J. Mol. Evol. 13, 1–14 (1979)

Journal of Molecular Evolution

Evolutionary Potential: A Mathematical Hypothesis of Mouse Hemoglobin Beta Chain Evolution

John G. Gilman*

Universitá di Milano, Cattedra di Enzimologia, Via G. Celoria, 2, I-20133 Milano, Italy

Summary. This paper examines the possibility that the linkage arrangements and regulatory properties of genes may be influenced by selection. A mathematical hypothesis is developed in order to show how selective properties of hemoglobin beta chains could have influenced the linkage and regulation of their structural genes. The hypothesis is applied to the case of mouse hemoglobin beta chains. In most mice, closely-linked pairs of loci (doublets) code for two structurally divergent beta chains in unequal amounts. Some mouse strains have singlet alleles, however, coding for another beta chain variant. With the mathematical hypothesis, one can show that selectively determined "evolutionary potentials" may have favored changes in proportions of major and minor chains produced by a doublet allele. In the extreme case, zero production of the minor chain may give a selective advantage, leading to a singlet; conversely, selection may favor linking another gene to the singlet locus to give a doublet. A specific prediction of the model is the stable maintenance under certain conditions of multiple alleles at regulatory loci. The concept of evolutionary potential thus suggests that selection could have influenced the evolution of genotypic fitnesses, in addition to causing changes in gene frequencies as in standard population genetics models.

Introduction

Studies of hemoglobin structure and genetics have revealed the existence of many cases of duplicated loci, or "doublets". In humans, there are two closely-linked loci for variant beta chains (β and δ) (Boyer et al., 1963) as well as duplicated alpha and gamma chain loci (Hollan et al., 1972; Schroeder and Huisman, 1974). In the laboratory mouse (Mus musculus), many strains have doublet beta chain loci, and doublet alpha chain loci have also been studied (Russell and McFarland, 1974; Hilse and Popp, 1968).

In adult laboratory mice, the doublet beta chain allele *Hbb*^{d} produces unequal quantities of its two beta chains, βdmaj and βdmin (80% and 20%, respectively) (Hutton et al., 1962a,b). A systematic change in proportions of the two chains occurs during mouse fetal development, in the direction of diminishing the amount of minor chain, βdmin (Whitney,

*Present Address: Naval Medical Research Institute, Bethesda, Maryland 20014, USA

0022-2844/79/0013/0001/$ 02.80

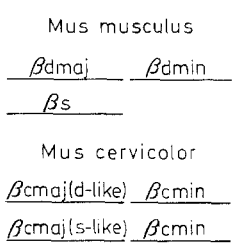

Fig. 1. The hemoglobin beta chain polymorphisms of Mus musculus and Mus cervicolor. Mus musculus has the Hbb^d-Hbb^s "doublet-singlet" polymorphism. The singlet codes for βs, while the doublet codes for unequal quantities of βdmaj and βdmin (about 80% and 20%, respectively). Mus cervicolor has a "doublet-doublet" polymorphism. Both doublets code for the minor chain βcmin, while they differ in their major chains – βcmaj (s-like) versus βcmaj (d-like). See Gilman (1976a,b) for a discussion of the sequence differences among the various mouse major and minor beta chains

1977). This suggests that the specific amount of minor chain βdmin produced by the adult animal could have adaptive significance.

The existence of a hemoglobin beta chain polymorphism in laboratory mice also suggests a certain adaptive relevance to the amount of minor beta chain produced: the alternative allele, Hbb^S, produces no minor beta chain in adult mice. Hbb^S produces only the βs chain, which is closely related to βdmaj (Gilman 1976b). I call Hbb^S a singlet locus, because it controls the production of only a single beta chain.

Another species of mouse, Mus cervicolor, has a genetic polymorphism for major beta chains similar to that for musculus chains βs and βdmaj, and yet both cervicolor alleles are doublets (Gilman, 1976b). Fig. 1 illustrates the beta chain breeding unit alleles for Mus musculus and Mus cervicolor.

Spofford (1972) has presented a mathematical hypothesis which relates the fitness of an organism to the microfitnesses of its various enzyme types. In this paper, that hypothesis is developed to show how selection could have caused the evolution of a doublet gene system in which variant beta chains were produced in unequal amounts. The hypothesis also illustrates how selection could have led to a switch from a doublet-singlet polymorphism (as in Mus musculus) to a doublet-doublet polymorphism (as in Mus cervicolor), or vice versa. Beta chain structural evolution is seen to lead to the development of "evolutionary potential", which may favor a switch in polymorphism type, or a change in the proportions of major and minor chains produced by a doublet.

