Proc. Indian natn. Sci. Acad., **B60**, No. 4, pp. 357-364 (1994)

On the Generation of Genetic Diversity in Microorganisms

WERNER ARBER

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

Bacterial genetics strongly influenced the development of molecular genetic strategies and techniques now available to study gene structure and functions of practically any living organisms. It also revealed natural processes of horizontal gene transfer (transformation, conjugation, phage-mediated transduction) as well as systems (e.g. restriction-modification systems) to hold such gene transfer in tolerably low frequencies to ensure a certain degree of genetic stability. Work with bacterial and bacteriophage systems has helped to unravel both homologous and non-homologous enzyme-mediated recombination processes at the molecular level. The acquired knowledge now helps to understand molecular processes contributing to the generation of genetic variation, especially DNA rearrangements resulting from transposition and from site-specific recombination which sometimes occurs at secondary crossing-over sites. Present knowledge on genetic plasticity of haploid microorganisms offers insights into the molecular basis for the natural interplay between mutagenesis and selection. This approach greatly profits from the short generation times and relatively small genome sizes of haploid microorganisms which allows one to investigate population genetic questions and to draw conclusions on the mechanisms of evolutionary processes.

Key Words: Genetic diversity, DNA rearrangement, microorganisms Site-specific recombination

Introduction

Microbial genetics is at the root of molecular genetics and gene technology. Their strategies and methods have become applicable to studies with any kind of living organisms. This has opened interesting possibilities to investigate and better understand mechanisms of interactions between biological macromolecules supporting life processes. Microbial genetics also offers deeper insights into the process of biological evolution. In this article we will concentrate on molecular mechamisms involved in microbial evolution, mainly at the level of spontaneous mutagenesis and the generation of genetic variation. Genetic diversity can be seen as a static picture of the dynamic process known as biological evolution. At least some of the principles identified in studies with microbial populations are likely to be of general relevance for biological evolution of all forms of life.

Important Roots of Molecular Genetics in the Decade Preceding the Publication of the Double Helix Model for DNA

The following discoveries largely based on work carried out between 1943 and 1953 with microorganisms were essential for the later development of molecular genetics (Arber 1993).

(i) It was realized that bacteria and bacteriophages have genes which can mutate,

and that spontaneous mutations normally arise independently of the presence of selective agents. It was also identified that the genetic information of bacteria and of some bacteriophages is carried in DNA molecules rather than in other biological macromolecules such as proteins.

(ii) The newly discovered phenomena of DNA transformation, bacterial conjugation and bacteriophage-mediated transduction demonstrated natural means of horizontal gene transfer between different bacterial cells.

(iii) It was seen that natural limits are set to horizontal gene transfer. Among these, several systems of host-controlled modification, today known as DNA restriction-modification systems, were independently described.

(iv) Mobile genetic elements were identified as source of genetic instability and to represent mediators of genetic rearrangements. Among such rearrangements is the integration of a bacteriophage genome into the genome of its bacterial host strain, which is thereby rendered lysogenic.

(v) Finally, structural analysis of DNA molecules lead to the double helix model offering an understanding of semiconservative DNA replication at the molecular level and thus of information transfer into progeny.

The Ease of Molecular and Population Genetic Investigations with Bacteria

Many classical microbial genetic investigations were carried out with *Escherichia coli* K-12. Its genome is a single circular DNA molecule (chromosome) of about 4,700,000 bp. It now has an extremely well established genetics with about 1,500 identified genes which represent, according to estimates, about half of all genes present. The haploid nature of *E. coli* brings about a rapid phenotypic manifestation of mutations. In periods of growth the rate of spontaneous mutagenesis is about 10^{-9} per bp and generation. This represents one new mutation per a few hundred cells in each generation. *E. coli* has several well studied bacteriophages and plasmids. This facilitates investigations on life processes in these bacteria.

Under good growth conditions the generation time of *E. coli*, measured between one cell division and the next, is very short, in the order of 30 min. Upon exponential growth this leads to a multiplication factor of 1000 every 5 hr. Thus, a population of 10^9 cells representing 30 generations is reached from an inoculum of a single cell in only 15 hr. This greatly facilitates population genetic studies and thus investigations on the evolutionary process.

Mechanisms and Effects of Spontaneous Alterations of Genomic DNA Sequences

On the filamentous DNA molecules of *E. coli* and its bacteriophages and plasmids, the genetic information is stored as linear sequences of nucleotides or base pairs. Genes depend on the presence of both open reading frames encoding specific gene products and expression control signals ensuring gene expression to occur at the relevant time with the needed efficiency. Mutations can affect reading frames as well as control signals, both of which represent specific DNA sequences. For simplicity, we will call here any type of alteration of a DNA sequence a spontaneous mutation.

