UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES DE LILLE

THÈSE

présentée pour l'obtention du grade de DOCTEUR DE L'UNIVERSITE DE SCIENCES ET TECHNOLOGIES DE LILLE

par

Catherine CARRIÈRE

PAX-QNR/PAX-6, UN GÈNE ESSENTIEL POUR LA FORMATION DES YEUX: STRUCTURE ET EXPRESSION DANS LA NEURORÉTINE D'OISEAU

Soutenance le 24 mars 1995 devant la commission d'examen

Président :Pr André VERBERTRapporteurs :Dr Brigitte ROYER-POKORADr Françoise POIRIERExaminateur :Pr Xavier DESBIENS

Directeur de thèse : Dr Simon SAULE

Je voudrais tout d'abord remercier Dominique Stéhelin de m'avoir accueillie dans son laboratoire, et Simon Saule de m'avoir gardée quand il m'a trouvée dans son équipe en rentrant de postdoc. Simon, j'ai énormément appris durant ces quatre années et c'est en grande partie grâce à toi.

Je voudrais remercier tout particulièrement Agnès qui m'a tout appris du clonage et de la vie du labo, pour sa gentillesse, son sourire, les aléas de la vie de Nicolas, les boites de nuit de la région, les chewing-gum sans sucre, les enzymes chères et pas chères mais c'est quand même pas une raison. T'inquiètes pas, au moindre problème, je t'appelle...

Brigitte, ma petite maman du nord, pour son sourire, sa confiance et son amitié.

Le groupe du Bunker ou EP56, à Manuella, Christine, Marie Claire, les Nathalies, Patrick, et récemment encore BQ, Carmen et Delphine, le nouvellement arrivé farfelu, Marc, pour l'enthousiasme et la passion partagée de ce drôle de gène. Tous les membres de l'équipe ont participé aux travaux présentés dans cette thèse. Plus particulièrement Serge qui m'a exaspéré presque aussi souvent qu'il m'a aidé et encouragé, ce qui n'est pas peu dire, et à qui je dois beaucoup. Et Jocelyne, parce que entre gens du Sud, on se comprend...

Françoise qui, bien qu'étant arrivée tout récemment, a fini par me faire comprendre qu'un traitement de texte, c'est quand même plus pratique qu'une machine à écrire et à qui je dois quasiment toutes les belles figures de ma thèse.

Nini et Marie Christine pour leur aide, leur efficacité et leur solidarité avec le labo, et... quand Agnès ne sait pas, peut être que Nini sait...

Daniel Lazarecki pour ses mille et une photos.

Les membres du labo, en général, à tous ceux qui m'ont souri et conseillée un jour.

Zou, pour son amitié et pour tous les bons moments que ce soit au labo ou ailleurs, pour ses crises de mauvaises humeurs et de fou rire et le fait qu'elle est toujours là quand il le faut.

Mes amis dans le désordre, Marie, BQ, Claire, Rachid, Marie Pierre, Fred, Anne, Frédérique, Didier, Sophana, Pad, Marie Claude, Annick et tous les autres non scientifiques... qui ont toujours cru que j'y arriverai...

Enfin, je suis très reconnaissante aux Professeurs A. Verbert et X. Desbiens, aux Docteurs B. Royer-Pokora, F. Poirier et S. Saule de me faire l'honneur de faire partie du jury de cette thèse, et tout spécialement aux Docteurs B. Royer-Pokora et F. Poirier d'avoir accepté d'en être les rapporteurs.

Je remercie également l'Association pour la Recherche contre le Cancer et l'institut Pasteur de Lille qui m'ont financé ces quatre années de thèse. "C'est quand même grâce aux progrès fantastiques de la science que désormais nous savons que, quand on plonge un corps dans une baignoire, le téléphone sonne."-Desproges.

"How often have I said to you, that when you have eliminated the impossible, whatever remains however improbable must be the truth." Sherlock Holmes.

A Patricia et Sylvie qui croient toujours que je deviendrai Prix Nobel... A ma famille et mes amis

SOMMAIRE

Abréviations	3
Liste des figures et tableaux	5
Résumé	6

INTRODUCTION

Introduction générale	7
I - Développement des insectes et gènes homéotiques	9
II- Conservation des homéogènes	11
III - Protéines à homéodomaine	13
1- Structure générale de l'homéodomaine	13
1-1 Région amino-terminale	14
1-2 Régions en hélice α	16
2- Mode de fonctionnement des protéines à homéodomaine	17
2-1 Interaction protéine-ADN	18
2-2 Interaction protéine-protéine	18
3- Classification des protéines à homéodomaine	20
3-1- Super classe organisée en complexe génomique	
Gènes homéotiques et gènes Hox	21
3-2- Super classe dispersée au niveau génomique	23
a) Classe avec hexapeptide	23
b) Protéines à homéodomaine	24
c) Protéines à plusieurs domaines de fixation à l'ADN	26
d) Protéines à domaines conservés ne fixant pas l'ADN	29
e) Protéines à homéodomaine atypique	32
IV - Famille des gènes Pax	33
1- Expression au cours du développement	34
a) Expression de Pax-1 et Pax-9	34
b) Expression de Pax-2, Pax-5 et Pax-8	35
c) Expression de Pax-3 et Pax-7	37

	d) Expression de Pax-4	38
	e) Expression de Pax-6	39
	2- Mutations associées	41
	a) Pax-1 et le phénotype Undulated	41
	b) Altérations du gène Pax-3 dans les souris Splotch et chez l'homme	
	dans les syndromes de Waardenburg 1 et 3 (WS1 et WS3)	42
	c) Altérations du gène Pax-6 dans les souris Small eye (Sey), chez l'homme	
	dans les cas d'Aniridia (AN) et chez la drosophile eyeless (ey)	45
	3- Fonctions des gènes Pax	48
	a) Analyse de l'expression	49
	b) Analyse moléculaire	49
	c) Potentiel oncogénique	53
V -]	La neurorétine	54
	1- Modèle biologique	54
	2- Développement de l'œil	54
	3- Gènes à boîte homéo exprimés dans l'œil	55
RÉ	SULTATS	57
I	Caractérisation d'un gène de caille contenant une boîte homéo	
	et une boîte paired exprimé dans la neurorétine	58
П	Structure génomique du gène Pax-QNR et propriétés de fixation	
	de son produit	69
III	Caractérisation des protéines codées par Pax-6 (Pax-QNR)	
	et exprimées dans la neurorétine	82
IV	Caractérisation des signaux de localisation nucléaires, des propriétés de fixation	
	à l'ADN et de transactivation des isoformes codées par Pax-6 (Pax-QNR)	96
DIS	SCUSSION	141
RÉ	FÉRENCES BIBLIOGRAPHIQUES	150

ABRÉVIATIONS

aa :	acide aminé
ADN :	acides désoxyribonucléiques
ADNc :	ADN complémentaire
AN :	Aniridie
ARN :	acides ribonucléiques
ARNm :	ARN messager
DNase :	Désoxyribonucléase
E :	jour embryonnaire
HOM-C:	complexe homéotique
N-CAM :	molécule d'adhésion cellulaire neurale
pb :	paire de bases
PRD :	paired
QNR :	neurorétine de caille
RMN :	résonance magnétique nucléaire
Sey :	Small eye
SNC :	système nerveux central
SNP:	système nerveux périphérique
Sp:	Splotch
Un ·	
011.	undulated

Gènes de drosophile :	Antp :	Antennapedia
	Abd-A:	Abdominal-A
	Abd-B:	Abdominal-B
	Bcd:	Bicoid
	Dfd :	Deformed
	Dls :	Distal-less
	Ems:	Empty-spiracle
	En :	Engrailed
	Ey:	Eyeless
	Ftz:	Fushi-tarazu
	Gsb et G	sbn : Gooseberry et Gooseberry neuro
	Lab:	Labial
	Otd :	Orthodenticle
	Pb:	Proboscipedia
	Poxm et	Poxn : Pox meso et Pox neuro
	Scr:	Sex comb reduced
	Ubx :	Ultrabithorax
	Zen :	Zerkhnült

Code des acides aminés :	1
--------------------------	---

Alanine	Ala	(A)		
Cystéine	Cys	(C)		
Acide aspartique	Asp	(D)		
Acide glutamique	Glu	(E)		
Phénylalanine	Phe	(F)		
Glycine	Gly	(G)		
Histidine	His	(H)		
Isoleucine	Ile	(I)		
Lysine	Lys	(K)		
Leucine	Leu	(L)		
Methionine	Met	(M)		
Asparagine	Asn	- (N)		
Proline	Pro	(P)		
Glutamine	Gln	(Q)		
Arginine	Arg	(R)		
Serine	Ser	(S)		
Threonine	Thr	(T)		
Valine	Val	(V)		
Thryptophane	Trp	(W)		
Tyrosine	Tyr	(Y)		

LISTE DES FIGURES ET TABLEAUX

Figure 1	Nomenclature des gènes Hox
Figure 2	Structure tridimensionnelle de l'homéodomaine
	A- Structure de l'homéodomaine comparé au répresseur λ
	B- Interaction de l'homéodomaine avec l'ADN
Figure 3	Séquence consensus de l'homéodomaine
Figure 4	Différentes classes de protéines à homéodomaine
Figure 5	Expression des gènes Pax dans le système nerveux central
Figure 6	Homologie entre eyeless et Pax-QNR
Figure 7	Structure des gènes Pax chez la souris
Figure 8	Structure bipartite du domaine PRD
Figure 9	Possibilité de transdifférenciation de la neurorétine
Figure 10	Organisation de la neurorétine
Figure 11	Organisation génomique du gène Pax-QNR
Figure 12	Schéma d'entrée des protéines dans le noyau
Figure 13	Tableau récapitulatif
Figure 14	Fixation du domaine PRD
Figure 15	Hypothèse de fixation par l'intermédiaire de l'homéodomaine
Tableau 1	Classification des gènes Pax
Tableau 2	Expression de Pax-1 au cours du développement de la souris
Tableau 3	Expression de Pax-2, Pax-5 et Pax-8 au cours du développement de la souris
Tableau 4	Expression de Pax-3 et Pax-7 au cours du développement de la souris
Tableau 5	Expression de Pax-6 au cours du développement de la souris
Tableau 6	Mutations dans les gènes Pax-1, Pax-3, Pax-6 et Pax-7
Tableau 7	Mutations dans Pax-6
Tableau 8	Gènes à boite homéo exprimés dans les tissus oculaires

<u>RÉSUMÉ</u>

Le gène Pax-QNR/Pax-6 code un facteur de transcription contenant 2 domaines de liaison à l'ADN : le domaine Paired (PRD) et l'homéodomaine. Des mutations dans ce gène sont responsables de graves anomalies du développement (Eyeless chez la drosophile, Small eye chez la souris et Aniridie chez l'homme). Pax-QNR est exprimé dans le tube neural, dans la neurorétine et dans le cervelet. L'unité transcriptionnelle de ce gène s'étend sur 17 Kbp d'ADN génomique et comprend au moins 15 exons. Le produit codé par *Pax-QNR* est capable de se lier à l'ADN sur une séquence reconnue par plusieurs membres de la famille Pax, la séquence e5, ainsi que sur son propre promoteur P0. Cette fixation est médiée à la fois par le domaine PRD et l'homéodomaine. Des anticorps, dirigés contre différentes régions du produit de Pax-QNR, ont permis de caractériser au moins 5 protéines produites in vivo dans la neurorétine aviaire. Les ADN complémentaires correspondants à ces différentes protéines ont été isolés. Ils sont générés par plusieurs événements d'épissage alternatif se produisant dans la région 5' non codante, dans le domaine PRD et dans la région carboxy-terminale. Ces protéines ont des localisations subcellulaires différentes : celles qui possèdent un domaine PRD complet sont nucléaires alors que celles qui ont un domaine PRD partiel ou absent, sont nucléaires et cytoplasmiques. Le transport dans le noyau est médié par au moins deux signaux de localisation nucléaires, l'un dans le domaine PRD, l'autre au début de l'homéodomaine. L'étude des propriétés de fixation à l'ADN de ces différentes protéines montre que toute altération du domaine PRD abolit sa capacité à se fixer sur la séquence e5 et sur le promoteur P0. Enfin, le domaine transactivateur est localisé dans le domaine carboxy-terminal. Il est constitué de deux régions séparées, en amino- et en carboxy-terminal de ce domaine et la présence de ces deux régions est nécessaire pour une activation optimale de la transcription.

INTRODUCTION

INTRODUCTION GÉNÉRALE

A partir d'une seule cellule, un organisme se développe en un grand nombre de types cellulaires, de tissus et d'organes. L'étude moléculaire des mécanismes permettant le développement des différentes parties de cet organisme a permis d'identifier un nombre toujours croissant de gènes ayant une importance majeure, en particulier ceux dont les produits peuvent activer ou réprimer d'autres gènes au cours des cascades conduisant à une expression cellule- ou région-spécifique.

Les premières étapes du développement ont lieu avant même la fertilisation de l'ovule. Ainsi, dans la plupart des espèces, excepté les oiseaux et les mammifères, le premier axe embryonnaire, l'axe antéro-postérieur, est établi par rapport à la position de l'ovule dans les ovaires comme chez la drosophile (Berleth *et al.*, 1988), ou par rapport à la position du noyau dans l'ovule chez les amphibiens (Heasman *et al.*, 1984). En effet, l'ovule accumule, avant la fécondation, des informations maternelles sous forme d'ARN messagers ou de protéines. Chez les mammifères et les oiseaux, le premier axe est l'axe dorso-ventral et sa mise en place n'a lieu qu'après la fécondation. Sa position est liée à la position de l'embryon au stade blastula par rapport au vitellus pour les oiseaux et au trophectoderme pour les mammifères. Le deuxième axe se forme perpendiculairement au premier. Chez la drosophile, sa position est également établie avant la fécondation par l'action de gènes d'origine maternelle. Dans les autres espèces, l'établissement du deuxième axe est du à des facteurs extérieurs comme l'entrée du spermatozoïde ou la gravité (voir la revue de Gurdon, 1992).

Après la fécondation, l'œuf va se diviser et de nombreux types cellulaires vont progressivement apparaître.

La diversité cellulaire est générée par deux mécanismes qui peuvent être complémentaires : une division asymétrique qui va répartir le cytoplasme de la cellule mère de façon différente entre les deux cellules filles et les interactions entre les cellules qui vont permettre l'induction de nouveaux types cellulaires. Ainsi, l'interaction initiale entre les deux cellules filles conduit, au cours des divisions cellulaires, à une amplification progressive de la diversité. Le premier mécanisme est très important chez les amphibiens et les insectes. Chez les mammifères et les oiseaux, il n'a pas été observé de phénomène de répartition asymétrique du cytoplasme au cours des divisions cellulaires et c'est essentiellement le deuxième mécanisme qui intervient. Ces interactions vont contrôler la prolifération, la migration et la différenciation cellulaire (voir la revue de Gurdon, 1992).

Les interactions cellulaires sont médiées par les membranes des cellules ainsi que par la matrice extra cellulaire. L'acquisition des propriétés membranaires est essentielle puisqu'elle permettra aux cellules de reconnaître leurs "voisines" et d'envoyer des signaux. Ces contacts se font par l'intermédiaire de molécules d'adhésion cellulaire, les CAM (Cell Adhesion Molecule), de récepteurs qui sont intégrés dans la membrane cellulaire, de molécules associées à la matrice extra-cellulaire et de facteurs de croissance sécrétés. Interviennent ensuite les molécules impliquées dans la transmission du signal intra cellulaire. Celui-ci peut agir soit directement sur le cytosquelette, soit par l'intermédiaire d'une cascade de second messagers qui vont en final activer ou réprimer la transcription dans le noyau. Ces interactions peuvent se faire de cellules à cellules comme les neurotransmetteurs, ou sur de plus longues distances, quand le système circulatoire est mis en place, comme dans le cas des hormones.

Ces interactions permettent à une cellule ou un tissu d'en induire un autre à se différencier, à proliférer et dans le cas de l'apoptose, à se détruire. Ces inductions peuvent être positives ou négatives et agir en combinaison d'où la génération de la variété cellulaire. Des cellules doivent être compétentes pour répondre à ces inductions, la compétence étant dépendante de la localisation des cellules et du stade de développement de l'embryon. Ces interactions permettent donc la mise en place des cellules avec leur identité de position spécifique et le développement des différentes structures appropriées à leurs positions.

Au niveau moléculaire, la cellule doit donc exprimer les facteurs nécessaires à l'expression ou l'inactivation d'autres gènes suite à la transmission du signal provenant des cellules voisines ou du système sanguin.

L'étude du développement segmenté chez la drosophile et de ces nombreux mutants a permis d'identifier un grand nombre de gènes dont les produits codant pour des facteurs de transcription, contrôlent des étapes "clés" du développement.

8

I - Développement de la drosophile et gènes homéotiques

L'étude des mutations homéotiques, en particulier, a permis de caractériser une famille de gènes dits homéotiques qui jouent un rôle essentiel dans la spécification de l'identité cellulaire le long de l'axe antéro-postérieur. Les embryons et larves porteurs de ce type de mutation ont des segments ou métamères manquants qui sont remplacés par d'autres, identiques aux segments plus postérieurs, d'où le nom homéo qui signifie semblable. Ces mutations peuvent ainsi générer des organes ectopiques comme des pattes à la place des antennes ou des éléments de la bouche comme dans le cas de la mutation *Antennapedia* (Gehring, 1966)(Garber *et al.*, 1983). Ces phénotypes suggèrent que les groupes de cellules à l'origine des différentes parties de l'organisme ne sont pas encore déterminés et peuvent être orientés entre plusieurs schémas de développement. Les produits des gènes homéotiques contrôlent ces décisions et donc la différenciation finale de ces cellules. Par la suite, il a été découvert que cette famille de gènes qui permet de définir les identités de position cellulaires le long de l'axe antéro-postérieur, existait également chez les autres espèces avec la même fonction.

Chez la drosophile, la définition d'un gène homéotique est basée sur deux critères. Le premier critère d'ordre fonctionnel, est qu'un gène homéotique quand il est défectueux, transforme un métamère ou un groupe de métamères, en la copie d'un autre. La fonction normale des gènes homéotiques est donc de spécifier l'identité des métamères. En l'absence d'information homéotique, la diversité morphologique disparaîtrait et tous les métamères se développeraient avec une même identité de base. Le deuxième critère, d'ordre moléculaire, est que les gènes homéotiques codent des protéines nucléaires qui possèdent un domaine de fixation à l'ADN hautement conservé de 60 acides aminés, l'homéodomaine. Ce dernier est codé par une séquence appelée la boîte homéo.

Ces gènes ont des profils d'expression restreints. Ils sont initialement activés aux stades précoces du développement et leur activité est maintenue durant l'embryogenèse et le stade larvaire. Toute une hiérarchie de gènes de contrôle est requise pour l'activation des gènes homéotiques. Ces gènes de contrôle sont mis en route progressivement avant même la fécondation.

Les premières étapes de la segmentation chez la drosophile sont initiées par quatre types de gènes: les gènes dits d'origine maternelle, les gènes "gap", les gènes "pair-rule" et les gènes de polarité de segments. Les gènes d'origine maternelle sont exprimés dans l'ovule avant même la fertilisation, sous forme d'ARN messagers, et vont définir l'axe antéro-postérieur de l'embryon. Par la suite, les produits codés par ces messagers vont diffuser à travers l'embryon qui est sous forme syncitiale, c'est à dire sans séparation cellulaire, pour générer des gradients de protéines qui vont permettre l'établissement des domaines d'expression des gènes "gap". Les interactions dans l'embryon syncitial, entre les différents gènes "gap" qui codent plusieurs types de facteurs de transcription, vont générer l'expression spécifique des gènes "pair-rule". Ces derniers codent exclusivement pour des facteurs de transcription. Ils seront les premiers à avoir une expression alternée et répétitive dans l'embryon qui va alors se cellulariser puis se segmenter en quatorze métamères. Ils sont, pour la plupart, exprimés dans sept segments sur quatorze avec une périodicité d'un segment sur deux dans l'embryon cellularisé. Les interactions entre les différents gènes "pair-rule" vont permettre la mise en place des parasegments qui ont une périodicité segmentale un peu déphasée si on les compare à la morphologie visible des segments. Les gènes "pair-rule" contrôlent l'expression ou activité des gènes de polarité de segment. Ces derniers, pour la plupart exprimés de façon périodique, maintiennent et précisent les limites définies par les gènes "pair-rule". Certains des gènes de polarité de segment codent des facteurs de transcription mais la majorité est impliquée dans la communication intracellulaire et dans la transduction du signal (voir la revue de Morata, 1993).

Cette cascade de facteurs qui agissent de façon combinée, va ensuite activer les gènes homéotiques qui vont permettre aux différents segments d'acquérir leur identité le long de l'axe antéro-postérieur.

Le nombre de gènes homéotiques s'est avéré être très réduit si l'on considère la variété de schémas de développement que ces gènes peuvent générer. Chez la drosophile, il y a seulement huit gènes qui vont contrôler la différenciation de quatorze segments différents. Ces huit gènes sont groupés en deux complexes génomiques, situés sur le même chromosome, appelés de façon

drosophile (Cf. Fig.1). Cela suggère que les différents complexes chez les mammifères proviennent d'un phénomène de duplication d'un même complexe ancestral. Cette organisation en complexe est retrouvée chez tous les métazoaires (Kappen et Ruddle, 1993).

La conservation au cours de l'évolution ne concerne pas seulement la boîte homéo qui code l'homéodomaine et l'organisation de ces gènes mais également leurs modes d'expression. En effet, ces gènes sont organisés le long des chromosomes dans l'ordre de leur expression le long de l'axe antéro-postérieur (Coldberg-Poley *et al.*, 1985)(Breier *et al.*, 1986). Les gènes homéotiques ont la même limite d'expression postérieure et c'est la limite antérieure qui varie. Leurs domaines d'expression vont en croissant en fonction de leur position dans les complexes génomiques sur l'axe 5'-3'. Les gènes les plus en amont des complexes sont ceux qui sont exprimés le plus tardivement et de la façon la plus restreinte. Les gènes en aval s'expriment plus précocement et dans les régions plus antérieures.

Cette organisation conservée n'est que partiellement comprise à l'heure actuelle. Une hypothèse est que ces gènes ont des éléments de régulation en *cis* communs. Les gènes adjacents doivent donc rester liés puisque la translocation de l'un d'entre eux le priverait lui ou un autre, de ses éléments de régulation en *cis*. Enfin, des régions intergéniques ont été découvertes qui pourraient être impliquées dans des modulations de la structure chromatinienne et donc dans la régulation des gènes qui les encadrent (Kennison *et al.*, 1993).

Cependant, la principale caractéristique structurale des gènes homéotiques, la boîte homéo, n'est pas spécifique de ce type de gènes. La majorité des gènes à boîte homéo n'est pas impliquée dans l'homéosis. Chez la drosophile et les vertébrés, on retrouve des gènes ayant des fonctions différentes à l'intérieur des complexes homéotiques, mais ils sont pour la plupart, localisés à l'extérieur des complexes.

Comme les gènes homéotiques, ces gènes à boîte homéo codent des facteurs de transcription. Ils peuvent posséder d'autres domaines de fixation à l'ADN en plus de l'homéodomaine. Ils sont également très conservés au cours de l'évolution et possèdent des homologues ou des gènes apparentés dans la plupart des espèces. Cette conservation peut s'étendre également de la structure du produit codé par ces gènes aux domaines d'expression. Ainsi, certains homologues de gènes exprimés dans les segments antérieurs des insectes, sont

collective le HOM-C (Homeotic Complex). Comme il y a plus de segments que de gènes homéotiques, ces derniers doivent fonctionner en combinaison et/ou être différentiellement exprimés, afin de rendre compte de toute la diversité morphologique. Les gènes homéotiques ont souvent des domaines d'expression chevauchant, ce qui permet des interactions entre leurs produits. Ils possèdent également des éléments de régulation en *cis* qui contrôlent leur expression au niveau spatial. Ces gènes codent des facteurs de transcription. Leurs produits possèdent donc un domaine impliqué dans la fixation de séquences spécifiques d'ADN, l'homéodomaine, et au moins un deuxième domaine qui est capable d'interagir avec un ou plusieurs composants de la machinerie transcriptionnelle pour activer ou réprimer la transcription. La fonction de ces protéines est de contrôler la transcription à partir de promoteurs cibles spécifiques.

Les homologues vertébrés de ces gènes sont également organisés en complexe génomique bien qu'il y ait 4 copies de ce complexe localisées sur des chromosomes différents (Duboule *et al.*, 1989).

II - Conservation des homéogènes

Le HOM-C de la drosophile est organisé en deux complexes génomiques, ANT-C (*Antennapedia*-complex, *Antennapedia* est le premier gène homéotique caractérisé au niveau moléculaire chez la drosophile et a donné son nom au complexe) et BX-C (pour *bithorax* complex) ces deux complexes étant séparés par plusieurs kilobases. Une telle dissociation en deux groupes n'est pas un fait essentiel car, chez d'autres insectes plus primitifs comme la blatte (ou*Tribolium*), les gènes du HOM-C sont regroupés en un seul complexe (Stuart *et al.*, 1991). Chez les mammifères comme la souris et l'homme, ces gènes sont organisés en 4 complexes (Hox-A, B, C, D) localisés sur différents chromosomes. Chaque gène d'un complexe donné possède un à quatre gènes paralogues, c'est à dire très proche au niveau séquence et position dans chaque complexe, et présente des homologies avec des gènes du HOM-C. En effet, en se basant sur l'identité de séquence de leurs homéodomaines et des régions adjacentes, sur la position dans le complexe ainsi que sur la conservation de fonction, les gènes Hox des mammifères peuvent être considérés comme les homologues des gènes homéotiques de la



Fig 1 : Nomenclature des gènes Hox.

La nouvelle nomenclature est encadrée et l'ancienne est indiquée en dessous de chaque gène. Les gènes des complexes ANT-C et BX-C sont placés au-dessus. La direction de la transcription est indiquée par des flèches. Les gènes en 3' (à gauche) sont exprimés plus précocément et plus antérieurement que les gènes en 5' (à droite). Ces gènes sont également plus sensibles à la stimulation par l'acide rétinoïque (RA). (d'après Duboule, 1994)

drosophile (Cf. Fig.1). Cela suggère que les différents complexes chez les mammifères proviennent d'un phénomène de duplication d'un même complexe ancestral. Cette organisation en complexe est retrouvée chez tous les métazoaires (Kappen et Ruddle, 1993).

La conservation au cours de l'évolution ne concerne pas seulement la boîte homéo qui code l'homéodomaine et l'organisation de ces gènes mais également leurs modes d'expression. En effet, ces gènes sont organisés le long des chromosomes dans l'ordre de leur expression le long de l'axe antéro-postérieur (Coldberg-Poley *et al.*, 1985)(Breier *et al.*, 1986). Les gènes homéotiques ont la même limite d'expression postérieure et c'est la limite antérieure qui varie. Leurs domaines d'expression vont en croissant en fonction de leur position dans les complexes génomiques sur l'axe 5'-3'. Les gènes les plus en amont des complexes sont ceux qui sont exprimés le plus tardivement et de la façon la plus restreinte. Les gènes en aval s'expriment plus précocement et dans les régions plus antérieures.

Cette organisation conservée n'est que partiellement comprise à l'heure actuelle. Une hypothèse est que ces gènes ont des éléments de régulation en *cis* communs. Les gènes adjacents doivent donc rester liés puisque la translocation de l'un d'entre eux le priverait lui ou un autre, de ses éléments de régulation en *cis*. Enfin, des régions intergéniques ont été découvertes qui pourraient être impliquées dans des modulations de la structure chromatinienne et donc dans la régulation des gènes qui les encadrent (Kennison *et al.*, 1993).

Cependant, la principale caractéristique structurale des gènes homéotiques, la boîte homéo, n'est pas spécifique de ce type de gènes. La majorité des gènes à boîte homéo n'est pas impliquée dans l'homéosis. Chez la drosophile et les vertébrés, on retrouve des gènes ayant des fonctions différentes à l'intérieur des complexes homéotiques, mais ils sont pour la plupart, localisés à l'extérieur des complexes.

Comme les gènes homéotiques, ces gènes à boîte homéo codent des facteurs de transcription. Ils peuvent posséder d'autres domaines de fixation à l'ADN en plus de l'homéodomaine. Ils sont également très conservés au cours de l'évolution et possèdent des homologues ou des gènes apparentés dans la plupart des espèces. Cette conservation peut s'étendre également de la structure du produit codé par ces gènes aux domaines d'expression. Ainsi, certains homologues de gènes exprimés dans les segments antérieurs des insectes, sont

12

exprimés, chez les vertébrés, dans l'encéphale. Cela suggère que, bien qu'ayant des modes de développement distincts, tous les organismes utilisent certains mécanismes de base qui sont pour la plupart contrôlés par les gènes à boîte homéo.

Il est quelque peu surprenant qu'aucune anomalie du développement n'ait pu être observée, chez les vertébrés, provenant d'une mutation spontanée dans les gènes Hox. Cependant, d'autres familles de gènes à boîte homéo ont pu être liées à ce type de mutation. Ainsi chez les mammifères, les gènes de la classe POU ou Pax ont pu être associés à de graves anomalies du développement. On peut supposer à partir des études faites sur ces mutants que ces gènes jouent un rôle régulateur essentiel dès les premiers stades du développement. Dans le cas du gène Pax-6, il est intéressant de noter que sa fonction est également conservée des insectes à l'homme puisque l'on peut observer le même phénotype chez les organismes mutés dans ce gène, l'absence d'organe de la vue.

III - Protéines à homéodomaine

1 - Structure générale de l'homéodomaine

La boîte homéo est une séquence de 180 pb qui code un homéodomaine de 60 acides aminés (aa). Ce domaine est capable de fixer l'ADN.

C'est sur la base d'une homologie avec les protéines MATa1 et MAT α 2 de levure et également d'une homologie plus faible avec des régulateurs de transcription bactériens que ce domaine a été supposé être capable de se lier à l'ADN (Scott *et al.*, 1984; MacGinnis *et al.*, 1984).

Les premières études ont été faites sur l'homéodomaine du gène Antennapedia. Ce dernier s'est avéré être effectivement capable de se lier à l'ADN sur des séquences spécifiques qui ont permis de définir des sites consensus de fixation contenant un motif ATTA (TAAT sur le brin complémentaire). Cependant, tous les homéodomaines connus peuvent reconnaître des séquences contenant ce motif, avec une affinité variable qui peut être fonction des bases qui l'encadrent. Cette différence d'affinité n'est pas suffisante pour expliquer la spécificité d'action



Fig.2 : Structure tridimensionnelle de l'homéodomaine

A - Structure de l'homéodomaine comparée au répresseur λ

B -Interaction homéodomaine/ADN (les principaux aa interagissant avec l'ADN sont indiqués : R, Arg; I, Ileu; N, Asn; Q, Gln)

(d'après Angrand et al., 1993)

de ces protéines à homéodomaine *in vivo* (Hoey et Levine, 1988)(revue de Hayashi et Scott, 1990). Mais les conditions dans lesquelles la plupart des expériences de fixation sur l'ADN ont été faites, souvent sur des homéodomaines isolés et produits en bactérie, restent éloignées du mode d'action *in vivo* (Chan et Mann, 1993).

L'analyse tridimensionnelle par résonance magnétique nucléaire (RMN) de cet homéodomaine (en fait l'homéodomaine d'Antennapedia plus sept résidus en carboxy-terminal) a montré qu'il était constitué d'un bras amino-terminal et de trois hélices α (Quian *et al.*, 1989). Le bras amino-terminal est constitué de neuf aa. L'hélice 1 de treize aa est séparée par une boucle de six aa de l'hélice 2 qui est constituée de onze aa et, un tour de trois aa sépare l'hélice 2 de l'hélice 3 codée par dix-sept aa (l'hélice 3 est en fait constituée de deux hélices, de onze aa et sept aa). La structure tridimensionnelle de l'homéodomaine est effectivement très proche de celle de certains répresseurs procaryotes sans qu'il y ait pour autant une homologie en acides aminés (Cf. Fig.2A).

L'analyse par cristallographie de l'homéodomaine du produit du gène de drosophile Engrailed fixé à l'ADN (Kissinger et al., 1990), ainsi que celle de l'homéodomaine de la protéine de levure MAT α 2 (Wolberger et al., 1991), a révélé une structure similaire, c'est à dire le même arrangement en 3 hélices et le contact avec l'ADN se fait de la façon suivante : La partie amino-terminale va contacter le petit sillon de l'ADN au niveau du site de fixation alors que la troisième hélice va contacter le grand sillon adjacent (Cf. Fig. 2B). La similarité de conformation entre des homéodomaines aussi divergents que MAT α 2 et Engrailed suggère que ces caractéristiques structurales sont conservées dans tous les homéodomaines. C'est dans l'hélice 3 dite "de reconnaissance" que l'on observe la plus forte conservation en aa.

1-1 Région amino-terminale

Le bras amino-terminal participe à la fixation à l'ADN et également à la spécificité de reconnaissance de la séquence cible. Le produit d'*Antp* a une spécificité d'action *in vivo* différente de celle du produit de *Sex comb reduced* (*Scr*) et leurs homéodomaines ne diffèrent que par quatre aa situés dans la partie amino-terminal. Si l'on remplace les quatre aa différents d'Antp par ceux de Scr, la protéine Antp modifiée est capable d'activer *in vivo* des gènes

normalement contrôlés par Scr (Zeng *et al.*, 1993). De même, le remplacement des six aa qui distinguent le bras amino-terminal de l'homéodomaine 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, et cela bien qu'il y ait d'autres aa qui diffèrent au niveau des structures en hélice α (Lin *et al.*, 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 des deux aa différents dans cette région entre Ubx et Antp est nécessaire mais ne suffit pas au changement de spécificité entre les deux protéines. Il faut également modifier trois des cinq aa qui diffèrent dans les régions en hélice α (Chan et Mann, 1993).

La façon dont les aa en amino-terminal déterminent la spécificité ne semble pas se faire par fixation à l'ADN puisque les deux aa (Arg 3 et 5) qui interagissent directement avec l'ADN dans le cas de l'homéodomaine d'Engrailed, sont fortement conservés entre plusieurs homéodomaines. Le rôle des autres aa pourrait être, soit d'influencer la conformation du bras amino-terminal, soit d'interagir avec d'autres cofacteurs qui contribueraient à la sélection de la séquence cible. Ainsi l'association entre deux protéines de drosophile à domaine POU, I-POU et Cf1-a, dépend de la séquence de la partie amino-terminale de l'homéodomaine (Treacy *et al.*, 1991).

La possibilité que l'interaction directe avec l'ADN soit la contribution principale du domaine amino-terminal pour la spécificité de fixation ne peut néanmoins pas être écartée. En effet, l'analyse par RMN de l'homéodomaine d'Antp fixé à l'ADN montre que la portion amino-terminale adopte une structure régulière seulement quand elle est en contact avec l'ADN (Otting *et al*, 1990). La conformation du bras amino-terminal pourrait être influencée par la séquence du site de fixation ou par les cofacteurs avec lesquels il peut s'associer.

Enfin, un travail récent montre que la partie amino-terminale est essentielle à la stabilisation de la structure tertiaire de l'homéodomaine. Dans ce travail, tous les aa non conservés de l'homéodomaine codé par le gène murin *Msx* ont été remplacés par des alanines. Cet homéodomaine modifié reste capable de fixer une séquence cible du produit de *Msx* mais avec une affinité faible et présente une structure tertiaire fluctuante. Un homéodomaine modifié excepté dans le bras amino-terminal fixe sa séquence cible avec une affinité très proche de la protéine native, et a une structure tertiaire stable. Ainsi, bien que les aa conservés de

15



Fig. 3 : Séquence consensus de l'homéodomaine.

A - Cette séquence consensus est déduite d'une comparaison entre 346 homéodomaines. Δ : un à deux aa trouvés dans cette position; \circ : trois à cinq aa dans cette position; * pas plus de

neuf aa différents trouvés dans cette position B - Représentation schématique des interactions intra-moléculaires et protéine-ADN pour chaque acides

aminés(résumé des contacts observés pour les homéodomaines de Antennapedia, engrailed et MATo2). Ces interactions ne se font pas dans tous les cas, mais cette représentation schématique permet d'estimer les contacts qui sont susceptible de se faire pour un homéodomaine de structure indéterminée. H : aa hydrophobe contribuant à la structure tertiaire de l'homéodomaine

B : aa contactant des bases situées dans le grand sillon et responsables de la liaison spécifique à l'ADN
M : aa interagissant avec le petit sillon

* : aa pouvant interagir avec les sucres et les phosphates.

(d'après Duboule, 1994)

l'homéodomaine soient suffisants pour la reconnaissance de l'ADN, le bras amino-terminal est nécessaire pour avoir une affinité de fixation optimale ainsi que pour le maintien de la structure tertiaire appropriée (Shang *et al.*, 1994).

1-2 Région en hélice α.

L'homéodomaine comporte trois hélices α . Les deux premières hélices sont en position antiparallèle au dessus du grand sillon et perpendiculaire à ce dernier. Ces deux hélices sont séparées de l'ADN par la troisième hélice dite hélice de reconnaissance, qui va, elle, interagir avec l'ADN (Cf. Fig.2). Seize aa sont conservés à travers la famille des protéines à homéodomaine (Cf. Fig. 3). Sachant que MATa2 et Engrailed se positionnent de la même façon par rapport à l'ADN, on peut supposer que ces seize aa sont essentiels à cette conformation. L'aa le plus conservé parmi toutes les protéines à homéodomaine connues est l'Asn 51 dans la troisième hélice. Il est supposé positionner l'hélice de reconnaissance sur l'ADN. D'autres aa. un peu moins conservés, peuvent dans certains cas interagir directement avec l'ADN, ce sont les aa 47, 50, et 53. Celui qui est en position 50 est très important pour la spécificité de reconnaissance et son identité peut faire varier la séquence de fixation, c'est à dire les nucléotides entourant le core ATTA, cela dans des expériences in vivo et in vitro (Treisman et al., 1989). Ainsi, l'homéodomaine de Fushi-tarazu (Ftz) possède un Gln en position 50 et reconnaît une séquence de type CCATTA. Celui de Bicoïd (Bcd) a une Lys 50 et reconnaît la séquence GGATTA. Si la séquence cible de Ftz est mutée en GGATTA, la fixation de Ftz à cette séquence diminue fortement. Par contre, si on remplace la Gln 50 de Ftz par une Lys 50, la fixation à l'ADN augmente à nouveau (Schier et Gehring, 1992). Une étude comparative de la fixation à l'ADN des homéodomaines d'Antp et TTF-1 sur différents consensus montre que l'homéodomaine de TTF-1 fixe préférentiellement le consensus CAAT par rapport au consensus TAAT et que la Gln 50 ne contacte l'ADN que lors de la fixation sur CAAT (Damante et al., 1994). L'utilisation de cet aa par différents homéodomaines dépend donc du contexte de séquence et suit une hiérarchie de contacts précise. Cet aa 50 permet de classer de façon générale certaines grandes familles d'homéodomaine. Ainsi, pour les homéodomaines de type Hox, Caudal, TCL, Msh, NK1 et 2, Engrailed et certains apparentés au type Paired, cet aa est une Gln, pour les homéodomaines de type Paired, c'est une Ser, pour les homéodomaines de type Bcd et certains apparentés au type Paired, une Lys, et pour les homéodomaines de type POU, il s'agit d'une Cys.

D'autres résidus qui n'interagissent pas directement avec l'ADN sont importants pour la spécificité de reconnaissance de l'homéodomaine. Ainsi, l'échange de trois aa qui diffèrent entre Antp et Ubx, deux situés entre les deux premières hélices de l'homéodomaine (aa 22 et 24) et un à la fin de la troisième hélice (aa 56), suffit à échanger la fonction de ces deux protéines (Chan et Mann, 1993). Les chaînes extérieures des aa 22 et 24 sont exposées au solvant et peuvent interagir avec d'autres facteurs. Cette interaction pourrait modifier les propriétés de fixation à l'ADN. Ainsi, dans le cas de l'interaction du produit du gène POU, *Oct-1*, avec la protéine VP16 du virus de l'herpes simplex, le résidu 22 est également essentiel pour la spécificité de cette interactions protéine-protéine. Ainsi, le produit du gène *HOXD8* peut réprimer l'expression de HOXD9, non pas en se fixant sur le promoteur de *HOXD9* mais en empêchant les produits des gènes *HOXD9* et *HOXD10* de se fixer. Seules la partie amino-terminale et la première hélice α de la protéine codée par *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).

Les homéodomaines atypiques qui possèdent des aa supplémentaires entre les différentes hélices ont une structure tertiaire identique à celle des homéodomaines classiques. Les aa entre les hélices seraient donc essentiels, non pas à la structure tridimensionnelle, mais à l'interaction avec d'autres facteurs.

2 - Mode de fonctionnement des protéines à homéodomaine

Les facteurs de transcription ont différents modes de fonctionnement. Ils peuvent agir soit de façon directe par fixation à l'ADN et interaction avec la machinerie transcriptionnelle, soit de façon indirecte sans fixer l'ADN, uniquement par des interactions protéine-protéine. Les protéines à homéodomaine peuvent fonctionner selon ces deux modes.

17

2-1 Interaction protéine-ADN

L'homéodomaine peut être associé à un ou plusieurs autres domaines de fixation reconnaissant des séquences cibles différentes. Ces derniers sont parfois apparentés à des domaines de fixation déjà connus pour d'autres facteurs de transcription n'appartenant pas à la famille des protéines à homéodomaine. Ainsi, on peut retrouver des domaines en doigts de zinc qui présentent de fortes homologies avec ceux d'autres facteurs de transcription comme le produit du gène GATA-1 de souris (Omichinski *et al.*, 1993). Ces domaines de liaison à l'ADN peuvent également être complètement différents comme le domaine Paired (PRD) et le domaine POU.

Les homéodomaines, quelque soit leur classe, ont des propriétés de fixation très similaires et reconnaissent la séquence consensus ATTA/TAAT. La combinaison de plusieurs domaines de liaison à l'ADN permet de restreindre, en un sens, les séquences cibles de ces gènes puisque plusieurs éléments de fixation seront nécessaires, et donc d'augmenter la spécificité de fixation.

La spécificité d'action des protéines à homéodomaine suggère cependant, que d'autres mécanismes que la reconnaissance de séquences spécifiques, entrent en jeu. Cette spécificité peut être augmentée par des effets synergisants entre protéines, soit par interaction directe, soit de façon indirecte via la machinerie transcriptionnelle et des cofacteurs. De plus, pour certaines classes de protéines à homéodomaine, les interactions protéine-protéine sont nécessaires pour la fixation à l'ADN.

2-2 Interaction protéine-protéine

Ces interactions peuvent passer soit par l'homéodomaine ou une partie de ce domaine comme dans le cas des protéines Hox (HOXD8/HOXD9). Dans le cas des protéines POU, tout le domaine POU est impliqué, c'est à dire le domaine POU spécifique et l'homéodomaine de type POU. D'autres domaines conservés peuvent également être impliqués comme les domaines LIM ou les domaines "leucine zipper" des protéines à homéodomaine d'*Arabidopsis thaliana* (Shena *et al.*, 1994). Il peut aussi s'agir de séquences spécifiques ne correspondant pas à un domaine conservé comme la partie carboxy-terminale de la protéine MATa1 qui est essentielle pour la dimérisation avec MAT α 2 (Stark et Johnson, 1994).

Ces interactions sont des événements essentiels pour la spécificité de fixation ainsi que pour le type de régulation positive ou négative qui en découle. De plus, l'expression différentielle des facteurs de transcription et la possibilité d'association avec des régulateurs tissu-spécifiques contribuent à leur multiplicité de fonction au cours du développement. Ces interactions peuvent se faire sous forme d'homodimères ou d'hétérodimères, phénomène existant également pour d'autres facteurs de transcription n'appartenant pas à la famille des protéines à homéodomaine. Ainsi, certains domaines d'interaction protéine-protéine, comme celui à "leucine zipper", ont été caractérisés chez plusieurs facteurs de transcription qui fonctionnent également en dimères comme les membres de la famille AP1 (Abate et Curran, 1990). Plusieurs protéines à homéodomaine fonctionnent en dimères. Chez la levure, l'hétérodimérisation des protéines MATa1 et MATa2 est essentielle au maintien du stade diploïde. Les protéines des familles POU et HNF1 peuvent former des homo et hétérodimères entre différents membres de la même famille, l'interaction étant partiellement médiée par l'homéodomaine. Les produits de la famille POU sont également capables d'interagir en hétérodimères avec des protéines virales ou avec des récepteurs aux hormones (revue de Wegner et al., 1993). Les protéines à domaine LIM ont été récemment montrées comme pouvant se lier à d'autres protéines, ce, par l'intermédiaire du domaine LIM (Schmeichel et al., 1994).

Dans le cas des homodimères, les séquences de fixation sont constituées de motifs symétriques contenant chacun le consensus de fixation plus ou moins conservé, la distance entre ces sites et leur divergence par rapport au consensus constituant des éléments de régulation en *cis* de l'activité transcriptionnelle. L'hétérodimérisation permet l'association de protéines différentes appartenant soit à la même famille de gènes, soit à des familles différentes. Il peut s'agir d'une interaction médiée par la fixation à l'ADN sur des sites adjacents différents comme dans le cas des interactions du produit du gène POU, *Pit-1*, avec le récepteur des oestrogènes (Day *et al.*, 1990). Pour une protéine donnée, différentes combinaisons de séquences cibles permettront une modulation de son activité transcriptionnelle. Il peut également s'agir d'une interaction directe entre protéines où chaque élément du complexe porte une partie de l'activité finale. Ainsi, la protéine à domaine POU, Oct-1, se fixe à l'ADN mais ne transactive que

faiblement. Quand elle est associée avec la protéine virale VP16, il y a, non seulement une modification des sites de fixation reconnus par Oct-1, mais également une augmentation des capacités transactivatrices suite à l'apport du domaine transactivateur fort de VP16 (Goding *et al.*, 1989).

Il a été également montré que l'homéodomaine, en tant que domaine de fixation capable de reconnaître des séquences spécifiques, n'est pas toujours indispensable pour toute l'activité biologique de la protéine. En effet, le produit du gène *Ftz* de drosophile conserve une partie de sa fonction quand il est dépourvu d'homéodomaine. Il fonctionne en synergie avec d'autres gènes de segmentation et cet effet synergique est maintenu en l'absence de l'homéodomaine et serait médié par la partie amino-terminale de la protéine. Cependant, ce produit dépourvu d'homéodomaine ne permet pas de restaurer un phénotype normal suite à une mutation supprimant l'expression du gène endogène (Fitzpatrick *et al.*, 1992)(Ananthan *et al.*, 1993). Dans le cas de Ftz, les deux modes d'action de la protéine à homéodomaine, via l'interaction avec l'ADN et via les interactions entre protéines, peuvent être partiellement dissociés.

3 - Classification des protéines à homéodomaine

Les protéines à homéodomaines sont divisées actuellement en 22 classes, 7 classes comprises dans la super classe organisée en complexe génomique et 15 classes dans la super classe dite dispersée, les gènes de cette dernière classe étant dispersés dans tout le génome. Cette classification se base sur des critères évolutifs : les homologies de séquence entre les homéodomaines et les séquences qui les encadrent, l'association avec d'autres motifs conservés (tels que le domaine Paired (PRD) ou le domaine POU) et sur la position des introns. Dans le cas de la super classe en complexe, la position du gène dans le complexe constitue également un critère de classification (Gehring *et al.*, 1994).

Cette classification peut aussi apporter des informations quant aux modes de fonctionnement de certaines familles de gènes. Dans certains cas, la fonction même peut être conservée et les résultats dans une espèce donnée peuvent être applicables à d'autres espèces.

20

3-1 Super classe organisée en complexe génomique : Gènes homéotiques et gènes Hox

Les gènes homéotiques codent des protéines contenant un homéodomaine de la classe Antp, c'est à dire possédant une homologie de séquence de l'ordre de 60% ou plus avec l'homéodomaine du gène *Antp* et de 100% au niveau de la troisième hélice. Ces gènes codent également un motif hexapeptide (IleTyrProTrpMetLys) juste en amont de l'homéodomaine, dont la fonction n'est pas connue, mais dont la présence est utilisée comme critère de classification.

Le HOM-C de drosophile contient huit gènes homéotiques: labial (Lab), proboscipedia (Pb), deformed (Dfd), Sex comb reduced (Scr), Antennapedia (Antp), ultrabithorax (Ubx) et abdominal-A (Abd-A). Le gène le plus en amont du HOM-C, abdominal-B (Abd-B), ne correspond pas exactement à la classe Antp et il code un hexapeptide très dégénéré sous la forme d'un Trp en position -6 ou -7 de l'homéodomaine. Mais c'est à cette classe qu'il reste le plus apparenté.

Les gènes du HOM-C n'ont pas tous d'homologue chez les vertébrés. Dans le cas de Scr, Antp, Ubx et Abd-A qui codent des homéodomaines sont très proches, il n'est pas possible de définir des homologues précis chez les vertébrés car il n'y a pas d'homologie en dehors de l'homéodomaine. Cependant, tous ces gènes restent apparentés (Cf. Fig.1).

Les gènes des 4 complexes Hox peuvent être alignés en 13 groupes fortement apparentés dits groupes paralogues. Dans chaque complexe, certains paralogues sont absents, suggérant que la duplication des complexes ne s'est produite qu'après la mise en place des 13 gènes et que certains gènes ont été perdus pendant ou après l'événement de duplication. Le complexe HoxA est sur le chromosome 6 chez la souris (7 chez l'homme), HoxB sur le chromosome 17 (11 chez l'homme), HoxC sur le chromosome 15 (12 chez l'homme) et HoxD sur le chromosome 2 (2 chez l'homme).

L'homologie entre les gènes de vertébrés et d'insectes est telle que certains gènes des complexes Hox peuvent mimer l'activité biologique de leurs homologues du HOM-C. Ainsi, la surexpression chez la drosophile du gène *Dfd* ou du gène de souris *HoxD4* provoque le même

phénotype (McGinnis *et al.*, 1990). De même, le gène *HoxB6* (*Hox-2.2*) de souris surexprimé chez la drosophile génère un phénotype semblable à celui produit par la surexpression de *Antp* (Malicki *et al.*, 1990).

Chez la drosophile, le HOM-C contient des gènes à boîte homéo qui n'ont pas de fonction homéotique et que l'on ne retrouve pas dans les complexes des vertébrés: *Fushi-tarazu* (*Ftz*) qui est un gène de segmentation, *bicoïd* (*Bcd*) qui est dit d'origine maternelle car présent sous forme d'ARNm dans l'ovule avant la fécondation et qui permet la définition de l'axe antéropostérieur et *zerknüllt 1* et 2 (*Zen1* et 2) qui sont impliqués dans la détermination de l'axe dorsoventral. De plus, si *Ftz* et *Zen* codent des homéodomaines de type Antp, ils ne contiennent pas la séquence hexapeptide. L'homéodomaine de Bcd est, lui, très divergent par rapport à la classe Antp. Ces gènes ne proviennent donc pas d'événement de duplication à l'intérieur du HOM-C mais ont été intégrés après la mise en place du complexe.

Chez les mammifères, deux gènes apparentés au gène de segmentation *even-skipped* (*Eve*) de drosophile, *Evx-1* et 2, sont liés aux complexes HoxA et D alors que chez la drosophile, *Eve* est localisé sur un autre chromosome que le HOM-C. Cependant, un gène de la classe Eve a été isolé chez les coraux à proximité d'un gène du type Antp. L'association des gènes *Evx* avec les complexes Hox, représenterait donc un arrangement ancestral qui a divergé chez les insectes (Miler et Miles, 1993). Ils ont, de plus, la particularité d'être transcrits dans le sens opposé à tout le complexe.

Comme il a été décrit précédemment, ces gènes ont conservé leur organisation spatiale sur les différents complexes (Duboule *et al.*, 1989). Et comme chez la drosophile, les domaines d'expression des gènes Hox sont fonctions de la position de ces gènes sur l'axe 5'-3' du complexe. On peut également corréler la position de ces gènes avec leur activation par l'acide rétinoïque à la fois *in vivo* (Cho *et al.*, 1990) et *in vitro* sur des modèles cellulaires tels que les cellules de carcinomes embryonnaires humains (Simeone *et al.*, 1990)(Stornaiuolo *et al.*, 1990)(Simeone *et al.*, 1991). En effet, il y a une activation progressive de ces gènes dans le sens 3'-5' du complexe sous l'effet de doses croissantes d'acide rétinoïque.

Enfin, au niveau fonction, de nombreux exemples montrent que les gènes Hox de mammifères peuvent générer, lorsqu'ils sont mutés, des altérations du développement, et



Fig 4 : Différentes classes de protéines à homéodomaine

.

certaines de type homéotique (Kessel *et al.*, 1990)(Kessel *et al.*, 1991). Dans le cas des mutations du type perte de fonction, il est assez difficile d'analyser ces effets sachant que chaque gène Hox possède un à quatre paralogues sur les autres complexes et qu'il peut donc y avoir compensation par ces autres gènes. Par contre, les mutations dites gain de fonction, peuvent générer de nombreuses transformations homéotiques, en particulier, au niveau des vertèbres (Voir la revue de McGinnis et Krumlauf, 1992).

3-2 Super classe dispersée au niveau génomique

Cependant, la majorité des gènes à boîte homéo ne font pas partie de ces complexes homéotiques et ils codent, pour la plupart, un homéodomaine divergent du type Antp. La super classe dispersée comprend actuellement 15 classes. La figure 4 regroupe les différentes classes en fonction de leur structure. Pour la plupart de ces classes, la divergence de l'homéodomaine par rapport au consensus général est de plus de 30%.

a) Classe avec hexapeptide

Les gènes du HOM-C/Hox ne s'expriment pas aux extrémités antérieures et postérieures de l'embryon. Plusieurs gènes sont requis pour la définition de la tête et de la queue. Deux de ces gènes chez la drosophile codent pour des protéines à homéodomaine de type Antp possédant également la séquence hexapeptide bien qu'ils soient localisés à l'extérieur des complexes. Cela suggère qu'ils proviendraient de la duplication d'un même gène ancestral qui aurait généré les gènes des complexes homéotiques et des gènes intervenant dans d'autres étapes du développement.