The Mathematical Hypothesis

The model of this paper derives from the two chain hybrid dimer hypothesis of Spofford (1972), and it extends the hypothesis to the case of N+1 chains produced by N doublet alleles at a single locus. Focus will be on three chains produced by two alleles. This three chain case is mathematically equivalent to the population genetic model for the three allele balanced polymorphism maintained by selection. A biological rationale for the hypothesis is not presented. It was used because it is simple, and lends itself readily to a discussion of the relationship between variability on a protein level and fitness on a genotypic level. This paper uses the hypothesis to show how selection for properties

Table 1. Application of the mathematical hypothesis of Spofford (1972) to the case of N doublet alleles[a]

Doublet allele	Major chain of doublet	Minor chain of doublet	Fraction of major chain[b]	Fraction of minor chain
B_1	β_1	β_{N+1}	x_1	$1 - x_1$
B_2	β_2	β_{N+1}	x_2	$1 - x_2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
B_N	β_N	β_{N+1}	x_N	$1 - x_N$

[a]A doublet is defined as two very tightly-linked genes which produce closely homologous polypeptide chains; [b]all doublet alleles produce the same total quantity of beta chain (major plus minor), equal to 1

of the hemoglobin tetramer could have influenced genetic organization and regulation at the hemoglobin beta chain locus.

Genotypic Fitness Values as Determined by Molecular Properties

Table 1 shows how the hypothesis of Spofford (1972) is applied to the case of N doublet alleles. Each doublet allele B_i is assumed to code for a major beta chain β_i, as fraction x_i of its total beta chain, and the minor chain β_{N+1} as fraction $1-x_i$.

This model assumes a type of dosage adjustment: each allele produces the same total quantity of beta chain. Thus, a singlet allele (producing only β_i) is defined by setting x_i equal to one.

In any given animal, only two alleles are present, B_i and B_j. In general, each red cell of a particular animal will have one alpha and three beta chains, β_i, β_j, and β_{N+1}. Each beta chain is found as the same fraction of total beta in every red cell: $X_1 = x_i/2$ (fraction of β_i), $X_2 = x_j/2$ (fraction of β_j) and $X_3 = (2-x_i-x_j)/2$ (fraction of β_{N+1}).

With up to three beta chains in an animal, up to six hemoglobin types $\alpha_2\beta_i\beta_j$ are formed by randomly combining chains β_i and β_j. Then, in this model, the fitness of the B_iB_j animal is given by

$$W_{ij} = \sum_{kl} w_{\mu_k\mu_l} X_k X_l$$

where $\mu_1 = i$, $\mu_2 = j$, $\mu_3 = N + 1$, and the summation goes from 1 to 3. The $w_{\mu_k\mu_l}$ may be considered "microfitnesses" of the hemoglobin types $\alpha_2\beta_{\mu_k}\beta_{\mu_l}$.

Applications to Real Polymorphisms

This model, which assumes variation only in the major beta chains β_i, is appropriate to a number of important human beta chain polymorphisms, such as that for sickle

Table 2. Genotypic fitnesses as a function of hemoglobin microfitnesses for a two-allele polymorphism[a]

Genotype	Genotypic fitness
B_1B_1	$W_{11} = w_{ss}(x_s)^2 + 2w_{ms}x_s(1 - x_s) + w_{mm}(1 - x_s)^2$
B_1B_2	$W_{12} = w_{dd}(x_d)^2/4 + w_{ss}(x_s)^2/4 + w_{mm}(2 - x_d - x_s)^2/4 + w_{dm}x_d(2 - x_d - x_s)/2 + w_{ms}x_s(2 - x_d - x_s)/2 + w_{ds}x_dx_s/2$
B_2B_2	$W_{22} = w_{dd}(x_d)^2 + 2w_{dm}x_d(1 - x_d) + w_{mm}(1 - x_d)^2$

[a]A doublet-singlet polymorphism is obtained by setting $x_s = 1$. The doublet-doublet polymorphism discussed further in this paper has $x_s = x_d$

cell hemoglobin. In Mus musculus, no minor beta chain structural variation is involved in the Hbb^d-Hbb^s polymorphism (Gilman, 1976b), which is widespread in North America (Selander, 1970). The same is true for the polymorphism of Mus cervicolor (see Fig. 1). A third Mus musculus allele, Hbb^p, will not be considered here, as it differs in structural variation of the minor beta chain (Gilman, 1976b); Hbb^p of present day laboratory mice may derive originally from Asian animals (Morton and Tobin, 1977).

To apply this model to Mus musculus and Mus cervicolor polymorphisms, allele B_1 will be considered to code for chains β_s and β_m, as fractions x_s and $1 - x_s$, respectively. The alternate allele, B_2, will be considered to code for β_d and β_m, as fractions x_d and $1 - x_d$, respectively.

Up to six hemoglobins may be found in a given animal: $\alpha_2\beta_s\beta_s$, with microfitness value w_{ss}, $\alpha_2\beta_d\beta_d$ (microfitness w_{dd}), $\alpha_2\beta_m\beta_m$ (microfitness w_{mm}), $\alpha_2\beta_d\beta_m$ (microfitness $w_{dm} = w_{md}$), $\alpha_2\beta_m\beta_s$ (microfitness $w_{ms} = w_{sm}$), and $\alpha_2\beta_d\beta_s$ (microfitness $w_{ds} = w_{sd}$). Table 2 gives the genotypic fitness values as functions of the microfitness values, for the two allele polymorphisms of both Mus musculus ($x_s = 1$) and Mus cervicolor ($x_s = x_d$).

