embryos and testes. *Mol Reprod Dev*, **44**, 305-314.
- Lee, C. H., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. et Wei, L. N. (1995a). Genomic structure, promoter identification, and chromosomal mapping of a mouse nuclear orphan receptor expressed in embryos and adult testes. *Genomics*, **30**, 46-52.
- Lee, H. J., Young, W. J., Shih, C. Y. et Chang, C. (1996b). Suppression of the human erythropoietin gene expression by the TR2 orphan receptor, a member of the steroid receptor superfamily. *J Biol Chem*, **271**, 10405-10412.
- Lee, M. S., Kliewer, S. A., Provencal, J., Wright, P. E. et Evans, R. M. (1993). Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding. *Science*, **260**, 1117-1121.
- Lee, S. L., Wesselschmidt, R. L., Linette, G. P., Kanagawa, O., Russell, J. H. et Milbrandt, J. (1995b). Unimpaired thymic and peripheral T cell death in mice lacking the nuclear receptor NGFI-B (Nur77). *Science*, **269**, 532-535.
- Lees, J. A., Fawell, S. E. et Parker, M. G. (1989). Identification of constitutive and steroid-dependent transactivation domains in the mouse oestrogen receptor. *J Steroid Biochem*, **34**, 33-39.

- Levin, A. A., Sturzenbecker, L. J., Kazmer, S., Bosakowski, T., Huselton, C., Allenby, G., Speck, J., Kratzeisen, C., Rosenberger, M., Lovey, A. et al. (1992). 9-cis retinoic acid stereoisomer binds and activates the nuclear receptor RXR alpha. *Nature*, **355**, 359-361.
- Li, H., Leo, C., Schroen, D. J. et Chen, J. D. (1997). Characterization of receptor interaction and transcriptional repression by the corepressor SMRT. *Mol Endocrinol*, **11**, 2025-2037.
- Lin, K. H., Parkison, C., McPhie, P. et Cheng, S. Y. (1991). An essential role of domain D in the hormone-binding activity of human beta 1 thyroid hormone nuclear receptor. *Mol Endocrinol*, **5**, 485-492.
- Lopes da Silva, S., Cox, J. J., Jonk, L. J., Kruijer, W. et Burbach, J. P. (1995). Localization of transcripts of the related nuclear orphan receptors COUP-TF I and ARP-1 in the adult mouse brain. *Brain Res Mol Brain Res*, **30**, 131-136.
- Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. et Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA [see comments]. *Nature*, **352**, 497-505.
- Lundback, T., Zilliacus, J., Gustafsson, J. A., Carlstedt, D. J. et Hard, T. (1994). Thermodynamics of sequence-specific glucocorticoid receptor-DNA interactions. *Biochemistry*, **33**, 5955-5965.
- Lundin, L. G. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics*, **16**, 1-19.
- Lutz, B., Kuratani, S., Cooney, A. J., Wawersik, S., Tsai, S. Y., Eichele, G. et Tsai, M. J. (1994). Developmental regulation of the orphan receptor COUP-TF II gene in spinal motor neurons. *Development*, **120**, 25-36.
- Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. et al. (1995). The nuclear receptor superfamily: the second decade. *Cell*, **83**, 835-839.
- Mangelsdorf, D. J., Umesono, K., Kliewer, S. A., Borgmeyer, U., Ong, E. S. et Evans, R. M. (1991). A direct repeat in the cellular retinol-binding protein type II gene confers differential regulation by RXR and RAR. *Cell*, **66**, 555-561.
