

Patrick Forterre : « **L'origine des génomes modernes** »

Le génome renferme la très grande majorité des informations nécessaires à la vie des organismes sous la forme d'une ou de plusieurs molécules géantes d'acide nucléique. L'origine des génomes se confond donc avec l'origine de la vie elle-même ; or, pour le moment, nous ne savons absolument pas comment celle-ci est apparue sur notre planète. Nous ne le saurons peut-être même jamais, car l'origine de la vie est un événement historique, et les historiens du vivant, comme les autres, sont complètement dépendants des reliques fossiles pour déterminer les événements du passé. En absence de telles reliques, certains de ces événements resteront toujours hors de notre portée. Je n'aborderais donc qu'en passant le problème de l'origine des tout premiers génomes, pour concentrer mon propos sur l'origine des génomes « modernes ». L'objectif est de reconstituer les dernières étapes de l'évolution qui ont abouti, il y a deux ou trois milliards d'années, aux êtres vivants et aux génomes tels que nous les connaissons actuellement. Cet objectif est moins ambitieux que celui de résoudre le problème mythique de nos origines, mais il est beaucoup plus réaliste (bien que toujours très ardu) et surtout, des progrès considérables ont été faits dans ce domaine ces dernières années, tout d'abord grâce aux avancées des connaissances en biologie moléculaire, et plus récemment, grâce au séquençage des génomes.

Les génomes modernes : cellules et virus

Passons tout d'abord en revue les génomes actuels ; ils présentent à la fois une grande uniformité et une grande diversité. Les génomes de tous les organismes sont constitués par des molécules géantes d'acides nucléiques. Ces molécules sont des polymères linéaires formés par l'association de briques élémentaires, appelés nucléotides, constitués d'une base aminée, d'un sucre, et d'un acide phosphorique (Figure 1). On distingue deux types d'acides nucléiques, l'ARN, acide ribonucléique, dont le sucre est le ribose (découvert au *Rockefeller Institute of Biochemistry*) et l'ADN, acide désoxyribonucléique, dont le sucre est le désoxyribose.

Les génomes de tous les organismes cellulaires sont constitués d'ADN sous forme dite « double-chaîne » dans laquelle deux polymères linéaires de nucléotides sont enroulés l'un autour de l'autre pour former la fameuse « double hélice ». L'ADN est formé par la polymérisation de quatre nucléotides distincts qui se caractérisent par quatre bases aminées différentes correspondant aux lettres A (adénine), T (thymine) G (guanine) et C (cytosine). Leur disposition linéaire dans la molécule d'ADN détermine le message génétique et les quatre lettres ATGC forment ce que l'on appelle parfois l'alphabet du vivant. Pour la diversité, l'ADN génomique des organismes cellulaires peut se présenter sous formes de molécules linéaires ou circulaires. Tous les organismes dit « eucaryotes » dont les cellules possèdent un noyau, et dont les gènes sont souvent morcelés en régions codantes et non codantes, ont des ADN linéaires, localisés à l'intérieur des chromosomes. Par contre, la plupart des organismes dont les cellules n'ont pas de noyau, Archées et Bactéries, traditionnellement appelés « procaryotes » (noyau primitif) ont des ADN circulaires – la double hélice d'ADN étant refermée sur elle-même.

La diversité des génomes est beaucoup plus grande dans le cas des organismes viraux. Si les génomes de nombreux virus sont composés d'ADN double brins, linéaire ou circulaire, tout comme les génomes cellulaires ; d'autres virus ont un génome composé d'un seul brin d'ADN circulaire, et d'autres encore, très nombreux, ont un génome composé d'un ou de deux brins d'ARN linéaire. L'ARN ressemble beaucoup à l'ADN, avec toutefois deux différences majeures, le désoxyribose est remplacé par le ribose, un

sucres apparentés, et l'une des quatre bases aminées, la thymine (T), est remplacée par une autre base, l'uracile (U) (Figure 1).