The Optimally Fit Single Animal

For a given fixed set of microfitness values w_{ss}, w_{ds}, and so on, the fittest possible animal may have one, two, or all three beta chains. With the model of this paper, one determines the proportions of each beta chain which optimize an animal's fitness by the same mathematical methods one used to determine equilibrium gene frequencies for an N allele polymorphism (see Crow and Kimura, 1970, pp. 270-277).

For example, if the optimally fit animal has only two chains, β_d and β_m, then the fraction x_o of β_d which optimizes its fitness is given by:

$$x_o = (w_{dm} - w_{mm})/(2w_{dm} - w_{mm} - w_{dd}) \quad (1)$$

For x_o to give a fitness maximum, one must have $w_{dm} > w_{mm}$ and $w_{dm} > w_{dd}$, by analogy with stability conditions for a two allele balanced polymorphism, as described below.

The Optimally Fit Population

For an N allele polymorphism, the equilibrium population fitness $\overline{W}$ may be calculated as

$$\overline{W} = \sum_{ij} p_i p_j W_{ij} \quad ,$$

assuming that p_i values represent equilibrium allele frequencies which are in Hardy-Weinberg proportions. The W_{ij} values are fitnesses of genotypes B_iB_j. For the two allele polymorphisms under consideration in this paper:

$$\overline{W} = W_{11}(1 - p)^2 + 2W_{12}p(1 - p) + W_{22}p^2 \qquad (2)$$

where p is the equilibrium frequency of allele B_2.

A stable equilibrium for the two allele polymorphism exists if $W_{12} > W_{11}$ and $W_{12} > W_{22}$ (Nagylaki, 1977, pp. 56-58). Population fitness will be maximal at the point of equilibrium, with $p = (W_{12} - W_{11})/(2W_{12} - W_{11} - W_{22})$.

Mutation Leading to Optimally Fitter Populations

For fixed values of microfitnesses w and major beta chain fractions x_S and x_d, the genotypic fitnesses W_{11}, W_{12} and W_{22} are fixed, and a stable two allele polymorphism for B_1 and B_2 may exist. However, a mutation in a beta chain structural gene could occur. This would introduce a new allele B_3 into the population, coding for a beta chain with altered structural and perhaps functional properties. It may also happen that a mutation occurs at a regulatory site, giving rise to an allele B_3 with altered proportions of the two beta chains produced by the doublet.

Population genetics theory (Wright, 1969, p. 44; Nagylaki, 1977, pp. 65-66) shows that the new allele B_3 will be maintained in the population if and only if

$$(W_{13} - W_{11})/(W_{12} - W_{11}) + (W_{23} - W_{22})/(W_{12} - W_{22}) > 1 \qquad (3)$$

Then, either B_3 will become the only allele in the population, or a new two allele polymorphism will arise (for B_1 and B_3, or B_2 and B_3), or the three allele polymorphism (for B_1, B_2, and B_3) will be stable.

As a guide to whether a new allele B_3 might lead to a greater average population fitness, one can examine how the equilibrium population fitness $\overline{W}$ would change if allele B_3 arose by mutation of B_2, and replaced it. For equilibrium fitness $\overline{W}$, $\partial\overline{W}/\partial p = 0$, so that equation (2) implies that the incremental change in equilibrium population fitness due to incremental changes in genotypic fitness is given by:

$$d\overline{W} = p^2 dW_{22} + 2p(1 - p)dW_{12} + (1 - p)^2 dW_{11} \qquad (4)$$

Using equation (4) and the formulas for genotypic fitnesses given in Table 2, one can show that $\overline{W}$ will always increase as microfitnesses w increase. However, if one holds microfitnesses constant and changes only x_s and x_d, $\overline{W}$ may reach a local maximum. For a simplified doublet-doublet polymorphism, with $x = x_d = x_s$, setting $d\overline{W}/dx = 0$ allows one to derive the following formula for the value of x which maximized equilibrium population fitness $\overline{W}$:

$$x_{\overline{W}max} = [8(w_{dm} - w_{mm})K + 4(w_{ms} - w_{dm})(K + 2L - 2M)]/[8KM - (K - 2L + 2M)^2] \quad (5)$$

where $K = 2w_{ds} - w_{dd} - w_{ss}$, $L = 2w_{dm} - w_{dd} - w_{mm}$, and $M = 2w_{ms} - w_{ss} - w_{mm}$. For this value of x to give a maximum for $\overline{W}$, the bracketed denominator in (5) must be positive.

For a doublet-singlet polymorphism such as *Hbb^d-Hbb^s* of Mus musculus, setting $d\overline{W}/dx_d$ equal to zero gives a cubic equation in x_d. However, by means of equation (4) one can show that letting $w_{ms} = w_{ds}$ leads to a simplification: The value of x_d which maximizes $\overline{W}$ is x_o, where x_o is given by equation (1).

Thus, the doublet-singlet population, with $w_{ms} = w_{ds}$, has maximal equilibrium population fitness $\overline{W}$, at fixed microfitnesses, when $x_d = x_o$. As shown above, x_o is also the value of x_d which maximizes the fitness of the doublet-doublet homozygote, and one can easily show that, for $w_{ms} = w_{ds}$, x_o also maximizes the fitness of the doublet-singlet heterozygote.

