Thermodynamic Approach to a Possible Theory of the Evolution of a Genetic Code

M. F. Macchiato and A. Tramontano

International Institute of Genetics and Biophysics

Z. Naturforsch. **37 c**, 1031–1037 (1982); received January 26, 1982

Codes, Information Value, Mutations, Efficiency, Evolution

One of the distinctive features of a biological system is its remoteness from equilibrium: for the treatment of the information exchanged by these systems with the external world the concept of information amount is not sufficient because for nonequilibrium systems we need further specifications about information, *i.e.* its value.

We defined the information value for a genetic code. The comparison of the results we had applying such a definition to the three genetic codes found in nature shows that our definition described the system of transmission and reception of information in a biological organism very well and that the information value of a genetic code can be considered as an index of its efficiency, *i.e.* of its ability to minimize the effects of a mutation of the genotype on the phenotype.

Otherwise, our results show that the information value, and hence the efficiency of the three known codes is the same and suggest that the prerequisite of the evolution of the codes is the preservation of this value.

Introduction

A biological system is a system that carries out some functions in an organized way: a function cannot be carried out unless information is present; that is, without systems which create, store, transmit and receive information [1].

The amount of information in a message is usually expressed as the number of bits in the message. The number of bits is defined [2] as $\log_2 K$ where K represents the number of different messages which could be formed with a given coding system; it is known [3] that the number of bits is connected with the entropy of the system, *i.e.* with the measure of the degree of molecular disorder. For example, for a DNA molecule formed by N nucleotides, the possible ways to arrange the four bases on N places is 4^N, hence the amount of information is $\log_2 4^N = 2N$ bits and the entropy of the system is $\ln_e 4^N$ cal/deg.

However the information theory, which only uses the amount of information, and the thermodynamics which uses the concept of entropy, only give an adequate description of equilibrium systems, namely of systems for which one could define state functions [4], but are not adequate for the treatment of nonequilibrium dynamic systems [5]. These are definitions of any biological system that burns free energy continuously to drive chemical-physical reactions which keep the system from fading away into the equilibrium or dead state.

Therefore it is difficult to correlate the amount of information stored, for example, in the DNA chains, with the structure and functions the DNA is coding for: if the information is contained in a message whose reception by a system produces as a consequence events occurring in the system, we say that information posses a value [1].

In biology the value is significant and not the amount of information. For example, as we have seen, the amount of information in a gene is linked only to its length and it is the same in an active gene as in a repressed one: instead the information value, that is the effect of the gene on the biological system, is different in these two cases. At the same time a small amount of information causes great consequences in a trigger nonequilibrium system: this is the case, for example, of insulin. In fact, the interaction of the hormone, whose chain is formed only by 51 amino acids (306 bits) with superficial structures of the cellular membrane, is sufficient to favour the admittance of glucose and some other sugar in the cell, to activate the anabolic processes and to inhibit the metabolic ones in muscolar tissue, in liver and in adipose tissue [6]. Therefore in these cases there is a small amount of information but it possesses a high value.

Please send corrispondence to: Dr. M. F. Macchiato, International Institute of Genetics and Biophysics, Via Marconi 10, 80125 Naples, Italy.

0341-0382/82/1000-1031 $ 01.30/0

Table I. The value of information for a genetic code in the hypothesis of equiprobable mutations*.

