S. Ohno

Department of Biology,
City of Hope Medical Center,
Duarte, Calif. 91010, U.S.A.

Received 28 October 1971

An Argument for the Genetic Simplicity of Man and other Mammals

There is a finite upper limit to the number of gene loci which an organism can afford to have. An organism having 3×10^6 or so gene loci would exterminate itself from an unbearable mutation load. Yet, the mammalian genome is large enough to accommodate that many gene loci.

Natural selection is essentially conservative. A new gene with a hitherto non-existent function can not be created unless a temporary escape from the relentless pressure of natural selection is provided by gene duplication. However, the chance of a redundant copy of an old gene emerging as a new gene is considerably smaller than of that copy becoming degenerate. Perhaps for this reason, evolution from simpler life forms to mammals appears to have been accompanied not only by successive additions of new gene loci, but also by accumulation in the genome of a great deal of degenerate "nonsense" DNA base sequences. At least 90% of the mammalian genomic DNA appears to represent "nonsense" DNA base sequence of various kinds.

The creation of additional regulatory systems contributed more to big leaps in evolution than did the creation of new structural genes. Yet, in order not to be burdened with an unbearable mutation load, the necessary increase in the number of regulatory systems had to be compensated by simplification of each regulatory system. It would not be surprising if each mammalian regulatory system is shown to have fewer components than the *lac*-operon system of *Escherichia coli*.

1. Introduction

Natural selection appears in two disguises. To the students of morphological traits, it appears as a great advocator of adaptive change. To the students of molecules, on the contrary, it appears as an extreme conservative which permits only trivial changes to the gene. Witness the fact that histone IV (110 amino acid residues) of cattle and garden peas differ from each other only by two amino acid substitutions (DeLange & Fambrough, 1968; McLaughlin & Dayhoff, 1969). The rate of evolution (amino acid substitution) of this protein has been incredibly slow (6×10^{-12} per amino acid site per year). It appears that an entire molecule of histone IV represents a functionally critical active site, so that almost any amino acid substitution makes a mutant protein delinquent in the performance of its assigned function. Accordingly, natural selection has permitted hardly any mutation to affect this locus since the creation of eukaryotes. In sharp contrast, fibrinopeptides A (15–19 amino acid residues) and B (14–21 residues) have undergone evolutional changes (Blomback, Blomback & Grondahl, 1965; Eck & Dayhoff, 1966) at a rate 1500 times faster than that of histone IV (Ohta & Kimura, 1971). During blood clot formation, fibrinopeptides A and B are split off from the inert fibrinogen molecule by the trypsin-like action of thrombin. As a result, fibrinogen is converted to the active fibrin. Their role being passive, natural selection has ignored many amino acid substitutions during the course of evolution as they are harmless.

The students of molecules have an advantage in that they deal with the direct products of genes. Have the students of morphological traits been under an illusionary spell with regard to the true nature of natural selection? The coat color of mammals immensely influences the survival value of individuals, for it can be used as camouflage or as a warning. Accordingly, one is dazzled by the adaptive changes of coat colors throughout evolution. Coat color changes are often due to functional deficiencies of the C-locus tyrosinase which is an enzyme concerned with synthesis of melanin pigments from the amino acid tyrosine. Hypoactive mutant tyrosinase produce dilute coat colors which characterize such phenotypes as *Chinchilla* and *Himalaya* (Searle, 1968). Such mutations are not deleterious because this tyrosinase is not involved in an important metabolic pathway. If this tyrosinase was involved not only in melanin synthesis but also in other metabolic pathways of tyrosine, such as synthesis of epinephrine, natural selection would not have permitted such mutations to persist.

It can be said that students of morphological traits are under no illusion only because outward appearances of organisms are often determined by the exceptional genes with little or no pleiotropic effects. So far as the rest of the genes are concerned, natural selection is too conservative to permit meaningful changes to occur in already existing gene loci. Accordingly, it is an escape from the relentless pressure of natural selection which is *conditio sine qua non* of big evolutional leaps. Without this escape, prokaryotes would have never evolved to eukaryotes, and fish would have forever stayed as fish. Only a redundant copy of a functional gene created by gene duplication can escape from surveillance by natural selection, and, while being ignored, it can accumulate formerly forbidden mutations to emerge as a new gene with a hitherto non-existent function. Evolution from fish to mammals has apparently been accomplished by a successive series of gene duplications (Ohno, 1970).

However, when a redundant copy of a functional gene escapes from natural selection, it has a greater chance of becoming degenerate than of emerging as a new gene. A successive series of gene duplications in the past must have produced more degenerate sequences than new genes. For this and other reasons, the genomes of higher organisms appear to contain a very large proportion of "non-informational" DNA base sequences.

Although the mammalian genome is large enough to accommodate a few million genes, the actual number of gene loci may only be one-fiftieth of that. The notion that mammals must have regulatory systems of infinite complexity to regulate the enormous number of structural genes they possess is largely an illusion.

2. From a Redundant Copy to a New Gene and the Cost of Success

The comparison of amino acid sequences revealed the common ancestry of many functionally related proteins, and, therefore, of the genes which specify them. The hemoglobin α-chain gene clearly arose from a redundant copy of the hemoglobin β-like chain gene, and the ancestry of both can be traced to the myoglobin gene (Ingram, 1963). Immunoglobulin light-chain (about 220 amino acid residues) as well as heavy-chain (about 550 residues) genes have apparently been created by repeated duplications and tandem fusions of an ancestral gene which specified a polypeptide chain made of about 110 amino acid residues (Lennox & Cohn, 1967; Black & Dixon, 1968).