- Gènes de type Caudal

Le gène *caudal* (*Cad*) permet la définition des segments les plus postérieurs de la drosophile. Il présente deux phases d'expression au cours du développement. Il est exprimé initialement, sous forme d'ARNm dans l'oocyte non fécondé. Après la fécondation, la protéine est traduite et forme un gradient dans l'embryon syncitial. Puis au stade cellularisé, la

transcription reprend dans le zygote (Mlodzyk *et al.*, 1987). Plusieurs gènes homologues à *Cad* ont été isolés chez les vertébrés et *Caenorhabditis elegans* (*C. elegans*). Ils s'expriment également dans la partie postérieure de l'embryon et un de ces gènes, *Cdx-1* a également une expression en deux phases (Duprey *et al.*, 1988).

- Gènes de type Empty-spiracle

Deux gènes de ce type ont été isolés chez la drosophile. Ils sont exprimés dans les segments les plus antérieurs. Les mutations gain de fonction génèrent des phénotypes de type homéotiques, contrairement aux mutations perte de fonction qui provoquent la disparition de certaines structures de la tête (Waldorf et Gehring, 1992). Ces gènes sont de type "gap" et pourraient être des cibles du gène *Bcd*. Chez les vertébrés, les deux gènes isolés sont exprimés dans le cerveau et les structures qui en dérivent (Simeone *et al.*, 1992a).

Les gènes de type *Caudal* et *Empty-spiracle* ont une expression similaire chez les insectes et les vertébrés. Comme dans le cas des gènes homéotiques, il y a une conservation au niveau structure et fonction.

- Gènes de type TCL

Bien que possédant un hexapeptide très divergent, ces gènes peuvent être rattachés au niveau phylogénétique à cette classe. Le plus connu est *TCL-3* ou *HOX-11* qui est impliqué dans des translocations chromosomiques qui conduisent à une dérégulation de ce gène et au développement de leucémies de type T (<u>T-C</u>ell <u>L</u>eukemia) (Lu *et al.*, 1991)(Hatano *et al.*, 1991). Un deuxième gène a été isolé chez la souris qui serait un autre membre de la famille, et un chez la drosophile.

b) Protéines à homéodomaine

- Gène de type Msh

Cette classe est très ancienne, un de ces membres ayant été isolé chez l'hydre de mer. Il y a, à l'heure actuelle, trois gènes de types de Msh connus chez les vertébrés (Msx-1 ou Hox-7, *Msx-2* ou *Hox-8* et *Msx-3*) et 4 types chez le poisson-zèbre (Holland, 1991). Ces gènes sont impliqués dans les interactions ectoderme-mesoderme, en particulier lors du développement des membres (Roberts *et al.*, 1991). Deux membres de cette famille, *Msx-1* et 2, semblent impliqués dans le développement de l'oeil chez les vertébrés. En effet, leur expression est très précoce et délimite les territoires qui se différencieront par la suite en rétine, en corps ciliés et en iris (Monaghan *et al.*, 1991).

- Gènes de type Distal-less

Chez la drosophile, *Distal-less (Dll)* est impliqué dans le développement des membres, en particulier, ceux des segments antérieurs (Cohen *et al.*, 1989). Quatre gènes de ce type, appelés *Dlx*, ont été isolés chez les vertébrés. Ils sont essentiellement exprimés dans le cerveau antérieur et les structures qui en dérivent, chacun de ces gènes ayant des profils d'expression restreints (Dollé *et al.*, 1992). Comme chez la drosophile, on trouve une faible expression de *Dlx-1* dans les membres en développement (Dollé *et al.*, 1992).

- Classe NK1

Cette classe d'homéodomaine a été isolée initialement chez *C. elegans*. Par la suite, des expériences d'amplification en chaîne (PCR) ont montré que ce type d'homéodomaine était très conservé des annélides aux vertébrés. Cependant, il y a très peu de données quant à la structure de ces gènes en dehors de l'homéodomaine.

- Protéines à homéodomaine apparenté aux homéodomaines de type PRD

Ces protéines ne possèdent pas de domaine PRD mais un homéodomaine qui a 55 à 75% d'homologie avec les homéodomaines de type PRD. On ne retrouve pas une Ser en position 50 comme dans les homéodomaines de type PRD mais en général une Gln ou plus rarement une Lys. De nombreux gènes ont été isolés chez la drosophile, chez *C. elegans* et chez les vertébrés mais un seul gène possède son homologue dans une autre espèce. Il s'agit d'*orthodenticle (Otd)* de drosophile qui a été isolé chez la souris (Simeone *et al.*, 1992b). Ce gène s'exprime dans les segments antérieurs de la drosophile. Chez les vertébrés, deux gènes homologues à *Otd* ont été isolés, *Otx-1* et *Otx-2*, et ils s'expriment dans le cerveau antérieur. D'après leurs profils

d'expression, ils pourraient également être impliqués dans le développement de l'oeil (Simeone *et al.*, 1992b)(Nucci *et al.*, 1993). D'autres gènes apparentés au type PRD ayant une expression spécifique dans la rétine, ont également été isolés chez le poisson rouge (Levine *et al.*, 1994) et la souris (Liu *et al.*, 1994).

c) Protéines à plusieurs domaines de fixation à l'ADN

- Classe Paired (PRD)

Isolés par homologie avec le gène *Paired* de drosophile, ces gènes possèdent une séquence très conservée, la boîte PRD qui peut être associée avec une boîte homéo. La boîte PRD s'étend sur 384 pb et code un domaine de 128 aa capable de fixer l'ADN (Treisman *et al.*, 1991). L'analyse de ce domaine prédit une structure en trois hélices α : une première hélice en amino-terminal de l'aa 23 à 31, et deux hélices très proches en carboxy-terminal du domaine PRD, de l'aa 80 à 89 et de l'aa 94 à 106. Cependant, cette structure n'a pas d'homologie avec les motifs de type hélice-tour-hélice ou hélice-boucle-hélice déjà connus, en particulier avec le motif hélice-tour-hélice des homéodomaines. L'homéodomaines classiques c'est à dire en hélice-tour-hélice. Ces homéodomaines sont caractérisés, entre autre, par une Ser en position 50. On retrouve également dans certains de ces gènes une séquence octapeptide dont le consensus est HisSerIleAspGlyIleLeuGly. Cette classe de protéines sera décrite de façon détaillée dans le chapitre suivant.

- Classe POU

Quatre facteurs de transcription, Pit-1, spécifique de la glande pituitaire (Ingraham et al., 1988)(Bodner et al., 1988), Oct-2, spécifique des cellules B(Scheidereit *et al.*, 1988)(Clerc *et al.*, 1988), Oct-1 dont le produit est ubiquitaire (Sturn *et al.*, 1988), et *unc-86* de *C.elegans* (Finney et al., 1990), isolés de façon indépendante et sur des critères biochimiques ont permis la caractérisation de cette famille de gènes qui contient un nouveau domaine de fixation à l'ADN, le domaine POU (pour <u>P</u>it-1, <u>O</u>ct, <u>unc-86</u>)(voir la revue de Rosenfeld, 1991). Le domaine POU est bipartite et comprend un homéodomaine divergent du type Antp, l'homéodomaine de
type POU (POU_{HD}) et une région de 76 à 78 aminoacides moins conservée, le domaine POUspécifique (POU_S) également capable de se fixer à l'ADN. Cependant, séparément, ces deux domaines se lient à l'ADN avec une très faible affinité. Ils sont séparés par une région peu conservée. Les protéines de cette famille sont capables de s'homodimériser entre elles et d'interagir avec d'autres membres de la famille POU ainsi qu'avec d'autres facteurs de transcription comme VP16 ou le récepteur aux oestrogènes (voir la revue de Wegner *et al.*, 1993).

L'importance de plusieurs membres de cette famille dans la détermination de phénotypes cellulaires spécifiques a été génétiquement établie. Ainsi, des mutations dans le gène du facteur de transcription Pit-1, sont responsables de deux cas de nanisme génétiquement transmis chez la souris. Les souris mutées dans *Pit-1* ne produisent plus d'hormone de croissance, plus de prolactine ni de TSH (Thyroïd Stimulating Hormone) et ne possèdent plus de cellule de type somatotrope, lactotrope et thyréotrope dans l'hypophyse antérieure. Une mutation dans le locus de *unc-86* de *C. elegans* a montré que ce facteur était nécessaire à l'engagement dans plusieurs voies neuroblastiques (Finney *et al.*, 1988)(Finney *et al.*, 1990). Et enfin, la suppression chez la souris du gène *Oct-2* par recombinaison homologue a montré que ce gène jouait un rôle clé dans la différenciation terminale et la survie des cellules B (Corcoran *et al.*, 1993). De nombreux membres de cette famille s'exprimant dans le système nerveux central (SNC) ont été isolés. Ils ont des profils d'expression distincts au niveau spatial et temporel, ce qui suggère que ces gènes pourraient jouer un rôle dans le développement du système nerveux central (voir les revues de Treacy et Rosenfeld, 1992a et Wegner *et al.*, 1993).

Ils sont classés en six familles basées sur la séquence de la région de jonction entre le POU_S et le POU_{HD} et celle de la partie amino-terminale de l'homéodomaine.

- Protéines à homéodomaines comportant des domaines dits en doigt de zinc

Cette classe code des protéines extrêmement grandes qui peuvent contenir jusqu'à dixsept domaines en doigts de zinc et quatre homéodomaines comme le facteur de transcription humain ATBF1 (Morinaga *et al.*, 1991) ou seize domaines en doigts de zinc et trois homéodomaines comme son homologue putatif chez la drosophile, Zfh-2 (Zinc <u>F</u>inger <u>Homeodomain</u>). Les homéodomaines de cette classe sont relativement proches de ceux des gènes de la classe LIM qui possèdent également un domaine capable de fixer le zinc. Le domaine en doigts de zinc de ces protéines est du type $(Cys)_2/(His)_2$ et présente des homologies avec les gènes *Hunchback* et *Krüppel* de drosophile qui sont des gènes "gap". Cette classe de protéines à homéodomaines pourrait provenir d'un événement de recombinaison entre deux ou plusieurs gènes de contrôle du développement.

- Classe CUT

Les protéines codées par les gènes de cette classe possèdent, en plus de l'homéodomaine, trois domaines conservés de 73 aa appelés répétitions cut, chacune de ces répétitions ayant plus de 50% d'homologie avec les deux autres. Cette séquence a été initialement trouvée dans le produit du gène *cut* de drosophile et ne correspond à aucun domaine connu (Blochlinger *et al.*, 1988). Par la suite, deux facteurs de transcription ont été isolés contenant un homéodomaine de type cut et trois répétitions cut en amont de l'homéodomaine. Il s'agit de la "CCAAT Displacement Protein", CDP, chez l'homme (Neufeld *et al.*, 1992) et de Clox, chez le chien (Andrès *et al.*, 1992). Les gènes codant ces facteurs de transcription sont les homologues vertébrés du gène *cut* de drosophile. Un gène possédant des séquences de type répétition cut a également été isolé chez *C. elegans*. (Waterston *et al.*, 1992).

Les répétitions cut ont été récemment montrées comme étant capables de fixer l'ADN de façon spécifique, en particulier sur des séquences de type CCAAT, ce qui serait en accord avec la fonction de CDP (Andrès *et al.*, 1994). Les séquences reconnues par chaque domaine cut sont fréquemment mais pas toujours reconnues par les autres répétitions. Cependant, leurs spécificités de fixation sont très proches (Harada *et al.*, 1993). Enfin, en l'absence d'ADN, les domaines cut peuvent interagir *in vitro* avec les homéodomaines de type cut. Cette interaction augmente l'affinité de liaison à l'ADN du complexe et est spécifique des homéodomaines de ce type (Andrès *et al.*, 1994)(Harada *et al.*, 1993). Il y a donc coopérativité entre ces différents domaines pour la liaison à l'ADN, phénomène déjà observé pour les autres protéines à domaine POU.

Plusieurs études chez les mammifères montrent que ces protéines de type cut agiraient en tant que régulateurs négatifs de la transcription. Cela a été montré sur plusieurs promoteurs : la

protéine cux, homologue murin de cut de drosophile, sur le promoteur de la N-CAM (Neural-Cell Adhesion Molecule)(Valarché *et al.*, 1993), CDP sur le promoteur de *c-myc* (Dufort *et al.*, 1994). La protéine CDP est exprimée dans des cellules indifférenciées et une diminution de son activité peut être corrélée avec une augmentation de l'expression de marqueurs de différenciation (Skalnik *et al.*, 1991). CDP pourrait réguler de façon négative ces marqueurs dans les cellules non différenciées. De plus, la protéine CDP, comme son nom l'indique, peut déplacer des facteurs fixés sur des séquences CCAAT. Ces séquences sont présentes dans certaines régions promotrices et fixent des facteurs régulateurs de la transcription. La CDP pourrait réprimer la transcription en empêchant la fixation de facteurs activateurs.

d) Protéines à domaines conservés ne fixant pas l'ADN

- Classe LIM

Le domaine LIM contient un motif riche en cystéine capable de fixer le zinc mais dont les capacités à lier l'ADN restent encore à étudier. Il a été initialement identifié dans les produits de 3 gènes: <u>lin-11</u> de *C. elegans* (Freyd *et al.*, 1990), <u>Isl1</u> chez le rat (Karlson *et al.*, 1990) et <u>mec-3</u> de *C. elegans* (Way *et al.*, 1988), lui même régulé par le produit du gène POU *unc-86* (Xue *et al.*, 1993). Ce domaine est très conservé au cours de l'évolution.

Trois classes de protéines à domaine LIM ont été découvertes. Les protéines de type LIM-HD contiennent deux domaines LIM en association avec un homéodomaine. La deuxième classe de type LIM-seul ne contient pas d'homéodomaine mais seulement un à trois domaines LIM. Une troisième classe a été définie récemment, LIM-PK, où 2 domaines LIM seraient associés à un domaine protéine kinase (Mizuno *et al.*, 1994).

L'importance de plusieurs gènes du type LIM-HD dans la spécification de type cellulaire a déjà été montré. Ainsi, *Isl1* joue un rôle essentiel dans l'expression cellule-spécifique de l'insuline (Ericson *et al.*, 1992)(German *et al.*, 1992) ou *mec-3* dans la différenciation de certains neurones mécanosensoriels (Way *et al.*, 1988).

Les gènes de la classe LIM-seul peuvent coder de un à trois domaines LIM. Seulement deux protéines avec un seul domaine LIM sont connues actuellement: CRIP (Birkenmeier *et al.*, 1986) et rEPS1 (Nalik *et al.*, 1989) chez le rat, et deux protéines à 3 domaines LIM, la zyxin de poulet (Sadler *et al.*, 1992) et la testin de rat (Sanchez-Garcia et Rabbitts, 1994). Il y a, par contre, de nombreuses protéines possédant deux domaines LIM. L'une d'entre elles, MLP (Muscle LIM Protein) joue un rôle essentiel dans la différenciation myogènique (Arber *et al.*, 1994). Trois protéines à deux domaines LIM ont été isolées chez l'homme, RTBN1, RTBN2 et RTBN3. Les gènes des deux premières sont impliqués dans des translocations chromosomiques t(11;14) et t(7;11) qui génèrent des leucémies aiguës de cellule T (revue de Sanchez-Garcia et Rabbitts, 1993). Les protéines LIM sont capables d'interagir *in vivo* avec des protéines à domaine basique et hélice-boucle-hélice (Wadman *et al.*, 1994) et avec des protéines à domaine POU (Xue *et al.*, 1993). Ces interactions sont partiellement médiées par le domaine LIM (Schmeichel *et al.*, 1994) et, dans certains cas, ce domaine peut réprimer la fixation spécifique à l'ADN d'homéodomaine de type LIM (Xue *et al.*, 1993).

- Classe NK2

Des gènes de la classe NK2 ont été isolés chez les annélides, chez la drosophile et chez les vertébrés. Chez ces derniers, ils sont exprimés dans la partie antérieure du cerveau embryonnaire.

Ils possèdent une séquence conservée codant un peptide de dix-sept aa, le CP (<u>C</u>onserved <u>P</u>eptide) en aval de la boîte homéo (Price *et al.*, 1992). Le premier gène vertébré isolé, *TTF1*, code une protéine qui a des propriétés de fixation à l'ADN très différentes des autres protéines à homéodomaine connues (Guazzi et al, 1990). Bien que possédant un homéodomaine similaire à celui d'Antp au niveau des résidus contactant le consensus classique ATTA/TAAT, cette protéine reconnaît un consensus différent de type CTCA (Guazzi *et al.*, 1990). Il est possible que les aa qui divergent par rapport au type Antp soient suffisants pour modifier la structure tertiaire de l'homéodomaine. Une autre hypothèse est que les aa qui encadrent les structures en hélice α , sont capables de modifier l'affinité de cette protéine pour le consensus ATTA.

- Gène de type Engrailed

Les gènes de la classe Engrailed sont caractérisés par, outre la boîte homéo de type Engrailed, quatre régions codant des domaines conservés. Le premier domaine conservé, EH1, d'environ treize aa, est situé dans la région amino-terminale. Les deux domaines suivants qui sont adjacents, EH2 et EH3, sont en amont de l'homéodomaine et font respectivement vingt et huit aa. Le quatrième domaine EH5 de vingt aa est situé directement en carboxy-terminal de l'homéodomaine. Chez la drosophile, on trouve deux gènes de cette classe, *Engrailed* et *Invected* (Kuner *et al.*, 1985) dont les homologues mammifères sont *En-1* et *En-2* (Joyner *et al.*, 1987). Chez le poisson-zèbre, 3 gènes de la famille Engrailed ont été isolés (Joyner et Hanks, 1991a).

Une des principales caractéristiques d'En est qu'il est capable de réprimer, *in vitro*, la transcription activée de ses gènes cibles mais pas le taux de base de transcription. Cette répression est médiée par un domaine situé en amont de l'homéodomaine. Quand ce domaine répresseur est cloné en fusion avec un domaine de fixation à l'ADN autre que celui d'En, il est capable de réprimer un promoteur activé. Ce domaine répresseur est riche en Ala comme celui d'autres répresseurs de drosophile tels que les produits des gènes *Krüppel* et *Eve* mais ces derniers sont capables de réprimer la transcription basale (Jaynes *et al.*, 1991).

Chez la souris, En-1 et -2 s'expriment dans le tube neural et le cerveau. En-1 est également détecté dans les cellules du dermomyotome qui vont former les vertèbres et les muscles du corps et de la tête (Davis et Joyner, 1988)(Davis *et al.*, 1991). En-2 est exprimé dans une population de myoblastes (Logan *et al.*, 1993). La délétion ciblée d'En-1 est létale (Wurst *et al.*, 1994) alors que celle de En-2 induit des anomalies au niveau du cervelet mais reste viable (Joyner et al, 1991b).

- Classe PBC

Cette classe de gènes code des protéines possédant un homéodomaine dit atypique car contenant des aa supplémentaires entre les hélices α par rapport à un homéodomaine classique, ainsi que des motifs conservés, les domaines PBC-A et -B, directement en amont de l'homéodomaine (Bürglin *et al.*, 1992). Ces domaines PBC possèdent de nombreux résidus basiques et il est possible qu'ils soient capables de fixer l'ADN.

Trois gènes de cette classe ont été isolés chez l'homme, *PBX-1*, -2 et -3 ainsi qu'un gène chez *C. elegans*, *ceh-20* (Wasterton *et al.*, 1992). Le gène *PBX-1* est impliqué dans des leucémies aiguës de type pré B, suite à des événements de translocations qui produisent des protéines de fusion avec E2a (Kamps et al, 1990)(Rabbits *et al.*, 1991). Cependant, les propriétés transformantes du produit de ce gène ne sont pas liées à la présence de l'homéodomaine (Monica

et al., 1994). L'effet oncogénique serait médié par des interactions avec d'autres facteurs de transcription.

- Classe à domaine dit "leucine- zipper"

Cette classe n'a été décrite que chez la plante *Arabidopsis thaliana*. Elle code des protéines connues sous le nom générique HAT (<u>Homeobox *Arabidopsis Thaliana*</u>) qui possèdent un motif "leucine zipper" lié à l'homéodomaine. Ce motif permet aux protéines de se dimériser. Il semble que cette classe soit apparue après la divergence entre les plantes et les animaux car aucun gène de cette classe n'a pu être mis en évidence chez ces derniers (Shena *et al.*, 1994). Ces gènes jouent un rôle important dans le développement et l'expression ectopique du gène *HAT4* provoque de nombreuses anomalies telles que des modifications de la forme des feuilles ou de la pigmentation (Shena *et al.*, 1993).

e) Protéines à homéodomaines atypiques

Ces homéodomaines sont caractérisés par des insertions ou des délétions dans les séquences entre les différentes hélices ainsi que par l'absence de certains résidus conservés à 100% dans les homéodomaines conventionnels.

MAT α 2 et HNF-1 constituent deux exemples d'homéodomaine atypique. Ainsi, il y a une insertion de trois aa entre les hélices 1 et 2 dans l'homéodomaine de MAT α 2. Cette protéine s'hétérodimérise avec un autre facteur du locus MAT, MATa1, au stade diploïde de la levure. Cet hétérodimère va réprimer les gènes spécifiques du stade haploïde. Dans l'homéodomaine de HNF-1 (ou HNF-1 α), il y a une insertion de vingt-trois aa dans ce qui est normalement un tour de deux aa entre la deuxième et la troisième hélice (Chouard *et al.*, 1990). Un deuxième gène codant un homéodomaine à 100% identique à celui du produit de *HNF-1* a été isolé, il s'agit de *vHNF-1* (ou *HNF-1\beta*). Les produits de la famille HNF-1 sont également capables de former des homodimères et des hétérodimères (Rey-Campos *et al.*, 1991)(Mendel *et al.*, 1991a et b).

Cependant, malgré une structure secondaire divergente, la structure tertiaire et le positionnement par rapport à l'ADN de ces homéodomaines atypiques sont équivalents à ceux d'un homéodomaine conventionnel, les modifications ne se produisant qu'entre les hélices.

	DROSOPHILE	POISSON-ZEBRE	CAILLE	SOURIS
Sous classe I PRD+ OCT+ HD-	pox meso (Poxm)			Pax-1 Pax-9
Sous classe II PRD+ OCT+ HD+	gooseberry (Gsb) gooseberry neuro (Gsbr paired (Prd)	1)		Pax-3 Pax-7
Sous classe III PRD+ OCT+ HD partiel	pox neuro (Poxn)HD-	Pax(zf-b)		Pax-2 Pax-5 (BSAP) Pax-8
Sous classe IV PRD+ OCT- HD+	eyeless (Ey)	Pax(zf-a)	Pax-QNR	Pax-6 Pax-4

<u>Tableau 1</u> : Classification des gènes Pax

PRD +/- : domaine paired présent ou absent; HD +/- : homéodomaine présent ou absent; OCT +/- : octapeptide présent ou absent

Les gènes de la classe PBC, déjà décrite dans la partie d), codent également un homéodomaine atypique.

IV Famille des gènes Pax

Les gènes contenant une boîte Paired (PRD), les gènes Pax (<u>Paired box</u>), ont été initialement caractérisés chez la drosophile (Bopp *et al.*, 1986). Par homologie avec le gène *paired (Prd)* qui exerce successivement la fonction d'un gène "pair-rule" puis d'un gène de segmentation, cinq nouveaux gènes ont été isolés :*gooseberry (Gsb), gooseberry neuro (Gsbn)* (Baumgartner *et al.*, 1987), *pox neuro (Poxn), pox meso (Poxm)* (Bopp *et al.*, 1989a et b) et *eyeless (Ey)* (Quiring *et al.*, 1994). La comparaison de séquence protéique entre ces différents gènes a permis de définir une nouvelle classe d'homéodomaine dite de type paired dans Prd, Gsb et Gsbn, ainsi qu'un nouveau domaine conservé, le domaine Paired (PRD), présent dans tous ces gènes.

Par la suite, les gènes de la famille Pax ont été isolés et étudiés chez de nombreuses espèces, en particulier chez le poisson-zèbre et les mammifères (Noll, 1993). Plusieurs anomalies de développement, que ce soit chez les mammifères ou chez la drosophile, sont dues à des mutations dans cette famille de gènes. Ces mutations sont létales au stade homozygote, et provoquent au stade hétérozygote, des désordres plus ou moins graves, en particulier au niveau du système nerveux central. L'étude des animaux mutés montre l'importance de ces gènes dans le développement et permet une approche de leurs fonctions *in vivo*.

Ces gènes peuvent être classés en 4 sous classes qui reflètent leurs homologies de structure ainsi que leurs modes d'expression souvent similaires à l'intérieur d'une même sous classe et cela d'une espèce à l'autre (Cf. Tableau 1).



<u>Tableau 2</u>: Expression de *Pax-1* au cours du développement chez la souris (d'après Tremblay et Gruss, 1994)

1- Expression au cours du développement

Chez la drosophile, les gènes de type "pair-rule" et plus particulièrement *paired*, contrôlent les gènes de segmentation *Gsb* et *Gsbn* qui sont impliqués dans la formation de la cuticule pour le premier, et dans le développement du SNC pour le second. Les gènes *Poxn* et *Poxm* sont contrôlés par les gènes de segmentation dont *paired* (qui a également la fonction de gène de segmentation). *Poxn* est impliqué dans le développement du système nerveux central et périphérique et *Poxm* dans la myogenèse. Les gènes Pax de drosophile jouent un rôle dans la segmentation. L'implication de ces gènes, chez les vertébrés, dans des phénomènes de segmentation n'est pas établie bien qu'ils jouent un rôle dans la spécification de l'information de position.

Chez les vertébrés, plusieurs gènes Pax s'expriment dans les tissus adultes mais l'expression initiale de tous les gènes Pax commence très tôt dans l'embryon. Ces gènes sont exprimés dans différents types cellulaires mais tous, excepté *Pax-1* et *Pax-9*, contribuent au développement précoce du système nerveux. De plus, contrairement aux gènes Hox qui ont une expression limitée le long de l'axe antéro-postérieur, les gènes Pax s'expriment sur toute la longueur du tube neural et cette expression diminue progressivement au cours du développement. Il y a bien une régionalisation de l'expression de ces gènes dans le tube neural mais sur l'axe dorso-ventral et dans les tissus d'origine mésodermique où l'expression semble être confinée dans les structures segmentées.

a) Expression de Pax-1 et Pax-9

(Cf. Tableau 2)

Chez la souris, on peut détecter l'expression de *Pax-1* à partir de E9 dans le sclérotome des somites en différenciation, dans le mésenchyme ventro-latéral par rapport à la notochorde à E10, dans les condensations périchordales à E12 et finalement dans les disques intervertébraux durant les derniers stades du développement. Chez l'adulte, on retrouve l'expression de *Pax-1* dans le thymus (Deutsch *et al.*, 1988) (Wallin *et al.*, 1994). Ce gène est également exprimé de façon transitoire dans les bourgeons de membres en développement (Timmons *et al.*, 1994). L'étude des mutations dans ce gène montre que l'expression de *Pax-1* est essentielle au



Tableau 3 : Expression de Pax-2, Pax-5 et Pax-8

au cours du développement chez la souris

(d'après Tremblay et Gruss, 1994)

développement du squelette axial et appendiculaire et plus généralement aux événements chondrogèniques.

Très proche au niveau peptidique de Pax-1 (80,1% d'homologie entre PAX-9 et PAX-1), Pax-9 présente des domaines d'expression similaires (Stapleton et al., 1993)(Wallin et al., 1993).

Pax-1 et Pax-9 sont les seuls gènes Pax à ne pas être exprimés dans le SNC. L'expression spatialement restreinte des autres gènes Pax suggère un rôle dans la régionalisation des différentes aires du cerveau. Cette hypothèse est soutenue par de nombreuses données, en particulier par l'observation que les domaines d'expression des gènes Pax respectent les limites anatomiques des neuromères pendant le développement. Dans le cerveau adulte, leur expression est plus restreinte. Cependant, la distribution des messagers chez l'adulte, le long de l'axe antéropostérieur reste similaire à celle dans le cerveau embryonnaire, suggérant un rôle pour ces gènes à la fois dans la différenciation neuronale et dans la maintenance des différents sous types neuronaux (Stoykova *et al.*, 1994) (Cf. Fig.5).

b) Expression de Pax-2, Pax-5 et Pax-8

(Cf. Tableau 3)

Pax-2 et Pax-8 sont initialement exprimés dans le tube neural et le rhombencéphale (Nornes et al., 1990)(Dressler et al., 1990)(Platchov et al., 1990). Dans le tube neural, cette expression est restreinte à un groupe de neurones différenciés post mitotiques de chaque côté de la limite dorso-ventrale (Cf. Fig.5). Ces cellules donneront les interneurones de la matière grise du tube neural. Dans le cerveau, Pax-2 et Pax-8 sont exprimés dans la zone intermédiaire entre le myélencéphale et le métencéphale, au dessus de la limite cerveau moyen-rhombencéphale, alors que des taux élevés de Pax-5 sont détectés dans cette limite et au niveau du tegmentum mésencéphalique (Cf. Fig.5)(Asano et Gruss, 1992). Les messagers de Pax-8 deviennent rapidement indétectables dans le cerveau alors que Pax-2 continue à être transcrit durant tout le développement. L'étude de l'expression de Pax(zf-b) chez le poisson-zèbre (Krauss et al., 1991b), qui présente de fortes homologies avec Pax-2 et Pax-5, montre que ce gène pourrait être impliqué dans la mise en place de la limite cerveau moyen-rhombencéphale. En effet, l'injection



Fig.5: Expression des gènes Pax chez la souris dans le système nerveux central.

(A) Distribution longitudinale à E12,5. (B) Expression sur une coupe de tube neural à E11,5. eo : épithélium olfactif, cs : corde spinale, my : myélencéphale, di : diencéphale, ms : mésencéphale, cv : cervelet, te : téléncéphale, pn : pons, TT : toît du tube neural, PA : plaque alaire, PB : plaque basale, PT: plancher du tube neural, SL : Sulcus limitans, ZI : zone intermédiaire, ZV : zone ventrale, IR : isthme rhombencéphalique, CP : commissure postérieure

(d'après Tremblay et Gruss, 1994)

d'anticorps anti Pax(zf-b) abolit la formation du mésencéphale (Krauss *et al.*, 1992). Le gène *Pax-2* est également exprimé durant les processus d'induction qui conduisent à la formation de l'oeil et de l'oreille (Nornes *et al.*, 1990).

A partir de E9-10 chez la souris, Pax-2 et Pax-8 sont exprimés dans le système d'excrétion en développement, en particulier dans le mésenchyme métanéphrique condensé et dans les tubules épithéliaux qui en résultent. Pax-2 est également détecté dans le canal de Wolff et dans les uretères qui se développent à partir de ce canal. L'expression de ce gène est nécessaire durant les premières étapes de la conversion du mésenchyme en épithélium pendant le développement des reins (Rothenpieler et al., 1993). Cependant, cette expression doit être transitoire car des souris transgéniques surexprimant Pax-2 présentent de graves anomalies rénales. L'expression continue de ce gène restreint le potentiel de différenciation des cellules épithéliales et sa répression est nécessaire pour un développement normal des reins (Dressler et al., 1993). Cela peut être lié au fait que, chez l'homme, les tumeurs de Wilms qui sont des tumeurs indifférenciées des reins, expriment Pax-2 et Pax-8 (Dressler et Douglas, 1992)(Poleev et al., 1992). Enfin, une mutation associée à une délétion complète du locus Pax-2 et des loci qui l'encadrent, provoque des anomalies importantes au niveau des reins. Il s'agit de la mutation Krd (Kidney retinal defects), induite par l'insertion d'un transgène. Les animaux mutants présentent également des anomalies dans la rétine où le nombre de cellules dans la couche ganglionnaire et dans la couche nucléaire interne est très fortement réduit (Keller et al., 1994). Pax-2 est exprimé dans la rétine au cours du développement et ce sont les seuls résultats qui associent ce gène avec une altération de l'oeil. Cependant, la surexpression de Pax-2 chez la souris ne provoque aucune anomalie des yeux (Dressler et al., 1993), et ces altérations peuvent être dues à la suppression des gènes voisins.

L'analyse de l'expression de *Pax-2* dans les tissus adultes révèle des taux élevés de messagers de *Pax-2* dans le système uro-génital: on peut le détecter dans les reins adultes, dans les uretères, dans la vessie, dans l'épididyme, les ovaires et l'utérus mais pas dans les testicules. Ce gène a donc une deuxième phase d'activité chez l'adulte (Fickenscher *et al.*, 1993). De même, *Pax-8* est exprimé dans la thyroïde des tous premiers stades jusqu'à chez l'adulte. Il constitue un des marqueurs les plus précoces du développement de cet organe (Zannini *et al.*, 1992).

Une analyse plus fine de l'expression de *Pax-2* et *Pax-8* montre que ces deux gènes génèrent par épissage alternatif plusieurs messagers codant des protéines différentes, ces phénomènes d'épissage pouvant être régulés de façon différentielle au cours du développement. Ainsi, le gène *Pax-8* génère au moins six isoformes. Quatre d'entre elles, proviennent d'événements d'épissage dans la région carboxy-terminale, qui code le domaine transactivateur, et elles ont des activités transcriptionnelles ainsi que des expressions différentes. Deux messagers ne sont exprimés que dans les ovaires et le placenta, un autre est exprimé dans ces organes, dans l'embryon durant tout le développement et chez l'adulte, les autres sont présent chez l'embryon et chez l'adulte. Le taux de tous ces transcrits varie au cours du développement (Kozmik *et al.*, 1993). *Pax-2* produit également plusieurs messagers par épissage alternatif. Ces épissages se produisent dans la portion carboxy-terminale de l'ARNm ainsi que dans la boîte PRD (Tavassoli *et al.*, 1994). Jusqu'à présent, il n'a pas été montré d'expression différentielle de ces messagers.

L'expression de *Pax-5* est détecté à partir de E13 dans le foie foetal quand commence la lymphopoïèse B. Il est ensuite exprimé dans les tissus lymphoïdes B et dans les cellules B à tous les stades de la différenciation. Cette expression persiste chez l'adulte dans tous les tissus lymphoïdes B en contraste avec son expression transitoire dans le SNC. Ses transcrits sont localisés dans les follicules de cellules B dans la rate et les ganglions lymphatiques. Il est également présent dans les testicules adultes (Adams *et al.*, 1992). La recombinaison homologue sur *Pax-5* provoque des altérations du cerveau moyen et du cervelet antérieur chez la souris. Il y a également blocage complet de la différenciation des cellules B à partir d'un stade très précoce. Curieusement, la délétion de *Pax-5* n'a pas d'effet semi-dominant comme les mutations dans *Pax-3*, *Pax-6* et *Pax-1*, et les souris hétérozygotes pour la délétion se développent normalement (Urbanek *et al.*, 1994).

c) Expression de Pax-3 et Pax-7

(Cf. Tableau 4)

Les transcrits de Pax-3 et Pax-7 sont détectés durant le développement précoce du cerveau et du tube neural avant le début de la différenciation neuronale. Chez la souris, à E8,5, Pax-3 est exprimé dans la zone ventriculaire dorsale du tube neural, incluant le toit du tube. Pax-



<u>Tableau 4</u> : Expression de Pax-3 et Pax-7 au cours du développement chez la souris (d'après Tremblay et Gruss, 1994)

7 est exprimé de la même façon dans le tube neural excepté dans le toit du tube et dans les crêtes neurales (Jostes *et al.*, 1991)(Cf. Fig.5). Son expression commence un peu plus tard que celle de *Pax-3* puisque cette dernière est détectée avant la fermeture du tube neural à E8,5 alors que celle de *Pax-7* n'est détecté qu'après. L'expression de ces deux gènes dans le cerveau et dans le tube neural peut être corrélée avec les zones de division cellulaire intense (Goulding *et al.*, 1991). Enfin des expériences de transplantation de notochorde montrent que l'expression de *Pax-3* et *Pax-7* est contrôlée par des signaux provenant de celle-ci (Goulding *et al.*, 1993b).

Pax-3 est exprimé également dans les cellules des crêtes neurales céphaliques et du tronc avant leurs migrations et plus tard dans des structures qui en dérivent comme les ganglions crâniens, les ganglions des racines dorsales et le mésectoderme facial (Gouding *et al.*, 1991). L'expression de Pax-3 dans tous les types de cellules dérivés des crêtes neurales tels que les précurseurs des mélanocytes ou des lignages sympatho-adrénaliens, n'a pas été clairement montrée bien que l'étude de souris mutées dans Pax-3 suggèrent une implication de ce gène dans le développement de ces tissus.

Pax-3 et Pax-7 sont exprimés dans la partie dorsale des somites, le dermomyotome, bien que Pax-3 soit exprimé avant Pax-7 dans la partie dorso-latérale des somites. On retrouve également l'expression de ces deux gènes dans les bourgeons de membres (Goulding *et al.*, 1994a). Plusieurs travaux récents montrent que Pax-3 joue un rôle crucial dans la spécification et/ou la migration des précurseurs myogéniques et donc dans le développement des muscles des membres, et il constitue un marqueur de la différenciation myogénique plus précoce que les membres de la famille MyoD (Goulding *et al.*, 1994a)(Bober *et al.*, 1994)(Williams et Ordahl, 1994)(voir la revue de Olson et Rosenthal, 1994).

Alors que l'expression de *Pax-3* chez la souris chute à E11,5, *Pax-7* continue à être exprimé jusqu'à E14,5 dans les muscles dérivés du myotome, suggérant que *Pax-7* pourrait être également impliqué dans la différenciation myogénique.

d) Expression de Pax-4

L'expression de ce gène n'a jamais été détectée.



<u>Tableau 5</u>: Expression de *Pax-6* au cours du développement chez la souris (d'après Tremblay et Gruss, 1994)

e) Expression de Pax-6/Pax-QNR

(Cf. Tableau 5)

Ce gène est très fortement conservé au cours de l'évolution. En effet, c'est le seul gène de la famille Pax pour lequel un homologue a été trouvé chez les insectes. Pour les autres membres de la famille, il n'y a que des gènes apparentés mais la conservation entre protéines est limitée au domaine PRD et/ou à l'homéodomaine. Pour Pax-6, la conservation s'étend sur la région de jonction entre les deux domaines conservés et sur la partie carboxy-terminale.

Son expression a été décrite dans de nombreuses espèces: la drosophile (Ey), le poissonzèbre (Pax(zf-a)), la caille (Pax-QNR), le poulet (Pax-6) et la souris (Pax-6). Les domaines et les stades d'expression sont conservés chez les vertébrés.

Comme la plupart des gènes Pax s'exprimant pendant la neurogenèse précoce, les domaines d'expression de Pax-6 coïncident avec des limites morphologiques du système nerveux embryonnaire en développement. C'est le premier gène de cette famille à être exprimé dans le cerveau. Pax-6 est exprimé dans des zones précises du cerveau antérieur en développement, qui vont donner le diencéphale et le télencéphale, mais pas dans les futures régions mésencéphaliques et métencéphaliques (Walther et al, 1991)(Krauss et al., 1991a). Il est exprimé dans le neuroépithélium dorsal et latéral du télencéphale et dans le thalamus ventral du diencéphale ainsi que dans la partie dorsale du sulcus diencephalus medius (Walther et al, 1991)(Cf Fig.5). Son expression s'arrête de façon nette au niveau de la commissure postérieure qui marque la limite entre le cerveau moyen et le cerveau antérieur (c'est à dire diencéphale et mésencéphale). Plusieurs travaux ont montré que les gènes En-1 et En-2 sont exprimés dans la zone d'exclusion de Pax-6, à peu près aux mêmes stades, ces domaines d'expression étant limités par plusieurs rangées de cellules n'exprimant aucun de ces deux gènes (Püschel et al., 1992). Par la suite, l'expression de *Pax-6* est détectée dans le mésencéphale caudal et dans le métencéphale. Il est également présent dans le myélencéphale. Son expression va persister dans ces structures pendant tout le développement. Une analyse plus fine au niveau du cervelet montre une expression différentielle de Pax-3 et Pax-6: dans le cortex cérébelleux, la couche de cellules granulaire exprime des taux de Pax-6 élevés, alors que la glie de Bergmann et les cellules qui entourent les cellules de Purkinje expriment Pax-3 (Stoykova et al., 1994).

Dans le tube neural, son expression commence un peu plus tard que celle de *Pax-3*, juste avant la fermeture du tube neural. Comme celle de *Pax-3* et *Pax-7*, son expression se fait dans des régions dorso-ventralement restreintes. *Pax-6* est détecté dans la zone ventriculaire de la partie basale et latérale du tube, excepté dans le plancher du tube, et chevauche les domaines d'expression de *Pax-3* et *Pax-7* (Cf. Fig.5). Très précocement, ces trois gènes sont exprimés dans des régions précises correspondant aux cellules neuroépithéliales en prolifération dans la zone ventriculaire du tube. Ces cellules vont ensuite migrer de façon radiale dans différentes couches et se différencier en neurones et cellules gliales (Walther *et al.*, 1991)(Goulding *et al.*, 1991)(Jostes *et al.*, 1991).

La première expression de Pax-6 dans l'oeil de vertébré est détectée au niveau de l'évagination latérale du cerveau antérieur. Un peu plus tard, cette évagination va former la vésicule optique en se rapprochant de l'ectoderme de surface. Pax-6 est exprimé dans la couche épithéliale de la vésicule, dans le pédoncule optique et dans l'ectoderme de surface qui va générer le cristallin. La vésicule optique va ensuite s'invaginer, formant la cupule optique, alors que le cristallin se forme à partir d'un détachement de l'ectoderme de surface. Pax-6 est fortement exprimé dans la couche interne de la cupule optique et dans le cristallin alors que la couche externe qui va donner la rétine pigmentée, montre seulement très peu d'expression de ce gène. De plus, les messagers de Pax-6 sont détectés dans l'ectoderme de surface qui va générer la cornée. Le cristallin et la cornée, une fois différenciés, continuent à exprimer Pax-6 dans sa partie postérieure et on peut détecter ce gène également dans la zone ventriculaire du pédoncule optique. En fin de différenciation, deux couches se sont développées dans la neurorétine et peuvent être clairement distinguées, une couche externe de grande densité cellulaire et une couche de faible densité cellulaire contenant essentiellement des cellules ganglionnaires. Les deux expriment Pax-6 avec la même intensité.

Son expression est également détectée dans le pancréas au cours du développement (Turque *et al.*, 1994) dès les tous premiers stades jusqu'à chez l'adulte. Cette expression est limitée au pancréas endocrine, essentiellement dans les cellules des îlots β mais il est possible qu'il soit également présent dans les cellules α puisque son expression peut être détectée dans des lignées dérivées de ces cellules.

GENE	DROSOPHILE	SOURIS	HOMME
Pax-1		Undulated (un)	
Pax-3	-	Splotch (Sp)	Syndrome de Waardenburg type I et III
			Rhabdomyosarcome : Fusion avec le gène Forkhead
Pax-6	eyeless (Ey)	Small eye (Sey)	Aniridie (AN) Anomalie de Peter
Pax-7			Rhabdomyosarcome : Fusion avec le gène Forkhead

Tableau 6 : Mutations dans les gènes Pax-1, Pax-3, Pax-6 et Pax-7

L'expression chez la drosophile reste similaire en tenant compte des différences dans le développement d'espèces aussi divergentes que les vertébrés et les insectes. Les transcrits Ey+ sont initialement détectés dans le système nerveux central, dans le cerveau et dans la corde nerveuse ventrale. Les premiers signes d'expression dans l'oeil en développement sont trouvés dans l'ébauche embryonnaire du disque imaginal de l'oeil. A partir de ce disque imaginal qui est initialement constitué d'une monocouche de cellules indifférenciées, l'oeil se développe. Il sera composé de 800 facettes ou ommatidies, chacune contenant des neurones photorécepteurs, des cellules accessoires et un cristallin. Pendant la morphogenèse, suite à une cascade d'inductions, ces cellules vont se différencier. La vague de différenciation se fait dans la direction postéro-antérieure, les cellules en position postérieure étant déterminées avant les cellules en position antérieure. Pendant les stades larvaires, la partie la plus antérieure du disque de l'oeil exprime Ey+, ce qui est compatible avec un rôle de déterminant précoce de la morphogenèse de l'oeil (Quiring *et al.*, 1994).

2- Mutations associées

Une des approches de la fonction de ces gènes passe par l'étude des nombreuses mutations qui leurs sont associés (Cf. Tableau 6). Le fait que ces mutations soient semi dominantes et donc que les animaux porteurs de la mutation sur un seul chromosome présentent un phénotype caractéristique par rapport à l'animal normal, a facilité leur isolement. Cela souligne également l'importance de ces gènes en tant que régulateurs essentiels du développement.

a) Altérations de Pax-1 dans la mutation Undulated

Les mutants undulated (un) chez la souris ont pu être associés à des mutations dans le gène Pax-1. Cette mutation spontanée est caractérisée par des malformations du squelette avec des disques vertébraux trop larges et des corps vertébraux trop réduits, ce qui génère un axe vertébral et une queue tordus (Balling *et al.*, 1988)(Timmons *et al.*, 1994). Tous ces défauts semblent provenir d'une mauvaise migration des cellules sclérotomales et de leur incapacité à former des concentrations périchordales correctes (Gruneberg, 1954). L'étude des mutants undulated montre que Pax-1 permet la différenciation de la partie ventrale des vertèbres en réponse à des signaux provenant de la notochorde (Koseki et al., 1993). Quatre mutants un ont été caractérisés: undulated, undulated extreme (un^{ex}), undulated minimal (un^m) et undulated short-tail (un^s). La mutation un est une mutation ponctuelle qui transforme un aa, la Gly 18 en Ser, cet aa étant situé dans une partie très conservée du domaine PRD (Balling et al., 1988). L'allèle un^{ex} génère un phénotype plus sévère que un, avec un taux de messagers Pax-1 très bas par rapport à la normale. Cela est du à une délétion de la partie 3' du gène incluant le dernier exon et le site de polyadénylation (Balling et al., 1993). Quant à l'allèle un^s , il est du à une délétion complète du gène Pax-1 (Balling, 1992). Les manifestations phénotypiques de cet allèle qui provoquent des malformations vertébrales à l'état hétérozygote et une mort embryonnaire ou néonatale à l'état homozygote, sont probablement dues en partie à la délétion des gènes voisins. Enfin, les bases moléculaires de la mutation un^m ne sont pas connues à l'heure actuelle.

La comparaison des propriétés de fixation à l'ADN des protéines Pax-1 normales et mutées a montré que la substitution d'un aa par un autre dans le domaine PRD, dans le cas de la mutation *un*, réduit ou supprime l'affinité de cette protéine pour des séquences cibles définies *in vitro*. Et dans des expériences de transfection transitoire, cette mutation abolit la capacité du produit de *Pax-1* à transactiver un promoteur minimal contenant ces séquences cibles (Chalepakis *et al.*, 1991). De plus, dans des expériences de fixation à l'ADN, la protéine mutée peut reconnaître des séquences cibles différentes. Ces résultats montrent donc que, dans le cas du mutant *un*, la protéine Pax-1 possède des propriétés de fixation à l'ADN différentes, ce qui l'empêche de se lier et de réguler ses gènes cibles. D'un autre coté, sur la base de ces nouvelles séquences cibles, on ne peut pas éliminer la possibilité qu'une partie du phénotype soit due à un effet du type gain de fonction avec la protéine Pax-1 mutée allant activer des gènes cibles non relevants. Cependant cette dernière hypothèse est peu probable vu la ressemblance entre les phénotypes des différents mutants.

b) Altérations du gène *Pax-3* dans les souris *Splotch* et chez l'homme dans les syndromes de Waardenburg 1 et 3

Des mutations dans le gène Pax-3 provoquent le phénotype *Splotch* chez la souris. Les animaux hétérozygotes pour *Splotch* ont des taches blanches sur le ventre et parfois sur la queue et les membres. Ces anomalies de pigmentation proviennent d'un défaut associé aux cellules dérivées des crêtes neurales : les futures cellules pigmentaires ne migrent pas dans certaines régions de la peau. Les embryons homozygotes meurent avant la naissance à E13-14 et ont des défauts de développement du tube neural caractéristiques comme l'exencéphalie et le *spina bifida*, c'est à dire pas de fermeture complète du tube neural, en plus d'anomalies de pigmentation, d'absence ou de réduction des ganglions des racines dorsales, de déficience en cellules de Schwann ou de défauts cardio-vasculaires. Enfin, ces souris homozygotes présentent des anomalies dans les tissus dérivés du mésoderme. Le mésoderme paraxial est désorganisé avec un plus faible nombre de cellules et une augmentation de l'espace intercellulaire. La musculature axiale est plus réduite en taille et les muscles des membres sont absents. Cette dernière déficience est due à l'incapacité des cellules qui expriment normalement *Pax-3* à envahir les bourgeons de membre (Goulding *et al.*, 1994a)(Williams et Ordahl, 1994).

Six allèles de Splotch ont été décrits. Deux d'entre eux sont des mutations spontanées, Splotch (Sp), l'allèle originalement décrit (Russel, 1947) et splotch-delayed (Sp^d) . Quatre autres allèles ont été produits par irradiation aux rayons X, Sp^{1H} , Sp^{2H} , Sp^{4H} et Splotch retarded (Sp^r) . La mutation Sp est due à une mutation dans un site d'épissage accepteur qui va produire un épissage anormal et la suppression d'un exon codant la fin du domaine PRD et la séquence octapeptide (Goulding et al., 1993a)(Epstein et al., 1993). La délétion dans PRD produit une forte diminution de l'affinité de ce domaine pour une séquence cible. L'octapeptide dont la fonction n'est pas connue, est localisé dans une région qui pourrait être impliquée dans des phénomènes de dimérisation (Chalepakis et al., 1994a). La mutation Sp^{2H} consiste en une délétion de 32 pb dans la partie 5' de la boîte homéo de Pax-3 qui conduit à un changement de cadre de lecture et à l'apparition d'un codon stop un aa plus loin. La protéine produite par ce mutant n'a plus d'homéodomaine fonctionnel ni de domaine carboxy-terminal, ce dernier étant nécessaire à l'activité transcriptionnelle (Epstein et al., 1991b). La séquence codante de Pax-3 est

entièrement supprimée dans les mutations Sp^{4H} et Sp^r . L'allèle Sp^r contient une amputation qui recouvre plusieurs marqueurs chromosomiques en plus du gène *Pax-3* (Epstein *et al.*, 1991a) alors que la délétion observée dans le cas de l'allèle Sp^{4H} est plus réduite. Enfin, la mutation Sp^d consiste en une mutation ponctuelle qui transforme la Gly 42 en Arg, à l'intérieur du domaine PRD, et l'on peut observer un taux de messagers de *Pax-3* cinq fois plus faible que chez la souris normale, ces découvertes corrélant le fait que cette mutation est la moins grave des mutations dans *Pax-3* (Vogan *et al.*, 1993).

Des syndromes humains ont été associés à des altérations dans le gène homologue chez l'homme à Pax-3, PAX-3, initialement caractérisé sous le nom HuP2 (Burri et al., 1989), les syndromes de Waardenburg de type 1 et 3 (Tassabehji et al., 1992)(Baldwin et al., 1992). Les syndromes de Waardenburg 1, 2 et 3 (WS1, WS2, WS3) constituent une série de désordres phénotypiquement définis et génétiquement transmis. Les manifestations cliniques majeures du WS1 consistent en une combinaison de défauts qui proviennent apparemment d'une déficience des cellules dérivées des crêtes neurales: une dystopia canthorum qui est un déplacement latéral du coin interne de l'oeil, une surdité, des déficiences de pigmentation. Le WS2 présente les mêmes caractéristiques excepté la dystopia canthorum mais aucune mutation dans PAX-3 n'a pu être associée à ce syndrome. Il serait lié à des mutations dans le gène microphtalmia, MITF, qui provoque quand il est muté, des anomalies de la pigmentation et de l'ouïe ((Tassabehji et al., 1994b). Le WS3 se manifeste par, en plus des anomalies présentes dans le WS1, des malformations des membres dues à une réduction de la taille des muscles. On peut également retrouver ces syndromes associés à des spina bifida et au désordre de Hirschsprung. Le WS3 est associé à des mutations dans PAX-3 et semble être une variation du WS1 plutôt qu'une maladie différente (Hoth et al., 1993). Aucun phénotype aussi drastique n'est observé chez l'homme en comparaison à la souris car les WS sont dus à des mutations sur un seul allèle et l'effet de dose ainsi que le fond génétique peuvent être la cause de ces différences (Tassabehji et al., 1994a).

Des WS1 présentant des phénotypes relativement similaires résultent de mutations dans le domaine PRD seul, dans l'homéodomaine seul, ou dans les deux domaines simultanément. Plusieurs mutations associées aux WS1 ont été analysées et ont permis de montrer de fortes analogies avec les mutations de type *Splotch* (Tassabehji *et al.*, 1994a). Ainsi, on peut retrouver des mutations ponctuelles qui vont résulter en un changement d'aa comme dans le cas de la mutation Sp^d ou des mutations dans les sites d'épissage qui vont provoquer des sauts d'exons comme dans la mutation Sp. Des délétions complètes du gène *PAX-3* ont été également isolées comme dans le mutant Sp^r . Certaines mutations ont été étudiées au niveau moléculaire (Chalepakis *et al.*, 1994a). Une mutation caractérisée par une délétion de 6 aa (de l'aa 62 à l'aa 67) dans une zone très conservée du domaine PRD, WS.05, va complètement supprimer la reconnaissance de séquences cibles spécifiques de Pax-3. Une autre, WS Brazil, consiste en un changement d'aa (Pro 50 en Leu) dans le domaine PRD. Ce domaine n'est plus capable de fixer l'ADN de façon indépendante bien que la protéine totale en soit toujours capable par l'intermédiaire de l'homéodomaine. Et enfin, la mutation WS.15 qui provoque également un changement d'aa (Gly 42 en Ala), fait chuter l'affinité de la protéine pour des séquences cibles de Pax-3. Toutes ces mutations sont considérées comme étant du type perte de fonction.

