J. Genet., Vol. 64, No. 1, July 1985, pp. 7–19. © Printed in India.

The role of compensatory neutral mutations in molecular evolution*

MOTOO KIMURA
National Institute of Genetics, Mishima, 411 Japan

MS received 14 February 1985

Abstract. A pair of mutations at different loci (or sites) which are singly deleterious but restore normal fitness in combination may be called compensatory neutral mutations. Population dynamics concerning evolutionary substitutions of such mutants was developed by making use of the diffusion equation method. Based on this theory and, also, by the help of Monte Carlo simulation experiments, a remarkable phenomenon was disclosed that the double mutants can easily become fixed in the population by random drift under continued mutation pressure if the loci are tightly linked, even when the single mutants are definitely deleterious. More specifically, I consider two loci with alleles A and A' in the first locus, and alleles B and B' in the second locus, and assign relative fitnesses 1, $1-s'$, $1-s'$ and 1 respectively to the four gene combinations AB, $A'B$, AB' and $A'B'$, where s' is the selection coefficient against the single mutants ($s' > 0$). Let v be the mutation rate per locus per generation and assume that mutation occurs irreversibly from A to A' at the first locus, and from B to B' at the second locus, where A and B are wild type genes, and A' and B' are their mutant alleles. In a diploid population of effective size N_e (or a haploid population of $2N_e$ breeding individuals), it was shown that the average time ($\bar{T}$) until joint fixation of the double mutant ($A'B'$) starting from the state in which the population consists exclusively of the wild type genes (AB) is not excessively long even for large $4N_es'$ values. In fact, assuming $2N_ev = 1$ we have $\bar{T} = 54N_e$ for $4N_es' = 400$, and $\bar{T} = 128N_e$ for $4N_es' = 1000$. These values are not unrealistically long as compared with $\bar{T} \approx 5N_e$ obtained for $4N_es' = 0$. The approximate analytical treatment has also been extended to estimate the effect of low rate crossing over in retarding fixation. The bearing of these findings on molecular evolution is discussed with special reference to coupled substitutions at interacting amino acid (or nucleotide) sites within a folded protein (or RNA) molecule. It is concluded that compensatory neutral mutants may play an important role in molecular evolution.

Keywords. Population genetics; stochastic process; the neutral theory; drift.

1. Introduction

Recently, much evidence has accumulated (see Kimura 1983 for review) supporting the neutral theory of molecular evolution which asserts that the great majority of evolutionary changes at the molecular level are caused not by Darwinian selection acting on advantageous mutants but by random fixation of selectively neutral or nearly neutral mutants (Kimura 1968). Furthermore, it is likely, as first seriously proposed by Ohta (1973, 1974), that very slightly deleterious mutations, in addition to neutral ones, play an important role in molecular evolution. In fact, the neutral mutations should be regarded as the limit of deleterious mutations (rather than advantageous ones) when the effect of the mutation on fitness becomes indefinitely small (Kimura and Ohta 1974).

In this paper, I intend to present a thesis that a pair of mutant genes which are individually deleterious but restore the normal function in combination also play an

*Contribution No. 1610 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.

important role in molecular evolution, particularly when they are tightly linked. Such counterbalancing mutations may be called compensatory neutral mutations, and I shall develop the population dynamics of such mutations in the next section.

2. Basic theory

Let us assume a random mating, diploid population of effective size N_e; roughly speaking, N_e is equal to the number of breeding individuals in one generation (for more details, see Crow and Kimura 1970, and Kimura and Ohta 1971).

We consider two linked loci (or sites), and denote by A the wild type allele at the first locus, and by B that at the second locus. Let us assume that, at the first locus, A mutates irreversibly to its allele A' at the rate v per generation. In reality, the mutant allele A' is not usually a single entity but a set of mutant alleles, but we designate them collectively as A'. Similarly, let us assume that, at the second locus, B mutates to allele B' irreversibly at the same rate v.

As to selection, we adopt the scheme of genic selection, and assume that the mutant alleles, A' and B', are individually deleterious, but that the double mutant $A'B'$ has the same fitness as the wild type AB. This is shown in table 1, where s' is the selection coefficient against the single mutants. In other words, we adopt a haploid selection model to simplify the mathematical treatment. Thus, although we consider a diploid population of effective size N_e, we can just as well consider a haploid population consisting of $2N_e$ breeding individuals.

I shall first consider the case of complete linkage between the two loci, and shall bring to light a remarkable phenomenon of the double mutant state ($A'B'$) easily evolving through a definitely deleterious intermediate state ($A'B$ or AB') by random drift under continued mutation pressure. I shall then examine how crossing-over between the loci interferes with this phenomenon.

