

Le fonctionnement du code des protéines *

par Ian C. Caldwell
(The Wellcome Research Laboratories)

Il était d'usage courant en biologie moléculaire d'extrapoler dans une large mesure les données obtenues dans la recherche sur un seul type de cellule. Depuis, une différenciation plus subtile entre les cellules munies d'un noyau et les cellules sans noyau a donné des résultats étonnamment fructueux dans l'élucidation du codage génétique. Dans le premier de deux articles, le Dr Caldwell passe en revue les récents progrès. Le second article sera consacré aux problèmes non résolus... et à l'avenir.

L'hypothèse, selon laquelle l'information génétique est emmagasinée et transmise au moyen d'un « code » porté par des molécules d'acide désoxyribonucléique (ADN) est largement admise depuis plus de 20 ans et l'étude du moyen de transmission et d'expression de cette information génétique — appelée communément biologie moléculaire — a certainement été l'un des principaux facteurs du développement de la recherche biologique de ces derniers temps.

On possède actuellement des connaissances étendues sur les réactions qui font l'objet de la biologie moléculaire, mais bon nombre de questions restent sans réponse. Beaucoup d'informations proviennent d'études entreprises sur des organismes procaryotes (ceux qui n'ont pas un noyau bien distinct), particulièrement des bactéries, et de tels systèmes ont été d'une aide immense.

Il a souvent été admis, du moins tacitement, que « ce qui s'applique au *Escherichia coli* s'applique à l'éléphant ». A bien des égards, ceci est probablement vrai. Avec des systèmes isolés et partiellement épurés, il est sûrement possible de mélanger des composants de sources différentes et d'obtenir des réactions similaires à celles que l'on observe dans des cellules entières. Toutefois, à mesure que l'on avance dans l'étude des eucaryotes (organismes possédant un noyau bien défini, tel par exemple la levure et les cellules végétales et animales), il apparaît qu'il y a bien des différences entre les divers types de cellules, et la prise en considération de ces différences joue un rôle de plus en plus important en biologie moléculaire.

Le but de mes articles est de broser le tableau de l'état actuel des connaissances, en termes de résultats moléculaires et d'applications biologiques plus générales, et de donner quelques indications sur les problèmes qui restent encore à résoudre et sur les orientations des recherches futures.

L'acide désoxyribonucléique

Le « code génétique » de la plupart des systèmes biologiques est enregistré dans des molécules d'ADN. Dans les cellules des procaryotes, l'ADN se trouve essentiellement dans la membrane cellulaire, alors que dans celles des eucaryotes la plus grande partie d'ADN se trouve dans le noyau, bien qu'il y en ait dans la mitochondrie, ces organites cellulaires « producteurs d'énergie ».

L'ADN est un polymère qui se présente presque toujours sous forme d'une double hélice à brins complémentaires. Chaque brin consiste en une chaîne d'unités de désoxyribose (sucre) raccordées par des groupes de phosphates. Connecté à chaque résidu de désoxyribose, il y a une base du groupe des purines ou des pyrimidines. Il s'agit d'adénine ou de guanine dans le cas des purines, et de thymine ou cytosine dans le cas des pyrimidines. (Un certain nombre de « bases mineures », chimiquement similaires à ces quatre, se trouvent également

* *Spectrum* 121.

dans l'ADN, en proportions variables et de provenances diverses.) La double hélice est comparable, en gros, à un escalier à vis avec la chaîne sucre-phosphate-sucre formant les rampes et des paires de bases complémentaires rattachées par des liaisons d'hydrogène formant les marches. Dans les dispositions complémentaires ayant une base de pyrimidine et l'autre de purine, l'adénine va de pair avec la thymine et la guanine avec la cytosine (voir figure 1). Cet appariement des bases contribue au maintien de l'intégrité structurale du complexe.

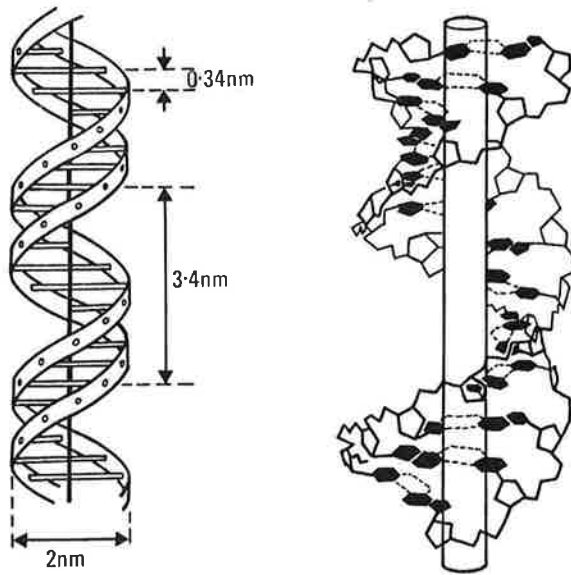


Figure 1. La structure de l'ADN.

