

Elements for a theory of molecular evolution[☆]

Werner Arber*

Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received 12 August 2002; received in revised form 15 February 2003; accepted 28 April 2003

Abstract

Biological evolution is known to be driven by the availability of genetic variants. Spontaneous genetic variation can be the result of a number of specific molecular mechanisms. These can be grouped into three qualitatively different natural strategies of generating genetic variations, namely local sequence changes, DNA rearrangement within the genome and horizontal gene transfer, which is referred to here as DNA acquisition. All of these strategies bring about alterations in the DNA sequences of the genome, thus corresponding to the molecular genetic definition of the term mutation. A detailed inspection of specific mechanisms of mutagenesis reveals on the one hand the impact of non-genetic internal and environmental factors, and on the other hand the specific involvement of gene products. The underlying so-called evolution genes can be classified into generators of genetic variations and into modulators of the frequency of genetic variation. These evolution genes are postulated to have themselves undergone biological evolution under the pressure of second-order selection. On the basis of a few selected examples of mutagenesis, elements for a theory of molecular evolution are collected without a claim for completeness. Philosophical dimensions as well as practical aspects of the advanced knowledge on specific molecular mechanisms involved in molecular evolution are also briefly discussed.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Deletion; DNA inversion; DNA rearrangement; Evolutionary theory; Evolution gene; Frequency modulator; Horizontal gene transfer; Insertion; Molecular evolution; Mutation; Natural selection; Nucleotide substitution; Recombination; Second-order selection; Transposition; Variation generator

1. Introduction

A comprehensive view of molecular evolution emerges from the interphase between evolutionary biology, genetics and nucleic acid biochemistry. These three fields of scientific investigations have their independent roots in a relatively short timespan after the middle of the 19th century. In 1859, Charles Darwin published the theory of natural selection. Its underlying concept helped to realize the role played by biological evolution in the building up of the present biodiversity and it enlightened the historical relatedness between different kinds of organisms. By describing the inheritance of phenotypic traits to progeny, Gregor Mendel set in 1866 the basis for classical genetics, which became widely recognized in the early decades of the 20th century. This development made it increasingly clear that

genetics and biological evolution are tightly interrelated. By about 1940, modern evolutionary synthesis brought the logical link between these two fields of science which had taken their independent origins. At that time, the chemical basis of inheritance was still unknown.

Nucleic acid biochemistry has its root in work accomplished in 1874 by Friedrich Miescher. But it is only in the 1940s that investigations made with microorganisms should reveal that deoxyribonucleic acid (DNA) is the carrier of genetic information of cellular organisms. Shortly thereafter, James Watson and Francis Crick discovered the double-helical structure of DNA molecules. Rapid advances in microbial genetics together with the knowledge on the biochemical basis of heredity then led to the development of molecular genetics including genetic engineering, DNA sequence analysis and eventually functional genomics. This allows one to study the molecular mechanisms of life processes. Life processes not only concern the functions carried out to the benefit of individual lives, but they may also make contributions to biological evolution. Therefore, the time has come to undertake another synthesis linking molecular genetics with the theory of biological evolution

[☆] Written version of the opening lecture given in June 2002 at the ISME/SMBE meeting on “Molecular Evolution” in Sorrento, Italy.

* Tel.: +41-61-267-21-30; fax: +41-61-267-21-18.

E-mail address: Werner.Arber@unibas.ch (W. Arber).

and leading to a theory of molecular evolution. It is the aim of this publication to discuss some elements for such a theory.

2. Molecular processes involved in biological evolution

Biological evolution resides on three pillars: (1) genetic variation, (2) natural selection, and (3) reproductive and geographic isolation.

Genetic variation, or the occasional spontaneous generation of a sequence alteration (mutation) in the genome of an organism, is a prerequisite for biological evolution to occur. Genetic variation thus drives biological evolution. Behind each mutational event, there is a molecular mechanism by which the DNA sequence alteration takes place. These events are molecular processes that can be studied individually with available research strategies. The results of such investigations reveal that several quite different molecular mechanisms are independently at work in spontaneous mutagenesis. Thus, each particular genetic variant has its own specific molecular origin which may differ from the one leading to the next genetic variation. These aspects will be given a closer look below.

Natural selection is exerted on mixed populations composed of organisms with the parental genetic makeup and of derivatives thereof carrying different genetic variations. Each of these variants will have its own characteristic ability to deal with the encountered living conditions, so that in a growing population, the fittest organisms will finally take over, while others may eventually become eliminated. Obviously, natural selection is exerted on expressed functions, not on bare sequence alterations in the genomic DNA. Natural selection, together with the variant phenotypes that are available in mixed populations, determine the directions taken by the evolution of the organisms in question.