- Martinez, D. E., Dirksen, M. L., Bode, P. M., Jamrich, M., Steele, R. E. et Bode, H. R. (1997). Budhead, a fork head/HNF-3 homologue, is expressed during axis formation and head specification in hydra. *Dev Biol*, **192**, 523-536.
- Maruyama, K., Tsukada, T., Ohkura, N., Bandoh, S., Hosono, T. et Yamaguchi, K. (1998). The NGFI-B subfamily of the nuclear receptor superfamily (review). *Int J Oncol*, **12**, 1237-1243.
- McCauley, D. W. (1997). Serotonin plays an early role in the metamorphosis of the hydrozoan *Phialidium gregarium*. *Dev Biol*, **190**, 229-240.
- Mitchell, P. J. et Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371-378.
- Miyajima, N., Kadowaki, Y., Fukushige, S., Shimizu, S., Semba, K., Yamanashi, Y., Matsubara, K., Toyoshima, K. et Yamamoto, T. (1988). Identification of two novel members of erbA superfamily by molecular cloning: the gene products of the two are highly related to each other. *Nucleic Acids Res*, **16**, 11057-11074.
- Müller, W. A. (1996). Pattern formation in the immortal Hydra. *Trends Genet*, **12**, 91-96.
- Näär, A. M., Boutin, J. M., Lipkin, S. M., Yu, V. C., Holloway, J. M., Glass, C. K. et Rosenfeld, M. G. (1991). The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell*, **65**, 1267-1279.
- Nagpal, S., Friant, S., Nakshatri, H. et Chambon, P. (1993). RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization in vivo. *Embo J*, **12**, 2349-2360.
- Nakshatri, H. et Chambon, P. (1994). The directly repeated RG(G/T)TCA motifs of the rat and mouse cellular retinol-binding protein II genes are promiscuous binding sites for RAR, RXR, HNF-4, and ARP-1 homo- and heterodimers. *J Biol Chem*, **269**, 890-902.
- Neuman, K., Soosaar, A., Nornes, H. O. et Neuman, T. (1995). Orphan receptor COUP-TF I antagonizes retinoic acid-induced neuronal differentiation. *J Neurosci Res*, **41**, 39-48.
- Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. et Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, **87**, 953-959.
- Ohno, S., Wolf, U. et Atkin, N. B. (1968). Evolution from fish to mammals by gene duplication. *Hereditas*, **59**, 169-187.
- Onate, S. A., Tsai, S. Y., Tsai, M. J. et O'Malley, B. W. (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science*, **270**, 1354-1357.
- Pani, A. K. et Anctil, M. (1994). Evidence for biosynthesis and catabolism of monoamines in the sea pansy *Renilla koellikeri* (Cnidaria). *Neurochem Int*, **25**, 465-474.
- Parker, M. G. (1990). Structure and function of nuclear hormone receptors. *Semin Cancer Biol*, **1**, 81-87.
- Pazin, M. J. et Kadonaga, J. T. (1997a). SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell*, **88**, 737-740.
- Pazin, M. J. et Kadonaga, J. T. (1997b). What's up and down with histone deacetylation and transcription? *Cell*, **89**, 325-328.
- Pereira, F. A., Qiu, Y., Tsai, M. J. et Tsai, S. Y. (1995). Chicken ovalbumin upstream promoter transcription factor (COUP-TF): expression during mouse embryogenesis. *J Steroid Biochem Mol Biol*, **53**, 503-508.