Le dessin de gauche montre schématiquement la structure à double hélice proposée par Watson et Crick. Le dessin de droite représente le modèle d'une partie d'une molécule d'ADN.

Les anneaux figurés sont les bases de purine et pyrimidine, les anneaux en blanc les unités de désoxyribose. Les traits pleins représentent les liaisons de phosphate entre les unités de désoxyribose et les pointillés des liaisons d'hydrogène entre les bases.

[Dessins de Davidson (1), page 137.]

La double hélice fut proposée pour la première fois par Watson et Crick en 1953 et est toujours largement admise, bien que des études ultérieures aient affiné les détails du modèle. Dans certaines molécules d'ADN, les deux extrémités se rejoignent pour former un ADN circulaire (chez certaines bactéries et virus et dans la mitochondrie). Les bases de purine et de pyrimidine sont disposées en séquences spécifiques qui forment les bases du code génétique. L'unité fondamentale consiste en une séquence de trois bases, ou triplet. Chaque triplet forme finalement le code pour un aminoacide dans une protéine déterminée.

Tout l'ADN d'une cellule n'intervient pas dans le codage pour les protéines. Une partie est très certainement affectée à « l'intendance », signalant le démarrage et l'arrêt d'une unité génétique particulière, fournissant les « entretoises » de séparation des unités génétiques individuelles et commandant l'utilisation de telle ou telle partie d'ADN à un moment donné. Certaines indications tendent à prouver que chaque cellule d'un organisme complexe contient toute l'information requise pour toutes les cellules de cet organisme. Mais toutes les cellules ne nécessitant pas l'ensemble des informations, une grosse part de l'ADN de certaines cellules pourrait être éliminée irréversiblement. Beaucoup d'ADN, particulièrement dans les cellules des eucaryotes, existe en complexe avec des protéines de divers types telles que les désoxyribonucléoprotéines. Ces protéines entrent très certainement en jeu dans la détermination de l'utilisation de l'ADN.

La duplication de l'ADN, par exemple avant la division cellulaire, est appelée « réplication ». Pour un certain nombre de raisons, principalement techniques ou opérationnelles, beaucoup de nos connaissances proviennent d'études d'organismes procaryotes, particulièrement de l'omniprésent *Escherichia coli*. Bien que l'on sache que la réplication de l'ADN est semi-conservative — c'est-à-dire que chaque molécule d'ADN nouvellement formée se compose d'un brin de l'ADN copié et d'un brin nouveau — le mécanisme détaillé de la synthèse est toujours sujet à controverse. Mais il apparaît de plus en plus que la synthèse d'au moins une des nouvelles chaînes est discontinue, avec formation initiale de fragments relativement courts, interconnectés par la suite.

Les enzymes qui interviennent dans la réplication sont généralement appelées polymérase d'ADN. La polymérase 1 (l'enzyme de Kronberg) était à l'origine tenue pour responsable de la réplication, mais des recherches plus poussées ont prouvé qu'elle n'intervenait pas dans toute la synthèse de l'ADN. Il semblerait plutôt que cette enzyme ait des fonctions de « réparateur », fermant des brèches dans l'ADN, ouvertes soit spontanément, soit par suite d'un endommagement ou pendant la réplication.

La polymérase III joue presque sûrement un rôle dans la réplication, et peut-être un rôle capital, alors que la fonction de la polymérase II

est moins bien définie. Toutes les trois peuvent avoir une incidence sur la synthèse, bien que la réplication puisse se faire — avec quelques changements — en l'absence de l'une d'elles.

