J. Mol. Evol. 12, 57–89 (1978)

Journal of Molecular Evolution

Multilocus Enzymes, Gene Regulation, and Genetic Sufficiency

Emile Zuckerkandl*

Linus Pauling Institute of Science and Medicine, 2700 Sand Hill Road, Menlo Park, California 94025, USA

Summary. Moderately diverged small sets of nonallelic genes that control one type of protein, such as genes for isozymes, are usually considered to be present in order to permit adaptations of protein function to local conditions in organismal time and space. However some irregularities, as well as some regularities, in the distribution of lactate dehydrogenase isozymes seem difficult to reconcile with this view. A complementary interpretation is based on selection for gene regulation. Irrespective of the relative importance of transcriptional control and of control through processing of the transcripts, regulatory patterns in higher organisms may be explored on the basis of formal relationships between molecular components interacting within and between "controller nodes". It is shown that an analysis in terms of controller node stability states may account for the variations in regulation of isozyme synthesis. Adaptive changes in regulatory sequences may insure the stability of the synthesis of a given type of protein in the face of developmental, as well as evolutionary, alterations in patterns of gene activity, correlated with the evolution of cell differentiation. It is proposed that gene duplication, beyond the addition of novel functions to the established ones, plays a further and rather opposite role. This role is to maintain the *constancy* of established protein functions in the presence of developmental and evolutionary change. To this effect the duplication of functional units of gene action, rather than merely of coding sequences, is likely to be required.

In a certain number of cases, higher organisms can apparently dispense with a given moderately diverged set of genes and use a single gene instead. This fact is, to some extent, at variance with both the protein-functional and regulatory interpretations of the existence of the set. The relevance of the concept of genetic sufficiency is suggested as a solution, and possibilities offered by this concept for quantitative investigations in population genetics are outlined in an appendix. Increases in fitness occur automatically and perpetually regardless of

*Directeur de Recherche, Centre National de la Recherche Scientifique, Paris, France.
Reprint requests should be addressed to the Linus Pauling Institute

0022-2844/78/0012/0057/$06.60

the fact that genetic sufficiency often has already been reached. This process is characterized as the fitness treadmill. The reason why higher organisms evolve from lower forms probably is that sufficiency in adaptedness in no way prevents the fitness treadmill from continuing to turn. Tolerance towards gene loss does not imply that the dispensable genes have not continuously been selected for. It is proposed that increases and decreases during evolution in haploid DNA content per cell occur according to whether the fitness treadmill alone operates constantly, or whether frequent episodes of population bottlenecks bring genetic sufficiency thresholds to the fore. Positive selection does not *per se* confer biological significance upon a mutant. The frequency of biologically important selection events is expected to depend on the variability of the environment. The approximately equal evolutionary rates of amino acid replacements in a given kind of protein in both evolutionarily relatively stable and unstable environments suggest that most increases in fitness occur under conditions of continued genetic sufficiency of the wild type that is being replaced by a mutant. In relation to the potential for survival of the species and lineage, most selected mutations probably are no more, nor less, important than neutral mutations.

Key words: Isozymes – Lactate dehydrogenase – Molecular evolution – Controller nodes – Fuga duplication – Regulatory differentiation – Natural selection – Fitness treadmill – Gene dispensability – Trends in DNA content.

On the basis of certain (stringent) operational criteria used in DNA reassociation and DNA-RNA hybridization studies, most structural genes appear in the compartment of "unique" sequences (cf. Lewin, 1975). Yet the study of molecular evolution, thanks to the application of appropriate mathematical tools (cf. Fitch, 1973), has shown that in a more fundamental sense structural genes are hardly ever, and probably never, unique. There may be no gene unaccompanied by homologous genes in the same genome. Indeed, as the data accumulate and techniques for their analysis become more refined, homology groups of proteins, formerly believed to be separate, fuse. The total number of homology groups shrinks spectacularly (Barker and Dayhoff, 1972; Dayhoff, 1973; Barker and Hunt, 1976; Zuckerkandl, 1975; Ycas, 1976; Rossmann and Argos, 1976; Strydom, 1977).

Some structural genes may remain without any *closely* related nonallelic relative in the same genome. It would seem that such genes represent a minority, however, and perhaps a small one. Vertebrate cytochrome c, for years, was a classical example of a protein controlled by such a gene (Margoliash and Fitch, 1968; Sherman and Stewart, 1971). Now it no longer is, as a cytochrome c present in the testis has been shown to be structurally distinct from cytochrome c in other tissues (Hennig, 1975; Goldberg et al., 1977). Yet there is no doubt that some genes are much more prone than others to appear in multiply and moderately diverged stets. This is illustrated by a comparison between hemoglobin and cytochrome c gene multiplicity, in spite of the recent development concerning cytochrome c.

The number of genes coding for the same kind of protein (endowed with the same kind of tertiary structure and function and therefore mostly carrying the same name) is, usually, 2 to 5, judging from studies on isozymes (Bailey et al., 1970; Markert

et al., 1975; Bailey et al., 1976). For globin genes of humans and their closest relatives, the count reaches at present 10 (β, δ, $G\gamma$, $A\gamma$, ϵ; $^1\alpha$, $^2\alpha$, $^3\alpha$, ζ; myoglobin) (Boyer et al., 1973; Nigon and Godet, 1976). Such genes may or may not be linked. Usually the least diverged among them are, and the most diverged are not (Petit and Zuckerkandl, 1976). Highly multiplied structural genes, such as those for the various histones (c.f. Elgin and Weintraub, 1975) and perhaps the feather keratin genes (Kemp, 1975) represent special cases that are not to be included in this group.

Moderate multiplicity of slightly or moderately diverged genes of essentially similar function is classically linked to a modulation of functional requirements in different tissues of the same organism as well as during its development. In this paper a complementary interpretation will associate moderate gene multiplicity with a multiplicity in requirements for gene *regulation.* This extension of the dominant theory will call for some discussion of the meaning of the concept of "functional requirement" or "functional necessity" and of the role of natural selection. Beforehand, a brief consideration of some pertinent – although only probable-features of the system of gene regulation will be necessary.

Selection for Protein Function in Moderately Diverged Gene Sets: Difficulties of the Concept

Lactate Dehydrogenase as a Material for Analyzing Isozyme Distribution

Isozymes are typical representatives of proteins controlled by multiple moderately diverged structural genes geared to one kind of protein function. They are defined as "multiple molecular forms of an enzyme in a species" (Markert et al., 1975). In the present context we are interested in isozyme molecules or subunits that are controlled by nonallelic genes. The best analyzed example of isozymes seems to be the lactate dehydrogenase (LDH) isozymes. Particular attention has been paid to three vertebrate genes, the A, B, and C[1] genes. Two of them, namely B and C, are known to be closely linked in pigeons (Zinkham et al., 1969). The close similarity in tertiary structure between the dogfish A gene and the porcine B gene on the one hand (Eventoff et al., 1975), and between the dogfish A gene and the mouse C gene[2] on the other hand (Musick et al., 1976) indicated that these A, B, and C genes are homologous. Homology of several vertebrate LDH chains was confirmed by sequencing (see Eventoff et al., 1977). It had also been supported by immunochemical crossreactivity involving bony fish, bird, and mammalian LDH A_4, bony fish and bird LDH B_4, and bony fish and mammalian LDH C_4 (Holmes and Scopes, 1974). On the basis of experience with many other proteins, we expect the homology not to be restricted to vertebrate LDH. The active LDH molecules are tetramers usually formed from various mixtures of whatever protomers are present in the tissue. The distribution of the activity

[1] Also designated ldhM, ldhH, and ldhC respectively by Bailey et al. (1976).

[2] Duplication-independent homology ("orthologous" homology) between the mouse "C" gene and the teleost C genes has not been proven (hence the quotation marks around "C"). It has been considered by some (Markert et al., 1975) that the mammalian gene in question, which is present exclusively in the spermatocytes of sexually mature mammals, and similarly in birds, might be orthologous to the C gene of bony fish (see, however, Holmes, 1972).

of the genes over different vertebrates from fish to man, and, within each organism, over different tissues and different periods of development, is complex (Markert et al., 1975). A large number of different activity combinations exist.

Isozyme separations are most commonly accomplished by electrophoresis. The relative electrophoretic positions of two isozymes (namely of homogeneous oligomers) controlled by two distinct genes may in some taxa be the reverse of what they are in others. The attribution of an electrophoretic component to a particular gene is achieved with high probability on the basis of combined criteria, primarily antigenic properties of the enzyme, but also its physical and kinetic properties and its tissue specificity (e.g. Shaklee et al., 1973; Markert et al., 1975).

For a number of aspects of the distribution of different isozymes over tissues and organisms, frequently a well motivated rationale can be presented in terms of catalytic function (e.g. Kaplan et al., 1968; Whitt, 1970, 1975; Somero, 1975; Lim et al., 1975; Salthe, 1975; Senkbeil and White, 1978). Sometimes such a rationale seems strained. Doubts of two opposite kinds are in order: a functional basis for "puzzling" isozyme distributions may yet be discovered; or such a functional basis, when asserted, may simply express the preconceived idea that natural selection is the cause of the observed distribution and that it must work as stated. In the present analysis we will not contend that this preconceived idea is generally wrong, but mean to put it in perspective.

Peculiarities of Isozyme Distribution: Problematic Irregularities

We shall focus here on the expression of one LDH gene, the C gene (ldh C of Bailey et al., 1976). The expression of this gene is quite variable in different tissues and different taxa of bony fish (Shaklee et al., 1973; Markert et al., 1975). On the other hand, the B gene, closely linked to the C gene at least in some birds, is much more generally and stably expressed, from fish to bird (Bailey and Wilson. 1968). Thus, whereas the tissue specificity of the expression of one gene is in general relatively constant over different animal groups, that of a homologous gene can be quite variable and even apparently erratic. The interpretation of such distributions in terms of protein function should not be automatic.

In spite of undeniable differences between isozymes with respect to various parameters of enzyme action, different isozymes often behave in evolution as though they were interchangeable or, when they are mutually exclusive, as though they excluded each other for reasons other than their enzymatic parameters. Even more than certain apparent irregularities in isozyme distribution over different tissues in related species, certain regularities in this distribution appear difficult to reconcile with the hypothesis according to which particular functional specializations of isozymes are "required" in certain tissues and organs.

Let us list first some irregularities in isozyme distribution.

Since liver LDH controlled by the C gene is mostly cathodal and the eye LDH apparently controlled by the same gene mostly anodal (Shaklee et al., 1973), the current interpretation is that the anodal enzyme is functionally required in the eye (see discussion of this function by Whitt, 1975) and the cathodal enzyme for some as yet unknown reason in the liver (Sensabaugh and Kaplan, 1972). Take however cypriniform teleosts. Most Cypriniformes display the cathodal C isozyme in the liver. In some species, e.g. *Epicytrus microlepis,* no cathodal LDH is expressed in

the liver. Instead, a relatively cathodal LDH is expressed in the eye. Apparently the *regulation* of the C locus has shifted to the eye phenotype, whereas the structure of the enzyme, at least as much as it is expressed in electrophoresis, has remained close to the liver form. Is one to assume that in a few species the presence of a cathodal enzyme in the eye has become functionally "necessary", whereas an anodal enzyme is "necessary" in the eye of related species? In other groups of fish the C isozyme in the eye likewise is more cathodal than anodal (e.g. in a Clupeiforme, *Alosa mediocris*, and in a Salmoniforme, *Osmerus mordax,* Shaklee et al., 1973). Furthermore, in some fish, especially in a Beryciforme, *Holocentrus rufus* an LDH isozyme controlled by the C gene does not occur in just the liver or the eye, but is found, in smaller quantities, in many tissues (Shaklee et al., 1973). Is one to suppose that this isozyme is functionally "necessary" in all the ten tissues examined, whereas it is not in most of these tissues in closely related fish? The paddlefish *Polyodon spathula* has much more C isozyme in the kidney than in any other tissue examined, including liver and eye (Markert et al., 1975). Although the kidney of most fishes contains no or little C isozyme, does kidney function in this particular fish necessitate the activity of the C gene, whereas eye function and liver function do not? A further question may be raised. The C isozyme seems to be adapted to highly aerobic environments (Whitt, 1975). One of the prominent localizations of the C isozyme, the retina, indeed has one of the greatest requirements for oxygen of any vertebrate tissue (see Whitt, 1970). On the other hand, the lens has perhaps the lowest oxygen consumption. However, the LDH C gene is functional in the lens of some fishes, notably of the weakfish *Cynoscion regalis* (Whitt, 1970). Clearly, functional adaptation would be a shaky explanation for the presence of the C isozyme in lens. In addition, if functional adaptation of the C isozyme to conditions in the retina is "important"; if high oxygen requirement in the retina is general; then how is one to explain the absence of this isozyme from the eye of a certain number of fish, such as herring (Klose et al., 1965), goldfish (Nakano and Whiteley, 1965; Whitt, 1970), carp (Markert et al., 1975), etc.?

