

2m11.2629.7

Université de Montréal

**CARACTÉRISATION DU GÈNE
DE LA CONVERTASE DES PROTOPROTEINES PC7
CHEZ LA SOURIS**

par

Fouad Habbouche

Sciences biomédicales

Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en sciences biomédicales

Décembre 1997

© Fouad Habbouche, 1997



W
4
U58
1998
V.067

Université de Montréal

CARACTÉRISATION DU GÈNE
DE LA CONVERTASE DES PROTEINES PCT
CHEZ LA SOURIS

par

Fouad Haboubche
Sciences biomédicales
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître en sciences (M.Sc.)
en sciences biomédicales

Décembre 1997



Fouad Haboubche, 1997

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:
**CARACTÉRISATION DU GÈNE DE LA CONVERTASE
DES PROPROTÉINES PC7 CHEZ LA
SOURIS**

Présenté par:
Fouad Habbouche

a été évalué par un jury composé des personnes suivantes:

Président rapporteur: Dr. Muriel Aubry
Directeur de recherche: Dr. Majambu Mbikay
Membre du jury: Dr. Ryszard Brzezinski

Mémoire accepté le:

20.04.1998

Sommaire

Un grand nombre de protéines et de peptides bioactifs nécessaires au développement normal et au maintien de l'organisme sont produits à partir de précurseurs inactifs qui subissent des coupures endoprotéolytiques au niveau de sites mono- ou dibasiques par les membres de la famille des convertases des proprotéines.

Le dernier membre de cette famille récemment découvert est la PC7. Cette convertase est une protéine trans-membranaire du type I composée de 747 acides aminés. Elle est exprimée de façon généralisée dans l'organisme, mais elle est particulièrement abondante dans le thymus, la rate et les testicules.

Peu était connu sur la structure du gène de la PC7 avant le travail présenté dans ce mémoire. On savait par contre, à partir des homologies de séquences dans la région catalytique, que ce gène est plus rapproché des kexines de levure que les autres convertases de mammifères. Sur cette base, nous avons émis l'hypothèse que la structure de son gène, quand à la position des introns, pourrait être différente de celles des autres convertases. Nous avons cherché à vérifier cette hypothèse par caractérisation du gène de la PC7 chez la souris.

Nous avons procédé, comme base de départ, en plaçant les introns de façon hypothétique sur l'ADNc de la PC7 de souris par comparaison avec les gènes d'autres convertases de proprotéines. Nous avons ensuite vérifié leur présence par séquençage à la PCR de l'ADN amplifié à l'aide d'amorces exoniques à partir de séquences clonées du gène de l'enzyme. Les résultats obtenus nous ont permis de positionner exactement tous les introns. Nous avons ainsi démontré que le gène de la PC7 compte dix exons séparés par neuf introns. En dehors de la région catalytique, la position relative des introns par rapport aux séquences codantes diffère sensiblement de celles des autres membres de la famille des convertases, appuyant ainsi la présomption initiale que le gène de la PC7 aurait divergé très tôt des autres gènes des convertases à partir d'un gène ancestral commun.

Par ailleurs, en comparant des séquences exoniques du gène à celle publiée pour l'ADNc de souris, nous avons relevé quelques polymorphismes de séquence sans effets probables sur la fonction de l'enzyme. Enfin, à la suite d'une réaction d'extension d'amorce, nous avons identifié un site majeur ainsi que deux sites mineurs d'initiation de la transcription .

La connaissance de l'organisation du gène de la PC7 permet désormais d'envisager la production de souris déficientes en cette enzyme, ce qui aiderait à déterminer son importance physiologique, à établir son implication éventuelle dans la fertilité et l'immunité, et à identifier ses substrats naturels. Des études ultérieures sur son promoteur permettront de mieux comprendre les bases moléculaires de son expression tissulaire et de sa régulation.

Table des matières

Sommaire	i
Liste des tableaux	vi
Liste des figures	vii
Liste des abréviations	viii
Remerciements	ix
1. Introduction	1
1.1. Les convertases des proprotéines.....	1
1.1.1. Découvertes des convertases.....	1
1.1.2. Substrats des convertases des proprotéines.....	3
1.1.3. Comparaison structurale des convertases.....	3
1.1.4. Activation des convertases des proprotéines.....	6
1.1.5 Les quatre classes des convertases des proprotéines et leur localisation tissulaire et cellulaire.....	9
1.1.6. Expression des convertases au cours du développement.....	14
1.1.7. Les convertases et leur association aux maladies.....	15
1.2. PC7, la dernière de la famille des convertases.....	16
1.2.1. Découverte de la PC7.....	16
1.2.2. Comparaison de PC7 aux autres convertases.....	18
1.2.3. Expression tissulaire et cellulaire de PC7.....	19
1.2.4. PC7 et le HIV.....	20
1.3. Objet de cette étude.....	21

2. Matériels et Méthodes	23
2.1. Bibliothèques génomiques.....	23
2.1.1. Les bibliothèques génomiques	23
2.2. Isolation des séquences génomiques de la PC7.....	23
2.2.1. Étalement de la bibliothèque.....	23
2.2.2. Transfert des plages de lyse sur membrane et hybridation.....	24
2.2.3. Purification des phages positifs.....	25
2.2.4. Préparation de l'ADN de phages positifs.....	25
2.3. Analyse par buvardage de Southern.....	26
2.4. Sondes moléculaires et leur marquage radioactif.....	26
2.4.1. Les sondes d'ADNc.....	26
2.4.2. Les sondes oligonucléotidiques.....	27
2.5. Réaction de polymérisation en chaîne.....	28
2.6. Sous-clonage des fragments d'ADN.....	29
2.7. Purification de plasmides.....	29
2.7.1. Mini-préparations.....	29
2.7.2. Maxi-préparations.....	30
2.8. Séquençage d'ADN.....	31
2.9. Extension d'amorce.....	32
2.10 Protocoles standards de laboratoire.....	33
3. Résultats	34
3.1. Contrôle de la sonde.....	34
3.2. Criblage de la bibliothèque génomique de souris 129SV.....	34

3.3. Criblage de la librairie génomique de souris Balb/c.....	37
3.4. Sous-clonage des bandes positives avec les sondes 5' et 3' du gène de la PC7.....	37
3.5. Cartographie du gène de la PC7 de souris par la PCR.....	38
3.6. Structure du gène de la PC7 de souris.....	46
3.7. Réaction d'extension d'oligodésoxynucléotide.....	49
4. Discussion.....	51
5. Conclusion.....	64
6. Bibliographie.....	65

Liste des tableaux

	PAGE.
TABLEAU I : Classification des substrats des convertases.....	4
TABLEAU II : Sites de coupure dans l'activation des pro-convertases.....	8
TABLEAU III: Classification des PCs selon leur distribution tissulaire et cellulaire.....	10
TABLEAU IV: Séquences et positions des oligodésoxynucléotides utilisés pour la cartographie.....	40
TABLEAU V : Tailles et jonctions des exons et des introns dans le gène de la mPC7.....	47
TABLEAU VI: Comparaison des tailles des gènes des différentes convertases.....	55
TABLEAU VII: Polymorphismes exoniques du gène de la PC7 de souris.....	61

Liste des figures

	PAGE.
FIGURE 1: Structure des convertases des proprotéines et leur comparaison avec la substilisine BPN et la kexine de levure.....	2
FIGURE 2: Phylogénie des convertases des proprotéines.....	7
FIGURE 3: Coupure séquentielle de la POMC par la PC1 et la PC2.....	13
FIGURE 4: Sonde d'ADNc de la PC7.....	35
FIGURE 5: Vérification de la qualité de la sonde.....	36
FIGURE 6: Séquence de la mPC7.....	41
FIGURE 7: Analyse par des produits de la PCR.....	45
FIGURE 8: Répartition des résidus fonctionnels dans les exons.....	48
FIGURE 9: Localisation des sites d'initiation de la transcription.....	50
FIGURE10: Alignement des introns dans les gènes des convertases.....	56

LISTE DES ABRÉVIATIONS

ACTH	Adrénocorticotropine (<i>adrenocorticotropic hormone</i>)
ADNc	Acide désoxyribonucléique complémentaire
ARNm	Acide ribonucléique messenger
BSA	Albumine de sérum bovin (<i>bovine serum albumin</i>)
RE	Réticulum endoplasmique (<i>ER, endoplasmic reticulum</i>)
gp160	Glycoprotéine de 160 kilodalton
IGF	Facteur de croissance du type insuline (<i>Insulin-like growth factor</i>)
L.B.	Bouillon de Luria (<i>Luria broth</i>)
ODN	Oligodésoxynucléotide
α -MSH	Hormone stimulatrice des mélanocytes α (<i>α-melanocyte stimulating hormone</i>)
PC	Convertases des pro-protéines (<i>proprotein convertase</i>).
P.C.	<i>Post coitum</i>
P.f.u.	Unité de formation de plaque (<i>plaque-forming unit</i>)
POMC	Pro-opiomélanocortine
proVWF	Pro facteur de von Willebrand
PTHrP	Peptide du type à l'hormone parathyroïdienne (<i>parathyroid hormone-related peptide</i>)
RT-PCR	Transcription inverse et réaction de polymérisation en chaîne (<i>reverse transcriptase-polymerase chain reaction</i>)
SDS	Sodium dodécyl sulfate
SNC	Système nerveux central
h	heure
min	minute
TGF β	Facteur de croissance transformant β (<i>transforming growth factor β</i>)
TGN	Réseau transgolgien (<i>Trans Golgi network</i>)
U.V.	Ultra-violet

Remerciements:

So many wonderful people have done so much and supported me to make this master possible. Words alone cannot express my appreciation....I'm truly grateful for their encouragement and belief and I'd like to thank them all.

I would first like to thank the Lord for having given me the strength to accomplish my work and for always being there for me.

To my parents, sister and brother, you have done so much for me, words can never be enough to express my gratitude , you are my breath of life and source of hope, may God bless you and always bring you as much joy as you bring to me. (Thank you for always pushing me to do my best Marianne, you are my inspiration)

To Majambu Mbikay, it is difficult to find words which can show my gratitude for everything you have done for me, for giving me the chance to work with wonderful people, for your total dedication and patience, thank you so much for putting all of your heart into what you do.

Thank you to Jim Rochemont, for your assistance your friendship and for always cheering me up .

To Haidy Tadros, thank you for always caring the way you do and for all the help through the years.

To Francine Sirois, thank you for always answering my questions with a smile and for sharing your talent with me.

To Gilles Croissandeau, thank you for giving your best in the critical reading of the manuscript.

To Sylvie Emond, thank you for your secretarial assistance, for your kindness and your cheerfulness.

To Nabil Seidah, thank you for letting me discover such a wonderful family.

To the jury, thank you for the time you have contributed to read my manuscript, I am truly grateful.

**To Roger, Doris,
Marianne and Jad,
you have always been my inspiration.**

1. INTRODUCTION

1.1. LES CONVERTASES DES PROPROTÉINES

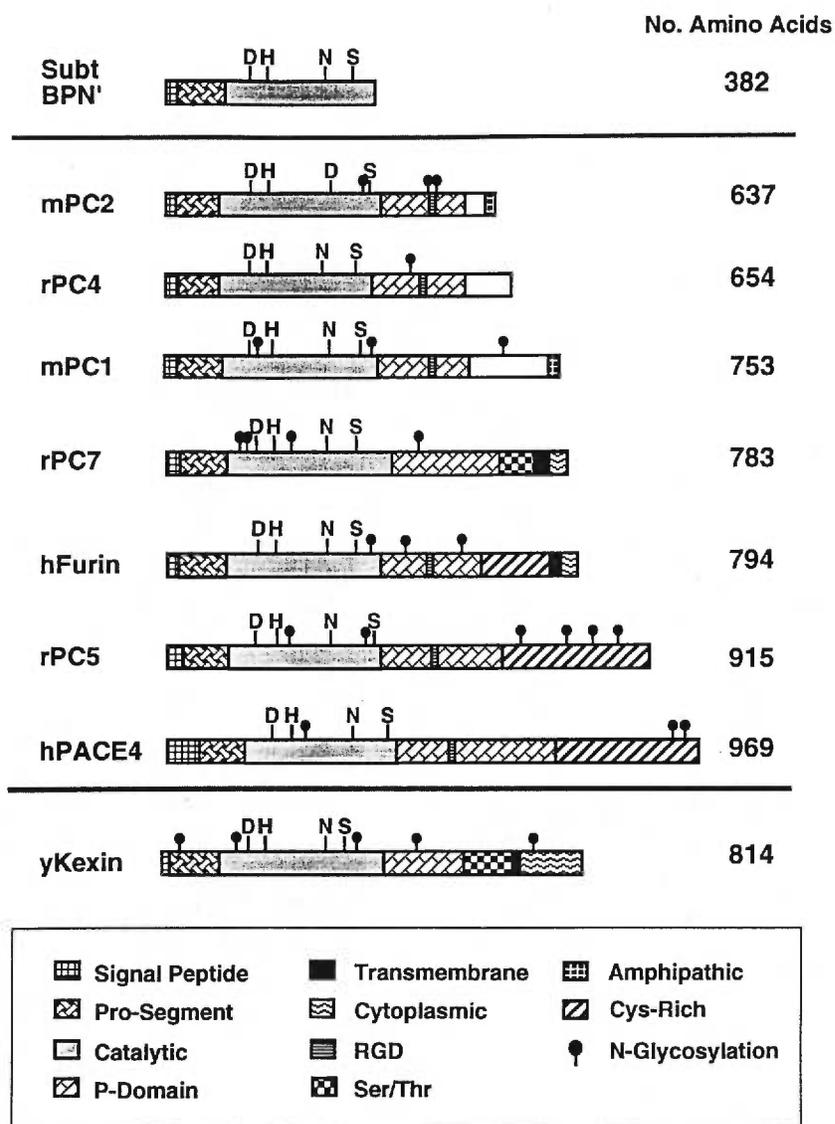
1.1.1. Découverte des convertases

De nombreuses protéines biologiquement actives sont synthétisées à partir de précurseurs qui subissent une ou plusieurs étapes de maturation. Ces étapes sont assurées par différentes familles d'enzymes, parmi lesquelles on trouve notamment les exoprotéases et les endoprotéases. Il est maintenant reconnu que les convertases des pro-protéines (*Proprotein convertases* ou PC) sont des endoprotéases qui coupent des précurseurs polypeptidiques après un résidu ou une paire de résidus basiques et souvent de façon spécifique au tissu (Devi, 1991; Seidah and Chrétien, 1992; Seidah *et al.*, 1997).

Les PCs de mammifères sont des sérines protéinases dépendantes du calcium qui ont une homologie de séquence avec les subtilisines bactériennes et les kexines de la levure (Seidah *et al.*, 1997). Kex2 a été la première convertase à être décrite. Elle a été obtenue et clonée à partir de la levure *Saccharomyces cerevisiae* (Julius *et al.*, 1984). Après l'obtention de cette convertase, plusieurs années se sont écoulées durant lesquelles aucune autre enzyme de cette famille n'a été découverte. Cependant, lors d'une étude sur l'oncogène *fes/fps*, Roebroek *et al.* (1986) ont cloné de façon inattendue la première convertase de mammifère qui a été nommée furine. Par la suite, à l'aide d'oligodésoxynucléotides (ODNs) dégénérés, utilisant des séquences provenant des abords du site actif de l'enzyme de la levure, d'autres convertases ont été découvertes (revu par Chrétien *et al.*, 1995).

À présent, la famille des convertases est constituée de sept membres nommés, PC1 (aussi connue sous le nom de PC3), PC2, furine (aussi dénommée PACE), PACE4, PC4, PC5 (aussi dénommée PC6) et PC7 (aussi nommée SPC7, LPC ou PC8) (revu par Seidah *et al.*, 1997) (figure 1).

Figure 1

Structure des Convertases des Proprotéines

Représentation schématique des convertases des proprotéines de mammifères et leur comparaison avec la subtilisine BPN ancestrale et la kexine de levure.

1.1.2. Substrats des proprotéines convertases.

Comme déjà mentionné ci-dessus, les convertases agissent sur des précurseurs protéiques. Ces substrats appartiennent à des familles de protéines différentes parmi lesquelles on dénombre les précurseurs d'hormones, des neuropeptides, des récepteurs peptidiques, des enzymes protéolytiques, des facteurs de croissance, des molécules d'adhésion cellulaire et des glycoprotéines de surface de virus et de bactéries pathogènes (revu par Seidah *et al.*, 1997)).

Seidah *et al.* (1997) ont regroupé les substrats des convertases en quatre catégories différentes selon leur type de site de coupure (tableau 1). Le premier groupe est composé des précurseurs qui ont un site de coupure de type R-X-(K/R)-R. Parmi ces précurseurs, on compte la gp160, le pro-7B2 et l'intégrine $\alpha 6$. Le deuxième groupe est constitué de précurseurs ayant un site de coupure de type (R/K)-(K/R), comme la majorité des précurseurs hormonaux, telles la pro-opiomélanocortine (POMC) et la pro-insuline. Le troisième groupe comprend des précurseurs coupés après un site composé d'un seul acide aminé basique, occupé par un résidu arginine, et la présence d'un second résidu basique à la position P4, P6 ou P8 est nécessaire. Enfin, le quatrième groupe est constitué de précurseurs coupés après un seul ou une paire de résidus basiques, mais diffère des autres par le fait que la séquence peptidique possède une arginine ou une lysine deux résidus en aval (position P2') du site de coupure.

1.1.3. Comparaison structurale des convertases.

Les étapes menant à la découverte de chaque PC ainsi que l'étude de leurs caractéristiques ont fait le sujet de plusieurs articles de revue (Chrétien *et al.*, 1995; Seidah *et al.*, 1994; Seidah *et al.*, 1997). En général, chaque convertase a des caractéristiques qui lui sont propres, mais elles possèdent aussi des régions d'homologie entre elles (figure 1). Tous les membres de la famille des PCs ont un peptide signal, un prosegment, une sous-unité catalytique contenant le site actif, dans lequel se trouve la triade Asp/His/Ser typique des protéases à sérine, et un domaine nommé domaine P (Wilcox and Fuller, 1991).

Tableau I
Classification des Substrats des Convertases

<u>Precursor protein</u>	<u>Cleavage site sequence</u>
<u>Type I precursors [R-X-(K/R)-R]</u>	P6 P5 P4 P3 P2 P1 P1' P2'
	[X- X- R- X- K/R- R ↓ X- X]
Pro-βNGF	Thr-His-Arg-Ser-Lys-Arg ↓ Ser-Ser
Leptin pro-receptor	Gln-Val-Arg-Glu-Lys-Arg ↓ Leu-Asp
Integrin α6	Asn-Ser-Arg-Lys-Lys-Arg ↓ Glu-Ile
HIV-1 gp160	Val-Gln-Arg-Glu-Lys-Arg ↓ Ala-Val
Pro-fertilin α	Gln-Ser-Arg-Met-Arg-Arg ↓ Ala-Ala
Pro-7B2	Glu-Arg-Arg-Lys-Arg-Arg ↓ Ser-Val
<u>Type II precursors [(K/R)-(K/R)]</u>	P6 P5 P4 P3 P2 P1 P1' P2'
	[X- X- X- X- K/R- R ↓ X- X]
POMC (αMSH/CLIP)	Pro-Val-Gly-Lys-Lys-Arg ↓ Arg-Pro
(γLPH/βEnd)	Pro-Pro-Lys-Asp-Lys-Arg ↓ Tyr-Gly
Pro-Insulin (B/C chain)	Thr-Pro-Lys-Thr-Arg-Arg ↓ Glu-Ala
(C/A chain)	Gly-Ser-Leu-Gln-Lys-Arg ↓ Gly-Ile
Pro-Renin	Ser-Gln-Pro-Met-Lys-Arg ↓ Leu-Thr
PACAP-RP	Ala-Pro-Leu-Thr-Lys-Arg ↓ His-Ser
Integrin α4	His-Val-Ile-Ser-Lys-Arg ↓ Ser-Thr
Pro-fertilin β	Ser-Cys-Lys-Leu-Lys-Arg ↓ Arg-Gly
<u>Type III precursors [Single R]</u>	P8 P7 P6 P5 P4 P3 P2 P1 P1' P2'
	(B)- X- (B)- X- (B)- X- X R ↓ X - X
Pro-Dynorphin (C-peptide)	Arg-Gln-Phe-Lys-Val-Val-Thr-Arg ↓ Ser-Gln
Pro-ANF	Arg-Ala-Leu-Leu-Thr-Ala-Pro-Arg ↓ Ser-Leu
Pro-Somatostatin (SS-28)	Glu-Met-Arg-Leu-Glu-Leu-Gln-Arg ↓ Ser-Ala
Pro-IGF-I	Leu-Lys-Pro-Thr-Lys-Ala-Ala-Arg ↓ Ser-Ile
Pro-IGF-II	Ala-Thr-Pro-Ala-Lys-Ser-Glu-Arg ↓ Asp-Val
Pro-EGF (N-terminal)	Glu-Asp-Gly-His-His-Leu-Asp-Arg ↓ Asn-Ser
Pro-EGF (C-terminal)	Asp-Leu-Arg-Trp-Trp-Glu-Leu-Arg ↓ His-Ala
<u>Type IV precursors [P2'=R/K]</u>	P8 P7 P6 P5 P4 P3 P2 P1 P1' P2'
	(B)- X- (B)- X- (B)- X- (B)-R/K ↓ X -R/K
PACAP-RP	Leu-Ala-Ala-Val-Leu-Gly-Lys-Arg ↓ Tyr-Lys
Pro-Mullerian Inhibiting Substance	Glu-Gly-Arg-Gly-Arg-Ala-Gly-Arg ↓ Ser-Lys
Pro-Glucagon	Gln-Trp-Leu-Met-Asn-Thr-Lys-Arg ↓ Asn-Arg
Pro-IGF-II	Glu-Ala-Phe-Arg-Glu-Ala-Lys-Arg ↓ His-Arg
Pro-PTP-μ receptor	Val-Glu-Glu-Glu-Arg-Pro-Arg-Arg ↓ Thr-Lys

(Tiré de Seidah *et al.*, 1997)

Classification des substrats des convertases selon leur site de coupure. Le B dans le consensus représente un acide aminé basique.