Récemment, une mutation de PAX-3 de type gain de fonction a été montrée être impliquée dans des rhabdomyosarcomes alvéolaires (Barr et al., 1993)(Galili et al., 1993)(Shapiro et al., 1993b). L'analyse cytogénétique des tumeurs révèle une translocation, t(2;13), et l'analyse moléculaire montre que cette translocation fusionne la partie 5' du gène PAX-3 (quasiment toute la séquence codante) à un fragment 3' du gène FKHR (Forkhead) qui code un facteur de transcription. Les transcrits générés par cette fusion contiennent tous les domaines conservés de PAX3, c'est à dire les domaines de liaison à l'ADN, et la moitié carboxy-terminale du domaine Forkhead de FKHR qui contient le domaine transactivateur. Un article récent montre que PAX-7 a été également isolé dans le même type de rhabdomyosarcomes et aussi en fusion avec FKHR (Davis et al., 1994b).

c) Altérations du gène *Pax-6* dans les souris *Small eye*, chez l'homme dans les cas d'Aniridia et chez la drosophile *Eyeless*

La mutation *Small eye* (*Sey*) (Roberts *et al.*, 1967) provoque, chez les souris homozygotes, une absence complète d'oeil et de nez, et ces animaux meurent à la naissance. Au stade hétérozygote, on observe une réduction de la taille de l'oeil. D'autres anomalies peuvent y être associées: une orbite oculaire réduite, une absence de chiasma optique, un amincissement des membranes externes de l'oeil, en particulier le cristallin et une dégénérescence cellulaire. On peut observer dans certains cas de délétions complètes du gène Pax-6, une hypoplasie de l'iris, une opacification de la cornée et une vacuolisation du cristallin. Il a également été montré chez le rat, dans le cas de la mutation r(Sey), une absence de migration des cellules dérivées des crêtes neurales provenant du cerveau moyen (Matsuo et al., 1993). Ces cellules doivent normalement coloniser la placode nasale, ce qui pourrait expliquer l'absence de développement du nez. Les cellules dérivées des crêtes sont arrêtées au niveau de l'oeil et cette accumulation anormale peut être responsable de la perturbation de l'induction de la placode du cristallin. Chez la souris Sey, la migration des pré-neurones corticaux est retardée d'une façon dépendante à la fois du stade, de la région et du dosage de Pax-6 et il y a des anomalies dans la différenciation neuronale et la croissance des axones (Schmahl et al., 1993). De plus, l'ectoderme n'est plus capable de se différencier en cristallin. Cela est du à des défauts dans les signaux provenant de la plaque neurale ou du mésenchyme sous jacent avant que la vésicule optique entre en contact avec l'ectoderme de surface (Fujiwara et al., 1994). Ces observations suggèrent que Pax-6, comme Pax-3, est impliqué dans le développement du système nerveux et en particulier, dans la migration des cellules dérivées des crêtes neurales.

Quatre allèles Sey sont connus chez la souris. L'allèle Sey est une mutation ponctuelle (sur la Glu 194) qui va introduire un codon stop entre la boîte PRD et la boîte homéo. L'allèle Sey^{neu} a été produit par mutagenèse chimique et, suite à une mutation ponctuelle dans un consensus accepteur d'épissage, présente une insertion de 116 pb d'intron. Une mutation non sens (aa 353) est introduite et résulte en une protéine tronquée en aval de l'homéodomaine. Cette dernière mutation suffit à supprimer toute activité transactivatrice du produit de *Pax-6* (Glaser *et al.*, 1994a), le domaine transactivateur du produit de ce gène se situant en aval de cette mutation dans le domaine carboxy-terminal (Czerny *et al.*, 1994). Enfin, les mutations *Small eye Harvey* (Sey^H) et Dickie's Small eye (Sey^{Dey}), la première produite par mutagenèse chimique et la seconde spontanément, correspondent à des délétions complètes du gène *Pax-6*. Les phénotypes de ces deux dernières mutations sont beaucoup plus sévères et les homozygotes meurent au cours du développement embryonnaire.

MUTATION	NOM (espèce)	POSITION	EFFET DE LA MUTATION	REFERENCES
DELETION	Sey ^H (S)		Deletion de Pax-6, wt-1 et d'autres genes	Hill et al., 1991
	Sey ^{Dey} (S)		Deletion de Pax-6	
	WAGR (H)		Deletion de Pax-6, WT-1 et d'autres genes	
	Syndrome PP(Peters')(H)		Deletion de Pax-6	Hanson et al., 1994
CHANGEMENT DE CADRE : DELETION OU		Insertion de 7pb dans exon 7, codons 123-124	Délétions des 8 derniers aa du domaine PRD et perte du cadre de lecture	Glaser et al., 1992
INSERTION	HZAMT (H)	Insertion de 2pb dans exon 10, codon 283	Produit tronqué, taille 317aa	Jordan et al., 1992
	AN1 (H)	Insertion de 2pb dans exon 5 codon 39	Produit tronqué dans domaine PRD	Davis et al., 1994a
	AN4 (H)	Insertion de 1pb dabs exon 10 codon 259	Produit tronqué dans homéodomaine	Davis et al., 1994a
	Vigma (H)	Délétions de 151pb dans exon 12, codons345-395	Produit tronqué dans la moitié du domaine carboxy-terminal	Hanson et al., 1993
Deletion Conservant Le cadre	NIKIT (H)	Délétion de 118pb dans exon6 codons84-119	Délétion de 36aa dans domaine PRD	Hanson et al., 1993
MUTATION SUR LES SITES	Sey ^{neu} (S)	SA, intron? Stop sur aa 353	Produit tronqué après l'homéodomaine	Hill et al., 1991
D'EPISSAGE	<i>rSey</i> (S)	Nouveau SD créé	Produit tronqué au début du domaine carboxy-terminal	Matsuo et al., 1993
	AN6 (H)	SD, intron 6 GT-► AT	Produit tronqué dans domaine PRD	Davis et al., 1994a
	RUBAI (H)	SA, intron 8 GT ➔ TT	Produit tronqué dans domaine FRD	Jordan et al., 1992
		SD, intron 3 AG 🔶 GG	Produit tronqué dans domaine PRD	Glaser et al., 1992
MUTATIONS FAUX SENS	JECK (H)	Exon 8 codon 208	Arg ✦► Trp entre domaine homéo et PRD	Hanson et al., 1993
	Famille 3 Syndrome de Peter (H)	Exon 5 codon 26	Arg + Gly aa 23 dans domaine PRD	Hanson et al., 1994
MUTATIONS NON SENS	Sey (S)	Exon 8 Codon 194	Produit tronqué avant l'homéodomaine	Hill et al., 1991
	AN2 (H)	Exon 6 codon 116	Produit tronqué dans homéodomaine	Davis et al., 1994a
		Exon 9 codon 240	Draduit transué dans	Glaser et al., 1992
	AN3 (H)	ldem	homéodomaine	Davis et al., 1994a
	CACHE (H)	ldem		Hanson et al., 1993
	AN5 (H)	Exon 11 codon 317	Produit tronqué après homéodomaine	Davis et al., 1994a
SA: site accepter	ur d'épissage; SD: si	te donneur d'épissage:	S: souris; H : homme	

Tableau 7 : Mutations dans Pax-6

 $\frac{1}{2}$

Des mutations dans le gène PAX-6 ont été identifiées chez plusieurs malades atteints d'aniridie (AN), une malformation des yeux qui peut aller d'une simple hypoplasie de l'iris jusqu'à une absence complète de ce dernier. D'autres anomalies oculaires peuvent s'y ajouter, cataracte, dislocation du cristallin, hypoplasie du nerf optique, nystagmus (oscillation rythmique du globe oculaire) et glaucome (augmentation de la pression oculaire). Un homozygote a été décrit qui présente des anomalies craniofaciales très sévères et, en particulier, une absence complète d'yeux et de bulbe olfactif ainsi que de nombreux défauts au niveau du développement du cerveau (Glaser et al., 1994a). Des mutations dans PAX-6 ont été également identifiées dans des cas d'anomalie de Peter. C'est un défaut congénital de la chambre antérieure de l'oeil, généralement une opacité du centre de la cornée générée par des adhésions entre le cristallin, la cornée et/ou l'iris. L'analyse plus fine des yeux des souris Sey montre qu'ils présentent des malformations du même type que celles observées dans l'anomalie de Peter (Hanson et al., 1994). Plusieurs types de mutation ont été trouvés dans de nombreux cas d'aniridie/anomalie de Peter, incluant des délétions conservant le cadre de lecture, des insertions ou des délétions supprimant le cadre de lecture, des mutations dans les sites d'épissage, des mutations non sens et plus rarement faux sens (Martha et al., 1994 a et b)(Cf. Tableau 7). De grandes délétions qui recouvrent PAX-6 et les gènes qui l'encadrent comme WT1, le gène impliqué dans la tumeur de Wilms, causent le syndrome WAGR (tumeur de Wilms, aniridie, anomalies urino-génitales et retard mental).

Comme chez les autres mutants PAX connus, la mutation de *PAX-6* est semi dominante. L'inactivation des deux allèles est létale, avec des anomalies craniofaciales très graves. La fonction normale de *PAX-6* apparaît être fortement dépendante du dosage de ses produits. La plupart des mutations génèrent des protéines tronquées et causent des perte de fonction. La sévérité du phénotype est également fonction du fond génétique de l'individu. Un cas d'aniridie a permis de corréler le phénotype observé avec l'activité en tant que facteur de transcription du produit de *PAX-6*. Le patient est porteur d'une mutation non sens dans la région qui code le domaine carboxy-terminal de la protéine PAX-6 qui constitue le domaine transactivateur, et les anomalies observées sont relativement réduites avec une cornée un peu petite et des problèmes de cataractes. Cette mutation provoque la perte d'un tiers de ce domaine et la protéine qui est générée conserve une activité transcriptionnelle partielle (Glaser *et al.*, 1994a).

Homologie entre eveless (949 aa) et Pax-ONR (416 aa):

Au niveau du domaine PRD :

	70	80	90	100	110	120	130
GHSGV	NOLGGVPVC	GRPLPDSTROP	UVELAHSGA	RPCDISRILO	VSNGCVSKILG	RYYETGSIR	PRAIGGSKPRVA
-===	*******	**********	********	*********	*********		**********
SHSGV	NOLCOVEVN	GRPLPDSTROP	IVELAHSGA	RPCDISRILO	VSNGCVSKILG	RYYETGSIR	RAIGGSKPRVA
	130140	150	160	170	180	190	200
140	150	160	170	180	190		
TAEVVSKI SOYKRECPSI PAWEIRDRLLOPNVCTNDNI PSYSSINRVLENLA							
TPEVVSKIAOYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLA							
21	0 2	20 23	0 2	40 2	50		

Au niveau de l'homéodomaine :

 420
 430
 440
 450
 460
 470
 480
 GEGENSNCGASNIGNTEDDQARLILERKLORNETSPTNDOIDSLEKEPERTHYPDVFARERLAGKIGLPPARIOVW

 GGGENTNS1SSNCEDSDEQARKLOLERKLORNETSPTNDOIDSLEKEPERTHYPDVFARERLAGKIGLPPARIOVW
 300
 320
 330
 340
 350
 360
 370
 380

 GGGENTNSISSNGEDSDEAQ

 310
 320

 490
 500

 <u>PSNRRAKWRR</u>EEKLRNQRR

 <u>PSNRRAKWRR</u>EEKLRNQRR

 390
 400

Homologie entre eyeless et Pax-ONR au niveau nucléotide : au niveau du domaine PRD :

20	0 2	10	220	230	240 2	250 260	
CACAAGGGT	CACAGTOGA	GTAAATCAG	TIGGTTGGC	TTTTTCTTCG	AGGAAGGCCT	TIGCCAGATTCAA	
** ** **		11 11 111			** *****	**** ** ** *	
CAGAACAGT	CACAGCGGA	GTGAACCAG	TCGGCGGGG	TGTTCGTCAA	COGGAGGCCG	TGCCCGACTCCA	
3	90	400	410	420	430	440	
27	0 2	80 3	290	300 :	310 :	320 330	
CACGGCAAA	AAATTGTCG	AACTGGCAC	ATTCTGGAG	TCGGCCATGI	ATATTTCTC	GAATTCTGCAAGT	
1 11111 1	: :: :: :	**** ** *					
CGCGGCAGA	AGATCGTGG	AACTCGCGC	ACAGCGGCG	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	JACATCTCCCC	GAATCCTGCAGGT	
4	60	470	480	490	500	510	
34	03	50 3	360	370	380	390	
ATCAAATGG	ATGTGTGAG	CAAAATTCT	CGGGAGGTA	TATGAAACAG	GAAGCAT-AC	JACCACGTGCTAT	

GICGAAIGG	ATGTGTGAG	TAAAATTIIN	GGCAGGTA	TACGAAACTC	CTCCATCAG	CCCAGG-GCGAT	
5	30	540	550	560	570	580	
4	10	420	430	440	450	460	
CGGAGGATC	CAAGCCACG	TGTGGCCAC	AGCCGAAGT	GTTAGCAAAA	TICGCAGTAC	AAACGCGAGTGT	
			******			***** *****	
CGGAGGTAG	TAAGCCGAG	AGTAGCGAC	ICCCGAAGT	CTAAGCAAAA?	TAGCGCAGTA	TAAACGAGAGTGC	
	600	610	620	630	640	650	
4	80	490	500	510	520	530	
CCTAGCATA	TTTGCTTGG	GAAATTCGG	ATAGATTAC	T-TCAGGAGA/	CGTTTGTACT	CAACGATAATATA	
	,,,,,,,,,,,	** *****			** *****	********	
CCCTCCATC	TITGEGIGG	GAGATTCGA	JACAGATTAC	TCTC-GGAGG	GGTCTGTAC	AACGATAACATA	
	670	680	690	700	710	720	
	550	560	570	580	590		
CCAAGTGTGTCCTCAATAAACCGTGTATTGAGAAACTTGGCT-GCGCAAAAGGAGCAG							
		*** * **	: : : : :				
CCCAGTGTGTCGATAAACAGAGTCCTCCGCAACCTGGCTAGCG-AAAAGCAACAG							
730	740	750	760	770	780		

au niveau de l'homéodomaine :

 1240
 1250
 1260
 1270
 1280
 1290
 1300

 GGTGAAAGGTGAAAACTCCAATGGTGAC-GGTTCAAATATAGGAA-ACACTGAGGATGATCAAGCTGGG-GG
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1

СПОВЕДСИЛИСТ С СОЛИТОВАТ С СОЛИСТИ С СОЛИСТИИ СОЛИСТИИ ПОВЕДСИ, СОЛИТИИ С СОЛИССАЛОВИТИ С СОЛИССАЛОВИТИ С СТОЛАВССИЛОСТИ СТОЛАВССИЛОСТИ С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ. С СОЛИСТИИ С СОЛИСТИИ. С

TCAGCGGAGA 1200

Fig.6 : Homologie entre Eyeless et Pax-QNR Les domaines Homéo et PRD sont soulignés

Chez la drosophile, la mutation *eyeless* est caractérisée par une réduction plus ou moins importante de la taille de l'oeil, qui dépend de la pénétrance de l'allèle muté. Environ vingt allèles Ey ont été isolés généralement produits par des mutations spontanées dues à l'insertion d'éléments transposables (Quiring *et al.*, 1994). Il n'y a pas encore de résultat d'analyse moléculaire de ces mutations. Cependant, l'expression très précoce de Ey+ dans les disques imaginaux de l'oeil suggère un rôle important de ce gène dès les tous premiers stades du développement de l'oeil. Mais d'autres expériences sont encore nécessaires telles que l'étude du comportement de marqueurs déjà connus, spécifiques de l'oeil, qui permettra la détermination des étapes de la différenciation où *Pax-6* intervient.

3 - Fonctions des gènes Pax

Les structures et les processus de développement communs entre espèces sont supposés être contrôlés par des gènes fortement conservés au cours de l'évolution. La boîte PRD est conservée dans un grand nombre d'organismes distincts tels que les nématodes, la drosophile, la souris et l'homme (Noll, 1993). Dans au moins un cas, la conservation s'étend sur toute la séquence codante. Ainsi la conservation entre *Pax-6* de souris et son homologue chez le poissonzèbre, Pax(zf-a), est de l'ordre de 97% et ils possèdent les mêmes domaines d'expression. Entre les produits des gènes *Pax-QNR* et *Ey* de drosophile, il y a 95% d'homologie au niveau du domaine PRD et 90% au niveau de l'homéodomaine, et une conservation de l'ordre de 25% dans la région de jonction entre le domaine PRD et l'homéodomaine ainsi que dans le domaine carboxy-terminal (Cf Fig.6). En comparaison, la conservation entre gènes Hox de différentes espèces apparaît plus limitée puisqu'elle ne dépasse pas 70%.

L'analyse de l'expression des différents gènes Pax suggère de nombreux rôles dans le développement embryonnaire, en particulier, dans le développement du SNC. Mais, la compréhension de la fonction de ces gènes est basée surtout sur l'étude des mutants spontanés de certains d'entre eux ainsi que sur des expériences de transgénèse. De plus, il a été montré pour certains de ces gènes que cette fonction pouvait être conservée même entre des espèces aussi divergentes que l'homme et la drosophile.



 \swarrow hélice α

.

Fig. 7 : Famille des gènes Pax chez la souris

a) Analyse de l'expression

Les gènes de la famille Pax sont exprimés en deux phases : une première phase d'expression très précoce dans des structures en développement qui suggère un rôle dans la mise en place de tissus, c'est à dire dans la multiplication et la différenciation cellulaire et une deuxième phase dans les tissus différenciés, ce qui impliquerait un rôle dans la maintenance de l'état différencié. L'étude de ces expressions et les anomalies graves qui résultent de mutations ou suppression d'expression dans ces gènes, montrent leur importance dans le contrôle de la prolifération et de la différenciation.

De plus, ces gènes s'expriment d'une part dans le système nerveux central, d'autre part dans différents organes ou tissus. Les gènes cibles sont différents. Les membres de la famille Pax sont donc capables de contrôler un éventail de gènes très large et constituent un groupe de gènes clé du développement.

b) Analyse moléculaire

Les produits des gènes Pax sont des facteurs de transcription. Ils possèdent un à deux domaines capables de se lier à l'ADN (Cf Fig.7). Excepté Pax-9 et Pax-4 pour lesquelles il n'y a aucune donnée, toutes les protéines Pax sont capables de se fixer sur des séquences spécifiques d'ADN. Ils sont également capables d'activer la transcription à partir de ces séquences spécifiques. Le domaine transactivateur qui est localisé en carboxy-terminal de ces protéines, est riche en Ser, Thr, et Pro. Plusieurs de ces produits ont une localisation nucléaire comme Pax-2 (Dressler *et al.*, 1992), Pax-5 (Adams *et al.*, 1992) et Pax-6 (Carrière *et al.*, 1994).

La protéine paired de drosophile reconnaît une séquence, au moins *in vitro*, qui est la séquence e5 présente dans le promoteur du gène *even-skipped*. Cependant, la régulation de ce dernier gène par paired est peu probable car des mutations perte de fonction ou gain de fonction dans le gène *paired* ne modifient pas la régulation de *even-skipped*. Cette séquence, cependant, est reconnue à la fois par le domaine PRD et par l'homéodomaine de *paired*. L'homéodomaine se fixe sur une séquence en amont du type ATTA/TAAT et le domaine paired se fixe sur un



Fig. 8: Structure bipartite du domaine PRD

(d'après Epstein., 1994)

consensus en aval du type GTTCC/GGAAC (Treisman *et al.*, 1989)(Treisman *et al.*, 1991). les produits de *Pax-1*, *Pax-3* et *Pax-6* reconnaissent également la séquence e5 par l'intermédiaire de leurs domaines de liaison à l'ADN. Pax-1 ne fixe l'ADN qu'avec le domaine PRD puisqu'il est dépourvu d'homéodomaine. Cependant, des nucléotides additionnels sont nécessaires à l'interaction avec l'ADN. Comme Pax-1 (Chalepakis *et al.*, 1991), Pax-2 qui possède en plus la première hélice de l'homéodomaine, se lie à un motif GTTCC. La fixation est médiée seulement par le domaine PRD mais, pour l'activation de la transcription, le domaine carboxy-terminal est nécessaire (Fickenscher *et al.*, 1993). De plus, les séquences reconnues avec une haute affinité par ces deux protéines sont assez divergentes de la séquence e5.

Dans le cas de Pax-3 (Goulding *et al.*, 1993) et Pax-6 (Dozier *et al.*, 1993), la fixation simultanée du domaine paired et de l'homéodomaine sur une séquence cible contenant les consensus ATTA et GTTCC est nécessaire pour l'obtention d' un complexe ADN-protéine stable et la suppression d'un des deux consensus suffit à faire chuter l'affinité de la protéine pour sa cible .

Une analyse récente du domaine PRD de Pax-5 montre que ce domaine est en fait composé de deux sous-domaines. En effet, le domaine PRD tronqué des 36 aa carboxyterminaux est capable de reconnaître un certain nombre de séquences nucléiques. L'alignement des séquences capables de fixer le produit de Pax-5 permet de déduire une séquence consensus contenant deux demi sites. Les deux sous-domaines se fixent sur les demi sites qui sont localisés dans les grands sillons adjacents du même coté de l'hélice d'ADN. Ces deux demi sites contribuent à l'affinité totale d'un site de reconnaissance pour Pax-5 et plus leurs séquences sont proches du consensus établi, plus l'affinité est élevée. Des expériences d'échange entre les sousdomaines du domaine PRD de Pax-5 et Pax-1, montrent que le produit de ce dernier gène possède également un domaine PRD bipartite (Czerny et al., 1993). De même, l'insertion d'un exon supplémentaire entre la première et la deuxième hélice du domaine PRD de Pax-6 permet de modifier les séquences cibles de ce domaine (Epstein et al., 1994a). La fixation du domaine PRD à l'ADN se ferait de façon prédominante par la partie amino-terminale de ce domaine, c'est à dire la première hélice, et l'insertion de l'exon permettrait de perturber cette structure et de démasquer le potentiel de fixation de la partie carboxy-terminale (Cf Fig. 8). Une structure bipartite serait donc commune à tous les domaines PRD, la spécificité de fixation de ce domaine étant déterminée par les deux sous-domaines amino- et carboxy-terminaux à la fois. Cette structure bipartite permettrait une grande flexibilité dans la reconnaissance de séquences cibles.

Des gènes cibles ont pu être isolés pour plusieurs membres de la famille. Le produit de *Pax-5* ou *BSAP* est exprimé dans les cellules B au cours de leur différenciation et au stade mature. Il régule le gène *CD19* qui code une protéine de surface spécifique des cellules B (Barberis *et al.*, 1990)(Kozmik *et al.*, 1992). *Pax-5* possède, par ailleurs, un homologue chez l'oursin de mer, *TSAP*, qui est capable de contrôler les gènes des histones H2A-2 et H2B-2 (Barberis et al, 1989).

Pax-8 est exprimé dans la thyroïde à tous les stades du développement et chez l'adulte. Il est également exprimé dans des lignées de cellules thyroïdiennes. Le produit de ce gène se fixe sur les promoteurs des gènes de la thyroglobuline et de la thyroperoxydase, qui sont exclusivement exprimés dans la thyroïde. Sur ces deux promoteurs, le site de fixation de Pax-8 recouvre celui de TTF-1 qui est une protéine à homéodomaine également impliquée dans l'activation de ces promoteurs (Civitareale *et al.*, 1989)(Francis-Lang *et al.*, 1992). Pax-8 peut activer la transcription de ces promoteurs, ce qui suggère que ce gène est impliqué dans la maintenance du phénotype différencié de la thyroïde (Zannini *et al.*, 1992) et, *in vivo*, son expression est corrélée avec celle des messagers de la thyroperoxydase, ce qui constitue un argument supplémentaire quant à sa fonction de régulateur de ce gène (Fabbro *et al.*, 1994). Pax-8 est également capable de fixer et d'activer le promoteur de la molécule d'adhésion cellulaire neurale ou N-CAM (Neural Cell Adhesion Molecule) (Holst *et al.*, 1994).

Le produit de *Pax-6/Pax-QNR* exprimé dans le SNC et dans l'oeil, est capable de se fixer et d'activer son propre promoteur (Plaza *et al.*, 1993). Il peut également interagir avec le promoteur du gène de α A-cristalline, protéine de structure du cristallin, et d'activer la transcription de ce gène (Cvekl *et al.*, 1994). Il a été récemment montré être capable de se fixer sur le promoteur du gène de la N-CAM L1 dont certains domaines d'expression coïncident avec ceux de *Pax-6* (Chalepakis *et al.*, 1994c). *Pax-6* est un gène essentiel pour le développement des yeux, et ce, dès les tous premiers stades. Qu'il puisse réguler des gènes impliqués dans les interactions cellulaires comme les N-CAM et dans la différenciation des cellules du cristallin comme les cristallines est cohérent avec sa fonction.
Plusieurs gènes Pax génèrent de nombreuses isoformes. L'existence de ces isoformes suggère que la régulation de la fonction de ces facteurs passe aussi par le contrôle de l'épissage alternatif. En effet, il a été montré pour Pax-8 que ses isoformes étaient exprimées de façon différentielle au cours du développement. Elles ont des propriétés transactivatrices différentes. les épissages alternatifs produisant des modifications dans le domaine carboxy-terminal ou domaine transactivateur (Kozmik et al., 1993). Il se produit le même type d'épissage alternatif pour Pax-2, c'est à dire dans la même région, Pax-2 et Pax-8 étant très proches au niveau séquence peptidique, mais jusqu'à présent aucun différentiel d'expression ou d'activité n'a été montré pour ce gène. Pax-6 génère plusieurs isoformes chez la caille et la souris, les épissages alternatifs se produisant essentiellement au niveau d'un des domaines de fixation à l'ADN, le domaine PRD. Un des épissages génère une protéine, contenant un exon de 14 aa supplémentaire dans le domaine PRD, qui est capable de reconnaître une séquence cible différente de celle reconnue par la protéine possédant un domaine PRD classique (Epstein et al., 1994a). Ces résultats suggèrent que les isoformes codées par Pax-6 sont potentiellement capables de réguler des gènes cibles différents. Il a déjà été montré pour de nombreux facteurs de transcription que leur fonction pouvait être régulée par des événements d'épissage, en particulier chez plusieurs gènes à boîte homéo. Ainsi, le gène POU, Pit-1, génère plusieurs isoformes qui sont fonctionnellement différentes avec, entre autre, des épissages dans le domaine POU spécifique qui modifient les propriétés de reconnaissance de l'ADN de ce facteur (Voss et al., 1993)(Theill et al., 1992).

Plusieurs résultats montrent que les produits des gènes Pax sont capables de se dimériser au moins *in vitro*. Les homéodomaines de type PRD sont capables de se fixer de façon coopérative en dimère sur des séquences cibles palindromiques de type ATTA/TAAT (Wilson *et al.*, 1993). Et sur ces mêmes séquences palindromiques, les protéines Pax-3 et Pax-7 peuvent se fixer et former des homo et hétérodimères. La dimérisation peut être également médiée par des séquences de fixation du domaine PRD. Ainsi, Pax-3, quand il est dépourvu de l'homéodomaine, peut se fixer en dimère (Chalepakis *et al.*, 1994). De même, le domaine PRD de Pax-6 est aussi capable de se fixer en dimère quand l'exon supplémentaire de 14 aa est présent (Epstein *et al.*, 1994). Cependant, aucun de ces résultats n'a pu être corrélé avec des observations *in vivo*.

c) Potentiel oncogènique

Les facteurs de transcription en général sont capables de transformer les cellules quand ils sont surexprimés. Tous les phénomènes observés sur les gènes à boîte homéo suggèrent que ces derniers régulent de nombreux processus de la différenciation mais également que leur expression inappropriée peut être oncogénique.

Plusieurs gènes Hox sont oncogèniques quand ils sont surexprimés dans des lignées immortalisées, ces dernières pouvant ensuite générer des tumeurs dans des souris nude (Maulbecker et Gruss, 1993a). Un nombre croissant de gènes à boîte homéo se trouve impliqué dans différents types de leucémies suite à des événement de translocation. Ainsi, chez l'homme, la translocation t(1;19) sur le gène *PBX* génère une protéine de fusion; cette dernière, produite *in vitro*, est capable également de transformer des cellules (Kamps *et al.*, 1991). De même, les gènes des protéines à domaine LIM, RBTN1 et RBTN2 sont également impliquées dans des translocations générant des leucémies aiguës de type T (Voir la revue de Rabbitts, 1991). D'autres protéines à homéodomaine, par coexpression avec des facteurs de croissance, peuvent transformer des cellules *in vitro* et induire des leucémies *in vivo* (Perkins *et al.*, 1990).

La surexpression des gènes de la famille Pax conduit également à la transformation de cellules immortalisées et ces cellules transformées sont capables de développer des tumeurs dans des souris nude (Maulbecker et Gruss, 1993b). De plus, les gènes *PAX-3* et *PAX-7* sont impliqués dans le développement de rhabdomyosarcomes alvéolaires suite à des translocations chromosomiques qui génèrent des protéines de fusion avec le gène *FORKHEAD* (Bar *et al.*, 1993)(Gallili *et al.*, 1993)(Shapiro *et al.*, 1993)(Davis *et al.*, 1994).

Le fait que les gènes Pax soient des oncogènes potentiels montre l'importance de ces gènes dans le contrôle de la prolifération cellulaire.

53



Fig. 9 : Possibilités de transdifférenciation de la neurorétine

V - La neurorétine

1 - Modèle biologique

L'oeil des vertébrés au cours du développement embryonnaire constitue un modèle biologique très intéressant pour l'étude des bases moléculaires de la différenciation. Plusieurs des tissus qui le constituent, sont capables de se transdifférencier à la fois *in vivo* et *in vitro*. Les cellules de la neurorétine, en particulier, sont capables de se transdifférencier en cellules du cristallin ou en cellules pigmentées (Okada *et al.*, 1976)(Cf Fig.9). 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) ou sous l'effet d'oncogène comme *v-myc* (Martin *et al.*, 1992).

Lorsque l'on met en culture des cellules de neurorétine, disséquée au stade de prolifération, elles cessent rapidement de se diviser et expriment plusieurs marqueurs de la différenciation neuronale (Crisanti-Combes *et al.*, 1977). La différenciation cellulaire est due à une expression différentielle de gènes et les facteurs de transcription codés par de nombreux oncogènes peuvent réguler une telle expression.

La neurorétine en tant que modèle expérimental a donc été développée dans le laboratoire afin d'étudier les effets des oncogènes sur le contrôle de la prolifération et de la différenciation cellulaire. Ce système cellulaire permet de révéler l'effet d'oncogènes jusqu'alors capable de transformer seulement un nombre très restreint de types cellulaires comme l'oncogène *myb* jusqu'alors limité au système hématopoïétique (Garrido *et al.*, 1992a et b).

2 - Développement de l'oeil et de la neurorétine

Le développement de l'oeil implique des interactions entre plusieurs tissus: le neuroectoderme du cerveau antérieur, l'ectoderme de surface et le mésenchyme dérivé des crêtes neurales. Les cellules dérivées du neuroectoderme vont former la rétine, le nerf optique et la partie épithéliale de l'iris. L'ectoderme de surface donnera le cristallin et l'épithélium de la



Fig 10 : Organisation de la neurorétine

- R : batonnets
 - H : cellules horizontales
- C : cones
- I : interneurones
- A : cellules amacrines
- G : cellules ganglionnaires
- M : cellule de Müller B : cellules bipolaires

cornée et les cellules dérivées des crêtes neurales, la sclérotique, le stroma de la cornée, la chambre antérieure et l'humeur vitrée.

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 (à E21). Le développement de la rétine suit un cours très similaire chez tous les vertébrés avec des différences temporelles liées à la durée totale du développement.

La neurorétine différenciée est constituée de plusieurs types cellulaires: les neurones et interneurones (photorécepteurs, cellules horizontales, bipolaires, amacrines et ganglionnaires) et de cellules gliales, les cellules de Müller (Cf Fig.10).

3 - Gènes à boîte homéo exprimés dans l'oeil

En plus des gènes Pax, d'autres groupes de gènes à boîte homéo semblent contribuer au développement de la partie antérieure du cerveau. Leurs expressions caractéristiques dans les régions antérieures de l'embryon les distingue des gènes des complexes Hox dont la limite antérieure d'expression se situe au niveau de la limite cerveau postérieur-cerveau moyen. Ces gènes sont importants pour le développement de l'oeil puisque celui-ci provient d'une évagination du diencéphale, subdivision du cerveau antérieur. De nombreux gènes à boîte homéo de drosophile, exprimés dans la partie rostrale, possèdent leurs homologues chez les vertébrés, ces derniers s'exprimant également dans le cerveau antérieur: *Dlx* (homologue de *distal-less*), *Emx* (homologue de *empty-spiracle*) et *Otx* (homologue *ortho-denticle*). On retrouve également l'expression de plusieurs gènes de la classe POU, tels que *Pit-1*, *Oct-1*, *Oct-2* et les gènes de type Brn. Plusieurs de ces gènes seront par la suite exprimés dans l'oeil (Cf Tableau 10). Leurs domaines d'expression correspondent aux limites entre différentes régions de l'oeil embryonnaire qui vont développer des structures distinctes. Le stade d'expression de ces gènes précède généralement la différenciation cellulaire et ils semblent jouer un rôle important dans la mise en place des différentes structures de l'oeil. Ainsi, les gènes POU, *Brn 3.0* et *Brn 3.2*, ne

Species

Clox Cut	Retina and not lens at E15*	Mouse
DIx <i>Distal-less</i>	Prospective retina; expression defines boundary between retina and inner layer of ciliary epithelium(mouse and frog). Retina of E3-4 chicken embryo.*	Mouse Chicken Xenopus
Emx-1 Empty-spiracles	Only in the lens at E17	Mouse
G6 (H Box8)	Adult retina*+ Anterior margin of optic cup, E3-4 embryo*	Goldfish Chicken
G10 (ceh-10)	Adult retina*+	Goldfish
HoxA-1, A-5, A-10, B-2, B-3, B-5, B-6, DD-10 <i>Antennapedia</i>	Adult retina*+	Goldfish
HoxD-3	Anterior margin of optic cup, E3-4 embryo.	Chicken
Msx-1 (Hox7) <i>Muscle-segment</i>	Periocular (neural crest) mesenchyme at E9.5. Mesenchyme between corneal epithelium and lens at E11.5. At E12.5 prospective inner layer of ciliary epithelium, persisting in this location until at least E19.5. Periocular mesenchyme in stage18 chicken embryo. Retina in adult goldfish.*+	Mouse Chicken Goldfish
Msx-2 (Hox8)	Prospective corneal epithelium, lens and inner layer of optic cup (future retina, ciliary epithelium, iris) at E9.5. Corneal epithelium and prospective retina (absent from margin of optic cup) at E11.5. Corneal epithelium only at E12.5. Not detected in eye at E13.5.	Mouse
Otx-1	Entire optic cup at E10; margin of optic cup at E12 (prospective iris and ciliary epithelium). Expression identifies the boundaries between these tissues and the retina and RPE. Ciliary epithelium only at E17 \bullet	Mouse
Otx-2	Entire optic cup at E10. In the RPE and sclera at E12. At E17 expression is limited to the REP and cells in the outer layer of the neural retina (photoreceptors?).	Mouse
Pax-6 Paired	Entire optic vesicle, ectoderm that will form the lens and the corneal epithelium.	Mouse Zebrafish Chicken Quail
Pax-2	Ventral half of the optic cup at E9-11. Optic stalk at 12h.	Mouse Zebrafish
Brn-3.0 (Brn3a)	Ganglion cell layer at E15.5. Adult retina.	Mouse
Brn-3.2 (Brn3b)	Ganglion cell layer at E15.5. Adult retina.	Mouse
SOHo-1	E2 and E4 retina. At E4, expression is high in the anterior and low in the posterior retina.	Chicken
XANF-1	Anterior neural folds, forebrain. Margin of optic cup of E3-4 chicken embryo*	Xenopus Chicken
XI.POU-1	Eye (optic vesicle?).	Xenopus
Chx10 Aristaless	Prospective neuroretina at E9.5. At E16.5, expression is limited to the inner nuclear layer (amacrine and interplexiform cells)	Rabbit Chicken Mouse

* The indicated tissue expresses this homeobox-containing gene. Due to the method used, it is not known whether expression is limited to this tissue of whether the gene is also expressed in other ocular tissues.

+ Goldfish eyes grow throughout life, continually adding new retinal cells. Therefore, homeobox genes expressed in the adult retina might be involved in retinal development, in regulating gene expression in diferentiated retinal cells, or both.

• The autors indicated that expression was limited to the iris at E17. However, the iris does not appear to have formed in mouse embryo at this stage.

<u>Tableau 8</u> : Gènes à boite homéo exprimés dans les tissus oculaires des Vertébrés (d'après Beebe, 1994)

sont exprimés qu'après la mise en place de la couche de cellules post mitotiques, et cette expression est limitée à la couche de cellules la plus interne des cellules ganglionnaires. Par la suite, l'expression de ces deux gènes reste confinée à la couche post mitotique, et cette expression continue après la naissance (Gerrero *et al.*, 1993)(Turner *et al.*, 1994). Cette expression suggère que ces gènes pourraient être impliqués dans la différenciation des cellules ganglionnaires mais également dans la maintenance de ces structures.

Deux gènes Pax sont exprimés dans la rétine: Pax-2 et Pax-6. Ce dernier est exprimé dès les tous premiers stades du développement de l'oeil (Martin *et al.*, 1992)(Walther *et al.*, 1991). L'importance de Pax-6 dans la formation des yeux suggère que d'autres gènes à boîte homéo pourraient être essentiels au développement de l'oeil. La caractérisation de ces facteurs et l'analyse de leurs interactions, fonctions et régulations permettront de mieux comprendre la mise en place de l'oeil.

RESULTATS

Les figures indiquées en gras correspondent aux figures des articles.

ÉTAT DES TRAVAUX

Après infection par le rétrovirus aviaire MC29 (portant l'oncogène *v-myc*), les cellules de neurorétine de caille (QNR), normalement quiescentes dans nos conditions de culture, se transforment morphologiquement et prolifèrent très rapidement. Après le 10ème passage, la culture se pigmente et les cellules mélanisées deviennent majoritaires. Ces cellules infectées par MC29 constituent un modèle biologique intéressant pour étudier les mécanismes moléculaires responsables de la prolifération et de la différentiation ainsi que le rôle de l'oncogène *myc* dans ces événements. Dans le but d'isoler et de caractériser les gènes potentiellement contrôlés par *myc* dans les cellules pigmentées infectées par MC29 (PMC29QNR), une banque d'ADNc a été construite à partir des messagers poly-adénylés codés par ces cellules et un criblage différentiel contre des ADNc de cellules embryonnaires de caille infectées par le même virus (MC29QEC) a été réalisé. Cette stratégie devait permettre d'isoler des gènes spécifiquement induits par l'oncogène *myc* dans les QNR et également des gènes spécifiquement exprimés dans la neurorétine et non réprimés par cet oncogène. Parmi les clones isolés, un clone, MC29QNR2, a retenu notre attention et a été analysé en détail.

ARTICLE I

"Characterization of a paired box- and homeobox-containing quail gene (*Pax-QNR*) expressed in the neuroretina."

Patrick Martin, Catherine Carrière, Christine Dozier, Brigitte Quatannens, Marie-Anne Mirabel, Bernard Vandenbunder, Dominique Stehelin & Simon Saule.

Oncogene, 7, 1721-1728, 1992

Caractérisation d'un gène de caille (*Pax-QNR*) contenant une boîte homéo et une boîte paired exprimé dans la neurorétine .

Le clone MC29QNR2 a été séquencé et la comparaison de cette séquence avec les banques de données a montré qu'il codait deux domaines très conservés: un homéodomaine de type PRD et un domaine PRD. Le gène correspondant appartient à la famille des gènes Pax et nous l'avons appelé *Pax-QNR*. Par la suite, ce gène s'est avéré être l'homologue aviaire du gène de poisson-zèbre *Pax(Zf-a)* (94,8% d'homologie en aa et 80,4% en nucléotides sur la séquence codante totale)(Krauss *et al.*, 1991), et du gène murin *Pax-6* (97,9% d'homologie en aa et 86% en nucléotides)(Walther *et al.*, 1991). Le domaine PRD est en position amino-terminale dans la protéine et il est à 100% homologue avec celui de Pax-6. La conservation au niveau de l'homéodomaine est plus importante car elle est de 100% entre différentes espèces. La présence d'une Ser en position 50, à l'intérieur de la troisième hélice qui est considérée comme l'hélice de reconnaissance, est caractéristique des homéodomaines de type PRD. Enfin, le domaine carboxy-terminal riche en Ser, Thr et Pro est du même type que celui de Pax-3 (Goulding *et al.*, 1991).

Le messager correspondant à cet ADNc est exprimé dans le tube neural et dans la neurorétine. Chez la caille, nous avons montré que *Pax-QNR/Pax-6* était exprimé essentiellement dans la neurorétine et dans le cervelet (**Fig. 3**). La neurorétine étant un modèle biologique développé dans le laboratoire, nous nous sommes surtout intéressés à son expression dans ce tissu. L'étude, par la technique de Northern blot, de l'expression des messagers codés par *Pax-QNR* montre que le taux de messagers augmente de E5 jusqu'à E8. A partir de ce stade, il atteint un plateau et reste constant jusqu'à l'éclosion (**Fig. 4**). Par hybridation *in situ*, nous avons montré que, comme chez le poisson-zèbre, *Pax-QNR* est exprimé à E3 dans le cerveau postérieur et dans les ébauches des yeux, au niveau de la rétine non différenciée, dans le cristallin et l'épithélium de la cornée. Dans la neurorétine différenciée en plusieurs couches, l'expression de *Pax-6* est restreinte à la couche des neurones ganglionnaires et à la partie interne de la couche nucléaire interne qui contient les neurones amacrines (**Fig. 5**). L'expression de ce gène semble

donc être nécessaire non seulement durant la période de prolifération mais également pendant la différenciation finale de la neurorétine.

L'ADNc MC29QNR2 ayant été isolé dans une banque de cellules de neurorétine infectée par le virus MC29, nous nous sommes également intéressés à une potentielle régulation de *Pax-QNR* par l'oncogène *v-myc*. Lorsque l'on infecte de la neurorétine avec le virus MC29, il y a une diminution de l'expression de *Pax-QNR* (**Fig.6**). Il ne s'agit pas d'une sélection, par ce virus, de population cellulaire n'exprimant pas ce gène car la majorité des cellules infectées continue de l'expressible d'isoler des clones infectés par MC29 qui n'expriment pas *Pax-QNR*, ce qui indique que les cibles cellulaires du virus sont hétérogènes et peuvent ne pas exprimer *Pax-QNR*.

Characterization of a paired box- and homeobox-containing quail gene (Pax-QNR) expressed in the neuroretina

Patrick Martin, Catherine Carriere, Christine Dozier, Brigitte Quatannens, Marie-Anne Mirabel, Bernard Vandenbunder, Dominique Stehelin & Simon Saule

CNRS URA 1160, Institut Pasteur de Lille, 1 Rue Calmette, 59019 Lille cedex, France

The retina is an integral part of the central nervous system, and consists of two layers, the outer pigmented layer and the inner sensory layer or neuroretina (NR). The NR layer contains several strata of cells (glial and neuronal) derived from proliferating neuroectodermal precursors that differentiate after terminal mitosis. In vitro, NR cells can differentiate not only into neuronal and glial types, but also into pigment and lens cells. Quail (Coturnix coturnix japonica) NR cells (QNR) infected with MC29 transforming retrovirus become pigmented after several passages in vitro. In order to characterize the genes expressed in these pigmented MC29 QNR, a cDNA library was prepared from these cells. After differential screening we have isolated a cDNA clone which identifies an RNA expressed in NR but not in the pigmented layer of the retina. This cDNA encodes a protein related to that of Drosophila, mouse and zebrafish paired box- and homeobox-containing segmentation genes and is called Pax-QNR. The expression of Pax-QNR in the NR is confined to the ganglionic cell layer and to the lower part of the inner nuclear layer containing the amacrine or correlation neurones.

Introduction

The vertebrate embryonic eye offers exceptional advantages for investigations of the molecular basis of differentiation, since it exhibits a wide repertoire of possible tissue type interconversions *in vivo* and *in vitro*. Cells of the neuroretina (NR) have the potential to transdifferentiate into lens cells or pigmented cells (Okada, 1976). Identification of the genes involved in these processes is required to understand these events at the molecular level.

Avian NR development proceeds through the proliferation of neuroectodermal precursors present in the anterior part of the neural tube. Three periods in the development of the avian retina have been defined: (1) proliferation from the second to the eighth day of incubation; (2) a period of cell readjustment, from the eighth to the tenth day; and (3) a period of terminal differentiation from the tenth day to the end of incubation (Romanoff, 1954). Cell differentiation is the product of differential gene expression, and transcription factors encoded by oncogenes (viewed as transcription

Correspondence: S. Saule

regulators) may regulate such events. The avian myelocytomatosis virus strain MC29, which has transduced the *myc* oncogene, has been used to study differentiation in a variety of cells. MC29 alters the competence for terminal differentiation of myogenic cells, does not suppress the appearance of a differentiated phenotype in chondroblasts or macrophages and induces the expression of catecholaminergic traits in neural crestderived cells (Gazzolo *et al.*, 1979; Graf *et al.*, 1981; Alema *et al.*, 1985; Falcone *et al.*, 1985; Fauquet *et al.*, 1990).

Quail NR (QNR) infected with MC29 proliferates much better than helper-infected QNR, and most of the cultures become pigmented. Therefore, NR cells infected with MC29 virus constitute a model to study mechanisms that regulate growth and differentiation, presumably under the control of the v-myc product. In order to isolate and characterize the genes potentially controlled by this oncogene in pigmented MC29 QNR (PMC29 QNR), we constructed a cDNA library from these cells, and used the molecular hybridization technique (differential screening) to isolate cDNAs corresponding to genes expressed in these cells but not expressed in quail embryonic cells (QEC) transformed by MC29.

Here, we describe the nucleotide sequence and expression pattern of a quail cDNA clone, MC29-QNR-2. The derived amino acid sequence of this cDNA includes paired box and homeobox sequences, conserved domains individually capable of sequencespecific DNA binding (Bopp et al., 1986; Treisman et al., 1989; 1991; Kissinger et al., 1990). This gene, which appears to be the quail homologue of the recently described Pax(zf-a) gene from zebrafish (Krauss *et al.*, 1991), has been named Pax-QNR. At least eight paired box genes (Pax-1 to Pax-8) have been identified in the mouse genome (Walther et al., 1991), and only three of these (Pax-3, Pax-6 and Pax-7) contain a paired-type homeobox in addition to the paired box (Goulding et al., 1991; Jostes et al., 1991; Walther et al., 1991). the Pax-6 paired sequence is 100% identical at the amino acid level to the Pax-QNR paired domain. Therefore, Pax-QNR, Pax-6 and Pax(zf-a) probably represent the same gene from distinct species. Pax-QNR is expressed in the neuroretina and in the cerebellum of chicken and quail embryos, but not in other tissues tested, including the pigmented retina, and is the first Pax gene to be isolated in avian species. The amount of Pax-ONR mRNA in the developing neuroretina increases regularly, reaching a maximal intensity at day 8 of embryogenesis and remaining constant thereafter. In situ hybridization performed in developing neuroretina reveals strong expression of Pax-QNR in the

Received 9 January 1992; accepted in revised form 26 March 1992

ganglionic cell layer and in the lower part of the inner nuclear layer.

Results

Isolation of Pax-QNR cDNA

Experiments leading to the isolation of *Pax-QNR* were initially performed to identify neuroretina-specific genes that may be involved in the proliferation or differentiation of MC29-infected QNR. After 15 serial passages *in vitro*, most of the MC29-infected QNR appeared black. The cDNA library constructed from these cells was screened by differential hybridization with radioactively labelled probes complementary to mRNA from MC29-infected QNR depleted of mRNA from MC29-infected QEC. We isolated two sets of cDNAs, both expressed in MC29 QNR but not in MC29 QEC. Since one of them (MC29-QNR-1) was found to contain a new paired-type homeobox, we focused on this set of cDNA. Using MC29-QNR-1 (found to be partial and to contain introns) as a probe, we isolated MC29-QNR-2, which contained a complete open reading frame corresponding to this *Pax* gene.

Nucleotide sequence analysis of MC29-QNR-2 cDNA

The nucleotide sequences of MC29-QNR-1 and MC29-QNR-2 and the sequence of part of the cellular equivalent as well as the sequence of the cDNA isolated from another cDNA library were determined. The sequence of MC29-QNR-2 (2509 bp) is presented in Figure 1 along with the derived sequence of the longest open

			1 AA	ттсст	TTTTT	TTTTA	TTGTC	ATTGA	CATT	AAACT	GTGTG	GCAGG	TTCCG	GCGTA	GGAAT	CGCCG	AGCGG	AGCCG	CAGG	FGCCT	CTCG	secce	20000	ccccc	GACA	GGACC
118	TCGG	CATCG	CACCG	00000	GCAGC	CTCAG	TGGAG	ACAAA	CTATT	ITGCG	GCGA	CACTT	TGGGA	GCACG	FCAGT	secco	CGGT	SCGCC	CCGTC	CTGCT	GACAG	CAGCGO	STGCC	CGGAG	GGGA	TATGC
247	CTCT	GCGAG	AGGGC	cacco	GGAGT	AGAGG	CACGC	AGATG	TGCGA	secce	CACGC	TGCAG	GAGAG	GACAG	cecec	ccGGC	TTCCA	ACCC	ICCAG	SCGAG	CCCG	CGGGAG	cccc	TGCC	ccccc	CCACC
377	ATG	CAG	AAC	AGT	CAC	AGC	GGA	GTG	AAC	CAG	CTC	GGC	GGG	GTG	TTC	GTC	AAC	GGG	AGG	CCG	CTG	ccc	GAC	TCC	ACG	CGG
1	Met	Gln	Asn	Ser	His	Ser	Gly	Val	Asn	Gln	Leu	Gly	Gly	Val	Phe	Val	Asn	Gly	Arg	Pro	Leu	Pro	Asp	Ser	Thr	Arg
																						•				
455	CAG	AAG	ATC	GTG	GAA	CTC	GCG	CAC	AGC	GGC	GCG	CGG	CCG	TGC	GAC	ATC	тсс	CGA	ATC	CTG	CAG	GTG	TCG	AAT	GGA	TGT
27	GIN	Lys	Ile	Vai	Glu	Leu	Ala	His	Ser	GIÀ	Ala	Arg	Pro	Cys	Asp	Ile	Ser	Arg	ile	Leu	GIN	Val	ser	A\$n	GTÀ	Cys
6 3 3	CTC	ACT		م	TTC	<u> </u>	200	T A T	TAC	C 3 3	NCT	<i></i>	тсс	ATC.	200		200		1.70	CC N	CCT	ACT	330	~~~	ACA	GTA
53	Val	Ser	Lys	Ile	Leu	Gly	Arg	Tyr	Tyr	Glu	Thr	Gly	Ser	Ile	Arg	Pro	Arg	Ala	Ile	Gly	Gly	Ser	Lys	Pro	Arg	Val
	1																									
611	GCG	ACT	ccc	GAA	GTT	GTA	AGC	ААА	ATA	GCG	CAG	TAT	AAA	CGA	GAG	TGC	ccc	TCC	ATC	TTT	GCG	TGG	GAG	ATT	CGA	GAC
79	Ala	Thr	Pro	Glu	Val	Vai	Ser	Lys	Ile	Ala	Gin	Tyr	Lys	Arg	Glu	Cys	Pro	Ser	Ile	Phe	Ala	Trp	GIU	Ile	Arg	Asp
c 0 0			676	.	C C C		~~~~		200		~ ~ ~			~~~	NOT	r		Tee		220	202		CTC.	~~~		CTC
105	AGA	Leu	Leu	Ser	Glu	Gly	Val	Cys	Thr	Asn	Asp	Asn	Ile	Pro	Ser	Val	Ser	Ser	Ile	Asn	Arg	Val	Leu	Arg	Asn	Leu
1																										
767	GCŤ	AGC	GAA	AAG	CAA	CAG	ATG	GGT	GCC	GAC	GGG	ATG	TAC	GAC	AAG	CTA	AGG	ATG	CTG	AAC	GGG	CAG	ACG	GGG	ACA	TGG
131	Ala	Ser	Glu	Lys	Gln	Gln	Met	Gly	Ala	Asp	Gly	Met	Tyr	Asp	Lys	Leu	Arg	Met	Leu	Asn	Gly	Gln	Thr	Gly	Thr	Trp
0.45									~~~								~~ `	~ 1 1	V V				a	~ ~	~ ~ ~	
157	Gly	Thr	Arg	Pro	Gly	Trp	TYE	Pro	Gly	Thr	Ser	Val	Pro	Gly	Gln	Pro	Ala	Gln	ASD	Gly	Cys	Pro	Gln	Gln	GAG	GLY
							- 4 -		1					1					L	-	-1					
923	GGA	GGG	GAG	AAC	ACC	AAC	тсс	ATC	AGC	тсс	AAT	GGG	GAA	GAT	тсс	GAT	GAG	GCC	CAG	ATG	AGA	CTT	CAG	TTG	ААА	AGG
183	Gly	Gly	Glu	Asn	Thr	Asn	Ser	Ile	Ser	Ser	Asn	Gly	Glu	Asp	Ser	Asp	Glu	Ala	Gln	Met	Arg	Leu	Gln	Leu	Lys	Arg
	,																			Ŧ						
209	AAG	CTG	CAG	AGA	AAT	AGG	ACA	TCC Ser	TTT Phe	ACC	Gln	GAG	CAA Gln	ATA	GAA	GCC	CTT	GAG	AAA Lvs	GAG	TTT	GAG	AGG Arg	ACC Thr	CAC	TAT
205	-1-	204							2110		01	014		110				•••	2/3	•10						-1-
1079	ccc	GAT	GTG	TTT	GCG	AGA	GAG	CGA	CTA	GCT	GCT	ААА	ATA	GAC	TTG	CCT	GAA	GCA	AGG	ATA	CAG	GTG	TGG	TTT	тст	AAC
235	Pro	Asp	Val	Phe	Ala	Arg	Glu	Arg	Leu	Ala	Ala	Lys	Ile	Asp	Leu	Pro	Glu	Ala	Arg	Íle	Gln	Val	Trp	Phe	Ser	Asn
										1																
261	AGA	AGG	GCC	AAG	TGG	AGA	AGG	GAG	GAG	AAG	CTG	CGG Ara	AAT	Gln	CGG Arg	AGA	Gln	GCC	AGC	AAC Asn	ACT Thr	Pro	AGC	CAC	ATC	CCC Pro
				273						515	Lea			9111			.		002				002			
1235	ATC	AGC	AGT	AGT	ттс	AGC	ACA	AGC	GTT	TAC	CAG	ccg	ATC	сса	CAG	CCG	ACC	ACC	ccc	GGT	тсс	ATG	TTG	GGC	AGG	ACA
287	Ile	Ser	Ser	Ser	Phe	Ser	Thr	Ser	Val	Tyr	Gln	Pro	Ile	Pro	Gln	Pro	Thr	Thr	Pro	Gly	Ser	Met	Leu	Gly	Arg	Thr
																										T
1313	GAC	ACG	GCG	CTC	ACA	AAC	ACG	TAC	AGC	GCG	CTG	CCG	CCC	ATG	CCC	AGT	TTC	ACC	ATG	GCC	AAC	AAC	CTG	CCT	ATG	CAA
212	Азр	TUL	AId	Leu	THE	Mall	1.11	τγι	26T	AIA	Leu	<i>P1</i> 0	FIO	met	FIG	261	rne	7115	rie c	ALA	ASI	ASII	Leu	<i>P10</i>	nec	GTU
1 3 9 1	ccc	ccc	GTA	ccc	AGC	CAG	ACC	тсс	тсс	TAC	тст	TGC	ATG	ста	ccc	ACC	AGC	сст	тса	GTG	AAC	GGC	CGG	AGC	тат	GAC
339	Pro	Pro	Val	Pro	Ser	Gln	Thr	Ser	Ser	Tyr	Ser	Cys	Met	Leu	Pro	Thr	Ser	Pro	Ser	Val	Asn	Gly	Arg	Ser	Tyr	Asp
																									T	
1469	ACC	TAC	ACC	ССТ	CCG	CAC	ATG	CAG	ACA	CAC	ATG	AAT	AGC	CAG	ccc	ATG	GGC	ACC	TCT	GGC	ACC	ACT	TCC	ACA	GGT	CTC
365	THE	TYP	inr	PIO	Pro	HIS	Mec	GIN	Thr	HIS	Met	Asn	ser	Gin	Pro	Met	CIÀ	TUL	ser	σtλ	rnr	ing	ser	inr	GΤΫ	Leu
1547	አጥጥ	TCC	~~ ~	66.2	GTC	ተር ነ	GTT	CC 2	CTT	6 8 8	GTT	CCT	aan	ACT	6 2 2	~~ *	ChT	ATC.	T.C.T.	C A C	тас	тас	CCA	101	ፐፐ ካ	C 3 C
391	Ile	Ser	Pro	Gly	Val	Ser	Val	Pro	Val	Gln	Val	Pro	Gly	Ser	Glu	Pro	Asp	Met	Ser	Gln	Tyr	Trp	Pro	Arg	Leu	Gln
1625	ТААА	CATCH	GTTA	TTOTO	TAAAT	GACTA	TATO	ACAAC	AGTTO	GATGT	TCAG	AGTAT	-	AAACO	AAGAA	TGGCT	CTCAC	AGTAC	TTCTC	TACTO	CAACT	GTGCC	CTACO	ATGTA	ACACA	TGGA
1754	34407	CTCT2	ACACC	CACCT	TTOT				ATCCC	TTOC		TOTTO		COTAT	CCARA	CTTTT	ATTC?		GTGT	TTATT	TOTAS	ATGOO	CATT	GTATO	TTATA	ATGA
1907	****												*****	ACT.					ACCAN			AAAC"		AC2 40		arca
2012	MACA								JUAIG		auror	GI FGG	ri uA i C	mno I'l		ANGAA	~~~~				~~~~					arra
2012	ACAAC		JUCACO	AA111	AAGAC	JTTT7	UCAAC	ATATA	CGAA	LACTI	CTACC	AATCI	GTTC	ALAGAT	LATGO	ATTGA	CUTTO	,CAAG1	TUTA	I CA CI		GLA (A	LAATI	MAACC	LGGAA	
2141	ACGCA	CTAG	TTTAI	GTAAC	AAAA	TATCI	GTTGC	TATT	CCAAA	GGTTC	TTAAC	GGACO	AGCTA	TGTGC	AAACA	AGGGT	TAAGA	GAGAA	CTTCG	ACGGA	GAAAA	GAGTT	TOATA	GATAA	AAGGT	AGAT
2270	TGTGT	CTTCC	ATATA	ATCCA	ATTTO	TTTT	TGTCC	AAATO	TAAGT	ATTTO	TCTTC	CCTAC	AAATO	TCCCA	GAATO	ATTTO	TATA	TAAAC	TTAAT	TTCAT	TTATA	TTTGA	CAAGA	ATATT	СТАТА	GATG
2399	TTTTA	TACAC	ATTT	CATGO	ATCAT	ACGTT	TCTT	TTGGT	CGGCA	AAAAC	TTAAT	TGTTC	TTAG	TATAG	TTGTA	CTACI	GTTC	CAGTO	CAATO	ATTT	TGTGC	АТСТА	G			

Figure 1 Nucleotide and deduced amino acid sequence of the MC29-QNR-2 cDNA. White arrowheads denote positions of introns found in the MC29-QNR-1 cDNA. Black arrowheads denote positions of introns found in the *Pax-QNR* genomic clone. The paired and the homeodomains are boxed

reading frame. This open reading frame codes for a protein of 416 amino acids. The 5' untranslated sequence contains 376 nucleotides and the 3' untranslated sequence contains 885 nucleotides. A putative polyadenylation signal AATAAA (Fitzgerald & Shenk, 1981) is found 149 nucleotides from the 3' end of the cDNA. Because Northern blot experiments revealed a 3.2-kb Pax-QNR transcript (see Figure 3), this cDNA most probably contains only part of the 5' and 3' untranslated region. The deduced protein of 416 amino acids revealed a paired domain and a paired-type homeodomain (boxed sequences in Figure 1). One in-frame translation start codon (Figure 1, CCACCAUGC) is present in the context of a favourable start sequence (CCPuCCAUGG, Kozak, 1986; 1987). Sequence homology analysis reveals that this gene is the quail counterpart of the Pax(zf-a) gene isolated from the zebrafish (Krauss et al., 1991, and Figure 2c) and (except Pax-6) is distantly related to the mammalian Pax genes already published.

