

Hargobind Khorana

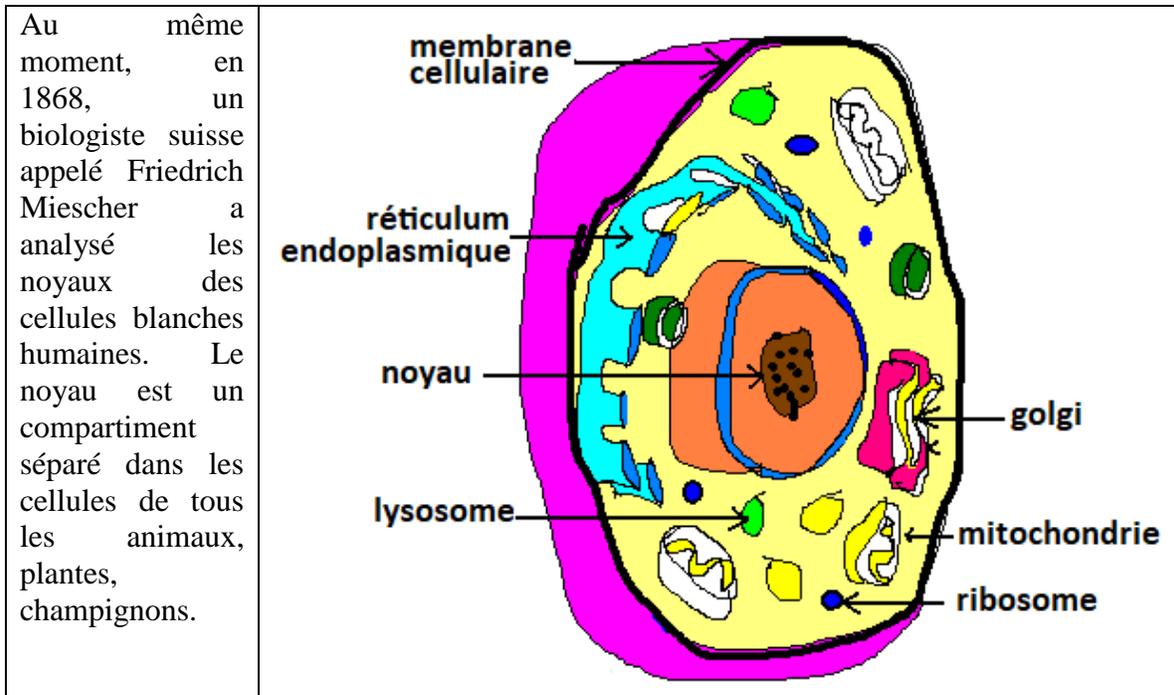
Dans un délai de 150 années court, le domaine de la génétique est passé d'un nouveau-né au domaine qui a le plus grand respect. L'Origine des Espèces, de Charles Darwin, publiée en 1859, fut le premier examen scientifique du mystère de l'héritage, et actuellement, la merveille est bien comprise, avec la documentation du génome humain. Le génome humain est le modèle du patrimoine génétique humain, des milliards d'unités d'informations stockées dans la structure d'une molécule géante qui se trouve dans chaque cellule vivante. L'histoire de la découverte a présenté de nombreuses personnalités et le Professeur Hargobind Khorana, lauréat du prix Nobel, était l'un des plus importants.

Avant 1859, on regardait les espèces comme immobiles et fixées, et qu'ils restaient toujours « comme Dieu les avait créées ». La découverte de Charles Darwin, que les espèces évoluaient, est basée sur l'histoire cachée dans des fossiles, et la répartition géographique, des espèces. La théorie de Darwin proposait que des changements dans les conditions de climat, de végétation, d'approvisionnement alimentaire avaient le résultat que des variations accidentelles particulières dans une espèce avaient un avantage pour la survie, et au fil des siècles, les espèces se changeaient vers ces variétés. La théorie était un défi théologique fort, mais il ne faisait aucune suggestion de la façon dont les variations pourraient se produire.