Evolution of x_d, The Fraction of Major Chain Produced by a Doublet

For the simplified doublet-singlet polymorphism just described, with $w_{ms} = w_{ds}$, one can show that evolution in the value of x_d will be in the direction of the value x_o, given by equation (1). Such evolution in the value of x_d for the doublet can only occur if mutant doublets with altered values of x_d arise in the population. Once a mutation of the old doublet allele B_2 to the new allele B_3 occurs, B_3 must become established in the population and replace B_2.

Once B_3 arises, it will become established in the population if and only if inequality (3) is satisfied. For the doublet-singlet population with $w_{ms} = w_{ds}$, this inequality will certainly be satisfied if allele B_3 has its value of x_d closer to x_o than allele B_2. This can be most easily seen by considering the population's genotypic fitnesses, given in Table 2, as expressed in a new coordinate system centered on x_o. In this new coordinate system, x_d for allele B_2 is equal to $b_2 + x_o$, while x_d for allele B_3 is $b_3 + x_o$. The genotypic fitnesses in the new coordinate system are given in Table 3.

Inspection of Table 3 shows that $(b_2)^2 > (b_3)^2$ implies that $W_{13} > W_{12}$ and $W_{23} > W_{22}$, if one assumes that $w_{dm} > w_{mm}$. This assumption, and the assumption that $w_{dm} > w_{dd}$, are necessary if the value of x_o in equation (1) is to give a maximum equilibrium population fitness; these assumptions will be made throughout this paper. Since stability of the original polymorphism required that $W_{12} > W_{11}$ and $W_{12} > W_{22}$, inequality (3) will be satisfied. Thus, B_3 will become established in the population if $(b_2)^2 > (b_3)^2$, which means that x_d for B_2 is farther from x_o than x_d for B_3.

Table 3. Genotypic fitnesses, in a transformed coordinate system[a], for three alleles (singlet B_1 and doublets B_2 and B_3)

Genotype	Genotypic fitness[b]
B_1B_1	$W_{11} = w_{ss}$
B_1B_2	$W_{12} = (W_{11} + W_{22})/4 + w_{ms}/2$
B_2B_2	$W_{22} = [(w_{mm} - w_{dm})/x_o]\,[(b_2)^2 - (x_o)^2] + w_{mm}$
B_1B_3	$W_{13} = (W_{11} + W_{33})/4 + w_{ms}/2$
B_2B_3	$W_{23} = [(w_{mm} - w_{dm})/x_o]\,[(b_2 + b_3)^2/4 - (x_o)^2] + w_{mm}$
B_3B_3	$W_{33} = [(w_{mm} - w_{dm})/x_o]\,[(b_3)^2 - (x_o)^2] + w_{mm}$

[a]The transformation is in x_d, from a coordinate system centered on zero (as in Table 2) to one centered on x_o: $x_d = b + x_o$. x_o is the value of x_d which maximizes equilibrium population fitness $\overline{W}$ for the doublet-singlet polymorphism, when $w_{ms} = w_{ds}$; x_o is given by equation (1) of the text. For allele B_2, $x_d = b_2 + x_o$, and for allele B_3, $x_d = b_3 + x_o$; [b]for these formulas, it is assumed that $w_{ms} = w_{ds}$

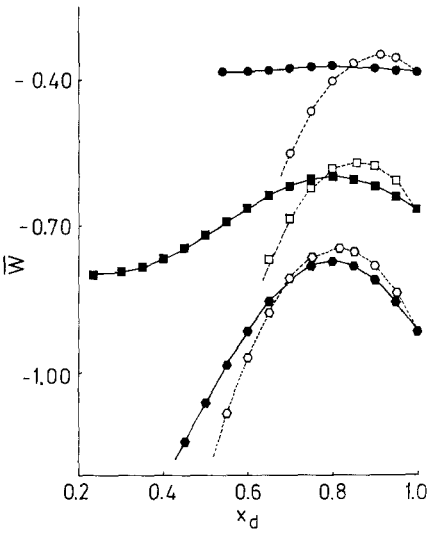

Fig. 2. Equilibrium population fitness $\overline{W}$ as a function of x_d, the fraction of major chain produced by a doublet. For three sets of microfitness parameters, $\overline{W}$ is plotted for both doublet-singlet ($x_s = 1$) and doublet-doublet ($x_s = x_d$) polymorphisms. All three sets of microfitnesses have $w_{dm} = w_{ds} = w_{ms} = 0$, $w_{dd} = -1$, and $w_{mm} = -4$, and differ only in w_{ss}: $w_{ss} = -0.3850$, ● (doublet-singlet), ○ (doublet-doublet); $w_{ss} = -0.8000$, ■ (doublet-singlet), □ (doublet-doublet); $w_{ss} = -1.662$, ⬢ (doublet-singlet), ⬡ (doublet-doublet). In every case, equilibrium population fitness $\overline{W}$ of the doublet-singlet is maximal when $x_d = 0.8$, and the frequency p of the doublet allele at $x_d = 0.8$ goes from 0.15 ($w_{ss} = -0.385$) to 0.50 ($w_{ss} = -0.8$) to 0.85 ($w_{ss} = -1.662$). For the middle and upper sets of curves, the doublet-singlet points were plotted for the range of x_d for which the polymorphism was stable ($x_d > 0.23$, and $x_d > 0.53$, respectively). For the doublet-singlet of the lower set of curves, and for all three doublet-doublet examples, the polymorphisms were stable for $0 < x \leqslant 1$