- Perlmann, T. et Jansson, L. (1995). A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1. *Genes Dev*, **9**, 769-782.
- Perlmann, T., Rangarajan, P. N., Umesono, K. et Evans, R. M. (1993). Determinants for selective RAR and TR recognition of direct repeat HREs. *Genes Dev*, **7**, 1411-1422.
- Perlmann, T., Umesono, K., Rangarajan, P. N., Forman, B. M. et Evans, R. M. (1996). Two distinct dimerization interfaces differentially modulate target gene specificity of nuclear hormone receptors. *Mol Endocrinol*, **10**, 958-966.
- Qiu, Y., Cooney, A. J., Kuratani, S., DeMayo, F. J., Tsai, S. Y. et Tsai, M. J. (1994). Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon. *Proc Natl Acad Sci U S A*, **91**, 4451-4455.
- Qiu, Y., Krishnan, V., Zeng, Z., Gilbert, D. J., Copeland, N. G., Gibson, L., Yang, F. T., Jenkins, N. A., Tsai, M. J. et Tsai, S. Y. (1995). Isolation, characterization, and chromosomal localization of mouse and human COUP-TF I and II genes. *Genomics*, **29**, 240-246.
- Qiu, Y., Pereira, F. A., DeMayo, F. J., Lydon, J. P., Tsai, S. Y. et Tsai, M. J. (1997). Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. *Genes Dev*, **11**, 1925-1937.
- Rastinejad, F., Perlmann, T., Evans, R. M. et Sigler, P. B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats [see comments]. *Nature*, **375**, 203-211.
- Renaud, J. P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H. et Moras, D. (1995). Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature*, **378**, 681-689.
- Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*, **21**, 327-335.
- Ruiz, M. S. et Anadon, R. (1991). Some considerations on the fine structure of rhabdomeric photoreceptors in the amphioxus, *Branchiostoma lanceolatum* (Cephalochordata). *J Hirnforsch*, **32**, 159-164.
- Rusconi, S. et Yamamoto, K. R. (1987). Functional dissection of the hormone and DNA binding activities of the glucocorticoid receptor. *Embo J*, **6**, 1309-1315.
- Schaller, H. C., Hoffmeister, S. A. et Dubel, S. (1989). Role of the neuropeptide head activator for growth and development in hydra and mammals. *Development*, **99**, 107.
- Schoonjans, K., Martin, G., Staels, B. et Auwerx, J. (1997). Peroxisome proliferator-activated receptors, orphans with ligands and functions. *Curr Opin Lipidol*, **8**, 159-166.
- Schule, R., Umesono, K., Mangelsdorf, D. J., Bolado, J., Pike, J. W. et Evans, R. M. (1990). Jun-Fos and receptors for vitamins A and D recognize a common response element in the human osteocalcin gene. *Cell*, **61**, 497-504.
- Schulman, I. G., Chakravarti, D., Juguilon, H., Romo, A. et Evans, R. M. (1995). Interactions between the retinoid X receptor and a conserved region of the TATA-binding protein mediate hormone-dependent transactivation. *Proc Natl Acad Sci U S A*, **92**, 8288-8292.
- Schummer, M., Scheurlen, I., Schaller, C. et Galliot, B. (1992). HOM/HOX homeobox genes are present in hydra (*Chlorohydra viridissima*) and are differentially expressed during regeneration. *Embo J*, **11**, 1815-1823.
- Schwabe, J. W., Neuhaus, D. et Rhodes, D. (1990). Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature*, **348**, 458-461.
- Scott, P. C., Maddox, J. F., Gogolin, E. K. et Brandon, M. R. (1992). The nucleotide sequence and evolution of ovine MHC class II B genes: DQB and DRB. *Immunogenetics*, **35**, 217.
- Seol, W., Choi, H. S. et Moore, D. D. (1995). Isolation of proteins that interact specifically with the retinoid X receptor: two novel orphan receptors. *Mol Endocrinol*, **9**, 72-85.
- Sharman, A. C. et Holland, P. W. (1998). Estimation of Hox gene cluster number in lampreys. *Int J Dev Biol*, **42**, 617-620.
- Shemshedini, L., Ji, J. W., Brou, C., Chambon, P. et Gronemeyer, H. (1992). In vitro activity of the transcription activation functions of the progesterone receptor. Evidence for intermediary factors. *J Biol Chem*, **267**, 1834-1839.
- Shenk, M. A., Bode, H. R. et Steele, R. E. (1993a). Expression of Cnox-2, a HOM/HOX homeobox gene in hydra, is correlated with axial pattern formation. *Development*, **117**, 657-667.
- Shenk, M. A., Gee, L., Steele, R. E. et Bode, H. R. (1993b). Expression of Cnox-2, a HOM/HOX gene, is suppressed during head formation in hydra. *Dev Biol*, **160**, 108-118.
- Shibata, H., Nawaz, Z., Tsai, S. Y., O'Malley, B. W. et Tsai, M. J. (1997). Gene silencing by chicken ovalbumin upstream promoter-transcription factor I (COUP-TFI) is mediated by transcriptional corepressors, nuclear receptor-corepressor (N-CoR) and silencing mediator for retinoic acid receptor and thyroid hormone receptor (SMRT). *Mol Endocrinol*, **11**, 714-724.
- Shimeld, S. M. (1997). Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. *Dev Biol*, **183**, 74-85.
- Shimeld, S. M., McKay, I. J. et Sharpe, P. T. (1996). The murine homeobox gene *Msx-3* shows highly restricted expression in the developing neural tube. *Mech Dev*, **55**, 201-210.
- Slack, J. M., Holland, P. W. et Graham, C. F. (1993). The zootype and the phylotypic stage. *Nature*, **361**, 490-492.
- Spruyt, N., Delarbre, C., Gachelin, G. et Laudet, V. (1998). Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome: relations to vertebrates. *Nucleic Acids Res*, **26**, 3279-3285.