It is well known that only relatively few mutations provide an advantage to the organism, can thus be considered as useful, become selected and succeed to eventually overgrow other members of microbial populations. More often, mutations either provide selective disadvantage or are lethal. In the longer term such mutations are eliminated from populations. Finally, some mutations are neutral or silent and do not immediately affect the life of the organism. Because of the relatively frequent occurrence

of lethal mutations and mutations providing selective disadvantage, a tolerable mutation rate of any haploid organism should be lower than one mutation per genome and generation. As we have already seen, this criterion is fulfilled by *E. coli.*

Molecular genetic studies have revealed that the overall spontaneous mutagenesis is brought about by a large number of specific molecular mechanisms. Not all of these mechanisms act with comparable efficiency and these efficiencies may depend on environmental conditions such as temperature. This renders studies on mutagenesis difficult. However, considerable knowledge has already been accumulated. On the basis of available data, we group the mechanisms of spontaneous mutagenes is, i.e. alterations of DNA sequences occurring without intervention of an investigator, into four categories as follows:

(i) Infidelity of DNA replication. An important source of such infidelities is likely to depend on tautomeric forms of nucleotides, i.e., a structural flexibility inherent to these organic compounds. Base pairing depends on specific structural forms and can result in a long term mispairing if shortliving, unstable tautomeric forms are "correctly" used in the synthesis upon DNA repliation. For these reasons, we do not consider mutations resulting in this process as errors but rather as infidelities. This process is a classical source of nucleotide substitution and it has its important role in the long-term development of new biological functions.

(ii) A wide variety of environmental mutagens, to which we count many external and internal chemicals, radiations and intrinsic metabolic effects, contribute in their specific way to the generation of genetic variation*.

(iii) Indeed, various recombination processes are well known to mediate DNA rearrangements which often represent new nucleotide sequences. While in haploid organisms general recombination is not essential for propagation, it has its role in various ways on the population level as generator of new sequence varieties. It can e.g. bring about sequence duplications and deletions by acting at segments of homology which are carried at different locations in a genome. Two other, widely spread types of recombination systems will be dealt with separately below: site-specific recombination and transposition. Both are known to contribute to genetic variation. Still other recombination processes, such as the one mediated by DNA gyrase, can perhaps best be grouped as illegitimate recombinations. This group may contain several different molecular mechanisms which act with low efficiency and have remained at least in part unexplained.

(iv) While the mutagenesis mechanisms listed in categories (i) to (iv) are exerted within the microbial genome and can affect any part of the genome, an additional category of spontaneous sequence alterations depends on an external source of genetic information. In DNA acquisition, genetic information indeed originates from another organism. This can occur either by transformation, by conjugation or by virus-mediated transduction. In the latter two strategies either a plasmid or a viral genome, respectively, acts as natural gene vector. DNA acquisition represents a particularly interesting source of new genetic information for the receptor bacterium, since the chance that the acquired DNA exerts useful biological functions is quite high in view of its likelihood to

*In part, sequence alterations brought about by replication infidelities and environmental mutagens are efficiently repaired by enzymatic systems. However, the efficiency of such repair is rarely 100%, so that evolutionarily relevant mutations persist. Some of the repair processes depend on genetic recombination systems, which by themselves can also contribute to spontaneous mutagenesis.

have already assumed the same functions in the donor bacterium. Besides the specific, already mentioned DNA transfer mechanisms, DNA acquisition largely also depends on DNA rearrangements belonging to the category (iii) processes. Indeed, some of the mechanisms of horizontal gene transfer depend on a recombinational interaction between the donor genome and the gene vector. Furthermore, the acquired genetic information must find a possibility in the receptor cell to become stably inherited. This can be insured either by recombination into the receptor genome or by its independent maintenance as a plasmid. The recombination mechanisms thereby involved can e.g. be general recombination in gene conversion, it can also be transposition or site-specific integration.

Site-specific DNA Inversion at Secondary Crossover Sites can be a Source of Novel Gene Fusions and Operon Fusions

In site-specific DNA inversion, a DNA segment bordered by specific DNA sequences acting as sites of crossing over becomes periodically inverted by the action of an enzyme called DNA invertase. Depending on the location of the crossover sites, DNA inversion can give rise either to gene fusion or to operon fusion. The underlying flip-flop system results in microbial populations of usually two, sometimes more, phenotypic appearances, e.g. phage populations with two different host ranges, if the DNA inversion affects the specificity of phage tailfibers, as is the case with phages P1 and Mu of *E. coli*.