Cases where it is difficult, if not currently impossible, to interpret the tissue distribution of an LDH isozyme in terms of enzyme function are not limited to the C gene. Thus, for instance, at fertilization, different species of fish contain only LDH B subunits or only LDH A subunits, or both subunits (Philip and Whitt, 1977). Do differences in physiological conditions at fertilization justify this peculiar distribution? Likewise, in the beef lens, the LDH B isozyme greatly predominates, whereas in the lenses of rat and rabbit the A isozyme predominates (see Whitt, 1970). This variation in expression occurs even though lens metabolism is likely to be rather similar in different mammals.

Peculiarities of Isozyme Distribution; Problematic Regularities

There seems to be a trend towards mutual exclusion among different tissues of the activity of the LDH C gene, such that strong activity in one tissue tends to exclude equally intense activity in another. Consider the manifestation of the C gene in the two tissues, eye and liver, in bony fish. Table 1, which is derived from Table 1 of Markert et al., (1975), correlates roughly evaluated quantities of C protomer in liver and eye among 16 orders of Osteichthyes (see also Shaklee et al., 1973). Fourteen

Table 1. Correlation of lactic dehydrogenase of type C in liver and eye of bony fish. The data, from Markert et al., 1975, relate to 16 orders, represented by between 1 and 10 (total 47) different species. The figures in the matrix indicate the numbers of orders displaying a given combination of quantities of isozyme found in the two tissues. For the set of species of each order, mean values were calculated by taking ± as 0.5, + as 1.0, ++ as 2.0, etc. The mean values for the orders were then rounded off to the next integer, and retranslated into the scale of "pluses". The null hypothesis is that the different amounts of enzyme in the two organs are not correlated. A χ^2 test, for 9 degrees of freedom, gives for this null hypothesis a probability of 0.02.

liver / eye	-	+	++	+++
-	0	1	1	1
+	0	2	0	0
++	0	1	0	1
+++	8	1	0	0

of the 16 orders are teleosts. When there is much enzyme in the eye, there nearly always is none in the liver. Small quantities of C enzyme in both tissues occur in two orders. The following are not found in this sample, however: (a) a medium quantity in both, (b) a large quantity in both, or (c) none in both. In the one case where a large quantity of C enzyme is found in the liver, there is none in the eye. In general, a trend in the distribution of C gene expression is characteristic of a teleost order and differences occur between orders, but marked variations do occur *within* some (relatively primitive) teleost orders. Thus, among the Osteoglossiformes, of the three species analyzed, two illustrate the mutual exclusion between C gene activity in liver and eye, in both directions; the third is an example of an instance where only a small quantity of C enzyme is found in both organs, in which case it seems that no trend towards mutual exclusion is manifest.

Beyond eye and liver, the tissue in which the enzyme controlled by the C gene predominates is different in different bony fishes. For instance, the spotted moray (Anguilliformes) has the highest amounts in the spleen, whereas the American eel (same order) displays the largest quantities of this enzyme in the kidney. Again, when this quantity is characterized by +++ in one tissue, this tissue tends to be the only one so characterized, out of ten. The evaluation was very rough and no strong conclusion can be drawn. Nevertheless, one might expect that, in some forms, more than one tissue of the ten would be recorded as displaying large (+++) amounts. This does not occur one single time in the 16 orders and 47 species listed.

In terms of enzyme function, in summary, a fish that relies heavily on C gene product in the eye "needs" none of it in the liver and vice versa; furthermore, a fish that "needs" a relatively large quantity of C gene product in one organ usually does not "need" an equally large quantity in some other organ. With respect to metabolic requirements, such mutual exclusions are rather implausible.

A Complementary Interpretation: Selection at the Level of Genic Regulation

Observations of this kind lead to the suggestion that, in many cases, the adoption of different gene duplicates by different tissues will be prompted, not by the functional differentiation of the isozyme molecules, but by a differentiation in the condi-

tions of the control of their rates of synthesis. Some general features of the control of gene expression in higher organisms must first be discussed in this connection.

Levels of Control of Protein Synthesis and Breakdown and their Relative Significance: some Comments applied to the Control of LDH

In investigations ultimately aimed at determining variations in isozyme gene expression, what has mostly been measured so far is relative enzyme activity or relative amounts of enzyme protein. Available data do not allow one to distinguish between the contributions to these traits of transcription and posttranscriptional events respectively.

Among the posttranscriptional processes that may interfere with the analysis of regulatory events closer to the source of gene activity is differential breakdown of isozymes coexisting in a given tissue. Its occurrence for LDH has been asserted (Fritz et al., 1969; Fritz et al., 1975), but later disproven (Lebherz, 1974; Nadal-Ginard 1978). In the case of aldolase isozymes no significant differential turnover rate has been found (Lebherz, 1975).

It has however been shown (Nadal-Ginard, 1978), for four tissues of the mouse, that the LDH A and B isozymes strongly depended on tissular conditions of turnover. These conditions vary from tissue to tissue, but, in every tissue, insure to all LDH isozymes the same half-life. For determining absolute amounts per cell of LDH in the four tissues, this posttranslational control seems to play a more prominent role than transcriptional control or any other control up to and including translation.

Even though LDH itself has apparently not been investigated in this respect, many lines of evidence, taken together, suggest that translational control itself is not, in general, a predominant factor in gene expression (Ross et al., 1974; Suzuki et al., 1974; Lanyon et al., 1975; McKnight et al., 1975; Harris et al., 1975; Gelinas and Kafatos, 1977; Paterson and Bishop, 1977; Kurtz and Feigelson, 1977).

Thus, if isozymes were in general regulated posttranslationally, our attention might focus on the pretranslation control of isozyme-level-controlling proteins (e.g. proteases). The genes to be considered for this pretranslation control would be different from the isozyme genes themselves, but the problem of the mechanism of differential regulation and of its biological driving force would not be fundamentally altered.

Moreover, a posttranslational control, like most translational control, can only be modulating quantitatively the steady state of the cellular concentration of a protein. The decision as to whether gene product can be synthesized or not belongs to the levels of transcription and processing, which we may continue to think of as fundamental.

Doubts are being expressed currently about the predominance of the role of transcriptional regulation, considering the impact, suggested by recent findings, of regulation through processing of the RNA transcripts. It has been shown, in sea urchin, that the profiles of nuclear transcripts are, to a certain approximation, constant at several if not all developmental stages and in several if not all tissues (Kleene and Humphreys, 1977). This extends the earlier findings of Mauck and Green (1973) and Johnson et al. (1976) to the effect that the rate of synthesis of only mRNA, not hnRNA, differs in resting and growing mouse fibroblasts.

In spite of interference by regulatory events further down the line, it is very probable that there is, in higher organisms, a significant general correlation between either transcription or processing and phenotype; between either transcription or processing and differentiation. Whether the correlations really are with the processing of transcripts more than with transcription itself, or vice versa, is at present an open question. A correlation with transcription is suggested by at least two lines of evidence taken together. One is apparent correlation between transcriptional activity and puffing in polytene chromosomes confirmed by Zhimulev and Belyaeva (1975). The other is the correlation between changes in puffing patterns during development and the appearance of morphological characteristics (Ribbert, 1972). Bonner and Pardue (1977 b), however, question the correlation between transcriptional activity and puffing[3].

It can be upheld that, irrespective of the importance of processing, transcriptional control is expected to conserve a key role in the control of rates of protein synthesis. This may be considered to be so because of an inbuilt circular relationship, to be postulated a priori, between control at the two levels, transcription and processing. If processing enzymes and other "processor" molecules[4] intervened in the regulation of the

[3] Bonner and Pardue (1977b) show that upon *in situ* hybridization of RNA, the labelling in puffs, in comparison with nonpuffed regions, is in general much weaker than expected on the basis of uridine incorporation autoradiograms. The authors consider that RNA storage in puffs might be the predominant factor in causing puffs to incorporate much more labelled uridine than nunpuffed regions. Further work is needed to decide the issue, since (1) Bonner and Pardue used a 10 min incubation period with labelled uridine, whereas Zhimulev and Belyaeva (1975) used 5 min. With labelling limited to 5 min, the proportion of nascent RNA in relation to labelled stored RNA should be greater, and the incorporation data of Zhimulev and Belyaeva thus be more representative of transcription than those of Bonner and Pardue. (2) The average amount of RNA hybridizing to puffs, as compared to the average amount hybridizing to nonpuffed transcribed regions ("micropuffs") has not yet been quantitated. The existence of a statistically significant difference in favor of RNA hybridizing to puffs has not been excluded. (3) As Dr. Pardue (personal communication) also emphasizes, the length of transcription units, the degree of gene reiteration, and the degree of polyteny all are expected to influence the hybridization autoradiograms.

Taking together the work of Bonner and Pardue (1977), the statement of Zhimulev and Belyaeva (1975) to the effect that nearly all polytene chromosome bands incorporate uridine at least weakly, and findings such as those of Johnson et al. (1976) and Kleene and Humphreys (1977) quoted above, one must indeed, at this point, consider the possibility that transcription occurs throughout the genome and that puffing is primarily the expression of protective storage of transcripts from certain functional units of gene action. In the absence of puffs, transcripts would be processed rapidly, but perhaps for a large part not processed to mRNA. Storage in puffs, associated with the binding to puff RNA of processor molecules of a kind that insures processing to mRNA, would lead, at the time the puffs regress (see table 1 in Bonner and Pardue, 1977b), to a release into the cytoplasm of large quantities of mRNA. Such a hypothesis would attribute to processing the maximal possible role in the control of gene expression. Against such an extreme eventuality is the demonstration by Bonner and Pardue (1977a) that ecdyson-induced puffs cease to transcribe as they regress. This suggests that puffing is associated with an activation of transcription and not just with a process of storage of transcripts that are made before as well as during puffing at approximately similar rates. Furthermore, Lewis et al. (1975) have shown that actinomycin D, which is effective if added to salivary glands at time 0 of puff-inducing heat shock treatment, when administered 5 min after the start of heat shock, no longer significantly prevents puff-specific protein synthesis from being induced. Maximal puff size is reached 20 to 30 min after the glands have been transferred to the higher temperature. Clearly, it is not RNA storage in puffs, nor processing of stored RNA that can be inhibited during only the first 5 min of heat shock treatment.

rate of transcription of regulator molecules[4], conversely regulator molecules necessarily control the synthesis of the different processor molecules. Even if future work should confirm and extend the role of processing in the control of gene expression, it would seem that the dependence of phenotype on transcriptional control in the case of any protein would thereby only be made more indirect, not abolished.

No statement can be made at this time, nor needs to be made here, about priorities in the coordinated partial processes of transcription and processing, a coordination that is in fact likely to be of the chicken - egg type.

The System of Controller Nodes

In spite of the differences in gene regulation between eukaryotes and prokaryotes, some of the basic processes of transcriptional regulation in higher organisms are probably shared with the prokaryote system, at least in a formal sense (Kourilsky and Gros, 1974; Zuckerkandl, 1976 a,b). Different formal components of the "regulatory kit" used by bacterial genomes quite likely occur in eukaryotes also: receptor sequences associated with the structural gene or genes [in eukaryotes the singular may most often apply (Newlon et al., 1975)], regulator molecules (protein or RNA) that combine with the receptor sequences for positive or negative control, and effector molecules, usually smaller organic molecules, such as hormones and metabolites, that combine with the regulator molecules and alter their affinities for the receptor sequences.

