

LES  
CAHIERS  
DE  
L'INSTITUT  
DE LA VIE

1968 No 15

---

# LES CAHIERS DE L'INSTITUT DE LA VIE

---

Siège: 89, B<sup>d</sup> Saint Michel, Paris V<sup>e</sup>  
Téléphone: 033-94-86  
Périodicité: trimestrielle

Prix du numéro: France 5 F. Étranger 6 F  
Abonnement: France 18 F. Étranger 22 F  
Conditions spéciales aux membres de  
l'Institut de la Vie.  
Renseignements au Siège.

---

## SOMMAIRE

---

CONFÉRENCE INTERNATIONALE  
PHYSIQUE THÉORIQUE ET BIOLOGIE  
Versailles 26-30 juin 1967

---

pages

*Journée du 29 juin 1967, 1<sup>ère</sup> séance*

SEVERO OCHOA (New York): Genetic Coding . . . . .	211
FRANÇOIS GROS (Paris): Remarques sur le Code Génétique . . . . .	215
ALFRED FESSARD (Paris): Les Problèmes du Code Nerveux . . . . .	230
MAARTEN A. BOUMAN (Soesterberg): Quantum Noise and Vision . . . . .	246
Discussions. . . . .	250

---

Les rapports et les discussions de la Conférence Internationale de Physique Théorique et Biologie, tenue à Versailles du 26 au 30 juin 1967, sont publiés dans les numéros 11 à 19 des Cahiers de l'Institut de la Vie.

Ils font d'autre part l'objet d'un ouvrage intitulé: Physique Théorique et Biologie — Theoretical Physics and Biology — publié en 1969 par la North-Holland Publishing Company, Keizersgracht 305-311, P.O.B. 3489, Amsterdam C (Pays-Bas). Par souci d'unité, nous avons respecté dans les présents cahiers, la numérotation des pages de l'ouvrage.

Les langues officielles de la conférence étaient le français et l'anglais. Nous reproduisons les textes dans les langues d'expression de leurs auteurs.

*Journée du 29 juin 1967*

---

## L'INFORMATION EN BIOLOGIE

*1ère séance*

PRÉSIDENT H. C. LONGUET-HIGGINS F.R.S.

---

S. OCHOA

Genetic Coding

F. GROS

Remarques sur le Code Génétique

A. FESSARD

Les Problèmes du Code Nerveux

M. A. BOUMAN

Quantum Noise and Vision

Discussions

## GENETIC CODING

S. OCHOA

*University School of Medicine, New York, U.S.A.*

### **Background**

#### *Genetic Expression*

The genetic information of living organisms and DNA viruses is contained in one of the two DNA strands. It is transcribed by transfer to a special messenger RNA through a DNA-directed synthesis of messenger. This RNA is an exact replica of the DNA strand that bears the genetic information and it programs the synthesis of proteins with features specified by the original DNA blueprint. Genetic and other experiments show that a linear sequence of deoxyribonucleotides in DNA specifies a corresponding sequence of ribonucleotides in messenger RNA and this in turn directs the synthesis of polypeptide chains with a unique sequence of amino acids. Thus, the four character (the four nucleotide bases) language of nucleic acids is translated into the twenty character (the twenty amino acids) language of the proteins.

#### *Genetic Code*

Clearly a linear sequence of several nucleotide bases must specify each of the twenty amino acids. A doublet code (two bases for one amino acid) would be insufficient to specify twenty amino acids for it would have only  $4^2 = 16$  doublets, but a triplet code, with  $4^3 = 64$  triplets, would contain enough information. There is evidence that the genetic code is a triplet code. Moreover, the code is non-overlapping and commaless. This means that in a sequence ABCDEFGHI...XYZ, ABC would specify one amino acid, DEF another one, and so forth.

#### *Molecular Mechanism of Translation*

Assembly of the polypeptide chains of proteins takes place on the ribosomes as they move along the messenger. The amino acids are taken to the site of synthesis in an activated form linked to special transfer RNA molecules each of which is specific for one of the twenty amino acids. Their alignment in a sequence prescribed by the nucleotide sequence of the messenger depends on the recognition of the various base triplets of the messenger (codons) by triplets of complementary base sequence (anticodons) of the amino acid-carrying RNA's. Codon-anticodon recognition and interaction are believed to occur through a Watson-Crick base pairing mechanism.

*Deciphering Of Genetic Code*

Cell-free systems of protein synthesis can be obtained from bacteria, reticulocytes, and other cells. They consist of ribosomes and supernatant fluid. The latter contains, among other things, various soluble enzymes required for the process. When supplemented with transfer RNA's, ATP, GTP, and messenger RNA these systems synthesize proteins characteristic of a given messenger, e.g., viral coat proteins, when viral RNA messengers are used. Natural messengers can be replaced by synthetic polyribonucleotides of known base composition. Poly U directs the synthesis of polyphenylalanine, poly A that of polylysine. Random polynucleotides such as poly UG, direct the synthesis of peptides containing phenylalanine, cysteine, valine, lysine and tryptophan among other amino acids. These observations opened the way for deciphering the genetic code for they showed that UUU and AAA are phenylalanine and lysine codons, respectively, whereas triplets containing 2 U and 1 G or 2 G and 1 U are codons for cysteine, valine, glycine and tryptophan.

With use of a variety of synthetic polynucleotides, the base composition of some fifty codons was established three years ago. Clearly there is more than one codon for each amino acid indicating redundancy of the genetic code. The base sequence of the individual codons has been recently established by studies of (a) the specific binding of aminoacyl-transfer RNA's to ribosomes in the presence of trinucleotides of known base sequence, and (b) the synthesis of polypeptides with artificial messenger polynucleotides of alternating base sequence. For example, poly (UG)<sub>n</sub> with UGU and GUG codons, promotes the synthesis of polypeptides containing strictly alternating cysteine and valine residues.

*Polarity of Translation*

Since polynucleotide chains have a polarity, it was of interest to know the direction in which the messenger RNA chain is read during translation. This question has been answered with use of short polyadenylic acid messengers with a unique triplet of known base sequence at either end of the chain. For example, polynucleotides such as AAAAAA.....AAAAC direct the synthesis of polypeptides of the structure lysine-lysine . . . . lysine-asparagine with NH<sub>2</sub>-terminal lysine and COOH-terminal asparagine (lysine codon, AAA; asparagine codon, AAC). Since polypeptides are assembled from the NH<sub>2</sub>- through the COOH-terminal end, the above results unequivocally establish the direction of reading of the message. The same conclusion has been reached from experiments on hybridization of insertion-deletion mutants of T2 bacteriophage affecting the synthesis of the phage-induced enzyme lysozyme.