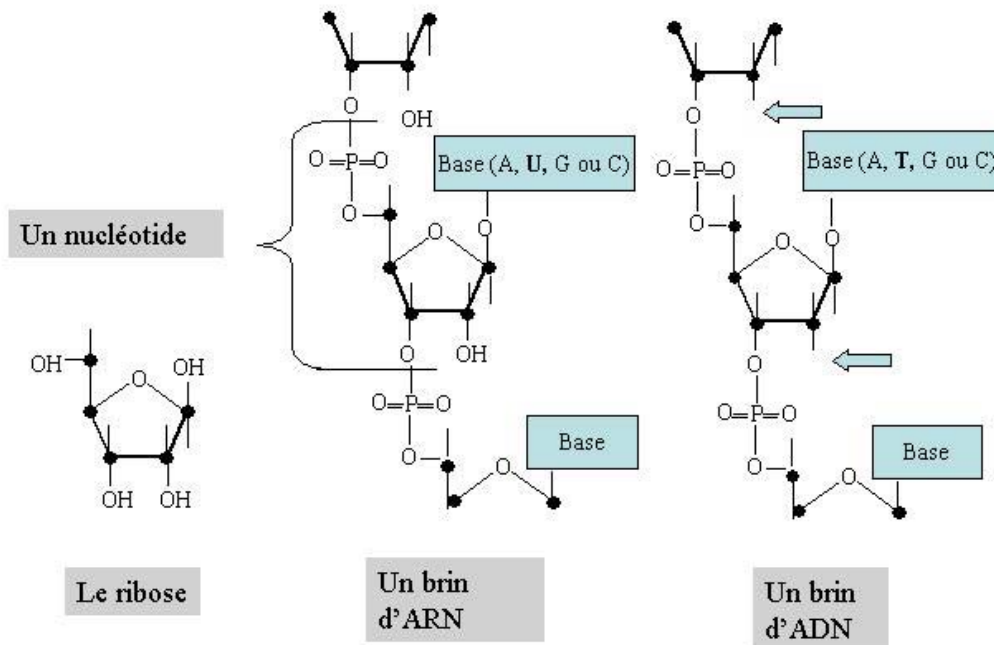


Figure 1

Les virus ne doivent pas être oubliés lorsque nous parlons de l'origine des génomes et de l'origine de la vie en général. Contrairement à une vision largement répandue, je pense (et c'est une idée de plus en plus partagée) que les virus sont bien des êtres vivants à part entière. Didier Raoult et moi-même avons ainsi proposé récemment de diviser le monde vivant en deux grands groupes : les organismes cellulaires, dont les génomes codent pour des ribosomes, et les organismes viraux, dont les génomes codent pour des capsides. Les ribosomes sont des complexes macromoléculaires, constitués de protéines et d'ARN, qui permettent aux organismes cellulaires de fabriquer leurs protéines. Les virus, qui ne possèdent pas de ribosomes, fabriquent leurs protéines en parasitant des organismes cellulaires pour avoir accès à leurs ribosomes. Ils se multiplient en produisant de nombreuses copies de leurs génomes qui sont empaquetés sous forme de virions. La capside, caractéristique des organismes viraux, correspond à la coque protéique qui protège le génome viral au sein du virion et permet sa dissémination dans la nature.

Comment ces deux formes de vie, organismes produisant des ribosomes et organismes produisant des capsides, sont-elles apparues ? Pourquoi existe-il des génomes à ADN, et d'autres à ARN ? Est-il possible d'imaginer des formes de vie plus anciennes qui auraient donné naissance aux formes de vie actuelles ? Voilà le type de questions auxquelles un petit nombre de chercheurs de par le monde essaient d'apporter des réponses.

Le monde à ARN

Si les génomes actuels sont composés d'ARN ou d'ADN, quels sont ceux qui sont apparus en premier ? Tous les évolutionnistes sont aujourd'hui d'accord sur la réponse à cette question : l'ARN a précédé l'ADN au cours de l'évolution. Cette idée s'est progressivement imposée depuis une trentaine d'année pour toute une série de raisons.

Notons tout d'abord que, sur le plan chimique, l'ADN est un ARN modifié. Le ribose, présent dans l'ARN, est un sucre « normal », comme le glucose ou le saccharose, il possède quatre fonctions chimiques alcool (OH) pour cinq atomes de carbone (Figure 1). Par contre, le désoxyribose (présent dans l'ADN) est un sucre atypique qui correspond à une forme dérivée du ribose ; il a perdu un atome d'oxygène et ne possède plus que trois fonctions alcools. De même, la thymidine (T), présente dans l'ADN, est une forme modifiée de l'uracile (U), présent dans l'ARN (un groupement méthyl, CH₃, a été ajouté sur la molécule d'uracile) (Figure 2). Il semble donc logique de penser que la forme chimique normale, l'ARN, est apparue au cours de l'évolution avant la forme dérivée, l'ADN.

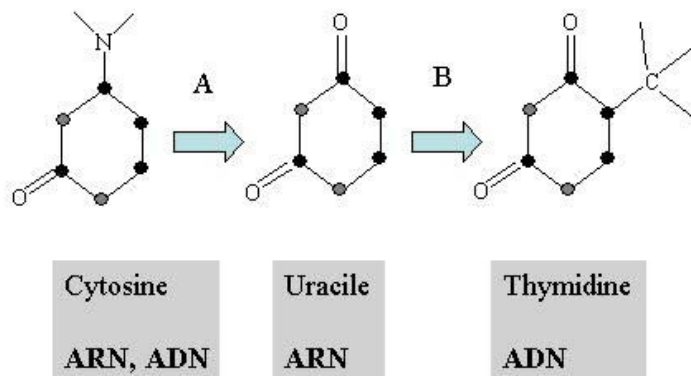
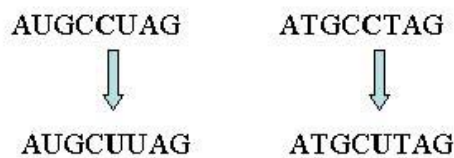


Figure 2



Le processus de transformation de l'ARN en ADN est en fait récapitulé par le métabolisme cellulaire ; celui-ci produit dans un premier temps les quatre ribonucléotides (rA, rU, rG, rC) qui vont s'assembler pour former l'ARN (Figure 3). Ces ribonucléotides vont ensuite être transformés en désoxynucléotides (dA, dT, dG, dC) précurseurs de l'ADN. Nous avons vu que les nucléotides (ribo ou désoxyribo) sont formés par l'addition d'un sucre, d'un acide phosphorique et d'une base azotée. Dans la molécule d'ARN, deux des quatre fonctions alcool du ribose sont utilisées pour associer entre eux les ribonucléotides par l'intermédiaire de l'acide phosphorique, et une troisième est utilisée pour fixer la base azotée. Il reste donc un oxygène libre pour chaque nucléotide, précisément celui qui est absent dans la molécule d'ADN qui ne possède donc pas d'atome d'oxygène libre (Figure 1). Nous verrons plus loin que cette différence chimique entre l'ADN et l'ARN est cruciale pour plusieurs raisons.