Il est intéressant de signaler que le domaine catalytique est le plus conservé des domaines entre les sept PCs. Dans la région centrale du domaine P, on trouve la séquence pentapeptidique RRGDL, hautement conservée dans la plupart des PCs excepté dans la PC7 où le Asp est remplacé par une Ser (RRGSL). Il est intéressant de noter que cette sérine pourrait être phosphorylée par une kinase qui reconnaît le motif Ser-X-Glu (Seidah *et al.*, 1996a). La fonction de cette séquence RRGDL ou RRGSL est toujours inconnue. Différentes hypothèses ont été émises à ce sujet. Cette séquence pourrait interagir avec des protéines apparentées aux intégrines dans la cellule lors du passage de l'enzyme à travers la voie sécrétoire comme l'avaient suggéré Seidah *et al.* (1990). De plus, la mutagenèse de cette séquence dans la PC1 diminue de façon très importante la coupure et l'activation du zymogène dans le réticulum endoplasmique (ER, *endoplasmic reticulum*) et détourne la protéine vers la voie de sécrétion constitutive et non vers les granules de sécrétion où réside normalement cette convertase (Lusson *et al.*, 1997).

La région la plus variable des PCs est le segment C-terminal. Dans le cas de la PC1 et de la PC2, on trouve une hélice amphipatique regroupant sur une face des résidus hydrophobiques et sur une autre des résidus hydrophiliques (revue dans Seidah *et al.*, 1997). La furine, la PACE4 et la PC5 ainsi qu'une variante de la PC5 nommée PC5B ont un domaine riche en cystéines dont le nombre se répète 2, 5 ou 22 fois respectivement (Seidah *et al.*, 1994; Seidah *et al.*, 1997). Il est intéressant de noter que ce même domaine riche en cystéines est présent dans les convertases homologues à la furine chez la drosophile (*Drosophila melanogaster*) (Roebroek *et al.*, 1992). La fonction de ce domaine est toujours inconnue. De toutes les PCs connues jusqu'à présent, seules la furine, un isoforme de la PACE4, la PC5-B et la PC7 possèdent un domaine transmembranaire qui sert à ancrer l'enzyme à la surface cellulaire (Bruzzaniti *et al.*, 1996; Constam *et al.*, 1996; Meerabux *et al.*, 1996; Seidah *et al.*, 1996).

Chaque convertase a une homologie plus ou moins importante avec la kexine (44%-46%) ou avec la substilisine BPN'(23%-28%) (Seidah *et al.*, 1994).

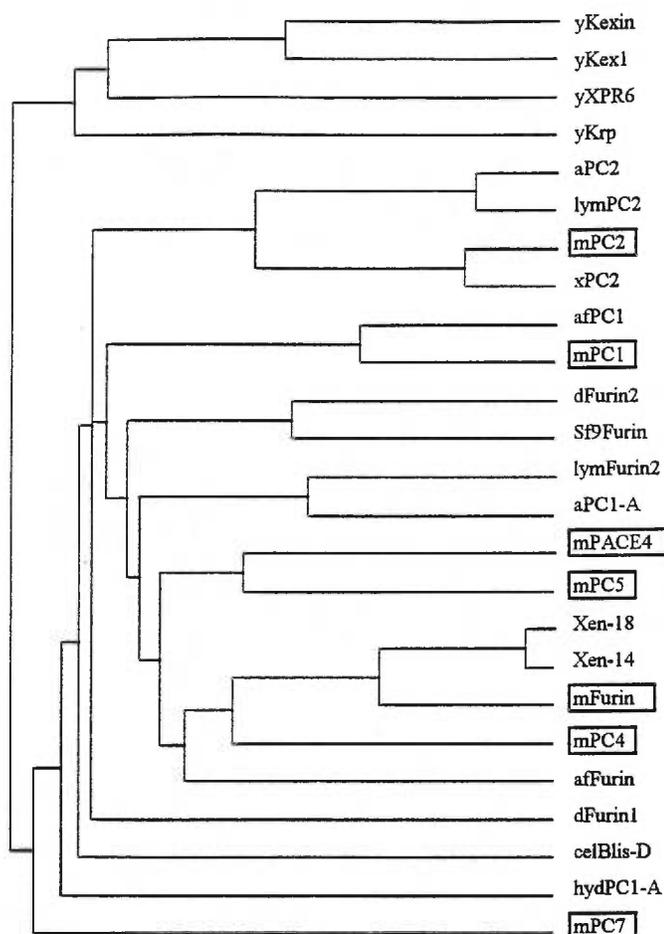
L'appartenance des PCs à une même famille est soutenue non seulement par leur homologie de séquence qui s'étend de 55 à 69%, particulièrement dans la région autour du site actif (60% au moins), mais aussi par les similitudes d'organisation intron-exon de leurs gènes (Ftouhi *et al.*, 1994; Mbikay *et al.*, 1994). Vue sous l'angle de l'évolution, l'existence d'une telle famille de gènes signifie qu'il devrait exister une diversification des fonctions de ces enzymes. En appliquant un programme informatisé d'alignement phylogénique des séquences (*Phylogenetic Tree Alignment Program of the GCG package*) aux domaines catalytiques des PCs, Seidah *et al.* (1996) ont démontré que la PC7 est la PC la plus proche des kexines, Krp, XPR6 et Kex1 de la levure. Ainsi, la PC7 aurait divergé très tôt des autres PCs (figure 2).

1.1.4. Activation des convertases des proprotéines.

Les convertases des proprotéines sont initialement produites sous formes inactives et ne peuvent agir sur leurs substrats que si elles sont activées. Ainsi, le prosegment doit être coupé afin de libérer la forme active de l'enzyme (voir tableau 2). La coupure se fait après la séquence R-X-(K/R)-R qui ressemble à la séquence contenue dans les précurseurs de type I reconnus par les convertases (Seidah *et al.*, 1997). Le devenir du prosegment après une coupure autocatalytique intramoléculaire dans le ER n'est toujours pas connu. Il a été proposé, entre autres hypothèses, qu'après coupure, le prosegment demeurerait dans le réticulum endoplasmique où, comme dans le cas de la subtilisine (Eder *et al.*, 1993), il agirait comme inhibiteur compétitif attaché de façon non covalente à l'enzyme jusqu'à ce que celle-ci atteigne le réseau transgolgien (TGN, *trans Golgi network*), ou les granules de sécrétion immatures; après quoi il se dissocierait de l'enzyme et serait détruit. Dans le cas de Kex2 et de la furine, il a été démontré que la présence du prosegment permet le bon repliement de la structure tertiaire du zymogène (Anderson *et al.*, 1997; Creemers *et al.*, 1995).

Figure 2

Phylogénie des Convertases des Proprotéines



(Tiré de Seidah *et al.*, 1997)

Dendrogramme représentant l'arbre phylogénétique des sept convertases des proprotéines de souris (préfixe m, *mouse*) connues, mPC7, mPC1, mPC2, mPC4, mFurine, mPC5 et la mPACE4 et leur comparaison à la kexine yKrp, yXPR6, yKexin et yKex 1 de la levure (y, *yeast*), la furine de la drosophile (d, *Drosophila melanogaster*), la blisterase-4 isoform D (celBlis-D) du *Caenorhabditis. elegans* (cel), la atPC1 de l'angler fish (at), la aPC2, la afurine et la aPC1-A de l'aplysie (a, *Aplysia californica*), la hydPC1-A de l'hydre, la lymPC2 du *lymnea stagnalis* et la xPC2, la Xen-14 et la Xen-18 de la grenouille (x, *Xenopus laevis*).

Tableau II

Sites de Coupure dans l'Activation des Pro-convertases

	1						60
mpace4	aLpppRPVY	TN...HWAVq	VI..GGp...ga	ADRV	AaahGYINLG	QIGnLDD..Y
mpc5rVY	TN...HWAVk	Ia..GGf...aE	ADRI	AsKyGFINvG	QIGaLkD..Y
mfurin	kIF	TN...tWAVh	IP..GGp...av	ADRV	AqKhGFhNLG	QI..FgD..Y
mpc4	sAQap.IY	vs...sWAVr	Vt..kGy...qE	AERI	ArKfGFVNLG	QI..FpDdqY
mpc2	E...RPVF	TN...HF1Ve	lhk.dGe...eE	ArqV	AaehGFg.vr	klpfaEg..l
mpc1	k...Rq.F	vN...eWAae	IP..GGq...ea	AsaI	AeelGYdlLG	QIGsLEn..h
mpc7	lsEAqGldIL	gtgglSWAVh	ldsleGerke	esltqq	ADaV	AqaaGLVNaG	rIGeLqg..h
Consensus	-----	-----	-----G-----	-----A---	A---	G-----	-----
	61				100		120
mpace4	YHFYH...Sk	tf.KRStLSs	.RGpHtflRm	DPqVKWLqQQ	evKRRvKR↓q.	..arsdsLYF	
mpc5	YHFYH...SR	tI.KRSvLSS	.RGtHSfiSm	EPkVeWiqQQ	vvKkRtKR↓dy	dlshaqstYF	
mfurin	YHFWH...R	aVtKRS.LSp	HRpRHSrLqR	EPqVKWLEQQ	vaKRRaKR↓dV	yqePt.....	
mpc4-a	FHLRH...R	gVAqqS.Ltp	HWGhr1rLkk	DPkVrWFEQQ	t1rRRvKR↓S1	v.vPt.....	
mpc2	YHFYHnglak	akArRS.L..	HhkRq..LeR	DPrIKmalQQ	egfdRkKR↓gy	rdineidinm	
mpc1	YlEkHkshpR	.rsrRSAL..	Hitkr..LSd	DdrVtWaEQQ	yeKeRsKR↓SV	qkdsaldL.F	
mpc7	YlFvqptghR	qameveAM..	rqqaeavLaR	heaVrWhseQ	t1lkRaKR↓SI	h.....F	
Consensus	-----	-----	-----	-----Q	----R-KR↓--	-----	

(Tiré de Seidah *et al.*, 1997)

Alignement des séquences des prosegments des sept convertases de proprotéines de mammifères. La flèche indique le site de coupure du prosegment.

L'analyse de la séquence d'acide aminé du prosegment montre que plusieurs résidus (Gln, deux Ala et Gly) sont conservés dans les sept membres connus de la famille des convertases des mammifères. Cela pourrait suggérer que ces acides aminés ont un rôle important dans la fonction du prosegment (Seidah *et al.*, 1997).

1.1.5 Les quatre classes des convertases des proprotéines et leur localisation tissulaire et cellulaire.

Les convertases des proprotéines, selon leur distribution et leur localisation intracellulaire, peuvent être divisées en quatre classes différentes (tableau 3) (Seidah *et al.*, 1997). Les membres de la première classe sont la furine et la PC7, exprimées dans la majorité des tissus. L'expression de la PC7 est relativement abondante dans les tissus des systèmes digestif et reproductif ainsi que dans le système nerveux central (SNC: cortex, striatum, hippocampe, thalamus, hypothalamus, cervelet, moëlle épinière) et l'hypophyse. Par contre, l'ARNm de la furine se trouve en plus grande quantité dans le foie mais est aussi retrouvé dans le SNC. Chez le rat adulte, les ARNm de la furine et de la PC7 sont exprimés fortement dans la rate et le thymus. De plus, l'expression de la furine est plus importante dans la pulpe rouge de la rate tandis que celle de la PC7 est plus élevée dans la pulpe blanche (Seidah *et al.*, 1996). Ainsi, même si ces enzymes sont exprimées de façon généralisée, les cellules qui les expriment en plus grande quantité sont différentes, ce qui suggère que la PC7 et la furine agiraient sur des précurseurs différents. Ces enzymes agissent sur les précurseurs qui rejoignent la surface cellulaire en passant par la voie de sécrétion constitutive. Certains des substrats de la furine sont le pro-TGF β et le récepteur de l'insuline (Blanchette *et al.*, 1997). La furine comme la PC7 coupe la gp160 du HIV (Decroly *et al.*, 1997; Decroly *et al.*, 1996).

Tableau III

Classification des PCs selon leur Distributions Tissulaire et Cellulaire

Convertase (PC)	Tissue expression	Cellular Localization
<u>Class I</u>		
Furin PC7	Ubiquitous , but very rich in a Number of Tissues Widespread , very rich in Lymphoid-Associated Tissues	TGN, Endosomes , Cell Surface TGN
<u>Class II</u>		
PC1	Neural and Endocrine Cells , rich in Peripheral tissues and CNS	TGN, Secretory Granules
PC2 (7B2)	Neural and Endocrine Cells , rich in Peripheral tissues and very rich in the CNS	Secretory Granules
<u>Class III</u>		
PC5-A	Endocrine & Non-Endocrine, Widespread , very rich in Adrenal Cortex, Endothelial & Sertoli Cells and Digestive System	TGN, Secretory Granules
PC5-B	Restricted , rich in Digestive System and Adrenal Cortex	TGN
PACE4	Endocrine & Non-Endocrine, Widespread , rich in Pituitary and Chondrocytes; Important in Bone Morphogenesis	TGN, Secretory Granules?
<u>Class IV</u>		
PC4	Testicular and Ovarian Germ cells , Important in Fertility	TGN?

Tiré de Seidah *et al.*, 1997

Les enzymes de la classe I se retrouvent initialement dans le TGN, mais peuvent circuler à la surface cellulaire pour être ensuite reprises par des compartiments endosomaux et retournées au TGN (Molloy *et al.*, 1994).

La deuxième classe d'enzymes comprend la PC1 et la PC2, deux enzymes exprimées dans les tissus neuronaux et endocriniens. Dans les systèmes nerveux central et périphérique des rongeurs, on retrouve les ARNm de la PC1 et de la PC2 dans les neurones de plusieurs centres cérébraux allant du lobe olfactif jusqu'à la moelle épinière et incluant l'hippocampe, les amygdales, le thalamus et l'hypothalamus.

Par ailleurs, on retrouve ces deux enzymes dans les péricaryons des différents centres et surtout dans les fibres nerveuses de l'éminence médiane (Marcinkiewicz *et al.*, 1993; Schafer *et al.*, 1993). La PC2 est aussi retrouvée dans la medullo-surrénale et en plus grande quantité dans le lobe neurointermédiaire de l'hypophyse (Day *et al.*, 1992; Seidah *et al.*, 1990). Ces convertases coupent les précurseurs aux niveaux de doublets d'acides aminés basiques et agissent sur ceux qui empruntent la voie de sécrétion régulée. Les produits de ces coupures sont accumulés dans les granules de sécrétion qui caractérisent les tissus neuroendocriniens, et sont libérés en réponse à des signaux sécrétoires spécifiques émanant de la membrane plasmique. La PC1 et la PC2 traversent le TGN pour se retrouver par la suite dans des granules de sécrétion matures (Benjannet *et al.*, 1993). Les quantités de transcrits PC1 et PC2 sont dans la plupart des cas co-régulés avec celles de leurs substrats (Day *et al.*, 1992). Un des substrats de la PC1 et de la PC2 est la POMC. Dans une étude (Benjannet *et al.*, 1991), la PC1 et la PC2 ont été exprimées avec la POMC dans des cellules à voie de sécrétion constitutive telle la lignée BSC-40, ou dans des lignées cellulaires endocriniennes (à voie de sécrétion régulée) telles les lignées PC12 et AtT-20. Il ressort de ces études que la PC1 coupe la POMC en adrénocorticotropine (ACTH, *adrenocorticotropic hormone*) et en β -lipotropine (β -LPH, *β lipotropic hormone*) et que ces deux substrats sont ensuite

coupés par la PC2 pour donner l'hormone stimulatrice des mélanocytes α (α -MSH, α *melanocyte-stimulating hormone*) et la β -endorphine. Toutefois, la PC2 peut aussi agir directement sur la POMC pour donner la α -MSH et la β -endorphine (figure 3).

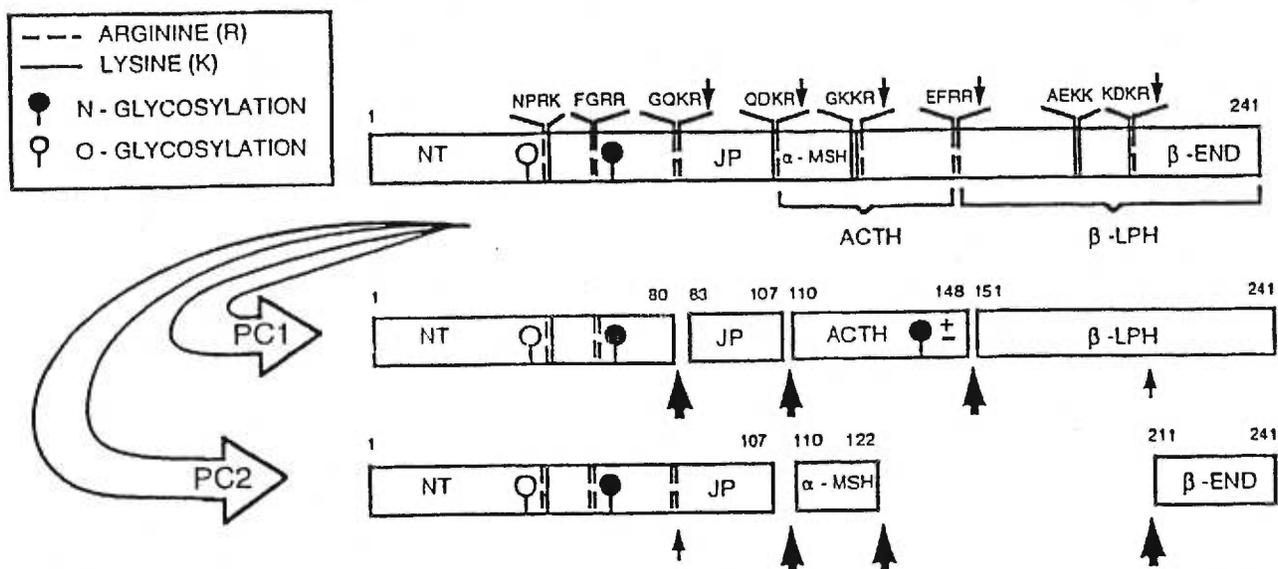
Enfin, il est intéressant de voir que la POMC peut aussi être coupée dans les cellules dépourvues de voie de sécrétion régulée, ce qui démontre que la coupure de la POMC ne dépend pas de la présence de granules de sécrétion (Seidah *et al.*, 1992a).

Le troisième groupe est composé des PCs exprimées autant dans les cellules endocriniennes que non-endocriniennes et qui vont donc agir aussi bien sur les précurseurs retrouvés dans la voie de sécrétion constitutive que sur ceux présents dans la voie de sécrétion régulée. Les enzymes en question sont la PC5 (A et B) et la PACE4 (Seidah *et al.*, 1997). L'ARNm de la PC5 est exprimé de façon généralisée dans les tissus de rat. Il est particulièrement abondant dans certains tissus du système digestif (estomac, duodénum, jéjunum, iléum), les surrénales, les poumons et les ovaires (Lusson *et al.*, 1993). L'ARNm de la PACE4 est plutôt abondant dans le cervelet et la moelle épinière. En terme de localisation intra-cellulaire, la PC5 se retrouve dans le TGN et dans les granules de sécrétion (De Bie *et al.*, 1996). La localisation intracellulaire de la PACE4 est toujours incertaine. Parmi les substrats établis de la PACE4, on compte le facteur von Willebrand (proVWF) (Wise *et al.*, 1990) et parmi ceux de la PC5, il y a la pro-rénine (Mercure *et al.*, 1996).

La PC4 est l'unique membre du quatrième et dernier groupe. Son ARNm se trouve en abondance dans les testicules et particulièrement dans les cellules germinales testiculaires (Seidah *et al.*, 1992b; Torii *et al.*, 1993). Sa localisation intracellulaire n'a pas encore été déterminée.

