



*Séance solennelle de l'Académie des sciences / 11 octobre 2011
Discours des nouveaux Membres sous la coupole de l'Institut de France*

L'ARN, des séquences à l'architecture

Éric Westhof

Monsieur le Président, Monsieur le Vice-Président,
Madame et Monsieur les Secrétaires Perpétuels,
Chères Consoeurs, Chers Confrères,
Chers Amis, Mesdames, Messieurs,

Permettez-moi d'abord de vous exprimer mes vifs remerciements pour l'honneur que vous m'avez fait et la confiance que vous m'avez témoignée en m'élisant membre de l'Académie des sciences.

Depuis mon élection comme Correspondant en 1999, j'ai pu apprécier l'extraordinaire chance de pouvoir côtoyer et enrichir ma pensée auprès des chercheurs dont j'admirais les articles et les ouvrages. Louis Leprince-Ringuet, lors de sa leçon terminale au Collège de France, a dit en parlant des chercheurs : « On n'a rien à montrer à nos amis ». Les amitiés et rencontres scientifiques sont donc particulièrement précieuses.

La biologie moléculaire s'est construite autour d'un concept central : la transcription de l'information génétique contenue dans l'ADN génomique en molécules intermédiaires, les ARN messagers, qui sont ensuite traduits en protéines. Les protéines mobilisent l'ensemble des rôles structuraux, enzymatiques, et régulent également l'expression des gènes en assemblant les différents composants de la machinerie cellulaire et en interagissant avec eux. Cependant, déjà dans les années 80, la découverte par Thomas Cech et Sidney Altman des propriétés catalytiques de l'ARN avait ébranlé cette vision des mécanismes cellulaires centrée sur les protéines. Ces ARN catalytiques, appelés ribozymes, ont clairement démontré que les molécules d'ARN étaient capables de se replier dans l'espace et de former des architectures aussi complexes que celles des protéines et qu'elles étaient également capables d'action catalytique sur d'autres molécules.

L'ARN est donc la seule molécule que nous connaissons à posséder ces deux propriétés : être dépositaire de l'information génétique et agir chimiquement sur d'autres molécules. L'ADN, quant à lui, n'est pas capable de catalyse, et les protéines ne sont pas utilisées pour coder l'information génétique. En réalité, l'état actuel des connaissances nous pousse à considérer l'ARN comme étant une molécule-clé aux origines de la vie (le monde de l'ARN il y a près de 3,8 milliards d'années). En effet, l'ARN est la seule molécule capable d'être à la fois mémoire, ou génome comme dans les rétrovirus, ainsi que catalyseur chimiquement actif.

On peut même considérer que l'ADN a été « inventé » au cours de l'évolution par l'ARN. Dans chacune de nos cellules, les nucléotides de l'ADN résultent de la réduction chimique de nucléotides de l'ARN (et non l'inverse). L'évolution a donc sélectionné l'ADN pour sa stabilité chimique et maintenu l'ARN pour sa réactivité chimique. Les stupéfiantes structures cristallographiques du ribosome, particule qui synthétise les protéines, ont démontré que le ribosome est bien un ribozyme ou enzyme à base d'ARN, confortant encore plus le rôle central de l'ARN aux origines de l'évolution biologique. Nos travaux récents, avec des collaborateurs du Max Planck Institut de Göttingen, pointent clairement vers un rôle catalytique des molécules d'ARN constitutives des complexes ribonucléoprotéiques nécessaires à l'élimination des séquences ne portant pas l'information pour coder les protéines dans les ARN messagers des eucaryotes.

Les nouveaux développements technologiques de séquençage de génomes (ADN) et de transcriptomes (ARN) ont apporté deux faits marquants : tout d'abord, la partie de l'ADN codant des protéines est très faible (moins de 2% chez l'homme) ; toutefois, la très grande majorité du génome est bien transcrite mais en une multitude de molécules d'ARN. Tous ces transcrits sont par conséquent appelés des ARN non codants. Ces observations ont de nombreuses conséquences sur notre compréhension de la notion de gène et d'expression génomique.

De nos jours, l'expression génomique ne peut être comprise que comme une résultante complexe et dynamique d'interactions multiples entre ADN et ARN, entre segments codants et non-codants ainsi qu'entre ARN codants et ARN non codants. Ces interactions moléculaires entre molécules d'ADN et d'ARN d'un même génome déterminent et régulent l'expression des gènes et conduisent également à des remodelages du génome.