The paired domain resides close to the amino terminus of the protein, as do all the paired box-containing genes isolated to date. Only three amino acids precede the 128 amino acid paired domain. The paired domain of *Pax-QNR* protein is 100% homologous to

the paired domain of Pax-6, 98.2% homologous to that of Pax(zf-a) and only 66.4% homologous to the paired domain of Pax-3. A similar percentage of homology is found within the Drosophila paired domain of the gene prd (62.7%). With the exception of the *Pax-1* paired box, which is not interrupted by any intron, all other murine paired boxes are encoded by three different exons, which is also the case for Pax-QNR (see Figure 1 for the position of the introns). The exon-intron boundaries have been determined by comparing cDNA and genomic sequences (the complete genomic structure of Pax-QNR will be presented elsewhere). Comparison of the different paired domains (Walther et al., 1991) shows five clusters of high homology: between amino acids 4 and 19; 35 and 41; 68 and 74; 95 and 100 and 115 and 120 (Walther et al., 1991, and Figure 2a). The octapeptide sequence found at the end of the paired domain in the murine Pax-1, Pax-2, Pax-3 and Pax-7 genes (Goulding et al., 1991; Jostes et al., 1991) and the Drosophila gooseberry proximal (gsbp) and gooseberry distal (gsb-d) genes is absent from the Drosophila paired (prd) gene as well as from this Pax-QNR gene.

The homeodomain contains a helix-turn-helix sequence specifying a DNA-binding motif and has

a						
	1	10	2 0	30 40	50 <u>60</u>	
Pax-QNR	SESGVNOI	LGGVTVNGRP1	PDSTRQKIVE	LLASGARPCDISRILQVSN	GCVSKILGRYYETGSIRP	
Der.3	COG 8 * * * *		* N H T * H * * * *	* # * * H * I * * * V * * * O * R * * H *	***********	
Pax-6	*******					
Pax-7	G Q G R + + + -		* N H I * H * * * *	• H • • E • I • • • V • • • Q • R • • E	* * * * * * * C * * Q * * * * * * *	
Prd	G Q G R * * * *	• • • • • 1 • • • •	* N N I * L * * * *	• H • A D • I • • • V • • • Q + R • • E	* * * * * * * N * * Q * * * * * * * *	
gsp-p	GQGR * * * *	• • • • • I • • • •	****I*I****	* M * A S * V * * * V * * * Q * R * * E *	* * * * * * * N * * Q * * * * * * * *	
gsp-d	GQGR • • •	• • • • • I • • • • •	* N H I * RQ * * *	* M * A A * V * * * V * * * Q * R * * B *	* * * * * * * N * F Q * * * * * * *	
	7 0	8 0	9 0	100 110	120	
			· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·		
Pax-QNR	RAIGGSKI	PR VATPEVVS	KIAQYKRECP	SIFAWEIRDRLLSEGVCTNI	DN125V351NRVLXNLX	
ZI-2						
Pax-5		- XQ - T - D - E -		·	**************************************	
Pax-0					стV * * * * * * 5 * * * * I К F	
Prd	GV+++++	** *******	R + E E + + + S S +	GH+S+++EX+IR++++DR	STA * * * * A * S * L V * G R D	
990-n	GV+++++	** ********	R. DELRO .N.	· · · · S · · · · E K · I K · · F A ·	P * * T * * * S * L * * G S D	
8-7-7-7 250-d	G V * * * * *	• • • • • • • • • • • • • • • • • • •	R*EEL+QSQ*	- G + • S * • • • A K • I E A • • • D K (Q N A * * * * * * S * L * * G S S	
D		1 10	2 0) 30 40	0 50 60	
Pre OND						
FIL-QAK	GHRLGLK		QEQIEREERE			
Pay-3		* ~ ~ * * * * * * * * * *	A + + L + F + + B A		K * T * * * V * * * * * * * * * R * * KOA	
Pax.7	FPD+P+++	********	A		K * T * * * F * * * * * * * * * * * * R * * KQA	
Prd	EPGIA	**OR*C**T*S	ASTLDETTRA		N * T * * * * * * * * * * * * R L * KQH	
2sb-d	EPGIP***	+ Q R + S + + T + +	A * * L * * * * R A	. • 5 • • Q • • • • YT • • E • • Q TT ;	A • T • • • • • • • • • • • • • R L • K H S	
gsb-p	E P S V * * * *	• • Q R • S • • T • S	N D * • D * * • R I	. • A • • Q • • • • • ¥T • • E • • Q S T C	G • T • • • V • • • • • • • • • R L • X Q L	
с						
Pax-QNR	PGFQTAPO	EPRGTPLPRP	THONSESOVN	QLGGVFVNGRPLPDSTRQK	IVELAHSGARPCDISRILQVSNGCVS	X
21-1	MPQKEYYN	IRATWE SGVA S	м • • • • • • • • •		<u></u>	•
	ILGRYYET	GSIRPRAIGG	SXPRVATPEV	VSKIAQYXRECPSIFAWEIF	RDRLLSEGVCTNDNIPSVSSINRVLR	. N
-			<u>•••••</u>			-
	T		*******		CROO FOCGENENSISSNOFOSOFIO	м
	LASERQOM	IGAUGMIDELH	MUNGQIGING.	,TRPGNIPGISVPGQPAQDGC		• •
	81.01 × 81.77	OBNETSFILLE	OTFALLERE		PEAR TO VWESN RRAXWAREEKLANOR	R
	*****					•
	L.,					
	QASNTPSE	IPISSSFSTS	VYQPIPQPTT	P GSHLGRTDTALTN	NTYSALPPHPSFTMANNLPHQPPVPS	Q
	* * * * 5 5 * *		• • • • • • • • • •	• V S S F T S • • • • • • S • • • • • •	• • • • • • • • • • • • • • • • • • • •	•
	TSSYSCHI	PTSPSVNGRS	YDTYTPPBHQ	TEMNSQPHGTSGTTSTGLIS	SPGVSVPVQVPGSEPDMSQYWPRLQ	
	• • • • • • • •	• • • • • • • • • • •	• • • • • • • • • • •	· . · · · · · · · · · · · · · ·	• • • • • • • • • • • • • • • • • • • •	

Figure 2 (a) Protein sequence comparison of the Pax-QNR paired box with those of other homeobox-containing Pax genes. Sequences were from *Drosophila* (Bopp *et al.*, 1986), mouse (Walther *et al.*, 1991) and zebrafish (Krauss *et al.*, 1991). (b) Comparison of the Pax-QNR homeodomains with other extended paired-type homeodomains. (c) Comparison of the Pax-QNRsequence with the Pax(zf-a) protein deduced from the sequence of Krauss *et al.* (1991). Paired and homeodomains are boxed. Asterisks indicate conserved amino acids

been identified in proteins encoded by a number of development-controlling genes (Kissinger et al., 1990). The 61 amino acid Pax-QNR homeodomain shows 100% homology with the Pax(zf-a) homeodomain (Figure 2b) and is distantly related (64%) to the homeodomain of the murine Pax-3 and Pax-7 genes, slightly nearer than to the Drosophila prd gene (54.1%). The third helix of the paired-type homeodomain is highly conserved among all Pax genes, and the serine at position nine, which confers sequence-specific recognition on the homeodomain (Treisman et al., 1989), is also present in Pax-QNR. Helix 1 and helix 2 are much less conserved. Recently the structure of the engrailed homeodomain has been described by crystallography (Kissinger et al., 1990). By comparison it is therefore possible to suggest a role for the amino acid residues of the Pax-ONR homeobox. The amino acids of the engrailed homeodomain interacting with phosphates or bases in DNA are conserved in the Pax-QNR homeobox. Arg-3 and Arg-5 interacting with bases in the minor groove of DNA are present, as well as all the amino acids interacting with phosphate backbone with the exception of Lys-57, which in the Pax-QNR homeodomain is an arginine. Amino acid differences are expected in the third helix, since amino acids interacting with the bases of the major groove of DNA are different between the engrailed homeodomain and the paired-type homeodomain (Treisman et al., 1989). These differences are observed, and among the amino acids only Asn-51 is conserved, Ile-47 is a valine and Gln-50 is a serine (Kissinger et al., 1990). The homology among the different paired-type homeodomains extends the homeodomain itself since the amino acid motif LKRK is found in all these domains and precedes the homeobox (Figure 2b). The Pax-QNR homeobox is also distinct from other Pax genes at the structural level since the Pax-QNR gene contains two introns (as shown with the MC29-QNR-1 cDNA, probably produced from a premessenger RNA, and as deduced from comparison with the genomic sequences; see Figure 1) in the homeobox. The second intron in the homeobox is at the same position as in the Evx-1, the mammalian homologue of the Drosophila evenskipped gene (Bastian & Cruss, 1990).

In addition to the paired domain and the homeodomain, Pax-QNR protein contains between these two domains a stretch of 78 amino acids, with 13 glycine residues scattered throughout this sequence, indicating the relative paucity of any helical structure within this part of the molecule. As also found in *Pax-3* (Gouding *et al.*, 1991), the 146 amino acid carboxy terminus of *Pax-QNR* is particularly rich in proline, serine and threonine residues.

Tissue specificity of Pax-QNR expression

We investigated the presence of *Pax-QNR* mRNA in various tissues from E14 quail and chick embryos (Figure 3a and 3b). Identical results were obtained for the chick and the quail (data not shown). This gene exhibits a restricted tissue specificity, since its mRNA was detected by Northern blot in neuroretina and, to a lesser extent, in the cerebellum (Figure 3b) but not in other tissues examined, including pigmented retina. This suggests that MC29 induces a pigmented phenotype in QNR cells, maintaining the expression of QNR



Figure 3 Northern blot analysis of Pax-QNR transcription. (a) expression of Pax-QNR in chicken embryonic tissues. Total RNA was extracted from E14 embryonic tissues indicated at the top of the figure. A 10-µg aliquot of RNA was hybridized with a ³²P-labelled MC29-QNR-2 probe. Hybridization of the RNA blots with the actin probe was used as control. Bars on the left side of the figure show the position of the 28 and 18S rRNA. (b) Expression of Pax-QNR in the brain of a 2-day-old hatched chick

genes already expressed before the transformation. Three periods in the development of the avian retina are defined: (1) cell division primarily at the margin of the optic cup, from E2 to E8; (2) cell migration and lamination of cell strata from E8 to E10, and (3) a period of terminal differentiation from E10 to hatching. We prepared a Northern blot experiment with RNA from dissected retinas starting from stage E5, at various intervals during development and in newly hatched chickens (Figure 4). The amount of *Pax-QNR* RNA increased from E5 and reached a plateau at E8. Densitometric scanning of the gel indicates that *Pax-QNR* is three times more abundant at E8 than at E5. This level of expression is maintained even after hatching, when the neuroretina is already differentiated.

Localization of Pax-QNR expression by in situ hybridization

To determine the cellular localization of Pax-ONR mRNA, we hybridized ³⁵S-labelled RNA probes to eye sections prepared at E3, E8 and E12. The hybridization signal was intense at E3 in the developing eyes and the hindbrain. No hybridization was found in the mesencephalon (Figure 5a). Similar results were found for Pax(zf-a) in the developing zebrafish embryo (Krauss et al., 1991). In the eye at this stage, the undifferentiated retina, the lens and the epithelium expressed Pax-QNR mRNA. In contrast, regionalization of the hybridization signal was observed at E8 and E12. The hybridization signal is clearly restricted to the neuronal ganglionic cell layer and to the lower portion of the inner nuclear layer, which probably corresponds to the amacrine neurones (Figure 5b and 5d). The control experiment with the sense probe was negative (Figure 5f).

Pax-QNR expression in MC29-infected QNR

Since the Pax-QNR cDNA was isolated from MC29infected neuroretina, we investigated the effect of viral infection on Pax-ONR expression. Surprisingly, infection of ONR cells with MC29 led to a decrease in the amount of Pax-QNR RNA detected by Northern blot analysis after several passages (Figures 4 and 6). Cells expressing the myc oncogene expressed the Pax-ONR gene as shown by in situ hybridization (Figure 5c), whereas most of the cells hybridized with the Pax-QNR probe. Therefore, the decrease in Pax-QNR mRNA in PMC29 QNR is not the result of the selection of a particular non-expressing cell. Figure 6, lane Cl 1, shows expression of Pax-QNR from a clone of MC29 QNR derived from a single soft-agar colony. However, it is possible to isolate clones from MC29 QNR which do not express this gene (lane Cl 2), indicating that cell targets for MC29 are heterogeneous with respect to Pax-QNR expression.

Discussion

By using quail neuroretina cells transformed by the myc-containing MC29 virus, we have isolated a new gene not yet described in mammals or in birds which is the quail homologue of the recently described Pax(zf-a)gene from zebrafish. Expression of this gene shows a tissue specificity restricted to the neuroretina and, to a lesser extent, in the cerebellum. Therefore, we called this gene Pax-QNR. Comparison of Pax-QNR with other known paired genes reveals that Pax-QNR is distantly related to the paired box- and homeoboxcontaining Pax genes already published (Pax-3 and Pax-7) but possesses the same paired box as Pax-6. Surprisingly, two introns split the homeodomain of Pax-QNR encoded by three exons. One intron is found after amino acid codon 46, at the same position as in Evx-1 gene, the murine homologue of the Drosophila even-skipped gene (Bastian & Gruss, 1990), whereas no intron in the homeodomain has been reported for Pax-3 and Pax-7 (Goulding et al., 1991; Jostes et al., 1991). This finding suggests that an ancient common ancestor exists for *Pax-QNR* and *Evx-1* (without a paired-type homeodomain) or, alternatively, that *Pax-QNR* arose



Figure 4 Northern blot analysis of *Pax-QNR* transcription in the developing retina. RNA was extracted from embryonic chick NR dissected at the indicated days, and 3 days after hatching (H)

after a recombination event between different ancient genes.

Expression of Pax(zf-a) has been described in the developing brain of the young zebrafish embryo (Krauss et al., 1991) in the diencephalon and the hindbrain but not in the mesencephalon. Preliminary results obtained for the expression of *Pax-QNR* in the young chick embryo suggest a similar pattern of expression, but with additional expression found in the neural tube (unpublished data), as for other Pax genes (Kessel & Gruss, 1990). Within the neuroretina, Pax-QNR expression appears to be confined to the neuronal cell type, the ganglionic cell layer and probably the amacrine cells. Expression of this gene is probably not associated with proliferation, but rather with differentiation, since maximal levels of Pax-QNR expression are found in the retina not when the cells proliferate (before E8), but later, when post-mitotic cells are differentiating. Alternatively, the Pax-QNR product could have two distinct effects on NR cells: one effect in cell proliferation in early stages and one effect later in cell differentiation. The fact that in MC29 QNR Pax-ONR mRNA is less abundant than in normal cultures may be the result of selection by the virus of a particular cell type expressing less of this gene or of a down-regulation of this gene by factors such as myc, involved in cell proliferation.

What is the function of the Pax-QNR gene product? Pax products containing two DNA-binding domains (Treisman et al., 1989; 1991) are presumably transcription factors. Molecular control of differentiation processes such as lineage commitment and cellular maturation is likely to hinge upon lineage-restricted transcription factors that induce or repress batteries of subordinate genes. Several vertebrate homeobox-containing proteins have recently been identified as transcription factors that are required for the expression of lineage-specific genes: Oct-2, a member of the Pit-Oct-Unc (POU) family is necessary for immunoglobulin expression by B lymphocytes (Müller et al., 1988) Pit-1, another POU member, is associated with expression of several genes in the pituitary gland (Ingraham et al., 1988); LF-BI is involved in the expression of



Figure 5 Detection of *Pax-QNR* RNA by *in situ* hybridization. (a) Localization of *Pax-QNR* in a parasagittal section of the eye at E3 (e) and in the hindbrain (h) with a ³⁵S-labelled antisense probe. Note that the mesencephalon (m) is not labelled. Bars, $50 \,\mu\text{m}$. (b) Hybridization of 8-day neuroretina with the antisense probe. A strong signal could be detected in the ganglionic cell layer and in the lower part of the inner nuclear layer. PCL, pigmented cell layer; ONL, outer nuclear layer; INL, inner nuclear layer; IPL, inner plexiform layer; GCL, ganglionic cell layer. (c) Expression of *Pax-QNR* in PMC29 QNR with the antisense probe. (d) Hybridization of 12-day neuroretina with the antisense probe. A hybridization pattern is similar to the one found at E8. Bars, $120 \,\mu\text{m}$. (e) Hybridization of PMC29 with a sense probe. No signal over the background is detected. (f) Hybridization of 8-day neuroretina with the sense probe. No signal over the background is detected. (f) Hybridization of 8-day neuroretina with the sense probe. No signal over the background is detected.

liver-specific genes (Frain *et al.*, 1989); *TTF-1* in that of thyroid-specific genes (Guazzi *et al.*, 1990); *Isl-1* is required for insulin expression in the pancreas (Karlson *et al.*, 1990); and *Hlx* has been found restricted to the haematopoietic myeloid lineage (Allen *et al.*, 1991). In the young zebrafish embryo, the expression of Pax(zf-a) coincides with morphological landmarks corresponding to the primary axon tracts generated in the embryonic brain (Krauss *et al.*, 1991). The *Pax-QNR* product may play a similar role in instructing a cell to become a particular neurone in both the developing brain and the developing retina. However, this gene is still expressed in the retina after neuronal

differentiation has occurred, suggesting that this gene product may also play a role as a tissue-specific transcription factor regulating gene expression in the retina.

The effects of mutations in homeobox-containing genes are particularly well described in *Drosophila*, and the phenotype of the mutated embryos has led to a threefold classification of these genes: gap, pair-rule and segment polarity genes (Scott & Carroll, 1987). In vertebrates, analysis of the effect of artificially introduced mutations in homeobox-containing genes leading to an abnormal embryonic development is now emerging (Chisaka & Capecchi, 1991; Lufkin *et al.*, 1991; Rosner *et al.*, 1991). For the *Pax* genes, a spontaneous



Figure 6 Pax-QNR expression in MC29-infected QNR. Sevenday-old QNR was infected with MC29 (RAV-1). RNAs from the infected cells were extracted after several passages *in vitro*, as indicated at the top of the figure. Cl 1 and Cl 2 were two clones selected in soft agar from the P7 MC29 QNR. PMC29 QNR refers to the pigmented MC29 QNR from which MC29-QNR-1 and MC29-QNR-2 cDNAs were isolated

mutation located in the paired domain of the Pax-1 gene has been associated with the mouse developmental mutant undulated, which exhibits malformations in the vertebral column. This point mutation leads to a decrease in the DNA-binding affinity of the mutated product (Chalepakis et al., 1991). Another mutant, splotch, affecting development of the mouse neural tube, has been associated with a mutation in the Pax-3 locus (Epstein et al., 1991). Recently a mutation in the Pax-6 gene has been associated with the mouse small eve mutation (Sey), strengthening the role of this gene in eye development (Hill et al., 1991). Therefore, the avian neuroretina cells cultured in vitro will be useful for studying the effect of oncogenes and growth modulators in the regulation of this important transcription factor.

Materials and methods

Cells and viruses

NRs were dissected from 7-day-old quail embryos. QNR cultures were maintained and passaged in Eagle's basal medium supplemented with 10% fetal calf serum. MC29 RAV-1 virus was obtained by co-transfection of DNA from the molecular clone pMC38 (Vennström *et al.*, 1981) containing the myelocytomatosis provirus, and Rous-associated virus type 1 (RAV-1) DNA into quail embryo cells.

Construction and differential screening of the cDNA library

Polyadenylated RNA $(5 \mu g)$ was prepared from pigmented MC29 QNR and used to construct a cDNA library as described by Gubler & Hoffman (1983). After addition of EcoRI linkers and subsequent digestion with EcoRI, the cDNAs

were inserted in the EcoRI site of the phage vector lambda NM1149. The library was packaged *in vitro* by using a Gigapack plus extract (Stratagene) and amplified in *Escherichia coli* C600 HflA.

The amplified library was differentially screened with ³²Plabelled cDNA probes prepared from poly(A)⁺ RNAs from MC29 ONR, enriched in specific cDNA by subtractive hybridization with an excess of RNA extracted from MC29 QEC. cDNA probes were synthesized from $5 \mu g$ of poly(A)⁺ RNA by using $10 \mu g$ of oligo(dT) (Pharmacia) under the conditions described by Maniatis et al. (1982). The radiolabelled cDNA was hybridized with 10 mg of total RNA from MC29 QEC. Hybridization was performed at a Crt value of 45 000 mol l⁻¹ as previously described (Saule et al., 1981). The cDNA remaining single stranded was recovered after chromatography on hydroxylapatite. Hybridization on replicas of the cDNA library was performed as described in Guermah et al. (1990). Plaques giving a signal with the single-stranded cDNA but not with the double-stranded cDNA probe were isolated and purified with the same probes. Phage DNA was extracted as described in Maniatis et al. (1982). Inserts were subcloned into the EcoRI site of the pUC vector (Promega Biotech). Two independent screenings of the cDNA library resulted in the isolation of the same set of cDNAs.

DNA sequencing of the Pax-QNR cDNA

We sequenced two *Pax-QNR* cDNAs (MC29-QNR-1 and MC29-QNR-2), as well as the corresponding genomic clones isolated from a genomic library constructed from peripheral blood DNA of E5 quail embryos inserted in phages EMBL4 and screened with the MC29-QNR-2 cDNA. One of the two cDNAs (MC29-QNR-1) was only partial and was found to contain introns. Nucleotide sequence was determined on both stands after subcloning restriction fragments into M13mp18 or M13mp19 vectors, using standard techniques and an Applied Biosystem 370A automatic sequencer. All sequence homology searches were done with software from Bisance (Dessen *et al.*, 1990).

Northern blot analysis

Total RNA was isolated by using the guanidinium thiocyanate-caesium chloride method (Chirgwin *et al.*, 1979) from quail or chicken tissues. 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 nitrocellulose filter in 20 × SSC (standard saline citrate) and hybridized to *Pax-QNR* labelled by nick translation or by random priming.

In situ hybridization

An EcoRI-PstI fragment from MC29-QNR-1 containing part of the paired-type homeobox (100 bp) and part of an intron (500 bp) was subcloned into pSP64 and pSP65 plasmid vector (Promega Biotec) and used to prepare high specific activity ³⁵S-labelled RNA sense and antisense probes. These probes were processed and used as previously described (Vandenbunder *et al.*, 1989; Queva *et al.*, 1992).

Acknowledgements

We thank Carmen Garrido, Kay MacLeod and Philippe Duprey for critical reading of the manuscript and Jean Coll for helpful discussions. This work was supported by grants from the Centre National de la Recherche Scientifique, Institut Pasteur de Lille and the Association Retinitis Pigmentosa.

References

- Alema, S., Tato, F. & Boettiger, D. (1985). Mol. Cell. Biol., 5, 538-544.
- Allen, J., Lints, T., Jenkins, N., Copeland, N., Strasser, A., Harvey, R. & Adams, J. (1991). Genes Dev., 5, 509-520.
- Bastian, H. & Gruss, P. (1990). EMBO J., 9, 1839-1852.
- Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. (1986). Cell, 47, 1033-1040.
- Chalepakis, G., Fristch, R., Fickenscher, H., Deutch, U., Goulding, M. & Gruss, P. (1991). Cell, 66, 873-884.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. & Rutter, W.F. (1979). *Biochemistry*, 18, 5294-5299.
- Chisaka, O. & Capecchi, M. (1991). Nature, 350, 473-479.
- Dessen, P., Fondrat, C., Valencien, C. & Mugnier, C. (1990). Cabios, 6, 355-356.
 Epstein, D., Vekemans, M. & Gros, P. (1991). Cell, 67,
- Epstein, D., Vekemans, M. & Gros, P. (1991). Cell, 67, 767-774.
- Falcone, G., Tato, F. & Alema, S. (1985). Proc. Natl. Acad. Sci. USA, 82, 426–430.
- Fauquet, M., Stehelin, D. & Saule, S. (1990). Proc. Natl. Sci. USA, 87, 1546-1550.
- Fitzgerald, M. & Shenk, T. (1981). Cell, 24, 251-260.
- Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R. & Cortese, R. (1989). *Cell*, **59**, 145-157.
- Gazzolo, L., Moscovici, C., Moscovici, M.G. & Samarut, J. (1979). Cell, 16, 627-638.
- Goulding, M.D., Chalepakis, G., Deutch, U., Erselius, J.R. & Gruss, P. (1991). EMBO J., 5, 1135-1147.
- Graf, T., von Kirchbach, A. & Beug, H. (1981). *Exp. Cell Res.*, **131**, 331-343.
- Guazzi, S., Price, M., DeFelice, M., Damante, G., Mattei, M.-G. & DiLauro, R. (1990). *EMBO J.*, **9**, 3631–3639.
- Gubler, U. & Hoffman, B.J. (1983). Gene, 25, 263-269.
- Guermah, M., Gillet, G., Michel, D., Laugier, D., Brun, G. & Calothy, G. (1990). Mol. Cell. Biol., 10, 3584–3590.
- Hill, R.E., Favor, J., Hogan, B., Ton, C.C., Saunders, G., Hanson, I., Prosser, J., Jordan, T., Hastie, N. & van Heyningen, V. (1991). Nature, **354**, 522-525.
- Ingraham, H.A., Chen, R., Mangaiam, H., Elsholtz, S., Flynn, C.R., Lin, D.M., Simmons, L., Swanson, L. & Rosenfeld, M.G. (1988). *Cell*, 55, 519-529.
- Jostes, B., Walther, C. & Gruss, P. (1991). Mech. Dev., 33, 27-38.

- Karlson, O., Thor, S., Norberg, T., Ohlsson, H. & Edlund, T. (1990). Nature, 344, 879-882.
- Kessel, M. & Gruss, P. (1990). Science, 249, 374-379.
- Kissinger, C., Liu, B., Martin-Blanco, E., Kornberg, T. & Pabo, C. (1990). Cell, 63, 579-590.
- Kozak, M. (1986). Cell, 44, 283-292.
- Kozak, M. (1987). Nucleic Acids Res., 15, 8125-8146.
- Krauss, S., Johansen, T., Korzh, V., Moens, U., Ericson, J.U. & Fjose, A. (1991). EMBO J., 10, 3609-3619.
- Lufkin, T., Dierich, A., LeMeur, M., Mark, M. & Chambon, P. (1991). Cell, 66, 1105-1119.
- Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982). Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
- Müller, M.M., Ruppert, S., Schaffner, W. & Matthias, P. (1988). *Nature*, **336**, 544-551.
- Okada, T.S. (1976). In: Tests of Teratogenicity in Vitro. Ebert, J.D. & Marois, M. (eds.). Elsevier: Amsterdam, pp. 91-105.
- Queva, C., Ness, S., Grässer, F., Graf, T., Vandenbunder, B. & Stehelin, D. (1992). Development, 114, in press.
- Romanoff, A. (1954). *The Avian Embryo*, Macmillan: New York, pp. 382-395.
- Rosner, M.h., De Santo, R., Arnheiter, H. & Staudt, L. (1991). Cell, 64, 1103-1110.
- Saule, S., Roussel, M., Lagrou, C. & Stehelin, D. (1981). J. Virol., 38, 409-419.
- Scott, M.P. & Carroll, S.B. (1987). Cell, 51, 689-698.
- Treisman, J., Gönczy, P., Vashishtha, M., Harris, E. & Desplan, C. (1989). Cell, 59, 553–562.
- Treisman, J., Harris, E. & Desplan, C. (1991). Genes Dev., 5, 594-604.
- Vandenbunder, B., Pardanaud, L., Jaffredo, T., Mirabel, M.A. & Stehelin, D. (1989). Development, 106, 265-274.
- Vennström, B., Moscovicci, C., Goodman, H.M. & Bishop, J.M. (1981). J. Virol., 39, 625-631.
- Walther, C., Guenet, J.-L., Simon, D., Deutch, U., Jostes, B., Goulding, M.D., Plachov, D., Balling, R. & Gruss, P. (1991). Genomics, 11, 424-434.

ARTICLE II

"Structure and DNA binding-Properties of *Pax-QNR*, a Paired Box- and Homeoboxcontaining Gene."

Christine Dozier, Catherine Carrière, Delphine Grévin, Patrick Martin, Brigitte Quatannens, Dominique Stehelin and Simon Saule

Cell Growth & Differenciation, 4, 281-289, 1993



Fig.11 : Epissages alternatifs du gène Pax-QNR

Structure génomique du gène Pax-QNR et propriétés de fixation de son produit.

Afin de pouvoir étudier la régulation du gène *Pax-QNR*, la structure génomique de ce gène a été déterminée. Cette étude nous a également permis de montrer que des exons alternatifs déjà caractérisés dans d'autres espèces, étaient présents chez la caille. Nous nous sommes ensuite intéressés aux propriétés de liaison à l'ADN du produit de ce gène, plusieurs gènes de la famille Pax étant capables de se fixer à l'ADN.

Le gène *Pax-QNR* s'étend sur 17 kilo paires de bases et comporte au moins 15 exons (Cf. Fig.11). La boîte PRD est codée par trois exons. Le premier code l'hélice 1 du domaine PRD qui a été montrée, dans le cas de Pax-1, être essentielle à la fixation à l'ADN (Chalepakis *et al.*, 1991). Les deux hélices suivantes sont codées par le deuxième exon. La région codée par le troisième exon n'a pas de fonction définie mais est très conservée entre les différents domaines PRD connus. L'homéodomaine est également composé par trois exons alors que dans les autres gènes Pax, il est codé seulement par deux exons.

Un deuxième ADNc, le clone B1, a été isolé avec la sonde MC29QNR2 dans une banque d'ADNc de moelle épinière d'embryon de caille. L'analyse de sa séquence montre qu'il diffère du clone MC29QNR2 en deux régions. En 5', l'extrémité diverge à partir du site d'épissage au début du domaine PRD et ce clone ne possède pas de codon initiateur. Le clone MC29QNR2 et les gènes *Pax-6* chez le poisson-zèbre, la souris et l'homme ont des cadres ouverts de lecture très longs en amont de la boîte PRD, le clone B1 également, mais le cadre ouvert de lecture de ce dernier ne présente aucune homologie avec les autres. Ce clone B1 possède également une insertion en carboxy-terminal de 18 nucléotides codant pour 6 aa et cette insertion est retrouvée dans les gènes *Pax-6* de souris et *Pax(zf-a)* de poisson-zèbre. L'analyse de la structure génomique montre que les clones MC29QNR2 et B1 sont produits par des événements d'épissage alternatif différents. L'extrémité 5' de MC29QNR2 comprend les exons 0, 1, 2, 3, l'AUG se trouvant dans ce dernier exon, alors que celle du B1 est codée par l'exon α (Cf. Fig11). La séquence de 18 nucléotides en carboxy-terminal est retrouvée dans la boîte PRD de certains ADNc de souris (Walther *et al.*, 1991). Elle est localisée à la fin du premier exon de cette boîte.

Cet exon est également présent dans l'ADN génomique de *Pax-QNR* bien qu'aucun messager possédant cet exon n'ait été jusqu'alors isolé chez la caille.

L'analyse de l'expression des clones MC29QNR2 et B1, par protection vis à vis de la RNase avec des sondes correspondant aux extrémités 5' de chacun de ces ADNc montre que ces deux gènes sont exprimés dans la neurorétine et dans le cervelet (Fig.4). Ces épissages alternatifs se produisent donc *in vivo*. Le fait que d'autres bandes protégées soient détectées dans ces expériences suggère que d'autres événements d'épissage se produisent en 5' du gène.

Afin d'étudier les propriétés de fixation à l'ADN du produit du clone MC29QNR2, celuici a été traduit *in vitro* en lysat de réticulocytes. Il code un doublet de protéines de 46/47 KDa où la p46 est majoritaire, et ce poids moléculaire correspond à la taille attendue pour une initiation sur le premier AUG en amont du domaine PRD (**Fig. 5A**).

Plusieurs protéines Pax fixent l'ADN et reconnaissent la séquence e5. Cette séquence contient les consensus de fixation de l'homéodomaine (ATTA/TAAT) et du domaine PRD (GGAA/TTCC). Nous avons montré que la p46 de Pax-QNR peut également se fixer de façon spécifique sur e5. Chaque domaine de liaison à l'ADN produit séparément en bactérie, est capable de se lier à e5 (Fig. 5C). De plus, le domaine PRD peut se fixer sur un oligonucléotide muté dans le consensus de fixation de l'homéodomaine (e5P) et l'homéodomaine peut se fixer sur une séquence dépourvue du consensus de fixation pour PRD (e5H) (Fig. 5C). Cependant la protéine totale ne reconnaît que très faiblement la séquence e5P alors qu'elle se fixe avec une affinité légèrement réduite à la séquence e5H (Fig. 5B). Dans le cadre de la protéine entière, la présence de l'homéodomaine serait donc nécessaire pour une fixation optimale des deux domaines. Le fait que le domaine PRD puisse reconnaître la séquence e5P quand il est produit séparément mais pas dans le contexte de la protéine totale suggère que des événements conformationnels pourraient empêcher la fixation du domaine PRD en le masquant, cela en l'absence de fixation de l'homéodomaine. Cependant, des résultats obtenus par la suite sont en contradiction avec cette interprétation. En effet, nous avons pu montrer que le produit de Pax-QNR dépourvu d'homéodomaine est capable de se lier à la séquence e5 alors que toute altération du domaine PRD abolit la liaison à l'ADN. La fixation du domaine PRD est donc essentielle. Il a été montré pour plusieurs protéines Pax que le domaine PRD contacte plusieurs bases encadrant le consensus GGAA/TTCC et que ces contacts sont nécessaires à la fixation. Il est donc possible que la séquence e5P ne permette plus ces contacts d'où un mauvais positionnement de la protéine sur cette séquence.

Structure and DNA-binding Properties of Pax-QNR, a Paired Box- and Homeobox-containing Gene

Christine Dozier, Catherine Carrière, Delphine Grévin, Patrick Martin, Brigitte Quatannens, Dominique Stéhelin, and Simon Saule²

Centre National de la Recherche Scientifique, URA 1160. Institut Pasteur de Lille, 1 Rue Calmette, 59019 Lille cedex, France

Abstract

After differential screening of a complementary DNA library constructed from quail neuroretina cells (QNR) infected with the v-myc-containing avian retrovirus MC29, we have isolated a complementary DNA clone which identifies a mRNA essentially expressed in the neuronal layer of the retina. This complementary DNA encodes a protein containing paired box and homeobox domains. This gene, called Pax-QNR, is homologous to the murine Pax-6 and human AN genes, which are mutated in the autosomal dominant mutation small eye (Sey) of the mouse and aniridia in humans. Here, we report the genomic exon-intron organization, as well as the existence of alternative splicing events taking place at both the 5' end and the middle part of the gene. A Pax-QNR clone translated in reticulocyte lysate directed the synthesis of a 46 kilodalton protein able to bind specifically to the e5 sequence present upstream of the Drosophila even-skipped gene and target of the Drosophila paired protein. The Pax-QNR paired and homeobox domains individually expressed in bacteria are both able to recognize the e5 sequence.

Introduction

The vertebrate embryonic eye offers exceptional advantages for investigations of the molecular basis of differentiation, since it exhibits a wide repertoire of possible tissue type interconversions in vivo and in vitro. Cells of the NR³ have the potential to transdifferentiate into lens cells or pigmented cells (1). Cell differentiation is the product of differential gene expression, and transcription factors may regulate such events. Homeobox-containing genes encoding developmentally regulated transcription factors defined domains in the developing eye (2-4), and previous studies of Drosophila homeobox-containing genes have shown that these genes are involved in the process of pattern formation (for reviews, see Refs. 5 and

To whom requests for reprints should be addressed.

6). Data from vertebrate systems are consistent with this role (7-9).

The homeobox encodes a conserved DNA-binding domain (10), and several vertebrate homeo proteins have been identified as transcription factors required for expression of lineage-specific genes (11–13). The paired motif found in Drosophila developmental genes, such as paired and gooseberry, has also been shown to be conserved in vertebrates and found to encode another DNAbinding motif (14, 15). This motif is present alone (in the murine genes Pax-1, Pax-2, and Pax-8) or together with the homeobox domain (Pax-3, Pax-4, Pax-6, and Pax-7) (14). In contrast to homeobox-containing genes whose expression is region specific, the Pax genes are expressed along the entire anteroposterior axis in the neural tube (16). Several developmental mutations in the mouse have been associated with mutations in the Pax genes (17). Mutations in the Pax-6 gene have been associated with the mouse mutation small eye (Sey) (18). The same gene has been found to be deleted in some cases of the human congenital disorder aniridia (AN), which is a similar disorder to small eye (19), strengthening the role of this gene in eve development.

We have previously isolated from MC29 (an avian retrovirus bearing the v-myc oncogene) infected quail NR cells (QNR-MC29), a cDNA clone expressed in the neuroretina and in the cerebellum of chicken and quail embryos, but not in the pigmented retina. In situ hybridization performed in developing neuroretina reveals strong expression of Pax-QNR in the ganglionic cell layer and in the lower part of the inner nuclear layer. The derived amino acid sequence of this cDNA shows paired box and homeobox sequences, and the corresponding gene, which has been named Pax-QNR, appears to be the quail homologue of the recently described Pax(zf-a) gene from zebrafish (20) and Pax-6 from the mouse (3). To understand the evolution of the Pax family and to facilitate the analysis of their transcriptional regulation, the chromosomal genes must be cloned. To date, no complete genomic organization of a Pax gene has been reported.

In this paper, we show that the quail Pax-QNR gene is composed of a transcriptional unit spanning more than 17 kbp of genomic DNA and comprising at least 15 exons. Both the paired box and the homeobox domains are encoded by three exons. Alternative splicing takes place both at the 5' end and in the middle of the gene. The expected products each have a distinct amino-terminus and, due to the use of an alternative splice acceptor, differ by 6 amino acids in the first exon following the homeobox domain. The Pax-QNR gene encodes at least a 46 kd protein that specifically binds to the e5 sequence present upstream of the Drosophila even-skipped gene. This e5 sequence is recognized in vitro by the Drosophila paired and the murine Pax-2, Pax-3, and Pax-5 proteins (15, 21-23) but not by Pax-1 and Pax-8 proteins (24, 25). When expressed in bacteria, both the Pax-QNR paired

Received 11/12/92; revised 1/6/93; accepted 1/11/93.

^{&#}x27;This work was supported by grants from the Centre National de la Recherche Scientifique, Institut Pasteur de Lille, and the Association Française Retinitis Pigmentosa.

³ The abbreviations used are: NR, neuroretina; QNR, quail NR; cDNA, complementary DNA; bp, base pair(s); kbp, kilobase pair(s); kd, kilotase dalton(s); ORF, open reading frame; 5'-UTR, 5'-untranslated region; Innucleotides; SDS, sodium dodecyl sulfate; EMSA, electrophoretic mobility shift assay; PCR, polymerase chain reaction.



1346 CCG ACC ACC CCC GTT TCC TCT TTC ACA TCA GGT TCC ATG TTG Pro Thr Thr Pro Val Ser Ser Phe Thr Ser Gly Ser Met Leu

Fig. 1. A, structure of the two Pax-QNR cDNAs isolated. White box, the B1-specific 5'-UTR; gray box, paired domain; hatched box, homeodomain; black box, the 6 amino acids found in clone B1 but absent in clone MC29-QNR2. Initiation and stop codons are indicated. Restriction enzymes used: S, SstII; Hc, HincII; Sm, Smal; Sp, SphI. B, nucleotide and deduced amino acid sequence of the B1-specific 5'-end. Arrowhead, position of intron found in the Pax-QNR gene. The first paired domain amino acid is boxed. Asterisks, the potential stem-loop structure arising from the inverted repeats. The potential translation codon ACG is underlined. C, sequence comparison between MC29-QNR2 and B1 cDNAs showing the 6-amino acid insertion (arrows) in the B1 clone due to an alternative splice acceptor site in exon 10.

domain and homeodomain peptides recognize the e5 sequence. Thus, despite the fact that *Pax-QNR* bears the most divergent paired box and homeobox sequences in the *Pax* family, its product shows the same *in vitro* DNA-binding specificity as the other *Pax* gene products already characterized.

Results

Isolation and Nucleotide Sequence Comparison of *Pax***-***QNR* **cDNAs.** We previously described the isolation of a cDNA clone (MC29-QNR2) corresponding to a *Pax* gene (*Pax-QNR*) expressed in the neuronal layers of the retina. The sequence of MC29-QNR2 (2509 bp) has been already published along with the derived sequence of the

longest open reading frame (4). The deduced protein (416 amino acids) revealed a paired domain (Fig. 1*A*, gray box) and a paired-type homeodomain (Fig. 1*A*, hatched box). The paired domain resides close to the amino-terminus of the protein, as do all the paired box-containing genes isolated to date.

Subsequently, by screening a quail spinal cord cDNA library using MC29-QNR2 as a probe, we isolated another cDNA clone, named B1 (Fig. 1A). Nucleotide sequence analysis of cDNA B1, as compared with the MC29-QNR2 clone, shows two differences. First, the 5' ends of the two cDNAs diverge, and no AUG can be found in the ORF used to translate the B1 deduced protein (Figs. 1B and 2). Second, in the middle of the gene, a stretch of 18 nucleotides, also found in zebrafish (20), murine (3), and human species (19), is present in clone B1. This insert encodes 6 amino acids (Val-Ser-Ser-Phe-Thr-Ser) inserted at amino acid 306 of the deduced MC29-QNR2 protein (Fig. 1C). No AUG was found to translate the B1 ORF, but an ACG in a nonoptimal in-frame sequence (GCTCCGACGA, underlined in Fig. 1B) could initiate the protein. This ACG is followed by a potential stem-loop present 7 nucleotides downstream. If the stem-loop hinders ribosome scanning appropriately to promote initiation at the in-frame ACG site (26), the polypeptide encoded by B1 cDNA, and starting with a threonine, would contain 24 amino acids upstream from the paired domain, whereas in the MC29-QNR2 cDNA, only 3 amino acids precede the paired domain. However, as observed for the other species (murine, fish, and human), the MC29-QNR2 ORF can be extended 5' of the putative translational start site (ATG). When compared to the murine Pax-6, the zebrafish Pax(zf-a), or the human AN deduced protein, the extended amino-terminus of Pax-QNR appears divergent (Fig. 2). Stretches of homology can be found with murine and human species, although no homology can be found with the reported zebrafish sequence in this extended ORF. The observed differences can be due either to distinct alternative splicing in the 5'-UTR of these genes or to the divergence accumulated during species evolution.

Nevertheless, differences observed between the two Pax-QNR cDNAs strongly suggested that cDNA B1 could represent an alternative splice product of the Pax-QNR gene. In order to clarify this point, we characterized the genomic Pax-QNR gene.

Isolation of Genomic Clones and Characterization of the *Pax-QNR* Locus. Three phages, named λ Pax-QNR1, λ Pax-QNR2, and λ Pax-QNR3, were isolated from two different quail genomic libraries using parts of the *Pax-QNR* cDNA or genomic fragments as probes (Fig. 3A). Inserts containing the different exons of *Pax-QNR* were isolated, subcloned, and sequenced. Results obtained show that the *Pax-QNR* gene is organized in at least 15

MC29-QNR-2	ESPSGAARCLLGPPRRRTGPRHRTAAAASVETNYFAGDTLGARQCPAVRPVLLTAAVPRTGYASARGPPGVEARRCASPHAAGEDSAPGFQTAPGEPRGTPLPRPTMQNSHSGVNQ
B1	RAASFILKTNVTE-HHIFPLFYIWSLFIAK-PGLSQPIYQEAPTRGERSERGAALLQ-SIP-AAAPG
Pax-6	-RA-SRCVRS-fileVityAPSSDSHVFEPRGIPHUTSSS
AN	LCARRCYRPHILMUTRUBHEDSHIFTEPRCIAHEASS
Pax-(zf-a)	TKNFSEDTKAVGTMPQKEYYNRATWESGVASH

B 1

Fig. 2. Comparison of the Pax-QNR protein sequence extended in the 5-UTR with those from other species. Sequences were from mouse Pax-6 (3), human AN (19), and zebrafish Pax(zf-a) (20). Boxes, conserved amino acids; arrowhead, the putative initiation Met; asterisk, the first paired domain amino acid differing in the two quail cDNA clones as the result of the alternative splicing.





Fig. 3. Schematic diagram of the structural organization of the *Pax-QNR* locus. *A*, genomic organization of *Pax-QNR*. At *top*, the three overlapping genomic phages are aligned with respect to *Pax-QNR* locus. Only part of the λ Pax-QNR3 is presented. The exon/intron distribution is shown below. Black boxes, exons numbered from 0 to 12; exon 4a represents an alternative exon identified in murine cDNA (3); hatched box, B1-specific exon denoted α . Restriction enzymes used: *X*, *Xbal*; *B*, *BamH*1; *Hc*, *Hinc*11; *K*, *Kpn*1; *H*, *Hind*111; *E*, *Eco*R1. *B*, alternative splicing of the *Pax-QNR* gene. The assumed initiation codon of *Pax-QNR* is in exon 3, and the TAA stop codon is in exon 12. *Gray box*, paired domain; *hatched box*, homeodomain; *small black box*, the 6-amino acid insertion in the B1 clone, due to an alternative splice acceptor in exon 10.

exons extending over 17 kbp of cellular DNA (Fig. 3A). The MC29-QNR2 cDNA contains 13 exons, and the B1 cDNA contains 10 exons (Fig. 3B). We have mapped three noncoding exons (named exons 0, 1, and 2) in the 5' part of MC29-QNR2 cDNA, assuming that the Pax-QNR protein synthesized from the MC29-QNR2 mRNA starts with the first ATG in exon 3. Thus, exon 3 is expected to be the first coding exon of the Pax-QNR gene. The first B1 exon (exon α), devoid of an ATG to initiate the Pax-QNR ORF, ends with a classical splice donor consensus sequence (see Table 1 for exon/intron boundaries) and is fused with exon 4, the first paired domain exon. As a result of the splicing process, the first amino acid of the paired domain encoded by B1 cDNA is a glycine (Fig. 1B) instead of a serine, as in all cDNAs isolated to date (see Fig. 2).

The sequence of the genomic *Pax-QNR* gene shows that the paired domain is encoded by three exons (exons 4, 5, and 6) (Fig. 3B). The first exon (exon 4) encodes the first helical structure of the paired domain (amino acids 23–31), which plays a major role in the DNA binding activity of the paired protein (15). The two helix-turnhelix motifs found in the carboxy-terminal part of the paired domain (amino acids 80–89 and 94–106) (14) are encoded by the second paired exon (exon 5). The conserved 115–120 amino acids encoded by the third exon

(exon 6) have no defined function but probably are important because of their complete conservation among the different paired domains. Thus, each exon may encode a distinct functional motif in the paired domain. An insertion of 42 bp encoding 14 amino acids has been found in the Pax-6 paired domain of some murine cDNAs (3). This insertion is located at the end of the first paired exon, suggesting that these 42 bp arise from an alternative splicing event. Therefore, we examined the nucleotide sequence of the intron between Pax-QNR exons 4 and 5 (paired exons 1 and 2) for the presence of this alternative exon. A sequence homologous to these 42 bp (ACCCATGCAGATGCAAAAGTCCAAGTGCTGGA-CAATCAAAAC) was found 150 bp upstream of the fifth exon (named exon 4a; Fig. 3B). Two nucleotide differences with the murine sequence were found in this sequence (underlined letters). At the protein level, these resulted only in the change of a Gln residue into an Asp residue.

