

## Lipocalines et transport de ligands hydrophobes

Les lipocalines sont des protéines qui transportent les ligands hydrophobes chez les vertébrés et les invertébrés. Les gènes codant pour les lipocalines s'expriment dans de nombreux tissus et les protéines correspondantes (à l'exception de deux d'entre elles) sont sécrétées dans les liquides biologiques : plasma, lait, salive, larmes, mucus nasal, liquide séminal, liquide amniotique. Ces protéines semblent appartenir à une même famille : leur structure tridimensionnelle, déterminée pour cinq d'entre elles, est très voisine : huit feuillets  $\beta$  antiparallèles forment un calice au fond duquel le ligand hydrophobe peut se nicher ; ce sont de petites protéines ayant un petit nombre d'acides aminés très conservés ; la structure de sept des gènes de lipocalines est étonnamment semblable : même taille des exons homologues, bonne conservation à travers les espèces de la position des jonctions intron-exon. De plus, six des huit gènes des lipocalines humaines connus sont situés sur le bras long du chromosome 9, confortant l'hypothèse d'un gène ancestral unique largement dupliqué.

---

**Anne-Marie Françoise  
Gachon**

---

ADRESSE

A.M.F. Gachon : maître de conférences des universités, praticien hospitalier, docteur ès sciences. Laboratoire de biochimie médicale, faculté de médecine, place Henri-Dunant, 63001 Clermont-Ferrand Cedex, France.

La reconnaissance et le transport des molécules hydrophobes sont assurés par une super-famille de protéines baptisées calyces ; ce terme, utilisé par Flower *et al* [1], provenant du grec *kalux*, calice, reflète la structure en forme de calice de ses membres. Deux grandes catégories de calyces ont été décrites à ce jour : les protéines fixant les acides gras (*fatty acid binding proteins* ou FABP), et les lipocalines. Les FABP sont de petites protéines, presque exclusivement intracellulaires, comprenant des

structures bien conservées de dix feuillets  $\beta$  antiparallèles ; leur chef de file est la CRBP (*cellular retinol-binding protein*). Les lipocalines sont presque toutes extracellulaires et possèdent un motif structural commun de huit feuillets  $\beta$  antiparallèles ; leur chef de file est la RBP (*retinol-binding protein*). Le système le mieux caractérisé et le plus étudié jusqu'à présent est celui d'une petite molécule hydrophobe, le rétinol, et de ses protéines de transport. Le transport du rétinol fait intervenir de nombreuses protéines ; parmi elles la RBP [2], qui est une lipoca-

m/s n° 1 vol. 10, janvier 94

line plasmatique et les CRBP qui sont des FABP intracellulaires. Les esters de rétinol arrivent dans le sang, incorporés dans les particules de chylomicrons, atteignent les cellules parenchymateuses hépatiques et sont hydrolysés. Le rétinol libéré se fixe à la RBP produite localement; ce complexe est ensuite sécrété et transporté jusqu'aux cellules de Kupffer puis sécrété à nouveau dans le plasma. La majeure partie du complexe rétinol-RBP est associée de manière réversible à la transthyrétine (préalbumine sérique). Le complexe rétinol-RBP, non associé à la transthyrétine, serait capté par les récepteurs membranaires des cellules cibles [3]. La RBP semble donc être une protéine bifonctionnelle permettant de diriger une «activité» vers une cible précise; elle possède deux sites de reconnaissance moléculaire, un pour le ligand hydrophobe (le rétinol), l'autre pour le récepteur de la cellule vers laquelle est transporté le

ligand (figure 1). De nombreuses protéines intracellulaires, fixant le rétinol (*cellular retinol-binding protein* ou CRBP), et l'acide rétinolique (*cellular retinoic acid binding protein* ou CRABP) interviennent ensuite, mais leur rôle n'est pas encore complètement connu.

Les travaux de Pervaiz et Brew [4] et de Godovac-Zimmerman [5] ont beaucoup apporté au concept de regroupement de la RBP et d'un certain nombre de protéines sécrétées au sein d'un même groupe de transporteurs de ligands hydrophobes: les lipocalines. Les ligands sont très variés: rétinol, phéromones, pigments biliaires, etc. (à noter que le ligand préférentiel de l'apolipoprotéine D n'est vraisemblablement pas le cholestérol mais la progestérone) [36]. Les valeurs des constantes d'affinité sont de l'ordre de  $10^{-8}$  à  $10^{-5}$  M, indiquant une affinité qui permet le relargage des ligands.

Une trentaine de lipocalines sont

actuellement connues (Tableau I) et de nombreuses lipocalines restent sans aucun doute à découvrir. Il faut noter que certaines protéines sont homologues. Parmi les protéines homologues (Tableau II), certaines ont un pourcentage de similitude élevé (par exemple, la RBP chez l'homme, les bovins, le lapin et le rat), plus faible pour d'autres (lipocaline lacrymale humaine et protéine de la glande de von Ebner chez le rat). L'*odorant binding protein* (OBP), l'orosomucoïde, les  $\alpha 2$  globulines urinaires, la prostaglandine D synthétase, l' $\alpha 1$  microglobuline, les protéines fixant les pigments, sont elles aussi retrouvées dans différentes espèces; la  $\beta$  lactoglobuline et la protéine placentaire PP14 sont analogues.