Isolation can be seen to modulate the evolutionary process, in the case of reproductive isolation by reducing the fertility between distantly related organisms. This is largely due to functional incompatibilities between the genomes of the involved partners.

3. Strategies to investigate processes of molecular evolution

An increasingly feasible approach to obtain information on evolutionary relatedness between different species, or also on the genetic polymorphism between members of the same species, is based on DNA sequence comparisons. Depending on the available sequence data and bioinformatic tools, such research can concern a single sequence domain, a given gene, a group of genes or an entire genome. In the latter cases, such studies can not only reveal the degree of sequence homology, but also identities and differences in the genomic organization of the involved

organisms. The data can help to postulate which kind of specific mechanisms could have brought about the observed genetic differences.

A more direct approach is to look at single events of genetic variation. Since these events occur rarely and usually randomly in space and in time they are not easy to observe. Small microbial genomes are the best objects for such studies, in which DNA sequences before the mutagenesis are compared with the homologous sequences after the event. As already mentioned, such studies point to a multitude of different specific molecular mechanisms contributing to the overall generation of genetic variants.

It is for a number of reasons that bacteria are convenient objects for studies on molecular evolution. Bacteria are haploid, unicellular microorganisms and propagate by cell division. Under good growth conditions, the generation time can be in the order of 30 min. Thus, in relatively short time periods, large progeny populations can be obtained from a single cell. This renders population genetic studies easy. Upon exponential growth and also in the stationary phase, single cells occasionally undergo a spontaneous genetic variation. In genetically well studied strains of *Escherichia coli* this mutagenesis affects between 0.1% and 1% of cells per generation. In view of the haploidy, the phenotypic manifestation of a new mutation is usually fast. Depending on the affected trait, this can help to detect products of single mutation events, which then can be submitted to comparative molecular analysis as mentioned.

4. Molecular genetic definition and generally observed effects of mutations

In classical genetics, a mutation is identified by a phenotypic trait which differs from that of the parental form and which becomes transmitted to the progeny. When DNA was identified as the carrier of genetic information, mutations were seen to be related to alterations in the DNA sequences. However, it rapidly became obvious that by far, not all spontaneous sequence alterations in the genome must lead to an altered phenotype.

In molecular genetics, and particularly in functional genomics, it has become common practice to define any DNA sequence alteration as a mutation. For example, a fruitful strategy to search for the functional relevance of a specific DNA sequence involves a designed alteration (site-directed mutagenesis) of the sequence followed by the experimental investigation of any phenotypic consequences of the sequence alteration after its reintroduction into the organism. In our description of molecular mechanisms generating genetic variations, we will follow this molecular genetic definition, since these mechanisms are likely to involve without discrimination any DNA sequences independent of the immediate functional relevance of the mutagenic change.

It is a common observation of geneticists working with either microorganisms or higher organisms that spontaneous DNA sequence alterations lead only rarely to an increase in fitness, i.e. turn out to be useful or beneficial for the individual in question, and by this, a contribution to the evolutionary progress. Most attempts have failed to identify evolutionary adaptation as a phenomenon by which changes in the environment would specifically induce useful mutations in appropriate genes. This kind of purpose-oriented mutagenesis could certainly not be a general natural mechanism. However, in rare exceptional cases one may expect to find specific causal explanations for such effects. Nevertheless, spontaneous mutations must occur largely at random. It is not the mutagenesis per se which determines the direction of evolution, but—as was already said—natural selection together with the chance availability of beneficial genetic variants. Indeed, many investigations reveal that such variants providing selective advantage are relatively rare and that by far more spontaneous sequence changes give a selective disadvantage to an organism and often lethality. As already outlined, many other sequence changes remain without immediate phenotypic effect, they are neutral or silent.

5. Molecular mechanisms of the generation of genetic variations follow three qualitatively different strategies

If we adopt the definition given by classical genetics to the term mutation, i.e. causing an alteration of a phenotypic trait, we know that such mutations can occur either spontaneously, without any deliberate interference of the investigator, or by induction such as by the addition of a chemical mutagen or by irradiation (a physical mutagen). Much knowledge on the mutagenic effects by such interferences has been accumulated. Very often, the results of induced mutagenesis are small local changes in the given DNA sequences such as the substitution of a single nucleotide, the deletion or additional insertion of one or a few nucleotides or else a scrambling of a few nucleotides. This knowledge became available almost 50 years ago. It was then taken more or less for granted that the same mechanisms are also responsible for the spontaneous mutagenesis. After all, some chemical and physical mutagens are omnipresent in nature. At about the same time, two additional sources for spontaneous mutagenesis were identified to relate to a certain degree of chemical instability and of structural flexibility of nucleotides. For example, cytosine can undergo oxidative deamination to become uracil. This later gives rise to an altered base pairing and results in nucleotide substitution. On the other hand, nucleotides can not only be present in their normal, most stable structural form, but they can, for short periods of time, assume different, so-called tautomeric forms. These again result in mispairings in the double-helical DNA and can often end up to give rise to a nucleotide substitution. In the literature, all