Les acides ribonucléiques

Les acides ribonucléiques (ARN) sont intimement liés à l'expression de l'information génétique enregistrée dans les molécules d'ADN, et ils sont synthétisés dans une large mesure sur le modèle de l'ADN. Il s'agit également de polymères ressemblant sensiblement à l'ADN, bien qu'il y ait quelques différences. Par exemple, l'ARN contient plutôt du ribose que du désoxyribose en tant que composant de sucre et de l'uracile de pyrimidine à la place de la thymine contenue dans l'ADN. L'ARN forme davantage un polymère à brin unique, bien que l'on trouve souvent des régions où un pliage de la chaîne rapproche des séquences complémentaires dans diverses parties de la molécule, résultant en des sections de nature bi-hélicoïdale qui contribuent au maintien d'une structure d'ARN partiellement ordonnée. On trouve aussi quelques molécules d'ARN en proche association avec des protéines telles que la ribonucléoprotéine, alors que d'autres existent à l'état libre.

La synthèse de l'ARN, appelée transcription, est effectuée par des polymérase d'ARN. Dans la plupart des cas, le module de transcription est l'ADN, bien que des molécules d'ARN puissent parfois servir. Il semblerait que les polymérase d'ARN ne copient qu'un des brins de l'hélice d'ADN pour produire des molécules d'ARN similaires à l'un des brins dans leur séquence de base et complémentaires à l'autre. Comme il n'y a transcription que de portions du module d'ADN, il doit exister des mécanismes spécifiques au démarrage et à l'arrêt de la synthèse d'ARN. Le facteur, un composant de polymérase d'ARN, joue un rôle dans le démarrage de la transcription au bon endroit, alors que l'on pense que les facteurs ρ et K ont une incidence sur son arrêt. La présence ou l'absence de protéines à lien d'ADN et la venue de séquences de base d'ADN, spécifiques jouent sans aucun doute un rôle dans la signalisation.

Dans les systèmes bactériels, il apparaît que les molécules d'ARN synthétisés sur le module de l'ADN ne sont que marginalement plus grands que les produits d'ARN fonctionnant finalement dans la cellule. Bien que les connaissances de la synthèse de l'ARN chez les eucaryotes soient bien moindres, il a été établi que les produits immédiats de la transcription sont souvent beaucoup plus grands que les produits définitifs et que des changements importants interviennent par la suite.

Les cellules contiennent un nombre d'espèces différentes d'ARN ayant des fonctions distinctes telles que l'ARN messager, l'ARN ribosomal, l'ARN transmetteur et d'autres.

L'ARN messager (ARNm)

Comme l'indique leur nom, ces molécules portent le « message » correspondant au code génétique, sous forme de triplets ou codons dérivés de ceux de l'ADN. On peut les assimiler à des bandes lues ou traduites pour produire les protéines spécifiques. Chez les organismes procaryotes, la synthèse, l'utilisation et la dégradation des molécules d'ARN semblent étroitement liées. L'existence de ces molécules peut être mesurée en secondes ou en minutes.

Dans les cellules des eucaryotes, le système semble toutefois bien plus complexe et la vie des molécules d'ARN est mesurée en heures. Les produits directs de la transcription d'ADN dans le nucléoplasme, partie principale du noyau, semblent être de très grandes molécules d'ARN nucléaires hétérogènes ou à dispersion hétérogène (ARNnh). Par un processus de maturation qui entraîne une perte allant jusqu'à 90 % de la molécule et une modification du restant, ils engendrent des molécules d'ARNm qui sont alors transportées dans le cytoplasme pour agir dans la synthèse des protéines.

L'ARNm des eucaryotes semble se composer de trois parties distinctes : — une « tête » comportant une séquence répétitive, similaire ou identique dans la plupart sinon dans tous les ARNm et reflétant probablement la présence de séquences répétitives dans l'ADN, — une partie centrale « unique », porteuse du code destiné à une protéine donnée, — une « queue » qui consiste en une séquence d'unités contenant de l'adénine (poly(A)).

Les fonctions de la tête et de la queue ne sont pas établies, mais elles pourraient intervenir dans le contrôle de la maturation de l'ARNnh en ARNm, dans la tâche de l'ARNm dans la synthèse des protéines ou dans le transport de l'ARNm vers le cytoplasme.

