

Histoire de la biologie moléculaire

P. BERCHE¹

RÉSUMÉ

Paradoxalement, c'est un physicien, Erwin Shrödinger, qui a montré la voie à la biologie au décours de la Seconde Guerre mondiale. La conséquence fut l'irruption des techniques de la chimie et de la physique dans les sciences du vivant qui ont métamorphosé le champ de la biologie. La biologie moléculaire commence par deux découvertes majeures : Oswald Avery montre en 1944 que les gènes sont constitués d'ADN, et Francis Crick et James Watson proposent en 1953 le modèle de la double hélice d'ADN à partir des données de diffraction des rayons X. On a d'abord identifié toute la machinerie enzymatique des cellules, notamment en utilisant des systèmes *in vitro* constitués d'extraits cellulaires : ADN et ARN polymérases, ARN messagers et de transfert, enzymes de restriction, transcriptase inverse... Conjointement, on a déchiffré le code génétique. Toutes ces avancées ont été à l'origine d'innovations technologiques qui ont joué un rôle majeur pour l'étude du vivant : clonage des gènes, séquençage de l'ADN des organismes, *Polymerase Chain Reaction*, synthèse des gènes, de virus et de bactéries, création de gènes nouveaux (*DNA shuffling*), clonage des organismes entiers (Dolly), *RNA silencing*, technologie CRISPR/Cas9. Ces progrès ont transformé profondément de nombreux domaines de la biologie, incluant le diagnostic médical et l'étude des maladies, la médecine légale, l'identification des individus et leur généalogie, la caractérisation de l'ADN ancien... La biologie moléculaire a été et sera la source de très nombreuses découvertes qui vont peut-être transformer notre vie.

MOTS-CLÉS : histoire, biologie moléculaire, clonage, séquençage, *Polymerase Chain Reaction*, synthèse des gènes, *DNA shuffling*, Dolly, *RNA silencing*, CRISPR/Cas9.

I. - INTRODUCTION

Les fondements de la biologie moderne reposent sur les idées de Charles Darwin, d'August Weismann et de Gregor Mendel. En 1859, Charles Darwin présente une nouvelle théorie de l'évolution des espèces vivantes apparues par sélection naturelle : les espèces s'adaptent continuellement aux différents environnements qu'elles rencontrent et évoluent par tâtonnements au cours de très longues périodes de temps (1). Travaillant sur le développement embryonnaire des œufs d'oursin comme modèle, August Weismann formule en 1885 la « théorie sur la continuité du plasma germinatif », où il distingue les cellules germinales qui transmettent la vie tout au long des générations, et les cellules somatiques qui constituent les corps « périssables » (2). L'hérédité serait portée par le « plasma germinatif », qui deviendra le génome au XX^e siècle. Un moine tchèque, Gregor Mendel, découvre les lois de l'hérédité, en travaillant sur les petits pois de jardin.

Il introduit le concept de caractères dominants et récessifs, qui sont à la base de la génétique naissante, mais ses travaux publiés en 1866 resteront ignorés jusqu'à l'orée du XX^e siècle (3). Par la suite, Hugo de Vries découvrira les mutations en 1900, et exhamera les travaux oubliés de Mendel. Suivra la théorie chromosomique de l'hérédité par Theodor Boveri et Walter Sutton en 1902, confortée par les travaux de Thomas Morgan, qui rattachera les gènes aux chromosomes observés lors des mitoses.

Jusqu'à la Seconde Guerre mondiale, on peut dire que ce sont surtout les naturalistes qui ont investi le champ de la biologie. La « biologie moléculaire » est un avatar de l'histoire de la biologie qui correspond en réalité à l'irruption des techniques de la chimie et de la physique dans le

¹ Institut Pasteur de Lille, 1 rue du Professeur Calmette, Lille 59000.



Fig. 1 - Erwin Shrödinger, prix Nobel de physique 1933, auteur de *What is life?*, 1944.

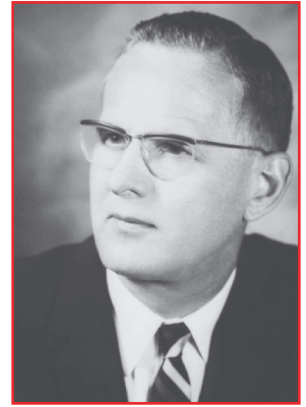
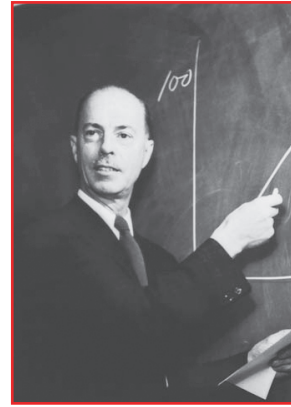
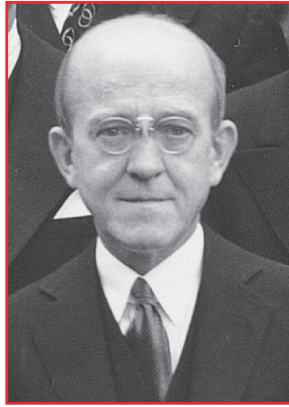


Fig. 2 - Oswald T. Avery, Colin McLeod et Maclyn McCarthy (de gauche à droite) au *Rockefeller Institute* (New-York) déterminent en 1944 la nature du « principe transformant » de Griffith, l'ADN (photos Wikimedia Commons).

domaine des sciences du vivant. La connaissance du vivant a été fortement influencée par la vision et les idées d'un physicien autrichien, prix Nobel et spécialiste de la mécanique quantique, Erwin Shrödinger (Figure 1), qui a publié en 1944 un petit livre intitulé *What is life ?* Cet ouvrage aura un impact profond sur plusieurs générations de biologistes (4). Shrödinger conçoit la vie comme un système très complexe, stockant et transmettant d'énormes quantités d'informations qui pourraient être compactées en un « code héréditaire » dans les molécules constituant les chromosomes. L'idée ancienne d'une force vitale qui animerait la matière organique disparaît. La voie est tracée. Il convient désormais de comprendre les stratagèmes qu'utilisent les cellules vivantes pour traiter les informations. La question cruciale est d'identifier les molécules du vivant permettant le traitement de ces informations.

II. - L'ADN SUPPORT DE L'HÉRÉDITÉ

De quoi sont faits les gènes ? C'est la question à laquelle les chimistes vont s'atteler à partir du début du XX^e siècle. Jusqu'aux années 1940, on a longtemps cru que les gènes étaient constitués de protéines, abondantes dans les cellules et douées d'extraordinaires propriétés catalytiques à l'origine de la synthèse de nombreux constituants cellulaires. En 1941, les travaux de George Beadle (1903-1989) et Edward Tatum (1909-1975) semblent conforter cette hypothèse : certaines mutations d'un champignon (*Neurospora crassa*) sont associées à la perte d'une enzyme spécifique, d'où le célèbre aphorisme : « un gène, une enzyme ».

Cependant, on savait depuis 1869 que les noyaux des cellules où sont localisés les chromosomes, sont formés de « nucléine », une substance acide découverte par un médecin suisse, Friedrich Miescher (1844-1895). Celui-ci l'isola des noyaux séparés de leucocytes provenant de plaies suppurées de patients récemment opérés (5). La nucléine (ou acides nucléiques) est constituée d'acide déoxyribonucléique (ADN), une grosse molécule compo-

sée de sucres et de phosphore, résistante aux protéases, et qui est le principal constituant du noyau cellulaire et des chromosomes. L'ADN est formé de très longues chaînes constituées d'une succession de quatre « briques » appelées nucléotides, l'adénine (A), la guanine (G), la thymine (T) et la cytosine (C), chacune reliée à un sucre, le désoxyribose, et au phosphore. Cet acide nucléique est « enroulé » de façon très compacte pour former les chromosomes et il est enrobé de protéines, les histones. Jusqu'en 1944, personne n'aurait misé sur l'ADN comme support de l'hérédité.