L'expression tissulaire de ces protéines a été généralement déterminée par des techniques d'immunohistochimie et d'hybridation *in situ*. Ces protéines sont exprimées dans des tissus très variés de différentes espèces. Presque toutes sont sécrétées; seules deux font exception: la probasine qui a une localisation intracellulaire (en plus de sa forme sécrétée), et la prostaglandine D synthétase qui a une localisation membranaire. Les fluides de sécrétion sont très divers: plasma, lait, salive, larmes, liquide séminal, liquide amniotique. Les protéines plasmatiques, de masses moléculaires inférieures à 40 000, sont présentes en faible quantité dans le plasma d'individus normaux; de telles protéines franchissent la barrière glomérulaire assez facilement, mais la plupart des protéines filtrées sont réabsorbées par les tubules rénaux et catabolisées; seule une petite fraction est excrétée dans l'urine finale. Les sujets présentant une fonction tubulaire déficiente excrètent en quantité considérable des protéines plasmatiques de faible poids moléculaire; c'est ainsi que la RBP et l' $\alpha 1$  microglobuline, lipocalines sériques, ont été isolées à partir d'urine de patients souffrant de protéinurie tubulaire. Les protéines plasmatiques de masse moléculaire supérieure à 40 000 sont purifiées à partir du plasma. Dans les sécrétions, les lipocalines sont, soit sous forme de dimère ( $\beta$  lactoglobuline des rumi-

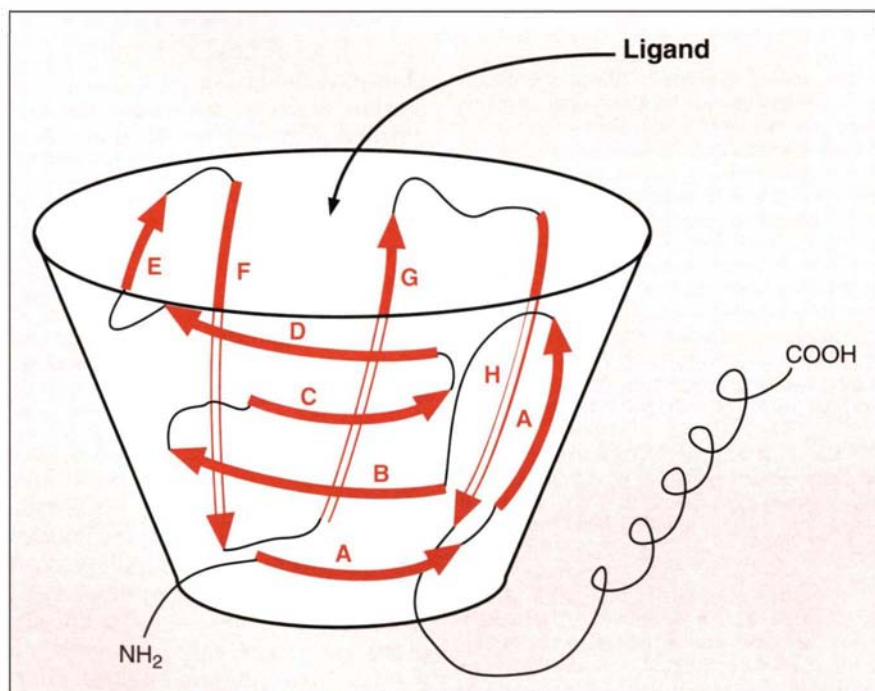


Figure 1. **Représentation schématique de la structure des lipocalines.** La structure en forme de « filtre à café » permet au ligand hydrophobe de venir se fixer. Certaines de ces protéines, comme la retinol-binding protein (RBP), pourraient posséder deux sites particuliers: un pour la fixation du ligand hydrophobe, l'autre pour être dirigées vers les récepteurs membranaires des cellules cibles.