Dans l'intervalle, Gregor Mendel, un moine augustinien et enseignant des lycéens en Autriche, a découvert des caractéristiques importantes de l'hérédité. Avec des enregistrements minutieux des résultats des plantes de pois croisés, Mendel a constaté que les traits passés par les parents à la progéniture étaient passés en forme de « facteurs » ou des unités - un pour chaque trait spécifique et avec une contribution égale de chaque parent. Ces facteurs héréditaires ne se sont pas combinés, mais ont été passés complets ; chaque parent transmet la moitié de ses facteurs héréditaires à chaque progéniture. Certains facteurs étaient « dominants » par rapport aux autres et les descendants différents des mêmes parents recevaient différents ensembles de facteurs héréditaires.

Le travail de Mendel, publié en 1900, a montré que l'hérédité traversait des unités discrètes, que nous appelons actuellement des gènes. Dans les années ultérieures, d'autres scientifiques ont raffiné la théorie de la façon dont les gènes ont été hérités, mais la nature du matériel génétique restait inconnue.

La découverte de la nucléine



Miescher a découvert que les noyaux contenaient une substance contenant du phosphore, qu'il appelait nucléine.

Il a constaté que les nucléines se composaient d'une partie acide et d'une partie protéinée. Nous savons maintenant que la partie acide était l'ADN et que la partie protéine contribue à coller l'ADN. Mais à l'époque, que le nucléine ou l'acide nucléique peut jouer un rôle dans l'héritage cellulaire n'était qu'un soupçon, presque rejeté face à l'apparente manque de diversité chimique dans les acides nucléiques.

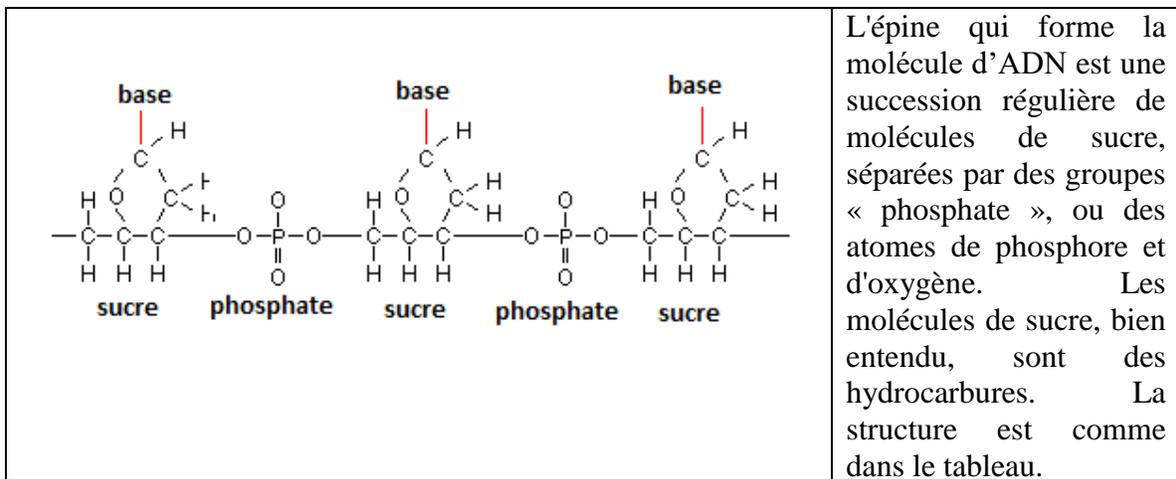
Ce n'est qu'en 1943 qu'il a été démontré que c'était l'acide nucléique de la cellule qui portait d'informations génétiques. Avery et d'autres au Rockefeller Institute ont montré que l'injection de l'acide nucléique d'une souche de bactéries dans une autre a transformé la bactérie receveuse en la souche du donneur. Il est apparu que la composante génétique dans le matériau injecté a été absorbée et est devenue une partie du destinataire. L'étude de la génétique, qui n'était qu'un sujet « décrivant et classifiant » est transitée actuellement à une science analytique. L'origine de la biologie moléculaire est née. En 1952, une expérience utilisant un marqueur radioactif pour identifier des atomes

individuels a montré que c'était en effet la partie acide nucléique, et non la couche protéique, d'un virus qui est entré dans la cellule hôte et fourni l'information génétique pour la réplication du virus à l'intérieur de l'hôte.