Dans les cellules actuelles, la transformation des ribonucléotides, précurseurs de l'ARN, en désoxyribonucléotides, précurseurs de l'ADN, se fait grâce à l'action séquentielle de deux enzymes, la première, appelée ribonucléotide réductase, va enlever l'un des oxygènes du ribose pour produire le désoxyribose (Figure 3). La deuxième, appelée thymidylate synthase, va ensuite transformer l'uracile du désoxyribonucléotide dU en thymidine, pour produire la lettre T (dT), spécifique de l'ADN (Figure 3). Ces deux modifications chimiques (dont nous verrons plus loin la signification), sont donc

catapultées par des enzymes spécifiques, qui sont absolument nécessaires pour passer de l'ARN à l'ADN. Il est donc encore une fois logique de penser que l'ARN a existé avant l'ADN (c'est-à-dire avant l'apparition des enzymes en question).

rA* = Adénine + ribose + trois acides phosphoriques = un nucléotide tri-phosphate

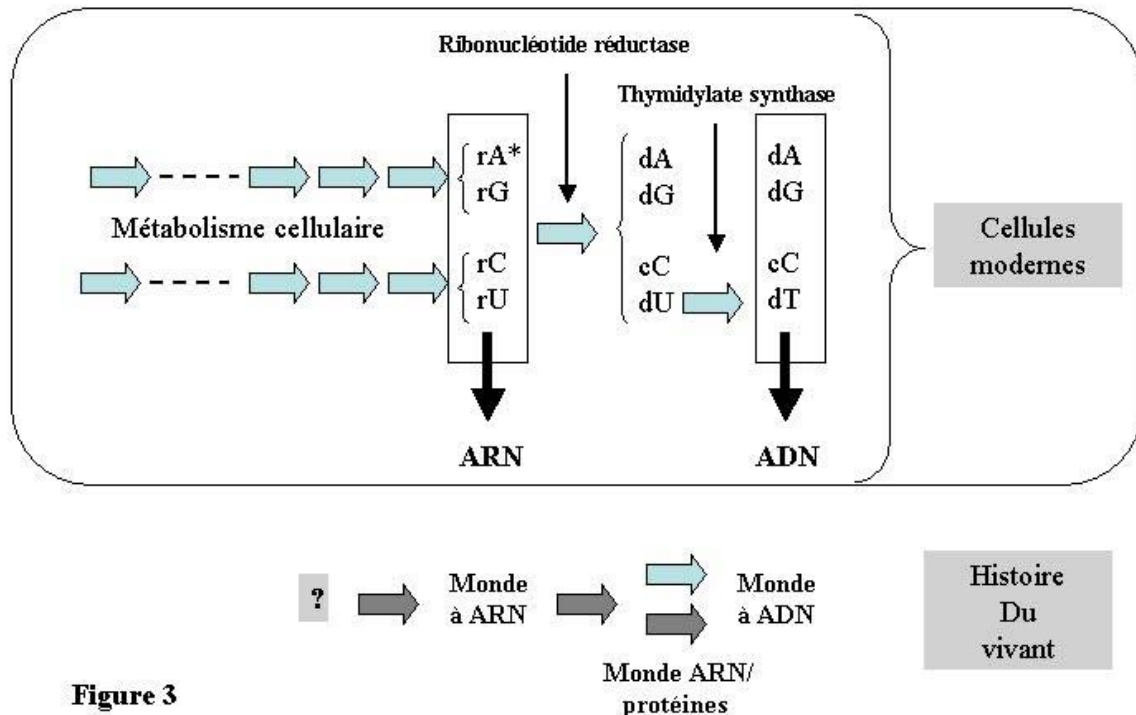


Figure 3

Un autre argument, tout à fait indépendant, va dans le même sens. Contrairement à l'ADN, l'ARN peut, dans certains cas, catalyser des réactions chimiques, exactement comme une enzyme. Les « enzymes » constituées d'ARN sont appelées ribozymes, pour les distinguer des enzymes classiques constituées de protéines. Un ribozyme est formé par un brin d'ARN qui se replie sur lui-même dans l'espace en trois dimensions pour former une structure globulaire, avec un site actif capable de fixer une autre molécule pour la transformer (ce qui correspond à une catalyse chimique). L'ADN ne peut pas catalyser de telles réactions chimiques, car l'oxygène du ribose qui a été éliminé dans l'ADN intervient de façon décisive dans la catalyse par l'ARN. En effet, cet oxygène est le seul atome réactif présent dans le ribose.

La découverte des propriétés catalytiques de l'ARN a résolu un problème qui a beaucoup perturbé les premiers biologistes moléculaires lorsqu'ils réfléchissaient à l'origine de la vie. Un problème qui rappelait celui de l'œuf et de la poule. Pour répliquer l'ADN, nous avons vu qu'il faut des protéines qui vont, entre autres, catalyser la formation des nouveaux brins (des polymérase). Or, ces protéines sont fabriquées sur la base de l'information génétique (le plan de montage) contenu dans l'ADN. Pour faire de l'ADN, il faut des protéines, pour faire des protéines il faut de l'ADN (pour faire un œuf il faut une poule et *vice versa*), comment s'en sortir ? L'ARN a fourni la solution : il est à la fois l'œuf et la poule, capable de porter une information génétique (tout comme l'ADN) et de catalyser des réactions chimiques (tout comme les protéines). On pouvait donc imaginer un « monde à ARN » ou des molécules d'ARN jouant le rôle de gènes, possédaient l'information pour fabriquer d'autres molécules d'ARN jouant le rôle d'enzymes, les secondes fabriquant les premières en se basant sur leur propre information. Pour certains auteurs, les virus à ARN seraient des reliques de ce monde à ARN.

On voit que dans le monde à ARN, non seulement l'ADN était absent, mais les protéines étaient facultatives. En fait, des travaux récents ont montré que les protéines telles que nous les connaissons à l'heure actuelle sont apparues après l'ARN. C'est la résolution de la structure du ribosome par cristallisation aux rayons X, au début des années 2000, qui a permis d'aboutir à cette conclusion. Cette structure a montré que l'association des acides aminés en longues chaînes linéaires, pour former les protéines, est catalysée par l'ARN du ribosome. Autrement dit, **le ribosome est un ribozyme** ! Du coup, les protéines modernes n'ont pas pu apparaître avant l'ARN, puisque leur formation dépend d'une molécule d'ARN.

