

DNA Amplification by Polymerase Chain Reaction (PCR)

The introduction of cell-free methods for multiplying DNA fragments of defined origin (DNA amplification) in 1985 ushered in a new era in molecular genetics (the principle of PCR is contained in earlier publications). This fundamental technology has spread dramatically with the development of automated equipment used in basic and applied research.

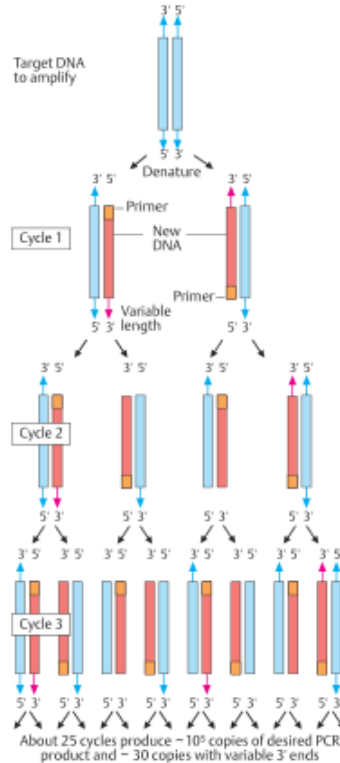
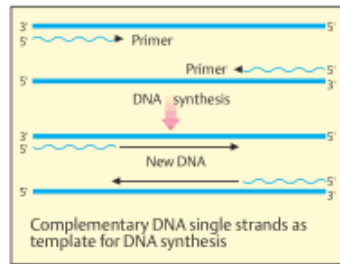
Polymerase chain reaction (PCR)

PCR is a cell-free, rapid, and sensitive method for cloning DNA fragments. A standard reaction and a wide variety of PCR-based methods have been developed to assay for polymorphisms and mutations. Standard PCR is an *in vitro* procedure for amplifying defined target DNA sequences, even from very small amounts of material or material of ancient origin. Selective amplification requires some prior information about DNA sequences flanking the target DNA. Based on this information, two oligonucleotide primers of about 15–25 base pairs length are designed. The primers are complementary to sequences outside the 3' ends of the target site and bind specifically to these. PCR is a chain reaction because newly synthesized DNA strands act as templates for further DNA synthesis for about 25–35 subsequent cycles. Theoretically each cycle doubles the amount of DNA amplified. At the end, at least 10^5 copies of the specific target sequence are present. This can be visualized as a distinct band of a specific size after gel electrophoresis. Each cycle, involving three precisely time-controlled and temperature-controlled reactions in automated thermal cyclers, takes about 1–5min. The three steps in each cycle are (1) denaturation of double-stranded DNA, at about

93–95°C for human DNA, (2) primer annealing at about 50–70°C depending on the expected melting temperature of the duplex DNA, and (3) DNA synthesis using heat-stable DNA polymerase (from microorganisms living in hot springs, such as *Thermophilus aquaticus*, Taq polymerase), typically at about 70–75°C. At each subsequent cycle the template (shown in blue) and the DNA newly synthesized during the preceding cycle (shown in red) act as templates for another round of synthesis. The first cycle results in newly synthesized DNA of varied lengths (shown with an arrow) at the 3' ends because synthesis is continued beyond the target sequences. The same happens during subsequent cycles, but the variable strands are rapidly outnumbered by new DNA of fixed length at both ends because synthesis cannot proceed past the terminus of the primer at the opposite template DNA.

cDNA amplification and RT-PCR

A partially known amino acid sequence of a polypeptide can be used to obtain the sequence information required for PCR. From its mRNA one can derive cDNA, and determine the sequence of the sense and the antisense strand to prepare appropriate oligonucleotide primers (1). When different RNAs are available in small amounts, rapid PCR based methods are employed to amplify cDNA from different exons of a gene. cDNA is obtained by reverse transcriptase from mRNA, which is then removed by alkaline hydrolysis (2). After a complementary new DNA strand has been synthesized, the DNA can be amplified by PCR (3). Reverse transcriptase PCR (RT-PCR) can be used when the known exon sequences are widely separated within a gene. With rapid amplification of cDNA ends (RACE-PCR), the 5' and 3' end sequences can be isolated from cDNA.



Polymerase chain reaction (PCR)

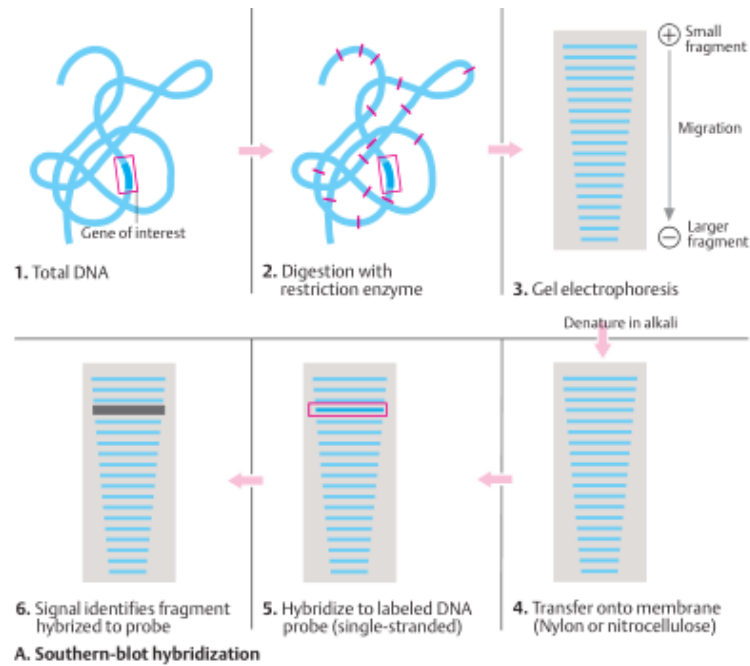
Restriction Analysis by Southern Blot Analysis

Restriction endonucleases are DNA-cleaving enzymes with defined sequences as targets. They are often simply called restriction enzymes. Since each enzyme cleaves DNA only at its specific recognition sequence, the total DNA of an individual present in nucleated cells can be cut into pieces of manageable and defined size in a reproducible way. Individual

DNA fragments can then be selected, ligated into suitable vectors, multiplied, and examined. Owing to the uneven distribution of recognition sites, the DNA fragments differ in size. A starting mixture of DNA fragments is sorted according to size. Two procedures detect target DNA or RNA fragments after they have been arranged by size in gel electrophoresis—the Southern blot hybridization for DNA (named after E. Southern who developed this method 1975) and the Northern blot hybridization for RNA (a word play on Southern, not named after a Dr. Northern). Immunoblotting (Western blot) detects proteins by an antibody-based procedure.

