

# **Introduction to Practical Biochemistry**

**György Hegyi  
József Kardos  
Mihály Kovács  
András Málnási-Csizmadia  
László Nyitrai  
Gábor Pál  
László Radnai  
Attila Reményi  
István Venekei**

---

# Introduction to Practical Biochemistry

by György Hegyi, József Kardos, Mihály Kovács, András Málnási-Csizmadia, László Nyitrai, Gábor Pál, László Radnai, Attila Reményi, and István Venekei  
Copyright © 2013 Eötvös Loránd University

This book is freely available for research and educational purposes. Reproduction in any form is prohibited without written permission of the owner.

Made in the project entitled "E-learning scientific content development in ELTE TTK" with number TÁMOP-4.1.2.A/1-11/1-2011-0073.  
Consortium leader: Eötvös Loránd University, Consortium Members: ELTE Faculties of Science Student Foundation, ITStudy Hungary Ltd.

National Development Agency  
www.ujszchenyiterv.gov.hu  
06 40 638 638



The project is supported by the European Union  
and co-financed by the European Social Fund.



---

# Table of Contents

|  |    |
|--|----|
| Foreword .....   | vi |
| 1. Common laboratory tools and equipment used in biochemistry and molecular biology .....  | 1  |
| 1.1. Biological samples and chemical substances in the laboratory .....  | 1  |
| 1.2. Plastic and glass tubes used for the storage of liquids .....   | 1  |
| 1.3. Beakers and laboratory flasks .....   | 4  |
| 1.4. Precise volumetric measurements with graduated cylinders and micropipettes .....  | 5  |
| 1.5. Mixing of liquids .....   | 8  |
| 1.6. Laboratory balances .....   | 9  |
| 1.7. Methods of sterilisation and in-house production of high-purity water .....   | 10 |
| 1.8. Working with cell cultures .....  | 12 |
| 1.9. Centrifuges .....   | 13 |
| 1.10. Other widely used laboratory techniques: spectrophotometry, electrophoresis, chromatography .....  | 16 |
| 1.11. Storage of biological samples .....  | 18 |
| 2. Units, solutions, dialysis .....  | 21 |
| 2.1. About units .....   | 21 |
| 2.2. Numeric expression of quantities .....  | 21 |
| 2.2.1. The accuracy of numbers, significant figures .....  | 21 |
| 2.2.2. Expression of large and small quantities: exponential and prefix forms .....  | 22 |
| 2.3. About solutions .....   | 23 |
| 2.3.1. Definition of solutions and their main characteristics .....  | 23 |
| 2.3.2. Quantitative description of solutions, concentration units .....  | 23 |
| 2.3.3. Preparation of solutions .....  | 24 |
| 2.4. Dialysis .....  | 25 |
| 2.4.1. The principle of dialysis .....   | 25 |
| 2.4.2. Practical aspects and applications of dialysis .....  | 26 |
| 3. Acid-base equilibria, pH, buffer systems .....  | 28 |
| 3.1. Ionisation equilibria of acids and bases in aqueous solutions .....   | 28 |
| 3.2. pH-stabilising acid-base systems (buffers) and the influence of pH on ionisation .....  | 31 |
| 3.3. Measurement of the pH .....   | 34 |
| 3.4. Demo calculations of charge and pI .....  | 35 |
| 3.4.1. Demonstration that pI is the average of the $pK_a$ values of the carboxylic acid and amino groups of an amino acid lacking an ionisable group in its side chain ..... | 35 |
| 3.4.2. Demonstration that the pI value of aspartic acid is the average of the $pK_a$ values of the two carboxylic acid groups in it .....                                    | 37 |
| 3.4.3. Demo calculation of the isoelectric point of a protein .....  | 39 |
| 4. Spectrophotometry and protein concentration measurements .....  | 42 |
| 4.1. Photometry .....  | 42 |
| 4.2. The UV-VIS photometer .....   | 43 |
| 4.3. Other possible uses of photometry .....   | 44 |
| 4.4. Frequently arising problems in photometry .....   | 45 |
| 4.5. Determination of protein concentration .....  | 45 |
| 4.5.1. Biuret test .....   | 46 |
| 4.5.2. Lowry (Folin) protein assay .....   | 46 |
| 4.5.3. Bradford protein assay .....  | 46 |
| 4.5.4. Spectrophotometry based on UV absorption .....  | 46 |
| 4.6. Spectrophotometry in practice: some examples .....  | 47 |
| 4.6.1. Absorption spectrum of ATP .....  | 47 |
| 4.6.2. Hyperchromicity of DNA .....  | 47 |
| 4.6.3. Absorption spectra and molecular structure of NAD and NADH .....  | 48 |
| 4.6.4. Absorption spectrum of proteins .....   | 49 |
| 4.6.5. Determination of the purity of DNA and protein samples .....  | 50 |
| 4.7. Fluorimetry .....   | 51 |
| 4.7.1. Physical basis of fluorescence .....  | 51 |
| 4.7.2. The fluorimeter .....   | 52 |

|   |     |
|---|-----|
| 4.7.3. Fluorophores .....   | 53  |
| 4.8. Appendix .....   | 57  |
| 4.8.1. Fluorescence, phosphorescence and chemiluminescence .....  | 57  |
| 4.8.2. Photobleaching .....   | 57  |
| 4.8.3. Fluorescence anisotropy and circular dichroism .....   | 57  |
| 4.8.4. Quenching and FRET .....   | 59  |
| 5. Cell disruption, cell fractionation and protein isolation .....  | 60  |
| 5.1. Cell disruption .....  | 60  |
| 5.2. Cell fractionation .....   | 61  |
| 5.3. Centrifugation .....   | 61  |
| 5.3.1. Differential centrifugation: cell fractionation based primarily on particle size .....             | 63  |
| 5.3.2. Equilibrium density-gradient centrifugation: fractionation based on density .....                  | 64  |
| 5.4. Low-resolution, large-scale protein fractionation .....  | 66  |
| 5.4.1. Fractionation methods based on solubility .....  | 67  |
| 5.4.2. Protein fractionation based on particle size .....   | 70  |
| 5.5. Lyophilisation (freeze-drying) .....   | 71  |
| 6. Chromatographic methods .....  | 72  |
| 6.1. Gel filtration chromatography .....  | 74  |
| 6.2. Ion exchange chromatography .....  | 77  |
| 6.3. Hydrophobic interaction chromatography .....   | 80  |
| 6.4. Affinity chromatography .....  | 81  |
| 6.5. High performance (high pressure) liquid chromatography (HPLC) .....                                  | 84  |
| 7. Electrophoresis .....  | 88  |
| 7.1. Principles of electrophoresis .....  | 88  |
| 7.2. About gel electrophoresis .....  | 89  |
| 7.3. Polyacrylamide gel electrophoresis (PAGE) .....  | 90  |
| 7.3.1. About the PAGE method in general .....   | 90  |
| 7.3.2. Native PAGE .....  | 94  |
| 7.3.3. SDS-PAGE .....   | 95  |
| 7.3.4. Isoelectric focusing .....   | 96  |
| 7.3.5. Two-dimensional (2D) electrophoresis .....   | 98  |
| 7.4. Agarose gel electrophoresis .....  | 99  |
| 7.5. Staining methods .....   | 99  |
| 7.5.1. General protein gel stains .....   | 99  |
| 7.5.2. General DNA gel stains .....   | 100 |
| 7.5.3. Specific protein detection methods: Western blot .....   | 100 |
| 7.5.4. Specific protein detection methods: In-gel method based on enzyme activity .....                   | 101 |
| 7.6. Typical examples of protein-separating gel electrophoresis .....                                     | 102 |
| 7.6.1. Native PAGE separation and detection of lactate dehydrogenase isoenzymes .....                     | 103 |
| 7.6.2. Molecular mass determination of myofibrillar proteins using SDS-PAGE .....                         | 104 |
| 8. Protein-ligand interactions .....  | 107 |
| 8.1. Biomolecular interactions .....  | 107 |
| 8.2. Reaction kinetics .....  | 107 |
| 8.3. Protein-ligand interactions .....  | 108 |
| 8.4. Relationship between the free enthalpy (Gibbs free energy) change and the equilibrium constant ..... | 109 |
| 8.5. Molecular forces stabilising ligand binding .....  | 110 |
| 8.6. Determination of the binding constant .....  | 112 |
| 8.7. Methods for the experimental determination of the binding constant .....                             | 115 |
| 8.7.1. Surface plasmon resonance (SPR) .....  | 116 |
| 8.7.2. Isothermal titration calorimetry (ITC) .....   | 117 |
| 8.7.3. Fluorescence depolarisation to characterise protein-ligand binding interactions .....              | 119 |
| 8.8. Test questions and problems .....  | 119 |
| 9. Enzyme kinetics .....  | 121 |
| 9.1 Thermodynamic interpretation of enzyme catalysis .....  | 121 |
| 9.2. Michaelis-Menten kinetics .....  | 125 |
| 9.3 Determination of initial reaction rates and principal kinetic parameters .....                        | 133 |
| 9.4. Enzyme inhibition mechanisms .....   | 135 |

|   |     |
|---|-----|
| 9.4.1. Competitive inhibition .....   | 136 |
| 9.4.2. Uncompetitive inhibition .....   | 138 |
| 9.4.3. Mixed inhibition .....   | 139 |
| 10. Recombinant DNA technology .....  | 141 |
| 10.1. Recombinant DNA techniques and molecular cloning .....  | 141 |
| 10.2. Plasmid vectors .....   | 141 |
| 10.3. Creation of recombinant DNA constructs .....  | 143 |
| 10.4. Introduction of recombinant DNA constructs into host cells and the identification of recombinant colonies ..... | 145 |
| 10.5. Isolation of plasmid DNA .....  | 148 |
| 10.6. Analysis of plasmid DNA by gel electrophoresis .....  | 149 |
| 10.7. Polymerase chain reaction (PCR) .....   | 152 |
| 10.8. Site directed <i>in vitro</i> mutagenesis .....   | 154 |
| 10.9. DNA sequencing .....  | 158 |
| 11. Bioinformatics .....  | 164 |
| 11.1. Introduction .....  | 164 |
| 11.2. Primary sequence and three-dimensional structure databases .....  | 165 |
| 11.2.1. GenBank .....   | 165 |
| 11.2.2. UniProt .....   | 168 |
| 11.2.3. Protein Data Bank (PDB) .....   | 170 |
| 11.3. Introduction to bioinformatics analysis of sequences .....  | 171 |
| 11.3.1. Bioinformatics tasks during molecular cloning .....   | 171 |
| 11.3.2. Sequence similarity search and sequence alignment .....   | 171 |
| 11.3.3. Bioinformatics analysis of protein sequences .....  | 174 |
| 11.4. Visualisation of protein structures by molecular graphics programs .....  | 174 |
| 11.4.1. RasMol .....  | 174 |
| 11.4.2. PyMOL .....   | 176 |
| 11.4.3. Jmol .....  | 177 |
| 12. Calculations and problem solving exercises .....  | 179 |
| 12.1. Useful preliminary information .....  | 179 |
| 12.2. Problems and exercises .....  | 179 |
| 12.2.1 Units of measure, solutions .....  | 179 |
| 12.2.2 Ionisation equilibria .....  | 182 |
| 12.2.3 Spectrophotometry of biomolecules .....  | 184 |
| 12.2.4 Cell disruption, cell fractionation and protein isolation .....  | 186 |
| 12.2.5 Peptides and proteins .....  | 187 |
| 12.2.6 Chromatographic methods .....  | 188 |
| 12.2.7 Electrophoretic methods .....  | 190 |
| 12.2.8 Protein-ligand interactions .....  | 191 |
| 12.2.9 Enzyme kinetics .....  | 191 |
| 12.2.10 Recombinant DNA technology .....  | 194 |
| 12.2.11 Bioinformatics .....  | 194 |
| 12.3. Solutions .....   | 195 |
| 12.3.1 Units of measure, solutions .....  | 195 |
| 12.3.2 Ionisation equilibria .....  | 196 |
| 12.3.3 Spectrophotometry of biomolecules .....  | 197 |
| 12.3.4 Cell disruption, cell fractionation and protein isolation .....  | 197 |
| 12.3.5 Peptides and proteins .....  | 197 |
| 12.3.6 Chromatographic methods .....  | 198 |
| 12.3.7 Electrophoretic methods .....  | 198 |
| 12.3.8 Protein-ligand interactions .....  | 199 |
| 12.3.9 Enzyme kinetics .....  | 199 |
| 12.3.10 Recombinant DNA technology .....  | 200 |
| 12.3.11 Bioinformatics .....  | 200 |
| 13. Epilogue .....  | 201 |
| Bibliography .....  | 204 |

---

# Foreword

by Attila Reményi

The “Introduction to Practical Biochemistry” e-book is mainly intended for B.Sc. students studying biology at Eötvös Loránd University. It is part of the course material for students attending the seminars run under the same title. As it covers a broad range of subjects on the basic as well as the practical aspects of biochemical and molecular biological work, it is likely that it will be also useful for any student attending different theoretical or practical biochemistry courses. The course material builds on pre-existing knowledge obtained at previous B.Sc. courses including General Chemistry, Physical Chemistry and Organic Chemistry. It assumes a solid background and experience in chemical calculations and the successful completion of the course entitled “Introduction to Biochemistry”, which is taught as part of the Biology B.Sc. program at Eötvös Loránd University, or that of another Biochemistry course at a similar level. The “Introduction to Biochemistry” e-book can be found [here](#).

The “Introduction to Practical Biochemistry” seminar series will prepare students for more advanced courses including the lectures on “Biochemistry and Molecular Biology”, and it is particularly indispensable for the third-year hands-on training course entitled “Practicals in Biochemistry”. The format of the course may be described as a “practical seminar”. This is a mixture of the classical seminar where the theoretical principles are further discussed interactively with students, and the classical practical where the same is accomplished by performing experiments and analyzing experimental data in a first-hand manner. On practical seminars, teachers present the basic principles of techniques broadly used in the biochemical and molecular biological laboratory practice, make some demonstrations on different techniques and show the use of some of the everyday laboratory instrumentation. We have put special emphasis on presenting demonstrations and problem sets that will make the students face “real-life” laboratory situations. Problem sets and biochemical calculations are to be solved interactively, with students working in groups on finding the solution and the teacher being involved only as a discussion moderator.

The e-book is not the description of different biochemical practicals, and it does not contain detailed experimental protocols to perform experiments. It rather contains a collection and description of principles that will help the students perform successful biochemical and molecular biological experiments on their own during their future carrier. The experience of the author team gathered during five years of practice resulted in a course material that enables students to efficiently use hands-on practicals in biochemistry and molecular biology later during their training. Moreover, the material also provides a solid background in biochemical calculations, a prerequisite for successful experimental design.

The e-book covers the course material for a one-semester B.Sc. course delivered in three hours per week. The course does not discuss all families of molecules that are subject to biochemical and molecular biological investigation. It mainly deals with techniques used to study proteins and nucleic acids. The methods on carbohydrates and lipids are discussed as parts of Organic Chemistry courses, and they are also discussed in lectures on “Biochemistry and Molecular Biology”.

By completing the “Introduction to Practical Biochemistry” course, students will acquire practical knowledge regarding the techniques used to investigate the properties of macromolecules. As mentioned earlier, the course puts a great emphasis on demonstrating how to solve “real-life” tasks and problems faced by the investigator in biochemical and molecular biological laboratories. Students will become familiar with commonly used labware and instrumentation of a biochemical and molecular biological laboratory, learn the requirements for sterile work, and will be able to store biological samples properly. They learn how to correctly use biochemical units of measure and to make solutions and buffers for basic biochemical experiments. They will be able to determine the charge of weak acids, bases and macromolecules by taking into account their chemical environment. They learn how to measure the macromolecular concentration of solutions by spectrophotometry. They will be able to design protocols for the purification of proteins from cell cultures or tissues. They will become familiar with the physico-chemical background of enzyme action and the fundamentals of enzyme kinetics. They learn how to characterise the interaction between macromolecules and their ligands quantitatively. They will become familiar and be capable of using basic recombinant DNA techniques for nucleic acid manipulation and the production of proteins. They will be able to use public bioinformatics databases to acquire information on the physical, chemical and biological properties of macromolecules.

The e-book comprises 11 chapters dealing with different topics. In addition, Chapter 12 contains more than 100 simple and more complex problems enabling students to constantly put their knowledge to a test by attempting to solve the problem sets belonging to specific chapters.

The author team of the Department of Biochemistry at Eötvös Loránd University wishes the students and all readers an enjoyable experience in entering the field of biochemical and molecular biological laboratory life!

---

# **Chapter 1. Common laboratory tools and equipment used in biochemistry and molecular biology**

by László Radnai

The aims of biochemical and molecular biological research are complex and diverse. Investigation of the network of chemical reactions taking place in living organisms and representing the most fundamental phenomena of life, identification of the molecules playing roles in biochemical processes, determination of their structure, function and interactions, examination of the molecular background of metabolism, the flow of energy and information within organisms are all among the common goals of biochemists and molecular biologists. In accordance with this diversity of problems, a high variety of tools, instruments and methods are required to answer scientific questions effectively. This chapter reviews the most common tools and instruments used very frequently in almost every laboratory. The appropriate handling and storage of biological samples and other chemical substances required for research will also be discussed.

## **1.1. Biological samples and chemical substances in the laboratory**

Tissue or cell samples from a living organism, different cell cultures grown in a laboratory incubator under controlled conditions, homogenates or extracts of cells and tissues, solutions of isolated and purified components (e.g. proteins, nucleic acids) can all be referred to as “biological samples”. As the medium of life is water, the majority of biological samples can be defined as aqueous solutions with one or more components, colloidal systems, or water-based suspensions (e.g. bacterial cells dispersed in a liquid medium). Consequently, most biochemical experiments also take place in aqueous environments. Therefore, laboratory vessels used to store liquids and laboratory tools required for the manipulation, transfer and accurate volume measurements of liquids will be introduced in this chapter. Different solids (e.g. chemical substances obtained from different companies, synthetic oligonucleotides or peptides) are also often necessary for biochemical research. In most cases, solids are dissolved in water (or sometimes in other solvents) prior to the experiments. Therefore, the methods of preparing solutions and measuring accurately the weight of the required solids will also be discussed below. Sometimes we use gases in the laboratory. These can be stored in gas cylinders (e.g. O<sub>2</sub>), in Dewar flasks in liquid state (e.g. liquid nitrogen), or dissolved in water (e.g. HCl or NH<sub>3</sub>). By working with gases it is very important to follow all safety instructions to avoid fire, explosions, frostbite or (in case of inhalation) asphyxia or poisoning.

## **1.2. Plastic and glass tubes used for the storage of liquids**

Vessels made of different transparent plastics are widely used in laboratory practice for the storage of liquids. Plastics are cheap and flexible. Containers, flasks and tubes are often equipped with lids, caps, or screw-caps. Moreover, plastic containers are ideal because they retain their flexibility in a wide range of temperatures, while glass can be more sensitive to temperature changes or can be broken easily. For the storage of liquids, probably the most important criterion is the air-tightness of the vessels. An air-tight cap can protect the sample from the evaporation of the solvent. It also protects against dust, bacteria, mould spores or other impurities originating from the environment. It blocks the dissolution of different gases into the sample. Dissolved gases can modify the biomolecules directly (e.g. oxygen reacts with sulfhydryl groups of cysteines within a protein, leading to the formation of disulphide bonds) or indirectly (e.g. carbon dioxide forms carbonic acid in water, which dissociates and decreases the pH of the solution, thereby affecting the protonation state and solubility of proteins). Plastics have some other advantages. One of these is their inertness against the majority of chemical substances used in most experiments. However, some experiments may require the use of organic solvents. In such cases one needs to check the compatibility of the given solvent with the plastic vessels prior to the experiment.

Liquid samples with volumes up to 50 millilitres can be stored in so-called Falcon tubes (also called Falcon centrifuge tubes) (Figure 1.1). Falcon tubes are manufactured with different nominal volumes (most typically, 15 mL and 50 mL) and are supplied with screw-caps. The conical bottom of the tube is particularly advantageous when there is only a small amount of liquid left in the tube. In this case all drops can be collected readily by centrifugation. These tubes have to be placed into appropriate racks. However, there are also free-standing tubes available having a plastic “skirt” around the conical bottom. These tubes can be placed on horizontal surfaces without any support, but must be handled carefully to avoid tipping up. Falcon tubes are graduated; thus the volume of the sample can be estimated easily. However, more precise volumetric measurements require other laboratory tools (e.g. graduated cylinders). It is crucial in all laboratories to mark the samples unambiguously. Unlabelled samples are generally (and rightfully) considered as litter. Falcon tubes can be labelled on the top of the screw-caps or on the side of the tube. (They often have an area with white background for this purpose.) Labels can be printed or hand-written by using a marker pen. Illegible writing must be avoided and the label must be protected against abrasion, e.g. by a piece of transparent cellophane tape. Components of the sample, concentrations, solvent(s), buffering component(s) and pH, possible toxicity are the most important parameters that generally need to be indicated on tubes. The name of the experimenter and/or the date of the sample preparation or the experiment are also often important, especially if the components are prone to degradation.



Figure 1.1. 50-mL and 15-mL Falcon tubes, glass test tube and Wasserman tube (left to right in blue rack) in plastic racks.

Test tubes and narrower Wassermann tubes (Figure 1.1) are usually made of glass and provided without caps. They have a U-shaped bottom. They are mainly used for temporary purposes (e.g. for preparing reaction mixtures or for collecting fractions during chromatographic separation of different components). The main advantage of glass is its high resistance against most chemical substances and solvents used in typical biochemical experiments (with the exception of concentrated strong bases). If the long-term storage of the sample is necessary, the openings of the tubes can be closed air-tightly by a piece of parafilm (Figure 1.2). Parafilm is a thin layer of paraffin, manufactured and supplied with a paper backing. It is ductile, flexible and cohesive. The opening of the test tube (or other laboratory container) can be covered by using a piece of parafilm of appropriate size. Overhanging ends must be wrapped around the neck tightly. Parafilm is often used also to provide additional sealing on laboratory vessels and tubes having a cap or lid, in order to protect the sample more effectively.

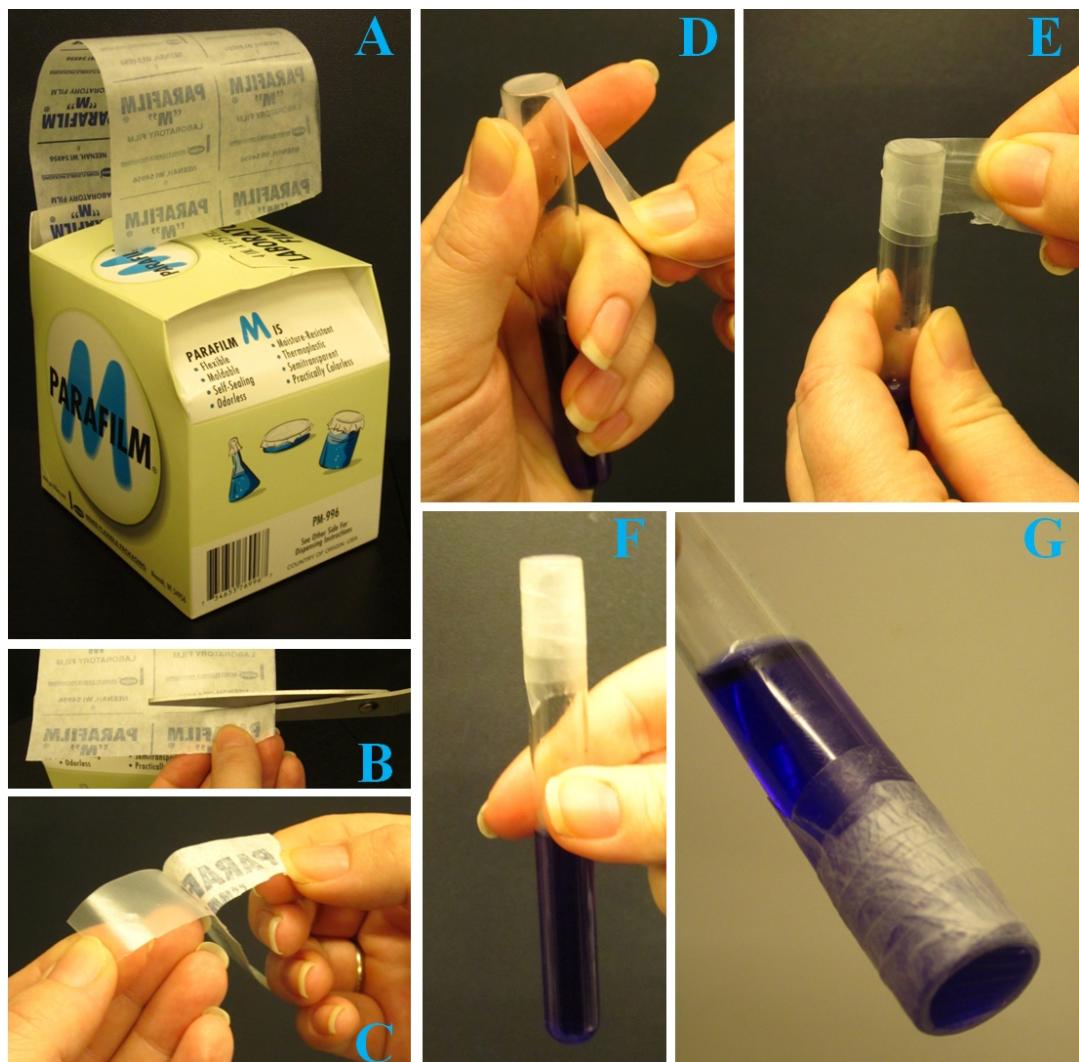


Figure 1.2. A, One roll of parafilm in a dispenser box. B, Cutting a piece of parafilm with scissors. C, Removal of the parafilm layer from the paper backing. D, Covering a Wasserman tube with a piece of parafilm. E, Wrapping the overhanging ends around the opening of the tube. F, Wasserman-tube sealed with parafilm. G, Parafilm ensures air-tight sealing.

For samples of small volume (ranging from several microlitres to several millilitres), Eppendorf tubes (or Eppendorf microcentrifuge tubes) are used (Figure 1.3). These tubes are available in different nominal volumes (e.g. 0.5 mL, 1.5 mL, 2 mL and 5 mL). The most common size is 1.5 mL. Eppendorf tubes have a conically-shaped bottom. Hence, they need to be placed in appropriate plastic racks. They are equipped with an attached plastic snap-lid. The connection between the snap lid and the tube is flexible enough to allow opening and closing many times. Labels can be placed on etched marking areas on the top of the lid and on the side wall of the tube.

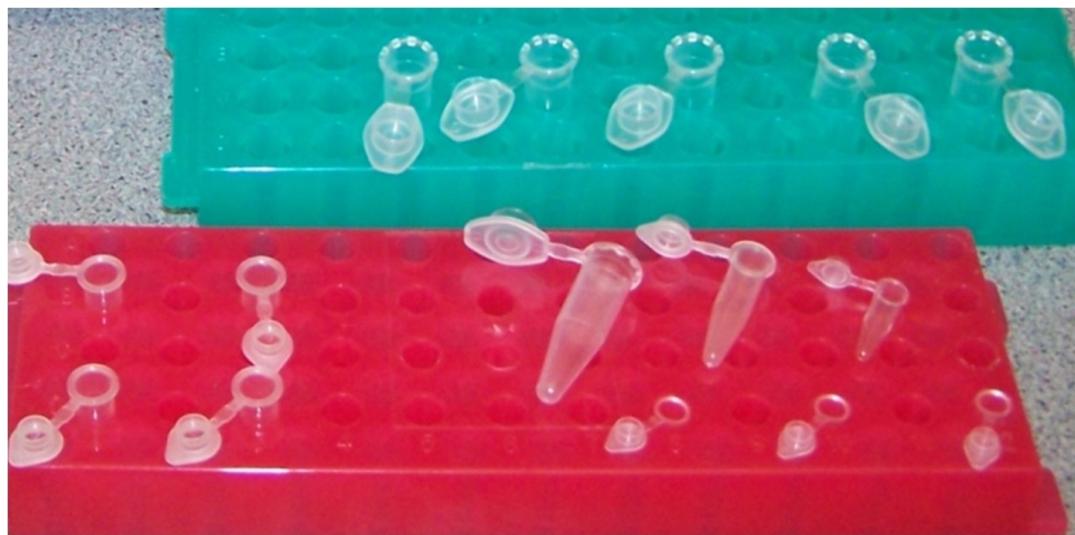


Figure 1.3. Eppendorf tubes with nominal volumes of 0.5 mL (left) and 1.5 mL (top green rack) and PCR tubes with a nominal volume of 200  $\mu$ L (bottom right corner).

PCR tubes (Figure 1.3) are named after their main purpose of usage, the polymerase chain reaction. PCR is one of the most commonly applied enzymatic reactions in recombinant DNA technology that is used frequently in the majority of laboratories to amplify a specified segment of linear double-stranded DNA. Nominal volumes of PCR tubes typically range up to 200  $\mu$ L. Similarly to Eppendorf tubes, PCR tubes are equipped with an attached snap-lid, and have an etched marking area on the top of the lid or on the side wall of the tube. Besides their application in PCR reactions, PCR tubes can be used for a variety of other enzymatic reactions.

## 1.3. Beakers and laboratory flasks

Beakers (or laboratory beakers) are simple cylindrical vessels with a flat bottom, typically used for the preparation and short-term storage of solutions and liquids (Figure 1.4). Nominal volumes of beakers vary between a few millilitres and several litres. Beakers are usually made of glass or plastics and are graduated, aiding the estimation of the actual volume of the sample. However, precise volumetric measurements are not possible with beakers. Beakers made of heat-shock resistant borosilicate glass are suitable for the heating or boiling of solutions by using a Bunsen burner.



Figure 1.4. Glass and plastic beakers of different size.

A variety of flasks are widely used in laboratories, mainly for the storage and preparation of solutions. Some flasks are used with silicone or rubber stoppers, some have standard taper joints equipped with glass stoppers that fit

tightly into the opening. Other flasks can be sealed by using a piece of parafilm. Nominal volumes of flasks typically vary between 50 millilitres and several litres. Like beakers, laboratory flasks can be made of glass or plastics. Precise volumetric measurements can only be performed by using volumetric flasks; however, this special subtype is rarely used in biochemical laboratories. The most commonly used flask is the so-called Erlenmeyer flask (also known as conical flask) (Figure 1.5), which has a conical body, a wide and flat bottom and a narrow neck. It is especially suitable for growing bacterial (or eukaryotic) cells in nutrient liquid media inside an incubator at a controlled temperature (Figure 1.5, panel B). Incubators provide continuous shaking of cultures to prevent sedimentation of cells and facilitate gas exchange (oxygen is required for efficient growth), also supported by the shape of the Erlenmeyer flask. (The surface of the liquid is relatively large due to the wide bottom of the Erlenmeyer flask.) The wide and flat bottom also helps in fixing the flasks into the holders of the plate of the incubator, while the narrow neck prevents the culture from spilling out. Openings are covered by a piece of aluminium foil allowing gas exchange while keeping out dust, other bacteria or spores from the environment.



Figure 1.5. A, Erlenmeyer flasks. B, *E. coli* bacteria growing in Erlenmeyer flasks.

## 1.4. Precise volumetric measurements with graduated cylinders and micropipettes

Graduated cylinders (Figure 1.6) are used to measure the volume of liquids precisely, ranging from a few millilitres to several litres. These laboratory vessels are essential for mixing or dispensing different liquids and for the preparation of solutions with pre-defined concentrations and volume. They are available in many different sizes ranging from 5 millilitres to a few litres.

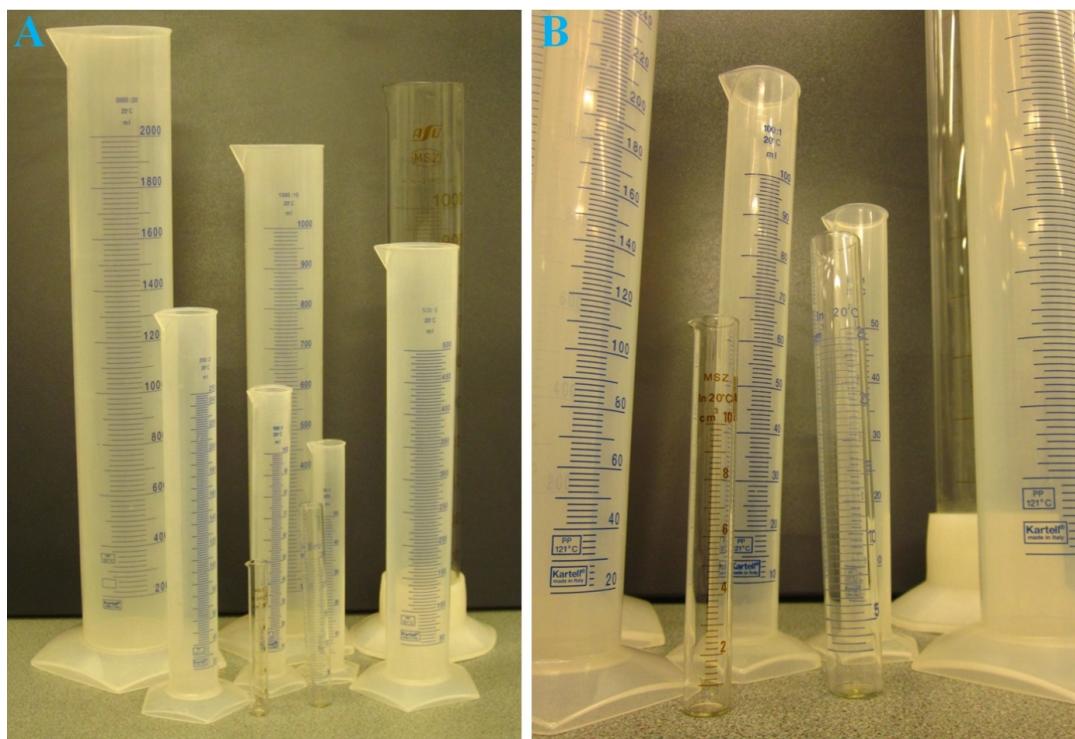


Figure 1.6. A, Glass and plastic graduated cylinders of various sizes. B, Small graduated cylinders (10 - 100 mL, middle of image).

While the smallest graduated cylinders can be used to measure volumes as small as 5 millilitres, it is often necessary—especially in biochemical and molecular biological work—to extend the range of accurate volumetric measurements below 1 millilitre or even below 1 microlitre. Micropipettes (also called piston-driven air displacement pipettes) are the most common laboratory tools applicable in this volume range (Figure 1.7). Within the plastic body of a micropipette, a piston is operated by pushing a button on the top of the device. The button is connected to the piston through a metal rod. The downward movement of the button causes higher pressure inside the airtight cylinder of the piston, thereby pushing out air from the device through a long, hollow plastic shaft, while upward movement generates vacuum. The vacuum is used to draw up liquid into a removable transparent plastic tip. Tips must be fixed in an air-tight manner onto the end of the plastic shaft of the pipette. Without an air tight connection, the volume being drawn will be reduced and undesired leakage will occur during the transfer of the liquid. Disposable tips are used to avoid cross-contamination of samples and stock solutions. Tips must be changed after each operation. (The synthesis and/or purification of chemical substances, proteins, enzymes, nucleic acids and other samples can be extremely laborious, time-consuming and expensive. Therefore, the avoidance of contamination is a crucial issue. Contaminated samples must be discarded and must not be used for any further experiments.) The desired pipetting volume can be set on the pipette, causing a controlled restriction of the movement of the piston. This can be done by turning the volume adjustment knob on the side of the pipette. (Alternatively, the plunger button of several pipettes is used for volume adjustment purposes.) The actual volume is indicated on a digital volume indicator on the side of the pipette body. Maximal and minimal allowed volumes are also indicated on the body or on the top of the pipette. Setting volumes beyond these limits should not be attempted because such operations will damage the pipette and lead to significant inaccuracies. There are pipettes available for different volume ranges. Large pipettes can be used between 1 mL and 5 mL. Below 1 mL, there are several pipettes with the following typical ranges: 200 µL - 1000 µL, 20 µL - 200 µL, 10 µL - 100 µL, 5 µL - 50 µL, 2 µL - 20 µL and 0.5 µL - 10 µL. The transfer of as little as 0.1 µL of liquid is possible with the smallest pipettes with a typical range of 0.1 µL - 2 µL.

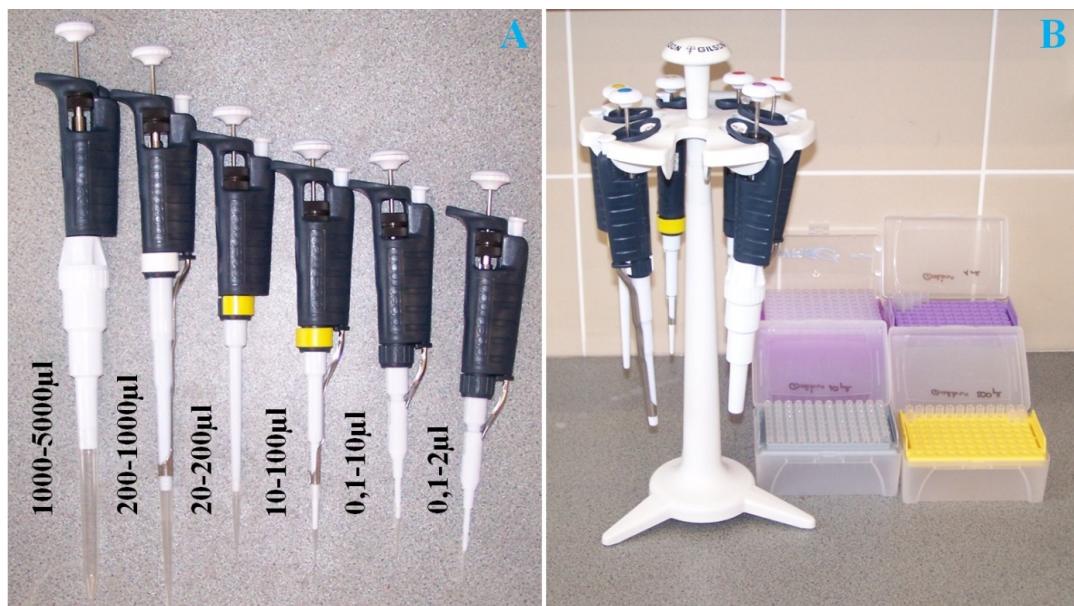


Figure 1.7. A, Micropipettes for different volume ranges. B, Pipettes hanging on a rotating pipette stand, and different tips in racks.

Different tips can be obtained from suppliers according to the volume range of the pipette. Tips can also meet some special criteria, e.g. there are tips supplied with an inner filter for sterile work, or long and narrow tips being very useful in situations where the bottom of deep and narrow tubes or wells must be reached (Figure 1.8). As tips are made of transparent plastics, the operation of the pipette can easily be checked by visual inspection (e.g. for the presence of undesired air bubbles entering the tip together with the liquid, which will cause inaccuracies). Tips can be removed from the shaft of the pipette by pushing the tip ejector button. This button is connected to a long arm surrounding the shaft and transmitting the force towards the upper end of the tip. Tips are also available pre-packed in appropriate racks. Racks protect tips from contamination and facilitate the fastening of the tips onto the pipette (Figure 1.8, panel B). (The shaft of the pipette can simply be pushed into the upper part of the tip.)

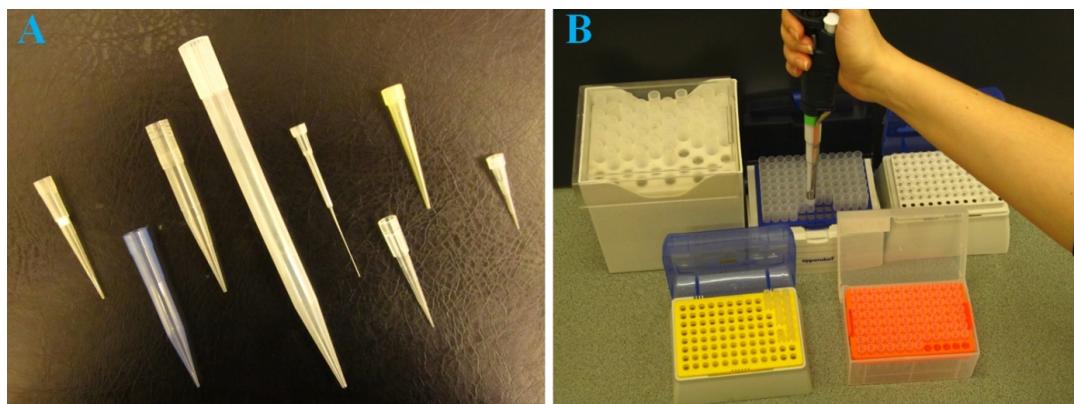


Figure 1.8. A, Pipette tips suitable for different pipettes and purposes (e.g. tips with sterile filters or elongated tips). B, Fastening a tip onto the pipette.

Micropipettes are user-friendly laboratory devices. However, the desired pipetting accuracy can only be achieved by some practice (Figure 1.9, panels A-H). The button of the pipette (and the connected piston) has three characteristic positions (Figure 1.9, panels A-C). In the resting position, the button is pushed up by a spring (first position). By pushing the button gently down with one's thumb (exerting a moderate force on it), air will leave through the previously fixed tip. The button stops in the second position when the desired volume (the one that has been set by the volume adjustment knob) has been reached. While holding the button in the second position, the end of the tip must be submerged into the solution and, subsequently, the button must be released very slowly to let the piston return to the first (resting) position, thereby drawing the desired amount of liquid into the tip. The liquid can then

be transferred to another location (e.g. into a test tube or other vessel). Pushing down the button again to the second position will release the liquid. However, a drop often remains at the end of the tip. To avoid inaccuracies, the entirety of the pipetted solution can be removed by exerting a higher force on the button and pushing it down to the third position. This way, some extra air will be blown out through the tip, thereby removing the remnants of the pipetted liquid. The button must not be released until the tip has been raised above the surface of the solution. (For complete removal of the remaining drop, the tip can be pulled out while the button is being pushed from the second to the third position.) The pipette must be held in a vertical position (with the tip pointing downwards) while transferring liquids. Tilting of the pipette may cause the liquid to leak into the inner parts of the device, causing corrosion and/or contamination. To prevent this, some pipettes have an inner filter (Figure 1.9, panels F-H). When transferring small volumes (e.g. less than 1  $\mu\text{L}$  of an enzyme solution), only the surface of the liquid should be touched with the end of the tip, as small drops can adhere to the outer surface of the tip and cause significant inaccuracies.



Figure 1.9. A-C, Three positions of the button of the pipette. D-H, Liquid handling with pipettes. D, Adjusting the volume. E, Fastening an appropriate tip. F, Drawing of the liquid. G, Transferring the liquid into an Eppendorf tube. H, Dispensing the liquid.

## 1.5. Mixing of liquids

Appropriate mixing of different solutions or liquids is a crucial issue in biochemical experimentation. (Diffusion can be very slow!) Pipettes are perfect tools for mixing volumes not exceeding far beyond their volume range. Submerging the tip into the solution and subsequent pushing and releasing the button (between the first and the second positions) several times will ensure extensive mixing.

The so-called Vortex mixer can be used as an alternative (Figure 1.10, panel A). By pushing down the rubber platform on the top of the device with the bottom of a tube containing the solution to be mixed, the platform will start to oscillate rapidly along a circular path. The liquid will start to shake and swirl along the walls of the tube,

ensuring quick and efficient mixing even if the volume of the solution is very small (e.g. a few microlitres). While vortexing, the tube must be held with fingers approximately at the upper third of its depth to avoid splashing out. (The liquid usually stays below the level of one's fingers.)

Magnetic stirrers are used for the mixing of large volumes (Figure 1.10, panel B). Magnetic stirrers contain an electric motor that rotates a strong permanent magnet. The angular speed of the motor can be set by a knob. The vessels (e.g. beakers, flasks, graduated cylinders) containing the liquid to be mixed are placed onto the top of the device. Some stirrers are equipped with a heater to warm up the solution. A stirring bar (a rod-like strong magnet with coating made of an inert plastic) is dropped into the liquid. The motion of the stirring bar follows the rotation of the magnetic field. Effective mixing is ensured by the spinning stirring bar within the solution. According to the volume and the geometry of the vessels, there are stirring bars available in different sizes and shapes (Figure 1.10, panel C). Magnetic stirrers are very useful tools for the facilitation of the dissolution of crystalline or solid chemical substances. They are also used widely if continuous, intensive mixing is necessary. A typical example is the adjustment of the pH of buffer solutions when a digital pH meter is used to continuously monitor the actual pH of the solution during titration with a strong base or acid until the desired pH value is reached. (Digital pH meters are equipped with a glass electrode that gives a voltage signal proportional to the concentration of oxonium ions in the solution.)

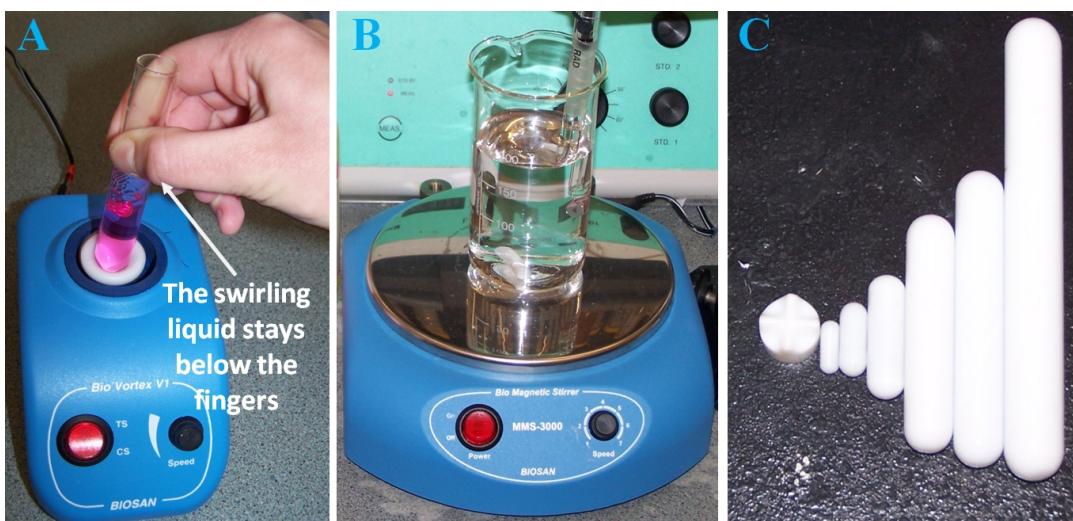


Figure 1.10. A, Mixing of a solution with a Vortex mixer. B, Mixing of a solution with a magnetic stirrer during pH adjustment (the glass electrode of the pH meter is immersed into the solution). C, Stirring bars of different size and shape.

## 1.6. Laboratory balances

Digital laboratory balances with of different capacity and accuracy are available for the measurements of the weight of samples and chemicals. Top-loading balances typically work with 0.1 g accuracy in the range from several grams to several hundred grams (Figure 1.11, panel A). For the precise measurement of small weights, analytical balances are used. These devices usually work with 0.01 mg accuracy in the range from 0.1 mg to several grams. A vibration-free, flat and horizontal surface is required for proper operation of such balances—usually a free-standing table with a marble surface plate. Moreover, the convection of the air around the balance should be avoided. (Air currents can exert forces on the pan of the balance comparable to the force exerted by the sample being measured.) For this purpose, analytical balances have a built-in enclosure with doors made of transparent plastic sheets (Figure 1.11, panel B). Measurements can be performed within this protective enclosure. Disposable weighing dishes and laboratory spatulas with spooned ends are important accessories of weight quantification. It is also possible to dispense solids directly into a beaker or a laboratory tube instead of a weighing dish—however, heavy vessels (with a mass above the upper limit of the balance) must not be applied. The mass of the empty dish must be determined first. Then, the display of the scale must be set to show zero by pushing the “tare” button. Next, the substance is added into the weighing dish. The display will thus show the pure mass of the substance.

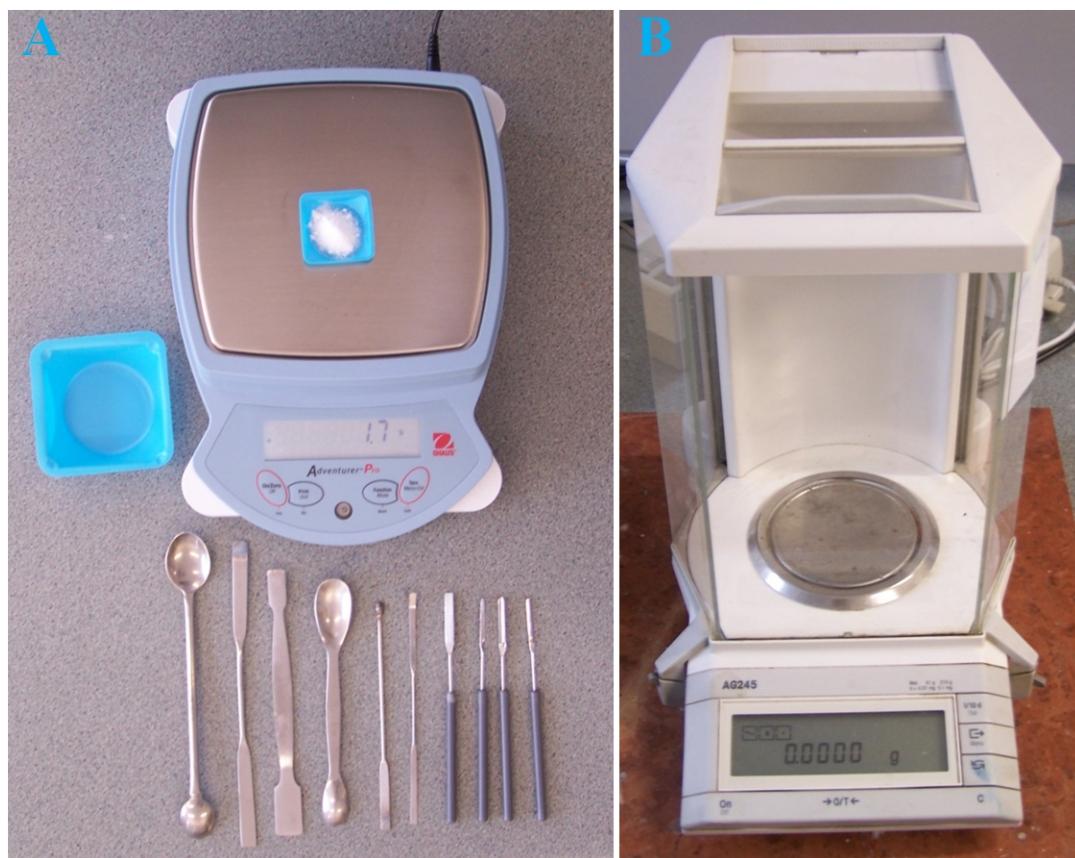


Figure 1.11. A, Top loading balance with disposable weighing dishes, laboratory spatulas and spoons. B, Analytical balance on a vibration-free surface.

## 1.7. Methods of sterilisation and in-house production of high-purity water

High-purity water suitable for biochemical and molecular biological experiments is a general need in all laboratories. Distilled or deionised water can be obtained from different suppliers. However, equipment for its in-house production is also available. Successive distillation is one of the applicable methods. Distillation is energy-consuming and expensive, and it raises various safety issues. A more common alternative is the filtration and deionisation of piped water with a combination of different filters and deionising resins (Figure 1.12, panel A). The applied filters have decreasing average pore diameters; thereby, they gradually remove the corresponding fractions of different contaminating floating particles, microorganisms and bacteria. Deionising resins exchange soluble ions (e.g. calcium, magnesium, sodium, bicarbonate, chloride or heavy metals) for oxonium or hydroxide ions. Clogged filters and used deionising resins must be replaced regularly. The quality of water must be constantly monitored by measuring its electric resistance by conductometers. (The resistance of pure water is above 18 MΩcm at room temperature.)

The importance of using high-purity water, high-quality chemicals and solvents cannot be overestimated. Moreover, having clean laboratory vessels, tubes and devices is also a basic necessity. Some experiments (especially if living cells or organisms are involved) require sterile laboratory vessels and equipment. We consider an experimental setup sterile if microorganisms originating from the environment are removed from it or killed. Depending on the properties of the given sample, chemical substance or device, there are many physical or chemical methods for sterilisation.

Laboratory suppliers sell sterile equipment (Eppendorf tubes, pipette tips, etc.) in sealed sterile bags. However, sterilisation can also be performed in-house, most commonly by using autoclaves (Figure 1.12, panel B). Autoclaves are equipped with a chamber with strong walls in which the process of sterilisation takes place. When all solutions, vessels and other laboratory equipment to be sterilised have been placed inside the autoclave, a defined amount

of water must be poured into the chamber. (The objects to be sterilised are standing on shelves or stages so that the water level in the chamber stays below them.) After closing the airtight door, the autoclave starts to heat up the water. Heating the water in a closed system results in the formation of hot steam and elevated pressure. Autoclaving for 20 minutes at 121°C results in sterility by destroying even the highly resistant endospores of bacteria. (In case of overpressure, safety valves are put into action to keep the process under control.) Equipment and chemicals are to withstand the high temperature in the autoclave; therefore, heat resistance must be checked prior to sterilisation. While laboratory glassware is heat resistant in the applied temperature range, some plastics melt at 121°C. When sterilisation by autoclave is necessary, laboratory tools made of heat resistant plastics must be obtained. Pipette tips must be placed into appropriate racks before sterilisation. Glassware, Eppendorf and Falcon tubes etc. must be placed into bigger vessels and covered with aluminium foil to avoid contamination after the sterilisation procedure. Solutions can also be autoclaved; however, if there are any components susceptible to thermal decomposition, other methods of sterilisation must be sought. Most inorganic compounds and some simple organic compounds (including typical buffers) are not prone to thermal decomposition at 121°C. Liquid media (e.g. for bacterial cell culture) can also be sterilised in an autoclave.



Figure 1.12. A, Filtration and deionisation of water. B, Autoclave.

If the thermal stability of the solution is low (when, for instance, glucose or other monosaccharide is present), sterile filtration can be a good alternative to autoclaving. Sterile filters (Figure 1.13) are pre-sterilised filters with pore sizes smaller than the diameter of bacterial cells (0.2 µm). They are packed in separate envelopes to avoid contamination. The solution can be filtered by using a syringe with an appropriate filter fixed onto it, or by using a vacuum chamber with a filter placed into a funnel on the top of the device.

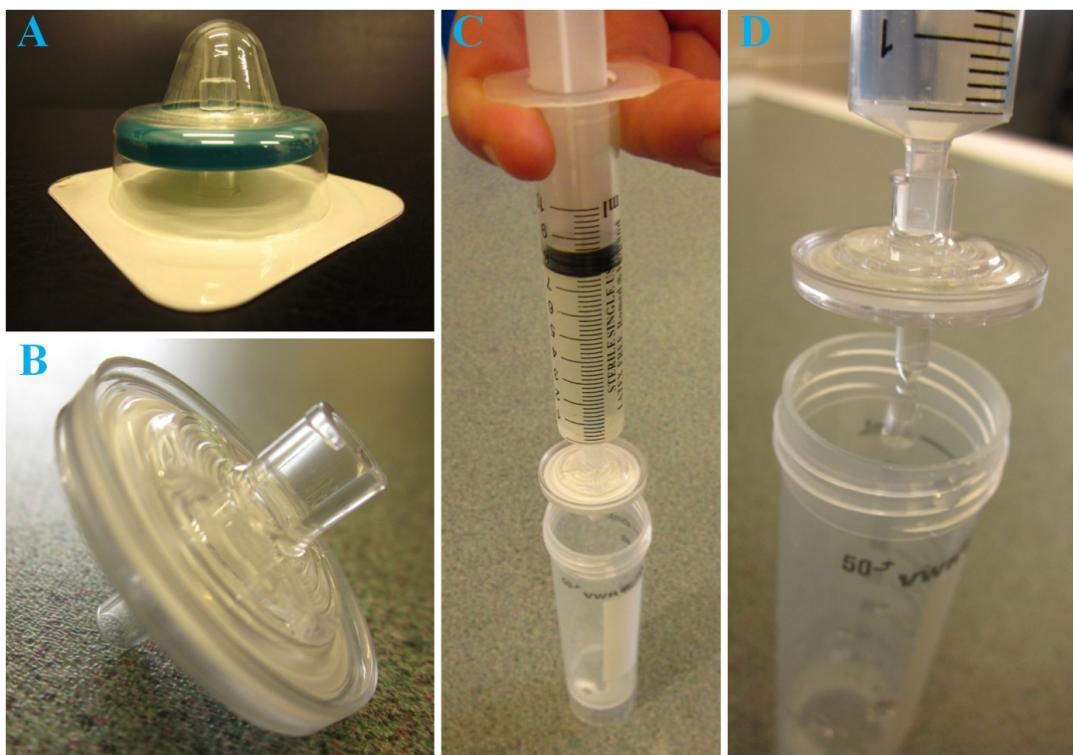


Figure 1.13. A, Sterile filter within its protective envelope. B, Close-up view of a sterile filter. C, Sterile filter fixed onto a syringe. D, Syringe-driven filtration.

## 1.8. Working with cell cultures

Sterility is a crucial issue during working with different cell cultures. Undesired microorganisms accidentally getting into a solution often start to grow by utilizing and degrading the available components. As these microorganisms synthesise the biomolecules of their cells, they release various metabolites into the solution. (Some algae can grow even in distilled water!) Cell cultures (e.g. the cultures of the *E. coli* bacterium used widely in recombinant DNA techniques) must be protected from such “invaders”. To achieve this, sterile laboratory tools, vessels and media must be used. A laminar flow cabinet (also known as tissue culture hood) is required for most operations performed with cell cultures (Figure 1.14, panel A). In the laminar flow cabinet, a constant air current is established. Air filtered through a HEPA (high-efficiency particulate air) filter is blown into the cabinet at the top, and leaves below the working bench. HEPA filters remove all microorganisms. The transparent door of the cabinet protects its inner contents from contamination. While the door is opened, work can be done through a narrow gap allowing only the hands of the operator to reach the equipment and the cell cultures inside. The operator must wear laboratory gloves. Ethanol and/or a Bunsen burner can be used to quickly sterilise equipment inside (e.g. inoculation loops, glass spreaders). Other accessories, including pipette tips and different tubes, are sterilised in an autoclave prior to use. It is very important to avoid situations in which the arm of the operator or any non-sterile equipment is directly placed above sterile cell cultures (or any sterile equipment), because dust particles or bacteria transferred by the air flow (pointing downwards) can easily contaminate the sterile components. Upon finishing the desired operations, the front door of the laminar flow cabinet must be closed. The inner contents can be sterilised by using a built-in UV lamp.