## RÉFÉRENCES

1. Flower DR, North ACT, Attwood TK. Structure and sequence relationships in the lipocalins and related proteins. *Protein Science* 1993; 2: 753-61.
2. Newcomer ME, Jones TA, Sundelin J, Eriksson U, Rask L, Peterson PA. The three dimensional structure of retinol-binding protein. *EMBO J* 1984; 7: 1451-4.
3. Sivaprasadarao A, Findlay JBC. The interaction of retinol-binding protein with its plasma-membrane receptor. *Biochem J* 1988; 255: 561-9.
4. Pervaiz S, Brew K. Homology and structure-function correlations between  $\alpha$ 1-acid glycoprotein and serum retinol-binding protein and its relatives. *FASEB J* 1987; 1: 209-14.
5. Godovac-Zimmermann J. The structural motif of  $\beta$ -lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic molecules. *TIBS* 1988; 13: 64-6.
6. Cowan SW, Newcomer ME, Jones TA. Crystallographic refinement of human serum retinol-binding protein at 2A resolution. *Proteins: structure, function and genetics* 1990; 8: 44-61.
7. Berni R, Stoppini M, Zapponi MC, Meloni ML, Monaco HL, Zanotti G. The bovine plasma retinol-binding protein: amino acid sequence, interaction with transthyretin, crystallization and preliminary X-ray data. *Eur J Biochem* 1990; 192: 507-13.
8. Pevsner J, Sklar P, Snyder H. Odorant-binding protein: localization to nasal glands and secretions. *Proc Natl Acad Sci USA* 1986; 83: 4942-6.
9. Pevsner J, Reed RR, Feinstein PG, Snyder SH. Molecular cloning of odorant-binding protein: member of a ligand carrier family. *Science* 1988; 241: 336-9.
10. Lee KH, Wells RG, Reed RR. Isolation of an olfactory cDNA: similarity to retinol-binding protein suggests a role in olfaction. *Science* 1987; 1053-6.
11. Delaire A, Lassagne H, Gachon AMF. New members of the lipocalin family in human tear fluid. *Exp Eye Res* 1992; 55: 645-7.
12. Lassagne H, Gachon AMF. Cloning of a human lacrimal lipocalin secreted in tears. *Exp Eye Res* 1993; 56: 605-9.
13. Gachon AMF. Lipocalins: do we taste with our tears? *TIBS* 1993; 18: 206-7.
14. Lassagne H, Ressot C, Mattei MG, Gachon AMF. Assignment of the human tear lipocalin gene (LCN1) to 9q34 by *in situ* hybridization. *Genomics* 1993; 18: 160-1.
15. Redl B, Holzfeind P, Lottspeich F. cDNA cloning and sequencing reveals human tear prealbumin to be a member of the lipophilic-ligand carrier protein superfamily. *J Biol Chem* 1992; 267: 20282-7.
16. Schmale H, Holtgreve-Grez H, Christiansen H. Possible role for salivary gland protein in taste reception indicated by homology to lipophilic-ligand carrier proteins. *Nature* 1990; 343: 366-9.
17. Vaisse C, Atger M, Potier B, Milgrom E. Human placental protein PP14 gene: sequence and characterization of a short duplication. *DNA Cell Biol* 1990; 9: 401-13.
18. Julkunen M, Seppala M, Janne OA. Complete amino acid sequence of human placental protein 14: a progesterone-regulated uterine protein homologous to  $\beta$ -lactoglobulins. *Proc Natl Acad Sci USA* 1988; 85: 8845-9.
19. Papiz MZ, Sawyer L, Eliopoulos EE, et al. The structure of  $\beta$ -lactoglobulin and its similarity to plasma retinol-binding protein. *Nature* 1986; 324: 383-5.
20. Ali S, Clark J. Characterization of the gene encoding ovine beta-lactoglobulin; similarity to the genes for retinol-binding protein and other secretory proteins. *J Mol Biol* 1988; 199: 415-26.
21. Böcskei Z, Groom CR, Flower DR, et al. Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature* 1992; 360: 186-8.
22. Gubits RM, Lynch KR, Kulkarni AB, et al. Differential regulation of  $\alpha$  2 micro globulin gene expression in liver, lacrimal gland, and salivary gland. *J Biol Chem* 1984; 259: 12803-9.
23. Shahan K, Denaro M, Gilmartin M, Shi Y, Derman E. Expression of six mouse major urinary protein genes in the mammary, parotid, sublingual, submaxillary, and lachrymal glands and in the liver. *Mol Cell Biol* 1987; 7: 1947-54.
24. Nagata A, Suzuki Y, Igarashi M, et al. Human brain prostaglandin D synthase has been evolutionarily differentiated from lipophilic-ligand carrier proteins. *Proc Natl Acad Sci USA* 1991; 88: 4020-4.

nants, odorant binding protein des bovins) soit sous la forme d'une sous-unité d'un hétéro-trimère (chaîne  $\gamma$  du facteur C8 de la cascade du complément), soit enfin sous celle de simple chaîne polypeptidique (retinol-binding protein, lipocaline lacrymale). Tous les stades de glycosylation sont observés, de protéines fortement glycosylées comme l'orosomucoïde et l' $\alpha$ 1 microglobuline à des protéines non glycosylées comme la RBP ou la lipocaline lacrymale.

Divers éléments ont cependant permis de regrouper ces protéines dans la famille des lipocalines: (1) les études de cristallographie aux rayons X; (2) les similitudes de séquences protéiques; (3) les analogies au niveau de l'organisation des gènes correspondants; enfin nous présenterons (4) les résultats récents obtenus sur les localisations chromosomiques de gènes de lipocalines humaines qui laissent supposer qu'ils dérivent d'un ancêtre commun.

