



par Bertrand JORDAN

*Les heurs et malheurs
du séquençage d'ADN
à grande échelle*

Un grand optimisme. A la fin des années 1980, l'on envisageait avec beaucoup d'optimisme le séquençage de grandes régions d'ADN : les « séquenceurs » commercialisés par Applied Biosystems, puis annoncés par Du Pont et Pharmacia/LKB, devaient permettre un grand bond en avant de la productivité tout en évitant l'utilisation d'isotopes radioactifs. Les systèmes robotiques allaient rapidement automatiser la plupart des procédures annexes, et des moyens informatiques sophistiqués devaient faciliter l'assemblage, la vérification des séquences et, dans un deuxième temps, leur interprétation. On envisageait donc sans trop d'inquiétude des programmes visant à établir des séquences sur une ou même plusieurs mégabases (1 mégabase : 1 000 kilobases), un million de nucléotides) : souvenons-nous des projets de séquençage du complexe majeur d'histocompatibilité (deux mégabases), de la bande q28 du chromosome X (dix mégabases) ou de l'ensemble des *loci* codant pour les gènes du récepteur T chez la souris et chez l'homme.

La stratégie adoptée. Rappelons la stratégie qui avait été adoptée, avec quelques variantes, par tous les groupes concernés. L'unité de séquençage était l'ADN cloné dans un phage (15 à 20 kb) ou dans un cosmide (40 kb). Le fractionnement de ce grand segment d'ADN en sous-clones était fait selon une tactique de *shotgun*, avec fragmentation aléatoire de l'ADN par sonication suivie du sous-clonage des fragments (il existe des méthodes plus systématiques, nous en reparlerons,

mais elles ont toutes paru trop lourdes pour être appliquées à grande échelle). Suivaient alors la préparation des matrices simple brin, les réactions de séquence avec des amorces ou des précurseurs fluorescents, et l'analyse des produits de réaction par un « séquenceur » automatique, en général de marque Applied. La capacité de ces machines, de l'ordre de 5 000 à 10 000 nucléotides de séquence brute par jour (24 pistes, 400 nucléotides lus par piste), devait permettre une « production » supérieure au million de nucléotides « bruts » par an et par machine et donc, avec plusieurs machines, plusieurs centaines de kilobases de séquence confirmés par an (notons qu'il faut cinq à dix kb de séquence brute pour une séquence de 1 kb confirmée). Les résultats donnés par ces machines devaient enfin être repris par un système informatique effectuant l'assemblage de toutes ces petites séquences.

On déçante... Aujourd'hui, début 1991, on est loin du compte : la plus longue séquence jamais établie reste celle du virus EBV, faite à la main en Angleterre, et les entreprises de mégaséquençage ne rencontrent plus le même enthousiasme. Bien peu de projets ont produit plus d'une centaine de kilobases de séquence confirmée d'un seul tenant, et les « séquenceurs fous » sont moins triomphalistes que par le passé. Ils se sont heurtés, en fait, à toute une série de problèmes techniques et organisationnels liés au saut quantitatif qu'ils tentaient d'effectuer et aux difficultés entraînées par une automa-

tisation partielle. Sur le plan technique d'abord, les machines à séquencer, qui n'automatisent en fait qu'une des étapes de la détermination de séquence (la séparation des fragments d'ADN sur gel et leur détection), se sont généralement avérées très exigeantes sur la *quantité* et la *qualité* des mélanges réactionnels qu'on leur donne à analyser. Là où la méthode manuelle s'accommode de petites quantités de préparations d'ADN peu purifiées, la machine exige une charge plus importante d'un échantillon quasiment parfait, et si l'on voit autant de « séquenceurs » inutilisés dans les laboratoires, c'est en grande partie à cause de ce problème. En dehors des questions de mise au point, il y a à cela une raison fondamentale. Dans la méthode classique, la détection finale est effectuée par un détecteur de grande taille (le film à rayons X) auquel on laisse tout son temps, des dizaines d'heures en général, pour enregistrer les données globalement sur toute la surface du gel. La machine, au contraire, doit détecter les fragments « au vol », en cours de migration et pendant un temps très court, ce qui impose une sensibilité extrême et laisse peu de marge pour compenser d'éventuelles imperfections de l'échantillon. La contrainte sur la qualité de l'ADN, commune à des degrés divers aux différentes variantes de ces machines, est particulièrement mal venue ici : on aurait aimé automatiser différentes étapes préalables comme la préparation d'ADN à partir des clones. Or les méthodes de préparation qui donnent de l'ADN très pur font presque toujours inter-

venir une étape de centrifugation qui est particulièrement difficile à automatiser : les techniques utilisant des filtrations sont plus facilement automatisables mais donnent en général un ADN de moins bonne qualité. On touche, là, à un problème général. Celui des goulets d'étranglement : une procédure complexe ne peut pas aller plus vite que la plus lente de ses étapes ; et il ne sert pas à grand-chose d'automatiser une partie du processus si l'autre reste manuelle et lente. L'expérience de ceux qui ont mis en place des laboratoires de mégaséquençage montre que, sitôt un goulet d'étranglement éliminé, un autre apparaît... de sorte que la mise en œuvre du projet ressemble à une course d'obstacles.