On a grander scale, most of the secreted vertebrate proteins, although widely divergent in function, appear to have originated from a few digestive enzymes secreted by entoderm of earlier and simpler forms of life (Adelson, 1971). Not only are the amino acid sequences of different peptide hormones similar, but also there exists recognizable homology in the amino acid sequences even between growth hormone, on one hand, and trypsin and chymotrypsin, on the other.

In my book (Ohno, 1970), I extensively discussed the subject of when in vertebrate evolution and by what means most of the new genes which characterize *Mammalia* as a class were created. I rather ignored, however, the relative inefficiency of the process of creating a new gene from a redundant copy of an old gene.

So long as a useful function is assigned to a single gene locus in the genome, natural selection would never permit an alteration in the active site amino acid sequence, for such an alteration is deleterious as it impairs the assigned function of that gene product. As a result, the lactate dehydrogenase gene would forever remain the lactate dehydrogenase gene, although lactate dehydrogenase would go through evolutional changes by accumulating amino acid substitutions at functionally less critical sites of the molecule.

If a new dehydrogenase gene for another substrate is to be created from the lactate dehydrogenase gene, mutations which change the active site amino acid sequence should be permitted to accumulate. Yet certain other sites, such as the site which recognizes NAD as a coenzyme, should be conserved. In the absence of natural selection, what is the chance of a redundant copy of the lactate dehydrogenase gene emerging as a new dehydrogenase gene? I must say that the chance is not very good. A drastic change, such as a *frameshift* mutation due to deletion or addition of a single DNA base, would so completely alter the amino acid sequence of the gene product that it would not likely function as an enzyme of any kind.

The three consequences of a single DNA base substitution are *samesense*, *missense* and *nonsense* mutations. The *samesense* mutation, such as a change from a codon UUG to CUG, does not alter the amino acid sequence of a gene product, for both codons specify leucine. A *samesense* mutation is largely harmless as well as meaningless. A *missense* mutation, such as a change from a phenylalanine codon (UUU) to a tyrosine codon (UAU), should primarily be responsible for the creation of a new gene when it affects the active site of a molecule. A *nonsense* mutation, such as a change from a cysteine codon (UGC) to a chain-terminating codon (UGA), would yield a short peptide chain due to premature chain termination. Such a peptide chain is likely to be useless under any circumstance.

Inasmuch as three (UAA, UAG and UGA) of the 64 codons are *nonsense* and since there are 61 codons for only 20 amino acids, we can calculate that random base substitutions would produce *samesense*, *missense* and *nonsense* mutations in the ratio of 6:17:1 (Whitfield, Martin & Ames, 1966). In other words, there is a one in 24 chance that a first base substituting mutation sustained by a redundant copy of a gene would be chain terminating. One can readily see that if a redundant copy is long ignored by natural selection, its likely fate is degeneration. In order to emerge as a new gene, a redundant copy before long should begin to specify a product which is useful in a new way, so that the new locus would again be placed under the surveillance of natural selection. Only by coming under the protection of natural selection can further accumulation of mutations which would endanger its newly acquired function be prevented.

Since the number of mutations a redundant copy can sustain before becoming totally degenerate is rather limited (average of no more than 20), there should remain a strong resemblance between a new gene and the ancestral gene in most instances. The creation of a peptidase gene from a dehydrogenase gene would be rather unlikely even by gene duplication. Indeed, genes which have shared a common ancestor usually perform different, but fundamentally similar, functions; e.g. trypsin and chymotrypsin, hemoglobin and myoglobin. Even so, for every redundant copy which has succeeded in becoming a new gene, there must have been a number of others which degenerated. The probability of success of such an especially daring experiment as creating a peptide hormone gene out of a proteolytic enzyme gene (Adelson, 1971) is so small, as it has to utilize deletions by chain terminating mutations and other means, that its success must have been accompanied by a great number of failures. Just as the surface of our earth is strewn with fossils of extinct species, our genome might contain a great deal of DNA base sequences which represent failures of past experiments with gene duplications.

3. The Number of Genes and the Genome Size

The mammalian genome (haploid) contains roughly 1000 times more DNA than prokaryotes (Nei, 1969), and about a 20-fold increase in the genome size appears to have occurred during the course of evolution from the tunicate-like primitive chordate to mammals (Atkin & Ohno, 1967; Ohno, 1970). Have the number of genes increased in proportion to the increase in the genome size?

The mammalian genome is represented by about $3{\cdot}0 \times 10^{-9}$ mg of DNA which conveniently corresponds to roughly $3{\cdot}0 \times 10^{9}$ base pairs. Taking the average gene size to be 1000 base pairs long which would specify a polypeptide chain made of 330 amino acid residues, we would conclude that there is room for 3×10^{6} genes. The figure of 3×10^{6} is good in relation to the estimated number of genes for *Escherichia coli* which is 4×10^{3}. But, unfortunately, no organism having 3×10^{6} genes can survive for very long!