Incubators are used to maintain controlled conditions for bacterial or eukaryotic cell cultures (Figure 1.14, panels B-D). These devices can heat or cool their inner space in order to provide constant temperature conditions for cells. (Simpler and cheaper incubators can only heat; hence, these can only be used above ambient temperature.) Cells can be cultured on different surfaces. Bacteria are often grown in Petri dishes on the top of a layer of gelatinous nutrient agar medium (Figure 1.14, panel B), while eukaryotic (e.g. human) cells are grown at the bottom of cell culture flasks, covered by liquid medium at an appropriate depth (Figure 1.14, panel C). Most of the commonly used eukaryotic cells are adherent: they settle and grow while attached to the surface of the cell culture dish.

Besides temperature, the most important environmental factor is the gas mixture in which the cells are being cultured (for instance, obligate anaerobic bacteria are killed by oxygen, and eukaryotic cells often require elevated levels of carbon dioxide). If necessary, a controlled atmosphere can be maintained by using appropriate incubators.

In general, cells grown in liquid cultures (suspended in liquid medium) can reach higher densities than adherent cultures, because in liquid cultures the entire culture volume can be utilised for growth. Liquid cultures are used if cells are required in large numbers (e.g. during recombinant protein expression in *E. coli* bacteria) (Figure 1.14, panel D). As mentioned above, Erlenmeyer flasks are the most widely used vessels for this purpose. Incubators equipped with a shaking platform are necessary to ensure uniform cell densities and growth conditions within the medium. Intensive shaking of the flasks by circular motion facilitates effective gas exchange. The attained cell density can be further increased by the use of fermentors.



Figure 1.14. A, Laminar flow cabinet. B, Incubator (37°C) with bacterial cultures on agar plates. C, Incubator (37°C) with cultures of eukaryotic cells. D, Incubator (37°C) with a shaking platform for liquid cultures of *E. coli* bacteria (in Erlenmeyer flasks).

## 1.9. Centrifuges

Cells can readily be harvested from liquid cultures by using different centrifuges (Figure 1.15). Similarly, any suspension or floating colloidal particles (e.g. precipitated proteins) in a solution can be separated into fractions

by spinning the sample in a centrifuge (see also Chapter 5 for more detail). The resulting fractions are referred to as “supernatant” (i.e. the solution) and “pellet” (i.e. the particles collected at the bottom of the centrifuge tube, pressed together into a compact mass). Centrifuges are relatively simple devices having stationary and rotary parts. The rotation generated by the electric motor of the centrifuge is transmitted to the rotor harbouring the samples contained within appropriate centrifuge tubes. Many biochemical samples are heat sensitive; for instance, proteins denature at elevated temperatures. Such samples require refrigerated centrifuges in which low temperature can be maintained during centrifugation. Centrifuges are available in different sizes ranging from simple bench-top centrifuges to preparative devices with much higher capacities (volumes up to several litres). Eppendorf tubes or Falcon tubes fit into some rotors of bench-top centrifuges, thereby simplifying the processing of many samples. Most centrifuges can be used in conjunction with several different rotors. This versatility allows users to adapt their centrifuges easily according to the actual requirements. Centrifuge tubes (Figure 1.16) must be chosen according to the manufacturer’s instructions. Importantly, the filled rotor must be counter-balanced during operation. To achieve this, tubes with equal weights must be placed into opposite buckets or holes of the rotor. The weight balance should always be checked by simple two-armed or digital scales. If the weights of the sample tubes are different, counter-balances must be prepared by filling similar tubes with water. Unbalanced rotors are subject to extremely high forces during rotation. Even a small asymmetry of the weights around the axis of rotation can result in the breakage of the rotor shaft at high angular speeds. In such cases, the rotor may also damage the whole device and even cause serious personal injuries.



Figure 1.15. A, Bench-top Eppendorf centrifuge with samples arranged symmetrically around the axis of the rotor. B, Semi-preparative, refrigerated centrifuge with a rotor for Eppendorf tubes. C, Preparative, refrigerated centrifuge with a rotor for the handling of volumes up to 6 litres. D, Semi-preparative rotor with lid. E, Preparative rotor with lid.



Figure 1.16. Centrifuge tubes.

The rotation speed of centrifuges is often specified as the number of revolutions per minute (RPM). However, as the force applied to the sample depends not only on the actual RPM value but also on the radius of the rotor, the relative centrifugal force (RCF) is more informative about a particular experiment. This defines acceleration according to the mass of particles floating in the sample. Therefore, RCF values are given as relative acceleration values (the centrifugal acceleration compared to  $g$ ,  $\sim 9.8 \text{ m/s}^2$ , the acceleration due to gravity on the surface of Earth). Thus, if the same sample is spun at equal RPM values in two centrifuges with different rotor geometries (different rotor radii), the results will be different. However, if equal RCF values are applied, sedimentation forces will be identical. Therefore, RPM values are only informative when specified together with the rotor type or radius.

Simple centrifuges can provide maximal accelerations around  $10^4 g$ . Ultracentrifuges (Figure 1.17) are capable of operating at maximal accelerations in the range of  $10^5$ - $10^6 g$ . To reach high angular speeds, ultracentrifuges generate vacuum around the rotor to decrease aerodynamic drag. Ultracentrifuges can be used for preparative (e.g. removal of cell debris from a lysate prior to the isolation of recombinant proteins) or analytical purposes (e.g. investigation of the interactions or the oligomerisation of different biomolecules). (The reader may remember that the first evidence of the semi-conservative replication of DNA also came from density gradient ultracentrifugation experiments.)

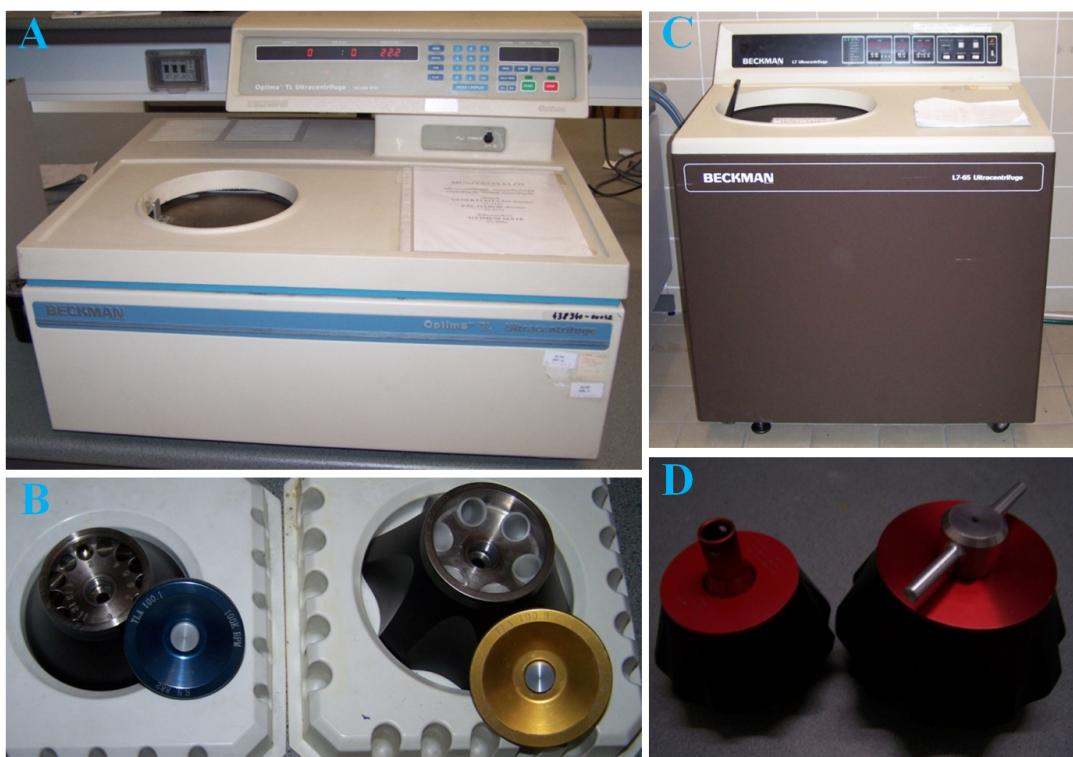


Figure 1.17. A, Micro-ultracentrifuge ( $\text{RPM}_{\text{max}} = 100000 \text{ min}^{-1}$ ,  $\text{RCF}_{\text{max}} = 300000 \text{ g}$ ). B, Rotors compatible with the micro-ultracentrifuge. C, Semi-preparative ultracentrifuge ( $\text{RPM}_{\text{max}} = 70000 \text{ min}^{-1}$ ,  $\text{RCF}_{\text{max}} = 350000 \text{ g}$ ). D, Rotors compatible with the semi-preparative ultracentrifuge.

## 1.10. Other widely used laboratory techniques: spectrophotometry, electrophoresis, chromatography

One of the most often used instruments in laboratories is the spectrophotometer (Figure 1.18). A photometer is a device measuring the intensity of light after passing through a sample (most often, a solution). Absorption of light by the sample will reduce the measured light intensity. A spectrophotometer is a photometer that can measure the absorbance of the sample at different wavelengths of the light. The absorbance depends on the presence and concentration of absorbing molecules (see Chapter 4 for more detail). Hence, spectrophotometers are widely used to determine the concentration of different biomolecules that show characteristic absorption (e.g. proteins, nucleic acids). Sample holders used in spectrophotometers are called cuvettes (Figure 1.18; see also Chapter 4). Depending on the wavelength range of the measurements, cuvettes made of transparent plastics, glass or quartz glass may be applicable. Plastic or glass cuvettes can be applied in the range of visible light (they are transparent as they do not absorb visible light). Quartz cuvettes have low absorption both in the UV and visible range of the spectrum. If the absorbance of many samples needs to be determined simultaneously, plate readers (Figure 1.18, panel C) can be used instead of spectrophotometers. These instruments can dramatically reduce the required working time. Samples are dispensed into the wells of plastic plates (e.g. those of a 96-well microplate with 8 rows and 12 columns of wells), and subsequently placed into the plate reader. Absorbance readings are transferred to and processed by a computer. Spectrophotometry and spectrophotometers will be discussed in more detail in Chapter 4.



Figure 1.18. A, Spectrophotometer. B, Different types of cuvettes: photometric sample holders made of plastic, glass or quartz. C, Plate reader with samples loaded onto a 96-well microplate.

It is often necessary to isolate and purify one or several components from a solution. Various methods are available for the separation of biomolecules. Electrophoretic and chromatographic methods represent two major families of separation techniques (Figure 1.19).

Charged particles in a homogeneous electric field experience different forces. Positively charged particles are repelled from the positively charged electrode (i.e. the anode), while they are attracted by the negatively charged electrode (i.e. the cathode). Negatively charged particles are attracted by the anode and repelled by the cathode. This phenomenon also appears in aqueous solutions. Electrophoresis (Figure 1.19, panel A) exploits the force exerted by an electric field on a charged biomolecule in solution. (For instance, DNA is negatively charged resulting from the deprotonated state of the phosphate moieties of its sugar-phosphate backbone.) Electrophoretic experiments are carried out in gelatinous substances in which the charged biomolecules can be separated according to the differences in their charge, molecular size and shape. Gels are three-dimensionally cross-linked systems of macromolecules with porous structures. Pores are filled by the aqueous solution. As the pore sizes are comparable to the sizes of molecules to be separated, smaller molecules migrate faster than large molecules within the gel. This effect is called molecular sieving. (Electrophoretic methods will be discussed in more detail in Chapter 7.)

Chromatography is another large family of separation techniques. In chromatography, the components of a mixture (e.g. a solution of biomolecules) are separated based on their differential partitioning between two phases: the so-called stationary and the so-called mobile phase. In biochemistry, the stationary phase is generally composed of small beads of a polymeric substance filled into a column (a tube made of glass, plastic or metal with a filter at the bottom), while the mobile phase (or eluent) is the solution carrying the analyte (i.e. the mixture of substances to be separated) through the stationary phase. Separation can be based on the size, shape, charge, isoelectric point, hydrophobicity or specific binding affinity or biological activity of the compounds. Different stationary phases are available to exploit these molecular features. The mobile phase must be chosen according to the properties of the sample and the stationary phase. Automated or semi-automated chromatographic systems typically include vessels for different buffers and solvents, pump(s) responsible for the delivery of the mobile phase, an injector responsible for loading the samples onto the column, various chromatographic columns, detector(s) monitoring the composition of the eluate (i.e. the solution leaving the column), a fraction collector, and a computer for controlling the units and analyzing data (Figure 1.19, panels B-C). Chromatography will be discussed in more detail in Chapter 6.



Figure 1.19. A, Power supply and buffer tank for gel electrophoresis. Gels are placed into the buffer tank. B, High-Performance Liquid Chromatography (HPLC) system. C, Fast Protein Liquid Chromatography (FPLC) system.

## 1.11. Storage of biological samples

Processing or purifying biological samples is often time consuming and labour-intensive. Hence, it is often necessary to store the samples for various time periods. As the samples can be highly divergent (e.g. living eukaryotic cells or bacteria, tissue samples, solutions of proteins or nucleic acids) and show different susceptibility to physico-chemical changes, optimal storage conditions have must be determined in each case experimentally. In an ideal case, the composition, chemical and physical properties and the biological activity of the sample all remain unchanged over time. Unfortunately, this is never true in practice. Therefore it is very important to optimise storage conditions and minimise the time of storage via good timing of the experiments.

A variety of spontaneous reactions and enzymatic processes can occur in any biological sample at different rates, transforming and/or degrading its key components. Some reactions require the components of air as reagents (for instance, oxygen reacts with sulphhydryl groups of proteins). It is relatively easy to protect the samples from such reactions by using a tube supplied with an air-tight cap, which is especially important during long-term storage. Further protection can be achieved by mixing additives directly into the sample. A good example is the use of 2mercaptoethanol as a reducing agent (protecting proteins against oxidation), or the addition of protease inhibitors to slow down proteolysis. Another example is sodium azide ( $\text{NaN}_3$ ), which blocks cellular respiration and thereby inhibits the growth of microorganisms in protein samples.

Another storage optimisation strategy is the removal of potentially reactive components from the sample. The simplest example is the removal of water by lyophilisation, also known as freeze-drying. Biochemical reactions mostly take place in aqueous solutions in which the reagents and catalysts are solvated molecules. Moreover, water is not only a solvent but is involved as a key reagent in many processes including the hydrolysis of peptide bonds in proteins and that of phosphodiester bonds in DNA or RNA. The simple removal of water from the samples can inhibit a variety of unwanted reactions. Lyophilisation is a widely used method for the dehydration of protein and nucleic acid solutions. Freeze-dryers (Figure 1.20) generate vacuum in one or more attached vacuum chambers. When frozen samples are placed into the chambers, ice starts to sublime slowly (solid-to-gas phase transition). Water released into the gas phase re-crystallises on a refrigerated surface outside the sample chamber. The water content of the sample can be very low after a prolonged lyophilisation cycle. However, lyophilisation may not be applicable in every situation. Some proteins lose their structure due to dehydration. Denatured proteins are neither soluble nor functional and, without an efficient renaturation protocol, their subsequent applicability is limited.



Figure 1.20. A, Freeze-dryer with a central vacuum chamber on the top. B, The central vacuum chamber can be expanded by tubes in order to enable the simultaneous handling of multiple samples. C, Freeze-drying of a sample within a Falcon tube.

The rate of chemical reactions can also be reduced simply by lowering the temperature. Samples can be stored in liquid nitrogen ( $-192\text{ }^{\circ}\text{C}$ ), in special laboratory freezers ( $-80\text{ }^{\circ}\text{C}$ ), in simple freezers ( $-20\text{ }^{\circ}\text{C}$ ; also used in households), on melting ice ( $0\text{ }^{\circ}\text{C}$ ) or in a refrigerator ( $4\text{ }^{\circ}\text{C}$ ). Melting ice is typically used for short-term storage of samples, e.g. during their manipulation. Liquid nitrogen is stored in Dewar flasks (Figure 1.21). Dewar flasks minimise heat exchange between the stored liquid and the environment. Liquid nitrogen constantly boils at atmospheric pressure. As boiling is an endothermic process, the temperature of the liquid phase remains constant at  $-192\text{ }^{\circ}\text{C}$  (the boiling point). The temperature can only rise if the whole amount of liquid has been converted to gas. If the Dewar flask was closed in an air-tight manner, evaporated nitrogen could not escape. Consequently, the pressure and the boiling point of the liquid would rise. Moreover, the walls of Dewar flasks are not designed to withstand high pressure. Therefore, samples must never be stored in an air-tight sealed container filled with liquid nitrogen, as this may result in an explosion. Samples are stored within cryogenic storage boxes in liquid nitrogen racks and canisters (Figure 1.21, panel B). In these applications, special cryotubes (Figure 1.21, panel D) are used instead of Eppendorf tubes, as ordinary plastics break easily at  $-192\text{ }^{\circ}\text{C}$ . Purified proteins and nucleic acids—or even living cells—can be stored in liquid nitrogen. Cells must be frozen quickly to avoid the formation of ice crystals, which would damage membranes. Flash-freezing in liquid nitrogen is excellent for this purpose. Cryoprotectants (e.g. glycerol) can also be used to ensure full protection against ice formation. Frozen cells remain viable for years. Flash-frozen samples can also be stored in laboratory freezers at  $-80\text{ }^{\circ}\text{C}$  (Figure 1.21, panel E). (This temperature can be appropriate for frozen bacterial cells, but not for eukaryotes.) Tissue samples, protein solutions, and other samples might be stored either at  $-80\text{ }^{\circ}\text{C}$  or  $-20\text{ }^{\circ}\text{C}$ , depending on the storage time period and the properties of the sample. Refrigerators ( $4\text{ }^{\circ}\text{C}$ ) are appropriate for buffer solutions and for some chemicals, but not for the long-term storage of biological samples.

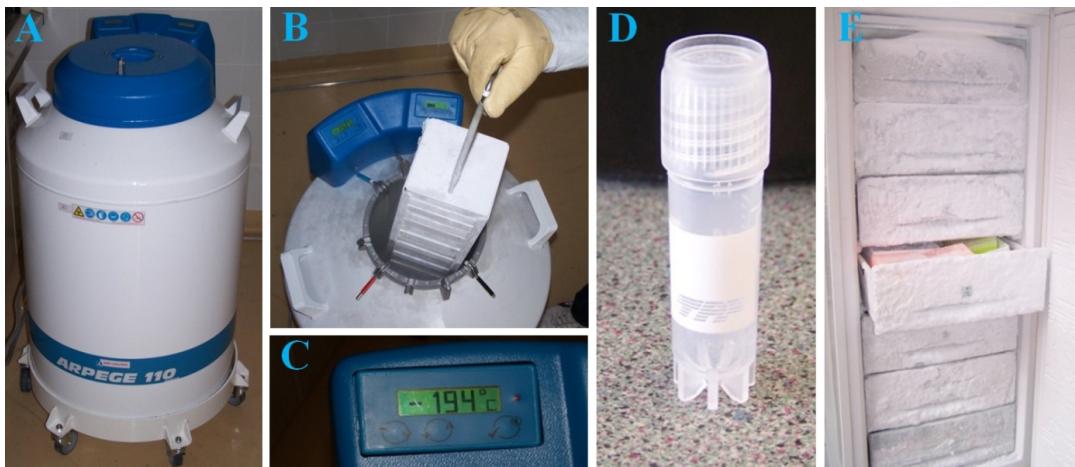


Figure 1.21. A, Dewar flask filled with liquid nitrogen. B, Liquid nitrogen storage rack with cryogenic storage boxes. C, Inner temperature indicated on the top of the Dewar flask. D, Cryotube for the storage of samples in liquid nitrogen. E, Laboratory freezer ( $-80^{\circ}\text{C}$ ).

---

# Chapter 2. Units, solutions, dialysis

by István Venekei

## 2.1. About units

Units/measures are used to specify quantities. Every unit has a well-defined standard basic value, which is used as a reference in measurements. The units used in scientific practice constitute a scientifically established system, SI (Système International). Although only SI units are “official”, numerous non-SI units are still widely used (calorie, Ångström, Celsius etc.).

We distinguish two types of quantities with two corresponding types of SI units. The base quantities cannot be derived from other quantities. Many of these are used in biochemistry (names of units and their symbols are given in brackets) to quantify the mass (gram, g), the length (meter, m), the time (second, s), the temperature (kelvin, K) and the charge (coulomb, C). Much larger is the group of derived quantities. Of these, the various units of concentration (to characterise the abundance of materials in solutions and other mixtures), volume and energy are the ones most frequently used in biochemistry.

It is important to note that the terms “mass” and “weight” are often used interchangeably as “alternatives”. Technically, however, they have different meanings. The mass is the total quantity of matter in an object, which comes from the mass of all of its protons and neutrons, although not simply additively. Weight is a measure of the gravitational force exerted on an object. As the mass of protons and neutrons is the unit mass, their total mass in an object (e.g. the mass of a molecule, the molar mass) is a unitless number (e.g. the molar mass of water is 18). In other words, the molar mass is a relative number that would reflect the number of protons and neutrons in an atom or molecule—if the masses of protons and neutrons were additive. We make mass measurable by expressing it in grams because this way we can handle it as weight and measure it with a balance, the device developed for this purpose. Thus, the measurement of weight means the measurement of mass of physical objects—not that of atoms or molecules, but e.g. that of 18 mL water. At the same time, this way we make the measured mass dependent on the place of measurement. (For example, 18 g of H<sub>2</sub>O is not the same amount (mL or mass) of water at the poles and the equator, although the difference cannot be demonstrated by a traditional lever-arm balance due to its principle of operation.) Thus the practical aspect of the relationship between mass and weight can be summarised as the following: we can measure mass only as weight, and weight is the effect of gravity exerted on mass. Weight is therefore determined by both the mass and gravitational forces, and an object can be weightless but never massless.

The mass unit to express the “size” of large molecules is Dalton (Da), used mainly in biochemistry. This shows how many times the mass of a macromolecule (e.g. a protein) is larger than the mass of a hydrogen atom (more precisely, a proton or a neutron).

An important quantity is the mole, which is special to chemistry. The atomic and molar masses expressed in grams correspond to one mole (e.g. 18 grams of water). One mole of a substance contains Avogadro’s number ( $6.022 \times 10^{23}$ ) of particles (atoms, molecules, ions or even photons). Due to the definition of the mole, and because the unit of mass (“weight”) is the gram, we express atomic and molecular masses (“weights”) in grams. These are called atomic weight and molecular weight (gram atomic weight, gram molecular weight). We get the number of moles by dividing the quantity of material present by the atomic or molar mass, both expressed in grams (number of moles (m) = mass/molar mass).

## 2.2. Numeric expression of quantities

### 2.2.1. The accuracy of numbers, significant figures

Except the numbers we obtain by counting (the number of items), all numbers are inexact because their values are determined with a certain degree of uncertainty. The source of uncertainty is the limited capacity of either the measuring device (due to the flaws in its construction or improper calibration) or the measuring person (due to improper skills). We know the degree of uncertainty only in the case of our own measurements. Without a detailed

discussion of the problem, the following is useful to know. In order to estimate and give the uncertainty of data, we need two (related) pieces of information: the order of magnitude and the significant figures.

The significant figures are determined by the achievable precision (exactness) of the measurement. Usually, the last figure of a measured value bears uncertainty, i.e. this figure should be considered as “estimated” (significant figures = all certain digits + one estimated digit). The value—and thus the significance—of zero digits in numbers is dependent on their position. Zeros are not significant if they are at the beginning of a number (leading zeros), or if they are at the end (trailing zeros) without a decimal point in the number (even though such zeros carry information about the magnitude). In contrast, zero digits are significant if they are at the end of a number containing a decimal point (because they show the exactness of measurement), as well as if they are inside the number (confined zeros, located between nonzero digits).

We must consider these aspects and the achievable precision of measurements when we simplify our numbers by rounding them up or down. We usually do this when we recalculate (mathematically transform) measured data or when we obtain them by calculation. (For example, a given amount of substance should be dissolved in a calculated solvent volume of 5.4786 mL. If our measuring device is calibrated with 0.1 mL division, then the figures “6” and “8” (corresponding to 0.0006 and 0.008 mL, respectively) are immeasurable, and figure “7” (corresponding to 0.07 mL) is uncertain. In this case, the desired volume can be approximated by pipetting 5.5 mL, considering the above uncertainties.) Carefully performed rounding is always recommended because series of digits of meaningless length make calculations unnecessarily difficult and bear the danger of calculation errors.

## 2.2.2. Expression of large and small quantities: exponential and prefix forms

In most cases, measured quantities differ from the unit of the given quantity by several orders of magnitude. In these cases, the length of the number cannot be substantially reduced by rounding. Two procedures are in use to avoid the writing of many zeros in the case of very large or small numbers, which would be uncomfortable and may be a source of error. During one of these procedures, large or small numbers are converted into their exponential form while keeping the unit of quantity (i.e. the scale). The other procedure replaces zeros with a prefix, i.e. it changes the unit of quantity (the scale). For example, we can write a quantity of 0.0000043 litre (L) as  $4.3 \times 10^{-6}$  L (exponential form) or 4.3  $\mu$ L (microlitre, prefix form) according to the first and the second procedure, respectively. In the latter expression, the “ $\mu$ ” sign (read as “micro”) is called a prefix, which reflects the degree of change in the scale relative to the basic unit of quantity (in this example, six orders of magnitude downward)—this way keeping the magnitude information that was originally conveyed by the “eliminated” (replaced) zeros.

The definition of prefixes specifies the relationship between the two procedures (Figure 2.1). Official SI prefixes are defined at steps of three orders of magnitude above or below the unit of the given quantity, i.e. at  $10^3$ ,  $10^6$ ,  $10^9$  and  $10^{-3}$ ,  $10^{-6}$ ,  $10^{-9}$  etc. Other (non-SI) prefixes including the deci ( $10^{-1}$ ), the centi ( $10^{-2}$ ) and the hecto ( $10^2$ ) are not official, although widely used in civil life outside the laboratory. A mixture of prefix and exponential forms within an expression should be avoided (e.g.  $5.2 \times 10^{-4}$   $\mu$ g is expressed correctly as 0.52 ng), similarly to the way in which the normal and exponential forms of numbers are used. (In exponential notation we do not write values larger than 10 or smaller than 1 such as  $3100 \times 10^4$  or  $0.12 \times 10^{-2}$ , because these numbers are correctly expressed as  $3.1 \times 10^7$  and  $1.2 \times 10^{-3}$ , respectively). The use of prefix or exponential forms of numbers is often insufficient to reduce their length. Therefore, rounding (see above) is also a necessary simplification.

| order of magnitude    | prefix | symbol |
|-----------------------|--------|--------|
| $10^{12}$             | Tera   | T      |
| $10^9$                | Giga   | G      |
| $10^6$                | Mega   | M      |
| $10^3$                | kilo   | k      |
| $10^0$ (=1, the unit) | –      | –      |
| $10^{-3}$             | milli  | m      |
| $10^{-6}$             | micro  | $\mu$  |
| $10^{-9}$             | nano   | n      |
| $10^{-12}$            | pico   | p      |
| $10^{-15}$            | femto  | f      |
| $10^{-18}$            | atto   | a      |

Figure 2.1. Magnitude scales, prefixes and their relationship

## 2.3. About solutions

### 2.3.1. Definition of solutions and their main characteristics

True solutions are homogeneous mixtures of two or more substances. The substance present in the largest quantity is the solvent, and all other components are called solutes. Homogeneous mixtures form a single phase by eye. (For example, we cannot discern two liquids within a solution, unlike in emulsions—which, therefore, are not true solutions.) In true solutions, the atoms, molecules or ions of the solute are dispersed in the solvent. Special cases are those solutions in which the dispersion of the solute is not molecular (or atomic or ionic) but we cannot visually discern the solute (e.g. micelles). These are not molecular dispersions; yet they do not form a distinct phase. Such solutions are called colloids. The other definition of colloid solutions is that the particle size of the solute is in the range of 1 to 1000 nanometres. The solutions of most proteins fall within this category; i.e. these are colloid solutions, albeit proteins in these solutions are molecular dispersions.

In laboratory practice, mixtures in a liquid state are usually called solutions. In the biochemical laboratory we generally prepare aqueous solutions, i.e. solutions in which the solvent is water. The main reason for this is that we usually work with proteins and other biomolecules that can stably adopt their native conformation in an aqueous environment. We often prepare solutions in organic solvents in order to dissolve smaller compounds (e.g. substrates of enzymes) or in the case of HPLC procedures. Most of the organic solvents used in biochemical experiments are miscible with water. As a semi-quantitative characterisation of solutions, we can distinguish unsaturated, saturated and supersaturated solutions. These solutions contain less, the same, or more than the maximum possible amount of solute, respectively. (These characteristics, together with solubility, are temperature dependent.)

### 2.3.2. Quantitative description of solutions, concentration units

We specify the quantity of components in a mixture quantitatively as their ratio. The precise expression of this can be achieved via concentration units. In liquid-state solutions, it is usually a certain quantity of the solution to which the quantity(ies) of its component(s) are compared, depending on the concentration unit. For instance, in the case of molar concentration (molarity, M, mol/dm<sup>3</sup> or mol/L), the most widely used concentration unit in biochemistry, we refer to the number of moles of a given substance in one litre of solution. (For instance, 0.2 M means 0.2 moles of a substance in one litre of solution.) Various percentage forms (%; reference to 100 units of something) are also widely used in biochemistry, although they are not SI units. In the case of weight-by-volume and volume-by-

volume percentages (w/v% and v/v%, respectively) we refer to 100 mL of solution, whereas in the case of weight-by-weight percentage (w/w%) we refer to 100 g of solution. In the case of v/v% the volume of the solute, while in the case of w/v% and w/w% the weight of the solute is the basis of reference. (For example, a 15 w/v% solution contains 15 g of solute in 100 mL of solution.) During conversions from or into w/v%, one must also consider the density of the solution (unless it is one g/mL). The density of dilute solutions (up to several %) generally deviates only negligibly from that of the solvent. Therefore, the w/w% and w/v% values of such solutions are practically identical (i.e. 3 w/v% corresponds to ca. 3 w/w%). It is important to note that, for practical reasons, the “biochemical” concentration units mg/mL and µg/mL are also used in the case of protein and other macromolecular solutions in biochemical experiments.

### 2.3.3. Preparation of solutions

As homogeneity is a basic criterion, it is important to know the solubility of solutes, which is a temperature-dependent property. Information on the solubility of compounds can be obtained from handbook tables (for inorganic compounds) or from the manufacturer (product catalogues of organic compounds). Besides temperature, solubility is influenced by the acidity (pH) of the solvent due to the acid-base character of the solute. Bases tend to dissolve better in acidic solutions, whereas acids do in basic solutions. Therefore, one often needs to change the pH of solutions accordingly. This can be most simply achieved by the addition of an inorganic acid or base (HCl, NaOH, KOH etc.) or by using a buffer system. In both cases, one must be aware of the fact that the prepared solution is a multi-component one, i.e. it contains a pH-setting substance in addition to the substance(s) to be dissolved.

In order to achieve the desired concentration precisely, the measuring range of the devices used (cylinders, pipettes, balances etc.) must be in the range of the quantity to be measured. (For example, a 70-mL solution should not be prepared in a measuring cylinder of a maximal volume of 250 mL, but in one with a 100-mL maximal volume.) As the most widely used concentration units (molarity, w/v% and v/v%) refer the amount of solute in a given volume of the solution (see above), it is the volume of the *complete solution* that must be set to the desired (calculated) value—thus, one must consider not only the solvent but the solute(s) as well. For instance, when preparing one litre of a 4-M CaCl<sub>2</sub> solution, the required amount of CaCl<sub>2</sub> must be added not to one litre of water but to much less (e.g. to 500-600 mL, regardless of the condition whether it can dissolve the applied amount of CaCl<sub>2</sub>), and subsequently complemented with additional water to fill up the volume to one litre. This is the only way to ascertain that the total volume of the solution (CaCl<sub>2</sub> and water together) will be one litre. During this procedure, one must take into account that the solute also contributes to the total (final) volume of the solution. Another reason why one must add the solute(s) to an amount of solvent which is less than the final volume might be that one will need to set the pH subsequently in order to enhance solubility and/or to reach a desired pH value. The volume of acid or base solutions used for pH setting is usually not known in advance—and is not negligible. If the solute was initially dissolved in the final desired volume, one would exceed this volume during pH setting (even if the volume of the solute is negligible) and, thus, the solution would not have the desired (calculated) concentration but a smaller (and mostly unknown) one. Thus, according to the correct procedure, one must take into consideration the fact that the final volume will contain the volume of the buffering component.

During mixing the components of a solution, it is often important to be aware of the appropriate and/or safe order of addition. The dilution of concentrated sulphuric acid is a good example of necessary precaution: it is sulphuric acid that must be added to water (and not *vice versa*) because otherwise water—which boils due to the high heat of solvation—would sprinkle around the hot and corrosive acidic solution, causing serious injury. The appropriate order of dissolving solid solutes is the addition of the solute into the solvent (preferably into less than the final volume aimed)—and not *vice versa*, pouring solvent onto the solute. This knowledge is of high practical value particularly in the case of dissolving proteins, which have a tendency to stick to the bottom of the dish, thereby making dissolution substantially slower.

Dissolution of substances can be enhanced in various ways including mixing, increasing the temperature, applying sonication or by performing dissolution in the appropriate order (e.g. in the case of the Coomassie solution used to stain proteins in acrylamide gels). Mixing is a commonly used and generally “harmless” method. However, special care must be taken during sonication and especially during heating: one must consider the (heat) stability of the dissolved compound. Proteins are notoriously sensitive molecules; so are a number of simpler organic compounds, too. If we need to set the pH, we must consider the pH sensitivity of the solute. Undesired events and reactions can be prevented by the avoidance of the addition of large quantities or high concentrations of acids or bases. Instead, it is better to use more dilute acid or base solutions and add these in small portions. Sometimes,

special requirements must be met such as sterility and/or the sensitivity of the solute to light or oxygen. If the solute is heat sensitive, the solution cannot be sterilised by heat. In these cases, filtration must be applied.

Prepared solutions must appear clear following the complete solution of the solute(s). If this is not the case, the solution must be filtered. (Obviously, one should check if it is not a part of the solute that fell out of the solution.) To this end, one can use traditional tools (filter papers, glass filters) or disposable filters (or filter units) of various materials and pore sizes according to the expected purity of the solution. For example, the solutions used in FPLC, HPLC or in optical devices (photometers, fluorometers) may not contain any floating particles because these are harmful to the device (HPLC, FPLC) or may severely interfere with the measurement.

Solutions are generally stored in glass or plastic containers (bottles, flasks or tubes) that must be tightly closed to prevent changes in concentration due to the slow evaporation of the solvent. (This is much more difficult to achieve in the case of more volatile organic solvents. Here, the use of glass screw-cap vials with Teflon cover is advisable.) For this reason, bakers, Erlenmeyer or other flasks and measuring cylinders (used during preparative procedures) are not useful for this purpose as they do not meet the above criterion of safe storage even in the case of aqueous solutions. Solutions of most of inorganic and organic compounds are generally stored at room temperature. However, in the biochemical laboratory, some solutions must be stored in a refrigerator or a freezer. The latter way of storage is needed to achieve chemical and/or microbiological stability. Unless they are sterilised, solutions of proteins and even those of simple inorganic compounds (e.g. phosphate salts) provide a favourable environment for the growth of bacteria and fungi. (The appearance of these microorganisms is indicated by the increased turbidity and unpleasant smell of the solution. Algae may even appear in distilled water!) Some enzymes (e.g. proteases) may also slowly lose their activity. Cooling—generally below freezing temperature—cannot completely prevent but at least substantially slow down such processes.

Dishes and containers of solutions must be provided with appropriate labels that must remain readable throughout the whole time period of the potential usage of the solution. In the case of solutions containing one or a few different solutes, the label should include the name (or the chemical formula) and the concentration of the solute(s). In many cases, specific names are given to more complex multi-component solutions. These names must be unambiguous at least within the laboratory (e.g. “activation solution”, “sonication buffer”, 10x reaction mixture). Names of many solutions are international if they are widely used during a common laboratory procedure. Such typical names include those of reagents (e.g. the Bradford reagent used to measure protein concentration) or culture media (e.g. LB or 2YT used to grow bacteria). The time of preparation of a solution may be of importance and thus it should be included in the label—the name of the person who prepared the solution is also an option.

## 2.4. Dialysis

### 2.4.1. The principle of dialysis

Dialysis is a procedure employed in a number of cases when a change in the concentration or composition of solutes is necessary. In the biochemical practice, dialysis is often used to alter the concentration of salts and/or small molecules in protein solutions—usually aimed at decreasing the concentration of these solutes. However, the composition of the solution can also be changed in additional ways.

Dialysis is based on diffusion during which the mobility of solute particles between two liquid spaces is restricted, mostly according to their size. (In rarely used versions of dialysis, restriction of diffusion via polarity or charge is also possible.) Size restriction is achieved by using a porous material, usually a semi-permeable membrane called dialysis membrane. This membrane is permeable only for particles below a certain size. In the biochemical laboratory, this membrane is mostly a hose made from transparent material (also called dialysis bag) that can be tightly closed (tied) at its ends (Figure 2.2). The solution to be dialysed (with a volume  $V_1$ ) is loaded into the dialysis bag. The dialysis bag is then placed into a dialysis solution (with a volume  $V_2$ ) that is stirred slowly to aid the diffusion of the subset of solutes that can be released through the bag membrane, in order to achieve equilibrium between solute concentrations in the two liquid spaces. If the difference in volume between the two spaces is large ( $V_2 \gg V_1$ , e.g.  $V_2 = 10 \text{ L}$  and  $V_1 = 0.1 \text{ L}$ , a 100-fold difference), the onset of the equilibrium will lead to a very significant dilution of the small solutes that were initially inside the bag (their concentration will change by a factor  $V_1/(V_1+V_2)$ , in this case  $\ll 1$ ), with only a slight change in the concentration of small solutes in the outside solution (by a factor  $V_2/(V_1+V_2) \approx 1$ ), whereas the concentration of the molecules inside the bag that cannot penetrate the membrane remains almost completely unchanged (see in details below).

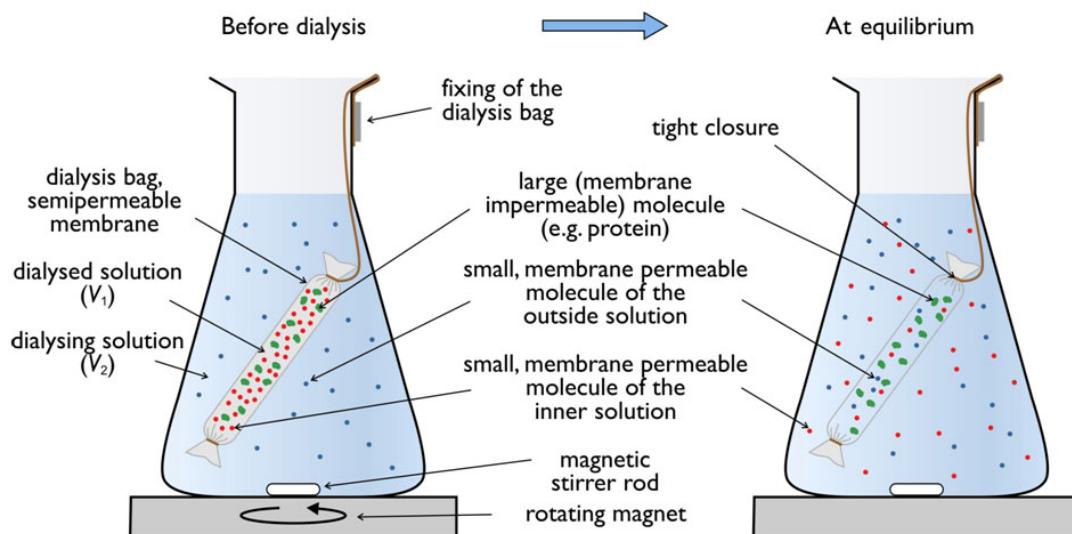


Figure 2.2. Dialysis in the biochemical laboratory practice

## 2.4.2. Practical aspects and applications of dialysis

The efficiency of dialysis, i.e. the extent to which the concentration and composition of the inside solution can be changed, is an important aspect. It follows from the above description of dialysis that the efficiency of dialysis largely depends on the difference between the volumes of the inside and outside liquid spaces. This is why we generally seek to use as large volume ( $V_2$ ) of the dialysing solution as possible. However, the efficiency of dialysis can be further increased by performing multi-step dialysis by exchanging the outer solution after the equilibrium has been reached. In this case, the attainable dilution of the inside solution will be  $[V_1/(V_1+V_2)]^n$  where  $n$  is the number of steps. It is easy to see that efficiency that can be achieved by applying a two-step dialysis at a 50-fold volume difference is much higher than the efficiency of a single-step dialysis at a 100-fold volume difference.

The speed of dialysis can be increased not only by stirring the outside solution but also by increasing the surface/volume ratio of the inside solution, as the flux of diffusion is linearly proportional to the cross-section. It is, therefore, more practical to choose a narrower and longer tube than a wider and shorter one.

The semi-permeable membrane can be crossed not only by salts and small molecules but also by solvent particles (in most cases, water). The direction and extent of the net solvent flow is determined by the difference between the total concentration of solutes in the inside and outside solutions such that the solvent migrates from the less to the more concentrated solution (with regard to solutes). This way the equilibrium concentration of the solute(s) of the inside solution that cannot cross the membrane will be influenced also by the diffusion of the solvent. As the solute(s) that cannot cross the membrane also contribute to the total concentration of the inside solution, the net direction of solvent migration will almost always point towards the inside solution. Therefore, the volume of the inside solution will increase, thereby selectively decreasing the concentration of the membrane-impermeable solute(s)—but not that of the membrane-permeable ones, even if the relative increase in the volume is large. However, the relative increase in the volume is generally not large because (i) the concentration of the large impermeable solutes is low (much lower than that of the small permeable ones) (ii) the dialysis tube is largely unable to increase its volume. The occasional small (5-20 %) volume increase of the inside solution is associated with the compression of air above the liquid phase that was originally enclosed in the bag. Taken together, the decrease in the concentration of the large solutes (proteins) is usually negligibly small. The increase in the volume of the inside solution is remarkable from a technical point of view because it is accompanied by a (sometimes substantial) elevation of the pressure. Therefore, if there is a hidden “weakness” somewhere in the material of the membrane, the elevation of pressure may lead to bursting of the bag and, as a consequence, the complete loss of the dialysed material (e.g. protein preparation). To avoid this “catastrophe”, it is recommended to perform a pressure test on the bag in its water-filled state. The other risk associated with pressure elevation occurs during the opening of the bag after completion of dialysis. In the absence of necessary care, the pressurised inside solution can sprinkle out, causing loss of material.

In the biochemical laboratory practice, solutions of proteins are generally dialysed following fractionated ammonium sulfate precipitation (detailed in Chapter 5) as well as before or after ion exchange chromatography (detailed in Chapter 6). A size selectivity (size exclusion or cut-off) specified as 4 or 11 kDa means that the pores of the dialysis membrane are impermeable for particles larger than 4 or 11 kDa, respectively.

Besides the biochemical laboratory, dialysis is utilised in the field of life sciences also for therapeutic purposes during haemodialysis, i.e. in artificial kidneys. The principal difference between these two applications is that, in the artificial kidney, dialysis is executed under continuous counter-flow of the two solution spaces: both the inside solution (the blood of the patient) and the outside solution are pumped. Thus, in such a setting, also the inside liquid space is “open”: it is not in a “bag” but flows inside a tube. Moreover, in order to increase the flux of diffusion, a large number of capillary tubes are employed in a bundle (which is actually the artificial kidney) by which the surface/volume ratio is increased enormously. The composition of the outside dialysing solution is very special as it must meet special requirements. In addition, the artificial kidney equipment is a very special apparatus because it must be able to ensure the appropriate pressure and temperature while the blood entering the body of the patient must be free of entrapped air bubbles that could lead to lethal consequences.

---

# Chapter 3. Acid-base equilibria, pH, buffer systems

by István Venekei

## 3.1. Ionisation equilibria of acids and bases in aqueous solutions

There are two main groups of solutes (dissolved substances). One group of solutes increases the conductivity of the solvent because it dissociates (releases ions) upon dissolution. Members of this group are called electrolytes. The other group (called non-electrolytes, e.g. glucose) does not increase conductivity. Such compounds do not dissociate. Electrolytes are classified into two groups: those of the strong and the weak ones. The dissociation of compounds in the former group is essentially complete (100 %). These are always ionic compounds (they contain ionic bonds). Weak electrolytes can be ionic or non-ionic compounds. The dissociation of the latter group is associated with the formation of ions; therefore, the dissociation of these electrolytes is called ionisation. At equilibrium, the dissociation of weak electrolytes is usually shifted towards the non-dissociated/non-ionised form. Weak acids release protons into the solution according to the following equation:



where AH and A<sup>-</sup> are the symbols of the weak acid and its anion, respectively.

The law of mass action (mass effect) describes the equilibrium in dilute solutions. In the above case this will be (square brackets indicate molar (mol/L) concentration):

$$\frac{[\text{A}^-] \times [\text{H}_3\text{O}^+]}{[\text{AH}] \times [\text{H}_2\text{O}]} = K_{\text{eq}} \quad (3.2)$$

or, more simply

$$\frac{[\text{A}^-] \times [\text{H}^+]}{[\text{AH}]} = K_a \quad (3.3)$$

where K<sub>eq</sub> is the equilibrium constant of the reaction, [H<sub>3</sub>O<sup>+</sup>] = [H<sup>+</sup>], and K<sub>a</sub> = K<sub>eq</sub> × [H<sub>2</sub>O] is the ionisation constant of the weak acid ([H<sub>2</sub>O] ~ 55,5 M, constant).

The K<sub>a</sub> values with their negative logarithms (pK<sub>a</sub> values) of some weak electrolytes are shown in Figure 3.1. As the K<sub>a</sub> values are known (Figure 3.1) and [A<sup>-</sup>] = [H<sup>+</sup>], the [H<sup>+</sup>] and, in turn, the pH can be calculated.

|   | K                      | pK    |
|---|------------------------|-------|
| H <sub>2</sub> CO <sub>3</sub> /HCO <sub>3</sub> <sup>-</sup>               | 4.30×10 <sup>-7</sup>  | 6.37  |
| HCO <sub>3</sub> <sup>-</sup> /CO <sub>3</sub> <sup>2-</sup>                | 5.60×10 <sup>-11</sup> | 10.25 |
| H <sub>3</sub> PO <sub>4</sub> /H <sub>2</sub> PO <sub>4</sub> <sup>-</sup> | 7.52×10 <sup>-3</sup>  | 2.12  |
| H <sub>2</sub> PO <sub>4</sub> <sup>-</sup> /HPO <sub>4</sub> <sup>2-</sup> | 6.23×10 <sup>-8</sup>  | 7.21  |
| HPO <sub>4</sub> <sup>2-</sup> /PO <sub>4</sub> <sup>3-</sup>               | 2.22×10 <sup>-13</sup> | 12.67 |
| NH <sub>4</sub> OH/NH <sub>4</sub> <sup>+</sup>                             | 1.77×10 <sup>-5</sup>  | 4.75  |
| CH <sub>3</sub> COOH/CH <sub>3</sub> COO <sup>-</sup>                       | 1.76×10 <sup>-5</sup>  | 4.75  |
| R-NH <sub>3</sub> <sup>+</sup> /R-NH <sub>2</sub>                           | 1.58×10 <sup>-11</sup> | 10.80 |

Figure 3.1. Ionisation constants (K) and pK values of several weak acids and bases

Weak bases undergo ionisation when they release  $\text{OH}^-$  ions into the solution:



where BOH and  $\text{B}^+$  symbolise the weak base and its cation, respectively.

The description of the equilibrium according to the law of mass action will yield

$$\frac{[\text{B}^+] \times [\text{OH}^-]}{[\text{BOH}]} = K_b \quad (3.5)$$

where  $K_b = K_{\text{eq}} \times [\text{H}_2\text{O}]$  is the ionisation constant of the weak base.

As the  $K_b$  values are known and  $[\text{B}^+] = [\text{OH}^-]$ , the  $[\text{OH}^-]$  and, in turn, the pOH can be calculated. The pH can be calculated by using the relationship

$$14 = \text{pH} + \text{pOH} \quad (3.6)$$

Another way of ionisation of weak bases occurs when they bind protons originating from the solution (e.g. released during the ionisation of water), thus shifting the equilibrium of the ionisation. This protonation is typical during the ionisation of organic bases including amines and imines. In the case of an amine with a general formula  $\text{R}-\text{NH}_2$ , the reaction will be (see also Exercise 1):



As it is apparent, this process is actually a hydrolysis. Thus, the equilibrium can be described by the following equation:

$$\frac{[\text{R}-\text{NH}_3^+] \times [\text{OH}^-]}{[\text{R}-\text{NH}_2]} = K_h \quad (3.8)$$

where  $K_h = K_{\text{eq}} \times [\text{H}_2\text{O}]$  is the hydrolysis constant.

According to the Bronsted-Lowry theory of acids and bases, an acid is a proton donor and a base is a proton acceptor. The acid with its anion or the base with its cation form a conjugate pair. Thus, in our example, the  $\text{R}-\text{NH}_2$  form is the base (proton acceptor) while the  $\text{R}-\text{NH}_3^+$  form is the acid (proton donor). With these considerations, we can describe the behaviour of weak organic bases as that of weak acids as follows:



By describing the equilibrium using the law of mass action, we get an equation that is essentially identical to Equation 3.3:

$$\frac{[\text{R}-\text{NH}_2] \times [\text{H}^+]}{[\text{R}-\text{NH}_3^+]} = K_a \quad (3.10)$$

where  $K_a$  is the acid dissociation constant, which is related to the hydrolysis constant of the basic form according to the equation  $K_a = 10^{-14} \times (K_h)^{-1}$  (see also Exercise 2).

Thus, according to Equations 3.9 and 3.10,  $\text{R}-\text{NH}_3^+$  is a weak organic acid. However, as seen in Figure 3.1, the protonated form of a weak organic base is usually a weaker acid than most of the weak organic acids (Figures 3.1 and 3.2). (The weaker an acid or a base, the smaller the dissociation constant.)

|               | $-\text{COOH}_{(\alpha)}$<br>( $\text{pK}_{\text{a}1}$ ) | $-\text{NH}_3^+_{(\alpha)}$<br>( $\text{pK}_{\text{a}2}$ ) | R<br>(residue) |
|---------------|--|--|----------------|
| glycine       | 2.34   | 9.60   | —              |
| arginine      | 2.17   | 9.04   | 12.48          |
| lysine        | 2.18   | 8.95   | 10.53          |
| tyrosine      | 2.20   | 9.10   | 10.10          |
| cysteine      | 1.80   | 10.80  | 8.30           |
| histidine     | 1.82   | 9.17   | 6.00           |
| glutamic acid | 2.19   | 9.67   | 4.25           |
| aspartic acid | 1.88   | 9.60   | 3.65           |

Figure 3.2. The  $\text{pK}_\text{a}$  values of glycine and amino acids containing an ionisable group in their side chain

Due to the principle of mass action, the degree of ionisation of a weak acid or a base can be influenced by the pH of the solution (e.g. via addition of a strong acid or base). On the addition of a strong acid (e.g. HCl) to a weak acid, the degree of ionisation of the latter will decrease. By the addition of a sufficient amount of a strong acid, we can reach a pH value at which the ionisation of the weak acid becomes negligible, i.e. essentially every molecule will be in the AH state.

If, in contrast, we add a strong base (e.g. NaOH) to the solution, we can achieve complete ionisation whereby all molecules of the weak acid will be in the  $\text{A}^-$  state.

On the addition of a strong acid or base, we actually perform titration: on acid or base addition, we titrate the  $\text{A}^-$  or the AH form, respectively. During this process, we can always set a pH where the degree of ionisation of the weak acid or base will just be 50 %. The equilibrium at this titration point is described by Equations 3.3, 3.5 or 3.10, depending on whether the substance is a weak acid or base. At this point of weak acid titration,  $[\text{AH}]$  will equal  $[\text{A}^-]$ . However, at the same time,  $[\text{A}^-]$  will not equal  $[\text{H}^+]$ , because the pH (the  $[\text{H}^+]$ ) will be determined by the strong acid or base added (since the amount of protons dissociated from AH is negligibly small relative to that originating from the strong acid). Due to this condition and also because  $[\text{AH}]/[\text{A}^-] = 1$ , Equation 3.3 describing the ionisation of a weak acid will be reduced to the following:

$$\text{K}_\text{a} = [\text{H}^+] \quad (3.11)$$

Transforming this equation to a logarithmic form and multiplying it by -1 we get:

$$\text{pK}_\text{a} = \text{pH} \quad (3.12)$$

since  $-\lg \text{K}_\text{a} = \text{pK}_\text{a}$ , and  $-\lg [\text{H}^+] = \text{pH}$ . From Equation 3.12 we can deduce a meaning and an interpretation of the value of  $\text{pK}_\text{a}$  (and  $\text{pK}_\text{b}$ , and thus, generally, a pK): the pK is a pH or pOH at which the degree of ionisation of a weak acid or base, respectively, is just 50 %. (The weaker the acid or base, the larger the pK.)

If the amount of the strong acid or base added to a weak acid or base sets the pH to a value at which the degree of ionisation of the weak acid or base is not 50 %, then—taking the case of a weak acid—we will not be able to reduce Equation 3.3 because in this case  $[\text{AH}]$  will not equal  $[\text{A}^-]$ . By transforming the equation logarithmically, multiplying it with -1 and then expressing the pH, we arrive at the following:

$$\text{pH} = \text{pK}_\text{a} + \lg \frac{[\text{A}^-]}{[\text{AH}]} \quad (3.13)$$

This relationship between the  $\text{pK}_\text{a}$  and pH is called the Henderson-Hasselbalch equation, which simplifies the calculation of pH in solutions that are not single-component acid-base systems—in other words, in which  $[\text{A}^-]$  does not equal  $[\text{H}^+]$  (see also Exercise 3). (Buffers are typically such systems, see below.) Applying the Bronsted-Lowry theory of acids and bases to Equation 3.13, we can write its more general form:

$$\text{pH} = \text{pK}_a + \lg \frac{[\text{base}]}{[\text{acid}]} \quad (3.14)$$

In our example, the two components of the acid-base system are  $\text{A}^-$ , the weak base, and  $\text{AH}$ , the weak acid conjugated to it. (It must be noted that Equation 3.14, although it is essentially identical to Equation 3.3, can be applied only in the case of two-component systems, see also Exercise 4).

## 3.2. pH-stabilising acid-base systems (buffers) and the influence of pH on ionisation

Buffer systems can be defined according to their composition and operation. *According to their composition*, buffers are two-component systems that contain both a weak acid and a weak base. (Of the above examples, such a system can be  $\text{AH}$  and  $\text{A}^-$ ,  $\text{BOH}$  and  $\text{B}^+$  as well as  $\text{R-NH}_3^+$  and  $\text{R-NH}_2$ .) *According to their operation*, they are acid-base systems that are capable of keeping the pH relatively constant because, upon the addition of either a strong acid or base, they convert these into weak ones. Taking the last conjugated pair above as an example, the operation can be illustrated using Equation 3.9 in the following way:

If we add  $\text{HCl}$ , it will react with the  $\text{R-NH}_2$  form (the weak base) and convert it into the  $\text{R-NH}_3^+$  form. In this reaction,  $\text{HCl}$  (the strong acid) is essentially consumed, while an equivalent amount of  $\text{R-NH}_3^+$  (the weak acid) forms. That is, the strong acid will be “exchanged” into a weak one.

If we add  $\text{NaOH}$ , it will react with the  $\text{R-NH}_3^+$  form (the weak acid) and convert it into the  $\text{R-NH}_2$  form. In this reaction,  $\text{NaOH}$  (the strong base) is essentially consumed, while an equivalent amount of  $\text{R-NH}_2$  (the weak base) forms. That is, the strong base will be “exchanged” into a weak one.

We can readily calculate the pH of a buffer and its changes upon the addition of a strong acid or base by using the Henderson-Hasselbalch equation (Equation 3.14). From the equation it is apparent that a buffer can provide the most efficient pH stabilisation against both acidic and basic shifts if the ratio of its two components is one or close to one. Accordingly, a buffer is effective in the pH range of its  $\text{pK}_a \pm 1$  pH unit (cf. Figure 3.3). The “capacity” of the buffering effect is also significantly influenced by the concentration of the buffer, which is the sum of the concentrations of the two components. (E.g. if  $[\text{BOH}] = 0.5 \text{ M}$  and  $[\text{B}^+] = 0.3 \text{ M}$ , then the concentration of the buffer will be  $0.8 \text{ M}$ , see also Exercise 5).

The pH dependence of the degree of acid-base ionisation provides the background of the buffering mechanism. Titration curves effectively illustrate the influence of the pH on the ionisation of weak acids and bases. Figure 3.3 shows the change in the pH of a  $0.1 \text{ M}$  acetic acid solution on its titration with a strong base. This is a monophasic titration curve because acetic acid has only one titratable (ionisable) group. The initial steep increase in the pH slows down quickly, and a pH plateau appears in the range of 4.0–5.8. As in this range  $[\text{CH}_3\text{COOH}]$  is comparable to  $[\text{CH}_3\text{COO}^-]$  (i.e. their ratio is within a tenfold range), we will have a two-component system containing both a weak acid and a weak base—i.e. a buffer. The centre of the plateau is at 50 % titration where  $[\text{CH}_3\text{COOH}] = [\text{CH}_3\text{COO}^-]$ . Therefore, according to Equations 3.13 and 3.14, the pH will just equal the  $\text{pK}_a$ . Following the plateau (upon the exhaustion of the buffer capacity), the pH will again increase steeply, and the top of the titration step will be reached, i.e. the point of neutralisation at which the degree of titration of the acetic acid will be 100 % (see also Exercises 6–8).