- Stock, D. W. et Whitt, G. S. (1992). Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. *Science*, **257**, 787-789.
- Stokes, M. D. et Holland, N. D. (1995). Ciliary hovering in larva lancelets. *Bio. Bull.*, **188**, 231-233.
- Sucov, H. M. et Evans, R. M. (1995). Retinoic acid and retinoic acid receptors in development. *Mol Neurobiol*, **10**, 169-184.
- Swaffield, J. C., Bromberg, J. F. et Johnston, S. A. (1992). Alterations in a yeast protein resembling HIV Tat-binding protein relieve requirement for an acidic activation domain in GAL4 [published erratum appears in Nature 1992 Dec 24-31;360(6406):768]. *Nature*, **357**, 698-700.
- Swaffield, J. C., Melcher, K. et Johnston, S. A. (1995). A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein. *Nature*, **374**, 88-91.
- Tanese, N. et Tjian, R. (1993). Coactivators and TAFs: a new class of eukaryotic transcription factors that connect activators to the basal machinery. *Cold Spring Harb Symp Quant Biol*, **58**, 179-185.
- Tansey, W. P. et Herr, W. (1997). TAFs: guilt by association? *Cell*, **88**, 729-732.
- Thompson, C. C. et Evans, R. M. (1989). Trans-activation by thyroid hormone receptors: functional parallels with steroid hormone receptors. *Proc Natl Acad Sci U S A*, **86**, 3494-3498.
- Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E. et Chambon, P. (1989). The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell*, **59**, 477-487.
- Tran, P., Zhang, X. K., Salbert, G., Hermann, T., Lehmann, J. M. et Pfahl, M. (1992). COUP orphan receptors are negative regulators of retinoic acid response pathways. *Mol Cell Biol*, **12**, 4666-4676.
- Tsai, S. Y. et Tsai, M. J. (1997). Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): coming of age. *Endocr Rev*, **18**, 229-240.
- Tugwood, J. D., Issemann, I., Anderson, R. G., Bundell, K. R., McPheat, W. L. et Green, S. (1992). The mouse peroxisome proliferator activated receptor recognizes a response element in the 5' flanking sequence of the rat acyl CoA oxidase gene. *Embo J*, **11**, 433-439.
- Ueki, T., Kuratani, S., Hirano, S. et Aizawa, S. (1998). Otx cognates in a lamprey, *Lampetra japonica*. *Dev Genes Evol*, **208**, 223-228.
- Umesono, K. et Evans, R. M. (1989). Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell*, **57**, 1139/1146.
- van der Wees, J., Matharu, P. J., de, R. K., Destree, O. H., Godsave, S. F., Durston, A. J. et Sweeney, G. E. (1996). Developmental expression and differential regulation by retinoic acid of *Xenopus* COUP-TF-A and COUP-TF-B. *Mech Dev*, **54**, 173-184.
- vom Baur, E., Zechel, C., Heery, D., Heine, M. J., Garnier, J. M., Vivat, V., Le, D. B., Gronemeyer, H., Chambon, P. et Losson, R. (1996). Differential ligand-dependent interactions between the AF-2 activating domain of nuclear receptors and the putative transcriptional intermediary factors mSUG1 and TIF1. *Embo J*, **15**, 110-124.
- Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D. et Fletterick, R. J. (1995). A structural role for hormone in the thyroid hormone receptor. *Nature*, **378**, 690-697.
- Wang, L. H., Tsai, S. Y., Cook, R. G., Beattie, W. G., Tsai, M. J. et O'Malley, B. W. (1989). COUP transcription factor is a member of the steroid receptor superfamily. *Nature*, **340**, 163-166.
- Webster, N. J., Green, S., Jin, J. R. et Chambon, P. (1988). The hormone binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function. *Cell*, **54**, 199-207.
- Weinziger, R., Salgado, L. M., David, C. N. et Bosch, T. C. (1994). Ksl, an epithelial cell-specific gene, responds to early signals of head formation in *Hydra*. *Development*, **120**, 2511-2517.
- Williams, N. A. et Holland, P. W. (1998). Gene and domain duplication in the chordate Otx gene family: insights from amphioxus Otx. *Mol Biol Evol*, **15**, 600-607.
- Wilson, T. E., Fahrner, T. J. et Milbrandt, J. (1993a). The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. *Mol Cell Biol*, **13**, 5794-5804.
- Wilson, T. E., Mouw, A. R., Weaver, C. A., Milbrandt, J. et Parker, K. L. (1993b). The orphan nuclear receptor NGFI-B regulates expression of the gene encoding steroid 21-hydroxylase. *Mol Cell Biol*, **13**, 861-868.
- Winston, F. et Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet*, **8**, 387-391.
- Wirtanen, L., Huard, V. et Seguin, C. (1997). Molecular cloning from neurulating *Ambystoma mexicanum* embryos of the cDNA encoding an orphan nuclear receptor (aDOR1) closely related to TR2- 11. *Differentiation*, **62**, 159-170.
- Wolffe, A. P. et Pruss, D. (1996). Targeting chromatin disruption: Transcription regulators that acetylate histones. *Cell*, **84**, 817-819.
- Wu, Q., Li, Y., Liu, R., Agadir, A., Lee, M. O., Liu, Y. et Zhang, X. (1997). Modulation of retinoic acid sensitivity in lung cancer cells through dynamic balance of orphan receptors nur77 and COUP-TF and their heterodimerization. *Embo J*, **16**, 1656-1669.
- Wurtz, J. M., Bourguet, W., Renaud, J. P., Vivat, V., Chambon, P., Moras, D. et Gronemeyer, H. (1996). A canonical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol*, **3**, 206.
- Yang, N., Schule, R., Mangelsdorf, D. J. et Evans, R. M. (1991). Characterization of DNA binding and retinoic acid binding properties of retinoic acid receptor. *Proc Natl Acad Sci U S A*, **88**, 3559-3563.