2.1 *Evolution under complete linkage*

Since we assume no crossing-over between the two loci (or sites), we may treat the four genotypes AB, $A'B$, AB' and $A'B'$ as if they were four alleles at a single locus. Furthermore, we may lump the single mutants, $A'B$ and AB', together because of equal fitnesses, and denote their frequencies collectively as p_1. Let us denote the frequency of the double mutant $A'B'$ by p_2, and that of the wild type by p_0, where $p_0 = 1 - p_1 - p_2$ (See table 2).

Our main problem is to determine how long it takes for the double mutant $A'B'$ to become fixed (i.e. to reach 100% in frequency) in the population, starting from the state in which the population consists exclusively of the wild type AB. This is a stochastic

Table 1. Table of fitnesses of four gene combinations at two loci.

	A	A'
B	1	$1-s'$
B'	$1-s'$	1

Table 2. Assignment of fitness and frequency parameters for 4 genotypes under complete linkage between the two loci (or sites).

Genotype	AB	$A'B$ or AB'	$A'B'$
Fitness	1	$1-s'$	1
Frequency	p_0	p_1	p_2

process, and, for the purpose of obtaining the average length of time until fixation, I shall make use of the diffusion equation method or "diffusion models" as they are often called (see Kimura 1964 for review; see also chapter 8 of Crow and Kimura 1970, and mathematical appendix of Kimura and Ohta 1971).

In the problem under consideration, two independent variables, p_1 and p_2 are involved, and therefore we must consider a two-dimensional diffusion process as in the case of "duplicate loci" investigated by Kimura and King (1979).

Let $u = u(p_1, p_2; t)$ be the probability that the double mutant $A'B'$ becomes fixed in the population by the t-th generation, given that the frequencies of the single and the double mutants are respectively p_1 and p_2 at the start ($t = 0$). Then u satisfies the following Kolmogorov backward equation

$$\frac{\partial u}{\partial t} = \frac{p_2(1-p_2)}{4N_e}\frac{\partial^2 u}{\partial p_2^2} - \frac{p_1 p_2}{2N_e}\frac{\partial^2 u}{\partial p_2 \partial p_1} + \frac{p_1(1-p_1)}{4N_e}\frac{\partial^2 u}{\partial p_1^2} + M_{\delta p_2}\frac{\partial u}{\partial p_2} + M_{\delta p_1}\frac{\partial u}{\partial p_1}, \tag{1}$$

where $$M_{\delta p_2} = v p_1 + s' p_1 p_2, \tag{2a}$$

and $$M_{\delta p_1} = 2v p_0 - s' p_1 (1-p_1) - v p_1, \tag{2b}$$

are respectively the mean changes per generation of p_2 and p_1 caused by mutation and natural selection. This partial differential equation is similar to equation [3] of Kimura and King (1979), but it contains an additional term (the second term on the right hand side) which represents the correlation between changes of p_1 and p_2 due to sampling of gametes.

The analytical solution of this equation appears to be very difficult to obtain. Even the numerical treatment seems to be more cumbersome in this case than in Kimura and King's equation [3].

Fortunately, however, we can apply a shortcut approximation for the biologically interesting case, where

$$s' \gg v > 0, \tag{3}$$

as I shall explain below.

Under the assumption that the magnitude of selective disadvantage (s') of a single mutant is much larger than the rate of its mutational production (v), it is reasonable to suppose that the frequency (p_1) of single mutants remains very low throughout the process, because of an approximate balance between mutational input and selective elimination at each locus. Therefore, we may assume that quasi-equilibrium

$$M_{\delta p_1} = 0 \tag{4}$$

holds approximately for the entire process in which the double mutant $A'B'$ gradually replaces the wild type AB. Here we note that under the assumptions of unidirectional mutation ($A \to A'$, $B \to B'$) and finite population size ($N_e < \infty$), the double mutant will eventually become fixed in the population, although the average time taken for such a fixation may be extremely long if the single mutants are highly deleterious. Assuming this quasi-equilibrium, and disregarding the second order term in p_1^2, we have, from equations (2b) and (4),

$$2v(1-p_1-p_2)-s'p_1-vp_1=0, \tag{5}$$

which leads to

$$p_1=\frac{2v}{s'+3v}(1-p_2). \tag{6}$$

Substituting this in (2a), we obtain

$$M_{\delta p_2}=(v+s'p_2)\frac{2v}{s'+3v}(1-p_2). \tag{7}$$

This means that, under the assumption of quasi-equilibrium, or the moving equilibrium as given by (6), we need only to consider the process of change in p_2. In other words, for our purpose of obtaining the average fixation time of the double mutant, the two-dimensional diffusion process can be approximated by a one-dimensional diffusion process.