La technique n'est pas seule en cause. Une autre difficulté, de nature différente celle-là, tient à l'interférence entre développement et production. L'objectif est l'obtention de données de séquence : il faut donc simultanément organiser leur établissement selon un certain processus et chercher en permanence à améliorer ce même processus. Il y a un conflit potentiel entre ces deux fonctions, surtout quand ce sont les mêmes personnes qui l'assument, et sur les mêmes appareils ; il suffit de quelques essais par semaine pour bloquer la moitié des machines et réduire de beaucoup la production. Inversement, si l'on ne fait pas de développement, on risque d'investir beaucoup de temps et d'argent de façon inefficace, avec des techniques en train de se périmer. L'équilibre n'est pas facile à trouver ! Un autre problème organisationnel est celui du personnel : le travail effectué dans un tel contexte n'est pas particulièrement passionnant (la situation est très différente de celle du chercheur qui séquence « son » gène) : à quel type de personne faut-il le confier ? Deux écoles à ce sujet : les uns pensent qu'il faut plutôt choisir des personnes d'un niveau d'études moyen que l'on formera « sur le tas » et qui accepteront de faire durablement ce travail répétitif. D'autres, au contraire, préfèrent viser haut en acceptant le fait que la personne engagée ne fera ce travail qu'un ou deux ans

et quittera ensuite le laboratoire pour passer à un emploi plus motivant. Ils considèrent que les inconvénients de cette instabilité seront compensés par la rapidité d'apprentissage et les capacités plus larges de ces employés. La deuxième option m'a paru majoritaire dans les laboratoires aux États-Unis, ce qui est sans doute lié (entre autres) à une tradition de mobilité d'emploi qui n'existe pas en France.

On repart sur des bases plus raisonnables. Les laboratoires de mégaséquençage ont maintenant tiré la leçon des difficultés rencontrées et affichent des buts plus modestes ; ils ont aussi pris conscience de la nécessité d'une organisation de type industriel. Cette organisation impose de séparer production et développement, avec, d'un côté, une unité de production et, de l'autre, des chercheurs ou ingénieurs qui perfectionnent les méthodes sans interférer au jour le jour avec cette unité. Une plus grande attention est portée à l'entretien et au dépannage des machines avec les moyens locaux, sans avoir à faire appel au constructeur, et, surtout, on se livre à une étude détaillée de toutes les étapes du processus pour évaluer de façon réaliste la production à chaque étape, rationaliser l'ensemble et faire porter l'effort maximal sur les étapes les plus limitantes. Toutes ces considérations ne seraient pas déplacées dans une usine d'automobiles ; et cela marque bien que le passage à cette échelle entraîne un changement qualitatif, qu'il ne s'agit pas simplement de faire cent fois plus, mais aussi de le faire autrement...

Et les « petits » laboratoires ? A côté de ce secteur du mégaséquençage, il existe de très nombreuses équipes qui ont de temps en temps besoin de déterminer la séquence de quelques kilobases ou dizaines de kilobases d'ADN. Ce besoin dépasse d'ailleurs largement la communauté de la génétique humaine et concerne peu ou prou tous les biologistes moléculaires. Pour le moment, ce secteur est resté fidèle aux méthodes traditionnelles de séquençage, et les machines actuelles sont à la fois trop chères, trop productives et trop « per-

turbantes » (par les modifications en amont et en aval qu'elles entraînent) pour s'y imposer. Il y a pourtant un potentiel d'innovations important dans ce secteur, et des possibilités commerciales auxquelles les constructeurs sont sensibles : car il y aura peut-être quelques dizaines de centres de mégaséquençage, mais il existe des milliers de laboratoires faisant un peu de séquence et intéressés à ce que cela devienne plus facile, plus rapide et moins cher... On peut citer, en vrac, parmi les innovations qui s'imposeront peut-être dans ce domaine, la détection directe de la radioactivité, ou au moins le remplacement des films à rayons X par des *imaging plates* que l'on expose (comme un film) pendant quelques heures au maximum pour les « développer » ensuite en quelques minutes dans un appareil qui utilise un faisceau laser pour leur lecture. Dans les deux cas on gagne en vitesse, et surtout l'image est directement enregistrée dans une mémoire d'ordinateur. Autre option, le remplacement des isotopes radioactifs par le système de marquage par photoluminescence qui pourrait enfin arriver à s'imposer dans les équipes de recherche. L'électrophorèse capillaire — qui a déjà fait un « tabac » dans les laboratoires de biochimie — peut être, elle, la base d'appareils de séquence plus modestes en production et en prix que les machines Applied ou Du Pont... On voit que la gamme des innovations possibles reste large, ce qui me permet de terminer cette chronique sur une note optimiste ! ■

Bertrand R. Jordan

Directeur de recherche au Cnrs, responsable du groupe génétique moléculaire humaine, CIML, Inserm/Cnrs, case 906, 13288 Marseille Cedex 9, France.

Remerciements

Je remercie Leroy Hood, Ben Koop et Tim Hunkapiller (Caltech, Pasadena, CA), Richard Gibbs (Institute of Medical Genetics, Houston, TX) et Paul Silverman (Beckman Instruments, Fullerton, CA) pour les discussions qui ont largement nourri cet article.