



## Puces-actualités

Bertrand Jordan

Un congrès spécialisé fournit souvent l'occasion de faire le point sur un domaine de recherche. Voici donc quelques informations tirées d'une récente réunion à Zurich, le colloque *Lab-on-a-Chip and Microarrays for Post-Genome Applications*, organisé pour la quatrième fois en janvier 2002 par le *Cambridge Healthtech Institute*. Cette raison sociale d'apparence académique recouvre, on le sait, une entreprise commerciale dont l'activité essentielle consiste à tenir chaque année une cinquantaine de colloques spécialisés. Leurs titres souvent évocateurs (*From proteins to profits, Effective Drug Discovery, Smarter Lead Optimization...*), tout comme les tarifs d'inscription (1 000 à 2 000 € pour les non académiques, non compris le logement dans les hôtels de luxe où ils ont systématiquement lieu) montrent bien qu'ils s'adressent à un public composé en grande partie d'industriels. Ils ne sont pas pour autant dénués d'intérêt, au contraire... J'ai choisi de concentrer cette chronique sur trois domaines : la qualité des données obtenues, qui devient (enfin !) une préoccupation générale, la description rapide de quelques nouveautés prometteuses et, enfin, l'état actuel plutôt décevant des *protein arrays* ou, pour parler français, des réseaux de protéines. Ce faisant, je passerai sous silence d'autres thèmes tout aussi intéressants et donnerai de ce domaine une vision partielle et personnelle, mais permettant au moins d'éclairer certaines tendances actuelles.

### Qualité, mon doux souci...

Dans le célèbre livre de Robert Pirsig *Traité du Zen et de l'entretien des motocyclettes* [1], le narrateur donne un jour à ses étudiants la dissertation suivante : « Qu'est ce que la qualité ? ». Tout le monde se retrouve fort embarrassé pour répondre à cette question, et il ressort finalement des discussions que chacun sait reconnaître la qualité – mais que personne n'arrive à définir précisément en quoi elle consiste. Pour la mesure d'expression par réseaux d'ADN, les éléments fournis ont généralement été très limités : une ou deux belles images, accompagnées de chiffres censés quantifier la sensibilité du système et le plus faible différentiel d'expression mesurable. La sensibilité était souvent exprimée par l'abondance minimale détectable pour une espèce donnée d'ARN messagers au sein du mélange extrait de la cellule ou du tissu analysés. Une valeur « magique » était d'ailleurs souvent avancée : 1/300 000, ce qui correspond approximativement à une molécule d'ARN messagers par cellule de mammifère. Cette « sensibilité » dépend en fait directement de la quantité d'échantillon utilisée (plus on « charge » l'expérience en ARN, plus une espèce rare est détectable), dont la valeur précise était souvent cachée au fin fond des *Materials and Methods*. Quant au plus faible différentiel mesurable, il était généralement évalué à un facteur 2 sur toute la gamme de mesure, alors que la précision dépend en fait très fortement de l'intensité

mesurée, les fortes valeurs étant naturellement les plus précises en valeur relative. Tout cela n'était pas terriblement sérieux...

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

Aujourd'hui, les préoccupations concernant la qualité des mesures sont beaucoup plus présentes.

La nouveauté des *microarrays* s'est émoussée, la pratique de nombreux laboratoires a montré la puissance mais aussi le coût de cette technologie... et la facilité avec laquelle l'on peut se noyer dans une mer de données de qualité souvent douteuse. Les revues deviennent plus exigeantes, et les efforts engagés par plusieurs organismes pour organiser des bases de données d'expression les amènent tout naturellement à définir des normes de documentation et de qualité. Ces soucis étaient illustrés lors du colloque de Zurich par plusieurs présentations, notamment une étude détaillée décrite par Sophie Wildsmith (*GlaxoSmithKline*, Royaume Uni) qui, dans le cadre de travaux de toxicologie sur le rat, montrait l'importance de la répétition des mesures. Une analyse de la variance globale entre expériences montre que la mise en évidence d'une différence significative entre les profils d'expression d'animaux témoins et traités (il s'agit de travaux de toxicologie sur le rat) réclame six répétitions de chaque hybridation. Je serais curieux de savoir combien des nombreux articles publiés dans ce domaine au cours des deux ou trois dernières années ont effectué autant de contrôles... Les avantages et inconvénients des différentes méthodes utilisables pour normaliser les données d'une expérience à une autre, la mesure de la déviation standard et de sa variation selon le domaine d'intensité considéré ont fait aussi l'objet de nombreuses présentations et discussions.