Thus, writing p for p_2, let us denote by $u=u(p,t)$ the probability that $A'B'$ becomes fixed in the population by the t-th generation, given that its frequency is p at time 0. Then, u satisfies the following one-dimensional diffusion equation

$$\frac{\partial u}{\partial t}=\frac{1}{2}V_{\delta p}\frac{\partial^2 u}{\partial p^2}+M_{\delta p}\frac{\partial u}{\partial p}, \tag{8}$$

where $M_{\delta p}$ and $V_{\delta p}$ are respectively the mean and variance of the change of p per generation, and in the present context, we have

$$V_{\delta p}=\frac{p(1-p)}{2N_e} \tag{9}$$

and

$$M_{\delta p}=s_1p(1-p)+v_1(1-p), \tag{10}$$

in which

$$s_1=\frac{2s'v}{s'+3v}, \tag{11a}$$

and

$$v_1=\frac{2v^2}{s'+3v}. \tag{11b}$$

Note that (10) is simply a rewriting of (7).

Let $\overline{T}(p)$ be the average time until fixation of $A'B'$, given that its initial frequency is p, so that

$$\overline{T}(p)=\int_0^\infty t\frac{\partial u(p,t)}{\partial t}\,\mathrm{d}t. \tag{12}$$

Then, using (8), we can show (Kimura 1980) that $\overline{T}(p)$ satisfies the ordinary differential

equation

$$\frac{1}{2}V_{\delta p}\frac{d^2\overline{T}(p)}{dp^2}+M_{\delta p}\frac{d\overline{T}(p)}{dp}+1=0, \tag{13}$$

where the appropriate boundary conditions are

$$\overline{T}'(0)=\text{finite, and } \overline{T}(1)=0. \tag{14}$$

Equation (13) and boundary conditions (14) are formally equivalent to [7] and boundary conditions [8] in my previous paper (Kimura 1980) where the average number of generations until fixation of the mutant allele under irreversible mutation pressure was investigated assuming a single locus in a finite population. In that paper, I assumed the same expression as (9) for the variance ($V_{\delta p}$), but assumed a more general expression than (10), that is,

$$M_{\delta p}=\{sp+h(1-2p)\}p(1-p)+v(1-p) \tag{15}$$

for the mean change. Then, I obtained the solution

$$\overline{T}(p)=4N_e\int_p^1 \exp[-B(\eta)]\eta^{-V}d\eta\int_0^\eta \frac{\exp[B(\xi)]\xi^{V-1}}{1-\xi}d\xi, \tag{16}$$

where

$$B(\xi)=(S/2)\xi^2+H\xi(1-\xi), \tag{17}$$

in which $S=4N_es$, $H=4N_eh$ and $V=4N_ev$. It is important to note here that $\overline{T}(p)$ depends on the products N_es, N_eh and N_ev but not on N_e, s, h and v separately. This solution can be applied directly to our problem since (15) reduces to (10) if we put $s=2s_1$, $h=s_1$ and $v=v_1$. Incidentally, the new selection coefficient s_1 given by (11a) represents a very weak selection, being of the order of mutation rate for $v \ll s'$.

The required solution for our present problem of how long it takes on the average for the double mutant to reach fixation starting from the population consisting exclusively of the wild type is then given approximately by (16) by letting $p=0$, and assuming $S=8N_es_1=4AB/(B+3A)$, $H=S/2$ and $V=4N_ev_1=2A^2/(B+3A)$ in which $A=4N_ev$ and $B=4N_es'$.

In figure 1, the average fixation time ($\overline{T}$), as computed by this approximation method, is illustrated by a solid curve for $4N_es'$ values ranging from 20 to 1000, assuming the mutation parameter $2N_ev=1$, which means that one mutant gene is fed into the population in each generation if the population consisted exclusively of wild type alleles. It is remarkable that the average fixation time here is much shorter than what would be expected from the previous studies (Li and Nei 1977; Kimura 1980) on the average fixation time of a deleterious mutant at a single locus. For example, in the single locus case, if the mutant allele A' is produced irreversibly at the rate v per generation from its wild type allele A and if A' is semidominant with selection coefficient s' against it, the average fixation time is roughly $100N_e$ generations if $4N_es'=9$ and slightly longer than $1000N_e$ generations if $4N_es'=13$, where we assume $2N_ev=1$. These are very much longer than $4N_e$ generation which is the corresponding average fixation time for neutral mutations ($4N_es'=0$) at a single locus. On the other hand, in the present case of compensatory neutral mutations at two loci, the average fixation time of the double mutant $A'B'$ assuming $2N_ev=1$ is as follows. It is only 7·4 N_e generations for $4N_es'=20$, 8·7 N_e generations for $4N_es'=30$, and about 54 N_e generations for $4N_es'=400$. Even for such a large selective disadvantage as $4N_es'=1000$, the average

