DARWINIAN EVOLUTION AT MOLECULAR LEVEL: A NEW METHOD OF TESTING THE HYPOTHESIS

R. BALASUBRAMANIAN AND G. RAGHUNATHAN
Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Madras 600 025, India

ABSTRACT

A mathematical parameter for measuring the amount of randomness in the sequences of nucleotides in a gene is considered. The homologous sequences of cytochrome c drawn from different species are used to evaluate this parameter for purposes of determining the trend in evolutionary changes. The results show that the values of the parameter tend to increase as the number of amino acid differences of various species (from human sequence) decreases, evidencing the existence of some kind of positive selection pressure in the evolution of species.

SEVERAL articles have contributed towards the understanding of the general trend in the evolutionary variations in the important class of biological macromolecules, *viz.* proteins.[1–20] They propound essentially two schools of thought on evolutionary changes at molecular level[21]—one holding the view that changes in the amino acid sequences of proteins that occur in the course of evolution are selected for, in the Darwinian sense and the other contending that they are selectively neutral, fixed as a result of random process. This issue has been analysed threadbare earlier[22].

It is well known that each protein, depending upon its functional role in the biological organisms is expected to have its own unique primary structure. As time progresses, mutational changes in the gene, coding for a protein take place altering the latter's primary structure. It is also well established that some of the amino acids along the sequence of a given protein are invariant in evolution and these invariant residues seem to be indispensable for the conservation of essential functions of the molecule in the various species.[10] Since these invariant residues mostly do not take part in the variation of the primary structure during evolution, it is logical to expect that in homologous proteins, variations in the amino acids in the non-invariant positions would hold the clue for understanding the trends in evolutionary changes. Hence the study of the amount of randomness (or non-randomness) introduced into the sequence as a result of these substitutions, as one advances along the evolutionary hierarchy would give a picture of the real trend in the evolution of species.

It is reasonable to expect that random mutations are taking place in the gene that code for a particular protein. Let us suppose that all such random changes are fixed in population with almost equal probability (except for mutations that correspond to invariant amino acid residues, since they are mostly indispensable for survival). This supposition corresponds to what may be termed as random mutation and neutral genetic drift. In such a case, non-invariant residues in homologous proteins are expected to drift towards a code-random composition* as one moves up the evolutionary hierarchy. In other words, if random mutations in the string of nucleotides of the gene are allowed to get stabilised in population (randomly fixed) without any selection pressure acting on them, the triplet nucleotides of this gene is expected to drift towards a uniform composition and hence, the protein coded by it will also drift towards code-random composition. The non-parametric χ^2-statistic may readily be used for estimating the deviation from code-random composition, of a protein drawn from a species near the bottom of the evolutionary scale and one may study the variation of the values of χ^2 for the same protein, derived from various species.

In our calculations, χ^2 has been defined as

$$X^2 = \sum_{i=1}^{20} \frac{(n_i - Nf_i)^2}{Nf_i}$$

where n_i is the observed frequency of one of the twenty amino acids, f_i is the expected fractional frequency for a code-random composition and N is the total number of *non-invariant* residues in the homologous sequences of a protein (see refs. 23, 6 and 24 for χ^2 including all residues).

Cytochrome c has been considered for our present study (see refs. 3, 9–11, 20 and 25–29 for studies of

*A random nucleotide sequence is expected to have equal probability of occurrence for all the possible (64) triplets and hence the amino acid composition of the protein, coded by such a gene would correspond to the frequency of amino acids as found in the Table of Genetic Code. Such a composition of amino acids is termed as code-random composition.

evolution using cytochrome c) since it is ubiquitous; moreover, it is the only protein for which sequences from a wide range of species are known (hemoglobin is another protein for which sequences from a large number of species are known; but this protein is found only in higher organisms). The homologous sequences of cytochrome c are taken from Dayhoff[30], which reports only those sequences which are at least 5% different from one another. Each species in their table is thus a representative of a set of closely related species and this data set would be enough for our calculations and inferences, as will be seen below.

If all the random mutations that are expected to take place in the gene coding for cytochrome c were given equal opportunity to be fixed in population (except for the mutations that happen to change the invariant amino acid residues, since they are essential for the functional characteristics of the molecule) one would expect that after sufficiently long time, the primary structure of nucleotides would tend towards a random sequence and hence code-random composition for the protein coded by the gene. Thus the value of χ^2 is expected to decrease as one goes up the evolutionary hierarchy, if evolution were a near-neutral process as discussed in refs. 31–34.