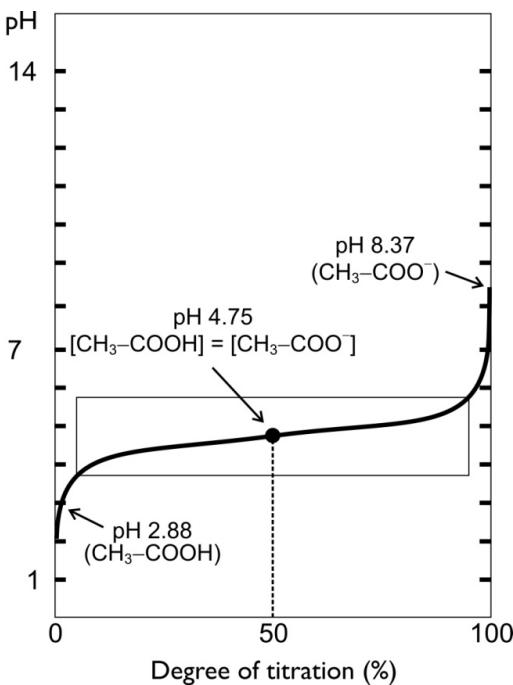


Figure 3.3. Titration curve of a 0.1 M acetic acid solution. The frame indicates the pH plateau (i.e. the buffered range).

The titration of a 0.1 M solution of  $\alpha$ -amino acetic acid ( $\text{H}_2\text{N}-\text{CH}_2-\text{COOH}$ , i.e. glycine) is shown in Figure 3.4. This titration curve is biphasic due to the presence of two titratable (ionisable) groups within the molecule: the carboxylic acid group and the amino group. (Accordingly, glycine has two  $\text{pK}_a$  values.)

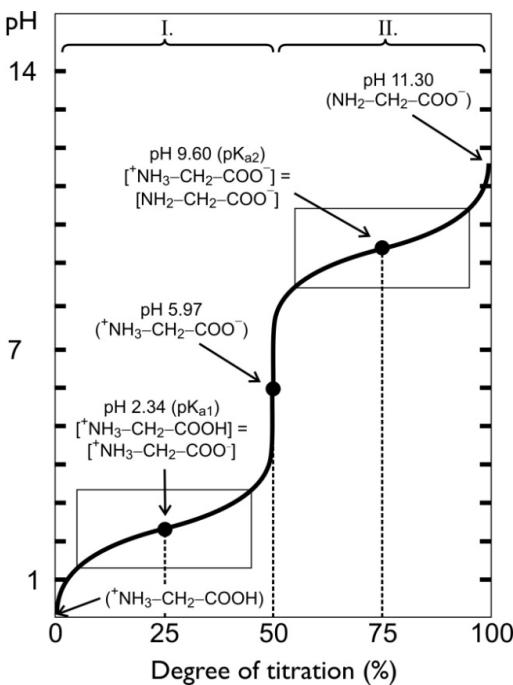


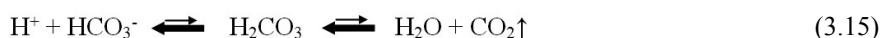
Figure 3.4. Titration curve of a 0.1 M glycine solution. Frames indicate pH plateaus (i.e. the buffered ranges).

The carboxylic acid group is a stronger acid than the amino group (it has a smaller  $\text{pK}_a$ , see Figure 3.2) and unloads its proton easier. Therefore, it is titrated first if the titration starts from the acidic range (phase I). The pH will equal the  $\text{pK}_a$  of this group ( $\text{pK}_{a1} = 2.34$ , see Figure 3.2) in the centre of the first pH plateau (at 25 % titration). On the top of the first titration step (at the end of phase I), the degree of titration (the neutralisation) of the carboxylic acid group will be 100 %, and the degree of titration of the entire compound will be 50 % because the amino groups

will not have been titrated. Continued addition of the strong base will lead to another plateau (in phase II), this time due to the buffering effect of the amino group. In the middle of this plateau (at 75 % titration) the pH will equal the  $pK_a$  of the amino group ( $pK_{a2} = 9.60$ , Figure 3.2, see also Exercise 9). The amino acids that possess an ionisable group in their side chain exhibit an additional, third titration step (with a total of three  $pK_a$  values, see Figure 3.2).

At the top of the first titration step of glycine, a full negative charge will be on the carboxylic acid group due to its complete dissociation. At the same time, the amino group will not have been titrated at all, thus it will be completely ionised (cf. Equation 3.9), lending it a full positive charge. Therefore, at 50 % titration, glycine will adopt the form  $^+H_3N-CH_2-COO^-$ . This form is called “twin-ionic” (or “Zwitter-ionic”). As the net charge of glycine is zero in this state and the electric field does not have an effect on molecules having zero net charge (i.e. the ones that are “actually uncharged”), this state of glycine is called the isoelectric state, and the pH at which this state is adopted is called the isoelectric point (pI). (See Demo calculation 3.4.1 for the pH-dependent charge of ionisable groups.) Generally, we can talk about the isoelectric point in the case of molecules containing two types groups together: those that undergo acidic and those that undergo basic ionisation (see also Exercises 10-11). All amino acids, peptides and proteins exhibit this property. Proteins contain ionisable, weakly acidic or basic groups in the side chain of many of their constituent amino acids (see also Exercise 12). The isoelectric point of an amino acid lacking an ionisable group in its side chain can be calculated as the average of the  $pK_a$  values of its  $\alpha$ -carboxylic acid and  $\alpha$ -amino groups (see Demo calculation 3.4.1). For an amino acid possessing an ionisable group in its side chain, the pI can be estimated as the average of the two  $pK_a$  values belonging to the same type of ionising group (i.e. both forming negative or both forming positive ions, see Demo calculation 3.4.2). The estimation or accurate calculation of the pI of proteins can be much more difficult (see Demo calculation 3.4.3). We can approximate the isoelectric point of an amino acid reasonably well by averaging its  $pK_a$  values. In the case of proteins, pI estimation is more difficult and its precise calculation is even more challenging. A prerequisite of both operations is the knowledge of the amino acid composition of the protein. In the absence of this knowledge, the isoelectric point can be determined experimentally by the method of isoelectric focusing (detailed in Chapter 7). Knowledge of the isoelectric point is of great help in protein purification by ion exchange chromatography (detailed in Chapter 6), while it is indispensable in performing native gel electrophoresis (detailed in Chapter 7) as this information is necessary to determine the net charge of a protein (positive or negative) at a given pH. The titration of glycine (Figure 3.4)—as well as Demo calculation 3.4.3—demonstrates the simple rule that the net charge is negative at  $pH > pI$  and it is positive at  $pH < pI$ . This results from the condition that, at  $pH > pI$ , the ionisation of the positively ionising group(s) begin(s) to decrease (unloading protons more) and become(s) lower than the ionisation of negatively ionising group(s). At  $pH < pI$  the situation is opposite: the negatively ionising group(s) become(s) less ionised (begin(s) to retain protons more) than positively ionising group(s). Proteins, as a rule, are the most stable and the least soluble at pH equal to their pI value.

Buffering the effect of acids and bases—i.e. stabilising the pH—is of enormous physiological significance as the structure of proteins and, in turn, their function is very sensitive to the degree of ionisation of their weakly acidic and basic groups. Two inorganic buffer systems and numerous organic molecules are present in living organisms that maintain the pH in a narrow range. Among the latter are amino acids and proteins. In particular, proteins possess ionisable groups in large numbers, providing them with a large buffer capacity. One of the inorganic buffer systems is the  $H_2PO_4^-/HPO_4^{2-}$  buffer, which has the advantage that its  $pK_a$  is close to the physiological pH (pH 7.4) and thus it can work close to its maximum capacity. The other inorganic system is the  $H_2CO_3/HCO_3^-$  buffer which, though, has the disadvantage of a  $pK_a$  value far from the physiological pH. However, it is present at a high concentration and—in contrast to any other system—its pH is tuneable with respiration, at least in vertebrates. The latter property has importance during acidic stress (acidosis), and is based on the instability of carbonic acid that decomposes into water and carbon dioxide, thereby affecting the ratio of the two components and, in turn, the pH of the buffer according to the following equation:



In acidosis, the organism has the possibility—via increased respiration (hyperventilation)—to shift both equilibria towards the right and, this way—actually, through increased  $CO_2$  unloading—to ultimately reduce the concentration of protons (see also Exercise 13).

The precise setting and stabilisation of the pH in solutions is very important in the biochemical laboratory practice, too, in order to keep the native, functional conformation of proteins, as well as during ion exchange chromatography

and many other applications (detailed in Chapters 5–7 and 9). Besides phosphate and acetate buffers, mainly synthetic organic buffer systems are used that have been developed to stabilise pH in various ranges. One of the most often used buffer systems is TRIS/TRIS-HCl (tris-hydroxymethyl-aminomethane,  $\text{H}_2\text{NC}(\text{CH}_2\text{OH})_3$ ). However, when working with enzymes, special factors should also be considered—buffers composed of chelating agents, e.g. citrate, should not be used with enzymes with metal ion cofactors, or phosphate buffers should be avoided with enzymes working with ATP.

### 3.3. Measurement of the pH

Traditional methods using pH-sensitive dyes (e.g. indicator papers) are suitable only for the estimation of the pH. Precise measurement is possible only on electrochemical basis. However, the hydrogen ( $\text{H}/\text{H}^+$ ) electrode is not among the simple possibilities because the use of this electrode is complicated and dangerous. In the laboratory practice, the measurement of pH is based on the ability of glass for fast and reversible  $\text{H}^+$  adsorption, which generates a  $[\text{H}^+]$ -dependent potential. pH electrodes are therefore made of glass. At the tip of the electrode, a globular glass membrane encloses an inner liquid. The pH of this liquid and, thus, the number of  $\text{H}^+$  ions absorbed on the inside surface of the glass membrane is constant. When the globular tip of the electrode is immersed into a solution to be measured (outside solution), the outer surface adsorbs or desorbs protons, the amount of which depends on the pH. The potential of the glass electrode is proportional to the difference between the amounts of protons adsorbed at the two sides. This, in turn, will depend only on the pH of the outside solution. The pH can be precisely determined via comparing this potential to the potential of a reference electrode (which is a non-polarisable electrode usually made of silver or calomel, built inside the glass electrode) following a pH calibration of potential differences.

#### Control questions and exercises

1. Explain the position of the equilibrium in the reaction of Equation 3.7. Why is it opposite to those in reactions of Equations 3.1 and 3.4?
2. Prove the connection between  $K_a$  and  $K_h$  and use it to interpret the relationship between the strength of a conjugate acid and base.
3. Derive Equation 3.13 for a buffer made of a weak base ( $\text{BOH}$ ).
4. Explain why the Henderson-Hasselbalch equation cannot be applied on single-component systems.
5. How does the pH of a buffer change upon dilution?
6. Explain why pH is not 7.0 at the neutralisation point (100 % titration) of acetic acid.
7. Draw the titration curve of a 0.1 M solution of a weak base ( $\text{BOH}$ ) with a pK of 9.8, and specify its ionisation states.
8. Does the concentration influence the pH at the centre of the titration plateau?
9. Draw the titration curves of lysine, arginine, glutamic acid, aspartic acid and histidine using the  $pK_a$  values specified in Figure 3.2.
10. Determine the isoelectric point of amino acids listed in Exercise 9.
11. Calculate the charge of a histidine side chain at pH 7.4 and pH 7.6.
12. Estimate the isoelectric point of the following peptide:  $\text{H}_2\text{N-Gly-Ala-Asp-Arg-Val-His-His-Glu-Met-COOH}$ .
13. Calculate the ratio of the components in the  $\text{H}_2\text{CO}_3/\text{HCO}_3^-$  buffer system at the pH of tissues (7.4) and that of the lungs (7.6).

## 3.4. Demo calculations of charge and pI

### 3.4.1. Demonstration that pI is the average of the $pK_a$ values of the carboxylic acid and amino groups of an amino acid lacking an ionisable group in its side chain

The correctness of the assumption in the title can be intuitively acknowledged by considering the precise interpretation of the correlation between charges and ionisation states as follows. According to the Henderson-Hasselbalch equation (Equation 3.14), the amount of charge on a group is proportional to the distance of the actual pH from the  $pK_a$  of the group. Thus, the net charge will be zero (i.e. the degree of deprotonation of the carboxylic acid will equal that of the protonation of the amino group) if the pH is at an equal distance from the  $pK_a$  values of these two groups.

We can prove this assumption by calculating the pH at which the amount of charges carried by the carboxylic acid and amino groups is identical, i.e. at which the degree of ionisation of these groups is the same. Using Equation 3.14, we can calculate the ratio of the concentrations of the ionised and non-ionised (i.e. charged and uncharged) forms of an ionisable group ( $[-COO^-]/[-COOH]$  or  $[-NH_2]/[-NH_3^+]$ ). According to the mole concept, this ratio is identical to the ratio of the number of groups in the ionised and non-ionised forms. This, in turn, is identical to the fraction of the form of a group adopting the charged state, while this fraction is identical to the number of ionised groups relative to the non-ionised ones. From this, the degree of ionisation of a group (i.e. the amount of charges it possesses) can be calculated. (See the procedure of such a calculation below in Demo calculation 3.4.2.) This tight correspondence between the charge on an ionisable group and the ratio of concentrations of the two forms of such a group simplifies the subsequent charge calculations because we can use the ratio of concentrations of the two forms of a group instead of its charge. According to this, we are looking for the pH at which

$$\frac{[-COO^-]}{[-COOH]} = \frac{[-NH_3^+]}{[-NH_2]} \quad (3.16)$$

The concentration ratios on the two sides of the equation can be obtained from the Henderson-Hasselbalch equation (Equation 3.14) applied to the carboxylic acid and amino groups. As the desired pH will be the pI, we can write the following equations:

$$pI = pK_{a(COOH)} + \lg \frac{[-COO^-]}{[-COOH]} \quad (3.17)$$

and

$$pI = pK_{a(NH_2)} + \lg \frac{[-NH_2]}{[-NH_3^+]} \quad (3.18)$$

From these:

$$10^{pI-pK_{a(COOH)}} = \frac{[-COO^-]}{[-COOH]} \quad (3.19)$$

and

$$10^{pI-pK_{a(NH_2)}} = \frac{[-NH_2]}{[-NH_3^+]} \quad (3.20)$$

and

$$10^{pK_a(NH_2)-pI} = \frac{[-NH_3^+]}{[-NH_2]} \quad (3.21)$$

Substituting into Equation 3.16:

$$10^{pI-pK_a(COOH)} = 10^{pK_a(NH_2)-pI} \quad (3.22)$$

that is:

$$pI - pK_a(COOH) = pK_a(NH_2) - pI \quad (3.23)$$

from which:

$$pI = \frac{pK_a(NH_2) + pK_a(COOH)}{2} \quad (3.24)$$

Thus, indeed, the pI is the average of the two  $pK_a$  values.

In the case of glycine it is:

$$\frac{9.60 + 2.34}{2} = 5.97 \quad (3.25)$$

Let us check by calculation whether the charge on the two ionisable groups is identical. Now Equation 3.16 will look as follows:

$$5.97 = 2.34 + \lg \frac{[-COO^-]}{[-COOH]} \quad (3.26)$$

and

$$5.97 = 9.6 + \lg \frac{[-NH_2]}{[-NH_3^+]} \quad (3.27)$$

From these:

$$\lg \frac{[-COO^-]}{[-COOH]} = 5.97 - 2.34 = 3.63 \quad (3.28)$$

and

$$\lg \frac{[-NH_2]}{[-NH_3^+]} = 5.97 - 9.6 = -3.63 \quad (3.29)$$

That is,

$$\frac{[-COO^-]}{[-COOH]} = 10^{3.63} = 4.27 \times 10^3 \quad (3.30)$$

and

$$\frac{[-\text{NH}_2]}{[-\text{NH}_3^+]} = 10^{-3.63} = 2.34 \times 10^{-4} \quad (3.31)$$

In the case of the carboxylic acid group, this result means that the fraction of the groups in the charged state is  $4.27 \times 10^3$ -fold relative to those in the uncharged state. In the case of the amino group, this fraction is  $1/2.34 \times 10^{-4} = 4.27 \times 10^3$ , i. e. the same as above. Thus, indeed, the net charge of glycine will be zero in a pH 5.97 solution. This pH is its isoelectric point. Notably, fraction values above a thousand mean that the charge of either ionisable group differs only negligibly from one, the charge at complete ionisation. This is natural since the pI differs more than three pH units from both  $\text{pK}_a$  values, which means that it is sufficiently basic for the complete ionisation of the carboxylic acid group and, at the same time, too acidic for the deprotonation of the amino group. Thus, the zero net charge is achieved by the molecule having a full negative and a full positive charge. In the next example, in which the pI will be much closer to the  $\text{pK}_a$  values, we will see how a zero net charge is achieved at fractional charges (partial ionisation).

### 3.4.2. Demonstration that the pI value of aspartic acid is the average of the $\text{pK}_a$ values of the two carboxylic acid groups in it

Before we set out to prove the above statement, it is worth considering what result can be expected. It is easily acceptable that the following conditions must hold at the pI value: (i) It is not possible for all three ionisable groups (the two carboxylic acids and the amino group) to be completely ionised because this would give a net charge of -1 to aspartic acid. (ii) The negative charges on the two carboxylic acid groups (that on the  $\alpha$ -carbon atom,  $-\text{COOH}_{(a)}$ , and that in the side chain,  $-\text{COOH}_{(R)}$ ) must together neutralise the positive charge on the amino group. (iii) For this reason, the combined charge of the two carboxylic acids cannot be larger than one, which implies a pI value at which the charge of both carboxylic acid groups is only fractional. (iv) Considering the data in Figure 3.2 ( $\text{pK}_{a(a)} = 1.88$  and  $\text{pK}_{a(R)} = 3.65$ ) as well as the meaning of  $\text{pK}_a$  (above), this is such a low pH where the amino group is completely protonated and thus it carries a full positive charge (see the interpretation of the previous pI calculation). Thus, the combined charge of the two carboxylic acid groups must be one at the pH of pI. We can formulate this condition as follows:

$$\frac{[-\text{COO}_{\alpha}^-]}{[-\text{COOH}_{\alpha}]} + \frac{[-\text{COO}_{\text{R}}^-]}{[-\text{COOH}_{\text{R}}]} = 1 \quad (3.32)$$

(In writing Equation 3.32 we again exploited the reduction allowed by the unambiguous correspondence between the charge on an ionisable group and the ratio of concentrations of that group's two forms.)

Since in this case both groups are of the same type (ionise by forming an anion), their different  $\text{pK}_a$  values do not allow a pH at which they carry the same amount of charge so that their combined charge is one. (This is contrary to the case of two groups that ionise by forming different (positive and negative) ions, in which case the same amounts of opposite charges neutralise each other—see the previous example.) At  $\text{pH} > 3.65$  (cf. Equation 3.12), the degree of ionisation of both carboxylic acid groups would be larger than 50 %, which would result in a combined charge larger than one. For the same reason, at  $\text{pH} < 1.88$  the degree of ionisation of neither carboxylic acid group would reach 50 %, and thus the combined charge would be less than one. Thus, the pI must be between the two  $\text{pK}_a$  values. In addition, at this pH it must hold that as much the charge of one group is short of 0.5, the charge of the other must exceed 0.5 to the same extent. (In other words, the absolute values of the charges must fall in equal “distance” from 0.5 on the two “sides” of it.) This can happen only if the  $\text{pK}_a$  values are in equal distance from the pI. Therefore, the pI will be the average of the two  $\text{pK}_a$  values. Now we will prove this.

As in Equation 3.17, we can write for both carboxylic acid groups that

$$pI = pK_a + \lg \frac{[-COO^-]}{[-COOH]} \quad (3.33)$$

From this, the ratio of concentrations of the ionised and non-ionised forms, which is proportional to the amount of charges, will be

$$10^{pI-pK_a} = \frac{[-COO^-]}{[-COOH]} \quad (3.34)$$

Therefore, we can transform Equation 3.32 as follows:

$$10^{pI-pK_{a(\alpha)}} + 10^{pI-pK_{a(R)}} = 1 \quad (3.35)$$

From which:

$$pI - pK_{a(\alpha)} + pI - pK_{a(R)} = 0 \quad (3.36)$$

Thus indeed:

$$pI = \frac{pK_{a(\alpha)} + pK_{a(R)}}{2} \quad (3.37)$$

Substituting the  $pK_a$  values we get the  $pI$  value of aspartic acid:

$$pI = \frac{1.88 + 3.65}{2} = 2.765 \quad (3.38)$$

As a check, we can calculate the charges on the two carboxylic acid groups. After substituting the appropriate values into Equation 3.34 we will get:

$$10^{2.765-1.88} = \frac{[-COO^-]_{(\alpha)}}{[-COOH]_{(\alpha)}} \quad (3.39)$$

and

$$10^{2.765-3.65} = \frac{[-COO^-]_{(R)}}{[-COOH]_{(R)}} \quad (3.40)$$

which are

$$7.67 = \frac{[-COO^-]_{(\alpha)}}{[-COOH]_{(\alpha)}} \quad (3.41)$$

and

$$0.13 = \frac{[-COO^-]_{(R)}}{[-COOH]_{(R)}} \quad (3.42)$$

Due to the identity of the ratio of the concentrations and the ratio of charged and uncharged groups, we can state that the fraction of charged  $\alpha$ -carboxylic acid groups is 7.67-fold relative to the uncharged ones, while this fraction is only 0.13-fold in the case of the carboxylic acid groups in the side chain. As these fractions reflect the number of charged groups relative to the uncharged ones (see above), we can determine the degree of ionisation—which corresponds to the charge possessed by an ionisable group—by dividing the number of charged groups by the total number of groups. In the case of the  $\alpha$ -carboxylic acid, for each non-ionised group there are 7.67 ionised ones. Therefore:

$$\text{degree of ionisation} = \frac{\text{number of ionised groups}}{\text{total number of groups}} = \frac{7.67}{7.67+1} = 0.885 \quad (3.43)$$

Thus the  $\alpha$ -carboxylic acid group has -0.885 charges. The case is just the opposite for the side chain carboxylic acid, in which there are  $1/0.13 = 7.67$  non-ionised groups for each ionised one. Therefore:

$$\text{degree of ionisation} = \frac{\text{number of ionised groups}}{\text{total number of groups}} = \frac{1}{7.67+1} = 0.115 \quad (3.44)$$

Thus the side chain carboxylic acid group has -0.115 charges. Indeed, the two carboxylic acid groups contribute unequally to the total negative charge of one and, as much the share of the  $\alpha$ -carboxylic acid is in excess of 0.5 ( $0.885-0.5=0.385$ ), this will be the same as the extent to which the share of the side chain carboxylic acid ( $0.5-0.115=0.385$ ) is short of 0.5.

### 3.4.3. Demo calculation of the isoelectric point of a protein

Let us calculate the pI of a protein that consists of 152 amino acids and has the following amino acid composition:

$\text{H}_2\text{N-Gly-(150 amino acids)-Gly-COOH}$ , containing the following amino acids with ionisable residues: 2Arg, 3Lys, 2Tyr, 1Cys, 5His, 3Asp.

As there are many ionisable groups in such a large molecule, it is obvious that the calculation of pI will be more a complex procedure. Therefore the procedures employed above are suitable only for an approximate estimation of pI. In some very rare cases, however, the calculation can follow that of the pI of aspartic acid (Demo calculation 3.4.2), and will not even be more difficult than that. In these cases the expected pI must be close to a  $\text{pK}_a$  so that, at the same time, it is at least two pH units away from all the other  $\text{pK}_a$  values. In order to determine whether this criterion is met, an estimation must be performed concerning the expected value of the pI. To this end, we must calculate the net charge of the protein at several different pH values. In order to facilitate these calculations, we can construct a charge calculation table listing the amino acids possessing ionisable residues in the order of their increasing  $\text{pK}_a$  values. In the case of our protein, this table will look as follows (Figure 3.5):

| amino acid/group            | $\text{pK}_a^*$ | factor** | the total charge of group(s) as the function of pH |      |     |
|-----------------------------|-----------------|----------|--|------|-----|
|                             |                 |          | pH= 7.0  | 6.0  | 6.4 |
| $-\text{COOH}_{(\alpha)}$   | 2.34            | 1        | -1   | -1   | -1  |
| Asp                         | 3.65            | 7        | -7   | -7   | -7  |
| His                         | 6.00            | 5        | +0.455   | +2.5 | +2  |
| Cys                         | 8.30            | 1        | -0.048   | 0    | ~0  |
| $-\text{NH}_3^+_{(\alpha)}$ | 9.60            | 1        | +1   | +1   | +1  |
| Tyr                         | 10.10           | 2        | 0  | 0    | 0   |
| Lys                         | 10.53           | 3        | +3   | +3   | +3  |
| Arg                         | 12.48           | 2        | +2   | +2   | +2  |
| Net charge:                 |                 |          | -1.593   | +0.5 | 0   |

Figure 3.5. Calculation of the net charge of a protein. \* The pK<sub>a</sub> values (from Figure 3.2) next to the names of the amino acids refer to the ionisable group in their side chain, while the values concerning the α-carboxylic acid and α-amino groups (C- and N-termini) are those of glycine. \*\* The number of amino acids (i.e. that of the given type of ionisable group) within the protein. (For example, a protein always has one N- and one C-terminus.)

In order to reach our goal more rapidly, prior to the calculations it worth considering the pH range into which the pI is expected to fall. As a conclusion from the reasoning employed at the pI calculation of aspartic acid (Demo calculation 3.4.2), the pI will fall in the acidic or basic range if the number of negatively or positively ionising groups exceeds that of the other, respectively. The more the ratio is shifted towards one type of group, the more the pI will be shifted towards the pH range of the pK<sub>a</sub> of these groups. In our protein, the number of groups of the two types is the same. Therefore, it is appropriate to start the calculation of charge at pH 7.0.

Calculation of the charge at pH 7.0

The charge of the α-amino group:

According to Equation 3.18:

$$pH = pK_a(\text{NH}_2) + \lg \frac{[\text{-NH}_2]}{[\text{-NH}_3^+]} \quad (3.45)$$

After substitution:

$$7.0 = 9.6 + \lg \frac{[\text{-NH}_2]}{[\text{-NH}_3^+]} \quad (3.46)$$

From this, following the earlier calculations:

$$\frac{[\text{-NH}_2]}{[\text{-NH}_3^+]} = 10^{7.0-9.6} = 10^{-2.6} = 0.0025 \quad (3.47)$$

According to the above considerations regarding the connection between the ratio of concentrations and the fraction of charges, for every charged group there are 0.0025 uncharged ones. Based on this and according to Equation 3.43, the degree of ionisation, i.e. the amount of charge will be:

$$\text{degree of ionisation} = \frac{\text{the number of ionised groups}}{\text{total number of groups}} = \frac{1}{1+0.0025} = 0.998 \quad (3.48)$$

This value differs negligibly from one. Therefore, we can take it as a full charge on the α-amino group. During approximate calculations we can always apply such rounding whenever the difference between the pH and the pK<sub>a</sub> is more than two pH units. This is allowed because in these cases the equilibrium of ionisation is shifted so much towards one of the forms of the group that the difference of the charge from one or zero will only be a few thousandths. Applying this simplification, without detailed calculations we can assign a full charge to the residues of aspartic acids, lysines and arginines and to the α-carboxylic acid, as well as zero charge to tyrosine residues.

Following the calculations above, the ratio of forms of the cysteine residue will be

$$\frac{[\text{Cys}^-]}{[\text{Cys}]} = 10^{7.0-8.3} = 10^{-1.3} = 0.05 \quad (3.49)$$

From this the charge of cysteine:

$$\frac{0.05}{0.05+1} \sim 0.048 \quad (3.50)$$

While the above ratio regarding histidine residues will be

$$\frac{[\text{His}]}{[\text{His}^+]} = 10^{7.0-6.0} = 10^{1.0} = 10 \quad (3.51)$$

and thus their charge:

$$\frac{1.0}{10.0 + 1} \sim 0.091 \quad (3.52)$$

By using the values of charges calculated above—and also taking into account the number of groups—we can fill the pH = 7.0 column of Figure 3.5. Summing up the charges, we can see that the charge of our protein at this pH will be -1.593. Thus, as it follows from the relationship between the net charge and the pH (see above), the pI must be below 7.0.

Continuing the estimation of the expected value of pI, we can now calculate the net charge of the protein at pH 6.0. As the minimally expected two-unit difference between the pH and the pK<sub>a</sub> values again holds for Asp, Lys, Arg and Tyr residues as well as for the  $\alpha$ -carboxylic acid and  $\alpha$ -amino groups, their charge can be taken the same as at pH 7.0. In contrast, the charges of cysteine and histidine residues will change. The former will decrease essentially to zero (because now the distance of pH from the pK<sub>a</sub> will be more than two units), while the latter will increase to +0.5 because now pH = pK<sub>a</sub> and, by the definition of pK<sub>a</sub>, here the degree of ionisation will be 50 %. Filling the pH = 6.0 column of Figure 3.5 and summing up the charges, we get that the net charge of the protein is +0.5. Thus, this pH is below the pI. Therefore, the pI must be between 6.0 and 7.0.

If we now assume (with some reason) that the pI will be close to 6, we can expect that even this value will differ at least two pH units also from the pK<sub>a</sub> of cysteine. Thus the criterion set above seems to be met, and its significance can now be understood. At several tenths of pH units above 6, the charge of every group—including cysteine—will only negligibly differ from that at pH 6.0. Therefore we can try to calculate the exact pI by employing the same logic that was used during the pI calculation of aspartic acid. Thus, at pH = pI, the net charge of the protein will be -2 without the charge of histidines. Therefore, the reasoning applied in writing Equation 3.32 will in this case appear as the following: we will look for the pH at which the combined charges of histidines can neutralise the -2 charge resulting from the other ionisable groups. That is,

$$5 \times \frac{[\text{His}^+]}{[\text{His}]} = 2 \quad (3.53)$$

thus

$$\frac{[\text{His}]}{[\text{His}^+]} = 2.5 \quad (3.54)$$

Writing the expression of pI as in Equation 3.18 and performing the appropriate substitutions we get:

$$\text{pI} = \text{pK}_a + \lg \frac{[\text{His}]}{[\text{His}^+]} = 6 + \lg 2.5 = 6.4 \quad (3.55)$$

Thus the isoelectric point we look for is 6.4. After the calculation and summation of the charge on every ionisable group, we must get zero. Performing this control calculation (see the pH = 6.4 column in Figure 3.5) indeed yields zero net charge.

# Chapter 4. Spectrophotometry and protein concentration measurements

by András Málnási-Csizmadia

## 4.1. Photometry

Spectrophotometry is one of the most widely used analytical procedures in biochemistry. The technique is well suited for simple routine determination of small quantities of materials. These measurements require that the examined material have an absorption maximum at one point of the spectrum. If the absorption maximum falls into the visible part of the spectrum, the material is coloured. Analyses can be performed in the visible spectrum also if the material in question has no colour by itself, but a chemical reaction can be performed that leads to the formation of a coloured product. Analyses in the ultraviolet spectrum are widespread, too, since many colourless materials exhibit intense absorption in this range (190-320 nm).

Quantitative measurements in spectrophotometry are based on the Lambert-Beer law of light absorption by solutions. The amount of light absorbed by the sample is defined as the ratio of the intensities of the incident and the transmitted light (Figure 4.1).

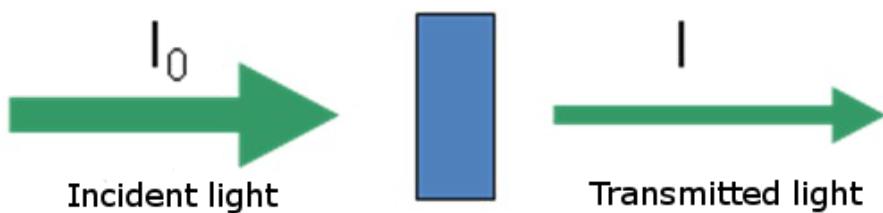


Figure 4.1. Incident and transmitted light. Light absorbed by the sample is characterised by the ratio of the intensity of the incident and transmitted light.

From the degree of absorption, the concentration of the absorbing component can be deduced by using the following equation. If the intensity of light,  $I_0$ , diminishes to  $I$  after passing through a path of length  $L$  in a medium, then, according to the Lambert-Beer law:

$$E = \log \frac{I_0}{I} = \varepsilon \cdot c \cdot L \quad (4.1)$$

The expression  $\log I_0/I$  is called extinction (E), absorbance (A) or optical density (OD). At a wavelength where the solvent does not absorb, according to the Lambert-Beer law, E will be proportional to the concentration of the solution (c) if the solute does not undergo molecular changes during dilution. The law is valid only for monochromatic light of a given wavelength. The degree of the absorption is characterised by  $\varepsilon$ , named extinction coefficient, a constant independent of the concentration of the solute. If the concentration is given in mol/L, then we speak of a molar extinction coefficient or molar absorptivity. Molar absorptivity gives the extinction of a solution of 1 mol/L (M) at the given wavelength and 1 cm path length; its unit is thus  $M^{-1}cm^{-1}$ . Its value is characteristic of the material in question, but also depends on the solvent and the temperature.

Using the above equation, the concentration can be calculated from the extinction as:

$$c = \frac{E}{\varepsilon \cdot L} \quad (4.2)$$

If  $\epsilon$  is the molar extinction coefficient, then  $c$  will be molar concentration.  $L$  is the optical path length of the cuvette (a transparent cell that holds the sample, usually made of glass, plastic or quartz), given in centimetres.

If the material to be measured does not follow the Lambert-Beer law exactly, a calibration curve can be established using a series of samples of increasing concentration, established using another exact method of measurement. The technique can then be used with the appropriate corrections.

Older photometers may measure the transmission (transmittance):

$$T = I/I_0 \text{ or } T \% = I/I_0 \times 100 \quad (4.3)$$

The relation of transmittance to extinction is (see also Table 4.I below):

$$E = \log \frac{100}{T\%} , \text{ can be written as } E = 2 - \log T\% \quad (4.4)$$

| E    | T %   |
|------|-------|
| 0.01 | 98 %  |
| 0.1  | 79 %  |
| 0.5  | 32 %  |
| 1    | 10 %  |
| 2    | 1 %   |
| 3    | 0.1 % |

Table 4.I. Some examples of the relation of extinction (E) to transmittance (T %). Note that an extinction value of 2 means that the sample absorbs 99 % of the incident light. At such a high level of absorption, the intensity of the light reaching the detector is very small. Thus, the accuracy of the measurement diminishes in the case of highly absorbent samples.

## 4.2. The UV-VIS photometer

The spectrophotometer is an apparatus devised for the measurement of absorbance. It can produce a beam of light of a given wavelength, direct it at a sample (usually in solution in a cuvette) and measure the intensity of the transmitted light. For all this, a light source, a monochromator, a sample holder, a detector and a display are required (Figure 4.2).

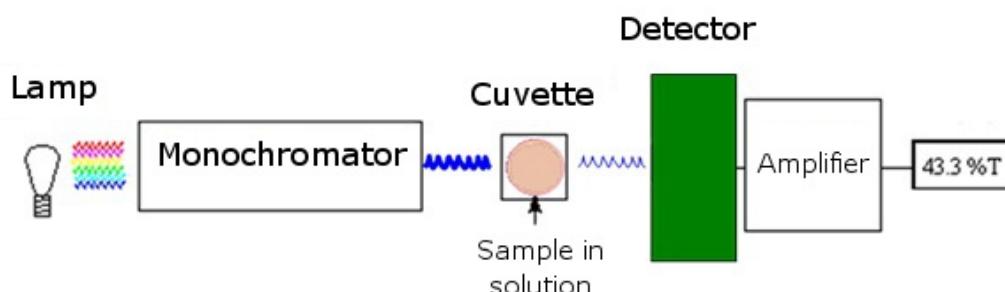


Figure 4.2. Schematic structure of a photometer

In the wavelength range of 200 to 320 nm, the light source can be either a deuterium, a high-pressure hydrogen or a high-pressure xenon lamp. To generate light in the visible or infrared spectrum, tungsten filament lamps are used. In a UV-VIS type photometer, both types of light sources can be found.

The role of the monochromator is to select light of a given wavelength from the continuous spectrum of the light source. In modern devices, a diffraction grating has replaced the previously used prism. The produced light is not strictly monochromatic, but its spectral bandwidth is relatively narrow. Light entering and leaving the monochromator passes through slits whose width determines the bandwidth. The wider the slits, the wider the bandwidth. However, narrowing the slits reduces the intensity of the incident light. Thus, one must find a compromise between spectral purity and intensity, and set the slits accordingly.

Light passes through the sample placed in a holder (cuvette) made of plastic, glass, quartz or other transparent material (Figure 4.3). Plastic and glass cuvettes are cheap but cannot be used at wavelengths under 280 and 320 nm, respectively, because their absorbance becomes too high. In the case of shorter wavelengths, quartz cuvettes are used. The quality and condition of the cuvette is critical for the measurements. Some devices are capable of keeping the temperature of the sample constant during measurement by a thermostatted cuvette holder. This is important, for example, when measuring the rate of chemical reactions.



Figure 4.3. Cuvettes for holding samples. Cuvettes are generally made of glass, plastic or quartz. Quartz cuvettes are used when absorbance is measured at a wavelength under 320 nm, as glass and plastic absorb most of the light in this range. The path length is 1 cm in most cuvettes whose typical volume is around 1.5 mL. There are narrowed cuvettes for measuring smaller volumes. It is recommended to use narrowed cuvettes with darkened sides when measuring absorbance.

Cheaper photometric devices are usually single-beam ones (such as the one depicted in Figure 4.2), meaning that they can measure one sample at a time. More sophisticated devices have a double beam, and thus they can take two measurements simultaneously. One cuvette holds the sample, i.e. the solution containing the substance of interest, while the other (the reference) holds a solution that is identical to the sample in all aspects except that it lacks the substance of interest. The device automatically subtracts the absorbance of the reference from that of the sample, and thus it records a differential spectrum. This kind of spectrum not only characterises the studied molecule better, but also corrects for the error stemming from the fluctuations in the intensity of light generated by the lamp.

The intensity of the transmitted light is measured by a light-sensitive electronic detector (photodiode or photomultiplier). Detector type and quality significantly affect the attributes of the system. Photomultipliers are characterised by high speed and high sensitivity in a wide spectral range. Photodiodes are small in size and have a lower sensitivity. They are used in so-called diode-array devices that measure a complete spectrum at a time. Such measurements are performed by illuminating the sample with multi-wavelength light and separating its components only after passing through the sample so that each diode measures at a given wavelength.

Besides measuring at a given wavelength, we can also determine complete spectra, i.e. the dependence of absorbance on wavelength by changing the wavelength. Up-to-date spectrophotometers can take spectra between preset values automatically. Simpler devices can measure in the range of 0-1 absorbance units (AU), whereas higher quality ones do in a range of 0-2 AU (or even wider). However, as mentioned above, absorbance values at or higher than 2 are prone to large error.

## 4.3. Other possible uses of photometry

Photometers are often used in combination with other devices. They are often used in conjunction with chromatographic equipment in which the sample leaving the column is conducted through a flow cell and the absorbance

of the solution is detected continuously. Plotting the measured absorbance against time will yield a curve called a chromatogram. If we use a diode-array detector, we can measure the absorbance of the conducted sample at multiple wavelengths simultaneously and can create a series of photometric spectra.

Another common practice is to measure the absorbance of not just one but multiple (tens or hundreds of) samples at a time by using a so-called plate reader (Figure 4.4). Plates holding 6, 12, 24, 48, 96 or 384 samples can be placed in such devices. Depending on the type of the plate, the volume of a sample can be between 10 µL and 10 mL. The most common type of plate has 96 wells with a maximal individual volume of 200 µL.



Figure 4.4. Plate reader with a 96-well plate

## 4.4. Frequently arising problems in photometry

- a. If the sample is turbid, this will result in an error as the light will be scattered and part of it will never reach the detector and thus will appear as absorbed.
- b. If the molecule is capable of association/dissociation and the absorbances of the two forms differ, the Lambert-Beer law will no longer be valid as the degree of dissociation will depend on the concentration.
- c. It is crucial for the cuvette to be clean and free of scratches. The sides of the cuvette in the light path should never be touched by hand.

## 4.5. Determination of protein concentration

Determining the exact quantity of proteins in a solution is very often necessary in the biochemical practice. There are many ways to measure protein concentration. In chromogenic methods, the absorbance of a coloured product formed by the protein and an organic molecule is measured. Protein concentration can also be determined from the protein's own (intrinsic) UV absorbance. Note, however, that these methods may give different results for different proteins of the same concentration. Also, different methods can yield somewhat different results for the same protein. There is no absolute photometric protein concentration assay. All methods have advantages and disadvantages and we must choose among them by taking the following aspects into consideration: specificity, sensitivity, the measurable range of concentration, the accuracy, the nature of the protein to be examined, the presence of materials interfering with the measurement, and the time required for the measurement.

### 4.5.1. Biuret test

Molecules with two or more peptide bonds react with  $\text{Cu}^{2+}$  ions in alkaline solution and form a purple complex. Nitrogen atoms of the peptide bonds form a coordination bond with the metal ion. The quantity of the complexes formed is proportional to the number of peptide bonds.

In practice, the determination of protein concentration is done using a calibration curve created using samples of known concentration. The protein treated with biuret reagent is measured at 540 nm after the purple product is formed.

The advantages of the method include that only few materials (e.g. Tris and amino acid buffers) interfere with it, it can be done in a short time and does not depend on the amino acid composition of the protein. Its disadvantages are its low sensitivity and that it requires at least 1 mg of protein.

### 4.5.2. Lowry (Folin) protein assay

In this sensitive technique, a coloured product is formed similarly to the biuret reaction, but a second reagent (Folin-Ciocalteu reagent) is used in addition to strengthen the colour. The strong blue colour is created by two reactions: (1) formation of the coordination bond between peptide bond nitrogens and a copper ion and (2) reduction of the Folin-Ciocalteu reagent by tyrosine (phosphomolybdic and phosphotungstic acid of the reagent react with phenol). The measurement is carried out at 750 nm.

As in the biuret reaction, a calibration curve is created (for example using BSA, bovine serum albumin), and the concentration of the unknown protein is determined from the curve.

The advantages of the method include that it is quite sensitive and is able to detect even 1  $\mu\text{g}$  of protein. Its disadvantages are that it takes rather long to carry out, is disturbed by various materials (including ammonium sulphate, glycine and mercaptans) and that the incubation time is critical. As different proteins contain different amounts of tyrosine, the amount of the coloured product will also be different. As a consequence, this method is more suited to compare the concentration of solutions of the same protein than to absolute measurement.

### 4.5.3. Bradford protein assay

Despite being relatively new, probably this is the most widely used protein assay. The method is based on the ability of the Coomassie Brilliant Blue dye to bind to proteins in acidic solution (via electrostatic and van der Waals bonds), resulting in a shift of the absorption maximum of the dye from 465 to 595 nm.

The advantages of the method include that it is highly sensitive, is able to measure 1-20  $\mu\text{g}$  of protein and is very fast. Only relatively few materials interfere with it (it works even in presence of urea or guanidine hydrochloride) but, importantly, detergents do. Even traces of detergent (e.g. cleaning products) can invalidate the results. Its disadvantages are that it depends strongly on amino acid composition and that it stains the cuvettes used.

### 4.5.4. Spectrophotometry based on UV absorption

This method is based on the fact that two of the aromatic amino acids, tryptophan and tyrosine, show a peak in absorbance around 280 nm. It has the advantage of being quick and easy. Since it needs no chemical reaction to be performed, it is widely used for detection of proteins or peptides during their separation by chromatography. As proteins contain different ratios of aromatic amino acids, *per se* it is more suited to the comparison of solutions of the same protein and less to absolute measurement. The latter requires the knowledge of the molar extinction coefficients of proteins. For many proteins, these were determined and can be found in the literature. Moreover, if we know the number of tyrosine and tryptophan amino acids in the protein of interest, since their absorption values are additive, it is possible to calculate the molar extinction coefficient.

This method has a moderate sensitivity with a material requirement of around 50  $\mu\text{g}$ . It is disturbed by anything that has an absorbance at 280nm—most commonly, DNA. (Nucleic acids have an absorption maximum at 260 nm and their absorption at 280 nm is still considerable.) By using the correction introduced by Warburg and Christian,

we can account for the error caused by nucleic acids. Absorption is measured at 260 and 280 nm and protein concentration can be calculated with the following equation:

$$c_{\text{prot}}(\text{mg/mL}) = 1.55 * A_{280\text{nm}} - 0.76 * A_{260\text{nm}} \quad (4.5)$$

Proteins and peptides also show high absorbance between 220 and 240 nm, originating from peptide bonds and carboxyl groups. This wavelength range can be used for quantitative assessment only if the solution is pure, because a large number of other substances also have high absorbance in this range.

## 4.6. Spectrophotometry in practice: some examples

### 4.6.1. Absorption spectrum of ATP

The absorption maximum of the molecule is at 260 nm due to its aromatic structure (Figure 4.5). As mentioned above, this peak at 260 nm is also present in the absorption spectra of nucleic acids.

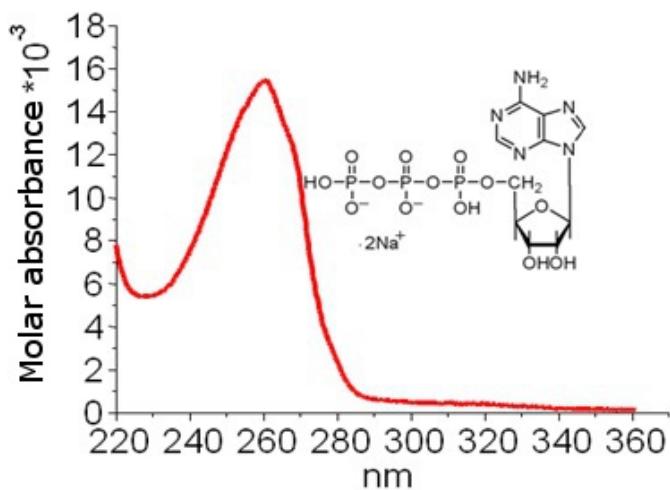


Figure 4.5. Absorption spectrum of ATP

### 4.6.2. Hyperchromicity of DNA

The absorption maximum of nucleic acids at 260 nm originates from their constituent aromatic groups. High temperature causes DNA to “melt”—hydrogen bonds connecting the bases start to break and absorbance at 260 nm rises. This phenomenon is known as hyperchromicity (Figure 4.6). The transition temperature is also called “melting temperature” ( $T_m$ ).

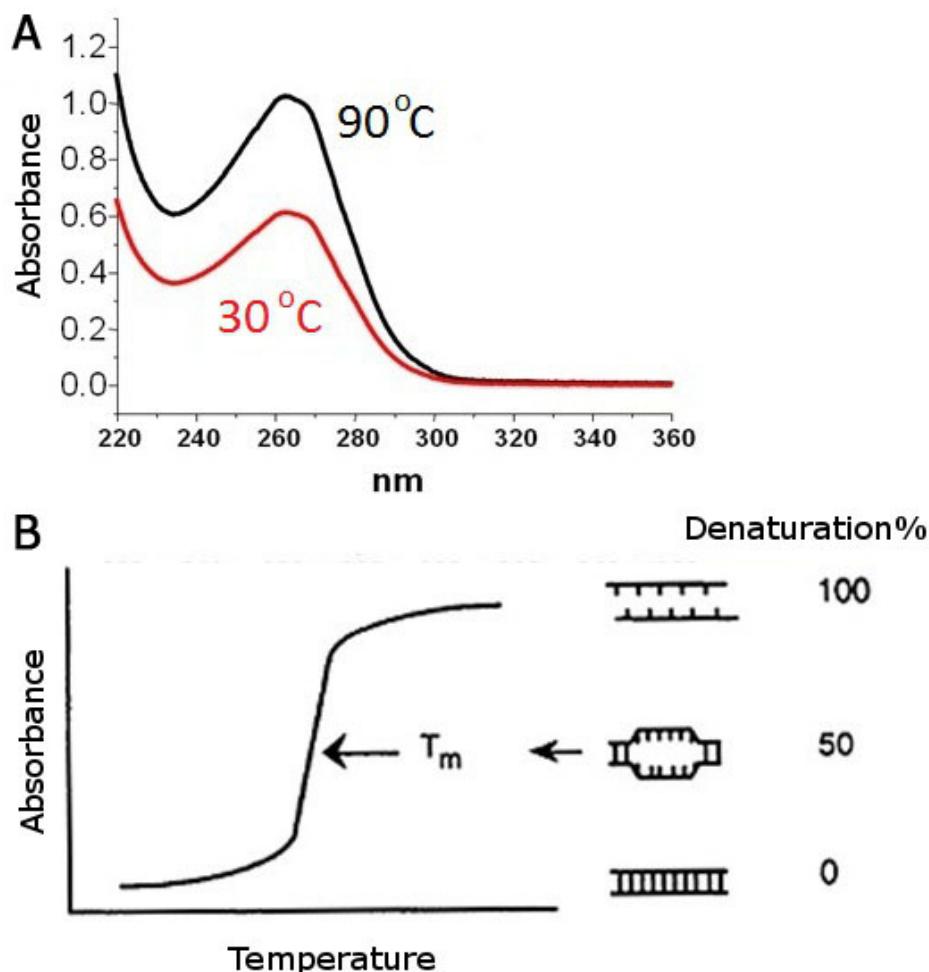


Figure 4.6. Hyperchromicity of DNA. A, Absorption spectra of a DNA molecule at 30°C (red) and 90°C (black). B, Temperature-induced change in absorbance at 260 nm, reflecting the denaturation (melting) of the double-stranded DNA structure.

### 4.6.3. Absorption spectra and molecular structure of NAD and NADH

Nicotinamide adenine dinucleotide has two aromatic groups (Figure 4.7). The adenine part is responsible for the absorption at 260 nm (as in ATP or nucleic acids, cf. Figures 4.5-4.6). The nicotinamide moiety has a specific absorption maximum at 340 nm, but only when protonated (NADH). The very high difference in the absorption between the oxidised and reduced forms ( $\epsilon_{\text{NADH},340} = 6220$ ) at this peak makes it a highly useful tool in analytical, enzyme kinetic and medical diagnostic measurements. For example, when measuring ATPase activity, we wish to measure the ADP generated from ATP. However, ATP and ADP cannot be distinguished photometrically—their absorption at 260 nm is the same. In such cases it is common to use a coupled reaction (Figure 4.8). We add enzymes (pyruvate kinase and lactate dehydrogenase) and substrates (PEP, NADH) to the solution so that the oxidation of NADH will be proportional (actually, equimolar) to ATP consumption during the examined reaction. NADH consumption can be readily followed at 340 nm.

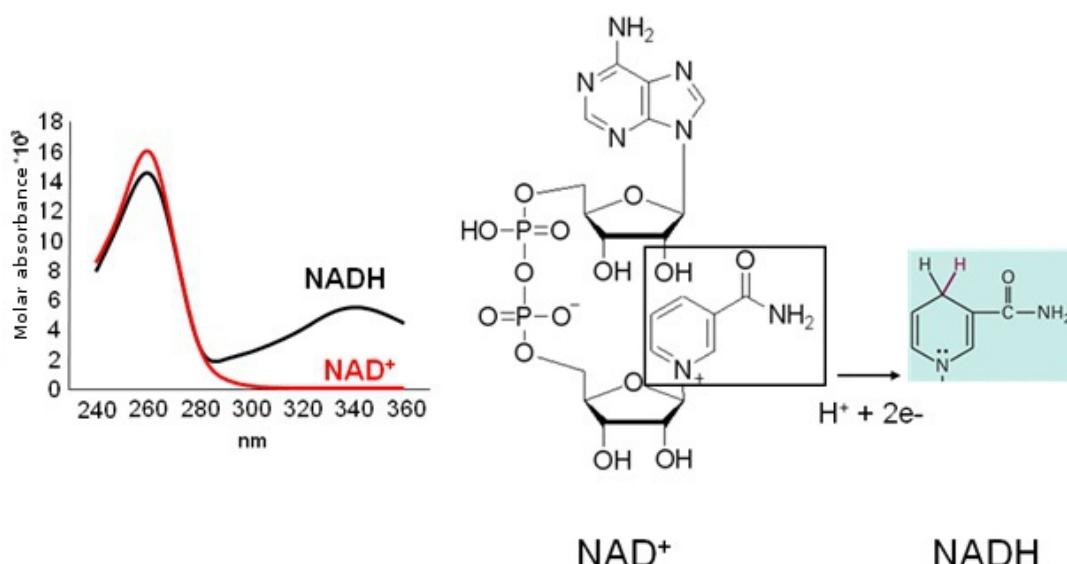
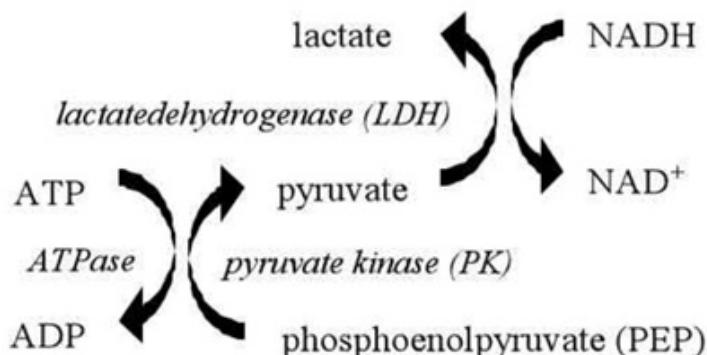
Figure 4.7. Absorption spectra (left) and molecular structure of NAD<sup>+</sup> (middle) and NADH (right)

Figure 4.8. Measuring ATPase activity via a coupled reaction

#### 4.6.4. Absorption spectrum of proteins

Proteinogenic amino acids with an aromatic group in their side chain have a peak in absorption at 280 nm (Figure 4.9). As mentioned earlier, the absorption of amino acids is additive and thus the molar extinction coefficient of proteins can be calculated from their amino acid composition. Non-aromatic amino acids do not absorb at this wavelength, except for the low absorbance of cystine (a pair of cysteines forming a disulfide bond). Absorption values calculated based on amino acid composition generally fall close to values determined experimentally, making them suitable for the determination of protein concentration. The side chains of tryptophan and, to a lesser extent, that of tyrosine are also fluorescent. Tryptophan can be selectively excited at 295 nm, as it can be inferred from the “shoulder” in its absorption spectrum (Figure 4.9).

In the spectrum of proteins, absorption at 220 nm is due to the peptide bonds and the peak at 280 nm is caused by the absorption of aromatic amino acids (Figure 4.10). The latter is better suited for concentration measurements because there are many other materials, including minor impurities in solvents, that exhibit absorption at 220 nm.

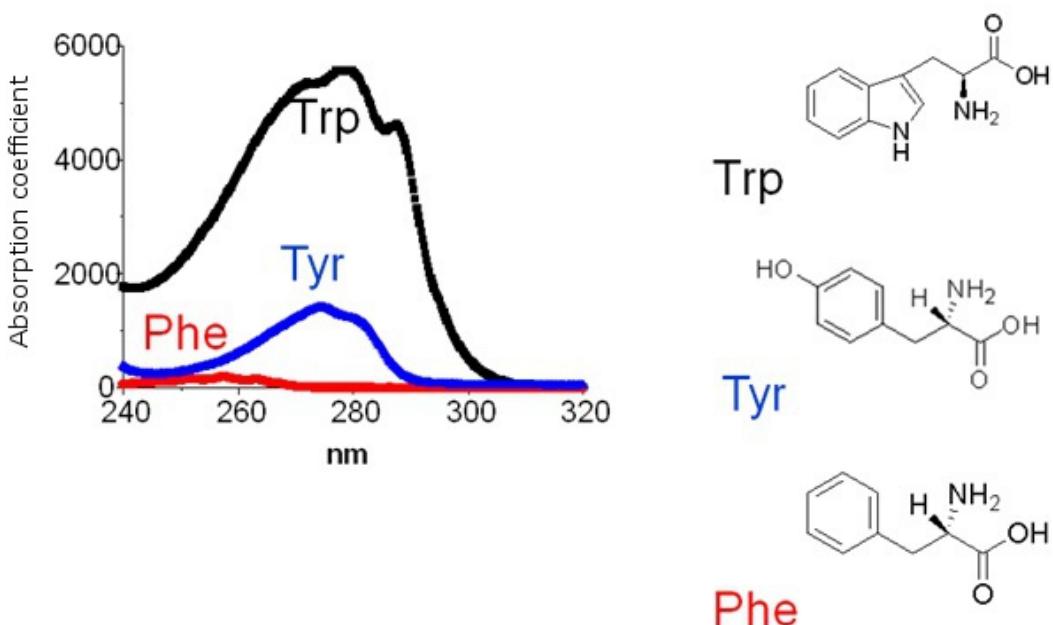


Figure 4.9. Absorption spectra of aromatic amino acids

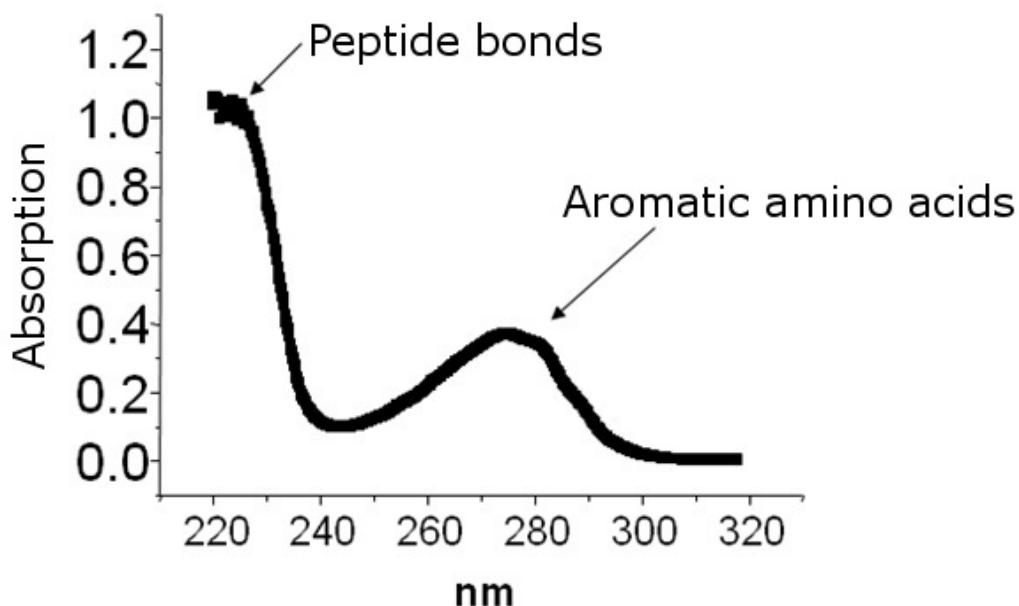


Figure 4.10. Absorption spectrum of a protein

#### 4.6.5. Determination of the purity of DNA and protein samples

The ratio of DNA and protein in biological samples can be estimated from the absorbance of the sample at 260 and 280 nm, resulting from the absorbance of nucleotide bases and aromatic amino acids. In the case of a pure solution of DNA,  $A_{260}/A_{280} = 1.8$ . The addition of protein to the solution will typically reduce this value. Figure 4.11 shows the absorption spectra of solutions with different DNA to protein ratios.