Other experiments indicate that the ribosomes start the reading of synthetic polynucleotide messengers at one end of the chain (the so-called 5'-end) and that a special codon (AUG) at this end sets the reading frame.

### *Beginning and End of Translation*

Synthesis of natural polypeptide chains appears to require signals for initiating and terminating translation of individual cistrons. Initiation involves codons (e.g., AUG) that direct the introduction of N-formylmethionine as the first (NH<sub>2</sub>-terminal) amino acid of the polypeptide chain. Two protein factors, normally associated with the ribosomes, are required for initiation of translation. Termination involves release of the completed peptide from the ribosomes preceded or followed by elimination of the transfer RNA attached to the peptide chain. The fact that certain mutations give rise to premature release of unfinished polypeptide chains suggests the existence of special codons for chain termination. Recent experiments with oligonucleotide messengers of specified base sequence have confirmed the prediction that UAA is a chain termination codon.

### **Translation of polycistronic messengers**

The main topic for discussion will center on some aspects of the translation of the RNA of RNA-containing bacteriophages. Its small size would seem to be advantageous for these studies.

### *Translation of MS2 RNA*

MS2 RNA has a molecular weight of about  $1 \times 10^6$  (about 3000 nucleotides) with potential information for the specification of 1000 amino acids if the full length of the RNA is translated.

Thus, this RNA could program the synthesis of from three (average mol. wt.  $\approx 30000$ ) to five (average mol. wt.  $\approx 20000$ ) virus-specific polypeptides. To date, genetic studies and studies of the viral proteins synthesized in *E. coli* cells infected with MS2 (or the related phages f2 or R17) both show the production of three viral polypeptides designated, in order of their electrophoretic mobility, as polypeptides I, II, and III. They have been characterized as viral RNA synthetase (s), a protein (maturation factor) required for the formation of viable phage particles, and viral coat protein, respectively. The possibility that one or more, as yet undetected, polypeptides are formed cannot be excluded at present.

### *Control of Translation*

Initiation of translation may be controlled by the already mentioned ribosomal factors. They function, at least in part, by enhancing the binding of formyl-methionyl-transfer RNA to ribosomes mediated by the AUG codon. Particularly intriguing are the mechanisms concerned with the relative rates of translation of the various genes of the polycistronic messenger. Considerably more coat than synthetase or maturation polypeptide is produced either *in vivo* or *in vitro*. In the

latter case it has been found that the ratio of coat to RNA synthetase polypeptide formed is at least 15 : 1. This appears to be due to repression of translation of the synthetase cistron by the coat polypeptide. Moreover the coat also appears to repress translation of the maturation cistron. Other observations suggest that the maturation factor may enhance the translation of the coat cistron.



## REMARQUES SUR LE CODE GENETIQUE

FRANÇOIS GROS

*Institut de Biologie Physico-Chimique, 13 Rue Pierre Curie, Paris 5ème*

On distingue aujourd'hui deux phases distinctes dans les mécanismes mis en oeuvre pour convertir l'information génétique présente dans le DNA en des structures protéiques définies. La première représente une transposition de la séquence constituant le texte génétique en une séquence chimique nouvelle, celle que représente l'alignement des nucléotides dans le RNA. Cette étape est généralement dénommée "transcription". La seconde est essentiellement la traduction en protéines spécifiques du texte représenté par les groupes de nucléotides qui forment la charpente du RNA. On appelle cette seconde étape la *traduction informative*.

### Transcription

Cette étape repose sur un mécanisme enzymatique relativement simple. Elle est "catalysée" en effet par une protéine oligomérique qui a pu être obtenue sous un très grand état de pureté chimique, l'enzyme RNA polymérase. Le système issu des cellules d'E. coli a été particulièrement bien étudié. Il s'agit d'une protéine dont le poids moléculaire est voisin de 500 000 daltons. Elle comporte probablement plusieurs protomères, lesquels sont séparément inactifs. Les conditions qui régissent l'équilibre entre le polymère et ses sous-unités ne sont que partiellement connues. De fortes concentrations en cations monovalents, particulièrement en ions  $K^+$  ou  $Na^+$  favorisent la dissociation en sous-unités. Il y a des raisons de penser que l'enzyme se dissocie en ses sous-unités avant de se fixer sur les sites d'initiation.

Puisque le DNA est composé de deux chaînes complémentaires, la possibilité s'offrirait a priori que chacune d'elles soit transcrite simultanément. Cette situation eut abouti à la production de deux chaînes de RNA de polarités opposées. Il est cependant établi que seule l'une des deux chaînes est transcrite (transcription asymétrique). Le RNA messenger est donc composé de molécules monofilaires. Ceci ne veut pas dire, toutefois, que des *portions distinctes* de chacune des deux chaînes de DNA ne puissent servir de modèle dans la synthèse des messagers formés. C'est ainsi que les RNA messagers formés au cours du développement d'un virus bactérien, le bactériophage lambda, sont transcrits à partir de territoires distincts présents sur chacune des deux chaînes complémentaires du DNA viral.

Ceci démontre aussi incidemment que le système transcripteur (la RNA polymérase) peut se "mouvoir" dans des directions opposées selon les phases considérées de l'expression génétique.

Bien qu'il soit possible de synthétiser des RNA messagers *in vitro* en mélangeant une préparation de DNA, de la polymérase et les substrats de la réaction (les nucléosides triphosphates), on ne possède pas encore la certitude que les molécules de RNA ainsi formées soient tout à fait conformes à celles des messagers naturels correspondants. La seule preuve tangible sera la démonstration que du RNA ainsi fabriqué *in vitro* peut induire à son tour la synthèse d'une protéine spécifique. Les expériences récentes du biochimiste allemand Zillig paraissent indiquer que le RNA produit par copiage "acellulaire" du DNA provenant du bactériophage T<sub>4</sub> est capable d'orienter la synthèse d'un enzyme fonctionnel, la cytosine désaminase, lorsqu'il est incubé avec les extraits adéquats.

Une autre inconnue qui paraît subsister a trait à la nature chimique de la ponctuation qui, sur le DNA, démarque les signaux de départ et d'arrêt dans la transcription d'un gène, lequel est, comme on sait, un segment de DNA de longueur définie. L'idée qui prédomine actuellement est que la ponctuation "départ"—ce que les généticiens dénomment à la suite des travaux de Jacob, un *promoteur*—serait constituée par une séquence "réitérative" de pyrimidines, telle que par exemple une succession de 3, 4, 5 résidus des bases thymine ou désoxycytosine, ou plus (voir fig. 1, tirée d'une publication de Szybalski).

