J.Mol.Evol.7,269-311 (1976)

Journal of Molecular Evolution

Evolutionary Processes and Evolutionary Noise at the Molecular Level

II. A Selectionist Model for Random Fixations in Proteins

EMILE ZUCKERKANDL*
Marine Biological Laboratory, Woods Hole, Mass. 02543, and
Department of Biological Sciences, University of Delaware,
Newark, Del. 19711, USA

Received November 25, 1974; January 1, 1976

Summary. On account, notably, of a competition between different component functions for individual sites in polypeptide chains, each protein molecule represents a functional compromise, with some functions optimized, but the overall state of the molecule "suboptimal". The proposal is made that the selection coefficient relating to a protein molecule under given conditions can in principle be broken down into partial selection coefficients relevant to the different functions that the molecule carries out. At general-function sites, each fixation improves some function, while others deteriorate, at first nonsignificantly, and the overall adaptive state of the molecule fluctuates around its maximum. A selective mechanism is described whereby kaleidoscopic changes in primary structure at variable sites are indefinitely promoted, independently of any environmental changes and with the molecule remaining close to a state of maximal overall adaptation. The paradoxical aspect of this proposal is analyzed. The implication of specific functions in substitutions at general-function sites is noted. Further, it is shown that a certain category of changes in the internal environment of the organism can be integrated into the constant-environment model for selection. Genetic sufficiency is considered a notion more adequate than genetic optimality for describing biological fitness and for providing a basis for the present model. On this basis selection occurs without genetic load. Multipolymorphism is one of the consequences. Several lines of evidence, in particular observations on polymorphism in deep sea organisms, seem to support the model. It is pointed out that it provides a theoretical foundation for a molecular evolutionary clock. The theoretical constancy of the clock depends on the constancy of functional density. The question of the evolution of functional density is examined. Comparisons of observed substitution fre-

*Directeur de Recherche at Centre National de la Recherche Scientifique, Paris.

quencies with values expected on a random basis are rejected as a measure of the contribution to evolution of nondetermination. They are considered to reflect a hierarchy in the resistance of the molecules to different amino acid residues as substituents. A limited component of "true" randomness, again accompanied by selection, is on the other hand provided by the model. Most amino acid substitutions are considered evolutionary noise, even though noise compatible with selection. It is proposed that evolutionarily significant substitutions may be identified by monitoring changes in functional density and weighted functional density.

Key words: Protein Evolution/Natural Selection/Random Substitutions/Neutral Substitutions/Partial Selection Coefficients/Genetic Sufficiency/Genetic Load/Polymorphism/Molecular Evolutionary Clock/Functional Density

On the basis of the analysis given in the preceding paper of general and specific functions in proteins, of functional density, and of the functional "degeneracy" of amino acids, a mechanism for evolutionary substitutions of amino acids can be formulated. This mechanism unites natural selection, a measure of randomness, and some aspects of neutrality. It is not a blend of these components with a fractional participation of each. Rather, it is altogether based on the hypothesis of natural selection, and the other aspects apply simultaneously. It is not meant, even if found acceptable, to integrally replace other theories and interpretations, of which several may apply to different extents in different situations. Yet the component mechanism to be proposed, if valid, should promote coherence in the understanding of evolutionary processes.

FUNCTIONAL COMPROMISE AND FUNCTIONAL SUBOPTIMALITY

Functional compromise should be a constant feature of any protein molecule, as well, probably, as of any larger biological unit.

In proteins, there must be compromise, because certain functional requirements are necessarily contradictory, and because others are overlapping with respect to the structures they involve.

Contradictory functional imperatives are manifest particularly among the general functions of proteins, for example: "Do decrease surface polarity in view of an interaction between protomers, but don't decrease the solubility of the protomers".

The overlapping of functions in proteins is very general, because of the following two circumstances. Firstly, functional "space" is limited: the set of different functional requirements has to be accomodated by a limited set of structural

sites. A maximal average limitation of the number of sites of the functional unit is expected to obtain as a feature of the trend toward maximal energetic economy compatible with survival. With respect to the supporting structure, in our case molecular structure, functions probably tend to be as crowded as compatible with their performance. Thereby the organism should reduce maximally the cost of maintaining structures.

Secondly, the functional "degeneracy" of amino acids, referred to in the preceding paper, implies that each amino acid residue will participate simultaneously in at least several general, if not also specific functions. A given evolutionarily effective amino acid substitution will operate an adjustment with respect to one function. But this substitution will have effects on other functions as well. Thus no substitution can be retained that is not a suitable, or tolerable, compromise with respect to all the functions involved.

At each accepted amino acid substitution, one advantage is traded for another. Not only a whole organism, in fact a single informational macromolecule is a big trading center. It is a general prerequisite for biological structures, at whatever level they be considered, that their functions must be compatible with all the other functions exercized by a superior organic unit. Proteins go one step further. In their case it is unavoidable that several functions compete for the same structural elements, namely for the same amino acid sites, and no amino acid site is likely to totally escape this generalization.

Considering macroscopic functional levels, Kermack (1954) already stated, a long time ago, on the basis of his biometrical study of a starfish, that "...a gene which, in any individual, produces a positive selective advantage in one set of characters must, at the same time, produce a negative selective advantage in another set. Thus the gene differences which survive in a random mating population will tend to affect characters of selective value in opposite senses, so that characters which make for fitness will be negatively correlated at any given time". We shall assume that, as far as informational macromolecules go, there is no difference with respect to the necessity of functional compromise, whether we consider the molecular species at a certain time or throughout evolution. This we believe to hold at least for the long period of past evolution when informational macromolecules had already attained a level of functional density close to the present level, - if we suppose that an initial significant increase in functional density occurred.

It is, in fact, an axiom of mathematics that two functions cannot be maximized simultaneously.

Basic to the mechanism for natural selection that will be described is the following postulate: no structural variant of a given type of protein is ever optimally adapted to all of its functions at once. Every protein is considered to be permanently in a state of slight inadaptation. Because of the number of functions that compete among themselves for reaching a better state of adaptation, the overall state of slight inadaptation fluctuates only moderately around a mean. An intrinsic *maximum* of overall functional fitness is defined by the level of compromise that is imposed by the competition between several functions for the same sites.

On the basis of the principle of concomitant ascent (Zuckerkandl, 1975), which implies that a less sophisticated protein is perfectly satisfactory within a system of less sophisticated proteins, a polypeptide chain whose functional density was well below its intrinsic maximum could still be near its optimum functional adaptation at very early times of evolution, - if we assume that the functions performed per unit polypeptide length were less numerous then and less specific. In that case it was in the course of further evolution that the optimal fitness of a protein caught up with its maximal fitness and indeed went beyond it, thus turning into a virtual, unattainable optimum.

At the present time, a protein molecule may still be optimally adapted to some of its *individual* general as well as specific functions. It is then assumed here, in compensation, to be suboptimally adapted to the remainder of at least the general functions.

Statements about biological optimality meet with the difficulty that something can be best only in relation to an objective. What exactly should be taken as the biological objective is not selfevident, nor is any definition of it likely to escape criticism. Survival of the species is an important possibility for such a definition, and one that will be considered in the present paper. In connection with this criterion, the use of the optimality concept of fitness of genotypes will be rejected.

On the other hand, the idea of an optimal constitution, or of alternate optimal constitutions of a machine does make sense, if it refers to a set of criteria relevant to the functioning of the machine. In a given environment, different alleles may be equally compatible with the survival of the species; but from the point of view of the functioning of the individual organism they may not be equally close to optimality. The optimal functioning of the individual would be described by the flawless performance, under constant environment, of the set of its functional activities (respiration, excretion, locomotion, reproduction, etc. etc.).

In referring in this paper to constant environment, we do not imply that the environment is not fluctuating; we mean an environment whose established fluctuations do not vary in kind, amplitude, and period. In the jargon of the ecologist, temporal variation is then "fine-grained". It is assumed that the organisms have had sufficient time for adapting to a stable environmental cycle so that genetic changes in the population are no longer necessary for insuring the survival of the species.

The physiological functions with respect to which an optimality concept of fitness is definable are ultimately based on the structures and functions of informational macromolecules They include many functions besides reproduction. However, the functional performances of the individual other than reproductive and, on the other hand, its reproductive performances must be statistically correlated. Through this correlation, individual functional optimality - to be traced down to the individual functional optimality of molecules - does have a bearing on natural selection.

PARTIAL SELECTION COEFFICIENTS

The numerous conservative substitutions that have occurred during the evolution of proteins have often puzzled us. Their functional impact, as the term conservative implies, appears to be limited, so much so that it has been possible to suggest that most of them are nearly neutral (Kimura, 1968a,b; King & Jukes, 1969; Ohta, 1972). Moreover, one has been wondering why such fixations should occur with a regularity which, although quite far from perfect, still remains striking in contrast to what utter and generalized irregularity would be. This quasi regularity was referred to as the molecular evolutionary clock (Zuckerkandl & Pauling, 1965; Sarich & Wilson, 1967; Derancourt et al., 1967; Kimura, 1968a; Dickerson, 1971; Fitch, 1975, and a number of other papers).

As stated in the preceding paper, sites reserved for general functions are the most variable sites and therefore those that account for the largest proportion of substitutions. Let us consider the hypothesis that any conservative substitution that becomes fixed at a site for general functions, when it improves the adaptive state of the protein with respect to one of these functions, most often will at the same time slightly impair its adaptive state with respect to other general functions.

What could cause such a process to recur repeatedly? I should like to propose tentatively the following mechanism:

Selection coefficients are established for organisms as a whole in relation to populations. With respect to a given mutation, such overall selection coefficients should, however, be the resultant of component factors, both positive and negative: $S = f(s_1, s_2 \ldots s_n)$.

It has often been emphasized that selection acts on the individual, not on individual genes, - let alone on individual nucleotides in genes. In effect, selection *eliminates* individuals, along with all their genes. However, in sexually reproducing species, individual genomes are established only once (neglecting identical twins). Such a unique event cannot be called selection, since it may as well be due to chance, and since selection deals with groups of individuals, populations, and gene frequencies. What is selected, or otherwise established in a population, is a certain frequency of individual genes. This frequency is, for every gene, the expression, among other factors, of the presence of the other genes. But it also is, for every gene, the expression of its own structure and function. Viewing any gene in relation to the other genes of the organism represented a conceptual progress over what Ernst Mayr (1970) calls "beanbag genetics". Yet, dissolving, as it were, the individual genes in the genome, as though they did not have an impact on selection in their own right, would seem to be as wrong as it always is to go too far in the right direction.