En 1928, un médecin de santé publique travaillant à Londres sur un vaccin contre le pneumocoque, Frederick Griffith (1879-1941), fait une observation qui va mettre sur la piste de la nature chimique des gènes. Cultivés à partir des crachats de patients atteints de pneumonie, les pneumocoques donnent des colonies lisses et sont virulents pour la souris à cause de la présence d'une capsule constituée de sucres qui les protègent de la phagocytose. Cette capsule est spontanément perdue au cours des repiquages *in vitro*, et les colonies prennent un aspect rugueux et perdent leur virulence pour la souris. Griffith a la surprise de constater qu'il peut restaurer la capsule et la virulence en mélangeant des bactéries vivantes avirulentes et sans capsule à des extraits bactériens tués par la chaleur provenant de souches capsulées virulentes. Cette « transformation » est donc liée à la présence d'un « principe transformant » provenant des extraits bactériens tués par la chaleur (6). C'est la première fois que l'on réussit à transmettre des caractères génétiques, en l'occurrence les gènes requis pour la fabrication d'une capsule sucrée, à l'aide d'une « substance » chimique contenue dans des bactéries mortes.

Ce phénomène facile à reproduire a été étudié de façon approfondie par des chimistes travaillant au *Rockefeller Institute* à New York, Oswald Avery (1877-1955), Colin MacLeod (1909-1972) et Maclyn McCarthy (1911-2005) (Figure 2). Pour déterminer la nature chimique du principe transformant, ces auteurs ont traité les extraits chauffés de bactéries capsulées avec diverses enzymes. En 1944,

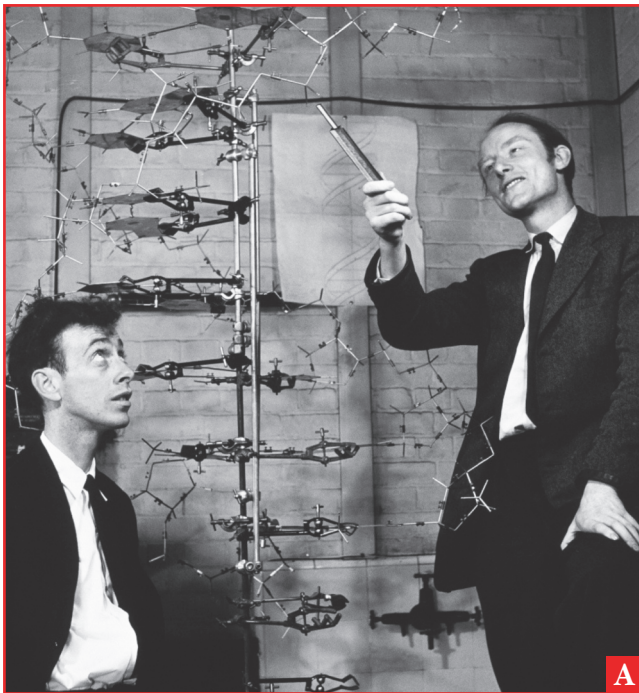


Fig. 3 - A. Francis Crick et James Watson et la structure de l'ADN par diffraction des rayons X. - B. Copie de l'article de Nature de 1953 sur le modèle de la double hélice d'ADN (photos Wikimedia Commons).

après des années de travail, ils mettent au jour que la capacité de synthèse de la capsule est réacquise malgré le traitement des extraits bactériens par des protéases et par des ARNases, enzymes détruisant les protéines et les ARN. En revanche, les enzymes détruisant l'ADN inhibent la transformation des pneumocoques. L'ADN est donc le support de l'hérédité. Cette découverte majeure sera rapidement confirmée, notamment en 1952 par les expériences d'Alfred Hershey (1908-1976) et Martha Chase (1930-2003), qui montreront que seul l'ADN de phages, et non leur coque protéique, permet à ces virus d'infecter des bactéries et de s'y multiplier. La découverte d'Avery est, à l'époque, un séisme conceptuel qui substitue l'ADN aux protéines comme support des gènes. Dès lors, la biologie moléculaire est née et la voie est ouverte à toutes les avancées technologiques permettant une meilleure compréhension du monde vivant.

III. - LA DOUBLE HÉLICE ET LE DOGME DE LA BIOLOGIE MOLÉCULAIRE

À l'instar de Louis Pasteur montrant que les cristaux d'acide tartrique dévient la lumière polarisée, le physicien allemand, Max von Laue (1879-1960), découvre en 1912 que les cristaux diffractent les rayons X, phénomène visualisé sur une plaque photographique par des taches symétriques. De complexes calculs permettent ensuite de reconstituer la structure tridimensionnelle des molécules cristallisées, c'est-à-dire leurs formes. Dès 1914, deux physiciens anglais, père et fils, Lawrence Bragg (1862-1942) et William Bragg (1890-1971), décrivent la structure du chlo-

ture de sodium par cette technique et ouvrent l'ère de la radiocristallographie X en créant à Cambridge un laboratoire spécialisé dans l'étude des molécules biologiques. Tous deux recevront le prix Nobel de physique en 1915.

On s'attachera d'abord à résoudre la structure des protéines cristallisées, puis des acides nucléiques. Le physicien britannique William Astbury (1898-1961) établit que la structure de l'ADN présente la forme d'un long filament comportant une succession d'unités répétitives empilées (les bases), régulièrement espacées de 0,34 nm. Dans les années 1950, Maurice Wilkins (1916-2004) et Rosalind Franklin (1920-1958) poursuivent ces travaux sur la structure de l'ADN au *King's College* (Londres) et obtiennent les premières images de l'ADN par diffraction des rayons X. L'interprétation de ces images sera l'œuvre du physicien britannique Francis Crick (1916-2004) et d'un jeune étudiant américain post-doctorant, James Watson (1928-), travaillant au prestigieux laboratoire Cavendish de l'Université de Cambridge (Figure 3). Impressionnés par la découverte sensationnelle de la structure hélicoïdale des protéines publiées en 1951 par le chimiste américain Linus Pauling (1901-1994), qui travaillait au *Cal Tech* (*California Institute of Technology*), Crick et Watson proposent en 1953, grâce aux données cristallographiques de Rosalind Franklin et de Maurice Wilkins, un modèle de structure de l'ADN dit en double hélice, où les deux brins d'ADN sont enroulés en spirale, formant un escalier en colimaçon. Ce modèle est basé aussi sur l'observation des biochimistes Erwin Chargaff (1905-2002) et James Norman Davidson (1911-1972), qui ont montré en 1949 que les nucléotides constituant l'ADN sont en quantités égales par groupes de deux : autant de thymine (T) que d'adénine (A), et autant de gua-

nine (G) que de cytosine (C). Crick propose l'existence d'un appariement des bases deux à deux, AT et GC, permettant ainsi la jonction entre les deux brins enroulés en double hélice. Le modèle de la double hélice fournit une explication lumineuse de la réplication de l'ADN au cours des divisions cellulaires, donnant une base moléculaire à l'hérédité des caractères transmis par la structure même de la molécule d'ADN. La découverte est publiée le 25 avril 1953 sous forme d'un très court article d'une seule page dans la revue *Nature* (7), conjointement avec les données de Rosalind Franklin (8). Crick postule l'existence d'un code génétique, avec un alphabet de quatre bases, expliquant la transmission conservée des caractères héréditaires à la descendance. Crick formule le « dogme central de la biologie moléculaire » (9) : l'information génétique est transmise à partir des acides nucléiques vers les protéines, l'ADN étant le support moléculaire d'une information qui s'exprime à travers des protéines, telles que les enzymes.

