



médecine/sciences 2000 ; 16 : 950-3

## Jusqu'où iront les puces ?

Les mesures d'expression grâce à l'emploi de réseaux d'ADN continuent à avoir le vent en poupe. Des articles majeurs montrent l'apport de ces méthodes dans des questions de recherche fondamentale [1] ou clinique [2]. Les colloques spécialisés fleurissent, tandis que l'offre commerciale d'*oligo-chips* et de *microarrays* prêts à l'emploi commence à s'élargir. De nombreux laboratoires français sont maintenant équipés – quoique pas toujours opérationnels, la mise en œuvre de ces méthodes se révélant en général plus laborieuses que prévu.

### Un rapide survol de la situation actuelle

Les différentes approches en présence sont maintenant bien connues [3,4]. Rappelons-les brièvement avant de passer à l'exercice de prospective scientifico-technique qui est l'objet de cette chronique. Dans tous les cas, la mesure d'expression repose sur l'hybridation entre un mélange d'ADNc marqués préparé à partir des ARN messagers de l'échantillon biologique, et un ensemble de segments d'ADN de séquence connue (appelés tantôt « cibles », tantôt « sondes » selon les auteurs) fixé sur un support. Après incubation et lavage, le taux d'hybridation est quantifié pour chaque élément du réseau. Il est proportionnel à l'abondance relative de l'espèce d'ARNm concernée dans l'échantillon de départ, et donne donc une mesure du niveau d'expression du gène correspondant.

La première famille de méthodes repose sur l'emploi de produits de PCR obtenus à partir de jeux de clones d'ADNc représentant chacun un gène. Elle se décline d'abord en *macroarrays* (ou « filtres à haute densité »), réalisés sur des membranes de nylon d'une centaine de cm<sup>2</sup> de surface ; dans ce cas l'échantillon est presque toujours marqué de manière radioactive. Cette approche, la plus ancienne, reste assez performante surtout pour des jeux de gènes pas trop nombreux et pour des échantillons peu abondants, et offre l'avantage d'être compatible avec le matériel déjà présent dans les laboratoires. Les *microarrays*, version miniaturisée de cette approche, sont généralement réalisés sur verre ou sur silicium, et peuvent porter une dizaine de milliers d'éléments sur quelques centimètres carrés ; l'échantillon est dans ce cas marqué directement ou indirectement par fluorescence. Ces réseaux miniature peuvent aussi être réalisés sur nylon et révélés par colorimétrie, ou par radioactivité avec un lecteur à haute résolution. Dans ce dernier cas, leur sensibilité permet la mesure à partir de très petits échantillons, ce qui est un avantage déterminant dans certaines situations [5]. *Macroarrays* et *microarrays* peuvent être confectionnés au laboratoire (pour autant que celui-ci dispose des robots *ad hoc* et de la compétence nécessaire pour les employer) ou achetés (fort cher) à des fabricants pour le moment peu nombreux. Notons qu'à part les réseaux sur support nylon, ces *microarrays* sont à

usage unique.

Le deuxième type d'approche, proposé pour le moment presque exclusivement par la firme Affymetrix, emploie des centaines de milliers d'oligonucléotides synthétisés directement sur la « puce » pour effectuer le même travail. Le relatif manque de spécificité des 20- ou 25-mères utilisés est compensé par une redondance élevée, chaque gène étant représenté par une vingtaine d'oligonucléotides auquel s'ajoutent autant de témoins présentant une différence de séquence et permettant l'évaluation du bruit de fond. La fabrication de ces *oligo-chips* réclame une technicité et des investissements hors de portée des laboratoires de recherche : la méthode photolithographique employée requiert, pour une puce comportant des 20-mères, quatre-vingts (20 x 4) masques permettant l'illumination sélective des différentes régions de la puce au cours des étapes de synthèse. Les *oligo-chips* sont donc obligatoirement achetés, à un prix unitaire de l'ordre de 5 000 à 10 000 – et l'expérimentateur s'engage à ne même pas essayer de les réutiliser...

Chacune de ces méthodes a ses avantages et ses inconvénients, en termes de performances, de coût, de commodité... Mais la situation est loin d'être figée, et des évolutions en cours permettent de penser que les procédures employées dans deux ou trois ans seront assez différentes de celles d'aujourd'hui. Aussi dans cette chronique vais-je me risquer – le jeu est dangereux... – à indiquer

les directions dans lesquelles il me semble que l'analyse du transcriptome va s'engager dans un futur relativement proche.