Once the general occurrence of these formal elements in eukaryotes is accepted, their mode of interaction, as determined by their general properties, leads to the derivation of a series of genetic, developmental and evolutionary phenomena that are in fact observed in higher organisms. This is not to be elaborated on in the present context.

Here we examine one presumed consequence of the structure of the system of gene regulation, – a structure derived from the interplay of the above components – namely the organism's use of multiple editions of proteins, notably isozymes.

For analyzing gene interactions, a useful concept is that of the controller node. The controller node pertaining to a structural gene consists of a particular set of components of the regulatory kit, namely the components directly involved in the regulation of the transcription of this gene. For instance, among many possible combinations, a controller node may include two receptor sequences which, along with the structural gene and other sequences belong to the same functional unit of gene action ("fuga", Zuckerkandl 1976 b; see below p. 71); two regulator molecules, one for positive and one for negative control; and two effector molecules that combine with the regulator molecules. The formal elements of all controller nodes thus are receptors, regulators, and structural modifiers of regulators.

Like DNA, RNA probably possesses receptor sequences that combine with macromolecules effective in the control of processing of gene expression. These macromolecules,

---

[4] One may call processor molecule or processor any macromolecule that, by combining with RNA or RNP (ribonucleoprotein) specifically intervenes in splitting or splicing of RNA or in the prevention of these events. Processors are so-to-speak the regulators of RNA. Regulator molecules, or regulators, are any macromolecules that, in combining with receptor sequences of DNA or DNP (deoxyribonucleoprotein) closely linked with coding sequences help determine the rate of transcription, or the nontranscription, of these coding sequences.

in turn, may be activated or inactivated by noncovalently bound effector molecules, or by covalently introduced, even though reversible, structural modifications. Such processing controller nodes are expected to be geared to the processing of transcripts whose synthesis is determined by molecular relationships within the transcriptional controller nodes. The two types of controller nodes are circularly interconnected, since processor molecules probably process transcripts coding for regulator molecules, and regulator molecules control the transcription of RNA coding for processor molecules.

Basic properties of controller node systems are discussed elsewhere (Zuckerkandl, 1976 a, 1978, and in preparation). Here we need to mention primarily the following:

- Controller nodes interact. A change in state in one will bring about a change in state in others, the effect becoming weaker as the interaction becomes more indirect, namely via interposed controller nodes, the set of which represents an interaction chain.

- The system of interacting controller nodes provides not only an off-on control of gene expression, but also a means of modulating it.

- Heritable changes in gene expression at the polynucleotide level are mostly incremental. They are brought about by changes in affinity constants pertaining to the interaction between components of the regulatory kit.

According to the model, each controller node in each tissue and at each developmental stage, is characterized by a stability state. The stability in question is that of certain processes (or of their absence), namely of transcription (or nontranscription) of a fuga, or of processing (or nonprocessing) of a fuga transcript. In either case, the stability state results from two interacting factors: the *concentration* of regulator (or processor) and effector molecules in the immediate neighborhood of the DNA (or RNA) target sequences and the *affinity* for each other of these components. The affinity referred to is that of effector for regulator (or processor) and of regulator for DNA or DNP (or of processor for RNA or RNP). The concentration of such a component may change or a mutation may affect its affinity for another component. In such a case the stability of the controller node with respect to transcription or nontranscription (processing or nonprocessing) will be closer to or further away from a reference stability state, which is the point at which "flip-over" between transcriptional (or processing) activity and inactivity occurs.

To picture the difference between a stable, buffered and an unstable controller node, consider two receptor sequences adjacent to a coding sequence. Suppose the two receptors combine, one with an activator of transcription, the other with a repressor of transcription. The species referred to of these two regulator molecules are the active ones, namely those that, upon combination with, or release of, an effector molecule have the highest affinity for the receptor. Two conditions will lead to highly stabilized and buffered transcription: a very high affinity of one receptor for the activator or a very low affinity of the other receptor for the repressor. In both cases the concentration of the regulator molecule can vary within wide limits without transcription being discontinued. Even low concentrations of activator and even high concentrations of repressor will not interfere with transcription. Only absence of activator or very high concentrations of repressor will. In this situation, assume that the concentrations of these regulators, vary for differenent tissues and different stages of development, but not in the extreme. Then the controller node pertaining to the fuga considered will be stabilized for transcription in many or all tissues and at many or all stages. Conversely, if the affinity of the competent receptor sequences for activator is low or for repressor is high, the fuga will

be transcribed only in those tissues in which the concentration of activator is particularly high, or that of repressor particularly low. Transcription will be tissue-selective and stage-selective. In such a system, mutations can change the tissue- and stage-selective (unstable) state into a nonselective (more stable) state and vice versa.

Although, as stated, a set of molecular relationships of the controller node type can, in principle, apply to both transcriptional control and control of processing, modulations in the rates of mRNA synthesis for different isozymes, or of proteins determining the turnover of isozymes in different tissues, are tentatively discussed here in terms of transcriptional control. If future research should warrant it, the terms of this discussion could be translated into those of formally analogous processing controller nodes.

Relevance of the Controller Node Stability Model

In a simple case, when two kinds of isozymic protomers are synthesized in one tissue under the control of nonallelic genes, but only one protomer in another tissue, such a difference can be due to a difference in affinity of the receptor sequences of the two fugas for one and the same regulator molecule (Fig. 1). Since different tissues are characterized by different distributions of gene expression, the concentration of a regulator molecule can be lower in one tissue than in the other. In the tissue in which this concentration is lower, it would be sufficient for the activation of only one of the isozyme structural genes. This is one among a number of possible interpretations, in terms of the controller node model, of a situation such as the absence of the B chain of LDH from the erythrocytes of most species in one suborder of rodents (Shows et al., 1969; Engel et al., 1972).

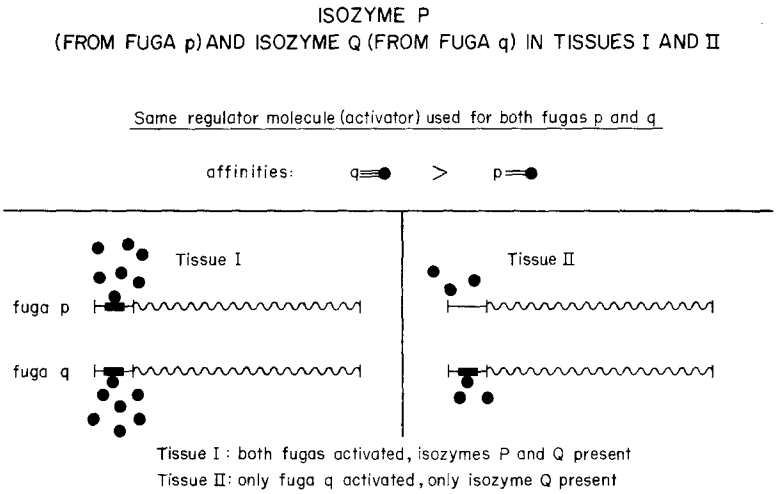

Fig. 1. An activator molecule reacts with two slightly different receptor sequences, each adjacent to a structural gene, of two different functional units of gene action (fugas), fragments of which are schematized. Ligation is indicated by a thickening of the line representing the receptor. In tissue I, the activator combines with both fugas. In the same organism, in tissue II, the concentration of activator is sufficient for combination with the receptor of one fuga but not of the homologous other fuga. Similar relationships can be envisaged for the control of processing rather than of transcription, the fuga being replaced by a fuga transcript, and the regulator molecule by a processor molecule.

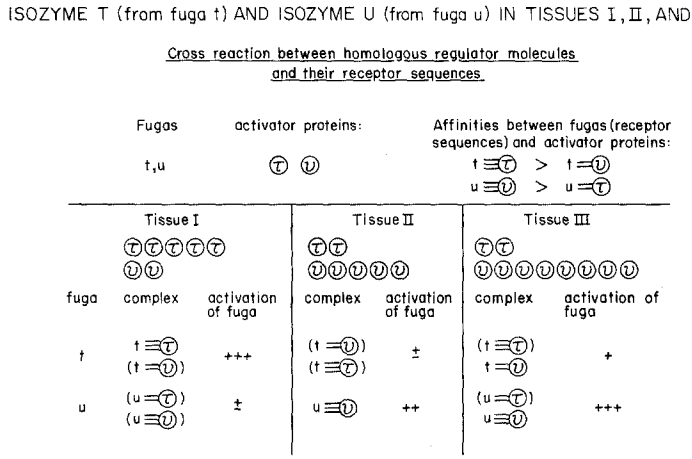

Fig. 2. Two kinds of homologous regulator molecules (activators) (T ,U) crossreact with receptor sequences on two homologous fugas (t and u) in three different tissues. The activator is assumed to be protein, but could also be RNA. In parenthesis, activator-receptor complexes of low effectiveness, namely *either of low affinity or low concentration*. Different intensities of resulting transcriptional activity are indicated. Similar relationships can be envisaged for the control of processing rather than of transcription (see caption for Fig. 1).

It seems also necessary to think in terms of cross reactions between regulator molecules and receptor sequences, because such cross reactions are a priori likely when we are dealing with duplicate fugas (see below section on gene duplication) of rather recent origin that have diverged only slightly. Figure 2 illustrates a case in which the activities of two fugas are distinct in three different tissues, a frequently encountered phenotype, illustrated by LDH. The figure shows how such differences might arise as a function of different relative and absolute concentrations of two variants of regulator molecules (protein or RNA), each of which has its highest affinity for one of the fugas but cross-reacts with the other. In parenthesis are the species of regulator-receptor complexes that have little or no effect in promoting transcription, either because a low affinity for a receptor is not being overcome by a relatively high concentration of a regulator molecule, or because a high affinity for receptor is not compensating for too small an amount of regulator. Such a case illustrates the interplay between quantities and mutual affinities of components of the regulatory kit, an interplay which may well represent one of the important relationships in biological processes.

Given the constitution of a fuga, only a small percentage of genetic variability is expected to be found in the structural genes. Most of the variability will be in the remaining parts of the fuga (Zuckerkandl, 1976a, and independently pointed out by Jack King, personal communication), firstly, because the remaining parts usually represent most of the fuga in terms of length, as judged from the ratio (close to four on the average in mice) between the percentages of DNA transcribed as hnRNA and as mRNA (see Lewin, 1975); secondly, because the transcribed "unique" sequences outside the structural genes have a rate of accepted mutations that is higher by roughly a factor of six than that of an

average structural gene[5]; thirdly because a fuga may, in addition to the transcribed sequences, contain nontranscribed sequences that presumably accept mutations and also play a role in gene regulation, perhaps through mechanisms involving high order DNP structures (Zuckerkandl, 1976b). Slight regulatory differences between individuals, with respect to all fugas, are therefore expected to be the rule rather than the exception (also emphasized by Jack King, personal communication). [Large differences in gene regulation between individuals probably are eliminated by selection in most cases (Holmquist, 1976)]. The predictions of individual and familial variability in gene regulation is even stronger if one considers the indirect effects of mutational events via the interactions between controller nodes that are postulated in the controller node stability model.

The controller node stability model also predicts tissue variability in gene expression. The difference in the spectrum of mRNAs in different tissues and at different stages (e.g. Galau et al., 1976) is the basis of the phenotypic aspects of differentiation. This tissular variation in transcriptional and processing background is liable to have distinctive effects on the transcriptional behaviour of any particular fuga or on the processing of its transcripts. In many cases the stability of a controller node will be significantly altered in one or the other tissue, and a difference in amounts per cell of the corresponding polypeptide chain will be detectable. This expectation is fulfilled in the case of LDH, since alterations in LDH subunit levels are not only specific for individuals (and presumably families), but in some cases are also tissue specific within one individual (Wilson et al., 1973).