of these mutagenic events are usually described as errors and accidents. On the basis of this knowledge, textbooks often generally define spontaneous mutations as small local changes in the genomic sequences. As we know today, these kinds of genetic alterations represent an important contribution to spontaneous mutagenesis, but they, by no means, comprise all mutagenic events.

Some high-energy radiations can give rise to the breakage of DNA molecules. Such damage can be repaired by specific enzyme systems (repair enzymes) which involve in this case recombinational events, for example between sister DNA molecules. This leads us to discuss a second general strategy to generate genetic diversity, by DNA rearrangement or recombinational reshuffling. It has long been recognized that general recombination between homologous segments of DNA is a common source for increased genetic diversity in sexual, diploid organisms. If we recall the molecular genetic definition of the term mutation, i.e. an event giving rise to an alteration of genomic sequences, we can with confidence consider all different events of recombinational reshuffling as a distinct natural strategy of mutagenesis. Recombinational events often occur spontaneously, without the intervention of an investigator. The example of high-energy irradiation given above may be considered as an exception to this statement if the study is based on experimental irradiation.

It is well established in microbial genetics that intragenomic homologous recombination occasionally also occurs between more or less long DNA segments of sequence homology. For example, specific insertion sequence (IS) elements can be present in several copies in a bacterial genome, at different locations. Therefore, general recombination affecting two differently located copies of a given IS element could cause deletion or DNA inversion, as well as a partial genome duplication if the recombination occurs between sister molecules.

IS elements are classical representatives of mobile genetic elements. These are omnipresent not only in microorganisms but also in the genomes of higher organisms. Besides providing portable sequence homologies for general recombination within the genome as already mentioned, mobile genetic elements can actively undergo transposition, which is often accompanied by further DNA rearrangement processes such as deletion formation and DNA inversion. Transpositional reshuffling is thus an important source of spontaneous mutagenesis. It can often be lethal, e.g. by an insertion occurring into an essential chromosomal gene or by the deletion of such a gene. Each distinct mobile genetic element has its own characteristic target specificity determining where insertion preferentially takes place. For example, some IS elements highly prefer a specific, short DNA sequence for insertion, but they may occasionally also use other sequences for that purpose. Other IS elements rather display a regional preference for insertion, upon which different actual integration sites within this DNA region may show no detectable homology among each other

(Sengstag and Arber, 1983). Remember that a relaxed target selection provides an increased chance for evolutionary innovation in the long term.

Similar statements can apply to another important class of enzymatically mediated recombination, the so-called site-specific recombination. Intensive studies were carried out with such bacterial systems. Here again, site-specificity determining the site of recombination on the DNA should not be taken strictly. As we will discuss later, exceptions from the rule can here again allow for evolutionary innovations.

We have so far seen that a number of different specific mechanisms contribute in parallel to the generation of small local changes in the DNA sequences and to an occasional reshuffling of segments of the genome. All the mechanisms belonging to these two strategies to produce genetic variations use the organism's genome as the substrate for the genetic alteration. In nature, we encounter still another, third general strategy to increase genetic diversity: this is by horizontal, or lateral, gene transfer. We call this strategy DNA acquisition. This is very well documented for bacteria and is receiving increasing evidence to also apply to higher organisms. In well-studied bacterial systems, horizontal gene transfer can occur (a) by either active or passive uptake of DNA fragments (having been liberated by other organisms) from the environment (transformation), (b) upon direct contact between the donor and recipient cells (conjugation), or (c) mediated by a viral particle serving as gene vector (transduction).

If one critically evaluates the contributions made by each of the three described general strategies to generate genomic sequence alterations, one can discern different qualities characteristic for each of the strategies.

The strategy of small local sequence change can bring about a stepwise improvement of available biological functions. In the long term, this strategy could also be a source for a new biological activity, but this is expected to come to bear only once a specific gene product gets expressed and starts to represent a substrate for natural selection. Therefore, the improvement process must be by far more efficient than the innovation function. Let us recall that the local sequence change provides to the investigator a basis to measure evolutionary distances between different organisms (molecular clock). Although local sequence change is only one of three general strategies involved in spontaneous mutagenesis, the molecular clock remains a very valid tool in evolutionary investigations involving phylogenies.

