Genet. Res., Camb (1992), **60**, *pp.* 61–67 *With 5 text-figures Printed in Great Britain*

The role of DNA replication and isochores in generating mutation and silent substitution rate variance in mammals

ADAM EYRE-WALKER

Institute of Cell Animal and Population Biology, University of Edinburgh, Edinburgh, EH9 3JT, Great Britain

(*Received 23 October 1991 and in revised form 26 February 1992*)

Summary

It has been suggested that isochores are maintained by mutation biases, and that this leads to variation in the rate of mutation across the genome. A model of DNA replication is presented in which the probabilities of misincorporation and proofreading are affected by the composition and concentration of the free nucleotide pools. The relationship between sequence $G+C$ content and the mutation rate is investigated. It is found that there is very little variation in the mutation rate between sequences of different $G+C$ contents if the total concentration of the free nucleotides remains constant. However, variation in the mutation rate can be arbitrarily large if some mismatches are proofread and the total concentration of free nucleotides varies. Hence the model suggests that the maintenance of isochores by the replication of DNA in free nucleotide pools of biased composition does not lead *per se* to mutation rate variance. However, it is possible that changes in composition could be accompanied by changes in concentration, thus generating mutation rate variance. Furthermore, there is the possibility that germ-line selection could lead to alterations in the overall free nucleotide concentration through the cell cycle. These findings are discussed with reference to the variance in mammalian silent substitution rates.

1. Introduction

There is considerable variation in the rate of silent substitution within all mammalian species so far studied (Li *et al.* 1987, Wolfe *et al.* 1989; Bulmer *et al.* 1991). Although selection is known to act at silent sites in a number of organisms (review by Sharp, 1989) there is little evidence of it doing so in mammals (Eyre-Walker, 1991). This suggests that the variation in silent substitution rate is caused by differences in the rate of mutation between genes.

Recently two mechanisms by which the mutation rate might come to vary across the genome have been proposed (Filipski 1988; Wolfe *et al.* 1989; Wolfe, 1991). Filipski (1988) suggested that the silent substitution rate variance was caused by differences in the level of repair between genes. Such variation in repair across the genome has been observed for various types of DNA damage, but not for base mismatches (Bohr *et al.* 1987).

The idea of Wolfe *et al.* (1989) came out of attempts to explain the isochore structure of the vertebrate genome (review by Bernardi 1989). They noted three observations: that the rate and pattern of mutation during DNA replication was dependent upon the free nucleotide concentrations (Phear and Meuth, 1989*a*, *b*; Meuth, 1989; Kohalmi *et al.* 1991), that the relative concentrations of the free nucleotides varied through the cell cycle (McCormick *et al.* 1983; Leeds *et al.* 1985) and that different DNA sequences were replicated at different times. As a consequence early and late replicating DNA should come to differ in their $G+C$ contents, and isochores (blocks of DNA with homogeneous $G+C$ content) should form. Wolfe *et al.* (1989) further suggested that isochores differ in their mutation rate as a result of this process.

The possible involvement of isochores and their maintenance in the generation of mutation rate differences is suggested by the fact that much of the variation in silent substitution rate can be explained by $G+C$ content (Filipski, 1988; Wolfe *et al.* 1989, Ticher & Graur, 1989; Bulmer *et al.* 1991), and that the $G+C$ contents of genes are correlated to that of the isochore in which they reside (Bernardi *et al.* 1985; Ikemura & Aota, 1988; Aissani *et al.* 1991). In quantitative terms there is approximately a twofold variation in silent substitution rate explained by $G+C$ content, with the overall variation being somewhat larger.

This paper considers whether the replication of DNA in free nucleotide pools of different composition and concentration produces variation in the rate of

mutation. A recent theoretical analysis of this problem by Wolfe (1991) suggested that this was indeed generally the case. In this paper his model is extended to more fully quantify the variation in mutation rate produced. The effects of changing the overall concentrations of free nucleotides are also investigated.

