



par Bertrand JORDAN

médecine/sciences 1994 ; 10 : 1154-6

*Fugu story*

C'est en 1991, au cours d'un déjeuner à Kyoto avec Mitsui Yanagida, que j'appris le nouveau thème de recherche choisi par Sydney Brenner : un poisson, *Fugu rubripes*, jusque-là surtout connu pour la place privilégiée qu'il occupe dans la cuisine japonaise. Je pensai d'abord à une plaisanterie : Yanagida, excellent biologiste moléculaire de la levure, est aussi un grand farceur, et nous étions justement en train de déguster de délicieuses tranches de Fugu cru... Mais l'information était bel et bien exacte. Cette option me parut saugrenue. Certes Sydney Brenner nous a déjà joué un tour de ce genre en décidant, il y a plus de vingt ans, de se consacrer à l'étude d'un animalcule jusque-là obscur, le nématode *Caenorhabditis elegans* ; l'énorme travail accumulé depuis a porté ses fruits, et ce petit ver sera très certainement le premier organisme multicellulaire à voir son génome (100 mégabases, soit la taille d'un chromosome humain) entièrement séquencé. L'œuvre de bénédictin qu'a représenté la détermination de toutes les connexions nerveuses du nématode et l'élucidation de leur ontogenèse n'a pas encore débouché sur des découvertes renversantes, mais c'est peut-être pour dans quelques années. En tout cas le pari de *Caenorhabditis* paraissait dès le départ sensé : prendre l'eucaryote multicellulaire le plus simple possible et le soumettre aux feux croisés de l'étude génétique, de la dissection anatomique et de l'analyse moléculaire au niveau de l'ADN, c'était à l'évidence une entreprise défendable [1]. Il n'en est pas de même pour *Fugu rubripes* : ce poisson inconnu, pour lequel n'existe au-

cun embryon d'étude génétique (d'ailleurs quasiment inabordable pour un ensemble de raisons pratiques), a pour seul atout un génome plus compact que le nôtre. Selon une série de mesures effectuées au début des années soixante, son génome haploïde correspondrait à 0,4 picogrammes d'ADN par cellule, contre 3 picogrammes pour l'homme. Détermination ancienne, mais effectuée de manière assez précise : les globules rouges étant nucléés chez les Poissons, il avait suffi de les dénombrer dans un échantillon de sang puis de doser, par les bonnes vieilles méthodes colorimétriques, la quantité d'ADN contenue dans un certain volume. Des mesures complémentaires sur le sperme avaient confirmé la valeur obtenue. A l'époque, ces chiffres n'avaient pas autrement surpris. Le génie génétique n'était pas né, nos connaissances sur les gènes étaient fort somnolentes, fondées pour l'essentiel sur quelques données obtenues chez les bactéries. La plupart d'entre nous pensaient, comme Jacques Monod, que ce qui était vrai pour *Escherichia coli* était vrai pour l'éléphant. L'on imaginait que, sauf exception, la quantité d'ADN indiquait le nombre de gènes, et personne n'était choqué à l'idée qu'un obscur petit poisson au minuscule cerveau ait huit fois moins de gènes que *Homo sapiens*. Nous n'en sommes plus là. L'intime parenté d'organismes apparemment fort différents est devenue manifeste : alors qu'à la fin des années soixantedix, après cinq ans de clonage et de séquençage, l'homme et la souris apparaissaient encore comme deux mondes séparés, nous savons mainte-

nant que pratiquement chaque gène humain a un homologue murin (et *vice versa*), et que leurs séquences codantes sont souvent les mêmes, à quelques pour cent près. Notion qui n'a d'ailleurs pas encore atteint le grand public, souvent surpris d'apprendre que la souris (sans parler des primates) est presque identique à l'homme de ce point de vue...