A gene can be defined as a stretch of DNA base sequence which specifies a product useful for an organism. Accordingly, the mutational deficiency of it would surely make an affected individual less fit. The point is that all gene loci are affected by mutations from time to time and mutations are as often deleterious as neutral or advantageous. For example, mutational deficiency of individual lysosomal enzymes cause 12 different lysosomal diseases in Man; many of them extremely severe (McKusick, 1970). Since the deleterious mutation rate per locus per organism generation is in the order of magnitude of 10^{-5}, an organism having 3×10^{6} genes would exterminate itself from an unbearable mutation load (the overall deleterious mutation rate per generation becomes 30). Because those deleterious mutations which are recessive are not readily eliminated from a population, they accumulate. Accordingly, Kimura (1968) as well as Ohta & Kimura (1970) estimated that if all the mammalian genomic DNA are functional genes, in order to maintain a stable population size, each mated pair would have to produce 10^{78} zygotes to assure that a few of them would survive.

Clearly, what Haldane (1957) termed "the cost of natural selection" imposes a finite upper limit to the number of gene loci an organism can afford to have. Muller (1948, 1967) observed that the total rate for deleterious mutations in man is at the most only 0·5 per generation. On this basis, he, as well as Crow & Kimura (1970), concluded that our genome must contain merely 4×10^{4} or so gene loci. This many gene loci account

for less than 2% of the genomic DNA if the average gene size is 1000 base pairs long.

Ohta & Kimura (1971) calculated that, if unhampered by natural selection, the mammalian DNA base sequence should be changing at the rate of $0{\cdot}33 \times 10^{-8}$ per base pair per year. As the observed rate of DNA base sequence evolution in mammals (Laird, McConaughy & McCarthy, 1969; Walker, 1968) is rather close to the calculated unhampered mutation rate, it was concluded that most of the mammalian genomic DNA being noninformational is totally ignored by natural selection. Their estimate is that only 6% of our genome represents functional gene loci. Is it possible that these calculations have grossly underestimated the number of gene loci?

Because of an extensive series of gene duplications in our past, the mammalian genome tends to contain more than one gene for each enzyme. For example, there are two separate gene loci for A_4 and B_4 isozymes of lactate dehydrogenase (Markert, 1964) and three separate gene loci specify phosphoglucomutase isozymes in Man (Harris, Hopkinson, Luffman & Rapley, 1967). On the surface, it may appear that the existence in duplicate and triplicate of genes with the same function should lower the deleterious mutation rate of each isozyme gene locus, since the deficiency of one isozyme can be covered by the presence of the other isozyme. This is true only if all the isozyme genes are expressed together all the time in every organ of the body. However, the *raison d'être* of isozymes being to provide each organ with an isozyme of a particular kinetic property to suit its needs, more often than not only one or the other of the isozyme genes is expressed in a given organ.

Inasmuch as *ribosomal* RNA's have to be made in great quantity in each cell, the vertebrate genome is endowed with hundreds of copies of the genes for *ribosomal* RNA's (Brown & Dawid, 1968). Similarly, the gene for each *transfer* RNA appears to exist in tens of copies (Ritossa, Atwood & Spiegelman, 1966*a*). A deficient mutation sustained by one of the multiple copies would certainly be ignored by natural selection. But for this very reason, deleterious mutations would accumulate slowly but surely in multiple copies.

Another difficulty is that deletions constantly afflict the chromosomal region made of tandemly arranged multiple copies. In such regions, unequal crossing-over occurs during meiosis between the homologues as well as during mitosis between two sister chromatids of the same chromosome. In the fruit fly, when a substantial number of copies of the gene for 18 s and 28 s *ribosomal* RNA are deleted, an affected individual manifests the *bobbed* phenotype characterized by markedly retarded growth (Ritossa *et al.*, 1966*b*). The fact that frequent reversion to wild-type is observed in the *bobbed* mutant stock indicates that the spontaneous rate of unequal crossing-over for this chromosomal region is alarmingly high. Thus, it appears that by having multiple copies of itself, a gene does not escape from paying its due to the cost of natural selection.

As previously mentioned, *ribosomal* and *transfer* RNA genes are exceptional in that their products have to be made in large quantities in every cell, since their services are needed for the translation of all the *messenger* RNA's present. As far as all other structural genes are concerned, there is neither the need nor the evidence for existence in multiple copies. The half-life of mammalian *messenger* RNA can become very prolonged. Hemoglobin *messenger* RNA's, for example, persist in enucleated erythrocytes for quite awhile, so that a single copy of a structural gene can suffice to produce a large amount of a polypeptide chain if need be.

All in all, it appears that the calculations made by Muller, Kimura and others are not far off the mark and that at least 90% of our genomic DNA is "junk" or "garbage" of various sorts.

4. Myriads of "Non-informational" DNA in our Genome

From the foregoing discussion, it would appear that while the genome size increased nearly 1000-fold in the transition from prokaryotes to mammals, the actual number of gene loci increased only 10- to 100-fold. Such disproportionate increase in the genome size is not surprising when one realizes that each success of creating a new gene out of a redundant copy of an old gene has invariably been accompanied by a number of failures.