A ce stade, nous avons beaucoup avancé dans notre reconstitution du passé. Nous pouvons en effet diviser l'histoire du vivant sur notre planète en trois périodes : le monde à ARN avant l'invention des protéines modernes (que j'ai tendance à appeler le premier âge du monde à ARN, en référence à JR Tolkien), le monde à ARN/Protéine, avant l'invention de l'ADN (le deuxième âge du monde à ARN) et le monde actuel (le troisième âge) avec son ménage à trois : ARN, protéines et ADN (Figure 3).

Comment sont apparus les premiers génomes à ARN ? C'est la question qui nous rapproche le plus des origines de la vie, et, bien évidemment, c'est celle dont nous ignorons encore la réponse ! L'origine des premiers ribonucléotides (les briques de l'ARN) reste mystérieuse, personne n'a encore réussi à les produire en laboratoire. Devant cette difficulté, certains chercheurs ont avancé l'idée selon laquelle l'ARN aurait été précédé par une autre macromolécule, plus simple, mais également capable de porter l'information génétique. Plusieurs candidats ont été proposés pour ces ancêtres de l'ARN ; mais, au final, ils ne semblent pas devoir être plus faciles à produire en conditions prébiotiques (au laboratoire) que l'ARN lui-même. Par ailleurs, quelques progrès ont été faits récemment dans la synthèse artificielle du ribose. On en revient donc plutôt aujourd'hui à l'idée d'une origine basée sur l'ARN.

Toutefois, l'idée d'un monde originel uniquement peuplé de molécules d'ARN est généralement abandonnée. Les premières molécules d'ARN, sans doute très courtes, étaient probablement déjà associées à des acides-aminoés. Ces derniers, tout au moins les plus simples d'entre eux, sont en effet assez faciles à fabriquer en conditions prébiotiques ; ils sont même présents dans l'espace interstellaire. La synthèse des premiers ribonucléotides, et leur assemblage en petits ARN, a pu être catalysé par des petites protéines « anciennes », elles-mêmes fabriquées par des mécanismes chimiques (et non biologiques). Ces premières synthèses ont pu être accélérées par des catalyseurs minéraux, et l'énergie nécessaire a pu être fournie par des polyphosphates, présent dans les milieux volcaniques. Le tout a pu se produire au sein de vésicules lipidiques, qui ont permis de concentrer les acteurs de ce premier bricolage moléculaire.

Des chercheurs américains ont montré récemment que l'addition d'ARN à des vésicules lipidiques leur confère des propriétés inattendues ; mises en présence de vésicules « vides », les vésicules à ARN capturent les lipides de ces dernières et grossissent à leur dépend, ce qui peut s'apparenter à une amorce de sélection naturelle ! Il a certainement fallu une très longue période évolutive (et une compétition féroce) au cours du premier âge du monde à ARN, pour passer de ces premières vésicules (proto-cellules) aux cellules à ARN qui ont hébergé l'ancêtre de nos ribosomes modernes, c'est à dire un ARN capable de fabriquer des protéines en lisant le message porté par un autre ARN (dans nos cellules modernes, c'est toujours un ARN, copie de l'ADN, l'ARN messager, qui vient porter le message de ce dernier aux ribosomes).

Les enzymes protéiques, composées de vingt acides aminés différents, aux propriétés chimiques variées, sont des catalyseurs beaucoup plus efficaces que les ribozymes à ARN. On comprend donc bien pourquoi elles ont progressivement remplacé les ribozymes comme catalyseurs dans la plupart des réactions du métabolisme au cours du second âge du monde à ARN. On pense que les premières protéines étaient de

petite taille, de séquence plus ou moins aléatoire, et qu'elles servaient principalement à stabiliser les ribozymes ou à augmenter le répertoire de leurs activités catalytiques. Au moment où l'ADN est apparu, on n'en était plus là, des protéines de grandes tailles, synthétisées avec une grande précision, devaient être présentes dans les cellules à ARN de cette époque. Nous pouvons l'affirmer, car les protéines qui sont nécessaires pour passer de l'ARN à l'ADN, ribonucléotide réductase et thymidylate synthases, sont de telles enzymes, complexes et précises. C'est particulièrement vrai pour les ribonucléotides réductases. Ces enzymes utilisent un mécanisme réactionnel très complexe pour enlever l'oxygène libre du ribose. Les enzymologistes qui étudient ces protéines considèrent que cette réaction n'a jamais pu être catalysée par un ribozyme, car elle fait intervenir un groupement chimique très réactif qui aurait attaqué et inactivé la molécule d'ARN formant le ribozyme. L'ADN n'a donc pu apparaître qu'après une longue période d'évolution du deuxième âge du monde à ARN, après l'apparition des enzymes modernes telles que nous les connaissons aujourd'hui.

Une autre réflexion nous conduit à la même conclusion. Les cellules à ARN de l'époque devaient déjà posséder toutes les enzymes des voies métaboliques complexes qui aboutissent aux ribonucléotides (AUGC) puisque, comme nous l'avons vu, les précurseurs de l'ADN sont formés à partir des précurseurs de l'ARN (Figure 3). Les cellules à ARN de la fin du deuxième âge devaient donc être très élaborées. Cette idée est difficile à admettre pour beaucoup de biologistes qui pensent que l'ARN est une molécule trop instable pour pouvoir porter une information suffisante pour la construction d'une cellule complexe. Ces biologistes se basent pour en arriver à cette conclusion sur l'observation des virus à ARN actuels qui ont tous de petits génomes, et qui répliquent ces génomes en faisant un assez grand nombre d'erreurs. Les plus grands génomes à ARN modernes connus, ceux du virus du SRAS par exemple, ne sont formés en effet que de 30 000 ribonucléotides environ, contre 160 000 désoxyribonucléotides pour le plus petit génome cellulaire connu. Je pense toutefois pour ma part que les virus à ARN actuels ne sont pas de bons modèles lorsque l'on essaye de se représenter les cellules du monde à ARN. Les virus à ARN modernes ont intérêt à répliquer leurs génomes de façon imprécise afin de muter à grande vitesse pour échapper aux défenses immunitaires de leurs hôtes. Des chercheurs ont toutefois montré que l'on pouvait modifier leurs enzymes en laboratoire pour les forcer à fabriquer de l'ARN avec une grande fidélité. Il ne faut pas oublier non plus que les virus à ARN produisent des protéines aussi complexes que les virus (ou les cellules) à ADN.