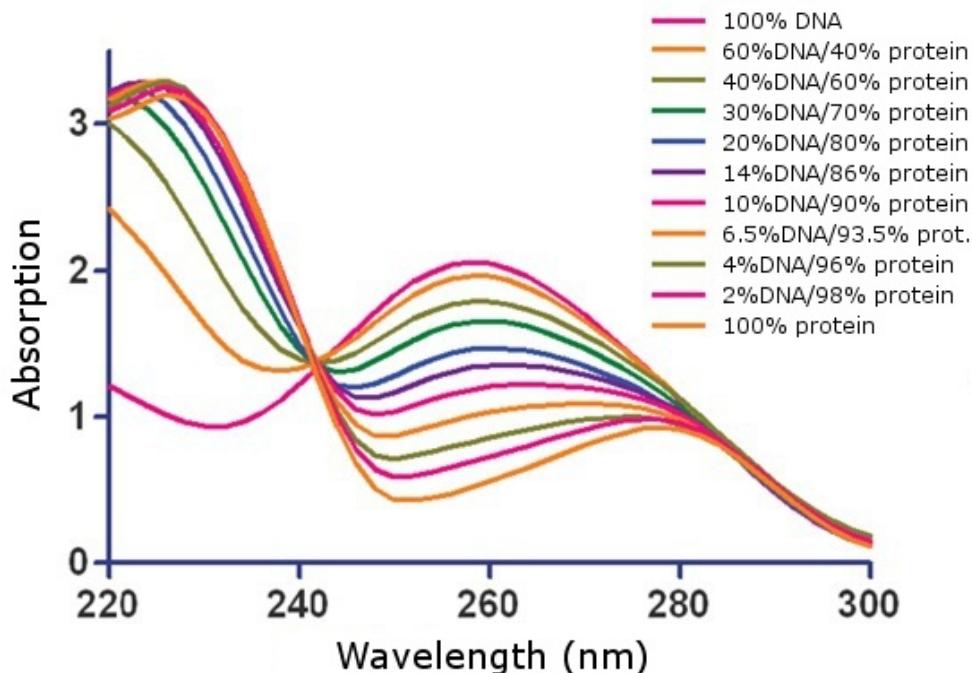


Figure 4.11. Determination of the purity of DNA and protein solutions

## 4.7. Fluorimetry

Fluorimetry is the quantitative study of the fluorescence of fluorescent molecules. Many biomolecules are fluorescent or can be labelled with fluorescent molecules, making fluorimetry a widely used tool in analytical and imaging methods. As the available photon-detecting devices are highly sensitive—even a single photon can be detected—and one fluorophore can emit millions of photons in a second, fluorimetry is suitable for and is often used in single-molecule experiments.

The phenomenon of fluorescence was discovered and published by Sir John Fredrick William Herschel in the mid-1800s. He observed that, when illuminated with white light, a solution of quinine emitted a strange blue light perpendicular to the direction of the illumination, even though it remained colourless when observed facing the light source.

Demonstrating the sensitivity of fluorescence measurements, such methods were used to prove that the rivers Danube and Rhine are connected by underground waterways. In 1877, researchers poured fluorescein (a fluorophore) into the Danube and could detect its green fluorescence 60 hours later in a small river flowing into the Rhine. Fluorescein is still used to aid the detection of space cabins that returned to Earth and fell into an ocean.

### 4.7.1. Physical basis of fluorescence

Photons of a given wavelength are absorbed by the fluorophore and excite some of its electrons. The system remains in this excited state for only a few nanoseconds and then relaxes into its ground state. (Note that light travels 30 centimetres in a single nanosecond.) When returning from the excited state to the ground state, the electron may emit a photon. This is known as fluorescent emission. The wavelength of the absorbed photon is always shorter than that of the emitted photon (i.e. the energy of the emitted light is lower than that of the absorbed one). This phenomenon, the so-called Stokes shift, is an important attribute of fluorescence both in theory and practice.

The relations between the wavelength ( $\lambda$ , nm), frequency ( $v$ , 1/s) and energy ( $E$ , J) of light are the following:

$$\lambda = c/v, \text{ where } c \text{ is the speed of light (approximately } 300\ 000 \text{ km/s}) \quad (4.6)$$

$$v = c/\lambda \quad (4.7)$$

$$E = hv, \text{ where } h \text{ is the Planck constant (approximately } 6.63 * 10^{-34} \text{ Js}) \quad (4.8)$$

## 4.7.2. The fluorimeter

The Stokes shift facilitates the creation of highly sensitive methods of detection of fluorescence. As the wavelengths of the exciting and detected (emitted) light differ, the background created by the exciting light can be minimised by using a proper setup. There are two ways to avoid that the exciting light get into the detector:

(1) Measurements are often carried out in a geometric arrangement in which the detection of emission is perpendicular to the exciting beam of light.

(2) Light filters are placed between the light source and the sample and also between the sample and the detector. Light of only a certain wavelength range can pass through these filters. Photons of the exciting light leaving the sample will not reach the detector as they are absorbed by the emission filter (Figure 4.12). In many cases, monochromators are used instead of filters. Their advantage is that the selected wavelength can be set rather freely and more precisely compared to filters that are set to a given interval and adjustments can only be made by replacing them (Figure 4.13).

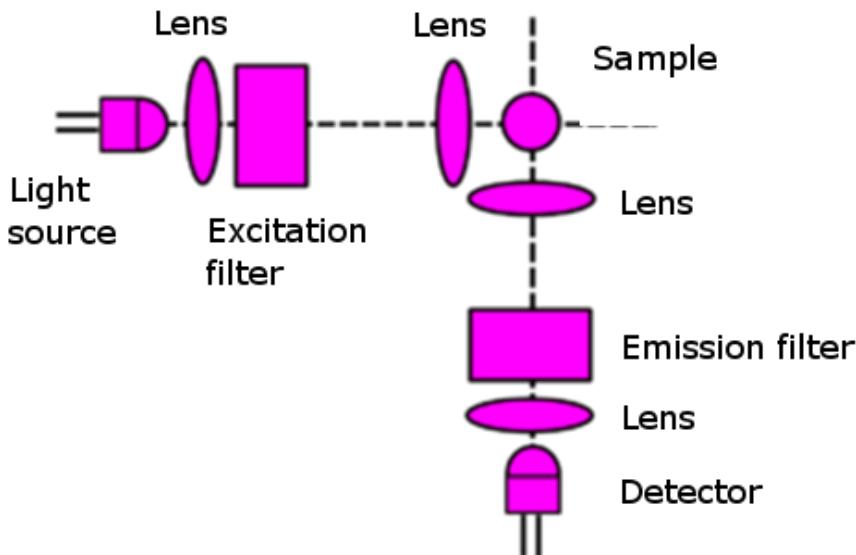


Figure 4.12. Schematic representation of the structure of a fluorimeter. The sample is placed between the light source and the detector, creating a perpendicular setup. The appropriate wavelength is selected using light filters.

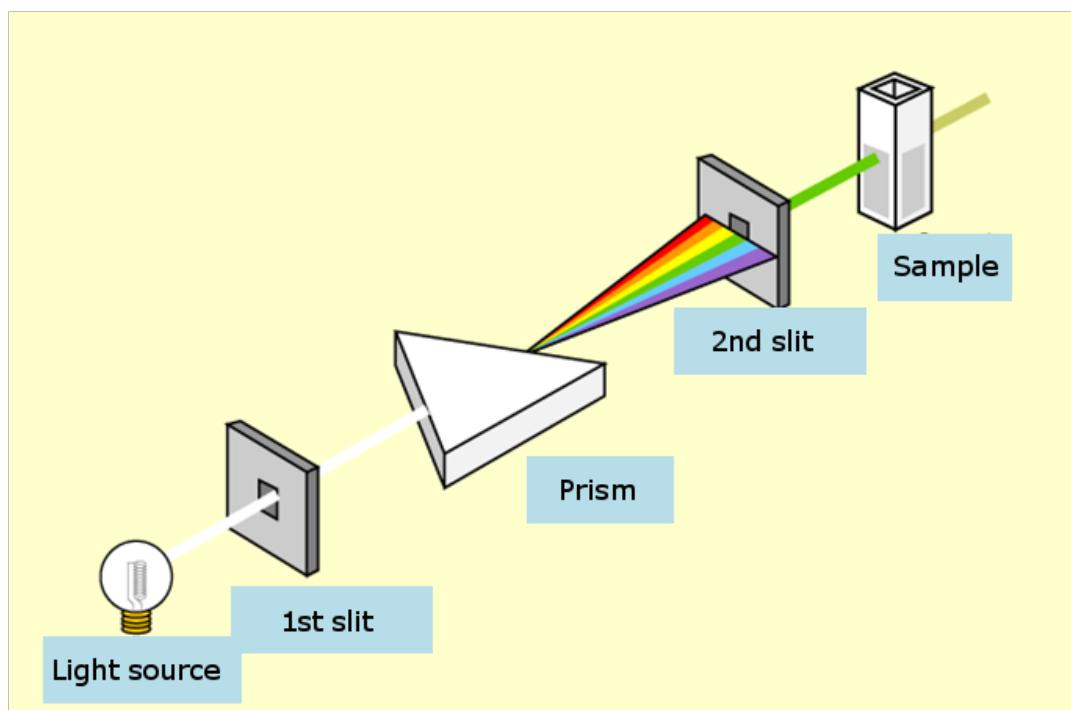


Figure 4.13. Scheme of a monochromator. From white (wide-spectrum) light, the monochromator is able to select light within a given narrow spectrum. White light is projected onto a prism splitting it to its components, effectively creating a rainbow behind it. On its way to the sample, light must pass through a small slit and therefore only a small part of the spectrum (a practically homogenous light beam) reaches it. Wavelength of the light leaving the monochromator can be changed by rotating the prism as this will let a different part of the rainbow through the slit.

This double protection of the detector from the exciting light is necessary due to the fact that the intensity of fluorescent light is usually two or three orders of magnitude smaller than that of the exciting light. This means that even if only 1 or 0.1 % of the exciting light reaches the detector, half of the detected signal intensity would arise from the exciting light and only the other half from the emission of the sample. This would result in a 50 % background signal level, as the detector is unable to distinguish photons based on their wavelength.

### 4.7.3. Fluorophores

Fluorophores are characterised by specific fluorescence spectra, namely their excitation (absorption) spectrum and emission spectrum. The excitation spectrum is recorded by measuring the intensity of emission at a given wavelength while the wavelength of excitation is continuously changed. The emission spectrum is recorded by measuring the intensity of the emitted light as a function of its wavelength while the wavelength of the exciting light is kept constant.

The shape of the excitation spectrum is usually the same as the shape of the emission spectrum. However, due to the Stokes shift, the emission spectrum is shifted towards red compared to the excitation spectrum, and usually the shape of the two spectra are mirror images of each other (Figure 4.14).

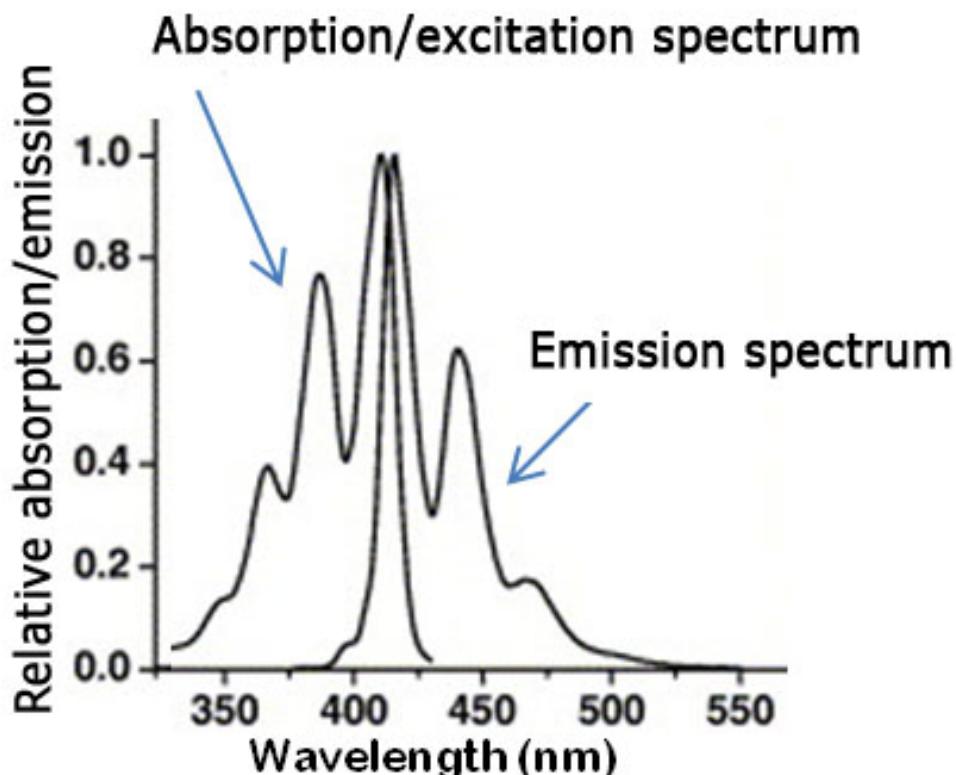


Figure 4.14. Absorption (excitation) and emission spectra of pyrene

The intensity of fluorescence of a molecule is sensitive to its environment. Emission intensity is significantly affected by the pH and the polarity of the solvent as well as the temperature. Usually, an apolar solvent and a decrease in temperature will increase the intensity. The immediate environment of the fluorophore is an important factor, too. Another molecule or group moving close to the fluorophore can change the intensity of fluorescence. Due to these attributes, fluorimetry is well suited to the study of different chemical reactions and/or conformational changes, aggregation and dissociation. In proteins, two amino acids have side chains with significant fluorescence: tryptophan and tyrosine (Figure 4.15). The fluorescence of these groups in a protein is called the intrinsic fluorescence of the protein. Tryptophan is a relatively rare amino acid; most proteins contain only one or a few tryptophans. Tyrosine is much more frequent; there are usually five to ten times more tyrosines in a protein than tryptophans. On the other hand, the fluorescence intensity of tryptophan is much higher than that of tyrosine.

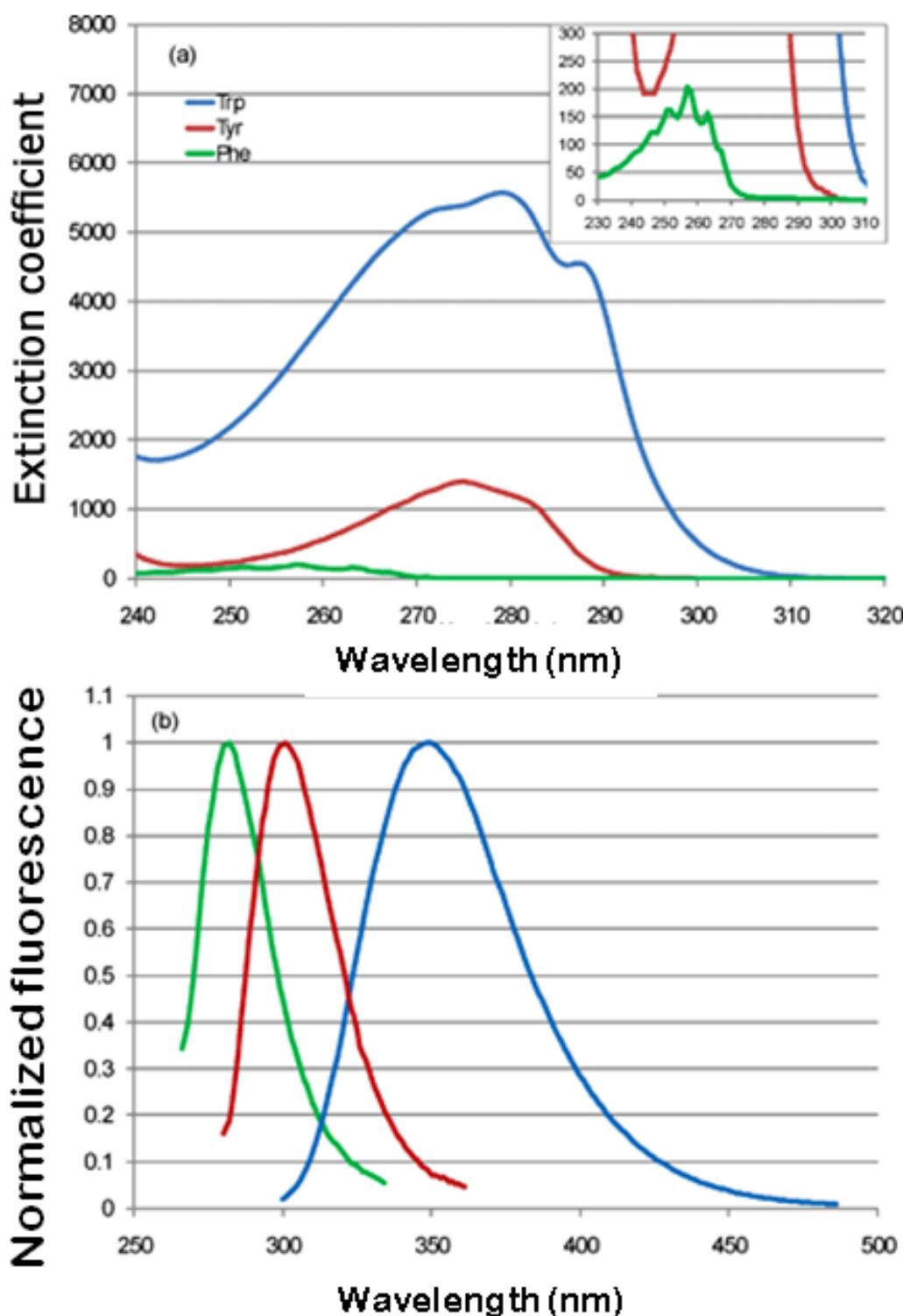


Figure 4.15. Extinction (A) and emission (B) spectra of tryptophan, tyrosine and phenylalanine. (Note that the three amino acids shown display markedly different fluorescence intensities. For visibility, emission spectra shown in panel B were normalised to their individual maxima.)

The spectra in Figure 4.15 clearly show that the fluorescence of tryptophan can be studied specifically even in the presence of tyrosines, since if the excitation is set to 295 nm and the detection of emission is set to 350 nm, the fluorescence of tyrosine can be neglected. Both the intensity of the fluorescence and the shape of the emission spectrum are sensitive to the surroundings of the side chain, which often changes upon conformational changes of the protein. Tryptophan fluorimetry is therefore suitable to detect conformational changes of enzymes and other

proteins. It can also be applied to detect the binding of ligands to proteins as well as the di- or multimerisation of proteins, provided that the reaction results in a change in the surroundings of a tryptophan side chain. The environment of tryptophans obviously changes on unfolding of proteins. Consequently, fluorescence is well suited also for following denaturation of proteins.

Tryptophan and tyrosine fluorescence is not the only way to detect and investigate proteins using fluorescence. There are proteins that undergo post-translational modifications including the covalent isomerisation of three amino acids that makes them fluorescent. The first such protein discovered was the green fluorescent protein (GFP), which is expressed naturally in the jellyfish *Aequorea victoria* (phylum *Cnidaria*) (Figure 4.16, left panel). Since then, fluorescent proteins were isolated from many other species. A large number of recombinantly modified forms of GFP were created in the last 20 years, all different in their fluorescence and colour (Figure 4.16, right panel). The intrinsic fluorescence of GFP can be used to label proteins. If we create a chimera from the genes of GFP and another protein of interest—in other words, we attach the gene of GFP to the 5' or 3' end of the gene encoding the other protein—this construct will be transcribed and translated into a protein that will have GFP fused to it at its N- or C-terminus. Thus, if using an appropriate vector we transform an organism and introduce this new gene into it, its product will show a green fluorescence when excited. Through this phenomenon, we can easily locate proteins on the tissue, cellular or subcellular levels. As a variety of differently coloured fluorescent proteins are at our disposal, we can even measure colocalisation of labelled proteins *in vivo*. The application of fluorescent proteins in biology was such a significant technological breakthrough that its pioneers were awarded a Nobel prize in 2008.

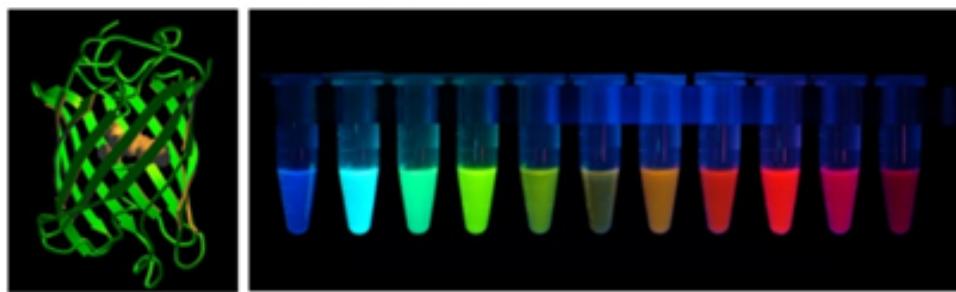


Figure 4.16. Left panel: Molecular structure of GFP. GFP has a typical  $\beta$ -barrel structure. The fluorophore is formed by the covalent isomerisation of three amino acids located in the centre of the protein (coloured orange in the figure). This core alone is not fluorescent, only when surrounded by the  $\beta$ -barrel in its native conformation. Right panel:

A variety of fluorescent proteins in Eppendorf tubes. Source: <http://www.tsienlab.ucsd.edu/>

Proteins and other biological molecules can also be made fluorescent by using extrinsic modifications. We can attach extrinsic fluorophores to biomolecules by either covalent or non-covalent bonds.

Covalent attachment of fluorophores is most often achieved by using the reactive side chains of cysteines. To this end, researchers use fluorophores that bear iodoacetamido or maleimido groups that alkylate the sulphydryl group of cysteine side chains under appropriate conditions.

Proteins can form complexes with fluorescent substrates or inhibitors also via non-covalent bonds. There exist also fluorophores that can bind to certain regions of proteins with a high affinity. For example, 8-anilinonaphthalene-1-sulfonic acid (ANS) binds to hydrophobic regions of proteins specifically and becomes strongly fluorescent when bound. We can take advantage of this phenomenon in experiments. A change in the amount of hydrophobic surfaces can occur in conjunction with structural changes induced by the binding of a ligand. Thus, addition of the ligand may cause the decrease of the amount of protein-bound ANS and thus the binding of the ligand can be studied by measuring the changes in the fluorescence of ANS. This way the binding constant of the protein and the ligand, as well as the kinetics of the binding can be examined in a simple yet quantitative manner (cf. Chapter 8).

Labelling of double-stranded DNA can also be achieved, for example, with ethidium bromide *in vitro*. When intercalated between the bases of DNA, the fluorescence of ethidium bromide will rise markedly. Formerly, visualisation of DNA in agarose gel electrophoresis was generally achieved using ethidium bromide. The dye was mixed into the agarose gel to form a complex with the DNA passing through it. When illuminated by ultraviolet light, the ethidium bromide accumulated in the DNA becomes visible due to its fluorescence. As ethidium bromide is carcinogenic, nowadays rather non-carcinogenic alternatives (e.g. SYBR Safe) are used.

Nucleic acids can also be labelled fluorescently through covalent modifications. Fluorophores can be bound to the 5' or 3' hydroxyl group. The most common way to create DNA labelled on its 5' end is to synthesise it in a PCR reaction with primers labelled at their 5' end.

There is a large number of different fluorophores available commercially that exhibit different fluorescent properties. We can make our choice based on the excitation and emission wavelengths. For example, fluorescein, one of the first fluorophores used, exhibits its absorption maximum at 494 nm and its emission maximum at 521 nm. The intensity of the fluorescence of a given fluorophore is determined by its absorption coefficient and emission efficiency (i.e. the probability that the electron emits a photon when it relaxes). This can provide another way to optimise our experiment. The extent of the Stokes shift is also an important aspect. Technically, fluorophores with a greater shift are more advantageous. The greater the difference between the excitation and detection wavelengths, the easier it is to prevent (by using filters or monochromators) the exciting light from getting into the detector. This significantly decreases the actual background.

## 4.8. Appendix

### 4.8.1. Fluorescence, phosphorescence and chemiluminescence

Even though the phenomena mentioned in the title are similar in many ways, it is important to make distinctions. In all three cases, the source of the light is an excited electron that, while returning to its ground state, can emit part of its excitation energy as a photon with some probability.

As mentioned earlier, in the case of fluorescence, excitation is performed using light of a wavelength that is optimally absorbed by the molecule. The transition between the excited and ground states is direct and fast (occurs on the nanosecond time scale).

In the case of phosphorescence, the difference lies in the manner and rate of emission. In this case, the electron does not „immediately” fall back from its excited state into its ground state, but is able to enter a particular alternative excited state. All transitions leading from this state to the ground state are so-called forbidden transitions. This does not mean that they do not happen at all, but their probability is rather low. This way, the lifetime of the excited state can increase from nanoseconds to milliseconds, minutes or even hours.

The difference between fluorescence and chemiluminescence is not in the way the emission occurs, but in how the excitation is achieved. In the latter case the energy necessary to excite the electron comes not from light but from a chemical reaction. This is the way some living organisms can produce light. We call this bioluminescence. Fireflies and anglerfishes of deep seas are well known examples of the occurrence of this phenomenon.

### 4.8.2. Photobleaching

Besides the effects listed above, excitation of fluorophores can also destroy them. This phenomenon is known as photobleaching. Photobleaching can occur during every excitation event with a given probability. Thus, its extent will be proportional to the intensity of the exciting light and the duration of the illumination. This effect can pose a serious problem when conducting long-term experiments. To counteract photobleaching, more durable fluorophores have been invented (although the rate of decay is not zero even in the case of these molecules).

Photobleaching can also be used to our advantage, as exemplified by the FRAP (Fluorescence Recovery After Photobleaching) technique. Using FRAP we can examine e.g. the movement of membrane-associated molecules in cells. After fluorescent labelling a molecule of interest in the membrane, we can induce the decay of fluorophores purposefully in the vicinity of the membrane. Subsequently we can assess how rapidly the fluorophores in the surrounding area enter the bleached patch.

### 4.8.3. Fluorescence anisotropy and circular dichroism

The properties of light are not completely characterised by its intensity and wavelength. It also has polarisation, which is the function of the direction and the fluctuation of the electric field vector. (All photons are polarised;

unpolarised light is the mixture of differently polarised photons). Polarisation can be linear (when the electric field vector changes in a given plane) or circular (when it changes in circles) (Figure 4.17). Both types of polarisation can be utilised in spectroscopy.

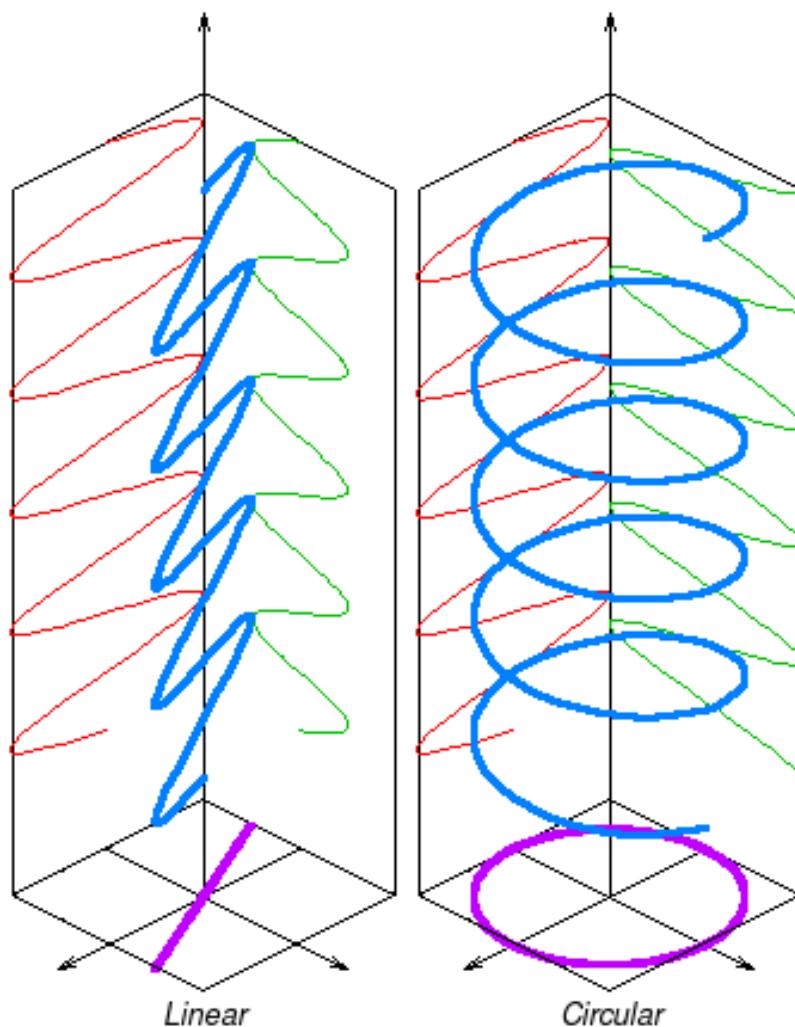


Figure 4.17. Linear and circular polarisation of light. The figure shows an electromagnetic wave (light), a transversal wave propagating vertically and the curve we get by following the electric field vector (blue line) creating it. It also shows what we get if we split the vector into its perpendicular components (red and green). Source: <http://scienceblogs.de/astrodicicum-simplex/wp-content/blogs.dir/28/files/2012/06/i-897a5d594e065d9046b61bb92a8d3d02-polarisation.png>

Linear polarisation has an effect on fluorescence through the efficiency of the excitation. In addition, the polarisation of the emitted light is a function of the position of the fluorophore. This phenomenon is known as fluorescence anisotropy. Using this effect, we can examine the rotation of a fluorophore, which can provide information on the size and environment of the labelled molecule (larger molecules or those located in more tightly packed environments will rotate slower). If the labelled molecule binds to something of a significant size (e.g. a protein), this event will appear in the experiment as a change in the size of the label-containing particle. This way, association/dissociation reactions can be studied.

Utilising circular polarisation is one of the possible ways to examine the chirality of molecules. Circularly polarised light is absorbed differently by the two forms of chiral molecules (e.g. L and D amino acids). This difference is called circular dichroism and is usually described by the difference between the extinction coefficients of the two enantiomers ( $\Delta\epsilon$ ). Plotting this parameter against the wavelength we get a so-called circular dichroism (CD) spectrum. It has been observed that polymers with a non-bilaterally symmetric structure also exhibit dichroism. Thus, via CD measurements performed in the UV range, the ratio of the different secondary structures in proteins and the double helix in nucleic acids can be studied. Although this measurement does not provide enough inform-

ation to determine the exact structure of molecules, we can gain important insights into their general features and detect structural changes. Consequently, circular dichroism measurements can be used to observe the folding and denaturation of proteins. It has also been shown that metal ions show CD in the visible spectrum when they are located in a chiral environment. As they lose this attribute when they are in solution, this effect can be utilised to assess the formation of metalloprotein complexes.

## 4.8.4. Quenching and FRET

Attentive readers may have noticed that we always described fluorescence emission as one of multiple possible ways to lose the energy of excitation. There are multiple reasons for this:

1. The entirety of the energy can be dissipated as heat.
2. Some electrons can enter a third, alternative state (cf. phosphorescence).
3. It is possible to transfer the energy to a nearby molecule called acceptor.

Phenomena in the last category are called quenching. If the acceptor molecule is not fluorescent, the transfer of energy will lead to loss of fluorescence. If the acceptor is fluorescent, by exciting the donor we will be able to detect acceptor emission, with the consideration of certain spectral requirements.

Quenching can occur through different mechanisms, of which the most common is FRET (Förster Resonance Energy Transfer). The transfer of energy happens in a non-radiative manner, i.e. there is no photon emission by the donor molecule. For this to happen, two prerequisites must hold: compatibility and physical proximity of the molecules involved. The emission spectrum of the donor and the excitation spectrum of the acceptor must overlap with each other. The efficiency of FRET is inversely proportional to the sixth power of the distance between the donor and the acceptor. Generally, if they are located farther apart than 10 nm, there is practically no interaction. This makes FRET a suitable effect for methods to study the distance and interaction of molecules or, with appropriate labelling, even parts of molecules.

In the basic approach, the distinction between the “near” and “far” states is made by observing the emission of the acceptor, based on its wavelength. This requires the donor and acceptor molecules to be different. But even if the two are identical, there is a possibility to detect FRET using an anisotropy measurement, as the latter also undergoes changes upon energy transfer.

---

# **Chapter 5. Cell disruption, cell fractionation and protein isolation**

by Gábor Pál

In this chapter we survey how individual proteins can be isolated in homogeneous form from biological samples—especially from tissues of multicellular organisms. Naturally, the more information we have on the protein to be investigated, the more straightforward it is to establish a well-suited and efficient isolation protocol.

The first question to consider is the distribution of the given protein among the various tissues of the organism. Obviously, the tissue in which the given protein is the most abundant should be used as the starting point of the isolation protocol. The next question is whether the protein is intracellular, extracellular (secreted) or membrane-bound.

If it is intracellular, the subcellular distribution of the protein should be considered. If the protein is associated to one of the many organelles of the eukaryotic cell, that organelle should be first isolated. The process through which individual organelles (plasma membrane, nucleus, mitochondria, etc.) are separated from one another is called cell fractionation. The process of cell fractionation starts with the disruption of the tissue and its cell constituents by a homogenisation procedure performed as gently as possible. Once the cells are opened up, individual organelle types can be separated from each other by various types of centrifugation techniques.

## **5.1. Cell disruption**

The technical details of the cell disruption procedure largely depend on the type of tissue or cells to be homogenised. In the case of multicellular organisms, the first aim is to disintegrate the tissue into individual cells by abolishing the connections that organise the cells into the given tissue. Then the plasma membrane and, in case of plants, fungi and bacteria, the cell wall need to be ruptured. The harshness of the treatment can greatly vary depending on the tissue and cell type. For example, in the case of blood cells that do not need to be disintegrated from a solid tissue, even a mild osmotic shock using a hypotonic solution can lead to the rupture of the cell membrane. In the case of cells having a cell wall or tissues stabilised by a strong extracellular matrix, simple osmosis-based treatments are inefficient. In such cases various mechanical methods applying shearing force on the cells can be used. The two most frequently applied tools are high-speed laboratory blenders and ultrasonic cell disruptors. While blenders can disrupt even highly structured strong tissues, ultrasonic cell disruptors applying ultrasound ( $\sim 20\text{--}50\text{ kHz}$ ) to the sample (sonication) are used mostly in the case of cell suspensions. The ultrasonic cell disruptor generates the high-frequency waves electronically. These shock waves are transmitted to the cell suspension via an oscillating metal probe. The oscillation causes large localised pressure inhomogeneity resulting in cavitation eventually disrupting the cells.

There are several other procedures that also use shearing force to open up cells. Some of these apply high pressure to pump the cell suspension through a very narrow channel or orifice into a low pressure container. Due to the sudden drop of pressure the cells “explode” in the container. Shearing force can be also generated by a pair consisting of a carefully designed glass tube and a tightly fitting glass pestle, called the Potter-Elvehjem homogeniser, or a potter in short. The diameter of the tube is just a little larger than that of the pestle. The sample is pushed into the very narrow space between the sides of the tube and the pestle. The shearing force is generated as the cell suspension squeezes up and past the pestle. This method is applied on the suspension of individual cells (already dissociated tissues, blood cells etc.).

Plant cells protected by cell wall are most often disrupted by various grinding methods. Manual grinding is the most common method. The tissue is usually frozen in liquid nitrogen and then crushed using a mortar and pestle.

Optimal cell disruption methods open up a high percentage of the cells in the sample while preserving the organelles or molecules to be investigated in their native state. This is not a trivial task. In order to preserve the native state of most organelles and molecules, the procedure should be quick and the heat generated by the disruption method should be dissipated by intensive cooling. This helps to avoid heat denaturation of proteins and also lowers the rate of unwanted chemical reactions such as oxidation or proteolytic cleavage of proteins. To further suppress these

chemical reactions, oxidation can be prevented by the addition of reducing agents and proteolysis can be controlled by the addition of a mixture of protease inhibitors, often referred to as a protease inhibitor cocktail. In order to extract the content of the cell into a native-like solution, the buffers used for cell disruption often mimic the cytosol in terms of pH and ionic strength. The sample might contain trace amounts of heavy metal ions. Such ions can form complexes with various amino acid residues of proteins. To prevent such complex formation, a chelating agent, most often ethylene-diamine-tetraacetic acid (EDTA), is added to the homogenisation buffer to sequester the heavy metal ion components.

## 5.2. Cell fractionation

The major goal of cell fractionation is to separate the various types of cell organelles from each other and from the cytosol. As mentioned, cell disruption should be intensive enough to open up a large fraction of the cells but still gentle enough to preserve the native state of the organelles and the soluble components of the cytosol. Upon cell disruption, the plasma membrane and the endoplasmic reticulum become disintegrated into small membrane vesicles. These vesicles—along with the native organelles of the cell and soluble molecules of the cytosol—compose the mixture from which the components need to be separated into different fractions. The most important fractionation technique applied to this mixture is centrifugation using specialised laboratory equipment. The following section reviews the principles and major types of laboratory centrifugation.

## 5.3. Centrifugation

When an object attached to a rope is whirled around, one can feel that the rope must be pulled inward towards the centre of the rotation in order to keep the object on the orbit. This force prevents the object from getting away and move with a constant speed along a straight tangential line. The inward force with which one has to pull the rope is called the centripetal force. One can also define the outward force, the centrifugal force, by which the object pulls the rope. This force is equal in magnitude to the centripetal force but has the opposite direction. The centrifugal force ( $F_c$ ) is a virtual, so-called fictional force emerging due to the inertia of the object. Yet, because it leads to a simpler mathematical formalism, equations describing the processes when solutions are centrifuged use the  $F_c$  force.

According to the well-known Newton equation:

$$F = m \times a \quad (5.1)$$

Upon centrifugation, acceleration equals the product of the radius of the orbit and the square of the angular velocity:

$$a = \omega^2 \times r \quad (5.2)$$

The fictive centrifugal accelerating force in vacuum is therefore:

$$F_c = m \times \omega^2 \times r \quad (5.3)$$

The product of the radius and the square of the angular velocity equals the centrifugal accelerating potential. Traditionally, and perhaps somewhat misleadingly, the magnitude of this potential is compared to the Earth's gravitational accelerating potential ( $g$ ), and has been expressed in "g" units. The reason is quite simple. Earth's gravitational potential, similarly to the accelerating potential provided by centrifugation, can also sediment particles dispersed in solution. This type of quantitation shows how many times centrifugation is more effective to sediment particles compared to the gravitational effect of Earth. In the fastest laboratory ultracentrifuges the applied accelerating potential can exceed 1 000 000 g.

When solutions are centrifuged, the particles are not in vacuum but in a solvent having a given density (mass/volume). Importantly, the centrifugal force acts not only on the particles, but on the solvent too. If the density of the particle equals the density of the solvent, the particle will not move relative to the solvent, and its velocity along the radius will be zero. If the density of the particle exceeds that of the solvent, the particle sediments (sinks), i.e. it moves outwards along the radius, while the displaced solvent molecules move inwards. In the opposite

case when the density of the particle is lower than that of the solvent, the particle floats—it moves inwards while the displaced solvent molecules move outwards.

In order to provide a simple mathematical description of this phenomenon, the buoyancy factor has been defined as follows:

$$(1 - \frac{\rho}{\rho_r}) \quad (5.4)$$

The numerator of the fraction contains the density of the medium (solvent) while the denominator contains the density of the particle.

Introduction of the buoyancy factor leads to the following equation:

$$F_c = m \times (1 - \frac{\rho}{\rho_r}) \times \omega^2 \times r \quad (5.5)$$

This equation clearly shows that, upon centrifugation, the force acting on a given particle is a function of the mass of the particle, the relative density of the particle (compared to that of the medium), the angular velocity of the rotation and the distance of the particle from the centre of the rotation (i.e. from the spindle of the centrifuge).

The first two of these parameters, namely the mass and the density, are characteristic of the particle and differences in these parameters can allow for the physical separation of different types of particles. As we will see, there are two major types of centrifugation-based separation techniques. In one technique called differential centrifugation, the separation is based on both particle mass and density. In the case of the other, called equilibrium density-gradient centrifugation, the separation is based strictly on the density of the particles.

As soon as the particles are accelerated by the centrifugal force and start moving towards the spindle, a dragging force ( $F_d$ ) called friction is exerted on them by the medium. This force, which has a direction opposite to the direction of the particle movement, is proportional to the velocity of the particle. At the typically very low speed of the sedimentation process, the  $F_d$  force is a linear function of the velocity. The ratio of the force and the velocity is defined as the frictional coefficient ( $f$ ). The value of  $f$  is a function of the viscosity of the medium and of the size and shape of the particle as described below by Stokes' law:

$$f = 6 \times r \times \pi \times \mu \quad (5.6)$$

In this equation "r" denotes the Stokes radius. If the particle is spherical, this equals the radius of the particle. If the particle is not spherical, "r" (a virtual value) denotes the radius of a spherical particle that has identical diffusion properties as the non-spherical particle in question and  $\mu$  denotes viscosity of the medium. Note that the value of the frictional coefficient is proportional to the radius of the particle. The larger the particle, the higher dragging force is exerted on it by the medium.

In the course of centrifugation, the velocity of each particle is increasing due to the accelerating force  $F_c$ . However, as the velocity increases, the dragging force also increases. Therefore, the velocity of each particle can increase to a given value where the value of the dragging force  $F_d$  reaches the value of the accelerating  $F_c$  value. The magnitude of the two opposing forces becomes equal in a very short time:

$$F_c = F_d = f \times v \quad (5.7)$$

Once the magnitude of the two opposing forces becomes equal, the resultant force becomes zero. Therefore, the particle will move with a constant velocity characteristic to that particle at the given accelerating potential and medium. (A similar phenomenon is described in Chapter 7 on electrophoresis. There, the accelerating force is proportional to the charge instead of the mass of the particle, but the friction force and the phenomenon of two opposing forces leading to a characteristic particle velocity is analogous.) Substituting  $F_c$  into the previous equation leads to the following equation:

$$f \times v = m \times (1 - \frac{\rho}{\rho_r}) \times \omega^2 \times r \quad (5.8)$$

If the above equation is rearranged by dividing particle velocity with the centrifugal acceleration potential, the resulting equation will lead to a useful parameter. This is the sedimentation coefficient (its unit of measure is one over seconds), which is usually expressed in Svedberg units. This coefficient describes the sedimentation propensity of the particle. It provides the characteristic sedimentation velocity of a particle triggered by a unit level of accelerating potential.

$$S = \frac{v}{\omega^2 \times r} = \frac{m \times (1 - \frac{\rho}{\rho_r})}{f} = \frac{M \times (1 - \frac{\rho}{\rho_r})}{Nf} \quad (5.9)$$

The numerator of the equation contains all parameters that favour sedimentation. The higher the mass and relative density (compared to the medium) of the particle, the higher its sedimentation velocity will be when unit accelerating potential is applied. The mass of the particle of a given density, of course, is linearly proportional to its volume. In other words, the mass is a linear function of the cube of the particle radius. The denominator contains the parameter that negatively influences sedimentation speed. The larger the frictional coefficient, the lower velocity will be triggered by unit level acceleration potential. As we have seen, the frictional coefficient is a linear function of the particle radius. As the accelerating force is a linear function of the third power of the radius, while the dragging force is a linear function of the first power of the radius, the velocity of the particle will ultimately be proportional to the second power of the particle radius. *If two particles have identical density, the larger particle will sediment faster and the ratio of the velocities will follow a square law with respect to the ratio of the particle radii.* This relationship provides the basis for the so-called differential centrifugation methods.

### 5.3.1. Differential centrifugation: cell fractionation based primarily on particle size

The density of the various organelles differs on a smaller scale than their size. Therefore, while both size and density affect sedimentation velocity, their size difference dominates when organelles are separated by centrifugation.

In the procedure of differential centrifugation, cell constituents are separated from each other by their Svedberg value. Several consecutive centrifugation steps are applied in the order of increasing accelerating potential. Each individual centrifugation step relies on the different sedimentation speed of the different cell constituents at the given acceleration potential. At a properly chosen acceleration potential, almost 100 % of the largest component will sediment in the time span of the centrifugation. The sedimented organelles form a pellet at the bottom of the centrifuge tube. The potential should be set so that in the same period of time only a small portion of all smaller constituents latch on to the pellet (Figure 5.1).

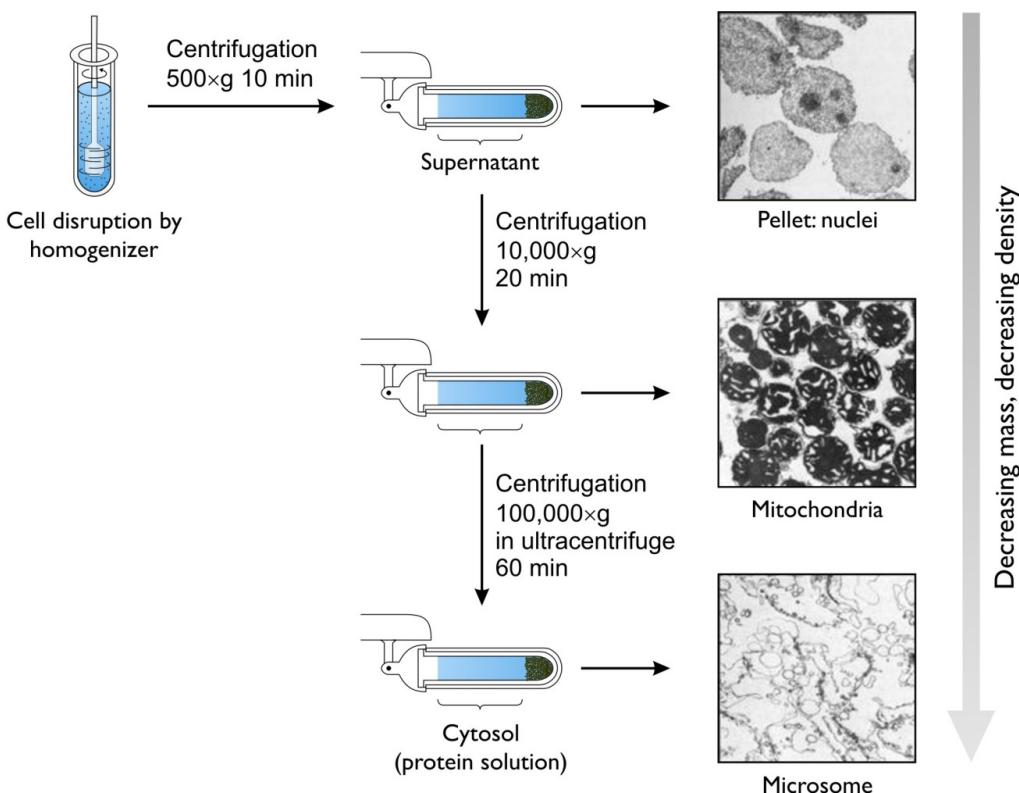


Figure 5.1. Differential centrifugation. In the course of differential centrifugation, consecutive centrifugation steps are applied. The consecutive centrifugation steps follow each other in the order of increasing centrifugal acceleration potential. During the first centrifugation, only the largest and/or heaviest cell constituents sediment in the time frame of the centrifugation. Typically, only nuclei and undisrupted whole cells form the pellet. The supernatant of the first centrifugation step is further centrifuged in the consecutive step at higher acceleration potential and typically for a longer period of time. Following this scheme, ever smaller and/or lower-density cell constituents can be sedimented.

The disrupted cell homogenate is centrifuged first at a relatively low accelerating potential of 500 g for 10 minutes. Under these conditions, only particles having the highest Svedberg value, intact cells and nuclei will form the pellet. All other cell constituents will sediment at a much lower rate and remain in the homogenate. The supernatant of the first centrifugation is transferred into an empty centrifugation tube and is subjected to another centrifugation step, now at a significantly higher accelerating potential of 10,000 g and for 20 minutes. These conditions favour sedimentation of mitochondria, lysosomes and peroxisomes having lower Svedberg values than nuclei. Many cell constituents still remain in the supernatant, which is again transferred into an empty tube. This tube is placed into an ultracentrifuge and, with an accelerating potential of 100 000 g in one hour, the so-called microsomal fraction sediments. This fraction contains mostly artificial vesicles with a diameter of 50–150 nm that originate mostly from the endoplasmic reticulum and are generated by the cell disruption procedure. Other natural cell constituents of the same size range will also contribute to this fraction. After this third centrifugation step, the supernatant contains mostly macromolecules and supramolecular complexes such as ribosomes. By applying an accelerating potential as high as several hundred thousand g, ribosomes and large proteins can also be sedimented.

### 5.3.2. Equilibrium density-gradient centrifugation: fractionation based on density

In the previous section we introduced the method of differential centrifugation. For simplicity, we stated that the constituents of the sample were separated in a medium of homogeneous density. This first approximation has didactical advantages as it makes the basic principle of differential centrifugation easier to comprehend. Nevertheless, it is sometimes advantageous to use a very shallow density gradient in the medium during differential centrifugation. This is done only to suppress convectional flows in the medium that could unsettle and mix layers of already separated cell constituents.

The essence of equilibrium density-gradient centrifugation is principally different. In this case, a rather steep density gradient is created in the medium—in such a manner that the density of the medium gradually increases towards the bottom of the centrifuge tube. This is achieved by using a very high-density additive, for example caesium chloride ( $\text{CsCl}$ ). The density gradient is created as follows. When the centrifuge tube is filled with the medium, a high concentration  $\text{CsCl}$  solution is added first. Subsequently, in the process of filling the tube, the concentration of  $\text{CsCl}$  is gradually decreased resulting in a  $\text{CsCl}$  gradient and, as a consequence, a density gradient in the tube. The sample is layered on the top of this special medium (Figure 5.2).

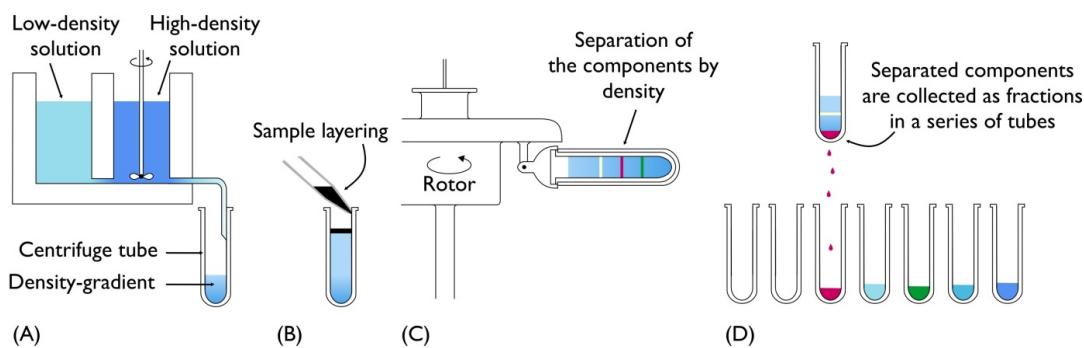


Figure 5.2. Equilibrium density-gradient centrifugation. In the course of equilibrium density-gradient centrifugation, a concentration gradient of a high density compound such as caesium chloride is generated. (The compound should not react with the biological sample.) The concentration gradient of this special additive creates a density gradient in the centrifuge tube. The density gradually increases toward the bottom of the centrifuge tube. The sample is layered on the low-density top of this gradient. As the centrifugation begins, each compound of the sample starts to sediment. By doing so, the compounds travel through layers of increasing density. As soon as a compound reaches the layer where the density equals its own density, the compound stops sedimenting. At this layer, no resultant force is exerted on the particle and thus it will float. As a result, equilibrium density-gradient centrifugation separates compounds from each other independently of their size, solely by their density, in a single run.

In the course of centrifugation, particles start to sediment moving towards the bottom of the centrifuge tube. By doing so, they travel through an increasing density medium. Each particle sediments to a section of the medium where its own density equals the density of the medium. At this section, the buoyancy factor becomes zero and, as a consequence, the accelerating force acting on the particle also becomes zero. The particle stops sedimenting. If it moved further towards the bottom of the tube, it would meet a higher density medium and a force opposing to its moving direction would be exerted on it, turning the particle back. If, by travelling backwards, it would meet a density lower than its own density, it would sediment again. As a consequence, this method separates particles exclusively based on their density. It is an equilibrium method in which, by the end of the separation, the system reaches a constant state. (In this aspect, this method shows an interesting analogy to the isoelectric focusing (IEF) method reviewed in Chapter 7. The two methods separate particles by entirely different characteristics (density versus isoelectric point), but in both cases, the separation leads to an equilibrium state. Both methods apply a gradient, but in the case of IEF a pH gradient is created.)

Note that the two centrifugation approaches introduced above separate particles by partially different characteristics. Consecutive combination of the two methods can lead to a more efficient separation than achieved by any of the methods alone. Therefore, to increase separation efficiency, fractions generated by differential centrifugation can be subjected to a subsequent density-gradient centrifugation step to further separate individual components (Figure 5.3).

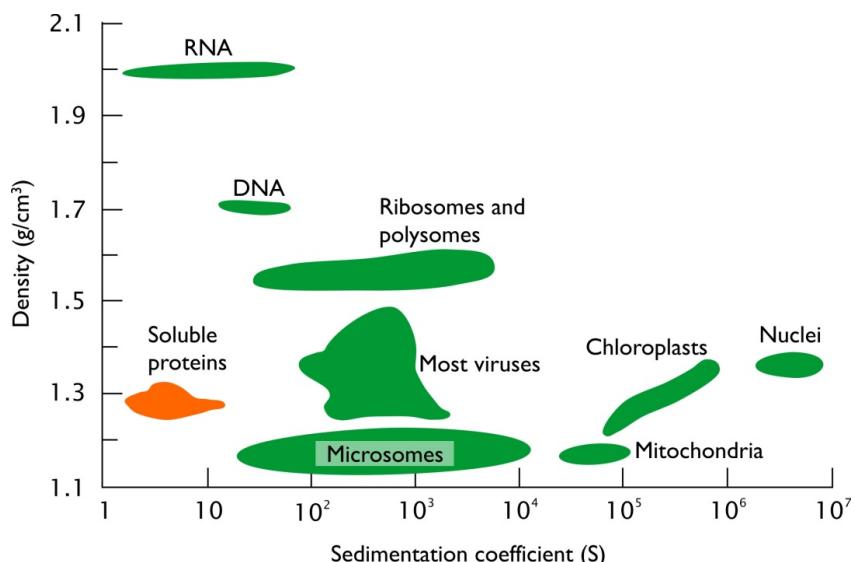


Figure 5.3. Combination of differential centrifugation and density-gradient centrifugation. Differential centrifugation separates compounds primarily based on their size, while density-gradient centrifugation separates compounds exclusively based on their density. Compounds that have different density but sediment in the same fraction during differential centrifugation can be separated by a subsequent step of density-gradient centrifugation. Two such consecutive steps of the two centrifugation methods can provide significantly higher separation efficiency than either procedure alone.

## 5.4. Low-resolution, large-scale protein fractionation

Once individual cell constituents have been separated to different fractions, proteins of these fractions can be further fractionated. The ultimate goal is to isolate the protein to be investigated in a homogeneous and functional form. In order to obtain a homogeneous protein solution, typically several purification procedures have to be combined consecutively. For each novel protein, the optimal isolation protocol needs to be established in an iterative trial and error process.

These multi-step protocols usually start with low-cost, large-scale but low-resolution fractionation steps. These steps are based mostly on characteristically different solubility of individual proteins depending on pH, ionic strength, temperature etc. The applied techniques usually include centrifugation or filtering to remove precipitated insoluble proteins. The components of the usually still crude fractions are then further separated by various smaller-scale but higher-resolution chromatography techniques (reviewed in Chapter 6.)

Obviously, individual proteins are separated from each other based on characteristic differences in their physico-chemical properties. Mostly, the following characteristics are important from this aspect. What is the temperature and pH range where the protein to be isolated preserves its native state? What is the isoelectric point (pI) of the protein? What is the molecular mass of the protein? Is the protein composed of a single polypeptide chain, or it is a multi-subunit protein? If it is composed of more than one polypeptide chain, what is the molecular mass of the individual chains? What is the subunit composition and the native molecular mass of the protein?

In order to establish an optimised protocol, one should be able to determine the quantity and, in an optimal case, the functionality of the protein to be investigated in the presence of an excess of irrelevant ‘contaminating’ proteins. If the protein to be isolated is an enzyme, a specific chemical reaction catalysed by the enzyme can be used for functional testing. If certain conditions apply (see later), this enzymatic reaction can accurately measure the amount of the functional enzyme molecules. With such a quantitation method, one can measure the initial quantity of the enzyme in the first sample and the actual quantity of the enzyme after each fractionation step. By dividing the quantity after fractionation with that before fractionation, we can obtain the yield of the given purification step. In general, one aims to choose purification methods that provide a high yield. It is equally important to apply fractionation techniques that, while retain a large proportion of the protein to be isolated, remove a large proportion of the

‘contaminating’ proteins. Thus, it is also important to measure the total amount of proteins in the sample after each purification step and to calculate the target protein versus total protein ratio. Dividing the ratio after fractionation by that before fractionation provides a useful descriptor of the purification called enrichment. An efficient purification method has a high yield (retains a large percentage of the target protein) and, at the same time, provides high enrichment (removes a large proportion of the contaminating proteins).

