1.3 Biomolecules in Analytical Chemistry

Chemical analysis of the biomolecules introduced in the preceding sections is often radically different from the "classical" analysis of relatively small organic molecules such as pesticides, drugs or explosives residues.

The analytical chemist is faced with a variety of tasks, which can be roughly separated into four categories: (1) qualitative analysis of a mixture of compounds, (2) qualitative analysis of a pure compound, (3) quantitative analysis of a selected compound in a mixture and (4) structure elucidation of a pure compound.

1.3.1 *Classical Analytical Chemistry*

A number of methods are used in classical analysis to perform these tasks. Qualitative as well as quantitative analysis of mixtures can be achieved by chromatographic methods such as gas chromatography (GC) and liquid chromatography (LC). Chemical sensors or biosensors can also be employed for selectively quantifying a compound in a mixture. However, such analysers have only been developed for a very limited number of analytes. Identification of pure compounds can be achieved by nuclear magnetic resonance (NMR) measurements, by mass spectrometry (MS), infrared spectroscopy (IR), UV/vis spectroscopy or X-ray crystallography, to name a few.

1.3.2 *Limitations of Classical Analytical Chemistry*

Most of these "classical" methods are not, or only to a limited extent, suitable for the analysis of nucleic acids and proteins. GC requires the analytes to be volatile and thermally stable, a property rarely exhibited by a biomolecule. MS is often used for molecular weight determination. However, the high molecular weight biomolecules are extensively fragmented by conventional ionisation methods, making molecular weight (MW) analysis impossible. NMR-spectroscopy works extremely well for structural elucidation of medium sized molecules. However, the sheer number of spin-active nuclei, especially ^{1}H, in any single DNA or protein molecule results in an enormous number of signals that make structural elucidation extremely difficult if not impossible. A similar problem holds true for structural elucidation via IR- and UV-spectroscopy. The number of functional groups and chromophores makes unambiguous identification impossible.

Biochemists are often interested in parameters that are not applicable for most classical analytes. Nucleic acids and proteins are both large biopolymers, consisting of sometimes thousands of monomers linked together. Methods are required to accurately determine the *sequence* of amino acids in a protein or the sequence

of bases in a DNA or RNA molecule. Changing a single amino acid in the protein might alter its folding pattern and biological activity. Changing a single base in a DNA strand might cause a genetic disorder. Not only must the sequencing methods be accurate, they should also have potential for automation, if large amounts of samples are to be analysed and compared. *Separation* methods for biomolecules must be extremely powerful. The number of proteins in a single cell can run into thousands. For sequencing, the biomolecules are often partially digested into smaller fragments, which have to be separated from each other. Separation methods are needed on both, preparative and analytical scales. The amount of sample available is often very small. Hence, high sensitivity is required. Very sophisticated methods are needed to elucidate the three-dimensional structure of a biomolecule with a molecular weight of possibly thousands of kDa.

1.3.3 *Bioanalytical Chemistry*

Without the development and improvement of bioanalytical methods over the recent decades, the enormous progress in genomics and proteomics would have been impossible. Methods for accurate determination of high molecular weights, for sequencing of DNA and proteins and for separating thousands of molecules in a single run have revolutionised analytical chemistry. Many of these methods have been transformed into commercially available bench top instruments that offer high-throughput, automated and computer controlled analyses. In Table 1.6, a number of methods employed for nucleic acid and protein analysis are summarised. The list is by no means exhaustive. It is intended to give an overview and to emphasise the difference between classical analysis and bioanalysis.

1.3.3.1 *Analysis of nucleic acids*

The field of *genomics* concerns the study of the entire genome of a cell or an organism, i.e. the complete DNA sequence, and the determination of all the genes within that sequence. The human genome consists of 3 billion base pairs. An estimated 30,000-40,000 genes are contained within this sequence. Scientists are trying to map the locations of genes to improve our understanding of genetic disorders, and to explore the organisation and interplay of these genes.

Before any sequencing reactions can be carried out, the nucleic acids must be isolated from the cell and purified. To obtain a sufficient amount of sample, amplification of the DNA molecules is usually required. Analysis of nucleic acids can, thus, be divided into the following steps: (1) isolation and purification, (2) quantification and amplification and (3) sequencing.

Table 1.6. Comparison of classical and bioanalytical chemistry.

