

BIOCHIMIE. — *Sur l'existence d'une certaine corrélation entre le poids moléculaire des acides aminés et le nombre de triplets intervenant dans leur codage.* Note (*) de MM. **Marcel-Paul Schutzenberger, Pierre Gavaudan et Jacques Besson**, présentée par M. Jean Roche.

On observe une relation globale de proportionnalité inverse entre le poids moléculaire des acides aminés et le nombre de triplets les codant ainsi qu'une relation de même sens dans l'étude statistique de la composition des protéines.

Des études théoriques antérieures consacrées à la structure du code génétique [(¹), (²)] nous ont conduits à rechercher au sein des classes de codage définies par les nombres de triplets affectés aux acides aminés qu'elles contiennent, des paramètres susceptibles d'individualiser simplement les acides aminés. En effet, quoique de composition connue les classes ne donnent que des informations statistiques faibles puisqu'au sein de chacune d'elles l'identité des acides aminés est ignorée. L'analyse par segments (²) indique dans certaines chaînes des fluctuations sensibles de la classe II, mais là encore l'anonymat des acides aminés fait perdre de précieuses informations. Nous avons pensé qu'il serait peut-être instructif de considérer ce paramètre très simple qu'est le poids moléculaire des acides aminés qui, à notre connaissance, a été laissé de côté jusqu'à ce jour.

Certes l'idée de choisir un paramètre de nature aussi triviale n'est-elle pas très attirante *a priori*, car il est bien certain qu'il fait abstraction de la forme et de la constitution des molécules ainsi que de leurs propriétés physiques et chimiques. Toutefois, représentant une sorte d'indice global de complexité, l'utilisation du poids moléculaire, facile du point de vue statistique, conduit à des résultats appréciables.

Nous disposons tout d'abord les poids moléculaires des acides aminés sur une échelle où ils croissent de gauche à droite ; ils figurent entre parenthèses auprès du numéro de la classe à laquelle appartient l'acide aminé correspondant. GLY (IV, 75) ; ALA (IV, 89) ; SER (VI, 105) ; PRO (IV, 115) ; VAL (IV, 117) ; THR (IV, 119) ; CYS (II, 121) ; LEU (VI, 131) ; ILE (III, 131) ; ASN (II, 132) ; ASP (II, 133) ; LYS (II, 146) ; GLN (II, 146) ; GLU (II, 147) ; MET (I, 149) ; HIS (II, 155) ; PHE (II, 165) ; ARG (VI, 174) ; TYR (II, 181) ; TRP (I, 204).

Sur cette échelle les neuf premiers acides aminés appartiennent aux classes VI, IV et III, respectivement codées par 6, 4 et 3 triplets à l'exception de CYS codée par 2 ; les onze acides aminés suivants appartiennent aux classes II et I à l'exception de ARG codée par 6.

L'attribution à chaque classe d'un poids moléculaire moyen permet de rejoindre et de préciser les précédentes conclusions.

TABLEAU

Classes.....	I	II	III	IV	VI
P. M. moyen	175	147	131	103	136

Dans chaque classe, à l'exception de la VI, le poids moléculaire moyen est en gros inversement proportionnel au nombre de triplets. La classe VI s'étend entre 105 (SER) et 174 (ARG) en passant par 131 (LEU) et sa moyenne est fortement alourdie à cause de l'Arginine.

Calculant d'abord la corrélation entre le poids moléculaire et le nombre des codons de chacun des vingt acides aminés on trouve un coefficient de corrélation d'une valeur absolue égale à 0,57 et significative au seuil de 99 %. Considérant ensuite les ordres des fréquences des acides aminés dans un échantillon de plus de 4 000 résidus appartenant à une trentaine de protéines on observe une corrélation encore plus forte égale à 0,77. De plus, à l'intérieur de chaque classe il existe une tendance des acides aminés les plus légers à être les plus fréquents. Enfin, à la nette dérive statistique de la classe IV qui, fortement majorée par rapport à la théorie, tend à être aussi abondante dans les protéines que la classe II, s'associe la différence sensible des poids moléculaires des deux classes. De plus, l'opposition des classes II et IV déjà remarquée (²), pour la redondance, les fréquences des acides aminés et le poids moléculaire, est soulignée par les différences des proportions dominantes des bases des codons (II : A + U et IV : G + C).