The fact that chromosomes of eukaryotes contain a number of genetically empty regions has long been known. The most conspicuous blocks of so-called heterochromatin (Heitz, 1933; Kaufman, 1934) are customarily seen around the centromere (centromeric heterochromatin), the terminal end (telomeric heterochromatin) and the nucleolar organizer of a chromosome. When more refined staining techniques are employed, additional blocks of heterochromatin (intercalary heterochromatin) can be seen (Darlington & La Cour, 1940; Haga, 1969). It has recently been shown that the centromeric heterochromatin is primarily made of short base sequences strung together in long tandem repeats (Yasmineh & Yunis, 1969; Jones, 1970; Pardue & Gall, 1970). Such extremely repetitious short base sequences appear as a satellite peak separated from the bulk of DNA in density gradients (Kit, 1961; Sueoka, 1961) and reassociate very rapidly after denaturation (Waring & Britten, 1966).

Clearly, this type of tandem repeat contains no genetic information. Even if they are transcribed and translated, they would produce nonsense oligopeptide. Southern (1970) has shown that the guinea pig α-satellite DNA is made of about 10^7 copies of the basic 6 base pair sequence. As there is no natural selection against random accumulation of mutations by these tandem repeats, copies seem to be diverging at a fast rate. Because of unrestricted random drift, it is not possible to estimate the proportion of the genomic DNA represented by extremely and moderately repetitious tandem repeats (Southern, 1971). Nevertheless, it would be safe to say that more than 50% of the mammalian genome is made of such base sequences which carry no genetic information. With improved staining techniques, it is beginning to be shown that intercalary heterochromatin of mammalian chromosomes is composed of moderately repetitious tandem repeats (Sumner, Evans & Buckland, 1971).

The genes reside in the euchromatic region of chromosomes. But in this region too, genic DNA base sequences seem to be interspersed with non-informational sequences. When the nucleolar organizing regions of a salamander (*Triturus viridescens*) were isolated and examined under the electron microscope, Miller & Beatty (1969) found that tandemly arranged copies of the gene for 18 s and 28 s *ribosomal* RNA's are spaced far apart from each other by a long stretch of non-transcribable DNA. Furthermore, each unit gene copy of man specifies a precursor RNA of enormous size (48 s), and during the maturation process almost half of the precursor molecule is destroyed; only the remainder giving rise to 18 s and 28 s *ribosomal* RNA (Penman, 1967). It would appear that although the gene copies for 18 s and 28 s *ribosomal* RNA's reside exclusively in the nucleolar organizing region, at least 70% of DNA in this region does not contribute to the formation of final gene products.

Even structural genes for enzymes and other proteins appear to be spaced far apart. The very fact that most of the reciprocal translocations do not reduce the fitness of heterozygotes as well as homozygotes indicates that when a chromosome is broken at a randomly chosen position, that position usually contains no genetic information. The semisterility of translocation heterozygotes is due merely to mechanical difficulty of homologous pairing encountered during meiosis.

At least a part or parts of nongenic DNA base sequence adjacent to a structural gene appears to be transcribed. Mammalian putative *messenger* RNA's isolated from polyribosomes appear to contain an extra piece of polyriboadenylic acid sequence some 150–200 nucleotides long, although it is not clear as to whether this piece is located at the 5′- or 3′-end of *messenger* RNA (Darnell, Wall & Tushinski, 1971; Edmonds, Vaughan & Nakazato, 1971; Lee, Mendecki & Brawerman, 1971).

Hemoglobin chains are 140 or 145 amino acid residues long, thus the informational part of *messenger* RNA for these peptide chains should be made of about 450 nucleotides. However, hemoglobin *messenger* RNA at least in the beginning appears to exist in a much longer form (17 s), although they may get trimmed down later on (Maroun, Driscoll & Nardone, 1971). The recent discovery of longer than usual human hemoglobin α- and β-chains indeed indicates that human hemoglobin *messengers* normally have extra nucleotides attached to their 3′ ends. A mutational change from a chain terminating *nonsense* codon to an amino acid specifying codon appears to have added 31 extra amino acid residues to the carboxyl end in the case of the α-chain (Milner, Clegg & Weatherall, 1971) and 10 extra amino acid residues in the case of the β-chain (Flatz, Kinderlerer, Kilmartin & Helmann, 1971).

All in all, the view that information for useful gene products is encoded in only a fraction of our genomic DNA appears to enjoy overwhelming support.

5. Transcriptional vs. Translational Control

Differentiated somatic cells of our body are, in a sense, slaves, for they are forced to make certain gene products gratuitously for the good of others. While gene products for their own "household" chores are made by all the cells of the body, there is a division of labor among differentiated somatic cell types with regard to the synthesis of gratuitous gene products. Therefore, the structural genes for the gratuitous products are genetically regulated. What allows immunoglobulin molecules to be made by lymphocytes, but not by liver cells?

At first glance, the type of transcriptional control originally elucidated by Jacob & Monod (1961) on the *lac*-operon system of *E. coli* appears to provide an answer. In order to be transcriptionally regulated, the head of a structural gene should be equipped with a stretch of particular base sequence (*operator*). The recognition of this *operator* region by a regulatory gene product (repressor or activator) determines whether or not the particular structural gene is transcribed into *messenger* RNA. However, the already discussed observations which suggest that a mammalian structural gene is transcribed in continuum with adjacent "non-informational" base sequences make us realize that precise transcriptional control of individual structural genes is rather unlikely to be operative in mammals. In fact, in spite of voluminous evidence from DNA–RNA hybridization studies which suggested transcriptional control as the basis of somatic cell differentiation,

Harris (1970) has long maintained that there can be no precise transcriptional control in mammalian somatic cells.