<table>
<tr><th rowspan="2">Analytical task</th><th>Classical analytical chemistry</th><th colspan="2">Bioanalytical chemistry</th></tr>
<tr><th>Small molecules</th><th>DNA, RNA</th><th>Proteins</th></tr>
<tr><td rowspan="2">investigate a mixture qualitatively</td><td>GC</td><td>CE</td><td>MALDI-TOF-MS</td></tr>
<tr><td>LC</td><td colspan="2">GE, 2D-GE</td></tr>
<tr><td rowspan="4">selectively quantify a compound in a mixture</td><td>GC</td><td>CE</td><td rowspan="3">bioassay</td></tr>
<tr><td>LC</td><td>Real-Time PCR</td></tr>
<tr><td rowspan="2">chemical or biosensor</td><td>DNA array</td></tr>
<tr><td colspan="2">biosensor</td></tr>
<tr><td rowspan="2">identify a pure compound qualitatively</td><td>MS</td><td>PCR</td><td>amino acid composition</td></tr>
<tr><td>NMR</td><td>DNA arrays</td><td>tryptic digest and
GE
MALDI-TOF-MS
ESI-MS</td></tr>
<tr><td rowspan="4">elucidate the structure of a pure compound</td><td>NMR</td><td>DNA sequencing</td><td>amino acid sequencing</td></tr>
<tr><td>MS</td><td colspan="2">NMR</td></tr>
<tr><td>IR</td><td colspan="2">X-ray crystallography</td></tr>
<tr><td>X-ray crystallography</td><td colspan="2">electron-microscopy</td></tr>
</table>

To isolate the DNA or RNA molecules from the cell, a number of conventional methods such as liquid-liquid extraction, precipitation and centrifugation (section 6.1) can be employed.

An isolated nucleic acid can then be quantified, for example by UV spectroscopy. The aromatic groups of the bases have an absorption maximum around $\lambda = 260$ nm. Alternatively, fluorescent or radioactive markers can be attached and quantitatively detected. A mixture of DNA molecules can be quantified by capillary electrophoretic methods (section 3.3).

A DNA molecule can be amplified by the polymerase chain reaction (PCR) (section 6.2), if part of its sequence is known. One DNA molecule is sufficient to generate millions of identical copies in a controlled amplification reaction. With real-time PCR, the DNA quantity can be measured during the amplification reaction (section 6.2.4). Other methods of DNA quantification include DNA arrays (section 5.3) and, if available, biosensors (section 5.2).

Sequencing methods for DNA include the Maxam-Gilbert method (section 6.3.2), the Sanger method (section 6.2.4), DNA arrays (section 5.3) and pyrosequencing (section 5.4). Usually, the DNA molecules are treated with a restriction enzyme (section 6.3.1) prior to sequencing. A number of fragments are thus generated, which are then separated from each other according to their molecular weight by gel electrophoresis (sections 3.2 and 3.3.5). The most efficient separation method is two-dimensional gel electrophoresis (2D-GE) (section 3.2.5).

1.3.3.2 *Analysis of proteins*

In *proteomics* research, the aim is to study all the proteins expressed in a cell, tissue or organisms to obtain an insight into the interplay of cells and organism. Protein analysis often involves isolation and investigation of one protein at a time.

Protein concentrations in cells are usually very low and an amplification reaction such as PCR for DNA molecules does not exist for proteins. Isolating a protein from a complex cell matrix with a high yield and without changing its biological functionality can be a difficult task. Some of the analytical methods involved are liquid-liquid extraction, precipitation and centrifugation. Often a protein or a group of proteins is separated from impurities by liquid chromatography (section 2) or polyacrylamide gel electrophoresis (PAGE) (section 3.2).

Quantitative analysis of proteins can be achieved by UV spectroscopy. The peptide bond has an absorption maximum around $\lambda = 205$ nm, the aromatic rings on the amino acids Tryptophan and Tyrosine absorb strongly around $\lambda = 280$ nm. Also commonly used are colorimetric assays, which contain reagents that specifically form coloured complexes with proteins. These quantitative methods usually measure the total protein concentration. Either the protein of interest has to be isolated prior to analysis, or a very specific method has to be found to quantify only the targeted protein. Very sensitive and specific analysis of antibodies and antigens can be achieved with bioassays (section 5.1) or biosensors (section 5.2).

The amino acid composition (section 7.5) of a protein can be determined by first completely hydrolysing the peptide bonds and then separating and quantifying the obtained amino acids. Ion exchange chromatography (IEC) (section 2.3.2), reversed phase liquid chromatography (RP-HPLC) (section 2.3.1) and capillary electrophoresis (CE) (section 3.3) can be employed for separation and quantification. If the protein is known and has been catalogued, the amino acid composition is often enough to unambiguously identify a protein in a sample.