Après ces premières remarques que nous développerons ultérieurement on est en droit de se préoccuper de la raison pour laquelle on peut observer une telle corrélation d'ensemble entre le poids moléculaire, le nombre des codons et la fréquence des acides aminés dans les protéines.

A l'évidence le poids moléculaire n'est qu'un paramètre grossier définissant seulement une sorte d'encombrement général, aussi ne voit-on pas comment ce facteur purement pondéral pourrait prescrire directement les nombres de codons affectés aux acides aminés. En dépit de la netteté de la corrélation on n'entrevoit pas la possibilité de l'existence d'une relation immédiate de cause à effet entre les deux séries de valeurs, l'explication devant être probablement de nature beaucoup plus complexe.

Sans ignorer la difficulté que posent les différences que l'on peut supposer devoir exister entre les systèmes biochimiques primitifs et actuels, l'examen des mécanismes modernes de synthèse des acides aminés pourrait fournir un modèle de raisonnement. Les synthèses de la Glycine, de la Sérine (famille du Triose) et de l'Alanine (famille du pyruvate) sont sans doute parmi les plus simples ; celles de la Méthionine, de l'Histidine, de la Phénylalanine, de la Tyrosine et du Tryptophane (à travers l'acide shikimique pour les trois derniers) sont parmi les plus complexes (³). On pourrait supposer que les premiers acides aminés synthétisés par les chaînes réactionnelles les plus simples auraient exercé une préemption sur un plus grand nombre de codons ; ceux de poids moléculaires plus élevés, et partant plus complexes, venus plus tard à la suite de la mise en place de nouveaux systèmes biochimiques, n'auraient alors disposé que d'un nombre réduit de codons, à la limite un seul pour la Méthionine et le Tryptophane. Notons que Jukes (⁴) a supposé qu'à l'origine il n'aurait existé qu'une quinzaine d'acides aminés primitifs engendrés par les synthèses abiogéniques. Codés par un code à 16 doublets ils auraient été ultérieurement complétés au cours de l'évolution biochimique par la Méthionine, le Tryptophane et la Tyro-

sine et les deux amides, avec passage au système de code à triplets. Bien que fondée aussi sur la supposition d'une augmentation évolutive du nombre des acides aminés, notre conception est indépendante de celle de Jukes puisque nous considérons le code à triplets comme primitif ⁽⁵⁾. Nous ne présentons ce modèle que comme une simple hypothèse sans prétendre lui attribuer une valeur explicative d'un caractère exclusif.

Sans doute, en effet, à travers l'indice de complexité de synthèse qu'il symbolise, le poids moléculaire n'est-il qu'un des facteurs ayant contribué à fixer les valeurs de redondance et à organiser les classes de triplets ; d'autres facteurs ont pu interférer avec celui de la complexité des synthèses, masquer son rôle et provoquer d'importants remaniements ; notamment l'organisation du code a pu tendre vers une minimisation des erreurs par groupement de triplets codant des acides aminés de propriétés analogues ⁽⁶⁾. Ces réserves étant faites, notre hypothèse sur la signification de la relation signalée entre poids moléculaire et codage offre l'intérêt de ne pas limiter les possibilités de compréhension de la structure du code à ses seules données internes et de lier son évolution à celle de l'ensemble des systèmes biochimiques.

(*) Séance du 24 février 1969.

(1) P. GAVAUDAN et J. BESSON, *Comptes rendus*, 264, Série D, 1967, p. 1919-1922.

(2) P. GAVAUDAN et J. BESSON, *Comptes rendus*, 268, Série D, 1969, p. 173-175.

(3) P. BERNFELD, *Biogenesis of natural compounds*, Pergamon Press, London, New York, 1963, 929 pages.

(4) Th. H. JUKES, *Molecules and evolution*, Columbia Univ. Press, New York and London, 1966, 282 pages, *Loc. cit.*, p. 64-69.

(5) A propos de l'hypothèse du changement de système de codage voir F. H. C. CRICK, *Journ. Mol. biol.*, 38, 1968, p. 367-379.

(6) A. L. GOLDBERG et R. E. WITTES, *Science*, 153, 1966, p. 420-424.

(Faculté des Sciences de Poitiers, Station Biologique de Beau-Site,
25, Faubourg Saint-Cyprien, 86-Poitiers, Vienne ;
Faculté des Sciences de Paris, Département de Mathématiques.)