- Yao, T. P., Oh, S. P., Fuchs, M., Zhou, N. D., Ch'ng, L. E., Newsome, D., Bronson, R. T., Li, E., Livingston, D. M. et Eckner, R. (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell*, **93**, 361-372.
- Yao, T. P., Segraves, W. A., Oro, A. E., McKeown, M. et Evans, R. M. (1992). Drosophila ultraspiracle modulates ecdysone receptor function via heterodimer formation. *Cell*, **71**, 63-72.
- Yasuo, H. et Satoh, N. (1993). Function of vertebrate T gene [letter]. *Nature*, **364**, 582-583.
- Yoshikawa, T., DuPont, B. R., Leach, R. J. et Detera, W. S. (1996a). New variants of the human and rat nuclear hormone receptor, TR4: expression and chromosomal localization of the human gene. *Genomics*, **35**, 361-366.
- Yoshikawa, T., Makino, S., Gao, X. M., Xing, G. Q., Chuang, D. M. et Detera, W. S. (1996b). Splice variants of rat TR4 orphan receptor: differential expression of novel sequences in the 5'-untranslated region and C-terminal domain. *Endocrinology*, **137**, 1562-1571.
- Zanaria, E., Muscatelli, F., Bardoni, B., Strom, T. M., Guioli, S., Guo, W., Lalli, E., Moser, C., Walker, A. P., McCabe, E. R. et al. (1994). An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature*, **372**, 635-641.
- Zechel, C., Shen, X. Q., Chambon, P. et Gronemeyer, H. (1994a). Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. *Embo J*, **13**, 1414-1424.
- Zechel, C., Shen, X. Q., Chen, J. Y., Chen, Z. P., Chambon, P. et Gronemeyer, H. (1994b). The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *Embo J*, **13**, 1425-1433.
- Zenke, M., Munoz, A., Sap, J., Vennstrom, B. et Beug, H. (1990). v-erbA oncogene activation entails the loss of hormone-dependent regulator activity of c-erbA. *Cell*, **61**, 1035-1049.
- Zhang, X. K., Hoffmann, B., Tran, P. B., Graupner, G. et Pfahl, M. (1992). Retinoid X receptor is an auxiliary protein for thyroid hormone and retinoic acid receptors. *Nature*, **355**, 441-446.
- Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M. A. et Rastinejad, F. (1998). Structural elements of an orphan nuclear receptor-DNA complex. *Mol Cell*, **1**, 849-861.
- Zilliacus, J., Dahlman, W. K., Carlstedt, D. J. et Gustafsson, J. A. (1992). Zinc coordination scheme for the C-terminal zinc binding site of nuclear hormone receptors. *J Steroid Biochem Mol Biol*, **42**, 131-139.