More recently, we realized that the interpretations of DNA–RNA hybridization data are reliable only so far as genes which exist in multiple copies are concerned. We still do not know whether or not immunoglobulin genes are transcribed in liver cell nuclei. Furthermore, Harris' original observation that most of the freshly transcribed RNA in mammalian cell nuclei are rapidly degraded *in situ* has been repeatedly confirmed. It appears that Harris was right from the very beginning.

Inasmuch as transcription appears to be rather indiscriminate, mammals perhaps have only a general sort of transcriptional control such as the activation of RNA polymerases resulting in increased production of "informational" as well as "non-informational" RNAs. The precise control is likely to be at the translational level. Harris emphasized "engagement" meaning that in order to be transported to the cytoplasma to serve as *messenger* RNA, a particular RNA species in the nucleus has to be engaged by ribosomal precursors in the nucleolus. Those which are not engaged are rapidly degraded within the nucleus.

6. The Number of Regulatory Systems and the Number of Components in each System

Drastic evolutional changes in organisms' appearances are usually due to changes in regulatory systems rather than in structural genes. Man uses all five digits while the horse stands on its middle-toes. Nevertheless, a digit is a digit. The same set of structural genes are mobilized for the formation of human fingers and equine cannons. It would be safe to say that the creation of additional regulatory systems contributed more to big evolutional changes than did the creation of new structural genes. There is little doubt that evolution from unicellular prokaryotes to multicellular eukaryotes of increasing complexity was accompanied by progressive increase in the number of regulatory systems. However, for this very reason, the often stated view that compared with the *lac*-operon system of *E. coli* and other regulatory systems of prokaryotes, genetic regulatory systems in mammals must be enormously complicated is unsound.

If a great increase in the number of regulatory systems was accompanied by proportional increase in the degree of complexity of each regulatory system (increase in number of components of each system), the total mutation load imposed upon the regulatory systems would have become intolerable and would have led to the extinction of organisms.

Let us consider the mutation load placed on the *lac*-operon system of *E. coli*. (1) The regulatory locus can mutate to i^s (*repressor non-inducible*) which represents a mutational loss by a repressor of the binding affinity to an inducer (IPTG). (2) The *operator* region can mutate to o^s (*operator non-inducible*) (Smith & Sadler, 1971). The mutant *operator* acquires too much binding affinity to the wild-type repressor. (3) Of three enzymes induced, transacetylase is trivial but β-galactosidase and permease are vital for lactose metabolism. Thus, mutational incapacitation of these two structural genes would also render *E. coli* incapable of utilizing lactose. All together, the *lac*-operon system contains four specific components which can be affected by deleterious mutations. Needless to say, mutations of other gene loci such as the RNA polymerase locus, the cyclic AMP-binding protein locus (Zubay, Schwartz & Beckwith, 1970) and the enzyme loci for

subsequent metabolism of glucose and galactose also affect *E. coli*'s capacity to utilize lactose. But it may not be fair to include these gene loci with pleiotropic effect.

Is each of the mammalian regulatory systems so complex that it contains 40 or possibly even 400 components? I have argued that because there is a finite upper limit to the number of gene loci which an organism can afford to have, evolutional increase in the number of regulatory systems could not have been accompanied by increased complexity of each regulatory system. Consequently, each mammalian regulatory system must consist of as few components as or even fewer components than the regulatory system of prokaryotes (Ohno, 1971).

In mammals, the dosage compensation for most, if not all, X-linked genes is accomplished in one step by inactivation of one or the other X in the female (Lyon, 1961). It would not be surprising if a single gene locus controls this inactivation process involving hundreds of structural genes. The beauty of this mechanism is in its simplicity.

The need to maintain simplicity must have been especially great for *inducible* regulatory systems of mammals. While substrates or metabolites of substrates serve as inducers of genetic regulatory systems in prokaryotes, it is a hormone which serves as an inducer in mammals. A certain number of genes have to be set aside for the production of each inducer. Accordingly, in addition to the components of regulatory systems in target cells, the genes concerned with synthesis and release of hormones in endocrine cells contribute to the mutation load. The observation that the actions of a multitude of peptide hormones are mediated through a small variety of intracellular inducers such as cyclic AMP (Robinson, Butcher & Sutherland, 1968) supports the view that natural selection favored economy and simplicity in mammalian regulatory systems.

