J. Hyg., Camb. (1985). **95**, 611–618
Printed in Great Britain

Escherichia coli as a genetic tool

By NAOMI DATTA*
Formerly Royal Postgraduate Medical School, London

SUMMARY

The study of *Escherichia coli* and its plasmids and bacteriophages has provided a vast body of genetical information, much of it relevant to the whole of biology. This was true even before the development of the new techniques, for cloning and analysing DNA, that have revolutionized biological research during the past decade. Thousands of millions of dollars are now invested in industrial uses of these techniques, which all depend on discoveries made in the course of academic research on *E. coli*. Much of the background of knowledge necessary for the cloning and expression of genetically engineered information, as well as the techniques themselves, came from work with this organism.

MUTATION AND MUTAGENESIS

The study of microbial genetics dates back to the 1940s when it was first proved that in *Escherichia coli*, development of resistance to phage resulted from spontaneous mutations that occurred at a measurable rate in the presence or absence of the phage. After the discovery of the structure and genetic function of DNA in the 1950s, research, often on *E. coli*, demonstrated the biochemical basis of mutations and how mutagens can produce them. Relative resistance to mutagenic agents depends on enzyme mechanisms that repair damaged DNA; when the mechanisms are faulty, the bacteria become highly sensitive to UV light and other mutagens. Bacteria are used to measure the mutagenic properties of chemicals (though applied in the Ames test, *Salmonella typhimurium* rather than *E. coli* is used).

Mutation implies changed sequence of DNA bases, substitutions of one base for another, deletions from or insertions into the molecule. It can result in an altered gene product or in altered regulation of production. An important kind of mutation is caused by the disruption of a gene by insertion of an extraneous DNA sequence. In bacteria, insertion sequences (and transposons) are moveable lengths of DNA that encode information for their own transposition into new genetic sites. Their relationship to the genetic control elements, discovered in maize in the 1950s, has been recognized and analogous sequences are found in animals. *E. coli* provides models for the study of these sequences, results of which may be relevant in other areas of research, such as oncology.

* Present address: 9, Dukes Avenue, London W4 2AA, U.K.

GENETIC CROSSES

Although bacteria are haploid organisms, they can undergo genetic recombination by cross-overs between homologous genetic regions, analogous to the crosses that occur between pairs of chromosomes in sexual reproduction. The *E. coli* chromosome is a single, circular DNA molecule. The species has no regular sexual cycle, but extraneous DNA can be introduced into living cells in nature or in the laboratory. When the introduced DNA is a chromosomal fragment of the same or a closely related species the genes it carries are aligned (by unknown mechanisms) with their alleles on the chromosome of the recipient cell, and recombination follows. This reciprocal exchange between homologous DNA sequences requires enzymes (Rec, or recombination, gene products) that are also important in DNA repair. Extraneous DNA is introduced into bacterial cells by three known mechanisms, transformation, conjugation and transduction.

Transformation

This results from uptake of DNA directly from the environment of the cell. The phenomenon was discovered in 1928 by Griffith in pneumococci; non-capsulated strains were transformed by contact with killed, capsulated cells, acquiring the capsular type of the dead cells. Avery's work, showing that the transforming principle was DNA, was an important step in solving the mystery of heredity. Much later it was found possible to transform salmonella and *E. coli*; they will take up DNA if made competent to do so by treatment with calcium ions and heat shock. This technique is important in DNA cloning.