*Conclusion  
générale*

## Discussion et conclusion générale

Pour un développement harmonieux de l'embryon, les gènes impliqués dans les processus du développement doivent non seulement être soumis à une régulation spatio-temporelle bien déterminée, mais aussi être exprimés selon des concentrations bien précises dépendant des régions et des territoires dans lesquels ils agissent. Nous avons vu dans la première partie de ce manuscrit, l'exemple du gène Pax-6, étudié dans le groupe de Simon Saule. Nous avons étudié dans le premier article présenté dans cette thèse, comment un élément *cis*, dénommé "enhancer  $\alpha$ ", intervient dans la régulation de l'expression spatio-temporelle du gène Pax-6 de caille. Cet enhancer  $\alpha$  active en effet l'un des promoteurs du gène Pax-6 de caille, le promoteur P0 et son action est spécifique des cellules de neurorétine et efficace, seulement à partir de 7 jours de développement embryonnaire. Nous avons montré que la fonction de cet enhancer était conservée chez la souris. La transcription de Pax-6 est par ailleurs activée, *via* ses deux promoteurs P0 et P1, par c-Myb, un gène, connu pour ses implications dans les lignées lymphoïdes et exprimé, fortement, dans le neurorétine et dans les neuroblastes n'ayant pas encore atteint leur stade de différenciation terminale. Nous pouvons supposer, d'après ces résultats, que c-Myb pourrait jouer un rôle non négligeable dans l'initiation de la transcription du gène Pax-6 dans la neurorétine et dans les régions du système nerveux central où ces deux gènes sont co-exprimés. Pax-6 est par ailleurs lui-même capable d'entretenir sa transcription par l'intermédiaire de l'une de ces protéines, la "P46<sup>Pax-6</sup>" qui transactive les promoteurs P1 et P0. De plus, la protéine Engrailed a été identifiée comme un facteur capable de réprimer la transcription du gène Pax-6 autoactivée par la P46. La protéine Engrailed interagit en effet avec la P46, et diminue alors l'affinité de cette dernière pour sa séquence de reconnaissance sur l'ADN. Cette étude a fait l'objet du second article présenté dans cette thèse.