I \ II	A			C			G			U			III
A	Lys	1.50	1.5	Thr	0.44	1.6	Arg	0.73	2.0	Met	1.30	2.5	A
	Asn	− 0.01	3.8	Thr	0.44	2.1	Ser	0.04	2.0	Ile	2.97	4.7	C
	Lys	1.50	3.0	Thr	0.44	1.7	Arg	0.73	2.0	Met	1.30	2.5	G
	Asn	− 0.01	3.8	Thr	0.44	2.1	Ser	0.04	2.0	Ile	2.97	4.7	U
C	Gln	− 0.10	4.1	Pro	2.60	3.8	Arg	0.73	2.4	Leu	2.42	2.1	A
	His	1.40	3.2	Pro	2.60	3.4	Arg	0.73	2.0	Leu	2.42	1.5	C
	Gln	− 0.10	4.1	Pro	2.60	3.8	Arg	0.73	2.5	Leu	2.42	2.1	G
	His	1.40	1.4	Pro	2.60	1.5	Arg	0.73	2.4	Leu	2.42	1.9	U
G	Glu	0.55	1.4	Ala	0.63	1.5	Gly	0.00	2.4	Val	1.69	1.9	A
	Asp	0.54	1.5	Ala	0.63	1.5	Gly	0.00	1.4	Val	1.69	2.3	C
	Glu	0.55	1.4	Ala	0.63	1.5	Gly	0.00	2.5	Val	1.69	1.9	G
	Asp	0.54	1.5	Ala	0.63	1.5	Gly	0.00	1.4	Val	1.69	2.3	U
U		–		Ser	0.04	3.9	Trp	3.00	5.3	Leu	2.42	2.2	A
	Tyr	2.87	3.7	Ser	0.04	3.1	Cys	0.65	3.6	Phe	2.65	2.2	C
		–		Ser	0.04	2.7	Trp	3.00	5.3	Leu	2.42	2.2	G
	Tyr	2.87	3.7	Ser	0.04	3.1	Cys	0.65	3.6	Phe	2.65	2.2	U
A						$L_{hm} = 2.56$							

I \ II	A			C			G			U			III
A	Lys	1.50	3.4	Thr	0.44	2.2	Arg	0.73	2.6	Ile	2.97	3.4	A
	Asn	− 0.01	3.8	Thr	0.44	2.1	Ser	0.04	2.0	Ile	2.97	4.1	C
	Lys	1.50	3.0	Thr	0.44	1.7	Arg	0.73	2.0	Met	1.30	3.1	G
	Asn	− 0.01	3.8	Thr	0.44	2.1	Ser	0.04	2.0	Ile	2.97	4.1	U
C	Gln	− 0.10	4.1	Pro	2.60	3.8	Arg	0.73	2.4	Leu	2.42	1.9	A
	His	1.40	3.2	Pro	2.60	3.4	Arg	0.73	2.0	Leu	2.42	1.5	C
	Gln	− 0.10	4.1	Pro	2.60	3.8	Arg	0.73	2.5	Leu	2.42	2.1	G
	His	1.40	3.2	Pro	2.60	3.4	Arg	0.73	2.0	Leu	2.42	1.5	U
G	Glu	0.55	1.4	Ala	0.63	1.5	Gly	0.00	2.4	Val	1.69	2.2	A
	Asp	0.54	1.5	Ala	0.63	1.5	Gly	0.00	1.4	Val	1.69	2.3	C
	Glu	0.55	1.4	Ala	0.63	1.5	Gly	0.00	2.5	Val	1.69	1.9	G
	Asp	0.54	1.5	Ala	0.63	1.5	Gly	0.00	1.4	Val	1.69	2.3	U
U		–		Ser	0.04	3.9	Trp	3.00	5.3	Leu	2.42	2.1	A
	Tyr	2.87	3.7	Ser	0.04	3.1	Cys	0.65	3.6	Phe	2.65	2.2	C
		–		Ser	0.04	2.7	Trp	3.00	5.3	Leu	2.42	2.2	G
	Tyr	2.87	3.7	Ser	0.04	3.1	Cys	0.65	3.6	Phe	2.65	2.2	U
B						$L_{ym} = 2.66$							

As we can see, it is very important to define quantitatively the information value, but, since this problem is connected with the thermodynamics of nonequilibrium systems (for a review see ref. [5]) it is still open.

However a definition of the information value must be correlated to the results of the reception of the message in the system and to estimate it we can value the consequences that a mutation in the message produces in the receptor system.

At present definitions of the information value can be formulated only in relatively simple situations, for example regarding replication [7] and translation [8]. All these attempts however assumed the existence of an universal code, which seemed obvious until a few years ago. Recently there has been instead evidence that human and yeast mitochondria have different genetic codes [9, 10].