Occasionally, a DNA sequence which may deviate considerably from the efficiently used crossover site can serve in DNA inversion. This results in novel DNA arrangements, many of which may not be maintained because of lethal consequences or reduced fitness. But a few new sequences may be beneficial for the life of the organism and be selectively favoured. This DNA rearrangement activity can thus be looked at as evolutionarily important. Since many different DNA sequences can serve in this process as secondary crossover sites, although at quite low frequencies, site-specific DNA inversion systems act as variation generators in large populations of microorganisms. We have thus postulated that this evolutionary role of DNA inversion systems may be more important than their much more efficient flip-flop mechanism which can at most help a microbial population to more readily adapt to two different, frequently encountered environmental conditions (Arber 1991). As a matter of fact, other strategies could be used as well for this latter purpose.

Transposition of Mobile Genetic Elements

Already nine different mobile genetic elements have been identified to reside in often several copies in the chromosome of *E. coli* K-12 derivatives. This adds up to about 1% of the chromosomal length as taken up by such insertion sequences, also called IS elements. At rates in the order of 10^{-6} per individual IS element and cell generation, these mobile genetic elements undergo transpositional DNA rearrangements. These include simple transposition of an element and more complex DNA rearrangements such as DNA inversion, deletion formation and the cointegration of two DNA molecules. Because of different degrees of specificity in the target selection upon transposition, the IS-mediated DNA rearrangements are neither strictly reproducible nor fully random. Again, transposition activities act as variation generators. In addition to DNA rearrangements mediated by the enzyme transposase, which is usually encoded by the mobile DNA element itself, other DNA rearrangements just take advantage of extended segments of DNA homologies, at

the sites of residence of identical IS elements at which general recombination can act. Altogether, IS elements represent a major source of genetic plasticity of microorganisms.

Transposition not only occurs in growing populations of bacteria, but also in prolonged phases of rest. This is readily seen with bacterial cultures stored at room temperature in stabs. Stabs are little vials containing a small volume of growth medium in agar. Stabs are inoculated with a drop of a bacterial culture taken up with a platinum loop which is stabbed from the top to the bottom of the agar. After overnight incubation the stab is tightly sealed and stored at room temperature. Most strains of *E. coli* keep viable in stabs during several decades of storage. That IS elements exert transpositional activities under these storage conditions is easily seen as follows:

A stab can be opened at any time, an aliquot of the bacterial culture picked and the bacteria well suspended in liquid medium. After appropriate dilution, aliquots are spread on solid medium. Individual colonies grown upon overnight incubation are then isolated. DNA from such subclones is extracted and fragmented with a restriction enzyme. The DNA fragments are separated by gel electrophoresis. Southern hybridization with appropriate hybridization probes can then show if different subclones reveal a restriction fragment length polymorphism (RFLP).

If this method is applied to subclones isolated from old stab cultures, and if DNA sequences from IS elements serve as hybridization probes, an extensive polymorphism is revealed. No or only little polymorphism is seen with hybridization probes from unique chromosomal genes. Good evidence is available that transposition represents a major source of this genomic plasticity observed in stabs, which at most allow for a very residual growth on the expense of dying cells. One can conclude that the enzymes promoting transposition are steadily present in the stored stabs. Indeded, the IS-related polymorphism increases linearly with time of storage for periods as long as 30 years. For a culture of *E. coli* strain W3110 it turned out to be difficult to guess which genome structure was at the origin of the bacterial population studied after 30 years of storage. In the average, each subclone had suffered about 12 RFLP changes as identified with hybridization probes from 8 different residential IS elements, of which IS5 was the most active. Nevertheless a very interesting pedigree of the analyzed individual subclones could be drawn and this offered an amazing insight into the genetic plasticity of *E. coli* (Naas et al. 1994).

Promotion and Limitation of Gene Acquisition

Transpositional activities and general recombination acting at IS elements, which are at different chromosomal locations, are a source of associating and dissociating chromosomal genes with natural gene vectors. These mechanisms have been well studied with conjugative plasmids and with bacteriophage genomes serving in specialized transduction. For example, composite transposons which are defined as two identical IS elements flanking a segment of genomic DNA with often more than one gene unrelated to the transposition process, are known to occasionally transpose into a natural gene vector and, after their transfer into the receptor cell, to transpose again into the receptor chromosome. Hence, together with other mechanisms, such as site-specific and illegitimate types of recombination, transposition also represents an important promoter of horizontal gene transfer.

One also finds several natural factors which seriously limit gene acquisition.