Cette découverte a fait l'effet d'un coup de tonnerre et marque un tournant majeur à l'origine de tous les développements de la biologie moléculaire. En 1962, James Watson et Francis Crick partagent le prix Nobel de physiologie/médecine avec Maurice Wilkins. La mort prématurée en 1958 à l'âge de 37 ans de Rosalind Franklin à la suite d'un cancer, ne lui permit pas de recevoir ce prix si mérité.

IV. - LA MACHINERIE ENZYMATIQUE DES CELLULES

Comprendre le fonctionnement de la machinerie enzymatique qui permet la synthèse des acides nucléiques et des protéines, sera le travail des biochimistes qui rafflèrent pendant plusieurs décennies une moisson ininterrompue de prix Nobel. En 1954, le biochimiste américain Paul Zamecnik (1912-2009) met au point un système de synthèse *in vitro* des protéines, en utilisant des extraits hépatiques riches de ribosomes et d'ATP, en présence des 20 acides aminés constituant les protéines, un système qui sera par la suite étendu aux extraits bactériens. C'est un progrès majeur qui entraînera de nombreuses découvertes. En 1956, les américains Severo Ochoa (1905-1993) et Arthur Kornberg (1918-2007), prix Nobel de physiologie/médecine en 1959, isolent des enzymes appelées polymérases capables de synthétiser de l'ADN et de l'ARN. La découverte de ces polymérases va permettre le développement de systèmes *in vitro* pour synthétiser des acides nucléiques et des protéines. À partir de 1961, on commence le décryptage du code génétique pressenti par Schrödinger. Tout est parti d'une observation inattendue de Marshall Nirenberg (1927-2010) et de son étudiant Johann Matthaei (1929-). Dans un système acellulaire préparé à partir d'extraits bactériens, l'ajout d'un ARN synthétique, polymère constitué uniquement d'uridine (polyU), induit la synthèse d'une protéine « monotone » constituée exclusivement de phénylalanine (polyPhe). Ainsi, le codon UUU reconnaît donc la phénylalanine. Dès lors, Nirenberg et Matthaei utilisent divers ARN synthétiques et le code sera déchiffré en moins

de cinq ans par ces deux chercheurs, et par Har Gobind Khorana (1922-2011) et Robert Holley (1922-1993), prix Nobel de physiologie/médecine en 1968. Toutes les protéines sont constituées de 20 « briques » ou acides aminés, organisés en chaînes plus ou moins longues. L'ADN est constitué de seulement 4 nucléotides (A, T, C, et G), qui sont les lettres du texte à décrypter. Il faut « trois lettres à la suite les unes des autres », un triplet (codon) de bases, pour former un mot « codant », c'est-à-dire un acide aminé. Le système de codons de 3 nucléotides à partir des 4 bases (ATCG) autorise 64 combinaisons différentes pour 20 acides aminés. Ceci signifie que le code génétique est « dégénéré », c'est-à-dire que plusieurs codons reconnaissent un même acide aminé. Chaque acide aminé est reconnu par un à 5 codons, au maximum. Enfin, il existe des codons de départ (*start*) et 3 codons d'arrêt (*stop*), qui ponctuent la lecture au début ou à la fin des gènes.

Ainsi, chaque cellule vivante possède un génome qui se présente comme un énorme livre de multiples molécules d'ADN, organisées en une succession ordonnée de « codons ». La lecture des codons de l'ADN est ainsi traduite en acides aminés séquentiellement ordonnés. De façon remarquable, le même code est retrouvé chez les bactéries, les champignons, les végétaux et les animaux, preuve éclatante de l'unicité du vivant. Apparus il y a 3,7 milliards d'années, les êtres vivants se sont développés avec une extraordinaire diversité selon leurs environnements. Tous sont régis par des gènes constitués d'ADN à l'origine de la fabrication de protéines cellulaires grâce au même code universel. Cette universalité du code qui apparente l'homme à toutes les espèces vivantes, permet aussi d'envoyer d'exprimer des gènes de végétaux ou d'animaux dans des bactéries ou des champignons.

V. - LA SYNTHÈSE DES PROTÉINES ET L'ARN MESSAGER

Comment l'ADN parvient-il à synthétiser des protéines ? On sait depuis 1934 que les protéines sont fabriquées dans le cytoplasme riche d'ARN, et non pas dans le noyau où réside la très grande majorité des molécules d'ADN. Dans les années 1940, est montrée l'existence d'une corrélation entre le taux de synthèse des protéines et la quantité d'ARN dans le cytoplasme. Le belge Jean Brachet (1909-1988) montre en 1955 que des cellules énucléées continuent à produire des protéines pendant plusieurs jours sans ADN. Dans le cytoplasme, l'ARN semble concentré dans de petits grains, extrêmement nombreux, les ribosomes observés au microscope électronique par le roumain George Palade (1912-2008), prix Nobel de médecine 1974.

Si la synthèse protéique a lieu dans le cytoplasme en l'absence d'ADN, il faut bien admettre que doivent exister des molécules intermédiaires permettant la synthèse des protéines. Évidemment, l'ARN abondant dans le cytoplasme est un bon candidat pour cette fonction. Cette hypothèse est confortée par la découverte, dans les années 1950, de virus uniquement composés d'ARN et de pro-



Fig. 4 - Jacques Monod et François Jacob (à droite), prix Nobel de physiologie/médecine 1965 (photos Wikimedia Commons).

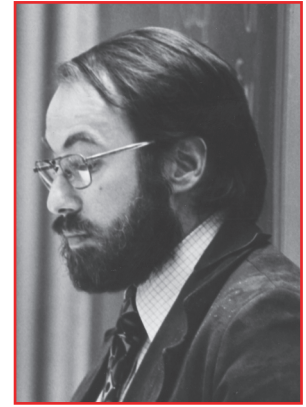
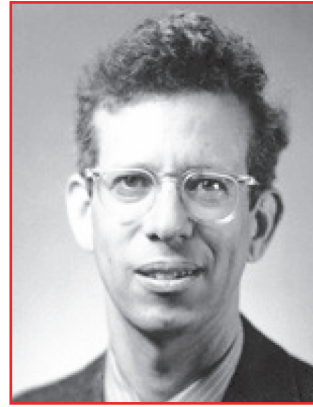


Fig. 5 - Les découvreurs de la transcriptase inverse : Howard Temin et David Baltimore (à droite) (photo Wikimedia Commons) prix Nobel de physiologie/médecine 1975.

téines. En 1956, Heinz Fraenkel-Conrat (1911-1999) et Gerhard Schramm (1910-1969) montrent que le virus de la mosaïque du tabac exclusivement constitué d'ARN, est capable de produire ses propres protéines qui lui confèrent un pouvoir infectieux, en l'absence d'ADN. Est lancée la quête de l'ARN intermédiaire entre l'ADN et la protéine synthétisée. En 1961, François Gros (1925-) montre l'existence d'une nouvelle famille d'ARN de taille variable, bien distincte de l'ARN des ribosomes (10). En même temps, François Jacob (1920-2013) et Sydney Brenner (1927-), dans le laboratoire de Mathew Meselson (1930-) en Californie, apportent la preuve directe qu'un ARN dit « messenger » s'associe aux ribosomes pour synthétiser les protéines (11). Ainsi, cet ARN messenger est l'intermédiaire entre l'ADN et les protéines. Il assure un recopiage fidèle de l'ADN et permet sa traduction en protéines. Le « dogme de la biologie moléculaire » devient alors : l'ADN est la matrice qui stocke l'information ; l'ADN est recopié (transcrit) en ARN messenger dont les codons sont lus séquentiellement et « traduits » en une séquence d'acides aminés ordonnés en protéines. Dans les années 1970, toute la signalétique de la lecture de l'ADN en ARN (« transcription » des gènes), puis de l'ARN en protéines (« traduction ») est dévoilée, avec notamment la découverte des ARN de transfert procurant les acides aminés à la chaîne peptidique en voie de formation.