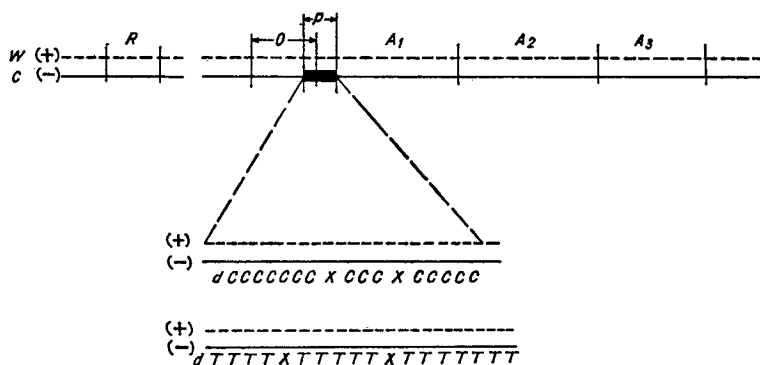


Fig. 1.

Deux arguments plaident en faveur de cette hypothèse: on sait en premier lieu que, si la RNA polymérase opère la transcription chimique d'un DNA préalablement dénaturé, en présence d'ATP comme seul substrat, de grandes quantités de polyribo A (poly A) sont formées. Une telle réaction s'observe également si l'on utilise comme matrice synthétique des oligonucléotides à chaînes courtes (tels que (oligo dT) 5-7). Le poly A formé peut comporter plus de 50

résidus d'adénine car l'enzyme "dérage" au contact des courts segments formés par l'union des nucléotides à thymine et "incorpore ainsi un très grand nombre de résidus adényliques en liaisons covalentes.

Un autre argument en faveur du rôle des pyrimidines dans la ponctuation de "départ" réside dans l'observation aujourd'hui bien étayée selon laquelle les premiers résidus nucléotidiques incorporés *in vitro* dans une chaîne de RNA messenger sont toujours des dérivés de purines. Selon l'origine du DNA servant de matrice, les purines incorporées en position proximale sont soit l'adénine, soit la guanine. Ceci est conforme à l'idée que les points d'attachement au niveau desquels l'enzyme commence (initie) sa réaction de copiage comportent des séquences de thymine ou de désoxycytosine (Hurwitz, Bremer et Konrad, etc.).

Egalement à l'appui de ces hypothèses viennent les expériences de Sheldrick et Szybalski. Ces auteurs ont observé que, si l'on dénature du DNA par la chaleur ou l'alcali, les chaînes monocaténaïres peuvent facilement former des complexes avec le poly A ou le poly G, résultat suggérant l'existence de faisceaux localisés (clusters) de thymine ou de désoxycytosine à l'intérieur de l'une ou l'autre de ces chaînes.

Le fait que la polymérase manifeste beaucoup plus d'affinité pour les chaînes de DNA (ou de RNA) monofilaires que pour les structures bihélicoïdales (DNA natif, complexe poly A, poly U, etc.) suggère que les sites au niveau desquels débute la transcription (les promoteurs de Jacob) pourraient fort bien être des régions partiellement dénaturées de la double hélice (voir fig. 1).

Certains biochimistes vont jusqu'à penser que les résidus de purines présents dans la "boucle" ainsi créée existeraient peut-être à l'état "méthylé" afin d'empêcher un réappariement local (Stent).

Depuis les travaux désormais classiques de Hurwitz, Bremer et Konrad, et de Hayashi, on a tout lieu d'admettre qu'au fur et à mesure que la polymérase s'éloigne de son point d'origine (le "start signal" des auteurs anglais), la double hélice de DNA se déroule sur son passage mais se reforme immédiatement après. Le DNA ayant servi de matrice à la formation de RNA messenger conserve en effet intactes ses propriétés de double hélice; il demeure "natif" tant en se référant aux critères d'hyperchromicité qu'aux tests biologiques (transformation). La chaîne de messenger "naissant" est probablement expulsée par reformation progressive de la double hélice. Elle demeure toutefois rattachée au DNA par la molécule même de l'enzyme en déplacement.

On ignore la nature des signaux chimiques qui, sur le DNA, déterminent l'arrêt de la transcription, signaux présents à l'extrémité distale des gènes. Certains auteurs contestent même l'existence d'une ponctuation d'arrêt en tirant argument du fait que, si l'on change les conditions ioniques au cours de la transcription *in vitro*, on peut synthétiser des chaînes de RNA messagers démesurément longues. Il se pourrait donc que les chaînes de RNA soient normalement "découpées" par des éléments du système traducteur.

**Traduction informative: Synthèse des protéines**

La seconde étape dans le mécanisme chimique qui permet de transférer l'information du DNA aux protéines est la traduction du RNA messenger. C'est ici qu'il convient de parler de décodage ou décryptage puisque nous voyons qu'une macromolécule (le RNA) dont la structure est constituée par l'assemblage de 4 types de lettres (les quatre bases du RNA) peut orienter l'assemblage colinéaire des protéines, lesquelles sont formées par l'union selon un enchaînement spécifique de 20 types distincts d'acides aminés.

Pour comprendre, même dans ses grandes lignes, le mécanisme de la traduction, il importe, d'une part, de saisir la nature même du code génétique (c'est-à-dire quels types de combinaisons de bases codent pour tel ou tel acide aminé) et de connaître les éléments de la machinerie cellulaire qui participent à la traduction.

*Considérations générales sur le code:* Puisqu'il existe 20 types distincts d'acides aminés et seulement 4 types principaux de bases, l'idée s'est de bonne heure imposée (Gamow) que le code ne pouvait être biunivoque. En d'autres termes, seules des combinaisons (des enchaînements) de bases selon une séquence définie, devaient exister en nombre égal ou supérieur à celui des acides aminés présents dans les protéines. Ce type de raisonnement devait conduire Crick et Brenner à rejeter un code de type binaire comme fort peu vraisemblable puisqu'il n'existe que 16 combinaisons de 4 lettres sous la forme de doublets. Un code à triplet semblait pouvoir convenir puisque le nombre de triplets possibles à partir de 4 types de lettres est de  $4^3 = 64$ , soit plus qu'il n'apparaissait nécessaire pour la représentation codée de chaque amino-acide. Aussi l'idée selon laquelle la plus petite unité d'information devait être un triplet composé de nucléotides a-t-elle été avancée sur des bases purement théoriques.

La démonstration de la nature ternaire du code a été fournie ultérieurement en étudiant, les protéines formées par recombinaison entre des mutants du bactériophage T<sub>4</sub>, lesquels avaient été obtenus séparément sous l'influence d'un agent particulier, la proflavine, un composé acridinique. Mais, avant de décrire le principe de cette très importante expérience, il importe de faire quelques remarques générales sur les modalités qui peuvent a priori se présenter dans le déchiffrement d'un code à 3 lettres.