### Les études de cristallographie aux rayons X

Les structures tridimensionnelles des cinq lipocalines déterminées à l'heure actuelle l'ont été par cristallisation et étude du diagramme de diffraction des rayons X. L'architecture particulière du complexe RBP-tout-trans rétinol, déterminée en 1984 par cristallographie aux rayons X [2], suggéra immédiatement que cette protéine pouvait être un représentant d'une nouvelle famille protéique. La structure est très particulière (figure 1): huit feuillets  $\beta$  successifs, disposés tête-bêche, formant deux parois d'un cône tronqué et aplati; deux entrées/sorties sont possibles mais l'une d'elles est bloquée par cinq cycles phénylalanine et une méthionine; les boucles qui connectent les feuillets sont courtes (deux à trois résidus) et la molécule de rétinol se niche au plus profond de cette cavité, entourée de résidus hydrophobes ou non chargés; du côté C-terminal se trouve une hélice  $\alpha$ . Le cycle  $\beta$  ionone est dans la partie la plus profonde de la poche et la chaîne isoprénique s'étire presque jusqu'à la surface de la protéine. La structure tertiaire est



Tableau I  
LIPOCALINES INDIVIDUELLES OU HOMOLOGUES ENTRE ESPÈCES

Protéine	Ligand	Distribution tissulaire	Sécrétion
<i>Retinol-binding protein</i> (RBP) (homme, bœuf, rat, lapin) [2, 6, 7]	rétinol	foie	plasma ; passage dans l'urine
<i>Odorant-binding protein</i> (OBP) (bœuf, rat, grenouille) [8-10]	molécules odorantes	épithélium olfactif, nasal	mucus nasal
Lipocaline lacrymale (homme) [11-15]	rétinol ?	glande lacrymale	larmes
Protéine de la glande de von Ebner (rat) [16]	molécules sapides ?	glandes de von Ebner	salive
Protéine placentaire PP14 (homme) [17, 18]	rétinol	placenta	liquide amniotique
$\beta$ lactoglobulin ( $\beta$ lac) (tous les mammifères sauf l'homme) [19, 20]	rétinol	glande mammaire	lait
$\alpha 2$ microglobuline (rat) [21, 22]	phéromones	foie	plasma, passage dans l'urine
MUP (souris) [21, 23]	phéromones	foie, glande lacrymale, glande salivaire	plasma, passage dans l'urine
Prostaglandine D synthétase (homme) [24, 25]	prostaglandine H2 bilirubine	cerveau	pas de forme sécrétée
Prostaglandine D synthétase (rat) [26]	prostaglandine H2	cerveau	pas de forme sécrétée
$\alpha 1$ acid glycoprotein (orosomucoïde) (homme, rat, souris) [27, 28]	progestérone, autres stéroïdes	foie	plasma
Protéine épидидymaire (rat, lézard) [29-31]	acide rétinolique	épididyme	fluide épидидymaire
Protéine fixant les pigments [32, 33]	biliverdine IX, biline	épiderme	hémolymphe
Crustcyanin Sous-unités C1 et A2 [34]	astaxanthine	carapace	?
$\alpha 1$ microglobuline (protéine HC) [35, 36]	rétinoïdes ?	foie	plasma
Chaîne C8 $\gamma$ du complément [37]	?	foie	plasma
Apolipoprotéine D [38, 39]	progestérone	foie	plasma
Probasine [40]	?	prostate : forme nucléaire et cytosolique	canaux prostatiques, prostate
24p3 [41]	?	rein	?
Ch21 [42]	?	tissu squelettique	matrice extracellulaire
Aphrodisine [43]	?	épithélium vaginal	mucus vaginal
Purpurine [44]	rétinol	rétine	interstitium rétinien
Protéine majoritaire des plexus choroïdes [45]	?	plexus choroïdes	fluide cérébrospinal

*Le cas de la protéine PP14 humaine et de la  $\beta$  lactoglobuline des autres espèces de mammifère est un peu particulier. La protéine PP14 est sécrétée chez la femme par l'endomètre, sous l'influence de la progestérone, durant le premier trimestre de la grossesse. La  $\beta$  lactoglobuline est sécrétée, sous contrôle de la progestérone, chez d'autres espèces de mammifère. La séquence des acides aminés de la PP14 présente 53 % de similitude avec ceux de la  $\beta$  lactoglobuline de cheval, ce qui est supérieur au pourcentage observé habituellement entre deux lipocalines non analogues. L'organisation des deux gènes est très voisine et ces deux protéines sont considérées comme analogues [35, 36].*

## RÉFÉRENCES

25. White DM, Mikol DD, Espinosa R, Weimer B, Le Beau M, Stefansson K. Structure and chromosomal localization of the human gene for a brain form of prostaglandin D2 synthase. *J Biol Chem* 1992; 267: 23202-8.

26. Igarashi M, Nagata A, Toh H, Urade Y, Hayaishi O. Structural organization of the gene for prostaglandin D synthase in the rat brain. *Proc Natl Acad Sci USA* 1992; 89: 5376-80.

27. Cooper R, Eckley DM, Papaconstantinou J. Nucleotide sequence of the mouse alpha-1-acid glycoprotein gene 1. *Biochemistry* 1987; 26: 5244-50.

28. Dente L, Ciliberto G, Cortese R. Structure of the human  $\alpha$ 1-acid-glycoprotein gene: sequence homology with other human acute phase protein genes. *Nucleic Acids Res* 1985; 13: 3941-52.

29. Girotti M, Jones R, Emery DC, Chia W, Hall L. Structure and expression of the rat epididymal secretory protein 1 gene: an androgen-regulated member of the lipocalin superfamily with a rare splice donor site. *Biochem J* 1992; 281: 203-10.