On the basis of our findings on the X-linked testicular feminization (*Tfm*) mutation of the mouse (Lyon & Hawkes, 1970), I have proposed that the entire response of mammalian target cells to testosterone is mediated by this single X-linked locus (Ohno, 1971). Testosterone normally induces different sets of specific enzymes in different target cells while causing hypertrophy of all the target organs. In sharp contrast, neither specific enzyme induction nor hypertrophy occurs in target cells of *Tfm*/Y mutants in response to administered testosterone. Consequently, the affected *Tfm*/Y develops as a phenotypic female despite the presence of testes. My thesis is that inside the target cells of testosterone a protein specified by the wild-type allele of the *Tfm* locus has a dual role. When it is not bound to testosterone or its intracellular metabolites, it stays in the cytoplasm and serves as a repressor of the translation of certain enzymes. Binding with testosterone changes its allosteric configuration so that the bound form detaches from these *messenger* RNA's, releasing a translational block, and moves into the nucleus where it activates RNA polymerase I inside the nucleolar region. The activation of RNA polymerase I is the cause of hypertrophy, as it results not only in increased production of ribosomes but also in increased transport of *messenger* RNA precursors from the nucleus to the cytoplasma by engagement. *Tfm* represents the mutational loss by this protein of the binding affinity to testosterone and its metabolites such as 5α-dihydrotestosterone and 5α-androstan-3α-17β-diol; hence, no specific enzyme induction and no hypertrophy in target cells.

The evidence upon which my thesis is based is as follows: (1) Specific enzyme induction is almost surely due to the removal of a translational block. 5α-androstan-3α-17β-diol which stays in the cytoplasma can substitute for testosterone as an inducer of alcohol dehydrogenase and β-glucuronidase in the wild-type mouse kidney proximal tubule cells (Ohno, Dofuku & Tettenborn, 1971) and an amino acid substituting mutation of the

β-glucuronidase locus gives it an o^s (*operator non-inducible*) character (Dofuku, Tettenborn & Ohno, 1971). In the transcriptional control, the *operator* region which occupies the head of a structural gene should not be translated. Thus, an *operator* mutation should not result in an amino acid substitution of the structural gene product. In the testosterone system, a repressor is apparently recognizing a translatable part of *messenger* RNA.

These two enzymes in the mouse kidney are under repressive rather than activating translational control, since in other cells where a product specified by the *Tfm* locus is absent, the enzymes are either made *constitutively* or not at all. For example, liver cells and fibroblasts of the mouse are not targets of testosterone. In liver cells, alcohol dehydrogenase is made *constitutively*. This suggests that in the absence of the *Tfm* protein, alcohol dehydrogenase *messenger* RNA is translated freely regardless of the presence or absence of testosterone. In fibroblasts, no alcohol dehydrogenase is made under any circumstance probably because no alcohol dehydrogenase *messenger* RNA is present.

(2) Liao & Stumpf (1968) have shown that testosterone-induced RNA synthesis occurs mainly in the nucleolar region of the target cell nuclei and that within 12 hr after the testosterone administration there occurs a threefold increase in the activity of a class of RNA polymerase which functions at a low ionic strength (Liao, Leininger, Sagher & Barton, 1965). This is the class of RNA polymerase which was later identified by Roeder & Rutter (1969) as a specific RNA polymerase for transcription of *ribosomal* RNA genes in the nucleolar organizing region of the chromosome (nuclear RNA polymerase 1).

After *in vitro* incubation of +/Y as well as +/+ mouse kidney cells, 10^8 nuclei incorporated as much as 50×10^{-15} M of tritiated 5α-dihydrotestosterone (Gehring, Tomkins & Ohno, 1971). We subsequently found that virtually all the protein-bound dihydrotestosterone concentrated in the RNA polymerase fraction extracted from these labeled nuclei. Thus, it appears probable that the so-called DHT-nuclear receptor protein complex (Bruchovsky & Wilson, 1968; Fang, Anderson & Liao, 1969) activates RNA polymerase 1 by becoming a sigma-factor-like subunit of this mammalian polymerase. We hope to get a definitive answer on this possibility shortly. In my view, the so-called nuclear receptor protein and the cytosol receptor protein are one and the same protein specified by the X-linked *Tfm* locus.

(3) The *Tfm* mutation does not seem to affect either the cytoplasmic uptake of testosterone or the subsequent intracellular metabolism of testosterone to 5α-dihydrotestosterone and androstandiols by target cells. Sixty min after the intravenous injection of 45 μc of tritiated testosterone to each effectively hepatectomized *Tfm*/Y mouse, the mutant kidney cytosol fraction contained 2000 cts/min/mg protein of testosterone, 100 cts/min of 5α-dihydrotestosterone and 400 cts/min of androstandiols. The same values were obtained on +/Y ♂ and +/+ ♀ (Bullock, Bardin & Ohno, 1971). The *Tfm* locus is neither the specific permease locus nor the 5α-steroid reductase locus.

(4) Despite the apparently normal entry of testosterone to *Tfm*/Y kidney cytoplasma, a major class of the cytosol testosterone-binding proteins was undetectable in these mutants (Gehring *et al.*, 1971). Consequently, there was little or no movement of testosterone-bound protein to the nucleus (Gehring *et al.*, 1971; Bullock *et al.*, 1971).

It is my belief that the *modus operandi* of other mammalian regulatory systems will prove to be just as simple as that postulated on the testosterone-regulon system. The mammalian regulatory system appears as an infinite complex merely because different regulatory systems which are often interlocked operate together in the same somatic cell type.

The work was supported in part by a grant (CA O5I38) from the National Cancer Institute, U.S. Public Health Service.

References

Adelson, J. W. (1971). Enterosecretory proteins. *Nature* **229,** 321–325.

Atkin, N. B. & Ohno, S. (1967). DNA values of four primitive chordates. *Chromosoma* **23,** 10–13.