Il est vrai que l'ADN est plus stable que l'ARN, en effet, l'oxygène libre du ribose présent dans l'ARN, très réactif, peut parfois s'attaquer aux liaisons entre nucléotides, ce qui aboutit à casser en deux la molécule d'ARN. Je pense toutefois que l'on a tendance à surestimer l'instabilité de la molécule d'ARN. Chez les Eucaryotes, dont les gènes possèdent souvent de nombreuses régions non codantes, les gènes peuvent être transcrits en très longues molécules d'ARN (plusieurs millions de ribonucléotides) qui sont suffisamment stables pour rester présentes plusieurs jours dans nos cellules. Il semble donc tout à fait envisageable d'imaginer un monde de cellules complexes dont les génomes étaient constitués de molécules d'ARN.

A quoi ressemblait le génome des cellules du monde à ARN ? Était-il constitué de molécules d'ARN simple ou double brins, en un seul morceau ou fragmenté comme chez certains virus modernes à ARN ? Je pense pour ma part que différents types cellulaires devaient sans doute co-exister à cette époque, avec différents types de génomes à ARN. Il nous faut sans doute imaginer toute une biosphère peuplée de cellules à ARN, et déjà composé de proies, de prédateurs et de parasites engagés dans une lutte sans merci dominée par la sélection naturelle. L'origine des virus pourrait dater de cette époque. Les premiers virus à ARN ont pu dériver de cellules à ARN parasites qui ont perdu leur propre capacité à fabriquer leurs protéines et sont devenus ainsi dépendants de leur cellule à ARN hôte. Ils ont pu au contraire dériver de fragments génomiques à ARN qui se sont autonomisés pour devenir infectieux. Dans les deux cas, l'étape clef

dans l'origine des virus a été l'apparition des capsides permettant la formation de virions pour le transfert des gènes viraux d'une cellule à l'autre.

L'origine de l'ADN

Comment l'ADN est-il apparu dans cette biosphère ancestrale de cellules à ARN complexes accompagnées de leurs virus ? Pourquoi, au bout du compte, l'ARN a-t-il été remplacé par l'ADN en tant que forme unique de génome cellulaire ? Pendant longtemps, la réponse à cette question semblait aller de soi pour les biologistes moléculaires ; l'ADN avait remplacé l'ARN parce qu'il est plus stable (grâce à la perte de l'oxygène libre du ribose) et parce qu'il peut-être réparé plus facilement, grâce au remplacement de la lettre U par la lettre T. Ce dernier point mérite une petite explication. Tous les jours, une réaction chimique se produit spontanément dans nos cellules, qui transforme au hasard et de façons spontanées certaines lettres C (cytosine) en U (uracile) (Figure 2). Des U apparaissent donc sporadiquement dans l'ARN et l'ADN. Ces transformations de C en U vont entraîner des mutations qui seront la plupart du temps nocives. Or nos cellules possèdent un mécanisme moléculaire sophistiqué capable de reconnaître les U dans la molécule d'ADN (ils n'ont rien à y faire) et de corriger l'erreur pour remettre un C à la place. Bien sûr, cela n'est possible que pour un génome à ADN. Dans un génome à ARN, le mécanisme moléculaire de réparation que je viens d'évoquer n'aurait aucun moyen de reconnaître un mauvais U (issu d'une modification d'un C) d'un U bien à sa place (Figure 2).

La stabilité intrinsèque des génomes à ADN, et la possibilité de corriger les mutations de C vers U, ont eu une conséquence très importante pour tout le reste de l'évolution. Ils ont rendu possible l'augmentation de la taille des génomes à ADN. En effet, la quantité d'information que peut porter un génome est directement proportionnelle au degré de fidélité avec laquelle cette information est reproduite. Or, cette augmentation de la taille de génomes était nécessaire pour que l'évolution biologique produise des organismes de plus en plus complexes, jusqu'à l'homme, « aboutissement de l'évolution ! ». L'invention de l'ADN a donc constitué un progrès indéniable, et, pour beaucoup, cette constatation servait aussi d'explication, le progrès n'est-il pas inéluctable ? Je caricature à peine, pour de nombreux biologistes moléculaires peu au fait des mécanismes de l'évolution, l'apparition de l'ADN était une nécessité, ce n'était pas vraiment la peine de chercher à comprendre pourquoi et comment il était apparu en premier lieu.

Et pourtant les choses ne sont pas si simples, ce n'est pas pour que des organismes extraordinairement complexes se promènent sur la terre aujourd'hui que l'ADN est apparu. De même, ce n'est pas parce que les oiseaux ont besoin de plumes pour voler que leurs ancêtres, les dinosaures, avaient des plumes ! Si nous voulons vraiment comprendre l'origine des génomes à ADN, il faut comprendre quel avantage sélectif le remplacement de l'ARN en ADN a apporté au premier mutant à ADN chez qui cette transformation s'est produite.