Besides the individual organism, another unit of selection that has been considered is represented by *groups* of functionally coordinated genes (Sved et al., 1967; King, 1967; Milkman, 1967; Maynard Smith, 1969; Clegg et al., 1972). In fact, it is well known that selection coefficients may also be determined at the level of individual proteins, as illustrated by sickle cell anemia, - in relation, of course, to the overall genetic and environmental conditions. We go one step further here in a direction that some have in the past held to be erroneous and consider component selection coefficients as they theoretically relate *to a single amino acid site and to a single substitution at that site.* For example, such a substitution may have a small but significant positive selection coefficient s_1^+ in relation to isoelectric point, and three smaller negative selection coefficients, one s_2^-, in relation to solubility, another, s_3^-, in relation to the structural stability of a helical segment, and a third one, s_4^-, relative to the overall ratio of polar to apolar amino acid residues in the protein. It should be noted that, unlike many macroscopic pleiotropic effects of mutations, different functional characteristics of proteins come into play in a constant and concomitant fashion. A gene that increases resistance to desiccation, but lowers heat resistance will be selected for or against according to environmental conditions with respect to desiccation and heat.

The internal environment faced by a protein is more constant. Therefore a change in a physico-chemical property associated with a *general* function of a protein has an effect definable in principle not only in direction, but also in magnitude without reference to variable effects as exerted by environmental conditions. Whereas specific functions of a protein are directed at once at the internal and the external environment of the organism (e.g. oxygen affinity of hemoglobin will be related, inter alia, to the oxygen tension of the external milieu), general protein functions are primarily defined by a constellation of factors of the internal milieu.

This is not the first time that it is proposed that the total selective value of a gene be broken down into component terms. Sewall Wright has done so in 1956. He considered different characters of an organism as determined by pleiotropic effects of a single gene. We consider here the pleiotropic effects of one amino acid substitution on different functional characteristics of a protein.

Let us assume that in absolute value

$$|s_1^+| < |s_2^- + s_3^- + s_4^-|$$

and that the situation is that of co-dominance of the character. In this situation the substituent should be eliminated, since the sum of the negative components of the selection coefficient outweighs the value of the positive component.

However, in the schematic example, each of the negative partial selection coefficients is assumed to be very small. It will then behave, practically, as though it were equal to zero. The only somewhat larger single partial factor, which is positive, will exert its effect and lead to a fixation of the substituent in the population. It is thus proposed that an amino acid substitution is not selected for in linear proportion to its virtual partial selection coefficients when these are very small.

As substitutions are being selected on account of their advantage to *one* of the general functions of the protein molecule, very small virtual and partial negative selection factors will accumulate in relation to other general functions. In due course, their sum will become sufficiently distinct from zero to be significant. The sum of the fixations that are each one negligibly deleterious with respect to some general function or functions eventually leads to some sort of selection vacuum. This vacuum waits to be filled by a substitution that is favorable to the general function(s) concerned. Such a substitution will be represented by *one out of a set* of possible amino acids and may occur at a number of possible molecular

sites. The first such substitution will be selected for. Since it will occur by chance, *the substitution will have a limited but definite random character.*

Component selection coefficients characteristic of the action of a given substitution in relation to several general functions of the protein may not be entirely metaphysical. The measurement of such component selection coefficients might be approached if, under constant genetic and environmental conditions, a large number of "conservative" substitutions in a single protein molecule could be compared from the point of view of their selective effect: first a set of substitutions at a highly variable site of a protein; then, for each member of this set, substitutions at other highly variable sites. (For instance, the selection for a lysyl as the substituent for an alanyl at the surface of a protein should be stronger if a positive charge has been lost elsewhere on the molecule than if it hasn't; and still stronger if two positive charges have been lost). At the present stage of the genetic art we are not in a position to carry out experiments of this type within an organism, and nature does not carry them out for us to a sufficient extent, especially not against a constant genetic and environmental background. Such experiments should allow one to define increments for partial selection coefficients in relation to different general functions of the protein molecule.

A SCHEMATIC REPRESENTATION OF THE MODEL

For the purpose of illustrating the present model (Figure 1), let us start out from a situation where all general functions except one are in a state of adaptation very close to optimal and where one function of lower fitness is not very far from optimal. Let us make the simplifying assumption that the effect of any one mutation on a general function other than the one it affects maximally is very small and negative (in the direction of deadaptation), and further, that such very small effects, when the function was very close to optimally adapted at the start, will remain *cryptic*, i.e. not lead to detectable natural selection. Let us also suppose that under these conditions a second mutation with a very small effect on the same functions will still be cryptic, whereas a third mutation will bring the functional property into a zone of deadaptation where selection will come into play detectably. The third mutation will thus have on the function in question an *overt* effect.

Larger overt mutational effects will either bring a function from a state of significant inadaptation to a condition close to optimal, or from a state close to optimal into a zone of significant inadaptation. These last kinds of mutation will be

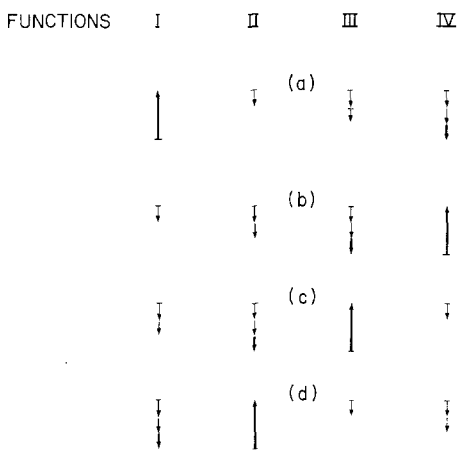

Fig.1
A schematic representation of the mechanism for selection of substituents under maximal global functional adaptation of the molecule.
I, II, III, IV represent different general functions of the protein. Under a, b, c, d the effects of four distinct successive amino acid fixations are symbolized. The fixations presumably occur at four different sites of the polypeptide chain. Each single substitution has effects on the four molecular functions I to IV and these effects are represented by arrows. Arrows pointing toward the top stand for favorable effects, arrows pointing toward the bottom for unfavorable effects. Small arrows represent small incremental effects, longer arrows larger effects. Thinly drawn arrows refer to selectively cryptic effects, thick arrows to selectively overt effects. For further explanation, see text

rapidly eliminated by natural selection. They are therefore taken out of the picture of the continuously "evolving" protein molecule. Thus the only mutations with large effects that remain part of this picture are those that take a function from a zone of significant nonadaptation to better adaptation.

The *overall* adaptation for the protein molecule that we consider is constantly close to its mean *maximum*, with a quasi-*optimal* adaptation realized for some of its individual functions. The environment of the molecule and that of the organism of which it is part must be postulated to be essentially *constant* (in the sense defined above), since otherwise the global adaptation of the molecule would not be maintained at or very close to its maximum.

Under these conditions, each mutation that will not eliminate the molecule will have three very small negative effects on three of the distinct general functions considered in Figure 1 and one larger positive effect on the fourth general function. Each of the successive mutations a, b, c, d, which occur presumably at different molecular sites, brings one significantly deadapted function back into the well adapted zone, whereas the three other functions will become very slightly less adapted. In one of these others, a third small increment is being added to two preceding cryptic increments. This third increment is then of overt selective significance. The mutation that brought the protein back into the adaptative zone from the point of view of one general function has at the same time deadapted it significantly from the point of view of another.

Yet, even though cryptic deleterious effects are constantly accumulating, in an overall way the molecule is maintained close to its state of maximal adaptation. A mean state of maximal overall adaptation is continuously approximated from above or below by mutational effects that take one or the other functional characteristic out of a selectively suboptimal zone, thereby pushing other functional characteristics towards this suboptimal zone. While the overall fitness of the molecule simply ondulates around a mean, the pictured cycle of events leads to a significant increase in fitness, at each evolutionarily effective mutation, with respect to *some* function of the molecule. It is with respect to this one functional characteristic that natural selection is assumed to act and to spread the mutation in the population.

DISCUSSION OF THE MODEL

The model thus presents an apparent paradox. In a sense functional optimality is constantly increasing; in a sense the best approximation to functional optimality is constantly maintained. The "normal" view would be that, in a constant environment, there should be no handle for selection whereby to spread in the population an amino acid substitution that happens to occur, unless this substitution brings the overall state of the molecule closer to optimal or maximal fitness. A previous, perhaps remote, change in environment may have lowered this fitness, and now the molecule is catching up gradually through a series of amino acid substitutions and is approaching optimality again.

This description certainly fits the real world, since real environments do change, at least from time to time, and thus should lower formerly optimal molecular adaptations besides optimizing formerly suboptimal adaptations, the overall effect on the molecule being in many cases a lowering with respect to maximal adaptation. This scrambling of partial molecular selection coefficients and lowering of the overall selective value of the molecule will then be mended progressively by incoming mutations.

Even though the present model can accomodate this realistic situation, it makes a point in going beyond. It purports to show how selection could proceed indefinitely, even at constant environment, and even at maximal *overall* molecular fitness. The rationale behind this proposal is the view that the balance between partial positive and partial negative molecular selection coefficients cannot be calculated by simple arithmetic because of the practical neutrality of functional effects that are very small. These only intervene with respect to natural selection when a further very small increment in decrease of

fitness brought the function into a weakly but overtly deleterious state. Although, in principle, in the mathematics involved here, 2 + 2 still are 4, in practice this is not so, because part of the contributions to the balance of partial molecular selection coefficients remain cryptic. The overt effects, which are alone instrumental, seem to negate these mathematics.

This paradoxical aspect is essentially removed, by pointing out that the model envisages a series of competitions between alleles, or rather the organisms that carry them, in which a negative interaction between alleles is possible. This will be discussed in a later section. At each turn of the kaleidoscope involving amino acid fixations at highly variable sites, the mutant will be significantly better than the wild type in (at least) one respect, and through a differential in reproductive capacity will replace the wild type. Though the new allele wins out, it also has its own disadvantage. Yet, at this point, its advantage is greater than its disadvantage and the latter thus does not decide the competition. The advantage gained will however decay during subsequent fixations. It will do so first in a nonsignificant way, namely not leading to a differential in reproductive efficiency. Thanks to these decays there is a basis for continuous positive selection at constant external environment. While the previously acquired advantage of the molecule is progressively reduced, the disadvantage gains in significance and a further new allele, apt to compensate for the latter will be selected. Fixations thus succeed each other in a circle around the different partial molecular functions involved.

The progression along this circle is evolutionary noise rather than a succession of events of evolutionary significance. We may conceive as a peculiarity of the biological world that selection and noise are not two categories to be opposed to each other. Selection itself can be noise. There is selection that is noise and selection that isn't. If the evolutionist remains interested in the noise component of molecular evolution, it is, in part, because of the value of this component as a tool. Without it, informational macromolecules would tell us less about evolutionary relationships, and therefore, probably, about the important evolutionary processes themselves. Also, the noise component deserves attention because it increases the potential for functionally significant processes through the multiplication of possible molecular pathways. Therein no doubt lies the significance of the "noise" itself.

The circle around partial molecular functions is of course an abstraction and is transgressed by reality. It is so, firstly, because the succession of the partial molecular function that selection addresses itself to will not be regular and the accumulation of cryptic and overt effects not sche-

matic like on the diagram of Figure 1. More importantly, the circle will be ruptured from time to time by significant evolutionary events. These will manifest themselves by changing the selected values of general-function parameters, values that, in the model, are assumed to be constant. The significance of the evolutionary events will however reside in modifications or innovations in specific functions. Changes in general-function parameters will presumably always be secondary to changes in specific functions, either of the polypeptide considered, or of other polypeptides in the molecular environment.