Figure 3

Coupure Séquentielle de la POMC par la PC1 et la PC2



(Tiré de Benjannet *et al.*, 1991)

Modèle représentant les produits finaux de la coupure de la POMC par la PC1 et la PC2. Les sites de coupures majeurs et mineurs après les paires de résidus basiques sont représentés par de grandes et de petites flèches respectivement. La séquence d'acides aminés et le site de coupure sont aussi montrés. Les numéros représentent le début et la fin des peptides produits, en se basant sur la séquence de la POMC (pPOMC) rapportée. Les sites de N- et O-glycosylation (CHO), ainsi que les résidus des doublets basiques sont représentés.

1.1.6. Expression des convertases au cours du développement

La croissance, la différenciation et le développement des tissus embryonnaires dépendent largement des interactions cellulaires et notamment des facteurs de croissance. Comme évoqué auparavant, de nombreux facteurs de croissance sont synthétisés sous forme de précurseurs inactifs et doivent être activés par des coupures protéolytiques. Il est supposé que leur disponibilité est contrôlée par l'activité de certaines convertases spécifiques (Bresnahan *et al.*, 1993). L'expression des convertases des proprotéines se fait à différentes étapes lors du développement. L'ARNm de la PC5 et de la PACE4 n'est pas exprimé de façon précoce dans l'embryon, mais il est détecté dans les structures extra-embryonnaires et l'utérus (Zheng *et al.*, 1997). Ainsi, l'ARNm de la PACE 4 est initialement détecté en faible concentration dans le côté mésométrial du décida au jour embryonnaire 6 (e6). Au jour e9, l'ARNm de la PACE 4 est observé partout dans le décida entourant l'embryon, mais est absent dans l'embryon jusqu'au jour e12,5 où il commence à être exprimé dans le SNC. Entre les jours e9-e10, l'ARNm de la PC5 est exprimé de façon plus restreinte à l'extérieur de l'embryon et en plus faible quantité dans les sommites de l'embryon. Au jour e10, on observe des transcrits de cette enzyme dans le SNC et dans les cellules du plafond recouvrant la région médiane du cerveau (Zheng *et al.*, 1997). La PC5 et la PACE4 pourraient donc être impliquées dans la maturation des précurseurs dans le système nerveux et dans les tissus périphériques dès le stade embryonnaire.

Dans le cas de la PC1 et de la PC2, leur expression est initiée dès le milieu de la gestation, au jour e13, et est restreinte aux systèmes neural et endocrinien. La PC1 et la PC2 seraient donc impliquées dans la maturation de neuropeptides lors de la formation du système nerveux.

L'expression de la furine est observée au jour e7. Son ARNm est distribué dans plusieurs tissus embryonnaires et fœtaux, surtout à l'extérieur du système nerveux en développement. La furine pourrait donc intervenir dans les coupures protéolytiques des proprotéines des tissus périphériques. Aux stades tardifs de

la gestation, on observe une diminution de l'expression de l'ARNm de la furine parallèle à celle de certains facteurs de croissance fœtaux et de leurs récepteurs (Zheng *et al.*, 1994). La forte expression de la furine durant la première partie de la gestation correspond à une phase critique de l'embryogénèse pendant laquelle des concentrations élevées de facteurs de croissance et de leur récepteurs sont détectées. Ainsi l'expression de la furine est nécessaire dès les premières étapes de la vie embryonnaire. En effet, l'inactivation du gène de la furine chez la souris entraîne la mort embryonnaire (Roebroek *et al.*, 1997).

En ce qui concerne la PC7, son expression est généralisée au cours du développement embryonnaire, mais devient plus localisée et enrichie dans les tissus lymphatiques chez l'adulte (Marcinkiewicz M., manuscrit en préparation)

1.1.7. Les convertases et leurs associations aux maladies.

Les convertases sont impliquées dans la production de neuropeptides et de protéines bioactifs qui sont importants pour un métabolisme normal. Il est donc cohérent de supposer que l'expression anormale, augmentée ou réduite, de ces enzymes pourrait entraîner des conséquences graves et contribuer à la pathophysiologie de certaines maladies. Parmi ces maladies, on peut citer les endocrinopathies telles la maladie de Cushing due à l'hypersécrétion d'ACTH ou l'hypocortisolisme dû à l'insuffisance de la même hormone, les maladies prolifératives tels le cancer et l'artériosclérose, les désordres neurologiques, l'hypertension ainsi que les infections virales (discuté par Chrétien *et al.*, 1995). Des expériences d'inactivation des gènes de convertases ont démontré l'importance des PCs sur la croissance et le développement. Chez la souris, l'inactivation de la PC4 par Mbikay *et al.* (1997) montre que cette enzyme a un rôle important non seulement dans la fertilité chez le mâle mais aussi dans le développement embryonnaire précoce. Des problèmes dans le métabolisme du glucose ont été démontrés par Furuta *et al.* (1997) chez des souris mutantes pour le gène de la PC2. Celles-ci montrent un léger retard de croissance, une hypoglucagonémie. Des problèmes d'obésité, d'hypogonadisme et

d'hypocortisolisme ont été signalés chez une femme qui a des mutations dans les deux allèles du gène de la PC1 (Jackson *et al.*, 1997; O'Rahilly *et al.*, 1995).

1. 2. PC7, LA DERNIÈRE DE LA FAMILLE DES CONVERTASES.

1.2.1 Découverte de la PC7

Des études effectuées dans les cellules LOVO (adénocarcinome du colon) dénuées d'activité furine, ont montré que ces cellules pouvaient toujours couper le précurseur du facteur de croissance apparenté à l'insuline (pro-IGF1, *pro-insulin-like growth factor*) en sa forme mature (Takahashi *et al.*, 1993; Takahashi *et al.*, 1995). De plus, ces cellules ont la capacité de couper la glycoprotéine de surface gp160 du virus HIV en gp120/gp41 (Ohnishi *et al.*, 1994). Ces résultats ont donc suggéré l'existence d'une autre convertase différente de la furine et qui devait être exprimée de façon plus ou moins généralisée dans le corps.

Lors d'études entreprises dans le but de comprendre le mécanisme de coupure post-traductionnelle du peptide apparenté à l'hormone parathyroïdienne (PTHrP, *parathyroid hormone-related peptide*), et afin de pouvoir identifier les membres de la famille des convertases qui ont un rôle dans cette coupure, Bruzzaniti *et al.* (1996) ont caractérisé plusieurs lignées cellulaires produisant le PTHrP, en regardant l'ADNc de leurs convertases. Cette caractérisation a été effectuée par RT-PCR en utilisant des ODNs dégénérés basés sur la séquence d'acides aminés (a.a.) conservée dans la région catalytique des PCs. Ceci leur a permis d'identifier dans la lignée BEN de carcinome de poumon humain un fragment d'ADNc dont la séquence était hautement apparentée aux ADNc des convertases, mais qui en était distincte. Ils ont ensuite cloné et obtenu la séquence complète de cette convertase qu'ils ont baptisée PC8.

Se basant sur la séquence partielle d'a.a. rapportée par ces chercheurs, Seidah *et al.* (1996) ont produit des ODNs dégénérés qui leur ont permis d'obtenir, par RT-PCR à partir d'ARNm des cellules BEN, un ADNc de 285 paires de bases (pb) qu'ils ont ensuite utilisé comme sonde pour cribler deux

librairies d'ADNc de rate de rat et de la lignée PC12 de phéochromocytome de rat. À partir de la librairie de rate de rat, ils ont isolé un clone contenant un ADNc de 1344 pb, et à partir de la librairie PC12, ils ont isolé un ADNc composite représentant les nucléotides 722-3468 de l'ADNc. Ils ont nommé cette convertase PC7.

Un autre groupe a localisé la séquence de cette convertase à un site de translocation du chromosome 11 survenue dans des lymphomes humains. Ils en ont cloné et séquencé l'ADNc et l'ont appelé enzyme LPC (*Lymphoma Proprotein Convertase*). La translocation n'affecte pas la séquence codante de l'enzyme, mais prend place dans un intron de la région 3' de ce gène. (Meerabux *et al.*, 1996).

La taille de l'ARNm de la PC7 varie entre 3,5 et 4,5 kilobases (kb) (Bruzzaniti *et al.*, 1996; Meerabux *et al.*, 1996; Seidah *et al.*, 1996). Ces tailles différentes pourraient provenir d'un épissage alternatif, de l'usage alternatif de promoteur ou d'un motif de polyadénylation différent au cours de la transcription. La séquence clonée de l'ADNc de la PC7 de rat est de 3485 pb; elle contient un cadre de lecture ouvert de 2349 nucléotides encodant une protéine de 783 a.a. Ce cadre est précédé de deux triplets ATG en phase, l'un en position 188 de l'ADNc et l'autre en position 199. Ces ATG initient de courts cadres de lecture qui se terminent respectivement aux positions 250 et 219. Par ailleurs, la séquence entourant la méthionine initiatrice probable n'est pas conforme à la séquence consensus de Kozak. En effet, selon Kozak, la sous-unité de 40S du ribosome s'attache au bout 5' de l'ARNm et migre jusqu'au premier AUG se trouvant dans une région avec un contexte favorable pour initier la traduction. Une telle région contient un A ou un G en position -3 de l'AUG et un G en position +4, en prenant en considération le fait que le A du AUG représente la position +1 (Kozak, 1989). La non-conformité de site d'initiation de la traduction de l'ARNm de PC7 suggère que la traduction de cette protéine est d'une efficacité réduite.

1.2.2. Comparaison de la PC7 aux autres convertases

La PC7, comme les autres convertases, contient un signal peptide, un domaine N-terminal, un domaine catalytique suivi d'un domaine P (ou domaine homo-B) et d'un domaine C-terminal. La coupure du prosegment de 104 a.a. de longueur libère l'enzyme active qui est constituée de 643 a.a. (Seidah *et al.*, 1996).

La région catalytique de la PC7 (289 a.a.) a une forte similitude avec celles de la furine (54 %), la PACE4 (56 %), la PC1 (53 %) et la PC2 (52 %) (Bruzzaniti *et al.*, 1996). De plus, dans le domaine catalytique, la PC7 contient la triade d'a.a. associés au site actif (Asp, His et Ser) ainsi que l'Asn catalytiquement important pour la stabilisation d'oxyanion (Seidah *et al.*, 1996).

En amont de la région catalytique, le degré de conservation de la séquence d'a.a. entre la PC7 et les autres PCs diminue de façon importante. Ainsi, à partir de la Met initiatrice au début du domaine catalytique, la PC7 a une identité de séquence de 17 % avec la furine, de 18 % avec la PACE4 et la PC1 et de 11 % avec PC2 (Bruzzaniti *et al.*, 1996)

Le domaine P qui suit directement la séquence après le domaine catalytique a une homologie plus élevée avec les autres PCs: 24 % d'identité avec la PACE4, 20 % avec la PC1 et la furine et 16 % avec la PC2 (Bruzzaniti *et al.*, 1996). La séquence en aval du domaine P varie énormément parmi les convertases. La PC7, une isoforme de la PACE4 et l'isoforme PC5/6 B possèdent un domaine transmembranaire. Celui-ci est de 17 a.a. dans la PC7. Il est suivi par une queue cytosolique de 100 a.a. Cette queue est plus longue que celle de la furine (56 a.a.), (Seidah *et al.*, 1996a) et celle de la PC5/6 B (84 a.a.). Cependant, dans les quatre enzymes, cette région contient un site de phosphorylation Ser/Thr qui pourrait jouer un rôle dans la régulation du temps de transit des enzymes entre le TGN et la membrane plasmique (Seidah *et al.*, 1996).

La séquence de la PC7 possède six sites de N-glycosylation dont un seul est dans une position homologue dans la furine. Elle possède aussi cinq résidus Cys à des positions homologues dans la furine, la PACE4, la PC1 et la PC2. Elle se

termine avec un résidu Cys comme celle de la furine et de la PACE4. Cependant, elle ne contient pas de région riche en cystéines comme ces deux dernières

1.2.3. Expression tissulaire et cellulaire de PC7

La distribution de la PC7 a été étudiée par hybridation *in situ* sur divers tissus (Seidah *et al.*, 1996). Dans le SNC, l'ARNm de la PC7 est exprimé à des niveaux relativement élevés dans le gyrus denté et l'hippocampe. Des quantités substantielles de l'ARNm de la PC7 sont aussi présentes dans les noyaux hypothalamiques dorso- et ventromédians, dans l'habenula ainsi que dans le noyau arqué. Les transcrits de la PC7 sont relativement abondants dans les tubules séminifères des testicules, ce qui suggère une expression dans les cellules spermatogéniques comme dans le cas de la PC4. Ils sont aussi abondants dans le thymus, la rate ainsi que dans les ganglions lymphatiques, ce qui suggère un rôle plutôt important dans le système immunitaire.

Des études de protection à l'ARNase chez le rat ont démontré que l'ARNm de la PC7 est exprimé partout dans l'embryon aux étapes post-implantatoires et tout au cours de la gestation sans changements observables en abondance (Constam *et al.*, 1996).

Dans les lignées cellulaires établies, le gène de la PC7 est exprimé autant dans les cellules ayant une voie régulée de sécrétion en plus de la constitutive que dans celles ayant la voie constitutive seulement. Son ARNm se retrouve en quantité relativement abondante dans les lignées cellulaires provenant de l'hypophyse telles la lignée gonadotrope α T3-1, la lignée corticotrope AtT20 et la lignée mammosomatotropes GH4C1. La lignée mammosomatotrope GH3 représenterait la seule exception. Des quantités relativement élevées d'ARNm de la PC7 ont aussi été observées dans les cellules RiNm5F et β TC-3 d'insulinome, dans les cellules Y1 de cortex surrénalien, dans les cellules PC12 dérivées d'un phéochromocytome ainsi que dans la lignée fibroblastique Itk.

Dans la lignée NG108 dérivée d'un hybride neuroblastoma-glioma, rat-souris, et la lignée HepG2 dérivée de carcinome hépatocellulaire, le niveau d'ARNm de la PC7 est relativement faible (Seidah *et al.*, 1996).

1.2.4. PC7 et le HIV

Le fait que l'ARNm de la PC7 est retrouvé en abondance dans le thymus suggère un rôle possible de cette convertase dans l'immunité et les maladies immunitaire tel le syndrome d'immunodéficience acquise (SIDA) causé par le virus d'immunodéficience humaine (HIV, *human immunodeficiency virus*). Les glycoprotéines de l'enveloppe des virus HIV-1 et HIV-2 sont synthétisées sous forme de précurseurs gp160 et gp140 puis coupés post-traductionnellement par les enzymes de la cellule-hôte en sous-unités de surface gp120 et gp125 et en sous-unités transmembranaires gp41 et gp36 respectivement. La coupure de ces glycoprotéines de surface est cruciale non seulement pour induire la fusion membranaire cellule-virus, mais aussi pour faciliter la formation de *syncytia* et la production de virions infectieux (Hallenberger *et al.*, 1997). Comme l'ARNm de 3,8 kb de la PC7 est abondant dans la rate, le thymus et les nodules lymphatiques, et afin d'avoir une idée sur l'expression de la PC7 dans des cellules individuelles du système immunitaire, des buvardages par Northern ont été réalisés sur des cultures primaires de cellules de sang et sur différentes cultures de lignées cellulaires dérivées de monocytes-macrophages ou de lymphocytes (Hallenberger *et al.*, 1997). La PC7 et la furine sont les seules PCs à être détectées dans les cultures de cellules primaires qui sont aussi des cibles du HIV. De plus, des populations homogènes de lymphocytes T primaires expriment les transcrits de la PC7 et de la furine mais pas ceux des autres PCs. En utilisant la technique de RT-PCR, il a été démontré que les lymphocytes T4 à l'état inactif expriment en faibles concentrations les ARNm de la PC7 et de la furine. En revanche, quand ils sont activés dans les conditions qui favorisent la réplication du HIV, il y a une augmentation de 5 à 10 fois des niveaux d'ARNm de la furine et de la PC7 (Decroly *et al.*, 1997).

La coexpression de la glycoprotéine de surface gp160 du virus HIV-1 avec différents membres de la famille des convertases a révélé que seules la PC7 et la furine étaient capables de couper ce précurseur glycoprotéique pour donner une protéine d'enveloppe biologiquement active. Les mêmes résultats ont été aussi observés dans le cas de la gp140 du HIV-2. On peut donc constater que PC7 et la furine sont toutes les deux capables d'activer la glycoprotéine de surface HIV *in vitro* et qu'ils ont peut-être un rôle dans le cycle d'infection de ce virus. La question qui se pose toujours est de savoir si, *in vivo*, la PC7 seule ou la furine seule ou les deux ensemble sont nécessaires pour cette conversion (Decroly *et al.*, 1996; Hallenberger *et al.*, 1997; Ohnishi *et al.*, 1994).

Dans une récente étude dans laquelle ils comparaient le rôle de la PC7 et de la furine, Decroly *et al.* (1997) ont suggéré que le fait que ces deux enzymes soient ancrées à la membrane favoriserait leur interaction avec la glycoprotéine de surface gp160 du HIV, tout en reconnaissant que la PC7 ou la furine pourrait activer une autre protéase encore non définie qui elle agirait sur la gp160

1.3. OBJET DE CETTE ÉTUDE

Comme la PC7 est la dernière des convertases des proprotéines à avoir été découverte, plusieurs caractéristiques sur la fonction et l'importance de ce gène ne sont pas encore connues. Comme la furine, l'expression étendue de cette convertase suggère qu'elle pourrait avoir plusieurs rôles, soit distincts soit se chevauchant avec ceux des autres convertases. Pour expliquer la variabilité tissulaire dans l'expression de la PC7, des analogies ont été effectuées avec l'expression de la furine. L'ARNm de cette dernière est produit à partir d'un promoteur à boîte TATA lors d'une expression réglée dans certains tissus et utilise un promoteur à boîte GC pour l'expression basale dans la plupart des tissus. Ceci pourrait être le cas de la PC7. Par ailleurs, il existe une multitude d'isoformes d'ARNm de la PC7 dans la rate de rat. Ces isoformes proviendraient d'épissages alternatifs dont le mécanisme restent encore obscurs (Seidah *et al.*, 1996).

Pour mieux en comprendre l'évolution, pour expliquer son profil d'expression tissulaire et sa régulation, pour élucider les déterminants structuraux des épissages alternatifs, il est important de déterminer l'organisation du gène de la PC7 et d'analyser la séquence et la fonctionnalité de son promoteur. Dans ce contexte, nous avons élucidé la structure exon-intron de gène de la PC7 de souris. Cette étude représente notre contribution à une meilleure compréhension de l'évolution des convertases.

2. MATÉRIELS ET MÉTHODES

2.1. LIBRAIRIES GÉNOMIQUES

2.1.1. Les librairies génomiques

La première librairie, construite dans le vecteur λ DASH (Stratagene, Lajolla, CA) (Elgen *et al.*, 1991), nous a été fournie par le Dr. Janet Rossant du Mount-Sinai Hospital Research Institute et le Département de Génétique Médicale de l'Université de Toronto. Cette librairie contient de l'ADN génomique de rein de souris femelle de la souche 129Sv. L'ADN a été digéré partiellement avec l'enzyme *Sau3A* I; les fragments de 10 à 22 kb ont été sélectionnés sur un gradient de sucrose et clonés dans le site *BamH* I du phage λ DASH.

La deuxième librairie a été obtenue de la compagnie Clontech (Palo Alto, CA). Elle contenait des fragments *Sau3A* I de 8 à 22 kb provenant d'une digestion partielle de l'ADN génomique de foie de souris adulte de la souche Balb/c, purifiés sur gradient de sucrose et insérés dans le site *BamH* I du vecteur λ EMBL-3 SP6/T7 (Stratagene Lajolla, CA) (Karn *et al.*, 1980).

2.2. ISOLATION DES SÉQUENCES GÉNOMIQUES DE LA PC7

2.2.1. Etalement de la librairie

Des bactéries de la souche NM359 (F^- supF hsdR ($r_k^- m_k^+$) lacY (P2)) de *Escherichia coli* (*E. coli*) cultivées à 37 °C dans du milieu LB (Luria broth) en présence de maltose jusqu'à une densité optique de 0,6 à 600 nm. L'addition de maltose permet aux cellules de développer à leur surface des récepteurs pour le maltose. Ces récepteurs seront éventuellement utilisés par les phages comme site d'accès à la cellule.

Pour déterminer le titre de chaque librairie, des dilutions sérielles de 100 μ l allant de 10^{-2} à 10^{-8} sont préparées dans du diluant λ (1 M Tris-HCl pH 7,5/1 M $MgSO_4$). Ces dilutions sont ajoutées à 0,6 ml de bactéries NM359 (pour la densité des bactéries voir paragraphe précédent). Le mélange est incubé à 37

°C pendant une heure pour favoriser l'adhésion des phages aux récepteurs de maltose à la surface des bactéries. Le tout est ajouté à de l'agarose mou (agarose à 0,6 % dans du LB) à 50 °C; le mélange est étalé dans des boîtes de Pétri sur milieu solide LB/agar. Après solidification de l'agarose mou, les boîtes sont incubées dans une étuve à 37 °C pendant 16 à 18 h. Le nombre de plages de lyses est compté et le titre de la libairie déterminé en utilisant l'équation pfu (*plaque forming unit*)/ml = $pfu \times volume\ utilisé \times dilution$)

Une fois le titre déterminé, la libairie est étalée dans 30 boîtes de Pétri à une densité de 5×10^4 pfu par boîte. La croissance des bactériophages est surveillée pendant 8-11 h pour pouvoir l'arrêter juste avant la confluence des plages de lyse. Les boîtes sont alors placées à 4 °C pour inhiber la croissance et durcir l'agarose.