Conjugation

In bacteria this implies the transfer of DNA direct from one cell to another. It depends upon the possession, by the donor cell, of transfer genes. The first known example of bacterial conjugation, and the one that is best understood, is that determined by the transfer genes of F (fertility) factor of *E. coli* K 12. After mixing two mutants of this strain with different nutritional requirements, Lederberg & Tatum, in 1946, isolated genetic recombinants, which seemed to imply that *E. coli* was capable of sexual reproduction. However, it was later found that one of the strains carried the infective agent, F (now classified as a conjugative plasmid), without which no recombination occurred. F itself is transferred at high frequency between cells of *E. coli*, so that if one F^+ cell is introduced into an F^- population, the whole culture may be F^+ after overnight incubation. But recombinants occur only if the F factor 'mobilizes' the chromosome of the donor cell, carrying it, or part of it, into the recipient cell. High frequency recombination (Hfr) occurs if F becomes integrated into the chromosome of the host cell, in which case the chromosome is transferred as if it were itself a conjugative plasmid. Hfr strains have played an important part in *E. coli* genetics. Because the type of recombinants they produce varies according to the site and the orientation of insertion of F, their study has allowed the mapping of genes all round the *E. coli* chromosome. Another important discovery was that, after insertion into the chromosome, F is sometimes inaccurately excised, taking with it one or more chromosomal genes. Such a plasmid, an F′, can be used to construct *E. coli* strains diploid for particular genes.

Using F'lac, mutations affecting the fermentation of lactose could be studied both *in cis* and *in trans*. It was shown that a repressor prevented the synthesis of β-galactosidase except in the presence of the substrate for the enzyme, lactose or other galactoside, which induced enzyme synthesis. By 1960, interactions of repressor, operator and inducer were elucidated and the concept of an *operon*, including regulatory and structural genes was introduced. Later the repressor of the lac operon was found to be a protein. This is just one early example of the use of *E. coli* to explore genes and genetic control. All the genes of an operon are transcribed into a single molecule of messenger RNA.

Transduction

In microbial genetics this means the injection, from a phage particle, of non-phage DNA into a bacterium. When bacteria are lysed by phage the cells are turned over from their normal functions into factories for production of new phage particles. In the final stages, preformed phage genomes are packed into protein coats to be liberated. Occasional particules contain, instead of phage DNA, random fragments of bacterial DNA. Such 'generalized' transducing particles, encountering a phage-sensitive cell, will inject their DNA into it: no lysis follows, but if homologous DNA is present in the new cell, crossover will occur, allowing the introduced DNA to be incorporated into the chromosome of the new cell.

Phage λ provides an example of 'specialized' transduction. As a prophage, it integrates at a specific place in the *E. coli* chromosome. To produce lysis, it must excise itself from its chromosomal locus, and in doing so is liable occasionally to take with it the bacterial genes located on one side or the other of its site, which are *gal* (galactose fermentation) and *bio* (biotin synthesis). Resulting transducing particles contain phage DNA linked to the bacterial genes *gal* or *bio*. When injected into a new *E. coli* cell, the DNA again integrates at the specific λ site and the relevant bacterial gene, *gal* or *bio*, is duplicated.

Genetic studies in *E. coli* (usually *E. coli* K 12) using these three methods of gene transfer, have permitted about 1000 genes to be mapped round the chromosome of *E. coli* whose total molecule consists of about 4 million DNA base pairs (4000 kb). The biochemical mechanisms of metabolic and synthetic pathways have been explored and the regulatory mechanisms controlling their expression investigated.

When phage or plasmid DNA is introduced into bacteria by any of these methods, recombination into the chromosome is not necessary, the DNA can establish itself extrachromosomally. In these cases therefore the Rec genes are not required.

NEW GENETICAL TECHNIQUES

The new genetics rests on the possibility of isolating genes, inserting them into a new DNA molecule and introducing that molecule into a living organism in which it will replicate and where the gene will be expressed. It has been made possible by a number of factors, briefly outlined below.

The genetic message

The new genetics is based on the series of discoveries about how DNA is replicated, how the message encoded in its base sequence is transcribed into