For evolution of the doublet-singlet polymorphism to occur through replacement of B_2 by B_3, neither the two allele polymorphism for B_2 and B_3, nor the three allele polymorphism for B_1, B_2, and B_3, can be stable. However, the two allele polymorphism for B_1 and B_3 must be stable; this is certainly possible, as Fig. 2 shows. Examination of Table 3 demonstrates that a two allele polymorphism for B_2 and B_3 would not be stable, since it is impossible to have both $W_{23} > W_{33}$ and $W_{23} > W_{22}$. One can also show that stability of the two allele polymorphisms, for B_1 and B_2, and B_1 and B_3, implies instability of the three allele polymorphism for B_1, B_2, and B_3, if b_2 and b_3 are of the same sign. If b_2 and b_3 have different signs (b_2 negative and b_3 positive, for example), then the three allele polymorphism may be stable.

The mathematical model therefore allows the possibility that the fraction x_d of major beta chain produced by a doublet, in a doublet-singlet population, may have evolved under the influence of selection.

Discussion

Mathematical models for evolution are often concerned with the effect of genotypic fitness values on gene frequencies at or near equilibrium, and the rates at which equilibrium is approached. The model of this paper, however, considers how parameters determined directly by the gene products might influence the evolution of the genotypic fitnesses. The model is based on that of Spofford (1972), which derives the genotypic fitnesses for a single locus from the microfitness values of dimeric enzymes whose production is controlled by that locus. I have applied this model to the case of mouse hemoglobin beta chains, in order to show how properties of the hemoglobins could have led to evolution in properties of the genes producing the hemoglobin beta chains.

Fig. 1 has illustrated the hemoglobin beta chain polymorphisms of two mouse species, the laboratory mouse Mus musculus, and the Asian species Mus cervicolor. Both species have doublet alleles producing a major beta chain and a minor beta chain. For the purpose of discussing the mathematical hypothesis, these chains are called βd and βm, respectively, for either species. The doublet allele controlling their production is called B_2. For both species, there is an alternate allele B_1, which produces a variant major chain βs. In Mus cervicolor, B_1 is a doublet, coding for βs and βm in approximately the same proportions as for βd and βm of B_2. In Mus musculus, B_1 is a singlet, controlling production of only βs; B_1 produces no minor chain βm.

In an arbitrary animal, up to six hemoglobins may be found (for simplicity, this model does not consider α chain variation): $\alpha_2\beta_s\beta_s$, $\alpha_2\beta_d\beta_s$, and so on. Associated with each hemoglobin a microfitness value: w_{ss}, w_{ds}, and so on. The fitnesses W_{11}, W_{12}, and W_{22}, for genotypes B_1B_1, B_1B_2, and B_2B_2, are determined by the microfitness values w, as well as by the fraction of major chain produced by B_1 (x_s) and B_2 (x_d). Genotypic fitnesses are given in Table 2, above.

Could Selection Have Favored Unequal Production of Two Beta Chains of a Doublet?

In the mathematical model of this paper, doublet-singlet populations will have different equilibrium fitnesses $\overline{W}$, depending on the fraction x_d of major beta chain produced by

the doublet. In a set of such populations with the same values for their microfitnesses, one population with a particular value of x_d may have greater fitness than any other. This is also true for the doublet-doublet polymorphism, and formula (5) above gives the value of x_d for maximal equilibrium population fitness $\overline{W}$ when $x_d = x_s$.

Fig. 2 illustrates this feature of the model. Fitnesses $\overline{W}$ for doublet-singlet populations are represented by filled symbols, and those symbols for populations with the same microfitness values but different x_d values are connected by solid lines. Corresponding doublet-doublet fitnesses (unfilled symbols) for populations with the same microfitness values are connected by broken lines. Three different sets of microfitness values are considered, which differ only in w_{ss} (w_{ss} = -1.662, -0.8, and -0.385, for the lower, middle, and upper sets of curves, respectively). For every curve of Fig. 2, there is one value of x_d which maximizes the equilibrium population fitness $\overline{W}$.

Early in the evolutionary history of mouse beta chain genes, one imagines that populations may have had doublet loci coding for equal amounts of two divergent beta chains. Suppose that the population has a doublet-singlet polymorphism, with the singlet allele B_1 (coding for βs) and the doublet B_2 coding for equal quantities of βd and βm. If a mutation occurs in B_2, generating the new doublet B_3 producing more βd than βm (for example, 70% βd and 30% βm), will B_3 replace B_2?