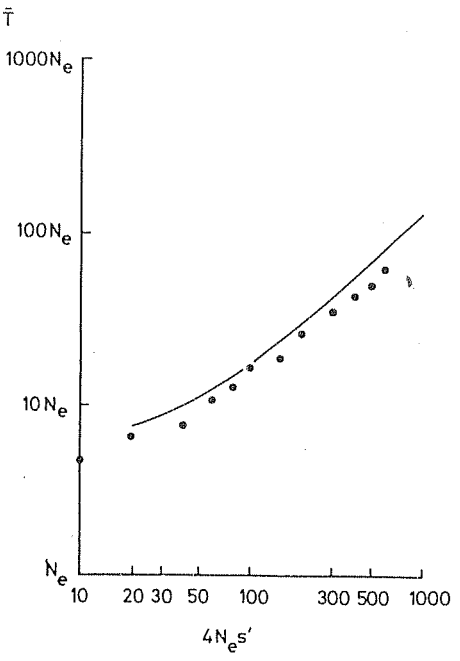

Figure 1. Relationship between $\bar{T}$, the average number of generations until fixation of the double mutant $A'B'$, and $4N_es'$, assuming $2N_ev = 1$, where N_e = effective population size, s' = the selection coefficient against the single mutants and v = mutation rate per locus. The curve represents the theoretical results while dots represent the results of Monte Carlo simulation experiments. For details, see text.

fixation time is about 128 N_e generations which is not an unrealistically long time for the fixation to occur in evolution.

2.2 *Effect of crossing-over in retarding the fixation*

In view of the remarkable phenomenon, as shown above, that rapid fixation of compensatory neutral mutations occurs under complete linkage, it is necessary to know how easily this is disrupted by crossing-over. Although an exact treatment of this problem is very difficult, I have developed a useful approximation method to treat this problem. As I shall demonstrate in the next section, this can predict quite well the simulation results for a low recombination fraction, c, between the loci. The approximation method is based on the premise that for a small c such that

$$v \leqq c \ll s', \tag{18}$$

the effect of crossing-over is mainly to decrease the frequency of the double mutant $A'B'$ through its crossing-over with the wild type AB. Thus we have, as a good approximation,

$$M_{\delta p_2} = vp_1 + s'p_1p_2 - p_0p_2c. \tag{19a}$$

This corresponds to (2a) in the case of no crossing-over. For the change of the frequency

of single mutants, the crossing-over contributes to increasing it by the amount $2cp_0p_2$, so that we have

$$M_{\delta p_1} = 2vp_0 - s'p_1(1-p_1) - vp_1 + 2cp_0p_2, \tag{19b}$$

which reduces to (2b) in the case of no crossing-over ($c = 0$). Then, the assumption of quasi-equilibrium, i.e., $M_{\delta p_1} = 0$, leads to the following expression

$$p_1 = \frac{(2v + 2cp_2)(1-p_2)}{s' + 3v}, \tag{20}$$

where a small term involving cp_2 has been omitted from the numerator. Substituting this in (19a) and writing p for p_2, we get

$$M_{\delta p} = \{s_2 p + h_1(1-2p)\}p(1-p) + v_1(1-p), \tag{21}$$

where

$$v_1 = \frac{2v^2}{s' + 3v}, \tag{22a}$$

$$h_1 = \frac{cv + 2s'v - cs'}{s' + 3v} \tag{22b}$$

and

$$s_2 = \frac{2c^2 + 2cv + 4s'v}{s' + 3v}. \tag{22c}$$

Note that (21) is formally equivalent to (15). Then, (16) can be used to obtain the average time until joint fixation of A' and B', namely, we can obtain from these formulae the approximate value for $\overline{T}$ by putting $p = 0$ and assuming $S = 4N_e s_2 = (2C^2 + 2CA + 4BA)/(B + 3A)$, $H = 4N_e h_1 = (CV + 2BA - CB)/(B + 3A)$ and $V = 4N_e v_1 = 2A^2/(B + 3A)$, in which $A = 4N_e v$, $B = 4N_e s'$ and $C = 4N_e c$.