Southern blot hybridization

The analysis starts with total DNA (1). The DNA is isolated and cut with restriction enzymes (2). One of the not yet identified fragments contains the gene being sought or part of the gene. The fragments are sorted by size in a gel (usually agarose) in an electric field (electrophoresis) (3). The smaller the fragment, the faster it migrates; the larger, the slower it migrates. Next, the blot is carried out: The fragments contained in the gel are transferred to a nitrocellulose or nylon membrane (4). There the DNA is denatured (made single-stranded) with alkali and fixed to the membrane by moderate heating (~ 80°C) or UV cross-linkage. The sample is incubated with a probe of complementary singlestranded DNA (genomic DNA or cDNA) from the gene (5). The probe hybridizes solely with the complementary fragment being sought, and not with others (6). Since the probe is labeled with radioactive ^{32}P , the fragment being sought can be identified by placing an X-ray film on the membrane, where it appears as a black band on the film after development (autoradiogram) (6). The size, corresponding to position, is determined by running DNA fragments of known size in the electrophoresis.

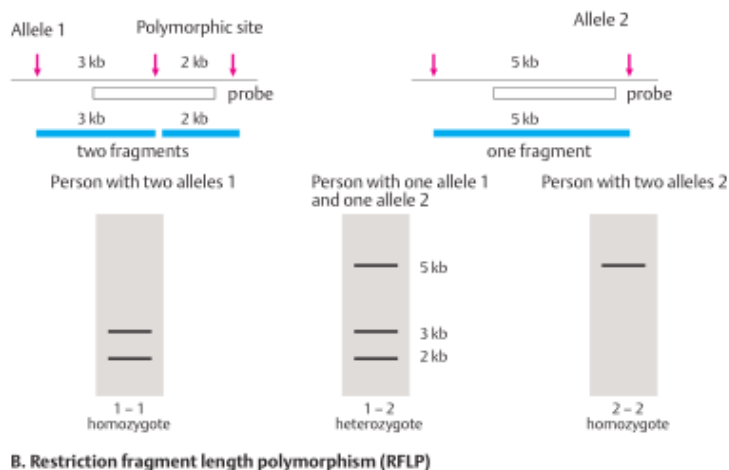


Southern blot hybridization

Restriction fragment length polymorphism (RFLP)

In about every 100 base pairs of a DNA segment, the nucleotide sequence differs in some individuals (DNA polymorphism). As a result, the recognition sequence of a restriction enzyme may be present on one chromosome but not the other. In this case the restriction fragment sizes differ at this site (restriction fragment length polymorphism, RFLP). An example is shown for two 5 kb (5000 base pair) DNA segments. In one, a restriction site in the middle is present (allele 1); in the other (allele 2) it is absent. With a Southern blot, it can be determined whether in this location an individual is homozygous 1-1 (two alleles 1, no 5 kb fragment), heterozygous 1-2 (one allele each, 1 and 2), or homozygous 2-2 (two alleles 2). If the mutation being sought lies on the chromosome carrying the 5 kb fragment, the presence of this fragment indicates presence of the mutation. The absence of this fragment would indicate that the mutation is absent. It is important to understand that the RFLP itself is unrelated to the mutation. It simply distinguishes DNA fragments of

different sizes from the same region. These can be used as markers to distinguish alleles in a segregation analysis. In addition to RFLPs, other types of DNA polymorphism can be detected by Southern blot hybridization, although polymerase chain reaction-based analysis of microsatellites is now used more frequently.



B. Restriction fragment length polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP)

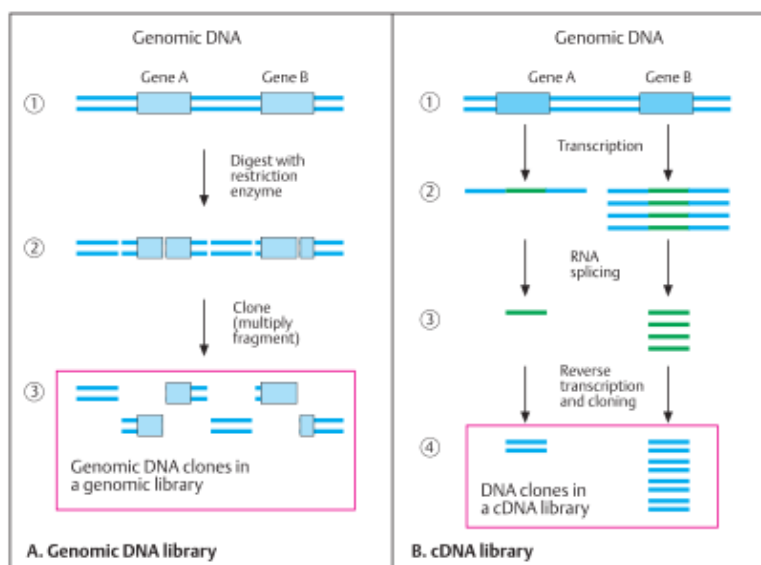
DNA Libraries

A DNA library is a collection of DNA fragments that in their entirety represent the genome, that is, a particular gene being sought and all remaining DNA. It is the starting point for cloning a gene of unknown chromosomal location. To produce a library, the total DNA is digested with a restriction enzyme, and the resulting fragments are incorporated into vectors and replicated in bacteria. A sufficient number of clones must be present so that every segment is represented at least once. This is a question of the size of the genome being investigated and the size of the fragments. Plasmids and phages are used as vectors. For larger DNA fragments, yeast

cells may be employed. There are two different types of libraries: genomic DNA and cDNA.

Genomic DNA library

Clones of genomic DNA are copies of DNA fragments from all of the chromosomes (1). They contain coding and non coding sequences. Restriction enzymes are used to cleave the genomic DNA into many fragments. Here four fragments are schematically shown, containing two genes, A and B (2). These are incorporated into vectors, e.g., into phage DNA, and are replicated in bacteria. The complete collection of recombinant DNA molecules, containing all DNA sequences of a species or individual, is called a genomic library. To find a particular gene, a screening procedure is required (see B).