The decoding of a message by means of a code is the first step in the transmission of the message

Table Ic (continued)

I \ II	A			C			G			U			III
A	Lys	1.50	3.4	Thr	0.44	2.2	Arg	0.73	1.9	Ile	2.97	2.4	A
	Asn	− 0.01	3.8	Thr	0.44	2.1	Ser	0.04	2.0	Ile	2.97	4.1	C
	Lys	1.50	3.0	Thr	0.44	1.7	Arg	0.73	2.0	Met	1.30	3.1	G
	Asn	− 0.01	3.8	Thr	0.44	2.1	Ser	0.04	2.0	Ile	2.97	4.1	U
C	Gln	− 0.10	4.1	Pro	2.60	3.8	Arg	0.73	1.8	Leu	2.42	1.9	A
	His	1.40	3.2	Pro	2.60	3.4	Arg	0.73	2.0	Leu	2.42	1.5	C
	Gln	− 0.10	4.1	Pro	2.60	3.8	Arg	0.73	2.5	Leu	2.42	2.1	G
	His	1.40	3.2	Pro	2.60	3.4	Arg	0.73	2.0	Leu	2.42	1.5	U
G	Glu	0.55	1.4	Ala	0.63	1.5	Gly	0.00	1.7	Val	1.69	2.2	A
	Asp	0.54	1.5	Ala	0.63	1.5	Gly	0.00	1.4	Val	1.69	2.3	C
	Glu	0.55	1.4	Ala	0.63	1.5	Gly	0.00	2.5	Val	1.69	1.9	G
	Asp	0.54	1.5	Ala	0.63	1.5	Gly	0.00	1.4	Val	1.69	2.3	U
U		–		Ser	0.04	2.6		–		Leu	2.42	2.8	A
	Tyr	2.87	3.7	Ser	0.04	3.1	Cys	0.65	3.2	Phe	2.65	2.2	C
		–		Ser	0.04	2.7	Trp	3.00	6.8	Leu	2.42	2.2	G
	Tyr	2.87	3.7	Ser	0.04	3.1	Cys	0.65	3.2	Phe	2.65	2.2	U
C						$L_n = 2.59$							

* a) mitochondrial human code (HM); b) mitochondrial yeast code (YM); c) code of the other organisms (N). The first number after the name of the residue is the hydrophobicity in Kcal/mol measured as the value of the change of the free energy per side group of the free aminoacid during its transition from ethanol into aqueous solution [11].

The second number is the information value of the codon, defined as:

$$V_K = \sum_{1j}^{3} \left(\left(\sum_{1}^{n_j} d_{ji} \right) n_j \right)$$

where n_j is the number of possible mutations, except the nonsense ones, of the base in the j-th position of the k-th codon and d_{ji} is the difference of hydrophobicity between the residue coded by the k-th codon and that arising when the base in the j-th position of the same codon changes by means of the i-th mutation. L is the information value of the code defined as:

$$L = \sum_{1K}^{M} V_k$$

where $M = 62$ for HM and YM codes and $M = 61$ for N code.

from DNA to the rest of the biological system, so the existence of different codes poses the problem of finding out how the transmission-reception system of the message is modified changing the code used. In other words one can ask which are the consequences of a variation in the code on the utilization of the information stored in DNA. The answer, clearly, is not related to the amount of information, but to the information value that we have defined for a genetic code.

In this paper we have defined the information value for each of the possible ways in which the message can vary and we have compared the values we obtained for the three codes examined.

Results

As we have seen, a definition of the information value must be related to the consequences of the variations of the message on the receptor system. In

Table II. Frequencies of mutation. A mutation can be only due to a transversion or to a transition. A transversion is the replacement of a purine (A or G) by a pyrimidine (C or U) and *vice versa*. A transition is the replacement of a purine by another purine or of a pyrimidine by another pyrimidine. The sample of population examined to measure the frequency of each mutation is not infinite, so the values are affected by an error, due to the limitation of the sample, which is about 1%. All the values obtained from these frequencies with algebraic algorithms, hence, are affected by the same error in per cent.

Mutation		Frequency ($\cdot 10^{-2}$)
Transitions	A → G	9
	G → A	34
	C → U	3
	U → C	3
Transversions	A → C	5
	C → A	13
	A → U	3
	U → A	2
	G → C	10
	C → G	11
	G → U	3
	U → G	3

Table III. The value of information for a genetic code in the hypothesis of non equiprobable mutations*.