If controller nodes differ indeed by the degree to which transcription or processing are buffered against changes in the interaction between node components (Zuckerkandl, in preparation), there is a prediction to be made on this basis. It applies to those multilocus homologous similifunctional[6] protein systems, duplicate member genes of which can be expressed in many or all tissues. (Genes, even though multiple, homologous, and similifunctional, whose expression is limited to one tissue, such as the hemoglobin non-α genes, are thus excluded from consideration). Among individual fugas belonging to such a group, those that are expressed in only one or very few adult tissues will be considered to display a limited stability range. Consequently they will be expected to be activated only locally and late in development, namely when the transcription and processing profile in the cell line leading to the competent differentiated tissue sufficiently approaches the differentiated state. Conversely, a controller node that manages to be active in many different adult tissues is considered as stable in the presence of varied RNA profiles. Therefore such a node is expected to be active also over a wider range of ontogenetic stages. In conformity with these predictions, in the teleost *Oryzias latipes*, the LDH B isozyme protomers, present in most adult tissues, are also present throughout embryogenesis. In contrast, the two other isozyme protomers, A and C,

---

[5] Britten and Davidson (1976) calculate that the protein sites characterized by the fastest evolutionary rates of amino acid replacements, as for instance the fibrinopeptide sites, are substituted at a rate that corresponds roughly to the rate of base substitution in the total "single copy" DNA compartment. The average rate of substitutions for proteins in general appears to be approximately six times slower than that for the fastest evolving protein sites (see e.g., Zuckerkandl, 1975).

[6] By similifunctional proteins are meant structurally distinct proteins that share the same basic functions.

which are more tissue specific in the adult, also appear later during embryogenesis (Philipp and Whitt, 1977). In several other species of teleosts, Philipp and Whitt confirmed that those isozymes present continually during embryogenesis also tended to be active in a wide variety of differentiated tissues in the adult fish and, conversely, that LDH isozymes which are active in a restricted number of adult tissues are detected only later in development. In other isozyme systems as well, the subunits that are more widely distributed in differentiated tissues are those present at all stages of development (Champion and Whitt, 1976). The meaning of the observations quoted is not unambiguous, however[7].

The controller node stability model leads to a further prediction that seems to be borne out. If the concept of regulatory interactions within and between controller nodes is correct, changes in temperature will affect these interactions and, in the case of unstable controller nodes, except if special compensatory mechanisms have been evolved, the resulting alteration will be expressed at the level of mRNA production. Since the stability state of controller nodes is assumed to be frequently different in different tissues, it is also predicted that a given change in temperature will have different effects in different tissues on the same controller node. Qualitatively, both predictions are in accord with observation (Shaklee et al., 1977), even though the findings have been interpreted in terms of functional adaptations, rather than, as we do here, in terms of physico-chemical effects of changes in temperature on equilibria between, or on the availability of, components of the regulatory kit. Obviously, such physico-chemical effects can be exploited by nature for functional adaptation, so that the two interpretations are not necessarily mutually exclusive. Aldolase furnishes an example

[7] A certain number of genes, such as the hemoglobin genes, are confined in their expression to a single tissue. This may often be due to a global process of regulation, affecting whole sets of fugas; one that might well be associated with the formation and melting of very high order structures in interphase chromatin. It has been called eurygenic regulation (Zuckerkandl, 1976a,b), as opposed to stenogenic regulation. In stenogenic regulation, individual fugas receive signals for transcription that can reach their target only when eurygenic regulation permits. Both types of regulation can be formally represented in a controller node system, although in eurygenic control equivalent receptor sequences might be highly multiplied (leading to "mass ligation" of regulator molecules), whereas in stenogenic regulation equivalent receptor sequences might be unique or few ("punctate ligation" of regulator molecules). Assuming this distinction to be meaningful, it can of course not now be decided whether the distribution of isozyme gene activities during ontogeny is a function primarily of eurygenic or stenogenic control. Possibly multilocus-controlled similifunctional proteins that play metabolic roles at all stages and in all tissues ("household" functions), such as LDH isozymes, are mostly not eurygenically controlled. Or else, most of the gene duplicates, active in different tissues, undergo identical eurygenic determination. If either is the case, then the distribution, during development, of the activities of the different duplicate genes (or of most of them), in relation to the distribution of the activities in the adult, indeed represents a test for the controller node stability model. But the alternative cannot at this point be excluded: the appearance late in development of proteins that are confined to few tissues in the adult may only reflect the progressive narrowing down of gene activities through successive binary (Kaufman, 1975) choices, characterizing eurygenic determination. Any gene activity that appears late would thus, per force, be confined to few tissues, without any other implications with respect to controller node stability. Against the generality of this latter interpretation one may cite some observations on isozymes, such as the case of esterase 5 in the green sunfish (*Lepomis cyanellus*) (Champion and Whitt, 1976). As in other cases, the confinement of the isozyme component to a few adult tissues is coordinated here with its confinement to a few ontogenetic stages. But the period of gene expression, during ontogeny, falls in early rather than in late phases.

of the temperature effects. Aldolase activity shows significant changes in three out of five tissues examined, and the changes occur in both directions (Shaklee et al., 1977). Changes in isozyme levels as a result of keeping organisms for protracted periods of time at different temperatures have also been observed, but the reports on widespread qualitative changes in the isozyme patterns of fish during temperature acclimation have been questioned (Shaklee et al., 1977; see literature quoted therein). On account of individual variability in, for example, LDH isozyme patterns, good quantitative measurements of isozyme components in different tissues of individuals from inbred strains appear necessary before a statement can be made about the extent to which the output of different isozyme protomers is affected by exposure to different temperatures and about the frequency of such effects among isozymes in general.

In summary, the observations reported here appear, at this time, in keeping with the controller node stability model. However, no quantitative test of the model is as yet available. Using the model for tentatively interpreting the distribution of the B and C gene products of LDH, we shall say that the controller node of the C gene, as a rule, either is closer to the flip-over threshold than the controller node of the B gene, or is submitted to larger variations in concentration of relevant components of the regulatory kit, as a function of tissue variation and interspecies variation in transcription spectra. When two tissues are likely to be characterized by partly very distinct mRNA spectra, such as presumably tissues in eye and liver, and when a controller node confined to a relatively narrow zone of stability is adapted for maximal expression in one of the two tissues, the chances are significant that the node is not suitable for expression in the other tissue. This can explain a trend towards mutual exclusion of gene expression in two tissues, such as the one described above.

A New Role for Gene Duplication

It has been obvious for a great many years that the duplication of structural genes permits new molecular functions to be evolved while the old functions can be maintained. This idea was implicit in the thinking of a number of investigators, such as Metz (1947) and even earlier workers. It apparently has been clearly spelled out for the first time by Lewis (1951) and also Stephens (1951).

This classical view is incomplete in two respects. Except in Ohno's (1970) work, it does not take into account the likelihood that a new protein function mostly implies new genic regulation. Nor does it take into account the likelihood that a train of modifications in genic regulation represents a "threat" to the regulation of those genes whose activity needs to remain what it was. The first of these circumstances is to be elaborated on elsewhere. Before considering the second circumstance, which concerns us here, the concept of gene duplication needs to be reformulated in terms of fuga duplication.

"Complete" and "Incomplete" Fuga Duplicates. In addition to the coding sequences and possibly some more or less neutral connecting sequences, a fuga comprises all those adjacent sequences that contribute to determining the rate of transcription and processing of the coding sequences. It supposedly includes individual short specific receptor sequences for various kinds of molecular signals, as well as, spread over more or less considerable lengths of DNA, multiple receptor sequences for macromolecules that contribute to the establishment, or prevention, of certain high order structures in chromatin.

(Some of the sequences in the fuga, without coding for polypeptides, may be emittor sequences acting in *trans*, rather than just receptor sequences acting in *cis*, and might code for small RNAs effective in the modulation of gene expression). A fuga, as a rule, will be at least several kilobases (kb) long and can be many kb long. It may be tentatively identified with an individual band plus one interband (or with a band plus parts of the two adjacent interbands) in polytene chromosomes (Zuckerkandl, 1976b).

During evolution, DNA sequence duplication and multiplication, through various mechanisms, apparently occurs at all possible levels of extensiveness, from complete genomes (polyploidy) downwards (Zuckerkandl and Pauling, 1965). Obviously, polyploidy or aneuploidy always lead to duplications of "complete" fugas or, rather, to fugas of identical extension.

The quotation marks acknowledge that the concept of completeness of a fuga may not be exactly definable. The limits of fugas probably are intrinsically imprecise. According to the definition of the fuga, adjacent fugas might even overlap. If sequences in one fuga, in the capacity of receptor sequences, have some influence on transcription rate in the adjacent fuga, they are also part, according to definition, of that adjacent fuga. Multiple receptor sequences functioning in mass ligation (in the binding of a protein by numerous sites spaced in propinquity over a significant length of DNA, Zuckerkandl, 1976b) might be added or subtracted from fugas with *some* moderate effect on transcriptional regulation. There may not be any particular length of sequences that can be designated as "the" complete fuga. A confirmation of this view must await extensive studies of DNA sequence and function.

In intrachromosomal duplication, the duplication process may or may not involve very extensive parts of fugas. A coding sequence alone may be duplicated, or a coding sequence plus only a short noncoding sequence, target, in the transcript, of a processing enzyme to insure the formation of monocistronic messengers. If, in such cases, both duplicates can remain functional, restrictions are likely to be set on the evolution of the duplicate stripped of all or nearly all its associated receptor sequences. Indeed, regulatory divergence is then excluded or limited.

In fact, the widespread differences in regulation between duplicate genes, as found in higher organisms – e.g. between hemoglobin non-α genes (e.g. Petit and Zuckerkandl, 1976) and between many isozyme genes (e.g. Markert et al., 1975) – suggest that evolutionarily effective duplication of structural genes is usually accompanied by the duplication of associated regulatory sequences.

The duplication of a "complete", identical fuga may, however, never be a necessity. In addition, in some cases duplication may overshoot the confines of a fuga to some extent without the effect being considerable. There probably are many variations in terms of the extensiveness of fuga regions involved in duplication, except of course when whole chromosomes or chromosome fragments are duplicated. The extensiveness of the duplication will no doubt contribute to determining whether or not evolutionary divergence in regulation can become maximal. One may consider as maximal regulatory divergence the inclusion of duplicate fugas in distinct eurygenic complexes. This term (Zuckerkandl, 1976a,b) relates to regulation affecting *groups* of (not necessarily adjacent) fugas whose activity characterizes different states of "determination", i.e. different tissues.

Different functions carried out by duplicates are expected to impose different limits on the required extensiveness of the fuga fragment involved in duplication.

Fuga Duplication as Correlated with the Preservation of Established Protein Functions. It was proposed many years ago (Zuckerkandl and Pauling, 1962) that the basic function of multiple "editions" of a polypeptide chain, each of which characterizes a different ontogenetic stage, such as the different non-α hemoglobin chains in humans, is to insure that at all times at least one of these editions will be on active duty. The same hypothesis of course applies to the ontogeny of isozymes (cf. Masters and Holmes, 1972). Different genes coding for the same kind of protein may indeed be thought to be needed during ontogeny for adaptation of regulation more fundamentally than for adaptation of protein function. Only distinct controller DNA sequences, in many cases, probably can be adjusted so as to respond identically when the spectrum of gene activities changes during development. Also, as mentioned, during ontogeny, different sections of the genome may be locked into transcriptionally inactive high order structures (Gersh, 1973; Zuckerkandl, 1976b). In this case, similar copies of a gene that is constantly needed may again be required in several locations in the genome.

Preserving an adequate output of a gene product in the face of the regulatory perturbations caused by differentiation probably is a challenge answered by gene duplication in the form of fuga duplication. Appropriate mutations may adjust the duplicate controller nodes to conditions in different tissues. In fact, two fugas may not suffice for filling this role of mutual proxies in all tissues. Several isozymic genes may often be at the disposal of an organism, allowing it to "play" with them in response to different states of the regulatory network.