La qualité doit naturellement être présente dès le départ : qualité des ADN et de leur dépôt sur le réseau. On se souvient des difficultés rencontrées avec la collection de clones d'ADNc IMAGE, ressource très utile mais comportant souvent un nombre d'erreurs inacceptable [2], même dans les jeux de clones *sequence-verified* vendus par certains fournisseurs. On sait aussi que l'entreprise *Affymetrix* a dû l'année dernière reconnaître que de nombreuses séquences présentes sur l'une de ses puces « souris » avaient été lues à l'envers dans les bases de données. Mais même lorsqu'elles sont correctement conçues, les puces *Affymetrix* (et les autres) peuvent présenter des défauts, comme le montrait une intéressante étude présentée par Andreas Hohn, de l'entreprise *Gene Data* (Bâle, Suisse). *Gene data*, comme de nombreuses autres firmes, commercialise un ensemble de logiciels destinés aux utilisateurs de *microarrays*. Cette offre comporte des programmes capables d'analyser les images d'hybridation et de repérer les défauts qu'elles peuvent présenter : imperfections locales dues à un accident de fabrication, rayures, poussières, zones ayant mal hybridé à cause de la présence d'une bulle dans la solution... Afin de montrer les possibilités de ces systèmes, Hohn a repris l'ensemble des données d'un article publié l'an dernier sur l'expression génique chez la drosophile [3] et pour lequel les données primaires (donc les images) étaient accessibles. L'analyse ainsi effectuée sur 14 hybridations montre que trois sont franchement mauvaises, sept présentent des défauts qui peuvent néanmoins être corrigés (grâce aux logiciels de *Gene Data*), le reste étant de qualité correcte. Les défauts graves sont pour l'essentiel inhérents aux puces, donc à leur fabrication ou à leur manipulation avant emploi (mais les puces *Affymetrix* étant scellées dans une cartouche, les fausses manœuvres possibles pour l'utilisateur sont assez limitées)... En somme, il ne faut pas faire une confiance aveugle,

même aux fabricants qui ont le plus d'expérience. Apparemment, il ne faut pas non plus s'obstiner à essayer d'utiliser toutes les expériences, même celles qui sont un peu douteuses : l'étude montre qu'en éliminant les images défectueuses, le rendement (le nombre de gènes dont l'expression s'avère différente de manière fiable) augmente très nettement.

Dans ce registre de la qualité, un mot des efforts de normalisation et d'archivage des données d'expression menés notamment à l'*European Bioinformatics Institute* (EBI) à Hinxton (Royaume-Uni). L'objectif (poursuivi actuellement par plusieurs institutions, et même par au moins une entreprise) est d'emmagasiner les résultats dans des bases de données accessibles à tous (ou moyennant finances pour les projets privés), un peu comme cela a été fait à partir de la fin des années 1970 pour les séquences d'ADN. Naturellement, c'est beaucoup plus compliqué : pour être significatives, les données d'expression doivent préciser le type de plate-forme, les conditions de mesure, la nature exacte de la sonde et les détails de la méthode de marquage... afin qu'un jour peut-être, il soit possible de consulter de telles bases pour en tirer, à partir de résultats archivés par des dizaines de laboratoires différents, le profil d'expression d'un gène dans un très grand nombre de tissus et de situations. La mise en place d'une base de données spécifique, appelée *MGED* (*Microarray Gene Expression Database*) est donc une tâche ardue, à laquelle s'est attelée une équipe d'une quinzaine de personnes assistée par plusieurs groupes de travail [4]. Chemin faisant, il s'est avéré nécessaire de définir les informations qui devaient obligatoirement être fournies pour permettre l'archivage d'une expérience. Ceci a été formalisé en une sorte de protocole appelé *MIAME* (*Minimum Information About a Microarray Experiment*) définissant les renseignements à fournir, qui devraient idéalement devenir obligatoires pour toute publication [5]. Mon

