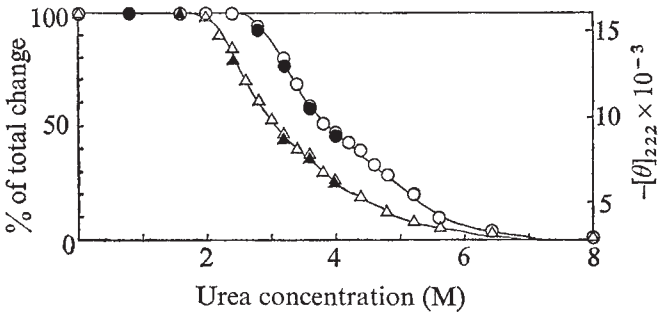


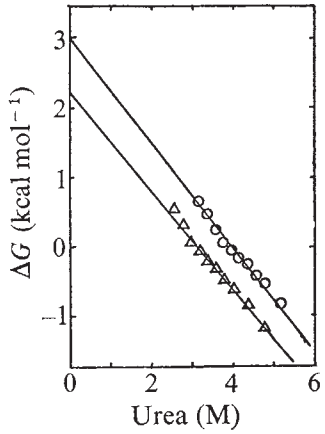
Fig. 1 Urea denaturation of the α-subunits of the wild type (△, ▲) and the mutant *trp*A*33* (○, ●). Temperature 25 °C, *p*H 8.0, 0.025 M Tris-HCl buffer. The solution contained 0.1 mM EDTA and 0.2 mM dithioerythritol. Open symbols represent the ellipticity at 222 nm after incubation for 60 min in urea solutions at various concentrations. Filled symbols represent the measurement of the samples diluted after exposure to 8 M urea solution to demonstrate the reversibility of the denaturation. CD measurements were carried out with a Jasco J-20 recording spectropolarimeter.

and enzymatic activity recovered completely. Urea denaturation is thus reversible in these conditions. The midpoints of urea denaturation curves at 25 °C were 3.1 and 3.9 M urea for the wild type and the mutant *trp*A*33*, respectively.

We analysed the data shown in Fig. 1 assuming that the urea denaturation can be approximately represented by an equilibrium between a unique native form and a completely unfolded form[8,9,15]. Figure 2 shows the dependence of the free energy change of unfolding on the urea concentration. By extrapolating to zero concentration, the free energy changes of unfolding in water for the α-subunit of the wild type and the mutant *trp*A*33* are calculated to be 2,290 and 3,040 cal mol^{-1}, respectively. The difference between the wild type and the mutant *trp*A*33*, 750 cal mol^{-1}, amounts to about 30% of the free energy change of unfolding for the wild type.

X-ray diffraction studies have shown that the replacement of Glu 49 by Met caused little alteration of the conformation of the α-subunit[16]. Our results, taken together with the above data, show that a single amino acid substitution increased the stability of the molecule without a gross change in conformation. The side chain contribution of an amino acid to free energy change for transfer from ethanol to water has been estimated to be 1,300 and 550 cal mol^{-1} for Met and Glu, respectively[17]. The difference between them, 750 cal mol^{-1}, coincides with the difference of the free energy change of unfolding of the wild type and the mutant *trp*A*33* proteins obtained here. This suggests that the increase in stability

Fig. 2 Effect of urea concentration on the free energy change of unfolding for α-subunits of the wild type (△) and the mutant *trp*A*33* (○). ΔG for two-state transition is given by the following equation; $\Delta G = -RT \ln K$. An equilibrium constant, K = (concentration of unfolded molecule)/(concentration of native one), is calculated from the data shown in Fig. 1, using only those points where K is between 0.2 and 5.0 (ref. 15). The lines in the figures were calculated by the method of least squares.



of the mutant *trp*A*33* protein compared with the wild type might be accounted for by the difference of hydrophobicity of Glu and Met, if we assume that the position 49 is somehow buried in the hydrophobic interior of the molecule. Although hydrophilic residues such as glutamic acid are usually located on the surface of protein molecules, there are exceptions, especially when the residues are in the active sites of enzymes[18]. In the case of Glu 49 of the α-subunit of tryptophan synthetase, its substitution by Met results in loss of activity with indole glycerolphosphate and serine as substrate, though not with indole and serine[14]. We may therefore assume that it is in the active site of the enzyme and hence in a largely hydrophobic interior environment, which may be simulated by ethanol or similar solvents, and in which partners for hydrogen bonding are at hand.