Examination of Fig. 2 suggests that this may be possible, since doublet-singlet population fitness $\overline{W}$ increases as x_d goes from 0.5 to 0.8, at which point $\overline{W}$ is maximal. The analysis of the mathematical hypothesis presented above shows that a more definite answer can be given. Applied to the doublet-singlet populations of Fig. 2, with $w_{ms} = w_{ds}$ and w_{ss} = -0.8 or -1.662, the analysis shows that B_3 (with $x_d = 0.7$) will replace B_2 (with $x_d = 0.5$). Then, once equilibrium is established for the doublet-singlet polymorphism with alleles B_3 and B_1, a mutation in B_3 creating B_4, with $x_d = 0.75$, will lead to a new doublet-singlet polymorphism for B_4 and B_1.

This process could continue until a doublet is generated with $x_d = 0.8$, as long as new doublet alleles have x_d greater than that of the old allele, and less than or equal to 0.8. If, however, a new doublet is generated with x_d on the other side of the maximum ($x_d > 0.8$), then the three allele polymorphism may be stable. For example, consider the doublet-singlet polymorphism for B_1 and B_4, with $x_d = 0.75$. Microfitness values for the lower set of curves will be assumed (w_{ss} = -1.662). Numerical calculations show that a new allele B_5, with $0.82 \leqslant x_d \leqslant 0.93$, will become established in the population as part of a three allele polymorphism for B_1 (the singlet), B_4 ($x_d = 0.75$), and B_5. For B_5 with $x_d = 0.82$, allele frequencies at equilibrium will be 0.15, 0.03, and 0.82, respectively.

This discussion of the mathematical model therefore suggests that selection could have favored evolution of the doublet from $x_d = 0.5$ to $x_d = 0.8$, which is the value for the *Hbb*^{d}-*Hbb*^{s} polymorphism of Mus musculus. With a fixed set of hemoglobin microfitness values, at $x_d = 0.5$ an "evolutionary potential" favored evolution of the value of x_d towards $x_d = 0.8$. The discussion also suggests that a real population might have many doublet alleles, with a distribution of x_d values centered approximately at 0.8. Evidence for multiple alleles at regulatory loci closely linked to hemoglobin structural loci has been reported for the deer mouse (Snyder, 1978).

Why Does Mus musculus have a Doublet-Singlet Polymorphism?

A singlet may be thought of as resulting from regulatory mutation of a doublet, so that x_S has evolved to a value of 1. In this sense, the question of why one mouse species has a doublet-singlet polymorphism and the other a doublet-doublet may be no different from the question of why a doublet produces unequal amounts of its two chains.

For the lower set of curves of Fig. 2 ($w_{ss} = -1.662$), one sees that at $x_d = 0.8$ the doublet-singlet population is less fit than the doublet-doublet ($x_s = x_d = 0.8$). Numerical calculations show that a mutation of singlet B_1 ($x_s = 1$) to doublet B_3 ($x_s = 0.8$) would lead to a new doublet-doublet polymorphism, for B_2 and B_3.

An evolutionary potential may also favor the change from a doublet-doublet to a doublet-singlet polymorphism: This is the case for a doublet-doublet population, with $x_d = x_s = 0.8$ and microfitnesses as for the lower curves of Fig. 2, except that $w_{mm} = -3.8$ and $w_{ms} = -1$ (instead of $w_{mm} = -4$ and $w_{ms} = 0$). Thus, ancestral Mus musculus populations with the doublet-singlet polymorphism may have been fitter than if they had possessed a doublet-doublet polymorphism. The mathematical model allows the possibility that selection determined polymorphism type in Mus species, given that the appropriate genetic variability was present in early populations.

The Development of an Evolutionary Potential

Examples have been presented of evolutionary potentials for change in the proportions of major and minor chains produced by a doublet. A special case of this is the evolutionary potential just described, for change from a doublet-doublet to a doublet-singlet polymorphism.

Evolutionary potentials such as those discussed may exist because genetic constraints prevent the population from achieving the fitness of an optimal single animal. For example, suppose that the optimal animal had only the two chains βd and βm, and no βs. Then x_o of equation (1) gives the fraction of βd in the optimal animal, as a function of w_{dd}, w_{dm} and w_{mm}. The only way population fitness could reach the optimal level of a single animal would be if all animals were homozygous for doublets producing fraction x_o of βd and βm. An evolutionary potential therefore exists for the emergence of such a doublet: if a genetic event occurred giving rise to it, and if its gene frequency initially increased due to selection, it would probably supplant preexisting loci.

A situation of stasis will never come about in this mathematical model, however, even though the population now has optimal fitness for its given microfitness values w_{dd}, w_{dm} and w_{mm}. The reason is that there is always an evolutionary potential for increases in microfitness values as the result of mutations in the structural genes coding for βd and βm. An increase in microfitness values is always translated into increased equilibrium population fitness $\overline{W}$. This increase in microfitness values may then lead to a new evolutionary potential for change in proportions of chains βd and βm. Structural gene mutation may also eventually lead to the existence of new beta chain polymorphisms and consequent new evolutionary possibilities.