Precise evolutionary hierarchy among various species has not yet been established, beyond reasonable doubts and disputes. But the following stands are fairly well established. (a) *Homo sapiens* are the latest in evolutionary manifestation. (b) Differences in amino acid sequences of homologous proteins of different species are perhaps the best measure[3] of the distance of the species from say, the *homo sapiens*. Hence we plot χ^2 as calculated above for cytochrome c's of various species against the differences in amino acids from the amino acid sequence of human cytochrome c_5 in order to get a picture of the trend in the evolutionary changes and this is shown in figure 1. The increasing tendency in the values of χ^2 in figure 1 as the number of differences decreases indicates a definitive selection pressure in evolution (whatever be its biochemical nature). In order to get a clear picture of this trend, the least square straight-line for this set of data is also shown in figure 1, by a continuous line. The equation for this line is $Y = -1{\cdot}15X + 107{\cdot}54$. Its negative slope of $-1{\cdot}15$ clearly shows the tendency for the values of χ^2 to increase, as one moves up the evolutionary hierarchy. This testifies that some kind of selection pressure is operating through evolution. Hence the selectionists seem to score a point from this new approach in looking at the evolutionary changes in protein molecules.

Out of curiosity, we calculated the values χ_1^2 for the sequences of the said species taking into consideration all the residues (both invariant and non-invariant residues). These are also given in table 1 and shown in figure 1 along with the data for the invariant residues alone. Both the least-square lines in figure 1 point towards the same tendency for the values of χ^2.

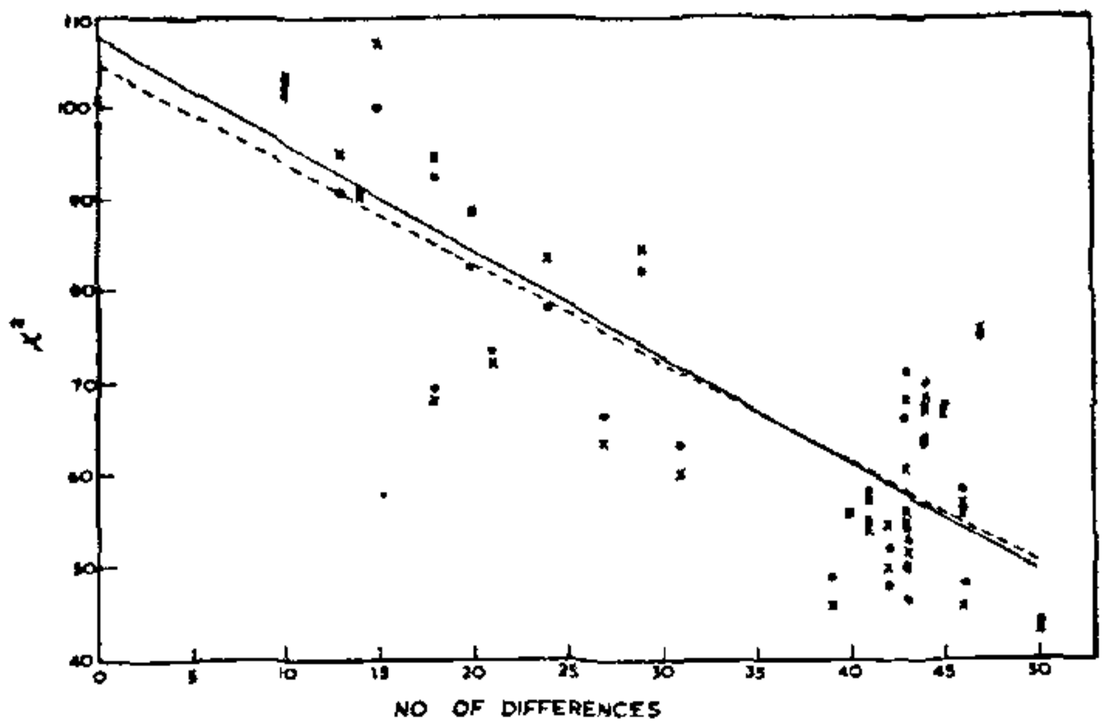

Figure 1. A plot of χ^2 versus number of differences in amino acids from the human sequence for various species of cytochrome c. The crosses correspond to the values of χ_1^2 for only non-invariant residues. The least-square line for this data is shown by a continuous line and has a slope of $-1{\cdot}15$. The dots correspond to the values of χ_1^2 for both variable and invariant residues. The least-square line for this data is shown by dotted line and has a slope of $-1{\cdot}08$.