As for the paired domain, the homeodomain is also encoded by three exons (exons 7, 8, and 9) (Fig. 3B). The homeodomain contains a helix-turn-helix sequence specifying a DNA-binding motif (27). Of the three helix motifs, helix 1 and helix 2 are the less conserved. Helix 1 is encoded by the end of exon 7 and the beginning of exon 8. Exon 8 also encodes the second helix and the

	Intron		Exon		Intron
			Exon 0	CGCCGGACAG	gtaacg
	ctgtctccttcccag	GACCTCGGCA	Exon 1	CCTCTGCGAG	gtgagt
	tctgctttgtcctag	AGGGCCGCCG	Exon 2	CGGCTTCCAG	gcaagt
	gtttcgttgccgcag	ACCGCTCCAG	Exon 3	ATGCAGAACA	gtaagt
			Exon $\alpha(B1)$	GCAGCCCCAG	gtcggc
	atttttgtgttatag	ACCCATGCAG	Exon 4a		gtaagc
	ctgtgcttcccgcag	GTCACAGCGG	Exon 4	AATCCTGCAG	gtgaag
	cgcttccttatccag	GTGTCGAATG	Exon 5	CATACCCAGT	gtaagt
	teetttgeettgeag	GTGTCGTCGA	Exon 6	CCCGCACAAG	gtgaga
	ctctccttcccacag	ACGGCTGCCC	Exon 7	CTTGAGAAAG	gtgagc
	tacttctccttgcag	AGTTTGAGAG	Exon 8	AAGGATACAG	gtacgg
	acgctgtgcttccag	GTGTGGTTTT	Exon 9	ACCACCCCCG	gtaatg
	tctattctattatag	TTTCCTCTTT	Exon 10(B1)	GCCTATGCAA	gtaage
	cctctttcacatcag	GTTCCATGTT	Exon 10 (MC29-QNR2)	GCCTATGCAA	gtaagc
	gctctatctttgcag	CCCCCGGTAC	Exon 11	ACTTCCACAG	gtgagc
	tttcgg	GTCTCATTTC	Exon 12		
Consensus sequences	$SA \begin{pmatrix} t \\ c \end{pmatrix} n^{t} ag$	6 ^T G			gt ^a agt

beginning of the third helix. The third helix of the pairedtype homeodomain is highly conserved among all *Pax* genes, and the serine at position nine, which confers sequence-specific recognition to the homeodomain (10), is encoded by exon 9.

Exon 10 differs in MC29-QNR2 and B1 cDNAs. A sequence of 18 bp coding for 6 amino acids (Val-Ser-Ser-Phe-Thr-Ser) is present in clone B1 precisely at the junction between exons 9 and 10 of cDNA MC29-QNR2 (Table 1; Figs. 1C and 3B). All other sequences reported to date contained the B1-like exon 10. Analyses of the genomic sequences reveal that this difference between the two cDNAs comes from the use of an alternative splice acceptor in the Pax-QNR exon 10 (Table 1). To examine whether Pax-QNR mRNAs expressed in neuroretina cells contained one, the other, or both types of exon 10, total RNA extracted from QNR was retrotranscribed using oligo(dT) as a primer. The single-strand DNA molecules obtained were used as templates for PCR experiments using oligonucleotide primers corresponding to sequences present in exons 9 and 10. The resulting bands were subjected to Hpall restriction enzyme hydrolysis. This enzyme was chosen because, due to the splicing event, an additional Hpall site is present within the fragment resulting from the PCR amplification of MC29-QNR2 mRNA. After enzymatic digestion, we found both B1-like (Hpall uncut) and MC29-QNR2-like (Hpall cut) fragments, suggesting that neuroretina cells expressed both types of exon 10 (data not shown). Part of the protein encoded by exons 9, 10, 11, and 12 is particularly rich in proline, serine, and threonine amino acids. The last identified exon of the Pax-QNR gene (exon 12) is more than 1 kbp long. Since the EcoRI site present at the 3' end of the genomic DNA is also present at the end of the longest cDNA isolated, the exact size of the noncoding sequence in exon 12 is unknown.

Alternative Splicing Takes Place in the 5' Part of the Pax-QNR Gene in Vivo. Since B1 cDNA was found without any AUG to initiate the Pax-QNR open reading frame, we performed RNase protection experiments in order to look for the existence of exon α -containing mRNA in vivo. We subcloned a Hincll fragment encompassing the 5' part of the B1 cDNA (exon α and part of exon 4) into the plasmid pGEM4 (Fig. 4A). Similarly, we subcloned

the 5' EcoRI-HincII fragment of the MC29-QNR2 cDNA (containing the exons 0, 1, 2, 3, and part of 4) (Fig. 4B). Antisense RNA probes (637 nt long for B1 cDNA and 455 nt long for MC29-QNR2 cDNA) were synthesized and hybridized to total RNA extracted from various quail tissues or from QNR cells. After RNase digestion of the hybridized probes, the mRNAs corresponding to B1 and MC29-QNR2 cDNAs could be detected by full-length protected fragments of 490 and 424 nt, respectively. Results obtained show that a mRNA containing exon α (Fig. 4A, arrow) is detected in neuroretinas (MC29 infected or not) and to a lesser extent in the cerebellum, but not in heart or brain (Fig. 4A). Although a similar tissue distribution is observed with the MC29-QNR2 mRNA, surprisingly, very little of this mRNA is detected in neuroretinas (Fig. 4B). However, with this probe, we detect several partially protected fragments (Fig. 4B, arrows). These fragments correspond to alternatively spliced Pax-QNR mRNAs that differ in their 5'-UTR (see Discussion").

Pax-QNR Encodes a DNA-binding Protein. In order to analyze the size of the protein encoded by MC29-QNR2 cDNA, *in vitro* synthesized MC29-QNR2 RNA was translated in rabbit reticulocyte lysate in the presence of [³⁵S] methionine. Results revealed a major labeled protein present in a doublet band of 46/47 kd specific for the MC29-QNR2 cDNA (Fig. 5A, Lane b). The apparent molecular weight of 46,000 for Pax-QNR protein in a SDSpolyacrylamide gel is consistent with the use of the first ATG in the cDNA to translate a protein of 416 amino acids with a predicted molecular weight of 45,800.

Drosophila paired protein has been shown to bind and footprint *in vitro* a region upstream of the Drosophila even-skipped gene, known as the e5 sequence (15). The 5' part of this sequence, containing the ATTA motif, binds the homeodomain of the protein, whereas the 3' part of the sequence binds the paired domain (15). The e5 sequence is also recognized by the murine Pax-2 (22), Pax-5 (23) (each containing only a paired domain), and Pax-3 gene products (21). To determine whether Pax-QNR (which differs in many positions from the other paired domain and homeodomain proteins) was able to specifically recognize the e5 sequence, EMSAs were performed using Pax-QNR protein synthesized *in vitro*



Fig. 4. RNase protection experiments. A, top: open box, B1 fragment subcloned in pGEM4 plasmid; bold line, linker sequence. The size of the expected RNase-protected fragment is indicated. Bottom: the probe was hybridized to 10 μ g of total RNA. Yeast tRNA was used as a control of RNase digestion; heart, 3-week-old quails; cerebellum and brain, 14-dayold quail embryos; QNR E7 and E14, neuroretinas from 7- and 14-dayold quail embryos, respectively; QNR MC29, quail neuroretina cells infected by MC29 virus; M, end-labeled Hinf-digested pBR322. B, same experiment as in A, but with a RNA probe derived from an MC29-QNR2 CDNA fragment. Arrows, the protected fragments.

and ³²P-labeled e5 DNA (Fig. 5*B*). Very little endogenous e5 binding activity was found in reticulocyte lysates containing the antisense Pax-QNR RNA (*Lane 2*). Expression of Pax-QNR protein resulted in an e5 DNA-binding complex (Lane 3). This complex was specific for the e5 sequence, since it disappeared in the presence of a 100fold excess of unlabeled e5 oligonucleotide (Lane 4). In contrast, a 100-fold excess of an oligonucleotide containing the AP-1 consensus recognition sequence failed to compete with the e5 sequence for Pax-QNR binding (Lane 5). To determine the nature of the Pax-ONR protein interaction with the e5 sequence, we performed mutational analysis of the homeo (e5P) or the paired (e5H) recognition sites. When each of the mutated oligonucleotides, e5P and e5H, was used as competitor, both competed for Pax-QNR binding on the e5 sequence (Lanes 6 and 7). These mutated oligonucleotides were then tested for binding to Pax-QNR protein. The oligonucleotide mutated in the paired binding site (e5H) produced a complex with Pax-QNR as efficiently as the e5 oligonucleotide (Lane 9). e5H binding was efficiently competed with a 100-fold excess of unlabeled e5 (Lane 10) or e5H (Lane 11) oligonucleotides. The oligonucleotide mutated in the homeodomain binding site (e5P) exhibited only negligible binding with Pax-QNR (Lane 14) and weakly competed with the e5H sequence for Pax-QNR binding (Lane 12).

Both the Paired Domain and the Homeodomain of Pax-QNR Are Individually Able to Bind DNA. The results obtained with the mutated e5 oligonucleotides suggested an influence of both the paired domain and the homeodomain on the DNA binding ability of the Pax-QNR protein. To ascertain the DNA binding ability of each domain, we expressed separately the paired domain and the homeodomain in bacteria and tested the DNA binding properties of each gel-purified peptide in EMSA (Fig. 5C). Results show that the paired peptide binds strongly to e5 (Lane 1), and this binding is competed with a 100fold excess of unlabeled e5 oligonucleotide (Lane 2). However, in contrast to the full Pax-QNR protein, the paired peptide binds efficiently to e5P (e5 oligonucleotide mutated in the homeodomain binding site; Lane 3) and very weakly to e5H (e5 oligonucleotide mutated in the paired domain binding site; data not shown). The homeodomain also binds to e5 (Lane 5), and this binding is competed with a 100-fold excess of unlabeled e5 (Lane 6) or e5H (Lane 8) oligonucleotides but is not competed with a 100-fold excess of unlabeled e5P (Lane 9) or AP-1 (Lane 7) oligonucleotides. As expected, the homeodomain peptide binds efficiently to e5H (Lane 10), and this binding is competed with e5H (Lane 11) but not with e5P (Lane 12) oligonucleotides. As a control, the carboxyterminal part of Pax-QNR protein was unable to bind the e5 oligonucleotide (Lane 13), demonstrating that the 99 MS2 amino acids fused with each of the peptide domains are devoid of DNA binding properties. Thus, the paired domain and the homeodomain are individually able to bind specifically to the e5 oligonucleotide.

Discussion

Genomic Organization of Pax-QNR**.** We have previously described the isolation of Pax-QNR (4), the quail homologue of the gene recently described as Pax(zf-a) in zebrafish, Pax-6 in mouse, and AN in humans. We have determined the exon-intron organization of this gene. The 5'-UTR of Pax-QNR appears variable, since we have isolated two cDNAs showing distinct 5'-UTRs. One of these cDNAs, clone B1, does not contain any AUG to

286 Structure and Properties of Pax-QNR



initiate the Pax-QNR protein. However, we believe that this cDNA is not an artifact because the nucleotide sequence of the first B1 exon (exon α) is identical in the cDNA and the genomic DNA, and because RNase protection experiments using a probe containing this exon demonstrate the presence of this latter sequence in mRNAs of neuroretinas. Moreover, experiments show that the MC29-QNR2 and B1 mRNAs are transcribed from two distinct promoters located, respectively, upstream from exons 0 and α . Recent reports suggest that non-AUG translation initiation occurs much more frequently in eukaryotic cells than was previously anticipated (26). In vitro translation in a reticulocyte lysate of B1 RNA indicates that this RNA is able to direct the synthesis of a Pax-QNR product (data not shown). However, the codon used to initiate this protein has not yet been defined. Further studies will be necessary to address this issue. We do not know whether other species also contain exon α or whether this particular transcript is avian specific. The previously isolated cDNA, MC29-QNR2, shows one in-frame AUG translation start site, which places the paired domain very close to the aminoterminus of the protein. Analysis of the sequence revealed that the open reading frame extended upstream from the first AUG. The conservation of this extension in all of the species studied (fish, guail, mouse, and human) suggests a possible role for the upstream part of the open reading frame, but only mouse and human cDNAs showed clear homology in the protein deduced from the extended reading frame. This finding implies that this part does not play a critical role in the biological function, or that the function played by this part of Pax-QNR is species specific. However there is no AUG codon that could be used to translate this extended open reading frame, a situation already found for other transcription factors (28). The 5'-UTR of MC29-QNR2 cDNA is organized in 4 exons (exons 0, 1, 2, and 3, the latter containing the initiation AUG). RNase protection experiments performed on neuroretina RNA using this 5'-UTR as a probe reveal the expected fully protected band, as well as several partially protected bands, the sizes of which could correspond to mRNAs generated from alternative splicing events arising in the 5' part of the Pax-QNR gene. This assumption is confirmed by the isolation of two other cDNAs, one showing an exon 0-exon 2 junction, and the other containing a 5' extended exon 1. The mouse and the human cDNAs also exhibited multiple 5'-UTRs (3, 19), suggesting a complex translational regulation of this gene in these species.

The paired box domain of *Pax-QNR* is encoded by three exons, like all other known murine paired boxes

Fig. 5. A, in vitro translated Pax-QNR protein. The MC29-QNR2 cDNA was transcribed in vitro in a sense and antisense orientation and translated in a rabbit reticulocyte lysate. Translated proteins were separated on a 10% acrylamide gel. Lane a, antisense RNA; Lane b, MC29-QNR2 sense RNA. B, Pax-QNR binding to the even-skipped e5 sequence. Reticulocyte lysate expressing the Pax-QNR protein (Retic) and ³²P-labeled double-stranded oligonucleotide were used for each EMSA reaction. s, sense; as, antisense. e5H, e5 oligonucleotide mutated in the paired binding site. Unlabeled competitor DNA was used in a 100-fold molar excess. C, binding of bacterially expressed paired domain and homeodomain peptides to e5 and mutated oligonucleotides. The peptides encoded by the paired, homeo-, and carboxy-terminal part (Ct) of the Pax-QNR protein were purified after electrophoresis of the bacterial proteins on a 15% SDS-polyacrylamide gel and used as in B.

with the exception of *Pax*-1. (14). The intron-exon boundaries are also conserved in the paired box domain between quail and mouse. The murine *Pax*-6 cDNA previously described (3) shows an insertion of 14 amino acids between the first and the second exon of the paired box domain with respect to the *Pax-QNR* sequence. We found a highly homologous sequence in the first paired box intron of *Pax-QNR* locus (between exons 4 and 5), and this sequence exhibits a classical exon-intron boundary. Therefore, we called this sequence exon 4a. However, we have not isolated any quail cDNA containing this exon. The genomic structure of the human *Pax*-6 gene exhibited similar organization (29).

The homeodomain of *Pax-QNR* is encoded by three exons. Thus, the *Pax-QNR* homeobox is distinct from *Pax-3* and *Pax-7* genes at the structural level, since no intron in the homeodomain has been reported for these genes (21, 30). Nucleotide sequence comparison of the two cDNAs isolated reveals that an alternative splicing event takes place in the exon following the homeodomain (exon 10). Indeed, a stretch of six amino acids (Val-Ser-Ser-Phe-Thr-Ser) was added in the protein encoded by cDNA B1. Serine and threonine amino acids are potential targets for cellular kinases, and phosphorylation events may modulate the activity of the two Pax-QNR products. Further studies will be necessary to understand the biological relevance of this particular splicing event in exon 10.

DNA-binding Properties of Pax-QNR Product. The Pax-QNR protein is able to bind to the Drosophila e5 sequence like Pax-2, Pax-3, and Pax-5 proteins (21-23), and this binding is most probably mediated by the paired domain and the homeodomain, as it as been demonstrated for the Pax-3 protein (21). From their degree of homology in conjunction with class-specific amino acids at certain positions, the paired domains of the Pax proteins fall into six different classes (14). Three distinct binding specificities of paired domains can be defined by the following DNA sequences: e5 (bound by Prd, Pax-3, Pax-2, and Pax-5), PRS-1 (bound by Pax-1), and CT (bound by Pax-8) (25). Prd and Pax-3 fall into class II, but Pax-2, Pax-5, and Pax-8 fall into class III, although showing a distinct DNA-binding recognition sequence. Pax-QNR/Pax-6 is the first member of the class IV-type paired domain, and we found that an efficient binding on e5 is obtained with Pax-QNR protein expressed in reticulocyte lysates. These data show that Pax proteins of different subclasses exhibit different sequence-binding specificities, due to amino acid changes in their paired domains, and that distinct sequence-binding specificities can be found in the same class as well as identical sequencebinding specificities in distinct classes. Therefore, the degenerated sequence recognition emphasizes the limited value of consensus sequences and core motifs for Pax proteins.

The fact that an efficient binding on e5 is obtained with Pax-QNR protein expressed in reticulocyte lysates suggests that Pax-QNR binding does not require the presence of additional proteins; similar results have been obtained for the Pax-5 protein (25). However, we have been unable to obtain DNA binding of the protein encoded by cDNA B1, translated in reticulocyte lysates, using e5 as probe (data not shown). This suggests that some modifications affecting the paired domain may be critical for the binding of this particular Pax-QNR protein isoform or that accessory proteins are needed for the DNA binding activity of this B1 Pax-QNR.

Modification of the homeodomain binding site in e5 results in a mutated oligonucleotide (e5P) unable to bind Pax-QNR, but able to compete with wild-type e5 for Pax-QNR binding. This indicates that Pax-QNR requires both homeodomains and paired domains for high-affinity binding, and that binding by the paired domain alone results in the formation of a transient complex, unstable under EMSA conditions. Modification of the paired domain binding site in e5 results in a mutated oligonucleotide (e5H), able to bind Pax-QNR. Thus, paired domains and homeodomains of Pax-QNR proteins exhibit distinct affinity for the e5 sequence. Such a situation also exists for another paired domain- and homeodomain-containing Pax gene, namely Pax-3. However, for Pax-3, oligonucleotides containing the paired domain recognition sequence alone exhibited weak binding, whereas the homeodomain recognition sequence alone gave no specific complexes with the Pax-3 protein (21).

That both domains are involved in Pax-QNR DNA binding is further indicated by the fact that paired domains and homeodomains individually expressed in bacteria are able to bind to e5. With bacterially expressed peptides, the e5P mutated oligonucleotide, unable to form stable complexes with Pax-QNR, forms stable complexes with the paired domain peptide, suggesting that conformational events involving other part of the Pax-QNR protein are involved in the DNA binding activity of the protein. Such interaction was also demonstrated for the Drosophila Prd protein (10, 15), for which the homeodomain binding appears to be masked by the COOHterminal part of the protein. In addition, the inhibitory effect of the COOH-terminal part of the protein appears to be sequence specific (15). This COOH-terminal modulation is not a general regulation of Pax gene DNA binding, since Pax-5 DNA binding is not modulated by this part of the protein (25). However, Pax-5 is devoid of homeodomain, and this particular type of DNA binding regulation may be specific to paired domain and homeodomain-containing proteins.

The core sequence of the Pax-1, Pax-2, and Prd paired domain binding site, TTCC, is also the core sequence of the binding site of the *ets* gene family products (31). We observed that certain types of *ets* target sequences, devoid of any ATTA homeobox binding sites, can be recognized by the MC29-QNR2 Pax-QNR protein, and that a stable complex can be formed under conditions in which the e5P oligonucleotide (also devoid of ATTA) does not allow the formation of a stable complex. Thus, the Pax-QNR binding on e5 demonstrates a cooperation between the two DNA binding domains in the protein, but the type of the sequence recognized is also involved in the stabilization of the DNA-protein complex.

Presently, it is unclear whether the e5 sequence is a target for Pax-QNR binding *in vivo* and whether this sequence will be instrumental in defining Pax-QNR as a transcription factor. Experiments performed with Pax-3 indicate that efficient binding on e5 is not correlated with any activation of transcription (21).

Materials and Methods

Cells and Viruses. NRs were dissected from 7-day-old quail embryos. QNR cultures from 7-day-old embryonic

. . .

tissue were maintained and passaged in Eagle's basal medium supplemented with 10% fetal calf serum. MC29 RAV-1 virus was obtained by cotransfection of DNA from the molecular clone pMC38 containing the myelocytomatosis provirus and Rous associated virus type 1 DNA into quail embryo cells (4).

Isolation of Quail Clones Containing Pax-QNR Homologous Sequences. Using MC29-QNR2 cDNA as probe, clone B1 was isolated from a cDNA library constructed in the pCDM8 vector from 15-day-old quail spinal cords (kindly supplied by C. Dulac). Clone λ Pax-QNR1 was isolated from a genomic library constructed from peripheral blood DNA of quail embryos in the EMBL4 vector, screened with the MC29-QNR1 cDNA probe (4). Clones λ Pax-QNR2 and λ Pax-QNR3 were isolated from a genomic library constructed from 14-dayold quail embryonic tissue DNA in the λ DASH vector, using as probes a 1.5-kbp BamHI genomic fragment containing exon 1 and a 94-bp EcoRI-HaeIII fragment representing the 5' part of MC29-QNR2 cDNA, respectively.

DNA Sequencing. Appropriate subclones of genomic DNA as well as B1 cDNA were ligated into M13mp18 and M13mp19. Nucleotide sequence was determined in both orientations by the dideoxynucleotide chain-termination method using standard techniques and an Applied Biosystem 370A automatic sequencer. All sequence homology searches were done with software from Bisance (32).

RNase Protection Assays. To obtain a B1-specific RNA probe, a 572-bp Hincll fragment containing 490 bp of cDNA B1 and 82 bp of linker sequences was cloned into the HincII site of the pGEM4 plasmid. The antisense RNA probe, 637 nt long, was transcribed in vitro from the HindIII-linearized plasmid using SP6 polymerase in the presence of $[\alpha^{-32}P]$ CTP under conditions provided by the riboprobe system (Promega). To obtain an MC29-QNR2-specific RNA probe, a 424-bp EcoRI-Hincl1 5' cDNA fragment was cloned into the EcoRI-HincII sites of the pGEM4 plasmid. The antisense RNA probe, 455 nt long, was transcribed in vitro from the EcoRI-linearized plasmid using T7 polymerase under the same conditions as described above. RNase protection assays were performed with 10 μ g of total quail RNA extracted from MC29-transformed neuroretina cells, 14-day-old embryonic tissues (cerebellum, brain, and QNR), QNR from 7day-old embryonic tissue, and heart from 3-week-old quails, as described previously (33). As a negative control, RNase protection assays were performed on yeast tRNA.

Expression of Paired, Homeo-, and Carboxy-terminal Domains of Pax-QNR in Bacteria. The appropriate fragments were obtained from the MC29-QNR2 plasmid used as template in PCR experiments using oligonucleotides corresponding to sequences present in paired, homeo-, and carboxy domains of Pax-QNR (the sequences of the oligonucleotides used are available upon request). The expected bands were subjected to BamHI-HindIII restriction enzyme digestion and inserted into the BamHI-HindIII sites of the pPLc24 vector. The subcloning resulted in the in-frame fusion of the desired sequence to the first 99 amino acids of the polymerase of phage MS2. For expression, plasmids were transferred into an Escherichia coli host which has a temperature-sensitive repressor of the PL promoter. Cultures of exponentially growing bacteria that carried 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-polyacrylamide gel. Proteins were visualized by staining with Coomassie brilliant blue and cut from the gel for purification.

In Vitro Transcription, Translation, and EMSA. Plasmids pSP64 and pSP65 MC29-QNR2 were linearized using the *Pvull* restriction site, and the sense (pSP65) and antisense (pSP64) RNA strands were synthesized with SP6 RNA polymerase. Half (5 μ l) of the RNA template was translated in a rabbit reticulocyte lysate. All reactions were performed according the manufacturer's specifications (BRL).

To prepare the DNA probes for gel shift analysis, an oligonucleotide containing the e5 binding site CTCAGCACCGCACGATTAGCACCGTTCCGCTTC and its complementary strand were synthesized, annealed, and used as DNA probe after 5'-end labeling. The recognition sequence bound by the homeodomain is underlined, and the paired domain recognition sequence is shown in boldface. Two other sets of oligonucleotides, e5P with mutations (bold letters) in the homeodomain binding site CTCAGCACCGCACGCGCTGCACCGTT-CCGCTTC, and e5H with mutations in the paired binding site CTCAGCACCGCACGATTAGCACCGTTT-GCTTC and their complementary strands, were synthesized, annealed, and used for the competition experiments.

One-fiftieth of the reticulocyte lysate (1 μ l) or 2 μ g of the gel-purified, bacterially expressed proteins were mixed with 1 ng of the DNA probe and ³²P labeled with T4 polynucleotide kinase in binding buffer [10% glycerol, 10 mм 4-(2-hydroxy-ethyl)-1-piperazineethanesulfonic acid (pH 7.9), 30 mм KCl, 0.5 mм dithiothreitol, 4 mм spermidine, 1.5 μ g polydeoxyinosinic-deoxycytidylic acid, 1 μ g salmon sperm DNA, protease inhibitors (0.5 mm phenylmethylsulfonyl fluoride, 1 μ g/ μ l each of leupeptin, antipain, and pepstatin, and 2 mм benzamidine] and phosphatase inhibitors (10 mm β -glycerophosphate, 2 mm levamisole, and 10 µм orthovanadate). Reactions were performed in 16 μ l final volume for 10 min on ice. DNA-protein complexes were analyzed by electrophoresis on 6% polyacrylamide gels (Tris-borate-EDTA buffer) at room temperature for 2 h (10 V/cm). The gels were dried and exposed to film.

Nucleotide Sequence Accession Number. The nucleotide sequence data reported in this paper will appear in the EMBL nucleotide sequence databases under the accession numbers X68168 and X68169.

Acknowledgments

We thank Kathryn Nason-Burchenal for critical reading of the manuscript.

References

1. Okada, T. S. *In:* J. D. Ebert, and M. Marois (eds.), Tests of Teratogenicity *in Vitro*, pp. 91-105. Amsterdam: Elsevier, 1976.

2. Monaghan, P., Davidson, D. R., Sime, C., Graham, E., Baldock, R., Bhattacharya, S., and Hill, R. The Msh-like homeobox genes define domains in the developing vertebrate eye. Development, *112*: 1053-1061, 1991.

 Walther, C., and Gruss, P. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development, 113: 1435-1449, 1991.
 Martin, P., Carrière, C., Dozier, C., Quatannens, B., Mirabel, M-A., Vandenburder, B., Cichelin, D., and Saula S., Characteristic of a pairof.

Vandenbunder, B., Stéhelin, D., and Saule, S. Characterization of a paired box- and homeobox-containing quail gene (Pax-QNR) expressed in the neuroretina. Oncogene, 7: 1721-1728, 1992. 5. Akam, M. The molecular basis for metameric pattern in the *Drosophila* embryo. Development, 101: 1–22, 1987.

6. Ingham, P. W. The molecular genetics of embryonic pattern formation in *Drosophila*. Nature (Lond.), 335: 25-34, 1988.

7. Izpisua-Belmonte, J-C., Tickle, C., Dolle, P., Wolpert, L., and Duboule, D. Expression of the homeobox *Hox-4* genes and the specification of position in chick wing development. Nature (Lond.), *350*: 585–589, 1991.

8. Murphy, P., Davidson, D. R., and Hill, R. E. Segment-specific expression of a homeobox-containing gene in the mouse hindbrain. Nature (Lond.), *341:* 156–159, 1989.

9. Tabin, C. J. Retinoids, homeoboxes, and growth factors; toward molecular models for limb development. Cell, 66: 199-217, 1991.

10. Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., and Desplan, C. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell, *59*: 553-562, 1989.

11. Ingraham, H. A., Chen, R., Mangaiam, H., Elsholtz, S., Flynn, C. R., Lin, D. M., Simmons, L., Swanson, L., and Rosenfeld, M. G. A tissuespecific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell, *55*: 519–529, 1988.

12. Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R., and Cortese, R. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. Cell, *59*: 145-157, 1989.

13. Karlson, O., Thor, S., Norberg, T., Ohlsson, H., and Edlund, T. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature (Lond.), *344*: 879–882, 1990.

14. Walther, C., Guenet, J-L., Simon, D., Deutch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R., and Gruss, P. Pax: a murine multigene family of paired box-containing genes. Genomics, *11*: 424–434, 1991.

15. Treisman, J., Harris, E., and Desplan, C. The paired box encodes a second DNA-binding domain in the paired homeo domain protein. Genes & Dev., *5*: 594–604, 1991.

16. Kessel, M., and Gruss, P. Murine developmental control genes. Science (Washington DC), 249: 374-379, 1990.

17. Hastie, N. Pax in our time. Dev. Genet., 1: 342-344, 1991.

18. 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.

19. Ton, C. C., Hirvonen, H., Miwa, H., Weil, M., Monaghan, P., Jordan, T., Van Heyningen, V., Hastie, N., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L. C., and Saunders, G. F. Positional cloning and characterization of a paired-box- and homeobox-containing gene from the aniridia region. Cell, *67*: 1059-1074, 1991.

20. Krauss, S., Johansen, T., Korzh, V., Moens, U., Ericson, J. U., and Fjose, A. Zebrafish *pax[zf-a]*: a paired box-containing gene expressed in the neural tube. EMBO J., *10*: 3609-3619, 1991.

21. Goulding, M. D., Chalepakis, G., Deutch, U., Erselius, J. R., and Gruss, P. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO J., *10*: 1135-1147, 1991.

22. Dressler, G. R., and Douglass, E. Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor. Proc. Natl. Acad. Sci. USA, 89: 1179–1183, 1992.

23. Adams, B., Dörfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I., and Busslinger, M. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. Genes & Dev., 6: 1589-1607, 1992.

24. Chalepakis, G., Fritsch, R., Fickenscher, H., Deutch, U., Goulding, M., and Gruss, P. The molecular basis of the *undulated/Pax-1* mutation. Cell, 66: 873-884, 1991.

25. Zannini, M., Francis-Lang, H., Plachov, D., and Di Lauro, R. 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, 1992.

26. Kozack, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA, *87*: 8301-8305, 1990.

27. Kissinger, C., Liu, B., Martin-Blanco, E., Kornberg, T., and Pabo, C. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. Cell, 63: 579–590, 1990.

28. Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R., and Charnay, P. The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptional activator. Mol. Cell. Biol., *10*: 3456–3467, 1990.

29. Glaser, T., Welton, D. S., and Maas, R. Genomic structure, evolutionary conservation and aniridia mutations in the human Pax-6 gene. Nat. Genet., 2: 232-239, 1992.

30. Jostes, B., Walther, C., and Gruss, P. The murine paired box gene, Pax 7, is expressed specifically during the development of the nervous and muscular system. Mech. Dev., 33: 27-38, 1991.

31. Macleod, K., Leprince, D., and Stéhelin, D. The ets gene family. Trends Biochem. Sci., 17: 251-256, 1992.

32. Dessen, P., Fondrat, C., Valencien, C., and Mugnier, C. Comput. Appl. Biosci., 6: 355–356, 1990.

33. Dozier, C., Ansieau, S., Ferreira, E., Coll, J., and Stéhelin, D. An alternatively spliced *c-mil/raí* mRNA is predominantly expressed in chicken muscular tissues and conserved among vertebrate species. Oncogene, 6: 1307–1311, 1991.

ARTICLE III

Characterization of Quail Pax-6 (Pax-QNR) Proteins expressed in the Neuroretina. Catherine Carrière, Serge Plaza, Patrick Martin, Brigitte Quatannens, Manuella Bailly,

Dominique Stéhelin & Simon Saule

Mol. Cell. Biol., 13, 7257-7266, 1993
Caractérisation des protéines codées par *Pax-6* (*Pax-QNR*) et exprimées dans la neurorétine.

Afin de pouvoir caractériser les protéines du gène *Pax-QNR*, nous avons produit des anticorps dirigés contre les différents domaines de la p46: contre le domaine PRD (sérum 11), contre la jonction entre le domaine PRD et l'homéodomaine (sérum 12), contre l'homéodomaine (sérum 13) et contre la partie carboxy-terminale (sérum 14). Ces sera nous ont permis de mettre en évidence au moins cinq protéines produites *in vivo* dans la neurorétine de caille. Trois protéines de 48, 46 et 43 KDa sont reconnues par les sera 11, 12, 13 et 14 et au moins deux protéines de 33 et 32 KDa ne sont reconnues que par les sera 12, 13 et 14 (**Fig.2 A et B**). Ces dernières protéines sont donc dépourvues de domaine PRD. Le fait qu'elles soient produites également lors de synthèse *in vitro* en lysat de réticulocytes à partir du clone MC29QNR2 (**Fig. 2D**), suggère qu'elles sont générées par des initiations internes en aval du domaine PRD. En effet, la mutagenèse dirigée transformant les ATG situés en aval du domaine PRD en ATC, supprime la production du doublet de protéines, ce qui confirme l'hypothèse que ces p33-32 sont produites par initiation interne (**Fig. 2E**).

Afin de contrôler que les cinq protéines caractérisées *in vivo* étaient toutes codées par le gène *Pax-QNR*, une cartographie à la protéase V8 de ces cinq protéines a été faite en parallèle avec celle des p46 et p33-32 produites à partir de l'ADNc MC29QNR2 en lysat de réticulocytes. Les cartes de digestion des différentes protéines sont superposables (**Fig. 3**). Ils ne s'agit donc pas de protéine reconnue par réaction croisée avec les anticorps ou suite à une association avec les produits de *Pax-QNR*. Les cinq protéines sont produites de façon indépendantes et ne proviennent pas d'événement de dégradation ou de maturation post-traductionnelle. Elles ont toutes la même durée de demi vie, de l'ordre de 6 heures, ce qui est particulièrement élevé pour un facteur de transcription (**Fig. 6**). Il n'y a pas de consensus de glycosylation dans la séquence peptidique et un traitement à la phosphatase ne modifie pas la taille des protéines, ce qui montre que les p48, 46 et 43 ne sont pas des formes différentiellement phosphorylées de la même protéine (résultats non présentés). Des événements de phosphorylation peuvent modifier l'activité des facteurs de transcription que ce soit au niveau de la translocation dans le noyau, de la fixation à l'ADN ou du contrôle de la transcription (voir la revue de Hunter et Karin, 1992).

Des expériences de marquage à l'orthophosphate, montrent que ces protéines sont phosphorylées *in vivo* sur des Ser et Thr (Fig. 5A et 5B) et des séquences consensus de phosphorylation pour la caséine kinase II et les MAP ("Mitosis associated proteins") kinases sont présentes dans la protéine.

L'ADNc que nous avons isolé initialement code une p46. Nous avons donc recherché les ADNc pouvant coder les p48 et p43. Par amplification en chaîne (PCR) avec des oligonucléotides encadrant le domaine PRD sur une rétrotranscription d'ARNm de neurorétine à E8, nous avons isolé un ADNc partiel plus court que le fragment normalement amplifié sur le clone MC29QNR2. La séquence de cet ADNc a montré qu'il contenait un épissage alternatif qui supprime partiellement l'exon 5 dans le domaine PRD (**Fig. 4A**). Ce fragment d'ADNc cloné en phase avec le reste de la séquence codante de MC29QNR2 code une protéine de 43 KDa (**Fig. 4B**). Enfin, la p48 pourrait être codée par un messager contenant l'exon 4a (équivalent aviaire de la forme publiée de Pax-6 chez la souris).

Les produits des gènes Pax codent des facteurs de transcription et la localisation subcellulaire de ceux-ci est essentielle à leurs fonctions. Les protéines codées par *Pax-QNR* diffèrent dans leurs localisations subcellulaires: des expériences de fractionnement cytoplasmenoyau montrent que les p48 et p46 sont nucléaires alors que les p43 et p33-32 sont essentiellement cytoplasmiques (**Fig.7B**). Cependant, transfectés en fibroblastes embryonnaires de caille ou en cellules COS, si l'ADNc MC29QNR2 produit des protéines localisées uniquement dans le noyau comme le montrent des expériences d'immunofluorescence réalisées avec le sérum 14, l'ADNc Δ Nhe qui code seulement les p33-32, produit des protéines ayant une localisation nucléaire et cytoplasmique (**Fig.7A**). Les protéines nucléaires générées par l'ADNc Δ Nhe sont retrouvées majoritairement dans le cytoplasme lors du fractionnement subcellulaire, ce qui montre que cette localisation est moins stable pour les protéines dépourvues de domaine PRD. Des expériences d'immunoprécipitation montrent que si la synthèse de grandes protéines se fait de façon constante de E5 à E15, celle des p33-32 chute entre ces mêmes stades (**Fig.8**).

Enfin, des expériences de fixation à l'ADN montrent qu'au moins une protéine produite *in vivo* est capable de reconnaître la séquence e5 (**Fig.9**). Après migration, le complexe ADN-protéine se trouve au même niveau que la p46 fixée à e5. De plus, ce complexe est reconnu par

le sérum 14, ce qui montre que *in vivo*, seule la p46 est capable de se lier à la séquence e5 (Fig.10).

Pax-QNR génère par épissage alternatif de nombreuses protéines dont la fonction reste à définir. Les protéines dépourvues de domaine PRD, bien que ne fixant pas la séquence e5, conservent un domaine de liaison à l'ADN, ce qui suggère qu'elles pourraient avoir des cibles différentes.

Characterization of Quail Pax-6 (Pax-QNR) Proteins Expressed in the Neuroretina

CATHERINE CARRIERE,¹ SERGE PLAZA,¹ PATRICK MARTIN,² BRIGITTE QUATANNENS,² MANUELLA BAILLY,¹ DOMINIQUE STEHELIN,² AND SIMON SAULE^{1*}

> CNRS EP 56¹ and CNRS URA 1160,² Institut Pasteur de Lille, 1 Rue Calmette, 59019 Lille Cedex, France

Received 8 February 1993/Returned for modification 25 March 1993/Accepted 31 August 1993

After differential screening of a cDNA library constructed from quail neuroretina cells (QNR) infected with the v-myc-containing avian retrovirus MC29, we have isolated a cDNA clone, Pax-QNR, homologous to the murine Pax-6, which is mutated in the autosomal dominant mutation small eye of mice and in the disorder aniridia in humans. Here we report the characterization of the Pax-QNR proteins expressed in the avian neuroretina. From bacterially expressed Pax-QNR peptides, we obtained rabbit antisera directed against different domains of the protein: paired domain (serum 11), domain between the paired domain and homeodomain (serum 12), homeodomain (serum 13), and carboxyl-terminal part (serum 14). Sera 12, 13, and 14 were able to specifically recognize five proteins (48, 46, 43, 33, and 32 kDa) in the neuroretina. In contrast to proteins of 48, 46, and 43 kDa, proteins of 33 and 32 kDa were not recognized by the paired antiserum (serum 11). Paired-less and paired-containing proteins exhibited the same half-life (6 h) and were phosphorylated mostly on serine residues. Immunoprecipitations performed with subcellular fractions of neuroretinas showed that the paired-containing proteins were located in the nucleus, whereas the 33- and 32-kDa proteins were found essentially in the cytoplasmic compartment. However, immunofluorescence experiments performed after transient transfections showed that p46 and p33/32 were also located in vivo into the nucleus. Thus, the Pax-QNR/Pax-6 gene can produce proteins with two DNA-binding domains as well as proteins containing only the DNA-binding homeodomain.

Cellular differentiation is the result of differential gene expression, and transcription factors are involved in the regulation of these events. Homeobox-containing genes encode developmentally regulated transcription factors. The homeobox encodes the homeodomain, a conserved DNAbinding domain with a helix-turn-helix motif (23), and several vertebrate homeodomain-containing proteins have been identified as transcription factors required for expression of lineage-specific genes (9, 19, 20).

The paired motif found in Drosophila developmental genes such as paired and gooseberry has also been shown to be conserved in vertebrates and found to encode another DNA-binding motif (32). This motif is present alone (in the murine 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 [1, 34]). In contrast to homeobox-containing genes whose expression is region specific, the Pax genes are expressed along the entire anteroposterior axis in the neural tube (22). Severe developmental alterations in the mouse have been associated with mutations in the Pax genes (13). Mutations in the Pax-6 gene have been associated with the mouse mutation small eye (14). The same gene has been found deleted in some cases of the human congenital disorder aniridia (30), confirming the importance of *Pax-6* in eye development. This gene is expressed in the optic cup, lens, and overlaying epithelium prior to morphological differentiation and later in the neuronal layer of the retina. Vertebrate eye development is mediated by a series of inductive interactions controlled in part by transcription factors (29). Other developmentally regulated transcription factors are also expressed in the developing eye. This is the case of msh-like

7257

homeobox gene Hox-8.1, expressed throughout the presumptive neural retina and ectoderm overlaying the optic cup (27), and another Pax gene, Pax-2, expressed in the ventral half of the developing eye and optic stalk (28).

We have previously isolated a cDNA clone from quail neuroretina (QNR) which has been named Pax-QNR and appears to be the quail homolog of Pax-6 (24, 33). To examine the potential biological role of Pax-QNR protein in the neuroretina, we have characterized and identified Pax-QNR proteins in vivo. We found that the avian Pax-QNR gene expresses five proteins recognized by rabbit antiserum directed against bacterially expressed Pax-QNR peptides. Three proteins of 48, 46, and 43 kDa were recognized by different antisera directed against the paired domain (serum 11), the region located between the paired domain and the homeodomain (serum 12), the homeodomain (serum 13), and the carboxyl-terminal part of the protein (serum 14). Two proteins of 33 and 32 kDa were not recognized by serum 11 but were immunoprecipitated by sera 12, 13, and 14. The amounts of these paired-less proteins decreased with the neuroretina differentiation. Paired-containing and pairedless proteins are phosphorylated in vivo and exhibit similar half-lives (6 h). Using immunofluorescence experiments, we found that paired-containing and paired-less proteins were located in the nucleus. The 46-kDa Pax-QNR protein produced in reticulocyte lysates bind specifically to the e5 sequence present upstream of the Drosophila even-skipped gene (8). Using UV cross-linking in solution with 5-bromodUTP (5-BrdUTP)-labeled e5 probe and neuroretina nuclear extracts followed by immunoprecipitation of DNA-protein complexes, we demonstrated that the only Pax-QNR protein involved in such complexes was the 46-kDa protein. After subcellular fractionation, paired-containing proteins were found in the nucleus but paired-less proteins were also

^{*} Corresponding author.

located in the cytoplasm. These results suggest that different developmentally regulated forms of the *Pax-QNR* transcription factors exist.

MATERIALS AND METHODS

Cells and plasmids used. Neuroretinas were dissected from 7-day-old (E7) or 14-day-old (E14) quail or chicken embryos. QNR and quail embryos cell (QEC) cultures were maintained and passaged in Dulbecco modified Eagle's medium supplemented with 10% fetal calf serum. COS-1 cells were maintained in the same medium.

Expression of Pax-QNR domains in bacteria and obtention of rabbit sera. The appropriate fragments were obtained from MC29-QNR2 cDNA as the template in polymerase chain reaction (PCR) experiments, using oligonucleotides corresponding to sequences present in the paired box, the fragment between the paired box and homeobox, the homeobox, and the carboxyl box of Pax-QNR containing the 5' BamHI restriction site and the 3' HindIII site (sequences of the oligonucleotides used are available upon request). The expected bands were subjected to BamHI-HindIII restriction enzyme digestion and inserted between the BamHI-HindIII sites of a pLC24-derived vector. The subcloning results in the in-frame fusion of the desired sequence to the first 99 amino acids of the polymerase of phage MS2. For expression, plasmids grown in LE392 (λ) were transferred into an Escherichia coli-host (SG4044) which has a temperature-sensitive repressor of the p_L 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 sodium dodecyl sulfate (SDS)-15% polyacrylamide gel. Proteins were visualized by staining with Coomassie brilliant blue and cut from the gel. Rabbit antisera were prepared as described previously $(\overline{10})$.

In vitro transcription. Plasmids pSP64 and pSP65 MC29-QNR2 were linearized by using the *PvuII* restriction site, and the sense (pSP65) and antisense (pSP64) RNA strands were synthesized with SP6 RNA polymerase. pSP64 and pSP65 MC29-QNR2 deleted up to the *NheI* site were also used. Half (5 μ l) of the RNA template was translated in a rabbit reticulocyte lysate. All reactions were performed as specified by the manufacturer (Bethesda Research Laboratories).

Cell labeling, immunoprecipitation, Western blotting (immunoblotting), and V8 analysis. Chicken neuroretina (CNR) cells were incubated for 45 min in the presence of 100 μ Ci of L-[³⁵S]methionine (specific activity, 1,000 Ci/mmol; Amersham) per ml, lysed in radioimmunoprecipitation assay (RIPA) buffer, and immunoprecipitated as previously described (2) with rabbit anti-Pax-QNR sera (serum 11 against the paired domain, serum 12 against the amino acids located between the paired domain and homeodomain, serum 13 against homeodomain, and serum 14 against the carboxylterminal part of the protein). Immunoprecipitated proteins were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) and fluorography.

For ${}^{32}P$ labeling, the cell cultures were preincubated for 4 h in a phosphate-free medium and further incubated for 4 h after the addition of 0.5 mCi of carrier-free [${}^{32}P$]H₃PO₄ per 100-mm-diameter plate. Phosphoamino acid determination was carried out as previously described (25).

Western blots were performed with 10^6 neuroretina cells boiled in loading buffer, subjected to SDS-PAGE, and electrophoretically transferred to Immobilon membranes (Millipore). Filters were treated for 60 min in blocking buffer (5% nonfat dry milk in phosphate-buffered saline [PBS]). Filters were then treated overnight in blocking buffer containing rabbit serum diluted 1:200 and washed for 30 min in blocking buffer–0.1% Tween 20. Bound antibodies were revealed with an enhanced chemiluminescence detection kit (Amersham).

Partial digestion by *Staphylococcus aureus* V8 protease in gel slices containing [³⁵S]methionine-labeled proteins was done essentially as described previously (4).

Immunofluorescence study and subcellular fractionation. Normal or transfected cells cultured on collagen-coated 12-mm-diameter microscope coverslips were fixed for 20 min with 3.7% paraformaldehyde in PBS. Transfected cells were fixed 78 h after the transfection and then treated as previously described (2) with sera 11 and 14. For subcellular fractionation, cells were labeled for 6 h with $[^{35}S]$ methionine, rinsed in PBS, and then scraped in lysis buffer (15 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid [HEPES; pH 7.5], 60 mM KCl, 15 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 0.15 mM spermidine, 0.5 mM spermine, 14 mM β -mercaptoethanol, 300 mM sucrose, 0.5% [vol/vol] Nonidet P-40). After a 3-min incubation on ice, the lysate was centrifuged. The cytoplasmic fraction contained in the supernatant was diluted in 10 volumes of RIPA buffer. The pellet which contained the nuclei was rinsed with lysis buffer and dissolved in RIPA buffer. Proteins immunoprecipitated with serum 14 were analyzed by SDS-PAGE and fluorography.

PCR experiment, DNA sequencing, and in vitro mutagenesis. Total RNA isolated from E7 QNR (15 μ g) was converted to cDNA by using 0.5 μ g of oligo(dT) primer (RT primer). Reverse transcription was carried out at 41°C for 2 h in a reaction buffer included in the kit and 0.5 mM each dATP, dCTP, dGTP, and dTTP, 500 U of Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories), and 60 U of RNAsin (Promega Biotec) in a final volume of 50 µl. Following heat inactivation at 95°C for 5 min, 2 µl of the reaction mixture was prepared for PCR (Eurogentec apparatus with Taq polymerase and Taq buffer [Amersham]) with 0.25 μ g of the appropriate oligonucleotide primers. The sequences of the primers were 5'-TAGGGATCCCCATGCA GAACAGTCACAGC and ATCAAGCTTAGATGGAGTTG GTGTTCTC for the junction between exons 3 and 4 and the beginning of exon 7, respectively. Amplifications were carried out in a Eurogentec thermojet for 40 cycles as follows: 1.0 min of denaturation at 94°C, 1.30 min of annealing at 50°C, and 1.0 min of elongation at 72°C. Reaction products were analyzed by electrophoresis through a 2% agarose gel. Amplified fragments were ligated between the BamHI and HindIII sites of pUC19. Nucleotide sequence was determined in both orientations by the dideoxynucleotide chain termination method, using a 35 S sequencing kit (Pharmacia). The amplified fragment was digested by restriction enzymes HindII-XmaI and inserted in the MC29-QNR2/B1 cDNA cloned in pSP65 vector digested with the same restriction endonucleases. Compared with MC29-QNR2, this clone contain an exon 10 18 nucleotides longer (8).

The mutagenesis on Pax-QNR conducted in order to remove the three internal AUGs was performed on the *HindII-XmaI* fragment inserted in M13 mp19 phage DNA digested with the same enzymes. The single-stranded directed mutagenesis was conducted with two mutated oligonucleotides complementary to the Pax-QNR sequence located between nucleotides 778 and 826 (24). The sequences of the oligonucleotides used were 5'-GTCGTAGATCCCGT CGGCACCGATCTGTTGC-3' and 5'-CCCGTTCAGGATC CTTAGCTT-3'. Letters in boldface are the mutated nucleotides. The mutation of an AUG to AUC is expected to result in the conversion of a Met to Ile in the protein.

Nuclear extract preparation and gel mobility shift assays. Nuclei were prepared from tissues. Crude nuclear extracts were prepared essentially as described previously (7). All buffers contained a cocktail of proteinase and phosphatase inhibitors (0.5 mM phenylmethylsulfonyl fluoride, 1 µg each of leupeptin, antipain, and pepstatin per ml, 2 mM benzamidine, 10 mM β -glycerophosphate, 2 mM levamisole, 10 μ M orthovanadate). Binding reactions were performed for 10 min on ice, using 1 ng of radiolabeled DNA probe and 4 µg of nuclear extract in 14 µl of 10% glycerol-10 mM HEPES (pH 7.9)-30 mM KCl-4 mM spermidine-0.1 mM EDTA-0.25 mM dithiothreitol-1 mM Na₂HPO₄ (pH 7.2)-1 µg of singlestranded salmon sperm DNA-1.5 µg of poly(dI-dC). For competition experiments, a 100-fold molar excess of unlabeled competitor oligonucleotides was added simultaneously with the probe. Samples were loaded on a 6% native polyacrylamide gel, run for 2 h at 180 V, and examined by autoradiography after exposure of the dried gel to a Kodak X-AR film at -70° C with an intensifying screen.

To prepare the DNA probes for gel shift analysis, an oligonucleotide containing the e5 binding site (CTCAGCAC <u>CGCACGATTAGCACCGTTCCGCTTC</u>) and its complementary strand were synthesized, annealed, and used as DNA probe after 5'-end labeling. The recognition sequence bound by the homeodomain is underlined, and the paired domain recognition sequence is shown in boldface.

UV cross-linking assays. DNA probes were prepared by 3'-end annealing of complementary 10-mer primers that were extended with Klenow fragment of DNA polymerase I and 50 μ Ci (3,000 Ci/mmol; Amersham) each of $[\alpha^{-32}P]$ dATP, $[\alpha^{-32}P]$ dCTP, and $[\alpha^{-32}P]$ dGTP and 5 mM 5-BrdUTP and purified on native 20% polyacrylamide gels. Cross-linking assays were performed essentially as described previously (15), using 5 ng of probe DNA and 30 μ g of nuclear extract (30 μ l, final volume). After the binding reaction, the solution was exposed to 302-nm UV light from a transilliminator. Cross-linked DNA-protein complexes were identified by autoradiography after immunoprecipitation with serum 14 in RIPA buffer.

RESULTS

Preparation of Pax-QNR antisera. The sequence of Pax-QNR from MC29-QNR2 cDNA (2,509 bp) has already been published along with the derived sequence of the longest open reading frame (24). The deduced protein sequence (416 amino acids) revealed a paired domain and a paired-type homeodomain (Fig. 1). MC29-QNR2 RNA translated in rabbit reticulocyte lysate in the presence of [³⁵S]methionine produced a doublet band of 46 and 47 kDa specific for this cDNA (8). To obtain rabbit antisera able to specifically recognize Pax-QNR protein, we expressed separately in bacteria distinct parts of Pax-QNR and injected rabbits with gel-purified peptides. We produced the paired domain (127 residues; rabbit 11), the junction between the paired domain and homeodomain (59 residues; rabbit 12), the homeodomain (81 residues; rabbit 13), and the carboxyl-terminal domain (146 residues; rabbit 14) as fusion proteins with part of MS2 phage polymerase (98 residues). The different anti-Pax-QNR sera were tested in immunoprecipitations using whole [35S]methionine-labeled MC29-QNR2 protein expressed in a reticulocyte lysates. SDS-PAGE analysis of anti-Pax-QNR immunoprecipitates revealed a predominant protein with a relative molecular mass of 46 kDa, which was

.....



FIG. 1. Characterization of the antisera directed against Pax-QNR protein. The MC29-QNR2 cDNA was transcribed in vitro in a sense orientation and translated in a rabbit reticulocyte lysate. Translated proteins were immunoprecipitated with rabbit antisera prepared against bacterially expressed peptides. The peptides encoded by the paired domain (serum 11), the junction between the paired domain and the homeodomain (serum 12), the homeodomain (serum 13), and the carboxy-terminal part (serum 14) of the Pax-QNR protein were eluted after SDS-PAGE (15% polyacrylamide gel) of the bacterial proteins and injected into rabbits (see Materials and Methods). The localization of these peptides with respect of the MC29-QNR-2 cDNA is shown in the diagram. Gray box, paired domain; hatched box, homeodomain. The start and stop codons are also indicated. Lanes i, serum from immunized rabbits; lanes p, serum from preimmune rabbits. Size is indicated in kilodaltons.

not recognized by the cognate preimmune serum (Fig. 1). The four sera immunoprecipitated the in vitro-synthesized Pax-QNR protein with equal efficiency. Pax-QNR protein was precipitated even when the serum was incubated with an excess of MS2 peptide, suggesting that the protein was recognized through specific determinants of the Pax-QNR segment of the immunogen (data not shown).