Suppose that only one fuga of a certain kind were present and that its regulation were affected in one tissue. A mutation that would neutralize the effect in that particular tissue might well offset the fuga's regulatory balance in another. An evolutionary change of regulation in a given tissue is thus likely to create a dilemma. An obvious way out would be so to stabilize the controller node of each unique gene that this gene will indeed be properly transcribed or processed in all tissues where its function is needed. But this may not be feasible, either because the background of transcriptive or processing activities in one tissue may be extreme with respect to the ensuing concentration of a component of the regulatory kit, or because the unique gene is situated at a juncture of the regulatory network such that its maximal stabilization (buffering) would compromise critically important stability states of other controller nodes. The only alternative, in such a case, is likely to be fuga duplication. If it were not for the existence of duplicate and closely similar fugas, appropriately different in their regulatory sections, differentiation and, therefore, progressive evolution might be considerably limited: each functional innovation might endanger too many established functional relationships. Evolution would be hampered in its most significant process of change, alteration of the patterns of gene control.

This generalization does not necessarily apply to the control of any particular protein, for which a "choice" between a single highly stable fuga and multiple less stable fugas may exist. But it may well apply to the genetic system as a whole, in which opposite trends towards stabilization and destabilization of controller nodes are likely to

either case functional divergence of the other type of fuga component is likely to follow suit, simply because the opportunity to follow suit exists (see, below, the remarks on the "fitness treadmill"). Of course only a very small fraction of all possible mutations and combinations of mutations can be tried out within the limits of evolutionary time. But similar functional properties in protein molecules are expected to result from a number of different amino acid sequence changes and combinations thereof that is also very large (Zuckerkandl, 1976c). Hemoglobins differing in a number of sequence characteristics may have similar oxygen affinities, the ones under identical physicochemical conditions, such as after dialysis against the same buffer, the others under the different physicochemical conditions that prevail in the red cells from different species (Allan and Jandl, 1960; cf. Zuckerkandl, 1974). Different sequences will not permit a comprehensive *set* of partial functions of a protein to be identical. But the values of any one or of a few partial functions (general and specific) are apparently compatible with a large array of different sequences. As indicated by the correlation between sequence and oxygen affinity in hemoglobins, such sequences have a significant chance of being "found" repeatedly. Thus one may suppose that, in general, any structurally possible functional change capable of increasing the organism's fitness is likely eventually to occur.

On the basis of the controller node stability model, selection for a regulatory variant in one tissue, e.g. selection for a high rate of synthesis of a particular mRNA in that tissue, will automatically determine the value of this rate in other tissues. A "puzzling" tissue distribution of isozymes may thus be the accidental effect of natural selection geared to isozyme function – and therefore regulation – in mainly one tissue. This adjustment of the rate to conditions in one tissue is expected to "carry" the different rates of synthesis of the same mRNA in other tissues. A given variant of such a situation would be maintained as long as the regulation of the isozyme fuga in the other tissues is not afflicted with negative selection coefficients whose sum more than compensates for the positive selection coefficient in the first tissue. When this happens in a species, the distribution pattern of this particular isozyme over different tissues could be profoundly modified, by virtue of a different tissue becoming predominant in the selection for the controller node's regulatory state.

The existence of an isozyme fuga duplicate would allow the predominant positive selection coefficient relating to this second isozyme to focus on some further tissue or tissues, with again a train of more nearly neutral consequences regarding the amounts of this second isozyme in the remaining tissues.

Thus we no longer need to expect each quantity of isozyme in each tissue to be functionally meaningful, even though the isozyme distribution, as found, is likely to be the consequence of a functionally meaningful, nonneutral evolutionary development.

Difficulty of both Selectionist Interpretations: The Use of Single Genes

We have considered the role duplicate genes play in exercizing basically identical functions within the same organism. In this respect duplicate genes may however appear superfluous, if indeed one and the same type of function can be carried out by unique as well as by multiple genes. We here use "unique" to mean the absence of any only moderately diverged gene duplicates. For instance, lamprey is reported to have

be at work. On the one hand, genomes are apt to display a trend towards stabilization (namely buffering against the effects of changes in concentration of components of the regulatory system) of a maximum proportion of controller nodes. This is expected (1) because higher stability is more economical in terms of the number of duplicate genes needed, and therefore of the energy expenditure that the management and activity of each gene implies, and (2) because the larger the proportion of unstable controller nodes is, the more vulnerable the organism must be to (a) environmental change, (b) mutational change. Nonetheless the mean stability of all controller nodes may be moderate. Another trend, indeed, is expected to run contrary to stabilization. If we consider a system whose components are supposed to interact, like the genetic system, the average stability of the components cannot be very high. Effective interaction implies the destabilization, by the component that is acting, of the target component that is acted upon. An illustration is furnished by the many fugas that must be so equilibrated that gene expression can be a function of graded positional information within a tissue. They must be sensitive to relatively small changes in concentration of components of the system of transcription and processing. In the case of obligatorily unstable controller nodes, the use by the same organism of differently equilibrated equally unstable controller nodes built around duplicate fugas may thus lead, in this organism, to increased *sensitivity* to environmental change and to a correlated *adaptability*, without concomitant loss of overall *stability* of the system. The simultaneous fulfillment of these three conditions appears to be a basic prerequisite for progressive evolution. A priori the problem of reconciling sensitivity with stability might seem difficult to solve. A solution seems to be offered by a set of unstable duplicate fugas with their narrow stability zones set to different values of a wider stability range.

Even without any regard to progressive evolution, the simultaneous presence in an organism of several closely similar fugas geared to essentially the same protein function may confer upon the organism an evolutionary, as well as developmental, flexibility that it would not otherwise possess. Such a set of similar, but not identical, fugas represents a buffer system with respect to local or temporal variations in the output of components of the regulatory kit. Evolutionary changes in a cell's RNA profile, if they interfere with the synthesis of a particular mRNA, may be neutralized by compensatory sequence changes in one of the duplicate fugas controlling this type of mRNA.

The potential that gene duplication offers to evolution has, so far, been considered in a somewhat onesided way. Gene duplication or, rather, fuga duplication, beyond making it possible to evolve new functions without abolishing the established ones, seems to have a further dimension, of opposite direction. Duplication not only is the condition of evolutionary novelty, but apparently also the condition of constancy in the presence of novelty. Certain evolutionary processes that have so far been understood as provisions for change may be basically provisions for constancy.

Functional Changes and Tissue Distribution Following Fuga Duplication. In many cases the functional specialization of homologous polypeptide chains may thus be superimposed only secondarily on a pattern of primary regulatory adjustments, e.g. in response to local changes in mRNA profiles during tissue evolution. In other cases it is the diversification in transcriptional control that is expected to be secondary. Whether the primary functional divergence is in a regulatory or a coding sequence, in

only one expressed LDH gene (Markert et al., 1975). It is apparent from this case that the different types of tissue of a complex animal can do with a single LDH gene. This seems to be at odds as much with the "necessity" for modulating protein function within an organism as with the "necessity" for modulating gene control.

An objection may be raised against this judgment. Requirements in regard to gene multiplicity may vary with the ecology and ethology of organisms, as correlated with selection pressures. Thus unique genes may do in some situations and not in others.

I believe that this remark does not really promise to rid us of the problem posed by unique versus multiple genes. Irrespective of ecological and ethological differences, we are, in all cases, considering highly complex organisms. The functions of their various organs are differentiated. An *internal* multiplicity of situations is present, whatever the environmental situations. Interior diversity can be considered to be as likely to be demanding of special adaptations as exterior diversity is. The fact is that intracellular differences in the environment of the cell nucleus and, within the nucleus, of chromatin are, in some species, correlated with multiple genes of a kind, whereas in other species one gene of the kind suffices for all intracellular situations. The use of unique genes at all or nearly all stages of development and in all or nearly all tissues suggests that it is indeed possible for a complex organism so to equilibrate the controller node pertaining to a structural gene that this gene will be adequately expressed as mRNA in spite of established variations in tissular transcription and processing profiles.

How can some genes, such as the cytochrome c genes, be active in nearly all tissues at nearly all times, whereas other genes are limited in the spatial and temporal distribution of their activity? One possibility, in terms of variation of controller node stability, has already been considered. Another possibility (see also footnote 7) is that the universally active genes are immune to some mechanism of control to which other genes are subjected. Two distinct levels of chromatin structure could be implicated in this reduction of regulation. Bands and interbands in polytene chromosomes can probably be considered to visualize traits of chromatin structure of general significance for eukaryote organisms (Hennig, 1974). Two types of bands may be distinguished in polytene chromosomes of flies: thick bands that contain a long enough segment of DNA duplex for this duplex to be, in principle, able to accommodate, at interphase, a very high order structure ("quintary" structure) of deoxyribonucleoprotein; and thin bands that contain too little linear DNA for a structure of such high order to be formed (Zuckerkandl, 1976b). Genes connected with thick bands might thus be subjected to an additional type of transcriptional control absent from thin bands. Genes that are always transcribed provided the *general* conditions for transcription be fulfilled (including the presence of the appropriate RNA polymerase) would thus be found in thin bands only.

At the structurally lower levels of regulation of transcription, which would be common to both thick and thin bands but would represent the only levels in thin bands, a reduction in regulation would imply the absence of shifting regulatory equilibria and, for all practical purposes, the absence of flip-over thresholds. This could be achieved if the functional units of gene action failed to include receptor sites for specific repressors and were the targets of activators produced in excess in all tissues. In this case, some variation in the rate of synthesis of activator would have at best a limited effect on rates of transcription in different tissues.

Although the molecular basis of the variations in the distribution of gene expression can be explored in future research along the lines indicated, the question will remain as to why organisms resort to using duplicate genes in some instances and refrain from using them in others. With respect to the LDH genes, the singularity of the gene in lamprey is not a unique case. The problem raised by "one-isozyme" organisms has been recognized some time ago (Markert, 1968). Whereas B subunits of LDH are present in most tissues of most vertebrates, and generally predominant in heart muscle and brain, some families of teleost fish, for instance among the Pleuronectiformes and Tetraodontiformes (Markert et al., 1975), display a very limited expression of the B gene. To proponents of the view that selection for enzyme function explains isozyme patterns, "this simplification of enzyme patterns is surprising and perplexing" (Markert et al., 1975).

A Solution: The Concept of Genetic Sufficiency

Genetic Sufficiency

There is thus an apparent contradiction between the existence, in many higher animals, of multiple closely similar genes whose products appear to be functionally adapted to different cellular conditions and the absence of this multiplicity in other higher organisms. Likewise, if we take the facts at face value, there is an apparent contradiction between the presence of multiple regulatory adaptations of fuga duplicates in one organism and the absence of this multiplicity in other related organisms. The concept of genetic sufficiency (Zuckerkandl, 1976c) resolves this opposition at both levels. Let us situate and define the concept.

The biological significance of a mutation could be evaluated a posteriori by its contribution either to the survival of a species and lineage, or to "progressive" (e.g. Zuckerkandl, 1976a) evolution. Progressive evolution is not to be considered here and is of course subordinate to survival. Survival is closely correlated with adaptedness. In his important paper on fitness of genotypes versus fitness of populations, Ayala (1969) shows that merely measuring a reproductive differential ("fitness") falls short of providing answers to the essential evolutionary question of adaptation. To put it in very general terms, the measure of fitness is therefore quite inadequate for defining the biological significance of a mutant. This implies that positive natural selection (which precisely consists of a mutant's relative superiority in reproductive efficiency) does not *per se* confer biological significance upon a mutant.

The absolute average number of adult individuals produced per unit time and the absolute average population size are in important respects more informative than fitness. They offer a means of measuring adaptedness, defined as the ability of an organism, a genotype, or a population to survive and reproduce in a given environment (Dobzhansky, 1968; Ayala, 1969). However, there are many levels of adaptedness over and above the survival threshold.

Genetic sufficiency precisely designates the domain above this threshold value of adaptedness, and genetic insufficiency, the domain below it. Genetic insufficiency is thus all-or-none. It characterizes the state of any gene in relation to a genome and to an environment, or of any genome in relation to an environment, at or above the minimum level of adaptedness compatible with the survival of the species. More precisely, are genetically sufficient all those variants of a fuga, including null mutations, that, in

the homozygous state, are compatible with the survival of the species under conditions of genome structure and-environmental variability found in the species.

