Letters to the Editor

Comments on
Counter-Examples to a Neutralist Hypothesis

by Lila L. Gatlin

J. Mol. Evol. 7, 185–195 (1976)

THOMAS H. JUKES

University of California, Space Sciences Laboratory,
Berkeley, Ca. 94720, USA

Gatlin (1976) has theoretically derived specific counter-examples to "the hypothesis that a random amino acid composition signifies a random evolutionary process". The "hypothesis" that she cites refers first to a publication by King & Jukes (1969) in which expected frequencies of amino acids, in terms of codon frequencies, were plotted against observed frequencies in 53 completely sequenced vertebrate proteins, and "fairly good" agreement was obtained *except for arginine.* Second, Gatlin refers to an article by Kimura & Ohta (1971) in which expected frequencies were replaced by "equilibrium frequencies" and "even arginine was found on the line" (Gatlin, 1976). During the period between 1969 and the date of submission of the manuscript by Gatlin, we published additional comments on selection against arginine, and on selection against the proportions of amino acids found in the genetic code (Jukes, 1973; Jukes et al., 1975). The latter was sent to Gatlin prior to publication. These findings led us to conclude that the amino acid composition of proteins was governed in part by selection for or against certain amino acids (e.g. for lysine, aspartic acid and glutamic acid, and against arginine), and in part by neutral or near-neutral changes. Our findings on the distribution of amino acids are compared with codon frequencies in Table 1. This also shows the results of two other summaries, which do not differentiate between aspartic and glutamic acid and their amides, and so are not useful for a complete evaluation of amino acid frequencies. However, all three summaries show with remarkable consistency that lysine occurs at twice the level expected, and arginine at less than half of the expected level. The summary by Smith (1966) was used by Kimura & Ohta (1971) to arrive at a conclusion that

Gatlin interprets as showing that arginine was "on the line".
This conclusion is misleading. Ohta & Kimura (1971), using
the same data as in Kimura & Ohta (1971), discussed "the sig-
nificantly lower frequency of Arg than is expected from random
base arrangement" and said "The biochemical explanation for
such selective constraint is not known, but it is conceivable
that, because Arg is unusually large and contains three amine
groups, its insertion might disturb normal configuration of
proteins".

Gatlin's demonstration is that a random amino acid distri-
bution in a protein could be produced from a highly non-random
DNA sequence. Presumably she is referring to a "protein" in
which the percentage of each amino acid is identical with its
percentage representation in the genetic code, but in which
the distribution of the amino acids is randomized. Her proposal
is self-fulfilling for the following reason: The same "protein"
could be coded by messenger RNA in which all the third codon
positions were C or G, or one in which all the third codon
positions (except those for methionine and tryptophan) were A

Table 1
Average amino acid distribution in certain representative proteins per 61
residues as compared with the genetic code

	Codons	(a)	(b)	(c)	
Ala	4	5.3	5.2	5.3	
Arg	6	2.6	2.7	2.8	
Asn	2	3.0			
Asp	2	3.6			
Asx	(4)	(6.6)	(6.5)	(6.3)	
Cys	2	1.3	1.4	0.7*	
Gln	2	2.4			
Glu	2	3.3			
Glx	(4)	(5.7)	(6.5)	(6.6)	
Gly	4	4.8	4.9	4.7	
His	2	1.4	1.3	1.2	
Ile	3	3.1	3.0	2.6	
Leu	6	4.7	3.9	5.1	
Lys	2	4.1	3.9	4.1	(a) Jukes et al. (1975)
Met	1	1.1	1.1	0.9	
Phe	2	2.3	2.3	2.4	(b) Reeck & Fisher (1973),
Pro	4	2.5	2.9	3.0	cited in (a)
Ser	6	4.5	3.8	4.7	(c) M.H. Smith (1966)
Thr	4	3.7	3.5	3.6	
Trp	1	0.8	0.8	0.5	
Tyr	2	2.3	2.0	2.2	* Smith apparently excludes
Val	4	4.2	4.1	4.2	cysteine from some Cys values (

or U. Such a "protein" would have the composition Ala_4 Arg_6 Asn_2 Asp_2 Cys_2 Gln_2 Glu_2 Gly_4 His_2 Ile_3 Leu_6 Lys_2 Met_1 Phe_2 Pro_4 Ser_6 Thr_4 Trp_1 Tyr_2 Val_4. The base composition of its mRNA could be A_{201} G_{177} C_{176} U_{205} or A_{175} G_{203} C_{208} U_{173}. Both are obviously non-random, and obviously, numerous other base compositions intermediate between these two extremes are also possible.

Gatlin states that the "original concept of King and Jukes... (was) that a random amino acid distribution constitutes evidence for the random nature of base changes in DNA". Actually, King and Jukes concluded that the distribution of amino acids in vertebrate proteins, with the exception of arginine, agreed fairly well with the distribution of codons among the twenty amino acids. King & Jukes (1969) also pointed out that the "third base" positions of codons could become "filled up" (by non-random base changes) leading to C and G replacing A and U, as implied by the pioneer work of Sueoka (1961) in microorganisms, and as actually shown in studies with the Treffers mutator gene in *E.coli* by Cox & Yanofsky (1967). King and Jukes did not discuss divergence from independence (D_2). This is derived from the sequence of bases in DNA, not from amino acid composition. The concept attributed to King and Jukes by Gatlin did not exist.

REFERENCES

Cox, E.C., Yanofsky, C. (1967). Proc.Natl.Acad.Sci.USA 58, 1895
Gatlin, L.L. (1967). J.Mol.Evol. 7, 185
Jukes, T.H. (1973). Biochem.Biophys.Res.Commun. 53, 709
Jukes, T.H., Holmquist, R., Moise, H. (1975). Science 189, 50
Kimura, M., Ohta, T. (1971). In: Proc.6th Berkeley Symp.on Math.Stat.and
   Prob., L.M. LeCam, J. Neyman, E.L. Scott, eds. Berkeley: U.C. Press
King, J.L., Jukes, T.H. (1969). Science 164, 788
Ohta, T., Kimura, M. (1971). Science 174, 150
Smith, M.H. (1966). J.Theoret.Biol. 13, 261
Sueoka, N. (1971). Proc.Natl.Acad.Sci.USA 47, 1141