In Van Valen's (1974) theory, which proceeds along lines initiated by Darwin and R.A. Fisher (1958), the improvement of fitness of one species with respect to its environment is necessarily accompanied by a decrease in fitness of other species. In a parallel fashion, the proposal here is that the improvement of fitness with respect to one function of a protein mostly implies the impairment of other functions of the same protein. Like Van Valen's, the present concept is selectionist. Even the most conservative substitutions in protein molecules are deemed explainable in terms of natural selection. At the same time, however, the proposed model explicitly retains randomness in molecular evolutionary change.

The concept according to which an (approximately) constant state of functional *inadaptation* is approached in every type of variable protein by ever different structural means might be designated, to use Van Valen's (1974) phrase, as a molecular "Red Queen" hypothesis: "'Now here, you see', the Red Queen said to Alice, 'it takes all the running you can do to keep in the same place'". Keeping in the same place of course is the opposite of evolving. All proteins "run" all the time, at a different pace. Few evolve, and rarely, if the word evolution is not to be a mere synonym of change. One may consider that *only substitutions that modify weighted functional density,* or some related parameter, *have evolutionary significance*. The evolutionary noise we referred to are the substitutions that do not result in a change in weighted functional density. Such a change may be a decrease, through the modification of a specific function or its substitution by another; or it may be an increase, accompanying the addition of a new function or the increase in specificity of a preestablished function.

In molecules with a very large majority of specific-function sites, such as cytochrome c, not to mention histone IV, Alice's running on the same spot is replaced by walking or creeping on the same spot, probably without any impairment to the organism. Indeed a slow optimization process of any one molecular function is as good as a fast one, or none, when maximal overall molecular fitness is already reached. To be sure, if a change in the environment (external and internal) entailed a redefi-

nition of functional optima, slowly "evolving" proteins would be slower than others in regaining maximal overall fitness. This disadvantage would however be a purely virtual one: no faster "evolving" molecule filling the same specific functions could be at hand, none could therefore turn the virtual disadvantage into a real one through competition with the more slowly evolving protein. The question is whether the species survives. If so, it does it on the basis of the "imperfect" molecule, and there may be no possibility of detecting and assessing the imperfection. The imperfect molecule would be *sufficient* (see below).

The present model ignores random drift. It does so, not in order to suggest that random drift is not important, but to show that it is not *necessary* for accounting for the many apparently trivial changes in primary structure of proteins during evolution. It is to be anticipated that the proposed mechanism, even if correct, accounts only for an undefined, and not easily definable, fraction of evolutionarily effective amino acid substitutions. An abundance of further mechanisms is available (Mayr, 1970; Lewontin, 1974). The model states that all mutations at general function sites in proteins are nearly neutral in *some* functional capacities. Kimura (1968a,b) and Kimura & Ohta (1971) consider mutations whose *overall* effect on the protein is practically neutral. The concept of a tight competition between several functions for the same amino acid sites would however make the chances of substitutions with overall neutral effects appear small.

Yet, in the preceding paper (Zuckerkandl, 1976a), the functional impact of a substitution such as aspartic acid for glutamic acid was considered. In the case of this and a few other substitutions, the effect of the small change in the number of apolar atomic groups should at times be so slight in relation to the whole molecule that it remains selectively cryptic. Charge is not modified and, the substitution occurring at the surface of a protein, possibly no other parameter is significantly modified. The substitution may thus be neutral when it occurs. If so, then, in later evolutionary history, it will contribute to the fact that another - perhaps identical - substitution will *not* be neutral. Indeed, any accumulation of small effects of the same kind may cause one further equally "insignificant" substitution to exert an overtly unfavorable effect.

Within the cycle of functional adjustment suggested by the model, a certain peppering of the selected mutations by neutral mutations must therefore not be excluded.

We have been dealing so far with successive adjustments of individual general functions through amino acid substitutions at highly variable sites. Is the mechanism for natural selection under constant environment to be related essentially to the general as opposed to the specific functions? In fact it is more basically related to a cognate pair of notions, namely to highly variable sites as opposed to sites displaying little or no evolutionary variation.

General functions furnished the best initial illustration of the model. There it no difficulty indeed in accepting the probable smallness and relative constancy of increments in general-functional properties as a result of amino acid substitutions at highly variable sites. Also the mutual involvement of general functions through the functional "degeneracy" of amino acid structure is patent, whereas such a mutual involvement between specific functions of the molecule should be less constant.

It must now be made clear, however, that changes in general functions necessarily involve specific functions as well, and that small and fairly reproducible incremental changes may occur also with respect to specific functions, provided the effects on specific functions are exerted via general-function sites. Specific functions should thus be considered to be themselves part of the cycle of functional adjustments.

It would not be realistic to suppose that a substitution that has any functional effect at all could exert that effect on general functions alone. The adequate general functional properties of a protein must be determined by its specific functions, as well as by the specific and general functions of other proteins. If this be true, then, conversely, any change in general functional properties must imply a change in some specific function(s) also. As an example, consider a reduction in solubility of a polypeptide chain by replacing polar by apolar residues on its surface. Although none of the residues carrying out specific functions may be affected, and although intrinsically the potential for carrying them out effectively may be intact, insolubility of the molecule would interfere with activity at the level of the specific functions.

Once we accept the generality of an interaction between general and specific functions in both directions, the question of the sensitivity of the specific functions to incremental changes in general functions remains. A conservative substitution - namely the fixation of an amino acid belonging to the set of those frequently accepted at a given general function site - may, even with respect to the most directly implicated specific function, have an effect so slight that it remains cryptic.

Yet the n^{th} substitution (with n small) affecting the same specific function should change the effect into an overt one. From this point of view, the specific functions will behave like the general functions.

They will, likewise, from a second point of view. Can a distinction between general-function sites and specific-function sites be maintained if the general functions react on the specific functions? It can be, on the grounds that the involvement of specific functions at general-function sites can be distinguished from their involvement at specific-function sites. At specific-function sites, characterized by lowered, very low, or zero variability, substitutions, even if of a type that would be characterized as conservative at general function sites, will have effects on at least one specific function ranging from moderately strong to very strong. Many mutations, even of the "conservative" type, will largely or totally impair a specific function. On the other hand, conservative substitutions at general-function sites should involve specific functions much in the same way they involve general functions. Like general functions, specific functions presumably will be altered by small increments. They will not be seriously impaired by any possible single substitution at general-function sites, if the substitution is of the conservative type, in the sense given of a frequently accomodated substitution at highly variable sites.

Under these conditions, specific functions, along with the general functions, participate in the kaleidoscopic cycle of functional adjustments, described by the present model as occurring even under constant environmental conditions and even though the overall functional performance of the molecule remains close to its adaptive maximum throughout.

The model thus is not specifically based on general-functional adjustments, but on adjustments of what may be termed *variable site functions.* The term refers to the set of most variable sites in every polypeptide chain, and the functions include the general functions and a particular class of effects - particular by its moderate and incremental nature - on specific functions as well.

It should be pointed out that adjustments in specific functions may be considered evolutionary noise just as much as adjustments in general functions, as long as they are part of a cyclic process that is occurring under constant environment, or would be occurring even if the environment remained constant.

THE POSTULATE OF CONSTANCY OF THE ENVIRONMENT

Critics may consider the present model indefendable because founded on an unrealistic postulate, that of the constancy of the external environment. This would not be a weighty objec-

tion. Not only can constant environments be found on earth, exceptionally (see below); but it may happen that aspects of reality can be isolated theoretically, and therefore perceived and understood, only on the basis of an unrealistic assumption, just as they can be isolated experimentally only by creating conditions that are not found in nature.

Another criticism of the model would seem more percussive at first sight. It may be pointed out that extending the postulate of constancy of the environment to the *internal* environment would not only be at odds with the facts, but contradict implications of the present model itself.

In contradistinction to the variability of the external environment, that of the internal environment of organisms, over time spans compatible with evolutionary change, apparently suffers no exception. The continuous and universal evolutionary divergence of the primary structures of DNA and proteins furnishes the central piece of evidence in this respect. Even the so-called living fossils continue to change with respect to the sequence of their proteins (Stenzel, 1974).

Postulating that sequence changes are continuously selected for, even under constant external environment, amounts to postulating that, under constant external environment, the internal environment is continuously modified.

This is so, but in one sense only. On the basis of the same model, one must point out that the changes in internal environment under constant external environment represent the collective molecular aspect of what we have been referring to as evolutionary noise. We do have continuous structural change and functional adjustment, but not functional change in the sense of a redefinition of functional states around which the cyclic adjustments would revolve. If we disregard noise, then we should say that the present model implies that, under constant external environment, the internal environment also remains constant. This consequence of the model can be confirmed or refuted by observation. One wants to look at closely related species or subspecies that contain and select different structural variants of proteins, yet share identical habitats and ecological conditions. Are the *physiological* functions controlled by these different protein variants identical or not? They should be, if the model is correct.

A typical and frequent change in the environment of a polypeptide, call it ppI, is a change in polypeptide II (ppII) with which ppI is to interact. If the substitution in ppII affects one of the sites instrumental in the interaction, both ppI and ppII will see their functional fitness decreased. The "selection vacuum" created here is relative to a specific function, namely to a contact function, and it is precisely around specific contact functions that the kaleidoscope of a changing internal milieu under external constancy is likely to primar-

ily revolve. This is then an example of the substantial participation of specific functions in evolutionary noise. Though with respect to ppI the selection vacuum to be filled was created by an external and not by an internal molecular event, this circumstance can be absorbed into a constant-environment model (featuring external *and* internal constancy) if the unit considered is not the isolated polypeptide chain ppI, but the set ppI - ppII of interacting polypeptide chains.

Generalizing, one may consider all interacting proteins together as forming a virtual "superprotein" from the evolutionary point of view. What was said about the individual protein molecule applies to the superprotein complex: under constant external environment (if that constancy has been in effect for a significant evolutionary time), the superprotein complex will have reached a state of maximal functional adaptation. Yet this state will not be optimal from the point of view of each separate interaction between pairs of interacting proteins. Improvements in individual interactions will constantly be selected for, while other molecular interactions will very slightly deteriorate, at first covertly, then overtly with respect to their significance for selection. The covert effects are of course synonymous with nearly neutral effects but, as stated before, the model does not consider (without excluding their existence) substitutions that are nearly neutral in *every* functional respect.

Because of the situation described, different structural editions of a given polypeptide chain may be selectively preferred in closely related organisms, as is in fact sometimes the case (Ayala & Anderson, 1973; see below). This does not imply, the point may be emphasized, that the specific functions to be carried out by these proteins are different in the two subspecies. It may only mean that the same functions are exercized by slightly different structural means, because the turns of the functional kaleidoscope with respect to individual proteins leads to the selection of different structural states. These different states would be appropriate for interactions that continue to fill the same functional roles in both organisms.

GENERAL REMARKS ON FITNESS AND ON SELECTION UNDER GENETIC SUFFICIENCY

Some peculiarities of the proposed mechanism prompted its detailed exposition and discussion. Yet the way it is supposed to work cannot be clarified sufficiently without examining what the notions of fitness or adaptation stand for. Even though the remarks that follow are general, they seem important for a proper focusing of the model itself.