DNA rearrangements within the genome can be seen as a playing around, or tinkering, with available sequence elements. Nature appears to do this unreflectedly, without a specific goal. However, the general goal is to produce once in a while, mostly by chance, a novel combination such as the fusion of two different functional domains or motifs. This can occasionally give rise to a new biological function or a different expression characteristic of a preexisting gene. In general, one can see DNA reshuffling as a

source for the improvement of the capacities of available elements.

DNA acquisition is the uptake of genetic functions developed by another kind of organism. It thus represents a sharing in the success of evolutionary developments made elsewhere. DNA acquisition usually occurs in small steps, which by the way increases its success, since a small genetic alteration goes along with an only small risk of disturbing the functional harmony of the recipient cell (action of natural selection). By "small steps" we understand the acquisition of a sequence domain, a gene or a group of a few genes, which represent a very small part of the produced hybrid genome. One can imagine that, rather exceptionally, acquisition can involve larger segments of DNA without seriously affecting the functional capacities of the resulting hybrid. One may assume that both the strategy of DNA rearrangement and that of DNA acquisition might explain rare evolutionary events of sudden appearance (emergence) of novel properties.

6. The thesis of the existence of evolution genes and their second-order selection

Upon a critical look, most of the molecular processes leading to spontaneous and induced mutagenesis depend on the action of particular enzymes, which are of course the products of genes carried in the genome or on accessory genetic elements of the organism. Can one therefore classify at least some of these genes as evolution genes having as a primary task to help biological evolution to proceed at the level of the population? This attitude of interpreting the available data has been defended before (Arber, 1993, 1995, 1999, 2000, 2002a). Evolution genes were thereby grouped into two classes: (a) generators of genetic variations and (b) modulators of the frequency of genetic variation by limiting genetic plasticity to tolerable, but evolutionarily useful levels. Transposable IS elements can be taken as representing the first class and DNA repair systems the second class of evolution genes. The reality appears to be a bit more complex. Indeed, specific enzyme activities sometimes contribute both to the generation of genetic variations and to the modulation of the frequency of genetic variation (e.g. restriction enzymes reducing the frequency of DNA acquisition and rendering the produced DNA fragments recombinogenic). Furthermore, while certain genes classifying as evolution genes can be deleted without any detectable loss of fitness of the concerned bacteria (at least under normal laboratory growth conditions), other genes clearly carry out functions serving both for the biological evolution and for the lives of each individual (e.g. DNA ligase and DNA topoisomerase). Another open question is whether a transducing virus can be seen as an element in the service of biological evolution (DNA acquisition strategy). From these considerations, we see that a strict classification of genes into those taking care

of the needs of each individual and those helping biological evolution to proceed would be too schematic. Rather, one can assume that a gene product is maintained under the pressure of natural selection as long as it carries out a particular, useful function or a set of such functions and does not seriously interfere with other functions.

This discussion raises the questions of whether evolutionary functions themselves also undergo evolution and if so, how selection is exerted. We assume that selection for evolution genes occurs at the level of populations and that the actual evolutionary activities are thereby the substrate for selection. Most evolutionary biologists would suppose that evolution genes are present because of favorable mutations they produced in individuals, whose greater fitness also lead to the fixation of the evolution genes through linkage. These assumptions postulate that a given bacterial strain found today in the biosphere owes its functional capacities largely to the action of evolution genes that had been present in its genome for a long time. The products of these genes would ensure on the one hand the generation of different genetic variants at the level of populations, and on the other hand a certain genetic stability favoring a long-term maintenance of that strain in the biosphere. These principles would exert an evolutionary fine-tuning for the different evolution functions ensuring a balanced life of the strain as such and of a majority of its individual cells. This kind of selection for evolution genes has been called second-order selection (Weber, 1996). Second-order selection is consistent with what had been said above.

In the context of the actual functional complexity of the bacterial cell and of populations of cells one could expect that a majority of the genes, such as housekeeping genes, would principally provide functions for the benefit of each individual cell and that a minority of genes would primarily serve the evolutionary progress. This latter group can be seen as formed by typical evolution genes. However, as noted, they are present because they benefit the individual too, although a small number of individuals. An additional number of genes might yield products helping both purposes, the accomplishment of individual lives and the evolution of the population. Such genes serving for more than one purpose might very well be submitted at once to direct selection and to second-order selection, which would bring about a coordinated fine-tuning for their different activities.