30. Brooks DE, Means AR, Wright EJ, Singh SP, Tiver KK. Molecular cloning of the cDNA for two major androgen-dependent secretory proteins of 18.5 kilodaltons synthesized by the rat epididymis. *J Biol Chem* 1986; 261: 4956-61.

31. Morel L, Dufaure JP, Depeiges A. The LESP, an androgen-regulated lizard epididymal secretory protein family identified as a new member of the lipocalin superfamily. *J Biol Chem* 1993; 268: 10274-81.

32. Huber R, Schneider M, Epp O, Mayr I, Messerschmidt A, Pflugrath J. Crystallization, crystal structure analysis and preliminary molecular model of the bilin binding protein from the insect *Pieris brassicae*. *J Mol Biol* 1987; 195: 423-34.

33. Holden HM, Rypniewski WR, Law JH, Rayment I. The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution. *EMBO J* 1987; 6: 1565-70.

34. Keen JN, Caceres I, Eliopoulos EE, Zagalsky PF, Findlay JBC. Complete sequence and model for the C(1) subunit of the carotenoprotein, crustacyanin, and model for the dimer, beta-crustacyanin, formed from the C(1) and A(2) subunits with astaxanthin. *Eur J Biochem* 1991; 202: 31-40.

Tableau II

A	
Hu RBP	ERDCRVSSFRVKENFDKARFS GTW YAMAKKDPEGLF
Bo RBP	ERDCRVSSFRVKENFDKARFA GTW YAMAKKDPEGLF
Rb RBP	ERDCRVSSFRVKENFDKARFA GTW YAMAKKDPEGLF
Rt RBP	ERDCRVSSFRVKENFDKARFS GLW YAIKKDPEGLF
B	
Hu LL	HHLLASDEEIQDVS GTW YLKAMTVDREFFMNLESV
Rt VE	FPTTEENQDVS GTW YLKAAAWDKVIPDKKFGS

**A** : Alignements de la partie N-terminale de la retinol-binding protein dans différentes espèces : homme (Hu RBP), bœuf (Bo RBP), lapin (Rb RBP), rat (Rt RBP). **B** : alignements de la partie N-terminale de la lipocaline lacrymale humaine (Hu LL) et de son homologue chez le rat (Rt VE). Un des fragments conservés (G-X-W), est encadré par deux lignes verticales.

Code à une lettre des acides aminés : A : Ala ; C : Cys ; D : Asp ; E : Glu ; F : Phe ; G : Gly ; H : His ; I : Ile ; K : Lys ; L : Leu ; M : Met ; N : Asn ; P : Pro ; Q : Gln ; R : Arg ; S : Ser ; T : Thr ; V : Val ; W : Trp ; Y : Tyr.

Tableau III

### ALIGNEMENTS DE LA PARTIE N-TERMINALE DE CINQ LIPOCALINES HUMAINES

1. Hu RBP	ERDCRVSSFRVKENFDKARFS GTW YAMAKKDPEGLF
2. Hu TL	HHLLASDEEIQDVS GTW YLKAMTVDREFF
3. Hu HC	GPVPTPPDNIQVQENFNISRIY GKW YNLAIGSTCPWL
4. Hu PP14	MDIPQTKQDLELPKLA GTW HSMAMATNNISL
5. Hu ApoD	QAFHLGKCPNPPVQENFDVNKYL GRW YEIE-KIPTTTFE

Le fragment le plus conservé est indiqué en rouge, encadré par les deux lignes verticales. RBP, retinol-binding protein ; Hu, humain ; TL, tear lipocalin ; HC, protéine HC ; PP14, protéine placentaire ; ApoD, apolipoprotéine D.

stabilisée par les liaisons covalentes de deux ponts disulfure.

L'année suivante, des études cristallographiques démontraient que la  $\beta$  lactoglobuline possédait une structure dans l'espace analogue à celle de la RBP [19]. Depuis lors, la structure de la RBP a été résolue avec une plus grande définition [6] et trois autres structures ont été déterminées : celle d'une protéine d'insecte, la *bilin-binding protein* [32], et celles de deux protéines sécrétées dans l'urine, l'une chez la souris et l'autre chez le rat [21]. Pour faciliter une comparaison structurale quantitative, Flower *et al.* [1] ont superposé manuellement, par modélisation sur ordinateur, les structures tridimensionnelles de ces lipocalines.

Les huit feuillets  $\beta$  sont bien conservés, liés entre eux par de courts motifs qui sont généralement très différents en conformation et en taille. Une trentaine de membres de la famille étant actuellement répertoriés, les structures tridimensionnelles de nombreuses lipocalines restent donc encore à établir.