I \ II	A			C			G			U			III
A	Lys	1.50	3.8	Thr	0.44	1.7	Arg	0.73	2.1	Met	1.30	2.3	A
	Asn	− 0.01	3.4	Thr	0.44	1.2	Ser	0.04	1.0	Ile	2.97	5.3	C
	Lys	1.50	2.3	Thr	0.44	1.3	Arg	0.73	1.7	Met	1.30	1.9	G
	Asn	− 0.01	2.8	Thr	0.44	1.2	Ser	0.04	1.1	Ile	2.97	4.8	U
C	Gln	− 0.10	3.5	Pro	2.60	4.2	Arg	0.73	1.8	Leu	2.42	1.9	A
	His	1.40	3.4	Pro	2.60	3.5	Arg	0.73	1.6	Leu	2.42	1.5	C
	Gln	− 0.10	2.0	Pro	2.60	4.2	Arg	0.73	1.6	Leu	2.42	2.2	G
	His	1.10	3.0	Pro	2.60	3.5	Arg	0.73	1.6	Leu	2.42	1.5	U
G	Glu	0.55	1.4	Ala	0.63	1.0	Gly	0.00	1.7	Val	1.69	2.1	A
	Asp	0.54	1.2	Ala	0.63	1.0	Gly	0.00	0.8	Val	1.69	2.4	C
	Glu	0.55	1.4	Ala	0.63	1.0	Gly	0.00	1.5	Val	1.69	1.8	G
	Asp	0.54	1.2	Ala	0.63	1.0	Gly	0.00	0.8	Val	1.69	2.4	U
U		–		Ser	0.04	4.3	Trp	3.00	4.1	Leu	2.42	2.3	A
	Tyr	2.87	4.1	Ser	0.04	3.2	Cys	0.65	2.2	Phe	2.65	2.5	C
		–9		Ser	0.04	4.1	Trp	3.00	4.1	Leu	2.42	1.7	G
	Tyr	2.87	4.1	Ser	0.04	3.2	Cys	0.65	3.5	Phe	2.65	2.5	U
A						$L'_{hm} = 2.376$							

I \ II	A			C			G			U			III
A	Lys	1.50	2.9	Thr	0.44	1.3	Arg	0.73	1.7	Ile	2.97	3.9	A
	Asn	− 0.01	3.4	Thr	0.44	1.2	Ser	0.04	1.0	Ile	2.97	4.3	C
	Lys	1.50	2.3	Thr	0.44	1.3	Arg	0.73	1.7	Met	1.30	3.0	G
	Asn	− 0.01	2.8	Thr	0.44	1.2	Ser	0.04	1.1	Ile	2.97	4.4	U
C	Gln	− 0.10	3.5	Pro	2.60	4.2	Arg	0.73	1.8	Lei	2.42	1.9	A
	His	1.40	3.4	Pro	2.60	3.5	Arg	0.73	1.6	Leu	2.42	1.5	C
	Gln	− 0.10	2.0	Pro	2.60	4.2	Arg	0.73	1.6	Leu	2.42	2.2	G
	His	1.40	3.0	Pro	2.60	3.5	Arg	0.73	1.6	Leu	2.42	1.5	U
G	Glu	0.55	1.4	Ala	0.63	1.0	Gly	0.00	1.7	Val	1.69	2.4	A
	Asp	0.54	1.2	Ala	0.63	1.0	Gly	0.00	0.8	Val	1.69	2.4	C
	Glu	0.55	1.4	Ala	0.63	1.0	Gly	0.00	1.5	Val	1.69	1.8	G
	Asp	0.54	1.2	Ala	0.63	1.0	Gly	0.00	0.8	Val	1.69	2.4	U
U		–		Ser	0.04	4.3	Trp	3.00	4.1	Leu	2.42	2.2	A
	Tyr	2.87	4.1	Ser	0.04	3.2	Cys	0.65	2.2	Phe	2.65	2.5	C
		–		Ser	0.04	4.1	Trp	3.00	4.1	Leu	2.42	1.7	G
	Tyr	2.87	4.1	Ser	0.04	3.2	Cys	0.65	3.5	Phe	2.65	2.5	U
B						$L'_{ym} = 2.376$							

our case the variation of the message is the replacement of a codon with a different one, coding for a different phenotype. Obviously we can have two kinds of substitution of a codon: in the first a triplet coding for an amino acid can be modified by means of a mutation, but it continues to code for an amino acid, in the second, a translated codon can be replaced by a codon which produces termination in the protein chain, *i.e.* by a terminal codon. A code, hence, will have an information value for each of these possible ways in which the message can change.

