

Séquençage du génome humain :

où en est-on réellement ?

La lecture de la grande presse ne peut que laisser perplexe le public, non spécialement averti, par les annonces successives concernant le séquençage du génome humain. Au mois de juin de l'an dernier c'était l'annonce de l'assemblage des fragments séquences, en février dernier certains journalistes n'hésitaient pas à annoncer le décryptage complet du génome humain ! Et nous savons bien que dans deux ans environ on annoncera la fin de cette colossale aventure? Qu'en est-il en

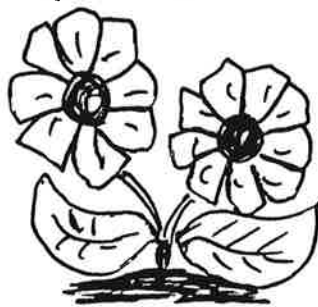
vérité d'une des plus grandes aventures de la biologie humaine ?

En fait le projet est poursuivi

simultanément par un consortium public (**Human Genome Project = HGP**) et

par une société privée américaine: **Celera**. Cette dernière a évidemment des besoins de communication, pour rassurer ses actionnaires et ses clients, ce qui entraine parallèlement des informations en provenance du consortium public qui ne peut laisser ignorer la progression de ses travaux dont bénéficie d'ailleurs Celera, qui utilise largement, pour l'assemblage des fragments séquences, les informations publiées en continu par HGP.

Le consortium public groupe les efforts d'une vingtaine de laboratoires dans 6 pays principaux : les USA, dont la contribution au séquençage est de l'ordre de 55 à 60%, l'Angleterre contribuant pour 33%, le Japon pour 10%, la France pour 2,5%, l'Allemagne pour 1,5% et la Chine pour 1%. Si la contribution de la France, qui pour l'essentiel se résume au chromosome 14, peut paraître faible dans le séquençage



n'oublions pas que l'assemblage des séquences, qui est une phase capitale, est beaucoup aidé par la carte des marqueurs satellites de **Généthon** publiée dans sa forme finale en 1996. La première carte physique, oeuvre des chercheurs du CEPH et de **Généthon** fut aussi un élément important dans cette tentative.

Ce qui a été publié jusqu'ici sur l'ensemble du génome, estimé à 3,2 gigabases (3,2 milliards de bases), constitue des brouillons, comportant quelques erreurs et beaucoup de « trous ». Ces trous, généralement dans des régions non codantes, sont au nombre d'environ 140000. La précision finale visée est d'une erreur par 10 000 bases soit une précision de 99,99% ce qui exige que chaque base soit séquencée 10 fois; or aujourd'hui la couverture est de l'ordre de 7 à 8. Au total on considère qu'environ 90% du séquençage du génome humain est déjà réalisé. L'achèvement en 2003, dans les conditions visées, est assurée grâce aux grands progrès techniques réalisés; la cadence de séquençage est aujourd'hui de 1 000 bases par seconde pour l'ensemble du consortium, ce qui veut dire que l'équivalent d'un génome entier est séquencé en moins de 6 semaines. Les difficultés encore à vaincre résident dans le clonage de certaines régions, ce qui explique les délais nécessaires pour la finition. Tel qu'il a été publié le premier « brouillon » de la séquence du génome humain a déjà fait la preuve de son utilité: une trentaine de gènes de prédisposition à des maladies ont déjà été clonés grâce à cette version imparfaite.

Dans l'état actuel de ce qui est déjà réalisé, beaucoup de travail peut être entrepris. L'annotation de ce qui est séquencé fait déjà l'objet de nombreux travaux et tend d'abord à déterminer les séquences codant pour des gènes, ce qui ne représente pas plus de quelques pour cent de l'ensemble. N'oublions pas que si un millier environ de gènes codant pour des protéines impliquées dans des maladies génétiques sont identifiés

il reste à en identifier encore 4000 (ou 5000?) non encore localisés, ou localisés de façon trop imprécise. Souvent pour des maladies rares il faut encore procéder à la collecte de familles atteintes ce qui est parfois très difficile. Dans le domaine neuromusculaire, la myosite ossifiante en est un bon exemple mais il en existe des centaines d'autres. Le but final, qui est de connaître le rôle physiologique des protéines intervenant dans des processus pathologiques est encore rarement atteint. Un énorme travail de physiologie cellulaire est devant nous, car, même si le génome ne comporte environ que 30.000 gènes, ce sont des centaines de milliers de protéines dont le rôle est à déterminer; un gène pouvant correspondre à plusieurs protéines. Par exemple le gène de la dystrophine, par le jeu de différents promoteurs correspond en fait à 6 protéines différentes. La comparaison avec d'autres organismes déjà totalement séquencés, et offrant des possibilités expérimentales (ce qui est évidemment exclu pour l'espèce humaine), est un précieux atout. La levure, le nématode *Coenorhabditis elegans*, la drosophile, bientôt la souris sont des auxiliaires précieux. La réalisation de modèles chez la souris pour les différentes maladies sont généralement indispensables pour la mise au point de thérapeutiques. Enfin la connaissance de la structure des protéines en trois dimensions offrent des perspectives considérables pour la pharmacologie, dans certains cas. Sans attendre la finition, dans deux ans, du séquençage, beaucoup de choses sont déjà possibles et de nombreux laboratoires à travers le monde sont à l'oeuvre.

Les solutions thérapeutiques à apporter aux maladies génétiques connaissent avec le séquençage du génome une formidable ouverture, mais c'est le premier acte d'une pièce qui en comporte beaucoup d'autres. Les découvertes qui chaque jour s'accumulent en thérapie génique liée aux greffes de cellules souches et aussi en chimie combinatoire associée au criblage à

haut débit, sont de formidables outils qui demain, sans aucun doute, aboutiront à la guérison de maladies jugées, il y a peu, incurables. Mais il ne faut pas dissimuler l'énorme travail qui reste encore à faire. On connaît le chemin à parcourir mais il faut méthodiquement le parcourir, des chemins de traverse restant toujours possibles à découvrir pour accélérer les choses.