The state of genetic sufficiency allows no prediction as to long term survival of the species. The survival value genetic sufficiency relates to is that of short term survival, namely survival under conditions of a continuation of quasi-"normal" environmental variations. However, the potential for survival of the species implied in the state of genetic sufficiency doesn't need to be specified in terms of any particular survival time. Indeed, the question of sufficiency or insufficiency of a wild type gene will be settled by comparison with a successful mutant gene, *under the environmental conditions under which the mutant gene has spread in the population*, irrespective of the (unpredictable) survival time the mutant gene (and, if it is shown to remain sufficient, the wild type gene) would actually confer upon the species if the genotypic composition of the species remained indefinitely invariant.

The method of detecting the superiority of an allele in terms of differential numbers of offspring of course implies the simultaneous presence of competing alleles in the tested population (Ayala and Anderson, 1973; Clarke, 1975). A test of genetic sufficiency, on the other hand, would exclude mutant or polymorphic alleles. Genetic sufficiency of an allele would be established if, in the homozygous state and under given environmental conditions, the number of individuals in the population increased, remained constant, or displayed a limited decrease, down to a new level of constancy, high enough not to significantly decrease the chances of survival of the population. Some guidelines for a general experimental exploration of genetic sufficiency are given in the appendix.

The selectionist interpretation of genetic change has so far been polarized by the concept of functional "requirements", functional "necessities" that are being filled by available mutations. Whenever observed genetic variation appeared to be poorly accounted for by the postulate that functional requirements differed in various circumstances and were being satisfied in various ways, the only alternative has been the view that molecular variation spread in populations, in the absence of any functional requirement, under selective neutrality. If a mutation was thought to have spread in the absence of a functional requirement, selection was not considered; only neutrality was. Beyond the usual selectionist view and the neutralist view, a third possibility, however, combines some components of the first two. This third possibility is that most mutations indeed spread in the absence of any functional requirement (this is the neutral mutation theory component), but nevertheless by selection. In that case the fixation in a population of a certain mutation or of a certain duplicate gene are brought about by means of a relative or absolute positive effective reproductive differential (= fitness increase). Yet the absence of these mutations or duplicate genes, when it is observed, is not a puzzle.

Below and up to the threshold of genetic sufficiency, adaptation, i.e. an *absolute* differential in effective reproductive capacity, is a necessity for survival, above the threshold it is not. Although natural selection certainly is critical for survival in a fraction of its operations, since, after some environmental changes, adaptation is necessary to bring the population back up to the survival threshold, the genetic sufficiency concept emphasizes, contrary to a commonly held view, that natural selection is not to be conceived as *generally* linked to the survival of populations.

The Luxury of Increased Fitness

Survival of the species is, of course, by itself, evidence of the "sufficiency" of any genetic situation that is generalized in the species. Thus, the variations in the number of duplicate genes endowed with different functional adaptations used by one organism suggest that in the case of any *particular* type of protein, this number is generally not critical for the survival of species and lineages. In the absence of the product of the LDH B gene, the product of the A gene appears to be "sufficient". Another example is furnished by human homozygotes for hereditary persistence of fetal hemoglobin who, as adults, have exclusively fetal hemoglobin and can be practically normal. Among the few cases reported, one was a 53-year-old truckdriver (Siegel et al., 1970). There is no indication to the effect that homozygotes for this genetic condition would not be sufficiently fertile to keep the human race in existence. If the condition were generalized (which presumably it isn't because of lower fitness in competition with carriers of the adult hemoglobin gene), mutations would no doubt soon occur to remove the slight anomalies noted. Furthermore, not only adults can live normally with fetal hemoglobin, fetuses can live normally with adult hemoglobins. In mammals such as the mouse and the horse, the same genes are used in the older fetus and in the adult (Craig and Russell, 1964; Stockell et al., 1961). A comparable case is that of a 64-year-old man who totally lacked one of the LDH isozyme components, the "heart" isozyme, LDH B (Kitamura et al., 1971). Some other dispensable proteins are listed by Wilson et al. (1977).[8]

Such cases throw considerable doubt on the "necessity" of many selected functional adaptations. Functional differentiation of multiple genes therefore may be considered not to be a condition for the survival of species and lineages but, in many individual cases, to be a luxury, the consequence of opportunities that happened to arise. The luxury is, for a population already adapted to survive, to become adapted to produce a larger number of offspring, or to create for the wild type conditions of reproductive inferiority that did not exist before the appearance of the mutant.

Selection of certain functional variants of gene duplicates will occur once the duplicates are available. Yet, in a significant proportion of cases, the spreading in a population of protein variants may be as gratuitous, in relation to the survival of the species, as the spreading of a neutral mutation. Selection itself may, in this sense, be a rather "neutral" process – an automatic mechanism of spreading unnecessary advantages, once they are offered.

This of course does not preclude that a truly neutral, or a gratuitously selected mutation can, at some future time, become critical for survival. Such a circumstance could arise through a change in conditions that neither an observer would foresee, nor nature "work for" in advance.

---

[8] A relatively little known example of apparent dispensability of a protein, even though not one of genetic change, is the following. The spidercrab *Maia squinado* is capable of living at least several months without any hemocyanin in its blood (Zuckerkandl, 1960). During that period, individuals deprived of hemocyanin through starvation after moulting seem to be normally reactive.

The Fitness Treadmill

There is then, in the biological world, a compulsively turning fitness treadmill. If there were such a thing as a "law of biodynamics", it would be the perpetual trend towards an increase in relative fitness, namely fitness of a mutant relative to the fitness of the wildtype in the presence of the mutant. This relative increase in fitness is often also absolute (i.e. the increase in rate of production of viable and fertile offspring is absolute). It is most of the time quite independent of the fitness to *survive*. Fitness continues to increase when adaptedness, with respect to prevailing conditions, has long been sufficient. The mechanical fitness increase is an inescapable consequence of the mutation process itself. This process continuously changes the conditions of effective fertility of the mutants as well as, frequently, of the nonmutants. Even if a relative increase in effective fertility is rather rare in comparison to the number of instances in which a relative decrease occurs, for increases to happen it is sufficient for the mutational process to continue. The two evolutionary trends discussed here, selection for functional diversification of proteins and selection for diversification in transcriptional or processing control appear as equivalent aspects of the fitness treadmill.

Organisms must remain imperfect because of the competition of functions for structures, because of mutually partly exclusive functions, and because of the resulting compromises (Zuckerkandl 1976 c). Nonetheless, it would seem that, thanks to the fitness treadmill, individual functions are mostly more nearly perfect than they need be from the point of view of the potential for survival of the species and lineage. Individual functions, however, are not alone to be considered.

The observation has been made, in a variety of organisms, that isozyme multiplicity, as an expression of nonallelic genes, is significantly correlated for different enzymes belonging to the same metabolic subsystem (Sing and Brewer, 1977[9], Senkbeil and White, 1978). Among the components of such a subsystem, the introduction of any *individual* gene duplicate as a functional variant, or the loss of one, may indeed be gratuitous from the standpoint of genetic sufficiency. However, the system of components as a whole may not be gratuitous. The *accumulation* of directional effects, whether exclusively selective or intermingled with drift, creates new structural and functional backgrounds, and finally a different organism -- a new basis for a test of genetic sufficiency.

Such building up of new sets of relationships, of new functional subsystems and eventually systems may be mainly a product of the fitness treadmill. Thus progressive evolution may directly result from a process that is unnecessary for survival. This may be a least a partial answer to the old question: why do higher forms of living systems evolve at all, since the lower forms are obviously well adapted? Higher forms evolve, because sufficiency in adaptedness in no way prevents the fitness treadmill from continuing to turn.

---

[9] A strong predominance, in the experimental sample, of gene duplication over polymorphism as the source of the observed enzyme multiplicity is probable, but not proven, as pointed out by the authors.

Dispensability, Selection, and Trends in c-Value

The dispensability of a protein is the genetic sufficiency of its null allele. It would be mistaken to think that the dispensability of a protein implies the absence of positive natural selection for the protein, or, conversely, that the presence of selection implies that the selected object is indispensable.

For example, the fact established by Ferris and Whitt (1977) that many variant duplicate genes cease to be expressed in tetraploid bony fishes (whether the duplicates are bodily eliminated or functionally silenced) by no means indicates that the lost protein variants had not been selected for. If these variants reappeared, they might be selected again. This would surely be the case of the dispensable genes mentioned above or in Wilson et al., (1977). For instance, it seems safe to assume that a serum albumin gene, if introduced into a human population lacking this gene, would be selected for. In general, null-states are rather likely to lose out in renewed competition with plus-states, if the plus-state has heretofore been preserved over millions of years and has thus indirectly demonstrated its effective functionality. In the absence of such competition, the null-state, however, may well be "sufficient".

It has been proposed (Allendorf, 1978) that "those proteins which tend to be polymorphic should retain gene duplication if they are variable because of their frequent association with some form of balancing selection. But if these proteins are ploymorphic because of a higher neutral mutation rate, then they should tend to lose gene duplication". Since, in the case of more highly polymorphic and heterozygous proteins, gene duplicates would seem, on the average, to be lost more rapidly than duplicates controlling less polymorphic proteins (Allendorf, 1978), the author feels that the neutrality theory has gained substantial support. In connection with the present treatment, it may be of interest to point out why this criterion for distinguishing between the selection theory and the neutral mutation theory is probably invalid.

1) In general, alleles differ from each other by only a few substitutions and in many cases, apparently, by just one. To my knowledge, the largest number so far found of sequence differences between polypeptide chains controlled by allelic genes is between the β^A and the β^B alleles of sheep hemoglobin (see Dayhoff, 1972), namely 7 amino acid differences. This figure appears to be much higher than the average sequence difference between alleles in general, probably including alleles for which balancing selection has been demonstrated. Structural divergence between contemporary alleles therefore has started in relatively recent evolutionary times. We have the choice between assuming that, in more ancient evolution, alleles did not diverge - a very unreasonable hyphothesis - or concluding that alleles are continuously lost, presumably even if maintained in populations by balancing selection. If such alleles can be lost, moderately diverged gene duplicates can be lost too. The loss by itself does not testify against positive selection of the gene. It is not necessary to assume that in all, nor even most, cases, first, a positive selection coefficient drops to zero or to a negative value and the loss of the gene then ensues. "Founder" episodes - tight populational bottlenecks followed by definitive reproductive isolation from the original population - are expected automatically to lead to the loss of alleles, whether they are selected for or not (Mayr, 1970; Petit and Zuckerkandl, 1976).

2) If a duplicate, before being lost, has been maintained in a functional state over a significant time span, selection against deleterious mutations must have occurred before the loss. In the absence of such negative selection, functional inactivation would

indeed be very rapid. From the phylogenetic tree of the tetraploid catostomid fish it can be deduced that the loss of duplicate isozyme genes occurred throughout time and thus, in some lines of descent, after the duplicates had been maintained for many millions of years (Ferris and Whitt, 1978). It is thus unwarranted to consider that these duplicates represented neutral genes, even if they accumulated more neutral mutations than other genes.

In case a significant proportion of the mutations that lead to accepted amino acid replacements were indeed neutral, this would only imply that neutral mutations can occur at *some* protein sites, - presumably at the most variable sites. It is common knowledge that some sites usually exclude all mutations and further sites exclude most mutations. In addition, even from the most highly variable sites in proteins, most amino acid substituents appear to be eliminated (Barnard et al., 1972; Zuckerkandl, 1975).

The "neutrality" of any polypeptide *site*, let alone of any functional gene as a whole, is thus excluded. Indeed the neutrality of any gene is excluded; only functional genes are detected as genes, and a nonfunctional allele cannot be detected if there is not a functional one. Admittedly, a functional duplicate of a gene might be neutral, but this hypothesis can no longer be entertained after the time required for functional inactivation through unchecked random mutation. Recorded duplicates usually have been in existence for much longer periods. Thus, nearly all genes are selected for. This does not preclude that null alleles of some continuously selected genes are sufficient. One may wonder if confusion can be avoided in this field, lest this be recognized.

Ferris and Whitt (1977) found that "advanced" catostomid fishes (i.e. forms morphologically more divergent from the ancestral form) have lost the expression of more duplicate genes than primitive species (i.e. whose morphology is more similar to that of the ancestral form). This would be expected on the basis of two assumptions. One is that morphologically divergent, specialized forms go through a larger number of population bottlenecks, founder episodes, episodes of ecological isolation than their less specialized relatives. This would seem to be a reasonable supposition, since functional specialization implies the limitation to special ecological niches and the isolation from populations in other niches. The second assumption is that such founder episodes, if repeated, tend to lead to the elimination, by a random process, of all DNA, coding and noncoding sequences, whose loss is compatible with genetic sufficiency, - perhaps a significant fraction of DNA is most higher organisms.