Black, J. A. & Dixon, G. H. (1968). Amino-acid sequence of alpha chain of human haptoglobins. *Nature* **218,** 736–741.

Blomback, B., Blomback, M. & Grondahl, N. J. (1965). Studies on fibrinopeptides from animals. *Acta chemica scandinavica* **19,** 1789–1791.

Brown, D. D. & Dawid, I. B. (1968). Specific gene amplification in oocytes. *Science* **160,** 272–280.

Bruchovsky, N. & Wilson, J. D. (1968). The intranuclear binding of testosterone and 5α-androstan-17β-ol-3-one by rat prostate. *Journal of Biological Chemistry* **243,** 5953–5960.

Bullock, L. P., Bardin, C. W. & Ohno, S. (1971). The androgen insensitive mouse: Absence of intranuclear androgen retention in the kidney. *Biochemical and Biophysical Research Communications* **44,** 1537–1543.

Crow, F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory.* New York: Harper & Row.

Darlington, C. D. & La Cour, L. F. (1940). Nucleic acid starvation of chromosomes in *Trillium. Journal of Genetics* **40,** 183–213.

Darnell, J. E., Wall, R. & Tushinski, R. J. (1971). An adenylic acid-rich sequence in messenger RNA *of* HeLa cells and its possible relationship to reiterated sites in DNA. *Proceedings of the National Academy of Sciences, United States of America* **68,** 1321–1325.

Delange, R. L. & Fambrough, D. M. (1968). Identical COOH-terminal sequences of an arginine-rich histone from calf and pea. *Federation Proceedings. Federation of American Societies for Experimental Biology* **27,** 392.

Dofuku, R., Tettenborn, U. & Ohno, S. (1971). Testosterone-regulon in the mouse kidney. *Nature New Biology* **232,** 5–7.

Eck, R. V. & Dayhoff, M. O. (1966). *Atlas of Protein Sequence and Structure.* Silver Springs, Md: National Biomedical Research Foundation, Maryland.

Edmonds, M., Vaughan, M. H & Nakazato, H. (1971). Polyadenylic acid sequences in the heterogeneous. nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: Possible evidence for a precursor relationship. *Proceedings of the National Academy of Sciences of the United States of America* **68,** 1336–1340.

Fang, S., Anderson, K. M. & Liao, S. (1969). Receptor proteins for androgens. On the role of specific proteins in selective retention of 17β-hydroxy-5α-androstan-3-one by rat ventral prostate *in vivo* and *in vitro. Journal of Biological Chemistry* **244,** 6584–6595.

Flatz, G., Kinderlerer, J. L., Kilmartin, J. V. & Helmann, H. (1971). Haemoglobin Tak: A variant with additional residues at the end of the β-chains. *Lancet* **i,** 732–733.

Gehring, U., Tomkins, G. M. & Ohno, S. (1971). Effect of the androgen-insensitivity mutation on a cytoplasmic receptor for dihydrotestosterone. *Nature, New Biology* **232,** 106–107.

Haga, T. (1969). Structure and dynamics of natural populations of a diploid *Trillium.* In (C. D. Darlington & K. R. Lewis, Eds), *Chromosomes Today* **2,** 207–217. Edinburgh: Oliver & Boyd.

Haldane, J. B. S. (1957). The cost of natural selection. *Journal of Genetics* **55,** 511–524.

Harris, H. (1970). *Nucleus and Cytoplasm.* 2nd ed. Oxford, Clarendon Press.

Harris, H., Hopkinson, D. A., Luffman, J. E. & Rapley, S. (1967). Electrophoretic variation in erythrocyte enzymes. In (E. Beutler, Ed.), *Hereditary Disorders of Erythrocyte Metabolism,* pp. 1–20. City of Hope Symposium Series, Vol. 1. New York: Grune & Stratton.

Heitz, E. (1933). Die somatische Heteropyknose bei *Drosophila melanogaster* und ihre genetische Bedeutung. *Zeitschrift fur Zellforschung und mikroskopische Anatomie* **20,** 237–287.

Ingram, V. M. (1963). *The Hemoglobins in Genetics and Evolution.* New York: Columbia University Press.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanism in the synthesis of proteins. *Journal of Molecular Biology* **3,** 318–356.

Jones, K. W. (1970). Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature* **225,** 912–919.

Kaufmann, B. P. (1934). Somatic mitoses of *Drosophila melanogaster. Journal of Morphology* **56,** 125–156.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217,** 624–626.

Kit, S. (1961). Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *Journal of Molecular Biology* **3,** 711–716.

Laird, C. D., McConaughy, B. L. & McCarthy, B. J. (1969). Rate of fixation of nucleotide substitutions in evolution. *Nature* **224,** 149–154.

Lee, S. Y., Mendecki, J. & Brawerman, G. (1971). A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells. *Proceedings of the National Academy of Sciences, United States of America* **68,** 1331.