In figure 2, the analytical results computed by using this approximation method are illustrated by a solid curve for $A = 4N_e v = 2$, $B = 4N_e s' = 40$ and assuming values $C = 4N_e c = 0 \sim 40$ (the dots in the figure represent the corresponding results from the Monte Carlo simulation experiments which I shall explain in the next section).

In figure 3, the average fixation time ($\overline{T}$) is shown by solid curves as a function of $4N_e c$ for various values of $4N_e s'$ ranging from 40 to 1000 (still assuming $2N_e v = 1$). It is interesting to note that the relationship between $\overline{T}$ and $4N_e c$ for a given $4N_e s'$ is almost linear in this range of $4N_e s'$ values. In the same figure, the dotted line is drawn by connecting $4N_e c$ values which retard the average fixation time by a factor of two as compared with the case of no crossing-over. For example, this is $4N_e c \approx 4{\cdot}7$ in the case of $4N_e s' = 1000$, but $4N_e c \approx 7$ when $4N_e s' = 100$.

3. Simulation experiments

In order to check the validity of the above approximation theory, extensive Monte Carlo simulation experiments were performed. Two types of Monte Carlo simulation methods were employed. One follows the standard practice of faithfully sampling $2N_e$ gametes in each generation. The other made use of the improved PSV (pseudo-sampling variable) method, which differs from the original PSV method (Kimura 1980) in that it incorporates a correction for low frequency classes, namely, if, at any generation, one of

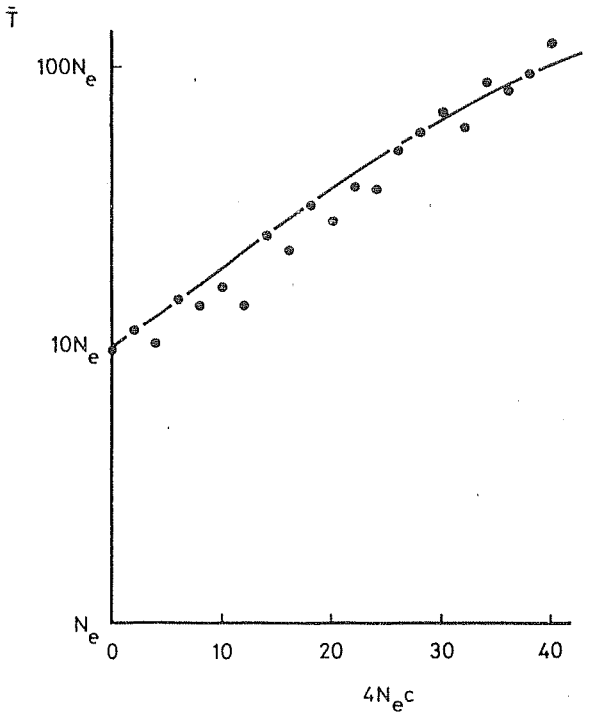

Figure 2. Relationship between the average fixation time ($\bar{T}$) of the double mutant and $4N_ec$, where N_e is the effective population size and c is the recombination fraction between the two loci. The solid curve represents the theoretical results assuming $2N_ev = 1$ and $4N_es' = 40$, where v is the mutation rate for the deleterious allele per locus per generation and s' is the selection coefficient against the individual mutant. The dots represent the corresponding results from Monte Carlo simulation experiments, which were performed to check the validity of the approximation theory that incorporates the effect of low rate crossing-over on retarding the average fixation time of individually deleterious but compensatory neutral mutants. For details see text.

the alleles happens to be represented in the population less than five times, a Poisson random variable is used to sample that allele. In either of these methods, changes of gene frequencies per generation due to mutation and natural selection were made following the standard deterministic theory of population genetics (see, Crow and Kimura 1970). The dots in figure 1 represent the results of Monte Carlo experiments which used the standard type simulation method. Each dot represents the average of 50 replicate trials assuming 250 breeding individuals ($N_e = 250$) with mutation rate $v = 0{\cdot}002$ so that $2N_ev = 1$. These experiments were performed in order to check the validity of the approximation method for the case of complete linkage with $2N_ev = 1$ as shown by the solid curve in the same figure. The agreement between the results of approximate treatment (assuming the quasi-equilibrium) and those of simulation experiments is satisfactory for the range $20 \leqq 4N_es' < 600$. There is, however, a possibility that the approximation method slightly overestimates the true value. Simulation experiments were also performed for $4N_es' = 0$ (completely neutral mutations) and it was found that the average fixation time for the double mutants is about $5N_e$ generations when $2N_ev = 1$.