Genomic DNA and cDNA library

cDNA library

Unlike a genomic library, which is complete and contains coding and non coding DNA, a cDNA library consists only of coding DNA sequences. This specificity offers considerable advantages over genomic DNA. However, it requires that mRNA be available and does not yield information about the structure

of the gene. mRNA can be obtained only from cells in which the respective gene is transcribed, i.e., in which mRNA is produced (1). In eukaryotes, the RNA formed during transcription (primary transcript) undergoes splicing to form mRNA. Complementary DNA (cDNA) is formed from mRNA by the enzyme reverse transcriptase (3). The cDNA can serve as a template for synthesis of a complementary DNA strand, so that complete double-stranded DNA can be formed (cDNA clone). Its sequence corresponds to the coding sequences of the gene exons. Thus it is well suited for use as a probe (cDNA probe). The subsequent steps, incorporation into a vector and replication in bacteria, correspond to those of the procedure to produce a genomic library. cDNA clones can only be won from coding regions of an active (mRNA-producing) gene; thus, the cDNA clones of different tissues differ according to genetic activity. Since cDNA clones correspond to the coding sequences of a gene (exons) and contain no noncoding sections (introns), cloned cDNA is the preferred starting material when further information about a gene product is sought by analyzing the gene. The sequence of amino acids in a protein can be determined from cloned and sequenced cDNA. Also, large amounts of a protein can be produced by having the cloned gene expressed in bacteria or yeast cells.

Screening of a DNA library

Bacteria that have taken up the vectors can grow on an agar-coated Petri dish, where they form colonies (1). A replica imprint of the culture is taken on a membrane (2), and the DNA that sticks to the membrane is denatured with an alkaline solution (3). DNA of the gene segment being sought can then be identified by hybridization with a radioactively (or otherwise) labeled probe (4). After hybridization, a signal appears on the membrane at the site of the gene segment (5). DNA complementary to the labeled probe is located here; its exact position in the culture corresponds to that of the signal on the membrane (6). A probe is taken from the

corresponding area of the culture (5). It will contain the desired DNA segment, which can now be further replicated (cloned) in bacteria. By this means, the desired segment can be enriched and is available for subsequent studies.

cDNA cloning

cDNA is a single-stranded segment of DNA that is complementary to the mRNA of a coding DNA segment or of a whole gene. It can be used as a probe (cDNA probe as opposed to a genomic probe) for the corresponding gene because it is complementary to coding sections (exons) of the gene. If the gene has been altered by structural rearrangement at a corresponding site, e.g., by deletion, the normal and mutated DNA can be differentiated. Thus, the preparation and cloning of cDNA is of great importance. From the cDNA sequence, essential inferences can be made about a gene and its gene product. <fn>Watson, J.D., et al.: Molecular Biology of the Gene, 3rd ed. Benjamin/Cummings Publishing Co., Menlo Park, California, 1987.</fn>

Preparation of cDNA

cDNA is prepared from mRNA. Therefore, a tissue is required in which the respective gene is transcribed and mRNA is produced in sufficient quantities. First, mRNA is isolated. Then a primer is attached so that the enzyme reverse transcriptase can form complementary DNA (cDNA) from the mRNA. Since mRNA contains poly(A) at its 3' end, a primer of poly(T) can be attached. From here, the enzyme reverse transcriptase can start forming cDNA in the 5' to 3' direction. The RNA is then removed by ribonuclease. The cDNA serves as a template for the formation of a new strand of DNA. This requires the enzyme DNA

polymerase. The result is a double strand of DNA, one strand of which is complementary to the original mRNA. To this DNA, single sequences (linkers) are attached that are complementary to the single-stranded ends produced by the restriction enzyme to be used. The same enzyme is used to cut the vector, e.g., a plasmid, so that the cDNA can be incorporated for cloning.

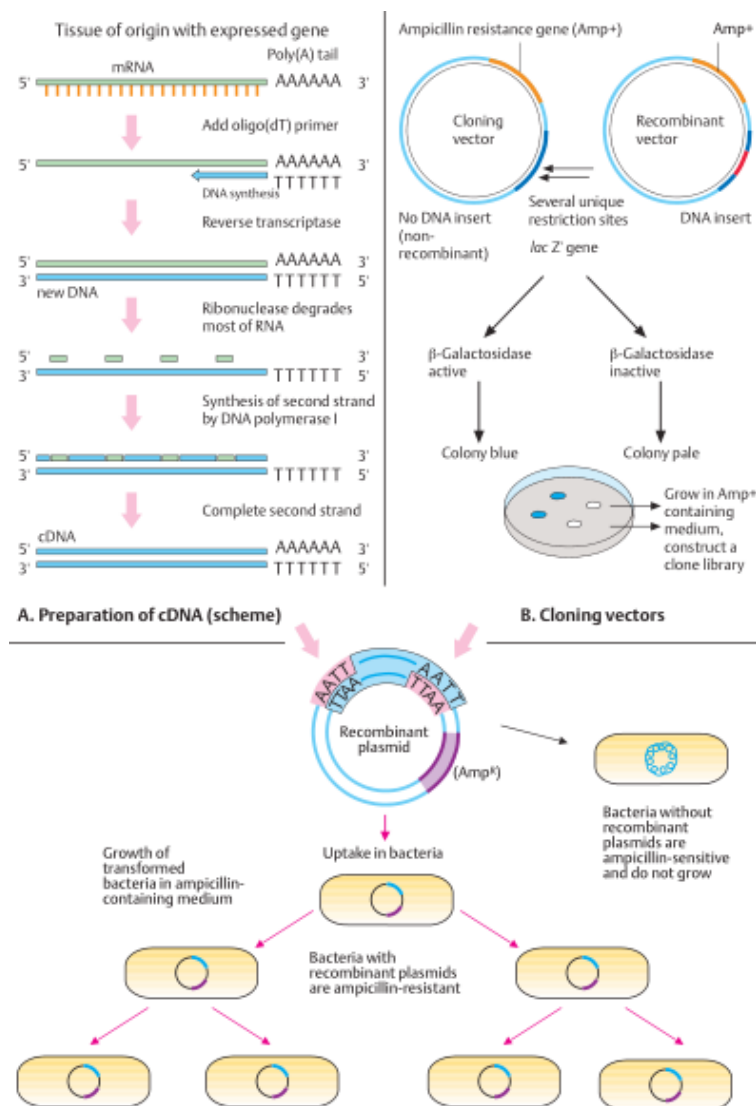
Cloning vectors

The cell-based cloning of DNA fragments of different sizes is facilitated by a wide variety of vector systems. Plasmid vectors are used to clone small DNA fragments in bacteria. Their main disadvantage is that only 5–10 kb of foreign DNA can be cloned. A plasmid cloning vector that has taken up a DNA fragment (recombinant vector), e.g., pUC8 with 2.7 kb of DNA, must be distinguished from one that has not. In addition, an ampicillin resistance gene (Amp⁺) serves to distinguish bacteria that have taken up plasmids from those that have not. Several unique restriction sites in the plasmid DNA segment where a DNA fragment might be inserted serve as markers along with a marker gene, such as the lacZ gene. The uptake of a DNA fragment by a plasmid vector disrupts the plasmid's marker gene. Thus, in the recombinant plasmid the enzyme β -galactosidase will not be produced by the disrupted lacZ gene, whereas in the plasmid without a DNA insert (non recombinant) the enzyme is produced by the still intact lacZ gene. The activity of the gene and the presence or absence of the enzyme are determined by observing a difference in color of the colonies in the presence of an artificial substrate sugar. Beta-Galactosidase splits an artificial sugar (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) that is similar to lactose, the natural substrate for this enzyme, into two sugar components, one of which is blue. Thus, bacterial colonies containing non recombinant plasmids with an intact lacZ gene are blue. In contrast, colonies that have taken up a recombinant vector remain pale white. The latter are grown in a medium containing ampicillin (the selectable marker for the uptake of

plasmid vectors). Subsequently, a clone library can be constructed. (Figure adapted from Brown, 1999).