Certains biologistes pensent que la plus grande stabilité de l'ADN suffit à expliquer la sélection de ce premier mutant. Cela me paraît peu probable, nous avons vu que l'ARN n'est pas si instable que cela ; de plus la durée de vie (entre deux divisions) des cellules à ARN était sans doute assez faible. Il me semble donc difficile d'admettre que l'augmentation de la stabilité de son génome ait été suffisante pour donner au premier mutant à ADN un avantage décisif dans le combat pour l'existence. Par ailleurs, la taille de son génome n'a pas pu augmenter instantanément ! De même, si l'on réfléchit au passage de l'ADN contenant la lettre U à l'ADN moderne (qui a dû se produire dans un second temps), on réalise que le premier mutant possédant la lettre T ne pouvait pas posséder le mécanisme de réparation dont nous avons parlé plus haut qui permettait de reconnaître les U provenant de C dans l'ADN et de corriger l'erreur. L'évolution ne

pouvait pas avoir mis en place un tel système, il aurait fallu qu'elle prévoie à l'avance l'apparition d'ADN-T !

Je pense donc que l'intérêt d'avoir un génome plus stable, et pouvant être réparé plus facilement, ne s'est manifesté que sur le long terme, lorsque des populations de cellules à ADN ont coexisté sur une longue période avec des populations de cellules à ARN. Il a fallu laisser du temps aux organismes à ADN pour que leurs génomes grandissent et pour qu'apparaisse chez certains d'entre eux un mécanisme de correction des mutations C vers U, mettant à profit l'absence de U dans l'ADN. Autrement dit, la possibilité pour les génomes à ADN d'augmenter leur taille n'explique pas pourquoi ils sont apparus, mais pourquoi, sur le long terme, ils ont permis aux cellules à ADN d'éliminer les cellules à ARN. Ce ne sont que des retombées (bénéfiques pour nous, c'est incontestable) de l'apparition de l'ADN. De même, le vol des oiseaux est une retombée du fait de posséder des plumes ; plumes dont le rôle devait être très différent chez les dinosaures de celui qu'elles ont acquit chez les oiseaux (les évolutionnistes parlent d'exaptation).

L'hypothèse virale

J'ai proposé, il y a quelques années, une solution possible au problème de l'origine de l'ADN. Selon moi, le premier « mutant » à ADN aurait pu être un virus. En effet, les virus ont un intérêt évident à modifier chimiquement leur génome ; cela leur permet d'échapper aux mécanismes de défense des cellules qui visent à détruire le génome viral. De très nombreux virus modernes utilisent en effet cette stratégie. Par exemple, le bactériovirus T4, qui s'attaque à la bactérie bien connue *Escherichia coli*, a modifié toutes les lettres C (cytosine) de son génome à ADN par les lettres HMC (hydroxymethyl-cytosine). L'enzyme virale qui réalise cette transformation est apparentée à l'une de celles qui modifient la lettre U en T chez tous les êtres vivants. Dans l'hypothèse virale pour l'origine de l'ADN, on voit bien que le premier mutant à ADN-U a immédiatement (au cours de son existence) obtenu un avantage sélectif immédiat sur ces congénères en se mettant à l'abri des enzymes cellulaires qui pouvaient dégrader son ARN. Le même phénomène a pu se produire pour la deuxième transition : un virus à ADN-U a obtenu un avantage immédiat en transformant son ADN-U en ADN-T, mettant ainsi son génome à l'abri des enzymes qui étaient apparues entre temps dans les cellules pour détruire l'ADN-U viral. On peut donc imaginer une période de l'évolution où des cellules à ARN étaient infectées par des virus dont les génomes étaient constitués soit d'ARN, soit d'ADN-U, soit d'ADN-T. Il faut noter à ce sujet, que certains virus actuels possèdent un génome à ADN-U, ces virus pourraient être des « fossiles vivants » qui témoignent de l'étape intermédiaire de la transition entre ARN et ADN (Figure 4).