Detection of Pax-QNR proteins in vivo. To detect the Pax-QNR products in vivo, 7-day-old CNRs were isolated and [³⁵S]methionine-labeled for 45 min. SDS-PAGE analysis of Pax-QNR products immunoprecipitated with serum 11 (directed against the paired domain) revealed a predominant doublet with relative molecular masses of 48 and 46 kDa (Fig. 2A, lane 1). A minor protein of 43 kDa was also reproducibly detected. The same proteins were recognized with immune sera 12, 13, and 14 but not with the cognate preimmune sera (Fig. 2A). Sera 12, 13, and 14 also recognized a doublet of 33 and 32 kDa (Fig. 2A, lanes 3, 5, and 7). These proteins were not detected by serum 11 (lane 1). To determine whether these proteins were Pax-QNR-associated proteins, we performed an immunoprecipitation experiment with serum 11 or 14, and then the proteins recovered from the immunocomplexes bound to protein A-Sepharose were immunoprecipitated again with either the same or a distinct Pax-QNR antiserum. If the 32- and 33-kDa proteins recovered with serum 14 were Pax-QNR-associated proteins and not Pax-QNR isoforms, we did not expect to find them in the second immunoprecipitation with serum 14, 13, or 12. Figure 2B demonstrates that this is not the case, since these proteins were recovered with serum 14 from serum 13 immunoprecipitates. From the serum 11 immunoprecipitate, the same triplet proteins (48, 46, and 43 kDa) can be recovered with all of the sera used for the second immunoprecipitation. To provide further evidence that the 33/32-kDa



FIG. 2. Expression and characterization of Pax-QNR proteins in vivo. (A) Seven-day-old CNRs were isolated and [³⁵S]methionine labeled for 45 min. Cell lysate was immunoprecipitated in RIPA buffer with sera 11, 12, 13, and 14 (lanes i) and with cognate preimmune sera (lanes p). (B) Immunoprecipitations were per-

doublet proteins are not unrelated proteins that strongly interact with one or more forms of Pax-QNR, we performed a Western analysis of unlabeled neuroretina proteins (Fig. 2C). This experiment clearly demonstrates that the 33/32kDa doublet proteins are not recognized by Pax-QNR antiserum through an association with Pax-QNR products. MC29-QNR2 RNA translated in rabbit reticulocyte lysate in the presence of [35S]methionine produced a doublet band of 46 and 47 kDa but also in some experiments a 33/32-kDa doublet protein presumably produced from an internal AUG. To test this hypothesis, we truncated the MC29-QNR2 cDNA to the NheI site in order to remove the first AUG (24) and translated this truncated cDNA in a rabbit reticulocyte lysate in the presence of [³⁵S]methionine. As shown in Fig. 2D, the expected short protein is produced from this cDNA (compare lane 6 with lanes 2 and 4). We also performed site-directed mutagenesis on MC29-QNR2 cDNA in order to convert the three internal AUGs to AUCs. Figure 2E show the nucleotide sequence of the mutated part of the cDNA. In vitro translation of the mutated cDNA resulted in the disappearance of the 33/32-kDa doublet protein (Fig. 2E, lanes 1 and 3). The nature of the 34-kDa protein translated from the mutated cDNA (suggesting that non-AUG initiation can occur in reticulocyte lysates) was not investigated further. Together, these results suggest that the paired-less 33/32kDa doublet proteins found in the neuroretina are produced from internal AUG initiation. It remain to be determined whether the same mRNA encodes the paired-containing and the paired-less proteins or whether a particular messenger is produced to translate the paired-less proteins.

To determine whether all of these five proteins detected in vivo were structurally related to Pax-QNR, we compared V8 protease fragments obtained from [³⁵S]methionine-labeled proteins generated from in vitro transcription-translation of MC29-QNR2 cDNA and from neuroretinas. Figure 3 shows that the labeled peptide patterns of the in vivo- and in vitro-generated proteins are superimposable, demonstrating that the 48-, 46-, 43-, and 33/32-kDa proteins recognized by anti-Pax-QNR sera are produced from the same gene.

formed in a sequential order with serum 11, 13, or 14, and then the proteins recovered from the immunocomplexes bound to protein A-Sepharose were reimmunoprecipitated with serum 11 or 14 as indicated above the lanes. (C) Western blotting was performed on neuroretina cell extracts with rabbit serum 14. Bound antibodies were detected with an enhanced chemiluminescence detection kit. (D) MC29-QNR2 RNA translated in rabbit reticulocyte lysate in the presence of [³⁵S]methionine produced a doublet band of 46/47 kDa, but also in some experiments a 33/32-kDa doublet protein immunoprecipitated by serum 14 (lane 2) but not by serum 14 blocked with an excess of bacterial immunogen (lane 1). To test the possibility that 33/32-kDa proteins found in the neuroretina (lane 4) were produced from internal AUGs, we truncated the MC29-QNR2 cDNA to the *Nhe*I site (Δ NheMC29-QNR2) in order to remove the first AUG and translated this truncated cDNA in rabbit reticulocyte lysate in the presence of L-[³⁵S]methionine. The resulting proteins was immunoprecipitated with serum 14 blocked with an excess of bacterial immunogen (lane 5) and serum 14 (lane 6). Lanes i, immune serum; lanes b, blocked serum. (E) We also mutagenized the three Pax-QNR internal AUGs to AUCs as shown by nucleotide sequence analysis. The sequence was made by the dideoxynucleotide-chain termination method as specified by the manufacturer (Pharmacia). Asterisks indicate the mutated nucleotides. The proteins translated in rabbit reticulocyte lysates from two distinct mutagenized clones (mt; lanes 1 and 3) were compared with proteins translated from MC29-QNR2 wild-type cDNA (wt; lane 2). Sizes are indicated in kilodaltons.



FIG. 3. One-dimensional V8 protease mapping of reticulocyte MC29-QNR2 cDNA-encoded and neuroretina Pax-QNR proteins. Seven-day-old CNR cells were labeled with 2 mCi of L-[³⁵S]methionine per ml, lysed in RIPA buffer, and immunoprecipitated with serum 14. Labeled proteins were excised from the gel and run on a 15% polyacrylamide gel after partial digestion for 1 h with either 20 or 200 ng of *S. aureus* V8 protease. Sizes of the purified Pax-QNR proteins are indicated above the lanes. The positions of prestained molecular weight standards are shown on the left in kilodaltons.

Since three paired-containing proteins (48, 46, and 43 kDa) can be detected in the neuroretina, it was important to determine the relationship between these proteins. They did not appear to be differentially phosphorylated forms of the same protein, since phosphatase treatment did not resolve the protein triplet in a single species (data not shown). In addition, we could find no evidence for a precursor-product relationship between the three proteins (see Fig. 6). We have previously reported that Pax-ONR is able to generate distinct RNA by alternative splicing (8). The paired-containing 48- and 46-kDa proteins could be produced from two already identified transcripts: the 48-kDa product may be encoded by the mRNA containing the additional paired exon 4a (8, 33), and the 46-kDa product may be encoded by the MC29-QNR2 mRNA. However, no mRNA has yet been described for the 43-kDa product. To identify a possible mRNA transcript for the paired-containing 43-kDa product generated by an alternative splicing which would reduce the size of the paired domain, we performed PCR amplifications using two oligonucleotides located, respectively, at the junction of exons 3 and 4 at the beginning of the paired domain and at the end of the junction between the paired domain and homeodomain (at the beginning of exon 7). As shown in Fig. 4A, in addition to the expected 600-bp fragment, we amplified in the neuroretina a fragment of 400 bp. The sequence of this fragment revealed a splicing event between the beginning of exon 5 and exon 6 resulting in a paired domain 66 amino acid shorter than that normally observed. The sequence located in exon 5 at this alternative splice junction (gtgagt) is identical to the sequence located at the splice junction in the intron between exons 1 and 2 (8); the unusual splice donor nucleotides GT used to produce the reduced exon 5 are found in the same location in the regular exon 5 (8) (Fig. 4A). To confirm that this truncated messenger can encode the 43-kDa protein, we replaced the HindII-XmaI fragment from the Pax-QNR cDNA by the same





FIG. 4. Characterization of a Pax-QNR messenger encoding the p43 isoform. (A) Sequence comparison and schematic representation of Pax-QNR paired exons previously described for MC29-QNR2 cDNA (8) and partial cDNA clone (CNR CL2) isolated by PCR from CNR RNA. (B) To test the possibility that the p43 protein found in the neuroretina was produced from the mRNA bearing the reduced paired domain, we truncated the MC29-QNR2/B1 cDNA (we used this particular cDNA because it contain an exon 10 six amino acids longer than that of MC29-QNR2 [8]) to the *Hind*II site (located in exon 4) and *Xma*I site (located in exon 6) and replaced this fragment with the *Hind*II-*Xma*I fragment produced from the CNR CL2 DNA. The reconstructed cDNA (p43 clone) was translated in rabbit reticulocyte lysate in the presence of L-[³⁵S]methionine. Proteins were immunoprecipitated with serum 14 from lysates indicated at the top. Lanes i, immune serum; lanes p, preimmune serum. Sizes are indicated in kilodaltons.

fragment prepared from the PCR amplifications described above. The resulting clone translated in rabbit reticulocyte lysate in the presence of [35 S]methionine produced a protein of 43 kDa (Fig. 4B, lane 3). We conclude that such truncation of the paired domain could explain the molecular weight of the 43-kDa paired-containing protein observed in neuroretina cell lysates.

Pax-QNR proteins are phosphorylated on serine and threonine residues. Many different types of stimuli that affect gene expression also lead to the activation of protein kinases, and most transcription factors are phosphoproteins. Phosphory-



1 2 3 4

FIG. 5. Phosphorylation of Pax-QNR proteins expressed in the neuroretina and phosphoamino acid determination. (A) Pax-QNR polypeptides were immunoprecipitated from ³²P_i-labeled CNR cells with immune sera 11 and 13 (lanes i) or cognate preimmune sera (lanes p). Sizes are indicated in kilodaltons. (B) CNR cells were labeled with ³²P_i, immunoprecipitated with serum 14, and subjected to electrophoresis. The Pax-QNR bands (p46 and p33/32) were cut out, hydrolyzed in acid, and subjected to two-dimensional electrophoresis on thin-layer chromatography plates.

lation has been shown to be important for subcellular localization, modulation of DNA binding, and interaction of transcription factor transactivation domains with the transcriptional machinery (16). To examine whether the Pax-QNR products identified above are phosphorylated, proteins were ³²PO₄ labeled and immunoprecipitations with anti-Pax-QNR sera were performed on labeled neuroretina cells (Fig. 5A). The results show that the 46-kDa and paired-less proteins were recovered. Because of the diffusion of the ³²P signal, it is not possible to determine whether the proteins of very similar molecular weight were all phosphorylated. However, both paired-containing and paired-less Pax-QNR products are phosphoproteins.

To determine the nature of the phosphorylated amino acid, we performed two-dimensional phosphoamino acid analysis on the excised ³²P-labeled proteins. Figure 5B shows that both proteins contain phosphorylated serines; however, the 46-kDa protein also exhibit a minor spot representing threonine.

Pax-QNR encodes stable proteins differing in their nuclear affinity and regulation during development. To examine the stability of the different Pax-QNR proteins, we performed pulse-chase analysis in neuroretina cells with a 45-min pulse with $[^{35}S]$ methionine. We found that paired-containing and paired-less proteins exhibited similar half-lives of 6 h (Fig. 6). In addition, the relative levels of these proteins were unaltered during the chase period, consistent with the idea that these proteins are independent translation products.

That Pax-QNR products are nuclear was demonstrated by immunofluorescence data (Fig. 7A and data not shown). To discriminate between paired-containing and paired-less proteins, we performed transient transfection experiments of paired-containing protein- and paired-less protein-encoding Pax-QNR expression vectors into COS cells as well as QEC. As expected, serum 11 is unable to detect any signal with the transfected pSG5-Nhe1 clone encoding the p33/32 proteins (Fig. 7A, panels e, g) but detected the nuclear p47/46



FIG. 6. Stability of Pax-QNR products analyzed by pulse-chase labeling. CNR cells were L-[³⁵S]methionine labeled for 45 min and chased for the indicated periods of time in complete medium. Cell lysates were immunoprecipitated with serum 14. Lane i, immune serum; lane b, serum blocked with an excess of the bacterial immunogen. Sizes are indicated in kilodaltons.

paired-containing protein encoded by the pSG5-SstII vector both in COS cells and QEC (Fig. 7A, panels a and c). Serum 14 recognized the nuclei of both cell types transfected with vectors encoding paired-less and paired-containing proteins (Fig. 7A, panels b, d, f, and h). However, a strong cytoplasmic staining was also observed with pSG5-Nhe1 vector (Fig. 7A, panels f and h). Since the immunofluorescence experiments do not give information on the localization of the different Pax-QNR products in the neuroretina, we performed neuroretina cell fractionation followed by immunoprecipitation with serum 14. The p46/48 paired-containing proteins are found in the nucleus, whereas the p43, p33, and p32 products were mostly cytoplasmic (Fig. 7B).

Next, we examined whether paired-containing and pairedless proteins are subjected to developmental switch. We found that the relative levels of the two types of Pax-QNR products varied dramatically with the age of the dissected neuroretina. Paired-less proteins were observed at E5 and hardly detectable at E15, whereas the p46/48 paired-containing proteins were detected in equivalent amount at E5 and E15 (Fig. 8).

Paired-containing but not paired-less Pax-QNR proteins bind the e5 DNA sequence. To determine whether pairedcontaining and paired-less Pax-QNR isoforms were able to specifically recognize the e5 sequence (as demonstrated for the p46 translated in reticulocyte lysates [8]) we performed electrophoretic mobility shift assays (EMSAs) using E6 and E15 neuroretina nuclear extracts and ³²P-labeled e5 DNA (Fig. 9). The e5 DNA-binding complex found in reticulocyte lysates is marked by an arrow. Similar complexes are also present in E6 and E15 neuroretina extracts but not in E6 liver nuclear extracts. This complex is specific for the e5 sequence, since it disappeared in the presence of a 100-fold excess of unlabeled e5 oligonucleotide (lane e5) but not in the presence of a 100-fold excess of unrelated AP3 binding site oligonucleotide (lane AP3). In addition to this complex,



FIG. 7. Nuclear localization of Pax-QNR products by immunofluorescence and subcellular fractionation. (A) Subcellular localization of Pax-QNR p46 and p33/32 products was assayed by indirect immunofluorescence on transfected fixed COS cells (a, b, e, and f) and QEC (c, d, g, and h). Anti-Pax-QNR immunoreactive proteins were detected with fluorescein isothiocyanate-labeled goat anti-rabbit immunoglobulin secondary reagent. (B) QNR cells (approximately 3×10^7) were labeled for 6 h with 0.2 mCi of L-[³⁵S]methionine. Scraped cells were separated into two equal aliquots. The first one was immunoprecipitated with serum 14 (unfractionated, T; lanes 1 and 2); the second one was further separated in cytoplasmic (C; lanes 3 and 4) and nuclear (N; lanes 5 and 6) fractions. Each fraction corresponding to the same number of cells was immunoprecipitated with serum; lanes i, immune serum). Sizes are indicated in kilodaltons.

a complex migrating with a slower electrophoretic mobility can be resolved. It is unlikely that this complex involves Pax-QNR proteins, since it exist in the liver extracts devoid of Pax-QNR products.

To confirm that Pax-QNR proteins are in the DNA-protein complex and to precisely define the isoforms involved, we performed UV cross-linking assays with the e5 probe simultaneously labeled with ³²P and 5-BrdUTP and neuroretina





FIG. 8. Developmental expression pattern of the Pax-QNR products in E5 and E15 CNR cells. Neuroretinas were isolated, labeled with $[^{35}S]$ methionine for 45 min, and immunoprecipitated with immune sera 11 and 14 (lanes i) and the corresponding preimmune sera (lanes p). Sizes are indicated in kilodaltons.

FIG. 9. Tissue-specific distribution of a factor that interacts with the e5 DNA probe. The arrow shows a specific DNA-protein complex found in reticulocyte lysates expressing the p46 Pax-QNR protein. The e5 probe was incubated without (lane probe) or with equal amounts of nuclear extracts from tissues indicated on the top. Competitions were carried out in the absence (-) or presence of a 100-fold molar excess of unlabeled competitor DNA e5 or AP3 binding site. Protein-DNA complexes were resolved on a 6% acrylamide gel.



FIG. 10. Determination of Pax-QNR isoforms involved in DNA-protein complexes. (A) EMSA with the e5 probe simultaneously labeled with ^{32}P and 5-BrdUTP and neuroretina nuclear extracts. (B) The DNA-proteins complexes formed after UV cross-linking between labeled e5 and p46 Pax-QNR (MC29-QNR2) or E5 QNR nuclear extracts (QNR) were immunoprecipitated in RIPA buffer with serum 14 (lanes i) or cognate preimmune serum (lanes p) and subjected to SDS-PAGE. (C) Binding to the e5 sequence of in vitro-translated Pax-QNR proteins. The e5 probe was incubated without (lane probe) or with 1 μ l of reticulocytes lysates. Other lanes: control, reticulocyte lysate expressing the paired-less p33/32 proteins; MC29-QNR2, reticulocyte lysate expressing the p46 Pax-QNR. (D) To determine the amounts of proteins involved in the EMSA, we performed a Western blot analysis of 5 μ l of reticulocyte lysate with rabbit serum 13. Sizes are indicated in kilodaltons.

nuclear extracts. The DNA-protein complexes formed in solution were then immunoprecipitated with serum 14 and resolved under denaturing conditions. As illustrated in Fig. 10A, the probe is recognized both by the 46-kDa Pax-QNR protein translated from both reticulocyte lysates and E6 neuroretina nuclear extracts. After immunoprecipitation with serum 14 in RIPA buffer followed by SDS-PAGE analysis, a protein of identical molecular weight is detected in MC29-QNR2 reticulocyte lysates and in nuclear extract (Fig. 10B), strongly suggesting that among the various Pax-QNR proteins, p46 alone is able to bind to the e5 sequence. Therefore, in the conditions used to allow DNA-protein complex formation, the paired-less Pax-QNR isoforms are unable to bind e5. This is further demonstrated by the lack of detectable complex in EMSAs performed with the reticulocyte lysates product of the Δ NheMC29-QNR2 clone encoding the paired-less proteins (Fig. 10C). Figure 10D shows the amounts of proteins used to perform the EMSAs shown in Fig. 10C.

DISCUSSION

We have previously described the isolation of *Pax-QNR* (24), the quail homolog of the *Pax-6* gene. To characterize the protein products of this gene in the neuroretina, we prepared rabbit antisera from bacterially expressed parts of the gene. The previously isolated cDNA, MC29-QNR2,

shows one in-frame AUG translation start site that places the paired domain very close to the amino terminus of the protein. This AUG is expected to initiate in vitro a major protein product of 46 kDa also found in vivo with all of the Pax-ONR sera used. In addition to this expected product, we detected in the neuroretina several other proteins translated from Pax-QNR, one of higher molecular mass (48 kDa) and three of lower molecular mass (43, 33, and 32 kDa). The murine Pax-6 cDNA previously described (33) shows an insertion of 14 amino acids between exons 1 and 2 of the paired-box domain with respect to the Pax-QNR sequence. We found the corresponding sequence in the first paired-box intron of the Pax-QNR locus. Therefore, we called this sequence exon 4a (8). An mRNA containing this additional exon may explain the 48-kDa protein identified in the neuroretina and immunoprecipitated with all sera used, since we found no evidence for posttranslational modification of p46 to explain the existence of this protein product. However, the cDNA corresponding to this particular splicing with an extended paired box has not yet been isolated in avian species, although we can amplify a DNA fragment in reverse transcription-PCR experiments conducted with neuroretina RNA and exon 4a-specific oligonucleotides. A minor p47 product was observed in vitro by MC29-QNR2 cDNA translation. Analysis of the sequence revealed that the open reading frame extended upstream from the first AUG. The conservation of this extension in all species studied (fish,

quail, mouse, and human) suggests a possible role for the upstream part of the open reading frame, but only mouse and human cDNAs showed clear homology in the protein deduced from the extended reading frame (8). It is tempting to assume that this extended reading frame presumably used in vitro from an alternative initiation codon may also be used in vivo to produce the p48 protein. However, the cellular p48 exhibits additional V8 peptides not found in the reticulocyte p47 (data not shown), suggesting that they are distinct proteins. The minor p43 protein recognized by serum 11 may arise from an alternative Pax-QNR transcript lacking most of paired exon 5, as demonstrated by PCR experiments and reconstructed Pax-QNR variant translated in vitro. This protein, which is 66 amino acids shorter in the paired domain, would still be recognized by serum 11, since this serum is made against the 121 paired residues. However, we have been unable to isolate a Pax-QNR cDNA corresponding to this particular messenger.

Paired-less p33 and p32 proteins are not processed products of the full-size proteins, since paired-less and pairedcontaining proteins exhibited similar half-lives, and we could not find evidence of alternative splicing either by PCR or cDNA library screening able to explain these short Pax-QNR products. An alternative explanation for the alternative splicing event (that we cannot formally rule out to date) is that these small paired-less proteins are generated by alternative translational initiations at internal AUGs in the full-length Pax-QNR mRNA. That the internal AUGs located after the paired box are functional in vivo is demonstrated by the expression of p33/32 proteins in transfection of avian or mammalian cells with Pax-QNR expression vectors. Alternative translational initiation is a process already found for the production of thyroid hormone receptor truncated forms (3), S-CREM proteins (5), and LAP/LIP proteins (6). Proteins generated by the use of internal AUG codons act as repressors of the activity of the full-size proteins (3, 5, 6). The size difference between p33 and p32 can be explained by the use of an alternative exon 10. Indeed, we found two Pax-QNR cDNAs able to encode two proteins differing by six amino acids encoded by exon 10 (8). Combination of all these alternative exons (4a, 5, and 10) and alternative translational initiations may explain the complexity of the Pax-QNR protein pattern found in the neuroretina.

Pax-QNR is expressed not only in the neuroretina during embryonic development but also in the developing neural tube and the pancreas (8, 13, 24, 33; unpublished data). We found that the developing neural tube expresses the p48 and p46 paired-containing proteins but no p33/32 paired-less isoforms. In contrast, the pancreas expressed the same Pax-QNR isoforms as did the neuroretina (32a).

The functions of all of these Pax-ONR isoforms remain to be defined. Paired-less proteins retain the homeodomain DNA binding and thus should behave like conventional homeodomain-containing proteins. In fact, genes encoding paired-type homeodomain-containing proteins have already been described (12, 21). However, these proteins differ from the Pax gene homeodomain in the ninth amino acid in the third DNA recognition helix of the homeodomain, which is an important determinant of DNA-binding specificity (31). This difference suggests that the DNA-binding specificities of the Pax-QNR p33/32 and the recently described pairedtype homeodomain products are not identical. The p46 Pax-QNR protein is able to bind to the Drosophila e5 sequence $(\bar{8})$, and this binding is mediated by the paired domain and the homeodomain, as for the Pax-3 protein (11), and both the homeodomain and paired domain are required for high-affinity binding (8). That both domains are involved in DNA binding is indicated by the fact that the paired domain and homeodomain individually expressed in bacteria are able to bind to e5 (8). However, we have been unable to demonstrate a DNA-binding activity for the p33/32 pairedless proteins expressed in reticulocyte lysates or in neuroretina nuclear extracts nor an association in vitro between the paired-containing and paired-less proteins. Unlike other homeodomain-containing proteins, hepatocyte nuclear factor 1α must dimerize in order to bind DNA, and a cofactor (DCoH) that regulates dimerization of hepatocyte nuclear factor 1α has been described (26). It is possible that the p33/32 proteins expressed in reticulocyte lysates need a cofactor in order to form DNA-binding dimers. It remains possible that the p33/32 paired-less Pax-QNR products possess specific DNA-binding activity and that the neuroretina nuclear extracts do not reveal any binding on e5 of these proteins because of their primarily cytoplasmic localization. However, cytoplasmic extracts prepared from the same cells were negative in EMSAs with the e5 probe (data not shown). Alternatively, it is possible that the carboxyl-terminal part of the protein masks the DNA-binding sequence of the p33/32 homeodomain as suggested for the Drosophila prd gene (31) and as demonstrated for p53 protein (17).

We found that the cellular Pax-QNR proteins were phosphorylated. Phosphorylation events are known to regulate the DNA-binding properties of some transcription factors (16), and phosphorylation is required for the DNA-binding properties of LR1/BSAP/Pax-5, a paired-box-containing transcription factor (35). Two types of kinase may be involved in phosphorylation of *Pax-QNR*. Based on the presence of consensus kinase sites in the *Pax-QNR* sequence, a potential casein kinase II phosphorylation site (Ser-Asn-Gly-Glu-Asp) is located at the beginning of the homeodomain, and two potential phosphorylation sites for mitosis-associated protein kinase (Pro-Thr-Thr-Pro and Pro-Thr-Ser-Pro) are located in the carboxyl-terminal part of the protein. It remain to be tested whether these sites are used in vitro or in vivo.

Because immunohistochemical localization of the proteins did not allow us to distinguish between paired-less and paired-containing proteins, we performed immunoprecipitations on subcellular fractions of neuroretinas. Surprisingly, the paired-less Pax-QNR proteins were found mostly in the cytoplasm and p48/46 was found in the nucleus. Transient transfection experiments with expression vectors encoding p46 or p33/32 revealed that both types of proteins are nuclear, even if p33/32 showed a significant cytoplasmic localization. It remains possible that paired-less p33 and p32 are recovered mostly in the cytoplasm after the fractionation process because strong nuclear affinity requires the presence of the paired domain. Since the ratio of paired-containing and paired-less proteins varied dramatically following neuroretina differentiation or tissue origin, the additional activity provided by the homeodomain may account for some of the specificity of action of Pax-QNR in development.

ACKNOWLEDGMENTS

We thank Christine Dozier and Kay MacLeod for critical reading of the manuscript.

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

REFERENCES

1. Adams, B., P. Dörfler, A. Aguzzi, Z. Kozmick, 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.

- Amouyel, P., V. Laudet, P. Martin, R. Li, B. Quatannens, D. 2. Stehelin, and S. Saule. 1989. Two nuclear oncogenic proteins, P135^{sug-myb-ets} and p61/63^{myc}, cooperate to induce transformation of chicken neuroretina cells. J. Virol. 63:3382-3388.
- 3. Bigler, J., W. Hokanson, and R. N. Eisenman. 1992. Thyroid hormone receptor transcriptional activity is potentially autoregulated by truncated forms of the receptor. Mol. Cell. Biol. 12:2406-2417
- 4. Cleveland, D. W., S. G. Fisher, M. W. Kirschner, and U. K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analyses by gel electrophoresis. J. Biol. Chem. 252:1102-1106.
- 5. Delmas, V., B. Laoide, D. Masquilier, R. de Groot, N. Foulkes, and P. Sassone-Corsi. 1992. Alternative usage of initiation codons in mRNA encoding the cAMP-responsive-element modulator generates regulators with opposite functions. Proc. Natl. Acad. Sci. USA 89:4226-4230.
- 6. Descombes, P., and U. Schibler. 1991. A liver-enriched transcriptional activator protein, LAP and a transcriptional inhibitory protein LIP, are translated from the same mRNA. Cell 67:569-579.
- 7. Dignam, J. D., R. M. Lebovitz, and R. G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11:1475-1489.
- 8. 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 pairedbox- and homeobox-containing gene. Cell Growth Differ. 4:281-289.
- Frain, M., G. Swart, P. Monaci, A. Nicosia, S. Stämpfli, R. Frank, and R. Cortese. 1989. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. Cell 59:145-157.
- 10. Ghysdael, J., A. Gegonne, P. Pognonec, D. Dernis, D. Leprince, and D. Stehelin. 1986. Identification and preferential expression in thymic and bursal lymphocytes of a c-ets oncogene encoded mr 54,000 cytoplasmic protein. Proc. Natl. Acad. Sci. USA 83:1714-1718.
- 11. Goulding, M. D., G. Chalepakis, U. Deutch, J. R. Erselius, and P. Gruss. 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO J. 5:1135-1147.
- 12. Grueneberg, D. A., S. Natesan, C. Alexandre, and M. Gilman. 1992. Human and Drosophila homeodomain proteins that enhance the DNA-binding activity of serum response factor. Science 257:1089-1095.
- 13. Gruss, P., and C. Walther. 1992. Pax in development. Cell 69:719-722.
- 14. 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.
- 15. Hirai, S.-I., and M. Yaniv. 1989. jun DNA-binding is modulated by mutations between the leucines or by direct interaction of Fos with the TGACTCA sequence. New Biol. 1:181-191.
- 16. Hunter, T., and M. Karin. 1992. The regulation of transcription by phosphorylation. Cell 70:375-387.
- 17. Hupp, T., D. W. Meek, C. A. Midgley, and D. P. Lane. 1992. Regulation of the specific DNA binding function of p53. Cell 71:875-886.
- 18. Ingham, P. W. 1988. The molecular genetics of embryonic pattern formation in Drosophila. Nature (London) 335:25-34.

- 19. Ingraham, H. A., R. Chen, H. Mangaiam, S. Elsholtz, C. R. Flynn, D. M. Lin, L. Simmons, L. Swanson, and M. G. Rosenfeld. 1988. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell 55:519-529.
- 20. Karlson, O., S. Thor, T. Norberg, H. Ohlsson, and T. Edlund. 1990. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and Cys- His domain. Nature (London) 344:879-882.
- 21. Kern, M., D. Witte, M. Valerius, B. Aronow, and S. Potter. 1992. A novel murine homeobox gene isolated by a tissue specific PCR cloning strategy. Nucleic Acids Res. 20:5189-5195.
- 22. Kessel, M., and P. Gruss. 1990. Murine developmental control genes. Science 249:374-379.
- 23. Kissinger, C., B. Liu, E. Martin-Blanco, T. Kornberg, and C. O. Pabo. 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.
- 24. Martin, P., C. Carriere, C. Dozier, B. Quatannens, M. Mirabel, B. Vandenbunder, D. Stehelin, and S. Saule. 1992. Characterization of a pairedbox- and homeobox-containing quail gene (pax-QNR) expressed in the neuroretina. Oncogene 7:1721-1728.
- 25. Martin, P., W. Vass, J. T. Schiller, D. R. Lowy, and T. Velu. 1989. The bovine papillomavirus E5 transforming protein can stimulate the transforming activity of EGF and CSF-1 receptors. Cell 59:21-32.
- 26. Mendel, D. B., P. A. Khavari, P. Conley, M. Graves, L. Hansen, A. Admon, and G. R. Crabtree. 1991. Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. Science 254:1762-1767.
- 27. Monaghan, P., D. R. Davidson, C. Sime, E. Graham, R. Baldock, S. Bhattacharya, and R. Hill. 1991. The Msh-like homeobox genes define domains in the developing vertebrate eye. Development 112:1053-1061.
- 28. Nornes, H. O., G. R. Dressler, E. W. Knapik, U. Deutch, and P. Gruss. 1990. Spatially and temporally restricted expression of Pax-2 during murine neurogenesis. Development 109:797-809.
- 29. Saha, M., M. Servetnick, and R. Grainger. 1992. Vertebrate eye development. Curr. Opin. Genet. Dev. 2:582-588.
- 30. 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.
- 31. Treisman, J., P. Gönczy, M. Vashishtha, E. Harris, and C. Desplan. 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59:553-562.
- 32. Treisman, J., E. Harris, and C. Desplan. 1991. The paired box encodes a second DNA-binding domain in the Paired Homeodomain containing protein. Genes Dev. 5:594-604.
- 32a. Turque, N., and S. Saule. Unpublished data.
 33. Walther, C., and P. Gruss. 1991. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development 113: 1435–1449.
- 34. 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.
- Williams, M., and N. Maizels. 1991. LR1, a lipopolysaccharideresponsive factor with binding sites in the immunoglobulin switch regions and heavy-chain enhancer. Genes Dev. 5:2353-2361.

ARTICLE IV

"Nuclear localization signals, DNA-binding and Transactivation properties of Quail Pax-6 (Pax-QNR) isoforms."

Catherine Carrière, Serge Plaza, Jocelyne Caboche, Christine Dozier, Manuella Bailly,

Patrick Martin & Simon Saule

soumis à Mol. Cell. Biol.



Fig.12 : Schéma d'entrée des protéines dans le noyau

Après synthèse dans le cytoplasme, les protéines contenant des signaux de localisation nucléaire s'associent avec le complexe constitué par le pore nucléaire (1),soit via les protéines associées au pore nucléaire pouvant fixer les NLS (NBP) ou via les nucléoporines (NP), soit en s'associant avec des NBP cytoplasmiques (1a) qui vont ensuite se fixer sur le pore (1b). Les protéines vont être transportées à travers le pore par un processus nécessitant de l'ATP (2), soi après relargage par la NBP (3), soit toujours associée à la NBP (4). Dans ce dernier cas, la Nbp sera libérée dans le noyau (5) et repassera dans le cytoplasme (6). Les protéines fixant les NLS, les nucléoporines, la fixation des protéines aux pores et le transport dépendant de l'ATP ont été démontrés. Les autres étapes restent hypothétiques. NBP: Nuclear Binding Protein; NLS: Nuclear Localisation Signal.

(d'après Silver, 1991)

Caractérisation des signaux de localisation nucléaire, des propriétés de fixation à l'ADN et de transactivation des isoformes codées par *Pax-6 (Pax-QNR)*.

Un facteur de transcription doit être transporté dans le noyau pour pouvoir agir sur la transcription. Ce transport se fait par l'intermédiaire de séquences peptidiques appelées signaux de localisation nucléaire (NLS : Nuclear Localization Signal). Ces séquences vont permettre la fixation de la protéine à d'autres protéines associées aux pores nucléaires au travers desquels le complexe sera ensuite transporté dans le noyau (Cf. Fig.12). Ces signaux de localisation nucléaire sont généralement riches en aa basiques et deux types de consensus NLS ont pu être établis : Lys-X-X-Lys/Arg et Lys/Arg-Lys/Arg-(X-X)₅-X-(Lys/Arg)₃-X où l'aa X est quelconque. Cependant, des NLS ont été caractérisés qui ne présentaient aucune homologie avec ces consensus tels que le NLS du virus de l'influenza qui ne contient pas de série d'aa basiques (voir revue de Dingwall et Laskey, 1991).

Nous avons montré précédemment que les produits de Pax-ONR avaient une localisation subcellulaire différente: les p48 (protéine contenant l'exon 4a supplémentaire dans le domaine PRD) et p46 sont nucléaires et les p43-33-32 sont nucléaires et cytoplasmiques (Table 1)(Fig.2A). Nous avons identifié deux régions de la protéine nécessaires au transport dans le noyau. La première (HNLS) se situe au début de l'homéodomaine (aa 206-212). Il s'agit d'une séquence riche en aa basiques, LeuLysArgLysLeuGluArg, qui correspond à un consensus de nucléarisation. Ce consensus est d'ailleurs conservé dans tous les gènes Pax codant un homéodomaine ou codant au moins la première hélice de ce domaine. Le HNLS est responsable de la translocation partielle des p43 et p33-32 dans le noyau car sa suppression suffit à rendre ces protéines complètement cytoplasmiques. Cette suppression n'a qu'un effet mineur sur le transport des p46-48 et ces protéines restent majoritairement nucléaires. La p43 et la p46 diffèrent au niveau du domaine PRD, l'exon 5 étant partiellement épissé dans la p43. Nous en avons déduit que cet exon contenait une séquence suffisante pour la nucléarisation des p48-46 bien que ne contenant pas de consensus basique. De plus, la suppression de cet exon 5 en plus de celle du HNLS, abolit également complètement le transport dans le noyau des grandes protéines (Fig.3A et C, Table 1). Cependant, un des mutants que nous avons construit dont la troncature recouvre la boîte homéo et une partie de la région carboxy-terminale, $\Delta 170-342$, produit une protéine qui a une localisation nucléaire et cytoplasmique alors que le mutant $\Delta 170-270$ dont la troncature s'arrête à la fin de la boîte homéo code une protéine nucléaire (**Table 2**). Ces résultats montrent qu'un troisième NLS est présent dans la protéine, au début du domaine carboxy-terminal, bien qu'il ne soit pas suffisant pour le transport dans le noyau en l'absence d'une des deux séquences caractérisées.

Enfin, les différentes expériences de transfections ne permettent que rarement d'obtenir des répartitions cytoplasme/noyau de 100% sur l'ensemble de la culture (**Tables 1** et **2**). Cela suggère que le transport dans le noyau doit être régulé par d'autres facteurs, en particulier l'état des cellules et leur position dans le cycle cellulaire.

Dans la séquence codante, nous avons identifié la boîte PRD comme étant une cible importante des mécanismes d'épissage alternatif. Nous avons donc étudié les capacités de fixation à l'ADN des différentes isoformes. Nous avions déjà montré que les p33-32 ne reconnaissaient pas la séquence e5. Excepté la p46, aucune des autres protéines n'est capable de se fixer sur cette séquence cible (**Fig.4A**). L'article de Plaza et al. (1993) montre que la p46 peut fixer et transactiver son propre promoteur, la fixation se faisant sur une séquence de 240 pb Sca1-Asp718. Aucune des autres protéines ne reconnaît cette dernière séquence ainsi que nous avons pu le montrer par des expériences d'immunosélection à l'aide du sérum 14 (**Fig.4B**). Cependant, une séquence cible du domaine PRD contenant l'exon 4a a été récemment décrite (Epstein et al., 1994). Nous montrons que la p48 peut se lier à cette séquence mais avec une affinité de l'ordre de 15% de celle de la p46 (**Fig.4C**) ce qui est surprenant car, avec des domaines PRD isolés, le domaine de la p48 reconnaît cette séquence plus efficacement que celui de la p46 (Epstein et al., 1994).

Un cas d'aniridie a été décrit qui serait du à une mutation ponctuelle dans la séquence codant le HNLS. Cette mutation provoque le changement de l'Arg 208 en Trp (Hanson *et al.*, 1993). Ayant tous les outils pour étudier un éventuel effet de cette mutation sur la localisation subcellulaire du produit de *Pax-QNR*, nous avons introduit cette mutation ponctuelle dans les différentes isoformes. Cependant, aucune modification de l'activité des protéines n'a été observée, que ce soit au niveau du transport dans le noyau (**Fig.3A** et **C**, **Table 1**), de la durée de

			Subcellular Localization	DNA Binding	Transactivating Activity
		p46	Ν	+	+
		p43	N, N/C	_	0
E4a	V/////	p48	Ν	-	0
E4a		p44	N, N/C		0
		p33/32	N/C	-	0
		p46 ∆21-47	Ν	-	0
		p46 ∆48-211	N/C, C		0
		p46 ∆206-212	Ν	+	+
		p46 R ²⁰⁸ /W	N	+	+
	W208	p46 ∆170-270	Ν	+	+
		p46 ∆170-342	N/C	+	0
		p46 ∆271-329	Ν	+	+
		p46 ∆331-379	Ν	+	+
	YIIII X	p46 ∆333-416	Ν	_	_

Fig. 13 : Tableau récapitulatif

demi vie, de la liaison à l'ADN (Fig.6D), ou de la capacité à transactiver les séquences promotrices P0 de Pax-QNR (Fig.6E).

Dans le but de définir le domaine transactivateur du produit de *Pax-QNR*, nous avons construit plusieurs mutants de délétion (**Fig.6A**) et leurs comportements ont été analysés (Cf. Fig. 13). Transfectés en cellules COS, ces mutants sont accumulés de la même façon dans les cellules excepté le mutant Δ BalI (Δ 333-416)(**Fig.6B**). Celui-ci, dont la troncature dans le domaine carboxy-terminal génère une fusion avec 10 aa provenant initialement de la région 3' non codante, a une demi vie réduite à environ 3 heures. (**Fig.6C**) Il est donc instable par rapport à la p46 (demi vie de 6 heures).

Comme nous l'avons montré pour les isoformes, toute altération du domaine PRD abolit la fixation à l'ADN. Ainsi, les mutants tronqués dans ce domaine, que ce soit dans la première hélice (Δ 21-47) ou dans les deux suivantes (Δ 48-211), ne peuvent plus reconnaître la séquence de 240 pb Sca1-Asp718 (**Fig.6D**). Par contre, les mutants tronqués dans l'homéodomaine comme le Δ (170-342) ou le Δ (170-270), ou dans la région carboxy-terminale, comme le Δ (331-379) ou le Δ (271-329), restent capables de se lier à cette séquence (**Fig.6D**). Seul, le mutant Δ BalI qui possède un domaine PRD complet, ne se fixe plus. Certes, la protéine codée par ce mutant est instable, mais produit en synthèse *in vitro*, elle reste incapable de se lier à l'ADN. Pour ce dernier mutant, il y a probablement un problème de conformation de la molécule qui peut être lié à la suppression partielle de la partie carboxy-terminale ou à l'ajout des 10 aa suite au changement de cadre produit par la mutation.

Les résultats de Glaser et al. (1994) montrent que le domaine transactivateur du produit de *Pax-6* se situe en aval de l'homéodomaine. En effet, ce domaine cloné en fusion avec le domaine de fixation à l'ADN de la protéine GAL4 est capable de transactiver une séquence cible de GAL4. De plus, ce domaine tronqué à partir de l'aa 352 (aa 346 pour notre ADNc qui ne contient pas l'exon 10 rallongé de 6 aa) conserve une activité transactivatrice partielle (Glaser *et al.*, 1994). Nous montrons que deux mutants tronqués de la fin de l'homéodomaine, aa 271, à l'aa 329 et de l'aa 331 à l'aa 379 transactivent aussi efficacement que la p46 (**Fig.6E**). Ces résultats suggèrent que le domaine transactivateur se situe en aval de l'aa 379. Cependant, le mutant Δ KpnI qui est tronqué en carboxy-terminal jusqu'à l'aa 342 ne transactive plus, bien qu'il soit toujours capable de se lier au promoteur de *Pax-6* (**Fig.6E**). La région située en aval de l'aa 342 n'est donc pas suffisante, seule, à l'activité transactivatrice. L'addition des aa 271 à 329 (construction $\Delta 170$ -270) restaure cette activité (**Fig.6E**).

Ces résultats montrent donc que deux régions sont impliquées dans la transactivation du promoteur P0 de *Pax-QNR* : la partie amino-terminale (aa 270 à 341) et la partie carboxy-terminale (aa 380 à 416). La différence entre nos résultats et ceux de Glaser *et al*. (1994) peut être liée au fait que nos expériences ont été faites dans le cadre de la protéine totale.

Nuclear Localization Signals, DNA-binding and Transactivation Properties of Quail Pax-6(Pax-QNR) Isoforms

Catherine CARRIERE, Serge PLAZA, Jocelyne CABOCHE[&], Christine DOZIER, Manuella BAILLY, Patrick MARTIN⁺, and Simon SAULE^{*}

CNRS EP 56 and ⁺URA 1160 Institut Pasteur de Lille, 1 Rue Calmette, 59019 Lille cedex. &Present address : Laboratoire de Neurochimie Anatomie, 9 quai S^t Bernard, Université Paris VI 75005 Paris.

*To whom correspondence should be addressed.

Phone number : 20 87 77 86, Fax number : 20 87 79 08.

Running Title : Properties of Quail Pax-6(Pax-QNR) Isoforms.

Key words : neuroretina / Pax-6 isoforms/ development / nuclear localization.

Abstract

We previously reported the characterization of Pax-QNR/Pax-6 products from a pairedbox and homeobox-containing gene expressed in the avian neuroretina. Five proteins (48, 46, 43, 33, and 32 kDa) were characterized, among which the 33 and 32 kDa are devoid of the paired domain. In contrast to the 48 (extended by the presence in the paired domain of the exon 4a) and 46 kDa proteins exclusively located in the nucleus, the 43, 33 and 32 kDa proteins were also found in the cytoplasmic compartment. We report the identification of the basic LKRKLQR region (amino acids 206-212) located in the N-terminus of the homeodomain as a nuclear localization signal used by the p43 and 33/32 kDa proteins. Deletion of this sequence in the p46 or 48 kDa proteins has only a minor effect in the protein localization, but renders the p43 and the 33/32 kDa proteins exclusively cytoplasmic. Since the difference between p46 and p43 lies in the paired exon 5 spliced out to generate the p43, this indicates that this exon contains a nuclear targeting signal dominant over the signal present in the homeodomain. A case of human Aniridia where the arginine 208 of the LKRKLQR is mutated into a tryptophane has been recently reported. We introduced this mutation in the Pax-QNR p46, p43 and p33/32 proteins. No effect on the nuclear localization, nor in transactivation potential of the proteins could be observed. Among the several Pax-QNR isoforms characterized, only the p46 exhibited DNA-binding and transactivating properties on the Pax-QNR promoter. Deletions of parts of the protein showed that the Pax-6 transactivation domain is located in the carboxyl terminus of the protein, and that the presence of the p46-paired domain is required for both the DNA-binding and transactivating properties on the Po Pax-QNR promoter.

Introduction

Cellular differentiation is the result of differential gene expression and transcription factors are involved in the regulation of these events. The paired-box encodes the paired domain, a conserved DNA-binding domain (28; 32), and nine vertebrate paired-containing genes have been identified (29; 32). Some of these genes (like *Pax-6*) contain, in addition to the paired domain, another DNA-binding motif, the homeodomain. The homeobox encodes the homeodomain, a conserved DNA-binding domain with a helix-turn-helix motif (20). Mutations in the *Pax-6* gene have been associated with the mouse mutant *Small eye* (18), the corresponding human gene (*AN*) has been found deleted or mutated in aniridia and other eye anomalies (10; 11; 15; 16; 27) and the *Drosophila eyeless* gene is *Pax-6* (25). The *Pax-6* gene is expressed in the developing central nervous system, the optic cup, lens, overlaying epithelium, in the neuronal layers of the retina and the endocrine pancreas (31; 22; 30). We have previously isolated the quail homologue of *Pax-6* termed *Pax-QNR* (7), and characterized five proteins synthesized *in vivo* (2). Three proteins of 48, 46 and 43 kDa contained the paired domain, but two proteins of 33 and 32 kDa are devoid of this DNA-binding domain.

The aim of this work was to further study the various isoforms of the Pax-6 proteins and more precisely to analyse the nuclear localization signals, DNA binding and transactivating properties of these isoforms. In a first step, we characterized the nuclear localization signals present in the Pax-6 proteins. Two distinct signals were found, one in the exon 5, in the paired domain, the other (LKRKLQR) in the exon 7 at the beginning of the homeodomain. The signal localized in the paired domain is dominant over the signal localized in the homeodomain since paired-containing proteins lacking the Homeo Nuclear Localization Signal, (HNLS, LKRKLQR) are still nuclear in most of the transfected cells. In the case of p43 or p33/32 this deletion rendered the proteins exclusively cytoplasmic. Since the R208 to W substitution in the LKRKLQR sequence has been described in an Aniridia patient, (16) we performed site directed mutagenesis of the 208R to W in the Pax-QNR gene. This mutation did not modify the nuclear localization of either the p46, 43 or paired-less 33/32 Pax-QNR proteins thus suggesting that this mutation is not acting through a wrong subcellular localization of the mutated Pax-6 proteins. We previously reported that the p46 was able to transactivate its own promoter (24). In a second step, we tested this property for the various isoforms with or without HNLS, and with or without the R208 to W mutation. In contrast to the p46 or the R208 to W mutated p46, none of these isoforms were able to transactivate this promoter, nor to bind the target DNA. Transactivation studies performed with deleted p46 proteins indicate that the carboxyl terminus part of Pax-QNR bears the transactivation domain.

Materials and methods

Cells and plasmids used. Neuroretinas were dissected from quail embryos from E4 to hatching. Quail embryos cells (QEC) cultures were maintained and passaged in Dulbeco Modified Eagles Medium supplemented with 10 % foetal calf serum. COS-1 cells were maintained in the same medium. Pax-QNR mutants were prepared from the *Pax-QNR* MC29-QNR2 DNA shortened at SstII site in 5' and inserted in the pSG5 expression vector. The mutant deleted from amino acids 48 to 211 resulted from a PstI fragment deletion ; The mutant $\Delta 170$ to 342 resulted from a Kpn1 fragment deletion ; The mutant $\Delta 331$ to 380 resulted from the Nco1 fragment deletion. The mutant $\Delta 333$ to 416 resulted from the Bal1 fragment deletion. In this mutant the 10 last amino acids were encoded by the *Pax-QNR* nucleotide sequence out of frame.

Cell labelling, immunoprecipitation, western blotting and RNAase protection. Transfected COS cells were incubated for 45 min. in the presence of 100 μ Ci ml⁻¹ of L-³⁵Smethionine (Amersham, specific activity of 1000 Ci mmol⁻¹), lysed in RIPA-buffer and immunoprecipitated with rabbit anti-Pax-QNR sera (serum 11 against the paired domain, serum 12 against the amino acids located between paired and homeo, serum 13 against homeo and serum 14 against the carboxyl-terminal part of the protein). Immunoprecipitated proteins were analyzed by SDS-PAGE followed by fluorography.

Western blots were performed with 10^6 neuroretina cells or $3\mu g$ of cellular COS extracts boiled in loading buffer, subjected to SDS-PAGE and electrophoretically transferred to Immobilon membranes (Millipore). Filters were treated for 60 min. in blocking buffer (5% non-fat dry milk in PBS). Filters were then treated over night in blocking buffer containing rabbit serum 13 diluted 1: 200, followed by 30 min. washing in blocking buffer, 0.1% Tween 20. Bound antibodies were revealed with an ECL detection kit (Amersham). RNase protection was performed as previously published with the probes described below (24).

Immunofluorescence study. Transfected cells (COS or QEC) 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 11 and 14.

PCR experiment, DNA sequencing and *in vitro* mutagenesis. Total RNA isolated from E7 QNR (15 μ g) was converted to cDNA using 0.5 μ g oligo dT primer (RT primer). Reverse transcription was carried out as described in ref.2, with 0.25 μ g of the appropriate oligonucleotide primers. Primers were 5'-ATCCTGCAGACCCATGCAGATGCAAAAGTC-3', and

5'-ATCAAGCTTAGATGGAGTTGGTGTTCTC-3' for the junction between exon 3 and 4a and the beginning of exon 7 respectively. Amplifications were carried out for 40 cycles in a Eurogentec thermojet : 1.0 min denaturation at 94°C, 1.30 min annealing at 50°C, 1.0 min elongation at 72°C. Reaction products were analyzed by electrophoresis through a 2% agarose gel. Amplified fragments were ligated between the Pst1-HindIII sites of pUC19. Nucleotide sequence was determined in both orientations by the dydeoxynucleotide chain-termination method using a ³⁵S sequencing kit (Pharmacia). The amplified fragments containing exon 4a (with and without exon 5) were digested by Pst1-Nhe1 restriction enzymes, and inserted after two steps, in the MC29-QNR2 cDNA cloned in pSG5 vector digested with the same restriction endonucleases. To obtain the Pax-QNR isoforms containing the extended exon 10, the MC29-QNR2 Nhe1-HindIII fragment was replaced in each construct, by the Nhe1-HindIII fragment from the B1 cDNA.

The mutagenesis on Pax-QNR conducted in order to replace the R208 to W has been made on the Asp718 fragment inserted in the M13 mp18 phage DNA digested with the same enzyme. The single stranded directed mutagenesis was conducted with a mutated oligonucleotide located between nucleotides 986 to 1015 (22). Oligonucleotide used is as follow :

5'-CTTCAGTTGAAATGGAAGCTGCAGAGAAAT-3'. In bold letters is the mutated nucleotide. The mutation of an AGG into TGG is expected to result in the conversion of an Arg into Trp in the protein. The mutagenesis on Pax-QNR conducted in order to remove the 206-LKRKLQR-212 has been made on the Asp718 fragment inserted in the M13 mp18 phage DNA digested with the same enzyme. The single stranded directed mutagenesis was conducted with a

mutated oligonucleotide located between nucleotides 974 to 1027. Oligonucleotide used is as follow :

5'-GCCCAGATGAGACTTCAGAATAGGACATCCTTT-3'. The mutagenesis on Pax-QNR conducted in order to remove the amino acids 21 to 47 has been made on the HincII-Xma1 fragment inserted in the M13 mp19 phage DNA digested with the same enzymes. The single stranded directed mutagenesis was conducted with an oligonucleotide complementary to the Pax-QNR sequence located between nucleotides 422 to 532. Oligonucleotide used is as follow :

5'-CAGTTGCCCTCCGGCCACAGCTTACCTACA-3'. The mutant $\Delta 170$ to 270 was produced as follow: the carboxyl terminus of Pax-QNR was amplified using the following oligonucleotide primers :

5'-AGGATGGTACCCCTGCGGAATCAGCGGAGA-3', containing a Kpn1 site and located between nucleotides 1187 to 1204 and the oligonucleotide reverse

5'-GATAAGCTTGTTGATCATGGCTTC-3', containing a HindIII site, and located between nucleotides 1199 to 2119. The amplified fragment digested by Kpn1-HinDIII was then inserted in the Pax-QNR mutant $\Delta 170$ to 342 digested by the same enzymes. The mutant $\Delta 271$ to 329 was produced as follow : the amino terminus of Pax-QNR from SstII to the end of the homeodomain was amplified using the following oligonucleotide primers:

5'-AGCCCCGCGGGGACCCCGCTGCCCC-3', containing a SstII site and located between nucleotides 345 to 368 and the oligonucleotide reverse

5'-GGCCATGGTCTTCTCCTCCCTTCT-3', containing a NcoI site, and located between nucleotides 1172 to 1186. The amplified fragment digested by SstII-NcoI was then inserted in the Pax-QNR digested by the same enzymes.

The probes for the RNAase protection were produced as follow. The paired fragments were amplified using the following oligonucleotides primers :

5'-TTCAGTCGACATCCTGCAGACCCATGC-3' for the p48 cDNA located between nucleotides 158 to 175 and

5'-TTCGTCAACGGGAGGCCGC-3' for the p43 cDNA located between nucleotides 73 to 91. The same oligonucleotide reverse is used for the two probes

5'-TACGGTCGACTTCGCTAGCCAGGTTGCG-3', located between nucleotides 450 to 467 for the p48 cDNA and between nucleotides 211 to 228 for the p43 cDNA. The three oligonucleotides contained a HincII restriction site. The amplified fragments digested by HincII were cloned in the pGEM3Z vector digested with the same enzyme.

Cellular extracts preparation and gel mobility-shift assays. $5x10^5$ COS-1 cells were transfected by 3 µg of ADN, using 18 µl lipofectamin (Gibco-BRL). Cells extracts were prepared essentially as described in ref.1. All buffers contained proteinase and phosphatase inhibitors. Binding reactions were performed for 20 min on ice, using 1 ng radiolabelled DNA probe and 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 Na₂HPO₄ (pH 7.2) and 3 µg poly dI-dC. For antibodies competition experiments, 1µl of serum 14 was incubated with the cellular extracts 20 min. on ice before adding the probe. Samples were loaded on a 6 % native polyacrylamide gel, run for 2 h at 180 V and examined by autoradiography after exposure of the dried gel to a Kodak X-AR film at -70°C with an intensifying screen.