Alternatively, the protein can be partially digested by an enzyme like trypsin. The fragments of this tryptic digest can then be separated and their molecular weights can be measured with a mass spectrometer. Methods available for such analysis are matrix assisted laser desorption ionisation (MALDI) time of flight (TOF)

mass spectrometry (MS) (section 4.1) as well as electrospray ionisation mass spectrometry (ESI-MS) (section 4.2) coupled to liquid chromatography (LC-ESI-MS) or capillary electrophoresis (CE-ESI-MS). Often the protein can be identified from the molecular weights of the tryptic digest fragments.

If the protein cannot be determined by either of these two methods, then sequencing of the amino acids becomes necessary. The strategies for this are outlined in chapter 7.

For proteomics analysis, all the proteins from a cell must be extracted and then separated from each other. Gel electrophoretic methods (section 3.2) are most powerful, especially two-dimensional gel electrophoresis (2D-GE), which is capable of separating thousands of proteins in a single run (section 3.2.5).

Three-dimensional structures of both proteins and nucleic acids can be obtained by sophisticated NMR experiments, by electron microscopy, and by X-ray crystallography, if a monocrystal can be obtained. Covering these techniques is beyond the scope of this book. The reader may refer to one of the textbooks in the references given at the end of this chapter.

Summary

The structure and main features of amino acids, proteins nucleotides and DNA were outlined in this chapter.

DNA is the hereditary molecule of all cellular life forms. It stores and transmits genetic information. DNA is a relatively simple molecule, composed only of four different nucleotides with the bases adenine, guanine, thymine and cytosine and β-D-deoxyribose as the sugar component. Millions of nucleotides can be linked together. Two complementary strands are twisted around each other in the form of a double helix. They are held together by hydrogen bonds between the base pairs adenine-tymine and cytosine-guanine. RNA is comprised of nucleotides with the bases adenine, guanine, uracil and cytosine and β-D-ribose as the sugar component. RNA is a single stranded molecule with base pairing occurring only in parts of this single chain.

Proteins are relatively complicated molecules made up from the 20 naturally occurring α-L-amino acids, which are linked to each other via peptide bonds. Fibrous proteins give mechanical strength to bones and muscles. Globular proteins such as antibodies and enzymes have specific functions in the immune system and in metabolism. The macromolecular chain of a protein is folded in a very specific way. This folding is essential for the protein's function and activity. The sequence of the amino acids is referred to as the primary structure of the protein. Parts of the amino acid chain form domains of regular structures such as α-helices and β-pleated sheets. These make up the secondary structure of the protein. The whole three-dimensional shape of the amino acid chain including interactions between different secondary domains is referred as the tertiary structure. Some proteins consist of

two or more chains and other, non-amino acid components. This is described as the quaternary structure of the protein.

The three dimensional structure of both, DNA and proteins, is only stable within a certain chemical and physical environment. Changes of temperature, pH and ionic strength may cause the biomolecule to denature. This denaturation is often irreversible.

In a cell, the DNA contains the genetic code. When a gene is “switched on”, it triggers the synthesis of a protein. This protein synthesis is achieved by the processes of transcription and translation.

The methods for analysing and identifying biomolecules are radically different from analysing relatively small organic molecules. Separation of biomolecules is commonly carried out by gel and capillary electrophoresis. Chromatography is used not so much as a separation method, but mainly as a method for purification and isolation of compounds. The molecular recognition that many biomolecules exhibit is used in many analytical tools including immunoassays, biosensors and DNA arrays. The structure of a biomolecule cannot easily be determined by spectroscopic methods. Determination of a protein as well as a nucleic acid structure involves a number of reactions and analysis steps to be carried out.

References

1. L. Stryer, J. M. Berg and J. L. Tymoczko, *Biochemistry*, 5th edition, W. H. Freeman and Co., 2002.
2. D. L. Nelson, A. L. Lehninger and M. M. Cox, *Principles of Biochemistry*, 3rd edition, Worth Publishers, 2000.
3. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, 4th edition, Garland Science Publishing, 2002.
4. H. Lodish, A. Berk, L. S. Zipursky, P. Matsudaira, D. Baltimore and James Darnell, *Molecular Cell Biology*, 4th edition, W. H. Freeman, 2000.