Our results suggest that the stability of an enzyme may be greatly increased by the increase in hydrophobicity by substitution of a few suitable amino acid residues. Stability of enzymes from thermophiles could be explained by similar mechanism.

We thank Drs K. Hamaguchi, T. Takagi, Y. Kyogoku and H. Ozeki for their valuable comments. Drs C. Yanofsky, I. P. Crawford and S. Hiraga provided mutant *E. coli* strains.

KATSUHIDE YUTANI
KYOKO OGASAHARA

Institute for Protein Research,

YOSHINOBU SUGINO

Faculty of Science,

AIZO MATSUSHIRO

Research Institute for Microbial Diseases,
Osaka University,
Osaka, Japan

Received 10 November 1976; accepted 20 January 1977.

1 Singleton, R. & Amelunxen, R. E. *Bact. Rev.* **37,** 320–342 (1973).
2 Ramachandran, G. N. & Sasisekharan, V. *Adv. Prot. Chem.* **23,** 283–437 (1968).
3 Spande, T. F., Witkop, B., Degani, Y. &Patchornik, A. *Adv. Prot. Chem.* **24,** 97–260 (1970).
4 Yutani, K. *Experientia.* Suppl. **26,** 91-103 (1976).
5 Tajima, M., Urabe, I., Yutani, K. & Okada, H. *Eur. J. Biochem.* **64,** 243–247 (1976).
6 Levy, P. L., Pangburn, M. K., Burstein, Y., Ericsson, L. H., Neurath, H. & Walsh, K. A. *Proc. natn. Acad. Sci. U.S.A.* **72,** 4341-4345 (1975).
7 Tanaka, M., Haniu, M., Yasunobu, K. T., Himes, R. H. & Akagi, J. M. *J. biol. Chem.* **248,** 5215–5217 (1973).
8 Alexander, S. S. & Pace, C. N. *Biochemistry* **10,** 2738-2743 (1971).
9 Puett, D. *J. biol. Chem.* **248,** 4623–4634 (1973).
10 Yanofsky, C. & Horn, V. *J. biol. Chem.* **247,** 4494–4498 (1972).
11 Maling, B. D. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **47,** 551–566 (1961).
12 Creighton, T. E. & Yanofsky, C. *Meth. Enzym.* **17,** part A, 365–380 (1970).
13 Heyn, M. P. & Weischet, W. O. *Biochemistry* **14,** 2962–2968 (1975).
14 Yanofsky, C. & Crawford, I. P. *The Enzymes,* 3rd edn **7** (ed. Boyer, P. D.) 1–31 (Academic, New York, 1972).
15 Tanford, C. *J. Am. chem. Soc.* **86,** 2050–2059 (1964).
16 Schulz, G. E. & Creighton, T. E. *Eur. J. Biochem.* **10,** 195–197 (1969).
17 Tanford, C. *J. Am. chem. Soc.* **84,** 4240–4247 (1962).
18 Dickerson, R. E. & Geis, I. *The Structure and Action of Proteins* 67–97 (Harper & Row, New York, 1969).

Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution

ACCORDING to the neutral mutation–random drift hypothesis of molecular evolution and polymorphism[1,2], most mutant substitutions detected through comparative studies of homologous proteins (and the nucleotide sequences) are the results of random fixation of selectively neutral or nearly neutral mutations. This is in sharp contrast to the orthodox neo-Darwinian view that practically all mutant substitutions occurring within species in the course of evolution are caused by positive Darwinian selection[3–5]. This paper shows that by comparative studies of messenger RNA (mRNA) sequences reliable estimates can be obtained of the evolutionary rates (in terms of mutant substitutions) at the third positions of the codon, and that the estimates conform remarkably well with the framework of the neutral theory.

