Milestones of Molecular Biology

The molecularization of biology since the middle of the 20th century has had a huge impact on research as well as on everyday applications. It consists of a synthesis of many biological, biochemical and medical disciplines. Whether in the diagnosis of diseases, the production of detergents, chemical basic materials or foodstuffs, biotechnological processes determine modern everyday life.

The advancement of knowledge made possible by the development of molecular biology has enriched the knowledge about the function of cells and of the whole organism in a revolutionary manner, and one begins to understand the complexity and diversity of the regulatory processes at the molecular level and to recognize the molecular structures involved.

In the following, some examples of important developmental steps and the scientists involved are presented.

The 1950s

In 1944, US American Oswald Avery hypothesized, based on transformation experiments with bacteria, that deoxyribonucleic acids (DNA) and not proteins are the carriers of the genetic material. In 1952, this hypothesis was finally confirmed by the observations made by Alfred Hershey and Martha Chase on the role of DNA in the infection of bacteria by phages. This was the start signal for an unprecedented scientific race to elucidate the molecular structure of the DNA.

The Englishman Francis H. Crick and the US American James D. Watson, who were both doing research in Cambridge in England, began in 1951 to work feverishly in order to decipher the structure of the DNA by means of the existing findings on its composition (Fig. 1). It was clear to them that they could be sure of receiving the Nobel Prize if they succeeded in puzzling out the structure of the DNA.

In 1953, they were able to decode the three-dimensional structure of the DNA, based on X-ray structure analysis of crystallized DNA performed by Rosalind Franklin and Maurice Wilkins at King's College, London, in the early 1950s. The X-ray diffraction images prepared by Franklin and Wilkins indicated not only a helical DNA structure, but also more than one polynucleotide strand per DNA molecule, either two or three strands.
In addition, the results allowed the conclusion that the readily water-soluble phosphate residues point outwards and the poorly water-soluble bases must lie inside the DNA. This was the only way to explain the overall good water-soluble property of the DNA. The structural model developed by Watson and Crick remains valid until today. They called this structure the DNA double helix. For their discovery, Watson and Crick received the Nobel Prize in Medicine and Physiology in 1962, together with Maurice Wilkins, who contributed the radiographic images. Rosalind Franklin, however, could not be honored by the Nobel Jury. The chemist had died four years before.

Other milestones

In 1955, Arthur Kornberg isolated the first DNA polymerase. These enzymes play a key role in DNA replication. For this discovery, he received the Nobel Prize in Physiology or Medicine together with Severo Ochoa in 1959.

In 1958, Francis Crick formulated the "sequence hypothesis" and the "central dogma" of molecular biology. According to the sequence hypothesis, the specificity of the nucleic acids lies exclusively in the sequence of their bases, which in turn determines the amino acid sequence of the proteins. The central dogma stated that the molecular information flows from the DNA to the proteins via the RNA and that the reverse path is excluded.

In 1959, Francois Jacob and Matthew Meselson showed that protein biosynthesis takes place on ribosomes.