2. The Model

Let us consider a very general model of DNA replication, in which a free nucleotide collides with the DNA polymerase and is then either incorporated into the growing DNA sequence or rejected. The probability that the base, z (where z can be A, T, C or G), collides with the polymerase is P_z, the relative concentration of the free tri-phosphate nucleotide dZTP. Given that a collision has occurred let the probability that the nucleotide is subsequently incorporated be j if the nucleotide is correct and k_a if it is incorrect (where a is i for a type I mismatch and ii for a type II mismatch). Type I mismatches are those which if left unproofread and unrepaired give mutations altering $G+C$ content: i.e. $C \leftrightarrow T$ and $A \leftrightarrow G$ transitions, and $C \leftrightarrow A$ and $G \leftrightarrow T$ transversions. If type II mismatches are left unaltered they give $C \leftrightarrow G$ and $A \leftrightarrow T$ mutations. The frequency with which y is misincorporated instead of z is then:

$$T_{zy} = k_i P_y / (jP_z + k_i P_m + k_i P_y + k_{ii} P_n)$$

Type I mutation, (1*a*)

$$T_{zy} = k_{ii} P_y / (jP_z + k_i P_m + k_i P_n + k_{ii} P_y)$$

Type II mutation, (1*b*)

where bases m and n also form mismatches. Under biologically realistic conditions where the probability of incorporating a mismatch is very small (i.e. $k_a \ll j$) equations 1*a* and 1*b* reduce to the form used by Wolfe (1991): $T_{zy} = \alpha P_y / P_z$ where α is a constant. It would of course be possible to further divide type I mutations into transitions and transversions, and to treat them separately. However, transitions and type I transversions turn out to have the same dynamics in this model, and were thus treated together.

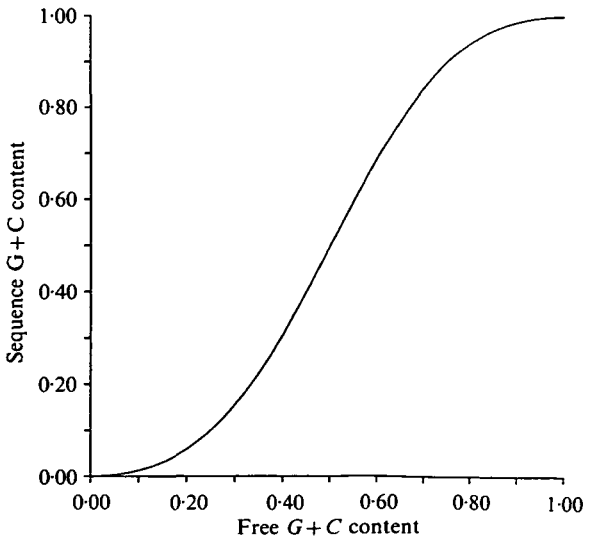

Fig. 1. The equilibrium sequence $G+C$ content.

Once a mismatch is incorporated it may be removed by proofreading or mismatch repair. We will only consider elimination by proofreading in this model since the interaction between repair and replication is likely to be non-trivial. Let the probability of not proofreading a mismatch be N_a. Then the probability with which z will mutate to y per cycle of replication will be:

$$M_{zy} = T_{zy} N_a \quad \text{where } a \text{ is } i \text{ or } ii \text{ as appropriate.} \quad (2)$$

(i) *Equilibrium G+C content*

Now consider the change in the frequency of a nucleotide in a sequence each time the sequence is replicated. For instance the change in the frequency of G:

$$\Delta f_G = -f_G(M_{GC} + M_{GA} + M_{GT}) + f_C M_{CG} + f_A M_{AG} + f_T M_{TG}, \quad (3)$$

where f_z is the frequency of nucleotide z in the sequence being replicated.