L'ARN ribosomat (ARNr)

Ces molécules, qui représentent entre 75 et 80 % de l'ARN cellulaire, se rencontrent dans les ribosomes, particules de ribonucléoprotéide qui interviennent dans la synthèse des protéines. Les ribosomes se composent de deux unités séparées de taille différente. Leur sous-unité la plus petite contient une molécule d'ARNr relativement grande, plus un certain nombre de protéines, alors que la plus grande contient une molécule d'ARNr légèrement plus grande, un nombre variable de molécules d'ARNr plus petites et un certain nombre de protéines. La taille des sous-unités ribosomales, des ribosomes et des molécules d'ARNr varie considérablement en fonction de leur provenance, les plus petites se trouvant dans la mitochondrie des cellules des eucaryotes.

ryotes, les plus grandes dans le cytoplasme des eucaryotes et celles qui sont de taille intermédiaire, chez les procaryotes. Chez les procaryotes, les différentes espèces d'ARNr sont probablement transcrites en tant qu'unité coordonnée et les produits immédiats de la transcription ne sont que marginalement plus grands que les molécules définitives d'ARNr. Chez les eucaryotes, et comme pour l'ARNm, le processus est bien plus complexe. La plupart des molécules d'ARNr des eucaryotes sont dérivées d'un précurseur unique très grand, nommé « ARN 45 », formé par transcription d'ADN dans le nucléole, petit corps dense dans le noyau. La maturation de cette molécule dans le nucléole engendre la majeure partie de l'espèce de l'ARNr, bien que le plus petit ARNr soit synthétisé séparément à l'extérieur du nucléole, dans le nucléoplasme. L'assemblage des ribosomes intervient probablement dans le nucléole et les particules sont alors transférées au cytoplasme, lieu de la synthèse.

L'ARN de transfert (ARNt)

Ces petites molécules d'ARN, qui interviennent pour environ 15 % dans l'ARN cellulaire, jouent un rôle majeur dans la synthèse des protéines en transportant des aminoacides « activés » sur le lieu cellulaire adéquat et en les transférant sur la chaîne de protéines croissante.

Il y a de nombreux types de molécules d'ARNt, au moins une (et souvent plusieurs) pour chaque aminoacide incorporé dans les protéines. On a pu déterminer des séquences complètes pour plusieurs espèces d'ARNt. Ces études ont prouvé que bien que chaque ARNt spécifique ait au moins un trait unique, toutes les molécules d'ARNt ont des propriétés communes, comme par exemple la forte teneur en bases modifiées et les régions à double hélice formées par l'appariement des bases. Une structure possible pour l'ARNt, basée sur le modèle généralement accepté de la feuille de trèfle est donnée par la figure 2A, alors que la figure 2B montre la forme simplifiée d'une éventuelle structure tridimensionnelle pour l'ARNt. Un trait caractéristique de chaque espèce est la présence d'un site anticodon spécifique, séquence de trois bases qui est complémentaire et qui se rattache au codon correspondant dans l'ARNm, assurant ainsi l'incorporation des aminoacides spécifiques dans les protéines.

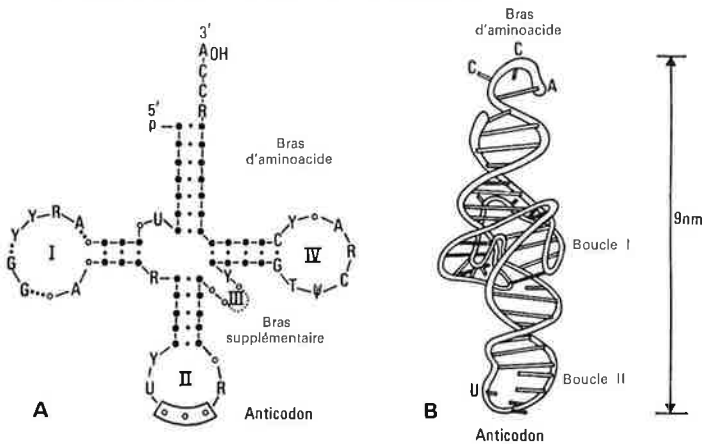


Figure 2. Proposition de structures d'ARNt.

La figure 2A montre une disposition basée sur le modèle « feuille de trèfle ». Les cercles continus représentent les bases dans les régions hélicoïdales et les points centraux les liaisons d'hydrogène; les cercles discontinus et les lettres symbolisent des bases dépareillées. La figure 2B est l'illustration schématique d'une structure tridimensionnelle possible d'une molécule d'ARNt.

[Dessins de Davidsohn (1), pages 124-125.]

Dans les cellules (chez les procaryotes comme chez les eucaryotes) la synthèse de l'ARNt semble se faire par le truchement d'une molécule un peu plus longue qui sert de précurseur (pré-ARNt), qui contient moins de bases modifiées et dont la structure est moins ordonnée que celle de l'ARNt. Pendant la maturation, les unités nucléotides excédentes sont éliminées, la modification d'un grand nombre de constituants s'opère et la structure tridimensionnelle compacte se forme.