On sait que l'activité de certains gènes peut varier selon les conditions environnementales. Comment cette modulation est-elle contrôlée ? Jacques Monod (1910-1976) s'est intéressé à cette question, à travers le phénomène de diauxie qu'il a décrit dans les années 1950. Des colibacilles cultivés dans un bouillon contenant deux sucres, vont utiliser préférentiellement l'un des deux sucres, puis après une phase de latence, consommer l'autre sucre. Il s'agit d'un mécanisme d'adaptation montrant que des gènes utilisant un sucre peuvent être activés ou réprimés selon les conditions de culture. Avec François Jacob, Jacques Monod va décortiquer ce phénomène. Les deux pasteuriens étudient l'activité du gène contrôlant la synthèse d'une enzyme, la β -galactosidase, qui scinde le lactose en glucose et galactose. En l'absence de lactose ou en présence d'une

source facilement disponible comme le glucose, la production de cette enzyme est inhibée par un gène « répresseur ». Monod et Jacob (Figure 4) introduisent le concept d'opéron qui associe des gènes dits « de structure », codant par exemple pour des enzymes nécessaires à la survie des bactéries, et des gènes régulateurs, « activateurs » ou « répresseurs » selon les conditions de l'environnement (12). Ces découvertes sont récompensées par le prix Nobel de physiologie/médecine en 1965, conjointement avec André Lwoff (1920-1994).

En 1970, le dogme de la biologie moléculaire de Crick est mis à mal par la découverte d'une enzyme « hérétique », la transcriptase inverse (ou rétrotranscriptase). Travaillant sur l'oncogenèse virale à l'université de Madison (Wisconsin), un jeune chercheur américain Howard Temin (1934-1994), âgé de 25 ans, observe que la répllication du virus du sarcome de Rous, un virus à ARN, est inhibée par l'actinomycine D, qui n'agit que sur l'ADN. Temin avance alors l'hypothèse hérétique que le cycle de multiplication du virus implique une étape utilisant l'ADN. En 1964, il propose un modèle de répllication de ce virus comprenant une phase pendant laquelle le virus est transformé en ADN et s'intègre dans les chromosomes de la cellule infectée sous la forme d'un « provirus », à l'instar du phénomène de lysogénie des bactéries infectées par certains bactériophages. Le virus intégré est alors transmis aux cellules-filles lors de la division cellulaire et peut être produit de façon intermittente par lecture de la copie insérée dans le chromosome cellulaire. Temin postule l'existence d'une enzyme capable de transformer l'ARN en ADN, enzyme qu'il met en évidence en 1970 : la transcriptase inverse du virus du sarcome de Rous. David Baltimore (1938-), un virologue spécialiste des polymérases cellulaires et virales travaillant au *Salk Institute* de La Jolla (Californie), découvre indépendamment cette même enzyme sur un autre rétrovirus, le virus de la leucémie murine de Rauscher. Cette découverte sera publiée conjointement dans l'édition du 27 juin 1970 du journal *Nature* (13, 14). Aujourd'hui, le phénomène de rétro-transcription de l'ARN (synthèse d'ADN à partir de l'ARN) apparaît universel dans le monde vivant. En 1975, Howard Temin et David Baltimore reçurent

le prix Nobel de physiologie/médecine pour leur contribution à la connaissance des virus oncogènes (Figure 5).

En 1977, une autre surprise sera rapportée par Richard Roberts (15) (1943-), Philip Sharp (16) (1944-) (Figure 6) et Pierre Chambon (17) (1931-), qui découvrent que, contrairement aux bactéries, les gènes des animaux et des végétaux sont fragmentés, avec des segments codant pour des séquences d'acides aminés (les exons) et des séquences non codantes (les introns). Un long ARN messager est recopié à partir de l'ADN, puis débarrassé de ses introns. C'est ce que l'on appelle « l'épissage » qui peut prendre diverses modalités pour un même gène, générant plusieurs ARN messager et donc plusieurs protéines pour un même gène.



Fig. 6 - Richard Roberts et Phillip Sharp (à droite) (photo Wikimedia Commons), les découvreurs des introns et de l'épissage alternatif, prix Nobel de physiologie/médecine 1993.

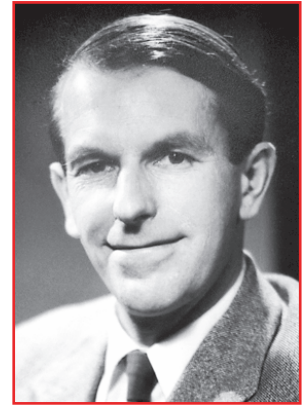


Fig. 7 - Frederick Sanger, pionnier du séquençage de l'ADN, prix Nobel de chimie 1958 et 1980 (photo Wikimedia Commons).

VI. SÉQUENÇAGE ET SYNTHÈSE DES MOLÉCULES DE LA VIE

En 1828, à partir d'acide cyanhydrique et d'ammoniac, Friedrich Wöhler réussit à synthétiser l'urée, une molécule organique très répandue dans le monde vivant. Cette découverte capitale marquera la naissance de la chimie organique qui se développera au cours du XIX^e siècle et sera à l'origine de la synthèse de très nombreuses substances chimiques. Cette discipline demeurera longtemps éloignée de la biologie. Comment envisager la synthèse des molécules complexes du vivant, comme les protéines, les acides nucléiques (ADN et ARN) ou les polysides (polymères de sucres), sans en connaître la structure chimique précise ? Il faudra attendre la première moitié du XX^e siècle pour comprendre la composition et la structure des molécules organiques, à partir de leur purification, de leur analyse et éventuellement de leur structure tridimensionnelle étudiée par cristallographie.