Let us make the simplifying assumption that the concentrations of dCTP and dGTP are the same (i.e. $P_C = P_G$) and the $P_T = P_A$. Quite clearly $f_C = f_G$ and $f_T = f_A$, so $P_A = (1-2P_G)/2$ and $f_A = (1-2f_G)/2$. If we let $p = P_G$ and $f = f_G$ then (3) simplifies to

$$\Delta f = 2k_i N_i \left[\frac{f(2p-1)}{2pB + 2k_i} + \frac{p(2f-1)}{2pB - j - k_{ii}} \right], \quad (4)$$

where $B = j - 2k_i + k_{ii}$. Solving $\Delta f = 0$ to get the equilibrium frequency of G (or C) in a sequence, gives under biologically realistic conditions (i.e. $k_a \ll j$),

$$f = 2p^2/(8p^2 - 4p + 1). \quad (5)$$

So the equilibrium frequency of G (or C) in the sequence is independent of proofreading and the probabilities of incorporation; essentially because of the symmetry in the model. As expected, when the free nucleotide pool is either all $A+T$, all $G+C$ or half and half ($p = 0$, $\frac{1}{2}$ and $\frac{1}{4}$) the equilibrium sequence $G+C$ content is equal to the pool $G+C$ content. Figure 1 shows the equilibrium frequency of $G+C$ plotted against the free nucleotide concentration of $G+C$. The relationship between the two variables is sigmoidal, so that at intermediate $G+C$ contents small changes in the free nucleotide concentrations have large effects on the equilibrium $G+C$ content of the sequence (e.g. a sequence of 80% $G+C$ is replicated in a pool of only 70% $G+C$). This non-

https://doi.org/10.1017/S0016672300030676 Published online by Cambridge University Press

linearity arises because the probability of misincorporating a nucleotide is dependent on the probability of a nucleotide being incorrect per collision, and the number of collisions that occur, both of which are dependent upon the pool composition.

(ii) *Proofreading*

The probability that a mismatch will be proofread depends on how long it takes to replicate the next position in the sequence. Once replication of the distal base has occurred the mismatch cannot be proofread, and must instead be corrected by other mechanisms which we shall not consider here. Let the average probability of proofreading a mismatch between collisions be V_a (where a is i or ii for proofreading type I and type II mutations respectively); and let us imagine that the polymerase is waiting to replicate nucleotide z distal to a mismatch. Assuming that proofreading cannot occur until the polymerase is ready to incorporate the next nucleotide, the probability that neither replication not proofreading has occurred after t collisions is $(1-V_a)^{t+1}(1-jP_z)^t$. So the probability that proofreading never occurs when a mismatch is followed by z is

$$jP_z(1-V_a)\sum_{t=0}^{\infty}((1-V_a)(1-jP_z))^t = \frac{P_z}{E_a+P_z}, \quad (6)$$

where $E_a = V_a/(j(1-V_a))$. Therefore the average probability of not repairing a mismatch is

$$N_a = \sum \frac{f_z P_z}{E_a+P_z} \quad (7)$$

as given by Wolfe (1991). In a sequence at equilibrium this becomes

$$N_a = \frac{p(2p-1)(8p^2-4p+1)+E_a(12p^2+6p-1)}{(8p^2-4p+1)(p(2p-1)+E_a(2E_a-1))} \quad (8)$$

E_a is a measure of the proofreading stringency. When there is no proofreading $E_a = 0$, and when proofreading is stringent $E_a \to \infty$. It is worth noting here that as proofreading becomes very stringent (i.e. $E_a \to \infty$)

$$(1+4E_a)N_a \to (24p^2-12p+2)/(8p^2-4p+1). \quad (9)$$

(iii) *Mutation rate*

The average mutation rate per nucleotide per replication of a sequence is

$$\mu = f_G \sum_{z \neq G} M_{Gz} + f_C \sum_{z \neq C} M_{Cz} + f_A \sum_{z \neq A} M_{Az} + f_T \sum_{z \neq T} M_{Tz}, \quad (10)$$

which for a sequence at equilibrium simplifies under biologically realistic conditions (i.e. $k_a \ll j$), to

$$\mu = \underbrace{\frac{8p(1-2p)(k_i N_i)}{j(8p^2-4p+1)}}_{\text{Type I mutations}} + \underbrace{\frac{k_{ii} N_{ii}}{j}}_{\text{Type II mutations}}. \quad (11)$$