De nombreux mécanismes impliquant l'ARN sont à l'origine de la dynamique évolutive des génomes : les rétrogènes, les rétrotransposons, les séquences dites *Alu* indispensables pour reconstruire l'évolution des primates. Un génome est tout sauf un plan de construction immuable transmis de génération en génération. Les génomes sont très loin du « rêve débilisant du statu quo » craint par Soljénitsine et récemment cité par Jean Dercourt. Un génome subit constamment la pression de processus actifs tendant à réarranger les segments par duplication, insertion, et délétion. Et, beaucoup plus que les mutations ponctuelles des protéines et enzymes, ce sont ces réarrangements génomiques qui ont été déterminants lors de l'évolution des espèces.

On sait maintenant que des molécules d'ARN sont à l'origine de nombreuses régulations dites épigénétiques et de nouveaux ARN non codants sont décrits dans la littérature tous les mois. Lors des interactions hôtes-pathogènes, les réseaux moléculaires se construisent entre molécules d'ARN provenant de génomes différents et participent ainsi à la détermination du degré de virulence du pathogène. La compréhension quantitative de l'établissement et de la propagation de ces réseaux d'interactions moléculaires constitue le thème de recherche central du laboratoire d'excellence obtenu dans le cadre des Investissements d'Avenir et que j'anime à Strasbourg.

Les résultats actuels de la bioinformatique structurale et de la génomique illustrent cette phrase étonnante de Friedrich Nietzsche dans *Aurore* : « Les mains de fer de la nécessité qui secouent le cornet à dés du hasard jouent leur jeu en un temps infini ».

De nos jours, nous disposons de très nombreux génomes et cela permet les comparaisons systématiques de séquences de molécules aux fonctions proches. De plus, nous disposons, grâce aux développements en cristallographie et microscopie électronique, de nombreuses structures et architectures de molécules biologiques. Ces deux ensembles de données, séquences et structures, nous permettent d'apprécier l'étendue de l'échantillonnage moléculaire du « jeu des possibles », ainsi que le décrivait notre éminent collègue François Jacob.

Ce qui est frappant est le degré auquel l'évolution biologique exploite les propriétés physico-chimiques des molécules jusque dans ses plus fines subtilités. Les ARN sont formés à partir de quatre bases, et donc de quatre briques constitutives : la cytosine, l'uracile, l'adénine et la guanine. Je ne pense pas qu'il y ait une seule potentialité offerte par la chimie de ces molécules qui ne soit

utilisée lors de l'un ou l'autre des divers processus biologiques. Au cœur de la génétique on retrouve la stabilité des bases par rapport aux changements de formes tautomères et l'exquise précision géométrique des interactions par liaison hydrogène qu'elles peuvent former. Par exemple, les quartets de guanine ont longtemps été considérés comme des curiosités cristallographiques ; ils sont maintenant au centre des processus de maintenance des télomères des chromosomes. Des états protonés ou déprotonés, très rares, de certaines bases sont au cœur des mécanismes catalytiques de l'ARN.

Nous comprenons donc de plus en plus finement l'articulation de ces « mains de fer » et leur ancrage dans la physico-chimie et la chimie quantique. Cependant, l'ensemble de ces contraintes, et cela avec une diversité époustouflante, permet des accommodements moléculaires, parfois nombreux et subtils, souvent inattendus, qui maintiennent des architectures sensiblement identiques pour un nombre considérable de séquences différentes avec un nombre très restreint de résidus invariants. Donc, à nouveau, les interactions moléculaires, par leur diversité et leur plasticité, sont à l'origine du formidable potentiel d'évolution de ces molécules grâce à leur pouvoir d'ajustement et d'acceptation « d'erreurs », garantissant ainsi la robustesse des architectures moléculaires et leur fonctionnement dans des conditions extrêmes.

C'est la conjonction des structures spatiales adoptées par l'ARN et des processus physico-chimiques sous-jacents aux subtilités de l'analyse génomique qui permet cette compréhension. J'ai eu la chance d'apprendre la stéréochimie et la cristallographie aux côtés d'un pionnier des structures des acides nucléiques, alors à l'Université de Madison au Wisconsin (USA), M. Sundaralingam, tragiquement disparu lors du tsunami de 2004. Il avait une incroyable compréhension intuitive des molécules et de leurs interactions. C'est en France, surtout avec François Michel, du Centre de Génétique Moléculaire à Gif-sur-Yvette, que j'ai appris les comparaisons de séquences et les analyses génomiques. Sans l'un et l'autre je ne serais pas ici.

J'aimerais terminer par une phrase de Bernard Dujon, collègue et ami. Il écrit ainsi qu'« un génome est à la fois mémoire du passé, déterminant du présent et moteur de l'évolution à venir ». Nous avons maintenant les outils pour disséquer le passé, comprendre certains déterminants du présent, et, de plus en plus, décomposer les rouages essentiels du moteur de l'évolution.

Je vous remercie pour votre aimable attention.