7. Site-specific DNA inversion exemplifying a generator of genetic variations by DNA rearrangement

Systems of site-specific recombination are quite widespread in bacteria and their viruses, the bacteriophages. One such system is known to periodically invert a segment of DNA flanked on each side by a 26-bp-long specific sequence carried in inverted orientations and serving as sites

for crossing over (Glasgow et al., 1989). The frequency of this DNA inversion is relatively high, if the flanking sequences conform well with a consensus sequence as determined from a few highly homologous specific sequences known to serve the purpose. Upon inversion, the two flanking crossing over sites become cut in their middle and differently religated. Since the crossing over sites have an imperfect dyad symmetry, their sequence structure is highly similar before and after the inversion. Therefore, DNA inversion can again occur, so that the original situation becomes restituted, and the process can occur again and again. This is known as a flip–flop system. In the bacteriophage P1 genome, such a flip–flop system affects a tail fiber gene (Iida, 1984). This gene consists of a constant domain (situated outside of the invertible DNA segment) and of a variable domain. For the latter, the invertible segment carries two distinct versions each of which is located near one of the ends of the invertible segment. Upon DNA inversion, the constant domain becomes separated from its variable partner and it is then fused to the other variable domain. As a consequence, preparations of the bacteriophage P1 contain a population of particles with either one of two different host ranges. So far, this has little to do with biological evolution. However, the capacity of a DNA inversion system to act as a generator of genetic variations can be shown in an experimental setup as follows.

A small bacterial plasmid is constructed in such a way that it carries the genetic information for a DNA inversion system, except that only one instead of two crossing over sites are contained on the circular DNA molecule. One can then expect that DNA inversion does not occur. In order to investigate this question, an antibiotic resistance gene without its usual expression control element is incorporated into the plasmid. The bacteria carrying this plasmid are thus not resistant to the antibiotic. The plasmid also carries two transcription promoters but these are oriented in the sense opposite to the reading frame for the antibiotic resistance. Under these conditions, very rare mutations providing antibiotic resistance can be easily detected upon spreading the bacterial culture on solid medium containing the antibiotic on which these mutants can form colonies. Most of these rare colonies reveal that their plasmid has undergone an event of DNA inversion, which has involved the single consensus sequence for crossing over and another site on the plasmid (Iida and Hiestand-Nauer, 1987). In this process, one of the promoters for gene expression has become placed in the correct orientation in front of the reading frame of the antibiotic resistance gene. In the particular experiment described here, a total of 22 independent antibiotic resistant derivatives were studied and 10 different so-called secondary sites of crossing over were found to have been used, 5 of them once, 3 twice, 1 five times and 1 six times (Arber, 1991). This distribution points to a statistical reproducibility of DNA inversion involving secondary sites of crossing over. Each specific site may have its own, very low probability to serve in inversion. Interestingly, there is no

obvious and prevailing common feature between the nucleotide sequences of the 10 analyzed secondary sites of crossing over. Identity of nucleotide positions carried in these sites with the corresponding positions in the consensus sequence varies between 50% and 25%, but the degree of homology does not strictly reflect the chance to become used.

A system like the one described here bringing about at very low frequencies one of a rather large number of possible, different, rearranged DNA structures is precisely what best characterizes a generator of genetic variations. Depending on the secondary sites of crossing over involved, such DNA inversion may occasionally result in the fusion of two hitherto unlinked functional domains. Reactions like this might explain how particular functional domains and specific motifs can become parts of otherwise unrelated genes. Similarly, the process can also bring a given reading frame under the expression control by a different promoter signal.

On a purely hypothetical basis, we postulate that recombination involving secondary sites of crossing over may depend on short-living structural variants of the involved recombination enzymes or the interacting DNA segments or both. This would be in line with the very low frequency of these events and with the statistical reproducibility of specific events.

8. Tautomeric forms of nucleotides contribute to the formation of small local sequence changes

The discussion of the possible evolutionary impact of short-living structural variants leads us to briefly consider the widely accepted interference of tautomeric forms of nucleotides with the fidelity of DNA replication (Goodman et al., 1993). This is a general source for substitution mutations. It is a matter of taste to consider these as errors of the replication process or alternatively to think that nature takes advantage of tautomeric variants to generate nucleotide substitutions of evolutionary relevance. Anyhow, since the genomic frequency of such substitution mutants obviously increases with increasing genome size, it must soon have become a necessity for organisms to develop DNA repair systems able to keep substitution mutation at low levels in order to ensure a certain degree of genetic stability. This is a good example for the close interplay of evolution gene products (in this case for DNA repair) with non-genetic factors such as the intrinsic structural flexibility of the nucleotides.