cDNA cloning

Only those bacteria become ampicillin resistant that have incorporated a recombinant plasmid. Recombinant plasmids, which contain the gene for ampicillin resistance, transform ampicillinsensitive bacteria into ampicillin-resistant bacteria. In an ampicillin-containing medium, only those bacteria grow that contain the recombinant plasmid with the desired DNA fragment. By further replication in these bacteria, the fragment can be cloned until there is enough material to be studied. (Figures after Watson et al., 1987).



cDNA cloning principle

DNA Cloning

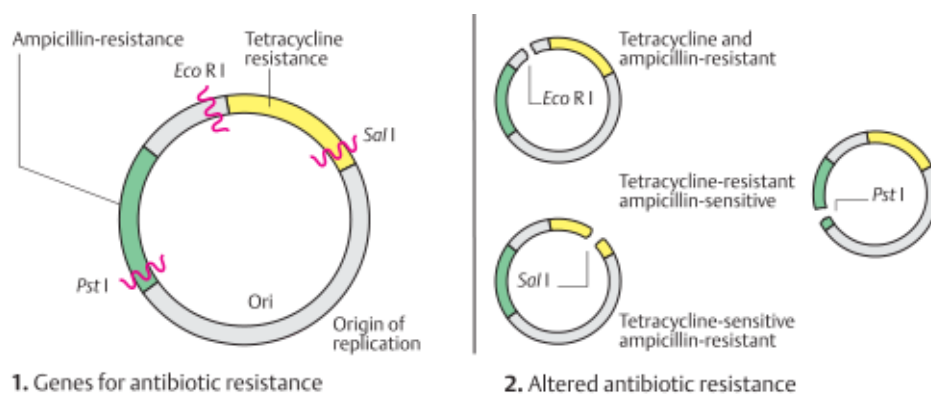
To obtain sufficient amounts of a specific DNA sequence (e.g., a gene of interest) for study, it must be selectively amplified. This is accomplished by DNA cloning, which produces a homogeneous population of DNA fragments from a mixture of very different DNA molecules or from all the DNA of the genome. Here procedures are required to identify DNA from the correct region in the genome, to separate it from other DNA, and to multiply (clone) it selectively. Identification of the correct DNA fragment utilizes the specific hybridization of complementary single-stranded DNA (molecular hybridization). A short segment of single stranded DNA, a probe, originating from the sequence to be studied, will hybridize to its complementary sequences after these have been denatured (made single-stranded, see Southern blot analysis). After the hybridized sequence has been separated from other DNA, it can be cloned. The selected DNA sequences can be amplified in two basic ways: in cells (cell-based cloning) or by cell-free cloning.

Cell-based DNA cloning

Cell-based DNA cloning requires four initial steps. First, a collection of different DNA fragments are obtained from the desired DNA (target DNA) by cleaving it with a restriction enzyme. Since fragments resulting from restriction enzyme cleavage have a short single-stranded end of a specific sequence at both ends, they can be ligated to other DNA fragments that have been cleaved with the same enzyme. The fragments produced in step 1 are joined to DNA fragments containing the origin of replication (OR) of a replicon, which enables them to replicate. In addition, a fragment may be

joined to a selectable marker, e.g., a DNA sequence containing an antibiotic resistance gene. The recombinant DNA molecules are transferred into host cells (bacterial or yeast cells). Here the recombinant DNA molecules can replicate independently of the host cell genome. Usually the host cell takes up only one (although occasionally more than one) foreign DNA molecule. The host cells transformed by recombinant (foreign) DNA are grown in culture and multiplied (propagation, 4). Selective growth of one of the cell clones allows isolation of one type of recombinant DNA molecule (5). After further propagation, a homogeneous population of recombinant DNA molecules is obtained (6). A collection of different fragments of cloned DNA is called a clone library (7, see DNA libraries). In cell-based cloning, the replicon-containing DNA molecules are referred to as vector molecules.

A plasmid vector for cloning



A plasmid vector for cloning

Many different vector systems exist for cloning DNA fragments of different sizes. Plasmid vectors are used to clone small fragments. The experiment is designed in such a way that incorporation of the fragment to be cloned changes the plasmid's antibiotic resistance to allow selection for these recombinant plasmids. A formerly frequently used plasma vector (pBR322) is presented. This plasmid contains recognition sites for the restriction enzymes PstI, EcoRI, and SalI in addition to genes for ampicillin and tetracycline resistance. If a

foreign DNA fragment is incorporated into the plasmid at the site of the EcoRI recognition sequence, then tetracycline and ampicillin resistance will be retained (2). If the enzyme PstI is used to incorporate the fragment to be used, ampicillin resistance is lost (the bacterium becomes ampicillin sensitive), but tetracycline resistance is retained. If the enzyme SalI is used to incorporate the fragment, tetracycline resistance disappears (the bacterium becomes tetracycline sensitive), but ampicillin resistance is retained. Thus, depending on how the fragment has been incorporated, recombinant plasmids containing the DNA fragment to be cloned can be distinguished from non recombinant plasmids by altered antibiotic resistance. Cloning in plasmids (bacteria) has become less important since yeast artificial chromosomes (YACs) have become available for cloning relatively large DNA fragments.

Automated DNA Sequencing

Large-scale DNA sequencing requires automated procedures based on fluorescence labeling of DNA and suitable detection systems. In general, a fluorescent label can be used either directly or indirectly. Direct fluorescent labels, as used in automated sequencing, are fluorophores. These are molecules that emit a distinct fluorescent color when exposed to UV light of a specific wavelength. Examples of fluorophores used in sequencing are fluorescein, which fluoresces pale green when exposed to a wavelength of 494 nm; rhodamine, which fluoresces red at 555 nm; and aminomethyl cumarin acetic acid, which fluoresces blue at 399 nm. In addition, a combination of different fluorophores can be used to produce a fourth color. Thus, each of the four bases can be distinctly labeled.