On savait de longue date que les protéines sont formées d'acides aminés organisés en une suite de briques, mais leur agencement est longtemps resté un mystère. Après la Seconde Guerre mondiale, un chimiste anglais de génie, Frederick Sanger (1918-2013), deux fois prix Nobel de chimie en 1958 et en 1980 (Figure 7), met au point une technique de fragmentation enzymatique des protéines suivie d'une chromatographie. Après dix ans d'effort, il détermine en 1954, la séquence d'une première protéine, l'insuline, une hormone pancréatique de 51 acides aminés (18). Grâce à cette technique (fastidieuse), on détermine la séquence de nombreuses protéines, que l'on rapporte aux données cristallographiques et à leur éventuelle fonc-

tion. En 1977, Sanger propose une méthode innovante de séquençage des bases qui constituent l'ADN (19). Le principe est d'initier la polymérisation de l'ADN à l'aide d'une ADN polymérase en utilisant une petite amorce oligonucléotide complémentaire d'une partie du fragment d'ADN à séquencer. Les fragments d'ADN synthétisés sont séparés par électrophorèse, ce qui permet d'en déduire la séquence. La séquence nucléotidique permet de déduire l'ordre des acides aminés de la protéine codée par le gène. À la même époque, Walter Gilbert (1932-) et son étudiant Allan Maxam (1942-) proposent une autre technique de séquençage fondée sur une dégradation chimique sélective de l'ADN (20). Les premiers gènes sont séquencés dès 1978, puis les premiers virus (ψ X 174) en 1978 (21, 22). Grâce au séquençage d'ADN, on va comprendre comment des changements (substitutions), des pertes ou des additions de quelques nucléotides dans un gène peuvent entraîner des maladies. Par exemple, dans l'anémie falciforme (drépanocytose), le britannique Vernon Ingram (1924-2006) établit, en 1957, que la substitution d'un seul acide aminé sur la chaîne β de l'hémoglobine induit cette maladie (23). C'est la naissance de la « médecine moléculaire ».

À partir du moment où l'on connaît la séquence des acides aminés des protéines et des nucléotides de l'ADN, rien ne s'oppose plus à utiliser les techniques de la chimie organique pour synthétiser les molécules du vivant. Un procédé de chimie organique a été mis au point en 1963 par Robert Merrifield (1921-2006) pour synthétiser des polymères (peptides ou oligonucléotides) à partir d'amorces fixées sur un support solide (24). Grâce à cette technique dite « en phase solide », on synthétise d'abord de petites protéines (peptides). Une fois encore, l'insuline sera la première protéine à être synthétisée, en 1965 (25). En 1979, Har Gobind Khorana (1922-2011) synthétise un premier gène de 207 nucléotides, qui code un ARN de transfert (*tyrosine transfer RNA*). Cette technique sera rapidement automatisée, permettant la synthèse et l'assemblage de courts fragments d'ADN, reconstituant ainsi progressivement des génomes entiers. En 2002, Eckard Wimmer

(1936-) réussit, à partir de la séquence *in silico*, la synthèse complète d'un virus pathogène, le virus de la poliomyélite (7 500 nucléotides). Suivront en 2003, celles du phage virulent ϕ X174 (5 386 nucléotides), puis en 2005, du virus de la grippe espagnole (13 500 nucléotides) (26). Craig Venter (1946-) réussit même, en 2008, à synthétiser le génome entier d'une bactérie artificielle (27), puis à lui redonner vie en 2010 (*Mycoplasma mycoides* avec un génome de 1,06 Mb) (28).



Fig. 8 - Stanley Cohen, Herbert Boyer (photo Wikimedia Commons) et Paul Berg (de gauche à droite), pionniers du clonage des gènes (1971).

VII. - LA COURSE AUX GÉNOMES

Les techniques de séquençage ont déclenché une « course au génome » en commençant par les virus dès 1978. De très nombreux virus ont été séquencés, notamment les virus pathogènes, jusqu'aux plus gros comme les poxvirus (186-192 kb), cytomégalovirus (229 kb) et mimi-virus (1181 kb). Conjointement, on détermine la séquence de l'ADN des mitochondries (6-200 kb) et des chloroplastes (> 100 kb). En 1992, l'américain Craig Venter crée la société TIGR (*The Institute for Genomic Research*) et réussit, avec sa femme Claire Fraser et Hamilton Smith (1936-), à séquencer (en quelques mois de 1995) le génome complet de *Haemophilus influenzae* (1 830 137 nucléotides ; 1 743 gènes), devançant les équipes qui tentent de séquencer le colibacille depuis plusieurs années. Le génome de *Escherichia coli* (> 4 000 kb) ne sera publié qu'en 1997. Suivront les publications des génomes de nombreuses autres bactéries et aujourd'hui, on connaît la séquence complète des génomes de la plupart des bactéries pathogènes, commensales ou saprophytes. Le génome de tous les organismes modèles est déchiffré : la levure de boulanger (*Saccharomyces cerevisiae*) en 1997, *Caenorhabditis elegans* (un petit nématode) en 1998, la mouche du vinaigre (*Drosophila melanogaster*) et une plante modèle, l'arabette des dames (*Arabidopsis thaliana*) en 2000...

Enfin, Craig Venter (Figure 8) épatera le monde en achevant la séquence complète du génome humain en 2003, après dix ans d'efforts, pour un coût de 3,7 milliards de dollars (29, 30). De façon étonnante, on n'a identifié que 23 688 gènes, contre les 100 000 attendus par les scientifiques. Le génome humain comporte 3,25 milliards de nucléotides et les gènes ne représentent que 1,2 % du génome. À quoi servent les 3 milliards de nucléotides restants ? Cet ADN non codant est constitué de multiples éléments (31) : introns ; régions régulatrices contrôlant la transcription des gènes ; très nombreuses séquences répétées (ADN satellites, mini-satellites, micro-satellites...) retrouvées tout au long du génome ; éléments mobiles (transposons, rétrotransposons, rétrovirus endogènes...)

disséminés sur le génome et amalgamés à nos chromosomes, et d'autres éléments au rôle inconnu... Aujourd'hui, grâce à la mise au point du pyroséquençage par Pål Nyrén en 1996 à Stockholm, technique commercialisée à partir de 2005, on peut réaliser le séquençage du génome humain en quelques jours et pour moins de 1 000 dollars.

VIII. - LES OUTILS DU GÉNIE GÉNÉTIQUE

Les biologistes moléculaires ont largement utilisé les microbes comme outils, des bactéries comme les colibacilles (*Escherichia coli*) et *Bacillus subtilis*, ou encore des champignons telle la levure de boulanger (*Saccharomyces cerevisiae*). À l'instar des couturiers, ils coupent l'ADN, introduisent dans le génome de nouveaux gènes ou même les recommandent. Leurs ciseaux sont des enzymes qui coupent l'ADN (endonucléases) à des endroits précis. Ils peuvent rabouter et coller des fragments d'ADN entre eux (ligases), ce qui permet les techniques de « clonage » pour isoler et étudier finement les gènes. L'étude du phénomène de restriction, une sorte d'immunité permettant aux bactéries d'éliminer les ADN étrangers, a permis au suisse Werner Arber (1929-), aux américains Daniel Nathans (1928-1999) et Hamilton Smith (1931-) de découvrir, vers 1970, des enzymes nouvelles, dites « endonucléases de restriction ». Produites en réponse à l'infection par un phage, ces enzymes reconnaissent et coupent de courtes séquences d'ADN (3-6 nucléotides), contrairement aux DNases, qui coupent l'ADN au hasard (32, 33, 34). Cette découverte leur valut le prix Nobel de physiologie/médecine en 1978. On peut désormais découper spécifiquement des fragments d'ADN portant les gènes à étudier.

Pour produire des fragments d'ADN en grande quantité, on s'est tourné vers les techniques de clonage utilisant des vecteurs dans lesquels on a inséré des gènes à étudier. Ce sont surtout des plasmides, molécules d'ADN circulaire présentes en grand nombre dans certaines bactéries. Ces plasmides peuvent être facilement introduits dans d'autres bactéries, et même dans les cellules animales ou végétales (Figure 9). C'est ainsi qu'en 1971, Stanley Cohen (1922-)

réussit une des premières manipulations génétiques. Il construit un plasmide hybride en collant deux plasmides portant chacun un gène de résistance à un antibiotique et l'introduit par « transformation » dans une bactérie commensale du tube digestif, *Escherichia coli*. L'année suivante à *Stanford University*, Paul Berg (1926-) crée la première molécule d'ADN recombinante entre un gène d'un virus simien oncogène et le phage λ . En 1973, Stanley Cohen et Herbert Boyer (1936-), réussissent à cloner le premier gène étranger dans *E. coli*, celui d'un ARN ribosomal d'un crapaud africain à griffes. Une véritable ingénierie génétique peut désormais se développer.

