

## Cartographie et séquence du génome humain

Chronique de « Cold Spring Harbor Laboratory »

Un récent colloque à Cold Spring Harbor Laboratory, près de New York (*Genome Mapping and Sequencing*, 27 avril au 1<sup>er</sup> mai 1988) a permis à près de deux cents spécialistes de ce domaine de faire le point sur les techniques de cartographie et de séquence du génome et sur les résultats récemment obtenus. Des impressions multiples et parfois contradictoires que l'on peut ressentir après quatre jours d'exposés très denses et le plus souvent d'excellente qualité surnagent des lignes de force (évidemment subjectives) que je vais essayer d'explicitier.

**La carte d'abord, la séquence après.** Malgré l'impact médiatique et même politique des projets de séquence totale du génome humain (trois milliards de nucléotides, alors que — rappelons-le — le total des séquences accumulées à ce jour après dix ans de génie génétique se compte en millions de nucléotides), il y a un consensus évident dans la communauté scientifique pour commencer par une cartographie physique détaillée avant de passer à la séquence. Les développements sur la séquence n'occupèrent d'ailleurs qu'une matinée sur les quatre jours du colloque, et firent ressortir, à côté de procédés techniques très intéressants comme le séquençage en multiplex (George Church, Harvard, MA, USA) ou l'automatisation des étapes précédant la séquence proprement dite, l'incapacité dans laquelle l'on est actuellement d'exploiter complètement les informations de séquence obtenues. Séquencer une centaine de kilobases, c'est un travail modeste à l'échelle du projet de séquence du génome (30 000 fois plus): pourtant l'interprétation de telles données, et par exemple la recherche informatisée de gènes présents dans la séquence, se heurte à de gros problèmes et demande le développement de nouveaux systèmes informatiques qui n'en sont qu'à

leurs balbutiements. Inutile évidemment d'accumuler des informations dont on ne sait que faire. En revanche, la cartographie permet à la fois (si l'on y tient vraiment) de séquencer de façon ordonnée, et, par différentes stratégies complémentaires, d'accéder aux gènes contenus dans la région parcourue, d'étudier leur expression dans différents tissus et leur implication éventuelle dans une maladie génétique.

**De bas en haut: les contigs.** Les contigs, ce mot barbare se rapporte à une stratégie de cartographie très britannique (développée en particulier par Alan Coulson au MRC de Cambridge, GB, et largement utilisée aux USA). A partir d'une cellule hybride hamster-homme contenant uniquement la région de chromosome humain que l'on souhaite étudier (en plus de chromosomes de hamster, bien sûr) on construit une très bonne banque de cosmides. Le sous-ensemble de cosmides contenant de l'ADN humain est ensuite repéré par hybridation avec de l'ADN humain total (via les séquences répétées spécifiques de l'homme). La sous-banque ainsi obtenue couvre, avec quelques milliers de cosmides, la région humaine étudiée. « Il ne reste plus qu'à »... cartographier tous ces cosmides pour les aligner et obtenir ainsi la carte complète de la région. La tâche n'est pas aussi insurmontable qu'on pourrait le penser grâce à l'utilisation des *fingerprints* (empreintes digitales) de cosmides et à l'automatisation de toute la procédure.

Les *fingerprints* de cosmides? En gros il s'agit de couper chaque cosmide avec une première enzyme qui va donner une vingtaine de fragments, de marquer leurs extrémités puis de recouper par une deuxième enzyme. Séparé sur un gel très résolutif (type gel de séquence nucléotidique), ce mélange donne une quarantaine de fragments caracté-

ristiques. Cette analyse répétée sur chacun des clones de la sous-banque finit par mettre en évidence les recouvrements: deux clones qui se recouvrent auront un certain nombre de fragments en commun. A partir de là la carte se construit assez rapidement de proche en proche, en laissant éventuellement quelques trous (séquences inclonables dans des cosmides par exemple) qu'il faudra remplir par d'autres méthodes.

