Biophysical Chemistry, 44 (1992) 79-80
Elsevier Science Publishers B.V., Amsterdam

BIOCHE 01677

Letter to the Editor

Reply to Wada's comments on the entropies of coding and noncoding sequences of DNA and proteins

Gordan Lauc [a] and Marija Heffer-Lauc [b]

[a] *Laboratory of Physical Chemistry, Faculty of Science, University of Zagreb, Marulićev trg 19, 41 000 Zagreb (Croatia)*
[b] *Department of Chemistry and Biochemistry, Faculty of Medicine, University of Zagreb, Šalata 3, 41 000 Zagreb (Croatia)*

(Received 26 March 1992)

In his recent comment [1] on our article [2] Professor A. Wada proposed an interpretation for the paradoxical relationship that exists between entropies of reading frame and the frameshifted versions (+1 and −1) in DNA and proteins. We can but only partly agree with the suggested explanation.

Assuming triplets as the coding units, the joint probability $P_{1\text{-}2,3}$ resulting from biased codon usage may lower the entropy of the correct reading frame (compared to the frameshifted counterparts). However, a part, if not the most, of the entropy difference seems to originate from evolutionary defined constraints to yield a functional protein. Even when single nucleotides are considered as coding units, the entropy of the coding region is lower than the entropy of the complete mRNA sequence (albeit with marginal statistical significance) [2], a fact which cannot be explained with biased codon usage.

There are reasons to oppose the suggested explanation for the observed highest entropy of the correct reading frame with amino acids as the codngs units. In his letter [1] Wada suggests that some amino acid species have no or very little chance to appear in the frameshifted translations because of the restricted choice of the third letter due to the usage of so called 'optimal codons', and that the absence of those amino acids lowers the entropy in the frameshifted sequences.

Although biased codon usage limits the selection of the third letter in the code, when averaged over a complete sequence, *all* four nucleotides appear in the third position with similar frequencies (with a deviation due to a different GC content). Combining this with the fact that neither the preceding amino acid, nor the codon

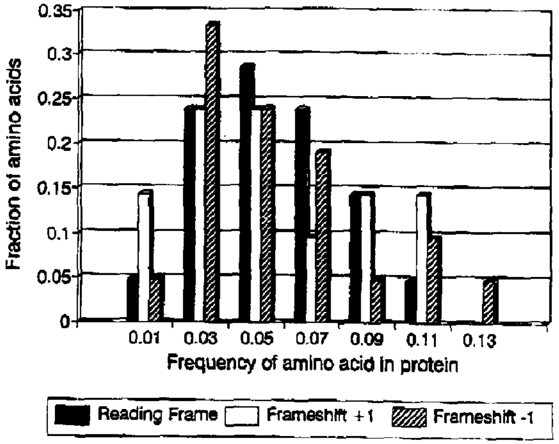

Fig. 1. Frequency distribution of the amino acids for 20 protein sequences from *E. coli*. Frameshifts −1 and +1 are distributions of amino acids obtained by translation of sequences with the shifted reading frame.

selection determine succeeding nucleotides we have anticipated that the distribution of amino acids must be similar in all three reading frames. An analysis of original sequence data, presented in Fig. 1, confirms the hypothesis (χ^2 test estimated probability of identical distributions above 0.99).

Thus, the suggestions offered by Wada may provide an explanation for the entropy difference concerning triplets as code units, but the 'entropy paradox' observed with the amino acids as code units still remains unexplained.

References

1 A. Wada, Biophys Chem, 44 (1992) 77.
2 G. Lauc, I. Ilić and M. Heffer-Lauc, Biophys. Chem. 42 (1992) 7.