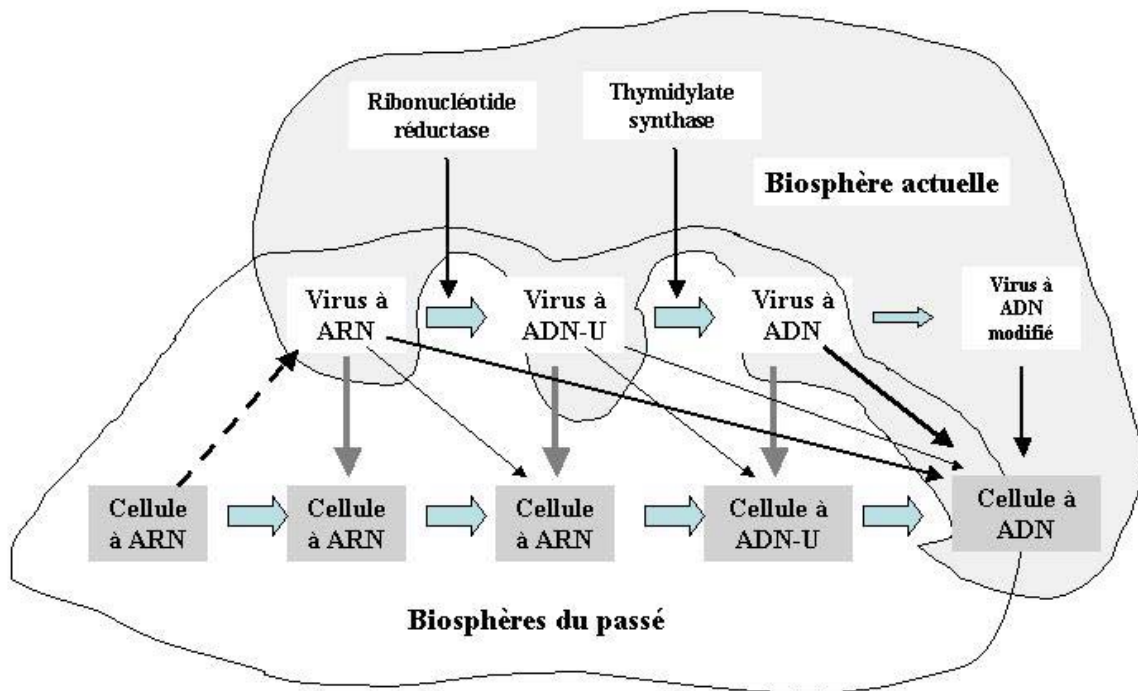


Figure 4

Si le premier organisme à ADN était bien un virus, il reste à imaginer comment ce virus a pu obtenir les précurseurs (désoxyribonucléotides) nécessaires à la fabrication de cet ADN. On peut penser que la ribonucléotide réductase est tout d'abord apparue dans le conflit entre virus et cellules comme un moyen d'inactiver les ribonucléotides de l'adversaire en les transformant en produits inactifs (les désoxyribonucléotides). La polymérase à ARN d'un virus aurait alors acquis par mutation la capacité de polymériser ces désoxyribonucléotides, réalisant ainsi la copie de l'ARN viral en ADN (les rétrovirus actuels possèdent toujours une enzyme, appelée transcriptase inverse capable de réaliser ce type de réaction.)

Pour moi, l'hypothèse d'une origine virale de l'ADN a le grand avantage d'expliquer pourquoi on observe aujourd'hui une grande diversité de mécanismes de réplication de l'ADN chez les virus et, en particulier, pourquoi de nombreux organismes viraux à ADN codent pour des enzymes de la réplication de l'ADN qui n'ont pas de parents proches (et parfois pas de parent du tout) chez les organismes cellulaires. En effet, si l'ADN est apparu en premier lieu dans le monde viral, l'évolution des mécanismes de réplication de l'ADN a dû se produire dans un premier temps essentiellement dans cette virosphère ancestrale. Différentes protéines qui auparavant répliquaient l'ARN ont pu être recrutées indépendamment dans différentes lignées de virus à ADN pour répliquer ce dernier. Une grande variété de mécanisme de réplication de l'ADN et leurs protéines associées ont dû ainsi s'élaborer progressivement dans la virosphère ancestrale. Par la suite, seul un sous-ensemble de tous ces mécanismes aurait été transmis aux organismes cellulaires, expliquant l'existence actuelle d'enzymes de la réplication que l'on ne trouve que chez les virus.

Si l'ADN est apparu en premier lieu dans le monde viral, il faut imaginer comment il a pu être ensuite transféré aux cellules. Mon hypothèse préférée est celle d'un virus à ADN infectant de façon chronique une cellule à ARN, et qui aurait progressivement pris le contrôle de cette dernière. Ce virus aurait tout d'abord perdu sa capsid (par

mutation) et se serait retrouvé sous forme de plasmides (petit chromosome circulaire) dans la cellule à ARN. Il aurait ensuite capturé progressivement les gènes à ARN de son hôte, par transcription inverse, et se serait ainsi transformé en génome cellulaire à ADN.

Le génome de LUCA et l'origine des trois domaines cellulaires

Dans le scénario que je viens d'évoquer, une question importante se pose, le transfert de l'ADN des virus aux cellules s'est-il produit avant ou après l'apparition du dernier ancêtre commun à tous les organismes cellulaires actuels (appelé LUCA en Anglais pour *Last Universal Cellular Ancestor*). Il a semblé pendant longtemps tout à fait évident que le génome de LUCA était constitué d'ADN double brins, comme ceux de tous les organismes cellulaires modernes. Cette certitude a été toutefois remise en question vers la fin du XXème siècle, suite à la comparaison des premiers génomes séquencés, celui d'une bactérie pathogène de l'homme, *Haemophilus influenzae* et celui d'un Eucaryote, la levure de boulangerie, *Saccharomyces cerevisiae*. Il est en effet apparu que les enzymes qui permettent la formation des nouveaux brins d'ADN au moment de la division cellulaire (on parle de réplication de l'ADN) étaient complètement différentes chez ces deux organismes. Elles étaient si différentes qu'elles ne pouvaient pas descendre d'enzymes ancestrales communes qui auraient répliqué l'ADN chez leur dernier ancêtre commun. Selon l'une des hypothèses visant à expliquer cette observation, l'ADN et les enzymes qui le répliquent, seraient apparues deux fois indépendamment au cours de l'évolution, une fois dans la lignée conduisant aux Bactéries et une seconde fois dans la lignée conduisant aux Eucaryotes. Dans cette hypothèse, le génome de LUCA était encore un génome à ARN.

Par la suite, il est apparu que les protéines de la réplication de l'ADN chez les Archées, le troisième domaine du vivant, étaient homologues de celles des Eucaryotes. Il existe donc deux types de protéines impliquées dans la réplication de l'ADN, un type bactérien et un type commun aux Archées et aux Eucaryotes. Si l'hypothèse virale de l'origine de l'ADN est correcte, deux transferts indépendants de l'ADN des virus aux cellules après LUCA pourraient expliquer cette observation, l'une dans la lignée bactérienne, l'autre dans une lignée commune aux eucaryotes et aux archées. On peut toutefois aussi imaginer un premier transfert avant LUCA et un second qui aurait remplacé les protéines de réplication ancestrales par de nouvelles protéines virales chez les Bactéries ou chez l'ancêtre des Archées et des Eucaryotes. Enfin, pourquoi la formation des génomes à ADN ne se serait-elle pas produite trois fois indépendamment, après LUCA ? Dans ce cas, les trois transitions des génomes à ARN vers les génomes à ADN auraient pu donner naissance aux trois domaines cellulaires actuels. Le transfert de l'ADN à partir de virus dont les génomes étaient très différents auraient ainsi abouti d'un côté aux génomes circulaires relativement simples des Archées et Bactéries, qui ressemblent en fait à de gros plasmides (certains plasmides actuels sont d'ailleurs beaucoup plus gros que les plus petits chromosomes bactériens) et de l'autre aux génomes complexes des eucaryotes, avec plusieurs chromosomes linéaires terminés par des structures curieuses appelées télomères. Il est difficile de choisir entre ces différents scénarios. Il existe en fait de nombreux points qui restent mystérieux ; par exemple, comment se fait-il que les protéines de réplication de l'ADN soient identiques chez les Archées et les Eucaryotes alors que leurs génomes sont si différents ? Nous avons ainsi en main plusieurs pièces d'un puzzle que nous avons encore du mal à rassembler.