Salser *et al.*[6] have presented a comparison of homologous parts from the fragments of the human and rabbit haemoglobin mRNA sequences. Among 53 nucleotide positions that can be compared, there are six base differences, of which only one leads to amino acid difference. Their Table 5 shows that among 17 third nucleotide positions of the codon which can be compared

between the two species, there are 5 nucleotide differences. To estimate the number of nucleotide substitutions per site (K) that have occurred in the course of evolution since the divergence of the rabbit and human lineages, I use the following formula which converts the observed difference into the estimate of evolutionary divergence

$$K = -\tfrac{3}{4}\log_e(1-4/3\,\lambda) \quad (1)$$

In this formula, λ is the observed fraction of the sites by which two homologous sequences differ from each other (for details, see ref. 7); K includes superimposed and reverted mutant substitutions. The rate of nucleotide substitution per site per year can then be obtained by $k_{nuc} = K/(2T)$, where T is the time (yr) since the divergence of the two lineages. Also, the standard error (σ_K) of K may be computed using a formula given in ref. 7. Letting $\lambda = 5/17$ in equation (1), we obtain $K = 0.373$. The corresponding standard error (σ_K) is 0.182. Thus, taking $T = 8\times10^7$, we obtain $k_{nuc} = (2.3\pm1.1)\times10^{-9}$ (the present estimate should be statistically more reliable than the corresponding estimate given in ref. 6). This is a very high evolutionary rate, comparable with that of the fibrinopeptides. If we denote by k_{aa} the rate of amino acid substitutions per site, it is known that for fibrinopeptides $k_{aa} = (4\sim9)\times10^{-9}$ (see refs 8, 9). This figure, when converted into the nucleotide substitution rate, is $k_{nuc} = (1.8\sim4.0)\times10^{-9}$. In addition, a direct comparison of human and rabbit fibrinopeptide A (using data given in ref. 8) gives $k_{aa} = 5.2\times10^{-9}$ leading to $k_{nuc} = 2.3\times10^{-9}$.

A similar but more interesting report comes from Grunstein *et al.*[10] who compared histone H4 (that is, histone IV) mRNA sequences of two sea urchin species *S. purpuratus* and *L. pictus*. It is well known that histone H4 is by far the most highly conserved protein (with $k_{aa} = 0.006\times10^{-9}$ (ref. 8)). Their report[10] now shows, however, that many synonymous changes (that is, the changes that do not lead to amino acid changes) have occurred in the gene coding for this protein in the course of evolution. From their Table 3, showing 84 nucleotide sites that can be compared between the messenger sequences of the two species, there are ten base differences, of which nine are located at the third positions of the codon. Actually, there are 27 third nucleotide positions which can be compared between these two species and there are nine nucleotide differences. Thus, we obtain $K = 0.440$ and $\sigma_K = 0.163$. According to Grunstein *et al.*, these sea urchins shared a common ancestor approximately 6×10^7 yr ago. Therefore, letting $T = 6\times10^7$, we obtain $k_{nuc} = (3.7\pm1.4)\times10^{-9}$ as an estimate of the average evolutionary rate at the third nucleotide position of the codon in the histone H4 gene. It is remarkable that, in this gene, synonymous mutant substitutions have occurred with the highest known rate, in spite of the fact that the amino acid substitutions have occurred in the corresponding protein with the lowest known rate.

How do these observations bear relation to the neutral theory? Let k be the rate by which mutant genes are substituted in the species in evolution. If a certain fraction f_0 of the molecular mutants are selectively neutral and the rest are definitely deleterious (assuming that definitely advantageous mutations are negligible in frequency), then for neutral alleles, the rate of mutant substitutions is equal to the mutation rate[1]. Therefore

$$k = v_T f_0 \quad (2)$$

where v_T is the total mutation rate. Now, among the prominent features[11] of molecular evolution the following two are particularly noteworthy: (1) for a given protein, the rate of evolution is roughly constant per year and (2) the molecules or parts of molecules that are subject to less functional constraints evolve (in terms of mutant substitutions) faster.