Cette analyse est maintenant de plus en plus automatisée: réaction de coupure effectuée en parallèle sur 96 clones par des robots, marquage à l'aide de bases fluorescentes, analyse automatique par électrophorèse et lecture directe sur les appareils développés pour la séquence automatique d'ADN (Applied Biosystems, Du Pont, etc.), recherche informatique des recouvrements... Une approche assez *high-tech* (haute technologie), qui intéresse beaucoup l'industrie et les grands centres comme Los Alamos et qui est utilisée entre autres par Alan Coulson et Sydney Brenner (MRC, Cambridge, GB) pour cartographier complètement le génome nématode (*C. elegans*), 80 mégabases (1 mégabase =  $10^3$  kb\* =  $10^6$  bases), par Maynard Olson (Saint-Louis, Mi, USA) pour la levure (14 mégabases), par Antony Carrano (Livermore, CA, USA) pour le chromosome humain 19 (100 mégabases) et par bien d'autres. Approche assez lourde bien sûr mais très puissante, bénéficiant d'une automatisation bien avancée, et qui a le gros avantage d'engendrer, en plus de la carte, une collection de clones de cosmides ordonnée, couvrant l'ensemble de la région et auxquels on peut revenir pour étudier plus en détail telle ou telle zone intéressante.

\* kb = kilobases.

## De haut en bas : gels pulsés et YAC.

A l'inverse, on peut essayer de cartographier à partir de grands morceaux d'ADN puisque ceux-ci peuvent maintenant être analysés grâce à l'électrophorèse en champ pulsé, et en principe clonés grâce aux chromosomes artificiels de levure, les *Yeast Artificial Chromosomes*. Comme on pouvait le prévoir, l'électrophorèse en champ pulsé se généralise. Les différentes variantes techniques proposées ne semblent pas encore départagées (à part la méthode consistant à inverser le champ électrique périodiquement, ou FIGE [*field inversion gel electrophoresis*] qui perd du terrain) et l'effort se porte sur l'étude de grands fragments de plusieurs mégabases, ce qui pose des problèmes de réactifs de coupure car les enzymes à site rare coupent encore trop souvent, et les digestions partielles sont difficiles à maîtriser. La mise au point des conditions de migration et l'obtention de témoins de masse moléculaire dans la zone de taille considérée restent également difficiles.

Néanmoins dès maintenant, dans les zones où la densité de sondes est suffisante, de nombreuses liaisons physiques sont obtenues lorsque deux sondes A et B hybrident à la même bande de 800 kb (par exemple) : on peut alors affirmer — moyennant quelques contrôles — qu'elles sont distantes au maximum de 800 kb sur le génome. De proche en proche se construisent ainsi des cartes de restriction couvrant quelques mégabases sur les chromosomes 3, 19, 21, dans différentes régions du chromosome X et en particulier la région pseudoautosomique à l'extrémité du bras court, étudiée de façon convergente par au moins trois groupes dont les résultats étaient rapportés par William Brown (Oxford, GB), Christine Petit (Institut Pasteur de Paris) et Gudrun Rappold (ICRF, Londres, GB). On peut d'ailleurs cartographier en champ pulsé de façon plus systématique en couplant l'analyse par une enzyme donnée (Not I par exemple) et la construction — grâce à divers artifices de clonage — d'une banque contenant tous les segments d'un chromosome

donné qui comportent un site Not I, et seulement ceux-ci. Un tel clone utilisé comme sonde en champ pulsé (après coupure par l'enzyme Not I) va révéler deux bandes (de part et d'autre du site) qui seront *ipso facto* liées physiquement. Plus fort encore, Nobuyoshi Shimisu (Keio School of Medicine, Tokyo, Japon) a montré qu'il était possible de séparer par tri de chromosome assez de chromosomes 21 (de l'ordre de 500 000) pour utiliser l'ADN correspondant en champ pulsé et que cet ADN était en assez bon état pour donner des fragments Not I de 1 000 kb et plus. Enfin l'auditoire a suivi avec intérêt les efforts de différents groupes de chimistes pour mettre au point des réactifs permettant de couper spécifiquement l'ADN en des sites plus rares que ceux clivés par les enzymes de restriction, à l'aide d'oligonucléotides (Claude Hélène, Muséum d'Histoire Naturelle, Paris) ou de protéines à site très spécifique (David Sigman, UCLA, CA, USA).

