Directional mutation pressure and neutral molecular evolution

(Guanine-plus-cytosine content/selective constraints/non-Darwinian evolution)

Noboru Sueoka

Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309

Communicated by Norman H. Horowitz, November 9, 1987 (received for review August 13, 1987)

ABSTRACT A quantitative theory of directional mutation pressure proposed in 1962 explained the wide variation of DNA base composition observed among different bacteria and its small heterogeneity within individual bacterial species. The theory was based on the assumption that the effect of mutation on a genome is not random but has a directionality toward higher or lower guanine-plus-cytosine content of DNA, and this pressure generates directional changes more in neutral parts of the genome than in functionally significant parts. Now that DNA sequence data are available, the theory allows the estimation of the extent of neutrality of directional mutation pressure against selection. Newly defined parameters were used in the analysis, and two apparently universal constants were discovered. Analysis of DNA sequence data revealed that practically all organisms are subject to directional mutation pressure. The theory also offers plausible explanations for the large heterogeneity in guanine-plus-cytosine content among different parts of the vertebrate genome.

Questions on relative roles of Darwinian selection and of non-Darwinian or neutral mutation on molecular evolution have been the focus of discussion for the last two decades between neutralists and selectionists. It is now widely accepted that the neutral concept of evolution propounded by Kimura since 1968 (1, 2) and by King and Jukes since 1969 (3) is as important in evolution as selection is. The neutral theory of evolution is based mainly on two premises: (i) functionally neutral changes of the genome by nucleotide substitution mutation escape the editing effects of Darwinian selection (1, 3) and (ii) these changes are fixed in the population by random genetic drift (1, 2) as formulated by Wright (4). The theory explains the much higher rates of DNA base substitution compared to the mutation rates previously estimated from phenotypically observable mutations (1). The two premises were reasonable; the first one had previously been pointed out (5, 6). These premises are not sufficient, however, to explain a number of the major features of the neutral evolution of the genome. For example, the neutral theory elaborated by Kimura (2) does not explain the main features of DNA base composition: the relatively small heterogeneity within a microbial species (7-10) in contrast to the extremely wide variation, ranging from approximately 25% to 75% guanine-plus-cytosine (G+C) content, among different species of bacteria, algae, and protozoa (11-13). By contrast, the quantitative theory of directional mutation pressure, proposed in 1962 (5, 6), can explain a number of evolutionary changes that have occurred in DNA base composition and nucleotide sequence that are not possible to understand by the principles of the neutral theory alone. Moreover, as shown in this paper, use of parameters previously defined in the directional mutation theory allows us to estimate the degree of neutrality of mutational change in terms of DNA base composition (G+C content) in bacteria or in different genes of individual vertebrates.

Existence of nonrandom overall mutation pressure—namely, directional mutation pressure toward the A pair (A-T or T-A) or toward the y pair (G-C or C-G), was first suspected from the variation of DNA base composition (expressed as G+C content, γ/(α + γ)] among DNAs of different species and heterogeneity within the genome of an individual species of bacteria (9, 10).

Variation of G+C content is also reflected in the total amino acid composition in bacteria and Tetrahymena (14, 15). As predicted, when the genetic code was deciphered, it became evident that the correlation of each amino acid content with DNA G+C content depends on the G+C content of the codon set for the particular amino acid. It is therefore important to note that the directional mutation pressure can exert some direction (nonrandom) changes on the amino acid composition of proteins as well. Results of some recent analyses of DNA sequence and codon usage have been explained by directional mutation pressure (16-23).

Theory of Directional Mutation Pressure. Based on the major features of variation and heterogeneity of DNA base composition, Sueoka (5) formulated a quantitative theory of directional mutation pressure in 1962. In the same year, Freese (6) also proposed a theory to explain some compositional features of DNA. The two theories are similar in principle and conclusions but are different in treatment. The directional mutation theory and rationale for it in the form presented by Sueoka (5) are briefly stated below.

The major cause for a change in DNA G+C content of an organism is the mutation (5) between an α pair and a y pair. When there are no selective constraints,

$$\gamma = \frac{\alpha}{1 - \beta} \left(p \frac{\alpha}{v} (1 - p)\right),$$

where $\rho$ is the fractional G+C content of DNA and $u$ and $v$ are mutation rates per generation per base. At equilibrium (5),

$$\rho = \frac{v}{u + v}, \text{ or } \frac{v}{u} = 1 - \beta.$$  

Here, $\rho$ is the G+C content at equilibrium. The large variation of G+C content among DNAs of different bacteria was explained mainly by differences in $v/u$ rather than by selection (5, 6).