At first we defined the information value of the code assuming all the mutations are equiprobable, *i.e.* an information value related only to the intrinsic structure of the code, independent, hence, of the mechanism of variation of the message.

Let us assume that the danger of a mutation causing the replacement of an amino acid residue with a different one for the biological function of the

Table III c (continued)

I \ II	A			C			G			U			III
A	Lys	1.50	2.9	Thr	0.44	1.3	Arg	0.73	1.5	Ile	2.97	3.9	A
	Asn	− 0.01	3.4	Thr	0.44	1.2	Ser	0.04	1.0	Ile	2.97	4.3	C
	Lys	1.50	2.3	Thr	0.44	1.3	Arg	0.73	1.7	Met	1.30	3.0	G
	Asn	− 0.01	2.8	Thr	0.44	1.2	Ser	0.04	1.1	Ile	2.97	4.4	U
C	Gln	− 0.10	3.5	Pro	2.60	4.2	Arg	0.73	1.4	Leu	2.42	1.9	A
	His	1.40	3.4	Pro	2.60	3.5	Arg	0.73	1.6	Leu	2.42	1.5	C
	Gln	− 0.10	2.0	Pro	2.60	4.2	Arg	0.73	1.6	Leu	2.42	2.2	G
	His	1.40	3.0	Pro	2.60	3.5	Arg	0.73	1.6	Leu	2.42	1.5	U
G	Glu	0.55	1.4	Ala	0.63	1.0	Gly	0.00	1.4	Val	1.69	2.4	A
	Asp	0.54	1.2	Ala	0.63	1.0	Gly	0.00	0.8	Val	1.69	2.4	C
	Glu	0.55	1.4	Ala	0.63	1.0	Gly	0.00	1.5	Val	1.69	1.8	G
	Asp	0.54	1.2	Ala	0.63	1.0	Gly	0.00	0.8	Val	1.69	2.4	U
U		–		Ser	0.04	3.0		–		Leu	2.42	2.9	A
	Tyr	2.87	4.1	Ser	0.04	3.2	Cys	0.65	4.1	Phe	2.65	2.5	C
		–		Ser	0.04	4.1	Trp	3.00	5.6	Leu	2.42	1.7	G
	Tyr	2.87	4.1	Ser	0.04	3.2	Cys	0.65	3.5	Phe	2.65	2.5	U
C							$L'_N = 2.379$						

* a) mitochondrial human code (HM); b) mitochondrial yeast code (YM); c) code of the other organisms (N). The first number after the name of the residue is the hydrophobicity in Kcal/mol (see legend of Table I). The second number after the name is the information value of the codon defined as:

$$V'_K = \sum_{1j}^{3} \left(\left(\sum_{1i}^{n_j} p_i \, d_{ji} \right) \Big\backslash n_j \right)$$

where n_j is the number of the possible mutations, except the nonsense ones, of the base in the j-th position of the k-th codon, d_{ji} is the difference of hydrophobicity between the residue coded by the k-th codon and that arising when the base in the j-th position of the same codon changes by means of the i-th mutation. L' is the information value of the code defined as:

$$L' = \sum_{1K}^{M} V'_k$$

where $M = 62$ for HM and YM codes and $M = 61$ for N code.

protein is higher, the greater the difference between the hydrophobicity of the residue and that of the residue obtained as a result of the mutational replacement [1]. We then defined the information value of a codon as the sum of the average differences in hydrophobicities arising in every single replacement of the first, the second and the third base of the codon itself. We did not take into account more bases than one substitution because their probability is negligible. The information value of a code (L) has been defined as the average of the information values of the codons. The results achieved are reported in Table I.

We then defined an information value of a code taking into account the way in which the variation of the message has happened. The analysis of the frequence of the mutations observed until now [12], in fact, demonstrates that mutations are not equiprobable. To account for the real consequences on the organism of a variation of the message, we assigned to each difference of hydrophobicity the respective statistical weight, assuming that the probability of a mutation is equal to the frequency with which we have observed it until now (Table II).