If additional purification steps do not further increase the enrichment value, this can be an indication that the sample has become homogeneous and contains only the target protein. If the target protein is not an enzyme, or it is an enzyme for which no selective activity assay exists, other specific detection methods can also be used. All of these rely on a selective binding interaction between the target protein and another—usually, labelled—molecule. If the target protein is an enzyme, the interacting molecule can be its specific inhibitor. If the target protein is a receptor, the interacting molecule can be a specific ligand. In a more general case, the interacting molecule can be a monoclonal antibody raised specifically against the target protein (see Chapter 7, Western blot). The interacting molecule can be labelled with a coloured or fluorescent dye to permit its detection by spectrophotometry or spectrofluorimetry. It can also be covalently linked to an enzyme for which sensitive spectrometric detection assays are available.

If no specific interacting molecule is available but the molecular mass of the target protein is known, a less reliable but almost universal method, SDS-PAGE (see Chapter 7) can be applied to monitor the purification process.

As already mentioned, the low-resolution, high-scale initial purification steps rely mostly on solubility differences and, in some cases, on size differences.

### 5.4.1. Fractionation methods based on solubility

The solubility of a protein is a measure of the maximal quantity of the given protein that a unit volume of solvent can keep in solution. Solubility is often expressed in units of mg/mL. Solubility depends largely on the composition of the solvent. Before going into details about the influence of solvent composition, it is important to clarify whether solubility refers to native or denatured proteins. A popular misconception based on oversimplification is that native proteins are always soluble while denatured proteins always precipitate from the solution. The following section focuses on the solubility of native proteins but later we will discuss the case of denatured proteins as well.

#### *pH-dependence of solubility*

The solubility of native proteins depends primarily on the pH and ionic strength of the solvent. It is widely known that proteins have the lowest solubility when the pH equals their pI value. At such pH value, the number of positive charges on the protein equals that of negative charges. Consequently, in this state the level of ionic repulsion between the protein molecules is minimal, while the number of possible intermolecular attracting ionic interactions is at its maximum (Figure 5.4).

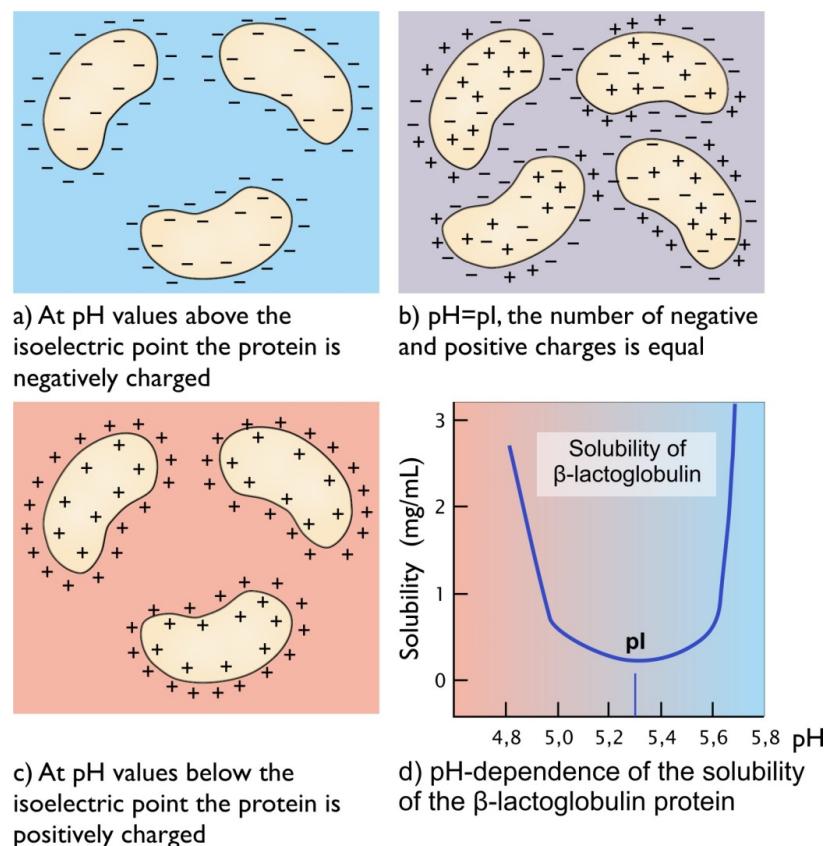


Figure 5.4. pH dependence of protein solubility. The solubility of proteins is at its minimum when the pH of the solution equals the pI value of the protein. At this pH, the number of the positive and negative charges of the protein is identical. As a consequence, the number of potential attracting intermolecular ionic interactions is at a maximum, which facilitates aggregation of the protein molecules. At pH values below or above the pI, the protein molecules have a net charge and so they tend to repel each other.

Naturally, a minimum in solubility does not mean that all molecules of a given protein precipitate when the pH equals their pI value. When the concentration of the given protein is below the solubility value, all of its molecules still remain in solution. When the protein concentration exceeds the solubility value, a dynamic equilibrium exists between the precipitated and the soluble set of molecules. Note that both the soluble as well as the insoluble subset contains native proteins—precipitation is not linked to denaturation. If one aims to precipitate native proteins, it is advisable to perform the experiment at a pH that equals the pI of the protein, since proteins are the most stable at pH values equal to their pI's.

#### *Dependence of solubility on ionic strength*

Solubility of proteins in distilled (deionised) water is significantly lower than in solutions containing ions. Accordingly, when salt is added to an ion-free protein sample in which the protein is dissolved above its solubility, the originally precipitated insoluble fraction becomes soluble. This phenomenon is called ‘salting-in’ (Figure 5.5). Apparently, ions added at a small concentration increase the solubility of proteins. As it has been discussed in the previous paragraph on pH-dependence, the solubility of proteins is lowered by attracting intermolecular ionic interactions between opposite surface charges of the protein molecules. Ions incorporated by the solvent act as counter ions and efficiently shield the intermolecular attraction between proteins. This effect readily explains the salting-in phenomenon.

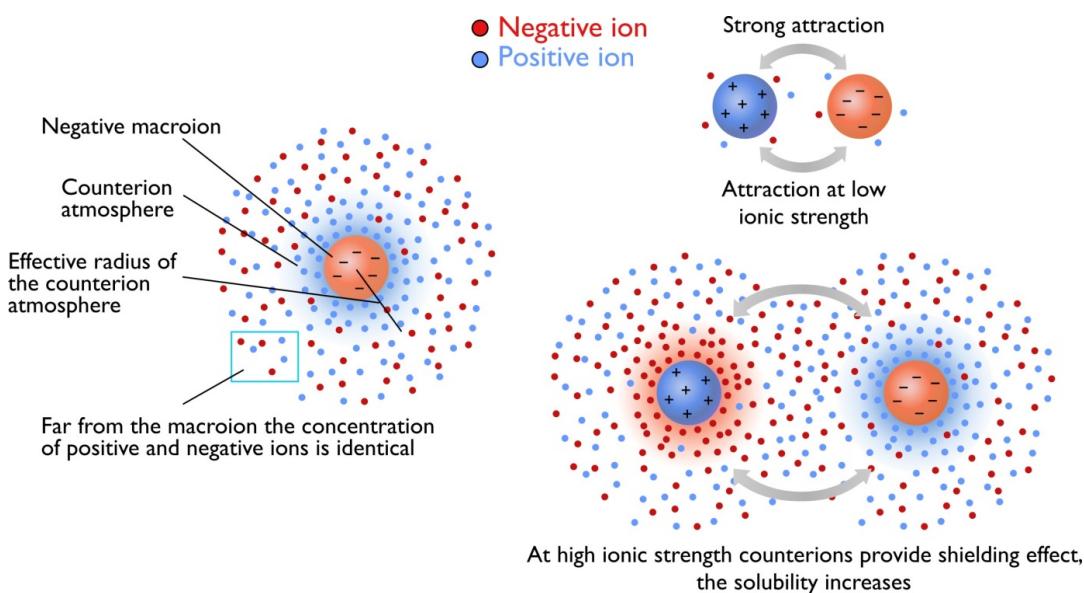


Figure 5.5. The ‘salting-in’ effect. The solubility of proteins is usually low in deionised water, but it increases with increasing ionic strength of the solution, up to a certain point. As the figure illustrates, dissolved salt (ions) facilitates solvation of proteins by shielding the exposed charges of protein molecules. This shielding effect hinders intermolecular attracting interactions between opposite charges of individual protein molecules and, by doing so, it also hinders protein aggregation.

When the concentration of ions is increased above a critical level, the solubility of proteins gradually decreases. Most importantly, the critical ion concentration level is characteristic of the identity of the protein. Accordingly, at high protein concentration, if the ionic strength of the solvent is increased, the excess of the proteins above the solubility level will precipitate while preserving their native state. This phenomenon is called salting-out. The salting-out effect can be explained as follows. When ions (salt) are added to aqueous solutions, ions become solubilised such that a multilayer water boundary, a hydration shell is formed around them. In chemically pure water, the concentration of water molecules is  $\sim 55.6\text{ M}$ . When the concentration of ions exceeds several mol/litre, a very high percentage of the originally free water molecules will become engaged in the hydration shell of ions. Formation of more and more new hydration shells around ions competes for water molecules hydrating proteins. As the hydration shells around proteins decrease intermolecular attracting forces, mostly apolar interactions between protein molecules take over leading to precipitation of native protein molecules.

Generally, various salts of univalent positive ions are used for this purpose, with the most common salt being ammonium sulphate. As the solubility of different proteins depends differently on salt concentration, the salting-out phenomenon can be applied as a protein fractionation method. Let us consider the following ammonium sulphate fractionation scheme. Let us suppose that we have a protein mixture and the protein to be isolated has a medium sensitivity to ammonium sulphate precipitation. When—very slowly and gradually, e.g. by adding with mild stirring a saturated ammonium sulphate solution—ammonium sulphate is introduced in the solution, many proteins precipitate. The ammonium sulphate concentration can be increased up to a level where the target protein remains in the solution while many other, more sensitive proteins precipitate. At this point, the precipitated proteins can be removed simply by centrifugation. (Clumps of precipitated proteins represent an enormous particle size and a very high corresponding Svedberg value. Therefore they sediment at a very high rate.) The supernatant will contain our target protein and many other proteins. The ammonium sulphate concentration can then be further increased to a level where the target protein and some other proteins precipitate, while many other proteins still remain in solution. After a second centrifugation step the supernatant can be disposed of and the pellet containing the target protein can be collected. The solution of the resolubilised pellet will contain a protein mixture in which the target protein will be enriched compared to the starting sample. Then the high concentration of ammonium sulphate can be decreased by dialysis (see below and in Chapter 2) or size exclusion chromatography (Chapter 6).

#### *Irreversible precipitation methods*

As already mentioned, the relationship between solubility and native/denatured state of proteins is much more complex than usually considered. Native globular proteins have most of their apolar residues buried in the hydro-

phobic core of the molecule. Nevertheless, the dominantly polar surface of the protein still contains hydrophobic patches of apolar residues. The size and distribution of these hydrophobic spots widely varies with individual proteins. Upon unfolding, all apolar residues become solvent-accessible. Water molecules form a highly structured, low-entropy clathrate structure around apolar residues. The free enthalpy level of the thermodynamic system can decrease if the amount of the clathrate structure decreases. This indeed happens, because this way the entropy of the system increases. The amount of the clathrate structures can decrease if the apolar residues form short-range, weak, apolar-apolar interactions with each other. These interactions are weaker than the apolar-polar interactions between the apolar residues and the water molecules, but the newly formed polar-polar interactions between water molecules partially balance this loss of interaction energy. As already mentioned, the breakdown of the clathrate structures has a major contribution to lowering the free energy of the system. As a result, the apolar residues are expelled from water and interact with each other rather than with water molecules. This phenomenon is denoted as the hydrophobic effect. Due to the hydrophobic effect, denatured globular proteins tend to precipitate from solution. However, it might be confusing that many popular denaturing agents such as urea or sodium dodecyl sulphate (SDS) are highly effective solubilisers of denatured proteins. When urea or SDS is used as denaturing agents, denaturation is not accompanied by precipitation.

Denaturation can be triggered by increasing the temperature or by applying extreme pH values. If the target protein is unusually resistant against certain denaturing effects (e.g. it is thermostable, acid-stable etc.), this property can be a good starting point for establishing an effective fractionation protocol. After high-temperature or extreme-pH treatment, many proteins unfold and precipitate, while the target protein (and, usually, several other resistant proteins) remains folded and stays in the solution. After a centrifugation step, the target protein will be in the supernatant and can be easily separated from the pellet of denatured ‘contaminating’ proteins.

## 5.4.2. Protein fractionation based on particle size

Several low-resolution fractionation methods separate molecules from each other based on their size. The two methods reviewed below apply semipermeable membranes to separate the sample in two fractions by size. Semipermeable membranes are special ‘sieves’ having pores with diameters in the size range of molecules. Molecules that are smaller than the pores of the membrane can permeate the membranes while molecules larger than the pores are retained.

In the course of *dialysis* (see also in Chapter 2), the sample is poured in a tube made of a semipermeable membrane. The two ends of the tube are then sealed, and the resulting sachet is immersed into a container containing an appropriate buffer. All molecules that are smaller than the pores of the membrane start to diffuse across the membrane in both directions. This process leads to an equilibrium. At equilibrium, the concentration of the small molecules on the two sides of the membrane becomes identical, while large molecules remain in the dialysis sachet. The time required for reaching the equilibrium state can be shortened by stirring the buffer in the container. Dialysis is usually applied in order to remove small molecules from the sample and/or to replace the buffer around the retained large molecules (usually proteins) in the sample. The equilibrium concentration of the small molecules can be calculated based on their original concentrations inside and outside the sachet and the ratio of the volumes of the sachet and the container. The larger the container, the lower the concentration of the small molecules to be removed will be in the sachet. However, instead of increasing the volume of the container to an impractical volume, a better strategy is to apply a relatively small container and change the buffer in the container several times. This way, the concentration of the small molecule in the sample will decrease exponentially. At the end of the previously mentioned salting-out experiment, the protein sample will contain high concentration of ammonium and sulphate ions. These ions are typically removed by dialysis.

Semipermeable membranes are applied in the course of *ultrafiltration*, too. However, in this case the sample solution is forced to flow through the membrane. This comes with two advantages. First, the process is faster than dialysis. The second, more important difference is that by applying an external driving force, the method does not rely on equilibrium. This way, the concentration of the molecules that are retained by the membrane can be increased tremendously. This method is often applied to concentrate protein samples. Furthermore, the solvent of the large molecules can also be easily changed by this method by adding the appropriate buffer to the sample between several consecutive ultrafiltration steps. The external force can be generated either by centrifugation of the sample in Centricon vials, or by a high-pressure inert gas (usually nitrogen) in Amicon systems.

## 5.5. Lyophilisation (freeze-drying)

Protein (or any other non-volatile molecule) samples can be concentrated by evaporating water and other volatile compounds from the sample. In principle, this could be achieved by simply heating the sample. However, most proteins would unfold during such a simple evaporation process. In order to prevent the denaturation of proteins, the sample is transferred into a glass container and is frozen as quickly as possible, usually by immersing the outside of the container into liquid nitrogen. Moreover, the container is rotated in order to spread and freeze the sample on a large surface area. The glass container with the sample is then placed into an extremely low-pressure space (vacuum) that contains a cooling coil as well. The cooling coil acts as a condenser. The temperature of the coil is usually lower than -50°C. Volatile compounds of the frozen sample will evaporate (sublimate) in the vacuum. The process of evaporation (in this case, sublimation) absorbs heat. This effect keeps the sample frozen. Evaporated molecules are captured from the gas phase by the cooling coil, forming a frozen layer on it. At the end of the process, proteins and other non-volatile compounds of the sample remain in the container in a solid form. This process does not cause irreversible denaturation of proteins. Thus, it is a method frequently used not only to concentrate but also to preserve proteins or other sensitive biomolecules for long-term storage. Such samples can usually be stored for years without a significant deterioration of quality. However, before lyophilisation it is very important to carefully consider all non-volatile compounds of the initial sample as these will concentrate along with the proteins. Non-volatile acids or bases can cause extreme pH, and the presence of salts can result in very high ionic strength when the sample is resolubilised.

# Chapter 6. Chromatographic methods

by György Hegyi and Mihály Kovács

Chromatography is the collective term for a set of separation techniques that operate based on the differential partitioning of mixture components between a mobile and a stationary phase. The mobile phase (a liquid or a gas) travels through the stationary phase (a liquid or a solid) in a defined direction. The distribution of components between the two phases depends on adsorption, ionic interactions, diffusion, solubility or, in the case of affinity chromatography, specific interactions. Depending on the experimental design, the separation in a liquid mobile phase may be carried out via column or planar chromatography, on analytical or preparative scales.

Chromatographic methods are important in the analytical and preparative separation of biological samples. Gel filtration chromatography (size exclusion chromatography) is often the method of choice to purify macromolecules, taking advantage of their different sizes and shapes. Ion exchange chromatography is also useful for the separation of macromolecules, operating based on the various net charges on their surface, which can be tuned via the pH of the medium. Biological specificity in enzyme-substrate, enzyme-inhibitor, receptor-ligand, antigen-antibody (and other) interactions is utilised in affinity chromatography. In this method, one interaction partner is immobilised on a solid surface (stationary phase) and can selectively bind its interacting partner from a mixture in the mobile phase. The other components of the mixture can then be removed by replacing the mobile phase (washing). The pure material is then eluted by applying a mobile phase that disrupts the specific interaction.

Of the quasi-infinite possibilities of analytical applications of liquid chromatography, only a few relevant ones will be mentioned in this chapter. The analysis of amino acid mixtures and that of the products of Edman degradation in the process of amino acid sequencing are both carried out using chromatography. Chromatographic methods coupled to on-line mass spectrometry are instrumental in current proteomics research.

## Quantification of separation

### (1) Resolution ( $R_S$ )

Resolution is a number describing the separation of chromatographic peaks. By definition, resolution is the distance between peak maxima (elution volumes) divided by the average of peak widths (Figure 6.1). The elution volumes and the peak widths are to be measured in identical units.  $R_S$  is therefore a dimensionless number. In the case of a constant flow rate, the quantification of volumes can be replaced with the more convenient measurement of time.

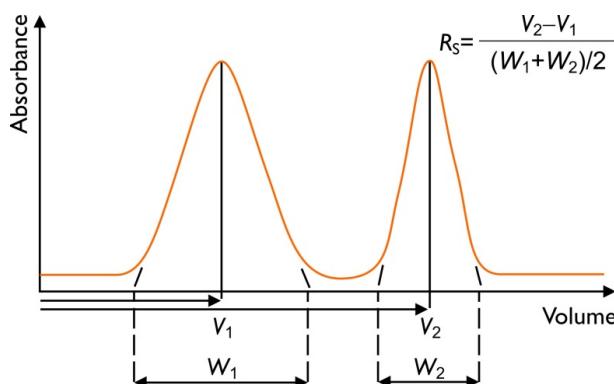


Figure 6.1. Determination of the resolution.  $R_S$  = resolution;  $V_1$  = elution volume at the maximum of the first peak;  $W_1$  = base width of the first peak;  $V_2$  = elution volume at the maximum of the second peak;  $W_2$  = base width of the second peak. In the case of a constant flow rate, it is more convenient to measure time instead of elution volumes. (In this case, time will be directly proportional to the elution volume.) Thus,  $R_S = 2(t_2 - t_1) / (W_2 + W_1)$ , where  $t_2$  and  $t_1$  are elution times corresponding to  $V_2$  and  $V_1$ , respectively.

The  $R_S$  value defines the extent of separation. The larger the  $R_S$  between two peaks, the more ideal the separation (Figure 6.2). (Note that even perfectly separated peaks may comprise impure materials. In many cases, two or more components may co-elute under a given set of chromatographic conditions.)

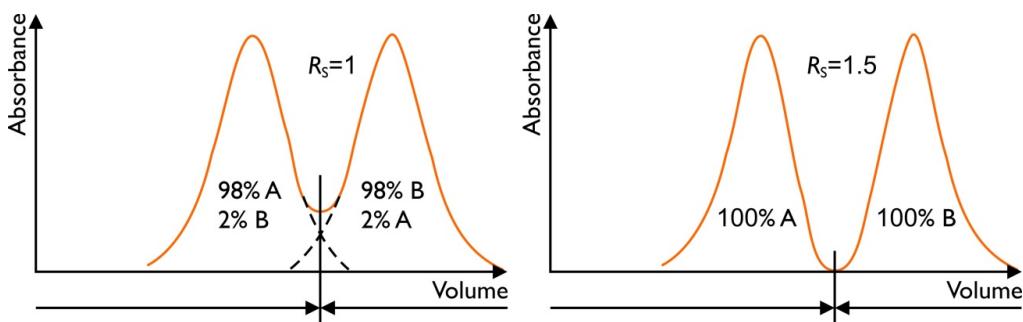


Figure 6.2. Relation between resolution and separation. It can be proven by calculation that, if  $R_S = 1$ , and both peaks have ideal shapes (i.e. Gauss curves) and identical sizes, then the two components can be isolated at 98 % purity. Perfect, so-called base-line separation can be achieved in cases where  $R_S > 1.5$ .

The chromatographic separation, the behaviour of the different peaks, and the efficiency of the chromatographic column can be described using the following parameters.

### (2) Retention

The retention factor or capacity factor,  $k'_n$ , is defined for the extent of retention of a compound under a given set conditions. The retention factor can be calculated for each peak.

$$k'_n = V_n - V_0 / V_0 \text{ or } k'_n = t_R - t_0 / t_0 \quad (6.1)$$

where  $V_n$  is the elution volume of component  $n$ ,  $V_0$  is the elution volume of a component proceeding through the column without any interaction, and  $t_R$  and  $t_0$  are the corresponding times, respectively (Figure 6.3).

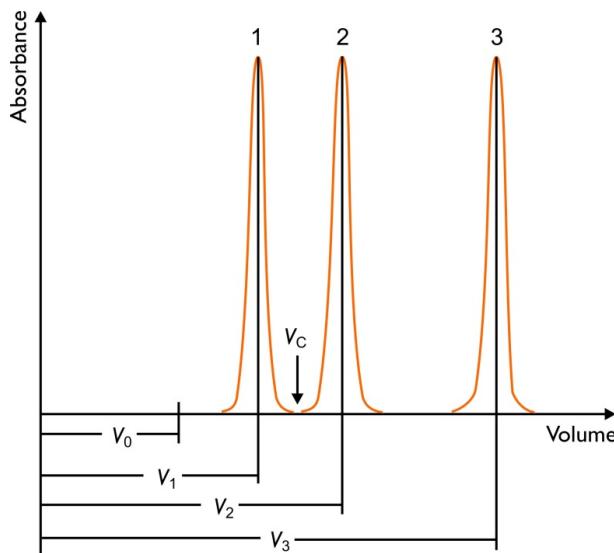


Figure 6.3. Idealistic chromatogram to demonstrate the calculation of the retention factor.  $V_0$  = elution volume of a material passing through the column without any interaction;  $V_C$  = total volume of the column;  $V_1-V_3$  = elution volumes corresponding to individual peaks. In the case of gel filtration,  $V_0$  equals the exclusion volume of the column, and separation takes place in the range between  $V_0$  and  $V_C$ . In other cases where the materials of interest bind to the stationary phase (based on ion exchange, adsorption, affinity etc.), the elution volumes can exceed the total volume of the column ( $V_C$ ) by several times.

The retention factor is characteristic of a component in a given composition of the mobile phase, in a given type of column, and at a given temperature. It is independent of the size of the column and of the flow rate of the mobile phase.

### (3) Efficiency ( $N$ )

The efficiency, N, or in other words, the number of theoretical plates characterises the spreading of the eluted compound. It can be calculated as follows:

$$N = (V_e/\sigma)^2 = 16(V_e/w)^2 = 5,54^*(V_e/w_{1/2})^2 \quad (6.2)$$

or

$$N = (t_R/\sigma)^2 = 16(t_R/w)^2 = 5.54^*(t_R/w_{1/2})^2 \quad (6.3)$$

where  $V_e$  and  $t_R$  are the elution volume and the retention time of the peak, respectively;  $\sigma$  is band broadening (to be determined by measuring the peak width,  $W = 4\sigma$ );  $W$  is peak width measured at the baseline;  $W_{1/2}$  is peak width measured at 50 % of the maximal peak depth. The sign \* marks a coefficient calculated from the ratio of  $W$  and  $W_{1/2}$  if the peak follows a Gaussian distribution.

The value of N is used to characterise the chromatographic column. As N largely depends on the experimental conditions including the flow rate of the mobile phase, the quantity and quality of the loaded sample, the determination of N is carried out in a standardised manner.

The main reason behind the spreading of a compound and the broadening of chromatographic bands is the longitudinal diffusion of molecules within the column. The effect of diffusion may be decreased by using smaller chromatographic beads and by enhancing the homogeneity of their size distribution.

Besides the particle size, efficiency is significantly influenced by the method of packing of the chromatographic media, especially when the column is home-made. Any inhomogeneity in the sedimented matrix (e.g. due to air bubbles or tunnels) will result in poorer efficiency and, in consequence, imperfect separation.

#### (4) Selectivity ( $\alpha$ )

The selectivity ( $\alpha$ ) for two neighbouring elution peaks at  $V_1$  and  $V_2$  is characterised by the quotient of their retention factors:

$$\alpha = k'_2/k'_1 \quad (6.4)$$

The overall quality of the chromatographic separation is influenced by both selectivity and efficiency. Nevertheless, selectivity is more definitive (Figure 6.4).

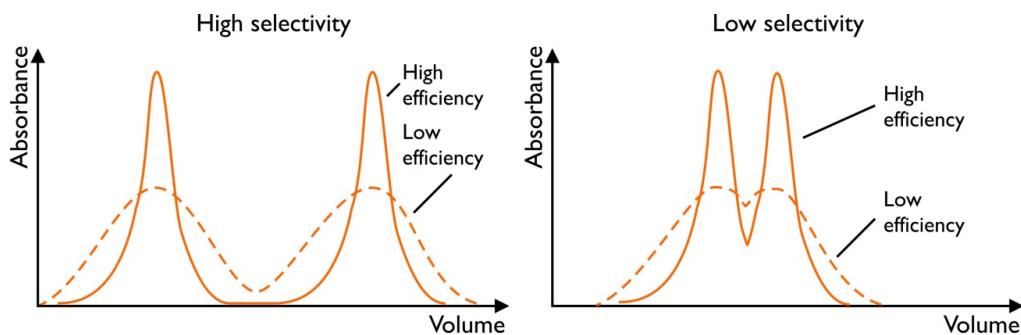


Figure 6.4. The effect of efficiency and selectivity on separation. The efficiency of a chromatographic column (N) is the qualitative property that reflects how thin and symmetrical the peaks are during elution. The separation of two materials is determined by the extent of selectivity. At high selectivity, even a smaller efficiency will provide sufficient separation. In an ideal case, a high selectivity is accompanied by a high efficiency.

## 6.1. Gel filtration chromatography

The chromatographic medium for gel filtration is a hydrophilic gel made up from porous, fine-grain spheres of 10-300 µm diameter. This type of medium defines two solution compartments within the column: one is the freely moving mobile phase outside the gel particles, while the other is the restricted liquid compartment inside the porous particles (Figure 6.5).

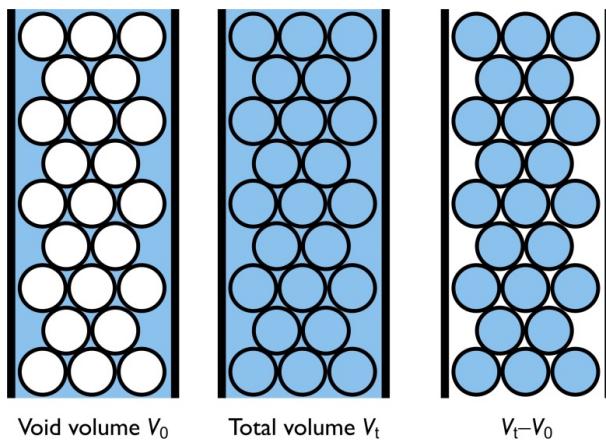


Figure 6.5. Liquid compartments inside a column packed with a porous gel. The individual liquid compartments are depicted as blue areas.  $V_0$  = exclusion volume (void volume);  $V_t$  = total volume of the column;  $V_t - V_0$  = combined volume of the liquid inside the gel particles and the material of the gel.

When a solution is moving through the gel filtration column, the movement of the solutes depends on two factors: the flow rate of the mobile phase and diffusion. Diffusion enables the molecules to explore the inside of the gel particles if their size so permits. The separation of a molecular mixture is based on the phenomenon that some molecules are excluded from the inside of the gel particles due to their size. These molecules travel quickly in the mobile phase of the column, which is the only compartment available to them. Smaller molecules, on the other hand, spend various amounts of time inside the particles (stationary phase) and flow through the column slower (Figure 6.6).

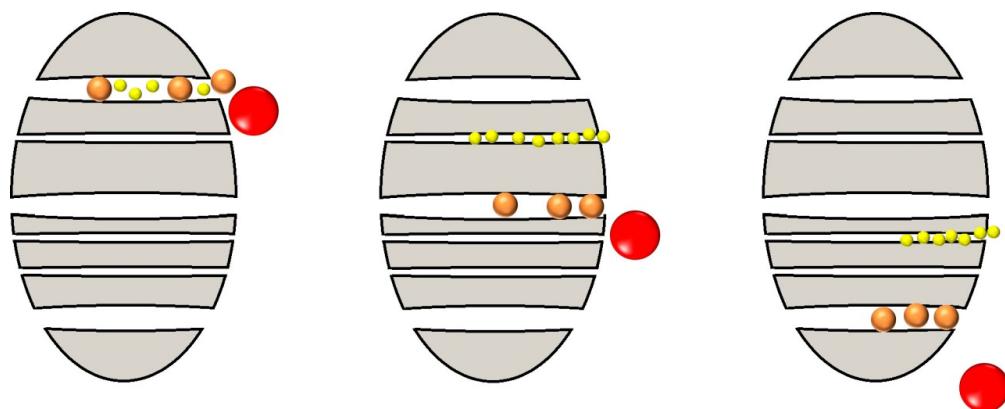


Figure 6.6. Travel of variably-sized molecules through a porous gel. During gel filtration (size exclusion) chromatography, molecules of different sizes will explore the available liquid spaces via diffusion. The largest molecules (red in the figure), due to their size, cannot enter the pores of the matrix at all. The movement of molecules with medium sizes (orange) is confined to the larger pores. The smallest molecules (yellow) can enter the gel particles through all pores. Therefore, the largest molecules will advance most rapidly through the column, whereas the smaller molecules will be retarded.

The result of a gel filtration experiment is usually depicted as an elution diagram. In this diagram, the concentration of the eluted compound is plotted against the volume of the eluent. The appearance of a given compound occurs at its elution volume ( $V_e$ ). As in other distribution chromatographic methods, the elution of a compound is best characterised by its distribution coefficient ( $K_d$ ):

$$K_d = (V_e - V_o)/V_s \quad (6.5)$$

where  $V_o$  equals the exclusion volume, i.e. the elution volume of a molecule that is larger than the largest pore size of the separating gel. Such a molecule therefore explores only the mobile phase, and is entirely excluded from the gel.  $V_s$  equals the volume of the stationary phase, i.e. the volume of the liquid inside the gel particles that is fully accessible only to molecules small enough to travel smoothly even through the smallest pores of the gel.  $V_s$

itself is difficult to determine. Therefore, in practice, it is replaced by the  $V_t - V_o$  term, also accounting for the non-negligible volume of the gel itself. As a result, a constant pertinent to an apparent volume ( $K_{av}$ ) is used instead of  $K_d$  (the latter would be valid only for real liquid volumes):

$$K_{av} = (V_e - V_o) / (V_t - V_o) \quad (6.6)$$

where  $K_{av}$  represents the portion of the gel volume that is accessible to a molecule of a given size. For a totally excluded macromolecule,  $K_{av} = 0$ ; whereas, for small molecules diffusing freely in the entire volume of the gel,  $K_{av} = 1$ .

### Planning a gel filtration experiment

#### (1) Choosing the gel type

Several different gel filtration media are available, which should be chosen according to the substance to be separated. These media differ in the chemical properties of the gel matrix, the pore size, the particle size, as well as the physical and chemical stability of the gel. The first developed and still widely used gel matrix is made of crosslinked dextran. Polymer beads made of dextran are known by the trade name Sephadex.

The pore size of the various Sephadex media is controlled by the number of crosslinks. The most popular ones are the entirely hydrophilic G-type gels. Numbers accompanying the G-type mark refer to the pore size and indicate the approximate molecular mass of excluded molecules in kDa. For example, Sephadex G-25 is used to separate relatively small molecules in the molecular mass range of 1000-5000 Da, including peptides. Alternatively, it can also be used to desalt larger proteins. To fractionate larger macromolecules up to 200-300 kDa, the G-150 or G-200 Sephadex gels are to be used. The mechanical properties of dextran gels having large pores are unfavourable due to the low density of crosslinks. These gels are easily compressible. Therefore, more rigid gels made of synthetic polymers are used to separate very large or elongated molecules.

If the size difference between the compounds to be separated is relatively large, e.g. during desalting of a macromolecule, it is practical to choose a gel in which the large-sized compound is eluted in the excluded volume ( $V_o$ ; thus  $K_{av} = 0$ ), while the small component elutes around  $V_t$  (thus  $K_{av} = 1$ ). In this case, the fraction containing the macromolecules appears sharply, with minimal band broadening and dilution, in the shortest possible elution time.

In case of fractionating macromolecules and if the molecular weight of the compound of interest is known, the gel should be chosen so that the component of interest will elute approximately at the half of the entire fractionation range. For example, if a 100-kDa protein is to be isolated from a protein mixture, the use of a gel that spans the 10-250 kDa fractionation range is recommended.

#### (2) Choosing the particle size of the gel

Fine-sized beads fill the available space within the chromatographic column more efficiently. Therefore, the volume of the mobile phase will be reduced. This will result in a similar reduction in dilution and band broadening and, in turn, will yield a better resolution. On the other hand, the flow rate in a compact gel column is also reduced. Therefore, larger pressure should be applied when using super-fine beads. Indeed, special pumps are needed below a particle size of 10 µm. Naturally, only rigid, non-compressible gel types can be used in these cases.

For most purposes, the Fine and Medium type particle sizes (20-150 µm) are suitable. For preparative purposes and desalting, where high flow rates are required and the compounds of interest separate well even at a poor resolution, Coarse type gels can be used too.

#### (3) Choosing the size of the column

During gel filtration, the distance between two zones of separation increases proportionally to the square root of the column length. Long columns (> 100 cm) are used when a high resolution is required, while shorter (< 50 cm) columns are more practical when the aim is desalting or the separation of compounds that can be eluted at markedly different volumes.

Columns with diameters of around 1 cm are used for analytical purposes. By increasing the diameter, the amount of the applied sample, i.e. the capacity of the column, can be increased.

#### (4) Choosing the sample volume

A narrow start zone (relative to the column length) is sought if maximal resolution is to be achieved, e.g. for analytical purposes or in the case of compounds whose separation is difficult. Therefore, the sample volume in this case should be chosen to be 1-5 % of the column volume. The resolution cannot be further increased using smaller sample volumes, while the dilution will be greater. The sample volume can be increased to as much as 15-20 % of the column volume if the compounds are readily separable, especially when working on a large scale.

#### (5) Choosing the eluent

The composition of the eluent does not directly influence the resolution of gel filtration. However, all components that have an effect on the molecules to be separated may influence the separation. The pH, ionic strength or the presence of detergents may influence the molecular state of the solutes. For instance, changes in molecular shape or the dissociation of multimeric proteins and enzyme-inhibitor complexes will change their chromatographic behaviour. In general, dilute (0.01-0.1 M) buffers are used that do not influence the structure of the compounds to be separated, but restrict the unwanted adsorption interactions between the gel matrix and the molecules of interest.

When the fractions containing the separated compounds are to be later concentrated, volatile buffers (e.g. ammonium bicarbonate) are practical to use that easily disappear during lyophilisation or film evaporation. The same considerations apply when the salt content of the gel filtration buffer should be subsequently eliminated.

#### (6) Choosing the flow rate of the eluent

During gel filtration, increasing the flow rate will deteriorate the resolution, because it prevents the formation of equilibrium between the mobile and the stationary phases. Generally, 5-10 mL/cm<sup>2</sup> x hour is recommended as an optimal flow rate, but in most cases, a several times excess of this will not deteriorate the separation significantly. When doing preparative work, or if the operation must be performed quickly for some reason, the advantage conferred by the higher flow rate may compensate for the deterioration of separation.

To achieve a higher flow rate, of course, a larger pressure must be applied. Therefore, in these cases, the mechanical stability of the gel matrix must be taken into consideration. Non-rigid gels may be compressed at pressures higher than allowed, which may lead to the complete clogging of the chromatographic column.

## 6.2. Ion exchange chromatography

Ion exchange chromatography is one of the most efficient methods for the separation of charged particles.

Ion exchange chromatography is most often performed in the form of column chromatography. However, there are also thin-layer chromatographic methods that work basically based on the principle of ion exchange. In the following, we will exclusively deal with column chromatographic applications.

Column materials used for ion exchange chromatography contain charged groups covalently linked to the surface of an insoluble matrix. When suspended in an aqueous solution, the charged groups of the matrix will be surrounded by ions of the opposite charge. In this "ion cloud", ions can be reversibly exchanged without changing the nature and the properties of the matrix.

The charged groups of the matrix can be positively or negatively charged. A positively charged matrix will bind negatively charged ions from the solution. Therefore, it is called an anion exchanger. Cation exchanger matrices have negative charges.

Based on the structure of the ion exchange matrix, we distinguish ion exchangers with hydrophobic and hydrophilic matrices. Ion exchangers with a hydrophobic matrix are most often highly substituted polystyrene resins. These are suitable for the binding of inorganic ions, e.g. in water softening applications. However, they tend to denature proteins due to the high hydrophobicity of their matrix and their high surface charge density.

Ion exchangers with hydrophilic matrices were first produced from modified cellulose. However, cellulose has disadvantageous mechanical properties: cellulose fibres are prone to break, making it difficult to create a well-utilisable column. This disadvantage has been partially remedied in Sephadex (dextran-based) ion exchange matrices.

In recent years, regular spherical and monodisperse matrices have been produced from synthetic hydrophilic polymers. The best known of such resins is the MonoBead-based ion exchange matrix.

Table 6.I summarises the charged groups linked to ion exchange matrices.

| Anion exchangers            | Functional group   |
|-----------------------------|--|
| diethyl-aminoethyl (DEAE)   | -OCH <sub>2</sub> CH <sub>2</sub> N <sup>+</sup> H(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>                                    |
| quaternary aminoethyl (QAE) | -OCH <sub>2</sub> CH <sub>2</sub> N <sup>+</sup> (C <sub>2</sub> H <sub>5</sub> ) <sub>2</sub> CH <sub>2</sub> CH(OH)CH <sub>3</sub> |
| quaternary ammonium (Q)     | -CH <sub>2</sub> N <sup>+</sup> (CH <sub>3</sub> ) <sub>3</sub>  |
| Cation exchangers           | Functional group   |
| carboxymethyl (CM)          | -OCH <sub>2</sub> COO <sup>-</sup>   |
| sulfopropyl (SP)            | -CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> SO <sub>3</sub> <sup>-</sup>  |
| methylsulfonate (S)         | -CH <sub>2</sub> SO <sub>3</sub> <sup>-</sup>  |

Table 6.I. Functional groups of ion exchangers

Ion exchangers containing sulfonyl and quaternary ammonium groups are called strong ion exchangers. These are practically completely charged between pH 3.0 and 11.0. The degree of dissociation of DEAE and CM groups—and thus their ion exchange capacity—depends on the pH of the medium.

### The theory of ion exchange

Most ion exchange experiments comprise five different phases (Figure 6.7).

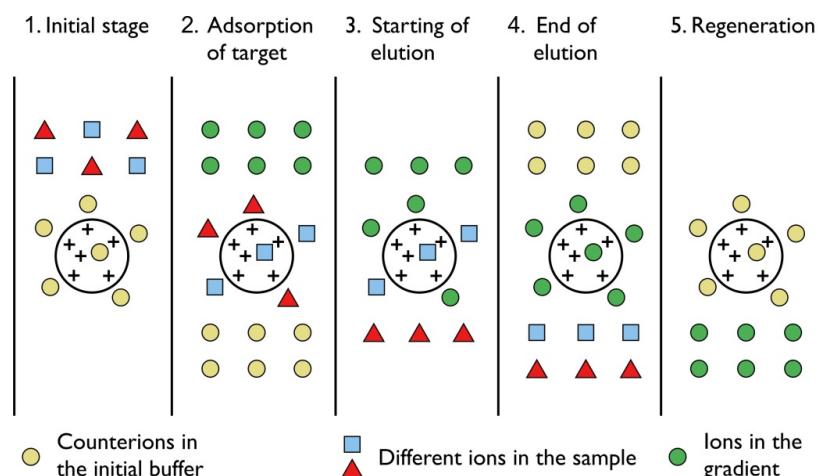


Figure 6.7. Phases of ion exchange chromatography (salt gradient elution). In the figure, a positively charged anion exchanger particle is shown, with counterions on its surface in the starting state (first phase). In the second phase, the binding of the ions to be separated takes place. At the start of the elution (third phase), weaker-binding ions are desorbed. At the end of the desorption, the stronger-binding ions are also desorbed (fourth phase). During regeneration (fifth phase), the starting state can be reconstituted via washing the column with the starting buffer.

The first phase is the equilibration of the ion exchange column with a so-called starting buffer, setting the conditions of the experiment (pH and ionic strength). In this phase, the charged groups of the ion exchanger will bind easily replaceable simple ions (e.g. chloride or sodium).

The second phase is the loading of the sample and its reversible binding to the column. If some of the contaminating materials do not bind to the column, these can be removed via washing the column with the starting buffer.

The third and fourth phases comprise the elution, i.e. the desorption of the bound molecules, which can be achieved via changing the composition of the elution buffer. The simplest form of elution is achieved via an increase in ionic strength, i.e. in the concentration of the counterions present. Another means of desorption is the change of the pH of the medium. The most effective method is the continuous change of the ionic strength or the pH, i.e. the

application of a gradient elution. During gradient elution, molecules with smaller net charges (i.e. the weaker-binding ones) will be the first to leave the column.

An important property of an ion exchange column medium is its *ion exchange capacity*. This parameter reflects the amount of counterions that can be bound to the column. Three types of capacity can be distinguished:

*Total capacity*: The number of charged groups per gram dry weight of the ion exchanger or per millilitre of swollen gel. This can be determined by titration with a strong acid or base.

*Free capacity*: Due to steric reasons, only a part of the full capacity is accessible for macromolecules (proteins, nucleic acids). This is the free capacity.

*Dynamic capacity*: The so-called dynamic capacity is determined when the binding of the given macromolecule to the column is measured during buffer flow.

The free and the dynamic capacity values are dependent on the properties of the material to be separated, the properties of the ion exchanger, and the applied experimental conditions.

With regard to separation, the important properties of the material to be separated are the size of the molecules and the pH dependence of their charge. This implies that the capacity of ion exchangers will be different for different proteins.

### Frequently used Sephadex-based ion exchangers

Via the modification of the Sephadex G-25 and G-50 gel filtration matrices with four different groups, eight different ion exchangers were created (Table 6.II). Matrices derived from G-25 (A-25 and C-25) contain more crosslinks. Therefore, they are more rigid and swell to a smaller extent than the ones derived from G-50.

| <b>Ion exchanger</b> | <b>Total capacity</b> | <b>Total capacity</b> | <b>Functional group</b>            | <b>Counterion</b> |
|----------------------|-----------------------|-----------------------|------------------------------------|-------------------|
|                      | <b>µmol/mg</b>        | <b>µmol/mL</b>        |                                    |                   |
| DEAE Sephadex A-25   | 3.5 ± 0.5             | 500                   | diethyl-aminoethyl                 | chloride          |
| DEAE Sephadex A-50   |                       | 175                   | diethyl-aminoethyl                 | chloride          |
| QAE Sephadex A-25    | 3.0 ± 0.4             | 500                   | diethyl-(2hydroxypropyl)aminoethyl | chloride          |
| QAE Sephadex A-50    |                       | 100                   | diethyl-(2hydroxypropyl)aminoethyl | chloride          |
| CM Sephadex C-25     | 4.5 ± 0.5             | 550                   | carboxymethyl                      | sodium            |
| CM Sephadex C-50     |                       | 170                   | carboxymethyl                      | sodium            |
| SP Sephadex C-25     | 2.3 ± 0.3             | 300                   | sulfopropyl                        | sodium            |
| SP Sephadex C-50     |                       | 90                    | sulfopropyl                        | sodium            |

Table 6.II. Properties of Sephadex-based ion exchangers

With the use of specially-treated polysaccharide and synthetic polymer matrices that are more pressure-resistant than Sephadex, it was possible to develop fine-grained regular spherical polymer beads for ion exchangers with significantly increased efficiency. Such matrices include monodisperse MonoBeads with a bead size of 10 µm, or MiniBeads with a 3-µm bead size. Such ion exchangers are used in FPLC techniques (see below).

### Important parameters to consider during the planning of ion exchange-based separation

#### 1. The charge of components present in the sample to be separated

In the case of proteins, this will depend on the isoelectric point. Below the isoelectric point, proteins are cations; above the isoelectric point, they are anions. The applied chromatographic buffer should be chosen in a way that the proteins to be separated should bind to the given anion or cation exchange column, from which they can be

eluted after the washout of the non-binding components. More rarely, it can occur that the unwanted contaminants are bound to the column, whereas the component to be isolated will freely flow through.

## 2. *The amount of the sample*

The size of the column should be chosen so that the dynamic capacity of the medium should somewhat exceed the amount of the sample. If the column is too small, the sample will saturate it, and part of the sample will not bind to the column. If the column is too large, a significant loss can occur during elution.

## 3. *The molecular mass of the protein to be isolated*

The pore size of the ion exchange matrix should be chosen in a way that gel filtration effects—i.e. size-based separation—do not occur during ion exchange.

# 6.3. Hydrophobic interaction chromatography

The method of hydrophobic interaction chromatography (HIC) is based on the observation that protein molecules can interact with fully hydrophobic adsorbents, and this interaction is dependent on the salt concentration of the solution. Similarly to the salting-out of proteins (see in details in Chapter 5) where the increasing salt concentration will lead to the aggregation and precipitation of protein molecules via the rearrangement of their hydrate shell, during HIC chromatography a high salt concentration (1-1.5 M neutral salt) facilitates the interaction between the hydrophobic chromatographic medium and the hydrophobic patches present on protein molecules. During separation, the decrease in salt concentration will lead to the elution of bound molecules.

Based on these observations, bead polymers were created that are suitable for HIC chromatographic purposes. The surface of the beads was modified by hydrophobic alkyl or aryl groups. Such media include the polysaccharide-based Butyl-Sepharose, Octyl-Sepharose and Phenyl-Sepharose materials that are derived from polymers that had proven to be suitable for chromatographic separation of proteins.

Here we draw attention to the fact that the so-called reverse-phase chromatography (RPC, see below) is in principle very similar to HIC chromatography. In both cases, the separation is based on the strength of interaction forming between hydrophobic surfaces. The main difference is that, in the case of media used in HIC chromatography, the concentration of the hydrophobic ligand bound to the solid matrix is 10-20 µmol/mL of column volume, whereas in RPC, the hydrophobic ligand concentration on the matrix surface is several 100 µmol/mL. Therefore, in the case of reverse-phase chromatography, the binding between the adsorbent and the molecules to be separated is very strong. Thus, for RPC elution, it is necessary to use solvents that are less polar than water (methanol, acetonitrile, etc.). Therefore, the isolation of native proteins is not always possible with RPC—the proteins can denature on the column as the amino acids forming their hydrophobic core can also strongly bind to the hydrophobic matrix. However, due to its high resolution, the RPC technique is very advantageous for the qualitative analysis of complex protein or peptide mixtures. During HIC, we can work in an aqueous, non-denaturing medium throughout the chromatographic procedure, from loading to elution.

## Factors affecting HIC chromatography

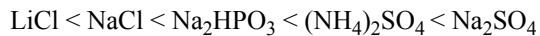
### 1. *The type and density of the hydrophobic ligand on the matrix surface*

The immobilised hydrophobic ligands will primarily determine the selectivity of the HIC adsorbent. Alkyl chains show purely hydrophobic interactions. In the case of aromatic ligands, beyond the hydrophobic effect, specific interactions between aromatic groups are also present. The choice between alkyl-containing versus aromatic-ligated matrices is mainly empirical and can be made based on preliminary binding experiments.

In the case of alkyl matrices, the protein-binding capacity of the HIC adsorbent will increase with the length of alkyl chains. In addition to alkyl chain length, the binding capacity will, of course, also depend on the concentration of these chains on the matrix surface. In the case of HIC, the optimal hydrophobic ligand concentration is around 20 µmol/mL of medium. The behaviour of the medium is, to some extent, also affected by the hydrophobicity of the polymer matrix. For instance, even in the case of the same hydrophobic ligands, the selectivity will differ in the case of agarose and synthetic polymer matrices.

### 2. *The quality of the salt used and its concentration*

The effect of salts used in HIC is similar to their salting-out efficiency. Both effects are associated with the effect of the given salt on the surface tension of water, which can be arranged in the following order:



Most often,  $(\text{NH}_4)_2\text{SO}_4$  is used, as the salting-out efficiency of this salt is about four times that of NaCl.

The initial salt concentration should be selected so that the protein to be isolated should bind to the column with sufficient efficiency. The determination of the salt concentration is empirical—a good approach may be testing the use of a 1-M  $(\text{NH}_4)_2\text{SO}_4$  solution.

We must also ensure that, at the applied initial salt concentration, protein precipitation does not occur due to salting-out.

### 3. The effect of pH

The pH of the medium changes the charge of ionisable groups of protein molecules. This effect will obviously affect the separation based on hydrophobic interactions.

It is found that, in general, the retention of protein samples changes dramatically below pH 5 and above pH 8.5. Thus, the states close to the zwitterionic state are advantageous for HIC. Therefore, similarly to the salt concentration, the pH of the solvent should also be optimised.

### 4. The effect of temperature

The hydrophobic interaction is temperature dependent. Increasing temperature will increase the strength of the interaction in most cases, but this phenomenon is complex—the opposite effect has also been observed. In practical terms, it must be taken into account that the outcome of a method developed at room temperature may not be reproducible in the cold room (at 4°C).

### 5. The effect of additives

Water-miscible organic solvents (alcohols, acetonitrile, dimethyl formamide, etc.) or added detergents can reduce the binding of proteins to be separated, even when present at low concentrations. These compounds “compete” with the protein for the adsorption sites on the matrix surface.

Therefore, when added in low concentrations, these substances can increase the efficiency of the elution. They can be thus applied e.g. when the decrease in salt concentration alone does not lead to satisfactory results. One must ensure, however, that the additives used do not denature the protein to be isolated. Additives can be used even at high concentrations for the purification or regeneration of HIC columns.

## 6.4. Affinity chromatography

By affinity chromatography, high-selectivity separation of biomolecules can be achieved through their specific interactions. This separation technique is special because it is based on the biological function or the unique chemical structure of a given biomolecule. During affinity chromatography, the interacting partner of the biomolecule is immobilised on a chromatographic resin. This ligand, fixed to the stationary phase, reversibly binds the desired biomolecule present in the multi-component mobile phase. The materials can be eluted from the column by changing the composition of the mobile phase.

The technique provides high selectivity, high resolution and generally high capacity for the desired protein. The degree of purification can be thousands of times, and the achieved yield can also be usually very good.

Affinity chromatography, as already mentioned, is unique in the sense that it is based on the specific biological function of the biomolecule of interest. This feature also makes affinity chromatography suitable for the selective separation of active biomolecules, and their isolation from the inactive or denatured forms.

Another significant advantage of the method is that, in many cases, it allows for single-step isolation of the desired biomolecule. However, it is required that the sample to be separated should be a clear solution free of large particles.

It is often advisable to prepare the sample for affinity chromatography via an initial partial separation. For instance, in the case of affinity isolation of very scarce components of the blood serum, it is advisable to perform an initial separation to eliminate serum albumin (which makes up more than 50 % of the serum protein content).

Affinity chromatographic purification is frequently of great importance in the case of recombinant proteins. Recombinant proteins are often produced in a way that they contain a fused “label” at their N- or C-terminus, resulting from genetic engineering. This way, if the label endows the protein to enter into affinity binding, the recombinant protein can be simply “fished out” of the cell extract via affinity chromatography.

One of the most widely used of such labels fused to protein termini is the oligo-histidine tag (His-tag), which binds reversibly to metal chelates (e.g. Ni chelate immobilised on the stationary phase). Another frequently applied tag is glutathione S-transferase (GST), a fusion protein that can be used to isolate the protein of interest using a glutathione-conjugate matrix. These specific affinity matrices are commercially available as pre-packed columns.

In other cases, the specific ligand is to be linked by the user to the chromatographic matrix. Various activated reactive chromatographic matrices are available for this purpose.

Some of the commonly used interactions in affinity chromatography are listed in Table 6.III.

|               |  |
|---------------|--|
| Enzyme        | Substrate analogue or inhibitor                |
| Antibody      | Antigen (virus, cells)                         |
| Nucleic acid  | Complementary nucleic acid                     |
| Nucleic acid  | Histone or other nucleic acid binding protein  |
| Hormone       | Hormone receptor                               |
| Glutathione   | Glutathione S-transferase (GST) fusion protein |
| Metal chelate | His-tag fusion protein                         |

Table 6.III. Commonly used interactions in affinity chromatography

The phases of affinity chromatography are shown in Figure 6.8.

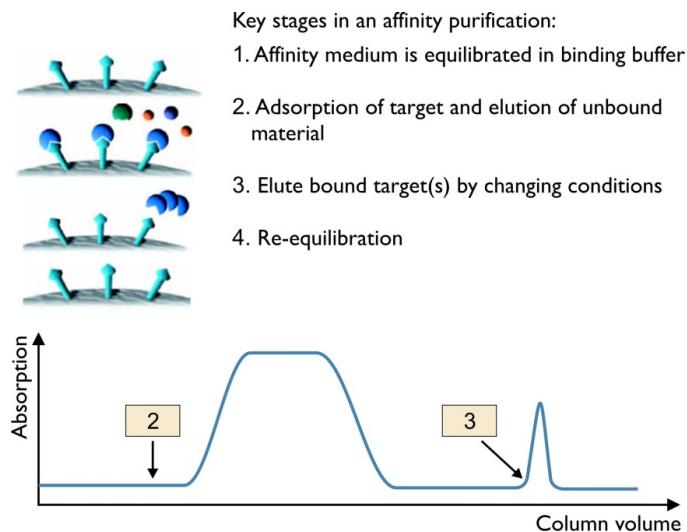


Figure 6.8. The phases of affinity chromatography. In the upper left side of the figure, an illustrative representation of affinity chromatography is shown. Of the molecules present in the sample, only the ones having a “matching” binding site can bind to the matrix-conjugated specific ligand molecules. Other molecules can be readily washed off. By changing the composition of the mobile phase, the molecules of interest can be isolated in a pure form.

The lower part of the figure shows a typical affinity chromatographic elution profile.

### 1. Sample preparation

The sample must be a clear solution free from solid particles. This can be achieved by centrifugation or filtration. Protein solutions should be centrifuged at at least 10000 g. Cell lysates should be centrifuged at 40-50000 g. A 0.45- $\mu\text{m}$  pore size filter can be used for filtration. (Such preparation of samples is also necessary in FPLC and HPLC methods.)

One must also consider how the solubility and stability of the sample or the desired protein can be influenced by the pH, the salt concentration, or the presence of any organic solvent. The factors affecting the interactions between the desired target protein and the matrix-bound ligand (pH, salt concentration, temperature) should also be determined. The composition of the initial binding buffer must be adjusted accordingly.

Sample components interfering with the target protein and/or the ligand (e.g. metabolites in cell lysates) should be removed before loading onto the column.

## 2. Equilibration with a buffer facilitating the specific interaction

The chromatographic column is washed with 3-4 column volumes of the starting (binding) buffer. The sample must also be equilibrated with this starting binding buffer (if necessary, via solvent exchange or dialysis).

## 3. Binding of the molecule of interest and wash-out of the unbound material

During sample loading, consider the strength of the interaction. In case of high-affinity samples, a high flow rate may be applied. In case of a weak interaction and/or a slow equilibration process, reduce the rate of sample loading. After sample application, the column should be further washed with binding buffer until all unbound components are removed.

## 4. Elution of the molecules of interest by changing the composition of the mobile phase

**Elution via pH and/or ionic strength changes:** One possible and simple means of elution is achieved through decreasing the interaction strength between the ligand and the target protein. Changes in the pH will change the ionisation state of charged groups of the ligand and/or the target protein, thereby changing the strength of the interaction. Similarly, increasing the ionic strength (usually by raising the NaCl concentration) will generally reduce the interaction strength. In either case, the solubility and stability of the target protein should be considered.

**Competitive elution:** For competitive elution, materials are applied that react with the target protein or the ligand, competing for the pre-existing interaction. For instance, His-Tag fusion proteins can be readily displaced from the metal chelate matrix by imidazole buffer (Figure 6.9). GST-tagged target proteins will detach from their column-conjugated glutathione ligand upon mixing excess glutathione into the elution buffer.

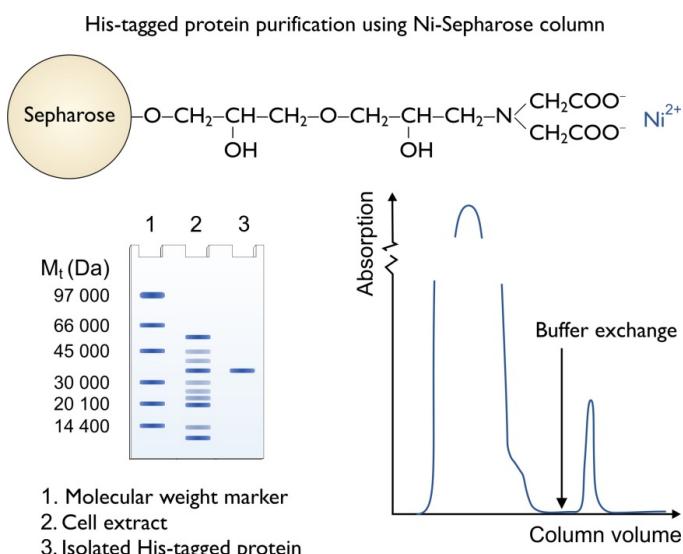


Figure 6.9. Isolation of a His-Tag fusion protein. The cell extract containing the His-Tag fusion protein is purified using a  $\text{Ni}^{2+}$  chelate column. The upper panel shows the structure of the immobilised  $\text{Ni}^{2+}$  chelate ligand. The sample is loaded onto the column in a neutral buffer. The fusion protein containing the His-Tag binds the  $\text{Ni}^{2+}$

chelate ligand. After washing off other proteins, competitive elution with imidazole buffer can be applied to isolate the pure fusion protein (lower right panel). Lower left panel shows the purity of fractions assessed by SDS gel electrophoresis, discussed in detail in Chapter 7.

