

4 - Recherche de gènes associés à des fonctions

La génomique fonctionnelle de la glande mammaire des ruminants

F. LE PROVOST, C. LEROUX, P. MARTIN

INRA, Laboratoire de Génétique Biochimique et de Cytogénétique,
78352 Jouy-en-Josas cedex

e-mail : Fabienne.LeProvost@biotec.jouy.inra.fr

Résumé. Dans la perspective de proposer de nouveaux critères d'amélioration génétique des caractères laitiers, il est essentiel d'identifier l'ensemble des gènes ayant un effet majeur sur l'expression de ces caractères et de rechercher ensuite une éventuelle relation entre le polymorphisme de ces gènes et les performances individuelles. La caractérisation fonctionnelle des gènes transcrits dans la glande mammaire des ruminants a été entreprise, dans notre équipe, par deux approches : l'approche globale, qui permet de répertorier les gènes exprimés dans ce tissu et d'estimer leur niveau d'expression, et l'approche différentielle, qui consiste à identifier les gènes dont l'expression est spécifiquement induite par un événement (changement de stade physiologique ...).

En raison de son importance dans l'alimentation humaine et des enjeux socio-économiques que suscitent sa production et sa transformation, le lait a fait l'objet de nombreuses recherches notamment sur la plan de sa biosynthèse, de sa composition et de ses propriétés technologiques. Si l'analyse structurale de ses constituants protéiques et des gènes qui les spécifient a connu un développement important (voir l'exemple des caséines), on connaît encore relativement mal les mécanismes qui concourent à la différenciation et au fonctionnement du tissu mammaire.

La lactation, qui correspond à la synthèse et à la sécrétion massive des constituants du lait (galactopoïèse), est l'aboutissement d'une succession de transformations qui affectent le tissu mammaire dès le début de la gestation (mammogenèse), et se poursuivent jusqu'à la parturition (lactogenèse) sous contrôle multihormonal (Jammes et Djiane 1988). De fait, cet organe est le siège d'une intense activité transcriptionnelle, traductionnelle et sécrétoire particulièrement efficace, mettant en jeu une multiplicité de gènes parmi lesquels doivent être recherchés des candidats sous-tendant la variabilité de caractères qui déterminent le potentiel laitier des animaux.

Pour proposer de nouveaux critères d'amélioration génétique des caractères laitiers, il est donc essentiel d'identifier ces gènes, en particulier ceux ayant un effet majeur sur l'expression de ces caractères. Ceux-ci devront faire ensuite l'objet d'une analyse structurale plus détaillée, destinée à mettre en évidence une éventuelle relation entre polymorphismes et performances individuelles.

La caractérisation fonctionnelle des gènes transcrits dans la glande mammaire des ruminants a été

entreprise par notre équipe. Le génome d'une cellule de mammifère contient entre 50 000 et 100 000 gènes différents, mais seuls 10 000 à 20 000 sont exprimés dans un type cellulaire donné. Les niveaux d'expression de ces gènes varient de 1 000 à 10 000 copies d'ARN messenger par cellule.

Nos objectifs consistent, à la fois, à répertorier l'ensemble des gènes exprimés dans la glande mammaire et estimer leur niveau d'expression (approche globale) et à identifier les gènes dont l'expression est spécifiquement induite par un événement (changement de stade physiologique, maladies, etc) (approche différentielle). L'approche globale a été abordée par des stratégies de séquençage systématique d'extrémités d'ADNc (EST : Expressed Sequence Tag) pour établir un premier profil transcriptionnel. L'approche différentielle a été abordée en utilisant la technique de 'RNA differential display', seule technique, à l'époque, qui permettait la comparaison de plus de deux échantillons simultanément. Avec l'émergence des techniques de filtres à haute densité ou cDNA arrays, un seul et même outil peut répondre aux deux questions (global et différentiel). C'est dans cet optique que le programme Roger-1 (pour Répertoire ordonné de gènes exprimés et régulés) a été développé au laboratoire.