2.2.2. Transfert des plages de lyse sur membrane et hybridation

Pour transférer les plages de lyse sur membranes de nylon (GeneScreen Plus, Dupont, Boston, MA.), ces dernières sont posées par dessus le milieu agar pendant 1 minute (min) pour la première réplique et 5 min pour la seconde réplique, avec des marques de position identiques sur les deux pour le repérage ultérieur des plages positives. Après le transfert, les membranes sont placées, phages vers le haut, sur deux feuilles de Whatman imbibées de 0,5 M NaOH/1,5 M NaCl pendant 5 min pour lyser les phages et dénaturer son ADN. Elles sont ensuite transférées sur deux autres feuilles Whatman imbibées de 1 M Tris-HCl/1,5 M NaCl pH 8 durant 15 min pour neutraliser le NaOH, puis séchées à l'air. Les membranes sont ensuite lavées dans du 3X SET/0,5 % SDS (1X = 0,15 M NaCl + 0,3 M Tris-HCl, pH 7,5 + 0,2 mM EDTA) à 68 °C 2 fois, pendant 1 h chaque fois, pour enlever les débris bactériens. La préhybridation et l'hybridation avec une sonde radioactive d'ADNc de rPC7 sont effectuées comme décrit dans la section 2.4.

Nous avons aussi utilisé une seconde méthode plus rapide dans laquelle les membranes portant les phages transférés sont placées dans un autoclave

(Amresco 3043-S) et les phages sont dénaturés et fixés à haute température (120 °C) et sous pression (18 psi), pendant 2 min. Les membranes sont séchées à l'air avant l'hybridation (Genescreen plus, Dupont, Boston, MA).

2.2.3. Purification des phages positifs

Les plages positives avec la sonde d'ADNc rPC7 sont repiquées et transférées dans 100 µl de diluant λ avec 3 gouttes de chloroforme. La présence du chloroforme inhibe la croissance bactérienne. On laisse diffuser les phages dans le diluant λ pendant 18 h pour ensuite en faire des dilutions sérielles dans 100 µl de la même solution. Ces dilutions sont ensuite ajoutées aux bactéries NM359 et étalées comme auparavant. Après transfert sur membrane et hybridation avec la sonde rPC7, les plages positives bien isolées sont repiquées et recriblées à nouveau. La pureté des clones phagiques est atteinte quand toutes les plages sont positives à la sonde. L'éluat du clone phagique pur est titré et étalé à semi-confluence pour amplifier le phage ou pour purifier son ADN.

2.2.4. Préparation de l'ADN de phages positifs

Les plages de lyse sont recouvertes de diluant λ et entreposées à 4 °C pendant 16 à 24 h. L'éluat est recueilli dans un tube. Du chloroforme y est ajouté à une concentration finale de 0,1 %. Après un brassage de 2 min, on laisse sédimenter le chloroforme et l'on recueille la suspension de phages. La purification de l'ADN du phage est alors effectuée à l'aide de la trousse "*Wizard Lambda Preps*" de (Promega Madison, WI). Selon le protocole de cette trousse, on traite la suspension de phages avec une solution contenant de la DNAase I et de la RNAase A pour dégrader les acides nucléiques provenant de la lyse bactérienne. Les phages sont ensuite précipités avec du polyéthylène glycol et resuspendus dans un tampon de lyse. L'ADN phagique est attaché à la résine hydrophobique et rincé pour être ensuite élué de la résine avec de l'H₂O à 80 °C

Pour obtenir une première idée de la taille de l'insert d'ADN génomique de souris dans le vecteur bactériophagique, l'ADN est digéré avec une ou plusieurs

enzymes de restriction et les fragments digérés sont analysés par la technique de Southern.

2.3. ANALYSE PAR BUVARDAGE DE SOUTHERN

L'ADN est digéré à l'aide d'enzymes de restriction et les fragments digérés sont séparés par électrophorèse dans un gel d'agarose. L'ADN est "coloré" dans une solution de bromure d'éthidium et photographié sous U.V. Le gel est traité avec de l'HCl 0,25 M puis avec du NaOH 0,4 M. Les fragments d'ADN sont ensuite transférés par capillarité sur membrane de nylon (GeneScreen-Plus, NEN, Boston, MA) pendant 4 h. Une fois le transfert terminé, la membrane est préhybridée puis hybridée dans les conditions appropriées pour la sonde utilisée (voir ci-dessus, section 2.5).

2.4. SONDÉS MOLÉCULAIRES ET LEUR MARQUAGE RADIOACTIF

2.4.1. Les sondes d'ADNc

La première sonde utilisée pour le criblage de la librairie génomique de souris couvre toute la longueur (3 kb) de l'ADNc reconstruite à partir d'un fragment 5' (nucléotides 1 à 913) provenant des cellules PC12 (phéochromocytome de rat) et d'un fragment 3' (nucléotides 1188 à 3271) provenant de la rate de rat (Seidah *et al.*, 1996a). Cet ADNc était inséré dans le vecteur pBluescript(KS⁺) (Stratagene, Lajolla, CA) entre les sites *Hind* III (5') et *EcoR* I (3') pour former le plasmide recombinant pB-rPC7. Ces deux enzymes ont donc été utilisées séquentiellement pour exciser l'insert du vecteur (figure 4). Le fragment libéré a été séparé du vecteur par électrophorèse dans un gel d'agarose 0,6 %. La bande de gel contenant l'ADNc a été fondue dans du 6 M NaI et l'ADNc en a été purifié par affinité sur des billes de verre (trousse GeneClean II, BIO/CAN, Mississauga, Ont.). La seconde sonde de 425 pb allant du nucléotide 131 au nucléotide 556 de l'ADNc du rPC7 est un fragment d'une digestion à la *Pvu* II de pB-rPC7 purifié comme la première.

Marquage Les sondes ADNc ont été marquées avec du [³²P] selon la méthode d'amorçage multiple de Feinberg et Vogelstein (1983). Dans une réaction typique de marquage, 100 ng d'ADN sont mélangés à 125 ng d'une solution contenant des oligodésoxynucléotides (ODN) hexamériques aléatoires (0.1 µg/ml) dans un volume final de 4 µl; le mélange est porté à ébullition pendant 5 min afin de dénaturer l'ADN et de permettre l'attachement des amorces hexamériques; 10 µl du tampon de marquage 10X (2 M HEPES pH 6,6/2 M Tris-HCl/1 M MgCl₂/12,5 M 2-β-mercaptoéthanol/100 mM dTTP/100 mM dGTP/100 mM dATP), sont alors ajoutés, suivis de 100 µCi de [α-³²P]-dCTP (activité spécifique: 3 KCi/mmol), de 6 unités de fragment de Klenow de l'ADN polymérase I et de H₂O pour compléter la volume à 25 µl; le mélange réactionnel est incubée pendant 2 h à 21 °C; le volume ajusté à 70 µl avec du 1X STE (0,1 M NaCl/10 mM Tris/1 mM EDTA, pH 8), puis l'ADN marqué est purifié par filtration sur gel dans une colonne NucTrap (Stratagene) avec du 1X STE. L'éluat est ensuite porté à ébullition durant 10 min et refroidi dans la glace. Le taux d'incorporation est déterminé avant et après purification par comptage de 1 µl dans une solution de scintillant.

Conditions d'hybridation. La pré-hybridation et l'hybridation avec une sonde d'ADNc sont effectuées à 65 °C dans un tampon contenant 0,5 M Na-PO₄ pH 7,3/ 7% SDS et 1,5 mM EDTA et le lavage dans du 0,1 % SDS/0,1X SSC (20XSSC: 3 M NaCl/0,3 M Na-Citrate) aussi à 65 °C. Les membranes sont ensuite exposées à des films XAR-5 (Kodak, Rochester, NY) durant 24 à 48 h dans des cassettes contenant des écrans intensificateurs.

2.4.2. Les sondes oligonucléotidiques

Les ODNs utilisées dans cette étude ont été dérivés de la séquence de l'ADNc de souris et commandés soit du Sheldon Biotechnology Centre (McGill University, Montréal, Qué), soit de la compagnie Perkin-Elmer/Applied Biosystems (Mississauga, Ont.). Ces ODNs ont été utilisés comme soit amorces pour la PCR, le séquençage et réaction d'extension d'amorce, soit comme

sondes radioactives pour l'analyse de Southern des clones génomiques de la PC7.

Marquage Dans une réaction typique, 20 pmol d'ODN sont phosphorylés en 5' avec du ^{32}P par la polynucléotide kinase (10 unités) à partir du $[\gamma\text{-}^{32}\text{P}]\text{-dATP}$ (100 μCi , activité spécifique: 3 KCi/mmol) dans le tampon de l'enzyme dans un volume final de 15 μl , à 37 °C pendant 30 min. La sonde est ensuite purifiée sur colonne NucTrap pour enlever la radioactivité non incorporée.

Conditions d'hybridation L'hybridation des membranes avec une sonde ODN est effectuée dans un tampon composé de 50 % formamide, 5X SSC, 5X Denhardt (50X: 1 % Ficoll/ 1 % polyvinyl pyrrolidone/ 0.02 % BSA), 1 % SDS, 50 mM Na-PO₄ pH 6,6 et 50 $\mu\text{g}/\text{ml}$ d'ADN de sperme de saumon. La température d'hybridation (T_h) utilisée est généralement 15 °C en dessous de la température de fusion (T_f) de l'ODN calculée selon la formule $T = 4X(\text{nombre de G et C}) + 2X(\text{nombre de A et T})$. L'hybridation se fait durant 2 h. Les membranes sont ensuite lavées dans le 0,1X SSC/0,1 % SDS trois fois pendant 15 min. Elles sont enfin exposées sur film XAR-5 (Kodak) ou phosphorimager pendant 16 à 24 h.

2.5. RÉACTION DE POLYMÉRISATION EN CHAÎNE

Les amplifications ont été effectuées dans un thermo-cycleur (Perkin-Elmer Cetus *DNA Thermal Cycler*). Une réaction-type de 20 μl contient 100 ng d'ADN, 1 μM d'ODN sens et antisens, le tampon TNK50 (10 mM Tris-HCl, pH 8,2/5 mM NH_4Cl /1,5 mM MgCl_2 /50 mM KCl), 100 μM de dATP, dCTP, dGTP, dTTP et 5 unités de la polymérase d'ADN Amplitaq (Perkin-Elmer/Cetus). Dans chaque tube, une couche de 40 μl de n-dodecane est ajoutée pour éviter l'évaporation durant les réactions d'amplification. L'ADN a été soumis à 30 cycles d'amplification comprenant chacun une dénaturation de 1 min à 94 °C, un appariement de 1 min à 66 °C ou 68 °C, et une élongation de 1 min à 72 °C, suivis d'une extension finale de 7 min à 72 °C.

2.6. SOUS-CLONAGE DES FRAGMENTS D'ADN

Certains produits de PCR ont été sous-clonés dans le vecteur PCR2.1 linéarisé et portant un T saillant aux extrémités 3' (Invitrogen, San Diego, CA.). Ce vecteur a été choisi car il offre des bouts cohésifs qui favorisent la ligation des produits de la PCR dont les extrémités portent généralement un A saillant en 3'. Une réaction-type (10 μ l) contient 50 ng de plasmide, une quantité équimolaire d'insert, 1 unité de T4 DNA ligase, le tout dans le tampon 1X (250mM Tris-HCl (pH 7.6)/50mM MgCl₂/5mM ATP/5mM DTT/25% (w/v) polyéthylène glycol-800) fourni par le fabricant de l'enzyme. La réaction est incubée à 14 °C pendant 16 h. Un aliquot est ensuite utilisé pour transformer les bactéries *E. coli* DH5 α (F⁻ (ϕ 80d Δ (lacZ)M15) recA1 endA1 gyrA96 thi hsdR17 (r_k⁻ m_k⁺) supE44 relA1 deoR Δ (lacZYA-argF) U169) comme suit: à 100 μ l de bactéries compétentes (obtenues de GIBCO/BRL), on ajoute 10 μ l de la réaction de ligation; on incube le mélange 30 min à 4 °C pour permettre aux plasmides d'adhérer à la surface cellulaire; on immerge le tube dans un bain à 37 °C durant 45 secondes (sec) et on le place immédiatement à 4 °C pendant 2 min; on ajoute au mélange 900 μ l de SOC (20g/l tryptone/5g/l extrait de levure/10 mM NaCl/2,5 mM KCl/20 mM glucose/10 mM MgSO₄/10 mM MgCl₂) et on incube à 37 °C durant 40 min avec agitation. Le mélange est ensuite étalé dans des boîtes de pétri sur du milieu LB-agar contenant 100 μ g/ml d'hydrochlorate d'ampicilline (Amp100), pré-recouvert de X-Gal (5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside) qui, hydrolysée par la β -galactosidase encodée le gène *lacZ* par le plasmide, rend les colonies bleues, alors que les colonies restent blanches quand ce gène est interrompu par un insert.

2.7. PURIFICATION DE PLASMIDES

2.7.1. Mini-préparations

Pour une préparation rapide de plasmide, une colonie bactérienne est inoculée dans 1 ml de LB/Amp100 (Ampicilline conc._{finale} 100 μ g/ml) et cultivée à 37 °C

sous agitation pendant 16 h. Après une centrifugation de 30 sec à 14.000 x g, le milieu de culture est enlevé et les bactéries sont ensuite resuspendues dans 170 μ l de tampon de lyse (3 % sucrose/50 mM EDTA/0,5 % Triton X-100/10 mM Tris-HCl pH 8/10 μ g lysozyme). Le mélange est porté à ébullition pendant 40 sec et est placé sur glace immédiatement après. Cette étape est suivie d'une centrifugation de 10 min à 4 °C pour enlever tous les débris bactériens. Au surnageant, 1/2 volume de 7,5 M NH_4 acétate est ajouté suivi de 2 volumes d'éthanol à 95 % pré-refroidi à -20 °C. Le tube est placé à -80 °C durant 10 min puis centrifugé à 4 °C pendant 15 min. Le surnageant est ensuite enlevé; le culot est rincé avec de l'éthanol à 75 %, centrifugé pendant 5 min, séché et resuspendu dans 25 μ l de H_2O .

Dans une autre méthode de mini-préparation de plasmides, la trousse *Wizard Plasmid Purification System* (Promega, Madison, WI.) a été utilisée. Selon la méthode du manufacturier, le plasmide est purifié par affinité pour une résine hydrophobique à partir du surnageant provenant d'une lyse alcaline des bactéries.

2.7.2. Maxi-préparations

Les bactéries sontensemencées dans 250 ml de milieu TB (Terrific Broth)/Amp100 et cultivées à 37 °C sous agitation durant 12 à 15 h. La purification de plasmides est ensuite effectuée en utilisant deux techniques. La première utilise une trousse de Qiagen, (Chatsworth, CA.). Cette procédure est basée sur la méthode de la lyse alcaline de Birnboim and Doly (1979) et suivie d'une purification sur une colonne qui sélectionne les plasmides surenroulés (*The Qiagen plasmid purification procedure*). En quelques mots, après sédimentation et resuspension, les cellules sont lysées dans une solution de 200 mM NaOH/1 % SDS en présence de la Rnase A. Le SDS dénature les protéines cellulaires et le NaOH dénature l'ADN chromosomal et plasmidique. Le mélange est ensuite neutralisé avec de l'acétate de potassium. Les fortes concentrations de sel tendent à faire précipiter les protéines dénaturées, l'ADN chromosomal,

les débris cellulaires et le SDS tandis que l'ADN plasmidique se renature et reste en solution. Après centrifugation, le lysat est passé à travers une colonne d'affinité et l'ADN plasmidique attaché est ensuite élué à l'aide d'une solution de 1,25 M NaCl à pH 8,5.

La seconde méthode utilisée est celle du BiggerPrep obtenu de 5 PRIME->3 PRIME Inc. (Boulder, CO.). Cette méthode suit les mêmes étapes que la méthode Qiagen, avec comme différence que la résine est ajoutée au surnageant pour capter l'ADN avant de le passer sur colonne.

2.8. SÉQUENÇAGE D'ADN

Trois techniques de séquençage ont été utilisées pour séquencer les bandes obtenues par la PCR ou sous-clonées dans les vecteurs choisis. Toutes les trois sont basées sur le principe de terminaison de chaîne comme décrit initialement par Sanger *et al.* (1975). La première technique utilise la polymérase d'ADN T7 (Pharmacia, Baie d'Urfé, Qué) selon le protocole du manufacturier. Cette technique comprend quatre réactions séparées, chacune contenant une amorce oligonucléotidique, l'ADN à séquencer, les quatre désoxynucléotides triphosphates dont du α -³⁵S-dATP ou du α -³²P-dCTP et un des quatre didésoxy-nucléotides triphosphates (ddATP, ddCTP, ddGTP ou ddTTP) pour une terminaison aléatoire des chaînes nouvellement polymérisées. Chaque réaction est chargée dans un puits distinct sur un gel de polyacrylamide à l'urée et les produits sont séparés par électrophorèse à haut voltage. Si l'isotope utilisé est le [³²P], le gel est directement exposé sur film XAR-5. Si l'isotope est le [³⁵S], le gel est incubé durant 40 min dans une solution contenant 10 % d'acide acétique et 12 % méthanol, transféré sur papier Whatman et séché pendant 1 h à 80 °C. Le gel est ensuite exposé à -80 °C sur film Biomax (Kodak). La séquence nucléotidique est déduite en lisant la succession de fragments séparés selon leur taille à travers les quatre pistes.

La deuxième technique a été utilisée pour séquencer les fragments d'ADN qui ne pouvaient être clonés à cause des réarrangements majeurs qu'ils

subissaient dans la bactérie. Dans cette méthode basée sur la PCR avec la Thermo-Sequenase, des α -³³P-didésoxynucléotides (ddNTP) sont utilisés pour marquer les chaînes à la terminaison, permettant ainsi de ne visualiser que les chaînes interrompues et non celles qui ne le sont pas. Cette méthode diminue de façon considérable le bruit de fond. Le protocole suivi est celui fourni avec la trousse *Thermo Sequenase Radiolabeled Terminator Cycle Sequencing* (Amersham, Cleveland, OH). Le fractionnement sur gel et la lecture manuelle de la séquence est effectuée comme dans la première méthode.

La troisième technique utilise une amorce ODN marquée à la fluorescéine en 5'. Les fragments fluorescents dont la polymérisation a été interrompue par l'incorporation de didésoxynucléotides dans les quatre réactions sont fractionnés dans un gel de polyacrylamide dénaturant et sont détectés lors de leur sortie du gel par un jeu de laser grâce à l'appareil ALF (LKB.A.L.F. DNA Sequencer). Cet appareil enregistre automatiquement le didésoxynucléotide correspondant aux puits (A, G, C, T) dont ils sortent. Cette méthode tend à générer une séquence fiable plus longue allant jusqu'à 500 nucléotides (comparée aux deux premières qui en génèrent 200 à 300); mais elle a l'inconvénient de requérir beaucoup de plasmide (16-20 µg) de très grande pureté.

2.9. EXTENSION D'AMORCE

Deux pmoles de l'ODN anti-sens 7595 (GGAGATGTCACCTTCACTGTG), dérivé de l'exon 1 de l'ADNc de la mPC7, sont marqués en 5' avec du ³²P selon la méthode de phosphorylation avec la polynucléotide kinase décrite dans la section 2.5.2. Cinq microgrammes d'ARN totaux de testicule dans 17 µl d'eau traitée au diéthyl pyrocarbonate (DEPC) sont chauffés à 70 °C pendant 5 min pour défaire les structures secondaires. A cette solution sont ajoutés 10 µl de l'ODN marqué et 3 µl de tampon d'hybridation 10X (3 M NaCl/0,1 M Tris-HCl, pH7,5/10 mM EDTA). Le mélange est incubé pendant 20 min à 60°C pour permettre l'appariement de l'amorce marquée. Sont ensuite ajoutés, 1 µl de RNA

guard (38,6 unités), 5 μ l de 1 mg/ml BSA, 10,5 μ l d'eau traitée au DEPC et 2 μ l (400 unités) de la transcriptase inverse Superscript II (Gibco BRL). Le mélange est incubé à 42 °C durant 60 min, puis 30 min à 50°C pour permettre la synthèse de l'ADNc. Après y avoir ajouté 1 μ l de RNase H (2,1 unités), il est placé à 37 °C durant 30 min pour dégrader l'ARN dans les hybrides ARN-ADNc. Il est ensuite déshydraté et désalé par extraction au n-butanol. L'ADNc du culot est dissout dans 5 μ l de solution stop (95 % formamide/20 mM EDTA/0,05 % bleu de bromophénol/0,05 % xylène cyanol FF), dénaturé à 80 °C pendant 3 min avant d'être analysé par électrophorèse dans un gel de séquençage et détecté par autoradiographie (Promega, TB 103: 2/91).