When an allele leads to more offspring than another, it will be said to have higher fitness, and it will replace the other allele. The term adaptation is sometimes used as a synonym. Perhaps it would have been better to apply to the moment thus designated a term with the fewest possible general and integrative implications, such as for instance reproductive differential or reproductive value. In their general biological usage, fitness and adaptation are fuzzy, yet important notions, difficult to handle, and a one-sided mobilization of the words may not favor clarification.

In their population genetical usage, fitness and adaptation designate the "success" of a certain genotype in a certain population as a sheer numerical success. Is numerical success to be equated with biological success generally?

If it is, then, from differential to differential, we shall reach higher absolute numbers per population and per species. The most successful genotype, from the point of view of evolution, would be the one that leads to the largest number of individuals in existence simultaneously and capable of reproducing. It is perceivable that essential biological realities are left aside in such a concept and that some are even contradicted.

If the highest reproductive rate of a genotype is the ultimate criterion for evolutionary success, I can see no valid reason to limit this criterion to intraspecific comparisons and not to extend it to genotypes generally. Prokaryotes would then appear as the most successful organisms in evolutionary history. Bacteria would be more successful than oysters, and oysters more than apes. In fact, man having become about as successful as an oyster in terms of population size, he is going to have ample occasion to reflect on how "successful" being more and more of one's own kind really is. Outside of humans, frequency dependent selection illustrates this point (e.g. Huang et al., 1971).

There may be an inverse relation between the level at which an organism finds itself on the scale of progressive evolution, anagenesis (cf. Zuckerkandl, 1976b), and the minimum number of individuals needed for insuring survival of the species. If, on the other hand, number of offspring are a measure of evolutionary success, we are caught in a contradiction: evolutionary progress, as it has been defined in serious attempts at biological objectivity, is then to be equated with lack of biological success.

Progress is, by its nature, a comparative notion. In a general, not in a population geneticist's sense, fitness and adaptation are *not* comparative notions, from a certain threshold on. In a wider biological acceptation, for which a word is needed as much as for differential reproductive

value, success, adaptation, fitness cannot be measured by number of offspring. It cannot be measured at all. It is an all-or-none effect: survival.

Obviously, it is not survival of the individual that is the useful criterion from the evolutionary point of view, but the permanence of the species. In the absence of a continuous trend toward reduction of population size, there is no such thing as a greater or lesser adaptation to a given environment. If the species survives, it is neither perfectly nor fairly adapted, it is sufficiently adapted and all surviving species are equally so. There certainly are countless ways of achieving *genetic sufficiency*. The threshold for this survival value that cannot be graded will be reached by any genotype that is sufficient for insuring the permanence of the species in a given and constant environment (defining constancy as constant variability, as stated), without any trend toward extinction ensuing. *Evolutionary* fitness of an allele is thus to be estimated in a binary system.

It is of more than verbal advantage to replace the notion of genetic optimality by the notion of genetic sufficiency. The notion of optimality of fitness was rejected also by Ernst Mayr (1970). According to a misconception of natural selection, the replacement of an allele by another under selection demonstrates that the eliminated allele had a lesser survival value for the species, that its genetic sufficiency, in a binary system, was zero. In fact, a large number of different alleles are sufficient in the sense that each of them, if alone present, would insure the permanence of the species in its environment.

Thus genetic diversification within species under constant environment, as well as the diversification between species, are to an important extent not adaptation at all, but the replacement of one solution of genetic sufficiency by another or the juxtaposition of different genetic sufficiencies. Much of the diversification of living forms, we postulate, was not called for by any necessity for insuring survival. It happened only, like the spreading of neutral alleles, by virtue of appearing on the scene, although it spread presumably through selection. Superimposed on this process was the selection of adapted variants as the environment changes, and, occasionally, the pressure towards anagenesis that I commented on elsewhere (Zuckerkandl, 1976b).

A human parallel to the replacement of some alleles by others, or of a change in their proportions, when the original alleles needed not to be improved and were "perfectly adapted", i.e., genetically sufficient, is offered by the situation of a native population that had the potential for indefinite survival, but was partly or totally destroyed by an invading race.

When two sufficient genomes, or two sufficient alleles are placed in competition with one another - still keeping the external environment constant, in the sense defined above - more often than not one of the two probably eliminates the other or becomes predominant under polymorphism. Selection for a particular genome or a particular allele thus does not necessarily mean the establishment of a better adapted organism. Realizing the possibility of such a situation, and of its potentially general significance, should lead us to accept more easily the idea, expounded above, that a slightly changed "edition" of a certain protein may replace another edition through selection without being really a "better" protein insofar as insuring the permanence of the species goes or even with respect to its reproductive rate.

Concerning this last point, for selection to lead to the replacement of one allele by another, the successful allele indeed does not necessarily have to determine an absolute increase in population size. A relative reproductive superiority at constant, or even at decreasing, population size would be sufficient. Such a situation arises if the appearance of the successful allele brings about an absolute decrease in the reproductive rate of the organism possessing the losing allele. A great many different effects of competition could lead to such a result, as well as some modification in the internal and external environment brought about by indirect effects of the action of the successful allele. The mutant may thus replace the wild type, not because it is reproductively better, but because it makes the environment for the wild type worse with respect to its reproductive efficiency. The mutant may of course also bring about an absolute increase in population size. But under constant external environmental conditions, this is of no advantage to the permanence of the species.

An increase in absolute population size *is* however of advantage to the survival potential of the species (up to certain limits) if the external environment is to change critically. Indeed, a larger population means a proportionately larger genetic variance and therefore an increased chance of the presence of genotypes of genetic sufficiency under stress conditions. The magnitude of this advantage will vary in proportion to the intensity of the stress and in inverse proportion to the size the population already has. In many cases the advantage of a small positive differential in reproductive value, sufficient to lead to the replacement of one allele by another, will be negligible with respect to the permanence of the species.

In some cases the advantage should, however, be critical. Perhaps the giant sauriens disappeared because their necessarily reduced numbers, given their sizes, did not provide

groups of genetic sufficiency under stress. It is only when a species is in danger of extinction that fitness, as defined by population geneticists, coincides with evolutionary fitness in the sense of survival value for the species and lineage. However, even very large population numbers, and consequently large genetic diversity, will not always insure survival of either species or lineage, as the disappearance, for instance, of the trilobites and ammonites demonstrates.

At any rate, most evolutionarily effective changes may win out, not because they are better for the species or the individuals that compose it, in any meaningful sense of the word "better", but for the exclusive reason that they occurred. Having occurred, the result is either that the reproductive rate of the mutant is greater than that of the wild type was, or the presence of the mutant results in reducing the reproductive rate of the wild type. Complete or incomplete substitution thus takes place, according to whether or not heterosis or frequency dependent balanced polymorphism, often with respect to yet other alleles, is superimposed on the situation or not. Yet by this complete or incomplete substitution the species in general is made no more capable of permanence than it was.

To fully understand the model which was described in the previous section, we may suppose that any of the alleles considered was genetically sufficient as long as the next successful allele had not appeared. Though we assume that the external environment has not changed, in the presence of the mutant the previously dominant allele is now at a disadvantage. The mutant is to share the same fate at the next turn of the functional kaleidoscope. There always is, in the model, a net effect with respect to relative reproductive potential of the genomes carrying the new allele, even though there is no absolute sense in which any of these alleles can be said to be "better adapted". In all these cases one allele becomes genetically insufficient exclusively through the appearance of another.

It is proposed that this obtains in nature in a significant proportion of cases of total or predominant fixation of an allele. It should obtain, at any rate, in all cases the model expounded in this paper addresses itself to. These cases, as stated, relate to evolutionary noise. To the extent nearly neutral substitutions do exist, they add to this noise. The total evolutionary noise, which thus appears as the sum of nearly neutral and of a category of selected substitutions, should represent a considerable fraction of the evolutionarily effective substitutions. It seems to be implicitly assumed that a substitution that is selected for is important, in

contradistinction to a neutral substitution. The point made here is that most selected substitutions are no more and no less important than the neutral ones.

Since the expression "selective advantage" is used in connection with any selected allele, the suggestion is that every selected allele has a functional advantage for the species over every eliminated allele. If, in fact, in many cases, the eliminated and the selected alleles, as long as they occur separately in isolated populations, may both be genetically sufficient, then the term selective advantage applies to only a fraction of the selectively successful alleles, and the phrase "selective *prevalence*" is to be preferred as an expression of general applicability.

In the light of these considerations on evolutionary fitness, a definition of genetic load as "the extent to which a population departs from a perfect genetic constitution" (Strickberger, 1968) is clearly inappropriate, since there is no perfect genetic constitution in relation to an environment, there are only genetically sufficient and insufficient constitutions in relation to it. Let us consider a definition of genetic load to which this criticism does not apply, namely "the average number of lethal equivalents per individual in a population" (King, 1968), a definition going back to the treatment of the problem by Haldane (1957), who wrote about the "cost of natural selection" in terms of deaths or equivalent number in lowered fertility. Since a genetically sufficient constitution of an organism may become insufficient through the mere presence of a new allele, the regression in or disappearance from a population of an allele is not convincing evidence that the overall genetic load has increased. As stated, many genes that lose out in the competition with certain alleles are by no means for this reason intrinsically deleterious for either the individual or the species. In fact, the mutant that replaces the wild type in a certain environment might lose out to this wild type when the environment changes: yet *both* alleles in *both* environments, as long as the other allele is not present, may be equally compatible with the survival of the species and even with similar reproductive rates.

In a situation of selection under constant environment, with one state of genetic sufficiency eliminating another, a substitution of one allele by another mostly occurs at no cost for the species. The allele that is being eliminated is replaced in proportion, or above proportion, by the substituent allele. Any fertility equivalents of deaths are compensated by fertility equivalents of births. It is true that the losing allele may, in some cases, be replaced below proportion. This would happen in the minority of cases when the effect of the new allele on the reproductive potential

of the old allele is notably deleterious, while the new allele has a smaller such potential than the old allele had before the new one appeared. In such cases the replacement would contribute indeed to genetic load.

If we are correct in assuming that such cases are a minority, perhaps a very small minority, then, according to the present concepts, it can no longer be said that, on account of genetic load, amino acid substitutions in proteins have occurred too fast for being attributable to natural selection[1] and that therefore the majority of alleles that spread in the population must be selectively nearly neutral (Kimura, 1968a). At the same time, in denying load as well as evolutionary advantage conferred by the substituent allele, what we describe is selection that behaves like neutrality.

If the present model of selection in protein molecules is founded, it may be anticipated that a large proportion of the conservative substitutions in proteins, i.e. the majority of all evolutionarily effective mutations in structural genes, do *not* contribute to genetic load, even though they are the target of selection and will replace each other in populations.