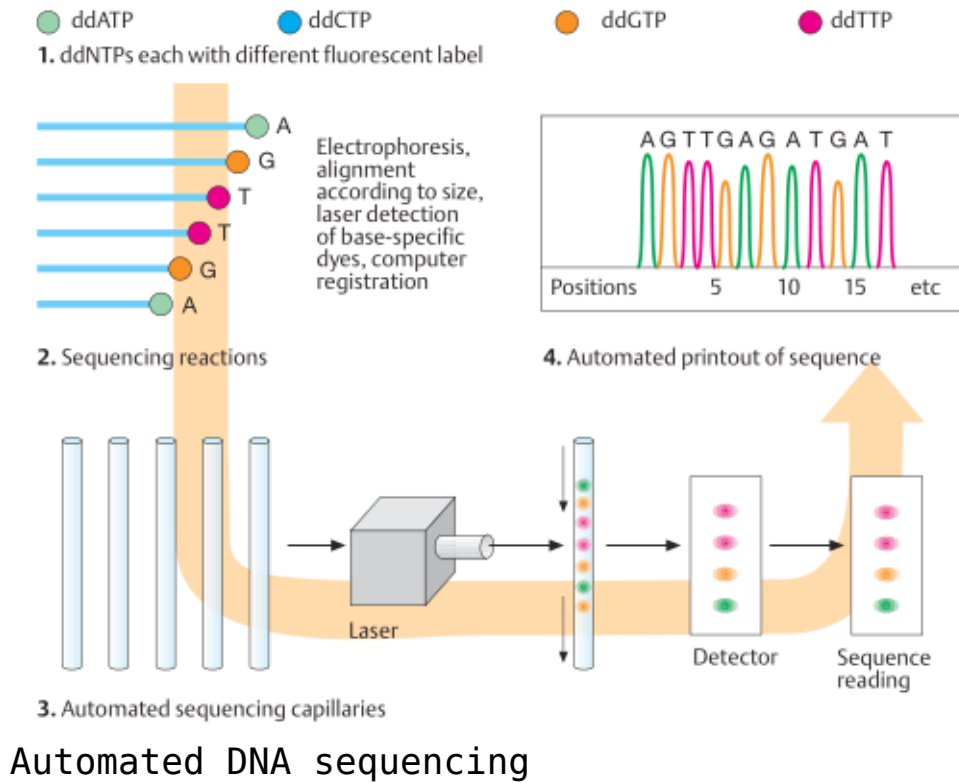
Another approach is to use PCR-amplified products (thermal cycle sequencing). This has the advantage that double-stranded rather than single-stranded DNA can be used as the starting material. And since small amounts of template DNA are sufficient, the DNA to be sequenced does not have to be cloned beforehand.

Thermal cycle sequencing

The DNA to be sequenced is contained in vector DNA <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. The primer, a short oligonucleotide with a sequence complementary to the site of attachment on the single-stranded DNA, is used as a starting point. For sequencing short stretches of DNA, a universal primer is sufficient. This is an oligonucleotide that will bind to vector DNA adjacent to the DNA to be sequenced. However, if the latter is longer than about 750 bp, only part of it will be sequenced. Therefore, additional internal primers are required. These anneal to different sites and amplify the DNA in a series of contiguous, overlapping chain termination experiments <fn>Rosenthal, N.: Fine structure of a gene—DNA sequencing. New Eng. J. Med. 332 :589–591, 1995</fn>. Here, each primer determines which region of the template DNA is being sequenced. In thermal cycle sequencing <fn>Strachan, T., Read, A.P.: Human Molecular Genetics. 2 nd ed. Bios Scientific Publishers, Oxford, 1999.</fn>, only one primer is used to carry out PCR reactions, each with one dideoxynucleotide (ddA, ddT, ddG, or ddC) in the reaction mixture. This generates a series of different chain-terminated strands, each dependent on the position of the particular nucleotide base where the chain is being terminated <fn>Wilson, R.K., et al.: Development of an automated procedure for fluorescent DNA sequencing. Genomics 6 :626–636, 1990.</fn>. After many cycles and with electrophoresis, the sequence can be read as shown in the previous plate. One advantage of thermal cycle sequencing is that double-stranded DNA can be used as starting material.

Automated DNA sequencing (principle)

Automated DNA sequencing involves four fluorophores, one for each of the four nucleotide bases. The resulting fluorescent signal is recorded at a fixed point when DNA passes through a capillary containing an electrophoretic gel. The base-specific fluorescent labels are attached to appropriate dideoxynucleotide triphosphates (ddNTP). Each ddNTP is labeled with a different color, e.g., ddATP green, ddCTP blue, ddGTP yellow, and ddTTP red <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. (The actual colors for each nucleotide may be different.) All chains terminated at an adenine (A) will yield a green signal; all chains terminated at a cytosine (C) will yield a blue signal, and so on. The sequencing reactions based on this kind of chain termination at labeled nucleotides <fn>Rosenthal, N.: Fine structure of a gene—DNA sequencing. New Eng. J. Med. 332 :589–591, 1995</fn> are carried out automatically in sequencing capillaries <fn>Strachan, T., Read, A.P.: Human Molecular Genetics. 2 nd ed. Bios Scientific Publishers, Oxford, 1999.</fn>. The electrophoretic migration of the ddNTP-labeled chains in the gel in the capillary pass in front of a laser beam focused on a fixed position. The laser induces a fluorescent signal that is dependent on the specific label representing one of the four nucleotides. The sequence is electronically read and recorded and is visualized as alternating peaks in one of the four colors, representing the alternating nucleotides in their sequence positions. In practice the peaks do not necessarily show the same maximal intensity as in the schematic diagram shown here. (Illustration based on Brown, 1999, and Strachan and Read, 1999).



Genome sequencing

Knowledge of the nucleotide sequence of a gene provides important information about its structure, function, and evolutionary relationship to other similar genes in the same or different organisms. Thus, the development in the 1970s of relatively simple methods for sequencing DNA has had a great impact on genetics. Two basic methods for DNA sequencing have been developed: a chemical cleavage method (A. M. Maxam and W. Gilbert, 1977) and an enzymatic method (F. Sanger, 1981). A brief outline of the underlying principles follows.