One may thus account, at least in part, for the general observation that more specialized members of a group tend to be lower in c-value (haploid DNA content per cell) than less specialized members (Stebbins, 1966; Hinegardner, 1976, and literature quoted therein).

On the other hand, if, as suggested above, the fitness treadmill is an important factor in the evolution from lower to higher organisms, we may expect that the fitness treadmill will tend to increase c-values. Indeed, higher organisms in general require more DNA per haploid genome than lower organisms, as can be seen by comparing the minimum amounts of DNA found in different higher taxa (Sparrow et al., 1974). In species that conserve large effective population sizes, the number of duplicate structural genes, and duplicates of noncoding DNA sequences, probably tend to increase constantly. The increase is thought to result from the "automatic" increase in fitness prompted by the functional offers made by protein variants and by noncoding DNA sequences

involved in various regulatory processes. In metazoa and metaphyta, the selective prevalence (to avoid the phrase selective "advantage", Zuckerkandl, 1976 c) of functional capabilities introduced by additional DNA appears to outweigh the selective disadvantage of an increased energy expenditure for DNA and RNA synthesis.

As we go up the ladder of progressive evolution, not only is more and more DNA generated through the workings of the fitness treadmill, but more and more also is necessary for genetic sufficiency. The minimum c-value of higher animals is not reducible to the level of c-values or much lower organisms.

Increases and decreases in DNA content may then occur alternately in related lineages, according to whether frequent episodes of population bottlenecks bring genetic sufficiency thresholds to the fore, or whether the fitness treadmill alone operates constantly.

Biologically Important and Unimportant Variants

The evidence on the tissular and organismal distributions of isozymes suggests the inadequacy of a widespread assumption, even though it is more often implicit than explicit; the assumption according to which, among accepted mutations, the dividing line between biologically important and unimportant variants, between required and gratuitous variants, passes between selection and neutrality. Under some circumstances this dividing line will indeed be so situated; but not generally so. These circumstances are expected to be a function of the variability of the external environment. Under constant environment, a new mutation may appear than confers upon the carriers an absolute advantage in numbers of viable offspring, or a relative advantage only, by depressing the fertility of the former wild type. Such mutations, even though selected, will be biologically gratuitous like neutral mutations. On the other hand, under environmental stress, the availability of a given mutation may be critical for the survival of the species. Selection and biological importance of a mutation will then go hand in hand.

Thus, the greater the frequency, intensity, and variability in kind of environmental stresses, the more necessary, in terms of survival of the species, the selection of variants may be expected to be on the average; the greater the stability of the environment, the more gratuitous this selection, on the average. Under the latter conditions, if there were a world without neutral mutations one wouldn't know the difference; - except of course for a distinction in terms of rates of spreading in populations of gratuitous variation.

The environment of the mammalian fetus is relatively stable in several ways. Cyclic "environmental" variations during fetal life are dampened, as compared to cyclic variations in external environment adult organisms are exposed to; the fetus is less subjected to noncyclic, "catastrophic" environmental events; and during evolution, the fetal environment changes as much as the internal environment of the mother-organism, which is certainly less (e.g. with respect to temperature or, chemically, with respect to "food") than the organism's external environment. If indeed environmental changes frequently bring about a state of genetic insufficiency, selection under conditions of genetic sufficiency is expected to be reponsible for a significantly larger fraction of substitutions in characteristically fetal than in characteristically adult macromolecules. In fact, however, the evolutionary rates of amino acid replacements

in fetal and adult hemoglobin chains are about the same (Derancourt et al., 1967; Fitch, personal communication). Given the relative stabilities of the environments in relation to which the two types of hemoglobins evolved, this equality of replacement rates does suggest that, even in adult hemoglobin and in a variable environment, most substitutions occurred under conditions of genetic sufficiency. It should be verified whether, in general, mammalian fetal and adult "editions" of proteins have indeed evolved at approximately equal rates since the time of the appearance of mammals.

It is likely to be so. Indeed, in spite of a number of recorded exceptions, in part of uncertain authenticity as "true" exceptions, [10] the molecular evolutionary clock (cf. Wilson et al., 1977) generally runs approximately (Fitch, 1976) on time, provided the functional differences between the protein "editions" remains small (as also stressed by Kimura and Ohta, 1974). This latter conditions is probably best expressed by linking the approximate constancy of the "clock" to a quasi-constancy of weighted functional density in the protein (Zuckerkandl, 1976d). The variants of a protein that obey the "clock" appear in organisms displaying differences in their internal environments and living in different external environments. These environments, obviously, have not changed to the same extent since the time of the common ancestor of different organisms, particularly if we compare lineages that have led to more progressive forms with lineages that have remained more primitive. In the latter the internal environments, in particular, have remained more constant (or the variations they are subjected to). Differences in rates of environmental change apparently do not significantly modify molecular evolutionary rates. "Vitally important" selection associated with greater environmental change cannot have been prominent. It appears rather likely, therefore, that, quite generally, amino acid replacements in proteins occur, for the most part, under conditions of genetic sufficiency, either by genetic drift, or as automatically generated products of the fitness treadmill.

Besides the environment of the mammalian fetus, the deep sea is another exceptionally stable environment (Gooch and Schopf, 1972; Ayala and Valentine, 1974; Ayala et al., 1975; Valentine and Ayala, 1975). The accumulation of polymorphisms in the deep sea may also be tentatively interpreted as due to selection under genetic sufficiency (see also Zuckerkandl, 1976c). A remark about genetic sufficiency in relation to polymorphism in general is relevant here. Cases of limited deleterious effects of inbreeding (Clarke, 1972) and cases of low degrees of polymorphism in some natural populations (Selander et al., 1971) suggest that polymorphism also may in most cases be considered as gratuitous, even when selected. After proving that selection established one allele in a population to the exclusion of or in balance with another (Clarke, 1975), it remains to be shown that, in the absence of a more "sucessful" allele, the species has a lesser potential for indefinite survival under realistic environmental conditions. Therefore, even if polymorphism could in most cases be explained in selectionist terms on the basis of solid experimental evidence, one might perhaps ask the potentially destructive question: and so what?

In summary, an alternative is offered here to the concept that variations in activity of different isozyme genes in different tissues and species represent functional adaptations of the proteins to different conditions. The alternative is that they represent

[10] inasmuch as genes whose relationship is treated as orthologous (duplication-independent) might in reality be paralogous (related through duplication-dependent homology).

functional adaptations of the regulatory system to different profiles of gene expression in different tissues. It is however pointed out that, in terms of their value for the survival of the species, most of these adaptations, whether in protein function or in regulation, are, individually, probably gratuitous and that, in this respect, most selected differences are not distinct from neutral ones.

Appendix

Guidelines for an experimental exploration of genetic sufficiency

A quantitative study of genetic sufficiency entails the exploration of the limits of sufficiency under various sets of environmental conditions. When the locus explored is polymorphic, the experiments are to be repeated for the different alleles, always tested in the homozygous state. For each set of environmental factors considered, the eventual extinction of the population or its indefinite short term survival is to be recorded as 0 or 1. Each gene is evaluated in relation to a genetic background with an average and a variance characteristic of the population studied. For each gene, a multidimensional spectrum of genetic sufficiency is to be established – rather laboriously – over a variety of environmental conditions in different combinations. One thus measures the ranges over which sufficiency varies in wild type and in mutant genes as a function of variation in single as well as compounded environmental factors. The total multidimensional "volume" of conditions of sufficiency obtained is to be compared with the average "volume" characteristic of wild type genes in general.

The range of values for the integrals of sufficiency spectra relative to wild type genes should reveal how small these multidimensional integrals over all conditions of sufficiency may be for individual genes without compromising the survival of the population, – in other words, how restricted the range of conditions compatible with adequate synthesis and function of a protein may be. One might decide somewhat arbitrarily to consider as sufficient any mutant gene for which the integral I_S over the multidimensional sufficiency spectrum is not below a value that itself is one standard deviation below the mean of I_S for wild type genes. A second criterion is also to be applied. Irrespective of how high I_S is, the narrowest range of sufficiency recorded for any single environmental parameter in comparison with the same value for average wild type genes may be the bottleneck of sufficiency. Therefore, for a mutant to be considered as sufficient, none of the tolerance ranges recorded for the mutant must be narrower than the narrowest range found in wild type genes. Alternatively, one may decide that the tolerance range of the mutant must not be below, say, two standard deviations from the mean of the corresponding wild type gene tolerance ranges for the same environmental factor.

The distinction between unique and multiple, functionally very closely related genes must however be heeded here. Comparisons between mutant and average wild type genes are to be made exclusively within corresponding multiplicity classes. For instance, if the mutant belongs to a gene family that controls a type of isozymes, and if there are two genes in the family active at the same developmental stage, e.g. the adult stage, then the narrowest sufficiency range found for the mutant within the multidimensional spectrum of the survival response to various environmental conditions is to be compared with a mean value for the range established in relation to the same environmental conditions for different wild type genes belonging to the same multiplicity class, namely the class for two genes active in the same tissue during adulthood. Indeed, the presence of two nonallelic genes with overlapping tolerance ranges will allow each of the genes to display a narrower tolerance range within the limits of genetic sufficiency than would be the case for single genes.

Such experiments and calculations would explore the precise relationship between survival on the one hand and, on the other hand, the mean, range, variance, and combinatorial properties of the functional response of proteins, or of DNA sequences, to environmental factors and against a given genetic background. Furthermore, the experiments and calculations would allow one to distinguish biologically significant processes of natural selection from those, probably majoritarian, whose short term significance in terms of a change in potential for survival is close to nil and whose long term significance is unpredictable.

Acknowledgments. I express my most special appreciation to Gregory S. Whitt, Stephen D. Ferris, and Suzanne E. Fisher for the exceptional attention they accorded this paper in reviewing and criticizing it and for their essential contribution to its revision. I also warmly thank Allan C. Wilson and his collaborators for important information and comments, Francisco J. Ayala for discussion of the concept of genetic sufficiency, Henri Dirren for assistance, and an anonymous referee for raising questions that needed to be clarified.