impression devant ces efforts est mitigée, et, pour tout dire, je crains que ce système n'entre jamais réellement en vigueur. L'exemple concernant *MIAME*, donné à Zurich par Helen Parkinson (EBI), comportait 11 pages de texte pour une série de cinq hybridations... Peut-être aurait-il été plus judicieux de mettre tout de suite en place un système d'archivage, même très imparfait, afin d'habituer les laboratoires à cette procédure, et de l'améliorer progressivement ensuite ? La mise au point d'une version *light* de *MIAME*, appelée *MIAME express*, montre que l'on est conscient du problème à l'EBI. En tous cas, l'idéal d'une base de données d'expression archivant les résultats de nombreux laboratoires et réellement utilisable de manière transversale semble - malheureusement - encore bien lointain...

### Puces à protéines : l'expectative

Le secteur des *protein arrays* occupait une place importante dans le programme de ce colloque : plus d'un tiers des présentations y étaient consacrées. Ayant assisté dans un passé récent à des conférences annonçant la construction de réseaux comportant des milliers de protéines, potentiellement utilisables pour étudier les interactions de systèmes complexes, j'étais fort intéressé par les nouveaux développements dans ce secteur. Je devais rester sur ma faim... Il existe bien des puces susceptibles de comporter dix mille plots, réalisées par microfabrication et comportant des surfaces spécialement mises au point pour fixer les protéines dans de bonnes conditions. David Wilson (*Zyomix*, États-Unis) en montrait des exemples - pour préciser ensuite que, ne disposant pas de dix mille protéines à placer sur ces réseaux, l'équipe s'était limitée à des réalisations beaucoup plus modestes (six unités identiques de 25 plots sur une puce). En fait, comme ceci est fort bien exposé dans un récent article de *Nature* [6], la fabrication de telles puces se heurte à au moins trois difficultés : disposer de nom-



breuses protéines pures à déposer sur le réseau, maîtriser leur interaction avec les entités qu'elles sont censées fixer, et avoir une méthode de détection sensible et quantitative. Ces problèmes sont résolus pour les réseaux d'ADN : l'amplification par PCR (ou la synthèse d'oligonucléotides) fournissent facilement l'ADN, l'interaction est assurée par l'hybridation, réaction bien connue et aux paramètres bien maîtrisés, et le marquage en masse de l'échantillon par différentes méthodes donne de bons résultats. Ils restent en revanche très présents pour les réseaux de protéines...

L'essentiel des réalisations présentées à Zurich concernait donc des puces de faible complexité (de trois à vingt cinq plots), sur lesquelles sont disposés des anticorps bien caractérisés destinés à capter une cible connue à l'avance et dont la présence est détectée ensuite par un deuxième anticorps apportant le système de détection – l'habituel *sandwich assay* des dosages immunologiques.