messenger RNA (mRNA) and how that message is translated into the specific sequence of amino acids when the mRNA reaches the ribosomes. *E. coli* and its phages played an important part in these discoveries. It was found that the genetic message was carried in the same code in man, *E. coli* and the tobacco mosaic virus. Indeed, the code is the same in all living things, in that each amino acid is specified by the same triplets of bases. Punctuation of the message is, up to a point, also the same, in that certain triplets indicate start and stop for translation from RNA to protein. However, the regulation of protein production is not the same in bacteria as in eukaryotes. In bacteria, most control is at the level of transcription of the message into mRNA, i.e. it is operon control. The greater the number mRNA molecules read off from the DNA gene, the more of the relevant protein will be made; the number of molecules depends on the promotor, that is the site, for that particular operon, at which RNA polymerase attaches and initiates mRNA synthesis. The promotor is controlled by a variety of agents, repressors or activators, that ensure that the right proteins are made according to the needs of the bacterial cell. The lac operon is required for lactose fermentation and is repressed unless lactose is available, an example of negative control. Amino acids need not be synthesized if they are available in the medium, as in nutrient broth, but in the absence of an amino acid, the genes for the synthesis of that amino acid are activated, an example of positive control. Many complex mechanisms of gene regulation are known in *E. coli* and its plasmids and phages. Regulation is different in eukaryotes in which the RNA message has to be processed to delete intron sequences and to pass out from the nucleus to cytoplasm.

Gene vectors

The first DNA molecules into which foreign genes were inserted for cloning were resistance (R) plasmids in *E. coli*. Small, non-conjugative plasmids (i.e. lacking transfer genes such as those of plasmid F) are the easiest to separate from lysed bacteria and the easiest to manipulate *in vitro*; they are less fragile than large molecules and carry less unwanted genetic information. A plasmid vector should be easily introduced into a living cell and once introduced should replicate satisfactorily, so that it is retained in all daughter cells, preferably producing multiple copies of itself in each cell. It must include a gene allowing the bacterial cell carrying it to be selected out from millions of cells lacking the vector. Antibiotic resistance genes of naturally occurring R plasmids are used for this purpose. The vector must also have appropriate cut sites for restriction endonucleases (see below). Plasmid pBR322 fulfils these requirements. It was derived from a colicinogenic plasmid, ColE1, a DNA molecule of about 6 kb pairs that produces many copies of itself in each *E. coli* cell. Into different sites of its molecule, genes for ampicillin and tetracycline resistance have been introduced. The tetracycline-resistance gene includes the only sequence of pBR322 that can be cut by the enzyme *HindIII*. Into this site, foreign DNA can be cloned, in which case the tetracycline gene is disrupted and no longer expressed. To clone DNA into pBR322 the method is (i) cut pBR322 with *HindIII* to produce linear molecules. Cut the DNA to be cloned with the same enzyme and mix it with the pBR322 molecules. (ii) Treat the mixture with DNA ligase to recircularize the molecules of pBR322, some of which will have incorporated a fragment of foreign DNA. (iii) Expose a 'competent' culture of *E. coli* to the DNA mixture. (iv) After a temperature shock

to encourage the culture to take up DNA (transformation) and allowing a little time for expression of their new genes, grow the cells in (or on) medium containing ampicillin. Only cells that have acquired pBR322 will grow. (v) Now screen them for tetracycline resistance. Tetracycline sensitive cells will be those in which pBR322 has acquired new DNA. Each such cell carries a cloned (i.e. isolated and replicated) DNA fragment. Plasmid DNA can be isolated from each clone and the introduced fragment separated from pBR322 DNA by cutting again with *Hind III*. The cloned fragment can then be purified by centrifugation or electrophoresis.

Phage λ is also used as a cloning vector in *E. coli*. The cloned gene can be obtained in quantity by using variants whose repressor of replication is temperature-sensitive. Raising the temperature of the culture to 42 °C inactivates the repressor and the phage enters the lytic phase, making hundreds of copies of its own genome and of the cloned gene.

Relatively small amounts of DNA can be cloned in plasmids or in phage λ . 'Cosmids' were devised to carry longer sequences, approximately the amount of DNA to fill a head of phage λ (50 kb). When λ lyses *E. coli*, preformed phage genomes are linearized for packing into the phage particles. The two ends of the linear molecule are single-stranded DNA, consisting of complementary base pairs. When injected into a new cell the phage genome is circularized by the joining up of the complementary (and therefore cohesive) ends. A cosmid is a plasmid, including plasmid replication and antibiotic-resistance genes and the cohesive ends of λ, but not the rest of the λ genome. This DNA can be packed into preformed λ heads *in vitro* to produce artificial transducing particles; DNA cloned in the cosmid is efficiently injected into *E. coli* cells.