Si les gènes sont définis, en première approximation, comme les segments délimités de la double hélice d'ADN, on doit pouvoir se représenter les ARN messagers qui en dérivent comme une succession de nucléotides au sein de laquelle des groupes de 3 nucléotides définissent un code. Il existe cependant de nombreuses façons d'entrevoir la nature d'un tel code ou, si l'on préfère, plusieurs cadres de lecture.

Considérons une séquence composée de 4 types de lettres, soit:

B C D A A B C D B A A C ...

Plusieurs dispositions des lettres peuvent être imaginées. Celles-ci peuvent être *adjacentes* comme dans l'exemple ci-dessous où la lecture débute par la première lettre:

(1)                    B C D    A A B    C D B    etc. . . .  
                           └──┘    └──┘    └──┘

Les triplets peuvent également être séparés par des lettres ne correspondant à aucun acide aminé précis mais servant par exemple de virgules:

(2)                    B C D , A, A B C , D, B A A    etc. . . .

Dans les deux types de codes décrits ci-dessus, chaque groupe de 3 lettres ne code que pour un seul aminoacide *à la fois*. On peut cependant imaginer un groupage tel que chaque lettre d'un triplet donné puisse être considéré comme faisant également partie du triplet adjacent. C'est le cas par exemple des codes dits "chevauchants":

(3)    A B C D E F G H                    le premier triplet déchiffré sera A B C  
          └──┘ └──┘                            le second                                    B C D  
          └──┘ └──┘                            le troisième                                C D E  
          └──┘ └──┘                            etc.

L'expérience démontre en fait que le code de lecture est bien un système à base 3 et précise en outre qu'il est *non* "chevauchant" et ne comporte pas de virgule. Il correspond donc à l'exemple N° 1.

Les évidences en faveur d'un code ternaire sont nombreuses et ne peuvent toutes être analysées ici.

Une des approches consiste à déterminer le nombre de nucléotides entrant dans la composition d'un ARN messenger spécifique ainsi que le nombre de résidus "amino-acides" présents dans la protéine dont la synthèse est codée par ce messenger. Il est malheureusement très difficile d'isoler d'une cellule l'ARN messenger particulier à une protéine. En outre, ainsi que nous le verrons, bon nombre d'ARN messagers cellulaires contiennent de l'information pour plus d'une protéine spécifique (messagers polycistroniques). Les ARN messagers naturels qui se prêtent le mieux à l'isolement et dont le contenu informatif correspond le plus souvent à une seule protéine (ou à un petit nombre d'entre elles) sont les acides ribonucléiques des virus animaux ou végétaux ou encore ceux des bactériophages. C'est ainsi que les acides nucléiques d'un grand nombre de petits virus ont des poids moléculaires proches de  $1,5 \times 10^6$  (ce qui correspond à environ 4 300 nucléotides) et ils contiennent fréquemment 3 types distincts de protéines dont les PM sont compris entre 30 000 et 40 000, c'est-à-dire qui renferment 400 résidus. Le rapport spécifique de codage doit donc être voisin de  $4\,300/400 \times 3 = 3,5$ . L'ARN du virus nécrotique des feuilles de tabac ren-

ferme 1200 nucléotides et code pour une seule protéine de capsid, laquelle comporte 400 résidus. D'où un rapport de 3 pour la plus petite unité de codage, etc.

Des données beaucoup plus précises sur les valeurs du rapport de codage (coding ratio) proviennent cependant des expériences de Crick et Brenner sur le brouillage du texte génétique par les acridines. Nous reviendrons sur leur principe dans un instant.

Sans doute non moins directe est la démonstration apportée par Nirenberg et Leder que, lorsque l'on mélange *in vitro* des trinuécléotides (généralement obtenus par hydrolyse enzymatique d'acides nucléiques suivie de fractionnements appropriés), des ribosomes et des ARN de transfert estérifiés par leurs aminoacides correspondants, on forme des complexes trinuécléotides-ribosomes-ARN de transfert spécifiques. Les mono et dinuécléotides ne permettent pas la "rétention spécifique" des ARN de transfert.

Ainsi, il s'avère que le cadre de lecture est bien élaboré par des groupes de 3 nucléotides.

Les expériences de Crick, Brenner et Watt-Tobins auxquelles nous nous sommes déjà référés plus haut démontrent en outre que la traduction des ARN messagers débute en un point précis, l'extrémité proximale de la chaîne, et chemine d'une manière unidirectionnelle. Le déchiffrement est réalisé par groupes de triplets et ne fait pas intervenir de virgules.

Ces expériences tirent parti de ce que les acridines causent facilement des mutations dans l'ADN des bactériophages en provoquant soit des micro-insertions (de bases nouvelles) dans la séquence d'un gène, soit au contraire en causant des micro-délétions.

Supposons un texte génétique dont la lecture par groupe de 3 lettres ait une signification précise. Pour simplifier, nous utiliserons ici des combinaisons de mots:

T O N A M I E S T I C I

Lu correctement, en commençant par le début, ce texte signifie "TON AMI EST ICI". Si, à la suite d'une mutation, nous "insérons" par exemple une lettre supplémentaire dans ce texte et que nous continuions à le déchiffrer sur le mode précédent, il est évident que nous faussons le cadre de lecture:

↓  
T O R N A M I E S T I C I

donne un message brouillé

T O R N A M I E S T I C I . . .

Si nous affectons d'un signe "+" une telle mutation qui allonge d'une lettre le texte génétique, il devient clair qu'il existe seulement deux moyens pour rétablir la signification du moins partielle du message:

— l'un est d'introduire une mutation de signe—c'est-à-dire déléter l'une des lettres dans le texte faussé:

$\begin{array}{ccccccccccc} & \downarrow & & & \downarrow & & & & & & & \\ & \dagger & & & \bar{\phantom{\dagger}} & & & & & & & \\ \text{T} & \text{O} & \text{R} & \text{N} & \text{A} & \text{M} & \text{E} & \text{S} & \text{T} & \text{I} & \text{C} & \text{I} \end{array}$

Dans un tel message, la fin du texte redevient compréhensible.

— l'autre consiste à introduire deux mutations supplémentaires de même polarité (+) que la précédente:

$\begin{array}{ccccccccccc} & \downarrow & & & \downarrow & & & & \downarrow & & & \\ & \dagger & & & \dagger & & & & \dagger & & & \\ \text{T} & \text{O} & \text{R} & \text{N} & \text{A} & \text{U} & \text{M} & \text{O} & \text{I} & \text{E} & \text{S} & \text{T} & \text{I} & \text{C} & \text{I} \end{array}$

Il est clair que, si une telle situation peut être génétiquement vérifiée, elle apporte la preuve que l'information est déchiffrée par groupes de 3 nucléotides, non séparés par des virgules, la traduction étant unidirectionnelle.