Directional Mutation Pressure. A definition of directional mutation pressure is necessary for the quantitative analysis of the effect of directional mutation on the G+C content of DNA. The directional mutation pressure ($\mu_D$) is now defined as (5)

$$\mu_D = \frac{v}{u + v}.$$
Table 1. Major statistics for the organisms referred to in this article

<table>
<thead>
<tr>
<th>Organism</th>
<th>No. of genes</th>
<th>Avg. no. of codons per gene</th>
<th>G+C content</th>
<th>v/u†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total*</td>
<td>P1</td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>10</td>
<td>308.5</td>
<td>0.42</td>
<td>0.516 ± 0.027</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>110</td>
<td>438.2</td>
<td>0.50</td>
<td>0.607 ± 0.045</td>
</tr>
<tr>
<td><strong>Fungus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>49</td>
<td>333.2</td>
<td>0.35</td>
<td>0.465 ± 0.051</td>
</tr>
<tr>
<td><strong>Protozoan</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hypotrichs</em></td>
<td>3</td>
<td>295.3</td>
<td>0.42</td>
<td>0.518 ± 0.096</td>
</tr>
<tr>
<td><strong>Invertebrates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>8</td>
<td>486.5</td>
<td>0.36</td>
<td>0.579 ± 0.096</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em></td>
<td>11</td>
<td>272.3</td>
<td>0.40</td>
<td>0.548 ± 0.047</td>
</tr>
<tr>
<td><em>Sea urchin</em></td>
<td>8</td>
<td>148.3</td>
<td>0.42</td>
<td>0.581 ± 0.053</td>
</tr>
<tr>
<td><strong>Vertebrates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Xenopus laevis</em></td>
<td>5</td>
<td>126.6</td>
<td>0.42</td>
<td>0.567 ± 0.018</td>
</tr>
<tr>
<td><em>Chicken</em></td>
<td>28</td>
<td>257.6</td>
<td>0.44</td>
<td>0.588 ± 0.060</td>
</tr>
<tr>
<td><em>Mouse</em></td>
<td>31</td>
<td>302.4</td>
<td>0.44</td>
<td>0.549 ± 0.049</td>
</tr>
<tr>
<td><em>Cat</em></td>
<td>43</td>
<td>247.8</td>
<td>0.43</td>
<td>0.560 ± 0.065</td>
</tr>
<tr>
<td><em>Human</em></td>
<td>90</td>
<td>311.4</td>
<td>0.43</td>
<td>0.552 ± 0.061</td>
</tr>
<tr>
<td><strong>Plant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>8</td>
<td>269.3</td>
<td>0.46</td>
<td>0.616 ± 0.039</td>
</tr>
</tbody>
</table>

Unless otherwise stated, codon-frequency data compiled by Maruyama et al. (27) were used. Each value for the three codon positions represents the average G+C content of genes with standard deviation.

*G+C content of the whole genome generally accepted (28).

†Ratio of mutation rates, calculated by Eq. 4b.

‡Ciliates (data collection assisted by David Prescott).

§Nematodes (data collection assisted by Thomas Blumenthal).

¶Different species are combined (27).

Here, μP2 > 0.5 indicates that the mutation pressure favors γ over α, and μP3 < 0.5 indicates a preference for α over γ. At equilibrium, the G+C content of a neutral nucleotide position equals mutation pressure μP2 (5); β is therefore termed the G+C content at "mutational equilibrium."

Analysis of directional mutation pressure is based on the relative value of mutation rates, u and v, not on their absolute values. Since the probability of fixation in the population is similar for individual neutral mutations, the values of v/u and μP2 also apply to the evolution (change in population) of P3, as far as P3 is in equilibrium with directional mutation pressure. On the other hand, the fixation of P12 and P1·23 in a population is subject to selective constraints, the extent of which will be examined in this paper, assuming that P12 and P1·23 are values in equilibrium with directional mutation pressure and selective constraints. In other words, the rate of directional change of DNA G+C content is not our concern in the present analysis.