Fig. 2 above, illustrates how change in the microfitness value w_{ss} leads to an evolutionary potential for change in polymorphism type. For the lower set of curves ($w_{ss} = -1.662$), at $x_d = 0.8$, the doublet-doublet population is fitter than the doublet-singlet. A

population initially in the doublet-doublet state will stay that way. As w_{ss} increases, Fig. 2 shows that the doublet-singlet population is eventually fitter, at $x_d = 0.8$; numerical calculations show that the singlet allele will become established, if introduced into the doublet-doublet population with $w_{ss} = -0.385$. The three allele polymorphism, for the singlet and both doublets, will be stable. Thus, an increase in the value of w_{ss} has led to an evolutionary potential for establishment of the singlet allele in the doublet-doublet population.

Fig. 3 shows how the development of an evolutionary potential due to increasing microfitness w_{ds} could have played a role in the evolution of the Mus musculus *Hbb*^{d}-*Hbb*^{s} polymorphism. Fig. 3a shows a schematic evolutionary pathway, in four stages, from an original singlet locus to the doublet-singlet polymorphism of today. Fig. 3b shows the effect of changing w_{ds} on two possible populations of stage 4.

Stage 1 of Fig. 3a shows that the original mouse beta chain gene is assumed to code for $\hat{\beta}$d, the ancestor of βdmaj of the present *Hbb*^{d} locus. In stage 2, the initial gene has undergone mutation so that there are now two alleles, coding for $\hat{\beta}$d and the variant chain $\hat{\beta}$m. The two alleles will be maintained as a balanced polymorphism, if one assumes that $w_{dd} = -1$, $w_{dm} = 0$ (as for Fig. 2), and $-3 < w_{mm} < -1/3$. If we let $w_{mm} = -2$, then equilibrium fitness $\overline{W}$ for the two allele singlet-singlet polymorphism of stage 2 is -0.958.

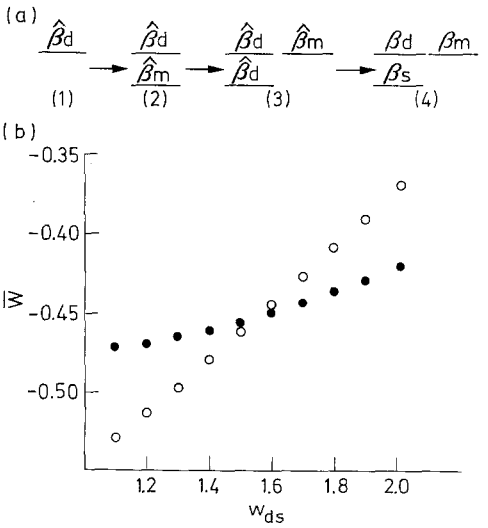

Fig. 3. A hypothetical scenario for the evolution of Mus musculus breeding unit alleles. (a) A possible evolutionary pathway from ancestral singlet to a doublet-singlet polymorphism, with the doublet producing equal amounts of both chains. A similar pathway, which bypasses the singlet-singlet stage (step 2), has already been suggested (Gilman, 1976a). See text for details. (b) This plot of equilibrium population fitness $\overline{W}$ versus w_{ds} illustrates how $\overline{W}$ would change if amino acid substitutions occurred in βs or βd, such that w_{ds} increased from 1.1 to 2.015. Values of the other microfitness parameters are assumed to remain constant ($w_{dm} = 0$, $w_{mm} = -0.897$, $w_{ms} = -0.9$, $w_{dd} = -1$, $w_{ss} = -1.414$). Plots are given for two populations, one with $x_d = 0.5$ (•), and one with $x_d = 0.8$ (○)

In stage 3, a genetic event has occurred linking genes for $\hat{\beta}$d and $\hat{\beta}$m on the same chromosome. As the simplest assumption, $\hat{\beta}$d and $\hat{\beta}$m are considered to be produced in equal amounts. Since the polymorphism of stage 2 was maintained by heterozygote superiority, it is fairly obvious that linking the genes for $\hat{\beta}$d and $\hat{\beta}$m ought to give the population increased fitness. It also happens that the new doublet coding for $\hat{\beta}$d and $\hat{\beta}$m, and the old singlet coding for $\hat{\beta}$d, will form a two allele doublet-singlet polymorphism, if $w_{mm} < -5/3$. As this condition holds ($w_{mm} = -2$), the doublet-singlet polymorphism is stable, and equilibrium population fitness $\bar{W}$ has increased to -0.740 in stage 3.

In stage 4 of Fig. 3a, evolution of beta chain sequences has occurred. The doublet codes for βd and βm in equal quantities, while the singlet now codes for βs. In consequence, microfitness values have evolved: w_{mm} has increased to -0.897, while w_{ms} = -0.9, w_{ss} = -1.414, and w_{ds} = 1.1. Equilibrium fitness for the population has risen to -0.472. Given these microfitness values, any deviation from equal production of both chains of the doublet would only lower the equilibrium population fitness $\bar{W}$. An evolutionary potential does exist for the generation of a new doublet coding for approximately equal quantities of βd and βs. However, crossover suppression is now assumed to be operative, as it may be in present day mice (Tiemeier et al., 1978). As the new doublet cannot easily be generated by unequal crossing over, further evolution must await changes in microfitness values.