Ce but a été atteint par les auteurs anglais en étudiant les effets mutationnels, simples ou combinés, de la proflavine sur la synthèse d'un enzyme spécifique, le lysozyme, par le gène correspondant présent dans l'ADN du bactériophage T<sub>4</sub>. Par le jeu des recombinaisons génétiques, on peut en effet introduire diverses mutations de signes + (insertion d'une base) ou — (délétion d'une base) dans le gène considéré et définir les situations qui *restaurent* l'activité de l'enzyme correspondant.

Les conclusions de Crick et ses collaborateurs ont d'ailleurs été étayées peu après grâce à l'analyse de la *séquence* peptidique du lysozyme normal, et des pseudo-lysozymes dont l'activité avait ainsi été rétablie. Il a été vérifié que, dans le cas d'un pseudo-enzyme de type (+ —), l'un des peptides de la chaîne renfermait un amino acide *de plus* que le peptide correspondant dans l'enzyme sauvage, tandis qu'un peptide plus distal était déficient en un amino acide par rapport à l'homologue normal.

### Nature du code génétique

Il n'a pas fallu beaucoup plus de 5 années pour établir la nature chimique du code génétique. Notre propos n'est pas de décrire ici les diverses phases des recherches qui ont permis cette remarquable réalisation de la Biologie Moléculaire.

Il est toutefois intéressant de rappeler quelles ont été les grandes voies d'approche.

Celle qui devait à coup sûr ouvrir le champ à ces investigations est sans nul doute l'observation due à Nirenberg et Matthaei, selon laquelle il est possible d'orienter *in vitro* la synthèse de polypeptides artificiels en substituant aux ARN messagers naturels des hauts polymères dont les biochimistes savent, depuis les travaux d'Ochoa et Grunberg-Manago, réaliser la synthèse. De tels polymères peuvent en effet être obtenus sous l'influence de la polynucléotide phosphorylase,

un enzyme capable de condenser selon un enchaînement statistique n'importe quel nucléotide, que celui-ci soit d'un type donné ou qu'il s'agisse d'un mélange de plusieurs types. On obtient ainsi toute une gamme d'homopolymères et de copolymères contenant 1, 2 et jusqu'à 4 types de bases.

Lorsque ces polymères synthétiques sont mis en présence de ribosomes purifiés extraits de la bactérie *E. coli*, et que le mélange réactionnel comporte en outre les aminoacides appropriés ainsi que divers autres effecteurs, on assiste à une synthèse d'homo ou de copolypeptides dont la composition en aminoacides dépend, fait remarquable, de celle des polynucléotides ajoutés.

Cette observation et son exploitation systématique par les laboratoires de Nirenberg et de Ochoa devait rapidement conduire à élucider la *composition* de presque tous les codons.

Restait à établir leur séquence. Il existe en effet 9 combinaisons de 3 lettres. Ici, deux méthodes ont principalement été utilisées. L'une déjà évoquée ci-dessus a consisté à synthétiser chimiquement ou à préparer par digestion enzymatique la quasi-totalité des trinucéotides de *séquences* connues et à identifier avec chacun d'entre eux la nature des amino-acyles-t-ARN dont ils permettent l'attachement aux ribosomes. Ceci peut être réalisé très commodément grâce à l'emploi de membranes millipores qui se laissent normalement traverser par les ARN adaptateurs, estérifiés ou non, mais qui retiennent ces derniers si un appariement correct est réalisé à la surface du ribosome entre le trinucéotide et l'anticodon spécifique du t-ARN (voir ci-après). Il est dès lors facile de préciser parmi les différents types de trinucéotides comportant une séquence de 3 bases données quel est celui, ou ceux, dont l'interaction avec un aminoacyl-t-ARN en permet (ou permettent) la fixation sur les ribosomes.

Une autre méthode a consisté à utiliser des copolynucléotides dont la séquence n'est pas distribuée au hasard (comme c'est le cas dans les copolymères formés par la polynucléotide phosphorylase) mais répond au contraire à une alternance définie telle que: A U A U A U A U etc. . . . ou A A C A A C A A C. Il est clair en effet que, si de tels polymères orientent la synthèse de copolypeptides *in vitro*, on peut, en déterminant l'alternance des aminoacides, définir du même coup la *séquence* des codons correspondants.

Toutes ces études ont en définitive abouti à préciser la signification des 64 trinucéotides pouvant normalement exister. Ceci a conduit à un certain nombre de conclusions générales.

1. — La première—et sans doute la plus importante—est qu'il existe toujours plus d'un triplet (encore appelé codon) capable de spécifier un aminoacide donné. En d'autres termes, à part trois exceptions que nous envisagerons dans un instant, *tous* les trinucéotides ont un sens dans l'alphabet nucléique. Il n'y a pas de correspondance biunivoque entre codons et aminoacides. Pour employer une terminologie consacrée, on dit encore que le code est dégénéré. L'importance de ce fait dans le cadre des mécanismes moléculaires de l'évolution est considérable:



qu'il existe plusieurs systèmes de codage pour un même aminoacide rend possible l'existence, au sein d'espèces dont les matériels génétiques (ADN) sont de compositions très différentes, de protéines-enzymes possédant des structures primaires suffisamment voisines pour y jouer le *même rôle* biologique. La dégénérescence du code permet donc à la cellule d'exercer une sorte d'*effet tampon* contre les variations que l'évolution physiologique et les mutations ne manqueraient pas d'apporter à la structure du matériel génétique.

Bien qu'il existe souvent deux ou trois codons distincts, voire davantage, pour un même aminoacide, les codons "homonymes" ont une particularité commune très importante: les deux premiers des nucléotides qui les constituent sont les mêmes. Seuls diffèrent les nucléotides situés en *troisième* position dans la séquence.

On était en droit de se demander s'il existe un ARN de transfert possédant une structure appropriée à *chacun des 64* codons possibles ou si l'ARN de transfert correspondant à un aminoacide défini pouvait servir d'adaptateur sinon pour tous, du moins pour plusieurs, les codons homonymiques. Cette dernière alternative a été suggérée par Crick dans sa théorie dite du "Wobble", d'après laquelle la spécificité de l'appariement du codon et de l'anticodon du t-ARN repose surtout sur la séquence des *deux premières lettres du triplet* (ou des deux dernières dans l'anticodon, puisque l'appariement messenger-t-ARN se fait de manière antiparallèle). En d'autres termes, un certain degré de "jeu" (Wobble) est permis dans les interactions entre le troisième nucléotide présent dans le triplet du messenger et le premier appartenant à la séquence de l'anticodon. Le code a donc une certaine *flexibilité* qui, jointe à son caractère de dégénérescence, le prémunit contre des changements mutationnels qui à tout coup risqueraient d'être léthaux.