There is no reason to suppose that the situation will be found different for controller sections of the genomes. Regulator genes that control the sequence of regulator protein molecules are, of course, at the same time "structural genes", in that they are at once transcribed and translated. Receptor genes, DNA sequences that interact specifically with such proteins, come under the heading of variable ligands (see paper I of this series). The mean rate of substitution in controller genes is likely to be quite different according to whether one considers regulator and receptor genes, or further segments of DNA presumably involved in the control of transcription (e.g. transconformational DNA, Zuckerkandl, 1974). This mean rate will differ also from the mean rate of substitution in structural genes. In other words, these different sectors of DNA are likely to be subject to a different amount of evolutionary noise. Yet the basic mechanisms of fixation of substitutions are probably the same, though the different mechanisms may contribute in different proportions. To establish these proportions is indeed a formidable experimental problem.

---

[1] This position was also rejected long ago on other grounds by King (1967), Sved et al. (1967), and Milkman (1967), in answer to a question raised by Lewontin & Hubby (1966). The latter authors wondered how their findings to the effect that there was very much more polymorphism in populations than expected by most geneticists could be reconciled with the genetic load argument. The concept of genetic load has also been criticized by Ernst Mayr (1963, 1970). Ways out of the "load paradox" have further been defined by Maynard Smith (1968) (see however O'Donald, 1969), Clarke (1972), Ewens (1972) and further authors (see, e.g., Ewens, 1972).

The very large amount of multipolymorphism that can be predicted for regions of DNA involved in regulation (Zuckerkandl, in preparation; predicted independently by Jack L. King, in press) as well as the high amount of polymorphism in structural genes are presumably compatible with selection without any excessive genetic load resulting, - in part on account of the considerations put forth here.

Ernst Mayr (1970), in describing the contradiction to which the genetic load argument on the one hand and the observation of multipolymorphism on the other has led, sees the solution to the problem in the fact that there is no selection except at the level of the individual. Here a different proposal is made, namely that there *is* selection at the level of the gene and even of the nucleotide, but that in most cases no net genetic load results.

THE TREND TOWARD POLYMORPHISM UNDER CONSTANT ENVIRONMENT

The concept of genetic sufficiency of individual alleles of course implies their sufficiency in homozygous condition. A sufficient allele will endow a monomorphic species with a potential of indefinite survival in a constant environment. This quality of sufficiency will be preserved against a certain number of different genetic backgrounds and, for each genetic background, against a certain number of different environmental conditions. It is not clear how individual alleles vary with respect to the span of conditions that are compatible with their sufficiency. Are there alleles that remain sufficient (in the homozygous state) under a wider variety of genetic and environmental circumstances than others? Or is the span of conditions compatible with sufficiency approximately constant, at a given locus, for all sufficient alleles in the homozygous state and varies significantly only as a result of a combination of alleles, namely of polymorphism?

Whatever the answer to this question, it appears that "sufficiencies" of individual genes will not add up to the sufficiency of a genome when environmental conditions are *changing*. A significant proportion of multipolymorphism probably is necessary for endowing a species with an increased potential for responding successfully to environmental change. This should hold for frequency (density) dependent selection as well as for heterosis and diversifying selection (see Pasteur, 1974). In changing environments, the question of genetic sufficiency thus is not the same at the level of the individual allele and of the genome as a whole. Even though given individual alleles may still be sufficient in the homozygous state,

their participation in polymorphism may contribute to increasing the life expectancy of the species, and thus alleles that are individually sufficient may not be collectively sufficient.

However, on the basis of the concepts presented, even constant environments would provide the opportunity for polymorphism to develop through natural selection in an inbred strain. Monomorphism no doubt limits the range of tolerated population densities more than polymorphism does (Clarke, 1972). At any rate, since significant shifts in population densities amount to significant shifts in the environment of the members of a species, major changes in population densities are excluded from the model. It is assumed that the necessary diverged gene duplicates and genetic regulatory systems have been evolved for a homozygous population to be able to cope with the different phases of a stable environmental cycle. Polymorphism, according to the model, is selected nevertheless and in spite of the genomes being genetically sufficient in its absence. Alleles that are being selected within the proposed kaleidoscopic functional cycle may be on their way to fixation. Their proportion must not be assessed on the basis of electrophoretically detectable variants alone. Other alleles will not have totally eliminated their older predecessors, due to various effects of balanced polymorphism. Such effects are as plausible under conditions of genetic sufficiency as the complete displacement of an allele by another. They are thought to lead to the reproductive differential whereby the previously monomorphic state of a locus becomes polymorphic, even though this is of no advantage to the survival, under continuance of prevailing conditions, of an already "sufficiently" adapted species. Like the complete displacement of alleles, polymorphism, in such circumstances, manifests selective prevalence rather than selective advantage.

Generally speaking, one might expect that the protein sites that are most highly variable in evolution will also be those that are most highly variable in populations. To what extent this is so depends however on how significant differences in *specific* functions must be for balanced polymorphism to arise. This extent is at present unknown. It will be of interest to learn whether heterosis, for instance, is more often obtained through single substitutions at the most variable general-function sites of proteins or at less variable sites; and whether substitutions at the most variable sites lead to heterosis more often when they are of the most conservative type (measuring conservatism by frequency of occurrence of a substituent at general function sites) or when they are less conservative. If the latter terms of the alternatives obtained,

it would mean that heterosis arises mainly when specific functions are more significantly affected than they are by a single conservative substitution at a general-function site.

There are usually many more polymorphic "editions" of a protein than can be expected to appear along a phylogenetic lineage. It is probably justified to consider each evolutionary fixation as a random choice among alleles that were successful as heteromorphs[2] (if the locus under consideration was polymorphic). Which of them is most likely to appear on the molecular phylogenetic tree will depend on which allele happened to be predominant in the population that insured phylogenetic descent. This presumed random choice may not significantly alter the degree of randomness that is introduced into molecular "evolution" independently of polymorphism (see above). Alleles that are present simultaneously in a population as heteromorphs might sometimes appear as evolutionary fixations in closely related species, if indeed the evolutionary "kaleidoscope" of sequence change plays predominantly with the same sequence elements, the same cycle of functional adjustments, as polymorphism does in populations.

It will be of interest to inquire about the existence of an inverse correlation between the functional density of a protein and its tendency toward polymorphism. The larger the number of general-function sites there is in a protein, the larger the number of accepted heteromorphs there might be in a species and in closely related sympatric species. It is possible that this applies to fish hemoglobins. Teleost as well as elasmobranch fish hemoglobin chains are apt to have a lower weighted functional density than hemoglobin chains from higher vertebrates, as suggested by the frequent presence of more than two different chains per tetrameric molecule (cf. Flynn & Sullivan, 1974; Martin, 1974). A lower specificity of subunit interaction seems to be implied. These hemoglobins are also particularly polymorphic (same authors), although a good deal of their structural multiplicity must be due to nonallelic genes.

As has become apparent now (Boyer et al., 1972; Bernstein et al., 1973), electrophoretically distinct alleles, the electromorphs of King & Ohta (1975), may actually represent a composite of heteromorphs. It cannot be decided at present to what extent such electrophoretically cryptic heteromorphs represent functional adjustments and thus belong to the category of cases covered by the present model, and to what extent they may be neutral. If, in proteins, there is a competition between function for sites, as postulated here, the first alternative will predominate.

---

[2]One may call heteromorphs structurally distinct variants of the same polypeptide chain controlled by allelic genes.

Even within the boundaries of the present selectionist model, the neutrality of alleles may be expected at times, in a special sense of the word. The organism may be neutral with respect to the decision which of two or several equivalent alleles to select. If, as stated, the kaleidoscope of perhaps nearly inexhaustible combinations of conservative fixations at sites predominantly involved with general functions is being turned indefinitely by a force independent of any changes in the environment; if this force expresses the permanence of molecular inadaptation in relation to even a constant environment; if, for each particular state of inadaptation, several different substitutions at several different molecular sites may in many cases bring one given function closer to its optimal state of adaptation, though in different ways; then natural selection will not "sense" the difference in these ways. If any amino acid substitution out of a certain number is equally effective in improving a state that has deteriorated with respect to a particular function, the selectively less prevalent allele may be replaced by, say, two different ones instead of by just one. (Of course the two substitutions are not supposed to occur in the same individual, and the effective population size must be large for the chances of their occurrence to be significant.) These alleles would be "neutral" with respect to one another. They would, of course, tend to progressively diverge. Their selective equivalence would be ephemeral. But evolutionary ephemerality may involve significant lapses of time.

Ernst Mayr (1970) saw in the selective equivalence of genotypes "perhaps the greatest single source of randomness and indeterminacy in evolution". He refers to different assortments of genes that react in an essentially identical manner to a given selection pressure. We consider here, from the same point of view, different assortments of nucleotide substitutions in a gene.

COMPATIBILITY OF THE MODEL WITH OBSERVATION

A direct check of the model requires observations in environments remaining constant over long evolutionary periods. Such environments are difficult to come by. There is, however, notably, the deep sea. It is considered "the most stable and homogeneous habitat of major ecological significance on earth" (Gooch & Schopf, 1972; Sanders, 1968). Remarkably, values for polymorphism in deep sea organisms are comparable to estimates for species inhabiting heterogeneous, unpredictable environments (Gooch & Schopf, 1972; Ayala et al., 1975; Valentine & Ayala, 1975).

There is, then, genetic variability of which it is difficult to say that it is maintained as an adaptation to environmental heterogeneity in time and space, although occasionally the existence of a pressure gradient has been successfully invoked. The general validity of this explanation is not established. On the other hand, the neutrality hypothesis does not seem to be a likely explanation of the extensive polymorphisms in the constant environment, in view of the fact that species with a very large effective population size do not seem to be more polymorphic than species with a smaller population size (Gooch & Schopf, 1972). The findings "suggest that the body of ecological-genetic theory that predicts a positive correlation of genetic variability with environmental instability is incorrect" (Valentine & Ayala, 1975). On the other hand, the observations on deep-sea organisms are consistent with the view proposed here that changes in the environment are not the only and are not a necessary ultimate motor for evolutionarily effective amino acid substitutions, and that this statement nevertheless does not imply the selective neutrality of such substitutions.

Another type of data must be referred to in connection with the present model, namely the geographical distribution of alleles. Since the demonstration by Lewontin & Hubby (1966), following the prediction by Wallace (1958), of a large amount of polymorphism in *Drosophila*, and since the similar finding of Harris (1966) of multipolymorphism in humans, studies in this field have multiplied. Nevertheless the respective contributions of the causal factors implicated in the observed distributions have not yet been assessed convincingly (cf. Lewontin, 1974), even though natural selection emerges from the work of Ayala and his collaborators (e.g. Ayala & Tracey, 1974) as a favored factor in *Drosophila* polymorphism (as revealed by electrophoresis). The work of Selander and his associates on the house mouse and related rodents has provided sets of data that are particularly suggestive in connection with the present model. They are relevant to this model by at least four aspects: the variation in degree of polymorphism; the absence of clear correlations between patterns of geographic distributions and physical factors of the environment; the patchiness of distributions; and the fixation of different alleles in subspecies sharing common ecological conditions.