Quant aux chromosomes artificiels de levure, les YAC, leur utilisation reste difficile. De fait, depuis la première publication de Burke, Carle et Olson (*Science* 1987 ; 236 : 806-12) aucun résultat utilisant ces vecteurs n'a encore été publié. En fait les premiers espoirs entretenus par la lecture de l'article sus-cité (pouvoir construire rapidement des banques génomiques complètes contenant des clones de 500 à 1 000 kb) étaient sans doute exagérés. Néanmoins à l'heure actuelle plusieurs groupes ont construit des banques de quelques milliers de clones contenant des fragments de 100 à 300 kb : banques partielles donc (un génome avec de tels clones représente au moins 20 000 colonies) mais déjà utilisables. Des clones YAC contenant un gène HLA, des gènes de l'ARN ribosomique 5S ou des régions variables d'immunoglobulines V<sub>K</sub> ont déjà été isolés (David Schlessinger, Maynard Olson, St Louis, Mi, USA) même si leur étude n'a pas été poussée très loin.

De nombreux groupes se lancent déjà dans une approche de type *contig* avec le système YAC : construction d'une banque YAC à

partir d'un hybride somatique, puis repérage des clones « humains » qui constituent une sous-banque. Il est évident qu'obtenir une région chromosomique de quelques bandes (20 ou 30 mégabases) sous forme d'une banque d'une centaine de clones YAC, *a priori* faciles à ordonner, est une perspective extrêmement séduisante. Ceci suppose cependant que la banque de départ soit très complète, ce qui n'est pas facile à obtenir vue la très faible efficacité de transformation de la levure par ces vecteurs : quelques centaines de transformants par microgramme d'ADN. Mais il suffit de se reporter aux premières années des banques dans les bactériophages ou les cosmides pour penser que la technique va rapidement évoluer et que les banques YAC vont devenir beaucoup plus maniables.

**Les télomères humains.** Un bon colloque n'est pas complet sans un ou deux *scoops*. Pas de déception de ce côté-là avec en particulier l'annonce par Bob Moysis (Los Alamos, CA, USA) du clonage de ce qui est vraisemblablement la structure commune à l'ensemble des télomères humains : une simple séquence répétée du type (TTAGGG)<sub>n</sub>, séquencée d'ailleurs il y a bien longtemps par E. Southern. Des hybridations *in situ* et des analyses par l'enzyme Bal 31 qui digère préférentiellement les télomères, très convaincantes, ne laissent guère de doute sur la solidité de ce résultat qui va avoir des conséquences importantes sur le plan théorique et pratique (cartographie des régions proches des télomères par digestion partielle par exemple).

**La carte et le territoire.** Certes comme l'a dit, dans un autre contexte, Mac Luhan, la carte n'est pas le territoire. La cartographie du génome ne nous dira pas tout et, comme ailleurs, après la génétique moléculaire pure et dure il faudra passer à l'étude fonctionnelle, à la biochimie fine et à la biologie cellulaire. L'orientation actuelle de la recherche sur la myopathie de Duchenne le montre bien : en quelques mois l'accent est passé du clonage de fragments du gène à la recherche du rôle exact de la protéine, la dystrophine, qui est son produit (*voir médecine/sciences*) ▶

■■■ BRÈVES ■■■

► 1988 ; 3 : 141-50).

Nous n'en sommes pas là et à l'heure actuelle l'analyse structurale du génome et l'identification des gènes qu'il contient constituent un préalable indispensable. Le brusque coup d'accélérateur que donne à cette recherche le développement de techniques très puissantes et l'ingéniosité que mettent les uns et les autres à exploiter ces nouvelles possibilités permettent de penser que d'ici un an ou deux nous aurons les réponses à beaucoup de questions actuellement posées, et sans doute de nouvelles questions que nous ne soupçonnons pas aujourd'hui !