It may be remarked that some of the large discrepancies (for example the large difference in χ^2 between candida krusei and debarymyces klockeri) may be due to various reasons. In some organisms, perhaps for special reasons applicable only to some individual species, a marked change in the protein sequence is necessitated by other local environmental requirements for the protein. Moreover, our knowledge of the origin and evolution of species, especially the lower ones, is sometimes vague, unreliable and incomplete.

In conclusion, we may remark that our main interest is in the overall nature or trend in the change of values of χ^2 as one moves along the course of evolution. Particularly χ^2 gives a very good picture of the path of evolution, as it has a direct bearing on the questions of mutation and fixation in population. Our study clearly shows that there are some selection pressures against the neutral drift towards a code-random composition.

ACKNOWLEDGEMENT

G. R. thanks the Council of Scientific and Industrial Research, New Delhi, for a fellowship.

1. Chirpich, T. P., *Science*, 1975, **188**, 1022.
2. Coutelle, R., Hofacker, G. L. and Levine, R. D., *J. Mol. Evol.*, 1979, **13**, 59.

TABLE 1

Number of amino acid differences from homo sapien sequence and values of χ^2 for cytochrome c's of different species

Species	No. of amino acid differences	χ^2 values for variable residues only	χ_1^2 for total (variable and invariant residues)
Human and chimpanzee	0	101·0	98·3
Pig, Bovine and sheep	10	100·1	101·6
Gray kangaroo	10	103·7	102·3
Chicken and turkey	13	95·2	91·7
Snapping turtle	15	107·3	100·1
Rattle snake	14	91·5	91·7
Bull frog	18	94·5	92·6
Bonito	21	72·0	73·6
Carp	18	68·0	69·4
Puget sound dogfish	24	83·6	77·9
Pacific lamprey	20	88·8	84·9
Garden snail	29	84·6	82·3
Screw-worm fly	27	63·4	66·7
Tobacco horn worm moth	31	60·2	63·3
Candida krusei	50	43·2	44·3
Debarymyces klockeri	45	66·7	67·7
Baker's yeast-Iso-I	44	66·9	70·3
Humicola languinosa	43	60·5	66·3
Neurospora crassa	43	68·1	72·2
Rust fungus	47	75·5	75·2
Rape and cauliflower	44	67·5	64·2
Hemp	46	45·9	48·2
Sesame	39	45·8	49·4
Castor	41	54·1	54·8
Elder	46	57·0	58·7
Box-elder	46	50·8	56·4
Leek	42	54·6	52·3
Love-in-a-mist	43	56·1	53·1
Nasturtium	44	68·4	63·5
Wheat	43	52·2	50·2
Sunflower	42	50·3	48·1
Parsnip	43	55·1	54·7
Buckwheat	40	55·3	55·7
Spinach	43	50·6	46·5
Ginkgo Biloba	41	57·1	58·3

3. Dayhoff, M. O., Park, C. M., and McLaughlin, P. J., *Atlas of protein sequence and structure*, National Biomedical Research Foundation, Washington, 1972, **7**, 16.
4. Fitch, W. M., *J. Mol. Evol.*, 1976, **8**, 13.
5. Gautier, C., *Science*, 1976, **194**, 642.
6. Holmquist, R., *J. Mol. Evol.*, 1978, **11**, 349.
7. Jukes, T. H., Holmquist, R. and Moise, H., *Science*, 1975, **189**, 50.
8. Jukes, T. H., Holmquist, R. and Moise, H., *Science*, 1976, **194**, 642.
9. Lyddiat, A., Peacock, D., and Boulter, D., *J. Mol. Evol.*, 1978, **11**, 35.
10. Margoliash, E., *Adv. Chem. Phys.*, 1975, **24**, 191.
11. Margoliash, E., Miller, S. F., Kang, C. H. and Brautigan, D. L., *Fed. Proc.*, 1976, **35**, 2124.
12. Moore, G. W., Goodman, M., Callagan, C., Holmquist, R. and Moise, H., *J. Mol. Biol.*, 1976, **105**, 15.
13. Ohta, R. and Kimura, M., *Science*, 1971, **174**, 150.
14. Rajan, S. S. and Srinivasan, R., *Curr. Sci.*, 1976, **45**, 859.