The geographic variation in degree of heterozygosity was studied in a species of the genus *Peromyscus* among continuously distributed populations (Selander et al., 1971). The authors conclude that the causal basis for the variation they observe is not readily apparent in an adaptational, evolutionary sense. They propose a correlation between degree of heterozygosity and width of the ecological niche: ecologically constrained populations would have lower levels or variation. If the con-

straint is not on numbers of individuals that can be accomodated in a type of niche, but on ecological diversity, it is difficult to sustain the hypothesis in the light of the high degree of polymorphism recently found in deep sea organisms (see above). According to the present selectionist model, erratic variations in degree of polymorphism would be expected in conformity with the local history of subpopulations and the degree of migration between subpopulations. A subpopulation that descends from an old subpopulation with high effective population size will have a high degree of polymorphism, since many mutations will have occurred and a number will have been selected. On the other hand, subpopulations that have emerged from local population bottle necks and that either have originated rather recently, or have emerged a long time ago but remained small, will show less polymorphism.

In some isolated areas the mouse *Peromyscus polionotus* has been shown to display levels of genic variation that are "by far the lowest reported for natural populations of organisms" (Selander et al., 1971). Yet there seems to be no sign that these mice that are so much more homozygous than other populations of the same species are in a state of biological insufficiency and threatened by local extinction, - at least as long as their environment does not change. In fact, experiments with several species do not fulfill the prediction of a high degree of inbreeding depression (Clarke, 1972). This checks with the idea that multipolymorphism, under constant (constantly variable) environment, is indifferent with respect to the permanence of the species, even though it is selected for as soon as the opportunity arises.

Little success was achieved, even in the best studied region (Texas), in attempting to relate geographic patterns of genic variation in the house mouse *Mus musculus* to patterns of variation in particular climatic factors. The primary correlates seemed to be latitude or longitude per se. Some correlations with, for instance, temperature, seemed to be secondary to the latitude/longitude dependency (Selander, Yang, and Hunt, 1969). In terms of the present concepts, this can be interpreted as diffusion of alleles rather than determination of alleles by environmental factors. Certain alleles get established as they reach any of the zones examined. They win out selectively when they happen to be present, but are not missed when they are absent.

If an old allele will be satisfactory at geographic sites where a new, successfully spreading allele has not appeared, one would expect that the distribution of alleles remains patchy over different and sometimes considerable periods of time, in proportion to the tightness of geographic or ethological separations of subpopulations. Mosaic gene-frequency distributions are predicted in the absence of selection

(Malecot, 1959, quoted by Lewontin, 1974; Kimura & Weiss, 1964). They are, however, also predicted on the basis of the present selectionist model. In nature, such distributions were found in the house mouse by Selander, Yang and Hunt (1969).

Selander, Hunt and Yang (1969) analyzed polymorphism in two subspecies of the house mouse occupying identical types of habitats in various regions of the Jutland peninsula. At 6 out of 17 polymorphic loci, different alleles were found to be fixed (or nearly so) in the two subspecies. This degree of genetic difference occurred in a region over which there is no apparent variation in the external environment. The authors propose, after Mayr (1963), that the difference is due to a different genetic background in the two subspecies: an allele that is selectively favored against one of these backgrounds is not so favored against the other. According to an alternate interpretation offered here the difference is accidental. One or the other allele happens not to have appeared by mutation in one of the subspecies. In both subspecies, the allele, whichever it is, has been selected, and both species carry on with the alleles they have, under genetic sufficiency.

The possibility that the genetic background reverses the ratio of selection coefficients relating to two different alleles is however equally compatible with the model, as was shown above. This possibility is strongly favored in the instance studied by Ayala & Anderson (1973). Here, like generally in this field, the question is not so much whether a proposed mechanism exists - several of them are likely - but in what proportion of cases it plays the predominant role. The authors showed that closely related species of *Drosophila*, reared for many generations in laboratory cages, will not select the same electrophoretically detected malate dehydrogenase allele as the dominant allele in the population. Taking this result at face value, it does of course not rule out the possibility that the nonselected allele would be "sufficient", in nature, in the absence of the selected allele. Even if and when the selection of a particular allele is highly sensitive to variations in the genetic background, as occurs between two closely related species, genetic sufficiency may be considerably less sensitive to such variations.

For a proper interpretation of results obtained in the course of work of the kind that was initiated by Ayala & Anderson (1973) several conditions are to be fulfilled. It must be shown that the results are not due to linkage disequilibrium, as mentioned by the authors. In each set of experiments, it must be known whether the so-called identical alleles are identical in sequence and not just identical in charge, and whether they represent unique heteromorphs or sets of heteromorphs. Also, a number of experiments of this type

should be done by working with alleles that do not differ in charge (detected, for instance, by the heat stability discrimination introduced by Bernstein et al., 1973). Indeed, as pointed out in paper I of this series, charge differences may represent particularly gross increments in a general functional property and thus, on the average, behave differently with respect to selection, as compared to increments in other general functional properties. Finally, the proportion of occurrence in nature of different types of situations is to be established. It would seem like a long road.

The Ayala-Anderson type of experiments might be supplemented in an interesting way by studying selection and neutrality at the molecular level alone, in the Qß phage RNA replicating system of Mills, Spiegelman and their associates (cf. Kramer et al., 1974). One of the many questions to be put to this system and one which is of particular relevance to the present model would be the following: When the replicating system is maintained under strictly constant conditions over long periods of time, does it happen that one RNA sequence is replaced by another?

A FOUNDATION FOR A MOLECULAR EVOLUTIONARY CLOCK

On this constant-environment model for natural selection, we assume that at time zero, from which on incoming mutations will be screened according to the partial selection coefficients they involve, a selection vacuum with respect to as least one molecular function already exists. The time that will elapse until a further mutation becomes fixed will then depend on the time necessary for a mutation to occur that will fill the particular selection vacuum in question. There should be an average value for the number of new amino acid fixations, probably somewhere between 1 and 3, that are necessary for a further, incipient selection vacuum to pass a threshold beyond which the selective value of an incoming mutation will be sufficiently distinct from zero so as to make fixation probable. The proportion of the randomly occurring mutations that are fixed in relation to functional adjustments at constant environment should remain constant, within limits of course, as long as the functional density of the protein is not significantly altered. Provided the mutation rate also remains nearly constant, this concept should allow one to define a basic rate of evolutionarily effective substitutions characteristic of every type of informational macromolecule. If no other factors intervened, a molecular evolutionary clock should exist. On the basis of the present model it is no longer necessary to postulate changes in the internal (genetic) and/or external (ecological) environment of organisms to account for the

occurrence of conservative substitutions in informational macromolecules as well as for the regularity of this occurrence.

It may be perceived that the introduction of changes in the environment - namely the situation in most of the real world - should mostly not change the rhythm of the molecular clock. What environmental changes will achieve as far as the status of the molecule's general functions is concerned is to shift around the functions in relation to which a selection vacuum exists. For instance, the isoelectric point may now be closer to, and the solubility further removed from optimal. *On the average* there should be a balance between the new disadvantages and the new advantages of the protein with respect to the new functional requirements that it has to meet. The effect of limited environmental changes will then be as follows:

In regard to general functions and correlated specific-function increments, the nature and the sites of the selectively favorable mutations may be altered, but on the average not the rate of their occurrence and fixation.

In regard to substitutions at specific function sites, bringing about more significant changes in specific functions, the rate of substitution should not significantly intervene in the overall rate, insomuch as such substitutions represent only a rather small fraction in comparison with substitutions involved with general functions predominantly.

There is however one type of specific functions, as mentioned, whose variation may involve a significant proportion of sites, namely the function of interaction with other molecules. When a change occurs in this respect, it may affect a number of sites and a temporary acceleration of the clock may take place. That this actually happens is made likely by the analysis of Goodman et al. (1975). From their reconstructed phylogenetic tree of globins, these authors deduce the observation that the rate of amino acid substitutions is significantly accelerated[3] as functional change occurs. The changes

[3] The most extreme changes in rate reported by Goodman et al. may not be accepted at face value at this time. Among factors leading to spurious conclusions in this respect (Zuckerkandl, 1976b), duplication dependent homology (Zuckerkandl & Pauling, 1965), called paralogous homology by Fitch (1970), should be one of the most important. We should like to know, for instance, whether among the several lamprey globin components that have not yet been sequenced there is one, perhaps a minor component, that resembles Gnathostome hemoglobin and myoglobin chains more than the sequenced lamprey chains do. If so, the most rapid rate of globin evolution listed by Goodman et al. (1975) would have to be lowered.

listed are all traceable to change in molecular interactions. It apparently is at the sites involved in these interactions that the rate of evolutionarily effective substitutions is altered.

A temporary acceleration of the rate of amino acid substitutions due to functional change in proteins was predicted by Pauling & Zuckerkandl (1963) (also Zuckerkandl & Pauling, 1965). A closely related idea was independently expressed by Goodman himself (1963, 1964). Goodman's idea is that a slowdown in rates would occur after new specific functions, especially contact functions, evolved in proteins. This idea related the supposed slowdown to an assumed increase, over evolutionary time, in the functional density of proteins (to use the terminology of the present paper).

Is such an increase to be expected?

FUNCTIONAL DENSITY DURING EVOLUTION

The issue deserves to be examined here, even though no certain conclusion can be offered at this time.

If we look at the evolution of proteins in a wide historical perspective, we perceive that functional density (FD), as well as specificity of contact functions as accounted for by weighted functional density (WFD), must have been considerably lower in very early proteins than it is to-day (Goodman, 1964; Woese, 1971; Yčas, 1974; Zuckerkandl, 1975; and surely others). The question is how early in evolution this ceased to be so. It is plausible that the evolution of FD and WFD should be represented by S-shaped curves. But we do not know what evolutionary periods were spanned by these curves nor how steep they were. As functional density increased in any one protein, a progressively more severe competition between different molecular functions for the same sites must have taken place. Increasing difficulties in successful coadaptation of the diverse functions should have resulted in a progressive slowdown in the average rate of increase in functional density. Concomitantly, the correlated slowing of the molecular clocks should have been attenuated and the clocks have been stabilized, inasmuch as this particular mechanism intervenes. Goodman (1976), however, locates in *recent* evolution major increases in functional density and an ensuing major decrease in the period of the molecular clocks. Maximal functional density should at any rate have been or be approached asymptotically. Many hundreds of millions of years may be required to accomplish the last, say, 10% of attainable average FD and WFD. This terminal process should no longer be significant in terms of the molecular clocks.

It seems likely that the evolutionary increase in functional density has been determined by two major driving forces. It was apparently of *energetic* as well as *organizational* advantage to crowd more functions into any one protein than could be satisfied optimally all at once; of energetic advantage because of the implied reduction in total length of messenger RNA to be translated per organism; of organizational advantage because multimolecular relationships are the condition for an orderly and specific performance of functions and for regulating this performance.

There is no reason why an inducement to economy, if it played a role ever, should not have done so from the earliest times of protein evolution on. As soon as the structure and length of polypeptide chains could be reproduced, it must have been of advantage for the primitive cell that functions be carried out on the basis of minimal lengths of polypeptide chains. On these grounds, proteins should have tended very early towards maximal functional density.