9. Genetic variations are coordinately generated by the impact of non-genetic factors and by the action of products of evolution genes

After our description of selected examples for the generation of genetic variations we can attempt to draw a

general picture of these processes. In the development of the living world, nature must have taken advantage of a number of intrinsic, non-genetic factors with mutagenic impacts. Such factors can be internal to the living organism and relate to the structural flexibility of biologically active molecules, to a certain degree of chemical instability of nucleotides or to ubiquitous mutagens present in a cell. Alternatively, external, environmental factors can also cause mutagenesis, e.g. chemical and physical mutagens. Another source of genetic variation in a given organism is its possible random encounter with a natural gene vector providing horizontal gene transfer. In many processes where non-genetic elements are involved in mutagenesis, products of genes are also engaged, in order to render a reaction bearable or to limit its efficiency and thus to ensure a required genetic stability. In summary, the spontaneous generation of genetic variants can be seen as the result of close interactions between non-genetic elements and the products of genes. Thereby, overall mutagenesis is the result of a number of different specific reactions with often different qualities with regard to their evolutionary contributions.

10. Rates of the generation of genetic variations

As we have already alluded to, the overall rate of spontaneous mutagenesis is the sum of the contributions made by all the different, specific molecular processes contributing to the generation of DNA sequence alterations. In general, such alterations indiscriminately affect functionally relevant gene sequences as well as intergenic sequences of no obvious biological relevance. Physically speaking, any genomic segment can be a possible target for spontaneous sequence alteration. However, mutagenesis appears not to occur fully at random in the DNA. Some sites and some specific regions are preferred targets for some of the mechanisms mediating the formation of sequence variations.

Experimental data are still relatively scarce on the rates by which each of the described three principal strategies and each of the mechanistically specific processes belonging to a given strategy of generation of genetic variations contribute to overall mutagenesis. These individual rates can be expected to vary from one type of organism to another as well as to depend on the particular physiological (internal) and the encountered living (external, environmental) conditions. One may also expect that best evolutionary fitness is provided to a population if its organisms are apt to carry out in parallel each of the three natural strategies to generate genetic variations, ideally by the presence of more than one enzyme system for each strategy. This situation seems to be reached in some of the genetically well studied strains of bacteria.

The primary causes for the generation of local sequence changes are mostly non-genetic factors such as the intrinsic

structural flexibility and the chemical instability of nucleotides, as we have already discussed. These factors would then in principle also influence the frequency of the involved kinds of mutagenesis. However, nature has developed enzymatic repair systems which are likely to have been fine-tuned in their activities to provide a sufficient genetic stability for the longer-term maintenance of a species as well as an evolutionarily useful frequency of mutagenesis. One can assume that genetic repair systems had been developed at early times in the evolutionary development of prokaryotes, and that such systems also served at later times to prevent too frequent somatic mutations in multicellular organisms.

In the strategy of DNA rearrangement to provide genetic changes, enzymes are generally involved. In many cases, the availability of recombination enzymes is under a tight direct control of the expression of the required enzymes. A good example is the poor availability of transposase for the activity of some mobile genetic elements. In the case of SOS repair in bacteria, it is DNA damage that induces a temporal intensification of the production of enzyme for homologous recombination. Alternatively, as in the case of DNA inversion enzymes described above to occasionally interact with secondary sites of crossing over, the rate-limiting factor may be the low affinity of the involved partner molecules to interact, possibly guided by a scarce availability of rare structural conformations.

DNA acquisition is a relatively rare, but often highly efficient strategy to bring about useful changes in the genome. This has, for example, been widely documented in studies of the spreading of antibiotic resistance genes under novel selective conditions. Several natural barriers are known to seriously limit the rates of DNA acquisition. Bacterial cell membranes act as such barriers as well as enzymatic restriction–modification systems. By their action, DNA acquisition is in general relatively rare and occurs in small steps, i.e. a relatively short segment of DNA becomes part of the recipient genome. As was already discussed, this small step principle increases the chance of evolutionary success of the process.

The latter statement has to do with the fact that, as a general rule, genetic variation does not have an a priori specific finality. Rather, it can be seen as a trial and error process. In other words, any novel genetic variant has to prove the functional compatibility between all of its genetic components as well as the success of its capacities to compete with the other organisms present. These aspects can be seen as an influence of natural selection.

Another interesting aspect influencing the rate of mutagenesis is related to the observation that genetically unstable subpopulations sometimes show up in large populations. We will briefly discuss here two examples. The first of these is related to DNA repair systems known to reduce the frequency of spontaneous formation of local sequence changes. If one of the enzymes of a repair system is affected in its activity by a mutation in its gene, the spontaneous mutation

frequency may become considerably higher and in other cases lower. Such genetic variants have been called mutators and antimutators, respectively, since phenotypically they give rise to altered rates of mutagenesis. However, the term mutator may lead the uninformed observer to a misinterpretation, if the word mutator is thought to stand for something producing mutations actively. The truth is the contrary. As a matter of fact, the task of the concerned enzyme is to inhibit the formation of mutations as a response to premutagenic damage. The genetic variation rendering the repair enzyme deficient in its inhibitory function may allow mutagenesis to occur more frequently, while other genetic variations of the gene may improve its mutation-avoidance functions (Kunkel and Bebenek, 2000).

