9. F. R. Hainsworth, *J. Comp. Physiol.* **88**, 425 (1974).
10. R. C. L. Perkins, *Fauna Hawaiiensis (Introduction)* (Cambridge Univ. Press, Cambridge, 1913).
11. F. L. Carpenter and R. E. MacMillen, *IBP Island Ecosystems IRP Tech. Rep. No. 61* (1975).
12. ———, *ibid., No. 63* (1975).
13. ———, *ibid., No. 33* (1973).
14. F. L. Carpenter, *Ecology*, in press.
15. R. E. MacMillen and F. L. Carpenter, in preparation.
16. R. C. L. Perkins, *Fauna Hawaiiensis (Vertebrata)* (Cambridge Univ. Press, Cambridge, 1903), vol. 1, pp. 365–466; D. Amadon, *Evolution* **1**, 63 (1947); H. T. Spieth, *Am. Nat.* **100**, 470 (1966).
17. We thank P. Johns and D. Wyatt for field assistance; E. Pacheco for providing access to one of our study areas; and P. Ewald, P. Kareiva, and L. Wolf for their discussions. This is Contribution 53, Island Ecosystems IRP/US IBP Hawaii, supported by NSF grant GB-23230.

27 May 1976; revised 27 July 1976

References

1. T. H. Jukes, R. Holmquist, H. Moise, *Science* **189**, 50 (1975).
2. M. O. Dayhoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972).
3. J. L. King and T. H. Jukes, *Science* **164**, 788 (1969).
4. C. Gautier, *Ann. Biol.*, in press.
5. R. Grantham, *science* **185**, 862 (1974).
6. A. D. McLachlan, *J. Mol. Biol.* **64**, 417 (1972).

26 September 1975; revised 3 February 1976

Average Proteins and the Genetic Code

Jukes, Holmquist, and Moise (*1*) create a new "average protein" which is different from that computed by Dayhoff (*2*), especially in proline and cysteine frequencies. Their sample description does not permit us to understand the reason for the differences. I suspect the sample change explains the contradictory conclusions found by King and Jukes in 1969 (*3*). They then showed serine, leucine, histidine, cysteine, and proline with "near expected frequencies," and this helped them to deduce that proteins reflect the structure of the genetic code. This time the same five amino acids are "significantly lower" than code levels (*1*). Furthermore, no proof is given for their statement that "Clearly, natural selection counteracts the genetic code to neutralize the charge on proteins," although on this and other points allusion is made to a "statistical analysis" and "significant" differences. Jukes *et al.* propose that "evolutionary selection maintains charge neutrality"; this predicts correlation between the amounts of basic and acidic amino acids to maintain neutrality in proteins. I have computed the linear correlation coefficient between frequencies of the two kinds of amino acids in each protein, taking as sample 53 fully determined proteins of length greater than 50 residues (one protein per family) from *Atlas of Protein Sequence and Structure* (*2*). The result, $r = -.07$, disagrees with the proposition of Jukes *et al.*

The biological interpretation given to their "statistical analysis" is questionable. Jukes *et al.* apparently explain all differences between observed and code frequencies for amino acids by selection of side chain properties. They ignore at least two biological possibilities: first, the probability of mutation may not be the same for all codons. Second, there may be selection at another level; for example, on the secondary structure of the messenger RNA, or on differences between the efficiency of the 61 codon-tRNA-enzyme systems. Such differences could be due to the cell concentration of tRNA (transfer RNA) and enzymes and the translation rate of each codon.

That Jukes *et al.* assign an evolutionary role to *some* of the amino acids is unwarranted. In a study of mutations among amino acids, I have found more replacements of lysine to arginine than the converse (*4*), which does not support the "evolutionary intruder" hypothesis (*1*). Also, the idea that arginine entered the code because it had more affinity for ornithine's tRNA than did ornithine itself seems physicochemically unmotivated. And, if alanine's relatively high frequency is due to the fact that its small side chain makes it useful as a "filler" in proteins, why does glycine, whose side chain has a molecular volume only about one-tenth that of alanine (*5*), have a lower frequency than alanine?

Finally, the model proposed for combining "selective and neutral mutations to give a picture of dynamic equilibrium in protein evolution" is obscure. The model becomes clearer by supposing the authors mean, as suggested by their valine-isoleucine example, that the amino acids in the expected range "pool" replace each other without any action of natural selection. But then the differences between amino acid substitution rates (*2, 4, 6*) will reject the hypothesis. Of course, the fact that frequencies of valine and isoleucine are approximately the same in the average protein and in the code is in no way a demonstration that mutations between them should be "neutral."

More data are needed for discriminating between all the influences leading to the present amino acid composition of proteins, particularly between the respective influences of the translation apparatus characteristics and of natural selection.