La découverte en 1970 de la transcriptase inverse, capable de recopier les ARN messagers en ADN, va fortement faciliter le clonage des gènes en mosaïque des animaux et des plantes. En 1977, Stanley Cohen et Herbert Boyer travaillant pour la société Genentech introduisent le gène humain de la somatostatine dans un colibacille qui produira en quantité l'hormone de croissance alors facile à purifier. C'est la première protéine humaine produite par génie génétique. Par la suite, on réussira à faire fabriquer en routine, par des colibacilles ou des levures, de nombreuses protéines humaines utilisées comme médicaments, telles que l'insuline, l'érythropoïétine, des hormones peptidiques, ou encore des facteurs de la coagulation. Suivra la création de plantes et d'animaux « transgéniques » exprimant des gènes étrangers pour le meilleur ou pour le pire. Le premier essai efficace de thérapie génique a été réalisé en 2001 par l'équipe d'Alain Fischer à l'hôpital parisien Necker-Enfants-Malades sur des enfants atteints de déficits immunitaires sévères dus à l'altération d'un seul gène (35). On peut espérer d'autres succès dans les prochaines années, notamment sur la thalassémie (36).

Dès leur découverte, ces manipulations génétiques posèrent le problème de leur utilisation. On ne tarda pas à mettre en question les nouvelles techniques de l'ADN recombinant. En 1974, une douzaine de biologistes célèbres signèrent avec Paul Berg, prix Nobel de chimie 1980, une lettre publiée dans la revue *Science* demandant un moratoire sur les manipulations génétiques pour se donner le temps de réfléchir sur ces extraordinaires innovations technologiques. Cela suscita en 1975 la conférence d'Asilomar en Californie rassemblant plus de 150 biologistes où l'on proposera des règles pour encadrer les manipulations génétiques, règles reprises l'année suivante par les *National Institutes of Health* aux États-Unis et par les autorités de contrôle dans les pays européens. Ce débat est aujourd'hui loin d'être clos avec, notamment, la rapide extension de la production et de la culture des plantes transgéniques.

IX. - NAISSANCE DE LA BIOINFORMATIQUE

Le décryptage du code génétique a permis de déterminer rapidement les séquences protéiques correspondant aux gènes, évitant l'étape très fastidieuse de purification et de séquençage des protéines. À partir de 1979, les données de séquences de l'ADN augmentent exponentiellement du



Fig. 9 - Craig Venter, pionnier du séquençage du génome humain (2003) (photo Wikimedia Commons).

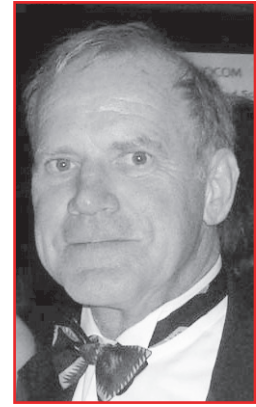


Fig. 10 - Kary Mullis, découvreur de la PCR, prix Nobel de chimie 1993 (photo Wikimedia Commons).

fait de l'automatisation des techniques, ce qui va faire naître une nouvelle discipline, la bioinformatique. Margaret Dayhoff (1925-1983) en fut la pionnière. Dans les années 1960, elle avait commencé à colliger les données de séquences des protéines. On lui doit le premier *Atlas des séquences protéiques* obtenu en compilant toutes les séquences publiées ou connues. Elle compare systématiquement celles-ci pour analyser leur ressemblance. C'est ainsi que naît le concept de famille des protéines, pouvant vraisemblablement provenir d'un ancêtre commun. À cette époque héroïque, les séquences peptidiques alignées sous forme de lettres imprimées étaient examinées à l'œil ! Rapidement, ces séquences deviennent impossibles à comparer. À partir de 1978, les données sont digitalisées pour être exploitées par informatique. Dayhoff développe alors les premiers logiciels de comparaison de séquences. À partir de 1988, de nombreux programmes informatiques apparaissent et sont encore utilisés, comme FASTA et BLAST.

En 1983, le *National Institute of Health* fonde une banque de données appelée PIR (*Protein Information Resource*), succédant à l'Atlas de Dayhoff, suivi en Europe de la création d'une autre banque appelée Swiss-Prot. La première banque de séquences d'ADN appelée GenBank est créée au Laboratoire National de Los Alamos en 1982, puis reprise en 1986 par le *National Center for Biotechnology Information* (NCBI). Les chercheurs purent rapidement envoyer et mémoriser leurs séquences par l'Internet, indépendamment de leurs publications. D'autres banques de données pour les séquences de virus, de bactéries, ou pour le génome humain, sont apparues, puis des banques colligeant les données tridimensionnelles des protéines obtenues par cristallographie et enfin des banques de « motifs structuraux », c'est-à-dire de fragments de séquences associés à une fonction, une activité catalytique par exemple. Aujourd'hui, ces banques de données de séquences de protéines et de gènes aident à la construction d'arbres de l'Évolution (dits « phylogéniques ») permettant de déterminer l'identification et le degré de parenté entre les espèces vivantes.

X. - LA POLYMERASE CHAIN REACTION (PCR)

Certaines découvertes technologiques ont eu un impact majeur sur les progrès de la connaissance du vivant. En 1983, Kary Mullis (1945-), un scientifique travaillant pour la Société Cetus (Figure 10), imagine une technique totalement nouvelle et innovante, la *polymerase chain reaction* (PCR) (37). En 1987, il dépose un brevet où la technique est adaptée et simplifiée grâce à l'utilisation de la *Taq* polymérase, une ADN polymérase stable à haute température provenant d'un microorganisme « extrémophile » vivant à 100° C, *Thermus aquaticus*, évitant des ajouts itératifs de la polymérase, thermosensible, après chaque cycle de chauffage. La PCR permet d'amplifier, à partir de doses infimes, de l'ADN provenant de tissus vivants ou morts. Cette technique a littéralement bouleversé la biologie, ouvrant des perspectives dans de nombreux domaines allant de l'archéologie à la médecine légale, l'écologie, le diagnostic médical, la génétique, l'identification des micro-organismes inconnus ou encore la généalogie. Véritable machine à remonter le temps, la PCR a permis aussi de travailler sur l'ADN « ancien », à partir d'ossements d'animaux ou de restes humains datant de milliers d'années, permettant par exemple de réaliser la séquence complète du génome de l'homme de Néanderthal disparu il y a 30 000 ans (38). De même, à partir d'une phalange de l'auriculaire d'une fillette morte il y a 80 000 ans dans la grotte de Denisova, en Sibérie, on a pu décrypter le génome d'une espèce inconnue d'hominidé (39). Kary Mullis recevra le prix Nobel de chimie en 1993.