Le gène Pax-6 est un gène dont les effets au cours du développement embryonnaire sont connus pour dépendre de la dose de protéines Pax produites. On observe effectivement des défauts du développements engendrés soit par la mutation d'un seul allèle du gène, soit par l'introduction d'une construction générant des copies surnuméraires du gène. Dans ce contexte, il est intéressant de

noter que d'autres mécanismes que les mécanismes de régulation transcriptionnels peuvent conduire à des variations dans les concentrations des différentes protéines isoformes du gène Pax-6. Nous avons notamment caractérisé un ARNm particulier du gène Pax-6, l'ARNm B1, pourvu d'un long 5'UTR, doté des exons codant les boîtes paired et homéo, mais ne traduisant que des protéines à domaine paired à partir d'AUG internes. Le rendement de la traduction de ces petites protéines est néanmoins nettement supérieur à partir de l'ARN B1 qu'à partir des autres ARNm du gène Pax-6. Ces derniers traduisent aussi des protéines à domaine paired et homéo, à partir de leur phase ouverte de lecture en 5'. Le long 5'UTR du B1 pourrait favoriser la traduction de ces protéines dépourvues de domaine paired, par un processus, connu sous le nom de "leaky scanning" : un mécanisme qui favorise le recrutement des ribosomes par des codons de traduction alternatifs situés dans le 5'UTR et facilite le glissement du complexe de traduction vers des codons meilleurs pour l'initiation, comme les AUG situés en amont de la boîte homéo, dans le cas de l'ARNm B1. Le rôle de ces petites protéines n'est pas encore connu chez les vertébrés. Chez le nématode *Caenorhabditis elegans*, elles jouent un rôle dans la détermination des organes sensoriels (cils vibratiles situés à la tête de l'animal et qui lui permettent de localiser sa nourriture). Elles sont cependant produites par un mécanisme de régulation différent chez le ver, puisqu'elles sont générées par l'existence d'un promoteur interne situé entre les boîtes paired et homéo. La fonction d'un gène est complexe et ne peut être assimilée dans son ensemble qu'en tenant compte des différentes protéines produites et des proportions de ces dernières dans les différentes régions d'expression du gène au cours du développement de l'embryon. S'ajoutent à ces effets d'autres facteurs de transcription qui interfèrent ou agissent en synergie avec le facteur de transcription considéré et appartiennent au même réseau de régulation et qui vont moduler son activité directement ou indirectement. Le gène Pax-6 peut permettre la formation d'un oeil dans des disque imaginaux de drosophile, pourtant, il n'induit pas la formation d'yeux dans toutes les régions où il s'exprime. D'où l'importance de prendre en compte le réseau de régulation auquel il participe pour comprendre sa fonction dans un tissu donné.

La fonction du gène Pax-6 dans l'oeil et dans le pancréas est assez bien documentée. Cependant, son rôle dans le développement du cerveau et même dans le cerveau adulte où il est exprimé, reste obscur. Comme je l'ai expliqué ci-dessus, pour comprendre son rôle dans le cerveau, il semble indispensable à l'heure actuelle, de replacer ce gène dans le contexte des régulations auxquelles il est soumis et celles auxquelles il participe à un temps du développement donné et pour un organe considéré. Parmi, les acteurs putatifs qui pourraient coopérer à sa fonction dans le système nerveux central, citons le cas des gènes COUP-TFs, des facteurs de transcription appartenant à la famille de récepteurs nucléaires. Les gènes COUP-TFs s'expriment notamment dans les mêmes territoires du cerveau que Pax-6, chez la souris, le télencéphale, le diencéphale, les neurones moteurs de la moelle épinière pour l'embryon, et dans le thalamus, pour la souris adulte. Ces deux gènes Pax-6 et COUP-TF, sont par ailleurs tous deux initialement exprimés à partir de 8,5 jours de gestation chez la souris, subissent tous deux un pic d'expression à 14 jours de développement embryonnaire, puis décroissent de façon similaire jusqu'à la naissance. Sont-ils régulés par des facteurs communs? Coopèrent-ils dans leur mécanisme d'action, par synergie ou au contraire par interférence? On retrouve notamment dans l'enhancer  $\alpha$ , des séquences qui ressemblent à des séquences de fixation pour les récepteurs nucléaires.