**Des similitudes restreintes de séquences peptidiques sont retrouvées dans des espèces très éloignées**

Ces protéines, malgré leur diversité d'origine tissulaire et d'espèce, présentent une homogénéité de taille (150 à 190 acides aminés, soit une

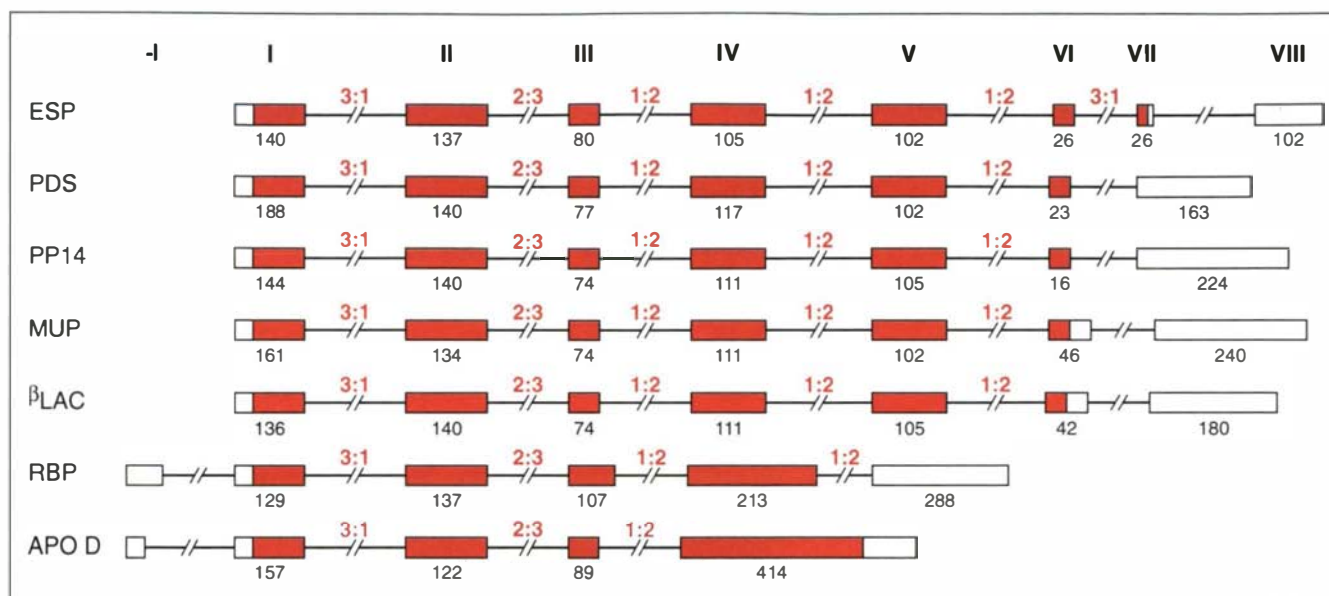


Figure 2. **Organisation exon/intron des gènes de différentes lipocalines.** Les exons sont représentés par des rectangles. Les zones codantes sont rouges, les zones non codantes sont claires. Les exons de la protéine sécrétoire épидидymaire (ESP) sont numérotés de I à VIII; ceux de la retinol-binding protein (RBP) de -I à IV. Les phases de lecture des introns, c'est-à-dire la position à l'intérieur du codon à laquelle l'intron interrompt l'exon, est notée de la manière suivante : 3:1 lorsque l'intron interrompt entre deux codons adjacents; 2:3 lorsque l'interruption se situe entre la deuxième et la troisième base d'un codon; 1:2 lorsque l'interruption se situe entre la première et la deuxième base d'un codon.

masse moléculaire de 17 000 à 25 000) et, au moins pour certaines, une structure tridimensionnelle commune. La comparaison par paire des protéines analogues chez l'homme et le rat indique un pourcentage d'identité élevé pour des protéines comme la RBP (supérieur à 86 % d'identité). Ce pourcentage est de 71 % pour la prostaglandine D synthétase et de 58 % pour la lipocaline lacrymale humaine et la protéine de la glande de von Ebner du rat. Cependant, les pourcentages d'analogies (comportant les pourcentages d'identité et les pourcentages de similarité) entre différentes lipocalines non analogues, obtenus par comparaisons par paire, sont faibles, ne dépassant pas 25 % à 30 %. Il faut noter qu'avec des pourcenta-

ges de similarité si faibles, les alignements peuvent perdre leur signification statistique et qu'il est difficile d'établir avec sûreté quels sont les acides aminés équivalents. Cependant, un pourcentage d'analogie faible est observé pour les membres de la famille dont les structures tridimensionnelles déterminées permettent de délimiter les feuillets β. Si l'on prend la numérotation des acides aminés de la RBP comme référence, les deux courts motifs protéiques qui caractérisent les membres de cette famille sont : Gly-X-Trp et Thr-Asp/Asn-Tyr (résidus 22-24 et 109-111 dans la RBP). Les ponts disulfures (entre Cys 4 et Cys 160, d'une part, et entre Cys 70 et Cys 174, d'autre part) ne sont pas présents dans toutes les lipocalines.

La famille présente les plus grandes variations dans la partie N-terminale (Tableau III).