In 1965, Cox and Yanofsky (24) reported experimental results indicating that the directional mutation pressure indeed changes the G+C content of *Escherichia coli*. They used a mutant strain, mutT (25); mutT enhances mutation rate 103-fold over the spontaneous level and all mutational changes due to mutT are A+T to C+G (26). The continuous culture of the mutT strain for 1200–1600 generations showed changes in G+C content by 0.2–0.5%.

**G+C Content of the Three Codon Positions.** In the present analysis, observed G+C contents of the first, second, and third codon positions (P1, P2, and P3, respectively) are corrected average G+C contents of the three codon positions that are calculated from 56 triplets out of 64. Because of the inequality of α and γ at the third codon position, the three stop codons (TAA, TAG, and TGA) and the three codons for isoleucine (ATT, ATC, and ATA) were excluded in calculation of P3, and two single codons for methionine (ATG) and tryptophan (TGG) were excluded in all three (P1, P2, and P3). The G+C contents are calculated for individual genes or for the sum of genes of an organism (Table 1). Unless otherwise specified, we analyze the data of codon-usage frequency recently calculated by Maruyama et al. (27) from the sequence data of GenBank* (29).

**Analysis of Equilibrium.** The G+C content of the third codon position (P3) was used for estimating the directional mutation pressure (μP3) in the present work. Because of the partially silent nature of the third codon position (30), P3 represents one of the most neutral nucleotides within the genome, so far as the G+C content is concerned (8, 31). The marked response of P3 to directional mutation pressure is also manifested in its extreme values in bacteria (9% and 95%; Fig. 1a), in some genes of vertebrates (95%; refs. 32 and 33), and in mitochondrial genes of *Drosophila yakuba* (6.2%; ref. 19). These results indicate close neutrality of the third codon positions in G+C content. The near neutrality of P3 does not mean that any change of the third codon position is neutral. Obviously, some transversions are not neutral, since they cause amino acid changes. However, transitions should be sufficient to bring P3 into equilibrium with μP3. Nevertheless, changes in P3 may not be completely neutral. If an independent method to assess the extent of neutrality of P3 becomes available in the future, small corrections may become necessary.

Unlike P3, P1 and P2 are subject to functional constraints against change because a mutation at these positions usually leads to an amino acid change, except between some codons of arginine, leucine, or serine. Other regions such as introns and intergenic spacers are abundant only in eukaryotes and their function is not well understood and may differ widely in their degree of functional neutrality.

Assuming that P3 is neutral with regard to directional mutation pressure and that a near equilibrium has been

*EMBL/GenBank Genetic Sequence Database (1985) GenBank (Bolt, Beranek, and Newman Laboratories, Cambridge, MA), Tape Release 38.0.*
attained for \( P_3 \), currently the best estimate of \( \mu_D \) and \( v/u \) can be obtained by approximating \( \hat{p} \) with \( \hat{P}_3 \) in Eqs. 2 and 3:

\[
\mu_D = \hat{P}_3 \tag{4a}
\]
\[
v = \frac{\hat{P}_3}{1 - \hat{P}_3} \tag{4b}
\]

where \( \hat{P}_3 \) is the G + C content of the third codon position at equilibrium with the directional mutation pressure. In reality, \( \hat{P}_3 \) can be approximated by \( P_3 \). In this article, \( P \) values with circumflex indicate values at equilibrium either with mutation alone (\( \hat{P}_3 \)) or with both mutation pressure and selection (\( P_1, P_2, P_12, \) and \( P_{123} \)).

**Extent of Directional Mutation Pressure and Equilibrium.** In order to analyze the relation between directional mutation pressure and selective constraints, the data used by Muto and Osawa (16) are presented in Fig. 1a with some modification. The G + C content of the first and the second codon positions (ordinate) are replotted against the G + C content of the third codon position (\( P_3 \); abscissa), instead of against the total DNA G + C content as originally reported (16). The total G + C content is not an ideal variable for quantitative arguments because it includes all three letters of the codon and also other noncoding areas. Their data on the coding areas and spacer regions between genes of the same set of bacteria are also replotted against \( P_3 \) (Fig. 1b).