In all cases, the flow rate of the buffer should be reduced during elution, thereby avoiding excessive dilution of the target protein.

In cases when the target-ligand interaction is very strong, the above elution methods may turn out insufficient for eluting the protein of interest. In these cases, chaotropic agents (urea, guanidine) can be used to wash off the target protein from the column. This naturally will involve the denaturation of the protein, which can then be renatured in some (lucky) cases under suitable conditions—in the case of urea or guanidine there is a good chance for this.

### 5. Regeneration

After successful completion of the elution, the column can be washed with several column volumes of binding buffer, and it can then be reused. For long-term storage, one must ensure that the column is not exposed to bacterial or fungal infection. The toxic compound sodium azide can be used to prevent such infections.

## 6.5. High performance (high pressure) liquid chromatography (HPLC)

In the discussion of gel filtration and ion exchange chromatography, we saw that the efficiency of chromatography increases with reducing the particle size of the gel matrix and with enhancing its size homogeneity. The efficiency reaches a new level of quality when a matrix grain size of 3–10 µm is applied. Liquid chromatography performed using such resins is called high performance liquid chromatography, abbreviated as HPLC.

However, at such particle sizes, the sufficient flow of the mobile phase (eluent) can be achieved only by applying a high pressure of around 10 MPa by using special precision pumps (Figure 6.10). HPLC thus also stands for high pressure LC—according to many researchers, the abbreviation also refers to the high price of the specialised equipment.

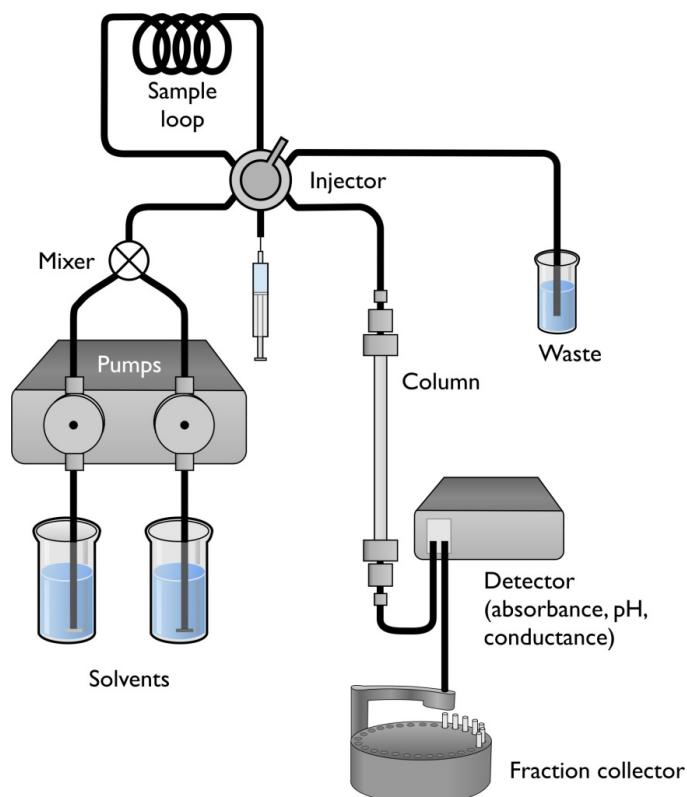


Figure 6.10. Schematic layout diagram of the HPLC equipment. The system consists of pumps that ensure the high-pressure delivery of a two-component mobile phase, a mixing unit, a sample injector, a chromatographic column and a detector.

Due to the application of high pressure, the primary requirement regarding HPLC columns is that they should be incompressible. Silica is predominantly used for this purpose. Under appropriate conditions, silica can be used to create homogeneous column media of sufficient strength and with a well-controlled particle size. For the hydrophilic silica stationary phase, only hydrophobic mobile phases can be applied. Therefore, HPLC was primarily suitable for the separation of hydrophobic organic solvent-soluble materials. Later, the chemical modification of the silica surface made possible the creation of hydrophobic silica gels. In this case, the hydrophilic-hydrophobic relation of the stationary and mobile phases became reversed, hence the term reverse-phase chromatography (RPC). Reverse-phase chromatography opened up the possibility of the separation of water-soluble substances, including the majority of molecules of biological origin. Large pore-size gels also allowed the separation of macromolecules. The hydrophobic surface is formed by long alkyl chains linked to the silica. These include octadecyl, octyl, butyl (labelled as C18, C8, C4) and also phenyl groups (see above at the description of HIC chromatography). Furthermore, gels containing charged groups can be used for ion exchange.

If the composition of the mobile phase is constant during chromatography, we speak of isocratic elution. Gradient elution is achieved via applying a linear or non-linear concentration gradient. In many cases, isocratic and gradient sections are combined in the elution profile. The gradient is most often created by using microprocessor-controlled, variable-speed pumps (two at least). In this case, the mobile phase components are mixed at the high-pressure side of the pumps.

With the help of precision valves, the gradient can also be made by mixing the buffers at low pressure. This way, the eluent can be transmitted onto the column by using only one pump. If unmodified silica is used, the mobile phase can be created by mixing organic solvents of different hydrophobicity.

In reverse-phase chromatography, a dilute aqueous solution (e.g. 0.1 % trifluoroacetic acid, few mM phosphate buffer) and a water-miscible organic solvent (e.g. acetonitrile, methanol, propanol) can be mixed to prepare a mobile phase for isocratic or gradient applications.

In the biochemical practice, light absorption and fluorescence detectors are most commonly used. However, refractive index, conductivity, optical rotation and electrochemical detectors are also frequently applied. For photometric detection, the eluent must be optically clear and must not absorb light in the applied wavelength range. The use of extremely pure solvents is generally required anyway in order to avoid both the contamination of the column media and the appearance of unexpected materials during chromatography.

A very important development is the appearance of mass spectrometric (MS) detectors. These are used to determine the mass of the separated components, which is a decisive parameter of a given substance. Actually, MS detectors do not determine the mass directly, but determine the mass/charge ( $m/z$ ) ratio. However, the mass can be easily derived for single- and multiple-charged values. In the past two decades, ionisation techniques have been introduced in MS measurement systems. Ionisation of high molecular weight biopolymers (such as proteins) can be efficiently achieved via ESI (electrospray ionisation) or MALDI (matrix-assisted laser desorption).

As a result, chromatographic methods play an increasingly important role in protein research. One- or multi-dimensional high performance liquid chromatography combined with in-line mass spectrometry allows the targeted identification of all proteins expressed in a cell at a given time, i.e. the “proteome”; or a specific subset of these proteins. This is significant because, unlike the genome, the proteome is not constant: it may vary by tissue or cell type and may also depend on the physiological state or developmental stage of the individual. The detection of post-translational modifications is also of importance. For instance, the specific detection of phosphorylated proteins in the proteome aids the understanding of various biological regulatory processes.

All components of the HPLC equipment must be pressure-resistant and chemically resistant. Thus, stainless steel and, in more special cases, resistant titanium alloys are applied. Moving parts, including pistons and valves, are also made of highly mechanoresistant materials (special ceramic, glass, industrial ruby etc.). Recently, the investigation of samples sensitive to trace metal contamination (e.g. some enzymes) has necessitated the application of particularly pressure-resistant plastic parts.

The FPLC (fast protein liquid chromatography) system was developed for the separation of native proteins. FPLC differs from the above-described HPLC chromatography in that the column resins are specially-treated dextran-based or synthetic polymeric materials (Superdex, Superose, Sephadex etc.), which are, due to their hydrophilic character and high porosity, particularly useful for the separation of biopolymers. These particles have a slightly larger size and lower pressure resistance than the silica-based HPLC media. However, FPLC media are suitable for ensuring sufficient fluid flow at pressures in the range of 0.5-1 MPa, and the efficiency of FPLC also meets most requirements of protein purification applications. FPLC columns are made of pressure-resistant borosilicate glass. Exposed metal parts are also avoided in the pumps and the piping systems.

Both HPLC and FPLC provide better efficiency and sensitivity as well as lower time requirement compared to conventional chromatographic applications.

### Some practical considerations

During sample preparation, one must take it seriously that the solution should be clear and free of dust or other particles. Otherwise, the apparatus may become blocked and the chromatographic column may become contaminated. The sample can be centrifuged at 20-40000 g and/or filtered through a 0.45- $\mu\text{m}$  filter. The sample is preferably dissolved in the starting mobile phase. It must be ensured that the sample is fully dissolved. The volume of the sample depends on the diameter of the column used (see below).

#### 1. Selection of the stationary phase (column medium)

In the case of nonpolar, water-insoluble materials, unmodified silica, or possibly diol media should be used. In the case of amino group-containing polar, water-soluble substances, reverse-phase C18, C8 or C4-modified silica media can be applied. Larger hydrophobic peptides and proteins bind too strongly to the C18 solid phase. In this case it is advisable to use C4 or C8 matrices.

For analytical purposes, smaller particle size matrices (around or less than 3  $\mu\text{m}$ ) should be used. This will increase the efficiency of separation, but has the disadvantage of increasing the pressure in the system. For semi-preparative and preparative purposes, 5-10- $\mu\text{m}$  particle sizes are suitable. One must also consider the porosity of matrix particles. In the case of low molecular weight metabolites, amino acids and small peptides, the commonly-used 100- $\text{\AA}$  pore-size media are suitable. In the case of macromolecules, high porosity (wide pore), 300- $\text{\AA}$  pore-size media are to be chosen (1  $\text{\AA}$  = 0.1 nm).

#### 2. Selection of the mobile phase

The most important characteristics of the mobile phase include purity, viscosity, UV transparency and miscibility with other solvents. HPLC techniques require special-purity ("HPLC grade") solvents, including water specially purified for this purpose. The selection of the correct column size is important in terms of the economic use of materials. High-viscosity solvents should be generally avoided as their use increases the system pressure.

In terms of detectability, it is important that the mobile phase should be optically pure. Given that the most commonly used chromatographic detectors operate in the UV range, the UV absorption of the mobile phase should be considered. The most commonly used HPLC reagents (e.g. acetonitrile) are available in various qualities. Highest-quality reagents enable photometric detection even at a wavelength of 200 nm. In the range of 280-340 nm, it is sufficient to use less expensive grades of acetonitrile.

In reverse-phase chromatography, the initial mobile phase is a dilute aqueous solution. The organic component used for the reduction of solvent polarity can be e.g. methanol, ethanol, propanol or acetonitrile. The initial aqueous solution can be, for instance, 0.1 % formic acid, acetic acid or few mM phosphate buffer. For ion pair formation, the commonly used agent is 0.1 % trifluoroacetic acid. Ion pair formation enhances the retention of highly charged molecules due to charge compensation. However, if chromatography is coupled to on-line MS measurements, trifluoroacetic acid should be avoided as it reduces the ionisability of molecules. In this case, the use of dilute formic acid or ammonium formate is recommended.

It must be taken into consideration that the mixing of the solvents during gradient elution will result in changes in the solubility of air in the mobile phase. This effect may result in air bubble formation when the solution leaves the column and the pressure is reduced. This will severely interfere with photometric detection. Mobile phase components must therefore be degassed prior to usage. Some chromatographic instruments contain a so-called

degasser unit, which applies a slight vacuum to keep the concentration of dissolved air continuously low. If no degasser unit is attached to the chromatographic instrument, mobile phase components should either be degassed by vacuum, or the very poorly water-soluble helium gas should be bubbled through the solutions to expel the dissolved air.

Before use, filtering of the mobile phase through a fine (0.45- $\mu\text{m}$ ) filter is recommended in order to get rid of fine dust contamination.

### *3. Selection of column size*

With regard to column size selection, it is crucially important whether the chromatographic column will be used for analytical or preparative purposes. For analytical purposes, microbore or minibore columns with an internal diameter of 1-3 mm should be used. In the case of microbore/minibore columns, the applicable sample volume is 5-25  $\mu\text{L}$ , which may contain 0.01-0.1 mg of material. In such applications, the amount of solvents used can be reduced significantly, which is advantageous for both financial and environmental reasons. The use of such thin columns is also recommended when only a small amount of sample is available. If a small sample is applied to a large column, the sample may “disappear”.

The most commonly used type of column is the standard column with a 4.6-mm diameter. These columns can be used for both analytical and semi-preparative purposes. The volume of standard columns is 10-50  $\mu\text{L}$  in the case of analytical uses; whereas for semi-preparative work, column volumes up to 1000  $\mu\text{L}$  can be used. The amount of material that can be applied is in the range of 0.1-2 mg.

Columns with a diameter of 10-20 mm can be used for preparative purposes. In such columns, the amount of material that can be applied may be 20-200 mg, and the sample volume can reach 5-10 mL.

Columns are expensive. It is advisable to insert a protective, few-millimetre-long “pre-column” before the main column, with the two columns having identical column media. The pre-column can be replaced at low cost if blockage occurs.

# Chapter 7. Electrophoresis

by Gábor Pál

## 7.1. Principles of electrophoresis

Electrophoresis is a method used to separate charged particles from one another based on differences in their migration speed. In the course of electrophoresis, two electrodes (typically made of an inert metal, e.g. platinum) are immersed in two separate buffer chambers. The two chambers are not fully isolated from each other. Charged particles can migrate from one chamber to the other (Figure 7.1). By using an electric power supply, electric potential ( $E$ ) is generated between the two electrodes. Due to the electric potential, electrons move by a wire between the two electrodes. More specifically, electrons move from the anode to the cathode. Hence, the anode will be positively charged, while the cathode will be negatively charged. As mentioned above, the two electrodes are immersed in two buffer chambers. Electrons driven to the cathode will leave the electrode and participate in a reduction reaction with water generating hydrogen gas and hydroxide ions. In the meantime, at the positive anode an oxidation reaction occurs. Electrons released from water molecules enter the electrode generating oxygen gas and free protons (which immediately form hydroxonium ions with water molecules). The amount of electrons leaving the cathode equals the amount of electrons entering the anode. As mentioned, the two buffer chambers are interconnected such that charged particles can migrate between the two chambers. These particles are driven by the electric potential between the two electrodes. Negatively charged ions, called anions, move towards the positively charged anode, while positively charged ions, called cations, move towards the positively charged cathode.

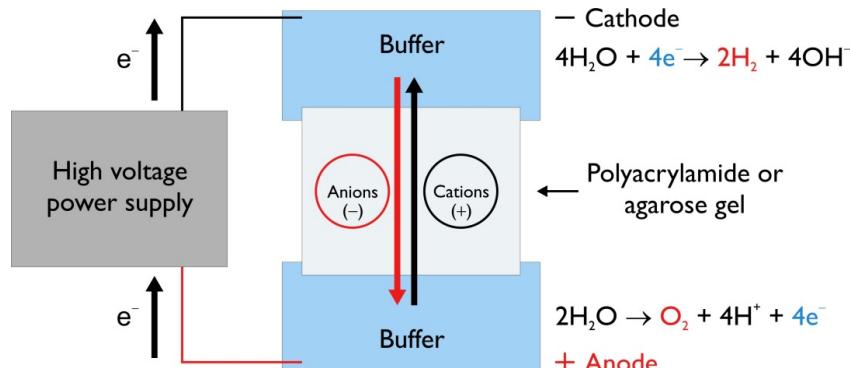


Figure 7.1. The principle of electrophoresis. In the course of electrophoresis, two electrodes are immersed in two separate buffer chambers. The two chambers are connected such that charged particles can migrate from one chamber to the other. By using a power supply, electric potential difference is generated between the two electrodes. As a result, electrons flow from one of the electrodes, the anode, towards the other electrode, the cathode. Electrons from the cathode are taken up by water molecules of the buffer, resulting in a chemical reaction which generates hydrogen gas and hydroxide ions. In the other buffer chamber, water molecules transfer electrons to the anode in another chemical reaction that generates oxygen gas and protons. (Protons are immediately taken up by water molecules to form hydroxonium ions.) As charged particles can migrate between the two chambers due to the electric potential difference, positive ions (cations) move towards the negatively charged cathode while negatively charged ions (anions) move towards the positively charged anode.

Different ions migrate at different speeds dictated by their sizes and by the number of charges they carry. As a result, different ions can be separated from each other by electrophoresis. It is very important to understand the basic physics describing the dependence of the speed of the ion as a function of the number of charges on the ion, the size of the ion, the magnitude of the applied electric field and the nature of the medium in which the ions migrate. By understanding these basic relationships, the principles of the many different specific electrophoresis methods become comprehensible. The fundamental principle of electrophoresis is illustrated in Figure 7.1.

The mathematical description of the force during electrophoresis is simple. An electric force  $F_e$  is exerted on the charged particle. The magnitude of the electric force equals the product of the charge  $q$  of the particle and the electric field  $E$  generated between the two electrodes:

$$F_e = q \times E \quad (7.1)$$

Dimensions of the electric field  $E$  are defined either in newton/coulomb or volt/cm units. During electrophoresis, the magnitude of the electric field  $E$  is defined in volt/cm units. It can be easily calculated using the value of the voltage (volt) set by the electric power supply and the distance of the two electrodes (cm).

As soon as the electric field is applied and the charged particles are accelerated by the electric force, a drag force ( $F_d$ ) called friction will also be immediately exerted on the particles by the medium. This force, whose direction is opposite to the direction of particle movement, is proportional to the velocity of the particle. At the typically very low speed of particle migration during electrophoresis, the force  $F_d$  is a linear function of the velocity ( $v$ ) of the particle, as described by Equation 7.2:

$$F_d = f \times v \quad (7.2)$$

The ratio of the force and the velocity is defined as the frictional coefficient ( $f$ ). The value of  $f$  is a function of the size and shape of the particle and the viscosity of the medium. The larger the particle and the more obstructing the medium, the higher the value of  $f$ .

When electrophoresis is started, particles accelerate instantaneously to a velocity ( $v$ ) at which the magnitude of the drag force equals the magnitude of the (opposite) accelerating electric force:

$$q \times E = f \times v \quad (7.3)$$

Once the magnitude of the two opposing forces becomes equal, the resultant force becomes zero. Therefore, each particle will move at a constant velocity characteristic of the given particle at the given accelerating potential and medium. (A similar phenomenon is described in Chapter 5 for centrifugation. There, the accelerating force is unrelated—being proportional to the mass instead of the charge of the particle—but the frictional force and the phenomenon of two opposing forces leading to a characteristic particle velocity is analogous.) A useful parameter, the electrophoretic mobility ( $\mu$ ) of the particle, defines the velocity of the particle in a given medium when one unit of electric field is applied. (This parameter is analogous to the Svedberg units defined for centrifugation.) Electrophoretic mobility is a linear function of the charge of the particle and it is a reciprocal function of the frictional coefficient (which depends on both the size of the particle and the nature of the medium):

$$\mu = \frac{v}{E} = \frac{q}{f} \quad (7.4)$$

Particles having different electrophoretic mobility, i.e. those that migrate at different speeds in the same medium and electric field, can be separated by electrophoresis.

In biochemical and molecular biological studies, the most typical charged molecules that are analysed and separated by electrophoresis are proteins and nucleic acids. Electrophoresis is always performed by using a special medium, most often a gel. The corresponding methods are therefore denoted as gel electrophoresis.

## 7.2. About gel electrophoresis

The principle of electrophoresis does not assume any particular requirements about the nature of the liquid medium in which the ions are separated. Yet, in the great majority of currently used electrophoretic applications, the medium has a three-dimensional network structure, i.e. the medium is a gel.

At the very beginning when the technique was invented, electrophoresis was performed without using a gel matrix. Charged particles were migrated in a homogeneous liquid phase. However, it soon became apparent that the use of a liquid medium raises at least three major difficulties.

One is that the separation of different ions in an ordinary liquid is rather inefficient. It is so because a significant factor of an effective separation should be a marked size-dependent drag force exerted by the medium on the

particles. Although even ordinary liquids do interfere with the migration speed of the particles in a size-dependent manner, this size dependence is quite moderate.

The other big problem has a simple technical origin. In liquid phase, even very small levels of temperature inhomogeneity trigger convection that significantly compromises the resolution of the separation. Finally, in an ordinary liquid phase, the extent of diffusion is high and, in the typical timeframe of the generally slow electrophoresis experiments, diffusion decreases the efficiency of the separation.

All three problems had been dealt with when, instead of ordinary liquids, gels were introduced as a medium.

The gel provides a three-dimensional molecular network structure to the liquid medium. It prevents convections and also lowers the rate of diffusion. Moreover, perhaps the most dramatic advantageous effect of the gel is that it acts as a molecular sieve: it interferes only slightly with the movement of small molecules, but drastically slows down the motion of large molecules. All gels are characterised by an average pore size. Molecules much smaller than the mean pore diameter are almost unaffected by the presence of the gel, while those that are larger than the pores practically do not migrate in the gel. When ions with sizes in the range of the pore size are migrated through the gel by electrophoresis, the gel exerts a pronounced size-dependent dragging force on them.

As a consequence, the pore size distribution of the gel determines an operational size range in which different ions can be separated. Looking at this from the opposite point of view, each separation problem defines an optimal pore size to be applied.

The gel has to fulfil several general criteria to be applicable for biochemical electrophoresis. It needs to be hydrophilic, chemically stable (should not participate in chemical reactions during electrophoresis), neutral (free of electric charges, otherwise it would act as an ion exchanger) and mechanically resistant (should not be too elastic or too rigid as such gels would be difficult to handle). Furthermore, as the separated ions (mostly proteins and nucleic acids) need to be visualised in the gel by some kind of staining procedure, the gel should be transparent, and should not strongly bind the dyes used for staining. Finally, and very importantly, the experimenter should be able to adjust the pore size during the preparation of the gel.

The size range of molecules (ions) studied in molecular biology is extremely broad. No single gel-forming compound is known that could cover the entire corresponding range of applicable pore sizes. Two compounds are dominantly used for gel electrophoresis: polyacrylamide and agarose. Polyacrylamide gels typically provide much smaller pores than do agarose gels. The polyacrylamide gel is formed by the radical polymerisation of acrylamide monomers. This process alone would lead to very long polymer chains instead of a three-dimensional gel. The three-dimensional network is brought about by the incorporation of *N,N'*-methylenebisacrylamide into the polymerising chains, which results in crosslinks between the long chains. The polyacrylamide gel is held together by covalent bonds. The pore size of polyacrylamide gels can be adjusted via the concentration of the acrylamide monomer and the ratio of the crosslinking agent, *N,N'*-methylenebisacrylamide. The pore size of polyacrylamide gels corresponds to a relatively low value (compared to that of agarose gels). Polyacrylamide gels are used typically for the electrophoresis of proteins and relatively small nucleic acids.

In comparison, the agarose gel is formed via non-covalent interactions between long polysaccharide chains. The pore size of agarose gels is much larger than that of acrylamide gels. Accordingly, agarose gels are used typically for the electrophoresis of large nucleic acids. The pore size of the agarose gel can be controlled via the concentration of the agarose solution. As the interaction between agarose molecules is non-covalent, the gel is formed by a physical (in contrast to a chemical) process. A suspension of agarose is heated up until the system reaches a sol state and then it is left to cool down to room temperature to reach the gel state. The following sections review the various polyacrylamide- and agarose-based electrophoresis methods.

## 7.3. Polyacrylamide gel electrophoresis (PAGE)

### 7.3.1. About the PAGE method in general

As mentioned previously, polyacrylamide gels can be used for the separation and analysis of proteins and relatively small nucleic acid molecules. For example, when it was first invented, Sanger's DNA sequencing method (see in details in Chapter 10) applied PAGE to separate linear single-stranded DNA molecules based on their length. The resolution of the PAGE method is so high that, in the size range of about 10-1000 nucleotide units, it is capable

of separating DNA molecules that differ in length only by a single monomer unit. In the case of single-stranded DNA, individual molecules are separated solely based on their length. This is due to the fact that, in the case of DNA (or RNA), the number of negative charges is a simple linear function of the number of monomer units (i.e. the length of the molecule). In other words, the specific charge (number of charges per particle mass) is invariant, i.e. it is the same for all DNA molecules. It is so because each monomer unit has one phosphate moiety that carries the negative charge. When an appropriate denaturing agent, such as urea, is added to the DNA sample and the gel is heated, the shape of the varying-length linear DNA molecules becomes identical. As a consequence, denatured molecules will be separated exclusively based on their size. (We will see the same principle at the SDS-PAGE method that separates denatured proteins almost exclusively based on their size (molecular weight)). There are several PAGE methods (SDS-PAGE, isoelectric focusing, 2D PAGE) that can be applied mostly for the separation of proteins based on distinct molecular properties.

At a given pH, different proteins carry different amounts of electric charge. Moreover, different proteins have different shapes and sizes, too. Consequently, during electrophoresis, proteins are separated by a complex combination of their charge, shape and size. PAGE separation of proteins provides high resolution. However, as three independent molecular properties simultaneously influence electrophoretic mobility, it will provide limited room for precise interpretation. For example, when two proteins are compared, it remains hidden what makes one of them migrate faster: a larger number of electric charges, a smaller size, or a more spherical shape. Nevertheless, even the simplest PAGE method, which will be referred to as native PAGE, provides many particular advantages (see below).

In order to increase the analytical applicability of the PAGE technology, several variations of the method have been established to separate proteins based on a single molecular property. As we will see, SDS-PAGE separates proteins based primarily on molecular weight, while isoelectric focusing separates proteins exclusively based on isoelectric point.

In the presence of suitable initiator and catalyst compounds, acrylamide can readily polymerise in a radical process. (Acrylamide is harmful by inhalation or skin contact, and thus it should be handled with care.) This reaction would lead to very long polyacrylamide chains, yielding a highly viscous liquid instead of a gel. As already mentioned, these long chains need to be cross-linked to form a three-dimensional network. This is achieved by mixing *N,N'*-methylenebisacrylamide into the acrylamide solution. In essence, *N,N'*-methylenebisacrylamide is composed of two acrylamide molecules covalently interconnected via a methylene moiety. When, during the polymerisation reaction, the acrylamide groups of *N,N'*-methylenebisacrylamide molecules become incorporated in the long polyacrylamide chains, cross-links are formed between the polyacrylamide chains leading to a gel (Figure 7.2). In the course of electrophoresis, ions (proteins or nucleic acids) are separated in this gel.

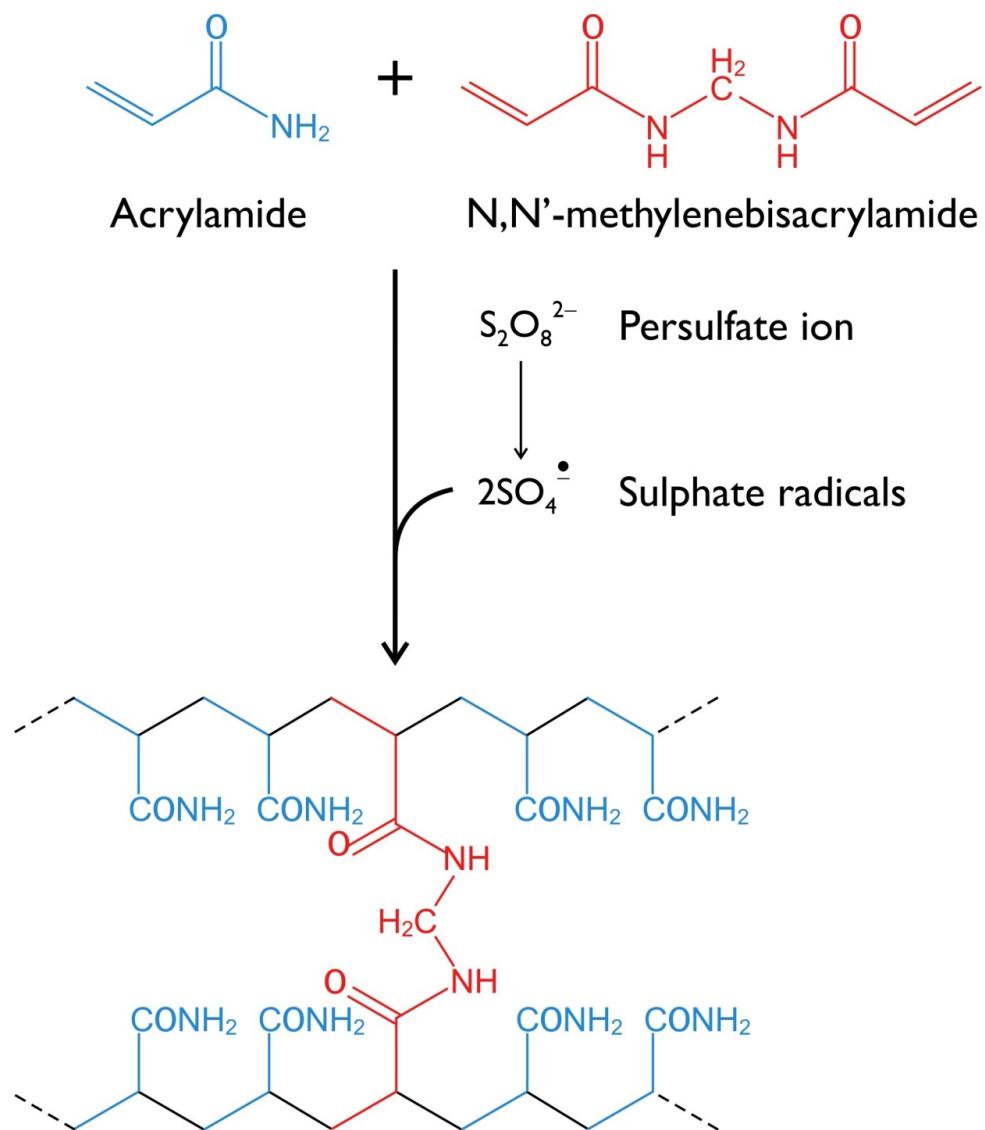


Figure 7.2. Molecular structure of the polyacrylamide gel. The three-dimensional molecular network comes into being by a radical polymerisation of acrylamide monomers and cross-linking *N,N'*-methylenebisacrylamide components.

Without any modification, polyacrylamide electrophoresis separates macromolecular ions based on a combination of charge, size and shape. Size (and shape) separation is due to the molecular sieving property of the gel. The size range in which molecules can be separated is dictated by the average pore size of the gel. In the case of polyacrylamide gels, this can be controlled through the concentration of the acrylamide monomer and the proportion of the cross-linking *N,N'*-methylenebisacrylamide. The acrylamide concentration can be set in the range of about 4-20 % as this is the range in which the mechanical properties of the gel are appropriate. Below this range the gel will be too soft and it will not keep its shape, while above this range it will be too rigid and prone to break. The optimal proportion of the *N,N'*-methylenebisacrylamide component is 1-3 % relative to the acrylamide component. The polyacrylamide gel possesses all advantageous properties necessary for a good electrophoresis medium, i.e. it is hydrophilic, free of electric charges and chemically stable. A further very important property of the polyacrylamide gel is that it does not participate in any non-specific or specific binding interaction with proteins. Furthermore, the polyacrylamide gel does not interfere with common protein staining reactions.

When electrophoresis is performed under native (non-denaturing) conditions, such as near neutral pH and ambient or lower temperature, many enzymes retain their native conformation and, in turn, their enzymatic activity. This way, many enzymes can be separated and specifically detected in the gel after electrophoretic separation.

In the course of creating the gel, a buffer with a properly chosen pH is mixed into the acrylamide/*N,N'*-methyl-enebisacrylamide solution. Radical polymerisation is subsequently triggered by suitable catalyst and initiator compounds. The catalyst is usually ammonium persulfate, which spontaneously decomposes in aqueous media, thereby generating free radicals. These free radicals in themselves cannot efficiently cleave the double bonds of the acrylamide molecule, but are able to excite the electrons of the initiator molecules. This leads to the generation of free radicals, originating from the initiator molecules, that are able to trigger radical polymerisation of acrylamide monomers. The most frequently used initiator is tetramethylethylenediamine (TEMED).

There are two types of gels according to their geometry. In early gel electrophoretic applications, gel tubes were used that allowed only a single sample to be run. Gel slabs were later introduced, allowing for many samples to be run at the same time in the same gel in parallel. Gel slabs became much more common than gel tubes. Gel slabs are created by pouring the gel-forming solution between two parallel glass sheets prior to polymerisation (Figure 7.3). Besides its higher throughput, this gel geometry provides another important advantage over gel tubes: samples are loaded side by side on such slabs and are run in the same gel at the same time. This allows for a more reliable comparison of the samples, facilitating the interpretation of experimental results.

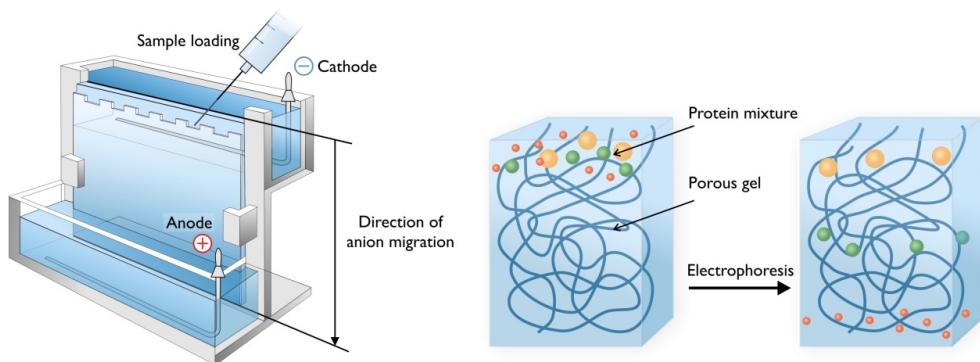


Figure 7.3. Separation of proteins in a polyacrylamide gel. As illustrated in the left panel, several samples can be run in parallel in a slab gel. Ions can move between the two electrodes only through the gel interconnecting the two chambers. The gel acts as a molecular sieve. The larger the molecule, the larger the drag force exerted on it by the gel.

Proper selection of pH and acrylamide concentration is instrumental for successful electrophoresis. For protein electrophoresis, the pH is set usually higher than the pI value of the proteins in the sample. At such a pH, all proteins will be negatively charged and will move towards the anode. The buffer in the medium serves two purposes. One is to set and maintain the proper pH during electrophoresis. The other function of the buffer is to establish the electric current in the medium.

The majority of the electric current is carried by the ions of the buffer. Normally, the protein-ions that are separated by electrophoresis have only a negligible contribution to the current. In other words, proteins have a low ion transport number. However, if the buffer concentration is set too low, the contribution of proteins to carrying the current will increase, and the protein molecules will migrate rapidly. This usually leads to smearing of the bands of migrating proteins. On the other hand, if the buffer concentration is set too high, the mobility of the proteins will be too low. In this case the electrophoresis process would take a very long time. Unnecessary lengthening of the process provides excess time for diffusion, which lowers the resolution of separation.

According to the applied buffer system, gel electrophoretic methods can be classified into two types: continuous and discontinuous. Continuous methods apply the same buffer in the gel and in the two buffer chambers containing the electrodes. The only advantage of this method lies in its simplicity. More complex discontinuous methods were introduced to provide higher resolution. SDS polyacrylamide gel electrophoresis (see later) is usually associated with such a discontinuous system.

The discontinuous system applies two gels of different pore size and three different buffers. One of the gels, the resolving gel, is polymerised at a higher acrylamide concentration. The pore size of this gel is set according to the size range of the proteins to be separated. Another gel, the stacking gel is created on top of the resolving gel. (The gels are mounted in a vertical format.) The stacking gel is polymerised from a more dilute acrylamide solution to provide larger pores. This pore size does not provide a molecular sieving effect.

As mentioned above, there are three buffers: different ones in each of the two gels and a third one, the so-called ‘running buffer’ in the buffer chambers containing the electrodes. In the gel buffers, the anion originates from a strong acid; it is usually chloride ion. Dissociation of strong acids does not depend on the pH: these acids always fully dissociate. Consequently, chloride ion is never protonated in the solution: its ionisation state is independent of the pH. On the other hand, the anion component of the running buffer is the conjugate base of a weak acid. Consequently, the ionisation state of this ion depends on the pH of the buffer. Glycinate ion is one of the most frequently used compounds for this purpose. The pH in the running buffer is set to 8.3.

The protein sample is layered on the top of the stacking gel. When an electric field is generated by the power supply, the protein ions and the ions of the running buffer enter the stacking gel. The pH in the stacking gel is set to 6.8. This value is only slightly higher than the pI value of glycine (6.5). At this pH, most glycine molecules are in a neutral zwitterionic state, and only a small portion of the molecules carry a net negative charge. In this state, glycine has a low electrophoretic mobility and a corresponding low transport number. The local sparsity of ions elevates the local electric resistance of the medium. As the electric current must be of the same magnitude at any segments of the electric circuit (there is no macroscopic charge separation), the voltage will increase according to Ohm’s law. Due to this effect, the migration speed of the proteins will be relatively high and the protein front will reach the chloride front in the stacking gel. The ion concentration in the chloride front is high and, therefore, here the electric resistance and the voltage are low. This slows down the protein front. This effect results in a very sharp protein front, with the protein molecules being crowded right behind the chloride ion front.

The protein sample will thus enter the resolving gel in a sharp band. The pH in the resolving gel is set to about 8.8. At this pH, almost all glycinate molecules are in the anionic state. Thus, the electric mobility of glycinate increases, and the concentrating effect applied by the stacking gel ends in the resolving gel. Different proteins will be separated in the resolving gel according to their charge, size and shape.

In most electrophoretic methods, a tracking dye is mixed in the sample. Usually, this dye is chosen to have a higher electrophoretic mobility than any of the components of interest (proteins or nucleic acids) in the sample. The function of the tracking dye is to visualise the running front and, in turn, the completeness of the run. The most popular tracking dye is bromophenol blue.

The following sections review the various PAGE methods listed from the simplest to the most complex one.

### 7.3.2. Native PAGE

Native PAGE is an electrophoresis method to separate native proteins. The conditions are set such that the migrating proteins are kept in their native state. The buffers provide a non-denaturing, native-like milieu, and the electrophoresis is performed at low temperature in order to dissipate heat. Many enzymes retain their native conformation and their enzymatic activities while running in the gel. If certain conditions apply, these enzymes can be highly selectively detected within the gel through a specific ‘staining’ reaction even in the presence of a large excess of ‘contaminating’ proteins. After completion of electrophoresis, the gel is soaked in a solution containing the substrate of the enzyme. As the substrate is usually a small molecule, it quickly diffuses into the gel while the large enzyme molecules do not diffuse out. In an optimal case, the natural product of the enzymatic reaction is a coloured and insoluble compound that precipitates inside the gel and marks the exact location of the enzyme. Of course, most enzymes do not have such natural substrates. However, once the molecular mechanism of catalysis is revealed, synthetic substrates can be designed that, on the one hand, mimic natural substrates and, on the other hand, lead to colourful insoluble products.

Native PAGE is also a useful method for checking the uniformity of the isolated protein. Even if the purified protein sample contains only a single type of protein, the sample might not be uniform. Some of the molecules might be unfolded or have undergone chemical modifications. Unfolding changes the overall shape of the molecule, while most chemical modifications change the electric charge of native molecules. These alterations can be detected after traditional staining of the purified sample. If no such side products are present, protein molecules will run in a single sharp band. Otherwise, multiple bands or smearing of the band is expected.

In addition, native PAGE can also be used to detect complex formation between proteins. If two (or more) proteins (or proteins and non-proteinous ligands) form a complex, the complex can be detected as an extra band in the gel. This is because in native-like conditions, many non-covalent (subunit-subunit, receptor-ligand, enzyme-inhibitor) interactions are maintained and the complex migrates apparently as a single molecule.

In the course of native PAGE, it is highly important to pay attention to the relationship of the pI values of the proteins or protein complexes and the pH of the gel buffer, as this will determine where individual proteins will migrate in the gel.

### 7.3.3. SDS-PAGE

SDS-PAGE is an electrophoresis method to separate proteins. However, unlike in the case of native PAGE, here the proteins migrate in their denatured state. As it was mentioned in the general introduction to traditional (native) PAGE, the migration velocity of proteins is a function of their size, shape and the number of electric charges they carry. As the velocity is a complex function of these properties, native PAGE cannot be used to estimate the molecular mass of proteins. The traditional native PAGE method is similarly unable to assess whether a purified protein is composed of a single subunit or multiple subunits. Even a multi-subunit protein may migrate in a single sharp band.

SDS-PAGE (Figure 7.4) was introduced to analyse such cases and to allow the estimation of the molecular mass of single-subunit proteins or those of individual subunits of multi-subunit proteins. SDS-PAGE is the most prevalent PAGE method currently in use.

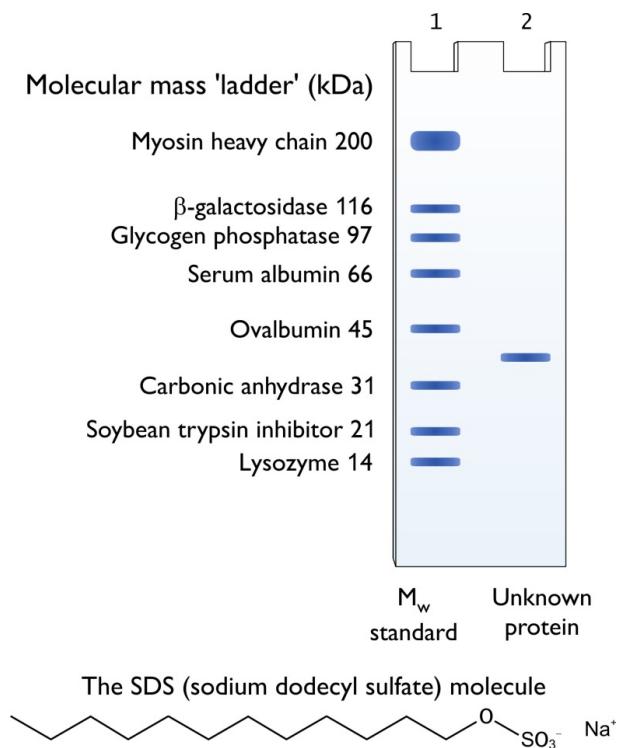


Figure 7.4. SDS polyacrylamide gel electrophoresis. SDS (sodium dodecyl sulphate) is an anionic detergent that unfolds proteins and provides them with extra negative charges. The amount of the associated SDS molecules—and therefore the number of charges—is proportional to the length of the polypeptide chain. The SDS gel separates individual polypeptide chains (monomeric proteins and subunits of multimeric proteins) according to their size.

The velocity of the proteins is an inverse linear function of the logarithm of their molecular mass. Proteins of known molecular mass can be used to establish a calibration curve (a descending line) along which the unknown molecular mass of other proteins can be estimated.

SDS (sodium dodecyl sulphate) is an anionic detergent. When proteins are treated with SDS at high temperature, radical conformational changes occur. The treatment breaks all native non-covalent intermolecular (inter-subunit) and intramolecular interactions. The subunit structure of multi-subunit proteins disintegrates and the proteins unfold. If the native structure is stabilised by disulfide bridges, reducing agents are also added to open up these connections. SDS molecules bind to unfolded proteins in large excess, providing extra negative charges to the molecules.

The amount of the bound SDS molecules is largely independent of the amino acid sequence of the polypeptide chain and it is roughly a linear function of polypeptide length—i.e. the molecular mass of the protein. Therefore,

upon SDS-treatment, the specific charge (the charge-to-mass ratio) of different proteins will become roughly identical. Another result of the treatment is that the shape of the different proteins becomes similar. The negatively charged SDS molecules repel each other, which lends a (presumably) rod-like shape to the SDS-treated proteins. These factors together result in a situation analogous to the one already discussed in this chapter for the PAGE separation of linear single-stranded (denatured) DNA molecules. Instead of being separated simultaneously by charge, shape and size, SDS-treated proteins—just like denatured linear DNA molecules—will be separated solely based on their size. As size is a linear function of mass, SDS-PAGE ultimately separates proteins based on their molecular mass.

SDS-PAGE is the most popular cost-effective method to estimate the molecular mass of protein subunits with considerable accuracy. The relative mobility (i.e. the running distance of the protein divided by the running distance of the tracking dye) of the SDS-treated protein is in inverse linear proportion to the logarithm of the molecular mass of the protein. By running several proteins of known molecular mass simultaneously alongside the protein of interest, a log molecular mass – relative mobility calibration curve (a descending linear graph) can be created. Based on the calibration curve, the estimated molecular mass of the protein in question can be easily calculated.

Table 7.I below shows the useful separating range of polyacrylamide gels as a function of acrylamide concentration. In the useful range, the log molecular mass – relative mobility relationship is linear.

| Acrylamide concentration (%) | Linear range of separation (kDa) |
|------------------------------|----------------------------------|
| 15                           | 12-43                            |
| 10                           | 16-68                            |
| 7.5                          | 36-94                            |
| 5.0                          | 57-212                           |

Table 7.I. Relation between acrylamide concentration and the molecular mass of optimally separated molecules

SDS-PAGE is a standard method for assessing whether the sample of an isolated protein is homogeneous. Besides that, SDS-PAGE is a robust method for the analysis of large supramolecular complexes such as multi-enzyme complexes or the myofibril, as discussed below. SDS-PAGE separates and denatures individual subunits of these complexes. Thus, all polypeptide chains will migrate separately in the gel. Via various staining procedures, all subunits can be visualised and the relative amounts of these proteins (subunits) can also be determined. This allows for the identification of each subunit of a complex and provides a good estimate of the stoichiometry of subunits, too.

### 7.3.4. Isoelectric focusing

In the course of isoelectric focusing, the conditions are set in a way that proteins will be separated exclusively based on their isoelectric point (Figure 7.5). The two termini and many side chains of proteins contain dissociable groups (weak acids or bases). The dissociation state of these groups is a function of the pH of the environment (as described quantitatively by the Henderson-Hasselbalch equation, see Chapter 3). Isoelectric focusing is based on the pH-dependent dissociation of these groups. Due to this pH-dependent phenomenon, the net electric charge of a protein molecule will be a function of the pH of the medium. If, in a given protein, the number of acidic residues (Asp, Glu) exceeds that of the basic ones (Arg, Lys, His), the protein will have a net negative charge at neutral pH. The isoelectric point (pI) of the protein—i.e. the pH at which the net charge of the protein is zero—will be in the acidic pH range. Such proteins are often denoted as acidic proteins. If the number of basic residues exceeds that of the acidic ones, the protein will be positively charged at neutral pH, and its pI value will be in the basic pH range. These proteins are often called basic proteins.

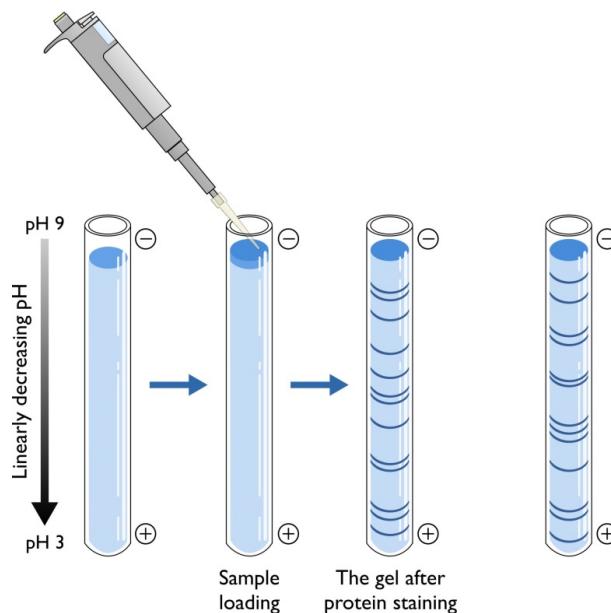


Figure 7.5. Isoelectric focusing. In the course of isoelectric focusing, a pH gradient is created in the gel (usually made of polyacrylamide, less frequently agarose). Upon electrophoresis, various proteins will accumulate in different narrow regions of the gel where the pH equals their individual pI value. At this pH, the number of positive charges equals that of the negative charges on the protein—the net charge will thus be zero. Consequently, no resultant electric force is exerted on the protein.

Isoelectric focusing is an efficient high-resolution method because the pI values of various proteins are spread across a broad range. If the pH is lower than the pI of the protein, the protein will be positively charged and will move towards the cathode during electrophoresis. If the pH is higher than the pH of the protein, the protein will be negatively charged and will migrate towards the anode. If the pH equals the pI value, the net charge of the protein will be zero and the protein will not migrate in the gel any further.

In the course of isoelectric focusing, proteins are placed in a gel representing a special medium in which the pH gradually decreases by going from the negative cathode towards the positive anode. As the protein migrates, it encounters a gradually changing pH and its net charge will also change accordingly. If it has a net negative charge and therefore moves towards the cathode, it will encounter a gradually decreasing pH, i.e. a more and more acidic environment. Consequently, the protein will take on more and more protons—up to a level where its net charge will be zero. This state is reached when the protein reaches a location where the pH equals its pI value. At this point, the protein will stop moving because no electric force will be exerted on it. If it spontaneously diffused further towards the anode, it would take on more protons, would become positively charged and would turn back to migrate towards the cathode. Following the same line of thinking, if a positively charged protein moves towards the cathode, it will encounter increasing pH and lose more and more protons. It will migrate to the place where the pH equals its pI value and will thus stop. If it diffused further towards the cathode, it would become negatively charged and would turn back towards the anode. As one can see, by performing electrophoresis in a medium in which the pH decreases from the cathode towards the anode, each protein will “find its place” according to its pI value and will become sharply focused at that location. In addition, it does not matter where exactly the proteins were introduced in the medium between the cathode and the anode.

A decisive component of this method is the usually linear pH gradient created inside the gel. There are two methods to create such a gradient. One of them applies carrier ampholytes (ampholyte is an acronym from the words amphoteric and electrolyte). Ampholytes or zwitterions are molecules that contain both weakly acidic and weakly basic groups. Just like in the case of proteins, the net charge of ampholytes is a function of the pH. In the course of isoelectric focusing, a mixture of various ampholytes is used such that the pI of the various ampholyte components will cover a range in which the pI values of the “neighbouring” ampholytes differ only slightly. This ampholyte mixture is soaked in the gel and an appropriate electric field is generated by a power supply. This leads to a process analogous to the one already explained for proteins. Each ampholyte will migrate to the location where its net charge becomes zero. As soon as this steady-state is achieved, ampholytes will function as buffers and keep the pH of their immediate environment constant. This establishes the pH gradient in which the proteins can be separated.

The other, more sophisticated method applies special ampholytes that can be covalently polymerised into the polyacrylamide gel. The appropriate ampholyte gradient is created before the gel is polymerised. This way, the gradient will be covalently fixed in the gel, providing an immobilised pH gradient. The appropriate pH range provided by the ampholyte mixture should be selected based on the pI values of the proteins to be separated.

Regardless of how the pH gradient was created, once the proteins reach the location in the gel where the pH equals their pI, they finally stop moving and the system reaches a steady-state.

One of the potential technical difficulties encountered during isoelectric focusing originates from the fact that the solubility of proteins is lowest at their pI value (see Chapter 5). This can lead to the precipitation of some proteins in the gel. To prevent this unwanted process, urea is most often applied in the gel as an additive. Urea denatures proteins and keeps denatured proteins in solution. As the pI value of proteins is largely independent of their conformational state, this modification does not compromise the method. The solubility of membrane proteins can be further promoted by the addition of non-ionic detergents.

Isoelectric focusing is aimed at separating proteins based exclusively on their pI value—thus, independently of their size. Therefore, the molecular sieving property of the gel in this method should be avoided. The only function of the gel is to prevent free convectional flows in the medium. Accordingly, for isoelectric focusing, polyacrylamide gels are made at very low acrylamide concentrations, and sometimes even agarose gels are applied when very large pores are needed. Isoelectric focusing is usually performed in a horizontally-mounted electrophoresis apparatus and by applying intense cooling.

### 7.3.5. Two-dimensional (2D) electrophoresis

The various separation methods are all aimed at separating complex systems to individual components. Separation is always based on at least one physicochemical property that shows diversity among the components. The general problem encountered in the case of complex mixtures is that not all components differ significantly from all other components when only one property is considered. Accordingly, separation based on a single property rarely results in single-component fractions. Some components will be efficiently separated from all others, while some other components will remain in the mixture.

The remaining mixtures can be further fractionated by another separation technique that relies on a different physicochemical property. The most effective separation can be achieved if the combined consecutive separation steps rely on absolutely independent physicochemical properties. A good example of this is the very high-resolution two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) that combines two already discussed electrophoresis methods, isoelectric focusing and SDS-PAGE (Figure 7.6).

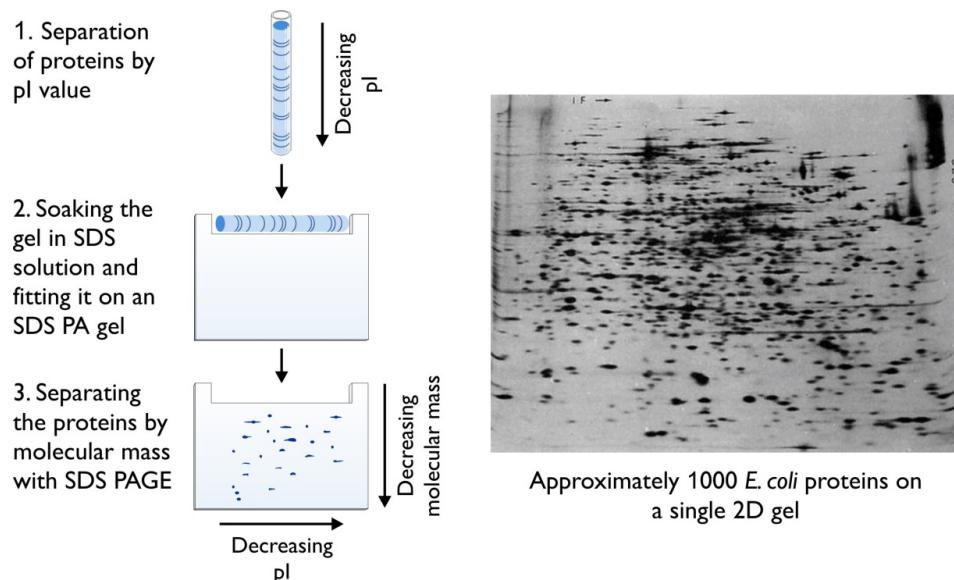


Figure 7.6. Two-dimensional (2D) electrophoresis. 2D electrophoresis is the combination of isoelectric focusing and SDS-PAGE. Proteins are first separated based on their pI values and then based on their molecular mass. As

these properties are completely independent, the combination of the two separation methods provides much higher resolution than either of the two methods alone.

As the first step of 2D gel electrophoresis, isoelectric focusing is performed to separate proteins based on their pI values. Only a single sample is loaded on a gel strip in this step. The sample is separated in one dimension both in a primary and in a figurative sense. In a primary sense because the components are separated along a single line, and in a figurative sense as the separation is based on a single well-defined property, the pI value.

After the first separation step has been completed in the first dimension, the gel strip is soaked in an SDS solution and is fitted tightly to one side of a “classical” SDS polyacrylamide gel. The second separation step is traditional SDS-PAGE, which separates proteins based on their molecular mass. This second step represents a second dimension in both a primary and a figurative sense. The second separation is performed in a second dimension in a direction rectangular to that of the first separation, and the property utilised in the second step (molecular mass) is completely independent of the one utilised in the first step (pI).

If, after the first step, some gel regions contain different proteins that coincidentally have identical pI values, these proteins will be separated from each other in the second step if their molecular mass is different. Note that every aspect discussed for SDS-PAGE also applies to the second separation step of 2D-PAGE. Van der Waals interactions that might have held protein subunits together in the course of isoelectric focusing will break and individual subunits will become separated. If disulfide bridges need to be opened up, some kind of reducing agent needs to be added. Accordingly, in the second separation step, single polypeptide chains will migrate in the gel. If isoelectric focusing collects a multimeric protein at a certain gel location, the second electrophoresis step will dissect it into individual chains. If the multimer contains subunits of different sizes, these subunits will be separated from each other in the second separation step.

## 7.4. Agarose gel electrophoresis

As discussed in the introduction, the pore size of the gel defines the operational molecular mass range in which effective separation can be achieved. The size of the DNA molecules most frequently handled in a typical recombinant DNA experiment is in the range of thousands of base pairs (see in Chapter 10). Such very large molecules cannot be separated in polyacrylamide gels, as even the largest pore sizes achievable in such gels are too small for these macromolecules. Naturally, a different type of gel matrix is needed.

For this purpose, agarose has become the most popular matrix. The agarose gel forms through non-covalent interactions between polysaccharide molecules. When such a gel is heated, it undergoes a phase transition from gel to sol state. The agarose powder is mixed with running buffer and the slurry is heated up to reach the sol state. The liquid is then poured in an appropriate, usually horizontally-mounted, template to solidify.

The pore size is set by the concentration of agarose monomers. The agarose gel meets all of the important requirements that make a good electrophoresis matrix: it is hydrophilic, does not carry charges, chemically inert and does not absorb the dyes used for nucleic acid staining.

## 7.5. Staining methods

Proteins and nucleic acids absorb light only in the ultraviolet region of the spectrum and are therefore invisible for the naked eye. UV spectroscopy in the gel is not straightforward due to the high background absorbance of the gel. Therefore, following electrophoresis, proteins or nucleic acids are visualised in the gel by using various staining methods.

### 7.5.1. General protein gel stains

Several dyes have been introduced for protein gel staining. These compounds bind tightly to proteins and absorb light in the visible region of the spectrum. There is no single absolutely best compound. The choice of the dye depends mostly on the quantity of the protein to be detected. The following dyes are listed in the order of increasing sensitivity: Coomassie Brilliant Blue R 250, Acidic Fast-Green, Amido black, and silver nitrate.

Coomassie Brilliant Blue is the most frequently used protein gel staining dye. Depending on the thickness of the gel and the properties of the protein, as low as 0.1 microgram protein can be detected using this dye.

## 7.5.2. General DNA gel stains

The most common methods for the in-gel detection of nucleic acids apply some kind of fluorescent dye. More accurately, the compounds alone have a low fluorescence level but, once they bind to the nucleic acid, the resulting complex will become highly fluorescent. The majority of dyes result in complexes that can be excited in the ultra-violet range and emit light in the visible region. To date, the most frequently used dye has been ethidium bromide. Ethidium bromide is an aromatic heterocyclic compound that is able to intercalate between the bases of double-stranded DNA. Such compounds can be mutagenic because they can increase the rate of mutations by interfering with the action of DNA polymerases. Accordingly, they should be used with extreme care.