The second example for a high genetic instability relates to subpopulations of bacteria giving a so-called burst of transposition. For example, IS30 only very rarely undergoes transposition. However, one of its transposase-mediated DNA rearrangements (site-specific deletion of DNA carried between two IS30 elements) can lead to a dimeric arrangement of two IS30. These forms have been seen to be highly unstable (Olasz et al., 1993, see also Arber et al., 1994). With frequencies nearly one million times higher than usual they mediate transposition events (thus the term burst of transposition). Even more frequently, the dimeric forms undergo still another event of site-specific deletion, and this involves the loss of one of the IS30 elements of the dimer. This process thus results again in the normal situation, where a single IS30 with its characteristic high degree of genetic stability is carried in the DNA segment in question. We can learn from this description that microbial populations may occasionally sort out a small subpopulation displaying high genetic instability. The advantage may be in the generation of rare beneficial mutants, while many other occurring mutations may be lethal. The frequent mutation “back to normal” is an interesting aspect of this phenomenon and can help to ensure the maintenance of the involved species or strain.

In this context, it may be relevant to refer to a study on the origin of lethal mutations in the genome of bacteriophage P1. Lethal mutations can be maintained and accumulated in the prophage state in lysogenic bacteria. Lethality becomes manifested when virus production is experimentally induced. Under these conditions, it was seen that 95% of independent lethal P1 mutations were due to the transposition of an IS element from the bacterial chromosome into the viral genome. Several different IS elements were involved in these mutagenic activities (Sengstag and Arber, 1983).

11. Evolution from prokaryotes to higher organisms

Much of the evidence underlying the postulates and proposals formulated here comes from microbial genetics. Prokaryotes are thought to have existed on our planet for

almost 4 billion years. In the course of long periods of time, they can be assumed to have developed and optimized their evolutionary fitness, specifically by developing the evolutionary genes still present and active today. This proficiency must also have allowed these single cellular organisms to sometimes experience multicellular cooperation, which can represent a source for division of labor and cell differentiation. Different forms of symbiosis must then have developed, which is a form of living together in mixed populations (Margulis, 1981). An extreme form of this is endosymbiosis which can give rise to permanent cohabitation. Cohabitation can highly favor occasional lateral gene transfer, a contribution to the evolutionary development by the strategy of DNA acquisition.

The coarse evolutionary path of the living world drawn here conforms well with the Darwinian theory of evolution. This theory receives more and more confirmation by our increasing knowledge on individual molecular processes contributing each in its own specific way to the evolutionary progress. Evolution is clearly a stepwise process in which the already approved genomic capacities can become either further improved or enriched by an additional, novel function. This ensures under constant selective pressure the evolutionary progress leading both to higher complexity and to a higher degree of biodiversity at all levels of complexity, including the microbial world.

12. Some philosophical, world view implications and practical aspects

The expansion of the Darwinian theory to the level of molecular processes can help to better explain the origin of the observed biodiversity. This development benefits from the rapidly accumulating data providing evidence for the different processes involved in the generation of genetic variations and in the actions exerted by natural selection on populations of different variants and of different organisms. A deepened understanding of molecular evolution can influence our world view.

Scientifically based world views should not a priori be considered as being in conflict with deeply anchored traditional wisdom. Rather, common features could guide the critical observer to find conformable interpretations for both scientific data and traditional texts. This process can result in an updated world view which may not be in conflict with its cultural roots (Arber, 2003).

The interpretation given to a set of scientific data is often heavily influenced by the general attitude of the investigator with regard to natural processes bringing about life and its evolutionary development. For example, some people see in the spontaneous generation of genetic variants the result of errors, accidents and illegitimate processes. An opposite attitude is based on the high appreciation of the inherent conceptual value of biological evolution. According to this latter view, which is defended in this paper, nature uses

both non-genetic intrinsic properties of matter and genetic tools to steadily produce genetic variations at frequencies not endangering the longer-term maintenance of functionally successful organisms. Nature is thus thought to care actively for the development of diversity of life manifestations. We marvel that life can exist under a large number of quite different living conditions, by having procured stepwise adaptations of the relevant organisms. This view contains a hopeful, positive message with regard to the future evolutionary progress of the living world.