As Muto and Osawa (16) pointed out, it is evident that bacteria with lower and higher values of G + C content have more extreme values in the third codon position than the total G + C content or than those of the first and second codon positions (Fig. 1a). It is also noted that even among bacteria whose DNA G + C content is well within the limits of 25% and 75% G + C, there is the tendency that species with a high G + C content have an even higher G + C content in the third codon position (G + C pressure) and species with a high A + T content have an even higher A + T content (A + T pressure). The consistently higher G + C content and the slope of the first codon position compared with the second come from the fact that more abundant amino acids have more \( \gamma \) pairs at the first codon positions than at the second codon positions (3, 14, 33).

The situation is shown diagrammatically in Fig. 2. The extent of discrepancy in G + C content between the third codon position and the average of first and second codon positions (\( P_{12} \)) (the average line in Figs. 1a and 2), is proportionately larger on both sides of the intersect (\( E_{p} \)), as indicated by the two thick open arrows, one toward higher A + T content and the other toward higher G + C content. This discrepancy is interpreted to mean that the directional mutation pressure is counteracted by selective constraints, thus forming a new type of equilibrium. As the parameter that represents the overall situation of G + C equilibrium, the mutation-selection equilibrium coefficient \( \epsilon \) is defined as the regression coefficient against \( P_3 \). The equilibrium coefficient \( \epsilon \) is 0 for no effect of directional mutation pressure (complete selective constraint) and 1 for the complete equilibrium with \( \mu_D \) (complete neutrality). Thus, \( \epsilon_{12} \) represents the extent of equilibrium of \( P_{12} \) by directional mutation pressure and selective constraints. The values of \( \epsilon \) and \( 1 - \epsilon \) are convenient measures of relative effects of neutrality and selective constraints, respectively.

**Mutation-Selection Equilibrium.** With the assumption of neutrality of \( P_3 \) and equilibrium between the directional mutation pressure and selective constraints, the relationship between \( P_3 \) and \( \hat{P}_{12} \) presented in Fig. 2 can be expressed as

\[
\hat{P}_{12} = E_p + \epsilon_{12}(P_3 - E_p), \tag{5a}
\]

or

\[
\hat{P}_{12} = \epsilon_{12}\hat{P}_3 + (1 - \epsilon_{12})E_p. \tag{5b}
\]

At the equilibrium point \( E_p \), the average \( \hat{P}_{12} \) of \( P_1 \) and \( \hat{P}_2 \) equals \( \hat{P}_3 \); note that if \( \epsilon \) equals 0 (complete selective constraint), \( \hat{P}_{12} \) equals \( E_p \) independent of \( P_3 \) or \( \mu_D \) (Eq. 5b). Note that exactly the same \( E_p \) value is applicable to \( P_{12} \) as well as the whole protein-coding area, including all three codon positions (\( P_{123} \); see Table 2 legend). \( E_p \) thus represents the proportion of G + C nucleotides in positions 1 and 2 or for the complete equilibrium.

![Fig. 1](image1.png)

**Fig. 1.** Regression of G + C content of the first and second positions with the third codon position among bacterial species. (a) Regression lines were drawn by linear least-squares analysis. G + C contents of the first codon position (\( \circ \)), those of the second codon position (\( \triangle \)), and those of the average of the first and second codon positions (\( \bullet \)) are shown. Data compiled by Muto and Osawa (16) were replotted against the G + C content of the third codon position (\( P_3 \)). Because of large differences in the number of genes for different bacteria, regression analysis was performed using the average value for each species without weighting by the number of genes. Regression coefficients (or \( \epsilon \) values) and their standard deviations for the first codon position, the second codon position, and the average of the two are 0.374 ± 0.013, 0.163 ± 0.012, and 0.269 ± 0.011, respectively; (b) Data of Muto and Osawa (16) on the G + C content of coding areas (\( \circ \)) and noncoding areas (spacers, \( \times \)) were replotted against the G + C content of the third codon position of bacteria. Regression coefficients for the coding areas and the spacers are 0.53 ± 0.02 and 0.66 ± 0.01, respectively.