1960s, the genetic code

After Watson and Crick had determined the structure of the DNA, scientists started thinking about how a sequence of nucleotides can be translated into a sequence of amino acids. On the basis of genetic investigations on phage mutants, Sydney Brenner and Francis Crick in Cambridge, England, in 1961 came to the important
conclusion that each amino acid coding element (codon) - that is the sequence of the base building blocks - consists of three bases (nucleotides).
In 1961, the messenger RNA (mRNA) was discovered by Matthew S. Meselson, François Jacob, and Sydney Brenner. The messenger molecules read the information from the DNA in the cell nucleus and are transported into the cytoplasm where the proteins are formed.
In 1961, the discovery that ribosomes are inherently unspecific and are programmed to synthesize specific proteins only by attached mRNA molecules prompted US American Marshall Nirenberg (Fig. 2), scientist at the NIH in Bethesda, and his German post-doctoral fellow Heinrich Matthaei to perform a historical experiment.
The two of them synthesized short mRNA pieces with known base sequences. In a test tube, they added everything that was necessary for the biological synthesis of proteins according to the state of knowledge at that time: all 20 amino acids, the protein factories of the cell - the ribosomes - and a synthetic mRNA. Each time, one of the 20 amino acids used was radioactively tagged by Matthaei and Nirenberg. After the ribosomes had had enough time to perform their "duty", they filtered them out together with the bound amino acids. The first mRNA they used had the base sequence UUU - hence the name of the experiment. It thus consisted exclusively of uracil, the base which occurs in the mRNA instead of thymine. The radioactive signal was found on the filter only when the amino acid phenylalanine was radiolabeled. Phenylalanine must therefore be bound to the ribosomes. If other amino acids were tagged, the filtrate was radioactive, but not the filter. This could only mean that the base triplet UUU encodes the amino acid phenylalanine.
In another experiment, an mRNA with the base sequence UCU was added to the system. Now radioactive serine was found on the filter (i.e., polyserine was synthesized), while radioactive phenylalanine was found in the filtrate (i.e., there was no synthesis of polyphenylalanine).
By 1966, all codons had been assigned and the code decrypted, including the three stop codons, which terminated the respective protein synthesis. In 1968, the Nobel Prize in Physiology or Medicine was given to Nirenberg, to Har Gobind Khorana, who had helped elucidate much of the code with his RNA synthesis methods, and to Robert Holley, who was the first to identify the sequence of a tRNA and to find the anti-codon in it. This scientific milestone finally made it possible to understand the processes of protein biosynthesis at the molecular level. Thus it forms the cornerstone for today's rapidly evolving genetic engineering.
In 1962, Werner Arber discovered the restriction endonucleases. With their help it was possible to cut isolated DNA into sections of a defined length since these enzymes can recognize specific sequence sections.
1970s, DNA sequencing

Until the early 1970s, DNA was the most difficult molecule to be analyzed for biochemists. The development of powerful, automated DNA sequencing methods is therefore one of the milestones in the development of gene technology. In the years 1975 to 1977, Frederick Sanger (Fig. 3) and co-workers developed the chain termination method, and Allan Maxam and Walter Gilbert developed chemical sequencing. Sanger and Gilbert received the Nobel Prize in Chemistry in 1980 for the development of their respective sequencing methods.

In the Sanger method, a labeled complementary DNA strand is synthesized in vitro with the use of so-called dideoxynucleotides that lead to a random chain termination and hence to DNA strands of different lengths. The resulting fragments are separated by polyacrylamide electrophoresis and evaluated in an autoradiogram. Newer methods work with fluorescence-labeled nucleoside triphosphates in which the fragments are determined in a fluorescence detector and visualized by means of colored peak diagrams.

In the method developed by Maxam and Gilbert, a DNA fragment is radioactively labeled at one end, denatured, and cleaved at specific bases via a chemical method. The cleavage products of different lengths are then separated by polyacrylamide gel electrophoresis for each base. The base sequence can be identified from the autoradiogram of the four lanes.

The chain termination method with dideoxynucleotides prevailed mainly due to its ability for automation, the quality of the sequences and the longer sequence reads. It is used above all in automatic sequencing, in which DNA fragments are labeled with fluorescent dyes. It was only by DNA sequencing that the era of genome research was initiated. Still, the isolation of the DNA and the cloning for multiplication of the test material were complex steps.

Other Milestones

In 1973, selective cloning of genes became possible thanks to the breakthrough experiments of Herbert Boyer, Stanley Cohen, Paul Berg and colleagues at Stanford University and at the University of California in San Francisco.

In 1975, Edwin Southern developed gel transfer hybridization to detect specific DNA sequences. In the same year, César Milstein, Georges Köhler and Niels Jerne published a principle for producing monoclonal antibodies. For their work, the three researchers received the Nobel Prize in Medicine in 1984.

1980s, Polymerase Chain Reaction
In 1985, Kary B. Mullis developed the concept of the polymerase chain reaction (PCR) during a boring car trip on a spring weekend. This method was to provide the solution to one of the most pressing biological problems of its time - the duplication of DNA. Whereas in the past large amounts of DNA had to be isolated and purified from tissue cells, bacteria or viruses in a complex procedure, a molecule can be copied by means of PCR in a repetitive process until sufficient material is available for further analysis. In this way, the genetic material can be multiplied so much that verifiable amounts are obtained even from minute samples.