Since we are interested in the relative, rather than absolute, mutation rate, let us divide μ by the rate of mutation in an equilibrium sequence of 50% $G+C$ content: i.e. $2k_i/(1+4E_i)+k_{ii}/(1+4E_{ii})$. The relative rate of mutation is

$$R = \frac{w_i(1+4E_i)4p(1-2p)N_i}{8p^2-4p+1} + w_{ii}(1+4E_{ii})N_{ii}, \quad (12)$$

where

$$w_i = \frac{2k_i/(1+4E_i)}{2k_i/(1+4E_i)+k_{ii}/(1+4E_{ii})},$$

and

$$w_{ii} = \frac{k_{ii}/(1+4E_i)}{2k_i/(1+4E_i)+k_{ii}/(1+4E_{ii})},$$

w_i and w_{ii} are the proportion of mutations which are of type I and type II in a sequence of 50% $G+C$ content.

3. Results

(i) *Type II mutations*

Let us consider the frequency of type II mutations alone. By setting $k_i = 0$ in equation 12 we obtain

$$R_{ii} = N_{ii}(1+4E_{ii}) \quad (13)$$

an expression which is solely dependent upon p, the concentration of dGTP or dCTP, and E_{ii} a measure of the proofreading stringency. R_{ii} is plotted against the equilibrium $G+C$ content of a sequence in figure 2 (dashed line) for various levels of proofreading. Remember that as proofreading becomes stringent ($E_a \to \infty$) the expression $(1+4E_a)N_a$ becomes independent of E_a (see equation 9).

When there is no proofreading sequences of all $G+C$ contents have the same rate of type II transversion mutations. However as the stringency of proofreading increases so sequences of extreme $G+C$ content have higher mutation rates than sequences of intermediate $G+C$ content. The reason: at extreme $G+C$ contents the polymerase only has to try on average two nucleotides before the correct one is found, compared to the four that must be tested at intermediate $G+C$ contents. Therefore, the probability of not proofreading (when proofreading is stringent) at extreme $G+C$ contents is twice that at intermediate $G+C$ contents. This means the mutation rate is twice as great at extreme $G+C$ contents.

https://doi.org/10.1017/S0016672300030676 Published online by Cambridge University Press

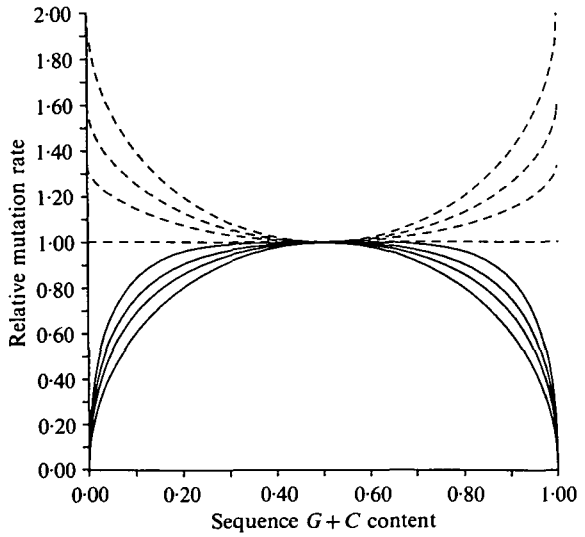

Fig. 2. The effect of proofreading on the relative rates of type I and type II mutations. Figure shows the relative mutation rates for type I (solid lines) and type II (dashed lines) mutations for various levels of proofreading. In each case from bottom to top E_a, the strength of proofreading, is 0, 0·25, 0·75 and ∞. Under these values of E_a the probability of proofreading a mismatch in a sequence of 50% $G+C$ content is 0, 0·5, 0·75 and $\rightarrow 1$ respectively.