### L'étude de l'organisation des gènes des lipocalines révèle des structures semblables

Les structures de trois gènes de lipocalines humaines et de quatre gènes de lipocalines d'autres espèces ont été comparées [20, 29, 24]. A l'examen attentif, des ressemblances apparaissent entre ces structures. Tout d'abord ces gènes sont relativement petits, de longueur toujours inférieure à 10 kilobases. Le morcellement le plus complexe semble être celui du gène de la protéine sécrétoire épидидymaire (ESP) : 8 exons



## RÉFÉRENCES

35. Lopez G, Grubb A, Soriano F, Mendez E. The complete amino acid sequence of human complex-forming glycoprotein heterogeneous in charge (protein HC). *Biochem Biophys Res Commun* 1984; 228: 544-54.
36. Salier JP, Simon D, Rouet P, Raguez G, Muscatelli F, Gebhard W, Guenet JL, Mattei MG. Homologous chromosomal locations of the four genes for inter- $\alpha$ -inhibitor and pre- $\alpha$ -inhibitor family in human and mouse: assignment of the ancestral gene for the lipocalin superfamily. *Genomics* 1992; 14: 83-8.
37. Ng SC, Rao AG, Howard OMZ, Sodetz JM. The eighth component of human complement: evidence that it is an oligomeric serum protein assembled from products of three different genes. *Biochemistry* 1987; 26: 5229-33.
38. Drayna D, Fielding C, McLean J, et al. Cloning and expression of human apolipoprotein D cDNA. *J Biol Chem* 1986; 261: 16535-9.
39. Simard J, deLaunoit Y, Haagensen DE, Labrie F. Additive stimulatory action of glucocorticoids and androgens on basal and estrogen-repressed apolipoprotein-D messenger ribonucleic acid levels and secretion in human breast cancer cells. *Endocrinology* 1992; 130: 1115-21.
40. Spence AM, Sheppard PC, Davie JR, et al. Regulation of bifunctional mRNA results in synthesis of secreted and nuclear probasin. *Proc Natl Acad Sci USA* 1989; 86: 7843-7.
41. Hrabá-Renevey S, Tuerler H, Kress M, Salomon C, Weil R. SV 40 induced expression of mouse gene 24p3 involves a post-transcriptional mechanism. *Oncogene* 1989; 4: 601-8.
42. Cancedda FD, Dozin B, Rossi F, Molina F, et al. The Ch21 protein, developmentally regulated in chick embryo, belongs to the superfamily of lipophilic molecule carrier proteins. *J Biol Chem* 1990; 265: 19060-4.
43. Henzel WJ, Rodriguez H, Singer AG, et al. The primary structure of aphrodisin. *J Biol Chem* 1988; 263: 16682-7.
44. Berman P, Gray P, Chen E, et al. Sequence analysis, cellular localization, and expression of a neuroretina adhesion and cell survival molecule. *Cell* 1987; 51: 135-42.
45. Achen MG, Harms PJ, Thomas T, Richardson SJ, Wettenhall REH, Schreiber G. Protein synthesis at the blood-brain barrier; the major protein secreted by amphibian choroid plexus is a lipocalin. *J Biol Chem* 1992; 267: 23170-4.
- séparés par 7 introns. Si, pour comparer des exons analogues, on numérote arbitrairement les exons du gène de l'ESP de I à VIII et ceux de la RBP de -I à IV, quelques caractéristiques communes sont notées. Les exons I ont des tailles voisines (de 129 à 188 paires de bases). Ils comportent tous une région 5' non codante; la partie codante code pour le peptide signal et la boucle N-terminale. Les exons II ont des tailles semblables (122 à 140 paires de bases) et codent pour les trois premiers feuilletts  $\beta$  du cône tronqué. Les exons III ont aussi des tailles du même ordre de grandeur (74 à 107 paires de bases) et correspondent approximativement aux feuilletts D et E. Pour les autres exons, les variations de taille sont plus importantes, la taille de certains exons pouvant se retrouver en supposant la disparition d'introns. Pour la RBP et l'apolipoprotéine D (Apo D), on trouve un exon (noté -I) en amont de l'exon I.
- La taille des différents introns est très variable, allant de 78 à 5 600 paires de bases pour l'intron 2 de la RBP et de l'apolipoprotéine D. Les phases de lecture des introns sont conservées à l'intérieur des membres de la famille; cela est mis en évidence dans la figure 2, où la position, à l'intérieur du codon, à laquelle l'intron interrompt l'exon est indiquée.

### **De nombreux gènes des lipocalines humaines ont une localisation chromosomique en regroupement (cluster)**

Les localisations chromosomiques des huit lipocalines humaines actuellement décrites, souvent obtenues par hybridation *in situ*, sont les suivantes. Le locus de la RBP, impliquée dans le transport du rétinol de son lieu de fabrication (le foie) jusqu'aux tissus cibles périphériques, est situé sur le chromosome 10. Le locus de l'ApoD est sur le chromosome 3. A côté de ces localisations dispersées sur le génome, des regroupements ont été observés. Les très récentes localisations chromosomiques des gènes de la prostaglandine D<sub>2</sub> synthétase (PDS) en