These cloning vectors, and related ones designed for specific purposes, are much used in *E. coli*. Vectors for gene cloning in other bacteria, especially *Bacillus subtilis* and in eukaryotic cells, yeast or mammalian cultures, have also been devised. These vectors, even in mammalian cells, usually include antibiotic resistance genes from bacterial plasmids as selective markers.

Enzymes for genetic engineering

(i) *Restriction endonucleases*, essential tools in gene cloning, were discovered because phage propagated on one strain of *E. coli* would not efficiently infect another strain – its host range was restricted (hence the term restriction enzyme). Many restriction enzymes have now been isolated; they are named according to their origin, e.g. *Hind III* came from a strain of *Haemophilus influenzae* type d, but defined according to the specificity of their action. They cleave both strands of the DNA double helix, either straight across to give blunt ends or at staggered sites, to give single-stranded ends. In the latter case, the ends, being complementary are sticky (like the cohesive ends of λ genomes) and will hybridize with one another or with any other DNA fragments having the same single-stranded base sequence. For cloning, the vector molecule is cut once with such an enzyme, converting it from the circular to the linear form. It is exposed to high concentrations of the DNA to be cloned, cut with the same enzyme. The chances of random collisions of the vector with the fragments are high, and the complementary ends hold them together. The recognition and cut sites for these enzymes are short palindromic base sequences; two examples are shown in Fig. 1.

(ii) *DNA ligase* is used to anneal the associated sticky ends. It is usually derived

Enzyme	Double-stranded cut site	Single-stranded sticky ends
Hind III	A|A G C T T T T C G A|A	– T C G A
Eco RI	G|A A T T C C T T A A|G	– T T A A

Fig. 1. Examples of restriction endonucleases.

from phage T4 in *E. coli*. Ligase will also join blunt ends of cut DNA molecules but efficiency is low.

(iii) *RNA polymerase* for transcription of DNA message *in vitro*.

(iv) *Reverse transcriptase* makes a DNA copy (cDNA) from an RNA message. Other enzyme systems, besides these, are used in genetic manipulations, many being available commercially.

Aids to the new genetics, commercially available

A great variety of restriction enzymes, also ligase, RNA polymerase etc. can be bought with instructions, ready to use. Plasmid and phage DNA can also be bought, either intact or pre-cut with the enzyme of choice. In case of absence of the right sequence for engineering molecules with restriction enzymes, DNA linkers and adaptors are available, that include the cut sites for chosen enzymes, so that molecules with sticky ends can be tailored to requirements. Reagents for labelling DNA to make probes, for sequencing DNA, for synthesizing DNA from an RNA molecule or synthesizing protein from a DNA molecule are all for sale, together with manuals explaining their use. Some fabricated plasmids and some techniques for genetic manipulation are patented, but this limits their industrial rather than experimental use.

USES OF THE NEW GENETICS

Gene cloning can be used to analyse genetic pathways, to investigate functions of normal or abnormal cells, to provide diagnostic tools, to make safe vaccines and to synthesize a variety of biochemical compounds such as hormones. The purpose of cloning may be to make use of the cloned DNA itself or to induce the production of its encoded protein: the DNA vector and cells used will depend on the purpose. For protein production in *E. coli*, special vectors are available with cloning sites so placed as to bring the cloned gene into an inducible operon with a strong promotor, for maximum and controllable expression. It may be desirable for the gene product to be excreted by the producing cells. This may be a reason for using other organisms than *E. coli*, and fabricating vectors effective in *B. subtilis* or *Saccharomyces*, to include useful sequences devised in *E. coli*. Much ingenuity goes into the development of such systems.

A genomic library for an organism is made by cloning its whole genome, as separate fragments. In such a library, a search can be made for individual genes, recognized either by their products or by their DNA sequence. A product can be recognized by its biological function such as toxicity or hormone activity, or by its serological specificity. (The uses of gene cloning interact with those of

monoclonal antibody production.) The DNA sequence can be recognized by using a gene probe (see below) or, if no such sequence has previously been isolated, by chemical analysis. From analysis of the sequence, making use of knowledge of the genetic code, the encoded amino acid sequence is discovered. The protein can even be synthesized *in vitro* by providing the required components, the gene, RNA polymerase, ribosomes, amino acids and tRNA in an artificial medium.