Lennox, E. & Cohn, M. (1967). Immunoglobulin genetics. *Annual Review of Biochemistry* **36,** 365–406.
Liao, S., Leininger, K. R., Sagher, D. & Barton, R. W. (1965). Rapid effect of testosterone on ribonucleic acid polymerase activity of rat ventral prostate. *Endocrinology* **77,** 763–765.
Liao, S. & Stumpf, W. E. (1968). Autoradiographic evidence for the selective enhancement of nucleolar ribonucleic acid synthesis in prostatic nuclei by testosterone. *Endocrinology* **83,** 629–632.
Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190,** 372–373.
Lyon, M. F. & Hawkes, S. G. (1970). An X-linked gene for testicular feminization in the mouse. *Nature* **227,** 1217–1219.
Markert, C. L. (1964). Cellular differentiation—an expression of differential gene function. In *Congenital Malformations,* pp. 163–174. New York: The International Medical Congress.
Maroun, L. E., Driscoll, B. F. & Nardone, R. M. (1971). Possible cytoplasmic precursor of haemoglobin messenger RNA. *Nature, New Biology* **231,** 270–271.
McLaughlin, P. J. & Dayhoff, M. O. (1969). *Atlas of Protein Sequence and Structure.* National Biomedical Research Foundation, Silver Springs, Maryland.
McKusick, V. A. (1970). Human genetics. *Annual Review of Genetics* **4,** 1–46.
Miller, O. L. & Beatty, B. R. (1969). Visualization of nucleolar genes. *Science* **164,** 955–957.
Milner, P. F., Clegg, J. B. & Weatherall, D. J. (1971). Haemoglobin-H disease due to a unique haemoglobin variant with an elongated α-chain. *Lancet i,* 729–732.
Muller, H. J. (1948). Evidence of the precision of genetic adaptation. *Harvey Lectures Series* **43,** 165–229.
Muller, H. J. (1967). The gene material as the initiator and the organizing basis of life. In (R. A. Brink, Ed.), *Heritage from Mendel,* pp. 419–447. Madison: The University of Wisconsin Press.
Nei, M. (1969). Gene duplication and nucleotide substitution in evolution. *Nature* **221,** 40–42.
Ohno, S. (1970). *Evolution by Gene Duplication.* Heidelberg: Springer-Verlag.
Ohno, S. (1971). An argument or the simplicity of mammalian regulatory systems: Single gene determination of male and female phenotypes. *Nature,* **234,** 134–137.
Ohno, S., Dofuku, R. & Tettenborn, U. (1971). More about X-linked testicular feminization of the mouse as a noninducible (i^s) mutation of a regulatory locus: 5α-androstan-3α-17β-diol as the true inducer of kidney alcohol dehydrogenase and β-glucuronidase. *Clinical Genetics* **2,** 1–13.
Ohta, T. & Kimura, M. (1971). Functional organization of genetic material as a product of molecular evolution. *Nature* **233,** 118–119.
Pardue, M. L. & Gall, J. G. (1970). Chromosomal localization of mouse satellite DNA. *Science* **168,** 1356–1358.
Penman, S. (1967). Ribonucleic acid metabolism in mammalian cells. *New England Journal of Medicine* **276,** 502–511.
Ritossa, F. M., Atwood, K. C. & Spiegelman, S. (1966*a*). On the redundancy of DNA complementary to amino acid transfer RNA and its absence from the nucleolar organizing region of *Drosophila melanogaster. Genetics* **54,** 663–676.
Ritossa, F. M., Atwood, K. M. & Spiegelman, S. (1966*b*). A molecular explanation of the bobbed mutants of *Drosophila* as partial deficiencies of "ribosomal" DNA. *Genetics* **54,** 819–834.
Robinson, G. A., Butcher, R. W. & Sutherland, E. W. (1968). *Cyclic AMP. Annual Review of Biochemistry* **37,** 149–174.
Roeder, R. G. & Rutter, W. J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* **224,** 234–237.
Searle, A. G. (1968). *Comparative Genetics of Coat Colour in Mammals.* London: Logos Press.
Smith, T. F. & Sadler, J. R. (1971). The nature of lactose operator constitutive mutations. *Journal of Molecular Biology* **59,** 273–305.
Southern, E. M. (1970). Base sequence and evolution of guinea-pig α-satellite DNA. *Nature* **227,** 794–798.
Southern, E. M. (1971). Effects of sequence divergence on the reassociation properties of repetitive DNAs. *Nature, New Biology* **232,** 82–83.
Sueoka, N. (1961). Variation and heterogeneity of base composition of deoxyribonucleic acids: A compilation of old and new data. *Journal of Molecular Biology* **3,** 31–40.
Sumner, A. T., Evans, H. J. & Buckland, R. A. (1971). New techniques for distinguishing between human chromosomes. *Nature, New Biology* **232,** 31–32.
Walker, P. M. B. (1968). How different are the DNA's from related animals? *Nature* **219,** 228–232.
Waring, M. & Britten, R. J. (1966). Nucleotide sequence repetition: A rapidly reassociating fraction of mouse DNA. *Science* **154,** 791–794.
Whitfield, H. J., Martin, R. G. & Ames, B. N. (1966). Classification of aminotransferase (C gene) mutants in the histidine operon. *Journal of Molecular Biology* **21,** 335–355.
Yasmineh, W. G. & Yunis, J. J. (1969). Satellite DNA in mouse autosomal heterochromatin. *Biochemical and Biophysical Research Communications* **35,** 779–782.
Zubay, G., Schwartz, D. & Beckwith, J. (1970). Mechanism of activation of catabolite-sensitive genes: A positive control system. *Proceedings of the National Academy of Sciences of the United States of America* **66,** 104–110.