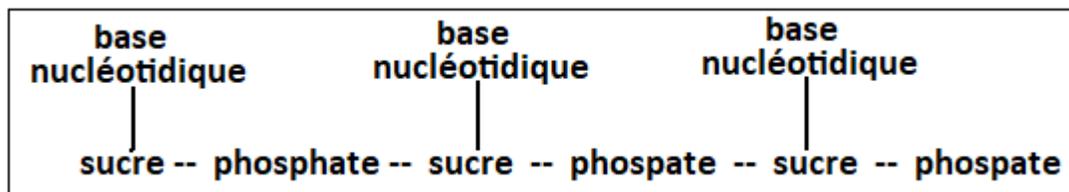
C'était en fait des progrès. De la première découverte de Mendel qu'il y avait des « facteurs » de chaque parent qui participaient aux traits de la progéniture, un certain nombre d'expériences ont montré que c'était l'acide nucléique qui portait l'information génétique dans toutes les cellules vivantes. Mais comment, en fait, cette information a été véhiculée et exprimée était encore une excellente question sans réponse.

C'était une question très difficile à l'époque. Comment le contenu d'un œuf et d'un spermatozoïde produit-t-il un être humain entier, différent de tout autre ? Comment les millions de cellules dans le corps d'une personne savaient-elles quel type de cellule elles étaient et quelles protéines produisaient-elles ? Comment l'information qui indiquait à une cellule rénale qu'elle est une cellule rénale, ou une cellule cérébrale de quelle cellule elle était, est garnis dans le noyau de la cellule?

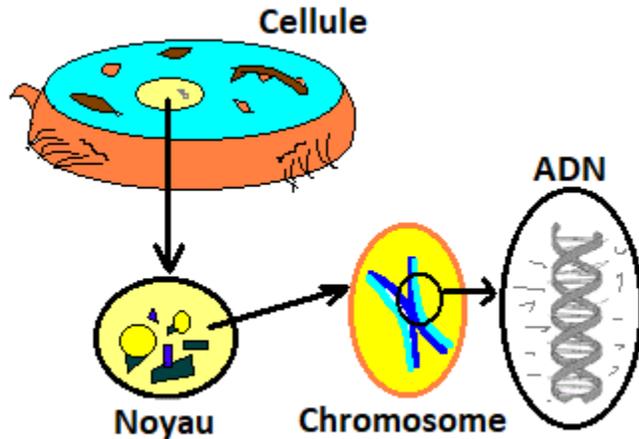
Le monde des scientifiques, bien sûr, s'est engagé à travailler et à toutes sortes d'études, à propos de la structure et des propriétés de l'acide désoxyribonucléique, ou de l'ADN, qui était ce support d'information génétique. Un peu d'information était que la molécule d'acide nucléique était un « épine » long attaché à des groupes appelés « bases » ou « bases nucléotidiques ».



Un moyen pratique de visualiser la molécule d'acide nucléique est comme ceci:



La molécule d'ADN se construit de cette façon et a des millions d'unités. Chaque molécule d'ADN forme un « chromosome », en la forme d'un fil, et il y a 23 paires de chromosomes dans le noyau de chaque cellule humaine. Le nom, « chromosome » est venu du fait que le matériau dans le noyau pourrait prendre une tache colorée, pour l'observation, et avait été nommé « chromatine ». L'ADN est « acide désoxyribonucléique », du « désoxyribose », la molécule de sucre trouvée dans l'épine de la molécule grande.



Le prochain élément d'information était que les bases, auxquelles chaque molécule d'acide nucléique était attachée, ne prenaient que quatre formes dans tous les échantillons d'ADN. Ces formes étaient l'adénine (A), la guanine (G), la thymine (T) et la cytosine (C). Mais ces quatre formes pourraient être présentes dans différentes proportions de l'ADN de différents organismes. Maintenant, une faite important, qu'on a remarqué à la fin des années 1940, était que les quantités de A et T ou G et C se révélaient toujours égales, c'est-à-dire, que chaque fois qu'il y avait un A, il y aurait un T, ou à chaque fois qu'il y avait un G, il y aurait un C. Cette observation était un indice important de la structure de l'ADN et plus tard aussi de sa fonction.

Études radiographiques

Le dernier indice qui a permis d'obtenir la structure de l'ADN est issu des études X-Rays de Franklin et Wilkins en Angleterre. L'utilisation des rayons X qui est bien connue est leur capacité de lancer l'ombre des os, pour détecter les fractures. Mais X Rays est également utile dans la science, car les rayons X consistent en des ondes dont les dimensions correspondent à la distance entre les atomes dans les cristaux et les molécules. Les caractéristiques régulières des molécules sont alors révélées par des motifs dans lesquels les rayons X traversant les molécules sont dispersés.