To prepare the DNA probes for gel shift analysis, an oligonucleotide containing the e5 binding site CTCAGC<u>ACCGCACGATTAGC</u>ACCGTTCCGCTTC and its complementary strand were synthesized, annealed and used as DNA probe after 5'end labelling. The recognition sequence bound by the homeodomain is underlined and the paired domain recognition sequence is shown in bold print. The 5aCON oligonucleotide and its complementary strand allowing the binding of the exon 4a containing paired domain were produced as follow : two oligonucleotides partially complementary were produced :

5'-AATAAATCTGAACATGCTCAGTGAATGTTC-3' and

5'-TGACCTCGAGAGTCAATGAACATTCACTGA-3', annealed, the remaining sequence was filled using Klenow polymerase. The probe was radiolabelled by kinasing the oligonucleotides with (γ^{32} P)ATP.

Cell transfection and CAT assays. QEC were seeded at 10^5 cells per 35mm diameter plates. Transfections were performed with the calcium phosphate method, with 1µg of pP0XACAT construct, pXA (24) and 5µg of pSG5 expression vectors encoding the different Pax-QNR isoforms or mutants. 48h after transfection the cells were lysed in the reporter lysis buffer (Promega). CAT assays were performed on cell extracts as described in ref. 13, using equal amounts of total proteins, which were determined by the Bradford method (Bio-Rad) and quantified with a phosphor imager.

Pax-QNR-DNA immunoprecipitation assays. The immunoprecipitation procedure was carried out as described previously (17) using the 214 bp ScaI-Asp718 probe, transfected COS-1 cells extracts (40 μ g), and 1 μ l of antiserum directed against the carboxy-terminal part of the Pax-QNR protein. The Protein A-sepharose beads used to collect the immunoglobulins were pre-incubated in Dignam buffer A (4) in the presence of 1% (P/V) Bovine Serum Albumine and 100 μ g/ml of sonicated salmon sperm DNA. After washing, the labelled DNA was recovered by phenol/chloroform extraction and loaded on a 6% polyacrylamide gel in TBE buffer.

Results

Identification of two putative nuclearization signals in Pax-QNR

As shown in Figure 1A and 1B, the Pax-QNR isoforms are differentially accumulated in the differentiating neuroretina *in vivo*. The p46 remains the major form of Pax-QNR protein throughout the neuroretina development, the paired-less 33/32 kDa doublet proteins rather accumulated when the neuroretina becomes post-mitotic. The localization of Pax-QNR proteins (paired-containing and paired-less products) has been studied by cell fractionation and immunofluorescence experiments (2). Neuroretina-cell fractionation of 7-day-old chicken neuroretinas (³⁵S) methionine-labelled for 45 min. followed by immunoprecipitation with serum 14 (recognizing the Pax-QNR carboxyl-terminus) showed that the p46/48 paired-containing proteins are found mostly in the nucleus (7% of the total amount is cytoplasmic), whereas the p43 (51% of the total amount) and the 33/32kDa products (79% of the total amount) were mostly cytoplasmic (Figure 1C).

The paired-containing 48 and 46 kDa proteins are produced from two already identified *Pax-QNR* transcripts : the 48 kDa product is encoded by the mRNA containing the additional paired exon 4a resulting in the insertion of a 14-amino-acid peptide in the paired domain (7; 11; 31) and the 46 kDa by the MC29-QNR2 mRNA (22). We recently identified the cDNA Q6 corresponding to a transcript for the paired-containing 43 kDa product (2). This cDNA is generated by an alternative splicing event in the 5'UTR between exons 0 and 2, and between the beginning of exons 5 and 6, resulting in a paired domain 66 amino acids shorter than the one usually observed (Figure 2A and 2B). RNase protection experiments performed with RNA probes

able to distinguish the variant paired domains show that the ratio between all the Pax-QNR isoforms-encoding messengers is constant throughout the neuroretina development (Figure 2C and data not shown), with an increase in expression following neuroretina development (22).

Thus, the substantial cytoplasmic localization of the p43 protein synthesized *in vivo* indicates that a nuclear localization signal is located in the paired exon 5. The paired-less 33/32 kDa doublet proteins found in the neuroretina is most probably produced from internal AUG initiation, this initiation codon is localized in the sequence between the paired and the homeodomain boxes (2). These proteins are mostly localized in the cytoplasm, but immunofluorescence data obtained from transient transfection experiments performed with expression vectors encoding the paired-less 33/32 kDa doublet proteins demonstrate that these proteins also contained a nuclear localization signal (2). Therefore, at least two distinct nuclear localization signals are present in the paired-containing Pax-QNR proteins.

Characterization of the paired-less p33/32 proteins nuclearization signal sequence

Nuclear targeting-like sequence LKRKLQR (5; 6; 26) is present in the exon 7 in the Nterminally extended paired-type homeobox and well conserved in all homeobox-containing *Pax* genes. To study the possible role of LKRKLQR as nuclearization signal sequence, we removed this sequence from p33/32, p43 and p46 (Figure 3A) by site directed mutagenesis in Pax-QNR isoforms-encoding vectors. Immunoexperiments (Figure 3C) showed that most of the QEC transfected with the Δ Nhe Δ HNLS (97%, see Table 1) exhibited a cytoplasmic labelling with serum 14 (Figure 3C, panel f), in contrast to the Δ Nhe where all the labelled cells exhibited a nuclear and cytoplasmic signal with serum 14 (Figure 3C, panel e and Table 1). Thus, this sequence is involved in the nuclear localization of the p33/32 paired-less proteins.

When exon 5 was removed from the *Pax-QNR* paired box, the resulting p43 protein was localized in both the cytoplasmic and nuclear compartment (Figure 3C panel d), in contrast to the exon 5-containing proteins p48 (Figure 3C panel a) or p46 (Figure 3C panel b). However, a significant amount of the p43 transfected cells (31%, see Table 1) exhibited a nuclear localization only. The p43 Δ HNLS showed that 80% of the transfected cells lacked a nuclear labelling above background (Figure 3C panel g), 20% of the cells still exhibited a cytoplasmic and nuclear labelling, but no cell with an exclusive nuclear localization could be found. In contrast, the p46 Δ HNLS showed that 99% of the transfected cells exhibit exclusively a nuclear fluorescence

and only 1% of the cells exhibit a cytoplasmic and nuclear labelling (Figure 3C panel h). Altogether, these results suggest that cell-dependent event is also involved in the Pax-QNR subcellular localization.

To study the possibility that exon 4a encoded 14 amino-acids could be involved in the nuclear localization process, we used an expression vector encoding the p44 Pax-QNR product. The 44 kDa protein is expected to correspond to the p43 containing the additional exon 4a. PCR experiment performed with neuroretina RNA suggests that this particular splicing in the paired domain occurs *in vivo*. As shown in Table 1, the QEC transfected with the p44 vector (Figure 1C) exhibited the same cytoplasmic and nuclear fluorescence staining with serum 14 than with the 43 kDa encoding vector (Figure 3C panel c).

Since the LKRKLQR sequence is clearly involved in the nuclear localization of the p33/32 paired-less proteins, we studied the subcellular localization of the protein encoded by the Δ Nhe vector containing the R208 to W substitution in the LKRKLQR signal (Δ NheW, see Figure 3B). Such mutation has been found in a case of Aniridia (16) and suspected to modify the activity and/or the localization of the mutated protein. However, the Δ NheW encoded proteins were localized in both the nuclear and cytoplasmic compartments (Table 1 and Figure 3C panel I), suggesting that the R208 to W substitution is not able to influence the sub-cellular localization of the p33/32 paired-less protein in the transfected QEC. Not surprisingly, no effect of such mutation could be observed in the subcellular localization of the p46 (Figure 3C panel j) or the p43 (Table 1) proteins.

The p46 but not the other Pax-QNR proteins can bind the e5 or P0 promoter DNA sequences

To determine whether the different Pax-QNR isoforms were able to specifically recognize the e5 sequence (as demonstrated for the 46 kDa translated in reticulocyte lysates, 7 or neuroretina nuclear extracts, 2) or the P0 promoter sequence (24), EMSAs and immunoselection assays were performed using the nuclear extracts of COS cells transfected with the various Pax-QNR isoforms encoding vectors (Figure 1C) and ³²P-labelled e5 or P0 DNA (Figure 4A and 4B). Negative results have been already published for the p33/32 isoforms, both on e5 and P0 DNA targets (2; 24). The e5 DNA-binding complex found in nuclear extracts is pointed out by an arrow (Figure 4A). This complex is specific since it disappeared in the presence of serum 14 (lanes +). This complex is seen only in the presence of the p46 and p46* (containing the alternative exon 10 (7)) coding vectors. No complex was detected with the nuclear extracts of the COS cells transfected with the others Pax-QNR isoforms encoding vectors. Immunoselection assay was performed with the *Pax-QNR* P0 promoter region. The labelled 214 bp ScaI-Asp718 DNA probe (24) was incubated with the nuclear extracts from the transfected COS cells and then with anti Pax-QNR serum 14. The DNA probe that was immunoprecipitated as a result of its association with the Pax-QNR protein was visualized on a polyacrylamide gel. As shown in Figure 4B, the probe was immunoprecipitated by the anti Pax-QNR serum only when p46 and p46* encoding vectors was used to transfect the COS cells. It has been recently reported that the paired domain containing the alternative exon 4a was able to bind a distinct DNA target, called 5aCON (8). We studied the binding potential of exon 4a containing proteins, the 48 kDa and the 44 kDa Pax-QNR products, on this particular DNA target. In contrast to what has been reported for the purified paired domain, the p48 isoform binds only slightly to the 5aCON labelled DNA, and the p48-DNA complex represents only 15% of the p46-DNA complex formed (Figure 4C). No binding was observed with the p44 protein.

To confirm that the Pax-QNR proteins are indeed present in similar amount in the nuclear extracts, we performed a western blot experiment with the same amount of proteins used to perform the EMSAs. Comparable amount of the Pax-QNR isoforms are found in these cells (data not shown).

The p46 but not the other Pax-QNR isoforms can transactivate the *Pax-QNR* P0 promoter DNA

To study the possibility that the Pax-QNR isoforms could transactivate the P0 *Pax-QNR* promoter by a DNA-binding independent mechanism, the different expression vectors were tested in transient co-transfection assays with the pXA reporter vector. This vector contains the 1.5 Kbp XbaI-Asp718 P0 fragment inserted upstream of the bacterial CAT gene of pBLCAT5 (24). The pXA DNA was co-transfected into QEC with 5µg of a Pax-QNR encoding vector. Figure 5 shows that only the p46 and p46* encoding vectors exhibited a transactivating property. This result demonstrates that Pax-QNR DNA-binding is required for Pax-QNR responsive transactivation.

Localization of the Pax-QNR transactivation domain in the p46 protein

Using the pXA reporter vector as a target to study the Pax-QNR transactivation domains, we prepared several Pax-QNR deletion mutants (Figure 6A). We first studied, by western blot analysis, the accumulation of the mutant proteins in the transfected cells. The Pax-QNR deleted constructs produced similar amount of proteins in the transfected COS cells (Figure 6B) excepted the Δ Bal I (Δ 333-416). The protein encoded by this construct ends at the amino acid 332 and is followed by 10 amino acids out of frame. Since we found that this protein is produced in reticulocyte lysate and could be detected by immunoprecipitation from the transfected COS cells, we tested the stability of this protein. After a chase analysis following a 45-min pulse with (^{35}S) methionine, the half-life of the Δ Bal I protein mutant was found significantly reduced, when compared to the wild type p46 half-life (Figure 6C) and this mutant is the only one encoding an unstable protein (half life of 3 hours). Next we studied the DNA-binding properties of the deleted proteins on the P0 promoter by immunoselection assays. As already shown for the Pax-QNR isoforms, when the paired domain is truncated (mutants $\Delta 21-47$ and $\Delta 48-211$) the proteins are unable to bind the 214 bp Sca1-Asp718 DNA target (Figure 6D). Surprisingly, the ΔBal I (Δ333-416) deletant which contained a complete paired domain is unable to bind the P0 DNA. The relative protein instability of this mutant is not sufficient to explain this negative result, since a similar lack of binding is observed with a $\Delta Bal I$ protein produced in a reticulocyte lysate (data not shown), suggesting that a conformational event is involved in this negative result. Deletion of the paired amino acids removing the first α -helical region produced a Pax-QNR mutant ($\Delta 21-47$) unable to bind the 214 bp Sca1-Asp718 DNA target or the e5 probe, although able to recognize the 5aCON DNA (Epstein et al., 1994) but with only 10% of the 46 kDa protein efficiency (data not shown).

Study of the transcriptional activity of the different Pax-QNR mutants (Figure 6E) indicates that the paired DNA-binding activity of the protein is required to transactivate (deletants Δ 21-47 and Δ 48-211 are devoid of activity) and as already suggested in ref (10), the carboxyl terminus of Pax-QNR contains the transactivating domain. A Pax-QNR protein deleted from the Leu 271 to Phe 329 transactivates as efficiently as the wild type 46 kDa protein. Another mutant, deleted from the Met 331 to the Pro 379 also transactivates as well as the positive control suggesting that the critical domain is located after the Met 380. However, the protein encoded by

the Δ Kpn I mutant, (Δ 170-342) deleted in the carboxyl terminus up to the Ser 343, gave no measurable increase in the CAT activity. Since this deleted protein is well accumulated in the transfected cells (Figure 6C, lane $\Delta 170-342$) and able to bind the 214 bp Sca1-Asp718 DNA target (Figure 6D, lane Δ 170-342), the observed lack of transactivation indicates that the Pax-QNR Pro Ser Thr rich (PST) domain downstream the Ser 343 is not sufficient alone to increase the CAT activity from the Pax-QNR promoter. Addition of the amino acids 271 to 342 to this mutant restored a full transcriptional activity (construct $\Delta 170-270$, Figure 6E). This mutant also demonstrates that the homeodomain is not required for transcriptional activity. Altogether, these results suggest that both the amino-terminus (from the amino acids 271 to 342) and the carboxyl terminus (from amino acids 380 to 416) of Pax-QNR PST domain are involved in order to induce the P0 transcriptional response. However, as shown in Table 2, the Δ Kpn I mutant encodes a protein ($\Delta 170-342$) located both in the cytoplasm and the nucleus (100% of the cells) whereas the mutant $\Delta 170-270$ encodes a protein exclusively located in the nucleus in 98% of the cells. Since they both contain the dominant nuclear localization signal encoded by the paired exon 5 and lack the homeodomain nuclear localization signal encoded by the exon 7, this suggests that an additional sequence plays a role in the protein nuclear localization. This sequence is expected to be in the 71 amino acids downstream the homeodomain, between the Leu 271 and the Val 342. It remains to be tested if the nuclear amount of the Pax-QNR Δ Kpn I mutant is sufficient to allow an efficient transactivation or not.

The Δ Bal I protein ending at the Ala 332, was found able to slightly inhibit the basal transcription of the Pax-QNR promoter (Figure 6E). These results are reminiscent of Pax-3 induced inhibition of the basal transcription of N-CAM promoter reporter constructs, by a DNA-binding independent mechanism (3). In addition, Figure 6E shows that the Δ HNLS as well as the R208 to W p46 mutants transactivate the *Pax-QNR* promoter as efficiently as the wild type p46 or p46* proteins. Figure 7 summarizes the results obtained with the different isoforms and mutants.

Discussion

We have previously described Pax-QNR (7; 22), the quail homologue of the Pax-6 gene. This gene encodes at least five distinct proteins *in vivo*, and is able to transactivate its own promoter *in vitro* (2; 24). We used the distinct Pax-QNR protein expression vectors in order to study the function of the different Pax-QNR isoforms and to characterize the different protein domains.

The subcellular localization of the different isoforms suggested that, at least two domains were involved in the nuclear localization of the protein. Transient transfection experiments with expression vectors encoding the p46 or the paired-less p33/32 proteins revealed that both types of proteins are nuclear, even if the 33/32 kDa proteins show a significant cytoplasmic localization. The N-terminally extended homeobox contains a sequence LKRKXXRXR, well conserved in all homeobox-containing Pax genes and matching well with a consensus nuclear localization signal (6). Deletion of the sequence encoding the 206-LKRKLQR-212 peptide in the Δ Nhe I vector (2) resulted in the synthesis of 33/32 kDa mutant proteins located in the cytoplasmic compartment only. However, this sequence (HNLS) was not sufficient to direct a non-nuclear protein (β galactosidase) to the nucleus if inserted either in its amino- or carboxyl terminus (data not shown) suggesting that other amino acids may be required to direct the protein into the nucleus. An Aniridia mutant in the PAX-6 gene with a R208 to W transition has been reported (16). However, a point mutation introducing a W to replace the R208 in the Δ Nhe I vector encoded protein is without effect on the subcellular localization or protein stability of the encoded protein. Therefore, the mutation responsible for the Aniridia could be outside the Pax-6 gene as reported for some patients (27) and the R to W transition could be an allelic variant. Alternatively, it remains possible that the mutated protein is no longer able to recognize some critical unknown target genes of Pax-6.

The nuclear localization of the paired-containing protein is only barely affected by the deletion of the HNLS sequence, demonstrating that the nuclear localization signal present in the paired domain exon 5 is dominant over the homeodomain signal. This is perhaps the result of an evolutionary process since several *Pax* genes are devoid of homeodomain (29). Deletion of the HNLS sequence in the 43 kDa encoding vector (paired-containing isoform devoid of exon 5)

resulted in the synthesis of a cytoplasmic protein in most of the cells, suggesting that no other part of the paired domain exert a dominant effect on the nuclear localization of Pax-QNR proteins. A protein deleted from most of the exon 4 (Pax-QNR $\Delta 21$ -47) exhibits a wild-type nuclear localization, despite the deletion of the 4 basic residues in the exon 4 encoded amino acids. As for the Δ HNLS, the exon 5 sequence is not sufficient to direct the β -galactosidase into the nucleus, even when associated with the HNLS sequence (data not shown). Therefore, another site of the Pax-QNR protein, unable alone to direct the protein into the nucleus, is expected to act in addition to the exon 5 and the HNLS motifs. Indeed, whereas the p46 Δ HNLS is nuclear, the protein $\Delta 170$ -342 encoded by the Δ Kpn 1 deletant and containing the paired exon 5 nuclear localization signal is nuclear and cytoplasmic, like the p43 or the p33/32 proteins devoid of this sequence. The Pax-QNR deletant $\Delta 170$ -270, devoid of the HNLS is nuclear like the p46 Δ HNLS mutant. Thus, the 71 amino acids 271 to 342 are expected to contain this third nuclear localization signal.

The functions of all the Pax-QNR isoforms remain to be defined. The p46 Pax-QNR protein is able to bind to the Drosophila e5 sequence (7) as well as to the Pax-QNR P0 promoter sequence (24). This binding is mediated by the paired and the homeodomain, as for the Pax-3 protein (3) and both homeo and paired domains are required for high-affinity binding (7). That both domains are involved in DNA-binding is indicated by the fact that both the paired and the homeo domain individually expressed in bacteria are able to bind to e5 (7) and both are able to footprint the 214 bp Sca1-Asp718 Pax-QNR P0 promoter DNA (24). However, we have been unable to demonstrate a DNA binding activity for the 48 kDa isoform either on e5 or on the Pax-ONR P0 promoter sequences. Thus, even if homeo domain binding sites are present on the DNA target, the presence of the 14 amino acids encoded by the exon 4a between the first and the second α -helical regions (8; 32) of the paired domain is sufficient to abrogate the p48 DNA binding, suggesting that the binding of the paired domain is required to allow the binding of the homeodomain. Recently, it has been demonstrated that the p48 paired domain recognized a distinct DNA-target sequence termed 5aCON (8). Pax-QNR p48 isoform was able to recognize this target sequence, but with 15% efficiency only when compared to the p46. With the purified paired peptides, the binding of the p48-type paired domain is more efficient (8) suggesting that structural constraints arising from other part of the molecule modulate the DNA-binding of the paired domain. Such constraints may explain why the Δ Bal I (Δ 332-416) mutant exhibiting a regular paired domain is unable to bind to a DNA target. Also, the paired-less proteins retaining the homeodomain should behave like conventional homeodomain-containing proteins and bind to ATTA rich domains. In fact, genes encoding paired-type homeodomain-containing proteins only have already been described (14; 19). It remains possible that the p33/32 paired-less Pax-QNR products possess specific DNA binding activity, and that the carboxy-terminal part of the protein masks the DNA-binding sequence of the p33/32 homeodomain as suggested for the *Drosophila prd* gene (28). Alternatively, the target sequence recognized by these proteins may require the presence of a palindromic sequence, in order to allow a cooperative dimerization (33). Therefore, the target genes regulated by the Pax-QNR isoforms are expected to be different.

The Pax-QNR/Pax-6 transactivation domain is localized in the carboxy-terminal Pro Ser Thr rich (PST) domain. This has been indicated by the use of a protein containing the yeast GAL4-DNA binding domain fused to the Pax-6 PST domain (10). A mutant PST domain which contains only the first 82 amino acids stimulated CAT expression through GAL4 binding sites at only 5-10% of the complete PST domain (10) suggesting that the transactivation domain lies downstream the Ser 347 (indicated as Ser 352 in ref.10, due to the presence of 6 additional amino acids coming from exon 10 in the human Pax-6 PST sequence, see ref. 7). Using the whole Pax-QNR/Pax-6 protein and the Pax-*QNR* PO promoter reporter, we observed that the PST domain is rather complex. First, deletion of the extreme carboxyl terminus results in the synthesis of an unstable protein. Second, a cooperation between the PST amino (amino acids 271 to 331) and carboxyl domains (amino acids 380 to 416) is required in order to obtain an optimal transactivation. This may explain why the PST domain truncated at the Ser 347 is only efficient at 10% when compared to the complete PST domain. Amino acids 221 to 329 is fully efficient, whereas a mutant deleted up to the amino acid 342 gave no measurable CAT activity.

The Pax-QNR protein deleted from the extreme carboxyl terminus is unable to bind the DNA target and inhibits the basal transcription of the reporter construct. This suggests that, as for Pax-3 (3), Pax-6 may contain transcription inhibitory domains. However, no precise domain can be found in the protein for this inhibitory function, and structural feature may rather be involved in such repression.
Alternative splicing of transcripts from a single gene is often used as a mechanism for generating protein variants with diverse functions (23). Several genes coding for transcription factors are subject to alternative splicing, and in many cases, it has been demonstrated that alternative splicing is responsible for expression of transcription factors with distinct or even opposing activities (9). Among the *Pax* genes two distinct ways are used to generate diversity. Pax-8 encodes isoforms possessing distinct transactivation potentials but identical DNA-binding properties, in order to differentially regulate Pax-8 target genes (21). In contrast, *Pax-6* bearing two DNA-binding domains encodes isoforms possessing distinct DNA-binding properties but identical transactivation domain. The complex regulation of *Pax-6* generates protein variants which may fulfil distinct functions during retinal development by differentially regulating target gene expression. Characterization of the different Pax-QNR/Pax-6 isoforms will now allow us to design retrovirus vectors changing the expression ratio of the different isoform after infection of the neuroretina *in vivo*.

Acknowledgements :

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

References :

1 - Andrews, N.C. and D. Faller. 1991. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res. 19 : 2499.

2 - Carrière, 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.

3 - Chalepakis, G., F.J. Jones, G. Edelman, and P. Gruss. 1994. Pax-3 contains domains for transcription activation and transcription inhibition. Proc.Natl.Acad.Sci. **91**: 12745-12749.

117

4 - Dignam, J.D., R.M. Lebovitz and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11 : 1475-1489.

5 - Dingwall, C. and R. Laskey. 1991. Nuclear targeting sequences-a consensus ? Trends In Biol. Sci. 16: 478-481.

6 - Dingwall, C. and R. Laskey. 1992. The nuclear membrane. Science 258 : 942-947.

7 - 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 pairedbox- and homeobox-containing gene. Cell Growth and Differentiation 4 : 281-289.

8 - Epstein, J. A., T. Glaser, J. Cai, L. Jepeal, D. Walton and R.L. Maas. 1994. Two independant and interactive DNA-binding subdomains of the Pax-6 paired domain are regulated by alternative splicing. Gene Dev. 8: 2022-2034.

9 - Foulkes, N.S. and P. Sassone-Corsi. 1992. More is better : activators and repressors from the same gene. Cell 68 : 411-414.

10 - Glaser, T., L. Jepeal, J.G. Edwards, S. Young, J. Favor and R.L. Maas. 1994. PAX-6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nature genet. 7: 463-471.

11 - Glaser, T., D. Waldon and R.L. Maas. 1992. Genomic structure, evolutionary conservation and aniridia mutations in the human Pax-6 gene. Nature genet. 2 : 231-239.

12 - Goulding, M.D., G. Chalepakis, U. Deutch, J.R. Erselius and P. Gruss.
1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO
J. 5: 1135-1147.

13 - Gorman, C.M., L.F. Moffat, and B.H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyl-transferase. Mol.Cell.Biol. 2: 1044-1051.

14 - Grueneberg, D.A., S. Natesan, C. Alexandre and M. Gilman. 1992. Human and Drosophila homeodomain proteins that enhance the DNA-binding activity of serum response factor. Science 257 : 1089-1095.

15 - Hanson, I.M., J.M. Fletcher, T. Jordan, A. Brown, D. Taylor, R.J. Adams, H.H. Punnett and V. van Heyningen. 1994. Mutations at the Pax-6 locus are

found in heterogeneous anterior segment malformations including Peters'anomaly. Nature Genetics 6: 168-173.

16 - Hanson, I.M., A. Seawright, K. Hardman, S. Hodgson, D. Zaletayev, G.
Fekete and V. van Heyningen. 1993. PAX-6 mutations in aniridia. Human Molec.Genet. 2
: 915-920.

17 - Hay, N., M. Takimoto, and J.M. Bishop. 1989. A FOS protein is present in a complex that binds a negative regulator of MYC. Genes Dev. 3 : 293-303.

18 - Hill, R.E., J. Favor, B. Hogan, C.C.Ton, G. Saunders, I. Hanson, J. Prosser, T. Jordan, N. Hastie and V. van Heyningen. 1991. Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature **354** : 522-525.

19 - Kern, M., D. Witte, M. Valerius, B. Aronow and S. Potter. 1992. A novel murine homeobox gene gene isolated by a issue specific PCR cloning strategy. Nucleic Acids Res. 20 : 5189-5195.

20 - Kissinger, C., B.Liu, E. Martin-Blanco, T. Kornberg and C.O. Pabo. 1990. Crystal structure of an angrailed homeodomain-DNA complex at 2.8 A resolution : a framework for understanding homeodomain-DNA interactions. Cell **63** : 579-590.

21 - Kozmik, Z., R. Kurzbauer, P. Dörfler and M. Busslinger. 1993. Alternative splicing of Pax-8 gene transcripts is developmentally regulated and generates isoforms with different transactivation properties. Mol.Cell.Biol. 13 : 6024-6035.

22 - Martin, P., C. Carriere, C. Dozier, B. Quatannens, M. Mirabel, B. Vandenbunder, D. Stehelin and S. Saule. 1992. Characterization of a pairedbox- and homeobox-containing quail gene (Pax-QNR) expressed in the neuroretina. Oncogene 7 : 1721-1728.

23 - McKeown, M. 1992. Alternative mRNA splicing. Annu. Rev. Cell Biol. 8 : 133-155.

24 - Plaza, S., C. Dozier and S. Saule. 1993. Quail Pax-6 (Pax-QNR) encodes a transcription factor able to bind and transactivate its own promoter. Cell Growth & Diff. 4: 1041-1050.

25 - Quiring, R., U. Walldorf, U. Kloter and W. Gehering. 1994. Homology of the eyeless gene of *Drosophila* to the small eye gene in mice and Aniridia in Humans. Science 265 : 785-789.

119

26 - Silver, P.A. 1991. How proteins enter the nucleus. Cell 64 : 489-497.

27 - 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.

28 - Treisman, J., E. Harris and C. Desplan. 1991. The paired box encodes a second DNA-binding domain in the Paired Homeodomain containing protein. Genes Dev. **5** : 594-604.

29 - Tremblay, P. and P. Gruss. 1994. Pax: Genes for Mice and Men.Pharmac.Ther. 61: 205-226.

30 - **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. Molecular Endocrinology **8**: 929-938.

31- Walther, C. and P. Gruss. 1991.Pax-6, a murine paired box gene, is expressed in the developing CNS. Development 113: 1435-1449.

32 - 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.

33 - Wilson, D., G. Sheng, T. Lecuit, N. Dostatni and C. Deplan. 1993. Cooperative dimerization of paired class homeo domains on DNA. Genes. Dev. 7: 2120-2134.



Figure 1 : Characterization of Pax-QNR proteins in vivo.

A - Western blot was performed on neuroretina cell extracts with rabbit serum 13. Bound antibodies were detected with an ECL detection kit.
B - A replica from the gel used in panel A was stained with commassie blue in order to evaluate the protein loading.



Figure 1C : Characterization of Pax-QNR proteins in vivo.

Subcellular localization of Pax-QNR isoforms. QNR cells (approximately $3x10^7$ cells) were labelled for 6 hours with 0,2 mCi of L-(³⁵S) methionine. Scrapped cells were separated in 2 equal aliquots. The first one was immunoprecipitated with serum 14 (unfractionated), the second one was further separated in cytoplasmic and nuclear fractions. Each fraction corresponding to the same number of cells was immunoprecipitated with serum 14 and the resulting autoradiogram was subjected to a densitometric scanning. The signal obtained for the different proteins in the unfractionated lanes represent the 100% value.



Figure 2 A: Characterization of Pax-QNR messengers and encoding vectors

Schematic representation of Pax-QNR isoforms encoding mRNA deduced from the genomic *Pax-QNR* structure and splicing events defined from cDNA cloning and PCR experiments. (*) refers to the 6 amino acids extended exon 10 (black box in the exon 10)



Figure 2B : Characterization of *Pax-QNR* messengers and encoding vectors

Immunoprecipitation with serum 14 of $L(^{35}S)$ methionine labelled COS-1 cells transfected with pSG5 vector encoding the different Pax-QNR isoforms



Figure 2C : Characterization of *Pax-QNR* messengers and encoding vectors

RNAase protection experiments. The structure of the (32P) cRNA probes and predicted fragments (open box) derived from the Pax-QNR isoforms paired domains cDNA subcloned in pGEM3Z shown diagrammatically at the top of the figure. Bold line, linker sequence. The 308 nts fragment should detect the p48-encoding mRNAs whereas the 242 nts fragment will detect the p46-encoding mRNAs. The 152 nts fragment should detect the p43-encoding mRNAs. 20µg of total RNA were used for each point. RNA were extracted from embryonic quail neuroretina.



Figure 3A : Subcellular localization of Pax-QNR isoforms Schematic representation of Pax-QNR isoforms and mutants



Figure 3B : Subcellular localization of Pax-QNR isoforms

Mutagenesis the R208 into a W as shown by the nucleotide sequence analysis. Sequence was made by dydeoxynucleotide-chain termination method according to the manufacturer specifications (Pharmacia). Asterisk and arrowhead indicate the mutated nucleotide



Figure 3C : Subcellular localization of Pax-QNR isoforms

Subcellular localization of Pax-QNR isoforms and mutated products was assayed by indirect immunofluorescence on fixed transfected QEF cells. Anti Pax-QNR immunoreactive proteins were detected with FITC-labelled goat anti-rabbit immunoglobulin secondary reagent. Panels a, p48; b, p46; c, p44; d, p43; e, p33/32; f, p33/32 Δ HNLS; g, p43 Δ HNLS; h, p46 Δ HNLS; i, p33/32W²⁰⁸; j, p46W²⁰⁸.



Figure 4A : Determination of Pax-QNR isoforms involved in DNA-protein complex

EMSA with e5 probe labelled with ³²P and transfected in COS-1 cell extracts supplemented or not (-) with antiserum directed against the carboxy-terminal part of the protein. The arrow shows specific DNA-protein complex.



Figure 4B : Determination of Pax-QNR isoforms involved in DNA-protein complex

The DNA-protein complexes formed between the labelled 214-bp Scal-Asp718 Pax-QNR P0 promoter fragment and the Pax-QNR isoforms present in the transfected COS-1 cell extracts were immunoprecipitated with serum 14, the probe was extracted and subjected to electrophoresis.



Figure 4C : Determination of Pax-QNR isoforms involved in DNA-protein complex

EMSA with the 5aCON probe labelled with 32P and transfected COS-1 cell extracts supplemented or not (-) with antiserum directed against the carboxy-terminal part of the protein.



Figure 5 : Transactivation of the Pax-QNR promoter by the Pax-QNR isoforms

The pXA contain the 1,5Kbp Xba1-Asp718 promoter fragment inserted 5' to the CAT gene (described in ref 24) and Pax-QNR expression vectors encoding the different isoforms were transiently transfected into QEC and CAT activities were measured with Phosphor Imager (Molecular Dynamic). CAT activities are expressed relative to that for the p46 set value 100. The results are the averages of three independant experiments performed in duplicate using two different DNA preparations, with the s.e.m. indicated.



Figure 6A : Definition of the Pax-QNR domains Schematic representation of Pax-QNR mutants used.



Figure 6B : Definition of the Pax-QNR domains

To compare the amount of proteins expressed in the transfected COS-1 cells, we performed a western blot experiment on lysates used in EMSA and immunoselection assay, with serum 13.



Figure 6C : Definition of the Pax-QNR domains

Stability of Pax-QNR \triangle BalI (\triangle 333-416) mutant analyzed by pulse and chase labelling. Transfected COS-1 cells were L-(³⁵S) methionine labelled for 45 min. and chased for the indicated period of time in complete medium. Cell lysates were immunoprecipitated with serum 14. Lane i, immune serum; lane B, serum blocked with an excess of bacterial immunogen. Left panel, p46; right panel, Pax-QNR \triangle BalI (\triangle 333-416) mutant.



Figure 6D : Definition of the Pax-QNR domains

Immunoprecipitation of the DNA-protein complexes formed between the ³²P labelled 214-bp Sca1-Asp718 Pax-QNR P0 promoter fragment and the Pax-QNR mutant proteins.



Figure 6E : Definition of the Pax-QNR domains

The pXA vector and mutant Pax-QNR-encoding expression vectors were transiently transfected into QEC and CAT activities were measured with a Phospho Imager. CAT activities are expressed relative to that of p46 set at a value 100 and that of the parental pSG5 plasmid set at value 0. The results are the averages of three independant experiments performed in duplicate using two different DNA preparations, with the s.e.m. indicated. The activation for the p46 was about 4 fold above the pSG5.



Figure 7 : Summary of the results obtained with the different Pax-QNR mutants and isoforms.

Subcellular localization, DNA-binding and transactivation properties of the proteins were indicated. The DNA used was the Pax-QNR P0 promoter.

Pax-QNR ISOFORMS and MUTANTS	CELLULAR LOCALIZATION(%)		
	N	N+C	С
p46	100		
p48	100		
p43	31	69	
p4 4	30	70	
p33/32		100	
p46∆HNLS	99	1	
p43∆HNLS		20	80
p33/32∆HNLS		3	97
p33/32W		100	
p43W	30	70	
p46W	100		

TABLE 1. Nuclear localization of Pax-QNR isoforms and mutants

•

Table 1

Pax-QNR MUTANTS	CELLULAR LOCALIZATION(%)			
	N	N+C	С	
p46	100			
∆21-47	100			
∆48-211		89	11	
∆170-270	98	2		
∆170-342		100		
∆271-329	100			
∆331-379	100			
∆333-416	100			

TABLE 2. Nuclear localization of Pax-QNR mutants

Table 2

DISCUSSION

Le gène *Pax-6* joue un rôle essentiel dans le développement des yeux d'espèces aussi divergentes au point de vue de l'évolution que les insectes et les vertébrés. Le fait que ce gène ait également pu être isolé dans des espèces possédant des systèmes visuels extrêmement primitifs tels que les oursins (Czerny *et al.*, 1994), suggère une fonction de base dans la mise en place des organes photosensibles. Chez les vertébrés, son expression étant observée dès les tous premiers stades de la différenciation de l'oeil et maintenue dans les tissus différenciés, il est probable qu'il soit impliqué dans la différenciation et également dans la maintenance du phénotype différencié. Ce gène aurait donc des fonctions multiples au cours du développement des yeux.

La neurorétine aviaire étant un modèle biologique développé dans le laboratoire, nous avons focalisé notre étude du gène *Pax-6/Pax-QNR* et de ses produits dans ce tissu.

Afin de mieux comprendre la fonction de ce gène, nous avons, dans un premier temps, étudié son expression globale, avec des sondes recouvrant la séquence codante, au cours du développement de l'embryon de caille. Comme chez la souris et le poisson-zèbre, il est exprimé très précocement dans le tube neural, dans le cerveau antérieur et postérieur. Au cours de la mise en place de l'oeil, son expression est détectée à la fois au niveau de la paroi du diencéphale qui va générer la vésicule optique et au niveau de l'ectoderme de surface qui donnera le cristallin. Une fois l'oeil formé, *Pax-QNR* sera exprimé dans la neurorétine au niveau de la couche ganglionnaire et de la couche nucléaire interne, dans le cristallin, la cornée et très faiblement dans la rétine pigmentaire. D'autres gènes à boîte homéo ont des domaines d'expression chevauchant ceux de *Pax-QNR* dans la neurorétine et seraient susceptibles d'interagir avec ce gène.

Plusieurs ADNc différents de *Pax-QNR* ont été isolés et l'étude de la structure génomique de ce gène a montré qu'ils provenaient d'événements d'épissage alternatif multiples. Ces épissages peuvent se produire à la fois dans la partie codante et dans la partie non codante du gène. Dans la région 5' non codante, au moins deux types d'épissage alternatif se produisent. Le premier génère le messager MC29QNR2 qui comporte les exons non codants 0, 1, 2, 3, l'AUG se situant à la fin de l'exon 3 très proche de la boîte PRD. Il code une protéine de 46 KDa qui est le poids attendu pour une protéine contenant toute la séquence codante. Le deuxième ADNc isolé, le clone B1, est codé dans la partie 5' non codante par l'exon α . Cet exon est dépourvu de codon initiateur et

présente un cadre de lecture ouvert en phase avec la boîte PRD de 69 codons. La présence de plusieurs autres fragments protégés par la sonde contenant les exons 0,1,2,3 suggère que d'autres événements d'épissage alternatif se produisent en 5' du gène. En effet, deux autres ADNc isolés proviennent de combinaisons d'exons en 5' encore différentes. Ces événements ne sont pas spécifiques du gène *Pax-QNR*. En effet, les messagers de *Pax-6* chez la souris, l'homme et la drosophile présentent également plusieurs extrémités 5' non codantes (Walther *et al.*, 1991)(Quiring *et al.*, 1994)(Ton *et al.*, 1991). La régulation transcriptionnelle de ce gène apparaît donc comme étant très complexe dans toutes ces espèces.

D'autres événements d'épissage alternatif se produisent également dans la partie codante de *Pax-QNR*. Ainsi, l'ADNc B1 contient un exon 10 alternatif de six aa plus long dans la partie carboxy-terminale de la protéine. Cet épissage alternatif a été caractérisé également chez la souris et l'homme (Walther *et al.*, 1991)(Ton *et al.*, 1991). En fait, le messager MC29QNR2 est le seul ADNc connu dépourvu de cet exon. Etant situé dans le domaine transactivateur de la protéine, cet exon est susceptible de modifier l'activité transactivatrice. Cependant, nous n'avons observé aucune modification de cette activité sur le promoteur P0 de *Pax-QNR*, entre la p46 et la p46* (article 4, Fig.5). Chez la souris, un petit exon supplémentaire de 14 aa a été caractérisé entre les deux premiers exons du domaine PRD, l'exon 4a, qui est présent également dans l'ADN génomique de caille (Walther *et al.*, 1991).

Nous nous sommes ensuite intéressés à l'expression des produits de *Pax-QNR*. A l'aide d'anticorps dirigés contre les différents domaines de la protéine codée par le messager MC29QNR2, nous avons montré que *Pax-QNR* codait *in vivo*, dans la neurorétine de caille, au moins cinq protéines : trois grandes protéines de 48, 46 et 43 KDa et deux petites protéines de 33 et 32 KDa. Les trois grandes protéines sont reconnues par tous les sera. L'ADNc MC29QNR2 code *in vitro* une protéine de 46KDa et l'on peut supposer que la p46 observée *in vivo* est initiée sur le même AUG. Un ARNm contenant l'exon 4a explique la p48. Nous avons pu isoler, par amplification en chaîne sur rétrotranscription d'ARNm de neurorétine de caille, un ADNc partiel contenant cet exon. Cloné en phase avec l'ADNc de MC29QNR2, il code une protéine de 48 KDa. La p43 est codée par un messager contenant un épissage dans le domaine PRD qui va supprimer une partie de l'exon 5, réduisant le domaine PRD à 62 aa, et l'exon 10 contenant les six aa

supplémentaires en position carboxy-terminale de la protéine. Enfin, les petites protéines de 33-32 KDa ne sont pas reconnues par les anticorps dirigés contre le domaine PRD mais par les autres sera. Elles sont donc dépourvues de ce domaine. Il ne s'agit pas de produit de dégradation des grandes protéines puisqu'elles ont la même durée de demi vie que ces dernières. Il peut s'agir d'initiation interne liée ou non à un événement d'épissage alternatif. Le fait que ces p33-32 soient également produites lorsque l'on transfecte l'ADNc MC29QNR2 dans des cellules aviaires ou de mammifères montre que le mécanisme d'initiation interne fonctionne in vivo. Des phénomènes d'initiation interne ont, par ailleurs, déjà été décrits pour les produits de S-CREM (Delmas et al., 1992) et les formes tronquées du récepteur des hormones thyroïdiennes (Bigler et al., 1992). Trois AUG situés en aval de la boîte PRD peuvent être utilisés comme codons initiateurs et la mutagenèse dirigée sur ces AUG suffit à supprimer la production des petites protéines en synthèse in vitro. La différence de taille entre les p33 et 32 peut être due soit à l'utilisation d'un codon initiateur différent soit à la présence de l'exon 10 alternatif. Par PCR avec des oligonucléotides encadrant différentes régions de la protéine sur des rétrotranscription d'ARNm de neurorétine de caille, nous avons recherché d'autres messagers alternatifs. Nous avons montré que dans la région carboxy-terminale, un seul type d'épissage se produisait (l'exon 10 rallongé)(Langlois, 1992). Nous n'avons isolé aucun ADNc généré par épissage alternatif dans la jonction entre la boîte PRD et la boîte homéo, et dans cette dernière région. Par contre, avec des oligonucléotides encadrant la boîte PRD, un ADNc supplémentaire a été isolé comportant l'exon 4a mais épissé dans l'exon 5. Cet ADNc, cloné en phase dans l'ADNc MC29QNR2, code une protéine de 44 KDa qui, si elle est exprimée in vivo, ne peut pas être séparée de la p43 dans nos conditions d'électrophorèse lors des expériences d'immunoprécipitation de protéines de neurorétine. Les événements d'épissage alternatif combinés aux phénomènes d'initiation interne peuvent expliquer le nombre de protéines caractérisées in vivo et codées par Pax-QNR.

Il a été montré pour plusieurs facteurs de transcription que des formes tronquées d'un facteur donné, générées par épissage alternatif ou par initiation interne, étaient capable de moduler sa fonction comme dans le cas du récepteur des hormones thyroïdiennes (Bigler *et al.*, 1992). Nous avons tenté de mettre en évidence un éventuel effet modulateur des p33-32 sur l'activité de la p46. Cependant, que ce soit pour la fixation sur e5 ou au niveau de l'activité transcriptionnnelle sur le promoteur de *Pax-QNR*, nous n'avons observé aucune modification de l'activité de la p46

en présence des protéines dépourvues de domaine PRD. De même, aucun complexe formé par ces différentes protéines n'a pu être mis en évidence. Cela ne permet pas d'éliminer la possibilité que ces protéines puissent interagir ou fonctionner de façon opposée mais suppose l'existence de cibles autres que le promoteur de *Pax-QNR*.

L'analyse de l'expression des ARNm de ces différentes isoformes par protection vis à vis de la RNase montre que les messagers des p48, 46 et 43 sont exprimés dans un rapport constant les uns par rapport aux autres durant le développement de la neurorétine. Par contre, on peut voir, par western blot, que le taux de p33-32 augmente au cours du développement bien que des expériences d'immunoprécipitation montrent que leur néosynthèse diminue. La demi vie élevée de ces protéines peut expliquer ces résultats. Cela peut également n'être lié qu'à un artefact de technique. En effet, les cellules de neurorétine sont cultivées au moins 24 heures avant le marquage à la (³⁵S)Met et l'immunoprécipitation alors que, pour les expériences de western, ces cellules sont lysées immédiatement après dissection. La mise en culture modifie fréquemment le comportement des cellules. Ainsi, des cellules tumorales se mettent à exprimer, lorsqu'elles sont mises en culture, différentes métalloprotéases alors que, *in vivo*, ces dernières ne sont pas détectées (Wernert *et al.*, 1994). Il est donc possible que, dans le cas présent, il y ait une modification de la production des p33-32. Une dernière possibilité est qu'il y ait une saturation du signal sur l'autoradiographie au niveau des p48-46-43 et dans ce cas là, l'augmentation observée du taux de p33-32 ne se ferait alors qu'en parallèle de celle du taux de grandes protéines.

La régulation de la synthèse des différentes protéines codées par Pax-QNR est importante. Le caractère semi-dominant des mutations Sey et AN montre que l'effet de dose est important dans l'activité des produits de Pax-6 (Glaser *et al.*, 1994). En effet, même si il n'y a plus qu'un seul allèle fonctionnant de façon normale, un phénotype est observé. Il peut y avoir plusieurs explications à cet effet de dose. Ainsi, à un certain nombre de molécules se fixant de façon coopérative sur une séquence cible donnée, correspondrait un taux de transactivation, et toute modification de la production des protéines altérerait la réponse. Il pourrait y avoir des formations d'homodimères ou d'hétérodimères avec d'autres protéines (isoformes différentes, autres produits de gènes Pax tels que ceux de Pax-2, autres protéines à homéodomaine ou cofacteurs) où une certaine stoechiométrie d'interaction serait nécessaire. Un cas d'AN décrit est du, suite à une mutation dans le site d'épissage entre l'exon 4 et l'exon 4a, à un changement de rapport entre le messager codant la p46 et le messager codant la p48 (Epstein *et al.*, 1994b). Ce dernier devient majoritaire et l'individu porteur de cette mutation présente des anomalies au niveau de l'iris. La p48 et la p46 ont au moins une séquence cible commune (5aCON) et la réduction du taux de la p46 permettrait à la p48 de se fixer à sa place mais avec un effet sur l'activité transcriptionnelle différent. Dans ce cas d'AN, il est également possible que l'effet observé ne soit pas du à l'augmentation du taux de la p48 mais seulement à la diminution du taux de p46. Cette protéine n'est alors plus en quantité suffisante pour pouvoir activer ses cibles spécifiques et la p48 n'est pas capable de la remplacer.

Nous avons montré que Pax-QNR code au moins un facteur de transcription, la p46. Les travaux de Epstein et al. (1994b) montre que la p48 est également capable d'activer la transcription à partir de séquences différentes de celles reconnues par la p46. De plus, des expériences préliminaires suggèrent que, dans des systèmes cellulaires différents, toutes les isoformes sont capables d'activer la transcription. L'hypothèse de départ étant que toutes les isoformes codées par ce gène étaient de potentiels facteurs de transcription, leurs capacités à fixer l'ADN ont été étudiées. Cette fixation se fait de façon complexe par l'intermédiaire de deux domaines de liaison, le domaine PRD et l'homéodomaine, qui reconnaissent des séquences cibles différentes. De plus, le domaine PRD a une structure bipartite qui peut être révélée, comme dans le cas de Pax-6, par des événements d'épissage alternatif. La première hélice du domaine PRD était considérée jusqu'alors comme étant essentielle pour la reconnaissance et la fixation à l'ADN (Chalepakis et al., 1991). La présence de l'exon 4a qui va séparer cette première hélice des deux suivantes perturbe la structure du domaine PRD. Ce domaine devient capable de reconnaître par l'intermédiaire des deux dernières hélices, une séquence cible (5aCON) autre que celles reconnues par le domaine PRD classique. Cette séquence est constituée en fait de deux demi sites très proches en séquence. Le domaine PRD contenant l'exon 4a peut se fixer en dimère sur 5aCON et en monomère sur un demi site (5aCON1/2) alors que la p46 ne reconnaît que la séquence complète et se fixe sous forme de monomère avec une affinité plus faible (Epstein et al., 1994a). Mais ces études ont été faites sur des domaines PRD purifiés. Dans le cadre de la protéine totale, la p46 reconnaît la séquence cible du domaine PRD de la p48 avec une meilleure affinité que cette

dernière. Cependant, la p48 peut se lier à cette séquence alors qu'elle ne reconnaît ni la séquence e5 ni le fragment de 240 pb du promoteur de *Pax-QNR*. De plus, les travaux de Epstein *et al.* (1994a) montrent que si la p48 et la p46 peuvent transactiver à partir de 5aCON, seule la p48 peut transactiver à partir de la séquence 5aCON1/2.

Aucune séquence cible n'a encore été définie pour la p43. Cependant, les travaux faits sur le domaine PRD de Pax-5 ont permis de caractériser deux consensus de fixation à l'intérieur des séquences reconnues par ce domaine, une pour la partie amino-terminale du domaine PRD contenant la première hélice et l'autre pour la partie carboxy-terminale comprenant les deux hélices suivantes (Czerny *et al.*, 1993). Ils montrent que la partie amino-terminale du domaine PRD est suffisante pour la fixation sur des séquences correspondant au premier consensus. Ainsi des mutants tronqués dans les hélices 2 et 3 restent capables de se lier à l'ADN. L'épissage dans le domaine PRD de la p43 supprime ces deux dernières hélices mais conserve la première. Il est donc envisageable que cette protéine soit capable de se fixer sur le même type de séquence que la partie amino-terminale du domaine PRD de Pax-5 (Cf. Fig.14).

Enfin, plusieurs travaux montrent que les homéodomaines de type PRD ou apparentés sont capables de se fixer en dimère sur une séquence palindromique séparée par deux ou trois pb contenant le consensus ATTA/TAAT (Wilson et al., 1993). Ainsi, les protéines Pax-3 et Pax-7 peuvent se fixer en homodimère et hétérodimère sur ce type de séquence palindromique (Schäfer *et al.*, 1994). Il sera donc intéressant d'étudier la fixation des produits de *Pax-QNR* sur ce type de séquence, en particulier la p43 et les protéines dépourvues de domaine PRD qui ne peuvent se fixer sur aucune des séquences reconnues par la p46. De plus, la formation d'hétérodimère entre produits de la famille Pax étant théoriquement possible, il peut être envisagé que les produits de *Pax-6* soient capables d'interagir avec d'autres protéines à homéodomaine en particulier celles dont l'homéodomaine est apparenté au type PRD (Cf. Fig.15). Un gène récemment isolé, *Chox10*, à homéodomaine de ce type, présente des territoires d'expression communs avec *Pax-6*, en particulier dans la couche nucléaire interne de la neurorétine au niveau des cellules amacrines et bipolaires (Liu *et al.*, 1994). Il serait donc susceptible d'interagir avec les produits de *Pax-6* dans la différenciation de l'oeil sont probablement multiples. Qu'il puisse agir de façon combinée avec d'autres facteurs semble logique. Ainsi, il a été montré sur plusieurs



Fig.14 : Fixation du domaine PRD

promoteurs de cristalline que son produit n'active la transcription que quand d'autres facteurs comme CREB sont présents (Cvekl *et al.*, 1994)(Cvekl *et al.*, 1995). Sa spécificité d'action serait donc médiée à la fois par les séquences cibles et par des interactions et/ou la présence d'autres facteurs. De plus, certains produits de *Pax-6* pourraient fonctionner en tant que répresseurs de la transcription. Les travaux de Plaza *et al.* (1994) montrent que la p46 codée par *Pax-QNR* peut se lier à des séquences reconnues par les protéines de la famille ETS (MacLeod *et al.*, 1992) et cette liaison est médiée par le domaine PRD. Il y a compétition au niveau de la fixation sur ces séquences entre le produit de *Pax-QNR* et différents produits de la famille ETS tels que la p68^{c-ets-1} et la p55^{erg}. De plus, si la p46 est incapable d'activer la transcription à partir de certaines de ces séquences, elle est, par contre, capable d'empêcher la transactivation par des protéines ETS à partir de ces mêmes séquences.