1 / L'approche globale

Un premier travail a consisté à inventorier les principaux transcrits synthétisés dans la glande mammaire de chèvres en lactation. Mille clones issus d'une banque d'ADNc synthétisée à partir d'ARN messagers extraits du tissu mammaire d'une chèvre lactante ont été isolés. Après soustraction des ADNc spécifiant les 6 lactoprotéines majeures (caséines α S1, α S2, β et κ , α -lactalbumine et β -lac-

toglobuline), qui représentent environ deux tiers des ARNm, le séquençage systématique de 435 clones, générant des ESTs, a permis d'identifier rapidement un nombre significatif de gènes exprimés dans ce tissu, soit 217 gènes dont 77 « inconnus » (Le Provost *et al* 1996).

Parmi l'ensemble des ADNc obtenus, l'expression de bon nombre d'entre eux était attendue, en raison des activités de synthèse et de sécrétion de cet organe. Parmi ceux-ci, citons des gènes dont les produits d'expression participent à la synthèse de protéines (protéines ribosomales, facteur d'élongation $\alpha 1$), des gènes spécifiant des protéines quantitativement mineures du lait (lactoferrine, lactopéroxydase, butyrophiline, lactophorine) ou encore des enzymes intervenant dans la lipogenèse (synthase des acides gras, stéaroyl-CoA désaturase). Nous avons également détecté la présence de transcrits dont l'expression n'avait pas encore été décrite au niveau mammaire (transcrit spécifiant la protéine PrP ou Prion) ou dont la présence est pour le moins inattendue (séquence d'insertion d'un élément génétique mobile procaryote : *Tn10*).

Cette approche nous a permis non seulement de mieux définir l'activité transcriptionnelle globale de la glande mammaire en lactation, mais aussi de disposer, pour chaque gène répertorié, d'outils moléculaires qui ont facilité et faciliteront leur analyse génétique (expression tissulaire, localisations cellulaire et chromosomique, analyse de polymorphisme ...). Certains ADNc ont été utilisés pour réaliser des études plus détaillées de gènes spécifiant des protéines constitutives du lait (lactoferrine : Le Provost *et al* 1994, lactophorine, lactopéroxydase) ou pour étudier les transcrits correspondant à des enzymes-clés de la lipogenèse (synthase des acides gras) dans le cadre de l'analyse génétique de la lipogenèse mammaire.

L'expression tissulaire de 18 gènes « inconnus » a été analysée dans trois tissus. Deux de ces gènes, révélant une expression préférentielle dans la glande mammaire, ont été plus finement étudiés. Leur expression a été analysée sur un plus grand nombre de tissus et à différents stades physiologiques du tissu mammaire. L'un des deux est détecté uniquement dans le tissu mammaire pendant la lactation. L'obtention de leurs ADNc complets (à l'aide de la technique de 5' RACE), puis leur séquençage a conduit à l'identification d'un cadre de lecture ouvert. L'un présente une similarité de séquence significative avec une protéine de la famille des « Heat-Shock » (protéines impliquées dans le repliement, l'assemblage et le transport de complexes protéiques), tandis que l'autre a une similarité de séquence avec un transporteur de cations organiques.

Par ailleurs, cette approche a permis la localisation chromosomique de gènes sur le génome bovin à l'aide d'un panel d'hybrides somatiques interspécifiques (hamster/bovin) (Le Provost *et al* 1996) et sur le génome caprin par la technique de FISH à l'aide de BAC (Le Provost *et al* 2000).

Actuellement, ce premier travail systématique est complété par l'établissement de profils transcriptionnels du tissu mammaire de chèvre à l'aide de la technique SAGE (Serial Analysis of Gene Expression, cf paragraphe 3).

2 / L'approche différentielle

Nous nous sommes attachés, en utilisant une autre stratégie, à mettre en évidence des gènes intervenant dans la différenciation terminale de la cellule épithéliale mammaire : étape-clé au cours de laquelle les cellules acquièrent les structures permettant une synthèse et une sécrétion intenses, et qui conditionne largement le niveau de production mais aussi de qualité du lait. La recherche de ces gènes différentiellement exprimés par comparaison de profils d'expression a été réalisée à l'aide de la technique de RNA differential display, qui s'appuie sur des techniques simples (transcription inverse, PCR, électrophorèse en gel de polyacrylamide dénaturant) et qui permet de comparer plusieurs échantillons simultanément.