Autres espèces d'ARN

Un bon nombre d'autres molécules ont été découvertes dans diverses cellules. Quelques-unes ont été isolées et étudiées en détail (par exemple les petites molécules d'ARNne qui se trouvent dans les noyaux), mais on ne sait que très peu sur leur fonction biologique et leur signification.

La synthèse des protéines

Le processus de synthèse des protéines met en jeu les blocs constructeurs de protéine, à savoir les aminoacides plus les ARNt et ARNm et des particules ribosomales ainsi qu'un grand nombre (sans cesse croissant) de facteurs accessoires. Nos connaissances actuelles sur

les différents stades de la formation des protéines sont fort étendues mais — comme c'est souvent le cas — elles comportent encore des lacunes.

Les débuts de la synthèse, du moins dans les organismes procaryotiques, impliquent la liaison de l'ARNm avec la petite unité ribosomale, puis la liaison avec l'unité aminoacide-ARNt « initiateur ». (Chez les eucaryotes, on a la preuve qu'il y a d'abord liaison du complexe aminoacide-ARNt puis de l'ARNm.) Chez les procaryotes, l'acidoamine initiateur est toujours la N-formylméthionine, et chez les eucaryotes c'est la méthionine. Dans les deux types de cellules, l'ARNt initiateur a la désignation $ARNt_{Met}^f$ et il est spécifique du démarrage des chaînes

de protéine. Il n'intervient nulle part ailleurs dans l'incorporation de méthionine dans les protéines.

Le « signal » de démarrage est l'apparition dans l'ARNm du codon AUG, séquence d'adénine-uracile-guanine, qui semble être le premier codon de la section « unique » de toutes les molécules d'ARNm. L'initiateur spécifique ARNt contient l'anticodon complémentaire à l'AUG et la liaison codon-anticodon assure le démarrage correct.

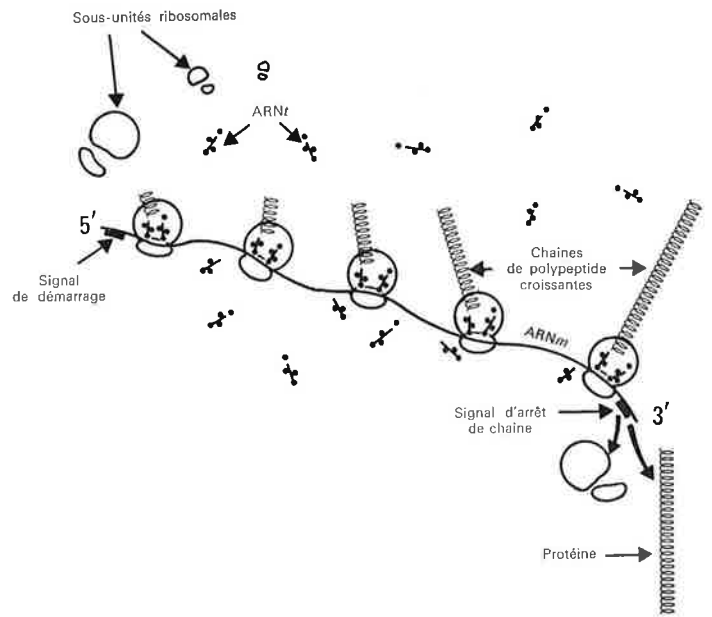
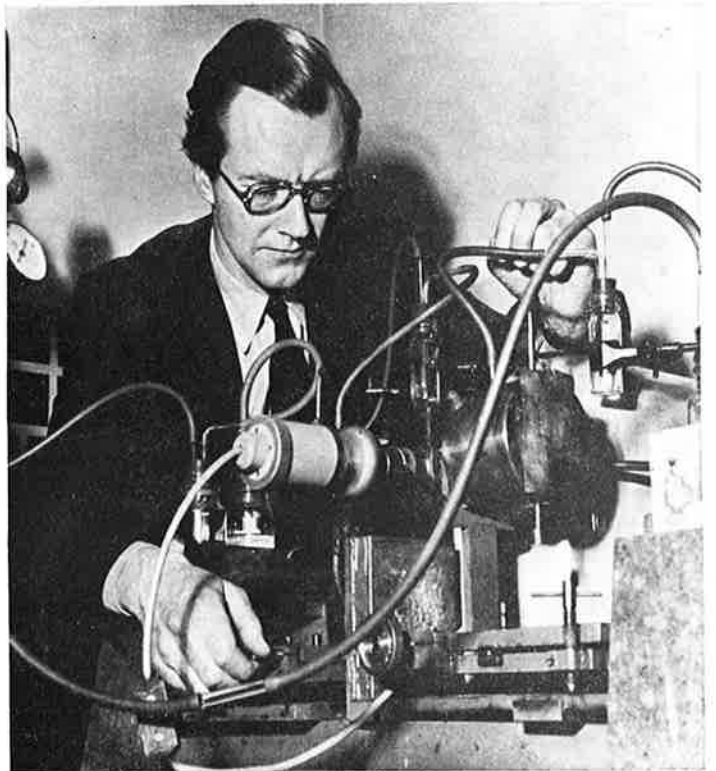