C. GAUTIER

Laboratoire de Biométrie, Université Claude Bernard, Lyon 1, 69621 Villeurbanne, France

Published compilations of amino acid composition are summarized in Table 1. The values found by Reeck and Fisher (*1*), Smith (*2*), and also Dayhoff *et al.* (*3*), substantiate our own findings. Reeck and Fisher and Smith, as well as we (*4*), find lower values for Cys (*5*) and Pro than does Dayhoff. Gautier's statement that we "create a new 'average protein' " is without basis. We did not accept Dayhoff's compilation in its totality because it included short peptides, as well as analyses that were incomplete with respect to Asn, Asp, Gln, and Glu.

King and Jukes (*6*) used the genetic code to calculate expected amino acid frequencies, assuming that each of the four nucleotides was used in proportion to its frequency as inferred from protein data. Jukes *et al.* (*4*) compared observed values with the amino acid frequencies in the genetic code table. The observed values of Ser, Leu, His, Cys, and Pro (Table 1) are lower than the expected values in column (a) of Table 1. When expected amino acid composition is calculated from the nucleotide base ratios inferred from the data in Table 1, the finding of King and Jukes (*6*) is confirmed.

Our statistical analysis is published (*7*) as was an examination of the basic and acidic amino acids in proteins (*8*). Electrical charge neutrality in proteins refers to the fact that some proteins are acidic, others basic, others neutral, and in such proportions as to maintain an organismal *p*H near 7. Correlations between acidic and basic amino acids in *individual* proteins are thus not necessarily expected. This is in accord with Gautier's low correlation coefficient.

Side-chain properties have a profound effect on selection (for example, sickle-cell hemoglobin). No evidence supports Gautier's unreferenced speculation that different codons fix mutations at different rates. The third codon position in partial sequences of hemoglobin messenger RNA (mRNA) sustains five times more fixations than the first two positions (*9*). This speaks against strong selection at the mRNA level. All amino acids have evolutionary roles.

Of the possible 380 directed amino

acid substitutions, Gautier (*10*) infers by various assumptions only 150. Of these 150, 63 are disqualified as elements of a valid transition matrix because they are multiple-step substitutions. Of the remaining 87, 34 are each based on a single *inferred* mutation. One such case is Arg → Lys. From sampling error, the true number of substitutions could well be two or three rather than one. The reverse mutation Lys → Arg occurs only eight inferred times with an associated error of about ± 3. This could explain entirely Gautier's abnormal result where the transition probabilities for Lys → Arg and Arg → Lys are listed as .0108 and .0048 (*10*), whereas Dayhoff finds (*3*) probabilities of .0023 and .0065, which are different both in magnitude and direction from the values calculated by Gautier. His finding that there are "more replacements of lysine to arginine than the converse" is therefore insupportable. Lysine is consistently present in most proteins at levels higher than those expected from the genetic code, while the reverse is true for arginine (Table 1). There is no observed phylogenetically dependent increase in arginine with respect to lysine in contemporary proteins. For a given protein family the proportions of these two amino acids is remarkably stable: for the cytochromes c isolated from 64 species, the percentage of lysine is 14.3 ± 2.5 (standard deviation) percent and of arginine 2.2 ± 0.6 percent. Gautier himself concludes (*10*) "Cependant le nombre trop faible de mutations décelées ne nous a pas permis d'obtenir des résultats certains sur ce problème des tendances." Gautier's discussion has no bearing on the "evolutionary intruder hypothesis," which refers to an event postulated to have occurred 2 billion or 3 billion years ago, before the present genetic code was in existence.

Glycine can be a helix-breaker, is less hydrophobic than alanine, and would be expected to be selected against relative to alanine in the interior of a molecule. Alanine is not a helix-breaker, is more hydrophobic, and is small enough to act as a "filler."

Our model for the interaction between "selective and neutral mutations to give a picture of dynamic equilibrium in protein evolution" can be obscure only if one has not read the literature (*11*). Gautier's supposition of what we mean is wrong: We do *not* believe and have not stated that the amino acids in the "expected range pool" replace each other without any action of natural selection. We did not state that all or most isoleucine and valine interchanges are neutral; some may be, and, for those, genetic drift would play a part.

Table 1. Amino acid distribution per 61 residues. The codons for Arg, Cys, Gln, Glu, Gly, Leu, Lys, Ser, Trp, and Tyr can each go to a chain-terminating codon by a single base change. The frequency of occurrence of Gln, Glu, Gly, Lys, and Tyr is incompatible with any hypothesis which states that "amino acids which possess codons that can go to a chain-terminating codon in one step are selected against."