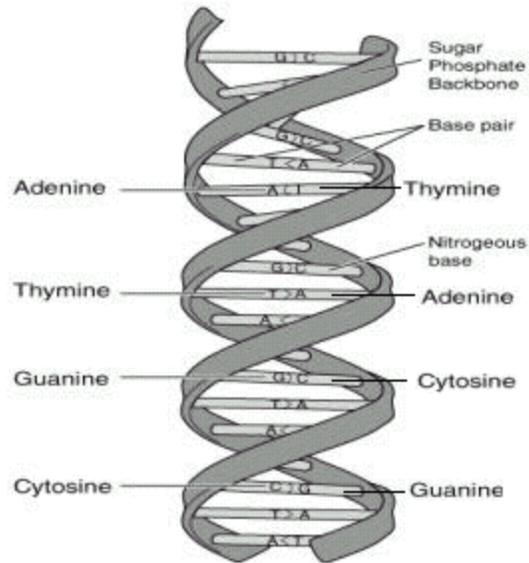
Le travail de Franklin et Wilkins a rapidement montré que la structure de l'ADN montre une nature « périodique » sur sa longueur, une période répétée tous les 0,34 nm et un secondaire tous les 3,4 nm (un nanomètre, nm, est un millionième de millimètre).

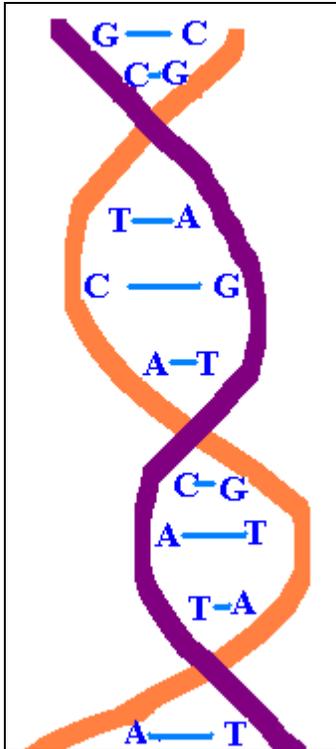
Watson et Crick

Watson et Crick ont rassemblé tous ces indices, du sucre et du phosphate répété, les A, G, T, C apparaissant avec des quantités égales d'A et T ou G et C et la périodicité sur la longueur. Et avec une méthode plutôt que de mettre en place un puzzle à trois dimensions, ils ont proposé le célèbre modèle « double hélice », de deux chaînes d'acide nucléique entrelacées et reliées par des liens d'A et T ou G et C (qui représentaient un A Pour chaque T et un G pour chaque C).

C'était une superbe percée. Les deux brins connectés par des paires complémentaires d'A-T ou G-C

s'assurent que si les deux brins étaient séparés, chaque moitié pourrait construire l'autre moitié en ajoutant exactement un A pour chaque T et un G pour chaque C et vice versa! Cela permet de créer des doublons exacts d'ADN, ou d'auto-réplication, en utilisant les deux moitiés en tant que modèles. L'ADN dans une cellule d'un organisme pourrait alors se diviser, avec la moitié de l'ADN dans la moitié de la cellule et l'autre moitié de l'ADN dans l'autre moitié de la cellule, puis génèrent deux cellules identiques et complètes. Watson et Crick ont partagé le Prix Nobel de 1962 pour la découverte.





Har Gobind Khorana, né en 1922, dans un village obscur au Punjab, dans l'Inde alors indivise, était un candidat improbable pour le stade central du drame qui se déroulait. Il était le plus jeune des cinq enfants du « patwari » du village, ou le greffier qui gardait les registres fonciers à l'extrémité inférieure du régime foncier indien. Le village lui-même avait une population d'environ cent personnes et le père de Khorana était pauvre. Mais il a tenu compte de l'éducation et les enfants ont été envoyés à l'école.