2. - Trois codons n'ont pas de signification définie dans le langage protéique. Il s'agit des triplets UAG, UAA et UGA. Ces triplets jouent le rôle de signes de ponctuation dans la traduction. Ce rôle apparaît des plus essentiels si l'on réalise que bien souvent, du moins chez les microorganismes ou leurs virus, plusieurs gènes *contigus* sont transcrits en un ARN messenger *unique* (transcription polygénique). Il est donc très fréquent qu'une même chaîne d'ARN contienne l'information pour plusieurs protéines distinctes. Cette organisation particulière du matériel génétique est liée en fait au mécanisme de la régulation (Jacob et Monod). Il est en effet plus facile de contrôler l'activité des gènes impliqués dans une même fonction physiologique si ces gènes sont exprimés d'une manière coordonnée, ce que rend possible leur "cotranscription".

Pour que le système de lecture (voir plus loin) puisse traduire un messenger unique en plusieurs protéines distinctes, il faut nécessairement une ponctuation qui délimite le début et la fin de la lecture. Nous dirons quelques mots ci-après des mécanismes d'initiation.

En ce qui concerne les signaux de terminaison de lecture, notons que les 3 codons UAA, UAG et UGA ont précisément ce rôle. Il semble cependant que dans les conditions normales l'arrêt de lecture ne soit signalé que par le seul

triplet *UAA*; *UAG* et *UGA* peuvent apparaître au sein d'un ARN messenger sous l'effet d'une mutation, ces deux triplets interrompant alors artificiellement la propagation des ribosomes et leur présence se manifestant par la synthèse de "fragments" protéiques. Les codons *UAA* et *UAG* sont souvent dénommés "ochre" et "amber" par les auteurs anglo-saxons.

3. - Le code est universel. Un trinuécléotide donné codera pour le même amino acide quelque soit l'organisme envisagé. Ceci peut être aisément démontré par l'emploi de systèmes acellulaires.

La signification du code peut cependant être perturbée de deux façons, soit par mutations, soit sous l'effet d'agents chimiques ou physiques qui modifient la "reconnaissance" d'un codon donné par les ribosomes et peuvent ainsi en changer le sens.

Une mutation est un changement dans la séquence chimique du texte génétique ADN. Certaines de ces mutations entraînent une délétion d'une région plus ou moins importante de l'ADN. Le plus souvent, il s'agit simplement du remplacement de l'une des bases d'un triplet par une base différente, perturbation apparaissant lors de la réplication. Lorsque l'on examine le tableau de code génétique, il est clair que, par le simple changement d'un nucléotide au sein d'un triplet, on peut modifier complètement le sens de ce triplet.

C'est un fait remarquable, lequel apporte à l'établissement du code une confirmation essentielle que toutes les substitutions d'aminoacides qui s'observent dans une chaîne à la suite d'une seule mutation, (mutation ponctuelle) s'expliquent parfaitement par le remplacement d'une seule des lettres dans un codon.

Cette conclusion est étayée par les remarquables études effectuées par l'école de Yanofsky sur les structures chimiques des tryptophane synthétases d'*Escherichia coli*, altérées par des mutations distinctes ainsi que par celles de Whitman sur les propriétés des protéines génétiquement modifiées que l'on isole des mutants du virus de la mosaïque du tabac.

Un codon de séquence définie peut cependant, nous l'avons souligné, changer de signification selon le contexte biochimique. C'est le phénomène d'ambiguïté ou d'infidélité dans la traduction. De telles ambiguïtés peuvent être créées par des changements thermiques, des variations de pH, la présence de certains agents dont l'activité repose sur une modification des propriétés des ribosomes. C'est le cas par exemple de certains antibiotiques, telles la streptomycine ou la néomycine, qui, ajoutés à faibles concentrations dans des extraits renfermant des ARN messagers artificiels, peuvent modifier la composition des polypeptides formés. L'ambiguïté "spontanée", c'est-à-dire le fait qu'un codon défini puisse être traduit d'une manière aberrante au sein d'une cellule, par ailleurs normale, est un phénomène très rare, encore que décelable (Von Ehrenstein).

*Comment s'effectue la traduction des codons du messenger en protéines? :* Telle est la question que nous envisagerons en dernier lieu. L'hypothèse a priori la plus

simple voudrait qu'un aminoacide soit *directement* attaché en regard de chaque codon présent dans la séquence d'un ARN messenger.

Une telle possibilité, pour économique qu'elle puisse apparaître à première vue, se heurte d'emblée à maintes objections théoriques. En effet, ce qui distingue chaque aminoacide de son voisin, c'est la nature du groupe latéral ou radical associé au  $(R)-CH \begin{matrix} \swarrow NH_2 \\ \searrow CO_2H \end{matrix}$ . Ces radicaux ont des structures chimiques extrêmement différentes (noyaux aromatiques, aliphatiques ramifiés ou non, etc. . . .). Par ailleurs, rien dans la structure linéaire de l'ARN messenger ne constitue un élément de stéréospécificité suffisant pour "reconnaître" chacun de ces radicaux. Même si tel était bien le cas, le positionnement des aminoacides au voisinage *direct* des codons disposerait ces aminoacides à des *distances* variables de la matrice et rendrait impossible la mise en jeu des enzymes nécessaires au cimentage de la liaison peptidique.

Ces considérations ont depuis plusieurs années conduit F. Crick à prédire l'existence d'adaptateurs spécifiques, eux-mêmes de nature nucléique, capables de "placer" chaque aminoacide en regard du codon correspondant. On sait que, peu de temps après, Hoagland a découvert l'existence de ces ARN adaptateurs encore appelés ARN de transfert ou en abrégé "t-ARN".

Ces t-ARN ont en quelque sorte une vertu double qui en fait des éléments clés dans le déchiffrement du code. Ce sont de petits polynucléotides (ils renferment environ 80 nucléotides et la structure complète de maints d'entre eux est connue). Ils abritent à l'intérieur de leur chaîne une séquence trinuécléotidique, l'*anticodon*,

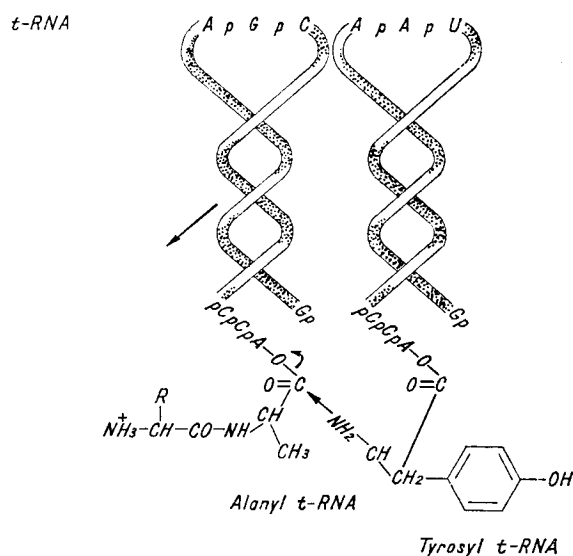


Fig. 2.

laquelle est capable de s'apparier directement à un triplet spécifique du messenger et ils sont porteurs à leur extrémité d'une séquence particulière (le résidu CCA) à laquelle se trouve rattaché l'acide aminé (fig. 2).