### Vers des puces « génome entier » ?

Avec l'obtention de la séquence humaine, le désir de disposer de réseaux permettant d'analyser simultanément l'expression de l'ensemble des gènes humains (ou murins) va logiquement s'exacerber, la disponibilité tant de la séquence que de grands jeux de clones facilitant en principe une telle réalisation (déjà effective dans le cas des 6400 gènes de la levure). Il faut pourtant souligner que la technologie actuelle n'a pas encore atteint le degré de miniaturisation permettant de représenter l'ensemble des gènes humains\* sur une surface inférieure à celle d'une lame de microscope. Les *microarrays* les plus denses comportent actuellement de l'ordre de 2 000 dépôts par cm<sup>2</sup>, soit environ 20 000 sur la surface utile d'une lame de microscope – sans compter les duplicats et les contrôles. Il s'agira donc plutôt d'un jeu de quatre ou cinq lames représentant l'ensemble des gènes humains – ce qui est déjà une performance remarquable. En fait, l'augmentation de cette densité ne poserait pas de problème insurmontable du point de vue de l'analyse puisque la résolution des systèmes confocaux employés pour la lecture atteint 10 ou même 5 micromètres. Il paraît toutefois difficile de réaliser en routine des dépôts d'ADN dont le diamètre soit inférieur à 100 micromètres\*\*, amenant en pratique à un « pas » de 200 micromètres qui correspond à la densité mentionnée plus haut. Des changements dans les systèmes de dépôt, et la mise au point de nouveaux types de surfaces, pourraient accroître cette densité. Il semble pourtant que la complexité et le coût de la prépara-

tion de dizaines de milliers de produits de PCR risquent en tout état de cause de freiner la diffusion de *microarrays* « génome entier ».

La situation n'est pas très différente actuellement pour les puces à oligonucléotides, mais elle est sans doute plus susceptible d'évolution. Les puces aujourd'hui commercialisées par Affymetrix portent environ 300 000 oligonucléotides. Néanmoins, la faible longueur de ceux-ci (conséquence du médiocre rendement des réactions photochimiques employées) impose d'en consacrer trente à quarante (contrôles compris) pour mesurer l'expression d'un seul gène, ce qui ramène le nombre de gènes analysables à moins de dix mille. Pour aller plus loin, il faudrait accroître la densité des puces et/ou réduire le nombre d'oligonucléotides par gène. Un meilleur choix des séquences de ces derniers, grâce à des progrès dans les algorithmes employés, peut permettre d'en réduire l'effectif – mais cette réduction ne sera que modérée afin de ne pas affecter la qualité de la mesure. L'augmentation du nombre d'éléments par puce est possible: les « plots » des puces Affymetrix sont actuellement de 24 par 24 microns, alors que l'industrie des microprocesseurs est descendue bien au-dessous du micromètre. Des puces comportant un ou plusieurs millions d'éléments sont donc possibles, mais c'est la lecture optique qui va alors atteindre ses limites en résolution et en sensibilité. De nouvelles techniques de synthèse *in situ* (voir plus loin) permettront sans doute de dépasser ces obstacles.

En tout état de cause, et au moins dans le futur immédiat, d'éventuels *microarrays* ou *oligo-chips* « génome entier » (humain ou murin) seront certainement très onéreux. De plus, les problèmes d'acquisition, de stockage et d'interprétation des résultats seront importants, compte tenu de la masse de données. De ce fait, l'emploi de réseaux d'ADN spécialisés représentant des jeux limités mais pertinents de gènes devrait rester tout à fait attractif.

### La victoire prévisible des oligonucléotides

Les nouveaux venus au monde des

*microarrays*, une fois qu'ils ont réuni les robots et les systèmes de lecture indispensables, découvrent souvent à leurs dépens la difficulté et le coût de la constitution de collections de milliers de clones d'ADNc validés, tout comme celui qu'entraîne la production de produits de PCR en qualité et quantité suffisantes. Divers problèmes (les erreurs dans la collection IMAGE, la contamination de nombreux clones par des bactériophages, sans parler des questions de propriété industrielle) compliquent encore la tâche. Sans doute pour les mêmes raisons, il existe encore peu de *microarrays* commercialement disponibles. De ce côté, le statut un peu ambigu des clones IMAGE\*\*\* tout comme la menace de procès de la part d'Affymetrix jouent sûrement un rôle de frein. En tout état de cause, les *microarrays* sont aujourd'hui chers (au moins 10 000 francs pour 2 000 gènes), et il est probable qu'ils le restent: les économies d'échelle que l'on peut attendre de la fabrication de milliers de *microarrays* sont modestes, puisque les coûts, tant pour l'obtention des produits de PCR que pour l'opération de dépôt, sont en gros proportionnels au nombre de réseaux produits.