Khorana a assisté à l'école DAV à Multan et plus tard est allé à l'Université de Punjab à Lahore, où il a pris son maîtrise en science. Il a eu de la chance, il y avait d'enseignants doués à l'école et au collège, et, en 1945, Khorana a reçu une bourse du gouvernement de l'Inde pour aller en Angleterre pour son doctorat. Il a pris son doctorat de l'Université de Liverpool en 1945 et a passé l'année prochaine avec le professeur Vladamir Prelog à Zurich. C'était une période où la pensée et la philosophie de Khorana envers la science, la recherche et le travail étaient fortement influencées et moulées dans la tradition européenne.

Après une brève visite en Inde en 1949, Khorana est retourné à Cambridge, jusqu'en 1952. Ceux-ci étaient trois années cruciales, lorsque Khorana s'intéressait aux protéines et à l'acide nucléique. Après Cambridge, Khorana a déménagé au British Columbia Research Council, à Vancouver, puis à l'Institute for Enzyme Research à l'Université du Wisconsin. De son côté, son intérêt était dans le domaine des protéines, des acides nucléiques et de l'ADN, les développements passionnants qui ont couru pour démêler le mystère de l'hérédité !

Lorsque Crick et Watson avaient proposé le modèle à double hélice en 1953, le cadre de transmission de l'information génétique avait été trouvé. Il s'agissait d'un mécanisme stable qui semblait avoir la complexité nécessaire, ainsi qu'une méthode pour se diviser en deux et pourtant être capable de reproduire l'original à partir des moitiés séparées. C'était élégant et polyvalent, mais il n'y avait aucune clarté sur le code lui-même, comment la chaîne de A, T, G et C a-t-elle conduit à des fonctions spécifiques et diverses des milliards de cellules dans un organisme vivant?

Les réponses à ces questions proviennent des travaux de trois, Holley, Khorana et Nirenberg. On a constaté que les fonctions des organismes étaient contrôlées par une sorte de protéine appelée enzymes, agents qui favorisent la croissance de certaines choses. La couleur des yeux d'une personne, par exemple, est le résultat d'une enzyme spécifique que son corps produit, ce qui favorise la production du colorant qui conduit à la couleur des yeux. Qu'une personne soit diabétique, encore une fois, est le résultat d'agents chimiques qui déterminent le taux de production d'insuline. Et la présence de ces agents et dans quelle quantité, dépend des instructions contenues dans les cellules qui produisent les agents. L'individualité d'une personne, en somme, était déterminée par les protéines qu'il avait programmées son corps à créer.

Les protéines, elles-mêmes, on avait trouvé, étaient composés de vingt blocs de construction, appelés acides aminés. Les acides aminés sont des molécules organiques contenant de dix à trente atomes. Différentes combinaisons, certaines fois des chaînes courtes, parfois des milliers d'unités, ces vingt composants construisent toutes les millions de protéines trouvées. Parmi ces vingt, seuls certains peuvent être synthétisés par le corps, le reste doit être ingéré comme nourriture. Mais à partir de ce stock de vingt acides aminés, le code dans l'ADN assure que des protéines compliquées sont construites. Comment cela se fait-il, c'est la façon dont le code fonctionne et pour encadrer les mots dans le code et créer les protéines que les mots épeautés, c'est à déchiffraient le code!

La façon dont il est fait est que la séquence des nucléotides, A, T, G, C, dicte les acides aminés à inclure. C'est facile à comprendre que quatre types de groupes ne peuvent coder que pour quatre acides aminés. Donc, clairement, plus d'un seul nucléotide est impliqué. Deux nucléotides pourraient coder pour $4 \times 4 = 16$ acides aminés, comme ceci:

A-A, A-T, A-G, A-C; T-A, T-T, T-G, T-C; G-A, G-T, G-G, G-C; C-A, C-T, C-G, C-C
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

OU

1 A A	5 T A	9 G A	13 C A
2 A T	6 T T	10 G T	14 C T
3 A G	7 T G	11 G G	15 C G
4 A C	8 T C	12 G C	16 C C

Ce n'est toujours pas suffisant, car nous avons vingt acides aminés à coder. Si nous utilisons trois ensembles de bases nucléotidiques, nous pouvons coder pour $4 \times 4 \times 4 = 64$ combinaisons, soit plus de vingt !