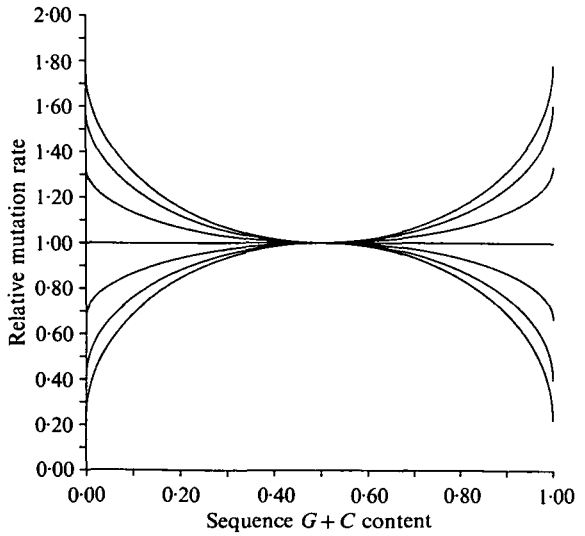

Fig. 3. The relative mutation rate for a sequence undergoing unproofread type I mutation and stringently proofread type II mutation. The curves represent different ratios of type I and type II mutations. The ratio of unproofread type I mutations to proofread type II mutations in a sequence of 50% $G+C$ content is, from top to bottom, 8:1, 4:1, 2:1, 1:1, 1:2, 1:4 and 1:8.

(ii) *Type I mutations*

Now consider type I mutations alone. Setting $k_{II} = 0$ in equation 12 we obtain

$$R_I = \frac{N_I(1+4E_I)4p(1-2p)}{(8p^2-4p+1)} \qquad (14)$$

an expression dependent upon p and E_I. Figure 2 (solid lines) shows R_I plotted against the equilibrium sequence $G+C$ content. For all levels of proofreading sequences of intermediate $G+C$ content always have higher rates of type I mutations than sequences of extreme $G+C$ content. The reason is that sequences of extreme $G+C$ content are replicated in nucleotide pools which are deficient in the free nucleotides required to make type I mismatches. As the stringency of proofreading increases the curves become much flatter so that eventually the differences in mutation rates of sequences at 50% and 70% (or 30%) $G+C$ content are negligible. Flattening occurs because proofreading elevates the mutation rate of sequences at extreme $G+C$ content compared to sequences at intermediate $G+C$ content (see figure 2 dashed lines).

(iii) *The overall mutation rate*

In graphical terms the overall mutation rate relative to that in a sequence of 50% $G+C$ content (equation 12), is simply the average of the curves shown in figure 2. For instance if at 50% $G+C$ content there are two unproofread type I mutations to every proofread type II mutation, then two of the bottom curves in figure 2 should be added to the top curve and the result divided by three. Thus the maximum variation in the mutation rate is achieved when all mutations arise via unproofread type I mismatches or stringently proofread type II mismatches.

Let us consider sequences in which both type I and II mutations can occur starting with the cases when type II mutations are not proofread. Since the rate of unproofread type II mutations is independent of the sequence $G+C$ content sequences of intermediate $G+C$ content will always have higher mutation rates than sequences of extreme $G+C$ content whether or not the type I mismatches are proofread. Qualitatively the curves will be similar to those given in figure 2 (solid lines) only flatter. Quantitatively the gradient at each point will be λ times the original gradient and the curve will bisect the abscissa at λ, where λ is the fraction of mutations which are type I mutations in an equilibrium sequence of 50% $G+C$ content.

Consider now the case when type II mutations are stringently proofread and type I mutations are not proofread at all. Under these conditions equation 12 reduces to

$$R = \frac{8w_I p(1-2p)+2w_{II}(12p^2-6p+1)}{8p^2-4p+1}, \qquad (14)$$

which can be shown to have a local maximum when more than half ($w_I > \frac{1}{2}$), and a local minimum when less than half ($w_{II} < \frac{1}{2}$), of the mutations in a sequence of 50% $G+C$ content are type I mutations. Equation 14 is plotted against the equilibrium sequence $G+C$ content in figure 3. The rate of mutation is only weakly dependent upon sequence $G+C$ content unless type I mutations are much more common than type II mutations, or vice versa. This is even more the

https://doi.org/10.1017/S0016672300030676 Published online by Cambridge University Press