Sequencing by chemical degradation

This method utilizes base-specific cleavage of DNA by certain chemicals. Four different chemicals are used in four reactions,

one for each base. Each reaction produces a set of DNA fragments of different sizes. The sizes of the fragments in a reaction mixture are determined by positions in the DNA of the nucleotide that has been cleaved. A double-stranded or singlestranded fragment of DNA to be sequenced is processed to obtain a single strand labeled with a radioactive isotope at the 5' end <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. This DNA strand is treated with one of the four chemicals for one of the four reactions. Here the reaction at guanine sites(G) by dimethylsulfate (DMS) is shown. Dimethyl sulfate attaches a methyl group to the purine ring of G nucleotides. The amount of DMS used is limited so that on average just one G nucleotide per strand is methylated, not the others (shown here in four different positions of G). When a second chemical, piperidine, is added, the nucleotide purine ring is removed and the DNA molecule is cleaved at the phosphodiester bond just upstream of the site without the base. The overall procedure results in a set of labeled fragments of defined sizes according to the positions of G in the DNA sample being sequenced. Similar reactions are carried out for the other three bases (A, T, and C, not shown). The four reaction mixtures, one for each of the bases, are run in separate lanes of a polyacrylamide gel electrophoresis. Each of the four lanes represents one of the four bases G, A, T, or C. The smallest fragment will migrate the farthest downward, the next a little less far, etc. One can then read the sequence in the direction opposite to migration to obtain the sequence in the 5' to 3' direction (here TAGTCGCAGTACCGTA).

Sequencing by chain termination

This method, now much more widely used than the chemical cleavage method, rests on the principle that DNA synthesis is terminated when instead of a normal deoxynucleotide (dATP, dTTP, dGTP, dCTP), a dideoxynucleotide (ddATP, ddTTP, ddGTP, ddCTP) is used. A dideoxynucleotide (ddNTP) is an analogue of

the normal dNTP. It differs by lack of a hydroxyl group at the 3' carbon position. When a dideoxynucleotide is incorporated during DNA synthesis, no bond between its 3' position and the next nucleotide is possible because the ddNTP lacks the 3' hydroxyl group. Thus, synthesis of the new chain is terminated at this site. The DNA fragment to be sequenced has to be single-stranded <fn>Brown, T.A.: Genomes. Bios Scientific Publ., Oxford, 1999.</fn>. DNA synthesis is initiated using a primer and one of the four ddNTPs labeled with ^{32}P in the phosphate groups or, for automated sequencing, with a fluorophore (see next plate). Here an example of chain termination using ddATP is shown <fn>Strachan, T., Read, A.P.: Human Molecular Genetics. 2nd ed. Bios scientific Publishers,</fn>. Wherever an adenine (A) occurs in the sequence, the dideoxyadenine triphosphate will cause termination of the new DNA chain being synthesized. This will produce a set of different DNA fragments whose sizes are determined by the positions of the adenine residues occurring in the fragment to be sequenced. Similar reactions are done for the other three nucleotides. The four parallel reactions will yield a set of fragments with defined sizes according to the positions of the nucleotides where the new DNA synthesis has been terminated. The fragments are separated according to size by gel electrophoresis as in the chemical method. The sequence gel is read in the direction from small fragments to large fragments to derive the nucleotide sequence in the 5' to 3' direction. An example of an actual sequencing gel is shown between panel A and B.

Eukaryotic gene structure

Eukaryotic genes consist of coding and noncoding segments of DNA, called exons and introns, respectively. At first glance it

seems to be an unnecessary burden to carry DNA without obvious functions within a gene. However, it has been recognized that this has great evolutionary advantages. When parts of different genes are rearranged on new chromosomal sites during evolution, new genes may be constructed from parts of previously existing genes.

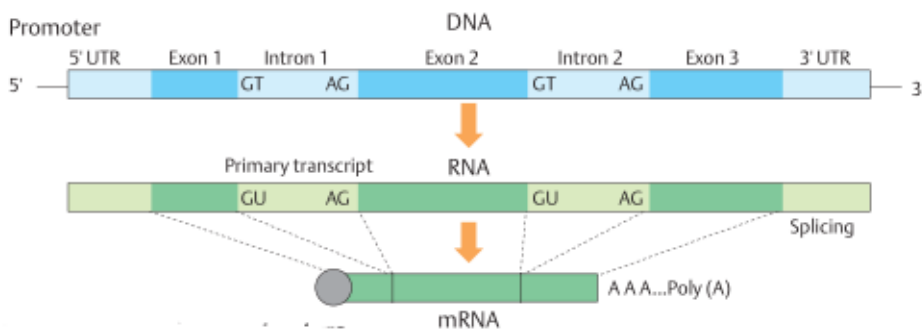
Exons and introns

In 1977, it was unexpectedly found that the DNA of a eukaryotic gene is longer than its corresponding mRNA. The reason is that certain sections of the initially formed primary RNA transcript are removed before translation occurs. Electron micrographs show that DNA and its corresponding transcript (RNA) are of different lengths (1). When mRNA and its complementary single-stranded DNA are hybridized, loops of single-stranded DNA arise because mRNA hybridizes only with certain sections of the single stranded DNA. In (2), seven loops (A to G) and eight hybridizing sections are shown (1 to 7 and the leading section L). Of the total 7700 DNA base pairs of this gene (3), only 1825 hybridize with mRNA. A hybridizing segment is called an exon. An initially transcribed DNA section that is subsequently removed from the primary transcript is an intron. The size and arrangement of exons and introns are characteristic for every eukaryotic gene (exon/intron structure). (Electron micrograph from Watson et al., 1987).

Intervening DNA sequences (introns)

In prokaryotes, DNA is colinear with mRNA and contains no introns (1). In eukaryotes, mature mRNA is complementary to only certain sections of DNA because the latter contains introns (2). (Figure adapted from Stryer, 1995).

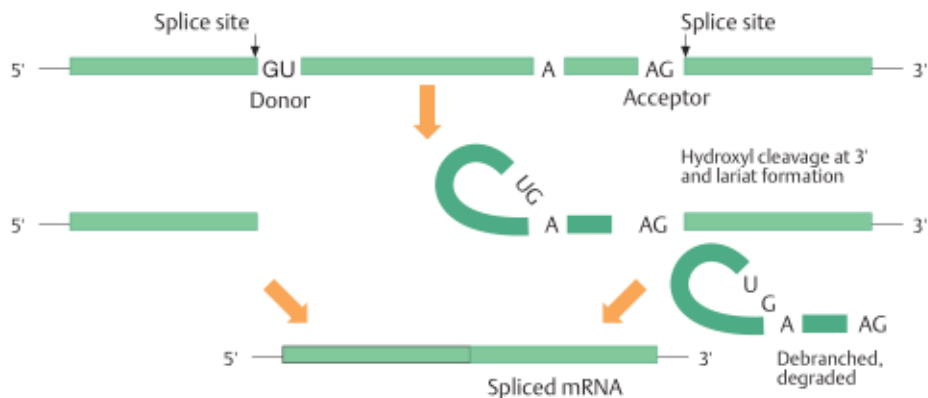
Basic eukaryotic gene structure



Basic eukaryotic gene structure

Exons and introns are numbered in the 5' to 3' direction of the coding strand. Both exons and introns are transcribed into a precursor RNA (primary transcript). The first and the last exons usually contain sequences that are not translated. These are called the 5' untranslated region (5' UTR) of exon 1 and the 3' UTR at the 3' end of the last exon. The non coding segments (introns) are removed from the primary transcript and the exons on either side are connected by a process called splicing. Splicing must be very precise to avoid an undesirable change of the correct reading frame. Introns almost always start with the nucleotides GT in the 5' to 3' strand (GU in RNA) and end with AG. The sequences at the 5' end of the intron beginning with GT are called splice donor site and at the 3' end, ending with AG, are called the splice acceptor site. Mature mRNA is modified at the 5' end by adding a stabilizing structure called a "cap" and by adding many adenines at the 3' end (polyadenylation).