Eh bien, c'est la façon dont le code est construit. Dans le déploiement des 64 codes, il existe plusieurs codes différents qui signifient le même acide aminé. Cela semble être la manière de la nature de jouer en toute sécurité, même si, lors de la transmission du code, il y a des erreurs, l'acide aminé correct est toujours transmis. Voici le code:

UUU UUC	phenyl alanine:	UCU UCC UCA UCG	serine	UAU UAC	tryosine	UGU UGC	cysteine
UUA UUG	leucine:			UAA UAG	finis	UGA UGG	finis tryptophane
CUU CUC CUA CUG	leucine:	CCU CCC CCA CCG	proline	CAU CAC CAA CAG	histidine: glutamine	CGU CGC CGA CGG	arginine
AUU AUC AUA	isoleucine	ACU ACC ACA ACG	thréonine	AAU AAC AAA AAG	asparagine: lysine	AGU AGC AGA AGG	serine: arginine
AUG	méthionine						
GUU GUC GUA GUG	valine	GCU GCC GCA GCG	alanine:	GAU GAC GAA GAG	acide aspartique acide glutamique	GGU GGC GGA GGG	glycine:

La plupart des acides aminés ont plus d'une manière d'être codés, les codes n'étant différents qu'en une partie de la séquence. Cela est l'évolution d'un moyen le plus stable d'éliminer d'erreurs, compte tenu du niveau de « redondance » disponible.

Ainsi, de cette façon, chaque groupe consécutif de trois nucléotides, ce qu'on appelle de « triades », code pour un acide aminé particulier. Les groupes de trois sont également appelés « codons ». Le codon AUG, qui code pour la méthionine, est le codon 'début', et on peut voir sur le graphique qu'il y a 3 codons 'arrêt'. Ceux-ci montrent où la séquence d'acides aminés pour une protéine commence et s'arrête. Les nucléotides fournissent ainsi un alphabet complet pour spécifier toutes sortes de protéines, en utilisant tous les 20 acides aminés dans un grand nombre de combinaisons. Chaque séquence de codons de « début » à « fini » identifie une protéine et s'appelle un « gène ». Et l'ADN entier est le code de milliers de protéines. Le complément de 23 paires de chromosomes dans la cellule humaine s'identifie sur toutes les protéines qui déterminent le design complet, comme des espèces humaines, ainsi que le patrimoine individuel de chaque personne vivante.

C'était le patrimoine dans lequel Khorana, avec Holley et Nirenberg, sont arrivés, pour accomplir leur travail le plus important. Alors que l'ADN contenait l'information pour la

synthèse des protéines, comment exactement l'information a été exploitée et utilisée dans la formation de protéines était tout autre technologie.

L'ADN entier, a-t-on trouvé, n'a pas participé à la fois à la formation de protéines. Des morceaux de l'ADN, des gènes qui représentent des protéines, ont été copiés sur des molécules similaires, appelées ARN et expédiées hors du noyau de la cellule, éventuellement pour protéger l'ADN contre les dommages. L'ARN est juste comme l'ADN, sauf qu'il y a une petite différence dans le sucre trouvé dans l'épine de la molécule et aussi dans l'une des bases. Le morceau de l'ADN est donc copié sur une forme d'ARN appelée « messenger » ou ARNm, dans un processus similaire à la réplication de l'ADN. L'ARNm s'attache alors à des structures appelées ribosomes dans le fluide en dehors du noyau de la cellule. Le point auquel l'ARNm 'attache' est l'endroit où le codon « AUG ». le codon « début ». ' est trouvé. L'ARN ne sait pas comment trouver son chemin, il s'attache juste lorsque, dans un mouvement fiévreux et aléatoire à l'intérieur de la cellule, les parties correctes de l'ARNm et du ribosome appartiennent à la proximité, et ils "se mettent" dans une position où ils s'adaptent, comme une serrure et une clé. C'est un peu comme une balle de golf errant sur le parcours, qui tombe dans le trou pendent qu'il se déplace.

Lorsque l'ARNm est en position, comme ceci, une autre classe d'ARN appelée « transfert » ou tRNA attache des acides aminés individuels à la succession de codons. L'ARNt est spécifique à chaque acide aminé et le premier à attacher à l'ARN est celui qui a le complément de « AUG », le codon « début ». '. La cellule contient des enzymes qui sont chimiquement du bon profil pour aider les ARNt particuliers à transférer des acides aminés à l'ARNm. Une fois que le premier acide aminé « début » a été attaché, l'ARNt portant l'acide aminé codé par le codon suivant se déplace, pour placer l'acide aminé suivant en position. Les deux acides aminés adjacents forment alors une liaison chimique, le début d'une chaîne d'acides aminés qui va devenir finalement la protéine.