Les puces à oligonucléotides ne souffrent pas des mêmes problèmes. Fondées sur la connaissance de la séquence (dont le rythme d'accumulation ne cesse de s'accroître), elles éliminent tout recours aux collections de clones d'ADNc. De plus, les économies d'échelle dans la fabrication peuvent être considérables. La firme Affymetrix pratique, à l'aide de machines automatiques, une série de 80 (20 x 4) cycles d'addition de bases afin de construire des jeux d'oligonucléotides (20-mères) sur des « gaufrettes » (*wafers*) de verre à partir des-

\*\*\* Les clones IMAGE sont librement accessibles et distribués à tout laboratoire moyennant une simple participation aux frais de stockage et d'envoi. Leur utilisateur s'engage en principe à rendre public tout résultat obtenu grâce à leur utilisation, ce qui rend par exemple difficile pour une entreprise la vente à des industriels de réseaux à façon réalisés à partir de clones IMAGE, ces firmes ne souhaitant généralement pas publier leurs résultats. D'autres collections de clones présentent également des restrictions d'usage qui rendent leur emploi pour la fabrication de *microarrays* commerciaux difficile.

\* Ce nombre est très discuté, puisque les estimations actuelles vont de moins de 30 000 à 120 000. Je raisonne ici sur un nombre d'environ 50 000.

\*\* Un dépôt d'un tel diamètre correspond à environ un nanolitre de solution. Les têtes des imprimantes à jet d'encre sont capables de projeter des gouttelettes bien plus petites, quelques picolitres; mais une solution d'ADN n'a pas la fluidité des encres spéciales mises au point pour cette application.

quelles sont ensuite découpées les puces individuelles. Les premières fabrications permettaient ainsi l'obtention, en quelques heures, de 49 *oligo-chips* à partir d'un *wafers*; aujourd'hui, un jeu de 400 petites puces à très haute densité – portant chacun autant, ou même plus, d'oligonucléotides différents que les anciennes – est produit durant la même période à partir d'une gaufrette de même taille. Les économies d'échelle correspondantes n'apparaissent pas de manière évidente dans le catalogue de la compagnie, mais cette situation changera sûrement si la concurrence s'intensifie.

De fait, plusieurs laboratoires et entreprises mettent actuellement au point la production d'oligonucléotides *in situ* par des approches différentes, reposant par exemple sur l'addition très rapide de réactifs aux différents sites de la puce. Ces additions sont réalisées grâce à des systèmes proches des têtes d'imprimante à jet d'encre. De telles procédures utilisent la chimie « classique » de synthèse d'ADN, dont l'excellent rendement permet d'envisager la synthèse de grands oligonucléotides, 50 ou même 100-mères (Edwin Southern, communication personnelle). La spécificité d'hybridation est alors telle que le niveau d'expression d'un gène peut être mesuré avec un tout petit nombre de plots (ou même un seul) : du coup une puce de dimensions raisonnables pourra représenter un grand nombre de gènes. De surcroît, cette approche est beaucoup plus flexible que celle d'Affymetrix. La fabrication d'une puce « à façon » requiert simplement une programmation différente de la tête d'impression, et non la confection d'une série de 80 masques destinés à cibler la synthèse par voie photochimique.

Le développement de ces technologies ne dépend pas seulement de facteurs scientifiques et techniques. La question des brevets est déjà très « chaude » dans ce domaine. La firme Affymetrix en détient une mul-

titude\* qui, selon son interprétation, lui permettent de s'opposer à la vente de pratiquement tout type de réseau d'ADN. Ces brevets sont néanmoins attaqués, notamment par Edwin Southern qui le premier décrivait l'emploi de « multitudes d'oligonucléotides attachés à des lames de verre » lors d'une communication au symposium de Cold Spring Harbor en 1989 [6]. Ce conflit – déjà pris en main par une pléiade d'avocats – sera, espérons-le, tranché d'une manière qui encourage la concurrence...