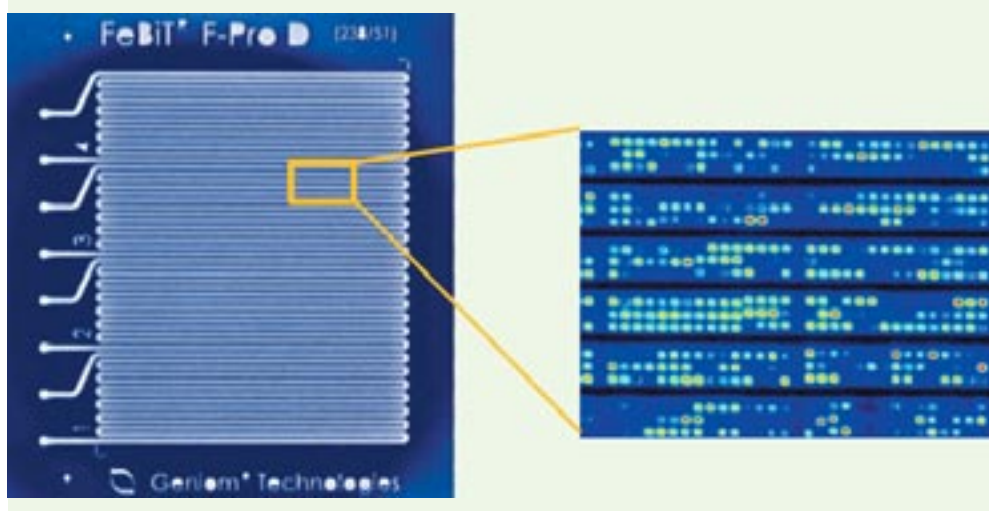
C'est, en somme, une transposition du système ELISA sur un support de type réseau, ce qui permet de multiplier le nombre de paramètres mesurés et de diminuer le coût de chaque dosage. Parmi les systèmes décrits, un réseau d'anticorps monoclonaux pour détecter des pesticides dans l'eau (Claus Christensen, Université de Copenhague, Danemark), une autre puce pour suivre les protéines caractéristiques de l'arthrite rhumatoïde (Sara Mangialaio, Novartis Pharma, Bâle, Suisse), ou encore (en renversant le système) un réseau constitué d'antigènes viraux afin de détecter les anticorps correspondants dans le sérum humain (sérodiagnostic) (Tito Bacarese-Hamilton, Imperial College, Royaume-Uni). Rien de grandiose donc (il manquait quelques ténors à ce colloque), mais des applications qui peuvent trouver leur place si le conservatisme des laboratoires d'analyse et la puissance des concurrents dont la technologie ELISA est bien installée ne les

réduisent pas au rôle de figurantes. Ce serait dommage, car ces approches préparent la voie à des puces à protéines plus complexes dont le rôle, malgré les difficultés rencontrées, sera sûrement important dans l'avenir.

### Quelques nouveautés

Un colloque *Cambridge Healthtech* serait décevant s'il ne comportait pas quelques innovations technologiques propres à faire rêver ceux que ce domaine inspire. Pour ma part, trois items (dont deux au moins ne sont pas tout à fait nouveaux) répondaient à ce critère. Pour commencer, le système de PamGene [7], firme hollandaise représentée par son vice-président Alan Chan. Il s'agit d'un réseau de faible complexité destiné aux applications cliniques, et réalisé sur un support très particulier, une sorte d'éponge d'alumine comportant d'innombrables canaux verticaux. L'ADN est fixé dans toute l'épaisseur de ce support, et la

goutte de solution d'hybridation déposée sur la puce la traverse de manière répétée sous l'influence d'une pression de gaz variable (la tension superficielle évite son arrachement du support). L'hybridation (avec une sonde fluorescente) est suivie en direct au microscope (le support mouillé est transparent) et est très rapide : une dizaine de minutes pour une identification bactérienne par hybridation d'un produit de PCR avec des oligonucléotides fixés sur le réseau. C'est me semble-t-il un bon exemple de puce à ADN adaptée aux applications cliniques pour lesquelles le nombre de plots nécessaire est faible, mais où les résultats doivent être