15. Reichert, T. A. and Wong, A. K. C., *J. Mol. Evol.*, 1971, **1**, 97.
16. Romero-Herrera, A. E., Lehmann, H., Joysey, K. A. and Friday, A. E., *Nature (London)*, 1973, **246**, 389.
17. Schwartz, R. M. and Dayhoff, M. O., *Science*, 1978, **199**, 395.
18. Srinivasan, R., *Indian. J. Biochem. Biophys*, 1978, **1**, 75.
19. Wilson, A. C., Carlson, S. S. and White, T. J., *Annu. Rev. Biochem.*, 1977, **46**, 573.
20. Yamanaka, T., *Adv. Biophys.*, 1972, **3**, 227.
21. Margoliash, E., *Fed. Proc.*, 1976, **35**, 2102.
22. FASEB Conference on Genetics and Biological Evolution, *Fed. Proc.*, 1976, **35**, 2073.
23. Cornish-Bowden, A., and Marson, A., *J. Mol. Evol.*, 1977, **10**, 231.
24. Laird, M. and Holmquist, R., *J. Mol. Evol.*, 1975, **4**, 261.
25. Boulter, D., *Curr. Adv. Pl. Sci.*, 1974, **8**, 1.
26. Dayhoff, M. O., *Atlas of protein sequence and structure*, Vol. **5**, 1, National Biomedical Research Foundation, Washington, 1973.
27. Fitch, M., *J. Mol. Evol.*, 1976, **8**, 13.
28. Fitch, W. M. and Margoliash, E., *Science*, 1967, **155**, 279.
29. Moore, G. W., Goodman, M. and Barnabas, J., *J. Theor. Biol.*, 1973, **38**, 423.
30. Dayhoff, M. O., *Atlas of Protein Sequence and Structure*, Vol. **5**, Suppl. 2, National Biomedical Research Foundation, Washington, 1976.
31. Harris, H., *Fed. Proc.*, 1976, **35**, 2079.
32. King, J. L. and Jukes, T. H., *Science*, 1969, **164**, 788.
33. Kimura, M. and Ohta, T., *Proc. Natl. Acad. Sci., U.S.A.*, 1974, **71**, 2848.
34. Fitch, W. M. and Langley, C. H., *Fed. Proc.*, 1976, **35**, 2092.

SYNTHESIS OF METHYL-METHACRYLATE GRAFTED NATURAL RUBBER IN BULK PHASE

J. S. P. RAI. S. K. JAIN AND G. N. MATHUR
Department of Plastics Technology, H. B. Technological Institute, Kanpur 208 002, India

ABSTRACT

Methyl-methacrylate grafted natural rubber was prepared at 40, 50, 60 and 70° C in the presence of benzoyl peroxide as thermal initiator and NN′ dimethyl aniline as activator. From the gross copolymerisation product, the rubber-PMMA grafted inter polymer fraction was isolated from the free rubber and homopoymer. The efficiency of grafting under varied conditions has been discussed.

INTRODUCTION

THE graft copolymerisation of methyl-methacrylate (MMA) onto natural rubber can be carried out in solution, in latex or in bulk rubber swollen with the monomer. Considerable amount of literature is available on grafting of MMA with natural rubber in *solution*[1–5] *and in* emulsion using the latex rubber[4,6–14]. However, very little is known about the reaction in bulk phase. Swift[15] studied the graft copolymerisation of MMA with natural rubber in solid phase at 80° and 100° C. Ceresa[16] has studied the grafting of acetone extracted and masticated natural rubber with MMA at 80° C and obtained grafted rubber. As dry rubber is easy to handle and transportation from place to place is not cumbersome when compared to latex, it is of interest to investigate the preparation of grafted rubber in a mass of dry rubber which has been allowed to imbibe the necessary quantity of monomer (MMA in our case). The aim of this investigation was to ascribe the conditions obtained during such polymerisations and to evolve a process which works within the limitations imposed by these conditions. In this communication we present the results on the compositions of natural rubber MMA graft copolymerisation products under varied reaction conditions.

EXPERIMENTAL

A known amount of rubber (smoked sheet) was allowed to swell in a fixed volume of the monomer (MMA-distilled and free from inhibitors) containing a known amount of the initiator benzoyl peroxide (recrystallized) and activator (NN' dimethyl aniline boiling range 191–195° C) for about 24 hr in dark at room temperature (23 ± 2° C) under inert atmosphere. The polymerisation was carried out in a constant