2.10. PROTOCOLES STANDARDS DE LABORATOIRE

Les protocoles de biologie moléculaire, tel l'étalement de la librairie, la préparation de gel d'agarose et acrylamide ainsi que la préparation de solution stock , ont été accomplis comme décrit dans le manuel de *Current Protocols in Molecular Biology* 1994 (Ausubel *et al.*, 1994).

3. RÉSULTATS

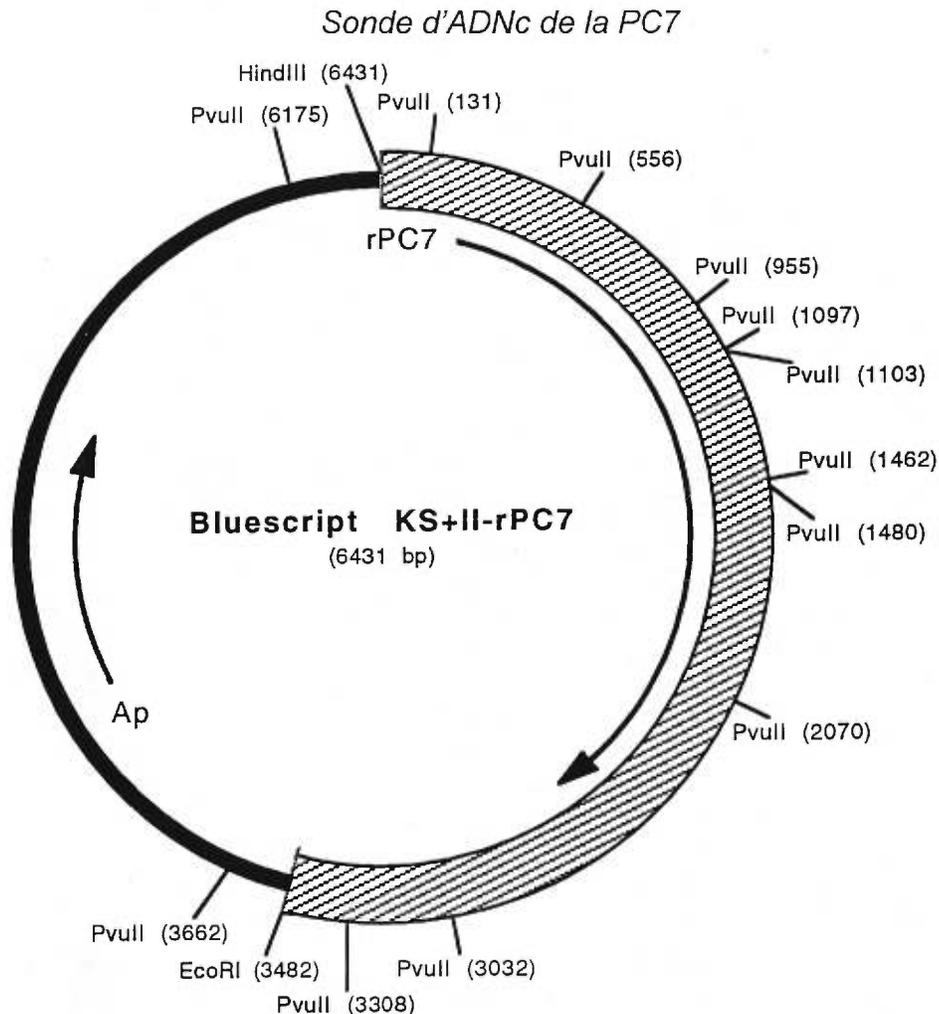
3.1. CONTRÔLE DE LA SONDÉ

La qualité de la sonde qui a été générée à partir de l'ADNc de la PC7 de rat (figure 4) (Seidah *et al.*, 1996) et qui a été utilisée pour le criblage des bibliothèques génomiques a été vérifiée dans une analyse préliminaire de buvardage de Southern de l'ADN génomique purifié du foie de souris digérés avec les enzymes de restriction *BamH* I, *Bgl* II, *EcoR* I, *EcoR* V, *Hind* III, *Hinf* I, *Kpn* I, *Msp* I, *Pst* I et *Sst* I. Les résultats de cette analyse sont présentés dans la figure 5. Ils montrent des bandes discrètes pour tous les fragments digérés exceptés ceux de *Hinf* I. Dans ce cas, les fragments qui auraient pu hybrider avec la sonde étaient de taille trop petite étant donné la haute fréquence de coupure de l'enzyme et ont migré hors du gel. Le fait que la sonde utilisée n'a donné aucune traînée indiquait qu'elle ne reconnaissait pas de séquences répétitives et qu'elle était donc adéquate pour le criblage de la banque d'ADN génomique.

3.2. CRIBLAGE DE LA BIBLIOTHÈQUE GÉNOMIQUE DE SOURIS 129SV

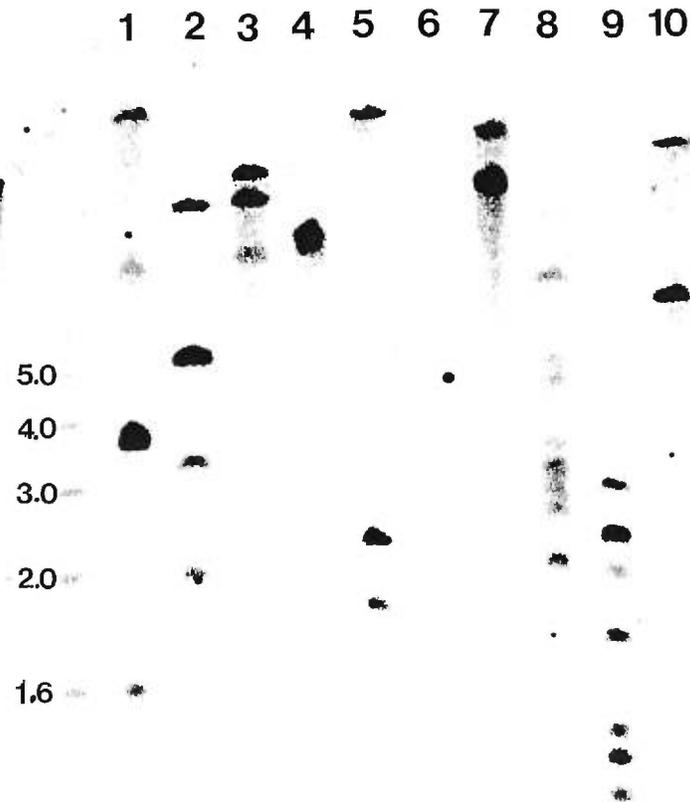
Nous avons criblé un total de $1,5 \times 10^6$ pfu de la bibliothèque de souris 129SV dans λ DASH. La taille du génome murin étant de 3×10^9 pb et celle des inserts de la bibliothèque étant de 1×10^4 pb en moyenne, le nombre de clones criblés représente environ l'équivalent de 5 génomes de souris. De ce criblage, nous avons obtenu un seul clone positif (λ DASH-9). Une fois le clone purifié, nous avons cherché à savoir quels exons du gène de la PC7 il contenait. Nous avons analysé l'ADN des clones par buvardage de Southern avec des sondes d'ADNc et d'ODN. L'ADNc de la rPC7 a été coupé en deux fragments 5' (nucléotides 1 à 913) et 3' (nucléotides 1188 à 3271). La sonde d'ADNc 3' a donné un signal positif. L'hybridation avec les sondes ODN dérivés de différentes régions de l'ADNc de rat a révélé que le clone λ DASH-9 contenait les exons de la région 3' du gène de la PC7, en aval du site catalytique (figure 1).

Figure 4



Représentation schématique du plasmide pBluescript II (KS⁺) contenant l'ADNc de rat. Le bout 5' de l'ADNc se trouve du côté de l'enzyme de restriction *Hind III* et le bout 3' se trouve du côté de l'enzyme de restriction *EcoR I*. Le fragment 5' utilisé comme sonde est généré par l'enzyme de restriction *Pvu II* et se trouve entre les sites *Pvu II* (131) et *Pvu II* (556).

Figure 5

Vérification de la Qualité de la Sonde

Buvarbage de *Southern* d'ADN génomique de souris digéré avec 1-*Bam*H I, 2-*Bgl* II, 3-*Eco*R I, 4-*Eco*R V, 5-*Hind* III, 6-*Hinf* I, 7-*Kpn* I, 8-*Msp* I, 9-*Pst* I, et 10-*Sst* I. On a utilisé pour l'hybridation un ADNc de PC7 de rat de 3kb pour détecter les fragments d'ADN génomique contenant de la PC7. Les digestions sont indiquées par les numéros 1-10. Le marqueur de poids moléculaire marqués au [³²P] se trouve à l'extrême gauche et représente le marqueur de 1kb de Gibco.

3.3. CRIBLAGE DE LA LIBRAIRIE GÉNOMIQUE DE SOURIS BALB/C

Un second criblage de la librairie λ DASH-129 avec le fragment 5' de l'ADNc de la PC7 n'ayant pas donné de clone positif, nous avons décidé de cribler une autre librairie contenant des fragments génomiques de souris Balb/c clonés dans λ EMBL3 SP6/T7. Du criblage initial de cette librairie avec comme sonde un fragment *Pvu* II de 425 pb, situé dans la région 5' de l'ADNc de la PC7 de rat (voir figure 4), nous avons isolé 5 clones positifs (λ EMBL3-6, 19, 22, 23 et 26s). Ces clones ont été analysés par buvardage de Southern avec plusieurs ODNs correspondant à différentes régions l'ADNc de la rPC7. Le clone λ EMBL3-6 a hybridé avec le plus grand nombre d'ODNs. Il contenait les séquences retrouvées dans les quatre autres clones, C'est sur ce clone, contenant le plus grand fragment du gène PC7, que nous avons donc concentré nos efforts.

Afin d'étudier les régions 5' et 3' du gène de la PC7 à partir de la même librairie génomique de souris, nous avons de nouveau criblé la librairie génomique Balb/c avec cette fois, une sonde 3' et nous avons obtenu un clone positif que nous avons dénommé λ EMBL3-5

3.4. SOUS-CLONAGE DES FRAGMENTS D'ADN HYBRIDANT AVEC LES SONDES 5' ET 3' DU GÈNE DE LA PC7

Le clone λ DASH-9 a été digéré avec *Bam*H I pour libérer le fragment inséré. De cette digestion, trois fragments de 2,9, 4,5 et 10 kb ont été obtenus. Ils étaient tous positifs par analyse par buvardage de Southern avec la sonde d'ADNc complet de la rPC7. Nous avons réussi à sous-cloner les fragments de 2,9 et 4,5 kb dans le site *Bam*H I du vecteur pBluescript, mais pas celui de 10 kb, probablement à cause de sa grande taille.

Après séquençage à partir des extrémités du site de clonage, seul l'insert de 4,5 kb a donné une séquence homologue à l'ADNc de la rPC7. Les séquences dérivées de l'insert de 2,9 kb ne montraient aucune homologie à la PC7 lors

d'une comparaison avec ce dernier et représenteraient donc des séquences introniques qui ne se trouvent pas dans l'ADNc.

Ces premiers résultats de séquençage nous ayant persuadés que nos clones contenaient bel et bien des fragments du gène de la mPC7, nous avons poursuivi la cartographie des exons et introns en utilisant la méthode de la PCR pour l'amplification aussi bien que le séquençage des différents domaines.

3.5. CARTOGRAPHIE DU GÈNE DE LA MPC7 PAR LA PCR

En utilisant un programme ALIGN, nous avons aligné les séquences des acides aminés de toutes les PCs (figure 10) et nous avons placé de façon approximative les positions des introns dans l'ADNc de la mPC7 par analogie avec les positions conservées des introns dans les gènes de la furine, de la PC1, de la PC2 et de la PC4. Douze introns ont été ainsi positionnés. Nous avons fait synthétiser des ODNs de part et d'autre des sites potentiels des introns, sur la base de la séquence d'ADNc de la rPC7 d'abord, puis sur celle de l'ADNc de la mPC7 quand celle-ci a été publiée (Constam *et al.*, 1996). Les réactions de PCR ont été effectuées en double, sur l'ADN cloné provenant des phages positifs d'une part, et, comme contrôle, sur l'ADN génomique de thymus de souris. En l'absence d'intron au site présumé, la taille du produit de la PCR devait correspondre à la distance en nucléotides entre les extrémités des amorces utilisés dans la séquence l'ADNc de la mPC7. Elle devrait être supérieure à cette dernière s'il y avait effectivement un intron au ou près du site présumé. Les séquences des amorces, leurs positions dans la séquence de la mPC7, leur température d'appariement lors de la PCR et la taille des produits amplifiés sont présentés dans le tableau 4. Elles sont aussi illustrées dans la figure 6. Avant de séquencer les produits de la PCR, nous avons confirmé leur relation avec la mPC7 par buvardage de Southern avec l'ADNc complet de la rPC7 comme sonde, dans des conditions d'hybridation sévères qui ne permettaient que l'hybridation longues régions d'homologie et excluait toute positivité due aux amorces utilisées au cours de l'amplification. Nous avons

aussi utilisé l'ADN génomique de souris comme contrôle positif à titre de comparaison avec les fragments amplifiés à partir de nos phages (figure 7).

Tableau IV

Séquences et Positions des ODNs Utilisés pour la Cartographie

Numéro de l'oligodésoxy nucléotide sens (s) et anti-sens (a.s)	Position dans la séquence nucléotidique mPC7	Séquence nucléotidique des ODNs	Température d'appariement	Taille attendue en l'absence d'intron selon ADNc en (pb)	Taille approximative des produits amplifiés sur gel en (kb)
7528	170-191	TGCTGTTCTGATGCCCGAAAGG	68	436	0.4
7565	589-609	GGGCAAGCGCAGTATCCACT	66		
7534	701-722	GAGCAGACCGTCCAGGACATT	68	132	0.6
7569	818-838	AGATCCCATGCCCCACCCTGA	68		
7566	79-100	CACAGTGAAGGTGACATCTCC	64	245	0.9
7563	302-323	TGAGGAGGGACGGATGCCTT	66		
7570	560-581	CTGGCACTCGGAGCAAACGCT	68	190	0.9
7567	729-749	GTGGTGGATGACGGAGTGAA	66		
7576	1080-1101	AGCTCGAAAGGCTGCCTTAC	66	181	0.8
7573	1241-1262	TGAGGAGGGACGGATGCCTTTTT	66		
7580	1402-1423	CTCCTCTGGCAGCTGGAATGA	66	165	0.4
7577	1546-1566	GCCACAGCCATCAGCATGGTT	66		
7582	1633-1654	GCCTCCTATGTCAGCCCCATG	68	144	0.1
7579	1755-1775	CATGTGGCAGTGACAGTCTCC	66		
7586	1980-2001	ATGGGCATCCTCCAGCAGTGG	68	574	1
7519	2484-2505	CTACGAGTCCTCTTACCAGGAC	68		
7594	818-840	AGATCCCATGCCCCACCCTGA	66	242	2
7571	1039-1059	GGGGCCCAGATGATGATGGGA	68		
7596	1242-1263	TGAGGAGGGACGGATGCCTTT	66	156	3
7575	1375-1397	AGGCCACACAGGAACCTCAGC	68		
7598	2327-2349	ACAGCACTAGAATCAATGCCA	60	1094	1
7597	3336-3360	TTTTTTTCTGCACAGCTTTCAAA	64		
7600	1894-2017	TGGCAGCTTAGAACTGAACTGTT	68	124	0.9
7581	1896-1916	TTTTCCACTGTGCGGTGCTGG	66		
7595	79-100	CAAAGCAACGATTTTGTGAA	64	-	Extension d'amorce

Liste des oligodésoxynucléotides sens (nombre pair) et anti-sens (nombre impair) utilisés pour la cartographie du gène de la mPC7. La position sur l'ADNc de souris, leur séquence nucléotidique, leur température d'appariement et la taille des fragments amplifiés sont aussi montrées.

Figure 6
Séquence de la mPC7

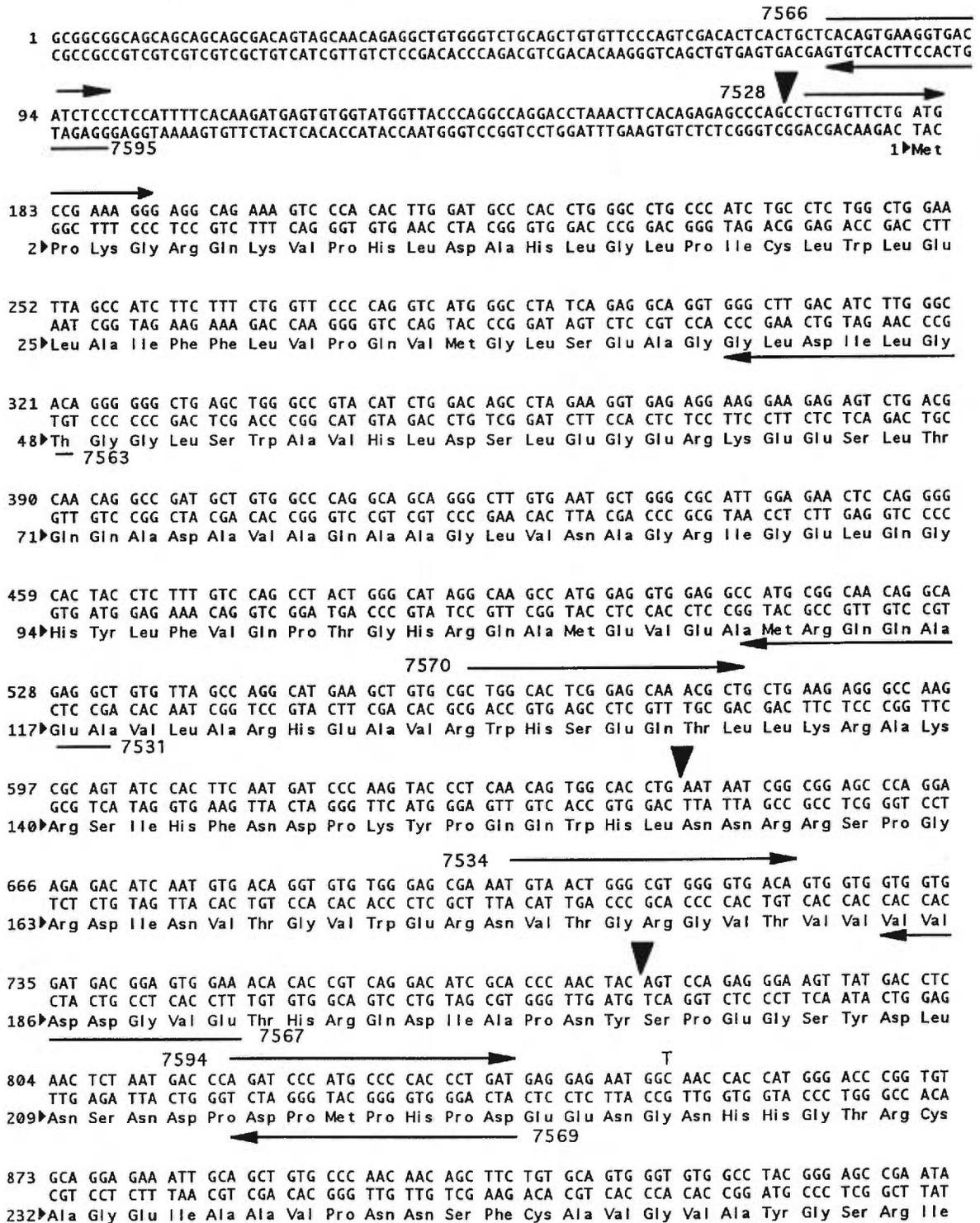


Figure 6 (suite)

T
▼

1839 TTG ATC GGC GCG CCC CGC AGC ATG GAC TCG GAT CCT AAT GGC TTC AAT GCT TGG ACC TTT TCC ACT GTG
AAC TAG CCG CGC GGG GCG TCG TAC CTG AGC CTA GGA TTA CCG AAG TTA CGA ACC TGG AAA AGG TGA CAC
554▶Leu Ile Gly Ala Pro Arg Ser Met Asp Ser Asp Pro Asn Gly Phe Asn Ala Trp Thr Phe Ser Thr Val
Leu

C

1908 CGG TGC TGG GGG GAA AGA GCA AGA GGC GTC TAC AGA CTG GTT ATC AGG GAT GTA GGA GAT GAG CCG CTC
GCC ACG ACC CCC CTT TCT CGT TCT CCG CAG ATG TCT GAC CAA TAG TCC CTA CAT CCT CTA CTC GGC GAG
577▶Arg Cys Trp Gly Glu Arg Ala Arg Gly Val Tyr Arg Leu Val Ile Arg Asp Val Gly Asp Glu Pro Leu
7581