References

Allen, D.W., Wyman, J., Jr., Smith, C.A. (1953). J. Biol. Chem. **203**, 81–87
Allendorf, F.W. (1978) Nature **272**, 76–78
Ayala, F.J. (1969). Can. J. Genet. Cytol. **11**, 439–456
Ayala, F.J., Anderson, W.W. (1973). Nature New Biology **241**, 274–276
Ayala, F.J. Valentine, J.W. (1974). Marine Biol. **27**, 51–57
Ayala, F.J., Valentine, J.W., Hedgecock, D., Barr, L.G. (1975). Evolution **29**, 203–212
Bailey, G.S., Wilson, A.C. (1968). J. Biol. Chem. **243**, 5843–5853
Bailey, G.S., Wilson, A.C., Halver, J.E., Johnson, C.L. (1970). J. Biol. Chem. **245**, 5927–5940; and see "The Isozymes," Vol. III and IV, Markert, C.L. ed. New York: Academic Press
Bailey, G.S., Tsuyuki, H., Wilson, A.C. (1976). J. Fisheries Res. Board of Canada **33**, 760–767
Barker, W.C., Dayhoff, M.O. (1972). In: Atlas of Protein Sequence and Structure, Dayhoff, M.O., ed., National Biomedical Research Foundation, Washington, DC pp. 101–110
Barker, W.C., Hunt, L.T. (1976) ibid., Suppl. 2, pp. 9–20
Barnard, E.A., Cohen, M.S., Gold, M.H., Kim, J.K. (1972). Nature **240**, 395–398
Bonner, J.J., Pardue, M.L. (1977a). Cell **12**, 219–225
Bonner, J.J., Pardue, M.L. (1977b) Cell **12**, 227–234
Boyer, S.H., Noyes, A.N., Boyer, M.L., Marr, K. (1973). J. Biol. Chem. **248**, 992–1003
Britten, R.J., Davidson, E.H. (1976). Federation Proc. **35**, 2151–2157
Champion, M.J., Whitt, G.S. (1976). J. Exp. Zool. **196**, 263–282
Clarke, B. (1972) Amer. Nat. **106**, 1–13
Clarke, B. (1975). Genetics 79, Suppl. 101–113
Craig, M.L., Russell, E.S. (1964). Develop. Biol. **10**, 191–201
Dayhoff, M.O., ed. (1972). Atlas of protein sequence and structure, Vol. 5, National Biomedical Research Foundation, Washington, D.C.
Dayhoff, M.O. (1973). In: Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 1, Dayhoff, M.O., ed., National Biomedical Research Foundation, pp. 51–58, Washington, DC
Derancourt, J., Lebor, A.S., Zuckerkandl, E. (1967). Bull. Soc. Chim. Biol. **49**, 577–607
Dobzhansky, Th. (1971). Genetics of the Evolutionary Process. New York: Columbia University Press
Elgin, S.C.R., Weintraub, H. (1975). Ann. Rev. Biochem. **44**, 725–774
Engel, W., Kreutz, R., Wolf, U. (1972). Biochem. Genet. **7**, 45–55
Eventoff, W., Hackert, M.L., Rossmann, M.G. (1975). J. Mol. Biol. **98**, 249–258

Eventoff, W., Rossmann, M.G., Taylor, S.S., Torff, H.J., Meyer, H., Keil, W., Kiltz, H.H. (1977). Proc. Natl. Acad. Sci. 74, 2677–2681
Ferris, S.D., Whitt, G.S. (1977). Nature 265, 258–260
Ferris, S.D., Whitt, G.S. (1978) Syst. Zool. (in press)
Fitch, W.M. (1973). Ann. Rev. Genetics 7, 343–380
Fitch, W.M. (1976). Molecular evolutionary clocks. In: Molecular evolution. F.J. Ayala, ed., pp. 160–178. Sunderland, Mass.: Sinauer
Fritz, P.J., Vesell, E.S. White, E.L., Pruitt, K.M. (1969). Proc. Natl. Acad. Sci. 62, 558–565
Fritz, P.J., White, E.L., Pruitt, K.M. (1975). In: "The Isozymes," Vol. III, Markert, C.L., ed., pp. 347–358, New York: Academic Press
Galau, G.A., Klein, W.H., Davis, M.M., Wold, B.J., Britten, R.J., Davidson, E.H. (1976). Cell 7, 487–505
Gelinas, R.E., Kafatos, F.C. (1977). Dev. Biol. 55, 179–190
Gersh, I. ed. (1973). Submicroscopic Cytochemistry, Vol. I., New York: Academic Pres
Goldberg, E. (1965). Arch. Biochem. Biophys. 109, 134–141
Gooch, J.L., Schopf, T.S.M. (1972). Evolution 26, 545–552
Harris, S.E., Rosen, J.M., Means, A.R., O'Malley, B.W. (1975). Biochem. 14, 2072–2081
Hennig, B. (1975). Eur. J. Biochem. 55, 167–183
Hennig, W. (1974). In: The cell nucleus, H. Busch, ed., Vol. 2. p. 333, New York: Academic Press
Holmes, R.S. (1972). FEBS Letters 28, 51–55
Holmes, R.S., Scopes, R.K. (1974). Eur. J. Biochem. 43, 167–177
Holmquist, R. (1976). Random and nonrandom processes in the molecular evolution of higher organisms. In: Molecular Anthropology. M. Goodman, R.E. Tashian, J.H. Tashian, eds., pp. 89–116, New York: Plenum
Hinegardner, R. (1976). Evolution of genome size. In: Molecular Evolution, Ayala, F.J., ed., pp. 179–199, Sunderland, Mass: Sinauer
Johnson, L.F., Levis, R., Abelson, H.T., Green, H., Penman, S. (1976). J. Cell. Biol. 71, 933–938
Kaplan, N.O., Everse, J., Admiraal, J. (1968). Ann. New York Acad. Sci. 151, 400–412
Kauffman, S. (1975). In: Cell Patterning. Ciba Foundation Symp. 29, 201–214
Kemp, D.J. (1975). Nature 254, 573–577
Kimura, M., Ohta, T. (1974). Proc. Natl. Acad. Sci. 71, 2848–2852
Kitamura, M., Iijima, N., Hashimoto, F., Hiratsuka, A. (1971). Clin. Chim. Acta 34, 419–423
Kleene, K.C., Humphreys, T. (1977). Cell 12, 143–155
Klose, J., Wolf, U., Hitzeroth, H., Ritter, H., Atkin, N.B., Ohno, S. (1968). Humangenetik 5, 190–196
Kourilsky, P., Gros, F. (1974). In: Fogarty Intern. Center Proc. No. 25, Harris, M. and Thompson B., eds., pp. 19–41. DHEW Pub. #74–648 NIH
Kurtz, D.T., Feigelson, P. (1977). Proc. Natl. Acad. Sci. 74, 4791–4795
Lanyon, W.G., Ottolenghi, S., Williamson, R. (1975). Proc. Natl. Acad. Sci. 72, 258–262
Lebherz, H.G. (1974). Experientia 30, 655–658

Lebherz, H.G. (1975). In: "The Isozymes," Vol. III, Markert, C.L., ed., pp. 253–279, New York: Academic Press
Lewin, B. (1975) Cell 4, 77-93
Lewis, E.B. (1951). Cold Spring Harbor Symp. Quant. Biol. **16**, 159–174
Lewis, M., Helmsing, P.J., Ashburner, M. (1975). Proc. Natl. Acad. Sci. **72**, 3604–3608
Lim, S.T., Kay, R.M., Bailey, G.S. (1975). J. Biol. Chem. **250**, 1790–1800
Margoliash, E., Fitch, W.M., (1968). Ann. New. York Acad. Sci. **151**, 359–381
Markert, C.L. (1968). Ann. New York. Acad. Sci. **151**, 14–30
Markert, C.L., Shaklee, J.B., Whitt, G.S. (1975). Science **189**, 102–114
Masters, C.J., Holmes, R.S. (1972). Biol. Rev. **47**, 309–361
Mauck, J.C., Green, H. (1973). Proc. Natl. Acad. Sci. **70**, 2819–2822
Mayr, E. (1970). Populations, species, and evolution. Cambridge, Mass.: Belknap Press of Harvard University Press
McKnight, G.S., Pennquin, P., Schimke, R.T. (1975). J. Biol. Chem. **250**, 8105–8110
Metz, C.W. (1947). Am. Nat. **81**, 81–103
Musick, W.D.L., Adams, A.D., Rossmann, M.G. (1976). J. Mol. Biol. **104**, 659–668
Nadal-Ginard, B. (1978). J. Biol. Chem. **253**, 170–177
Nakano, E., Whiteley, A.H. (1965). J. Exp. Zool. **159**, 167–179
Newlon, C., Gussin, G., Lewin, B. (1975). Cell 5, 213–226
Nigon, V., Godet, J. (1976). Inter. Rev. Cytol. **46**, 79–176
Ohno, S. (1970). Evolution by gene duplication. Berlin, Heidelberg, New York: Springer
Paterson, B.M., Bishop, J.O. (1977). Cell **12**, 751–765
Petit, C., Zuckerkandl, E. (1976). Evolution, Génétique des Populations. Evolution Moléculaire. Paris: Hermann
Philipp, D.P., Whitt, G.S. (1977). Devel. Biol. **59**, 183–197
Ribbert, D. (1972). Res. Probl. Cell. Diff. **4**, 153–179
Ross, J., Gielen, J., Packman, S.I., Kawa, Y., Leder, P. (1974). J. Mol. Biol. **87**, 697–714
Rossmann, M.G., Argos, P. (1976). J. Mol. Biol. **105**, 75–95
Salthe, S.N. (1975). In: "The Isozymes," Vol. IV, Markert, C.L., ed., pp. 665–678, New York: Academic Press
Selander, R.K., Smith, M.H., Yang, S.Y., Johnson, W.E., Gentry, G.B., (1971). In: Studies in Genetics VI, Wheeler, M.R., ed., pp. 49–90. University of Texas Pub. # 7103, Austin, Texas
Senkbeil, E., White, H.B.III, (1978). J. Mol. Evol. **11**, 57–66
Sensabaugh, G.F., Kaplan, N.O. (1972). J. Biol. Chem. **247**, 585–593
Shaklee, J.B., Kepes, K.L., Whitt, G.S. (1973). J. Exp. Zool. **185**, 217–240
Shaklee, J.B., Christiansen, J.A., Sidell, B.D., Prosser, C.L., Whitt, G.S. (1977). J. Exp. Zool. **201**, 1–20
Sherman, F., Stewart, J.W. (1971). Ann. Rev. Genetics 5, 257–295
Shows, T.B., Massaro, E.J., Ruddle, F.H. (1969). Biochem. Genet. **3**, 525–536
Siegel, W., Cox, R., Schroeder, W.A., Huisman, T.H.J., Penner, O., Rowley, P.T. (1970). Ann. Internal. Med. **72**, 533–536
Sing, C.F., Brewer, G.J. (1971). Biochem. Genet. **5**, 243–251
Somero, G.N. (1975). In: "The Isozymes," Vol. II, Physiological function, Markert, S.L., ed., p.221, New York: Academic Press

Sparrow, A.H., Price, H.J., Underbrink, A.G. (1974) (ed. by H.H. Smith) Brookhaven Symp. Biol. **23**, 451–485
Stebbins, G.L. (1966). Science **152**, 1463–1469
Stephens, S.G. (1951). Adv. Genet. **4**, 247–265
Stockell, A., Perutz, M.F., Muirhed, H., Glauser, S.C. (1961) J. Mol. Biol. **3**, 112–116
Strydom, D.J. (1977). J. Mol. Evol. **9**, 349–361
Suzuki, Y., Suzuki, E. (1974). J. Mol. Biol. **88**, 393–407
Valentine, J.W., Ayala, F.J. (1975). Deep Sea Res. **22**, 37–44
Whitt, G.S. (1970). J. Exp. Zool. **175**, 1–36
Whitt, G.S. (1975). In: Vision in Fishes, Ali, M.A., ed., p. 459, New York: Plenum Press
Wilson, A.C., Carlson, S.S., White, T.J. (1977). Ann. Rev. Biochem. **46**, 573–639
Wilson, F.R., Whitt, G.S., Prosser, C.L. (1973). Comp. Biochem. Physiol. **46B**, 105–116
Ycas, M. (1976). J.Mol. Evol. **7**, 215–244
Zhimulev, I.F., Belyaeva, E.S. (1975). Chromosoma **49**, 219–231
Zinkham, W.H., Isensee, H., Renwick, J.H. (1969). Science **164**, 185–187
Zuckerkandl, E. (1960). Ann. Inst. Oceanogr. **38**, 1–122
Zuckerkandl, E. (1974). In: Ecole de Roscoff, Editions du Centre National de la Recherche Scientifique, Paris, pp. 69–79
Zuckerkandl, E. (1975). J. Mol. Evol. **7**, 1–57
Zuckerkandl, E. (1976a) In: Molecular Anthropology: Genes and Proteins in the Evolutionary Ascent of Primates, Goodman, M. and Tashian, R.E. (eds.), pp. 387–447. New York: Plenum Press
Zuckerkandl, E. (1976b) J. Mol. Evol. **9**, 73–104
Zuckerkandl, E. (1976c) J. Mol. Evol. **7**, 269–311
Zuckerkandl, E. (1976d) J. Mol. Evol. **7**, 167–183
Zuckerkandl, E. (1978). Zeitschr. Morph. Anthropol. **69**, 117–142
Zuckerkandl, E., Pauling, L. (1962). in Horizons in Biochemistry, Kasha, M., Pullman, B., eds., pp. 189–225, New York: Academic Press
Zuckerkandl, E., Pauling, L. (1965). In: Evolving Genes and Proteins. Bryson, V., Vogel, H.J., eds., pp. 97–166, New York: Academic Press

Received November 6, 1977; Revised May 10, 1978