Transformation, conjugation and transduction depend on surface compatibilities of the receptor bacteria. Furthermore, upon penetration of donor DNA into receptor cells the DNA is very often confronted with restriction endonucleases. These cause a fragmentation of the invading foreign DNA, which is subsequently completely degraded. However, before fragments become degraded they are recombinogenic and may find a chance to incorporate all or part of their genetic information into the host genome. Therefore, we interpret the role of restriction systems as follows: they keep the rate of DNA acquisition low, and at the same time they stimulate the fixation of relatively small segments of acquired DNA to the receptor genome. This strategy of acquisition in small steps can best guarantee to microbial populations to occasionally extend their biological capacities without an extensive risk to mess up the functional harmony of the receptor cell by an acquisition of too many different functions at once. These considerations have their relevance at the level of selection of hybrids resulting from horizontal DNA transfer. This selection is exerted as one of the last steps in the acquisition process.

Genetic Diversity Reflects the Present State of Biological Evolution

Biological evolution is a steady, dynamic process. The four already described categories of molecular mechanisms of mutagenesis continuously exert their interactions on microbial populations, so that genetic diversity should steadily increase. However, while some of the newly arising DNA arrangements redisappear rapidly by appropriate repair processes, others may first remain but are then submitted to selection pressure. This, together with sampling due to the limited size of the biosphere of our planet, which offers space to only about 10^{30} living cells, may normally keep genetic diversity at a more or less constant level. This may hold as long as the overall environmental conditions of life do not undergo drastic alterations, which would of course seriously affect selection. It is for this reason that a relatively large pool of gene functions and life forms, together with some genetic plasticity, are important in order to insure good chances to genetic adaptation in times of changes of living conditions.

Gene acquisition can be seen as a very efficient mean of a microbial strain to extend its genetic capacities. This may be of particular relevance, if selective pressure undergoes drastic changes, as we have e.g. testified it since the wide use of antibiotics in human and veterinary medicine. Much of our present knowledge on gene acquisition stems indeed from studies on the spreading and selection of drug resistance determinants. DNA acquisition, which we can consider as a strategy of sharing in the success of others, should be accounted for in drawing the evolutionary tree of bacteria by adding horizontal shunts between individual branches. While, as we have discussed, usually only small DNA segments flow through such shunts at one time or another in horizontal gene transfer, in the vertical flux of genes from one generation to the next in the growing branches of the tree, the entire genome is of course steadily target of anyone of the mechanisms belonging to the first three categories of spontaneous mutagenesis. In comparing with DNA acquisition and in considering efficiencies per single event, DNA rearrangements internal to a genome may be evolutionarily less efficient and nucleotide substitution still less. However, all of these processes make their important contibutions to the evolutionary process. Nucleotide substitution is a major source of new biological functions and it also contributes to the amelioration of existing functions. DNA rearrangements can bring

about important improvements of existing capacities, e.g. by the fusion of functional domains and of DNA sequence motifs. In considering the evolutionary process, we should keep in mind that it largely depends both on the diversity and on the kind of genetic information already available, since time spans needed to develop completely new functions without making use of existing sequences are very long, keeping in mind the extremely large number of different specific sequences that are possible in the linear arrangement of nucleotides of a gene and of a genome.

Conclusions

Rather than to be the result of an accumulation of errors, biological evolution appears to depend on many different specific biological functions. Enzymes acting as variation generators make fundamental contributions to spontaneous mutagenesis, including the rearrangement of existing sequences. In addition, more or less complex enzyme systems and organelles promote horizontal gene transfer and regulate its efficiency. With only a few exceptions, enzymes involved in these processes are not required for life of individuals in populations, and their biological significance becomes manifested only at the population level. As a matter of fact, spontaneous mutagenesis more often hampers the life of an affected individual than it would bring it benefit. The benefit only becomes obvious at the population level by its contribution to a steady evolution. It is a very interesting concept for our world view to know that among the genes encoded in the genomes of living organisms there are, besides a large number of genes essential for each individual life, also other genes, the products of which can insure a long-term development of life in a wide variety of phenotypic forms. Molecular genetic studies carried out with haploid microorganisms have brought about good evidence for these conclusions. It is quite likely that they might also apply to higher organisms, but these are much less accessible than microorganisms to experimental approaches in search of evidence.

References

Arber W 1991 Elements in microbial evolution; *J. Mol. Evol.* **33** 4-12

——— 1993 Evolution of prokaryotic genomes; *Gene*, **135** 49-56

Naas T, Blot M, Fitch W M and Arber W 1994 Insertion sequence-related genetic variation in resting *E.coli* K-12; *Genetics* **136** 721-730