Les premiers stades dans la synthèse des protéines n'ont donc d'autre but, on s'en doute, que d'activer thermodynamiquement les acides aminés pour en permettre l'attachement à l'extrémité CCA. Cet attachement est, bien entendu, catalysé par un enzyme spécifique (il en existe un pour chaque acide aminé). Une fois fixés à leurs t-ARN adaptateurs, on pourrait penser que les acides aminés sont disposés le long de la chaîne du messenger et peuvent y être condensés en polypeptide. A peu de chose près, c'est en gros ce qui se passe mais il faut apporter quelques touches supplémentaires à ce tableau. En effet, bien que l'interaction primaire des amino-acyl-t-ARN et du messenger soit assurée par un appariement entre les deux triplets complémentaires, les liens hydrogènes qui en sont cause sont si fragiles qu'un tel complexe fondrait très rapidement s'il ne se trouvait renforcé par d'autres facteurs. La cellule doit donc mettre en oeuvre un dispositif spécial qui assure le renforcement de ces liaisons. C'est ici qu'interviennent les ribosomes, ces granules cytoplasmiques qui renferment la quasi totalité de l'ARN stable de la cellule.

Les ribosomes sont en effet des dimères, en ce sens qu'ils comportent une grande sous-unité (50 S) et une plus petite (30 S), l'ensemble sédimentant à 70 S. Leur rôle essentiel est de renforcer transitoirement l'association du t-ARN et du messenger et de contenir les enzymes cimentant le lien peptidique entre deux acides aminés voisins. D'une part, en effet, ils se fixent par la sous-unité 30 S à l'ARN messenger (Takanami). D'autre part, leur sous-unité 50 S contient deux

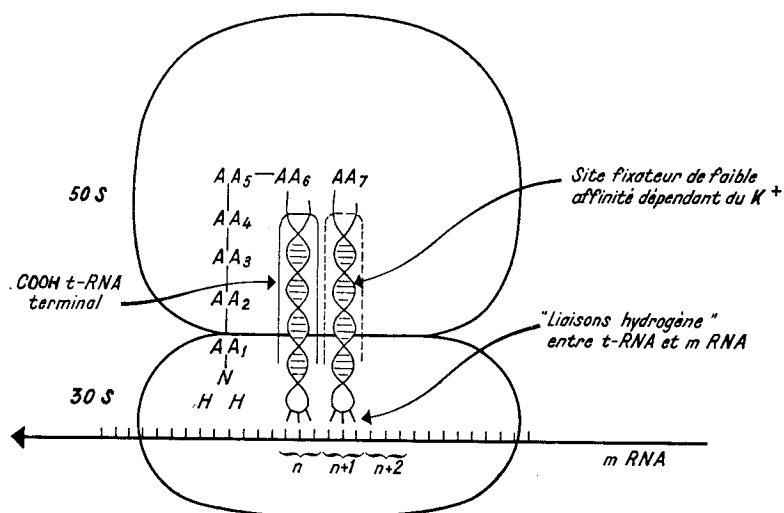
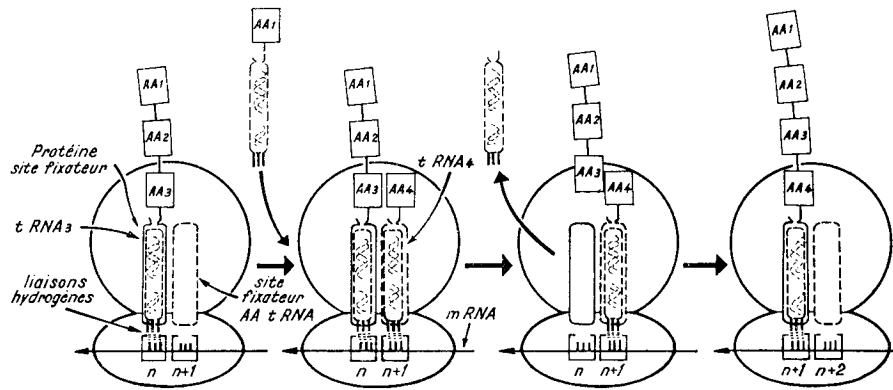


Fig. 3.



Croissance graduelle de la chaîne polypeptidique

Fig. 4.

logettes spécifiques et contigües dans lesquelles peuvent s'insérer les amino-acyl-t-ARN correspondant à deux codons adjacents (voir fig. 3).

Une peptidase synthétase (Traut et Monroë) qui fait partie intégrante de l'unité 50 S scelle les deux amino-acyles voisins par un lien peptidique (fig. 4). Un dipeptidyl-t-ARN est alors formé. L'étape suivante peut être définie comme une "contraction" du ribosome qui déplace le dipeptidyl-t-ARN de la logette où il se trouvait dans celle renfermant le t-ARN libre. Ce dernier est expulsé et un nouvel amino-acyl-t-ARN vient occuper la place laissée libre par le dipeptidyl-t-ARN. A chaque contraction (laquelle est catalysée par un enzyme) la translocase, et par un cofacteur allostérique, le GTP, le ribosome se meut d'une distance équivalente à un triplet du messenger: ainsi se trouve "tissé" le long ruban polypeptidique qui doit constituer la protéine finale. Au fur et à mesure que les premiers ribosomes attachés à la chaîne du messenger se sont suffisamment déplacés vers les régions "distales", d'autres ribosomes s'attachent en début de chaîne et commencent un nouveau cycle de synthèse. A l'état de régime, une même chaîne de messenger peut donc être "parcourue" par un train de ribosomes se déplaçant dans une même direction et tissant, chacun pour son propre compte, une chaîne polypeptidique. On donne le nom de polyribosomes (ou polysomes) à ces formations, lesquelles sont facilement décelables au microscope électronique. Une fois arrivé au niveau d'un codon UAA, signal de fin de lecture, le ribosome "relargue" le polypeptidyl-t-ARN. Cette opération est préalablement catalysée par un facteur protéique qui vient d'être identifié par Capecchi. Une hydrolase spécifique (Cuzin et Chapeville) clive alors le lien entre le t-ARN et le polypeptide tandis que celui-ci revêt sa configuration secondaire et tertiaire.