7586 →

1977 CAG ATG GGC ATC CTC CAG CAG TGG CAG CTG ACC CTG TAT GGC TCC ATG TGG AGT CCA GTA GAC ATC AAG
GTC TAC CCG TAG GAG GTC GTC ACC GTC GAC TGG GAC ATA CCG AGG TAC ACC TCA GGT CAT CTG TAG TTC
600▶Gln Met Gly Ile Leu Gln Gln Trp Gln Leu Thr Leu Tyr Gly Ser Met Trp Ser Pro Val Asp Ile Lys

▼

2046 GAC AGA CAA AGT CTC TTA GAA AGT GCT ATG AGT GGA AAA TAC CTG CAT GAT GGC TTC ACC CTG CCT TGC
CTG TCT GTT TCA GAG AAT CTT TCA CGA TAC TCA CCT TTT ATG GAC GTA CTA CCG AAG TGG GAC GGA ACG
623▶Asp Arg Gln Ser Leu Leu Glu Ser Ala Met Ser Gly Lys Tyr Leu His Asp Gly Phe Thr Leu Pro Cys

2115 CCA CCT GGG CTG AAA ATC CCT GAG GAG GAT GGC TAC ACC ATT ACC CCT AAC ACA CTC AAG ACC CTT GTG
GGT GGA CCC GAC TTT TAG GGA CTC CTC CTA CCG ATG TGG TAA TGG GGA TTG TGT GAG TTC TGG GAA CAC
646▶Pro Pro Gly Leu Lys Ile Pro Glu Glu Asp Gly Tyr Thr Ile Thr Pro Asn Thr Leu Lys Thr Leu Val
7605

2184 CTG GTA GGC TGC TTC AGT GTC TTC TGG ACT ATT TAC TAC ATG CTA GAA GTG TGC CTG AGC CAG AGG AAC
GAC CAT CCG ACG AAG TCA CAG AAG ACC TGA TAA ATG ATG TAC GAT CTT CAC ACG GAC TCG GTC TCC TTG
669▶Leu Val Gly Cys Phe Ser Val Phe Trp Thr Ile Tyr Tyr Met Leu Glu Val Cys Leu Ser Gln Arg Asn

2253 AAG GCT TCC ACC CAT GGC TGC AGG AAG GGA TGC TGC CCC TGG GCC CCA CGA AGG CAA AAC TCC AAG GAT
TTC CGA AGG TGG GTA CCG ACG TCC TTC CCT ACG ACG GGG ACC CGG GGT TCC GTT TTG AGG TTC CTA
692▶Lys Ala Ser Thr His Gly Cys Arg Lys Gly Cys Cys Pro Trp Ala Pro Arg Arg Gln Asn Ser Lys Asp
7598 →

2322 GCA GGG ACA GCA CTA GAA TCA ATG CCA CTG TGC AGC AGC AAG GAC CTG GAT GGA GTG GAT TCA GAG CAT
CGT CCC TGT CGT GAT CTT AGT TAC GGT GAC ACG TCG TCG TTC CTG GAC CTA CCT CAC CTA AGT CTC GTA
715▶Ala Gly Thr Ala Leu Glu Ser Met Pro Leu Cys Ser Ser Lys Asp Leu Asp Gly Val Asp Ser Glu His

2391 GGG GAC TGC ACA ACT GCA TCT AGC TTC CTG GCC CCA GAG CTG GAT TGT CCT CCA CAC CAA CCC CCA GAC
CCC CTG ACG TGT TGA CGT AGA TCG AAG GAC CGG GGT CTC GAC CTA ACA GGA GGT GTG GTT GGG GGT CTG
738▶Gly Asp Cys Thr Thr Ala Ser Ser Phe Leu Ala Pro Glu Leu Asp Cys Pro Pro His Gln Pro Pro Asp

2460 CTG CTG CAG GGG AAG AGC GGG CAG ATT TGC TGA CCCAGAGCCCAGCTTCCCCATGTACAACAGGCCACTTCTAAACTCGG
GAC GAC GTC CCC TTC TCG CCC GTC TAA ACG ACT GGGGTCTCGGGTCGAAGGGGGTACATGTTGTCCGTGAAGGATTTGAGCC
761▶Leu Leu Gln Gly Lys Ser Gly Gln Ile Cys

2542 TTATGAGGCTTTCAATCGTGGAAGCTCAGGAGGGAATGGTCTGATAGTGACATCTGGGACCCAGGCCTCTGAAGCATTTCGAGCCTCTCA
AATACTCCGAAAGTTAGCACCTTCGAGTCTCCCTTACCAGGACTATCACTGTAGACCCTGGGGTCCGGAGACTTCGTAACAGTTCGGAGGAGT

2635 GGAGGGTGAGGGCCATCTTAATGCAGTGGTAAGGGACTGGTGTATCATGGCTTCTCTAGCTGAACCTTGGCTGAGAACCCTTGACAAGGGATTGG
CCTCCCACTCCCGGTAGAATTACGTACCATTCCCTGACCACTAGTACCGAAGAGATCGACTTGAACCGACTCTTGAAACTGTTCCCTAAC

2728 CACCAAGACCAGGCGAGCCCTTGATTTCTCGCTCTCTGGAGCTAGTCACTGTCTAGACCACCGAGCCTCTACTCATTTCATGAGTTAGTTAT
GTGGTTCTGGTCCGTCCGGAACTAAAGGAGCGAGAGACCTCGATCAGTACAGGATCTGGTGGTCTGGAGATGAGTAAGTACTCAATCAATA

Figure 6 (suite)

2821 GCAGCAAGTCCCTATATACCTAGAAATGCTCTTGTGAGAAGTGTCACTGCCAAGGGACAGAGAGAGGCTGCTAAGAGCCTGGACATGCCTAC
 CGTCGTTCCAGGGATATATGGATCTTTACGAGAACAGTCTTACAGTGACGGTCCCTGGTCTCTCCGACGATTCTCGGACCTGTACGGATG

2914 CCTGAAAGCAGCTGCCTTCATTACCCCTTTCTGTGCGCTGTTCTGAAGCCAACTAAACCTGTCCACATCTTCACTACCATTGCGCCTTCTG
 GGACTTTCGTGACGGAAGTAATGGGAAAAGACACGCGGACAAGACTTCGGTTGAATTTGGACAGGTGTAGAAGTGATGGTAAGCGGGAAAGAC

3007 CTGAAGGAATGGTGACGTGTTCCAGGAGAACTGGGAGGCAACAGGATCCCCAGACCCTTGCTGCCTCATTGGGAAGTTGGCCAGCTGGACT
 GATTCCTTACCCTGCACAAGGGTCTCTTGACCCCTCGTTGTCTAAGGGGTCTGGGAACGACGGAGTAACCCCTCAACCGGTCGACCTGA

3100 TCAGAGAGCTAAGACACAGCATTTTCAAAGATAGCCTGGGAGCACTGTCTTGACCACCTGTCTTCTCTGCACCGTGCTCAGGGAAATGGCC
 AGTCTCTCGATTCTGTGTCGTAAGTTTCTATCGGACCCCTCGTGACAGAGACTGGTGGACAGAAGGAGACGTGGCACGAGTCCCTTTACCGG

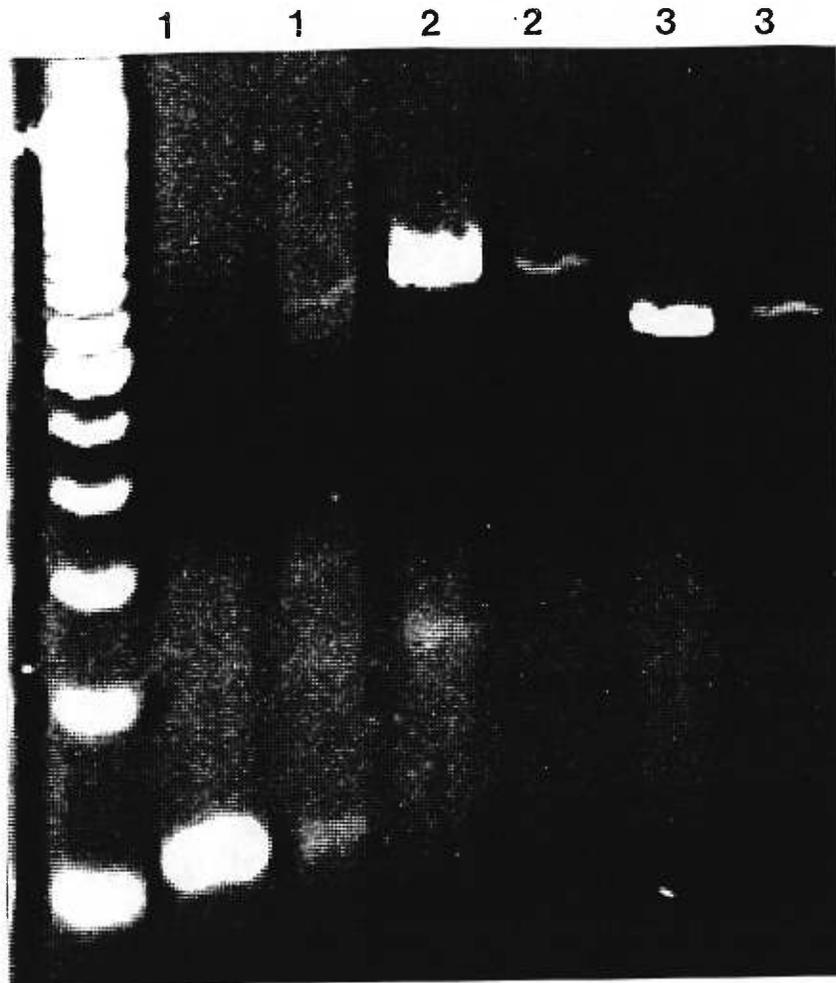
3193 TTCATTCCAGAGGCCAGCTGTCTGCCTGACTTTTCTCCGTTCTTGGCCTTCTCCCTTCCCATCTTCTGAGCTAATTGTTAATATGAATTTT
 AAGTAAGGTCTCCGGTCGACAGACGGACTGAAAAGGAGGCAAGAACCGGAAGAGGGGAAGGGTAGAAGACTCGATTAACAATTATACTTAAAA

3286 TTAATGCTTAAGATTTGATTTTACTTTTCAAAGCAACGATTTTGTGAAATTTTTCTGCACAGCTTCCAAAATAAAAAGCAGAAGTATAA
 AATTACGAATCTAAACTAAAATGAAAAGTTTCGTTGCTAAAACAACCTAAAAAAGACGTGTGAAAGGTTTTATTTTTCGTCTTCATATT
 ← 7597

3379 AAACCCAAAAA
 TTTGGGTTTTT

Les flèches au dessus de la séquence représentent les ODN sens et celles au dessous représentent les ODN anti-sens. Les triangles noirs représentent la position des introns telle que déduite après séquençage et les nucléotides superposés à la séquence nucléotidique représentent les polymorphismes exoniques obtenus lors du séquençage. L'ODN 7595 a été utilisé pour la réaction d'extension d'amorce.

Figure 7
Analyse des Produits de la PCR



Photographie de gel d'agarose montrant, à titre d'exemple, la co-migration des ADNs phagique et génomique respectivement, amplifiés par la PCR. L'ADN génomique a été utilisé comme contrôle positif et correspond à la bande de faible intensité à la droite de l'ADN phagique, de forte intensité. Les bandes numéros 1 représentent une région de l'exon 2, les bandes numéros 2 représentent l'intron 2 et les bandes numéros 3 représentent l'intron 3.

Le séquençage des produits de la PCR nous a permis de positionner les introns de la mPC7 (figure 6). Nous avons constaté que certains introns se trouvaient plus ou moins aux sites prédits par le programme ALIGN. D'autres introns, part contre, s'en écartaient de façon frappante (figure 10). Ainsi, le premier intron qui était présumé se trouver en position 440 sur l'ADNc de la mPCs se positionne en fait à la position 168 et donc avant l'ATG initiateur. Les introns 2, 5 et 9 prédits en positions respectives 556, 857 et 1667 de l'ADNc n'existent pas. L'intron 6 prédit en position 1181 se situe légèrement en aval en position 1232. De même l'intron 8 est en position 1499 au lieu de la position 1445 prédite.

La comparaison des séquences exoniques que nous avons obtenues à celle de la séquence de l'ADNc de la mPCs déjà publiée (Constam *et al.*, 1996), montrent des homologies de 98,2 % et 100 % (figure 6). Ces différences seront analysées plus en détails dans la partie Discussion.

3.6. STRUCTURE DU GÈNE DE LA MPC7

La combinaison du criblage des deux bibliothèques et les réactions PCR effectuées sur l'ADN phagique, nous ont permis de définir la carte physique, la taille des introns et les jonctions introns/exons du gène de la mPC7. Les résultats finaux qui sont résumés dans le tableau 5 montrent que le gène de la mPC7 est composé de 10 exons séparés par 9 introns. La longueur des introns varie de 0,44 à 2,74 kb. Tous les introns commencent et finissent avec les dinucléotides attendus, GT en 5' et AG en 3' (Breathnach *et al.*, 1978).

L'ATG initiateur se situe dans l'exon 2 ainsi que la totalité du prosegment qui se termine avec la séquence Arg-Ala-Lys-Arg (figure 8). Les codons des résidus importants de la région catalytique sont situés sur des exons différents: Asp dans l'exon 3, His dans l'exon 4, Asn dans l'exon 5, et Ser dans l'exon 7. Les quatre sites potentiels pour la N-glycosylation se trouvent sur les exons 3, 5 et 8. Enfin, la séquence conservée RRGSL (Seidah *et al.*, 1996), se situe dans l'exon 8 et la région transmembranaire dans l'exon 10.

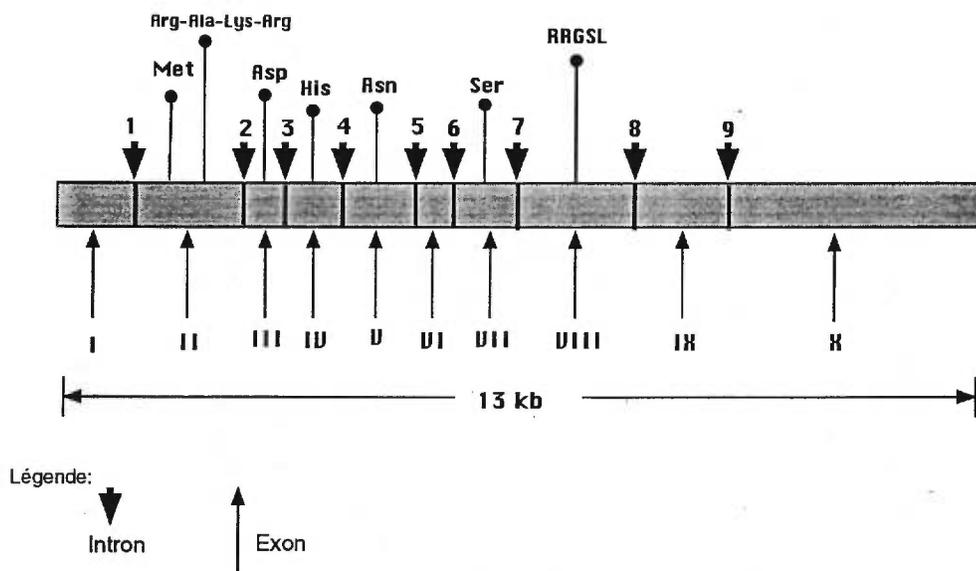
Tableau V

Tailles et Jonctions des Exons et des Introns dans le Gène de la mPC7

Exon				Intron			
no	5'	pb	3'	5'	kpb	3	no
1	CAGCGC	155	GCCCAG	gtggga	0,6	ttgcag	I
2	CCTGCT	477	CACCTG	gtaagt	0,6	ctacag	II
3	AATAAT	135	AACTAC	gtgagt	0,5	ttcag	III
4	AGTCCA	167	TAGCAG	gtaact	1,6	acctag	IV
5	GTATCC	285	CCATAG	gtagtg	0,7	tgccag	V
6	GAGCTG	101	AGCATT	gtgagc	2,7	ttgcag	VI
7	GTGACC	168	ATCCAG	gtgaga	0,4	ttcag	VII
8	TATGAA	367	GGACTC	gtatgt	0,7	aactag	VIII
9	TGGATC	190	ACAAAG	gtaggt	0,8	cctgag	IX
10	TCTCTT	1334	CCCAAA				

Jonctions intron/exon du gène de la PC7 chez la souris. Les séquences exoniques 5' et 3' représentent les nucléotides qui entourent l'intron. Les séquences introniques 5' et 3' représentent le début et la fin de l'intron respectivement. La taille des exons et des introns est aussi indiquée.

Figure 8

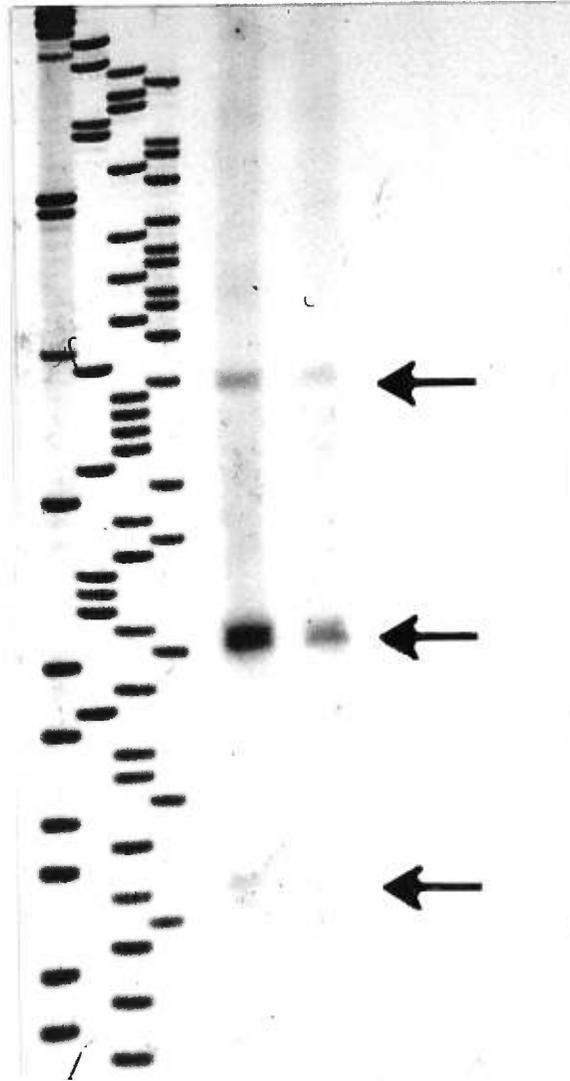
Répartitions des Résidus Fonctionnels dans les Exons

Représentation schématique de la PC7 de souris et l'emplacement des résidus fonctionnels. La Met est encodée par l'ATG initiateur, l'Asp, l'Asn, His et Ser représentent les a.a. du site catalytique et la séquence RRGSL représente la séquence conservée chez les PCs des mammifères.

3.6. Réaction d'extension d'ODN

Pour définir les différents sites d'initiation de la transcription du gène de la mPC7, des réactions d'extension d'amorce ont été faites sur des ARNm totaux isolés de testicules de souris, en utilisant un ODN situé à 80 pb en aval de l'ATG initiateur dans l'exon 1. Les résultats de la réaction d'extension d'amorce sont présentés dans la figure 9. Trois sites d'initiation de la transcription y sont révélés. Le site majeur se trouve à 166 pb de l'ATG initiateur, et les deux sites mineurs à 181 et 155 pb de cet ATG.

Figure 9

Localisation des Sites d'Initiation de la Transcription

Autoradiographie de la réaction d'extension d'amorce utilisant l'ODN 7595. L'échelle de droite correspond au promoteur de la PC2 et est utilisée comme marqueur de poids moléculaire. Le site d'initiation majeur se trouve à la position 166, les deux autres sites mineurs observés se trouvent en amont et en aval du site majeur, aux positions 181 et 155 respectivement.

4. DISCUSSION

La caractérisation des activités enzymatiques des convertases des proprotéines et l'analyse de leur distributions tissulaire et cellulaire permettent d'associer des substrats spécifiques à ces enzymes et de se faire une idée de leurs rôles biologiques. L'étude comparative des gènes des convertases permet de dériver leur relation évolutive et de déterminer les éléments structuraux qui contribuent à la spécificité tissulaire de leur expression et à leur régulation. C'est dans cette perspective que nous avons entrepris la caractérisation du gène de la PC7.

Evaluation à posteriori des banques génomiques

À partir du criblage de la librairie génomique de souris λ DASH129, nous avons obtenu un clone positif pour la région 3' du gène de la PC7 et aucun pour la région 5'. Une hypothèse que nous avons émise pour expliquer ce résultat est que le gène de la mPC7, comme est le cas dans la plupart des gènes déjà caractérisés, aurait un dernier exon relativement long (400 à 800 pb) en 3' qui favorise l'hybridation, alors que les exons intermédiaires qui sont de taille plus courte n'ait pas hybridé avec notre sonde. Comme nous avons utilisé une sonde d'ADNc complet de la rPC7, il était fort possible que, dans nos conditions d'hybridation, cette sonde n'ait pas d'hybridé avec ces petits exons. De plus, même si l'hybridation avait eu lieu, les appariements auraient été déstabilisés par les boucles dues aux introns et seraient plus aisément défaits lors des lavages des membranes. Du côté 3' par contre, le long exon aurait permis la formation d'hybrides plus stables. Cette présomption a été confirmée par la cartographie finale, puisque le dernier exon du gène de la mPC7 est de 1395 pb.