A gene probe is a DNA fragment labelled, usually, with ^{32}P. Denatured DNA (i.e. single stranded) to be probed is exposed to the denatured probe and if the base sequences are the same, it will re-associate with the probe to make double-stranded, labelled DNA, separable and recognizable. Probes can be used with DNA fragments blotted on to filters (Southern blot technique) or on bacterial colonies, also on filters, lysed to expose their DNA (colony hybridization). Fragments or colonies that take up the probe are identified by autoradiography.

Cloning of microbial genes has potential medical as well as experimental uses. It should be possible to make safe, effective vaccines by cloning the genes for protective antigens without virulence factors. This requires genetic analysis of pathogenic organisms, bacteria, viruses or protozoa, that can now be achieved, and in some cases is already achieved. *E. coli*, for example, with the appropriate cloned gene, can synthesize fragment A of diphtheria toxin: a gene for the surface antigen of hepatitis B virus, cloned not in *E. coli* but in vaccinia virus, produces antibodies in rabbits: recombinant non-toxigenic *Vibrio cholerae* strains have been developed as possible vaccines.

Gene probes will probably be used more and more for the rapid diagnosis of infection. Already probes have been used to identify *E. coli* enterotoxin (or, more accurately, enterotoxin gene) in crude samples of human faeces. Research to identify DNA sequences specific for particular pathogens, together with the development of a labelling method that does not require radioactive reagents, will bring new methods to diagnostic microbiology laboratories.

A number of human genes have been cloned and expressed in *E. coli*. The isolation of a human gene requires much ingenuity and hard work but brings important rewards. Fewer human than *E. coli* genes have so far been mapped, the difficulties being obviously much greater. The cloning of human and animal genes can help in solving problems of many kinds in biochemistry, immunology, oncology, tissue differentiation, as well as in identifying genetic lesions that produce hereditary diseases. Probes would allow prenatal diagnosis, and are already possible for the haemaglobinopathies such as sickle cell anaemia. Ultimately, perhaps, therapy to replace an abnormal with a normal gene may be possible.

Therapeutic agents from bacteria are already in production. Human insulin was the first to be marketed. Production of other hormones and human interferons is in progress, the interferons already undergoing clinical trials. Factor VIII, for the treatment of haemophilia, has been cloned and no doubt production will be accelerated to protect haemophiliacs from exposure to blood-borne viruses. Vaccines have been mentioned. Many other possibilities are being explored by the pharmaceutical industry.

EXPERIMENTAL CONDITIONS

When the first vertebrate DNA was cloned in *E. coli* about 10 years ago, fears of unforeseen dangers were expressed, for instance that virulent new pathogens would emerge or even that *E. coli* carrying accidentally cloned oncogenes might escape into human communities. Scientists and governments tried to devise protective rules for DNA manipulations, based on estimates of possible risks, and these are now laid down. Where work involves genes that might be harmful, levels of physical containment are applied, as for the handling of known pathogens. In addition, cloning vectors and bacteria are used that would not spread if they should accidentally escape from the laboratory. Vectors are non-transmissible, and could not transfer themselves to wild strains of bacteria. Bacterial hosts are 'disabled' in having mutations, such as *recA*, which prevents DNA repair and limits viability. Containment, biological and physical, is strictly enforced especially in the large-scale cultures of recombinant DNA used in industry.

REFERENCES

The following books and papers will allow the primary sources for this review to be found.

ENGLEBERG, N. C. & EISENSTEIN, B. I. (1984). The impact of new cloning techniques on the diagnosis and treatment of infectious disease. *New England Journal of Medicine* **311**, 892–901.

GLASS, R. E. (1982). *Gene Function:* E. coli *and Its Heritable Elements.* London: Croom Helm.

LANCET (1984). DNA in Medicine. (Series of ten weekly articles by different authors, 13 Oct.–22 Dec.)

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning* (*Laboratory Manual*). New York: Cold Spring Harbor Laboratory.