Ces gènes sont recherchés en comparant les profils d'expression d'un modèle cellulaire, à savoir un clone stable de la lignée de cellules épithéliales mammaires murines (HC11) surexprimant le récepteur de la prolactine de lapin (établi par J. Paly, INRA Unité de Biologie cellulaire et moléculaire, Jouy-en-Josas). Ces cellules, stimulées ou non par la prolactine, permettent d'étudier les gènes précocement induits par cette hormone qui est essentielle à la différenciation terminale (collaboration avec M. Charlier et J. Djiane, INRA Unité de Biologie cellulaire et moléculaire, Jouy-en-Josas).

La comparaison des ARNm extraits du clone HC11 stimulé ou non par la prolactine pendant différentes durées de temps, nous a permis d'isoler plusieurs fragments d'ADNc. Leur expression a ensuite été étudiée sur un modèle *in vivo* : glandes mammaires de souris à différents stades de gestation et de lactation. Ce travail a permis de mettre en évidence un gène (spécifiant l'histone H2AZ) dont l'expression forte pendant la gestation diminue significativement en période de lactation. L'homologue de cette protéine chez la drosophile semble jouer un rôle essentiel au cours du développement (van Daal *et al* 1992). D'autres travaux (Wu *et al* 1982) sur lignée de cellules d'ovaire de Hamster en culture, ont montré que le taux de transcrits H2AZ est plus faible dans des cellules quiescentes que dans des cellules en prolifération. Ces résultats suggèrent que ce gène pourrait intervenir dans le processus de différenciation terminale de la cellule épithéliale mammaire. Nous avons donc entrepris de mieux caractériser ce transcrit en étudiant sa localisation cellulaire par hybridation *in situ* et la régulation de son expression chez des brebis vierges chez lesquelles une lactation artificielle a été induite par un traitement hormonal approprié.

Les recherches de gènes différentiellement exprimés ont été complétées par l'étude d'un modèle biologique plus complexe, c'est-à-dire en comparant des glandes mammaires de chèvres à différents stades physiologiques (mi-gestation, fin de gestation et lactation). En ce qui concerne ce modèle, la majorité des bandes différentielles obtenues contient des transcrits spécifiant les six lactoprotéines majeures qui, en raison de leur abondance, masquent en partie les autres différences que nous recherchions. Cette approche nous a tout de même permis de détecter une dizaine d'ADNc. Leur expression différentielle reste toutefois à confirmer par d'autres techniques telles que Northern blot et RT-PCR semi-quantitative.

Pour compléter ce travail d'approche différentielle et en raison des difficultés rencontrées avec la technique RNA differential display pour la comparaison des glandes mammaires de chèvres, nous avons développé de nouvelles stratégies (filtres à haute densité, SAGE).

3 / Le projet Roger-1

Pour, à l'aide d'un même outil, établir des profils transcriptionnels et identifier des gènes impliqués dans l'expression de caractère d'intérêt, nous avons entrepris, en collaboration avec trois équipes (2 INRA et 1 CNRS) de produire un premier Répertoire Ordonné de Gènes Exprimés et Régulés (Roger-1). Cette action, initiée en 1997, concerne l'espèce bovine et notre intérêt s'est d'abord porté sur trois tissus : l'embryon (équipe de J.-P. Renard, Jouy-en-Josas), le muscle (équipe de J.-F. Hocquette, Theix) et la glande mammaire (notre équipe). L'équipe CNRS de C. Auffray et G. Piétu assure le soutien logistique et nous fait bénéficier de son savoir-faire.

Au cours de la première année, notre objectif était, d'une part, de produire l'outil : trois membranes (une par tissu) sur lesquelles 3 600 ADNc (1 200 par tissu) seraient ordonnés. D'autre part, nous souhaitons utiliser ces filtres pour analyser différentes sondes complexes, correspondant aux tissus étudiés et représentatives de différents stades physiologiques et de développement.

En ce qui concerne le tissu mammaire, un filtre sur lequel ont été ordonnés 400 ADNc a été réalisé. Les produits de PCR déposés sur ce filtre ont été obtenus à partir des clones de la banque d'ADNc de glande mammaire de chèvre (voir plus haut), des ADNc correspondant à des gènes connus recherchés dans la banque génomique caprine de BAC (Schibler *et al* 1998) et des gènes localisés pour les programmes de cartographie comparée (projet ComparEST, P. Laurent).