Residue	a*	b†	c‡	d§	e‖
Ala	4	5.3	5.2	5.3	4.9
Arg	6	2.6	2.7	2.8	2.4
Asn	2	3.0			2.7
Asp	2	3.6			2.9
Asx	(4)	(6.6)	(6.5)	(6.3)	0.5
Cys	2	1.3	1.4	0.7	2.1
Gln	2	2.4			2.2
Glu	2	3.3			2.9
Glx	(4)	(5.7)	(6.5)	(6.6)	0.6
Gly	4	4.8	4.9	4.7	4.6
His	2	1.4	1.3	1.2	1.3
Ile	3	3.1	3.0	2.6	2.8
Leu	6	4.7	3.9	5.1	4.5
Lys	2	4.1	3.9	4.1	4.3
Met	1	1.1	1.1	0.9	1.0
Phe	2	2.3	2.3	2.4	2.1
Pro	4	2.5	2.9	3.0	3.4
Ser	6	4.5	3.8	4.7	4.8
Thr	4	3.7	3.5	3.6	4.0
Trp	1	0.8	0.8	0.5	0.7
Tyr	2	2.3	2.0	2.2	2.1
Val	4	4.2	4.1	4.2	4.2

*From the genetic code table. †From (*4*). ‡From (*1*). §From (*2*). ‖From (*3*).

Because of the paucity of data or references to the literature in Gautier's comment, and the variance of his theoretical inferences from a significant body of experimentally established fact, we cannot credit his conclusions.

We observed in preparing Table 1 (footnote) that five of the ten amino acids which possess codons that can mutate to a chain-terminating codon by a single nucleotide replacement are present in frequencies in excess of expectation (Table 1). One might have anticipated a selection against these ten amino acids, but, as a class, such codons are clearly not strongly selected against.

THOMAS H. JUKES
RICHARD HOLMQUIST
HERBERT MOISE
Space Sciences Laboratory, University of California, Berkeley 94720

References and Notes

1. G. R. Reeck and L. Fisher, *Int. J. Protein Res* **5**, 109 (1973).
2. M. H. Smith, *J. Theor. Biol.* **13**, 261 (1966).
3. M. O. Dayhoff, R. V. Eck, C. M. Park, in *Atlas of Protein Sequence and Structure* **5**, 92 (1972), figure 9-7. The values of Dayhoff *et al.* are probabilities for two substitutions distributed among 100 residues, whereas Gautier's are for one substitution among 20 residues. Adjustment to a common unit would not alter our conclusions.
4. T. H. Jukes, R. Holmquist, H. Moise, *Science* **189**, 50 (1975).
5. Abbreviations for the amino acid residues are as follows: Ala, alanine; Arg, arginine; Asp, aspartic acid; Asn, asparagine; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, iosleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine.
6. J. L. King and T. H. Jukes, *Science* **164**, 788 (1969).
7. R. Holmquist and H. Moise, *J. Mol. Evol.* **6**, 1 (1975).
8. R. Holmquist, *ibid.* **4**, 277 (1975).
9. W. Salser, S. Bowen, D. Browne, F. El Adli, N. Fedoroff, K. Fry, H. Heindell, G. Paddock, R. Poon, B. Wallace, P. Whitcome, *Fed. Proc. Am. Soc. Exp. Biol.* **35**, 23 (1976).
10. C. Gautier, thesis sustained at Lyon (24 June 1975), p. 38, table 1: "Les taux de mutations orientés: une mésure de l'évolution de protéines." This thesis was sent to us in response to a request for reference 4 in Gautier's comment.
11. References to this literature are in G. W. Moore, M. Goodman, C. Callahan, R. Holmquist, T. H. Jukes, *J. Mol. Biol.* **105**, 15 (1976); and in R. Holmquist, T. H. Jukes, H. Moise, M. Goodman, G. W. Moore, *ibid.*, p. 39.

27 September 1976

Acid Precipitation: Strong and Weak Acids

In a recent report Frohliger and Kane (*1*) have made various assertions concerning the acid character of rain. Their concluding remarks (*1*) perhaps best summarize these assertions: "The presence of weak acid species in precipitation casts additional doubt on the idea that the *p*H of rainfall was ever controlled by the solubility of CO_2 in the precipitation and discredits the assumption that strong acids account for the presently observed *p*H value." Considering the scientific and environmental implications of this statement, a careful and complete examination of the facts and their interpretation is in order.

Frohliger and Kane report on samples from 26 selected precipitation events collected during a 5-month period (December through May) in Pittsburgh, Pennsylvania (*2*). The data are in the form of two chemical parameters per sample, *p*H and total acidity (integrated over a *p*H range of < 5 to $\simeq 9$). They report *p*H values ranging from 4.12 to 5.78, and a buffer intensity 2 to 100 times greater than anticipated for pure water or an unbuffered acid, and then infer that the proton donors in their precipitation samples are weak acids. From these data, they discount the presence and function of strong acids in precipitation. They arrive at the latter conclusion by means of an incomplete syllogism. Their arguments run