How does this consideration of economy tie in with the second process mentioned, the development of multiple molecular interactions? The evolution of these interactions was definitely not confined to very early protein evolution. The logical way to reconcile the economic and the organizational imperatives would be to develop multiple interaction functions through polypeptide accretion, essentially without further increase in functional density, assumed to be already close to maximal. This could be done by covalent bonding (fusion between polypeptide chains) or noncovalent bonding (formation of quaternary structure) between subunits each of which had already nearly attained maximal functional density. This consideration is in accord with the suggestion repeatedly made that proteins were originally smaller and that they indeed grew by accretion (Cantor & Jukes, 1966; Woese, 1971; Dickerson, 1971). It also implies the tendency of polypeptide chains to form higher order structures composed of functionally distinct subunits.

The force of the argument according to which the essential increases in functional density must have been confined to very early evolution is however mitigated by another consideration. Even with functional densities well below maximal, polypeptide structures could not be reduced below a certain size, because a number of sites were required merely for the stabilization of the molecular architecture. Such stabilization involves specific contact functions, and thus a number of the sites referred to were specific-function sites. But others, at the molecular surface, probably qualified as general-function sites and may have remained available for the development of further specific functions.

The introduction of new specific functions into proteins is known to have occurred during the last 600 million years, for instance in hemoglobins, and was no doubt quite common. Yet it is no proof of increases in functional density. The rates of amino acid fixations in the single subunit of vertebrate myoglobin and in Gnathostome hemoglobins endowed with quaternary structure seem to be similar (of the order of 20 amino acid fixations per hundred million years; compare Romero-Herrera et al. (1973) with Goodman et al. (1975), though the variability of the figures precludes any strong statement about identity or difference). If the evolution of interchain contact sites had involved the mobilization of a large proportion of former general-function sites, to be used henceforth for a specific function, the rate of evolution of the myoglobins should be distinctly faster than that of the Gnathostome hemoglobins. Perhaps some difference in this direction does exist. But it also seems possible that, in the muscle cell, myoglobin is actually interacting with itself or with other molecules in a way that has not yet been understood, and that consequently the evolution of quaternary structure in hemoglobin did not entail, with respect to myoglobin, a notable increase in WFD. It may be that, during more recent evolution, the switch of single protein subunits to oligomeric structures involved in many or most cases a functional switch between sites rather than a committment of a supplementary fraction of general-function sites to specific functions. An approach is thus defined for future investigations.

Change in specific molecular functions, as it continued to take place during relatively recent evolution, may also have led sometimes to a certain *decrease* in functional density. This could occur, notably, when the formation of a dimer between two distinct protomers results in the loss of the main function of one of the protomers. An example is α-lactalbumin. It is one of the protomers in the dimeric lactose synthetase and seems to act solely in regulating the activity of its partner chain (Hill et al., 1969). α-Lactalbumin is homologous with mammalian lysozyme (cf. Dayhoff, 1972). It has preserved the cleft characteristic of the substrate binding region of lysozyme, but in an apparently nonfunctional state. Some of the residues in this region probably continue to carry out specific structure-functions inasmuch as they are needed for preserving the overall structure of the molecule. Other sites in this cleft may be examples of specific-function sites turned into general-function sites. The new specific contact function with the partner chain may or may not have made up for this loss in specific-function sites.

Changes during more recent evolution in specific functions thus represent either functional additions, or functional substitutions, or functional losses. The average net effect on functional density in proteins of these three processes cannot now be predicted. It may well be small.

Should at least an increase in WFD be expected during recent evolution? Even this is not clear. Some proteins have been nearly invariant for a very long period. Histone IV is at present the most outstanding example (cf. Dayhoff, 1972). The unicellular eukaryote that was the common ancestor of pea and man must have had a histone IV nearly identical with the contemporary molecules. If all proteins, with few exceptions, are extremely "old" (Zuckerkandl, 1975) and if one of them was able to attain its maximal weighted functional density about a billion years ago, moreover at a value not far from 1.0 (see paper I of this series), it should follow that all proteins had in principle time enough to reach maximal WFD at about that time or earlier.

Nevertheless, in many proteins the requirements for higher specificity in molecular interaction may have evolved gradually (Yčas, 1974). This does not imply, on the other hand, that these requirements have increased significantly in recent evolutionary times. They may have been very high more than a billion years ago, judging from comparisons between contemporary prokaryote and eukaryote enzymatic chains. Thus, the most significant increase in WFD is likely to have taken place, again, in earlier rather than later evolution. As to shifts in interaction specificity, such as occurred and continue to occur frequently, they by no means necessarily imply changes in WFD. Citri & Pollock (1966) show how mutations in the penicillinase gene modify the "substrate profile" of the enzyme. As long as a new substrate interacts as specifically with the enzyme as the former preferred substrate did (and the change is one adopted by natural selection), WFD is not altered.

In recent as well as ancient evolution, we do expect temporary fluctuations in WFD to accompany substitutions of specific functions the ones by the others. At each such substitution, WFD should pass through a trough. This should be accompanied by a temporary acceleration of the ticking of the molecular evolutionary clock.

On the basis of the present discussion a marked slowdown of the clock during recent evolutionary times is not expected. If such a slowdown nevertheless took place in recent vertebrate evolution - in the case of cytochrome c, during the last 300 million years - as the results of Goodman (1976) and his associates now clearly suggest, the question remains to what extent this is due to the mechanism invoked by Goodman and to

what extent to other factors. An increase in generation time may be considered (e.g. Ohta & Kimura, 1971; see however Sarich & Wilson, 1973).

Even though the end result is the same, the mechanism of slowdown in fixation rate is not indifferent with respect to the question whether or not the rates of evolutionarily effective amino acid substitutions, if not linearly related to time after proper statistical treatment of the data, are nevertheless a simple function of time. This is what the question of the clock is about.

RANDOMNESS AND NONRANDOMNESS

Some evidence is available to suggest that "the more restricted the selectively acceptable opportunities for changes are, the more nonrandom is the distribution of these mutations among the nucleotide positions of the codon". This was found by Fitch (1973) as he checked randomness and nonrandomness at the first and second coding positions of hemoglobin α- and β-chains and of cytochrome c. Hemoglobin β-chains evolve faster than α-chains, as predicted by Ingram (1961), and as shown first by Derancourt et al. (1967), and α-chains evolve much faster than cytochrome c. Fitch observes that the degree of randomness of base substitutions decreases along the same series. Zuckerkandl et al. (1971) had found by likewise examining frequencies during evolution of base substitutions at the different coding positions that, for all types of base substitutions taken together, the overall result was nonrandomness in the case of the globins as well as of the cytochromes c (see Appendix). However, different types of globin chains had been pooled. There is no contradiction with the results of Fitch, since by combining his data on the hemoglobin α- and β-chains, the overall result for this group of chains is also nonrandom[4]. About one half of the individual kinds of base substitutions (taking their direction into account) was found by us (Zuckerkandl et al., 1971) to be clearly "random".

[4] A discrepancy does exist, however, between Fitch's results and our own, in that according to our results the degree of nonrandomness of the base substitutions is less pronounced in cytochrome c than it is in globin, - although, as mentioned, it is significant in both cases. In evaluating the discrepancy it should be noted that, at the time the computations were done, instead of Fitch's 49 cytochrome c sequences we had 21 sequences at our disposal, and that among these 21 the sequence of the prokaryote cytochrome c_{551} of *Pseudomonas* was included. The latter circumstance may have resulted in some randomization.

Fitch's observations obviously fit in with the present model. In more slowly evolving proteins a larger proportion of sites is expected to be concerned with specific functions (higher functional density) and if not, the variability of certain specific-function sites is reduced (higher weighted functional density), thereby preventing them from intervening to some extent in the adjustment of general functional states. Thus the random character of substitutions, which is supposed to relate mainly to general functions, is reduced.

In the case of base substitutions above expectancy, we have shown (Derancourt et al., 1967) that they correspond to a number of the most conservative, in the sense of frequent, amino acid substitutions. This was noted then, in particular, for the substitution between G and A at the first coding positions in globin messenger, and can be generalized as demonstrated by Fitch (1972). It thus appears that a rate above expectancy in a particular base substitution can in general be traced to the exchange between two or very few amino acid residues, even though each type of base substitutions involves a somewhat larger total number.

The latter cases of abnormally frequent, nonrandom substitutions give us a clue as to what "randomness" really means in this connection. The expected frequency of a certain type of base substitution had been calculated on the basis of the amino acid composition of the protein, the corresponding messenger composition at the first two coding positions, assuming mutation frequencies in proportion to this base composition, and the structure of the genetic code (distribution of one-step mutations), as was also done by King & Jukes (1969). An observed frequency in accord with the expected was considered random, and higher or lower frequencies, nonrandom. The base substitutions retained with a frequency higher than expected cannot be construed to correspond to a higher *mutation* rate (Zuckerkandl et al., 1971). Rather, mutations between, say, A and G at the first position of globin codons, will be retained in a higher proportion of occurrences than any other base substitution. Other types of base substitutions (e.g. C by G and G by U at the first and second coding positions in globins and cytochromes c, Zuckerkandl et al., 1971) happen to be eliminated, in relation to the frequency of their occurrences, at rates higher than the preceding but approximately equal among themselves. This quantitatively important group of base substitutions represents some sort of mean selective stringency, measured by the number of times certain base changes have to appear by mutation, on the average, before a fixation occurs. It is this group of types of base substitutions that will contribute predominantly to determining the total number of evolutionarily effective mutations that occurred in a certain sector of a molecular phylogenetic tree,

the more so as fixations more frequent and less frequent than this intermediate value will tend to yield a mean in approximate accord with it. Therefore the calculated expectancy for each type of base substitution will be related to this figure. It is indeed a function of the total number of base substitutions that occurred within the phylogenetic tree. A near-conformity with this figure will be taken as indicating randomness of fixations, when in fact it only indicates near-conformity with a certain predominant evolutionary rate.

"Randomness" thus only means here that the percentages of certain mutationally occurring base substitutions that are eliminated from the population of organisms before a fixation takes place are close to a certain mean. When a larger percentage of a certain type of base substitution is eliminated, another zone of "nonrandomness" is entered, with a fixation of mutations "below expectancy". Going from the most frequently fixed base substitutions to the least frequently fixed, the passage from "nonrandomness" through "randomness" to "nonrandomness" again does not in reality entail any qualitative change in the process, but only a progressive decrease in the fractions of the mutations that become fixations. This decrease reflects a hierarchy in the resistance to different types of amino acid substitutions. Such a resistance exists even where it is weakest, at the most variable sites. Even at these sites there is indeed reason to believe that the fixations are significantly below mutation rate: an amino acid that is frequently accepted is not accepted every time it presumably appears by mutation (Zuckerkandl, 1975).

Under these conditions, the so-called randomness of substitutions, which of course remains valid as an objective observation, does not have certain implications the term suggests. It establishes in no way the participation in molecular evolution of nondetermined events. One may be entitled to consider as the only "true" random events those that are nondetermined within the boundaries of the system one is considering. It is precisely such "true" random events that the present model implies, to a certain degree. Ironically, the zone where this nondetermination comes into the picture mainly coincides with one of the nonrandom zones as defined by the expectancy/observation comparisons referred to above, namely with the zone of base substitutions occurring with a frequency above random expectation.