**Figure 1. Le Geniom 1.** Cette figure, reproduite avec l'aimable autorisation de FeBit, montre l'élément central de la machine Geniom 1, le *DNA processor*. C'est une plaque transparente (à gauche) comportant un ensemble de quatre canaux indépendants ; la synthèse d'oligonucléotides se fait directement sur la paroi plate du canal par voie photochimique, l'illumination étant assurée par un système comportant 40 000 micromiroirs commandés individuellement, et les réactifs étant apportés *via* les canaux. L'ensemble dure environ six heures pour la synthèse de 25-mères. L'hybridation est ensuite effectuée dans les mêmes canaux, la solution *ad hoc* circulant à l'intérieur de ceux-ci ; la lecture du résultat d'hybridation se fait grâce à une caméra CCD, toujours dans le *DNA processor* (image de droite). Dans le modèle actuel, le nombre total d'éléments dans les quatre canaux est de 32 000 (taille individuelle 32  $\mu\text{m}$ ) ; la longueur des oligonucléotides peut atteindre 60 bases.

acquis rapidement – et, de préférence, par des méthodes qui ne dépayseraient pas trop le laboratoire d'analyse. Son application à des mesures d'expression est en cours de mise au point, mais va sans doute se heurter à des problèmes de sensibilité. Il existe au moins une autre entreprise développant une technologie analogue, *MetriGenix* (États-Unis), mais elle n'était pas présente à Zurich.

Autre nouveauté dans le domaine de l'analyse : le système des *Nanobarcodes* présenté par Michael Natan de *SurroMed* (États-Unis) [8]. Il s'agit de minuscules barreaux métalliques longs de quelques micromètres et constitués de deux ou trois métaux (argent, or, cuivre). Ils sont fabriqués par synthèse électrochimique sur une sorte de moule (microscopique), ce qui permet d'alterner les couches déposées et d'obtenir des barreaux comportant jusqu'à une dizaine de bandes de différents métaux et d'épaisseur variable. On peut alors fixer sur chaque type de barreau un réactif (protéine ou ADN), puis les mélanger pour analyser un échantillon complexe, la lecture du résultat se faisant au microscope (automatisé) ou par un système de détection en flux. Cette technologie, qui montre l'alliance de la microfabrication et de la

biologie, peut avoir des applications dans des domaines très variés.

Terminons cette revue des nouveautés (il y en avait d'autres) par un appareil plus conventionnel mais néanmoins assez étonnant, le « Geniom 1 » de l'entreprise allemande *FeBit* (Mannheim, Allemagne) [9] (Figure 1). J'y avais fait une rapide allusion dans une précédente chronique [10], mais ai pu constater cette fois la réalité de cette machine et les progrès accomplis depuis une année. Il s'agit d'un appareil d'un seul tenant, qui est capable de synthétiser des puces à oligonucléotides, puis d'effectuer l'hybridation et la lecture du résultat – tout le matériel nécessaire tenant à l'intérieur d'un cube dont le côté mesure un peu moins d'un mètre ! L'élément central est une sorte de cartouche transparente à l'intérieur de laquelle la synthèse est réalisée par voie photochimique (en utilisant des réactifs différents de ceux d'Affymetrix, et un multimiroir programmé de *Texas Instruments*). La même cartouche sert à l'hybridation, et les résultats sont lus ensuite par un système optique à base de caméra CCD. Cela semble trop beau pour être vrai... Pourtant les machines existent réellement, une série de dix prototypes a été

vante – qui risque néanmoins de se heurter à la concurrence de puces prêtes à l'emploi et produites en grande série à des prix plus doux que les tarifs actuels..

### Pour conclure...

Bien d'autres aspects des *microarrays* étaient présentés lors de ce colloque, notamment leur emploi pour analyser l'expression génique dans différents types de cancer. Ce genre d'étude, activement poursuivi par de nombreuses équipes et entreprises, donne des résultats prometteurs en termes de classification des cancers. Il fournit des informations dont on peut penser qu'elles permettront d'ajuster le traitement administré à la nature précise de chaque cancer et donc d'améliorer à la fois le pronostic et le confort du malade. Il y a donc là une importante possibilité d'application clinique dans un avenir proche, et ce thème mériterait à lui seul une chronique ou même un article de fond. Je n'ai pas non plus traité autrement qu'en passant la très importante question de la bio-informatique associée aux mesures d'expression, que ce soit pour acquérir et vérifier les mesures, pour les analyser par différents modes de regroupement (pour ne pas dire *clustering*), et pour essayer d'en tirer du sens du point de vue biologique. Elle faisait d'ailleurs l'objet d'un deuxième colloque *Cambridge Healthtech* et mériterait, elle aussi, une chronique entière...