Albeit with lower sensitivity, ethidium bromide is also able to detect single-stranded DNA molecules. Single-stranded DNA molecules can have partially complementary segments that form short double-stranded sections in the molecules. In addition to ethidium bromide, several other dyes of higher sensitivity have also been introduced and many of these also provide a lower risk of mutagenicity.

## 7.5.3. Specific protein detection methods: Western blot

If, instead of detecting all proteins, the goal is the highly selective detection of a certain protein, two different methods are available.

One method called Western blot can be used if a highly selective antibody is available against the protein of interest (Figure 7.7). The principle of the procedure is as follows. In the first step, proteins are separated based on their size by using SDS-PAGE. Subsequently, a nitrocellulose membrane is laid tightly on one side of the gel. This “sandwich” is placed into a special electrophoresis apparatus. By performing electrophoresis in a direction perpendicular to the plane of the gel, the proteins will be transferred from the gel onto the membrane. This step preserves the relative spatial arrangement of individual protein bands.

The membrane carrying the tightly bound proteins is then incubated in a solution containing the specific antibody against the target protein, usually called the primary antibody. The primary antibody will bind only to the target protein. To visualise the sites where the antibody has bound to the membrane, a secondary antibody is added that had been conjugated (covalently bound) to an enzyme. The secondary antibody recognises the constant region of the primary antibody. Therefore, a given secondary antibody conjugate can be used with many primary antibodies.

After this step, the membrane is soaked in a solution containing a substrate of the conjugated enzyme. Catalytic conversion of the substrate yields a colourful insoluble product that precipitates at the site where it was generated and, ultimately, labels the site of the target protein.

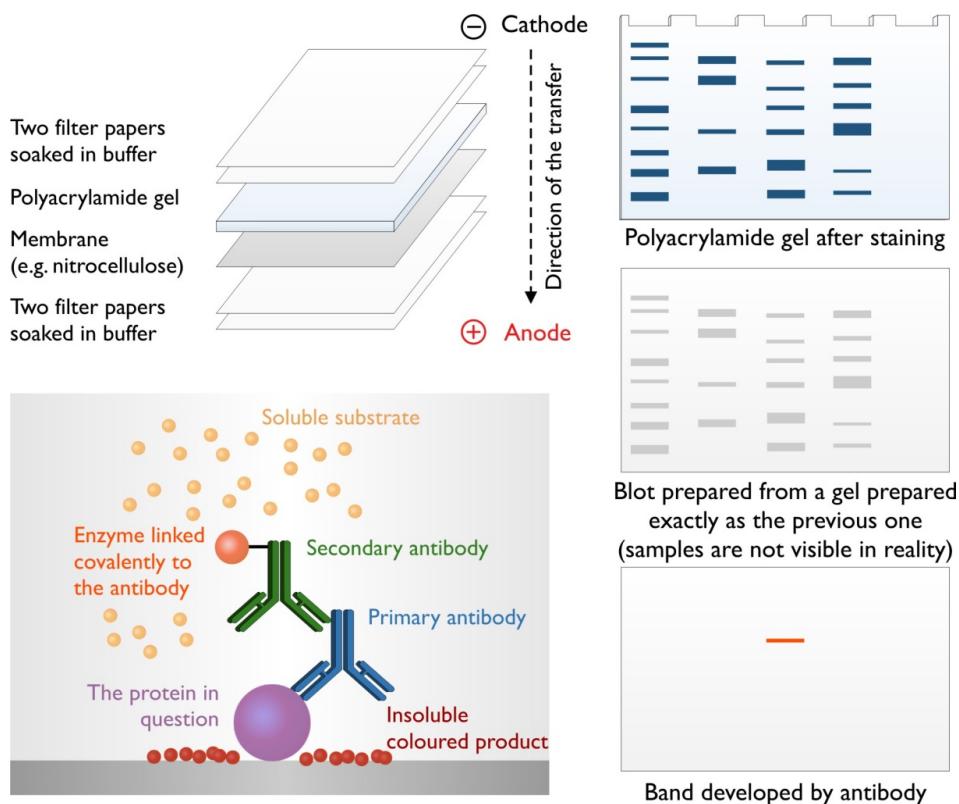


Figure 7.7. The scheme of Western blot. After (usually SDS-) PAGE separation, proteins are transferred to a membrane (usually made of nitrocellulose) by a second electrophoresis performed in a direction perpendicular to the plane of the gel. The membrane tightly binds the proteins. The target protein can be specifically detected among hundreds or thousands of other indifferent proteins by a method analogous to the sandwich ELISA assay discussed in Chapter 8. A critical reagent for this detection method is a highly selective antibody raised against the target protein. This antibody “finds” and binds specifically to the target protein even in the presence of a large excess of unrelated proteins. Visualisation of the location of the bound antibodies is usually achieved via enzymatic methods. In principle, the enzyme could be covalently linked (conjugated) to the target-recognising, so-called primary antibody. However, in this case, such antibody-enzyme conjugates would need to be prepared for each new target protein. A more straightforward approach has been the use of a secondary antibody that recognises the constant region of many different primary antibodies. In the course of the enzyme reaction, a soluble substrate is enzymatically converted into a coloured insoluble product that precipitates around the target protein.

#### 7.5.4. Specific protein detection methods: In-gel method based on enzyme activity

When electrophoresis is performed in native conditions, many enzymes retain their catalytic activity in the gel. There are robust and highly selective in-gel activity detection methods for several types of enzymes (dehydrogenases, ATPases, proteases etc.). For such specific detection, the gel is soaked after electrophoresis in the solution of the proper substrate. Usually, synthetic substrates are used that have a common built-in property: upon being converted by the enzyme, they generate a colourful and insoluble product. As a consequence, the end product stays in the gel very near to the enzyme and colours the spot where the enzyme is located (Figure 7.8).

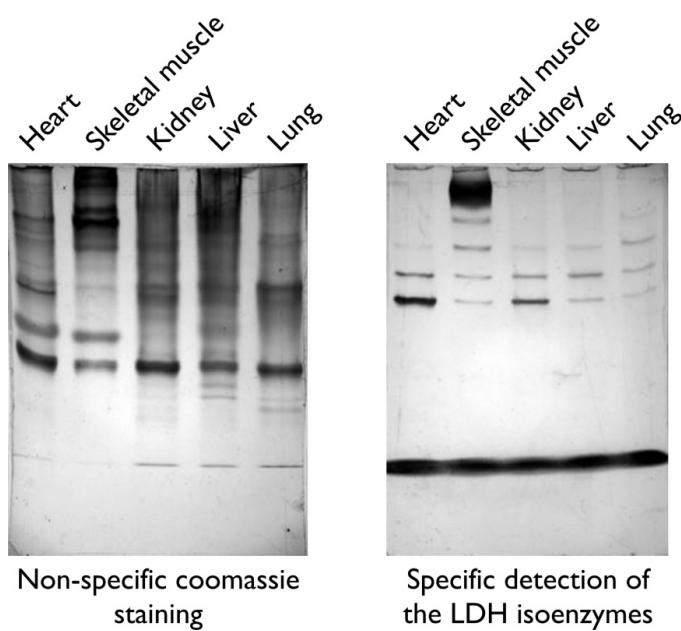


Figure 7.8. In-gel assay to detect lactate dehydrogenase (LDH) enzymes. The five different LDH isoenzymes ( $M_4$ ;  $M_3H$ ;  $M_2H_2$ ;  $M_1H_3$ ;  $H_4$ ) can be separated in a native gel such that they retain their catalytic activity. Left panel: crude protein extracts from various animal tissues were run in a native gel, and the gel was stained by a general protein-staining dye. As hundreds of different proteins were loaded on the gel, the separation of individual components was not achieved: there are overlapping signals instead of individual separate bands. Right panel: the same samples were loaded but the staining method was principally different. Instead of using a dye that binds to all proteins, a selective chemical reaction catalysed only by LDH enzymes was performed. The product of the multi-step reaction is a coloured insoluble compound that precipitates in the gel around the LDH enzymes.

Zymography, a special variant of activity detection methods, has been introduced for the detection of proteinases. While in the above example a native gel was run, in the case of zymography usually a modified SDS-PAGE is performed. When the resolving gel is prepared, some kind of large molecular weight protein (casein, denatured collagen etc.) is mixed into the buffer. This protein is so large that it practically does not migrate in the gel. The sample to be loaded onto the gel is SDS-treated, but no reducing agent is added and the sample is not treated at high temperature. Unlike in the case of normal SDS-PAGE, the goal here is to denature the proteins reversibly. After the proteins are separated by gel electrophoresis, the gel is soaked in a non-ionic detergent solution in order to completely remove SDS. As the SDS is removed, at least some percentage of the proteinases (and other proteins) can regain their native conformation. The refolded proteinases start to digest the protein substrate in the gel. The products of the proteolytic degradation are small peptides that diffuse out of the gel. After an appropriate time period, the gel is stained by one of the general methods. The gel will be intensely stained in most parts except where it contained proteinases. Around the proteinase molecules the staining will be lighter. This is a kind of negative staining approach in which the specific signal is actually the lack of signal.

## 7.6. Typical examples of protein-separating gel electrophoresis

DNA-separating gel electrophoresis methods are discussed in Chapter 10. In this section we show two specific examples of two protein analysis methods, native PAGE and SDS-PAGE. In the previous sections we already explained the theoretical background of these methods. The examples shown here take a step closer to everyday applications and illustrate the practical significance of these methods.

## 7.6.1. Native PAGE separation and detection of lactate dehydrogenase isoenzymes

Enzymes that catalyse the same reaction in the same organism but have different chemical structure are called isoenzymes. The very first enzyme shown to be present in animal tissues in the form of isoenzymes was lactate dehydrogenase (LDH). It turned out that most tissues contain five different isoenzyme forms.

LDH is a tetrameric protein. It has two types of subunits called M (muscle) and H (heart). The M subunit is characteristic mostly of skeletal muscle, while the H subunit is characteristic of cardiac muscle.

The two different subunits are encoded by two different genes. In the quaternary structure of the enzyme, the subunits are arranged as the four vertices of a tetrahedron. As the two subunits can combine randomly, the tetrameric enzyme can be of five isoenzyme forms:  $M_4$ ,  $M_3H$ ,  $M_2H_2$ ,  $MH_3$  and  $H_4$ . Due to random combination of the four subunits, the proportion of the five isoenzymes in different tissues will depend on the tissue-specific relative level of expression of the two encoding LDH genes. The shape and molecular mass (34,000 Da) of the two subunit types is identical. Therefore, the shape and molecular mass (136,000 Da) of the five different tetrameric isoenzymes will also be identical. However, the electric charge of the two subunits is different. Therefore, the charge of the five isoenzymes will also be different. H-type subunits carry more negative charges than M-type ones.

When LDH is subjected to native PAGE, the subunits remain associated and tetramers migrate in the gel. As the five different tetramers have identical shape and size but distinct charge, the native PAGE in this case will separate the isoforms based exclusively on charge differences.

The larger number of H subunits present in the isoenzyme, the more negative charges it will carry. Accordingly, the  $H_4$  isoenzyme will have the highest and the  $M_4$  the lowest electrophoretic mobility. Based on the principles of electrophoresis, the relative mobility of the isoenzymes can be easily deduced.

The magnitude of the electric force ( $F_e$ ) exerted on a given isoenzyme is linearly proportional to the number of H subunits it contains. On the other hand, due to their identical size and shape, the frictional coefficient ( $f$ ) is identical for all isoenzymes. As a consequence of this, the electrophoretic mobility ( $\mu$ ) of the isoenzymes will be a linear function of the number of their H subunits. Let us consider that, at a given electric field, the velocity of the  $M_3H$  isoenzyme exceeds that of the  $M_4$  type by a value of X cm/h. This velocity change is due to the replacement of a single M subunit with a single H subunit. The same increase in velocity should apply to all additional M-to-H replacements. As a result, the electrophoretic mobility will be linearly proportional to the number of H subunits. As a remarkable consequence of this simple relationship, at the end of the separation, the distances between neighbouring isoenzyme bands will be identical.

The fact that the proportion of the various isoenzymes is tissue-specific provides a great diagnostic value. Several diseases (cardiac attack, hepatitis, tumours, certain muscle dystrophies etc.) are accompanied by massive cell destruction. Thus, components of the cytoplasm of the damaged cells including the LDH isoenzymes enter the blood stream. The normal LDH isoenzyme distribution of the serum will therefore be shifted towards the distribution characteristic of the damaged tissue.

Serum analysis of the altered LDH isoenzyme distribution would be a very complex and laborious procedure if the isoenzymes had to be purified to homogeneity. This is circumvented by applying native PAGE combined with a subsequent in-gel enzymatic assay (Figure 7.8). Although hundreds or thousands of different proteins are run in the gel, the specific enzymatic reaction will detect only the LDH isoforms.

LDH catalyses the reversible chemical reaction shown in Figure 7.9. LDH can be detected by a special chromogenic reaction. The assay is based on the reducing power of NADH. After electrophoresis, the gel is incubated in a buffer solution containing lactate,  $NAD^+$  coenzyme, phenazine methosulfate (PMS, an electron acceptor) and nitro-blue-tetrazolium-chloride (NBTC, a redox dye). In the multi-step reaction, lactate is oxidised to pyruvate, while  $NAD^+$  is reduced to NADH. NADH will reduce PMS, which then will reduce the soluble, yellowish oxidised NBTC. The reduced form of NBTC is blue and has a very low solubility. The position of LDH enzymes in the gel will be detected through the formation of a dark blue precipitate in the gel.

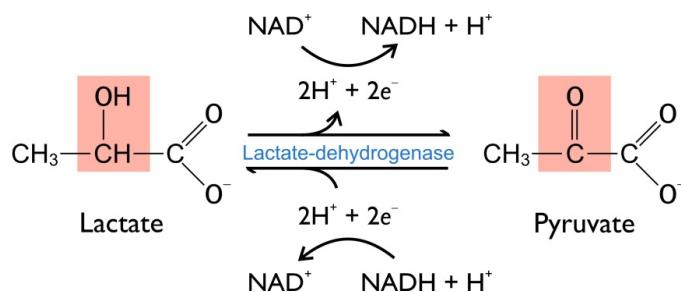


Figure 7.9. The chemical reaction catalysed by lactate dehydrogenase. In the reversible reaction, depending on the actual direction of the reaction, either the oxidised or reduced coenzyme ( $\text{NAD}^+$  or  $\text{NADH}$ , respectively) is produced. In the assay illustrated in Figure 7.8, the gel was soaked in a buffer containing lactate and oxidised  $\text{NAD}^+$ . Two other consecutive chemical steps utilise the reduced  $\text{NADH}$  state to produce a coloured insoluble product that precipitates around the LDH enzymes (see main text for details).

## **7.6.2. Molecular mass determination of myofibrillar proteins using SDS-PAGE**

The myofibril is the cytoplasmic contractile element of skeletal and cardiac muscles. Myofibrils are the loci of spatially organised force generation. Each fibril is composed of a large number of parallel filaments of myofibrillar proteins. Thick filaments are composed mostly of myosin, while thin filaments contain chiefly actin. Myosin is composed of six subunits: two identical heavy chains (HC) and four light chains (LC) (Figure 7.10). Myosin molecules associate through their superhelical portion, thus forming the thick filament.

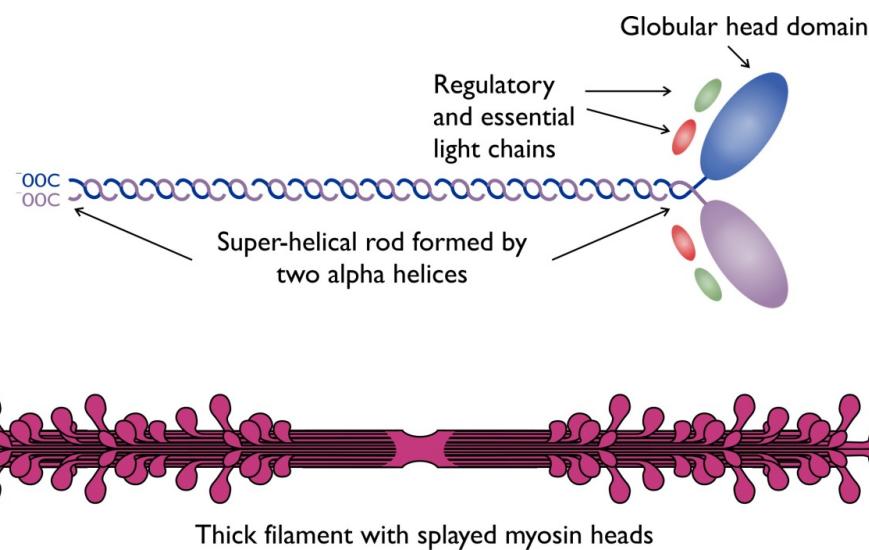


Figure 7.10. Schematic structure of skeletal muscle myosin and the thick filament. Skeletal muscle myosin is composed of three different kinds of protein subunits: one kind of heavy chain and two different kinds of light chains. The heavy chain contains a head, a neck and a tail domain. Two heavy chains associate via their tail domains through a coiled-coil structure forming a rod region. Each heavy chain binds two light chains: an essential and a regulatory light chain. The globular head domains provide the actin binding and the ATPase activities of myosin.

The thick filament is formed by the association of the rod regions of myosin molecules.

The structure of the thin filament is established via the polymerisation of globular actin monomers into actin filaments. In vertebrates, several regulatory proteins (tropomyosin and heterotrimeric troponin) are associated to this filament (Figure 7.11).

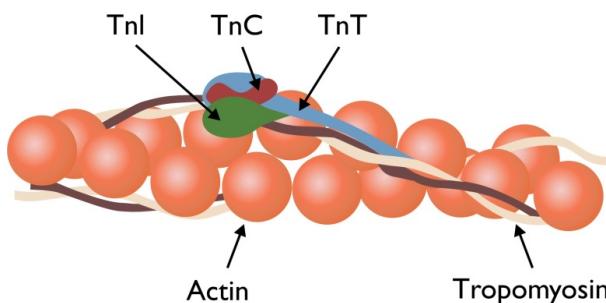
**The troponin complex:**

Figure 7.11. Schematic structure of the thin filament. The thin filament is formed by the polymerisation of globular actin monomers into actin filaments. Regulatory proteins, tropomyosin and subunits of the troponin complex (TnI, TnC, TnT), bind to the actin filament.

The essence of SDS-PAGE can be nicely illustrated via the gel electrophoresis of myofibrils (Figure 7.12). Myofibrils prepared from white skeletal muscle of rabbit contain all the myofibrillar proteins arranged in complex structures described above. These proteins bind to each other through many well-organised van der Waals interactions. All of these interactions are disrupted by SDS treatment at high temperature. The individual myofibrillar proteins and, moreover, the individual subunits of the proteins will be separated in the gel based on their molecular mass.

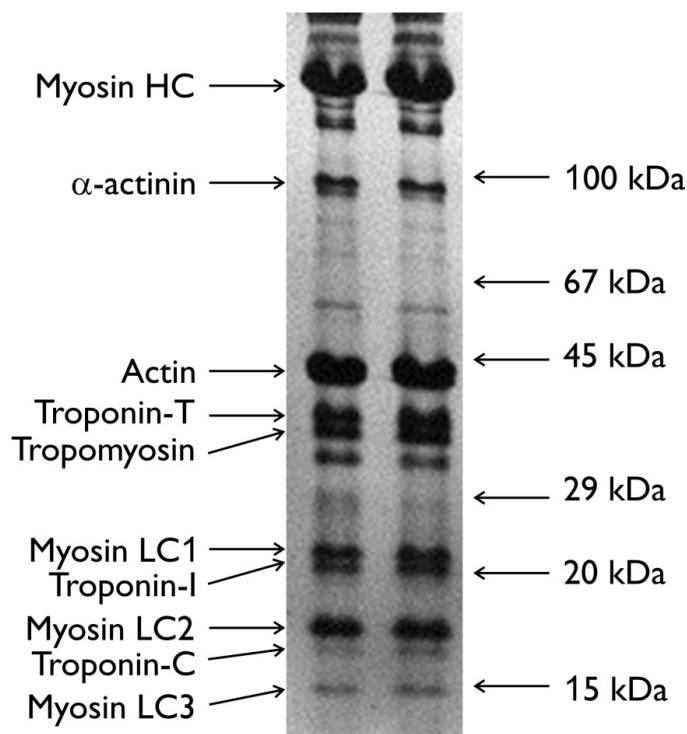


Figure 7.12. SDS-PAGE separation of myofibril components. Individual proteins and protein subunits of the supramolecular myofibril complex are liberated via SDS treatment. Individual unfolded proteins and protein subunits are separated in the SDS polyacrylamide gel according to their molecular mass, which also allows for their identification. Molecular mass estimation is achieved by running a mixture of proteins of known molecular masses. For clarity, these calibration proteins are not shown, but their positions in the gel and their molecular masses are indicated by arrows on the right side of the gel.

A mixture of properly chosen proteins of known molecular mass is run in one of the lanes of the same gel. By using these control proteins, a calibration curve can be established based on the logarithm of their molecular mass and their relative mobility. A molecular mass value can thus be estimated for each band in the lane in which the my-

ofibrillar proteins were run. These molecular mass values allow for the assignation of the individual bands to individual proteins or subunits (as the molecular mass of these had already been identified).

# Chapter 8. Protein-ligand interactions

by József Kardos

## 8.1. Biomolecular interactions

All functions of living systems ranging from primitive bacteria to higher-order organisms, as well as their interactions with their environment, are realised through macromolecular interactions. These interactions might be simple or rather complex—with at least one of the partners being a biological macromolecule, usually a protein. In this chapter we briefly discuss the various types of interactions of proteins, their molecular background, theory and some practical applications.

Protein interactions can be classified into different groups regarding the molecular properties and functions of the interacting partners. (These groups are intertwined in several cases.) Some examples include:

- i. The interactions of proteins with other proteins, small molecules, carbohydrates, lipids or nucleic acids;
- ii. Receptor-ligand interactions;
- iii. Antigen-antibody interactions;
- iv. Enzymatic interactions, enzyme-inhibitor interactions.

In the following, these reactions will be denoted as protein-ligand interactions. Such interactions often play key roles in the initiation of complex multi-step reactions. As an example, ligand binding—such as that of steroid hormones to their cytoplasmic or nuclear receptors or the binding of secreted peptide ligands to transmembrane receptors—might activate a receptor molecule, which induces a cellular response.

## 8.2. Reaction kinetics

Factors that determine a particular protein-ligand interaction include the concentrations of the partners, the binding affinity, and the rate constants of association and dissociation (these are also discussed in Chapter 9). As a theoretical introduction, we will first provide a brief and simplified description of reaction kinetics.

Let us consider the following interaction:



The reaction rate will be the sum of the rates of product formation and the rate of the reverse reaction:

$$v = k_+ [A]^m [B]^n - k_- [C][D] \quad (8.2)$$

where  $k_+$  and  $k_-$  are the rate constants,  $m$  is the order of reaction with respect to A,  $n$  is the reaction order for B, and  $m+n$  provides the order of the overall reaction. A few simple general reactions are listed in Figure 8.1.

|               | reaction                              | reaction rate                | reaction order |
|---------------|---------------------------------------|------------------------------|----------------|
| monomolecular | $A \longleftrightarrow B$             | $k_+ [A] - k_- [B]$          | first order    |
| bimolecular   | $A + B \longleftrightarrow C + D$     | $k_+ [A][B] - k_- [C][D]$    | second order   |
| trimolecular  | $A + B + C \longleftrightarrow D + E$ | $k_+ [A][B][C] - k_- [D][E]$ | third order    |

Figure 8.1. Examples for reaction rates and reaction orders

In equilibrium, the overall reaction rate is zero, i.e. the rates of product formation and dissociation are equal and the concentrations of the reactants and products are constant, having equilibrium values:

$$k_+ [A]^m [B]^n = k_- [C][D] \quad (8.3)$$

The equilibrium dissociation (and association) constants can be defined as the ratio of the products of concentrations of the reacting molecules and the products of the concentrations of the resulting molecules, with the concentrations raised to the power reflecting the stoichiometry of the components:

$$K_D = k_- / k_+ = ([A]^m[B]^n) / ([C][D]) \quad (8.4)$$

The association constant is the reciprocal of the dissociation constant:

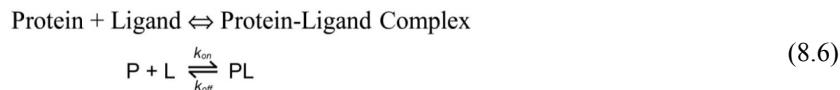
$$K_A = 1 / K_D \quad (8.5)$$

The equilibrium association and dissociation constants describe the extent to which the reaction is shifted towards the formation of the products. In the case of protein-ligand interactions (see below), these characterise the binding affinity between the components.

Far from the equilibrium state, the “on” or the “off” rate can be dominating. In such cases, the rate constants can be determined directly by rapid kinetic experiments.

## 8.3. Protein-ligand interactions

After the general discussion, we will take a simple second-order protein-ligand interaction model. From now on, we will denote the reactants by their initials,  $P$  and  $L$ , and their protein-ligand complex as  $PL$ :



where  $k_{on}$  and  $k_{off}$  are the rate constants of binding and dissociation, respectively.

The equilibrium association and dissociation constants are as follows:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}} \quad (8.7)$$

$$K_A = \frac{[PL]}{[P][L]} = \frac{k_{on}}{k_{off}} \quad (8.8)$$

where  $[P]$  is the concentration of free ligand molecules and  $[PL]$  is the concentration of the protein-ligand complex. In this simple reaction, the unit of the dissociation constant is concentration (M, mol/litre). The lower the dissociation constant, the stronger the binding. In biological systems, “tight binding” corresponds to a dissociation constant in the order of  $10^{-9}$  M (nM) or less. In some high-affinity enzyme-inhibitor interactions, such as the formation of the bovine trypsin inhibitor (BPTI)-trypsin complex, the  $K_D$  value is in the order of  $10^{-13}$  M (sub-picomolar).

## 8.4. Relationship between the free enthalpy (Gibbs free energy) change and the equilibrium constant

As a consequence of the second law of thermodynamics, at constant pressure and temperature, a reaction will spontaneously take place if the Gibbs free energy (free enthalpy) difference between the final state and the initial state is negative, i.e. if  $\Delta G < 0$ . During the approach to equilibrium, the free enthalpy decreases, reaching a minimum at equilibrium where  $\Delta G = 0$ . As a reminder, the definition of the Gibbs free energy is:

$$G = H - TS \quad (8.9)$$

where  $H$  is the enthalpy,  $T$  is the absolute temperature and  $S$  is the entropy. In solutions, the standard free enthalpy,  $G^\circ$ , is defined for a standard state in which the concentration of each component is 1 M. With respect to this reference value, the free enthalpy of the system of a solution of molecule  $A$  can be expressed in the form:

$$G = G^\circ + RT \ln a \quad (8.10)$$

where  $a$  is the activity (effective concentration),  $R$  is the universal gas constant, which has the value  $8.31447 \text{ J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$ . In dilute solutions,  $a$  approximately equals the molar concentration of molecule  $A$  ( $[A]$ ). Therefore:

$$G = G^\circ + RT \ln [A] \quad (8.11)$$

In solution phase, in the case of a reaction with several components, the free enthalpy change of the reaction can be expressed as:

$$\Delta G = \Delta G^\circ + RT \ln \frac{\prod[\text{products}]}{\prod[\text{reactants}]} \quad (8.12)$$

where, in the fraction, the products of the concentrations appear. In equilibrium,  $\Delta G = 0$  and the concentrations in the fraction are the equilibrium concentrations. Thus, the fraction gives exactly the equilibrium association constant. Therefore, the standard free enthalpy change can be determined directly from the equilibrium constant:

$$\Delta G^\circ = -RT \ln K_A = RT \ln K_D \quad (8.13)$$

It is worth noting that, in biochemical systems, we use aqueous solutions in which the standard state is defined at neutral pH (pH 7.0).

$\Delta G^\circ$  describes the stability of the products in a reaction, or the stability of the complex in a protein-ligand interaction in the equilibrium, relative to the reactants. Figure 8.2 shows the relationship between the dissociation constant and  $\Delta G^\circ$  for different binding affinities, compared to the case at  $K_D = 100 \text{ nM}$  (second line).

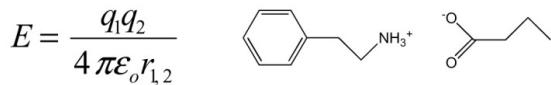
| Ligand | $K_D$ (nM) | $\Delta G^\circ$ (kJ/mol) | $\Delta \Delta G^\circ$ (kJ/mol) |
|--------|------------|---------------------------|----------------------------------|
| X      | 1000       | -34.2                     | 5.7                              |
| Y      | 100        | -39.9                     | 0                                |
| W      | 10         | -45.6                     | -5.7                             |
| Z      | 1          | -51.3                     | -11.4                            |

Figure 8.2. Comparison of  $\Delta G^\circ$  at different binding affinities at 25°C

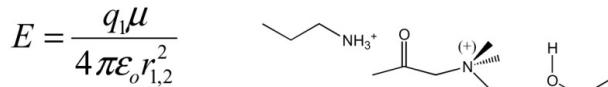
## 8.5. Molecular forces stabilising ligand binding

The strength of protein-ligand binding is usually determined by a variety of non-covalent interactions (Figures 8.3 and 8.4). These interactions differ from one another in their strength (binding energy) and their range. These interactions are as follows:

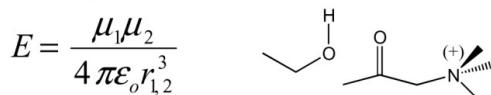
- Covalent bonding. The strongest interaction resulting irreversible binding. Energy:  $\sim -500 \text{ kJ/mol}$ .
- Ionic (Coulomb) interaction:  $-20 - -40 \text{ kJ/mol}$ , distance dependence:  $1/r$ .



- Charge-dipole interaction:  $-10 - -30 \text{ kJ/mol}$ ,  $1/r^2$



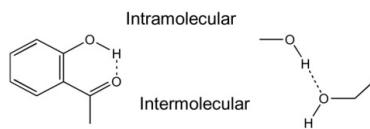
- Dipole-dipole interaction:  $-5 - -20 \text{ kJ/mol}$ ,  $1/r^3$



$q$ , charge;  $\mu$ , dipole moment;  $r$ , distance,  $\epsilon_0$ , dielectric constant

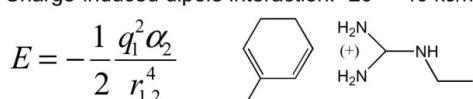
- Hydrogen bond:  $-4 - -30 \text{ kJ/mol}$

H-bond donor ... acceptor



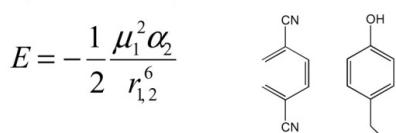
Hydrogen bond energy is largely dependent on the donor-acceptor distance, which is shorter 3,5 Å, and the orientation

- Charge-induced dipole interaction:  $-20 - -40 \text{ kJ/mol}$ ,  $1/r^4$

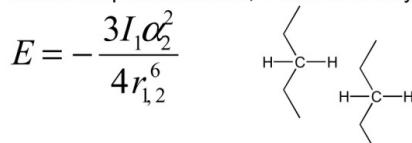


$q_1$ , charge;  $\alpha_2$ , polarizability;  $r_{1,2}$ , distance

- Dipole-induced dipole interaction:  $-5 - -30 \text{ kJ/mol}$ ,  $1/r^6$



- London-dispersion forces,  $-2 \text{ kJ/mol}/\text{methylene group}$

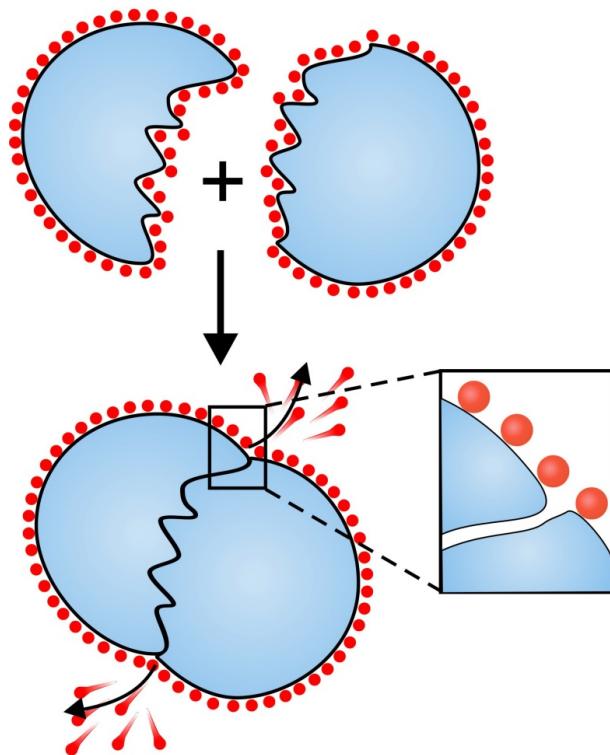


$\mu$ , dipole moment;  $\alpha$ , polarizability;  $I$ , ionizability;  $r$ , distance

In the case of clashing atoms a strong repulsion is observable, which is explained by the Pauli exclusion principle and is usually taken into account with  $1/r^{12}$  distance dependence.

Figure 8.3. Types of intermolecular interactions

## Hydrophobic effect



**Two hydrophobic surfaces come together to exclude water**

Figure 8.4. The hydrophobic effect

Beyond the above ones, the hydrophobic effect is a complex, noncovalent interaction in which the hydrophobic part of the molecule becomes buried from water and thus the water-accessible apolar surface area of the macromolecule decreases. The interaction is additive; its energy is roughly proportional to the buried surface area:  $\sim -80\text{-}100 \text{ J/mol}/\text{\AA}^2$ . This represents approximately  $-3 \text{ kJ/mol}$  energy for the burial of a methylene group.

The strong binding affinity and high specificity of protein-ligand interactions result from a high amount of weak noncovalent interactions. Figure 8.5 shows an enzyme-inhibitor complex with a dissociation constant of  $10^{-13} \text{ M}$ .

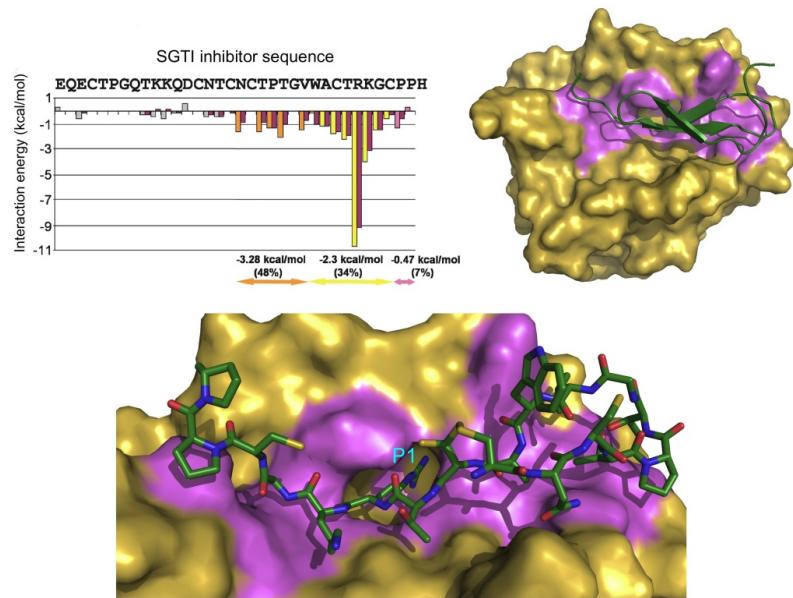


Figure 8.5. The interaction of a trypsin inhibitor with crayfish trypsin. Upper left panel: distribution of the interaction energies with bovine trypsin among the residues of the inhibitor. Upper right panel: 3D structure of the complex. Bottom panel: close-up view of the binding interface. Note the Arg side chain in the P1 position fitting in the substrate binding site of the protease.

## 8.6. Determination of the binding constant

Consider the ligand binding interaction described in section 8.3. The dissociation constant can be written by introducing total protein concentration  $[P]_T$  as follows:

$$P + L \rightleftharpoons PL \quad (8.14)$$

$$K_D = \frac{[P][L]}{[PL]} = \frac{([P]_T - [PL])[L]}{[PL]} \quad (8.15)$$

where  $[P]$ ,  $[L]$  and  $[PL]$  are the concentrations of the free protein, the free ligand and the protein-ligand complex, respectively, and  $[P]_T$  is the total protein concentration.

The above equation can be transformed by multiplying by  $[PL]$  and dividing by  $[L]$  and  $K_D$  to reach the following formula:

$$\frac{[PL]}{[L]} = \frac{[P]_T - [PL]}{K_D} \quad (8.16)$$

Note that this is a linear equation, i.e. the values  $y = [PL]/[L]$  (complex concentration over free ligand concentration) are located on a straight line as a function of  $x = [PL]$ . This is called the Scatchard plot (Figure 8.6, Equation 8.17).

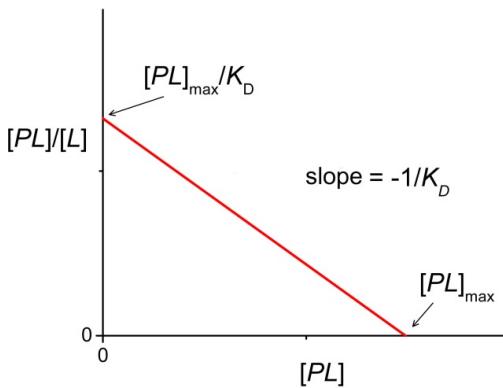


Figure 8.6. Scatchard plot for a simple binding reaction

$$y = -\frac{1}{K_D}x + \frac{[P]_T}{K_D} \quad (8.17)$$

Using the Scatchard plot, it is easy to determine the binding constant because the slope of the line is  $-1/K_D$ . The drawback of this representation is that it distorts experimental errors. It is rarely used nowadays because, using currently available computational capacity, we can easily perform non-linear fitting and, thus, linearisation is no longer a necessity.

In a general case when the number of binding sites is more than one, instead of  $[PL]$ , we use the fraction of the concentrations of the bound ligand and the protein in complex. In this case, the line intersects the  $x$  axis at the value of  $n$ , which is the number of binding sites on the protein (receptor) molecule. This way, the Scatchard plot is suitable for the determination of the number of binding sites.

Now let us rearrange Equation 8.16:

$$[PL] K_D = ([P]_T - [PL]) [L] \quad (8.18)$$

After another rearrangement:

$$[PL] (K_D + [L]) = [P]_T [L] \quad (8.19)$$

From this, we can obtain the saturation of the protein ( $[PL]/[P]_T$ ), which describes the fraction of the protein being in complex with the ligand:

$$\frac{[PL]}{[P]_T} = \frac{[L]}{K_D + [L]} \quad (8.20)$$

The equation describes a hyperbola. The saturation curve, as a function of the free ligand concentration, converges to 1, which is the asymptote of the hyperbola (see also Chapter 9). Figure 8.7 represents saturation curves representing different  $K_D$  values. It is worth noting that the value of  $K_D$  is equal to the free ligand concentration producing 50 % saturation.

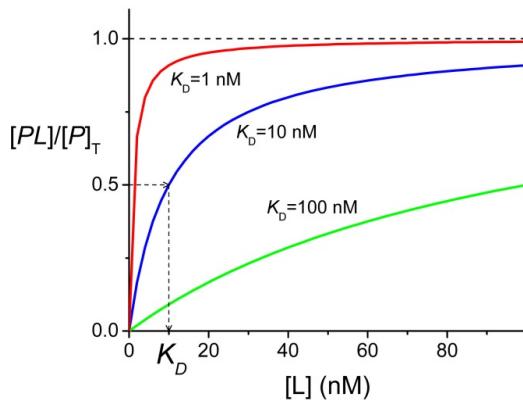


Figure 8.7. Hyperbolic saturation curves with different dissociation constants

The dissociation constant can be determined easily by fitting a hyperbola to the saturation curve. To obtain the saturation curve, we need to know the free (unbound) ligand concentration. Often, we can experimentally measure the bound ligand concentration ( $[PL]$ ), thus the free ligand concentration can be calculated from the total ligand concentration ( $[L]_T$ ) by subtraction:

$$[L] = [L]_T - [PL] \quad (8.21)$$

Then, the value of  $[L]$  will contain the experimental error of the determination of  $[PL]$ . Instead,  $K_D$  is often expressed using the total ligand concentration, which results in a second-order polynomial equation. The problem related to the determination of the free ligand concentration can be solved if the protein concentration is significantly less than both  $K_D$  and  $[L]_T$ :

$$[P]_T \ll K_D, [L]_T \quad (8.22)$$

In such a case, the number of molecules in the complex can be neglected compared to the total ligand concentration. Therefore, the free ligand concentration can be taken as equal to the total ligand concentration:

$$[L] \approx [L]_T \quad (8.23)$$

A classical biochemical example of a hyperbolic saturation curve is the reversible oxygen binding by myoglobin. Hemoglobin shows high similarity to myoglobin—however, it is a tetramer of four subunits with oxygen binding ability. Oxygen binding to hemoglobin is brought about by a cooperative process. The first oxygen molecule binds with low affinity, but induces an allosteric effect, which increases the binding affinity of further oxygen molecules and results in a sigmoid-like saturation curve. Figure 8.8 shows the oxygen saturation curves of myoglobin and hemoglobin as a function of the partial pressure of oxygen. The main role of hemoglobin is oxygen transport in the blood of vertebrates. By analysing the saturation curves, we can understand the mechanism of oxygen transport. In the capillary of the lung, the partial pressure of oxygen is approximately 100 Hgmm. At this pressure, hemoglobin is 97 % saturated. Red blood cells are transferred to various tissues where oxygen is consumed. In these tissues, the partial pressure of oxygen is no more than 40 Hgmm. At this pressure, only half of the hemoglobin subunits can bind oxygen. As a consequence, the saturated hemoglobin arriving from the lungs will release oxygen. Importantly, in tissues using more oxygen, the partial pressure will be lower, thus hemoglobin will release more oxygen. The system is thus efficiently controlled: oxygen is released where it is needed, and this occurs to the necessary extent. In contrast, myoglobin binds oxygen with high affinity even at lower partial pressures. Its function is probably the storage of oxygen, as indicated by the fact that it can be found at high concentrations in striated muscle tissue, especially in the muscles of sea mammals that spend long time periods underwater.

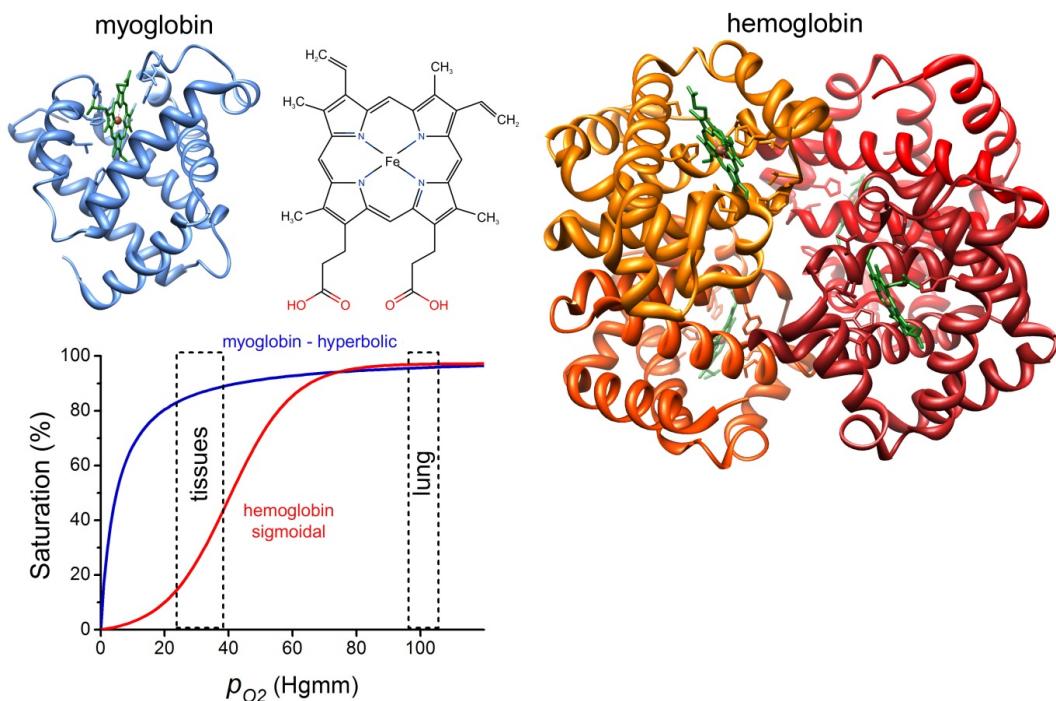


Figure 8.8. The structure of myoglobin, haem and hemoglobin and their reversible oxygen binding as a function of the partial pressure of oxygen

## 8.7. Methods for the experimental determination of the binding constant

Understanding the molecular mechanisms and interactions that govern various processes and reactions in living systems is a central issue in molecular biology, biochemistry, medicine and in the new fields of proteomics and genomics. To reach this goal, it is indispensable to identify the interacting partners and characterise their interactions. A variety of techniques are available for screening and measuring protein-ligand interactions. Methods include affinity chromatography, cross-linking, gel filtration, co-localisation, two-hybrid methods, spectroscopic methods, 3D structure determination, equilibrium dialysis, radioactive labelling, sedimentation velocity measurements, isothermal titration calorimetry, surface plasmon resonance, microarray, immunoblotting and ELISA.

Below we discuss some of the frequently used methods. The radioimmunoassay (RIA) is suitable for the determination of the concentration of various antigens, hormones and drugs in different body fluids such as the human blood. During the experiment, a known amount of a radioactively-labelled antigen is mixed into the solution of its corresponding antibody. Usually, gamma-radiating isotopes such as tyrosine-bound iodine are used. Subsequently, the unlabeled (“cold”) antigen sample is added to the solution and the liberation of the labelled antigen is measured. The unlabelled material competes with the labelled one and may “chase it off” from the antibody, depending on the concentrations. The material bound in complex is separated from the free antigen. This is a technique with high sensitivity; however, the handling of radioactive material needs special care. Nowadays, the ELISA technique is more popular, which measures the antigen-antibody interactions by applying colour reactions.

In ELISA (enzyme-linked immunosorbent assay), the antigen is bound to a surface and then recognised by a specific antibody. Subsequently, a second antibody is applied that is specific to the first one and usually carries a covalently-bound enzyme molecule. By the addition of the substrate of the enzyme, a colourful and/or fluorescent product will form. This way, the presence of the antigen—or even its concentration—can be determined. A more sophisticated variant of ELISA is the “sandwich” ELISA. In this case, a “capture” antibody is immobilised on the membrane to which the antigen binds in the second step. Then a second antibody is added that also binds the antigen but—similarly to the above procedure—it carries a covalently-bound enzyme. Thus a “sandwich” is being formed, and then detection is carried out by adding substrate. The sandwich ELISA is much more sensitive than the common ELISA technique. Pregnancy tests are usually based on this method.

In the following, we provide an introduction to modern methods for the detailed characterisation of protein-ligand interactions, including surface plasmon resonance (SPR), isothermal titration calorimetry and fluorescence depolarisation.

### 8.7.1. Surface plasmon resonance (SPR)

During a surface plasmon resonance experiment, one of the interacting partners is immobilised onto the surface of a chip and then a solution containing the other partner (analyte) is flown over the surface. The association and dissociation rate constants ( $k_{on}$  and  $k_{off}$ ) can be determined in the system described in detail below. The instrument measures the material (mass) bound to the surface upon the binding reaction by a spectroscopic phenomenon, the surface plasmon resonance. The chip is a glass slide whose surface is covered by a thin layer of metal, usually gold. Over this surface, micro-flow channels are constructed, making possible the controlled flow of solutions containing the interacting partners over the surface. Usually, the golden layer is covered with one or more further layers (such as dextran) to ensure the immobilisation of one of the partners, which usually occurs via covalent linkage. After successful immobilisation, the other component is applied in the solution flowing over the surface and can bind to the immobilised partner on the surface, forming the complex. Figure 8.9 depicts an SPR system. The chip is placed on a prism. During the measurement, the chip is irradiated from the bottom with a beam of a wide angle range within that of total internal reflection. The beam is reflected from the glass surface holding the golden layer and projected onto a detector. It is a known physical phenomenon that, during total internal reflection, the incident electromagnetic wave penetrates to the other side by an intensity exponentially decreasing with distance. This is the so-called evanescent field, which is localised in the metal layer and a narrow range above it. Importantly, it is sensitive to the material bound to the surface. The incident light at a certain angle interacts with the free electrons in the metal layer and their electromagnetic field, exciting them, leading to the phenomenon of surface plasmon resonance. The light at the resonance angle will be absorbed, thus the intensity is decreased. The detector will determine the angle of the intensity decrease. Upon binding of any material to the surface, the local refractive index will change and, through the evanescent field, it will have an effect on the plasmon waves, which results in the shift of the resonance angle. This shift is linearly proportional to the amount (mass) of the material bound to the surface. Its value is given in angle change or, in the case of some SPR instruments, in resonance units (RU). 1 RU = 1 pg/mm<sup>2</sup>—in other words, it is a 1-pg mass deposited per 1-mm<sup>2</sup> surface. An angle shift of 1 millidegree equals ~6 RU.

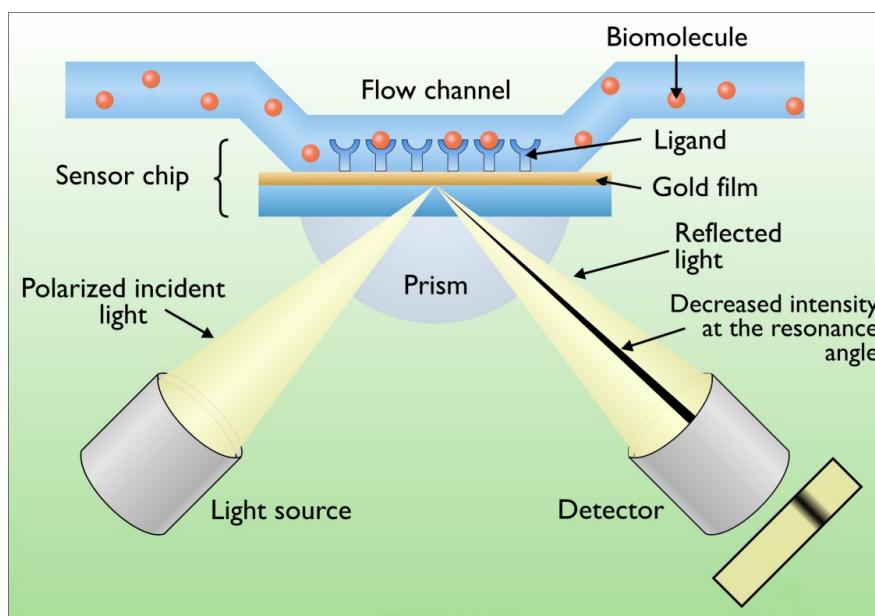


Figure 8.9. The method of surface plasmon resonance

The  $k_{on}$  association rate constant of binding can be determined by the kinetics of the SPR signal change. The  $k_{off}$  dissociation constant can be measured after binding, by washing the surface with ligand-free buffer. The ratio of the two will give the  $K_A$  equilibrium association constant *i.e.*, the binding affinity, and through this,  $\Delta G^\circ$ . Figure 8.10 shows the steps of an SPR measurement.

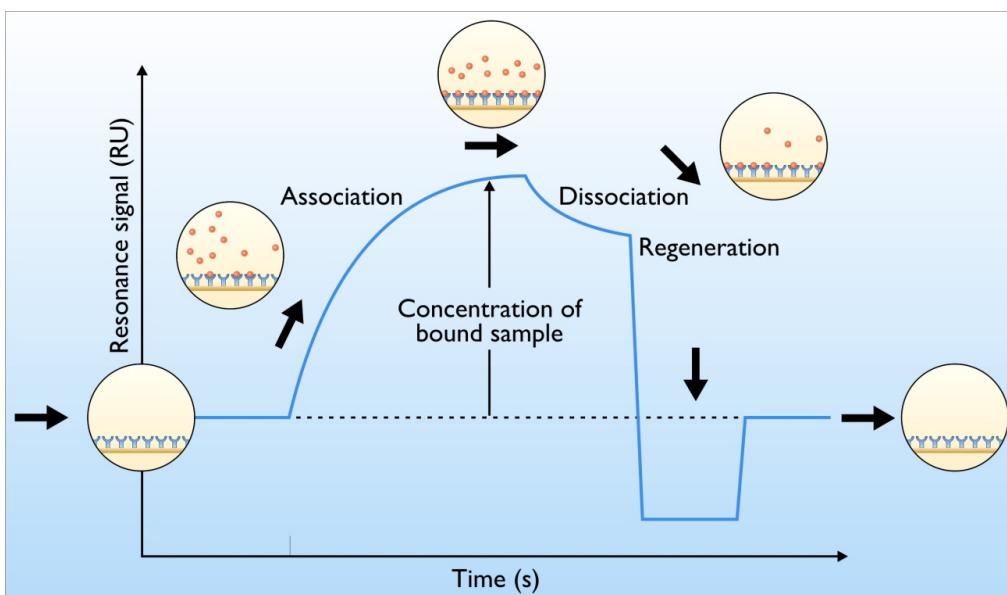


Figure 8.10. Representation of an SPR measurement

During immobilisation, the ligand is usually bound to the surface of the chip. The covalent link is very stable and usually does not require the modification of the ligand before attachment. However, the ligand must contain some reactive groups (such as  $-NH_2$ ,  $-SH$ ,  $-COOH$ ), and its binding orientation is not well-defined. Most often, thiol and amino groups are used for immobilisation; however, the streptavidin-biotin linkage is also popular. First, the dextran surface is activated by modifying its carboxyl groups with N-hydroxysuccinimide (NHS) and 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC). Subsequently, the ligand can be linked to the surface in the required concentration. Finally, the remaining free reactive surface can be blocked by injecting ethanolamine. The advantage of the SPR technique is that it does not require the modification or labelling of the molecules of interest, keeping their properties unchanged. The binding kinetics can be followed in real time, and the parameters of the interaction can be determined. A further advantage of the method is that it requires a low amount of material. Moreover, the chip can be regenerated with the bound ligand and can be reused several times, enabling the measurement of different binding partners. The turbidity of the solution does not interfere with the measurement. Although the partners do not need to be modified or labelled for the SPR measurement, the immobilisation can interfere with the binding reaction by orienting the binding region towards the chip surface or if the immobilisation occurs through the same groups as those involved in ligand interaction. This disadvantage of the technique may be a serious issue in some cases. To avoid such a problem, a longer linker molecule can be used for immobilisation. Another problem might arise from the condition that the binding interaction occurs on a surface instead of the solution phase. In the case of strong binding or high concentration of immobilised molecules, the dissociated binding partners will be present in an increased local concentration and may rebind immediately after dissociation. In this case, an artificially increased binding affinity will be observed. In case of multiple binding sites, the interpretation of the results might become complicated. It is worth noting that molecules with sizes of  $M_w < 300-500$  might be too small for a quantitative measurement.

## 8.7.2. Isothermal titration calorimetry (ITC)

Protein-ligand interactions can be measured by ITC in solution phase without the need of any immobilisation or modification. The number of binding sites, the equilibrium constant ( $K_a$ ), and the thermodynamic parameters including enthalpy, free enthalpy, entropy and the partial specific heat capacity ( $\Delta H$ ,  $\Delta G^\circ$ ,  $\Delta S^\circ$ ,  $\Delta C_p$ ) can be determined directly via the measurement of the reaction heat. (As far as the thermodynamic parameters are concerned, all other methods, such as the spectroscopic methods, are indirect: calculations are based on spectral or other differences between the complex and the free ligand.) In the ITC measurement, the turbidity and the absorption of the solution are indifferent. Another advantage of the technique is its cheap maintenance cost. The evaluation of the data is simple and computer-controlled. Compared to SPR, the disadvantage of ITC is that it needs a higher amount of material and, sometimes, solubility problems make it difficult to establish suitable experimental conditions.

The calorimeter has two cells: a reference and a sample cell. The reference cell is filled with water. The sample cell contains the protein or ligand solution, which is titrated stepwise with its binding partner through a computer-controlled injector. During the experiment, the reference cell is heated with a small power, which does not increase the temperature of the cell significantly during the measurement. The essence of the measurement is that the temperature of the sample cell is kept the same as that of the reference cell with high accuracy. For this purpose, a variable heating power should be introduced on the sample cell, driven by feedback depending on the studied reaction. In the absence of reaction, the sample cell needs to be heated with a power similar to that of the reference cell. In the case of an endothermic reaction that consumes heat, the sample cell needs to be heated with higher power than the reference. When the reaction is exothermic, less heating power is needed to keep the temperature constant. Small aliquots of the titrant are injected into the cell at defined time intervals, and the reaction heat is measured. The reaction heat is proportional to the bound fraction of the injected molecules. At the beginning, a larger fraction of is bound, whereas no binding occurs upon injection when the binding sites have become saturated at the end of the titration.

The measured heat will be:

$$Q = V_0 \Delta H_b [M]_t K_a [L] / (1 + K_a [L]) \quad (8.24)$$

where  $V_0$  is the cell volume,  $\Delta H_b$  is the enthalpy change corresponding to the binding of one mole of ligand,  $[M]_t$  is the total macromolecule concentration in the cell,  $K_a$  is the association constant, and  $[L]$  is the free ligand concentration. In the case of multiple binding sites, the equation is summed up for all binding sites.

The accurate determination of concentration is of high importance in ITC. After subtracting the baseline (which is the heating power with no reaction), the areas under the observed peaks give the enthalpy changes that belong to the reactions occurring upon injection steps. Figure 8.11 shows a schematic representation of the calorimeter and a characteristic titration profile with its evaluation. It is very important to have the protein and the injected ligand exactly in the same buffer solution at identical pH to avoid a large dilution and ionisation enthalpy. To account for such effects, control measurements should be carried out by injecting buffer to buffer, titrant to buffer, and buffer to the protein solution. The measured enthalpy effects can be used for correction.