Il est en tout cas très probable que les puces à ADN à base d'oligonucléotides, fondées uniquement sur la connaissance de la séquence, seront de plus en plus utilisées dans le futur. Cela sera certainement le cas pour les réseaux « standard », dont l'exemple actuel est le jeu complet de gènes de la levure. Le règne des oligonucléotides pourra s'étendre au-delà, pour les puces spécialisées, à condition que les nouvelles méthodes de fabrication flexibles tiennent leurs promesses, facilitent la confection de réseaux sur mesure et en réduisent le coût.

### De l'artisanat à l'industrie

Les puces à oligonucléotides discutées dans le paragraphe précédent, nécessitant outillage et compétence spécialisés, ne seront pas confectionnées par les laboratoires de recherche mais achetées auprès d'industriels. Même pour les *microarrays* fondés sur l'emploi de clones d'ADNc et de produits de PCR, une tendance vers l'emploi de produits commerciaux se fera sentir. Il n'est pas rentable pour des équipes ou même des instituts de recherche d'investir des ressources importantes pour construire des réseaux standard : une fois encore, le cas de la Levure est un bon exemple. Une telle tâche peut être assurée de manière plus efficace par l'industrie ou, dans certains cas, par des centres de ressources publics. Ce n'est pas pour autant que la fabrication de *microarrays* va disparaître des laboratoires : des réseaux à façon permettant de tester des jeux limités mais choisis de gènes resteront nécessaires dans beaucoup de situations expé-

mentales, et la flexibilité maximale sera atteinte en les réalisant au laboratoire. Il existera sûrement d'autres possibilités : certains fabricants disposant d'une technologie flexible pourront offrir la fabrication de jeux *ad hoc*, d'autres vendront des jeux de produits de PCR « prêts à déposer ». La situation finale sera sans doute complexe, les jeux standard étant réalisés industriellement tandis que les réseaux spécialisés seront produits dans le cadre de différents arrangements entre le monde académique et l'industrie. Dans ce contexte, il est à l'évidence vital de standardiser les formats et les systèmes de détection afin que chaque réseau produit par l'industrie n'impose pas l'emploi de son système de lecture « propriétaire » coûtant au bas mot un demi-million de francs...

### Du réseau isolé à la « puce-laboratoire »

La technologie des puces biologiques n'est pas limitée aux réseaux d'ADN. L'intégration de diverses fonctionnalités sur des *chips* dont les dimensions se mesurent en centimètres est bien engagée : ces systèmes peuvent gérer des liquides, effectuer des filtrations, des réactions de PCR et même réaliser des électrophorèses capillaires [7]. Leur mise au point est fortement poussée par les besoins des industries pharmaceutiques qui doivent aujourd'hui tester littéralement des millions de composés issus de la chimie combinatoire (*high throughput screening*). Il leur faut effectuer ces essais très rapidement, de manière massivement parallèle et en utilisant le volume minimum de réactifs afin de limiter les coûts – d'où l'intérêt évident de la miniaturisation. A terme, et au moins pour les applications cliniques et industrielles, les mesures d'expression – portant probablement sur un nombre limité de gènes – auront vocation à être intégrées dans de tels systèmes. C'est par exemple la forme sous laquelle cette approche pénétrera dans les services d'oncologie clinique – si du moins les indications montrant l'intérêt clinique de telles données se confirment [2].

\* Une interrogation du site du US Patent and Trademark Office (<http://www.uspto.gov>) sur les mots clefs Affymetrix et oligonucléotide fait apparaître pas moins de 58 brevets...

## Du marquage préalable à la détection électrique

Le marquage fluorescent de l'échantillon est relativement complexe, interfère avec l'hybridation et impose des systèmes de détection très sensibles et coûteux; l'emploi de la radioactivité pose problème dans de nombreux environnements et sa résolution reste limitée même avec les détecteurs les plus performants (et les plus chers). Il est souhaitable de pouvoir détecter l'hybridation, et d'en quantifier le degré, à l'aide d'une autre méthode. De préférence, ceci devrait passer par la mesure d'un signal électrique et, dans l'idéal, n'imposerait aucune modification de l'échantillon avant l'expérience. De nombreux groupes déploient des efforts dans cette direction [8, 9]. Les méthodes envisagées vont de la détection d'un changement subtil des propriétés électriques après hybridation jusqu'à des approches très « exotiques » comme l'emploi de microbalances « pesante » la masse supplémentaire du matériel hybridé, ou la détermination du nombre de molécules double-brin (donc hybridées) par microscopie à force atomique. La faisabilité de certaines de ces méthodes a été démontrée; reste à savoir si elles pourront atteindre la sensibilité requise et le débit nécessaire. Il est probable que leurs premiers emplois se situeront dans des applications comme la détection de bactéries ou de mutations, pour lesquelles une réponse qualitative peut suffire; elles seront éventuellement étendues ensuite à la mesure d'expression pour laquelle une quantification précise est requise.