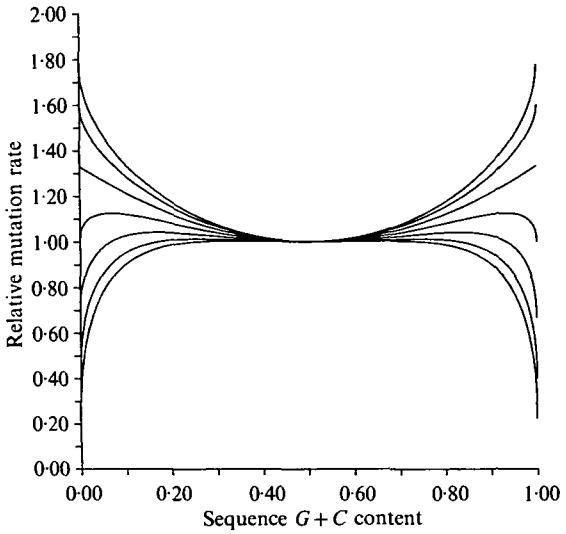

Fig. 4. The relative mutation rate for a sequence undergoing stringently proofread type I and type II mutation. Each curve represents a different ratio of type I and type II mutations. The ratio of proofread type I mutations to proofread type II mutations in a sequence of 50% $G+C$ content is, from top to bottom, 8:1, 4:1, 2:1, 1:1, 1:2, 1:4 and 1:8.

case when all mismatches are stringently proofread (figure 4).

(iv) *Changing the overall concentration of free nucleotides*

Up till now we have only considered the effect of varying the relative concentrations of the free nucleotides. However variations in the overall concentration of free nucleotides will affect the probability of proofreading, under this model, if the polymerase is not saturated by the free nucleotides.

Let the rate at which correct nucleotides collide with the polymerase be α and the rate at which proofreading occurs be β. The time t between collisions is distributed exponentially and the probability of proofreading during a particular time interval is a Poisson process with mean βt. Therefore the probability of not proofreading a mismatch between collisions is

$$1-V=\int_0^\infty e^{-\beta t}\alpha e^{-\alpha t}\,dt=\frac{\alpha}{\alpha+\beta} \qquad (18)$$

and $E=\beta/(j\alpha)$. α is proportional to the overall concentration of the nucleotides, so doubling the concentration leads to a doubling of α. When proofreading is slow (β and E small) changes in the overall concentration of free nucleotides have little effect on the probability of proofreading (see equation 7). However, when proofreading is stringent (β and E large) the probability of proofreading becomes a linear function of the free nucleotide concentration: i.e. $N\approx \alpha j(12c^2-6c+1)/(\beta(16c^2-8c+2))$. Hence it is possible for fluctuations in the overall concentration of nucleotides to cause large variations in the probability of proofreading, and the rate of mutation.

4. Discussion

It is important to appreciate that the $G+C$ values of the model do not necessarily correspond to the $G+C$ contents of the real world in either value or scale. This is because the mutation pattern was highly simplified and several assumptions were made about the free nucleotide pools. However, the general lack of mutation rate variance that DNA replication produces across sequences of different $G+C$ contents in the model is expected to be a robust result. Note in particular how the rates of misincorporation and proofreading act against each other for type I mutations. If there is an increase in the number of free nucleotides the polymerase has to try before finding the correct one, then the rate of misincorporation increases. However, the rate of DNA replication slows and the probability of proofreading rises. Also note how the relationship between $G+C$ content and the frequency of type II mutations opposes or dampens the relationship for type I mutations.

(i) *Isochore $G+C$ content*

Isochore $G+C$ contents vary from $\sim$ 38% to $\sim$ 53% in humans, with the range being somewhat narrower in mice and rats (Bernardi *et al.* 1985). Over this sort of range, assuming that the $G+C$ content scale of the model corresponds roughly to that in the real world, there is very little variation in the rate of mutation unless the total concentration of free nucleotides changes. Note in particular that if the frequency of type I and type II misincorporations are within an order of magnitude of each other the variation in the mutation rate across sequences of all $G+C$ contents is extremely limited. The variation appears to be too limited to explain either the $G+C$ content related, or total, silent substitution variance.