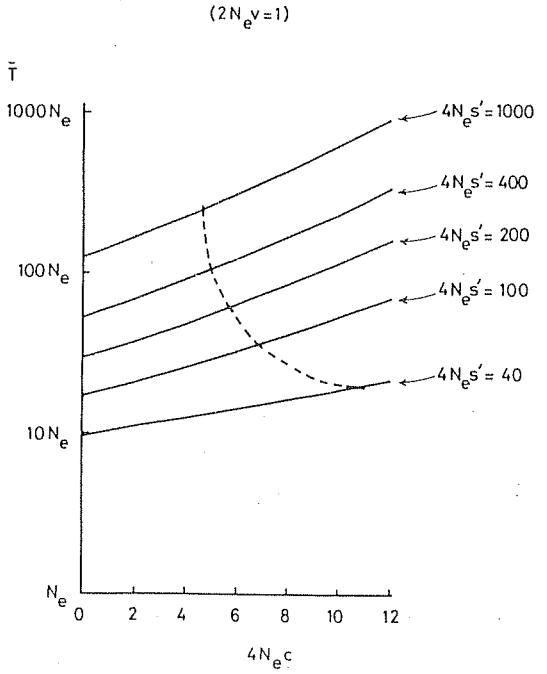

Figure 3. Relationship between the average fixation time ($\bar{T}$) and $4N_ec$, as computed by using the analytical (approximation) theory, is illustrated by solid lines for various values of $4N_es'$. The intersection between the dotted line and each of the solid lines gives the values of $4N_ec$ which retard the average fixation time by a factor of two as compared with the case of no crossing-over ($c = 0$).

The approximate analytical treatment is based on the assumption of quasi-equilibrium, $M_{\delta p_1} = 0$, namely, it is based on the premise that the "moving equilibrium" as given by (6) holds as a good approximation (on the average) in the process of fixation of $A'B'$. To check the validity of this assumption, I examined the quantity

$$\theta = \left(\frac{p_1}{1-p_2}\right) \Big/ \left(\frac{2v}{s'+3v}\right), \tag{23}$$

using the improved PSV method, to see if the average value of θ is reasonably close to unity. It turned out that this is indeed so, although the actual values of θ were often slightly larger than unity, such as, for example, $\theta \approx 1{\cdot}1$. In figure 4 an example of a sample path leading to the fixation of $A'B'$ starting from a very low frequency is illustrated with intervals of 5 generations, assuming the parameters $N_e = 50$, $s' = 0{\cdot}1$ and $v = 0{\cdot}01$, so that $2N_ev = 1$ and $4N_es' = 20$. The broken line represents the line of moving equilibrium as computed by equation (6), and the figure shows that this is a reasonable assumption.

To check the validity of the approximation theory for estimating the effect of low rate crossing-over in retarding the fixation, Monte Carlo simulation experiments were performed using the standard type sampling method. The results are plotted by dots in figure 2. Each dot represents the average of 50 replicate trials assuming the parameter values, $N_e = 100$, $v = 0{\cdot}005$ and $s' = 0{\cdot}1$, so that $2N_ev = 1$ and $4N_es' = 40$. By