Tout diodon qu'il soit, le Fugu fait partie des vertébrés, et, à ce titre, nous pensons maintenant qu'il doit avoir un jeu de gènes à peu près équivalent au nôtre. Voilà donc un génome particulièrement compact, puisque – si les chiffres cités plus haut sont exacts – il comporte en tout 400 mégabases d'ADN au lieu de nos 3000. C'était bien le motif avancé par Brenner pour le choix de ce système biologique. L'argument me semblait pourtant mince, et je soupçonne que je n'étais pas le seul. Lancer ainsi un lourd projet sur un organisme quasiment inconnu, sans autre justification que la compacité présumée de son génome, était-ce bien raisonnable ? Les données acquises sur le Fugu seraient-elles vraiment transposables aux autres vertébrés, ou faudrait-il décrypter l'embryologie, le comportement et la génétique de cet organisme pour leur donner un sens ? Pour tout dire, ce programme m'apparut un peu comme la dernière lubie d'un grand esprit habitué à avoir raison contre tout le monde et voulant frapper un dernier coup avant sa retraite.

C'est à Singapour que j'eus l'occasion de voir de plus près ces travaux – et de réviser quelque peu mon jugement. Un congrès organisé par la

*Asian Pacific Society of Bioscientists* y rassemblait à la mi-août 1994 quelques centaines de chercheurs provenant des quatre « petits dragons » (Singapour, Taiwan, Corée du Sud, Hong Kong) ainsi qu'une poignée d'Australiens (mais aucun Japonais); ce colloque réunissait également des stars comme Robert Gallo, Michael Bishop ou Robin Weiss. Sydney Brenner était lui aussi de la partie – en tant que membre de la maison car, outre son laboratoire à Cambridge et son groupe à San Diego, ce diable d'homme dirige également une équipe de six personnes à Singapour. Dans cette « Suisse de l'Asie » au développement foudroyant (dix pour cent de croissance annuelle, et un niveau de vie maintenant supérieur à celui de la Grande-Bretagne, son ancienne métropole), la recherche biologique fondamentale était restée insignifiante jusqu'à la fondation il y a sept ans de l'*Institute of Molecular and Cellular Biology*. Somptueusement installé, mené par un directeur énergique, Chris Tan, et bénéficiant d'un budget à faire rêver un directeur des Sciences de la Vie (cent vingt millions de francs par an, salaires compris, pour deux cent trente personnes), cet institut a réussi à attirer d'excellents chercheurs qui publient dans *Cell*, *EMBO J*, *Nature*, *PNAS*... et dont les travaux devaient dominer le colloque.

L'exposé que fit Sydney, et quelques discussions avec ses collaborateurs, m'ont permis de mieux mesurer les premiers résultats du travail sur le génome du Fugu et d'en apprécier les retombées possibles. Les deux équipes impliquées dans le projet (celle de Cambridge, une dizaine de personnes, et celle de Singapour; le groupe de San Diego travaille sur d'autres thèmes) ont commencé par vérifier la taille du génome de ce poisson. La méthode a consisté tout simplement à compter les clones révélés lorsqu'on criblé une banque génomique de Fugu avec des sondes déjà caractérisées provenant du même organisme et dont on sait (par une analyse en *Southern blot*) qu'elles correspondent à un gène unique. Une banale règle de trois incluant le nombre de clones examinés dans le criblage, la taille moyenne des seg-

ments insérés et le nombre de positifs trouvés indique une valeur de 384 mégabases. La précision de l'accord avec les 400 mégabases déduites des anciens dosages est sûrement fortuite, mais cela confirme bien la taille du génome de notre diodon. Pour évaluer l'effectif total des gènes, une approche astucieuse combinant séquençage partiel et exploitation des bases de données a été employée [2]. Six cents clones pris au hasard dans une banque génomique (construite à partir d'ADN fragmenté par ultrasons) ont subi un séquençage partiel, deux ou trois cents nucléotides. On peut alors comparer ces « étiquettes » (bien qu'il s'agisse ici d'ADN génomique et non d'ADNc) à l'ensemble des séquences géniques de mammifères contenues dans les bases de données. Une partie d'entre elles correspond à des séquences répétées : le génome du Fugu contient des minisatellites, des microsatellites et bien sûr de multiples copies d'ADN ribosomique, mais pas d'éléments de type ALU ou L1. Pour les autres, l'on observe, soit des analogies très fortes, indiquant une excellente conservation des séquences codantes, soit aucune ressemblance, signifiant que le segment de Fugu considéré n'est pas une séquence codante, ou contient un gène dont l'équivalent n'a pas encore été séquencé chez les mammifères. Finalement les 130 kb séquencés (en 600 petits morceaux) s'avèrent contenir au total environ une kilobase de séquence codante connue, soit près de un pour cent. Or l'ensemble des séquences géniques connues chez les mammifères représente trois mégabases (une fois les redondances éliminées) sur un génome de 3000 mégabases. Si nous avions séquencé au hasard des fragments d'ADN humain selon les modalités décrites ci-dessus, nous devrions donc, après comparaison, retrouver dans cet échantillon un pour mille de séquences codantes connues. Chez le Fugu, la proportion de séquences géniques est dix fois plus grande, ce qui indique un nombre total de gènes équivalent puisque le génome est dix fois plus petit. Le calcul exact suggère une valeur proche de 80 % du chiffre humain, soit de 50 à 80 000 gènes. Les hypothèses de départ sont donc