![Fig. 2](image2.png)

**Fig. 2.** Diagrammatic representation of the equilibrium coefficient, \( \epsilon_{12} \), the equilibrium point (\( E_p \)), and the minimal (\( A_{\text{min}} \)) and maximal (\( A_{\text{max}} \)) values of the average G + C content of the first and second codon positions. Solid line represents the regression line of the average G + C content of the first and the second codon positions of various bacteria (Fig. 1a). Broken line is drawn parallel to the regression line through the origin to visualize the equilibrium coefficient, \( \epsilon_{12} \) (degree of neutrality), and the disequilibrium coefficient, 1 - \( \epsilon_{12} \) (degree of selective constraints). Open arrows indicate deviation of the regression line from complete equilibrium with the directional mutation pressure, \( \mu_D = v/(u + v) \) (dotted line).
Table 2. Relative contributions of neutrality (\(e \)) for \(P_{12} \) and \(P_{123} \), \(E_p \) common for \(P_{12} \) and \(P_{123} \), and \(A_{min} \) and \(A_{max} \) for \(P_{123} \) for coding areas of various organisms

<table>
<thead>
<tr>
<th>System</th>
<th>Neutrality, *%</th>
<th>(E_p )</th>
<th>(A_{min} ) for (P_{123} ), %</th>
<th>(A_{max} ) for (P_{123} ), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial species†</td>
<td>(26.3 \pm 6.6 )</td>
<td>0.464 ± 0.076</td>
<td>22.8 ± 4.6</td>
<td>73.6 ± 4.6</td>
</tr>
<tr>
<td>Vertebrates‡</td>
<td>(50.8 \pm 4.4 )</td>
<td></td>
<td>20.7 ± 4.6</td>
<td>77.5 ± 4.6</td>
</tr>
<tr>
<td>Xenopus</td>
<td>(19.3 \pm 12.9 )</td>
<td>0.473 ± 0.114</td>
<td>25.4 ± 1.5</td>
<td>71.6 ± 1.5</td>
</tr>
<tr>
<td>Chicken</td>
<td>(15.4 \pm 5.0 )</td>
<td>0.464 ± 0.044</td>
<td>26.2 ± 2.3</td>
<td>69.8 ± 2.3</td>
</tr>
<tr>
<td>Mouse</td>
<td>(27.4 \pm 5.4 )</td>
<td>0.442 ± 0.051</td>
<td>21.4 ± 2.8</td>
<td>73.0 ± 2.8</td>
</tr>
<tr>
<td>Rat†</td>
<td>(22.6 \pm 5.8 )</td>
<td>0.428 ± 0.051</td>
<td>22.1 ± 2.6</td>
<td>70.5 ± 2.6</td>
</tr>
<tr>
<td>Human</td>
<td>(19.4 \pm 2.0 )</td>
<td>0.447 ± 0.018</td>
<td>24.0 ± 2.3</td>
<td>70.3 ± 2.3</td>
</tr>
</tbody>
</table>

Coding areas represent the areas of DNA that encode amino acid sequences, including all three codon positions. For a visual presentation of these parameters, see Fig. 2.

*\(e_{123} \) is the equilibrium coefficient for the coding area. Relative effects are expressed in percent neutrality (\(e_{123} \)) and selective constraints (1 - \(e_{123} \)) on the effective directional mutation pressure (\(P_{12} - E_p \)) (Fig. 2; Eq. 5a).

\(E_p \) values were calculated from the regression of \(P_{12} \) or \(P_{123} \) (\(P_{12} + P_{12} + P_{12} \)) against \(P_{12} \) without weighting by the number of genes for each species of bacteria and without weighting by the number of codons for each gene for vertebrates. The \(E_p \) value for the whole coding area (\(P_{123} \)) is the same as the \(E_p \) value for \(P_{12} \), because at \(E_p \) for \(P_{12} \), \(P_{12} = P_{12} \), and consequently, at \(E_p \), \(P_{123} = (2P_{12} + P_{12})/3 = P_{12} \), which is the same for \(P_{12} \). Standard deviations for \(E_p \) were calculated as \((1 - \epsilon)^{-1} Var(A_{min}) + (A_{min})^2 (1 - \epsilon)^{-1} Var(e)^{-1} \), where \(Var(A_{min}) \) and \(Var(e) \) are variances for \(A_{min} \) and for \(e \), respectively (see ref. 34).

†Data compiled by Muto and Osawa (16) were used for calculation (Fig. 1b).

‡Genes of each vertebrate (27) were used for calculation.