Afin de mieux comprendre le mode de fonctionnement des produits de Pax-QNR, nous avons défini leurs domaines fonctionnels. Ces protéines ont une localisation subcellulaire différente suite à la présence ou non des deux séquences de localisation nucléaire. Les p48 et p46 sont nucléaires, les p43 et p33-32 sont nucléaires et cytoplasmiques. L'analyse des mutants tronqués dans différentes régions de la protéine, présentés dans l'article 4, suggère qu'une troisième séquence de localisation nucléaire est présente en aval de l'homéodomaine bien qu'elle ne soit pas suffisante pour transporter la protéine dans le noyau quand les deux autres séquences sont supprimées. De plus, nous avons réalisé des protéines de fusion entre une protéine cytoplasmique, la ßGalactosidase, et les deux séquences de nucléarisation que nous avons caractérisées. Aucune de ces séquences, que ce soit séparément ou ensemble, n'est capable de médier le transport dans le noyau de la ßGalactosidase. La troisième séquence de localisation nucléaire est peut être nécessaire à ce transport. La translocation dans le noyau constitue une étape de contrôle essentielle pour l'activité d'un facteur de transcription (voir la revue de Silver, 1991). Le fait que une de ces séguences de nucléarisation, HNLS, soit à proximité d'une séquence consensus de phosphorylation par la caséine kinase II (CKII), laisse supposer que la phosphorylation par cette kinase pourrait contrôler la translocation dans le noyau. In vivo, les protéines codées par Pax-QNR sont phosphorylées sur des Ser et Thr. En plus du consensus de phosphorylation par la CKII, deux séquences consensus de phosphorylation par les MAP kinases



HETERODIMERES ENTRE LES ISOFORMES ET D'AUTRES PROTEINES A HOMEODOMAINE DE TYPE PRD OU APPARENTE



Fig.15: Hypothèses de fixation par l'intermédiaire de l'homéodomaine

(Mitosis Activated Protein kinase) sont présentes dans le domaine carboxy-terminal. Ce dernier constitue le domaine transactivateur de Pax-QNR et des événements de phosphorylation pourraient moduler son activité transactivatrice. Ce domaine est riche en Ser, Thr et Pro et des domaines transactivateurs de ce type ont déjà été caractérisés (Ingraham et al., 1988). Deux régions du domaine carboxy-terminal sont nécessaires à l'activation de la transcription. La région en aval de l'aa 346 est essentielle car quand elle est supprimée, il n'y a plus qu'une activité transactivatrice résiduelle comme l'ont montré Glaser et al. (1994). Nous avons généré plusieurs protéines tronquées dans le domaine carboxy-terminal qui montrent que cette région se situe en fait en aval de l'aa 380. Une deuxième région en amino-terminal de ce domaine est également nécessaire. Elle est révélée par le mutant Δ170-342 qui, bien que capable de se lier à l'ADN, n'active plus la transcription. Cette activité est restaurée quand les aa de 271 à 329 sont réinsérés. Une coopération entre ces deux régions du domaine carboxy-terminal serait donc nécessaire pour une activation optimale de la transcription. Le mutant Δ Ball (Δ 333-416) bien que ne se liant pas à l'ADN, réprime faiblement la transcription à partir du promoteur de Pax-ONR. Il a déjà été montré que le produit de Pax-3 était également capable de réprimer la transcription à partir du promoteur du gène de la N-CAM L1 et le domaine impliqué dans la répression se situe en amino-terminal de la protéine recouvrant en partie le domaine PRD (Chalepakis et al., 1994c). Il est possible que l'amputation de la fin du domaine carboxy-terminal dans le mutant $\Delta Ball$ dévoile un potentiel domaine répresseur. Et comme dans le cas du produit de Pax-3, la répression se fait de façon indirecte sans liaison à l'ADN. Il a été montré pour de nombreuses protéines à homéodomaine que l'activité biologique n'est pas uniquement médiée par la fixation à l'ADN (Fitzpatrick et al., 1992)(Zappavigna et al., 1994). Une partie de la fonction passe par les interactions protéineprotéine et l'effet répresseur du mutant ABall peut être du à ce type d'interaction avec la machinerie transcriptionnelle ou d'autres cofacteurs. L'amputation d'une partie du domaine carboxy-terminal modifierait donc la conformation de la protéine qui pourrait alors s'associer avec des facteurs différents de ceux avec lesquels interagit la protéine normale. Les p33-32 ne sont pas capables de se lier à l'ADN, alors que l'homéodomaine seul le peut, ce qui suggère que le domaine carboxy-terminal pourrait bloquer la liaison à l'ADN. Cela a également été observé pour la protéine paired de drosophile (Treisman et al., 1991). Plusieurs événements conformationnels semblent moduler l'activité de la protéine, en particulier au niveau de la région carboxy-terminale.

Le gène *Pax-6* est très conservé au cours de l'évolution tant au niveau de la séquence que pour le phénotype observé quand il est muté. Ainsi, même si le développement des yeux des insectes est très éloigné de celui des vertébrés, en plus de l'intérêt purement évolutif que constitue la conservation de la fonction de *Pax-6*, les études chez la drosophile peuvent apporter des informations quant à sa fonction chez les vertébrés. La drosophile constitue un modèle biologique extrêmement riche du fait de l'existence d'un nombre élevé de mutants et l'étude de mutants possédant des altérations dans le développement des yeux peut permettre l'identification de gènes cibles de *eyeless* (Quiring *et al.*, 1994). On peut supposer que certaines de ces cibles sont conservées chez les vertébrés.

De même, l'étude de l'expression de ce gène et de ces produits chez des vertébrés ayant un développement des yeux particuliers tels que les poissons cavernicoles ou certains amphibiens pourrait constituer une approche originale de cette fonction. Ces animaux développent des yeux de façon normale au cours de l'embryogenèse. Par la suite, ces yeux vont se kératiniser et s'atrophier. Dans certains cas d'aniridie, des patients développent précocement une kératinisation de la cornée (Glaser *et al.*, 1994). L'étude du gène *Pax-6* chez ces animaux pourrait apporter des informations: y-a-t-il une expression différente et donc une régulation différente de ce gène ? Ce gène n'est peut être exprimé que de façon transitoire lors de la mise en place du système visuel de ces organismes. Est-il porteur de mutations particulières à ces espèces qui vont modifier sa fonction spécifiquement au niveau des yeux ? Ces mutations pourraient supprimer l'existence de certaines isoformes, modifier les propriétés de fixation à l'ADN, l'activité transactivatrice ou encore altérer des interactions.
REFERENCES BIBLIOGRAPHIQUES

Abate, C. and T. Curran. (1990). "Encounters with Fos and Jun on the road to AP-1." Sem. Cancer Biol., 1, 19-26.

Adams, B., P. Dörfler, Z. Kozmik, P. Urbànek, 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.

Ananthan, J., R. Baler, D. Morrissey, J. Zuo, Y. Lan, M. Weir and R. Voellmy. (1993). "Synergistic activation of transcription is mediated by the N-terminal domain of *Drosophila* fushi-tarazu homeoprotein and can occur without DNA binding by the protein." *Moll.Cell.Biol.*, 13, 1599-1609.

Andrès, V., B. Nadal-Ginard and V. Mahdavi. (1992). "Clox, a mammalian homeobox gene related to the drosophila cut, encodes DNA-binding regulatory proteins differentially expressed during development." *Dev.*, **116**, 321-334.

Andrès, V., M. D. Chiara and V. Mahdavi. (1994). "A new bipartite DNA-binding domain: cooperative interaction between the cut repeat and homeodomain of the cut homeo proteins." *Genes & Dev.*, 8, 245-257.

Angrand, P. O. (1993). "Les domaines de liaison à l'ADN des facteurs de transcription eucaryotes." m/s, 9, 725-736.

Arber, S., G. Halder and P. Caroni. (1994). "Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differenciation." *Cell*, **78**, 221-231.

Asano, M. and P. Gruss. (1992). "Pax-5 is expressed in the midbrain-hindbrain boundary during mouse development." Mech. Dev., 33, 27-38.

Baldwin, C. T., C. F. Hoth, J. A. Amos, E. O. da-Silva and A. Milunski. (1992). "An exonic mutation in the *HuP2* paired domain gene causes Waardenburg's syndrome." *Nature*, 355, 637-638.

Balling, R., U. Deutsch and P. Gruss. (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.

Balling, R., C. F. Lau, S. Dietrich, J. Wallin and P. Gruss. (1992). "Development of the skeletal system" *Ciba Found. Symp.*, 165, 132-143.

Balling, R., C. Ebensperger, N. I. Hoffman, K. Imai, H. Koseki and Y. W. Mizutani J. (1993). "The Genetics of Skeletal Development" Ann. Genet., 36, 56-62.

Barberis, A., G. Superti-Furga, L. K. Vitelli I. and M. Busslinger. (1989). "Developmental and tissue-specific regulation of a novel transcription factor of the sea urchin." *Genes & Dev.*, **3**, 663-675.

Barberis, A., K. Widenhorn, L. Vitelli and M. Busslinger. (1990). "A novel B-cell lineagespecific transcription factor present at early but not late stages of differentiation." *Genes & Dev.*, 4, 849-859.

Barr, F. G., N. Galili, J. Holick, J. A. Biegel, G. Rovera and B. S. Emanuel. (1993). "Rearrangement of the PAX3 paired box gene in the paediatric solid tumor alveolar rhabdomyosarcoma." *Nature Genet.*, 3, 113-117.

Baumgartner, S., D. Bopp, M. Burri and M. Noll. (1987). "Structure of two genes at the *gooseberry* locus related to the *paired* gene and their spatial expression during *Drosophila* embryogenesis." *Genes & Dev.*, **1**, 1247-1267.

Beebe, D. C. (1994). "Homeobox Genes and Vertebrate Eye Development." Invest. Ophthal. Vis. Sci., 35, 2897-2900.

Berleth, T., M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll and C. Nüsslein-Volhard. (1988). "The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo." *EMBO J.*, 7, 1749-1756.

Bigler, J., W. Hokanson and R. N. Eisenman. (1992). "Thyroid hormone receptor transcriptionnal activity is potentially autoregulated by truncated forms of the receptor." *Moll.Cell.Biol.*, 12, 2406-2417.

Birkenmeier, E. H. and J. J. Gordon. (1986). "Developmental regulation of a gene that encodes a cysteine-rich intestinal protein and maps near the murine immunoglobuline heavy chain locus." *Proc. Natl. Acad. Sci. USA*, 83, 2516-2520.

Blochlinger, K. R., R. Bodmer, J. Jack, L. Y. Jan and Y. N. Yan. (1988). "Primary structure and expression of a product from *cut*, a locus involved in specifying sensory organ identity in drosophila." *Nature*, 333, 629-635.

Bober, E., T. Franz, H. H. Arnold, P. Gruss and P. Tremblay. (1994). "Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells." *Dev.*, **120**, 603-612.

Bodner, M., J. L. Castrillo, L. E. Theill, T. Deerinck, M. Ellisman and M. Karin. (1988). "The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein." *Cell*, 55, 505-518.

Bopp, D., M. Burri, S. Baumgartner, G. Frigerio and M. Noll. (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., E. Jamet, S. Baumgartner, M. Burri and M. Noll. (1989a). "Isolation of two tissuespecific *Drosophila* paired box genes, *Pox meso* and *Pox neuro*." *EMBO J.*, 8, 3447-3457.

Bopp, D., M. Burri, S. Baumgartner, G. Frigerio and M. Noll. (1989b). "Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*." *Cell*, 47,

Breier, G., M. Bucan, U. Francke, A. M. Colberg-Poley and P. Gruss. (1986). "Sequential expression of murine homeobox genes during F9 EC cell differentiation." *EMBO J.*, 5, 2209-2215.

Burri, M., Y. Tromvoukis, D. Bopp, G. Frigerio and M. Noll. (1989). "Conservation of the paired domain in metazoans and its structure in three isolated human genes." *EMBO J.*, 8, 1183-1190.

Bürglin, T. R. and G. Ruvkun. (1992). "A new motif in PBX genes." Nature Genet., 341, 239-243.

Carrière, C., S. Plaza, P. Martin, B. Quatannens, M. Bailly, D. Stéhelin 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., M. Goulding, A. Read, T. Strachan and P. Gruss. (1994a). "Molecular basis of Splotch and Waardenburg *Pax-3* mutations." *Proc.Natl.Acad.Sci. USA*, **91**, 3685-3689.

Chalepakis, G., J. Wijnholds, P. Giese, M. Schachner and P. Gruss. (1994b).

"Characterization of Pax-6 and Hoxa-1 binding to the promoter region of the neural cell adhesion molecule L1." DNA & Cell Biol., 13, 891-900.

Chalepakis, G., F. S. Jones, G. M. Edelman and P. Gruss. (1994c). "Pax-3 contains domains for transcription activation and transcription inhibition." *Proc.Natl.Acad.Sci. USA*, **91**, 12745-12749.

Chan, S. K. and R. S. Mann. (1993). "The segment identity functions of Ultrabithorax are contained within its homeodomain and carboxy-terminal-sequences" *Genes & Dev.*, 7, 796-811.

Chida, K. and P. K. Vogt. (1992). "Nuclear translocation of viral Jun but not cellular Jun is cell cycle dependent." *Proc. Natl. Acad. Sci. USA*, **89**, 4290-4294.

Cho, K. W. Y. and E. M. De Robertis. (1990). "Differential activation of *Xenopus* homeobox genes by mesoderm-inducing growth factors and retinoic acid." *Genes & Dev.*, 4, 1910-1916.

Chouard, T., M. Blumenfeld, I. Bach, J. Vandekerckhove, S. Cereghini and M. Yaniv. (1990). "A distal dimerization domain is essential for DNA-binding by the atypical HNF1 homeodomain." *Nucleic Acids Res.*, 18, 5853-5863.

Civitareale, D., R. Lonigro, A. J. Sinclair and R. Di Lauro. (1989). "A thyroid-specific nuclear protein essential for the tissue specific expression of the thyroglobuline promoter." *EMBO J.*, 8, 2537-2542.

Clerc, R. G., L. M. Corcoran, J. H. LeBowitz, D. Baltimore and P. A. Sharp. (1988). "The B-cell-specific Oct-2 protein contains POU box-and homeo box-type domains." *Genes & Dev.*, 2, 1570-1581.

Cohen, S. M., G. Brönner, F. Küttner, G. Jürgens and H. Jäckle. (1989). "Distal-less encodes a homeodomain required for limb development in Drosophila." Nature, 338, 432-434.

Coldberg-Poley, A. M., S. D. Voss, K. Chowdhury and P. Gruss. (1985). "Clustered homeo boxes are differentially expressed during murine development." *Cell*, 43, 39-45.

Corcoran, L. M., M. Karvelas, G. J. Nossal, Z. S. Ye, T. Jacks and D. Baltimore. (1993). "Oct-2, although not required for early B-cell development." *Genes & Dev.*, 7, 570-582.

Crisanti-Combes, P., A. Privat, B. Pessac and G. Calothy. (1977). "Differenciation of chick embryo neuroretina cells in monolayer cultures. An ultrastructural study. I. Seven day retina." *Cell Tissue Res.*, 185, 159-173.

Cvekl, A., C. M. Sax, E. H. Bresnick and J. Piatigorsky. (1994). "A Complex Array of Positive and Negative Elements Regulates the Chicken αA-Crystallin Gene: Involvement of Pax-6, USF, CREB and/or CREM, and AP-1 Proteins." *Mol. Cell. Biol.*, **14**, 7363-7376.

Cvekl, A., F. Kashanchi, C. M. Sax, J. N. Brady and J. Piatigorsky. (1995). "Transcriptionnal regulation of the mouse α A-crystallin gene : activation dependent on a cyclic AMP-responsive element (DE1/CRE) and a Pax-6 binding site." *Mol. Cell. Biol.*, **15**, 653-660.

Czerny, T., G. Schaffner and M. Busslinger. (1993). "DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site." *Genes & Dev.*, 7, 2048-2061.

Czerny, T., M. Busslinger. (1994). "Three amino acids determine the DNA binding specificity of Pax-6." *The First EMBL Meeting on Transcription*. Heidelberg.

Damante, G., D. Fabbro, L. Pellizzari, D. Civitareale, S. Guazzi, M. Polycarpou-Schwartz, S. Cauci, F. Quadrifoglio, S. Formisano and R. Di Lauro. (1994). "Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain." *Nucleic Acids Res.*, 22, 3075-3083.

Davis, C. A. and A. L. Joyner. (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.

Davis, C. A., D. P. Holmyard, K. J. Millen and A. L. Joyner. (1991). "Examining pattern formation in mouse, chicken and frog embryos with an *En-2* specific anti serum." *Dev.*, 111, 287-298.

Davis, R. D., C. M. D'Cruz, M. A. Lovell, J. A. Biegel and F. G. Barr. (1994). "Fusion of *PAX7* to *FKHR* by the Variant t(1;13)(p36;q14) Translocation in Alveolar Rhabdomyosarcoma." *Canc. Res.*, **54**, 2869-2872.

Davis, A. and J. K. Cowell. (1994a). "Mutations in the PAX6 gene in patients with hereditary aniridia." *Hum. Mol. Genet.*, 2, 2093-2097.

Day, R. N., S. Koike, M. Sakai, M. Muramatsu and R. A. Maurer. (1990). "Both Pit-1 and the estrogen receptor are required for estrogen responsiveness of the rat prolactine gene." *Mol. Endocrinol.*, 4, 1964-1971.

Delmas, V., B. Laoide, D. Masquilier, R. de Groot, N. Foulkes and P. Sassone-Corsi. (1992). "Alternative usage of initiation codons in mRNA encoding the cAMP-responsive-element modulator generates regulators with opposite functions." *Proc. Natl. Acad. Sci. USA*, **89**, 4226-4230.

Deutsch, U., G. R. Dressler and P. Gruss. (1988). "*Pax 1*, a member of a paired box homologous murine gene family, is expressed in segmented structures during development." *Cell*, 53, 617-625.

Dingwall, C. and A. R. Laskey. (1991). "Nuclear targeting sequences - a consensus ?" TIBS, 16, 478-481.

Dollé, P., M. Price and D. Duboule. (1992). "Expression of the mouse *Dlx-1* homeobox gene during facial, ocular and limb development." *Differentiation*, **49**, 93-100.

Dozier, C., C. Carrière, D. Grévin, P. Martin, B. Quatannens, D. Stéhelin and S. Saule. (1993). "Structure and DNA-binding properties of *Pax-QNR*, a pairedbox- and homeobox-containing gene." *Cell Growth Differ.*, **4**, 281-289.

Dressler, G. R., U. Deutsch, K. Chowdhury, H. O. Nornes and P. Gruss. (1990). "*Pax-2*, a new murine paired-box-containing gene and its expression in the developing excretory system." *Dev.*, **109**, 787-795.

Dressler, G. R. and E. C. Douglass. (1992). "Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor." *Proc.Natl.Acad.Sci. USA*, **89**, 1179-1183.

Dressler, F. R., J. E. Wilkilson, U. W. Rothenspieler, L. T. Patterson, L. Williams-Simons and H. Westphal. (1993). "Deregulation of *Pax-2* expression in transgenic mice generates severe kidney abnormalities." *Nature*, 362, 65-67.

Duboule, D. and P. Dollé. (1989). "The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes." *EMBO J.*, **8**, 1497-1505.

Dufort, D. and A. Nepveu. (1994). "The human Cut homeodomain protein represses expression from the c-myc promoter." *Mol. Cell. Biol.*, **14**, 4251-4257.

Duprey, P., K. Chowdhury, G. T. Dresler, R. Balling, D. Simon, J. Guenet and P. Gruss. (1988). "A mouse gene homologous to the *drosophila* gene *caudal* is expressed in epithelial cells from the embryonic intestine." *Genes & Dev.*, **2**, 1647-1654.

Epstein, D. J., D. Malo, M. Vekemans and P. Gros. (1991). "Molecular characterization of a deletion encompassing the *Sploch* mutation on mouse chromosome 1." *Genomics*, **10**, 89-93.

Epstein, D. J., M. Vekemans and P. Gros. (1991). "Splotch (Sp^{2H}) , a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax-3*." Cell, **67**, 767-774.

Epstein, D. J., K. J. Vogan, D. G. Trasler and P. Gros. (1993). "A mutation within intron 3 of the *Pax-3* gene produces aberrantly spliced mRNA transcripts in the *splotch* (*Sp*) mutant." *Proc. Natl. Acad. Sci. USA*, **90**, 532-536.

Epstein, J. A., T. Glaser, J. Cai, L. W. Jepeal D.S. and R. L. Maas. (1994a). "Two independent and interactive DNA-binding subdomains of the Pax-6 paired domain are regulated by alternative splicing." *Genes & Dev.*, 8, 2022-2034.

Epstein, J., J. Cai, T. Glaser, L. Jepnal and R. Maas. (1994b). "Identification of a *Pax* Paired Domain Recognition Sequence and Evidence for DNA-dependent Conformational Changes." *J. Biol. Chem.*, **269**, 8355-8361.

Ericson, J., S. Thor, T. Edlund, T. M. Jessell and T. Yamada. (1992). "Early Stages of Motor Neuron Differentiation Revealed by Expression of Homeobox Gene *Islet-1*." *Science*, 256, 1555-1550.

Fabbro, D., C. Di Loreto, C. A. Beltrami, A. Belfiore, R. Di Lauro and G. Damante. (1994). "Expression of Thyroid-specific Transcription Factors TTF-1 and PAX-8 in Human Thyroid Neoplasms" *Canc. Res.*, 54, 4744-4749.

Fickenscher, H. R., G. Chalepakis and P. Gruss. (1993). "Murine Pax-2 protein is a sequence-specific transactivator with expression in the genital system." DNA & Cell. Biol., 12, 381-391.

Finney, M., G. Ruvkun and H. R. Horvitz. (1988). "The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein with a homeodomain and extended similarity to transcription factors." *Cell*, 55, 757-769.

Finney, M. and G. Ruvkun. (1990). "The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*." *Cell*, 63, 895-905.

Fitzpatrick, V. D., A. Percival-Smith, C. J. Ingles and H. M. Krause. (1992). "Homeodomain-independent activity of the fushi tarazu polypeptide in *Drosophila* embryos." *Nature*, 356, 610-612.

Francis-Lang, H., M. Price, M. Polycarpou-Schwarz and R. Di Lauro. (1992). "Cell-type-specific expression of the rat thyroperoxydase promoter indicates common mecanisms for the thyroid-specific gene expression." *Mol.Cell.Biol.*, **12**, 576-588.

Freyd, G., S. K. Kim and H. R. Horvitz. (1990). "Novel cysteine-rich motif and homeodomain in the product of the *C.elegans* lineage gene *lin-11*." *Nature*, 344, 876-879.

Frigerio, G., M. Burry, D. Bopp, S. Baumgartner and M. Noll. (1986). "Structure of the segmentation gene *paired* and the *Drosophila P R D* gene set as part of a gene network." *Cell*, 47, 735-746.

Fujiwara, M., T. Uchida, N. Osumi-Yamashita and K. Eto. (1994). "Uchida rat (*rSey*): a new mutant rat with craniofacial abnormalities resembling those of the mouse *Sey* mutant." *Differentiation*, **57**, 31-38.

Gallili, N., R. J. Davis, W. J. Fredericks, S. Mukhopadhyay, F. J. Rauscher, B. S. Emanuel, J. Rovera and F. G. Barr. (1993). "Fusion of a fork head domain gene to PAX3 in the solid tumor alveolar rhabdomyosarcoma." *Nature Genet.*, 5, 230-235.

Garber, J., A. Kuroiwa and W. Gehring. (1983). "Genomic end cDNA clones of the homeotic locus Antennapedia in Drosophila." EMBO J., 2, 2027-2036.

Garrido, C., F. Grasser, J. Lipsick, D. Stehelin and S. Saule. (1992). "Protein truncation is not required for *c-myb* proto-oncogene activity in neuroretina cells." *J. Virol.*, **66**, 6773-6776.

Garrido, C., D. Leprince, J. S. Lipsick, D. Stehelin, D. Gospodarowwicz and S. Saule. (1992). "Definition of functional domains in p135 gag-myb-ets and p48v-myb proteins required to maintain the respons of neuroretina cells to basic fibroblast growth factor." J.Virol., 66, 160-166.

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.

Gehring, W. J., M. Affolter and T. Bürglin. (1994). "Homeodomain proteins." Ann. Rev. Biochem., 63, 487-526.

German, M. S., J. Wang, R. B. Chadwick and W. J. Rutter. (1992). "Synergistic activation of the insulin gene by a LIM-homeodomain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex." *Genes & Dev.*, 6, 2165-2176.

Gerrero, M. R., R. J. McEvilly, E. Turner, C. R. Lin, S. O'Connell, K. J. Jenne, M. V. Hobbs and M. G. Rosenfeld. (1993). "Brn-3.0: a POU-domain protein expressed in the sensory, immune and endocrine systems that functions on elements distinct from known octamer motifs." *Proc. Natl. Acad. Sci. USA*, **90**, 10841-10845.

Glaser, T., D. S. Walton and R. L. Maas. (1992). "Genomic structure, evolutionary conservation and aniridia mutations in the human *PAX6* gene." *Nature genet.*, 2, 232-239.

Glaser, T., L. Jepeal, J. G. Edwards, S. R. Young, J. Favor and R. L. Maas. (1994). "PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects." *Nature Genet.*, 7, 463-471.

Goding, C. and P. O'Hare. (1989). "Herpes simplex virus Vmw65-octamer binding interaction: a paradigm for combinatorial control of transcription." *Virology*, **173**, 363-367.

Goulding, P. D., G. Chalepkis, U. Deutsch, J. R. Erselius and P. Gruss. (1991). "Pax-3, a novel murine DNA binding protein expressed during early neurogenesis." *EMBO J.*, 10, 1135-1147.

Goulding, M., S. Sterrer, J. Fleming, R. Balling, J. Nadeau, K. J. Moore, S. D. M. Brown, K. P. Steel and P. Gruss. (1993a). "Analysis of the *Pax-3* gene in mouse mutant *splotch*." *Genomics*, 17, 355-363.

Goulding, M. D., A. Lumsden and P. Gruss. (1993b). "Signals from the notochord and floor plate regulate the region-specific expression of two Pax genes in the developing spinal cord." *Dev.*, **117**, 1001-1016.

Goulding, M., A. Lumsden and A. J. Paquette. (1994). "Regulation of *Pax-3* expression in the dermomyotome and its role in muscle development." *Dev.*, **120**, 957-971.

Grüneberg, H. (1954). "Genetical studies on the skeleton of the mouse. XV. Tail kinks." J. Genet., 53, 536-560.

Guazzi, S., M. Price, M. De Felice, G. Damante, M.-G. Mattei and R. Di Lauro. (1990). "Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity." *EMBO J.*, 9, 3631-3639.

Gurdon, J. B. (1992). "The generation of diversity and pattern in animal development." Cell, 68, 185-199.

Hanson, I. M., A. Seawright, H. K., S. Hidgson, D. Zaletayev, G. Fekete and V. van Heyningen. (1993). "PAX6 mutations in aniridia." Hum. Molec. Genet., 2, 915-920.

Hanson, I. M., J. M. Fletcher, T. Jordan, A. T. Brown D., R. J. Adams, H. H. Punnett and V. van Heyningen. (1994). "Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peter's anomaly." *Nature Genet.*, 6, 168-173.

Harada, R., D. Dufort, C. Denis-Larose and A. Nepveu. (1993). "Conserved Cut repeats in the human Cut homeodomain protein function as DNA-binding domains." J. Biol. Chem., 269, 2062-2067.

Harada, R., G. Bérubé, O. J. Tamplin, C. Denis-Larose and A. Nepveu. (1995). "DNAbinding specificity of the cut repeats from the human cut-like protein." *Mol. Cell. Biol.*, 15, 129-140.

Hatano, M., C. M. W. Roberts, C. Minden, W. M. Crist and S. J. Kosmeyer. (1991). "Deregulation of homeobox gene, HOX11, by the t(10;14) in T-cell leukemia." *Science*, 253, 79-81.

Hayashi, S. and M. P. Scott. (1990). "What determines the specificity of action of drosophila homeodomain proteins?" *Cell*, 883-894.

Heasman, J., J. Quarmby and C. C. Wylie. (1984). "The mitochondrial cloud of Xenopus oocytes/ the source of gerrminal granule material." *Dev. biol.*, 105, 458-469.

Hill, R. E., J. Favor, B. L. M. Hogan, C. C. T. Ton, G. F. Saunders, I. M. Hanson, J. Prosser, T. Jordan, N. D. Hastie and V. van Heyningen. (1991). "Mouse *Small eye* results from mutations in a paired-like homeobox-containing gene." *Nature*, 354, 522-525.

Hoey, T. and M. Levine. (1988). "Divergent homeobox proteins recognize similar DNA sequences in *drosophila*." *Nature*, 322, 858-861.

Holland, P. W. H. (1991). "Cloning and evolutionary analysis of *msh*-like homeobox genes from mouse, zebrafish and ascidian." *Gene*, 98, 253-257.

Holst, B. D., R. S. Goomer, I. C. Wood, G. M. Edelmn and F. S. Jones. (1994). "Binding the Activation of the Promoter for the Neural Cell Adhesion Molecule by Pax-8." J. Biol. Chem., 269, 22245-22252.

Hoth, C. F., A. Milunsky, N. Lipsky, R. Sheffer, C. S.K. and C. T. Baldwin. (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)" *Amer. J. Hum. Genet.*, **52**, 455-462.

Hunter, T. and M. Karin. (1992). "The regulation of transcription by phosphorylation" Cell, 70, 375-387.

Ingraham, H. A., R. Chen, H. Mangaiam, S. Elsholtz, C. R. Flynn, D. M. Lin, L. Simmons, L. Swanson and M. G. Rosenfeld. (1988). "A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype." *Cell*, 55, 519-529.

Jaynes, J. B. and P. H. O'Farrel. (1991). "Active repression of transcription by the Engrailed homeodomain protein" *EMBO J.*, 10, 1427-1433.

Jordan, T., I. Hanson, D. Zalatayev, S. Hodgson, J. Prosser, A. Seawright, N. Hastie and V. Van Heyningen. (1992). "The human *PAX-6* gene is mutated in two patients with aniridia." *Nature Genet.*, 1, 328-332.

Jostes, B., C. Walther and P. Gruss. (1991). "The murine paired box gene, *Pax7*, is expressed specifically during the development of the nervous and muscular system." *Mech. Dev.*, 33, 27-38.

Joyner, A. L. and G. R. Martin. (1987). "En-1 and En-2, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis." Genes & Dev., 1, 29-38.

Joyner, A. L., K. Herrup, B. A. Auerbach, C. A. Davis and J. Rossant. (1991). "Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the *En-2* homeobox." *Science*, 251, 1239-1243.

Joyner, A. L. and M. Hanks. (1991). "The engrailed genes: evolution of function." Sem. Dev. Biol., 2, 435-445.

Kamps, M. P., C. Murre, H.-X. Sun and D. Baltimore. (1990). "A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in Pre-B ALL." *Cell*, 60, 547-555.

Kappen, C. and F. H. Ruddle. (1993). "Evolution of a regulatory gene family: HOM/HOX genes." Curr. Opin. Genet. Dev., 3, 931-938.

Keller, S. A., J. M. Jones, A. Boyle, L. L. Barrow, P. D. Killen, D. G. Green, N. V. Kapousta, P. V. Hitchcock and R. T. Swank. (1994). "Kidney and retinal defects (Krd), a transgene-induced mutation with a deletion of mouse chromosome 19 that includes the Pax-2 locus." *Genomics*, 23, 309-320.

Kennison, J. A. (1993). "Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements." *TIG*, 9, 75-78.

Kessel, M., R. Balling and P. Gruss. (1990). "Variations of cervical vertebrae after expression of a *Hox-1.1* transgene in mice." *Cell*, **61**, 301-308.

Kessel, M. and P. Gruss. (1991). "Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid." *Cell*, 67, 89-104.

Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg and C. O. Pabo. (1990). "Crystal Structure of an engrailed Homeodomain-DNA Complex at 2.8 A Resolution: A Framework for Understanding Homeodomain-DNA Interactions." *Cell*, **63**, 579-590.

Koseki, H., J. Wallin, J. Wilting, Y. Mizutani, A. Kispert, C. Ebensperger, B. G. C. Herrmann B. and R. Balling. (1993). " A role for Pax-1 as a mediator of notochord signals during the dorsoventral specification of vertebrae." *Dev.*, **119**, 649-660.

Kozmik, Z., S. Wang, P. Dörfler, B. Adams and M. Busslinger. (1992). "The promoter of the *CD19* gene is a target for the B-cell-specific transcription factor BSAP." *Molec. Cell. Biol.*, 12, 2662-2672.

Kozmik, Z., R. Kurzbauer, P. Dörfler and M. Busslinger. (1993). "Alternative Splicing of *Pax-8* Gene Transcripts Is Developmentally Regulated and Generated Isoforms with Different Transactivation Properties." *Mol. Cell. Biol.*, **13**, 6024-6035.

Krauss, S., T. Johansen, V. Korzh, Y. Moens, J. A. Ericson and A. Fjose. (1991a). "Zebrafish *Pax* [zf-a]: a paired box-containing gene expressed in the neural tube." *EMBO J.*, 10, 3609-3619.

Krauss, S., T. Johansen, V. Korzh and A. Fjose. (1991b). "Expression of the zebrafish paired box gene *pax[zf-b]* during early neurogenesis." *Dev.*, **113**, 1193-1206.

Krauss, S., M. Maden, N. Holder and S. W. WIlson. (1992). "Zebrafish *pax*[*b*] is involved in the formation of the midbrain-hindbrain boundary." *Nature*, **360**, 87-89.

Kuner, J., M. Nakanishi, Z. Ali, B. Drees, E. Gustavson, J. Theis, L. Kauvar, T. Kornberg and P. O'Farrel. (1985). "Molecular cloning of *engrailed*, a gene involved in the development of pattern in *Drosophila melanogaster*." *Cell*, 42, 309-316.

Lai, J. S., M. A. Cleary and W. Herr. (1992). "A single amino acid exchange transfers VP16induced positive control from the Oct-1 to the Oct-2 homeodomain." *Genes & Dev.*, 6, 2058-2065.

Langlois-Ullah, M.C. (1992) "Etude de l'expression de l'ADNc B1, copie d'une ARN messager alternatif du gène Pax-QNR." *Diplôme d'Etudes Approfondies "Sciences de la Vie et de la Santé"*, Université des Sciences et Techniques de Lille.

Levine, E. M., P. F. Hitchcock, E. Glasgow and N. Schechter. (1994). "Restricted expression of a new paired-class homeobox gene in normal and regenerating adult goldfish retina." *J. Comp. Neurol.*, 348, 596-606.

Lin, L. and W. McGinnis. (1992). "Mapping functional specificity in the Dfd and Ubx homeodomains" Genes & Dev., 6, 1071-1081.

Liu, Y. S. C., J. 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 th neuroretinaand inner nuclear layer." *Neuron*, 13, 377-393.

Logan, C., W. Khoo, D. Cado and A. L. Joyner. (1993). "Two enhancer regions in the mouse *En-2* locus direct expression to the mid-hindbrain region and mandibular myoblast." *Dev.*, 117, 905-918.

Lu, M., Z. Gong, W. Shen and A. D. Ho. (1991). "The TCL-3 proto-oncogene altered by chromosomal translocation in T-cell leukemia codes for a homeobox protein." *EMBO J.*, 10, 2905-2910.

MacLeod, K., D. Leprince and D. Stehelin. (1992). "The ets gene family" 17, 251-256.

Malicki, J., K. Schughart and W. McGinnis. (1990). "Mouse Hox-2.2 specifies thoracic segmental identity in drosophila embryos and larvae." *Cell*, 63, 961-967.

Martha, A., R. E. Ferrell, H. Mintz-Hittner, L. A. Lyons and G. F. Saunders. (1994). "Paired Box Mutations in Familial and Sporadic Aniridia Predicts Truncated Aniridia Proteins." *Amer. J. Hum. Genet.*, 54, 801-811.

Martha, A. D., R. E. Ferrell and G. F. Saunders. (1994). "Nonsense Mutation in the Homeobox Region of the Aniridia Gene." *Human Mutation*, 3, 297-300.

Martin, P., C. Carrière, C. Dozier, B. Quatannens, M.-A. Mirabel, B. Vandenbunder, D. Stéhelin and S. Saule. (1992). "Characterization of a paired box- and homeobox-containing quail gene (*Pax-QNR*) expressed in the neuroretina." *Oncogene*, 7, 1721-1728.

Matsuo, T., N. Osumi-Yamashita, S. Noji, H. Ohuchi, E. Koyama, F. Myokai, N. Matsuo, S. Taniguchi, H. Doi, S. Iseki, Y. Ninomiya, M. Fujiwara, T. Watanabe and K. Eto. (1993). "A mutation in the Pax-6 gene in rat *Small Eye* is associated with impaired migration of midbrain crest cells." *Nature Genet.*, **3**, 299-304.

Maulbecker, C. and P. Gruss. (1993). "The oncogenic potential of deregulated homeobox genes." *Cell Growth Differ.*, **4**, 431-441.

Maulbecker, C. C. and P. Gruss. (1993). "The oncogenic potential of Pax genes." *EMBO J.*, 12, 2361-2367.

McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa and W. J. Gehring. (1984). "A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans." *Cell*, 37, 403-408.

McGinnis, N., M. A. Kuziora and W. McGinnis. (1990). "Human Hox-4.2 and Drosophila Deformed encode similar regulatory specificities in Drosophila embryos and larvae." Cell, 63, 969-976.

McGinnis, W. and R. Krumlauf. (1992). "Homeobox and axial patterning." Cell, 68, 283-302.

Mendel, D. B. and G. R. Crabtree. (1991a). "HNF-1, a member of a novel dimerizing homeodomain protein." J. Biol. Chem., 266, 677-680.

Mendel, D. B., L. P. Hansen, M. K. Graves, P. B. Conley and G. R. Crabtree. (1991b). "HNF-1 α and HNF-1 β (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers *in vitro*." *Genes & Dev.*, **5**, 1042-1056.

Miler, D. J. and A. Miles. (1993). "Homeobox genes and the zootype." Nature, 365, 215-216.

Mizuno, K., I. Okano, K. Ohashi, K. Nunoue, K. Kuna, T. Miyata and T. Nakamura. (1994). "Identification of human cDNA encoding a novel protein kinase with two repeats of the LIM/double zinc finger motif." *Oncogene*, 9, 1605-1612.

Mlodzik, M. and W. J. Gehring. (1987). "Expression of the *caudal* gene in the germ line of Drosophila : formation of an RNA and protein gradient during early embryogenesis." *Cell*, **48**, 465-478.

Monaghan, A., D. Davidson, C. Sime, E. Graham, R. Baldock, S. Bhattacharya and R. Hill. (1991). "The msh-like homeobox genes define domains in the developing vertebrate eye." *Dev.*, 112, 1053-1061.

Monica, K., D. P. Lebrun, D. A. Dedera, R. Brown and M. L. Cleary. (1994). "Transformation properties of the E2a-Pbx1 chimeric oncoprotein: fusion with E2a is essential, but the Pbx1 homeodomain is dispensable." *Moll. Cell. Biol.*, 14, 8304-8314.

Morata, G. (1993). "Homeotic genes of Drosophila." Curr. Opin. Genet. Dev., 3, 606-614.

Morinaga, T., H. Yasuda, T. Hashimoto, K. Higashio and T. Tamaoki. (1991). "A human a-fetoprotein enhancer-binding protein, ATBF1, contains four homeodomains and seventeen zinc-fingers." *Mol. Cell. Biol.*, **11**, 6041-6049.

Nalik, P., M. Panayotova-Heiermann and O. Pongs. (1989). "Characterization of an oestradiol-stimulated mRNA in the brain of adult male rats." *Moll.Cell. Endocrinol.*, **62**, 235-242.

Neufeld, E. J., D. G. Skalnik, P. M. J. Lievens and S. H. Orkin. (1992). "Human CCAAT displacement protein is homologous to the drosophila homeodomain protein, *cut.*" *Natur Genet.*, 1, 50-55.

Noll, M. (1993). "Evolution and role of Pax genes." Curr. Opin. Genet. Develop., 3, 595-605.

Nornes, H. O., G. R. Dressler, E. W. Knapik, U. Deutsch and P. Gruss. (1990). "Spatially and temporally restricted expression of *Pax-2* during murine neurogenesis." *Dev.*, **109**, 797-809.

Nucci. (1993). IOVS Suppl., 34, 876-.

O'Hara, B. F., C. Bendotti, R. H. Reeves, G. M. Oster, J. T. Coyle and J. D. Gearhart. (1988). "Genetic mapping and analysis of somatostatin expression in Snell dwarf mice." *Brain Res.*, 464, 283-92.

Okada, Y. and C. Tanigushi Shimada. (1976). "High concentration of GABA and high glutamate decarboxylase activity in rat pancreatic exocrine and endocrine cells." *Science*, 194, 620-622.

Olson, E. N. and N. Rosenthal. (1994). "Homeobox genes and muscle patterning." Cell, 79, 9-12.

Ominchinski, J. G., C. Trainor, T. Evans, A. M. Gronenborn, G. M. Clore and G. Felsenfeld. (1993). "A small single "finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex." *Proc. Natl. Acad. Sci. USA*, 90, 1676-1680.

Otting, G., Y. Q. Quian, M. Billeter, M. Muller, M. Affolter, W. J. Gehring and K. Wuthrich. (1990). "Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopie in solution." *EMBO J.*, 9, 3085-3092.

Perkins, A., K. Kongsuwan, J. Visvader, J. M. Adams and S. Cory. (1990). "Homeobox gene expression plus autocrine growth factor production elicits myeloid leukemia" *Proc.Natl.Acad.Sci.USA*, **81**, 8398-8402.

Pittack, C., M. Jones and T. A. Reh. (1991). "Basic fibroblast growth factor induces retinal pigment epithelium to generate neural retina *in vitro*." *Dev*, **113**, 577-588.

Plachov, D., K. Chowdhury, C. Walther, D. Simon, J.-L. Guénet and P. Gruss. (1990). "*Pax8*, a nurine paired box gene expressed in the developing excretory system and thyroid gland." *Dev.*, **110**, 643-651.

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, 1-10.

Plaza, S., D. Grevin, K. MacLeod, D. Stehelin and S. Saule. (1994). "Pax-QNR/Pax-6, a paired- and homeobox- containing protein, recognizes Ets binding sites and can alter the transactivating properties of Ets transcription factors." *Gene Express.*, **4**, 43-52.

Poleev, A., H. Fickenscher, S. Mundlos, A. Winterpacht, B. Zabel, A. Fidler, P. Gruss and D. Plachov. (1992). "*PAX8*, a human paired box gene: isolation and expression in developing thyroid, kidney and Wilms' tumors." *Dev.*, **16**, 611-623.

Price, M., D. Lazzaro, T. Pohl, M. G. Mattei, U. Rüther, J. C. Olivo, D. Duboule and R. Di Lauro. (1992). "Regional expression of the homeobox NKx2-2 in the developing mammalian forebrain." *Neuron*, 8, 241-255.

Püschel, A. W., P. Gruss and M. Westerfield. (1992). "Sequence and expression pattern of Pax-6 are highly conserved between zebrafish and mice." *Dev.*, 114, 643-651.

Quian, Y. Q., M. Billeter, G. Otting, M. Muller, W. J. Gehring and K. Wüthrich. (1989). "The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparaison with prokaryotic repressors." *Cell*, **59**, 573-580.

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

Rabbits, T. H. (1991). "Translocation, master genes, and differences between the origins of acute and chronic leukemia." *Cell*, 641-644.

Rey-Campos, J., T. Chouard, M. Yaniv and S. Cereghini. (1991). "vHNF1 is a homeoprotein that activates transcription and forms heterodimers with HNF1." *EMBO J.*, **10**, 1445-1457.

Roberts, R. C. (1967). "Small eye, a new dominant mutant in the mouse." Genet. Res., 9, 121-122.

Rosenfeld, M. G. (1991). "POU-domain transcription factors : POU-er-ful developmental regulators." *Genes & Dev.*, 5, 897-907.

Rothenpieler, U. W. and G. R. Dressler. (1993). "Pax-2 is required for mesenchyme-to-epithelium conversion during kidney development." Dev., 119, 711-720.

Russel, W. L. (1947). "Splotch, a new mutation in the house mouse Mus musculus." Genetics, 32, 107.

Sadler, I., A. W. Crawford, J. W. Michelsen and M. C. Beckerle. (1992). "Zyxin and cCRP: two LIM domain proteins associated with the cytoskeleton." *J. Cell Biol.*, **119**, 1573-1587.

Sanchez-Garcia, I. and T. H. Rabbitts. (1993). "LIM domain proteins in leukemia and development." Semin. Cancer Biol., 4, 349-358.

Sanchez-Garcia, I. and T. H. Rabbits. (1994). "The LIM domain : a new structural motif found in zinc-finger-like proteins" *TIG*, **10**, 315-320.

Schäfer, B. W., T. Czerny, M. Bernasconi, M. Genini and M. Busslinger. (1994). "Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes." *Nucleic Acids Res.*, 22, 4574-4582.

Scheidereit, C., J. A. Cromlish, T. Gerster, K. Kawakawi, C. G. Balmaceda, R. A. Currie and R. G. Roeder. (1988). " A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein." *Nature*, 336, 551-5557.

Schmahl, W., M. Knoedlseder, J. Favor and D. Davidson. (1993). "Defects of neuronal migration and the 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.

Schmeichel, K. L. and M. C. Beckerle. (1994). "The LIM domain is a modular protein-binding domain." *Cell*, **79**, 211-219.

Scott, M. P. and A. J. Weiner. (1984). "Structural relationships among genes that control development : sequence homology between the *Antennapedia*, *Ultrabithorax* and *fushi-tarazu* loci of *Drosophila*." *Proc. Natl. Acad. Sci. USA*, **81**, 4115-4119.

Shang, Z., V. E. Isaac, H. Li, L. Patel, C. M. Catron, T. Curran, G. T. Montelione and C. Abate. (1994). "Design of a "minimAL" homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure." *Proc. Natl. Acad. Sci. USA*, **91**, 8373-77.

Shapiro, D. N., J. E. Sublett, B. Li, J. R. Downing and C. W. Naeve. (1993). "Fusion of *PAX3* to a member of the forkhead family of transcription factors in human alveolar rhabdomyosarcoma." *Canc. Res.*, 53, 5108-5112.

Shena, M., A. M. Lloyd and R. W. Davis. (1993). "The HAT4 gene of Arabidopsis encodes a developmental regulator." *Genes & Dev.*, 7, 367-379.

Shena, M. and R. W. Davis. (1994). "Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families." *Proc. Natl. Acad. Sci. USA*, 91, 8393-8397.

Silver, P. A. (1991). "How the proteins enter in the nucleus." *Cell*, 64, 489-497.

Simeone, A., D. Acampora, L. Arcioni, P. W. Andrews, E. Boncinelli and F. Mavilio. (1990). "Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells." *Nature*, **346**, 763-766.

Simeone, A., D. Acampora, V. Nigro, A. Faiella, M. D'Esposito, A. Stronaiuolo, F. Mavilio and B. E. (1991). "Differential regulation by retinoic acid of the homeobox genes of the four HOX loci in human embryonal carcinoma cells." *Mech. Dev.*, 33, 215-228.

Simeone, A., M. Gulisano, D. Acampora, A. Stornaiuolo, M. Rambaldi and E. Boncinelli. (1992a). "Two vertebrate homeobox genes related to the *Drosophila empty-spiracles* gene are expressed in the embryonic cerebral-cortex." *EMBO J.*, **11**, 2541-2550.

Simeone, A. (1992b). "Nested expression domains of four homeobox genes in the developing brain." *Nature*, 358, 687-.

Skalnik, D. G., E. C. Strauss and S. H. Orkin. (1991). "CCAAT displacement protein as a repressor of the myelomonocytic-specific gp91-phox gene promoter." *J.Biol.Chem.*, 266, 16736-16744.

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.

Stark, M. R. and A. D. Johnson. (1994). "Interaction between two homeodomain proteins is specified by a short C-terminal tail." *Nature*, 371, 429-432.

Stornaiuolo, A., D. Acampora, M. Pannese, M. D'Esposito, F. Morelli, E. Migliaccio, M. Rambaldi, A. Faiella, V. Nigro, S. A. and B. E. (1990). "Human HOX genes are differentially activated by retinoic acid in embryonal carcinoma cells according to their position within the four loci." *Cell Differ. Dev.*, **31**, 119-127.

Stoykova, A. and P. Gruss. (1994). "Roles of *Pax*-Genes in Developing and Adult Brain as Suggested by Expression Patterns." J. Neurosci., 14, 1395-1412.

Strachan, T. and A. P. Read. (1994). "PAX genes" Curr. Opin. Genet. Develop., 4, 427-438.

Stuart, J. J., S. J. Brown, R. W. Beeman and R. E. Denell. (1991). "A deficiency of the homeotic complex of the beetle Tribolium." *Nature*, 350, 72-74.

Sturn, R. A., G. Das and W. Herr. (1988). "The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeobox subdomain." *Genes & Dev.*, 2, 1582-1599.

Tassabehji, M., A. P. Read, V. E. Mewton, R. Harris, R. Balling, P. Gruss and T. Strachan. (1992). "Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired box gene." *Nature*, 355, 635-636.

Tassabehji, M., V. E. Newton, K. Leverton, E. Seemanova, J. Kunze, K. Sperling, T. Strachan and A. P. Read. (1994a). "PAX3 gene structure and mutations: close analogies between Waardenburg syndrome and the *Splotch* mouse." *Hum. Mol. Genet.*, **3**, 1069-1074.

Tassabehji, M., V. Newton and A. P. Read. (1994b). "Waardenburg syndrome type 2 caused by mutations in the human microphtalmia (MITF) gene." *Nature Genet.*, 8, 251-255.

Tavassoli, K., A. Poleev, S. Zannini, A. Feliciello, A.M. Musti and D. Platchov. (1994). "Functionnal analysis of the human PAX-2 and PAX-8 transcription factors." *The First EMBL Meeting on Transcription*. Heidelberg.

Theill, L. E., K. Hattori, D. Lazzaro, J. L. Castrillo and M. Karin. (1992). "Differential splicing of the GHF-1 primary transcripts gives rise to two functionally distinct homeodomain proteins." *EMBO J.*, 11, 2261-2269.

Timmons, P. M., J. Wallin, P. W. J. Righby and R. Balling. (1994). "Expression and function of *Pax-1* during development of the pectoral girdle." *Dev.*, **120**, 2773-2785.

Ton, C. C. T., H. Hirvonen, H. Miwa, M. M. Weil, P. Monaghan, T. Jordan, V. van Heyningen, N. D. Hastie, H. Meijers-Heijboer, M. Dreschler, 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.

Treacy, M. M., X. He and M. G. Rosenfeld. (1991). "I-POU: a POU-domain protein that inhibits neuron-specific gene." *Nature*, 350, 577-584.

Treacy, M. N. and M. G. Rosenfeld. (1992a). "Expression of a family of POU-domain protein regulatory genes during development of the central nervous system." *Ann. Rev. Neurosci.*, 15, 139-165.

Treacy, M. N., L. I. Nelson and e. al. (1992b). "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.

Treisman, J., P. Göncsy, M. Vashishtha, E. Harris and C. Desplan. (1989). "A single amino acid can determine the DNA binding specificity of homeodomain proteins." *Cell*, **59**, 553-562.

Treisman, J., E. Harris and C. Desplan. (1991). "The paired box encodes a second DNAbinding domain in the Paired homeo domain protein." *Genes Dev.*, 5, 594-604.

Turner, E. E., K. J. Jenne and M. G. Rosenfeld. (1994). "Brn-3.2: a Brn-3-related transcription factor with distinctive central nervous system expression and regulation by retinoic acid." *Neuron*, **12**, 205-218.

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." *Molec. Endocrinol.*, **8**, 929-938.

Urbanek, P., Z.-Q. Wang, I. Fetka, E. F. Wagner and M. Busslinger. (1994). "Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax-5/BSAP" *Cell*, **79**, 901-912.

Valarché, I., J. P. Tissier-Seta, M. R. Hirsch, S. Martinez, C. Goridis and J. F. Brunet. (1993). "The mouse homeodomain protein Phox2 regulates Ncam promoter activity in concert with Cux/CDP and is a putative determinant of neurotransmitter phenotype." *Dev.*, **119**, 881-896.

Vogan, K. J., D. J. Epstein, D. G. Trasler and P. Gros. (1993). "The splotch-delayed (Sp^d) mouse mutant carries a point mutation within the paired box of the *Pax-3* gene." Genomics, 17, 364-369.

Voss, J. W., L. Wilson, S. J. Rhodes and M. G. Rosenfeld. (1993). "An alternative Pit-1 RNA splicing product reveals modular binding and nonmodular transcriptional activities of the POU-specific domain." *Molec. Endocrinol.*, 7, 1551-1560.

Wadman, I., J. Li, R. O. Bash, A. Forster, H. Osada, T. H. Rabbitts and R. Baer. (1994). "Specific in vivo association between the bHLH and LIM proteins implicated in human T cell leukemia." *EMBO J.*, 13, 4831-4839.

Waldorf, U. and W. J. Gehring. (1992). "Empty spiracle, a gap gene containing a homeobox involved in Drosophila head development." EMBO J., 11, 2247-2259.

Wallin, J., Y. Mizutani, K. Imai, N. Miyashita, K. Moriwaki, M. Tanigushi, H. Koseki and R. Balling. (1993). "A new Pax gene, Pax-9, maps to the mouse chromosome 12." *Mamm. Genome*, 4, 354-358.

Wallin, J., J. Wilting, H. Koseki, R. Fritsch, B. Christ and R. Balling. (1994). "The role of *Pax-1* in axial skeleton development." *Dev.*, **120**, 1109-1121.

Walther, C. and P. Gruss. (1991). "Pax-6, a murine paired box gene, is expressed in the developing CNS." Dev., 113, 1435-1449.

Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson, L. Hillier, R. Durbin, P. Green, R. Shownkeen, N. Halloran, M. Metzstein, T. Hawkins, R. Wilson, M. Berks, Z. Du, K. Thomas, J. Thierry-Mieg and J. Sulston. (1992). "A survey of expressed genes in *Caenorhabditis elegans.*" *Nature Genet.*, 1, 114-123.

Way, J. C. and M. Chalfie. (1988). "mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans." Cell, 54, 5-16.

Wegner, M., D. W. Drolet and M. G. Rosenfeld. (1993). "POU-domain proteins: structure and function of developmental regulators." *Curr. Opin. Cell Biol.*, 5, 488-498.

Wernert, N., F. Gilles, V. Fafeur, F. Bouali, M. B. Raes, C. Pyke, T. Dupressoir, G. Seitz, B. Vandenbunder and D. Stehelin. (1994). "Stromal expression of c-Ets-1 transcription factor correlates with tumor invasion." *Cancer. Res.*, 54, 5683-5688.

Williams, B. A. and C. P. Ordahl. (1994). "Pax-3 expression in segmental mesoderm marks early stages in myogenic cell specification." Dev., 120, 785-796.

Wilson, D., G. Sheng, T. Lecuit, N. Dostatni and C. Desplan. (1993). "Cooperative dimerization of Paired class homeo domains of DNA." Genes & Dev., 7, 2120-5124.

Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson and C. O. Pabo. (1991). "Crystal structure of MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions." *Cell*, **67**, 517-528.

Wurst, W., A. B. Auerbach and A. L. Joyner. (1994). "Multiple developmental defects in the Engrailed-1 mutant mice: an early mid-hindbrain deletion and patterning defects in the forelimb and the sternum." *Dev.*, **120**, 2065-2075.

Xue, D., T. Yuan and M. Chalfie. (1993). "Cooperative interactions between the *Caenorhabditis elegans* homeoproteins UNC-86 and MEC-3." *Science*, 261, 1324-1328.

Zannini, M., H. Francis-Lang, D. Plachov and R. Di Lauro. (1992). "Pax-8, a paired domaincontaining 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., D. Sartori and F. Mavilio. (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., D. J. Andrew, L. D. Mathies, M. A. Horner and S. M.P. (1993). "Ectopic expression and function of the *Antp* and *Scr* homeotic genes: the N-terminus of the homeodomain is critical to functionnal specificity." *Dev.*, **118**, 339-352.