Ces filtres ont été criblés avec des sondes complexes correspondant aux ARNm de glandes mammaires de chèvres en gestation (milieu et fin) et en lactation. Ce travail a permis d'étudier l'expression de ces 400 gènes dans la glande mammaire aux trois stades physiologiques. Les données analysées au moyen du logiciel développé par l'équipe du CNRS ont révélé, notamment, une surexpression, au cours de la lactation, d'une série de gènes au premier rang desquels figurent fort logiquement ceux spécifiant

les caséines, l' α -lactalbumine, deux protéines de la membrane des globules gras (lactophorine et butyrophiline), mais aussi trois EST dont la fonction est pour l'heure inconnue, ainsi que quelques gènes connus (antigène CD36 et sous-unité régulatrice de la calpaïne II).

Pour passer de 400 à 1 200 « cibles », nous avons mis en œuvre une stratégie comportant deux volets :

- la construction d'une banque d'ADNc mammaire normalisée (étape qui permet de minimiser les variations du nombre de copies de chaque transcrit) afin d'isoler 30 000 clones. Ces clones seront repiqués en plaques de 96 puits et organisés en pools et superpools selon un schéma de 'pooling 3D' permettant l'adressage, par PCR, d'un clone parmi l'ensemble. Idéalement, 1 000 clones seront séquencés de façon partielle de manière à les identifier et de poursuivre l'inventaire transcriptionnel mammaire ;
- le séquençage systématique de deux banques SAGE (glandes mammaires en milieu et en fin de gestation). Cette technique permet de générer des séquences courtes de 14 nucléotides (= tags) qui constituent des signatures transcriptionnelles de gènes exprimés dans un tissu. L'obtention d'un grand nombre d'étiquettes (environ 15 000) permettra d'estimer le niveau d'expression des transcrits, dont l'identification pose toutefois encore quelques problèmes. La comparaison des profils transcriptionnels de ces deux banques permettra de détecter les gènes différemment exprimés entre ces deux états physiologiques. Les ADNc complets correspondant aux tags pertinents seront recherchés par PCR dans la banque normalisée mammaire.

Conclusion

En identifiant les gènes impliqués dans le développement et le fonctionnement du tissu mammaire, nous contribuons à développer les bases génétiques qui devraient permettre, à terme, de disséquer les mécanismes gouvernant le fonctionnement mammaire et fournir les outils nécessaires à la prise en compte de nouveaux critères de sélection. Ces travaux qui nous ont amenés à créer et utiliser différents outils vont être poursuivis dans le cadre du programme Roger-2 et complétés par le programme Asteroger (Hatey *et al* 2000, cet ouvrage). Ce dernier a pour but de construire et d'exploiter, pour chacune des principales espèces animales d'élevage (vache, porc, poulet et truie) un répertoire ordonné de gènes correspondant à une fraction significative de leur transcriptome.

Références

- Hatey F., Martin P., Douaire M., Le Gac F., Dambrine G., Herpin P., Monget P., 2000. Le programme Asteroger : vers un outil multifonctionnel pour les productions animales. INRA Productions Animales, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 175-180.
- Jammes H., Djiane J., 1988. Le développement de la glande mammaire et son contrôle hormonal dans l'espèce bovine. INRA Productions Animales, 1, 299-310.
- Le Provost F., Nocart M., Guérin G., Martin P., 1994. Characterization of the goat lactoferrin cDNA: assignment of the relevant locus to bovine U12 synteny group. Biochemistry Biophysics Research Communications, 203, 1324-1332.
- Le Provost F., Lépling A., Martin P., 1996. A survey of the goat genome transcribed in the lactating mammary gland. Mammalian Genome, 7, 657-666.
- Le Provost F., Schibler L., Oustry-Vaiman A., Martin P., Crihiu E.P., 2000. Cytogenetic mapping of 25 goat mammary gland ESTs. soumis pour publication.
- Schibler L., Vaiman D., Oustry A., Guinec N., Dangy-Caye A.-L., Billault A., Crihiu E.P., 1998. Construction and extensive characterization of a goat Bacterial Artificial Chromosome library with threefold genome coverage, Mammalian Genome, 9, 119-124.
- van Daal A., White E.M., Gorovsky, M.A., Elgin S.C.R., 1992. A histone variant, H2AvD, is essential in *Drosophila melanogaster*. Molecular Cell Biology, 3, 593-602.
- Wu R.S., Tsai S., Bonner W.M., 1982. Patterns of histone variant synthesis can distinguish G0 from G1 cells. Cell, 31, 367-374.