The results are reported in Table III and they show that the weighted information values (L') are the same for the three codes.

Now, regarding the intrinsic structure of the codes, the probability that a mutation changes the physical-chemical characteristics of the aminoacid residue is different for the three codes and, particularly, it is higher in yeast mitochondria (see in Table I the values of L); the equality of the weighted information values (see in Table III the values of L') for the three codes demonstrates, however, that the real probability of given difference of hydrophobicity between the wild type protein and the protein coded by the gene after a mutation is constant and that it is less than the difference of hydrophobicity that there would be if mutations were random and

equiprobable, the value of L (Table I) being less than the value of L' (Table III). This means that higher differences of hydrophobicity are, on the average, due to less probable mutations and *vice versa* so that the dangerous effect of a mutation are kept to minimum.

Regarding the codons whose mutations can transform them into terminal codons (nonsense mutations), we must, once again, distinguish between the intrinsic structure of the code and the utilization of its structure by the organism, owing to different probabilities of each mutation. In fact, if all the mutations are assumed to be equiprobable, as the three terminal codons of N code can arise from 23 single substitutions and the two terminal codons of mitochondrial codes from 15 single substitutions, the intrinsic structure of the codes is done in such a way as to make a nonsense mutation more probable in N code than in HM and YM codes. If, instead, we use the probabilities of mutation given in Table II we get different results.

Let us normalize to unity the probability of a mutation. A mutation can be only due to a transition or to a transversion (Table II): the probability of a nonsense mutation is, hence, a compound one and its value is the product of the value w of the probability of a transversion with the relative probability P that a nonsense mutation is due to a transversion added to the product of the probability P' that a nonsense mutation is due to a transition with the probability $(1 - w)$ of a transition.

Now we must calculate the values of P, P' and w for the three codes.

The value of P, *i.e.* the relative probability of a nonsense mutation caused by a transversion is the ratio between the number of favourable cases, that is the number of nonsense mutations due to a transversion, and the number of possible cases, that is, the total number of nonsense mutations either transitions or transversions. The value of P' can be calculated in the same way. Regarding the value of w, *i.e.* the probability of a transversion, it must be equal to the ratio between the number of observed transversions and the number of possible transversions one could observe with the examined code.

Hence, clearly, the product between w and the number of possible and observable transversions is equal to the number of transversions observed until now. Likewise the product between the probability $(1 - w)$ of a transition and the number of possible and observable transitions must be equal to the number of observed transitions.

Using the probability of mutations reported in Table II we can calculate the ratio q between the number of observed transversions and the number of observed transitions by dividing the sum of the observed frequencies of transversions and the sum of the observed frequencies of transitions. In such a way we have:

$$\frac{w \cdot N^\circ \text{ of observable transversions}}{(1 - w) \cdot N^\circ \text{ of observable transitions}} = q$$

from which we can obtain w.

Hence we defined the information value of a code for nonsense mutations as: $L'' = P \cdot w + P' \cdot (1 - w)$.

The observed values are reported in Table IV.

As one can see, although the three codes are structurally different the information values L'' related to the probability of a nonsense mutation are equal within the experimental errors due to the sampling of the frequencies (see legend of Table II).

We can conclude, hence, that each of the codes is such that the characteristics of the phenotype change in the same way when any mutation changes the genotype (equality of the values of L' in Table III and of the values of L'' in Table IV).

Furthermore it is important to underline that the average differences in hydrophobicity of the coded

Table IV. The information value for a genetic code for nonsense mutations taking into account the different probability of each mutation.

Code	N	HM	YM
N_t = Total number of nonsense mutations	23	15	15
N_1 = Total number of nonsense mutations due to a transition	5	3	3
N_2 = Total number of nonsense mutations due to a transversion	18	12	12
P = Probability to have a nonsense mutation for transition (N_1/N_t)	0.78	0.80	0.80
P' = Probability to have a nonsense mutation for transversion (N_2/N_t)	0.22	0.20	0.20
N_3 = Number of possible and observable transitions	276	288	288
N_4 = Number of possible and observable transversions	116	118	118
q = Rate between the observed frequencies of transversions and transitions	1	1	1
w = Probability of a transversion $(q/(q + N_3/N_4))$	0.30	0.29	0.29
$L'' = P \cdot w + P' \cdot (1 - w)$	0.38	0.37	0.37

proteins due to mutational replacements are less than the differences arising if one supposes that all the mutations are equiprobable, as it is evident from the comparison between the value of L in Table I and the value of L' in Table III.