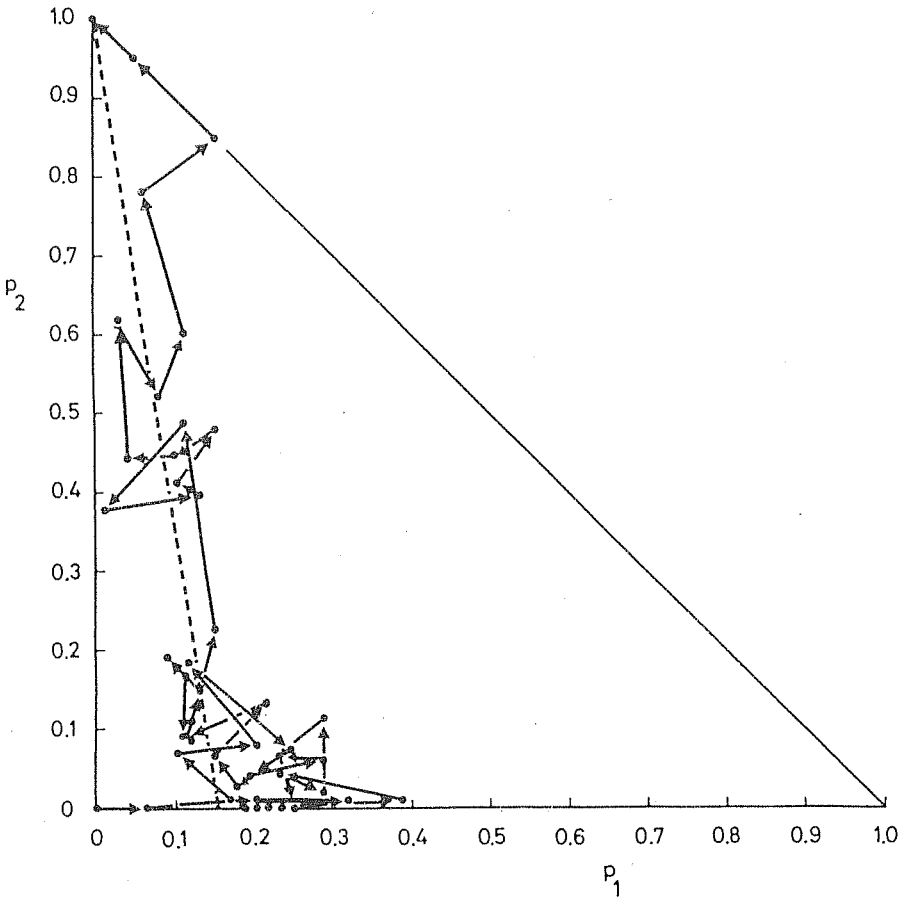

Figure 4. An example of a sample path obtained from the Monte Carlo simulation experiments as illustrated with intervals of 5 generations. Parameter values are $N_e = 50$, $s' = 0{\cdot}1$ and $v = 0{\cdot}01$ so that $2N_e v = 1$ and $4N_e s' = 20$. The broken line represents the quasi-equilibrium or "moving equilibrium" computed by using equation (6). In this figure, the abscissa (p_1) represents the frequency of single mutants ($A'B$ or AB'), while the ordinate (p_2) represents the frequency of the double mutant ($A'B'$).

comparing these results with the corresponding analytical results (solid curve), it is evident that both agree with each other sufficiently well, suggesting the validity of the analytical treatment based on approximations.

4. Discussion

In the preceding sections, I have demonstrated the remarkable property that "compensatory neutral mutations" can easily become fixed in the population by random drift under continued mutation pressure when the genes are tightly linked. By compensatory neutral mutations I mean a pair of mutations at different loci (or sites) which are individually deleterious but restore the normal fitness in combination. A

similar model of individually deleterious but jointly advantageous mutations at two loci was investigated long ago by Haldane (1931). His treatment, however, was deterministic and differs fundamentally from my present treatment which is stochastic.

By using Wright's terminology, it may be said of the above phenomenon of compensatory neutral evolution that, for such a system of genes, a species can readily move from one adaptive peak to a nearby, equally adapted peak passing through a deep valley by mutation pressure and random drift. This resembles Wright's shifting balance process (Wright 1932; see Wright 1977, for review), but, in his theory, intergroup selection plays a crucial role, whereas in our case, very tight linkage is essential.

Whether evolution can really proceed through an intermediate deleterious state is an interesting and provocative question, but, so far, no definite example of it seems to have been reported from classical evolutionary studies.

I now intend to show that compensatory neutral mutations have played an important role in the evolution of protein (and RNA) molecules. Probably the most appropriate circumstance in which this is suggested is the coupled substitutions of amino acids within molecules, which Ohta (1973, 1974) referred to in relation to her hypothesis that very slightly deleterious as well as neutral mutations play an important role in molecular evolution. Wyckoff (1968), in his comparison of rat and bovine pancreatic ribonucleases, noted that "a number of changes are paired." For example, in bovine RNase, amino acid positions 57 and 79 are occupied respectively by valine and methionine, while in rat RNase, these positions are occupied by isoleucine and leucine. What is important is that these two amino acid sites are close to each other in the three-dimensionally folded structure, although they are relatively far apart in the linear sequence.

A similar example was found by Tsukihara *et al* (1982) in their study of [2Fe – 2S] ferredoxins isolated from various plants and algae. According to them, in *Equisetum* (horsetail) species, when two duplicated genes (I and II) of this protein are compared in terms of amino acid sequence, a change from threonine to arginine at position 25 correlates with change from arginine to glutamine at position 42 in the molecule. Again, these two amino acid positions are close to each other in the three-dimensionally folded structure.