(ii) *The mutation pattern*

There is very little data on the pattern of misincorporation in mammals. Some data are available from the analysis of substitution patterns in pseudogenes (Gojobori *et al.* 1982; Li *et al.* 1984). However, the probability of repair appears to differ amongst mismatches (Brown and Jiricny, 1988, 1989) and across the genome (Bohr *et al.* 1987; Filipski, 1988) so substitution patterns may not correspond to the patterns of misincorporation.

(iii) *Overall changes in the free nucleotide concentrations*

The analysis above suggests that the maintenance of isochores by replication in biased free nucleotide

https://doi.org/10.1017/S0016672300030676 Published online by Cambridge University Press

pools does not *per se* generate mutation rate variance. However, changes in the overall concentration of free nucleotides can affect the mutation rate (under the present model) if at least one type of mismatch is proofread and the polymerase is unsaturated by free nucleotides. These conditions appear to be met in the real world, although proofreading may be restricted to certain mismatches and only one polymerase (Matthews & Slabaugh 1986; Kunkel *et al.* 1987; Meuth, 1989).

Changes in the overall concentration of free nucleotides might come about as a consequence of variations in the composition maintaining isochores, or via germ-line selection (Hastings 1989). Germ-line selection might lead to changes in the concentration because genes expressed in a tissue appear to be replicated early in that tissue (Holmquist, 1987, 1989; Goldman, 1988). So a decrease in the overall concentration might be favoured by selection to reduce the mutation rate in those genes being expressed in the germ-line; or alternatively an increase in the concentration might be selected for to increase the dosage of the early replicating genes by speeding up replication.

(iv) *Replication and repair*

The range of gene $G+C$ contents, especially at the 3rd position thereof ($\sim 30\%$ to $\sim 95\%$ in humans), is considerably greater than that across isochores. It seems likely that the mechanism responsible for generating this vast range of gene $G+C$ contents is also in part responsible for the variation in silent substitution rate. In particular the case has been recently made that biased DNA repair is the responsible party (Filipski, 1988). Although the interaction between repair and replication is non-trivial it seems unlikely that the maintenance of isochores by DNA replication would be a major contributor to mutation rate variance unless there are large changes in the total free nucleotide pool concentrations. Such overall changes in concentration might accompany changes in free nucleotide pool composition, or be produced by germ-line selection.

I am very grateful to Ken Wolfe for giving me access to his work prior to publication, for helpful discussion and encouragement. I thank Bill Hill, Peter Keightley and Prof A. J. Jeffreys for comments on this manuscript and to SERC for their financial assistance.

References

Aissani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K. & Bernardi, G. (1991). The compositional properties of human genes. *Journal of Molecular Evolution* **32**, 493–503.

Bernardi, G. (1989). The isochore organization of the human genome. *Annual Review of Genetics* **23**, 637–661.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. & Rodier, F. (1985). The mosaic genome of warm blooded vertebrates. *Science* **228**, 953–958.

Bohr, V. A., Phillips, D. H. & Hanawalt, P. C. (1987). Heterogeneous DNA damage and repair in the mammalian genome. *Cancer Research* **47**, 6426–6436.

Brown, T. C. & Jiricny, J. (1988). Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* **54**, 705–711.

Brown, T. C. & Jiricny, J. (1989). Repair of base–base mismatches in simian and human cells. *Genome* **31**, 578–583.

Bulmer, M., Wolfe, K. H. & Sharp, P. M. (1991). Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationships of the mammalian orders. *Proceedings of the National Academy of Science USA* **88**, 5974–5978.

Eyre-Walker, A. (1991). An analysis of codon usage in mammals: selection or mutation bias. *Journal of Molecular Evolution* **33**, 442–449.