L'origine du noyau des cellules eucaryotes, qui renferme leurs génomes, reste un grand point d'interrogation. La encore, certains auteurs ont proposé récemment des hypothèses qui donnent aux virus le premier rôle dans cette histoire. Selon eux, un virus à ADN, apparenté au virus de la variole, serait à l'origine de leur noyau. Ces virus forment en effet dans les cellules qu'ils infectent des mini noyaux entourés d'une membrane construite selon le même principe que la membrane nucléaire (à partir des membranes intracellulaires du réticulum endoplasmique). La découverte récente du

Mimivirus a relancé cette hypothèse, en effet ce virus géant, apparenté au virus de la variole, produit des usines virales qui ont la même taille que le noyau de la cellule qu'ils infectent (une amibe). Je pense pour ma part que les virus ont pu jouer un autre rôle, plus indirect, dans l'origine du noyau. Il est possible que celui-ci soit apparu en premier lieu pour protéger le génome cellulaire de l'attaque des virus. Peut-être une cellule (l'ancêtre des Eucaryotes) a-t-elle finalement réussi à retourner contre eux une stratégie d'origine virale, en transformant un mini noyau viral (destiné au départ à protéger l'ADN viral) en noyau cellulaire. La cellule eucaryote est en fait si compliquée, que je pense nécessaire de faire intervenir plusieurs gros virus pour expliquer son apparition. L'existence de multiples ADN et ARN polymérases nucléaires chez les Eucaryotes pourrait être une relique de ces multiples infections fondatrices.

Si les données de la génomique comparée n'ont pas encore permis de résoudre complètement l'énigme de l'origine des génomes, elles ont toutefois fait avancer le problème en soulevant de nouvelles questions, comme celle posée par l'existence de deux types de protéines de réplication dans le monde cellulaire. La biologie moléculaire a joué un rôle majeur en nous permettant de diviser l'histoire ancienne des génomes en plusieurs étapes bien définies: premier et deuxième âge du monde à ARN, apparition du monde à ADN. De façon inattendue pour beaucoup de biologistes, les virus, longtemps négligés par les évolutionnistes, se sont invités avec force dans ce débat. S'il s'avère qu'un, ou même plusieurs gros virus à ADN, sont à l'origine de notre génome, nous serons bien obligés de finir par considérer les organismes producteurs de capsides avec un peu plus de respect, Le culte de l'ancêtre nous conduira peut-être à placer une photo de virus sur notre autel domestique.

Références

RAOULT, D and FORTERRE, P. Redefining viruses : lessons from Mimivirus Nature Reviews Microbiology. 6 :315-319 (2008)

FORTERRE, P. and GRIBALDO, S. The origin of modern terrestrial life The HFSP Journal, 1, 156-168 (2007)

FORTERRE, P. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. 117, 5-16 (2006)

FORTERRE, P. Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc. Natl. Acad. Sci. 103, 3669-3374 (2006)

FORTERRE, P. The two ages of the RNA world, and the transition to the DNA world : a story of viruses and cells. Biochimie. 87, 793-803 (2005)

FORTERRE, P. L'Origine du génome La Recherche, (2004)

FORTERRE, P. Les virus ont-ils inventé l'ADN? Pour la Science, Juillet (2008)

FORTERRE, P. Microbes de l'enfer, Collection Regards, Edition Belin (2007)

Légendes des figures

Figure 1 : Structures chimiques schématiques du ribose, de l'ARN (un brin) et de l'ADN (un brin). Les traits représentent des liaisons chimiques. Les boules noires représentent des atomes de carbones, les atomes d'hydrogènes fixés aux carbones n'ont pas été représentés. Les flèches pointent vers la position de l'oxygène absent dans l'ADN

Figure 2 : Structures chimiques schématiques des bases dites pyrimidiques : C (cytosine), U (uracile) et T (thymine). Les boules noires représentent des atomes de carbones et les boules grises des atomes d'azote. Les atomes d'hydrogènes n'ont pas été représentés. Les flèches représentent des transformations chimiques ; A : déamination des cytosines, une réaction chimique spontanée qui transforme C en U, elle se produit en présence d'eau (lentement mais sûrement) chez les êtres vivants, B : transformation de U en T, cette transformation est catalysée par une enzyme, la thymidylate synthase, lorsque U est sous la forme du désoxyribonucléotide dU. Deux petites séquences d'ARN et d'ADN sont représentées en bas de la figure. Un changement de C vers U passe inaperçu dans l'ARN, mais pas dans l'ADN.

Figure 3 : Formation de l'ADN dans les cellules modernes et au cours de l'évolution des cellules ancestrales. Les flèches bleues indiquent des transformations enzymatiques catalysées par des protéines-enzymes.

Figure 4 : Les différentes étapes de la transition ARTN vers l'ADN dans l'hypothèse d'une origine virale de l'ADN. La flèche grise symbolise le transfert de l'ADN des virus aux cellules.