

### OBSERVED CORRELATION BETWEEN COMPOSITION AND MOLECULAR WEIGHT OF AMINO ACIDS IN PROTEINS

THE composition of various proteins, in particular, the distribution of amino acids has been a topic of considerable interest. Early work<sup>1,2</sup> quoted in Meister<sup>3</sup> would appear to yield some general conclusions, although some of these need revision in the light of recent analysis (Dayhoff<sup>4</sup>) which is based on primary sequence analysis. The latter analysis has yielded the relative frequency of occurrence of the amino acids according to the classification of polar, apolar, heterocyclic (and aromatic) and ionisable types.

The randomness or otherwise of the occurrence of these amino acids has been treated statistically<sup>5-7</sup> and tables are available to assist in assessing whether or not their occurrence in a particular protein may be considered as random. Other statistical studies<sup>8-16</sup> are aimed at the analysis of the relative frequency of occurrence of residues in relation to the characteristic secondary structure of the protein chain, with a view to using them in algorithms for predicting protein structure from the given sequence. So also analysis from the point of view of volume occupied by side chain is also available<sup>17</sup>. However as far as we are aware of, no study seems to have been made on the relative abundance of the amino acids in relation to the molecular weight of the constituents.

We have done this and the simple linear correlation that has been observed, is felt to be sufficiently interesting to merit reporting here. Though Dayhoff's Table D 23<sup>4</sup> includes 108 proteins, a few fibrous proteins and also some small peptides have been included in it. In view of this we have selected 32 proteins whose structure by x-ray crystallographic

methods have been reported and hence the sequence is uniquely known. This we believe is a sufficiently good sample to yield statistically meaningful result. Figure 1 shows the relation between relative frequency of occurrence and the molecular weight.

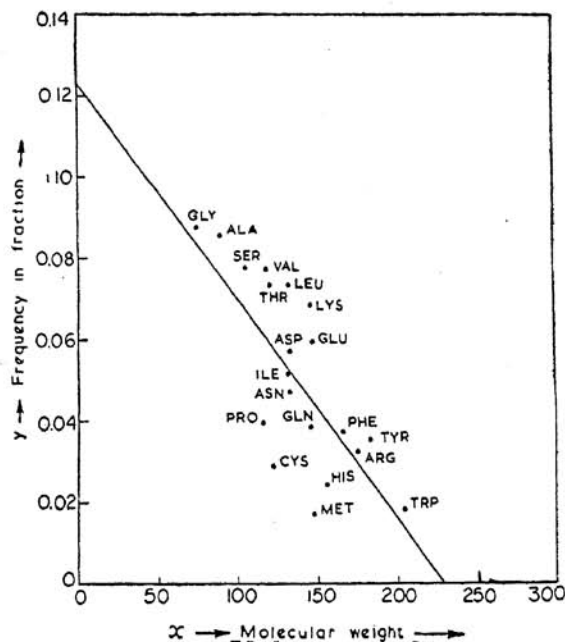


FIG. 1. Plot of molecular weight vs. the relative frequency of occurrence of amino acids in 32 proteins.

It is interesting to note that there is an approximately linear relation. The largest deviation from the above line is noticed for methionine and cystine and it is interesting that these two are amino acids containing sulphur. Among the rest, leucine (non-polar), lysine (polar) and histidine (heterocyclic) show somewhat large deviations.

The least squares equation to the line in Fig. 1 :  
$$y = -0.00054x + 0.12445.$$

The equation leads to the result that nearly 67% of the amino acids are those having molecular weights less than 137 which is the mean of the 20 amino acids. At the two ends, the relative frequencies work out to be 8.4% for glycine and 1.4% for tryptophan. These are respectively about 70% more and 70% less than the mean (5%) to be expected if all the 20 amino acids are randomly distributed with equal probability.

Further analysis from this angle is in progress and results will be reported in due course.

Dept. of Crystallography and  
Biophysics,  
University of Madras,  
Guindy Campus, Madras 600 025,  
India, July 30, 1976.

S. S. RAJAN.  
R. SRINIVASAN.

1. Tristram, G. R., *The Proteins* (Eds. Neurath, H. and Bailey, K., Academic Press, London, 1953), Vol. 1, Part A, p. 181.
2. —, *Advances in Protein Chemistry*, 1963, 18, 227.
3. Meister, A., *Biochemistry of Amino Acids*, Academic Press, New York, London, 1965, Vol. 1, Chap. I.
4. Dayhoff, M. O. and Hunt, L. T., *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Maryland, 1972, Table D 23, D-356.
5. Holmquist, R., *J. Mol. Evol.*, 1975, 4 (4).
6. Laird, M. and Holmquist, R., *Ibid.*, 1975, 4 (4), 261.
7. Black, J. A., Stenzel, P. and Harkins, R. N., *J. Theor. Biol.*, 1975, 50 (1), 161.
8. Chou, P. Y. and Fasman, G. D., *J. Mol. Biol.*, 1973, 74, 263.
9. — and —, *Biochemistry*, 1974, 13, 211.
10. — and —, *Ibid.*, 1974, 13, 222.
11. Burges, A. W., Ponnuswamy, P. K. and Scheraga, H. A., *Israel J. Chem.*, 1974, 12, 239.
12. Lim, V. I., *J. Mol. Biol.*, 1974, 88, 857.
13. —, *Ibid.*, 1974, 88, 873.
14. Nagano, K., *Ibid.*, 1973, 75, 401.
15. —, *Ibid.*, 1974, 84, 337.
16. Ptitsyn, O. B., *Ibid.*, 1969, 42, 501.
17. Chothia, C., *Nature*, 1975, 254, 304.