Ce qui apparaît en tous cas de manière évidente, c'est que la logique des *arrays* fait tache d'huile dans la biotechnologie actuelle. Logique dont les éléments essentiels sont l'emploi de réseaux (d'ADN, de protéines, ou même de cellules comme dans le cas des *phenotype arrays* présentés par Barry Bochner [*Biology*, États-Unis]) autorisant un grand nombre de mesures simultanées, et la miniaturisation qui réduit les besoins en échantillons et réactifs, diminuant ainsi les coûts. Le mode de réalisation de ces réseaux semblait, lors de ce colloque, avoir atteint une certaine stabilité. C'est sans doute trompeur, car

## RÉFÉRENCES

1. Pirsig RM. Zen and the art of motorcycle maintenance. Bodley Head, 1974. Édition française la plus récente : Collection Points. Paris : Seuil, 1998.
2. Halgren RG, Fielden MR, Fong CJ, Zacharewski TR. Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res* 2001 ; 29 : 582-8.
3. De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci USA* 2001 ; 98 : 12590-5.
4. [www.ebi.ac.uk/microarray/](http://www.ebi.ac.uk/microarray/)
5. Brazma A, et al. (30 auteurs). Minimum information about a microarray experiment (MIAME) : toward standards for microarray data. *Nat Genet* 2001 ; 29 : 365-71.
6. Alison Abbott Betting on tomorrow's chips. *Nature* 2002 ; 415 : 112-4.
7. [www.pamgene.com](http://www.pamgene.com)
8. [www.surromed.com](http://www.surromed.com)
9. [www.febit.com](http://www.febit.com)
10. Jordan BR. Chroniques génomiques. Puces à ADN, les brevets contre le progrès. *Med Sci* 2001 ; 17 ; 893-6.
11. [www.BioArrayNews.com](http://www.BioArrayNews.com)
12. [www.bmi-heidelberg.com/BioChipNet/](http://www.bmi-heidelberg.com/BioChipNet/)

de nombreuses firmes parmi les plus innovantes n'étaient pas représentées et l'on peut s'attendre à de nouvelles surprises durant les mois qui viennent ; mais le principe même des réseaux est, lui, bien installé et va certainement continuer à étendre son règne sur la biologie à grande échelle comme sur de nombreuses applications cliniques.

### Post-scriptum

Quelques nouvelles sur le front des brevets et de la concurrence, pour donner une suite à ma récente diatribe sur le sujet [10]. *Affymetrix* a fait la paix avec *Oxford Gene Technologies*, l'entreprise liée à Edwin Southern que je considère comme le vrai inventeur des puces à oligonucléotides, il lui en a coûté près de vingt millions de dollars. La paix a aussi

été conclue avec l'entreprise *Incyte* (les deux compagnies s'étaient réciproquement attaquées depuis plusieurs années), qui s'est retirée en octobre 2001 du marché des *microarrays* tout comme *Corning*, autre concurrent potentiellement dangereux. *Motorola*, *Agilent* et un certain nombre de *start-ups* (certaines mentionnées dans cette chronique) restent dans le domaine, mais *Affymetrix* (qui emploie maintenant plus de 900 personnes) est clairement la force dominante. Pour ceux qui souhaiteraient suivre de près ce paysage changeant, une revue spécialisée qui semble bien informée mais est malheureusement fort chère, *BioArray News* [11], et un site tendant à rassembler l'ensemble des informations, qui était payant mais va devenir gratuit, *BioChipNet* [12]. ♦

**The latest on chips**

---

### TIRÉS À PART

B. Jordan