## Amélioration des méthodes d'analyse et centralisation de données normalisées

Les aspects informatiques et bio-informatiques sont très importants et n'avaient pas été suffisamment pris en compte au début de la révolution des réseaux d'ADN. Même aujourd'hui, la validation et l'analyse des données d'expression restent relativement grossières [10]. En

outre, la plus grande partie des résultats reste indisponible en dehors du laboratoire d'obtention (parfois même en son sein...), et les jeux de données présentés par certaines équipes sur leurs sites Web ne sont pas directement comparables entre eux faute d'un format commun. De grands efforts sont faits actuellement pour améliorer la situation en mettant au point des logiciels d'analyse plus sophistiqués. Ces derniers comportent à la fois des analyses statistiques, de corrélation et de *clustering*, et des liens directs vers les informations constamment actualisées disponibles sur le réseau. De surcroît, un travail important est en cours pour définir un format de données standard qui devrait permettre d'archiver les données d'expression et de les rendre disponibles à tous – à la manière dont ont été traitées les données de séquences. Il faut néanmoins avoir conscience que le problème des résultats d'expression est bien plus complexe, compte tenu des différentes méthodologies employées et de la qualité assez variable – et délicate à évaluer – de données obtenues. Malgré ces difficultés, les progrès vont certainement être rapides. Il devrait devenir possible de tirer plus d'informations des résultats d'expression (*data mining*) mais aussi d'obtenir un profil d'expression assez complet pour tout gène d'intérêt à partir d'une interrogation de quelques sites sur le réseau.

## Le futur des mesures d'expression

On peut, sans risque, prévoir que la mesure d'expression va rester d'actualité. Certes, d'autres méthodes susceptibles de donner une information fonctionnelle vont être améliorées, rendues plus rapides et applicables à un plus grand nombre d'objets: études d'interaction de protéines, protéomique en général, inactivation de gènes dans différents modèles animaux. La mesure d'expression à grande échelle, facilitée par la disponibilité générale des informations de séquence et poussée par l'amélioration de la technologie des réseaux d'ADN, va certainement rester une approche majeure en bio-

logie durant plusieurs années. Elle va sans doute aussi se banaliser dans un certain nombre d'applications cliniques ou industrielles, jouant dans ce cas un rôle tantôt complémentaire, tantôt alternatif par rapport à l'identification par l'ADN ou la recherche de mutations. ■

## RÉFÉRENCES

1. Iyer VR, Eisen MB, Ross DT, *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* 1999; 283: 83-7.
2. Alizadeh AA, Eisen MB, Davis RE, *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503-11.
3. Jordan BR. Voyage au pays des puces. *Med Sci* 1998; 10: 1097-102.
4. Granjeaud S, Bertucci F, Jordan BR. Expression profiling: DNA arrays in many guises. *BioEssays* 1999; 21: 781-90.
5. Bertucci F, Bernard K, Loriod B, *et al.* Sensitivity issues in DNA array-based expression measurements: performance of Nylon microarrays for small samples. *Hum Mol Genet* 1999; 8: 1715-22.
6. Southern EM, Maskos U. Synthesis of oligonucleotides tethered to a glass surface: applications in the analysis of nucleic acid sequences. *Genome Mapping and Sequencing Meeting*. New York: Cold Spring Harbor University Press 1989: 136.
7. Talary MS, Burt JP, Pethig R. Future trends in diagnosis using laboratory-on-a-chip technologies. *Parasitology* 1998; 117 (suppl): S191-203.
8. Souteyrand E, Cloarec JP, Martin JR, *et al.* Direct detection of the hybridization of synthetic homo-oligomer DNA sequences by field effect. *J Phys Chem B* 1997; 101: 2980-5.
9. Wang J, Jiang A, Mukherjee B. New label-free DNA recognition based on doped nucleic-acid probes within conducting polymer films. *Anal Chim Acta* 1999; 402: 7-12.
10. Ermolaeva O, Rastogi M, Pruitt KD, *et al.* Data management and analysis for gene expression arrays. *Nat Genet* 1998; 20: 19-23.

## Bertrand Jordan

Marseille-Génopole, Case 901, Parc Scientifique de Luminy, 13288 Marseille Cedex 9, France.

## TIRÉS À PART

B. Jordan.