validées, et le Fugu présente bel et bien un assortiment de gènes équivalent au nôtre mais codé par dix fois moins d'ADN. Les travaux menés depuis deux ou trois ans montrent comment cette compacité est atteinte. Pour les cinq ou six gènes déjà séquencés chez le Fugu, on retrouve presque toujours la même organisation que chez les mammifères: nombre d'exons et position des introns sont respectés. Mais ces derniers sont beaucoup plus petits, en général longs seulement d'une centaine de nucléotides; de sorte qu'un gène « moyen » mesure de deux à trois kilobases, introns compris. Quant aux distances intergéniques, elles semblent osciller autour d'une valeur analogue. Du coup, le Fugu devient immédiatement utile pour l'étude de gènes humains nouvellement découverts. Ces derniers s'étendent en effet souvent sur des distances décourageantes: sans atteindre les trois millions de bases du gène de la dystrophine, nombreux sont ceux qui occupent cent ou deux cents kilobases, à raison de dizaines de petits exons perdus dans un vaste désert intronique. Le déchiffrement d'un tel ensemble est malaisé, et le raccourci du passage par l'ADNc moins utile qu'il n'y paraît: il est souvent difficile d'obtenir un clone complet. De plus, les phénomènes d'épissage alternatif sont fréquents dans ces grands gènes, l'on n'est donc jamais sûr d'avoir trouvé tous les exons. Une tactique payante peut donc être d'isoler et de séquencer l'homologue Fugu du gène en question. La séquence codante ainsi obtenue sera très proche de celle de l'homme ou de la souris, et la disposition des exons généralement identique. Une fois ces derniers définis chez le poisson, il sera facile de vérifier leur présence chez l'homme et de détecter d'éventuelles différences – tout comme d'obtenir la séquence humaine grâce à un déchiffrement sélectif guidé par la connaissance de l'organisation du gène. Le gène de la Huntingtine mesure ainsi 25 kb chez le Fugu, contre plus de 300 chez l'homme; quant à celui de la dystrophine il dépasse un peu les 100 kb – ce qui est énorme pour le diodon mais autrement plus abordable que

les trois mégabases de l'homme ! Notons que, jusqu'à maintenant, tous les gènes cherchés dans ce système ont pu être trouvés, et que même une entité apparemment aussi spécialisée que la tétrahydrocannabinol hydroxylase (qui métabolise le produit actif du hachisch) a son gène chez le Fugu...

Un autre aspect de ce travail concerne la conservation des relations de proximité entre les gènes [3] - ce que Sydney Brenner appelle *adjacency* par répugnance à employer le terme de synténie qui tend à désigner la conservation de grandes régions. On sait qu'entre l'homme et la souris, séparés par cent quarante millions d'années dans l'évolution, les homologues entre zones chromosomiques s'étendent en général sur dix à vingt mégabases - c'est-à-dire que dans ces intervalles on retrouve les mêmes gènes, souvent dans la même disposition, comme si les chromosomes d'une espèce étaient un *patchwork* de ceux de l'autre (ou *vice versa*). C'est environ mille cinq cents millions d'années qui nous séparent du diodon, on peut donc imaginer que les segments homologues s'étendent sur une à deux mégabases - soit cent à deux cents kilobases en ADN de Fugu, puisque ce dernier est dix fois plus compact. Cela semble effectivement être le cas, et a fait l'objet d'une communication lors du dernier *Cold Spring Harbor Genome Mapping and Sequencing Meeting*. Le nombre d'exemples étudiés est certes encore insuffisant, mais cela ouvre d'intéressantes perspectives pour la recherche de gènes impliqués dans une maladie. L'analyse génétique «localise» en général cette dernière à quelques centimorgans, quelques mégabases près ; il faut alors étudier en détail la région ainsi délimitée et faire l'inventaire des gènes qu'elle contient, pour examiner ensuite l'état fonctionnel de chacun d'eux chez les malades. Une telle zone, en l'état actuel de nos connaissances sur le génome humain, contiendra en