There is wide agreement that genetic variation is in general not per se strictly adaptive. As a matter of fact, a majority of new genetic variants turn out to be unfavorable under the encountered prevailing living conditions. This, together with the fact that the underlying evolution genes are carried side by side with the genes fulfilling the essential needs of each individual life, forms the basis of intrinsic dualities anchored in the genome. The genetic information reveals a juxtaposition of favorable and unfavorable elements with respect to the needs of the fulfillment of each individual life. This situation can be seen as a sacrifice brought about for the harmonious progress of biological evolution at the level of populations.

The specific involvement of evolution genes driving biological evolution by the activity of their products as generators of genetic diversity prompts us to make a brief comment on the definition of the term gene or genetic determinant. Particularly in circles of non-experts, this is often too strictly taken to suggest that on the basis of an advanced knowledge of a gene sequence and of its biological activities, the scientists might also precisely predict life activities. A genetic program is thereby supposed to reliably guide life processes. This is of course true for some biological activities, but by no means for all. Enzymatically mediated generators of genetic variation will interact with the DNA, but the specific outcome of such interactions will and cannot be reproducible from case to case. At most, reproducibility will be statistical. And the long-term consequence of the activity of a variation generator can clearly not be foreseen in specific terms. A conclusion to be drawn from these considerations is that gene actions are not strictly programmed, they rather serve a purpose.

Another practical aspect of relevance concerning the knowledge of molecular evolution relates to the assessment of long-term risks of genetically modified organisms as produced by genetic engineering. In this strategy for basic research and biotechnological development, a deliberate modification is introduced into the genetic information of an organism, usually in small steps, as it is the case in the spontaneous, natural processes of generation of genetic variants. With increasing knowledge on the natural processes, these can be compared with those produced in genetic engineering (Arber, 2002b). Such a comparison can help to draw valid conclusions on conjectural risks of genetically deliberately modified organisms, which can of

course participate in the evolutionary process as all living organisms do.

References

- Arber, W., 1991. Elements in microbial evolution. *J. Mol. Evol.* 33, 4–12.
- Arber, W., 1993. Evolution of prokaryotic genomes. *Gene* 135, 49–56.
- Arber, W., 1995. The generation of variation in bacterial genomes. *J. Mol. Evol.* 40, 7–12.
- Arber, W., 1999. Involvement of gene products in bacterial evolution. In: Caporale, L.H. (Ed.), *Molecular Strategies in Biological Evolution*. Annals New York Academy of Sciences, vol. 870, pp. 36–44.
- Arber, W., 2000. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol. Rev.* 24, 1–7.
- Arber, W., 2002a. Evolution of prokaryotic genomes. In: Hacker, J., Kaper, J.B. (Eds.), *Pathogenicity Islands and the Evolution of Pathogenic Microbes*. *Curr. Top. Microbiol. Immunol.*, vol. 264/I, pp. 1–14.
- Arber, W., 2002b. Molecular evolution: comparison of natural and engineered variations. *Pontif. Acad. Sci. Scr. Varia* 103, 90–101.
- Arber, W., 2003. Traditional wisdom and recently acquired knowledge in biological evolution. *Proceedings of UNESCO Conference on “Science and the Quest for Meaning”*, Paris, 2002, in press.
- Arber, W., Naas, T., Blot, M., 1994. Generation of genetic diversity by DNA rearrangements in resting bacteria. *FEMS Microbiol. Ecol.* 15, 5–14.
- Glasgow, A.C., Hughes, K.T., Simon, M.I., 1989. Bacterial DNA inversion systems. In: Berg, D.E., Howe, M.M. (Eds.), *Mobile DNA*. American Society of Microbiology, Washington, DC, pp. 637–659.
- Goodman, M.F., Creighton, S., Bloom, L.B., Petruska, J., 1993. Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* 28, 83–126.
- Iida, S., 1984. Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome. *Virology* 134, 421–434.
- Iida, S., Hiestand-Nauer, R., 1987. Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific *Cin* recombinase of phage P1. *Mol. Gen. Genet.* 208, 464–468.
- Kunkel, T.A., Bebenek, K., 2000. DNA replication fidelity. *Ann. Rev. Biochem.* 69, 497–529.
- Margulis, L., 1981. *Symbiosis in Cell Evolution*. Freeman, San Francisco, USA.
- Olasz, F., Stalder, R., Arber, W., 1993. Formation of the tandem repeat (IS30)₂ and its role in IS30-mediated transpositional DNA rearrangements. *Mol. Gen. Genet.* 239, 177–187.
- Sengstag, C., Arber, W., 1983. IS2 insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: non-random distribution of target sites. *EMBO J.* 2, 67–71.
- Weber, M., 1996. Evolutionary plasticity in prokaryotes: a Panglossian view. *Biol. Philos.* 11, 67–88.