Splicing pathway in GU–AG introns



Splicing pathway in GU – AG introns

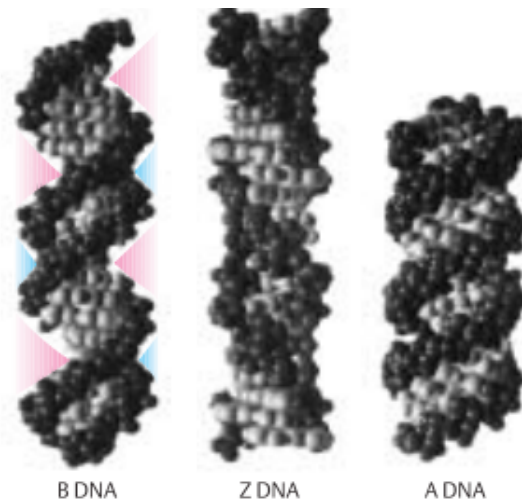
RNA splicing is a complex process mediated by a large RNA-containing protein called a spliceosome. This consists of five types of small nuclear RNA molecules (snRNA) and more than 50 proteins (small nuclear riboprotein particles). The basic mechanism of splicing schematically involves autocatalytic cleavage at the 5' end of the intron resulting in lariat formation. This is an intermediate circular structure formed by connecting the 5' terminus (UG) to a base (A) within the intron. This site is called the branch site. In the next stage, cleavage at the 3' site releases the intron in lariat form. At the same time the right exon is ligated (spliced) to the left exon. The lariat is debranched to yield a linear intron and this is rapidly degraded. The branch site identifies the 3' end for precise cleavage at the splice acceptor site. It lies 18–40 nucleotides upstream (in 5' direction) of the 3' splice site. (Figure adapted from Strachan and Read, 1999)

Alternative DNA Structures

Gene expression and transcription can be influenced by changes of DNA topology. However, this type of control of gene expression is relatively universal and non specific. Thus, it

is more suitable for permanent suppression of transcription, e.g., in genes that are expressed only in certain tissues or are active only during the embryonic period and later become permanently inactive.

A. Three forms of DNA



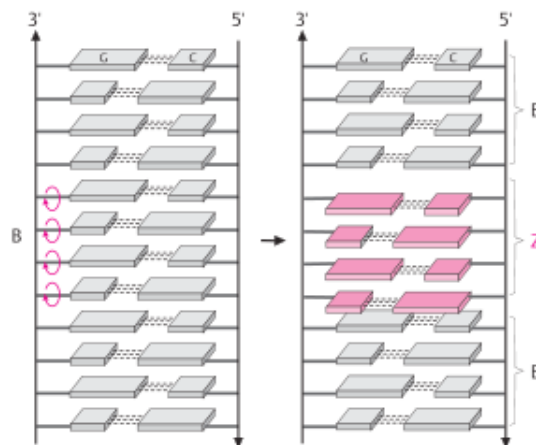
Three forms of DNA

The DNA double helix does not occur as a single structure, but rather represents a structural family of different types. The original classic form, determined by Watson and Crick in 1953, is B-DNA. The essential structural characteristic of B-DNA is the formation of two grooves, one large (major groove) and one small (minor groove). There are at least two further, alternative forms of the DNA double helix, Z-DNA and the rare form A-DNA. While B-DNA forms a right-handed helix, Z-DNA shows a left-handed conformation. This leads to a greater distance (0.77nm) between the base pairs than in B-DNA and a zig zag form (thus the designation Z-DNA). A-DNA is rare. It exists only in the dehydrated state and differs from the B form by a 20-degree rotation of the perpendicular axis of the helix. A-DNA has a deep major groove and a flat minor groove (Figures from Watson et al, 1987).

B. Major and minor grooves in B-DNA

The base pairing in DNA (adenine–thymine and guanine–cytosine) leads to the formation of a large and a small groove because the glycosidic bonds to deoxyribose (dRib) are not diametrically opposed. In B-DNA, the purine and pyrimidine rings lie 0.34 nm apart. DNA has ten base pairs per turn of the double helix. The distance from one complete turn to the next is 3.4 nm. In this way, localized curves arise in the double helix. The result is a somewhat larger and a somewhat smaller groove. <fn>Stryer, L.: Biochemistry, 4 th ed. W.H. Freeman & Co., New York, 1995.</fn>

C. Transition from B-DNA to Z-DNA



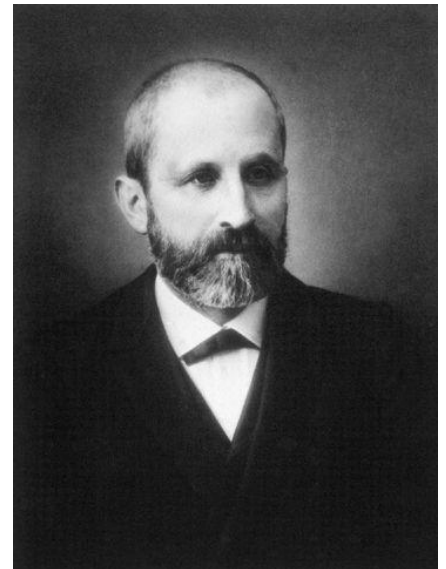
Transition from B-DNA to Z-DNA

B-DNA is a perfect regular double helix except that the base pairs opposite each other do not lie exactly at the same level. They are twisted in a propeller-like manner. In this way, DNA can easily be bent without causing essential changes in the local structures. In Z-DNA the sugar–phosphate skeleton has a zigzag pattern; the single Z-DNA groove has a greater density of negatively charged molecules. Z-DNA may occur in limited segments in vivo. A segment of B-DNA consisting of GC pairs can be converted into Z-DNA when the bases are rotated