général quelques dizaines de gènes dont seulement un ou deux connus. Leur mise en évidence demandera beaucoup d'efforts, compte tenu de la dispersion des exons dans ce grand intervalle. L'alternative maintenant offerte est d'isoler le segment correspondant chez le Fugu (une banque de YAC, avec des segments insérés allant de cent à deux cents kb, serait idéale de ce point de vue, et des efforts sont en cours pour la réaliser) : l'inventaire des séquences codantes contenues sur ce segment relativement court et maniable sera bien plus facile que chez l'homme.

L'on peut enfin spéculer - et Sydney Brenner ne s'en est pas privé, au cours d'une intervention prévue pour trente minutes et qui devait durer près d'une heure et demie - sur l'équivalence fonctionnelle entre gènes de Fugu et gènes de mammifères, et sur les façons d'en tirer parti. Imaginons, en effet, que l'on remplace dans une souris un gène par son homologue tiré du diodon, sous forme de clone génomique contenant les régions adjacentes 5' et 3'. Si ce gène fonctionne chez la souris (ce qui est très vraisemblable), et s'il y est correctement contrôlé (pari plus hasardeux mais non stupide), cela voudra dire qu'il porte toutes les séquences régulatrices en *cis* nécessaires. L'identification de ces dernières pourra alors être réalisée en comparant les séquences des régions non codantes du gène murin et de celui du poisson : on peut compter sur mille cinq cent millions d'années pour faire diverger toutes les régions non directement impliquées dans la régulation ou le codage de la protéine.

Revenons, pour terminer, sur la signification générale de cet ensemble de résultats. Il devient de plus en plus scabreux d'attribuer un rôle essentiel aux 95 % du génome humain consacrés aux introns et aux séquences intergéniques : le poisson en contient dix fois moins, ce qui ne l'empêche nullement de faire fonctionner un jeu de gènes à peu près aussi com-

plexe que le nôtre et, d'ailleurs, très proche de lui. Difficile dans ces conditions d'attribuer à notre *junk DNA* une fonction cachée - à moins d'y loger la conscience ou l'âme, ce qui semble peu probable, d'autant que nous le partageons avec la vache et la souris. Non, il faut se rendre à l'évidence : notre ADN est encombré d'un magma de séquences superflues mais suffisamment peu nuisibles pour que l'évolution ne les ait pas éliminées. Cet exemple parmi beaucoup d'autres\* nous montre que, contrairement aux apparences, la nature n'est pas parfaite. Notre plongée dans les mystères du vivant le confirme : décidément, selon l'expression profonde de Richard Dawkins, le « Grand horloger » est aveugle... ■

\* On pourrait aussi parler des voies de transduction et de leur incroyable complexité, fruit d'un bricolage prolongé avec acharnement durant des milliards d'années...

## RÉFÉRENCES

1. Labouesse M. C. *Elegans*, les promesses d'un petit animal intelligent : « small is beautiful ». *médecine/sciences* 1994 ; 10 : 337-41.
2. Brenner S, Elgar G, Sanford R, Macrae A, Venkatesh B, Aparicio S, Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* 1993 ; 366 : 265-8.
3. Little P. Small and perfectly formed. *Nature* 1993 ; 366 : 204-5.

## Bertrand R. Jordan

Directeur de recherche au Cnrs, responsable du groupe « Structure du Génome et fonctions immunitaires ». CIML, Inserm/Cnrs, case 906, 13288 Marseille Cedex 9, France.

## TIRÉS A PART

B.R. Jordan.