However, when Mullis, who was working for Cetus Pharmaceuticals at the time, carried out the first PCR experiments, he also used the enzyme DNA polymerase from enterobacteria, which, like most proteins, is destroyed at temperatures above 45°C. Since each PCR cycle begins with a heating step, in each duplication cycle, after the sample had been heated, another enzyme had to be added for the multiplication. Finally, Mullis had the brilliant and strikingly simple idea of using a DNA polymerase isolated from thermophilic bacteria, which are not harmed by the DNA denaturation process. The use of these heat-stable polymerases, which are still active at temperatures of approximately 100°C, greatly simplified the hitherto cumbersome procedure. In the years 1985 to 1989, Mullis further developed the PCR technique and received the Nobel Prize in Chemistry together with Michael Smith in 1993. Since then, an almost unmanageable variety of PCR variants has been developed.

Other Milestones

In 1982, the first recombinant drug, human insulin produced by Genentech, was approved.

In 1986, the first genome sequencer came onto the market. The device presented by Applied Biosystems made the previous use of radioactivity unnecessary in sequencing. The sequence was determined by the detection of fluorophores specific for each base.

1990s, The Human Genome Project (HGP)

In 1990, a project was launched, the extent of which was supposed to dwarf everything that had previously existed in the field of biosciences: the mapping of the entire genetic information of humans.

After first establishing the Human Genome Organization (HUGO) as an independent association of scientists and genome research institutes in 1988 to coordinate the work groups distributed all over the world, the Human Genome Project (HGP) began its work in 1990 as a public, mainly American large-scale research project.
This quickly resulted in a loose network of national genomic research projects in more than 30 different countries. Approximately 60% of the work was done by various sequencing centers in the USA. The British Sanger Center accounted for a quarter of the task. The remaining sequences were mainly examined by genome researchers from France, Japan, China and Germany. By 2005 the work should be done, so the plan.

After French and American scientists had published a complete map of the human genome in 1994, a year earlier than planned, HUGO and the US company Celera Genomics announced the decoding of the human genome: the blueprint of the human body, an exact sequence of 3.2 billion gene letters. Founded by Craig Venter, Celera Genomics had joined the race in 1998 as a commercial competitor to the HGP. In the end, however, the project could only be successfully completed by cooperating with HUGO. The total cost of the HGP was around $3 billion, or about $1 per base pair.

Other Milestones

With quantitative PCR (qPCR), the kinetic analysis of product formation during amplification, Higuchi and co-workers succeeded in developing a significant advancement of PCR in 1993. The qPCR is a multiplication method for nucleic acids, which also offers the possibility of quantification in real time, so it is also referred to as "real-time PCR". The quantification is carried out with the aid of fluorescence measurements at the end of or during a PCR cycle. The fluorescence results from the incorporation of the dye into a DNA double strand, therefore increasing proportionally with the amount of the PCR products and making quantification possible. Since the amount of product directly correlates with the amount of DNA used, the qPCR allows calculation of the initial concentration of the nucleic acid molecule to be examined.

qPCR is considered to be one of the most sensitive and most versatile methods in DNA quantification. However, their sensitivity is also the cause of methodological problems. Inaccurate pipetting is the most common cause of conflicting results. In case of serial dilutions of small volumes and several pipetting steps, the error rate is particularly high. The first real-time PCR device was launched in the second half of the 1990s.

In 1991, Hood and Hunkapillar introduced a new automated DNA sequencing technique.

In 1996, Ian Wilmut and associated researchers from the Scottish Roslin Institute produced the first cloned mammal - the sheep Dolly (Fig. 5).

In 1996, David J. Lockhart and colleagues as well as Joseph DeRisi and colleagues presented DNA microarrays, which allowed the simultaneous examination of
thousands of genes. In a so-called biochip, a great deal of information (in this case several thousand proteins, RNA or DNA fragments) is present on a very small space. Thus, even with a small amount of biological sample material, a large number of molecular biology investigations can be performed simultaneously, and for the most part automated.

In 1998, US researcher James Thomson succeeded in isolating stem cells from human embryos and cultivating them as embryonic stem cells (ES cells) in the laboratory. The ES cells can be transformed into almost all human cell types. Thomson thus created an important experimental base for stem cell research and regenerative medicine.

In 1998 Andrew Fire and Craig Mello discovered the method of RNA interference: genes are silenced with small RNA bits.