Figure 3. Diagramme schématique d'un polysome, montrant les ribosomes avec les chaînes de protéines croissantes se déplaçant le long de la lignée d'ARNm.



Le modèle d'ADN à double hélice, proposé en 1953 par Watson et Crick, fut confirmé au cours de recherches ultérieures par le Professeur M. H. F. Wilkins. Il se servait d'un diffractomètre à rayons X pour la prise d'images qu'il analysa pour déterminer la structure. Cette photographie, prise en 1961, montre Wilkins à l'œuvre.

Le démarrage est parachevé par l'addition de la sous-unité ribosomale plus grande et le processus de translocation, connu de façon incomplète à ce jour, entraîne la mise en position dans l'ARNm du codon suivant. L'unité suivante d'acide-amino-ARNt, liée par interaction spécifique codon-anticodon, forme un lien de peptide entre les deux acides-amino. Le premier ARNt est alors libéré. Après translocation, l'unité suivante d'acide-amino-ARNt se lie et le processus se répète maintes et maintes fois, avec la chaîne de peptide croissante rattachée à chaque fois à l'ARNt porteur du dernier acide-amino d'apport. Lorsque le dernier acide-amino a été ajouté à la chaîne, un codon d'arrêt est mis en position dans l'ARNm par translocation. (Les trois codons déterminés comme signaux d'arrêt de chaîne sont d'une part l'UAA et l'UAG, désignés en génétique respectivement par codon « ocre » et « ambre » et, d'autre part, l'UGA. Ces trois codons n'ont d'incidence sur aucun acide-amino et sont souvent appelés les « codons du non-sens ».)

Une fois le codon d'arrêt en position, la chaîne terminée quitte le ribosome, processus suivi d'un relâchement de la dernière molécule d'ARNt et de la dissociation du ribosome en deux sous-unités. Dans les organismes procaryotes, le groupe de formyle est séparé de la méthionine à l'extrémité de démarrage de la chaîne. La méthionine résiduelle est souvent supprimée et dans le système procaryotique aussi bien chez les procaryotes que chez les eucaryotes.

Pendant la synthèse de protéines, plusieurs ribosomes synthétisent

en même temps, des protéines sur une molécule d'ARNm. Le complexe d'ARNm et des ribosomes, avec leurs chaînes de peptide croissantes, est désigné comme polysome (voir fig. 3). Dans les cellules des procaryotes, la synthèse intervient alors que l'ARNm est toujours rattachée à l'ADN et probablement pendant que les dernières sections d'ARNm se synthétisent. Dans les cellules des eucaryotes, la synthèse de protéines se fait évidemment dans le cytoplasme, séparée du lieu de la synthèse d'ARNm dans le noyau.

Ainsi, le système susmentionné produit finalement la séquence d'acides-amino qui forme la première structure de la protéine. Les liaisons chimiques et physiques responsables des structures secondaire et tertiaire — le mode d'interconnexion des différentes chaînes ou parties de chaîne aboutissant à la forme tridimensionnelle des protéines — sont sans aucun doute formées ultérieurement par modification post-translacionnelle.

Un article de cette nature rend le développement détaillé du thème impossible, mais des indications supplémentaires se trouvent dans les textes spécialisés tels que l'excellent ouvrage de Davidson (*).

Dans le second article, il sera question de quelques possibilités actuelles et futures d'application de nos connaissances en biologie moléculaire.

(1) J. N. Davidson, *La biochimie des acides nucléiques*, Chapman and Hall, et Sciences Paperbacks, Londres 1972.