XI. - ARN INTERFÉRENT

Une autre découverte importante est celle des ARN interférents. En cherchant à obtenir, dans les années 1990, la couleur plus mauve des fleurs de pétunias, Richard Jorgensen de l'université de Tucson (Arizona), introduit dans le génome de la plante des copies supplémentaires du gène responsable de la pigmentation mauve des pétales. À sa grande surprise, les fleurs deviennent blanches ou mauves tachetées de blanc. L'addition du gène entraîne paradoxalement la perte de la pigmentation des pétales (40). Le mécanisme de cet énigmatique phénomène sera résolu par Andrew Fire (1959-) et Craig Mello (1960-). Travaillant sur *Coenorabditis elegans*, un minuscule ver modèle, ils réalisent une expérience remarquable, publiée en 1998, qui démontre que l'ARN double brin inhibe l'action du gène correspondant. Ils injectent à ce petit nématode des ARN purifiés du gène *unc-22* impliqué dans le fonctionnement de ses muscles. Aucun effet n'est obtenu en injectant chacun des deux brins complémentaires purifiés de l'ARN correspondant au gène ; seule l'injection de l'ARN double brin entraîne un dysfonctionnement musculaire avec de violentes convulsions. Ils mettent au jour les mécanismes moléculaires de cette « interférence » de l'ARN (*RNA silencing*), qui inactive spécifiquement l'activité du gène correspondant (41). Le gène mauve instillé par Jorgensen était certainement contaminé par l'ARN bicaténaire. Fire et Mello ont montré que l'ARN double brin introduit dans



Fig. 11 - Les découvreurs de l'ARN interférent, Andrew Fire (à gauche) et Craig Mello (à droite), prix Nobel de physiologie/médecine 2006 (photos Wikimedia Commons).

les cellules du ver est coupé en petits fragments de 25 nucléotides par l'enzyme Dicer. Ces petites molécules, appelées ARNsi (*small interferent*), interagissent ensuite avec un complexe protéique appelé Risc qui permet la dégradation uniquement du brin « sens », le brin « anti-sens » étant dirigé par le complexe Risc jusqu'à l'ARN messenger complémentaire qui sera alors dégradé, bloquant l'expression du gène correspondant. Cette découverte leur valut le prix Nobel de physiologie/médecine en 2006 (Figure 11). Ce phénomène d'interférence de l'ARN jouerait un rôle important dans la régulation de l'expression génétique lors du développement. Le *RNA silencing* est actuellement mis à profit pour des manipulations génétiques et à visée thérapeutique. L'interférence de l'ARN est observée dans l'ensemble du monde vivant, à l'exception des bactéries. Ce serait une innovation des eucaryotes apparus il y a 1,6 milliard d'années. C'est un moyen d'inactivation de l'ARN étranger à l'intérieur même des cellules.

XII. - LES RECOMBINAISONS ALÉATOIRES DE L'ADN (DNA SHUFFLING)

Wilhem Stemmer, en 1994, a mis au point une technique originale permettant de créer des gènes artificiels présentant des propriétés préalablement définies. Le *DNA shuffling* consiste à couper en petits fragments (50-100 bp) un gène ou une famille de gènes très apparentés, puis à les réassembler au hasard, créant ainsi des gènes complètement nouveaux exprimant des propriétés totalement originales (42). Cette technique est développée notamment par Maxygen, une petite société de biotechnologie de Redwood City en Californie, spécialisée dans le brassage génétique. Ainsi, Stemmer a réussi à créer un nouveau gène bactérien rendant les bactéries très résistantes aux pénicillines de dernière génération. Ce gène code une « super-enzyme », une β -lactamase 32 000 fois plus active pour détruire les antibiotiques que l'enzyme codée par le gène bactérien d'origine. Par comparaison, les procédés traditionnels créant *in vitro* des mutations permettent d'augmenter l'activité enzymatique tout au plus d'un facteur 16. En 2000, Stemmer a appliqué cette technique à la création de virus nouveaux, l'appelant « l'élevage

moléculaire des virus » (*molecular breeding of viruses*). À partir de 6 souches de rétrovirus de souris, il a pu sélectionner un virus totalement nouveau infectant des cellules jusque-là insensibles aux virus sauvages (43). Cette technique « d'évolution dirigée » a pu être étendue avec succès aux bactéries. Ainsi, on est aujourd'hui en mesure de créer des bactéries et des virus aux propriétés totalement nouvelles. On peut ainsi produire des gènes originaux mimant dans un tube à essai ce que la nature fabrique en des centaines de milliers d'années.

XIII. - LE CLONAGE DES ORGANISMES ENTIERS

Le 23 février 1997, un nouvel événement va ébranler la communauté scientifique et l'opinion internationale : l'annonce du clonage d'une brebis surnommée Dolly, née le 5 juillet 1996, obtenu à partir d'un noyau d'une cellule somatique provenant de la glande mammaire d'une brebis, injecté dans un ovule énucléé de brebis. L'équipe de Keith Campbell et Ian Wilmut chez PPL Therapeutics, en collaboration avec le *Roslin Institute* à Édimbourg, a ainsi créé 277 cellules-œufs, donnant naissance à 30 embryons. Seul l'un d'entre eux a réussi à se développer jusqu'à l'âge adulte, donnant naissance à Dolly (44). L'annonce de ce premier clonage réussi d'un mammifère a suscité une émotion considérable du fait de la crainte, en filigrane, d'un clonage humain avec tous les problèmes éthiques que cela pose. Dolly a présenté des signes de vieillissement accéléré à partir de janvier 2002. Atteinte d'une maladie pulmonaire progressive, elle sera euthanasiée le 14 février 2003. Ce vieillissement précoce a été attribué à l'âge assez avancé de la brebis donatrice des ovules (6 ans). En 2000, on réalise avec la même technique le premier clonage chez un singe, puis 2002 chez un porc à l'Institut Roslin. Cette année-là, la Société Clonaid et la secte des Raéliens prétendent avoir réussi le premier clonage humain, allégations qui se révéleront mensongères. Beaucoup de pays d'Europe vont interdire les manipulations génétiques sur les cellules germinales, alors que cela demeure possible aux États-Unis et en Chine. En revanche, les manipulations génétiques sur les cellules somatiques sont admises dans la plupart des pays.

XIV. - CRISPR/CAS9

En 2012, une nouvelle technologie permettant de manipuler rapidement le génome est apparue. Dénommée CRISPR-Cas9, ce système dérivé des bactéries a été décrit par Jennifer Doudna (1964-) (UC Berkeley) et Emmanuelle Charpentier (1968-) (*Max Planck Institute for Infection Biology*). Il utilise, d'une part des molécules d'ARN pour cibler une séquence spécifique d'ADN, et d'autre part la nucléase bactérienne Cas-9 (45). Il permet ainsi de remplacer n'importe quel gène dans n'importe quel organisme, bactéries, plantes, animaux, y compris les primates. Il est désormais entré dans l'arsenal des techniques de correction de l'ADN. Avec une précision de scalpel fondée sur la com-

plémentarité ADN-ARN, on peut facilement corriger une mutation, détruire ou remplacer un gène déficient ou néfaste, ou encore stimuler l'activité d'un gène normal. Ce pourrait être un outil puissant pour réaliser une thérapie génique de certaines maladies humaines, telles que le sida, l'hémophilie, la thalassémie et certaines formes de cancer. Des essais de thérapie génique sur les cellules musculaires somatiques ont été entrepris avec succès pour traiter la myopathie de Duchenne chez l'homme dans des modèles de rongeurs (46, 47).