That such physical proximity of sites within a folded protein is the basis of compensatory fitness interaction is strongly suggested by the mutation studies of Yanofsky *et al* (1964) on tryptophan synthetase A protein. According to them, position 210 in the wild type protein is occupied by glycine, but, if this is replaced by glutamic acid by mutation, the enzyme becomes nonfunctional. However, if a further change occurs at position 174, changing tyrosine of the wild type to cysteine, the activity of the enzyme is recovered, although a mutation (Tyr → Cys) at the second position alone causes loss of function. By extensive reversion studies of this sort, these authors came to the conclusion that the interacting amino acid sites (as in positions 210 and 174) are close to each other in the folded protein. Similar compensatory changes were also observed in a paired region of t RNA molecules by Ozeki and his associates (Ozeki *et al* 1980) in their study of nonsense suppressor mutants of *E. coli*.

Thus, we can envisage the two sites A and B assumed in our model (see table 1) as representing two amino acid sites within a protein (see figure 5). These two codon positions, being in the same cistron must be very tightly linked, approximating the complete linkage model assumed in table 2. Even if crossing-over tends to interfere with this phenomenon, it is likely that the intra-cistronic crossing-over of 10^{-4} or less per

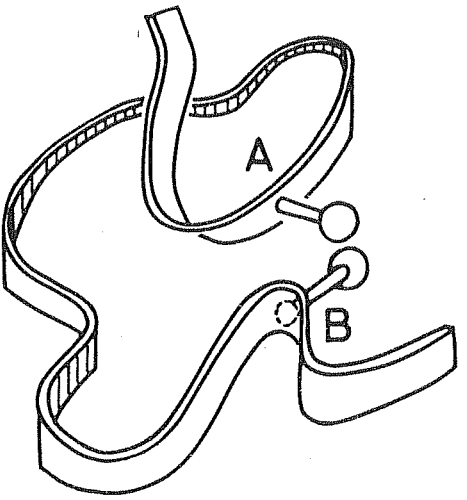

Figure 5. A diagram illustrating two interacting amino acid sites (A and B) within a folded protein whose evolutionary pattern may conform to the model of "compensatory neutral mutations".

generation does not invalidate the present model as applied to interacting sites within a molecule.

I conclude that if this model of compensatory neutral mutations is realistic, it will not only lend support to Ohta's (1973, 1974) concept of very slightly deleterious mutations, but also help us to widen the applicability of the neutral theory for understanding the mechanism of evolution at the molecular level.

Acknowledgement

I thank Drs Tomoko Ohta, James F Crow and Tomitake Tsukihara for stimulating discussions, and Dr. Kenichi Aoki for having gone over the manuscript to suggest improved presentation. I am also grateful to Dr. Ohta for her help in carrying out simulation experiments. This work is supported in part by a grant-in-aid from the Japanese Ministry of Education, Science and Culture.

References

Crow J F and Kimura M 1970 *An introduction to population genetics theory* (New York: Harper and Row)
Haldane J B S 1931 *Proc. Cambridge Philos. Soc.* 27: 137–142
Kimura M 1964 *J. Appl. Probab.* 1: 177–232
Kimura M 1968 *Nature (London)* 217: 624–626
Kimura M 1980 *Proc. Natl. Acad. Sci. USA* 77: 522–526
Kimura M 1983 *The neutral theory of molecular evolution* (Cambridge: University Press)
Kimura M and King J L 1979 *Proc. Natl. Acad. Sci. USA* 76: 2858–2861
Kimura M and Ohta T 1971 *Theoretical aspects of population genetics* (Princeton: University Press)

Kimura M and Ohta T 1974 *Proc. Natl. Acad. Sci. USA* 71: 2848–2852
Li W-H and Nei M 1977 *Genetics* 86: 901–914
Ohta T 1973 *Nature* (*London*) 246: 96–98
Ohta T 1974 *Nature* (*London*) 252: 351–354
Ozeki H, Inokuchi H, Yamao F, Kodaira M, Sakano H, Ikemura T and Shimura Y 1980 in *Transfer RNA: biological aspects* (eds.) D Soll, J Abelson and P Schimmel (Cold Spring Harbor Laboratory) pp. 341–362
Tsukihara T, Kobayashi M, Nakamura M, Katsube Y, Fukuyama K, Hase T, Wada K and Matsubara H 1982 *Biosystems* 15: 243–257
Wright S 1932 *Proc. VI Int. Congr. Genet.* 1: 356–366
Wright S 1977 *Experimental results and evolutionary deductions. Evolution and the genetics of populations* (Chicago and London: University of Chicago Press) Vol. 3, Chap. 13
Wyckoff H W 1968 *Brookhaven Symp. Biol.* 21: 252–257
Yanofsky C, Horn V and Thorpe D 1964 *Science* 146: 1593–1594