Conclusions

A biological systems is always far from equilibrium states. So, the description of the mechanism of transmission and reception of information in a biological system needs the introduction of the concept of information value. In other words it is necessary to define a parameter taking into account the effects of a variation of the genotype on the receptor system.

We defined the information value for a genetic code in such a way as to take into account the possible changes of the message, *i.e.* of the information stored in DNA chains.

The information value has been calculated either supposing equiprobable mutations, *i.e.* regarding the intrinsic structure of the codes, or assigning to each mutation its own probability, that is taking into account the utilization of the codes by the receptor system.

In applying these definitons to the three different codes we can draw some conclusions about the characteristics of a genetic code. In fact, in spite of the structural differences that ought to make the consequences of a mutation different in the systems utilizing different codes, the information value is constant for each code and it is significantly less than one would expect if the mutations were equiprobable: *i.e.* each genetic code provides a decreased probability of dangerous mutations whereas higher probability mutations cause smaller variations in the phenotype and this peculiarity cannot be statistically random.

So we can conclude that the information value of a code is an index of its efficiency concerning the utilization of the information stored in DNA. Restrictions imposed to a code by the conservation of this efficiency could contribute to explain the small number of different codes one finds in nature because they limit the possible random changes on could fix in a code.

At least one can ask oneself if the three known codes are the most efficient. Clearly this is an unanswerable question until it is a qualitative one: our quantitative defintion of efficiency, *i.e.* of information value of a code, instead, allows us to formulate the question in a resolvable form. In fact, one could calculate, by means of a computer, the information values of all the possible codes with four bases and twenty amino acids and could analyze the distribution of these values.

From the comparison of this distribution and the observed information values of the three examined codes, one could obtain some information about the characteristics of a code really suitable to a biological organism. It is possible, for example, that the observed information values of the three known codes are as low as possible, *i.e.* that they are the most efficient codes, or that they are the most probable one, *i.e.* that they had been casually selected.

Such a work is, obviously, very long and laborious, but we think that our results are already sufficiently indicative and cogent to encourage us to embark on it.

[1] M. V. Volkenstein, Foundations of Physics, **7,** 97 (1977).
[2] M. Apter, Cybernetics and Development, Pergamon Press, New York 1966.
[3] L. Brillouin, Science and Information Theory, Academic Press, New York 1962.
[4] C. E. Reid, Principles of Chemical Thermodynamics, Reynold Publishing Corporation, New York 1960.
[5] M. Eigen, Naturwissenschaften **58,** 465 (1971).
[6] P. Cuatercasas, Proc. Nat. Acad. Sci. **63,** 450 (1969).
[7] I. Prigogine and J. Nicolis, Quart. Rev. Biophys. **4,** 149 (1971).
[8] M. Volkenstein, Molecular Biophysics, Academic Press, New York 1976.
[9] B. G. Barrel, A. T. Bankier, and J. Drouin, Nature **282,** 189 (1979).
[10] T. D. Fox, Proc. Nat. Acad. Sci. USA **76,** 6534 (1979).
[11] J. Tanford, J. Am. Chem. Soc. **84,** 4240 (1962).
[12] I. H. I. C. Comprehensive Sickle Cell Center, USA 1979.
[13] S. R. de Groot, Thermodynamics of Irreversible Processes, North Holland Publishing Corporation, Amsterdam 1962.
[14] J. L. Dobb, Stochastic Processes, J. Wiley, New York 1953.
[15] M. Dutta, Physics Today **21,** 75 (1968).
[16] M. Eigen and D. Porschke, J. Mol. Biol. **53,** 123 (1970).
[17] P. Glansdorff and I. Prigogine, Thermodynamic Theory of Structure, Stability and Fluctuations, Wiley Interscience, New York 1971.
[18] J. M. Klein, Phys. Rev. **103,** 17 (1956).
[19] J. Meixner, Ann. Phys. **43,** 470 (1943).