Fig. 3b shows how change in βd or βs, leading to change in w_{ds}, could give rise to an evolutionary potential for change in the fractions of major and minor chains produced by the doublet. All other microfitness values are assumed to remain constant. Plots of equilibrium population fitness $\bar{W}$ versus w_{ds} are given for two populations, the existing one with $x_d = 0.5$ (filled symbols), and an alternate doublet-singlet population, $\bar{W}$ increases as w_{ds} goes from 1.1 to 2.015. However, for the alternate population with $x_d = 0.8$ (unfilled symbols). For the existing population, $\bar{W}$ increases as w_{ds} goes from 1.1 to 2.015. However, for the alternate population with $x_d = 0.8$, $\bar{W}$ increases much faster. When $w_{ds} = 2.015$, the fittest doublet-singlet population is one for which the doublet produces 80% βd and 20% βm, as in the present day Hbb^d-Hbb^s polymorphism of Mus musculus.

One can show that such evolution in w_{ds} can occur, because a new allele in the doublet-singlet population, giving a higher value of w_{ds} than the old allele, will replace the old allele. Once $w_{ds} = 2.015$, an evolutionary potential exists for replacement of the old doublet ($x_d = 0.5$) by one with $x_d = 0.8$. If a new doublet arises which produces 80% βd and 20% βm, it will replace the old doublet. A population similar to that of present day Mus musculus will have evolved.

Conclusion

The mathematical model described here related the fitness of an animal to its hemoglobin microfitnesses. As applied to mouse hemoglobin beta chain genes, the model has suggested that a population may be able to maximize its fitness by several means: Evolution of sequences of proteins and consequently of their functions, regulation of the relative amounts of major and minor chains produced by a doublet, and change in types of alleles (doublet or singlet) present in a polymorphic population. These processes may interact with each other, with the result that one type of alteration may lead to an evolutionary potential for another type of change. From a population genetics standpoint, evolutionary potentials are seen as leading to changes in genotypic

fitnesses. These changes are then reflected in increased population fitness at equilibrium.

While the model of this paper assumed an ideal "Darwinian" situation, it is conceivable that even "non-Darwinian" evolution could lead to development of evolutionary potentials for selectively advantageous change. It may also be possible to apply some of the ideas concerning evolution of genetic systems to the evolution of the protein molecule itself. Since the protein is a constrained system, molecular evolution of parts of it could lead to an evolutionary potential for change in other parts. Coates (1975) used the term "preadaptation" to describe such a process.

It may therefore be necessary to view evolution of an organism, or even of a protein molecule, as a dialectical process, in which one type of evolutionary change may interact with or facilitate another.

Acknowledgements. I thank Dr. L. Rossi-Bernardi for his hospitality during the work on this paper, and Dr. C. Denniston and Dr. J.F. Crow for discussions. Dr. Gilman was supported in part by the National Institutes of Health (1 FO 2 HL55564-02) and the Consiglio Nazionale delle Ricerche (Italy) (Decreto No. 197213).

References

Boyer, S.H., Rucknagel, D.L., Weatherall, D.J., Watson-Williams, E.J. (1963). Am. J. Hum. Gen. **15**, 438-448

Coates, M.L. (1975). J. Mol. Evol. **6**, 285-307

Crow, J.F., Kimura, M. (1970). An Introduction to Population Genetics Theory, pp. 270-277, New York: Harper & Row

Gilman, J.G. (1976a). Biochem. J. **155**, 231-241

Gilman, J.G. (1976b). Biochem. J. **159**, 43-53

Hilse, K., Popp, R.A. (1968). Proc. Nat. Acad. Sci. U.S.A. **61**, 930-936

Hollan, S.R., Szelenyi, J.G., Brimhall, B., Duerst, M., Jones, R.T., Koler, R.D., Stocklen, Z. (1972). Nature **235**, 47-50

Hutton, J.J., Bishop, J., Schweet, R., Russell, E.S. (1962a). Proc. Nat. Acad. Sci. U.S.A **48**, 1505-1513

Hutton, J.J., Bishop, J., Schweet, R., Russell, E.S. (1962b). Proc. Nat. Acad. Sci. U.S.A. **48**, 1718-1724

Morton, J.R., Tobin, G. (1977). Biochemical Genetics **15**, 101-108

Nagylaki, T. (1977). Lecture Notes in Biomathematics, Vol. 15: Selection in One- and Two-Locus Systems, pp. 56-58, 65-66, New York: Springer-Verlag

Russell, E.S., McFarland, E.C. (1974). Ann. N.Y. Acad. Sci. **241**, 25-38

Schroeder, W.A., Huisman, T.H.J. (1974). Ann. N.Y. Acad. Sci. **241**, 70-79

Selander, R.K. (1970). Am. Zoologist **10**, 53-66

Snyder, L.R.G. (1978). Genetics **89**, 531-550

Spofford, J.B. (1972). Brookhaven Symposia in Biology **23**, 121-143

Tiemeier, D.C., Tilghman, S.M., Polsky, F.I., Seidman, J.G., Leder, A., Edgell, M.H., Leder, P. (1978). Cell **14**, 237-245

Whitney, J.B., III (1977). Cell **12**, 863-871

Wright, S. (1969). The Theory of Gene Frequencies, Vol. 2, p. 44, Chicago: University of Chicago Press

Received January 15, 1978/Revised November 25, 1978