**Bertrand Jordan**

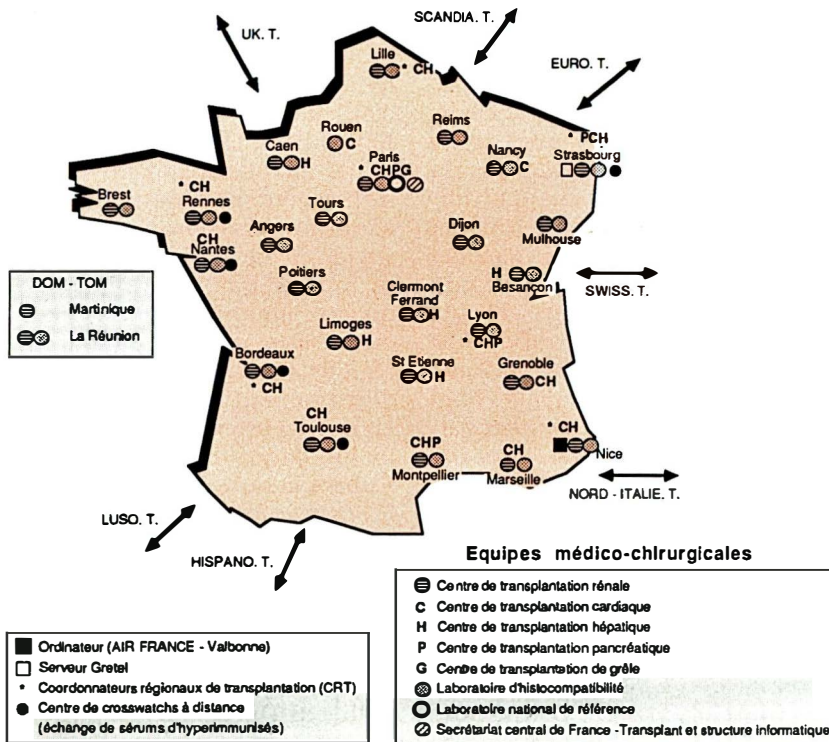
■■■ Une isoenzyme de phosphatidylinositol-kinase catalyse la synthèse de phosphatidylinositol 3 phosphate (Ptd-Ins (3) P). Le Ptd-Ins (4, 5) P2 est le précurseur bien connu du diacylglycérol et de l'Ins P3, seconds messagers de nombreuses actions hormonales ou stimulées par les facteurs de croissance. Le Ptd-Ins est phosphorylé par deux isoenzymes de Ptd-Ins kinase, appelées I et II. La forme I est associée à plusieurs types de tyrosine-kinases, par exemple le récepteur du facteur de croissance PDGF et le complexe entre l'antigène moyen T du virus polyome et le produit P60<sup>c-src</sup>, de l'oncogène

*c-src*. En fait, il vient d'être démontré que cette isoenzyme I catalysait la synthèse de Ptd-Ins (3) P et non de Ptd-Ins (4) P. Le dérivé phosphorylé en 3 est un composant mineur représentant moins de 10 % du Ptd-Ins (4) P et rien n'est connu de son métabolisme ultérieur. Néanmoins, l'association de la Ptd-Ins kinase I avec des produits d'oncogènes suggère que le Ptd-Ins (3) P pourrait être impliqué dans le contrôle des phénomènes de prolifération cellulaire.

[Whitman M, *et al. Nature* 1988 ; 332 : 644-6.]

**FLASH**

**Structure de France-Transplant**



France-Transplant est une association Loi 1901, fondée le 23 septembre 1969 par Jean Dausset, reconnue d'utilité publique le 18 décembre 1978 et exclusivement subventionnée par le ministère de la Santé et la CNAMTS.

Elle regroupe l'ensemble des équipes de transplantation et de prélèvement, les laboratoires d'histocompatibilité et les coordonnateurs de transplantation, répartis sur le territoire national et reliés entre eux par un réseau télématique, initialement de type Télex puis récemment par Minitel et micro-ordinateurs.

Au 31 décembre 1987, 91 équipes participent à l'effort national de transplantation : 39 pour les greffes de rein, 25 pour le cœur, 20 pour le foie, 5 pour le pancréas et 2 pour l'intestin grêle.

La réalisation des tests d'histocompatibilité (groupes HLA du donneur et du receveur, crossmatch en urgence) est assurée par 26 laboratoires.

**J. Hors**