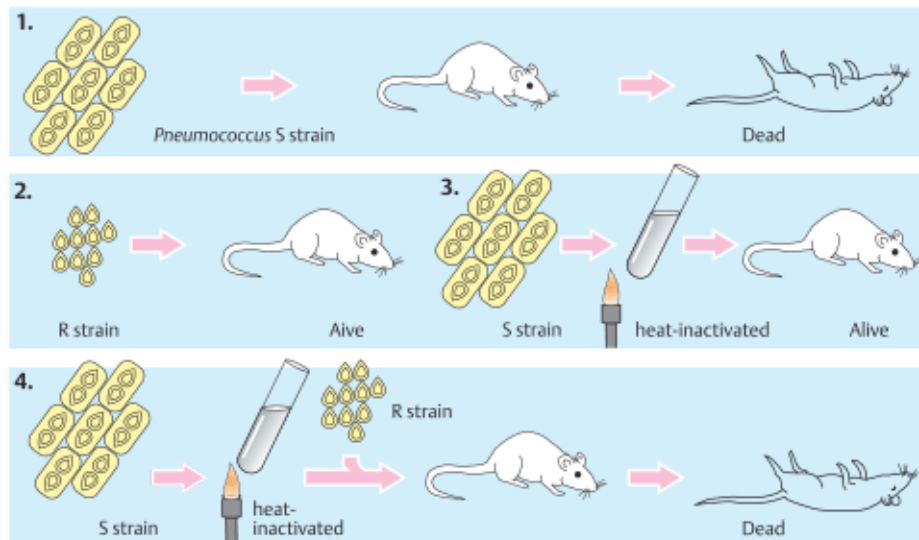
180 degrees. Normally, Z-DNA is thermodynamically relatively unstable. However, transition to Z-DNA is facilitated when cytosine is methylated in position 5 (C5). The modification of DNA by methylation of cytosine is frequent in certain regions of DNA of eukaryotes. There are specific proteins that bind to Z-DNA, but their significance for the regulation of transcription is not clear. <fn>Watson, J.D. et al.: Molecular Biology of the Gene. 3rd ed. Benjamin/Cummings Publishing Co., Menlo Park, California, 1987.</fn>

DNA as Carrier of Genetic Information

Although DNA was discovered in 1869 by Friedrich Miescher as a new, acidic, phosphorus containing substance made up of very large molecules that he named "nuclein", its biological role was not recognized. In 1889 Richard Altmann introduced the term "nucleic acid". By 1900 the purine and pyrimidine bases were known. Twenty years later, the two kinds of nucleic acids, RNA and DNA, were distinguished. An incidental but precise observation (1928) and relevant investigations (1944) indicated that DNA could be the carrier of genetic information.



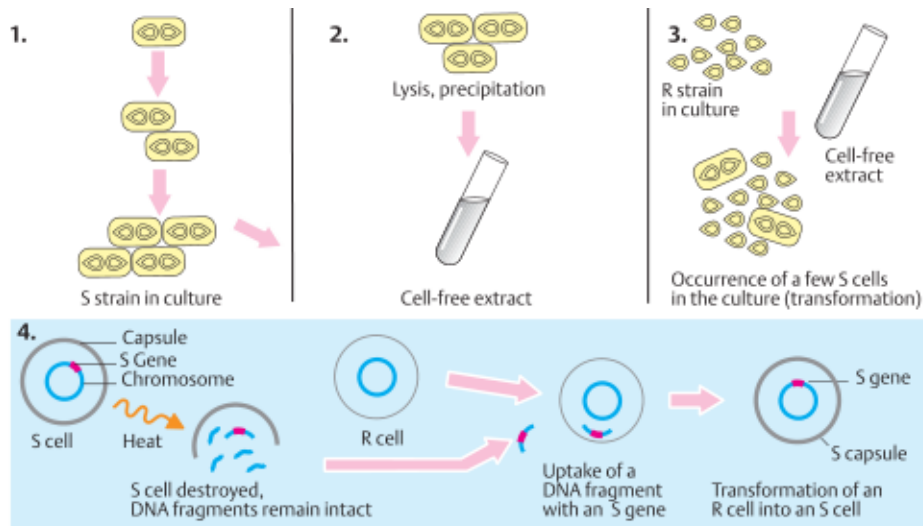
A. The observation of Griffith



The observation of Griffith

In 1928 the English microbiologist Fred Griffith made a remarkable observation. While investigating various strains of Pneumococcus, he determined that mice injected with strain S (smooth) died (1). On the other hand, animals injected with strain R (rough) lived (2). When he inactivated the lethal S strain by heat, there were no sequelae, and the animal survived (3). Surprisingly, a mixture of the nonlethal R strain and the heat-inactivated S strain had a lethal effect like the S strain (4). And he found normal living pneumococci of the S strain in the animal's blood. Apparently, cells of the R strain were changed into cells of the S strain (transformed). For a time, this surprising result could not be explained and was met with skepticism. Its relevance for genetics was not apparent.

B. The transforming principle is DNA



The transforming principle is DNA

Griffith's findings formed the basis for investigations by Avery, MacLeod, and McCarty (1944). Avery and co-workers at the Rockefeller Institute in New York elucidated the chemical basis of the transforming principle. From cultures of an S strain (1) they produced an extract of lysed cells (cell-free extract) (2). After all its proteins, lipids, and polysaccharides had been removed, the extract still retained the ability to transform pneumococci of the R strain to pneumococci of the S strain (transforming principle) (3). With further studies, Avery and co-workers determined that this was attributed to the DNA alone. Thus, the DNA must contain the corresponding genetic information. This explained Griffith's observation. Heat inactivation had left the DNA of the bacterial chromosomes intact. The section of the chromosome with the gene responsible for capsule formation (S gene) could be released from the destroyed S cells and be taken up by some R cells in subsequent cultures. After the S gene was incorporated into its DNA, an R cell was transformed into an S cell(4).

C. Genetic information is

transmitted by DNA alone

The final evidence that DNA, and no other molecule, transmits genetic information was provided by Hershey and Chase in 1952. They labeled the capsular protein of bacteriophages (see p. 88) with radioactive sulfur (^{35}S) and the DNA with radioactive phosphorus (^{32}P). When bacteria were infected with the labeled bacteriophage, only ^{32}P (DNA) entered the cells, and not the ^{35}S (capsular protein). The subsequent formation of new, complete phage particles in the cell proved that DNA was the exclusive carrier of the genetic information needed to form new phage particles, including their capsular protein. Next, the structure and function of DNA needed to be clarified. The genes of all cells and some viruses consist of DNA, a long-chained threadlike molecule.