Pour obtenir une sonde continue en 5', nous avons décidé d'utiliser une sonde de 360 pb obtenue par la PCR sur de l'ADN génomique de souris et contenant l'intron 1. La vérification de cette sonde par analyse de buvardage de Southern sur de l'ADN génomique de souris avait montré que cette sonde contenait des séquences répétitives et n'était donc pas indiquée pour le criblage.

Comme alternative, nous avons choisi de produire une sonde en 5' à partir de l'ADNc cloné de la rPC7, en nous basant sur le positionnement hypothétique des introns dans les alignements des séquences des PCs dont les gènes sont déjà caractérisés. Nous avons ainsi purifié un fragment *Pvu* II de 425 pb entièrement contenu dans l'exon 1. Comme sonde, ce fragment a donné des bandes discrètes dans l'analyse d'ADN génomique par buvardage de Southern. Il a été donc utilisé pour un nouveau criblage de la librairie λ DASH-129. Aucun clone positif n'a été obtenu de ce criblage, suggérant que cette librairie n'était pas représentative du génome entier de la souris.

Nous avons donc recouru à une seconde librairie d'ADN génomique de souris Balb/c. Celle-ci s'est avérée plus utile car dès le premier criblage avec le fragment *Pvu* II décrit ci-haut, nous avons obtenu cinq clones positifs. L'utilisation des ODNs comme sondes dans l'analyse de l'ADN phagique par buvardage de Southern nous a permis d'évaluer rapidement l'étendue de ces clones en 5' et 3'. Le clone λ EMBL3-6 recouvrait les exons 1 à 4 et se prolongeait en 5', alors que, le clone λ DASH-9 ne couvrait qu'une partie du 3' du gène de la mPC7. De part la taille de l'insert, le clone λ Dash 9 devait contenir plus que les exons 10 et 11, mais une cartographie plus poussée avec des sondes ODNs de part et d'autre de ces exons indiquait que ce clone serait hybride et qu'il contiendrait des séquences non-apparentées à celles du gène de la mPC7. Le recriblage de la librairie Balb/c de souris avec une sonde dérivée d'une région 3' de l'ADNc de rPC7 nous a permis d'isoler un meilleur clone représentatif de cette région. Ce sont ces clones d'ADN génomique de souris Balb/c que nous avons utilisés pour la cartographie finale du gène de la mPC7.

Il ressort de cet effort de criblage que la librairie génomique λ DASH-129 serait incomplète et contiendrait des clones hybrides qui pourraient compliquer la cartographie d'un gène. Ceci est d'autant plus sérieux que cette librairie a été construite afin d'en tirer des séquences génomiques pour la production des vecteurs servant à l'inactivation des gènes par recombinaison homologe dans

les cellules embryonnaires dérivés de la souris 129Sv (Joyner, 1991). Sans une cartographie attentive des gènes tirés de cette librairie, on risquerait fort de connaître des échecs inexplicables dans cette entreprise

Avantages de la cartographie et du séquençage par la PCR

Traditionnellement, la cartographie d'un gène se fait par le sous-clonage dans des plasmides des fragments du gène qui se recoupent et leur séquençage. Nous avons initialement utilisé cette approche pour le clone λ DASH-9. Nous en avons sous-cloné des fragments de 4,5 kb et de 2,9 kb et nous avons pu tirer du premier fragment des séquences de la mPC7. Cette méthode est longue et exigeante à bien des égards. Une des difficultés que nous avons rencontrées est la non-clonabilité de certains fragments du gène. Ces fragments subissaient constamment des réarrangements après amplification du plasmide dans la bactérie, probablement à cause des séquences répétitives qu'ils contenaient

L'amplification des séquences de plus en plus longues par la PCR (Cheng and Kolmodin, 1997) et le séquençage direct du produit amplifié par la même méthode permettent maintenant de réaliser la cartographie d'un gène sans sous-clonage préalable. Cette méthode est fiable et rapide. Elle a récemment servi à la cartographie du gène de la PC1 humaine (Ohagi *et al.*, 1996). Nous l'avons utilisée dans notre étude avec autant de succès

Caractérisation du gène de la PC7

Les résultats de l'analyse de l'ADN génomique de souris par buvardage de Southern avec l'ADNc complet de la rPC7 comme sonde avait révélé un nombre restreint de fragments hybridants avec les fragments de digestion générés par la plupart des enzymes de restriction (voir figure 5). Nous en avons conclu que le gène de la mPC7 ne serait pas très long. En effet, après cartographie, nous estimons la taille du gène de la mPC7 à environ 13 kb. Elle est supérieure à celle de la mPC4 (9.5 kb) (Mbikay *et al.*, 1994) et de la hfurine (10 kb) (Barr *et al.*, 1991), mais inférieure à celle de la mPC1 et de la hPC1 (42kb) (Ftouhi *et al.*,

1994; Ohagi *et al.*, 1996), celle de la hPC2 et de la mPC2 (140-150 kb) (Ohagi *et al.*, 1992); Raffin-Sanson M.L. and Mbikay M, résultats non publiés) et à celle de la PACE4 (Tsuji *et al.*, 1997) (tableau 6).

La longueur relativement courte du gène de la mPC7 est un reflet du petit nombre d'introns qu'il contient et de la faible taille de ses introns. En fait, le gène de la mPC7 contient moins d'introns que tous les autres gènes des PCs caractérisés à date. Il compte neuf introns alors que celui de la mPC4 en compte quatorze, celui de la hfurine quinze, celui de la mPC1 quinze, celui de la hPC2 douze et celui de la PACE4 vingt-cinq (tableau 6). Bien que le gène de la PC7 n'a que deux introns de moins que celui de la mPC2, aucun de ces introns ne dépasse les 3 kb, alors que ceux de la mPC2 sont tous supérieurs à 5 kb (dont 3 > 20 kb, Raffin-Sanson M.L. et Mbikay M, résultats non publiés), d'où la grande taille de ce dernier gène

La comparaison de la position des introns de la PC7 avec celle des autres PCs (figure 10), notamment la hfurine (Barr *et al.*, 1991), la hPC2 (Ohagi *et al.*, 1992), la mPC1 (Ftouhi *et al.*, 1994) et la mPC4 (Mbikay *et al.*, 1994), montre que l'organisation exon-intron du gène de la mPC7 diffère de celui des autres PCs à plusieurs égards. La région catalytique montre la plus forte homologie dans la position des introns avec les autres gènes des PCs. En dehors de cette région, le gène de la mPC7 n'a aucun intron en position analogue avec les autres gènes de PCs. Il est aussi intéressant de noter par exemple que dans le gène de la mPC7, il n'y a pas de correspondants des introns 2, 5 et 9 retrouvés dans les autres gènes de convertases (figure 10). L'intron 1 de la mPC7 se situe en amont de l'ATG initiateur et en aval dans les autres gènes de PCs.

Tableau VI

Comparaison des Tailles des Gènes des Différentes Convertases

convertase de Proprotéines	Taille (kb)	Nombre d'intron
mPC7	13	9
mPC4	9,5	14
hfur	10	15
mPC1	42	14
mPC2	>140	11
PACE4	250	25

Figure 10

Alignement des Introns dans les Gènes des Convertases

	1				50
mpc4	~~~~~	~~~~~	~~~~~MRPSQ	TELWLGLTLT	LALLAVRWAS
mfurin	~~~~~	~~~~~	~~~~~MELRS	WLLWVVAAG	AVVLLAADAQ
mpc1	~~~~~	~~~~~	~~~~~MEQRGW	TLQCTAFAFF	CVWCALNSVK
mpc2	~~~~~	~~~~~	~~~~~MEGGC	GSQWKAAGF.	.LFCVMVFAS
mpc7	▼MPKGRQKVPH	LDAHLGLPIC	LWLELAIFFL	VPQVMGLSEA	GGLDILGTGG
	51				100
mpc4	AQAPIYVSSW	AVRV.TKG.Y	QEAEERLARKF	GfVNLGQIFP	DDQYFHLRHR
mfurin	GQ.KIFTNTW	AVHI.PGG.P	AVADRVAQKH	GfHNLGQIFG	D..YYHFWRH
mpc1	AKRQ.FVNEW	AAEI.PGG.Q	EAASATAEEL	GyDLLGQIGS	LENHYLFKHK
mpc2	AERPvFTNHf	LVELHKDG.E	EEARQVAAEH	GfG.VRKLpF	AEGLYHFYHN
mpc7	LSWAVHLDSL	EGERKEESLT	QQADAVAQAA	GLVNAGRIGE	LQGHYLFVQP
	101				150
mpc4	GVAQQSLTPH	WGHRLR...L	KKDPKVRWFE	QQTlRRRVKR	slV.VP....
mfurin	AVTKRSLSPH	RPRHSR...L	QREpQVKWLE	QQVAKRRAKR	DVYQEP....
mpc1	SHPRRSRRSA	LHITKR...L	SDDDRVTWAE	QQYEKERSKR	SVQKDSALDL
mpc2	GLAKAKARRS	LHhKRQ...L	ERDPRIKMAL	QQEGFDRKKR	GYRDINEIDI
mpc7	TGHRQAMEVE	AMRQQAeAVL	ARHEAVRWHS	EQTLlKRAKR	SI.....
	151				200
mpc4	..TDPWFSKQ	WYMNKEIQQ.DLNIL	KAWNQGLTGR	GVVISILDDG
mfurin	..TDPKFPQQ	WYLSGVTQR.DLNVK	EAWAQGFTGH	GIVVSILDDG
mpc1	.FNDPMNWQQ	WYLQDTRMTA	ALPKLDLHVI	PVWEKGITGK	GVVITVLDDG
mpc2	NMNDPLFTKQ	WYLFNTGQAD	GTPGLDLNVA	EAWELGYTGK	GVTIGIMDDG
mpc7	HFNDPKYPQQ	WHLNN...R	RSPGRDINVT	GVWERNVTGR	GVTVVVVDDG
	201				250
mpc4	IEKDHPDLWA	NYDPLASYDF	NDYDPDPQPR	YTPNDENRHG	TRCAGEVSAT
mfurin	IEKNHPDLAG	NYDPGASFDV	NDQDPDPQPR	YTQMNDRHG	TRCAGEVAAV
mpc1	LEWNHTDIYA	NYDPEASYDF	NDNDHDPFPR	YDLTNENKHG	TRCAGEIAMQ
mpc2	IDYLHPDLAY	NYNADASYDF	SSNDPYPYPR	YTDDWFNSHG	TRCAGEVSAA
mpc7	VETHRQDIAP	NYSPEGSYDL	NSNDPDPMPH	PDEENGNHHG	TRCAGEIAAV

Figure 10 (suite 1)

	251		6		300
mpc4	ANNGFCGAGV	AFNARIGGVR	MLD.GAITDI	VEAQSLSLQP	QHIHIYSASW
mfurin	ANNGVCGVGV	AYNARIGGVR	MLD.GEVTDA	VEARSLGLNP	NHIHIYSASW
mpc1	ANNHKCGVGV	AYNSKVGGIR	MLD.GIVTDA	IEASSIGFNP	GHVDIYSASW
mpc2	ASNNICGVGV	AYNSKVAGIR	MLDQPFMETDI	IEASSISHMP	QLIDIYSASW
mpc7	PNNSFCAVGV	AYGSRIAGIR	VLD.GPLTDS	MEAVAFNKHY	QINDIYSCSW
	301		7		350
mpc4	GPEDDGRTVD	GPGLLTQEAF	RRGVTKGRQG	LGTLFIWASG	● NGGLHYDNCN
mfurin	GPEDDGKTVD	GPARLAEEAF	FRGVSQGRGG	LGSIFVWASG	NGGREHSDCN
mpc1	GPNDGKTVE	GPGRLAQKAF	EYGVKQGRQG	KGSIFVWASG	NGGRQGDNCN
mpc2	GPTDNGKTVD	GPRELTLQAM	ADGVNKGRGG	KGSIVVWASG	DGGSY.DDCN
mpc7	GPDDDGTVD	GPHQLGKAAL	QHGV MAGRQG	FGSIFVWASG	NGGQHNDNCN
	351				400
mpc4	CDGYTNSIHT	LSVGSTTRQG	RVPWYSEACA	STFTTTFSSG	VVTDPO..IV
mfurin	CDGYTNSIYT	LSISSATQFG	NVPWYSEACS	STLATYSSG	NQNEKQ..IV
mpc1	CDGYTDSIYT	ISISSASQQG	LSPWYAEKCS	STLATSYSYG	DYTDQR..IT
mpc2	CDGYASSMWT	ISINSAINDG	RTALYDESCS	STLASTFSNG	RKRNPEAGVA
mpc7	YDGYANSIYT	VTIGAVDEEG	RMPFYAEECA	SMLAVTFSSG	DKMLRSIVTT
	401		9		450
mpc4	TTDLHH..QC	TDKHTGTSAS	● APLAAGMIAL	ALEANPLLTW	RDLQHLVVRA
mfurin	TTDLRQ..KC	TESHTGTSAS	APLAAGIIAL	TLEANKNLTW	RDMQHLVVQT
mpc1	SADLHN..DC	TETHGTGTSAS	APLAAGIFAL	ALEANPNLTW	RDMQHLVVWT
mpc2	TTDLYG..NC	TLRHSGTSAA	APEAAGVFAL	ALEANLDTW	RDMQHLTVLT
mpc7	DWDLQKGTGC	TEGHTGTSAA	APLAAGMIAL	MLQVRPCLTW	RDVQHIIVFT
	451		10		500
mpc4	SR..PAQLQA	EDWRINGVGR	QVSHHYGYGL	LDAGLLVDLA	..RVWLPKTP
mfurin	SK..PAHLNA	DDWATNGVGR	KVSHSYGYGL	LDAGAMVALA	..QNWTTVAP
mpc1	SEYDPLASN.	PGWKKNAGL	MVNSRFGFGL	LNAKALVDLA	DPRTWRNVPE
mpc2	SKRNQLHDEV	HQWRRNGVGL	EFNHLFGYGV	LDAGAMVKMA	..KDWKTVPE
mpc7	A..IQYEDHH	ADWLTNEAGF	SHSHQHGFG	LNAWRLVNAA	..KIWTSVPY

Figure 10 (suite 2)

	501		11		550
mpc4	QKKCAIR.VV	HTPTPILPRM	▼	LVPKNVTACS	DGSRRRLIRS LEHVQVQLSL
mfurin	QRKCIVE.IL	VEPKDIGKRL	▼	EVRKAVTACL	GEPNH..ITR LEHVQARLTL
mpc1	KKECVVKDNN	FEPRALKANG	▼	EVIVEIPTRA	CEGQENAIKS LEHVQFEATI
mpc2	RFHC.VGGSV	QNPEKIPPTG	▼	KLVLTLKTNA	CEGKENFVRY LEHVQAVITV
mpc7	LASYVSPMLK	ENKAVPRSPH	▼	SLEVLWNVSR	TDLEMSGLKT LEHVAVTVSI
	551		12		600
mpc4	SYSRRGDLEI	FLTSPMGTRS	▼	TLVAIRPL.D	ISGQGYNNWI FMSTHYWDED
mfurin	SYNRRGDLAI	HLISPMGTRS	▼	TLAARPH.D	YSADGFNDWA FMTTHSWDED
mpc1	EYSRRGDLHV	TLTSAVGTST	▼	VLLAERER.D	TSPNGFKNWD FMSVHTWGEN
mpc2	NATRRGDLNI	NMTSPMGTKS	▼	ILLSRRPRDD	DSKVGFDKWP FMTTHTWGED
mpc7	THPRRSLEL	KLFCPSGMS	▼	LIGAPRSM.D	SDPNGFNAWT FSTVRCWGER
	601		13		650
mpc4	PQGLWTLGLE	N.KGYFFNTG	▼	TLYYITLLLY	GTAEDMTARP QAPQVTSRAR
mfurin	PAGEWVLEIE	N.TSEANNYG	▼	TLTKFTLVLY	GTAPEGLS.. .TPPESSGCK
mpc1	PVGTWTLKIT	DMSGRMQNEG	▼	RIVNWKLILH	GTSSQPEHMK QPRVYTSYNT
mpc2	ARGTWLLEL.	GFVGSAPQKG	▼	LLKEWTLMLH	GTQSAPYIDQ VVRDYQS.KL
mpc7	ARGVYRLVIR	DVGDEPLQMG	▼	ILQQWQLTLY	GSMWSPVDIK DRQSLLSAM
	651		14		700
mpc4	ACVQRD..TE	GLCQESH..	▼	PLSILAGLCL	ISSQQWWWLY SHPQQPVTEG
mfurin	TLTSSQ..AC	VVCEEYSLH	▼	QKSCVQHCPP	GFIPQVLDTH YSTENDVEII
mpc1	VQNDRRGVEK	MVNVVEKRPT	▼	QKSLNGNLLV	PKNSSSSNVE GRRDEQVQGT
mpc2	AMSKKQEELE	ELDEAVERS	▼	QSILRKN~~~~	~~~~~
mpc7	SGKYLHDGFT	LPCPPGLKIP	▼	EED..GYTIT	PNTLKTLLV GCFSVFTIY
	701				750
mpc4	QASCHPPVTP	AAAA~~~~~	▼	~~~~~	~~~~~
mfurin	RASVCTPCHA	SCATCQGPAP	▼	TDCLSCPSHA	SLDPVEQTCS RQSQSSRESR
mpc1	PSKAMLRLLQ	SAFSKNALSK	▼	QSPKKSAPAK	LSIPYESFYE ALEKLNKPSK
mpc2	~~~~~	~~~~~	▼	~~~~~	~~~~~
mpc7	YMLEVCLSQR	NKASTHGCRK	▼	GCCPWAPRRQ	NSKDAGTALE SMPLCSSKDL

Figure 10 (suite 3)

```

          751                                     800
mpc4  ~~~~~~
mfurin PQQQPPALRP EVEMEPRLQA GLASHLPEVL AGLSCLIIVL IFGIVFLFLH
mpc1  LEGSEDSLYS DYVDVFYNTK PYKHRDDRLL QALMDILNEE N~~~~~
mpc2  ~~~~~~
mpc7  DGVDSEHGDC TTASSFLAPE LLGEADWSLS QNSKSELDCP PHQPPDLLQG

          801                                     850
mpc4  ~~~~~~
mfurin RCSGFSFRGM KVYTMDRGLI SYKGLPPEAW QEECPDSEED DEGRGERTAF
mpc1  ~~~~~~
mpc2  ~~~~~~
mpc7  KIC~~~~~

          851
mpc4  ~~~~~~
mfurin IKDQSAL
mpc1  ~~~~~~
mpc2  ~~~~~~
mpc7  ~~~~~~

```

Alignement des séquences de la mPC4, mFurine, mPC1, mPC2 et mPC7 à l'aide du programme ALIGN. Les positions des introns sont indiquées par des têtes de flèches. Les points noirs représentent les a.a. du site actif; La flèche perpendiculaire représente le site de coupure du zymogène, la séquence pentapeptidique hautement conservée chez les PCs est soulignée, enfin, les introns de la mPC4 sont numérotés pour titre de comparaison.

L'évolution des gènes passe par des délétions, des insertions et des mutations autour des domaines fonctionnels (Ftouhi *et al.*, 1994; Roebroek *et al.*, 1995). La disparition d'introns durant l'évolution dans certains gènes d'une même famille après leur introduction, est un phénomène bien documenté, comme l'illustre la différence positionnelle d'introns dans les deux gènes d'insuline de rat (Bell *et al.*, 1980; Perler *et al.*, 1980). Les similitudes et les différences d'organisation entre le gène de la mPC7 et ceux des autres gènes des PCs indiquent que tous ces gènes partageraient un gène ancestral commun, mais que celui de la PC7 aurait divergé des autres assez tôt au cours de l'évolution. La comparaison de la structure des gènes de convertases semble donc confirmer la phylogénie de ces gènes telle que dérivée de leur homologie de séquences (Seidah *et al.*, 1996) (voir figure 2)

Le gène de la PC7 (symbole du locus *Pcsk7*) a été localisé sur le chromosome (Chr) 9 chez la souris (Seidah *et al.*, 1996), alors que celui de la PC1 (*Pcsk1*) sur le Chr 13, celui de la PC2 (*Pcsk2*) sur le Chr 2, celui de la furine (*Pcsk3*) et de la PACE4 (*Pcsk6*) sur le Chr 7, celui de la PC4 (*Pcsk4*) sur Chr 10 et celui de la PC5 (*Pcsk5*) sur le Chr 9 (revu dans Mbikay *et al.*, 1995). Ainsi à l'exception des loci *Pcsk3* et *Pcsk6* qui semblent être issus d'une duplication sur le Chr 7, les autres loci *Pcsk* seraient dérivés de translocations d'un chromosome à l'autre. Ces réarrangements auraient eu lieu avant la divergence des lignées conduisant à la souris et à l'homme, puisqu'il existe une synténie des loci avoisinants entre les deux espèces (discuté dans Mbikay *et al.*, 1995; Seidah *et al.*, 1996).