Cette technologie révolutionnaire a récemment été utilisée en Chine sur des embryons humains, suscitant un grand trouble à travers le monde et un débat scientifique et éthique. Des chercheurs chinois ont tenté de réparer le gène *HBB* qui code la protéine de la β -globine humaine, responsable de la β -thalassémie. Ils ont utilisé des embryons fertilisés d'une seule cellule, mais non viables (48). Parmi les 86 embryons inoculés par le mélange enzymatique, 71 embryons ont survécu jusqu'au stade de 8 cellules et 54 ont pu être testés : seulement 28 ont subi une coupure par la nucléase, mais seule une fraction de ceux-ci contenait le matériel génétique de remplacement. Fait inquiétant, l'équipe a trouvé un nombre surprenant de mutations hors de la cible introduite par la technologie CRISPR/Cas9. Ce taux de mutations est beaucoup plus élevé que celui observé lors d'études de corrections de gènes réalisées chez les embryons de souris et sur des cellules somatiques humaines adultes. Les essais de manipulation des embryons de singe avec cette technique montrent qu'au moins la moitié des 10 grossesses ont été suivies d'avortement. Pour les nouveau-nés, tous n'avaient pas les changements désirés. Cette absence de maîtrise des mutations secondaires, en plus de la manipulation de cellules germinales, a suscité un grand débat en 2015.

XV. - CONCLUSION

Nous assistons aujourd'hui à une explosion des connaissances dans tous les domaines et particulièrement dans celui des sciences du vivant. Les extraordinaires progrès de la biologie nous interrogent sur ses futures évolutions, en particulier dans le champ de la génétique humaine quand on voit apparaître les nouveaux outils permettant de ciseler le génome humain. Comme souvent en science, les progrès peuvent revêtir deux visages. D'un côté, le but est de comprendre le fonctionnement des êtres vivants, de prévenir leurs maladies, et de réparer les imperfections de la nature. D'un autre côté, on pressent les dangers potentiels de ces avancées, notamment si l'on manipulait des cellules germinales fécondées. Ne risque-t-on pas un jour de produire des individus sélectionnés ou manipulés génétiquement pour les rendre « parfaits » selon des critères bien incertains ? Ces manipulations génétiques pourraient éliminer le hasard, cette force qui guide l'évolution de tous les êtres vivants jusqu'à l'homme, depuis l'apparition des premières cellules vivantes il y a plus de 3 milliards d'années.

Conflit d'intérêt : aucun

RÉFÉRENCES BIBLIOGRAPHIQUES

- (1) Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. *John Murray*, London ; 1859.
- (2) Winther RG. August Weismann on germ-plasm variation. *J Hist Biol* 2001 ; **34** : 517-55.
- (3) Mendel JG. Versuche über pflanzenhybriden. *Verhandlungen des Naturforschenden vereines in Brünn* 1865 ; **4** : 3-47.
- (4) Schröninger E. *What is Life? Macmillan Co*, New-York ; 1946.
- (5) Miescher F. Ueber die chemische zusammensetzung der eiterzellen. *Medizinisch-chemische Untersuchungen* 1871 ; **4** : 441-60.
- (6) Griffith F. The significance of pneumococcal types. *J Hyg (Lond)* 1928 ; **27** : 113-59.
- (7) Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 1953 ; **171** : 737-8.
- (8) Franklin RE, Gosling RG. Molecular configuration of sodium thymonucleate. *Nature* 1953 ; **171** : 740-1.
- (9) Crick F. Central dogma of molecular biology. *Nature* 1970 ; **227** : 561-3.
- (10) Gros F, Hiatt H, Gilbert W, Kurland CG, Risebrough RW, Watson JD. Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*. *Nature* 1961 ; **190** : 581-5.
- (11) Brenner S, Jacob F, Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 1961 ; **190** : 576-81.
- (12) Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 1961 ; **3** : 318-56.
- (13) Temin HM, Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 1970 ; **226** : 1211-3.
- (14) Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 1970 ; **226** : 1209-11.
- (15) Berk AJ, Sharp PA. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* 1977 ; **12** : 721-32.
- (16) Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 1977 ; **12** : 1-8.
- (17) Breathnach R, Mandel JL, Chambon P. Ovalbumin gene is split in chicken DNA. *Nature* 1977 ; **270** : 314-9.
- (18) Sanger F. Chemistry of insulin; determination of the structure of insulin opens the way to greater understanding of life processes. *Science* 1959 ; **129** : 1340-4.
- (19) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977 ; **74** : 5463-7.
- (20) Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 1977 ; **74** : 560-4.
- (21) Godson GN, Barrell BG, Staden R, Fiddes JC. Nucleotide sequence of bacteriophage G4 DNA. *Nature* 1978 ; **276** : 236-47.
- (22) Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, et al. The nucleotide sequence of bacteriophage ϕ X174. *J Mol Biol* 1978 ; **125** : 225-46.
- (23) Ingram VM. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* 1957 ; **180** : 326-8.
- (24) Merrifield RB. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* 1963 ; **85** : 2149-54.
- (25) Kung YT, Du YC, Huang WT, Chen CC, Ke LT. Total synthesis of crystalline bovine insulin. *Sci Sin* 1965 ; **14** : 1710-6.
- (26) Tumpey TM, Basler CF, Aguilar PV, Zeng H, Solórzano A, Swayne DE, et al. Characterization of the reconstructed 1918 spanish influenza pandemic virus. *Science* 2005 ; **310** : 77-80.
- (27) Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 2008 ; **319** : 1215-20.
- (28) Gibson DG, Glass JL, Lartigue C, Noskov VN, Chuang RY, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 2010 ; **329** : 52-6.
- (29) Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001 ; **291** : 1304-51.
- (30) Venter JC. A part of the human genome sequence. *Science* 2003 ; **299** : 1183-4.
- (31) ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012 ; **489** : 57-74.
- (32) Arber W, Linn S. DNA modification and restriction. *Annu Rev Biochem* 1969 ; **38** : 467-500.
- (33) Smith HO, Wilcox KW. A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J Mol Biol* 1970 ; **51** : 379-91.
- (34) Danna K, Nathans D. Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc Natl Acad Sci USA* 1972 ; **68** : 2913-7.
- (35) Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, Nussbaum P, et al. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 2000 ; **288** : 669-72.
- (36) Cavazzana-Calvo M, Payen E, Nègre O, Wang G, Hehir K, Fusil F, et al. Transfusion independence and HMG2 activation after gene therapy of human β -thalassaemia. *Nature* 2010 ; **467** : 318-22.
- (37) Mullis KB. Process for amplifying nucleic acid sequences, 1987, United States Patent 4683202.
- (38) Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neanderthal genome. *Science* 2010 ; **328** : 710-22.
- (39) Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 2010 ; **468** : 1053-60.
- (40) Napoli C, Lemieux C, Jorgensen R. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* 1990 ; **2** : 279-89.
- (41) Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998 ; **391** : 806-11.
- (42) Stemmer WP. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc Natl Acad Sci USA* 1994 ; **91** : 10747-51.
- (43) Soong NW, Nomura L, Pekrun K, Reed M, Sheppard L, Dawes G, et al. Molecular breeding of viruses. *Nat Genet* 2000 ; **25** : 436-9.
- (44) Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. Viable offspring derived from fetal and adult mammalian cells. *Nature* 1997 ; **385** : 810-3.
- (45) Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012 ; **337** : 816-21.
- (46) Long C, Amoasii L, Mireault AA, McAnally JR, Li H, Sanchez-Ortiz E, et al. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* 2016 ; **351** : 400-3.
- (47) Nelson CE, Hakim CH, Ousterout DG, Thakore PI, Moreb EA, Castellanos Rivera RM, et al. *In vivo* genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* 2016 ; **351** : 403-7.
- (48) Liang P, Xu Y, Zhang X, Ding C, Huang R, Zhang Z, et al. CRISPR/Cas9-mediated gene editing in human triploid nuclear zygotes. *Protein Cell* 2015 ; **6** : 363-72.