Dans le processus d'élongation du polypeptide, on sait que le premier amino acide incorporé dans la chaîne est celui qui correspond au résidu  $\alpha$ -aminé présent

dans la chaîne protéique terminale, le dernier correspondant au résidu carboxyle libre de la protéine (Dintzis-Goldstein).

Il semble que chaque messenger naturel débute par une séquence particulière qui servirait de point d'attache au premier ribosome, lequel serait d'abord fixé à l'état de sous-unité 30 S (Nomura). Cet attachement paraît catalysé par un facteur protéique spécial, dénommé facteur C, dont Revel et Gros ont récemment démontré l'existence. La séquence de fixation ne serait pas à proprement parler "traduite" en protéines. La traduction ne commencerait qu'un peu plus loin vers l'intérieur de la chaîne. Des précisions ont été apportées sur ce dernier point:

a - le premier codon traduit (chez les microorganismes) est le codon AUG (Capecchi; Khorana);

b - le premier aminoacyl-t-ARN incorporé est un dérivé découvert par Marcker et Sanger: le formyl-méthionyl-t-ARN. Ainsi, *toutes* les chaînes protéiques chez les bactéries (et chez de nombreux virus) débutent par le résidu formyl-méthionine. Ce résidu est ensuite clivé par une peptidase spécifique (Waller-Capecchi). Ainsi, le mécanisme qui assure l'initiation des synthèses protéiques revêt-il un caractère de généralité tout à fait remarquable. C'est au niveau de la sous-unité 30 S qu'a lieu la fixation du formyl-méthionyl-t-ARN, et elle est catalysée par deux facteurs protéiques particuliers (appelés F<sub>1</sub> et F<sub>2</sub>) (Wabba, Ochoa, Brawerman-Revel et Gros).

Il existe en somme une polarité stricte dans le processus traducteur qui aboutit à la synthèse des protéines. De remarquables travaux dus à Ochoa, d'une part, ainsi qu'au groupe de Streisinger, d'autre part, ont clairement établi que l'extrémité du messenger codant pour le résidu  $\alpha$ -aminé de la chaîne polypeptidique était la terminaison 5'-phosphate de l'ARN. Ainsi la traduction progresse selon une polarité 5'p  $\rightarrow$  3'OH.

Si l'on songe que la transcription (synthèse du messenger) se déroule également dans la direction 5'P  $\rightarrow$  3'OH, on voit du même coup que les deux phases de l'expression génétique (transcription et traduction) évoluent selon la même polarité. Tout porte d'ailleurs à croire que ces deux étapes sont, dans la cellule, intimement couplées (Stent-Gros). Ce sont les ribosomes qui, en s'attachant aux premières séquences transcrites du messenger, détacheraient cet ARN de sa matrice génétique. Ce point est certainement d'une grande importance dans les phénomènes cellulaires qui ajustent le taux de formation des messagers au taux de synthèse des protéines.

**Bibliographie**

Les références se rapportant à ce problème sont trop nombreuses pour pouvoir figurer *in extenso*.

Le lecteur pourra cependant se reporter aux revues suivantes:

- [1] J. Monod, F. Jacob et F. Gros, Structural and rate determining factors in the biosynthesis of adaptive enzymes, *Biochem. Soc. Symp.* **21** (1961) 104.
- [2] J. D. Watson, *Molecular biology of the gene* (Benjamin, 1965).
- [3] V. Ingram, *The biosynthesis of macromolecules* (Benjamin, 1965).
- [4] M. Grunberg-Manago, Le code génétique, *Ann. Pharmac. Franc.* **25** (1966) 147.
- [5] J. Tavlitski, Le code génétique, *Sciences* **34** (1964) 19 et **35** (1965) 14.
- [6] *The genetic code*, Symposia on quantitative biology, Vol. **31** (1966), Cold Spring Harbor Laboratory of Quantitative Biology.

## LES PROBLEMES DU CODE NERVEUX

ALFRED FESSARD

*Laboratoire de Neurophysiologie Générale, Institut Marey, Paris, France*

Les rapports qui s'établissent entre la physique—théorique ou expérimentale—et la physiologie du système nerveux, ou neurophysiologie, sont de deux ordres. On peut distinguer ces deux ordres de la même façon que l'on ne confond pas, si l'on y prête attention, les termes *fonctionnement* et *fonction*. Je m'explique:

Le système nerveux, en particulier le cerveau des vertébrés supérieurs—qui nous intéresse au premier chef en ce qu'il nous aide à comprendre le cerveau de l'homme—est sans doute le produit organisé le plus complexe de la nature. Nous admettons qu'il obéit comme tout autre organe dans son fonctionnement, aux strictes lois de la physique et de la chimie, et le neurophysiologiste cherche tout d'abord à expliciter ces lois de fonctionnement. Lorsqu'il attaque les problèmes en physicien—on le dit alors biophysicien—les concepts et les grandeurs qu'il manie sont ceux des propriétés physiques de la matière et des divers aspects de l'énergie, avec leurs deux composantes—de tension et de quantité—qu'il s'agisse de chaleur, d'électricité ou d'énergie mécanique. On les rencontre à différentes échelles, des molécules constitutives au neurone, et de celui-ci au système nerveux tout entier. Un des mécanismes nerveux fondamentaux est celui que révèle et qu'explique assez bien aujourd'hui la biophysique des membranes cellulaires, ou membranes "plasmiques", avec leurs propriétés électriques, passives et actives, parmi lesquelles celle qui se manifeste sous la forme d'une brève impulsion, le "potentiel d'action" (en anglais "nerve impulse"). Celui-ci, selon nous, pourrait être appelé en français neuroquantum ou neuropulsion. Il est cependant toujours désigné entre spécialistes par le terme anglais imagé de "spike".

Bien entendu, la seconde tâche du neurophysiologiste—c'est même la plus proprement "physiologique"—consiste à établir quelles sont les *fonctions* de ce système nerveux, c'est-à-dire les opérations auxquelles il participe majoritairement avec d'autres organes (les organes des sens, les muscles, les glandes endocrines, etc . . .) pour accomplir certaines catégories d'actions auxquelles nous reconnaissons (conventionnellement, mais avec beaucoup de bonnes raisons!) une certaine utilité biologique, utilité (préférons ce mot à "finalité") pour la survie de l'organisme total, pour son adaptation aux conditions changeantes du milieu, pour sa croissance et, éventuellement sa reproduction. Les fonctions accomplies par le système nerveux sont des performances qui, lorsque tout va bien, sont remarquablement ordonnées. En langage moderne, nous dirions qu'elles exécutent un