Comparaison des séquences exoniques du gène de la mPC7 avec celle de l'ADNc

En comparant les séquences exoniques que nous avons obtenues à celles publiées par Constam *et al.* (1996), nous avons relevé quelques différences qui méritent une certaine attention (voir figure 6).

Tableau VII
Polymorphismes exoniques dans la mPC7

Codons ADNc de souris	Codons génomique de souris	Site de restriction perdu	Site de restriction généré
<p style="text-align: center;"><i>Asn Gly Asn</i></p> <p>846 <u>AAT GGC AAC</u> 854</p>	<p style="text-align: center;"><i>Asn Gly Asn</i></p> <p>AAT <u>GGT AAC</u></p>	-----	<i>BstEII</i>
<p style="text-align: center;"><i>Tyr Ala Asn</i></p> <p>1200 TAT GCC AAC 1208</p>	<p style="text-align: center;"><i>Tyr Thr Asn</i></p> <p>TAT ACC AAC</p>	-----	-----
<p style="text-align: center;"><i>Glu Val Leu</i></p> <p>1695 <u>GAG GTG CTA</u> 1703</p>	<p style="text-align: center;"><i>Glu Val Leu</i></p> <p>GAG <u>GTC CTA</u></p>	-----	<i>Avall</i>
<p style="text-align: center;"><i>Asp Ser Asp</i></p> <p>1863 <u>GAC TCG GAT</u></p>	<p style="text-align: center;"><i>Asp Leu Asp</i></p> <p>GAC TTG GAT</p>	<i>Hinfl</i>	-----
<p style="text-align: center;"><i>Asn Ala Trp</i></p> <p>1884 AAT GCT TGG 1892</p>	<p style="text-align: center;"><i>Asn Ala Trp</i></p> <p>AAT <u>GCC TGG</u></p>	-----	<i>BstnI</i>

Séquences "exoniques" autour des codons différents (voir figure 6), entre la séquence publiée l'ADNc de la souris et celle de l'ADN génomique déterminée dans cette étude. Dans certains cas, un sites de restriction est soit perdu, soit gagné. Ces sites sont soulignés.

Une transition convertit le codon GGC₂₂₄ au GGT tous les deux spécifiant une Gly. Une autre change le codon GCC₃₄₂ d'une Ala en ACC pour une Thr et le codon TCG₅₆₃ d'une Ser au TTC d'une Leu, enfin, le codon GCT₅₇₀ est muté en GCC, tous deux spécifiant une Ala. Une transversion convertit le codon GTG₅₀₇ en GTC, tous les deux spécifiant une Val (voir figure 6).

Il est à noter que les mutations Ala342Thr et la mutation Ser563Leu, amèneraient un changement de polarité. Nous ignorons si elles produiraient de perturbations majeures dans la structure et l'activité de la mPC7. Très probablement, toutes ses mutations seraient de simples polymorphismes entre souches de souris. De façon intéressante, certaines de ces substitutions génèrent des sites nouveaux ou des sites reconnus par différents enzymes de restriction (tableau 7) qui pourraient s'avérer utiles dans d'éventuelles études des polymorphisme de longueurs de fragments de restriction (RFLP, *restriction fragment length polymorphism*) chez les souris

Sites d'initiation de la transcription

L'analyse des sites d'initiation de la transcription par extension d'amorce (*primer extension*), indique que le produit majeur de transcription du gène de la PC7 donne un ARNm avec une région 5' non traduite (UTR, *untranslated region*) longue de 166 nucléotides. Toutefois, la longueur de l'UTR peut varier dans certains isoformes de la mPC7 puisque nous avons observé des sites mineurs d'initiation de la transcription en positions 181 et 155 en amont et en aval du site majeur. La multiplicité des sites d'initiation de la transcription est une caractéristique des promoteurs qui n'ont pas de boîte TATA, mais qui sont riches en GC, comme le sont ceux des gènes des convertases étudiés à ce jour (Ftouhi *et al.*, 1994). Le site majeur que nous avons identifié est interne aux termini 5' rapportés pour les ADNc de la rPC7 et de la mPC7 (Constam *et al.*, 1996; Seidah *et al.*, 1996). Les séquences 5' rapportées sont dérivées des produits de l'amplification rapide des bouts des ADNc (RACE, *rapid amplification of cDNA ends*). Elles représenteraient les plus longues séquences obtenues par cette

méthode, et n'identifient pas, comme l'extension d'amorce, l'isoforme majeure d'ARNm de la mPC7. Toutefois, même nos résultats d'extension devraient être dans le futur corroborés par l'analyse de protection à RNase (RPA, *RNAase protection assay*) après hybridation de l'ARN tissulaire avec une sonde d'ARNc dérivé d'un fragment génomique du gène de la mPC7 recouvrant le 5'-UTR de son ARNm

CONCLUSIONS ET PERSPECTIVES

Ce travail de caractérisation de gène de la PC7 chez la souris constitue notre contribution à la compréhension de l'évolution des gènes des convertases des proprotéines. La connaissance de sa structure permettra de concevoir de façon rationnelle des vecteurs servant à l'inactivation éventuelle de ce gène par recombinaison homologe, afin de produire des mutants sans PC7 chez la souris. La production d'une souris déficiente en PC7 non seulement permettra d'évaluer son importance pour le développement et la physiologie normale, mais aussi de rechercher et d'identifier ses substrats naturels

Une analyse plus approfondie de son promoteur reste à faire. Elle permettra d'élucider des éléments structuraux qui déterminent son expression tissulaire et sa régulation ainsi que les facteurs nucléaires qui s'y attachent. Il serait intéressant de savoir par exemple quelles régions du gène facilitent l'expression élevée de la PC7 dans les testicules et le thymus et stimulent cette expression dans les lymphocytes activés (Decroly *et al.*, 1997). Il serait aussi d'intérêt de savoir si et pourquoi la translocation réciproque (11:14) (q23:q32) dont la jonction se situe dans la région 3' du gène de la PC7 humaine serait à la base du lymphome à partir duquel la hPC7 a été identifiée (Meerabux *et al.*, 1996). De même la surexpression de la PC7 récemment observée par M. Marcinkiewicz et N.G. Seidah (communication personnelle) autour des plaques-amyloïdiques des cerveaux de patients morts de la maladie d'Alzheimer pourrait constituer une expression du gène aux facteurs paracriniens produits par ces tissus lésés. L'étude des éléments et des facteurs qui contrôlent cette réponse permettrait de cerner la contribution de cette enzyme à la pathogenèse et la pathophysiologie de cette maladie et d'identifier les points éventuels d'intervention pour son traitement.

Des recherches futures sur le gène de la PC7 permettront de le dire.

BIBLIOGRAPHIE

- Jackson, R. S., Creemers, J. W. M., Ohagi, S., Raffin-Sanson, M. L., Sanders, L., Montague, C. T., Hutton, J. C., and O'Rahilly, S. (1997). Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat. Genet.* 16, 303-306.
- Anderson, E. D., VanSlyke, J. K., Thulin, C. D., Jean, F., and Thomas, G. (1997). Activation of the furin endoprotease is a multiple-step process: requirements for acidification and internal propeptide cleavage. *EMBO J.* 16, 1508-1518.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K. (1994). *Current Protocols in Molecular Biology*.
- Barr, P. J., Mason, O. B., Landsberg, K. E., Wong, P. A., Kiefer, M. C., and Brake, A. J. (1991). cDNA and gene structure for a human subtilisin-like protease with cleavage specificity for paired basic amino acid residues. *DNA Cell Biol.* 10, 319-328.
- Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer, E., and Goodman, H. M. (1980). Sequence of the human insulin gene. *Nature* 284, 26-32.
- Benjannet, S., Rondeau, N., Day, R., Chrétien, M., and Seidah, N. G. (1991). PC1 and PC2 are proprotein convertases capable of cleaving proopiomelanocortin at distinct pairs of basic residues. *Proc. Natl. Acad. Sci. U.S.A.* 88, 3564-3568.
- Benjannet, S., Rondeau, N., Paquet, L., Boudreault, A., Lazure, C., Chrétien, M., and Seidah, N. G. (1993). Comparative biosynthesis, covalent post-translational modifications and efficiency of prosegment cleavage of the prohormone convertases PC1 and PC2: glycosylation, sulphation and identification of the intracellular site of prosegment cleavage of PC1 and PC2. *Biochem. J.* 294, 735-743.
- Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7, 1513-1523.

- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U.S.A.* 75, 4853-4857.
- Bresnahan, P. A., Hayflick, J. S., Molloy, S. S., and Thomas, G. (1993). Endoproteolysis of growth factors and other endocrine precursor proteins. In "Mechanisms of intracellular trafficking and processing of proproteins" (Y. Peng Loh, Ed.), pp. 225-250. CRC Press, Boca Raton.
- Bruzzaniti, A., Goodge, K., Jay, P., Taviaux, S. A., Lam, M. H., Berta, P., Martin, T. J., Moseley, J. M., and Gillespie, M. T. (1996). PC8, a new member of the convertase family. *Biochem. J.* 314, 727-731.
- Cheng, S., and Kolmodin, L. A. (1997). XL PCR amplification of long targets from genomic DNA. *Meth. Mol. Biol.* 67, 17-29.
- Chrétien, M., Mbikay, M., Gaspar, L., and Seidah, N. G. (1995). Proprotein convertases and the pathophysiology of human diseases: prospective considerations. *Proc. Assoc. Am. Physicians* 107, 47-66.
- Constam, D. B., Calfon, M., and Robertson, E. J. (1996). SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone morphogenetic proteins at distinct sites during embryogenesis. *J. Cell Biol.* 134, 181-191.
- Creemers, J. W., Vey, M., Schafer, W., Ayoubi, T. A., Roebroek, A. J., Klenk, H. D., Garten, W., and Van de Ven, W. J. (1995). Endoproteolytic cleavage of its propeptide is a prerequisite for efficient transport of furin out of the endoplasmic reticulum. *J. Biol. Chem* 270, 2695-2702.
- Day, R., Schafer, M. K., Watson, S. J., Chrétien, M., and Seidah, N. G. (1992). Distribution and regulation of the prohormone convertases PC1 and PC2 in the rat pituitary. *Mol. Endocrinol.* 6, 485-497.
- De Bie, I., Marcinkiewicz, M., Malide, D., Lazure, C., Nakayama, K., Bendayan, M., and Seidah, N. G. (1996). The isoforms of proprotein convertase PC5 are sorted to different subcellular compartments. *J. Cell Biol.* 135, 1261-1275.

- Decroly, E., Benjannet, S., Savaria, D., and Seidah, N. G. (1997). Comparative functional role of PC7 and furin in the processing of the HIV envelope glycoprotein gp160. *FEBS Lett.* 405, 68-72.
- Decroly, E., Wouters, S., Di Bello, C., Lazure, C., Ruyschaert, J. M., and Seidah, N. G. (1996). Identification of the paired basic convertases implicated in HIV gp160 processing based on in vitro assays and expression in CD4(+) cell lines. *J. Biol. Chem* 271, 30442-30450.
- Devi, L. (1991). Consensus sequence for processing of peptide precursors at monobasic sites. *FEBS Lett.* 280, 189-194.
- Eder, J., Rheinhecker, M., and Fersht, A. R. (1993). Folding of subtilisin BPN': role of the pro-sequence. *J. Mol. Biol.* 233, 293-304.
- Elgen, E. *et al.* (1991). *Strategies* 4, 8-9.
- Feinberg AP. and Vogelstein B. (1983). A technique for radiolabeling DNA restriction endonuclease fragment to high specific activity. *Anal. Biochem.* 132, 6-13.
- Ftouhi, N., Day, R., Mbikay, M., Chrétien, M., and Seidah, N. G. (1994). Gene organization of the mouse pro-hormone and pro-protein convertase PC1. *DNA Cell Biol.* 13, 395-407.
- Furuta, M., Yano, H., Zhou, A., Rouille, Y., Holst, J. J., Carroll, R., Ravazzola, M., Orci, L., Furuta, H., and Steiner, D. F. (1997). Defective prohormone processing and altered pancreatic islet morphology in mice lacking active SPC2. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6646-6651.
- Hallenberger, S., Moulard, M., Sordel, M., Klenk, H. D., and Garten, W. (1997). The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells. *J. Virol.* 71, 1036-1045.
- Julius, D., Brake, A., Blair, L., Kunisawa, R., and Thorner, J. (1984). Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. *Cell* 37, 1075-1089.

- Karn, J.M., Brenner, S., Barnett, L., Casareni, G. (1980). Proc. Natl. Acad. Sci. 77, 5172.
- Kozak, M. (1989). The scanning model for translation: an update. *J. Cell Biol.* 108, 229-241.
- Lusson, J., Benjannet, S., Hamelin, J., Savaria, D., Chrétien, M., and Seidah, N. G. (1997). The integrity of the RRGDL sequence of the proprotein convertase PC1 is critical for its zymogen and C-terminal processing and for its cellular trafficking. *Biochem. J.* 326, 737-744.
- Lusson, J., Vieau, D., Hamelin, J., Day, R., Chrétien, M., and Seidah, N. G. (1993). cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a candidate proprotein convertase expressed in endocrine and nonendocrine cells. *Proc. Natl. Acad. Sci. U.S.A.* 90, 6691-6695.
- Marcinkiewicz, M., Seidah, N. G., and Chrétien, M. (1993). Les convertases des prohormones et le système nerveux. *Médecine-Sciences* 9, 553-561.
- Mbikay, M., Raffin-Sanson, M. L., Tadros, H., Sirois, F., Seidah, N. G., and Chrétien, M. (1994). Structure of the gene for the testis-specific proprotein convertase 4 and of its alternate messenger RNA isoforms. *Genomics* 20, 231-237.
- Mbikay, M., Tadros, H., Ishida, N., Lerner, C. P., De Lamirande, E., Chen, A., El-Alfy, M., Clermont, Y., Seidah, N. G., Chrétien, M., Gagnon, C., and Simpson, E. M. (1997). Impaired fertility in mice deficient for the testicular germ-cell protease PC4. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6842-6846.
- Meerabux, J., Yaspo, M. L., Roebroek, A. J., Van de Ven, W. J., Lister, T. A., and Young, B. D. (1996). A new member of the proprotein convertase gene family (LPC) is located at a chromosome translocation breakpoint in lymphomas. *Cancer Res.* 56, 448-451.
- Mercure, C., Jutras, I., Day, R., Seidah, N. G., and Reudelhuber, T. L. (1996). Prohormone convertase PC5 is a candidate processing enzyme for prorenin in the human adrenal cortex. *Hypertension* 28, 840-846.

- Molloy, S. S., Thomas, L., VanSlyke, J. K., Stenberg, P. E., and Thomas, G. (1994). Intracellular trafficking and activation of the furin proprotein convertase: localization to the TGN and recycling from the cell surface. *EMBO J.* 13, 18-33.
- Ohnishi, Y., Shioda, T., Nakayama, K., Iwata, S., Gotoh, B., Hamaguchi, M., and Nagai, Y. (1994). A furin-defective cell line is able to process correctly the gp160 of human immunodeficiency virus type 1. *J. Virol.* 68, 4075-4079.
- O'Rahilly, S., Gray, H., Humphreys, P. J., Krook, A., Polonsky, K. S., White, A., Gibson, S., Taylor, K., and Carr, C. (1995). Impaired processing of prohormones associated with abnormalities of glucose homeostasis and adrenal function. *N. Engl. J. Med.* 333, 1386-1390.
- Roebroek, A. J. M., Pauli, I., Van Leuven, F., and Van de Ven, W. J. M. (1997). Targeted inactivation of the FUR gene in mice. In "Processing of Peptide Hormones, Neurotransmitters, Growth Factors and Viral Proteins" (K. Symposia, Ed.), pp. 60. Keystone Symposia, Taos, New Mexico.
- Roebroek, A. J., Creemers, J. W., Pauli, I. G., Kurzik-Dumke, U., Rentrop, M., Gateff, E. A., Leunissen, J. A., and Van de Ven, W. J. (1992). Cloning and functional expression of Dfurin2, a subtilisin-like proprotein processing enzyme of *Drosophila melanogaster* with multiple repeats of a cysteine motif. *J. Biol. Chem* 267, 17208-17215.
- Roebroek, A. J., Schalken, J. A., Leunissen, J. A., Onnekink, C., Bloemers, H. P., and Van de Ven, W. J. (1986). Evolutionary conserved close linkage of the c-fes/fps proto-oncogene and genetic sequences encoding a receptor-like protein. *EMBO J.* 5, 2197-2202.
- Schafer, M. K., Day, R., Cullinan, W. E., Chrétien, M., Seidah, N. G., and Watson, S. J. (1993). Gene expression of prohormone and proprotein convertases in the rat CNS: a comparative in situ hybridization analysis. *J. Neurosci.* 13, 1258-1279.

- Seidah, N. G., and Chrétien, M. (1992). Proprotein and prohormone convertases of the subtilisin family: recent developments and future perspectives. *Trends Endocrinol. Metab.* 3, 133-140.
- Seidah, N. G., Chrétien, M., and Day, R. (1994). The family of subtilisin/kexin like pro-protein and pro-hormone convertases: divergent or shared functions.. *Biochimie* 76, 197-209.
- Seidah, N. G., Day, R., Benjannet, S., Rondeau, N., Boudreault, A., Reudelhuber, T., Schafer, M. K.-H., Watson, S. J., and Chrétien, M. (1992a). "The prohormone and proprotein processing enzymes PC1 and PC2: structure, selective cleavage of mouse POMC and human renin at pairs of basic residues, cellular expression, tissue distribution and mRNA regulation." U.S. Department of Health and Human Services, Washington.
- Seidah, N. G., Day, R., Hamelin, J., Gaspar, A., Collard, M. W., and Chrétien, M. (1992b). Testicular expression of PC4 in the rat: molecular diversity of a novel germ cell-specific Kex2/subtilisin-like proprotein convertase. *Mol. Endocrinol.* 6, 1559-1570.
- Seidah, N. G., Day, R., Marcinkiewicz, M., and Chrétien, M. (1997). Precursor convertases: an evolutionary ancient, cell-specific, combinatorial mechanism yielding diverse bioactive peptides and proteins. *Ann. N.Y. Acad. Sci.* in press.
- Seidah, N. G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M., and Chrétien, M. (1990). cDNA sequence of two distinct pituitary proteins homologous to Kex2 and furin gene products: tissue-specific mRNAs encoding candidates for pro-hormone processing proteinases. *DNA Cell Biol.* 9, 789.
- Seidah, N. G., Hamelin, J., Mamarbachi, M., Dong, W., Tadros, H., Mbikay, M., Chrétien, M., and Day, R. (1996). cDNA structure, tissue distribution and chromosomal localization of rat PC7: a novel mammalian convertase closest to yeast kexin-like proteinases. *Proc. Natl. Acad. Sci. USA* 93, 3388-3393.
- Takahashi, S., Kasai, K., Hatsuzawa, K., Kitamura, N., Misumi, Y., Ikehara, Y., Murakami, K., and Nakayama, K. (1993). A mutation of furin causes the lack

- of precursor-processing activity in human colon carcinoma LoVo cells. *Biochem. Biophys. Res. Comm.* 195, 1019-1026.
- Takahashi, S., Nakagawa, T., Kasai, K., Banno, T., Duguay, S. J., Van de Ven, W. J., Murakami, K., and Nakayama, K. (1995). A second mutant allele of furin in the processing-incompetent cell line, LoVo. Evidence for involvement of the homo B domain in autocatalytic activation. *J. Biol. Chem* 270, 26565-26569.
- Torii, S., Yamagishi, T., Murakami, K., and Nakayama, K. (1993). Localization of Kex2-like processing endoproteases, furin and PC4, within mouse testis by in situ hybridization. *FEBS Lett.* 316, 12-16.
- Wilcox, C. A., and Fuller, R. S. (1991). Posttranslational processing of the prohormone-cleaving Kex2 protease in the *Saccharomyces cerevisiae* secretory pathway. *J. Cell Biol.* 115, 297-307.
- Wise, R. J., Barr, P. J., Wong, P. A., Kiefer, M. C., Brake, A. J., and Kaufman, R. J. (1990). Expression of a human proprotein processing enzyme: correct cleavage of the von Willebrand factor precursor at a paired basic amino acid site. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9378-9382
- Zheng, M., Seidah, N. G., and Pintar, J. E. (1997). The developmental expression in the rat CNS and peripheral tissues of proteases PC5 and PACE4 mRNAs: comparison with other proprotein processing enzymes. *Dev. Biol.* 181, 268-283.
- Zheng, M., Streck, R. D., Scott, R. E., Seidah, N. G., and Pintar, J. E. (1994). The developmental expression in rat of proteases furin, PC1, PC2, and carboxypeptidase E: implications for early maturation of proteolytic processing capacity. *J. Neurosci.* 14, 4656-4673.