As to the first feature, the neutralists assume that for a given molecule f_0 is roughly constant. Recently, some authors, such as Fitch and Langley[12], have emphasised the non-constancy, but their results show, in agreement with our earlier estimate[13], that the variance of evolutionary rates among lines is only about 2.6 times as large as that expected from chance fluctuations. It is possible that there is delicate fluctuation of intrinsic evolutionary rate around the mean, as caused by shifting of the molecular constraint as amino acids are substituted one after another in various parts of the molecule. This means that the probability f_0 of a mutation being neutral is not strictly constant but fluctuates around its characteristic mean. Also, as pointed out by Ohta[14], if nearly neutral but very slightly deleterious mutations are prevalent, the evolutionary rate fluctuates as the population goes through a series of bottlenecks. As to the difference of evolutionary rates among molecules or parts of molecules, the neutralists assume that the probability of a mutational change being neutral depends on functional constraints. Namely, the weaker the functional constraint, the larger the probability (f_0) of a random change being selectively neutral, with the result that k in equation (2) gets larger. According to this explanation, the maximum evolutionary rate is attained when $f_0 = 1$, that is, when all the mutations are neutral. Now, the high evolutionary rates observed at the third position of the codon can be explained from the neutral theory by assuming that the majority of synonymous changes are selectively neutral. Note that roughly 2/3 of random nucleotide substitutions at the third position of the codon are synonymous. The possibility of synonymous nucleotide changes being selectively neutral was discussed extensively in an early paper[2] on the neutral theory. Of course, it is possible that not all synonymous mutations are completely neutral, but the possibility is very high that, on average, synonymous changes are subject to natural selection very much less than the mis-sense mutations. On the other hand, if we adhere to the selectionist position that practically all the mutant substitutions in evolution are caused by positive natural selection, there can be no upper limit to the evolutionary rate at the molecular level (as directly set by the mutation rate v_T), and there is no reason to believe that the rate of evolution is uniform even approximately for a given molecule among different species.

In my opinion, various observations suggest that as the functional constraint diminishes the rate of evolution converges to that of the synonymous substitutions. If this is valid, such a convergence (or plateauing) of molecular evolutionary rates will turn out to be strong supporting evidence for the neutral theory.

MOTOO KIMURA

National Institute of Genetics, Mishima, 411, Japan

Received 22 December 1976; accepted 27 March 1977.

1 Kimura, M. *Nature* **217**, 624–626 (1968).
2 Kimura, M. *Genet. Res. Camb.* **11**, 247–269 (1968).
3 Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
4 Fisher, R. A. *Proc. R. Soc.* **B121**, 58–62 (1936).
5 Mayr, E. *Animal Species and Evolution* (Harvard University Press, Cambridge, Massachusetts, 1965).
6 Salser, W. *et al. Fedn Proc.* **35(1)**, 23–35 (1976).
7 Kimura, M. & Ohta, T. *J. molec. Evol.* **2**, 87–90 (1972).
8 Dayhoff, M. O. *Atlas of Protein Sequence and Structure 1972* (National Biomedical Research Foundation, Washington, D.C., 1972).
9 Barnard, E. A., Cohen, M. S., Gold, M. H. & Kim, J-K. *Nature* **240**, 395–398 (1972).
10 Grunstein, M., Schedl, P. & Kedes, L. *J. molec. Biol.* **104**, 351–369 (1976).
11 Kimura, M. & Ohta, T. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2848–2852 (1974).
12 Fitch, W. M. & Langley, C. H. *Fedn Proc.* **35**, 2092–2097 (1976).
13 Ohta, T. & Kimura, M. *J. molec. Evol.* **1**, 18–25 (1971).
14 Ohta, T. *Nature* **252**, 351–354 (1974).

Inheritance of recombinant HLA–GLO haplotype suggesting the gene sequence

THE HLA region in man has been assigned to the short arm of chromosome 6 (ref. 1). Close genetic linkage has been shown between *HLA* and genes involved in the synthesis of the second[2–3] and fourth[4] components of complement as well as structural variation of the glycine-rich B-glycoprotein (GBG or Bf)[5]. Two enzymes, phosphoglucomutase-3 (PGM_3)[6–7] and glyoxalase I (GLO)[1,8–12], and two blood group antigens, Chido[13] and Rodgers[1], have been shown to be in this linkage group. Here we describe a