A parotid gland proline-rich protein gene was eliminated from the calculation.

all three codon positions; where \(P_{12} \) and \(P_{123} \) are at equilibrium with directional mutation pressure (\(\mu_{DP} \)) and are not subject to selective constraints. The \(E_p \) values calculated for different bacteria and the \(E_p \) values for different genes of individual vertebrates are similar (Table 2). \(E_p \) may, therefore, be a universal constant for \(P_{12} \) and for \(P_{123} \) (the entire coding area).

**Heterogeneity of \(P_3 \) Within Vertebrate Species.** As is evident from Fig. 3 and Tables 1 and 2, the intraspecific

![Fig. 3.](image)

**DISCUSSION**

Both the directional mutation theory and the neutral theory base their arguments on the selectively neutral changes by mutation that play a major role in changing DNA base sequence. In that sense, the two theories belong to the same category. The most fundamental difference between the two theories lies in the treatment of the mutation effect itself. The directional mutation theory treats mutations as bidirectional events as indicated in Eqs. 1 and 2. In addition, the theory incorporates the effect of selective constraints as an integral part. For each mutational event, the principle of random drift should be applicable. In the neutral theory, however, directionality and uniformity of mutation within species or within a domain are not taken into account.

The three constants \(e_{12} \), \(e_{123} \), and \(E_p \) appear to be universal as shown by the fact that, in the coding areas, they are similar both among different bacterial species and among different genes of individual vertebrates (Table 2).
Evolution: Sueoka

![Figure 4](image)

FIG. 4. Combined distribution of $P_{12}$ of yeast, nematode, and Drosophila. Distributions of $P_{12}$ values of yeast ($\circ$), nematode ($\times$), and D. melanogaster ($\triangle$) are combined. $e_{12}$, $e_{123}$, and $E_p$ values calculated by the least-squares method without weighting each datum point with the number of codons are $0.137 \pm 0.003$, $0.425 \pm 0.020$, and $0.428 \pm 0.033$, respectively. The point marked with an arrow represents the average value of collagen genes (col-1 and col-2) of C. elegans (37); this point was eliminated from the calculation of the regression coefficient.

The constancy of these values suggests that $e_{12}$, $e_{123}$, and $E_p$ reflect the nature of the code and the necessary physicochemical constraints of proteins to be stable and functional in the cytoplasm. This point is further supported by the data on yeast, nematode, and Drosophila presented in Fig. 4, where combination of the data for the three organisms forms another set of $e_{12}$ ($0.137 \pm 0.003$), $e_{123}$ ($0.425 \pm 0.020$), and $E_p$ ($0.428 \pm 0.033$) that are within the range found in bacteria and vertebrates (Table 2). The extreme values (25% and 75%) in the G+C content of the coding area in bacteria, protozoa, and algae (10–13) are those corresponding to $\mu_{G+C} = 0$ and $\mu_{G+C} = 1$. This can be explained as the natural consequence of the universal values of $e_{123}$ and $E_p$ and the relatively small proportion of noncoding areas in these organisms (Fig. 1).

To explain the G+C heterogeneity, Cox (38) wrote in 1972 that mutation rates may vary in different parts of the chromosome and the genes may be organized on the chromosome by selection according to the extent of their functional constraints. Two plausible explanations based on directional mutation pressure for the wide intraspecific heterogeneity of $P_3$ among proteins of higher vertebrates are as follows. (i) There might be several different directional mutation pressures (not necessarily mutation rates) in different locations on the genome, and the cause for this difference might reside in the local structural elements of the chromatin. Thus, major mutagenic events (DNA replication and repair) may act differently in different domains. (ii) DNA replication and DNA repair synthesis may make replication errors differently and the extent of DNA repair synthesis may vary among various domains of the chromatin because of the different susceptibility of DNA to damage and repair due to differences in chromatin structure. Clarification of chromosome domains is an important future consideration. Whether or not domains are equivalent to loop structures of the chromatin (39) remains to be seen.

I am grateful to those people who have helped me by supplying data as acknowledged in this article and to Shawn Elliott, Charles Hambleton, and Todd Devine for their patient assistance in data calculation. I am much indebted also to those in my laboratory and in the Department for their help in improving the manuscript. Constructive comments by Drs. James F. Crow and Norman H. Horowitz on this paper are gratefully acknowledged. This work came out of the research supported by a National Science Foundation Grant G-15080 (to N.S.) during the period 1959–1962.