**The 2000s, Next Generation -Sequencing (NGS)**

For a long time, DNA sequencing developed at a very leisurely pace. Even in the late 1980s, the chain termination method presented by Sanger in 1977 had to be handled with huge, unstable sequencing gels and painstakingly evaluated bands. Finally, the capillary electrophoresis sequencers introduced in 1990, which automated the Sanger sequencing (CE sequencing), simplified the method. Modern CE sequencers achieve read lengths of about 1000 base pairs and read the DNA sequence very precisely. To this day their accuracy is therefore considered a gold standard. However, they can perform only one read operation at a time and therefore work very slowly.

In 2005, new DNA sequencing techniques were introduced called "Next Generation Sequencing" (NGS) techniques. Most ultra-high-throughput methods no longer use a separation of the DNA via capillary electrophoresis, as in the Sanger method, but a coupling of molecules to surfaces and recordings of rows of high-resolution images. These innovative technologies have the potential to revolutionize biological and biomedical research by significantly accelerating genome analyses and by reducing their costs.

However, due to technical limitations, a genome cannot be read in a single approach from the beginning to the end in a linear reaction, even by applying the novel sequencing techniques. It has to be subdivided into smaller pieces during sequencing. However, the read lengths of all commercially available NGS devices are considerably shorter than those of the Sanger sequencing method. After sequencing the pieces therefore have to be combined to a complete genome using bioinformatics.

Still, NGS has some limitations: its accuracy is less than that of the Sanger method. Moreover, the equipment for NGS is still very expensive. This is due to the high
equipment prices and the extremely complex data analysis which have to be managed and processed in specially equipped data centers. But these problems are of a technical nature and they will probably be solved in the foreseeable future and with more competition between manufacturers. Compared to the former methods, however, the modern ones are extremely cheap per base pair and a perfect example of the rapid development of molecular biology: the sequencing of a human genome. In 2005 it needed 10 years and $ 3,000,000,000, today it is one day and $ 1,000.

**More Milestones**

In 2002, Eckhard Wimmer synthesized the first complete genome of a poliovirus in a test tube. The work is considered to be a milestone for the emerging field of synthetic biology.

In 2006, Shinya Yamanaka and fellow Japanese researchers, succeeded for the first time in bringing back differentiated skin cells from mice into the state of a pluripotent cell, that is, they behaved like embryonic stem cells. The artificially generated result they called induced pluripotent stem cells (iPS cells).

In 2007, Japanese researchers succeeded in achieving the same rejuvenating cure with human skin cells. Shinya Yamanaka's work and findings were so influential in stem cell research that the Japanese researcher was honored with the Nobel Prize in Medicine as early as in 2012.

**2010s, DNA Sequencing**

In 2010, a new sequencing method based on semiconductor technology was presented. The DNA to be sequenced is deposited in micro-reaction chambers on a semiconductor chip. These reaction chambers contain the DNA polymerase, the various nucleotides are added sequentially. The method follows a sequencing-by-synthesis approach in that a DNA template is complemented by sequential nucleotide incorporation. However, the incorporated nucleotides are not detected through an optical signal. The incorporation of a nucleotide involves the forming a covalent bond to release a pyrophosphate and a positively charged hydrogen ion. The released hydrogen ions are detected by a change in the pH value of an ion-sensitive layer under the reaction chambers.

In 2012, the so-called single-molecule sequencers were ready for use. Sequencing techniques based on nanostructures or nanopores make it possible to read individual nucleic acid molecules, base by base. Used massively in parallel on chips, high throughput rates can be achieved and costs saved.

**Other Milestones**
In 2010, Hamilton Smith and colleagues succeeded in incorporating a synthetic genome into a bacterium. The bacterium with the transplanted genome develops and passes it on to its daughter cells.

Outlook

The dynamic development in molecular biology research is still far from complete. Existing methods will be further refined and new procedures developed. New microscopic, spectroscopic and biotechnological methods will continue to gain importance at the same pace for quite some time. So far, especially in biology and medicine, only the surface has been touched. We are still far away from a deeper understanding of the complex processes that take place in the body of a living being. There is still much to discover. The G.I.T. Laboratory Journal will continue to accompany this exciting development and will keep you, dear readers, informed about new technologies and the research done with them.

Authors

Ralf Kempf