Filipski, J. (1988). Why the rate of silent substitution is variable within the a vertebrate's genome. *Journal of Theoretical Biology* **134**, 159–164.

Gojobori, T., Li, W.-H. & Graur, D. (1982). Patterns of mutation in pseudogenes and functional genes. *Journal of Molecular Biology* **18**, 360–369.

Goldman, M. A. (1988). The chromatin domain as a unit of gene regulation. *Bioessays* **9**, 50–55.

Hastings, I. M. (1989). Potential germline competition in animals and its evolutionary implications. *Genetics* **123**, 191–197.

Holmquist, G. P. (1987). Role of replication time in the control of tissue specific gene expression. *American Journal of Human Genetics* **40**, 151–173.

Holmquist, G. P. (1989). Evolution of chromosome bands: molecular ecology of non-coding DNA. *Journal of Molecular Evolution* **28**, 469–486.

Ikemura, T. & Aota, S. (1988). Global variation in $G+C$ content along vertebrate genome DNA: possible correlation with chromosome band structures. *Journal of Molecular Biology* **203**, 1–13.

Kohalmi, S. E., Glattke, M., McIntosh, E. M. & Kunz, B. A. (1991). Mutational specificity of DNA precursor pool imbalances in yeast arising from deoxycytidylate deaminase deficiency or treatment with thymidylate. *Journal of Molecular Biology* **220**, 933–946.

Kunkel, T. A., Sabatino, R. D. & Bambara, R. A. (1987). Exonucleolytic proofreading by calf thymus DNA polymerase delta. *Proceedings of National Academy Science USA* **84**, 4865–4869.

Leeds, J. M., Slabaugh, M. B., Mathews, C. K. (1985). DNA precursor pools and ribonucleotide reductase activity: distribution between the nucleus and the cytoplasm of mammalian cells. *Molecular and Cellular Biology* **5**, 3443–3450.

Li, W.-H., Wu, C.-I. & Luo, C.-C. (1984). Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution* **21**, 58–71.

Li, W.-H., Tanimura, M. & Sharp, P. M. (1987). An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *Journal of Molecular Evolution* **25**, 330–342.

Matthews, C. K. & Slabaugh, M. B. (1986). Eukaryotic DNA metabolism: are deoxyribonucleotides channelled to replication sites? *Experimental Cell Research* **162**, 285–295.

McCormick, P. J., Danhauser, L. L., Rustim, Y. M. & Bertram, J. S. (1983). Changes in ribo- and deoxyribonucleoside triphosphate pools within the cell cycle of a synchronised mouse fibroblast cell line. *Biochima Biophysica Acta* **755**, 36–40.

Meuth, M. (1989). The molecular basis of mutations induced

by deoxyribonucleoside triphosphate pool imbalances in mammalian cells. *Experimental Cell Research* **181**, 305–316.

Phear, G. & Meuth, M. (1989*a*). A novel pathway for transversion mutation induced by dCTP misincorporation in a mutator strain of CHO cells. *Molecular and Cellular Biology* **9**, 1810–1812.

Phear, G. & Meuth, M. (1989*b*). The genetic consequences of DNA precursor pool imbalance: sequence analysis of mutations induced by excess thymidine at the hamster *aprt* locus. *Mutation Research* **214**, 201–206.

Sharp, P. M. (1989). Evolution at 'Silent' sites in DNA. In *Evolution and Animal Breeding: Reviews on Molecular and Quantitative Approaches in Honour of Alan Robertson.* (ed. W. G. Hill, T. F. C. Mackay). Wallingford CAB International 1989, pp. 23–31.

Ticher, A. & Grauer, D. (1989). Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein coding genes. *Journal of Molecular Evolution* **28**, 286–298.

Wolfe, K. (1991). Mammalian DNA replication: mutation biases and the mutation rate. *Journal of Theoretical Biology* **149**, 441–451.

Wolfe, K., Sharp, P. M. & Li, W.-H. (1989). Mutation rates differ among regions of the mammalian genome. *Nature* **337**, 283–285.