These base substitutions depend on relatively frequent amino acid substitutions, of which either the first out of a small set, as it appears by mutation, has a high chance of becoming fixed, or the n^{th} that appears by mutation, n being thought to be relatively constant and not a function of the amino acid in the set.

In many cases the chances of fixation of the substituents belonging to a set of amino acids frequently fixed at highly variable sites are not really equal. These chances may differ by a factor of two or more (see Table 1 in Zuckerkandl, 1975). In other cases, members of small sets (mostly very small, namely sets of no more than two) may have an equal chance of being fixed at a level of frequency that happens to correspond with the level of "randomness" according to the expectancy/observation calculations. In such cases the two kinds of randomness coincide.

Indeed, equality of chance of fixation of two or more amino acids can come into play at any level of frequency of fixation, including levels corresponding to base fixations at frequencies "below expectation". As one moves towards these levels, however, the sets of amino acids with equal chances of substitution become ever smaller, ever rarer, and the quantitative impact of the phenomenon on evolutionary change ever smaller.

On the whole, a limited measure of nondetermination presumably remains.

CONCLUSION

The molecular processes we have been predominantly concerned with here have only limited impact on evolution. The impact lies mainly in the creation of structural variety within set structural boundaries. At any one time such variety of accepted mutations is present only to the extent polymorphism is. Nevertheless it may occasionally offer raw materials for significant functional change, next to substitutions at specific-function sites. Variety at the level of controller genes is likely to have more frequently far reaching consequences. Conservative substitutions at general-function sites mostly resemble Brownian motion. The word "evolve", "evolution" was sometimes put in quotation marks in the present two papers, to underscore that most of the fixations that occur in genes and proteins are not evolution, but evolutionary noise. The noise, as stated, is of a peculiar kind, in that it is interpretable in terms of selection, if the present model has any pertinence. The question as to whether this particular noise is "Darwinian" or "non-Darwinian" continues to be hotly debated. Perhaps one may wonder to what extent it is a fundamental question for evolutionary theory and for biology at large.

Acknowledgements. I am greatly indebted to Drs. Morris Goodman, Jack L. King, Leigh Van Valen, and Helmut Vogel for stimulating and important comments that led to a revision of this paper. This work was begun at Centre de Recherches de Biochimie Macromoleculaire (CNRS) and Groupe de Recherche U 67 (INSERM), Montpellier. It was subsequently supported by a grant from the American Philosophical Society.

APPENDIX

Errata

Since the data of Zuckerkandl et al. (1971) have some bearing on the topic treated here, the opportunity is taken to apologize to readers and to the *Journal of Molecular Biology* for the following errors in this paper that have remained uncorrected:

Figure 1, p. 478, which gives an example of the derivation of amino acid substitutions: at node 8, serine should have been listed along with glycine as a possibility of equal likelihood.

Legend to Table 2, p. 477. Instead of "Same as Table 1, except that the data on globin and cytochrome c have been pooled", read: "Data on cytochrome c, treated like in Table 1". (The table was correctly referred to in the text, but the caption was erroneous).

Finally, the lower part of Table 3 (p. 480) should read as follows:

		Globin					
		low			high		
cytochrome c	low	AU_1	CG_1	GU_1	AG_1	CU_1	CU_2
		AC_2	CG_2	GU_2	GC_2	UA_2	UC_2
	high	AG_2	CA_1	CA_2	AC_1	AU_1	GA_1
		UA_1	UC_1	UG_1	GA_2	GC_1	UG_2

REFERENCES

Ayala, F.J., Anderson, W.W. (1973). Nature New Biology 241, 274

Ayala, F.J., Tracey, M.L. (1974). Proc.Natl.Acad.Sci. 71, 999

Ayala, F.J., Valentine, J.W., Hedgecock, D., Barr, L.G. (1975). Evolution 29, 203

Bernstein, S.S., Throckmorton, L.H., Hubby, J.L. (1973). Proc.Natl.Acad.Sci. 70, 3928

Boyer, S.H., Noyes, A.N., Timmons, C.F., Young, R.A. (1972). J.Hum.Evol. 1, 515

Cantor, C.R., Jukes, T.H. (1966). Proc.Natl.Acad.Sci. 56, 177

Citri, N., Pollock, M.F. (1966). Advan.Enzymol. 28, 237

Clarke, B. (1972). Am.Nat. 106, 1

Clegg, M.T., Allard, R.W., Kahler, A.L. (1972). Proc.Natl.Acad.Sci. 69, 2474

Dayhoff, M.O. (1972). Atlas of protein sequence and structure, Vol. 5. Washington, D.C.: Natl.Biomedical Research Foundation

Derancourt, J., Lebor, A.S., Zuckerkandl, E. (1967). Bull.Soc.Chim.Biol. 49, 577

Dickerson, R.E. (1971). J.Mol.Biol. 57, 1
Ewens, W.J. (1972). Am.Nat. 106, 273
Fisher, R.A. (1958). The genetical theory of natural selection, 2nd edition. New York: Dover Publications
Fitch, W.M. (1970). System.Zool. 19, 99
Fitch, W.M. (1972). Haematologie und Bluttransfusion 10, 199
Fitch, W.M. (1973). Ann.Rev.Genet. 7, 343
Fitch, W.M. (1975). An evaluation of molecular evolutionary clocks. In: Molecular study of biological evolution, F.J. Ayala, ed. Sunderland, Mass.: Sinauer Associates
Flynn, U.E.H., Sullivan, B. (1974). Biochem.Genet. 11, 373
Gooch, J.L., Schopf, T.J.M. (1972). Evolution 26, 545
Goodman, M. (1963). Man's place in the phylogeny of the primates as reflected in serum proteins. In: Classification and human evolution, S.L. Washburn, ed., p. 204. Chicago: Aldine publishing Co.
Goodman, M. (1964). The specificity of proteins and the process of primate evolution. In: Protides of the biological fluids, H. Peeters, ed., p. 70. Amsterdam: Elsevier
Goodman, M. (1976). Towards a genealogical description of the primates. In: Molecular anthropology, M. Goodman, R.E. Tashian, eds. New York: Plenum (in press)
Goodman, M., Moore, G.W., Matsuda, G. (1975). Nature 253, 603
Haldane, J.B.S. (1957). J.Genet. 55, 511
Harris, H. (1966). Proc.Roy.Soc.(London) ser.*B* 164, 298
Hill, R.L., Brew, K., Vanaman, Th.C., Trayer, J.P., Mattock, P. (1969). Brookhaven Symp.in Biol. 21, 139
Huang, S.L., Singh, M., Kojima, K.I. (1971). Genetics 68, 97
Ingram, V.M. (1961). Nature 189, 704
Kermack, K.A. (1954). Phil.Trans.Roy.Soc.London *B* 237, 375
Kimura, M. (1968a). Nature 217, 624
Kimura, M. (1968b). Genet.Res. 11, 247
Kimura, M., Ohta, T. (1971). J.Mol.Evol. 1, 18
Kimura, M., Weiss, G.H. (1964). Genetics 49, 561
King, J.L. (1967). Genetics 55, 483
King, J.L., Jukes, T.H. (1969). Science 164, 788
King, J.L., Ohta, T. (1975). Genetics 79, 681
King, R.C. (1968). A dictionary of genetics. New York: Oxford University Press
Kramer, F.R., Mills, D.R., Cole, P.E., Nishihara, T., Spiegelman, S. (1974). J.Mol.Biol. 89, 719
Lewontin, R.C. (1974). The genetic basis of evolutionary change. New York: Columbia University Press
Lewontin, R.C., Hubby, J.L. (1966). Genetics 54, 595
Martin, F. (1974). Etude de l'hémoglobine d'un Sélacien, *Scylliorhinus canicula*. Thèse de doctorat ès sciences physiques, Université des Sciences et Techniques du Languedoc, Montpellier
Maynard Smith, J. (1968). Nature 219, 1114
Mayr, E. (1963). Animal species and evolution. Cambridge, Mass.: Harvard Belknap Press

Mayr, E. (1970). Populations, species, and evolution. Cambridge, Mass.: Belknap Press of Harvard University Press
Milkman, R.D. (1967). Genetics 55, 493
O'Donald, P. (1969). Nature 221, 15
Ohta, T. (1972). J.Mol.Evol. 1, 305
Ohta, T., Kimura, M. (1971). J.Mol.Evol. 1, 18
Pasteur, G. (1974). Mém.Soc.Zool.France 37, 473
Pauling, L., Zuckerkandl, E. (1963). Acta Chem.Scand. 17, S9
Romero-Herrera, A.E., Lehmann, H., Joysey, K.A., Friday, A.E. (1973). Nature 246, 389
Sanders, H.L. (1968). Am.Nat. 102, 243
Sarich, V.W., Wilson, A.C. (1967). Proc.Natl.Acad.Sci. 58, 142
Sarich, V.W., Wilson, A.C. (1973). Science 179, 1144
Selander, R.K., Hunt, W.G., Yang, S.Y. (1969). Evolution 23, 379
Selander, R.K., Smith, M.H., Yang, S.Y., Johnson, W.E., Gentry, G.B. (1971). Biochemical polymorphism and systematics in the genus *Peromyscus*. In: Studies in genetics VI, M.R. Wheeler, ed., p. 49. Austin, Texas: University of Texas Publ.No.7103
Selander, R.K., Yang, S.Y., Hunt, W.G. (1969). Polymorphism in esterases and hemoglobin in wild populations of the house mouse (*Mus musculus*). In: Studies in genetics V, M.R. Wheeler, ed., p. 271. Austin, Texas: University of Texas Publ.No.6918
Stenzel, P. (1974). Nature 252, 62
Strickberger, M.W. (1968). Genetics. New York: Macmillan
Sved, J.A., Reed, T.E., Bodmer, W.F. (1967). Genetics 55, 469
Valentine, J.W., Ayala, F.J. (1975). Deep Sea Res. 22, 37
Van Valen, L. (1974). J.Mol.Evol. 3, 89
Wallace, B. (1958). Evolution 12, 532
Woese, C.R. (1971). J.Theoret.Biol. 33, 29
Yčas, M. (1974). J.Theoret.Biol. 44, 145
Zuckerkandl, E. (1974). Biochim. 56, 937
Zuckerkandl, E. (1975). J.Mol.Evol. 7, 1
Zuckerkandl, E. (1976a). J.Mol.Evol. 7, 167
Zuckerkandl, E. (1976b). Programs of gene action and progressive evolution. In: Molecular anthropology, M. Goodman, R.E. Tashian, eds. New York: Plenum (in press)
Zuckerkandl, E., Derancourt, J., Vogel, H. (1971). J.Mol.Biol. 59, 473
Zuckerkandl, E., Pauling, L. (1965)..Evolutionary divergence and convergence in proteins. In: Evolving genes and proteins, V. Bryson, H.J. Vogel, eds., p. 97. New York: Academic Press