9q34.2-9q34.3 [40] et de la lipocaline lacrymale (LCN1) en 9q34.1-q34.3 [14] confirment en effet la présence d'un regroupement sur le bras long du chromosome 9. Quatre autres membres de la famille des lipocalines ont été repérés dans cette région: l'orosomucoïde (ORM, 9q32), la chaîne légère de l'inter  $\alpha$  trypsin inhibitor (ITIL, 9q32-q33), la protéine placentaire 14 ou protéine PP14 (PAEP, 9q34), la chaîne  $\gamma$  du facteur C8 du complément (C8GC, 9q). Les localisations homologues de gènes codant pour des lipocalines d'espèces très différentes sont en faveur de l'hypothèse d'une séquence « fondatrice »; une partie des séquences amplifiées serait restée groupée et une autre partie dispersée en divers loci du génome. Dans un travail récent, Salier et al. ont conforté cette hypothèse en démontrant la présence de synténies\* conservées entre des espèces aussi différentes que l'homme et la souris [36].

D'autres gènes ont été repérés dans la même zone, en particulier l'oncogène *c-abl* (ABL 9q34), le gène des groupes sanguins (ABO, 9q34.1-q34.2), le gène de la dopamine  $\beta$  hydroxylase (DBH, 9q34.3). Les loci de la sclérose tubéreuse de Bourneville (TSC1, 9q34.1-9q34.2), de *Xeroderma pigmentosum* (XPA, 9q32-q34.1) et de la dystonie musculaire déformante (9q34) sont groupés dans la même région; de plus, les études de positionnement indiquent que le gène de la prostaglandine D<sub>2</sub> synthétase est un gène candidat pour la dystonie musculaire déformante [25]. Chez l'homme, six des huit lipocalines connues jusqu'à présent sont donc localisées sur le bras long du chromosome 9. Cette localisation en regroupement, non loin des loci de la sclérose tubéreuse de Bourneville et de la dystonie musculaire déformante, donne beaucoup d'intérêt à l'étude de cartographie génétique et physique de cette région.

\* Synténie: localisation sur un même chromosome, soit de plusieurs loci différents d'une même espèce, soit d'un même locus dans différentes espèces.

## Des molécules à structures modélisées pour des fonctions précises

Le rôle exact de nombre de lipocalines connues ne sera prouvé qu'après examen attentif des relations structure-activité. Cette famille est un excellent sujet d'étude, tant les aspects qui peuvent être abordés sont divers et fondamentaux. La biologie structurale des macromolécules biologiques, autrement dit l'élucidation de leur structure tridimensionnelle, est un outil majeur dans tous les champs de la recherche biologique et médicale concernés par ces molécules. La connaissance à l'échelle atomique des structures tridimensionnelles des protéines permet une ouverture vers la compréhension des mécanismes d'interaction avec les ligands, d'interaction avec des agents pharmacologiques, et se révèle indispensable au remodelage (« *design* moléculaire ») des protéines. En effet, la recherche des relations structure-fonction des protéines est fondamentale pour, à terme, concevoir et proposer des molécules qui, tout en mimant l'activité d'une protéine, soient mieux adaptées que la macromolécule naturelle aux besoins médicaux ou industriels. A partir de la localisation et de l'orientation spatiale des acides aminés dont on sait qu'ils sont essentiels à la reconnaissance du ligand, les laboratoires d'ingénierie génétique ont trois types d'approche : (1) réaliser des mutations ponctuelles, délétions, substitutions, fusions de séquences au niveau d'un ADNc recombinant qui permettent d'opérer des changements variés au niveau des protéines produites dans des vecteurs d'expression et des cellules de plus en plus performants ; (2) prédire une « structure modèle » reposant sur l'accumulation de données qui proviennent de l'étude de familles de protéines et de leur évolution. Après avoir défini pour les membres de la famille les régions de structures conservées, celles-ci sont modélisées et l'emploi de la mécanique et de la dynamique moléculaires permet d'affiner le modèle ; (3) synthétiser totalement *de novo* des polypeptides permettant de tester des modèles simples ■

m/s n° 1 vol. 10, janvier 94

## Summary

### Lipocalins : hydrophobic-molecule transporters

Lipocalins are a group of proteins present in both vertebrates and invertebrates, sharing a common function: they transport hydrophobic ligands. There are three main criteria suggesting that they belong to the same family: (1) the three dimensional structure of five of them shows eight  $\beta$  strands which form a calyx (antiparallel  $\beta$  barrel) enclosing an internal ligand binding site; (2) they are predominantly small secreted proteins (about 200 amino-acid residues) and possess a few highly conserved amino-acid positions; (3) the gene structure of seven lipocalins presents surprising similarities; the size of the corresponding exons are about the same and the positions of exons/introns junctions are well conserved through species. Furthermore, six out of eight lipocalin genes, in human, have been localized to the long arm of chromosome 9. Given the tendency of the lipocalin genes to duplicate, it is possible that a large series of duplications took place from a lipocalin ancestral gene. Genes encoding lipocalins are expressed in a variety of tissues and, for all but two, the corresponding proteins are secreted in various fluids as blood serum, milk, saliva, tears, nasal mucus, seminal fluid and amniotic fluid. Their exact role will be proven only after extensive structure-activity relationship studies.

## TIRÉS A PART

A.-M. Gachon.