Le processus d'addition d'acides aminés successifs continue jusqu'à ce que le codon « finis ». soit atteint. Lorsque cela se produit, les enzymes s'activent et l'ARNm est libéré de la protéine qui a été synthétisée.

Cette compréhension de la séquence des événements s'est déroulée à la suite d'une série d'expériences pour déduire ce qui se passait au niveau moléculaire invisible. Nirenberg et Heinrich Matthaei ont d'abord développé un moyen de créer des protéines simples dans un tube à essai, en utilisant un ARNm qui avait été extrait des cellules. Khorana a avancé sur le travail de Nirenberg avec la synthèse de l'ARN, qui, à son tour, pourrait créer des protéines. Holley a suivi avec le transport de l'ARNm et la synthèse actuelle.

Le premier ARN synthétisé n'était qu'une chaîne de 'U' (l'équivalent dans les ARN, de « T » de l'ADN). On a trouvé que cela code pour une protéine qui était une chaîne de l'acide aminé, la phénylalanine. On peut voir à partir du diagramme de code génétique que le codon, 'UUU' code pour cet acide aminé. La répétition d'une chaîne de «CA» dans l'ARN a conduit à la chaîne histidine-thréonine, et ainsi de suite.

Les composants individuels des codons ont dû être identifiés par des expériences soigneusement conçues. Par exemple, pour un élément particulier apparaissant dans l'un des groupes, une forme radioactive de l'élément pourrait être utilisée. Une fois que cela a été fait, un brin particulier d'ADN, lorsque l'ADN a été brisé, pourrait être tracé en suivant le « chemin » radioactif. Essentiellement, en utilisant des méthodes comme celle-ci, le sens de chaque codon a été élaboré et la signification du code caché de la nature a été reconstituée. La formule concise et persistante, qui permettait les individus d'une espèce à développer, comme une expression des protéines de leurs cellules produites, a été découverte.

Khorana a fait la majeure partie du travail lui-même et a finalement complété le décodage, l'identification des acides aminés que chacun des six codons représentait. Dans le processus, Khorana a développé des techniques de localisation, de navigation et de marquage d'endroits sur toute la longueur de la molécule d'ADN, les moyens de couper et épurer les brins de l'ADN, les enzymes qui pourraient le faire, la manière de travailler avec eux. Ce sont les pinces et les scalpels de dissection et d'assemblage des composants fondamentaux de la génétique.

La communauté scientifique a rapidement compris l'importance du travail accompli. Le génome humain, ou le code entier des vingt-trois chromosomes, se compose de près de trois milliards de gènes. Une poignée a été identifiée comme étant impliquée dans des affections humaines spécifiques. Les techniques de Khorana ont donné naissance à une industrie générant des morceaux génétiques spécifiques pour aider un individu qui n'a pas ces gènes spécifiques à récupérer une fonction perdue. La communauté et l'industrie ont immédiatement vu l'immense potentiel de cartographier l'ensemble du génome humain. La tâche était évidemment trop importante pour les individus. Les biophysiciens, les microbiologistes, les experts de l'informatique, dans le monde entier, ont été répartis dans le cadre du projet mondial du génome humain.

Aujourd'hui, le génome humain est cartographié, tant par un groupe public que par une entreprise privée. Par conséquent, l'humanité possède le modèle le plus fondamental de sa nature. Les cellules humaines, en principe, peuvent être programmées pour éliminer la maladie ou la tendance à la maladie. Il y a la possibilité de créer la super race, il y a des craintes d'abus. Les applications en droit et la détection du crime sont devenues communes. Il y a des craintes d'une invasion de la vie privée.

La compréhension du mystère de l'hérédité, sondée hésitamment il y a 150 ans, peut être la partie la plus importante de la connaissance au XXI^e siècle.

Khorana est aujourd'hui citoyen naturalisé des États-Unis. Il est marié avec Esther Elizabeth Sibler et a trois enfants. À partir de 1970, il a été professeur au Massachusetts Institute of Technology, où il a continué à apporter d'importantes contributions dans le domaine de la biologie moléculaire. Parallèlement à Holley et Nirenberg, Khorana a reçu le prix Nobel de médecine en 1968.