L'approche choisie au cours de mon travail de thèse, nous a permis d'apporter des éléments d'informations sur la formation du système nerveux central au cours de l'évolution par l'isolement et la caractérisation de COUP-TF chez l'hydre et l'amphioxus, et TR2/4 chez l'amphioxus, des récepteurs nucléaires impliqués dans la mise en place du système nerveux central.

Les gènes COUP-TF isolés chez l'hydre et l'amphioxus dans le groupe de Vincent Laudet, sont remarquablement bien conservés avec leurs homologues de vertébrés. On retrouve notamment, les acides aminés impliqués dans la fixation à l'ADN dans le gène COUP-TF, chez un organisme aussi précoce que l'Hydre. Ce gène est par ailleurs, également exprimé dans les cellules de type neuronal chez ce diploblaste (cf. discussion de l'article 5). TR2/4 est quant à lui exprimé dans la vésicule cérébrale de l'amphioxus en début de formation. De façon intéressante, COUP-TF n'est pas exprimé dans la vésicule cérébrale chez l'amphioxus; il est cependant exprimé de manière segmentée dans la corde neurale, l'équivalent du

tube neural des vertébrés (cf. article 6). Ces résultats suggèrent d'une part, l'existence d'une segmentation du tube neural de l'amphioxus, ce qui n'était jamais apparu clairement. Et d'autre part, une acquisition après la divergence amphioxus/vertébrés d'une fonction de segmentation du diencéphale pour COUP-TF. Il est à noter que Pax-6, lui-même impliqué dans la segmentation du cerveau mais à un niveau moins "raffiné" : il aide notamment à la séparation diencéphale/mésencéphale est exprimé par contre, dans la vésicule cérébrale de l'amphioxus, comme chez les vertébrés. Ces résultats suggèrent que la fonction ancestrale des gènes COUP-TFs était de participer à la formation des neurones et que plus tard, après la duplication des gènes COUP-TFs, survenue après la divergence amphioxus/vertébrés (cf. discussion de l'article 6), ces gènes, notamment COUP-TFII, ont acquis la capacité à participer à la segmentation du diencéphale. En effet, l'inactivation de COUP-TFI chez la souris n'altère pas la segmentation du diencéphale, mais génère un nerf crânien X fusionné au nerf IX.

Par ailleurs, nous avons montré que le récepteur COUP-TF chez l'amphioxus possédait déjà sa capacité d'interférence avec le signal de transduction médié par les acides rétinoïques car nous avons montré que lorsqu'il est co-transfecté en cellules COS, avec les récepteurs RAR et/ou RXR en présence d'acides rétinoïques, il inhibe la transactivation du gène rapporteur tkCAT *via* un élément de réponse de type DR5. L'acquisition de la fonction d'interférence des gènes COUP-TF avec la voie des rétinoïdes est par conséquent située entre les protostomiens, comme les drosophiles qui ne possèdent ni récepteur RAR ni acides rétinoïques, et les céphalocordés.

Les résultats de l'article 4 suggèrent par ailleurs, une acquisition récente de la capacité à fixer un ligand par les récepteurs nucléaires. Dans ce cas, sans ligand et sans récepteurs classiques à réprimer, comment pouvaient donc agir les récepteurs orphelins ? Nous avons vu que des événements de phosphorylation dans le domaine transactivateur AF-2 peuvent stimuler son activité et sa capacité à interagir avec des cofacteurs (cf. introduction de la partie récepteurs). COUP-TF, notamment, peut être phosphorylé sous l'effet de la dopamine, et sa capacité à transactiver est alors augmentée. D'autre part, chez l'hydre la métamorphose est assurée par des neurotransmetteurs, tandis que chez les mammifères, ce sont les hormones thyroïdiennes qui jouent ce rôle. Il est tentant d'imaginer qu'avant

que n'existent les hormones et les organes endocriniens des organismes les plus tardifs de l'évolution animale, les neurotransmetteurs suffisaient à assurer l'homéostasie chez les organismes les plus précoces de l'échelle évolutive, en coopération avec les récepteurs orphelins.