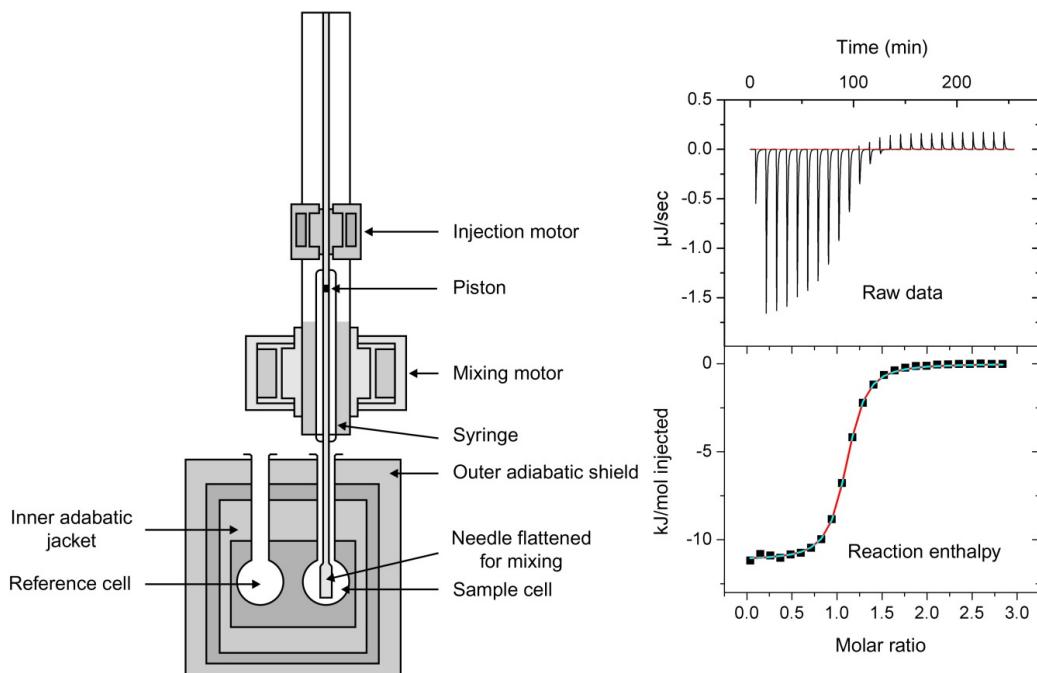


Figure 8.11. Schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment (upper right) with its evaluation (lower right)

The sensitivity of the ITC instrument is extremely high: heat changes in the range of  $10^{-8}$  W can be measured. Association constants can be determined accurately in the range of  $K_a = 10^2 - 10^9 \text{ M}^{-1}$ .

### 8.7.3. Fluorescence depolarisation to characterise protein-ligand binding interactions

In fluorescence depolarisation (FD) experiments, the sample is excited by a plane-polarised light, and the polarisation level of the emitted light is measured by recording the intensity in two polarisation planes: one parallel to that of the exciting light and the other being perpendicular (see also Chapter 4). Polarisation is defined by the measured intensities in the parallel and perpendicular polarisation planes:

$$P = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + I_{\perp}) \quad (8.25)$$

The level of polarisation is often expressed in the form of another related parameter, the anisotropy:

$$A = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2I_{\perp}) \quad (8.26)$$

When a fluorescent group is oriented and rigid, i.e. not moving, the emitted light will be polarised. In case the group is moving and/or rotating, the polarisation level of the emitted light is decreased.

The absorption of light by a chromophore occurs extremely rapidly in the order of  $10^{-15}$  second (femtosecond). In contrast, fluorescence emission is a substantially slower process in the order of  $10^{-8}$  second (10 nanoseconds). When the sample is excited by a plane-polarised light, the movement and rotation during absorption is negligible. During the longer time necessary for emission, the molecules might move and rotate—making the emitted light more depolarised. Large molecules move slower in solution; thus, they emit more polarised light, while small molecules and rapidly rotating fluorophores show higher degree of depolarisation.

The size and shape of the molecular complex formed in a protein-ligand interaction is different from that of the individual partners. Therefore, their diffusion and mobility will also differ. The local mobility of a fluorophore can be altered inside a molecule upon complexation if the packing of their environment is changed or they participate in interactions. All of these conditions affect the level of depolarisation, which can be measured easily and thus used for the characterisation of protein-ligand binding.

The fluorescent groups or probes applied in FD measurements might be natural inner chromophores such as tryptophan and tyrosine, NADH and FAD. Synthetic molecules can be used to label, modify or substitute amino acid residues. External fluorescent probes exhibit fluorescence upon binding to target molecules.

Fluorescence depolarisation can be measured in steady-state or time-resolved modes. In steady-state measurements, the fluorescence intensities are measured in the parallel and perpendicular polarisation planes at constant excitation. In time-resolved mode, the fluorescence lifetime or anisotropy decay is measured after excitation with a short pulse. This is an accurate measurement to determine the translation and rotation of fluorophores.

Fluorescence depolarisation is a simple and rapid technique to measure protein interactions in solution phase. Using inner fluorophores, we can measure without modification of the protein molecules; however, we often need to use fluorescent labelling, which needs circumspection because it might change the properties of the system and can interfere with ligand binding. Site-specific information can be obtained by linking the fluorescent label to an appropriate group of the molecule.

## 8.8. Test questions and problems

1. What is the definition of  $K_D$  and what is its unit in a simple binding reaction?
2. Describe a second-order reaction.
3. What will determine the direction of a reaction?

4. What is the relationship between the free enthalpy and the binding constant?
5. What kind of noncovalent interactions contribute to the binding affinity?
6. What kind of techniques can be used to study protein-ligand interactions?
7. Describe the receptor-ligand saturation curve.
8. What are the methods for the graphical determination of  $K_D$ ?
9. What percentage of the receptor molecules is in complex if the free ligand concentration equals  $3K_D$ ?
10. By what factor will the stability ( $\Delta G$ ) change if  $K_D$  decreases by three orders of magnitude?
11. In the case of small ligand molecules, which of the rate constants ( $k_{on}$  or  $k_{off}$ ) should be altered to increase binding affinity? In which direction should the rate constant be changed?
12. Explain the differences between oxygen binding by myoglobin and hemoglobin.
13. Compare the advantages and disadvantages of the SPR, ITC and FD techniques.

---

# Chapter 9. Enzyme kinetics

by Gábor Pál

One of the common fundamental characteristics of all living organisms is that thousands of chemical reactions occur in them at relatively low temperature. Moreover, these reactions happen at very high rates and in a highly regulated manner.

High rates and regulation have a common origin: these reactions are catalysed by enzymes, which are mostly proteins and, in some important cases, RNA. Enzymes, either protein- or RNA-based, are all macromolecules. These macromolecules bind to the interacting chemical compounds in a highly specific manner using large complex binding sites. Regulation of enzyme-catalysed reactions is primarily achieved by direct regulation (activation or inactivation) of the enzyme.

The discipline of enzyme kinetics studies—among many things—how the rate of enzyme-catalysed reactions depends on the concentration of the compounds directly interacting with the enzyme and what is the highest rate achievable by the enzyme. Further questions of interest include how the rate of the catalysed reaction depends on the temperature, pH, ionic strength etc. of the medium. All related studies provide pieces of information that serve as input data to establish a mechanistic model of the enzymatic reaction.

Enzyme kinetics, as its name implies, studies primarily the rates of reactions and all factors that affect these rates. However, before introducing fundamental rate equations of enzyme kinetics, we need to review how the rate-increasing capacity of enzymes can be interpreted in the framework of thermodynamics.

## 9.1 Thermodynamic interpretation of enzyme catalysis

Let us first examine what determines the rate of a *non-catalysed* chemical reaction. The reaction scheme described in Equation 9.1 corresponds to a chemical reaction leading to equilibrium. The reacting compounds are denoted as A and B, while P is the product of the reaction.



In the above reaction, as long as only A and B are present (no product has yet been generated), the reaction rate can be described as shown in Equation 9.2. “V” is the rate (velocity) of the reaction, square brackets indicate molar concentrations, whereas “t” denotes time:

$$V = \frac{d[P]}{dt} = k[A][B] \quad (9.2)$$

In order to react, the two molecules need to encounter each other and collide. It is therefore intuitively comprehensible that the rate of the reaction will be linearly proportional to the concentration of both molecules. The factor of proportionality, “k” is denoted as the reaction rate constant. This constant is a central parameter of the reaction. What is the deeper meaning of “k” and what determines its value?

The simplest mechanistic model that successfully interpreted the meaning of “k” is illustrated in Equation 9.3:



The interpretation of Equation 9.3 is as follows. Let us suppose that the reaction between A and B is instantaneous and results in a compound  $X^*$  called the transition state. (Note that, while the unusual name describes a “state”,  $X^*$  is in fact an actual albeit very short-lived compound.) While compounds A and B are in ground state,  $X^*$  is in an activated state, i.e. it is at a higher free enthalpy level. Let us also suppose that this first reaction step leads to

equilibrium (or, rather, quasi-equilibrium) between the reactants and the transition state. In this model, the transition state will convert into the product in a separate step. The equilibrium constant (importantly, this is *not* a rate constant!) is denoted as  $K^*$ , while the *rate constant* of the  $X^* \rightarrow P$  reaction step is defined as  $k'$ . The reaction scheme of Equation 9.3 will thus lead to rate Equation 9.4:

$$V = \frac{d[P]}{dt} = k[A][B] = k'[X^*] \quad (9.4)$$

Ultimately, the rate of product formation is the product of the concentration of  $X^*$  and the newly introduced rate constant,  $k'$ . Introduction of a new rate constant introduces additional complexity but, as we will see, it will eventually simplify the case.

Let us now examine what determines the concentration of  $X^*$ . This will lead us to the topic of thermodynamics. As shown in Equation 9.5, the equilibrium between the reactants and the transition state can be defined by the  $K^*$  equilibrium constant:

$$K^* = \frac{[X^*]}{[A][B]} \quad (9.5)$$

In thermodynamics, the equilibrium constant can be interchangeably defined by the standard free enthalpy change of the reaction leading to the given equilibrium. For the given case, it is shown in Equation 9.6:

$$-RT \ln K^* = \Delta G^* \quad (9.6)$$

In Equation 9.6,  $\Delta G^*$  defines the free enthalpy difference between the activated state of compound  $X^*$  and the ground state of the reactants. A more familiar term for the  $\Delta G^*$  parameter is as follows: it is the activation free enthalpy of the chemical reaction. As we will see, it is important that  $\Delta G^*$  corresponds to a free enthalpy change. Yet, for the sake of brevity, it is often referred to—rather incorrectly—as “activation energy”.

From Equation 9.6,  $K^*$  can be expressed by two simple algebraic transformations according to Equations 9.7 and 9.8:

$$\ln K^* = \frac{-\Delta G^*}{RT} \quad (9.7)$$

$$K^* = e^{\frac{-\Delta G^*}{RT}} \quad (9.8)$$

Equations 9.8 and 9.5 lead to Equation 9.9 that ultimately results in Equation 9.10:

$$V = k'[X^*] = k'(K^*[A][B]) = k'e^{\frac{-\Delta G^*}{RT}} [A][B] \quad (9.9)$$

$$V = k'e^{\frac{-\Delta G^*}{RT}} [A][B] \quad (9.10)$$

Let us summarise where we have got so far. The rate of the reaction is proportional to the concentration of the reactants and also to the  $k'$  rate constant. It is also a function of the  $\Delta G^*$  activation free enthalpy. As  $\Delta G^*$  represents an activated state, its value cannot be negative. If  $\Delta G^*$  is 0 (no activation barrier exists), the value of  $e^{-\Delta G^*/RT}$  will be 1. The larger positive number the activation free enthalpy has, the lower positive value the  $e^{-\Delta G^*/RT}$  factor will have, the latter approximating zero. The value of the  $e^{-\Delta G^*/RT}$  factor can therefore be between 0 and 1.

This can be interpreted by several similar phrasings of the physical meaning of this factor. The  $e^{-\Delta G^*/RT}$  factor in the  $e^{-\Delta G^*/RT} [A][B]$  product provides a measure of the fraction of A-B collisions that actually lead to the generation of  $X^*$ . If this factor is 1, all collisions will generate a transition state. If this factor is 0, none of the collisions will generate  $X^*$ .

In the first case, the reaction rate is maximal, while in the second case it is, naturally, zero. In general, the higher the reaction-specific activation free enthalpy, the lower the rate of the reaction.

Moreover, the above model provides a comprehensible explanation of the physical meaning of the  $k'$  rate constant as follows. In order for the transition state to convert into product, at least one chemical bond must be broken. The  $k'$  rate constant describes the rate, i.e. frequency, of the breakage of the chemical bond.

This has been put in mathematical terms in Equation 9.11 in which “ $\nu$ ” (Greek letter nu) denotes the vibration frequency of the chemical bond to be broken and “ $\kappa$ ” (Greek letter kappa) provides a measure of the proportion of  $X^*$  decaying forward (generating product) versus decaying backwards (regenerating the reactants). If the value of kappa is 0.5, half of the decays will generate product, whereas the other half will generate reactants. If the value of kappa is 1,  $X^*$  will decay exclusively towards the product.

$$k' = \kappa \nu \quad (9.11)$$

For simplicity, let us consider the case when the value of kappa is 1. The energy of chemical bond vibration is described in Equation 9.12, according to Planck's law:

$$\nu = \frac{E}{h} \quad (9.12)$$

In the above equation,  $E$  is the energy of the vibrating bond, while  $h$  is the Planck constant. The same energy can be expressed using Equation 9.13 in which  $k_b$  is the Boltzmann constant and  $T$  is the temperature on the Kelvin scale. This describes the average kinetic energy of molecule  $X^*$  at the given temperature:

$$E = k_b T \quad (9.13)$$

By combining Equations 9.11-9.13 and keeping the value of kappa at 1, Equation 9.14 is obtained:

$$k' = \frac{k_b T}{h} \quad (9.14)$$

Combining Equations 9.4, 9.10 and 9.14 leads to Equations 9.15 and 9.16. The latter equations ultimately explain the dependence of the rate of non-catalysed chemical reactions on all already mentioned factors:

$$k = \frac{k_b T}{h} e^{\frac{-\Delta G^*}{RT}} \quad (9.15)$$

$$V = \frac{k_b T}{h} e^{\frac{-\Delta G^*}{RT}} [A][B] \quad (9.16)$$

As evident, the rate constant “ $k$ ” of the reaction is inversely proportional to the reaction-specific activation free enthalpy and is a complex function of the temperature. Temperature appears twice in the equation: it is present in the pre-exponential term as a multiplying factor and it is also present in the denominator of the exponential term. Both terms indicate that raising the temperature will increase the rate of the reaction. In the pre-exponential term, which corresponds to the  $k'$  rate constant, a higher temperature results in a higher vibration frequency of the

chemical bond to be broken. In the exponential term, a higher temperature corresponds to a larger proportion of transition state-generating reactant collisions.

Based on Equation 9.4, the rate of the chemical reaction is a linear function of the rate constant and the concentration of the reactants.

The thermodynamic model of the chemical reaction rate is illustrated in Figure 9.1.

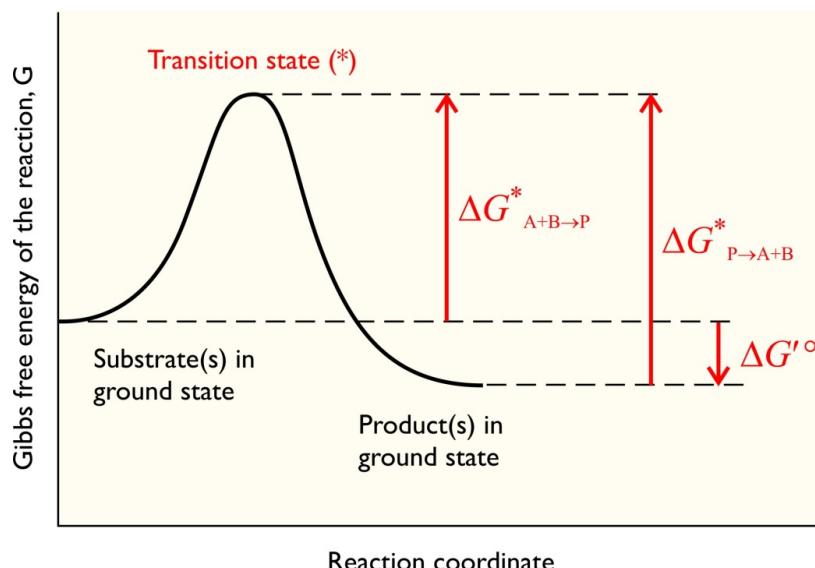


Figure 9.1. Thermodynamic description of the rate of the chemical reactions

Figure 9.1 provides a clear graphical illustration of the centrally important fact that the equilibrium proportion of the reactants described by the standard free enthalpy change ( $\Delta G^\circ$ ) of the chemical reaction is completely independent of the rate of the reaction expressed by the activation free enthalpy ( $\Delta G^*$ ) term.

The thermodynamic description of the rate of chemical reactions unequivocally points out how enzymes (and all catalysts) can increase the rate of chemical reactions: they do so by *decreasing the activation free enthalpy of the chemical reaction* (Figure 9.2).

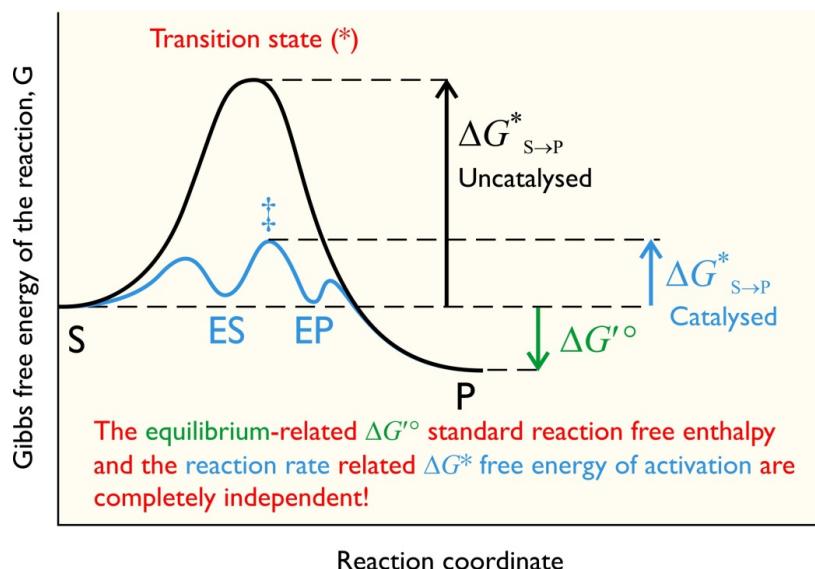


Figure 9.2. Enzymes increase reaction rates by decreasing the activation free enthalpy of a chemical reaction  $S \rightarrow P$  ( $S$ , substrate;  $P$ , Product;  $ES$  and  $EP$  are their complexes with the enzyme, respectively).

This statement does not tell much about the molecular mechanism of action of enzymes. The question is how, in general, the  $\Delta G^*$  term can be decreased. To answer this, let us recall that free enthalpy can be decomposed to two different terms: an enthalpy-related one and an entropy-related one, as shown in Equation 9.17:

$$\Delta G^* = \Delta H^* - T\Delta S^* \quad (9.17)$$

An unfavourably high  $\Delta H^*$  decreases the rate of the reaction, as it corresponds to a high-energy transition state. This can be due, for example, to a highly unfavourable electron distribution in the transition-state compound, which represents a high potential energy state. As a consequence of this, the proportion of the transition-state reactants to the ground-state ones will be low. An enzyme can decrease  $\Delta H^*$  by binding to—and thus stabilising—the transition state, providing attracting interactions through a proper binding site. Accordingly, the enzyme must possess a binding pocket that is spherically and electrostatically complementary to the transition state.

A negative  $\Delta S^*$  term is also unfavourable and thus decreases the reaction rate. Such a negative term corresponds to a more ordered transition state than ground state. This means that, by moving from the ground state towards the activated state, the system must get more ordered, which comes with an entropy-related energetic price, sometimes also called penalty, expressed in the  $-T\Delta S^*$  term. Enzymes can increase the rate of chemical reactions by positioning and orienting the reactants into an arrangement that is optimal for the reaction.

## 9.2. Michaelis-Menten kinetics

After the thermodynamic interpretation, we will now review the fundamental relationships of enzyme kinetics. As we focus on enzymatic reactions, the reactant will be denoted as substrate and will be referred to as “S”. The product will be denoted as P as previously. For simplicity, instead of the previous  $A + B \rightarrow P$  scheme, we will introduce the simplest  $S \rightarrow P$  case.

When the topic of enzyme kinetics first emerged, almost nothing was known about the physical nature of enzymes and the possible mechanisms of rate enhancement.

Let us start with a thought experiment considering the dependence of the rate of a non-catalysed chemical reaction as a function of reactant concentration. In the case of the simplest first-order reaction, the rate of the non-catalysed  $S \rightarrow P$  reaction can be written as  $V = d[P]/dt = k[S]$ . In other words, the rate of the reaction is linearly proportional to the concentration of the reactant S. In principle, the rate could be increased to “infinity”—the only limit would be set by the solubility of S.

In typical cases, when an enzyme catalyses the same reaction, the rate is enhanced by orders of magnitude. However, very importantly, the plot of the  $[S]$ -V function would be principally different in this latter case. At a constant enzyme concentration, [E], and in the range of low  $[S]$  values, increasing  $[S]$  would result in an almost linear increase of the rate, V. However, as  $[S]$  is increased even further, V would not increase to the same extent and it would ultimately approximate a maximal value limit (Figure 9.3).

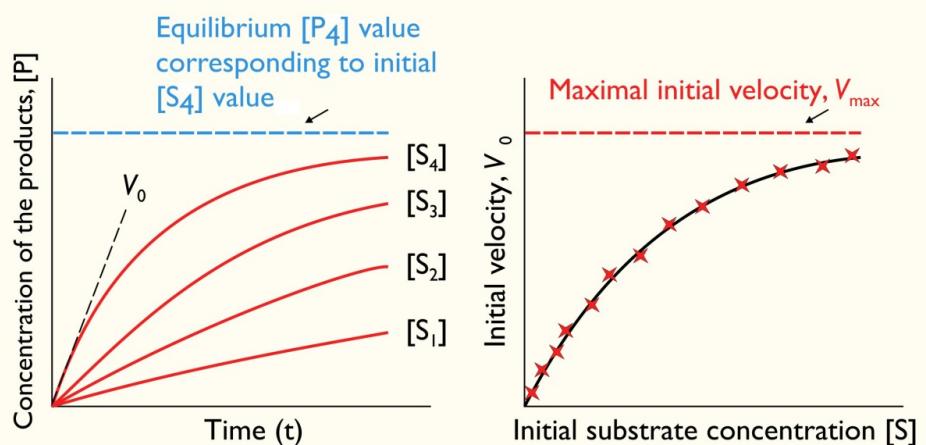


Figure 9.3. Initial rates ( $V_0$ ) in enzyme kinetic experiments

The first kinetic model that successfully explained this phenomenon was introduced by Leonor Michaelis and Maud Menten. Their presumption, which nowadays might seem trivial, was revolutionary in their time. They assumed that the enzyme directly interacts with the substrate in a stoichiometric manner, the interaction results in a well-defined intermediate complex, and the interaction leads to thermodynamic equilibrium. This scheme is illustrated by Equation 9.18 in which ES denotes the complex. As a tribute to this first successful model, ES has been named the Michaelis complex.



The above simplest scheme is based on the following assumptions. The interaction between the substrate and the enzymes that generates the ES complex leads to a (quasi-)equilibrium; and the reaction is instantaneous, i.e. so fast that the rate constants corresponding to this step do not restrict the overall rate of the reaction. Accordingly, for this first reaction step, the simple model introduces only an equilibrium constant,  $K_S$ , and it is not concerned with the two rate constants that determine  $K_S$ .  $K_S$ —which, in the case of non-covalent E-S interaction, is a dissociation constant—is defined by Equation 9.19:

$$K_S = \frac{[E][S]}{[ES]} \quad (9.19)$$

According to this model, the catalytic rate constant,  $k_{cat}$  that corresponds to the rate of the decomposition of ES towards the product, is much lower than the (non-defined) rate constants corresponding to  $K_S$ . Accordingly, the rate of ES decomposition towards the product is so low that (at least in the time frame of the measurement) it does not affect the quasi-equilibrium concentrations of [E], [S] and [ES].

Let us see how the initial rate of the reaction depends on substrate concentration if the starting assumptions apply.

The rate equation corresponding to the scheme introduced in Equation 9.18 is shown in Equation 9.20. This equation refers to a first-order reaction in which the rate of the reaction is proportional to the concentration of only a single entity, in this case the ES complex:

$$V_0 = k_{cat}[ES] \quad (9.20)$$

The scheme does not consider the opposite reaction, i.e. the one in which the interaction of the enzyme and the product would regenerate the ES complex. This is because the model focuses on the very beginning of the reaction when the concentration of the product is negligible. It is therefore of utmost importance that the rates defined in this model are always initial rates that correspond to the (theoretical) zero time point of the reaction.

The concentration of ES is not pre-set by the experimenter, but it can be determined experimentally once the right model is established. In the following steps, we will transform Equation 9.20 into a derived one that contains pre-set enzyme and substrate concentration parameters. To do so, we need to express ES concentration as a function of the pre-set enzyme and substrate concentrations. Let us start with Equation 9.19 that defined the  $K_S$  dissociation constant. Then let us consider a self-explanatory relationship shown in Equation 9.21, between the total enzyme concentration, ( $[E]_T$ ), the free enzyme concentration and ES concentration:

$$[E] = [E]_T - [ES] \quad (9.21)$$

By combining Equation 9.21 with Equation 9.19, we get equation 9.22:

$$K_S = \frac{[E][S]}{[ES]} = \frac{([E]_T - [ES])[S]}{[ES]} = \frac{[E]_T[S] - [ES][S]}{[ES]} \quad (9.22)$$

Multiplying both sides by ES concentration yields equation 9.23:

$$K_s[ES] = [E]_T[S] - [ES][S] \quad (9.23)$$

In Equation 9.24, the ES-containing terms are rearranged to be side by side:

$$K_s[ES] + [ES][S] = [E]_T[S] \quad (9.24)$$

Then, in Equation 9.25, [ES] is multiplied out from the sum of the products:

$$[ES](K_s + [S]) = [E]_T[S] \quad (9.25)$$

Finally, both sides are divided by the multiplying factor of [ES], which results in Equation 9.26:

$$[ES] = \frac{[E]_T[S]}{(K_s + [S])} \quad (9.26)$$

By these algebraic transformations, ES concentration has been expressed as a function of the experimentally preset enzyme and substrate concentrations and that of the equilibrium (dissociation) constant.

If the starting conditions are set such that the total substrate concentration exceeds the total enzyme concentration by orders of magnitude, the amount of substrate getting into the ES complex will be negligible compared to the total amount of substrate. Consequently, the free substrate concentration (at the beginning of the reaction) will practically equal the total substrate concentration. This way, both the total enzyme concentration,  $[E]_T$ , and the free substrate concentration,  $[S]$ , will be experimentally-set known parameters.

In the next step, based on Equation 9.20, both sides of Equation 9.26 are multiplied by the  $k_{cat}$  rate constant to yield the initial rate, according to Equation 9.27:

$$V_0 = k_{cat}[ES] = \frac{k_{cat}[E]_T[S]}{(K_s + [S])} \quad (9.27)$$

Note that the highest achievable initial reaction rate, denoted as  $V_{max}$ , will be achieved when all enzyme molecules are incorporated into the ES complex. In this case, the substrate saturates the enzyme molecules. Then, and only then,  $[ES] = [E]_T$ . According to this, the  $k_{cat} [E]_T$  product in Equation 9.27 will be in fact the value of  $V_{max}$ . By taking this into account, we can formulate Equation 9.28, which is the final equation of the simplest enzyme kinetic model:

$$V_0 = \frac{V_{max}[S]}{(K_s + [S])} \quad (9.28)$$

This equation is a so-called rectangular hyperbola function that has the following general description:  $Y = P_1 X / (P_2 + X)$ , where  $X$  is the independent variable, in our case the substrate concentration,  $Y$  is the dependent variable, in our case the initial reaction rate, while  $P_1$  and  $P_2$  are the two parameters of the function, in our case the  $V_{max}$  and the  $K_s$ , respectively.  $P_1$  is also the horizontal asymptote of the hyperbola, the maximal value of  $Y$  that the graph of the function approaches as  $X$  tends to infinity. The  $P_2$  parameter is  $K_s$ .

Note that this equation is in a perfect accordance with the experimental observations regarding the  $[S]$ - $V_0$  relationship illustrated in Figure 9.3: when  $[S] \ll K_s$ ,  $[S]$  becomes negligible in the denominator, leading to Equation 9.29:

$$V_0 = \frac{V_{\max}}{K_S} [S] \quad (9.29)$$

In Equation 9.29, the multiplying factor of the substrate concentration is the quotient of two constants and, as such, it is also a constant. Accordingly, in the  $[S]$  range where the substrate concentration is orders of magnitude lower than the value of  $K_S$ , the initial reaction rate will be linearly proportional to substrate concentration, exactly as the experiments show. In other words, in this substrate concentration range the reaction is a (pseudo)-first order reaction in respect of the substrate.

In the other extreme case when  $[S] \gg K_S$ , it is  $K_S$  that will be negligible compared to the value of  $[S]$  in the denominator of Equation 9.28. Accordingly, if we consider only this  $[S]$  range, we get equation 9.30:

$$V_0 = \frac{V_{\max} [S]}{[S]} = V_{\max} \quad (9.30)$$

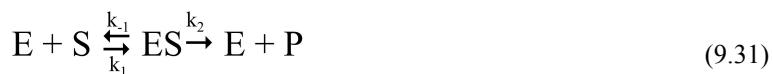
In the substrate concentration region where  $[S]$  exceeds  $K_S$  by orders of magnitude, the substrate saturates the enzyme, i.e. all enzyme molecules will be in the ES complex, and the initial reaction rate reaches a maximal value, i.e. it cannot be further increased by increasing the concentration of the substrate. Accordingly, in this substrate concentration range, the initial reaction rate is practically independent of the substrate concentration, (i.e. it is zero-order in respect of the substrate). This is why the  $[S]$ - $V_0$  plot illustrated in Figure 9.3 is often referred to as saturation curve.

When the substrate concentration equals the value of  $K_S$ , the initial reaction rate is the half of the  $V_{\max}$  value.

In a simple descriptive way, the deduced Equation 9.28 appears to be in accordance with the observations. Yet, it leads to serious theoretical contradictions. The more efficient the enzyme, the less rational the initial assumptions of the above model. If the enzyme is highly efficient, the rate of ES conversion into product should be very high. If so, that process should interfere with the presumed quasi-equilibrium between the enzyme, the substrate and the ES complex. Therefore, the assumption of equilibrium for the first step of the reaction renders the model ill-suited for describing the action of genuinely efficient enzymes.

The second—and equally significant—problem is that this first model also contradicts the thermodynamic bases of catalysis.  $K_S$  is a dissociation constant and, as such, it defines the affinity, i.e. the strength of the binding interaction. In the first model,  $K_S$  describes how strongly the enzyme binds the substrate or, in other words, how stable the ES complex is. The lower the  $K_S$ , the more stable the complex. Moreover, based on the model, the lower the  $K_S$ , the more effective the enzyme. This is because a low  $K_S$  means that the enzyme reaches half-maximal reaction rate at low substrate concentration. But there is a discrepancy here. By increasing the stability of the interaction between the enzyme and the substrate, the reaction rate should decrease because the enzyme would stabilise the substrate in the ground state. Naturally, the enzyme must bind the substrate, but it should not bind it too tightly. Instead, the enzyme should bind tightly the transition state, thereby decreasing the activation free enthalpy of the reaction.

Due to these contradictions, the first kinetic model had to be developed further. The improved model accounts for the two additional kinetic rate constants that were left out from the first model. One of these, denoted as  $k_1$ , corresponds to the formation of the ES complex from free enzyme and substrate. The other, denoted as  $k_{-1}$ , corresponds to the reverse reaction, the dissociation of the ES complex towards enzyme and substrate. The improved scheme is illustrated in Equation 9.31:



According to the improved model, when the solutions of the enzyme and the substrate are mixed, instead of a dynamic equilibrium, a steady-state will develop very quickly, almost instantaneously. Depending on the exact initial conditions, the steady-state may last long, i.e. the concentration of the ES complex can remain practically constant

for a long period of time (exactly,  $d[ES]/dt$  is not zero, but much smaller than either  $d[P]/dt$  or  $d[S]/dt$ ). The steady-state requires equal rates for ES generation and ES decomposition. This is illustrated in Figure 9.4. The reaction is triggered by adding the substrate. (Proportions of the figure are not realistic, as otherwise various parts of the figure could not be shown on the same page. In reality, the steady-state can be reached in milliseconds and it can last for minutes. Moreover, in reality,  $[S]$  starts from a much higher level because  $[S] \gg [E]$ .)

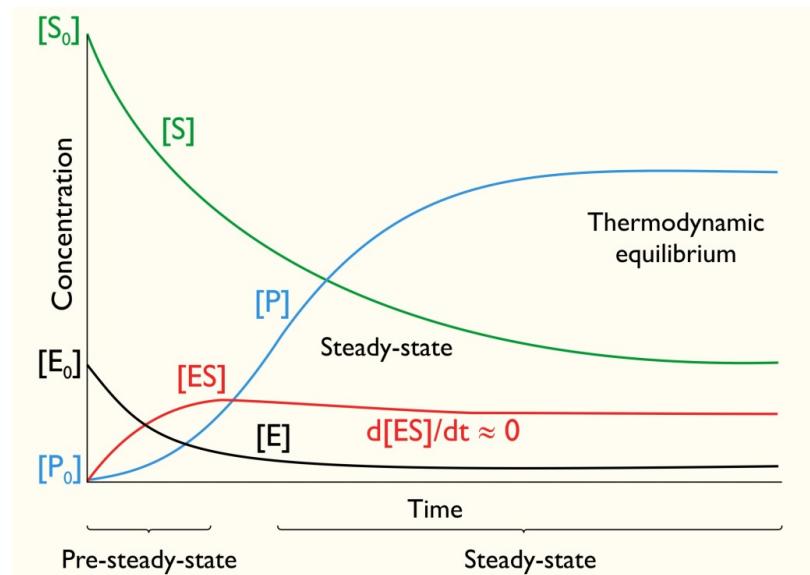


Figure 9.4. The steady-state

Decomposition of the ES complex can happen on two different routes: it can occur towards product formation with rate constant  $k_2$ , and also in the opposite direction towards substrate regeneration with rate constant  $k_{-1}$ . The development of the steady-state is illustrated in Figure 9.4, while the mathematical requirements of steady-state formation are formulated by Equation 9.32:

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES] = 0 \quad (9.32)$$

Although the steady-state modification of the model was suggested by George Edward Briggs and John Burdon Sanderson Haldane, the improved model is still being referred to as the kinetic model of Michaelis and Menten and the derived equation as the Michaelis-Menten equation. In this equation, the ES complex forms at a rate  $k_1[E][S]$ ; it decays back towards the substrate at a rate  $k_{-1}[ES]$  and decomposes towards the product at a rate  $k_2[ES]$ . In the steady-state, the rate of ES formation and the sum of the two types of ES decomposition rates are equal in magnitude and, thus, the concentration of the ES complex does not significantly change.

In the next section, Equation 9.32 will be transformed in several steps in order to yield a final equation containing the experimentally pre-set enzyme concentration and substrate concentration. The improved model has the same initial requirement: the substrate concentration must exceed the enzyme concentration by orders of magnitude. The rate of the reaction can be easily formulated, as shown in Equation 9.33, which is analogous to Equation 9.20 of the simpler model:

$$V_0 = k_2[ES] \quad (9.33)$$

As the upgraded model contains three rate constants instead of one, the earlier model's  $k_{cat}$  is replaced with the  $k_2$  rate constant introduced in the improved scheme. By rearranging Equation 9.32 we get Equation 9.34:

$$k_1[E][S] = k_{-1}[ES] + k_2[ES] \quad (9.34)$$

Just as it was done in the first model, the free enzyme concentration is expressed as the difference of the total enzyme concentration and the concentration of the ES complex (Equation 9.35):

$$k_1([E]_T - [ES])[S] = k_{-1}[ES] + k_2[ES] \quad (9.35)$$

Simple algebraic transformations result in Equation 9.36:

$$k_1[E]_T[S] - k_1[ES][S] = (k_{-1} + k_2)[ES] \quad (9.36)$$

Then the [ES] term-containing parts are organised to the same side, resulting in Equation 9.37:

$$k_1[E]_T[S] = (k_{-1} + k_2)[ES] + k_1[ES][S] \quad (9.37)$$

By multiplying out the [ES] factor, we get equation 9.38:

$$k_1[E]_T[S] = (k_{-1} + k_2 + k_1[S])[ES] \quad (9.38)$$

In Equation 9.39, the [ES] term is arranged to the left side:

$$[ES] = \frac{k_1[E]_T[S]}{k_{-1} + k_2 + k_1[S]} \quad (9.39)$$

Then, on the right side of the equation, both the numerator and the denominator are divided by the  $k_1$  rate constant, which results in Equation 9.40:

$$[ES] = \frac{\frac{[E]_T[S]}{(k_{-1} + k_2)}}{k_1} \quad (9.40)$$

In the denominator of Equation 9.40, there is a complex composed of three rate constants. This quotient has been defined as the Michaelis constant, with the abbreviation  $K_M$ . Replacing the quotient for  $K_M$  yields Equation 9.41:

$$\frac{(k_{-1} + k_2)}{k_1} \equiv K_M \quad (9.41)$$

Note that  $K_M$  is defined as the ratio of the two-direction decay rate of the ES complex and the one-direction formation rate of the complex. In other words, it quantifies the instability of the ES complex.

Building Equation 9.41 into Equation 9.40 leads to Equation 9.42:

$$[ES] = \frac{[E]_T[S]}{K_M + [S]} \quad (9.42)$$

By combining Equations 9.33 and 9.42, we get Equation 9.43:

$$V_0 = k_2[ES] = \frac{k_2[E]_T[S]}{(K_M + [S])} \quad (9.43)$$

As already mentioned, in the case of Equation 9.27 of the simpler model, the highest initial rate (at a given total enzyme concentration),  $V_{\max}$ , is achieved when all enzyme molecules are part of an ES complex, i.e.  $[ES] = [E_T]$ . Therefore, the  $k_2[E]_T$  product in Equation 9.43 is in fact the  $V_{\max}$ . Taking this into consideration, we get Equation 9.44, which is the final and most commonly used rate equation of the improved Michaelis-Menten model:

$$V_0 = \frac{V_{\max} [S]}{(K_M + [S])} \quad (9.44)$$

Note that the mathematical forms of the corresponding final equations of the simple (Equation 9.28) and the improved (Equation 9.44) models are identical. Consequently, just like Equation 9.28, Equation 9.44 is also consistent with experimental results in terms of the dependence of the initial rate on substrate concentration. In the substrate concentration range where  $[S] \ll K_M$ ,  $V_0$  is linearly proportional to  $[S]$ ; while in the range where  $[S] \gg K_M$ , the reaction rate does not depend on the substrate concentration, and it has a maximal value.

When the numerical value of  $[S]$  equals that of  $K_M$ , the rate of the reaction is exactly half of the maximal one.

In spite of the many formal similarities, there are principal differences in the interpretations of the two models. Note that the  $K_S$  constant of the simple model and the  $K_M$  constant of the improved model have different meanings. Equation 9.45 illustrates how the dissociation constant-type  $K_S$  is derived from the  $k_1$  and  $k_{-1}$  rate constants.

$$K_S = \frac{k_{-1}}{k_1} \quad (9.45)$$

As already mentioned,  $K_S$  is a kind of affinity descriptor indicating how tightly the enzyme binds the substrate. It is readily apparent from the comparison of Equations 9.45 and 9.41 that  $K_M$  will equal  $K_S$  only if the  $k_2 \ll k_{-1}$  requirement is fulfilled. This requirement would mean that the rate of ES decomposing towards the product would be orders of magnitude lower than the rate of ES decomposing back to substrate. However, for efficient enzymes, the very opposite situation, i.e.  $k_2 \gg k_{-1}$ , may apply. As already mentioned, low values of  $K_S$  would—by mathematical formalism—indicate an efficient enzyme. However, from a thermodynamic aspect, the very same characteristics would indicate an inefficient enzyme.

The more efficient the enzyme, the less correct it is to interpret  $K_M$  as indicating affinity of the enzyme-substrate interaction.

Clearly,  $K_M$  and  $K_S$  of the two models are not interchangeable. On the other hand, the  $k_2$  rate constant of the improved model, if the ES complex decays to enzyme and product in a single step, is identical to the  $k_{\text{cat}}$  rate constant of the simpler model. (If the final product is formed in a complex multi-step process, the  $k_{\text{cat}}$  rate constant is derived from an equation containing the rate constants of all individual steps.)

Let us examine the meaning of the  $k_{\text{cat}}$  rate constant based on Equations 9.46 and 9.47:

$$V_{\max} = \left( \frac{d[P]}{dt} \right)_{\max} = k_{\text{cat}} [E]_T \quad (9.46)$$

$$k_{\text{cat}} = \frac{\left( \frac{d[P]}{dt} \right)_{\max}}{[E]_T} \quad (9.47)$$

As it was shown, the maximal rate is the product of  $k_{\text{cat}}$  and the total enzyme concentration. Equation 9.47 is obtained when the rate of product concentration change is divided by the total enzyme concentration. As the product and the enzyme are present in the very same solution having the same volume, the ratio of concentrations is identical to the ratio of the numbers of molecules. Therefore, Equation 9.47 will provide a value on the number of product molecules generated by a single enzyme molecule in a unit period of time.

Accordingly, the  $k_{cat}$  rate constant is also referred to as the turnover number of the enzyme. It has the dimension of reciprocal of time. Naturally, the higher the value of  $k_{cat}$ , the more efficient the enzyme as a “chemical” catalyst, working after the ES complex had already been formed. Let us also note that, in cases when in the improved model  $k_{cat} = k_2$ , this rate constant is present in the nominator of the quotient defining  $K_M$ . This means that a high  $k_{cat}$  will increase the value of  $K_M$ . On the other hand, it is also clear that a lower  $K_M$  means that the half-maximal reaction rate is achieved at a lower substrate concentration, which is another measure of being an efficient enzyme. What would then be the best parameter to describe enzyme efficiency?

The most effective enzymes are expected to catalyse the reaction at a high rate even at low substrate concentration and, naturally, their turnover number should also be high. As already explained, at low substrate concentration where  $[S] \ll K_M$ , the reaction rate is a linear function of substrate concentration. This is shown in Equation 9.48:

$$V_0 \cong \frac{k_{cat}}{K_M} [E][S] \quad (9.48)$$

The  $k_{cat}/K_M$  quotient in the equation is the rate constant of a second-order reaction, as the rate of this reaction depends on the concentration of two compounds. This  $k_{cat}/K_M$  quotient illustrates how efficiently the enzyme performs in the most challenging situation when the substrate concentration is very low. Accordingly, the  $k_{cat}/K_M$  quotient can be called *catalytic efficiency*, although its commonly used denomination is *specificity constant*.

Let us examine what sets the limit of the catalytic efficiency of enzymes by reviewing the case of the most efficient enzymes. Starting with Equation 9.48, we can generate Equation 9.49 by replacing the  $k_{cat}$  term with  $k_2$  according to the simple one-step scheme of the reaction, and by replacing  $K_M$  (by definition) with the  $(k_{-1}+k_2)/k_1$  quotient:

$$V_0 = \frac{k_2}{(k_{-1}+k_2)} [E][S] = \frac{k_2 k_1}{k_{-1}+k_2} [E][S] \approx k_1 [E][S] \quad (9.49)$$

In the right side of the resulting equation, and in the special case when  $k_2 \gg k_{-1}$ , the  $k_{-1}$  term becomes negligible in the denominator. By ignoring the  $k_{-1}$  term, both the numerator and the denominator of the resulting quotient can be divided by the  $k_2$  term. The  $k_2 \gg k_{-1}$  situation means that the ES complex decays practically exclusively towards the product instead of returning towards the substrate. When this assumption is valid, the  $k_{cat}/K_M$  rate constant of the reaction approximates the value of  $k_1$ . This means that the limit for the rate of the enzymatic reaction will be set by the rate at which the enzyme and the substrate encounter. Naturally, the steady-state rate of the reaction cannot exceed the rate of ES formation.

At first approximation, the limit for the most effective enzymes is set by the diffusion that limits the speed of the enzyme-substrate encounters. The diffusion rate can be accurately calculated based on the size of the diffusing molecules and the viscosity of the medium. In the case of some extremely efficient enzymes, it turned out that the rate of the enzymatic reaction exceeded the calculated diffusion rate.

There are at least two situations that can explain this apparent discrepancy. If the substrate carries large number of electric charges of one type and the substrate binding site of the enzyme carries complementary (opposite) electrical charges, then the electrostatic attraction can steer the substrate towards the enzyme. The electrostatic interaction between full charges is a long-range interaction (compared to the size of small molecules). This condition, together with the orienting capacity of the interaction, significantly increases the frequency of enzyme-substrate encounters relative to the simple diffusion-limited case. A second important exception from the simple diffusion limit is related to multi-enzyme complexes. Many “final products” of metabolic or other chemical pathways are formed by a series of consecutive chemical reactions that proceed through many intermediate compounds. This means that the product of a given enzymatic reaction is the substrate of another enzyme catalysing the subsequent reaction. In a simple case, the product of the first reaction must encounter the second enzyme by random diffusion. If, on the other hand, these enzymes are organised in a proper arrangement in multi-enzyme complexes, the intermediate compounds can be “handed” directly from one enzyme to the other. This channelling effect can greatly enhance the overall rates of multi-step pathways.

## 9.3 Determination of initial reaction rates and principal kinetic parameters

As it has been mentioned several times in this chapter, the initial reaction rate of an enzyme-catalysed chemical reaction ( $V_0$ ) is by definition the change of the product concentration as a function of the elapsed time:  $V_0 = d[P]/dt$ . In the general case of an S→P reaction scheme, an identically valid rate definition is the one that accounts for the decrease in substrate concentration:  $V_0 = -d[S]/dt$ . When the initial rate  $V_0$  is to be measured, the concentration of either the product or the substrate needs to be measured as a function of time.

A chemical reaction is inherently accompanied by changes in electronic structure. Alteration of the electronic structure alters the excitability of the electrons, which in most cases can be detected by various spectroscopic methods. In the simplest case, the optical properties of the product will measurably differ from those of the substrate (and all other components of the solution). In such cases, the formation of the product or the diminution of the substrate can be measured in real time—in other words, a “progress curve” can easily be determined. In an appropriate concentration range (see Chapter 4), the magnitude of the absorbance—or, in the case of fluorescence, the intensity of the emitted light—is linearly proportional to the concentration of the light-absorbing or light-emitting molecule. Based on this, changes in the concentration of the product (or that of the substrate) can be precisely determined in real time.

If neither the product nor the substrate possesses readily measurable spectroscopic properties, there is still an opportunity for a spectroscopic measurement. It can be done if the product can be driven to a second (not necessarily enzymatic) chemical reaction that is instantaneous and results in an optically active (light-absorbing or fluorescent) compound. Moreover, if the second chemical reaction is compatible with the conditions (pH, temperature etc.) of the enzymatic reaction of interest, and its product does not interfere with the enzyme, this more complex approach can also establish a real-time measurement. On the other hand, if the second (coupled) reaction is not compatible with the enzymatic reaction, then—instead of a real-time measurement—a different approach, end time analysis can be applied. In such cases, the enzymatic reaction is usually started in a relatively large volume, and aliquots are withdrawn from the reaction chamber at different time intervals. These aliquots are treated to instantly stop the enzymatic reaction (usually by applying extreme pH or temperature, or strong protein denaturing reagents, e.g. trichloroacetic acid) and then the second, signal-generating chemical reaction is performed.

The principle of real-time initial rate measurements is illustrated in the left panel of Figure 9.3 showing progress curves with different initial substrate concentrations. The reaction is started by adding either the substrate or the enzyme to the solution and, after thorough mixing of the sample, detection is started as soon as possible. A quick start is very important as initial rates are to be measured (cf.  $V_0$ , a tangent to the curve at the start of the reaction with substrate concentration  $S_4$  in Figure 9.3). Naturally, during the enzymatic reaction, the concentration of the substrate will gradually decrease. To this end, the measurement should be performed so that this decrease in substrate concentration should be negligible. This is important because, as illustrated in the left panel of Figure 9.3, decreasing substrate concentration is accompanied by a decreasing reaction rate. Moreover, to make the situation more complicated, the significance of this unwanted effect is itself a function of substrate concentration, as it can be seen on the saturation curve in the right panel of Figure 9.3. At substrate concentrations well below the value of  $K_M$ , the  $[S]-V_0$  function is approximately linear; while at substrate concentrations well above the  $K_M$ , the  $V_0$  is almost independent of  $[S]$ . As a consequence, the lower the substrate concentration, the larger the error in the measured rate value at an identical (for example 10 %) percentage of substrate concentration change. For example, at  $[S]$  well below  $K_M$ , a 10 % decrease in  $[S]$  will cause an about 10 % decrease in  $V_0$ ; while at  $[S]$  well above  $K_M$ , a 10 % decrease in  $[S]$  will cause a much smaller error. As a rule of thumb, the initial rate should be measured so that, during the measurement, the concentration of the substrate does not decrease by more than 10 %. If this requirement is fulfilled, the initial rates will be determined with an error lower than 10 %.

In order to determine the main kinetic parameters,  $V_{max}$  and  $K_M$ , the initial reaction rates are measured at various initial substrate concentrations. If substrate concentrations are dispersed in the proper range,  $V_0$  values plotted as the function of  $[S]$  trace out a typical saturation curve illustrated in the right panel of Figure 9.3.

For accurate determination of the parameters, the values of  $[S]$  should cover a wide range. Naturally, as long as no estimate exists on the value of the  $K_M$ , iterative trial-and-error sampling of the right range needs to be performed. Once the  $K_M$  is estimated, the measurements should be set such that the pre-set  $[S]$  values cover the ~0.2  $K_M$ –5  $K_M$  region with at least 8 evenly placed substrate concentrations.

Once a satisfactory number of  $[S]$  -  $V_0$  data are collected, all information is available to determine the values of the  $V_{\max}$ ,  $k_{cat}$  and  $K_M$  parameters.

The very first kinetic measurements date back to times when computers did not exist. At the time, the evaluation of the data presented a considerable challenge. In the right panel Figure 9.3, the  $[S]$  -  $V_0$  plot corresponds to a saturation curve, actually a part of a rectangular hyperbola. Most naturally, a simple visual analysis of such a curve would not allow for a reliable determination of the maximum value,  $V_{\max}$ , to which the curve converges. For the very same reason, the substrate concentration belonging to the half-maximal initial rate, i.e. the value of  $K_M$ , cannot be reliably estimated either.

For quantitative analysis of such non-linear relationships, the following procedure was implemented. The mathematical equation describing the model of the relationship was transformed such that the transformed equation had a linear form. This way, when derived values of the experimental data were plotted, it resulted in a straight line instead of a curve. Therefore, the problem was simplified to the task of a linear regression, which could be performed without a computer.

Nowadays, kinetic parameters are calculated using computer programs and non-linear regression. Nevertheless, in the field of enzyme kinetics, linear plotting of derived values still remained didactically useful, particularly in studies of enzyme inhibitors. Enzyme inhibitors can be classified into several groups based on their mechanism of action. As we will see later, the linearised plots clearly report the type of mechanism of action of the inhibitor.

The most widespread linearised version of the Michaelis-Menten equation (Equation 9.44) was introduced by Hans Lineweaver and Dean Burk. The transformation they proposed leads to Equation 9.52 that was later named after them the Lineweaver-Burk equation. The transformation is done as follows. The reciprocal of both sides of original Equation 9.44 is taken, leading to Equation 9.50:

$$\frac{1}{V_0} = \frac{K_M + [S]}{V_{\max}[S]} \quad (9.50)$$

From this, Equation 9.51 is generated by a simple algebraic transformation that resolves the fraction on the right side of the equation into the sum of two fractions:

$$\frac{1}{V_0} = \frac{K_M}{V_{\max}[S]} + \frac{[S]}{V_{\max}[S]} \quad (9.51)$$

Then, on the right side of the equation, the fraction in the second term can be reduced by dividing both the numerator and the denominator by  $[S]$ . This results in Equation 9.52:

$$\frac{1}{V_0} = \frac{K_M}{V_{\max}[S]} + \frac{1}{V_{\max}} \quad (9.52)$$

The plot corresponding to Equation 9.52 is the double reciprocal form of the original saturation curve introduced in Figure 9.3. While, in the original plot,  $V_0$  data were plotted as a function of  $[S]$ , in the double reciprocal Lineweaver-Burk plot,  $1/V_0$  values are plotted as a function of  $1/[S]$ . As illustrated in Figure 9.5, the plot corresponding to Equation 9.52 is a straight line.

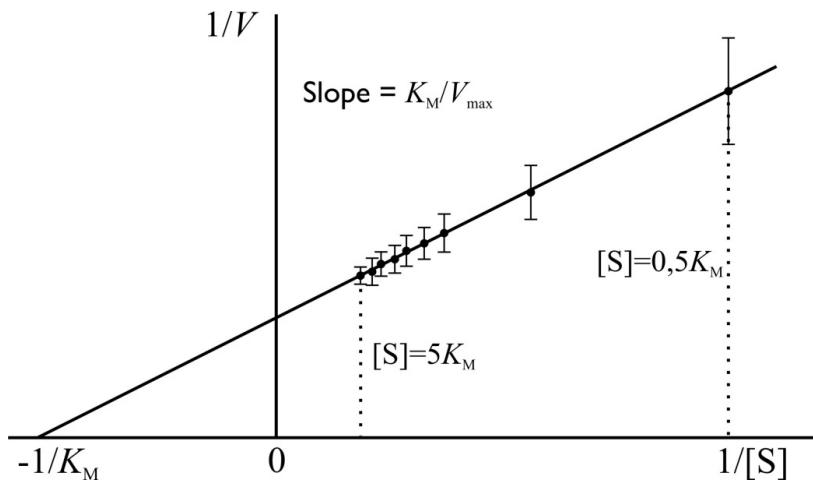


Figure 9.5. The Lineweaver-Burk double reciprocal plot

According to the general equation of a line in slope-intercept form,  $Y = aX + b$ . In our specific case,  $Y = 1/V_0$ ;  $X = 1/[S]$ ;  $a = K_M/V_{\max}$  and  $b = 1/V_{\max}$ . The slope of the line, „ $a$ ” (which is the tangent of the angle between the line and the x axis), provides the value of  $K_M/V_{\max}$ ; the intercept of the y axis (where the value of X and therefore the value of  $1/[S]$  is 0) provides the value of  $1/V_{\max}$ ; whereas the intercept on the x axis (where the value of Y, therefore that of  $1/V_0$  is 0) provides the value of  $-1/K_M$ . Based on the slope and any of the two intercepts, both  $V_{\max}$  and  $K_M$  can be determined.

This approach is elegantly simple, but this type of data evaluation, without using proper weighing of the individual data, can lead to significant errors. Because of experimental detection limitations,  $V_0$  data measured at the lowest  $[S]$  are the most likely to be inaccurate. Due to the double reciprocal analysis, these least reliable values will contribute the highest (reciprocal) values, and these will thus have the highest impact on the calculated  $K_M$  and  $V_{\max}$  parameters.

In the era of computers, data analysis is performed by non-linear regression. In the particular case, an appropriate algorithm based on Equation 9.44 is used. The algorithm iteratively searches for the  $V_{\max}$ - $K_M$  parameter pair that, when substituted in the rectangular hyperbolic equation, provides the best (lowest-deviation) fit to the experimental  $[S]$  -  $V_0$  dataset.

Most fitting programs require initial (estimated) values for the parameters (in this case  $V_{\max}$  and  $K_M$ ) to start the iterative search.

If the total concentration of the enzyme,  $[E]_T$ , is known during the measurements, then—based on Equation 9.43—not only  $V_{\max}$  but also the turnover number,  $k_{\text{cat}}$ , can be determined by dividing  $V_{\max}$  by  $[E]_T$ .

It is worth noting that, if the aim is the determination of only the catalytic efficiency (the  $k_{\text{cat}}/K_M$  quotient), a simple direct measurement can be applied. If the value of  $V_0$  is measured at a substrate concentration orders of magnitude lower than the value of the  $K_M$  and if the total concentration of the enzyme,  $[E]_T$ , is known, then—based on the same line of thinking that led to Equation 9.48— $k_{\text{cat}}/K_M$  will equal  $V_0/[E]_T[S]$ . Nevertheless, this type of measurement is only advisable for particular enzyme-substrate pairs. The substrate should convert to a product that provides a highly intense signal, while the enzyme should have high turnover number on that substrate. This is required because, in this kind of measurements, very low levels of product concentration need to be detected.

## 9.4. Enzyme inhibition mechanisms

Enzymes play central roles in life processes. It holds for most enzymes that their function is needed only in certain conditions. When those conditions do not apply, the activity of a given enzyme can be futile or even harmful. Accordingly, the activity of most enzymes is under strict control. Enzymes can be regulated at multiple levels, ranging from transcriptional regulation of the expression of the enzyme-encoding gene through the direct regulation of the activity of the enzyme molecule by effector molecules to the controlled proteolytic decomposition of the enzyme. In this chapter, only those inhibitors will be reviewed that reversibly and specifically bind to enzymes through

non-covalent interactions and inhibit the substrate-binding and/or catalytic apparatus of the given enzyme. These inhibitors can be classified into three mechanistic groups based on their mechanism of action: competitive, uncompetitive and mixed inhibitors. The type of inhibition can be determined through enzyme kinetic measurements.

In the most frequent procedure, the kinetic parameters  $V_{max}$  and  $K_m$  are determined as already described in Section 9.3—however, in this case, in the presence of several pre-set inhibitor concentrations. When the experimental data are examined in the form of double reciprocal Lineweaver-Burk plots, the deviation of the obtained line from that of the uninhibited case will be diagnostic of the type of inhibition.

### 9.4.1. Competitive inhibition

Competitive inhibitors compete for the substrate-binding site of the enzyme with the substrate, because the substrate and the inhibitor bind to identical or overlapping sites. Due to the overlapping nature of the binding sites, a ternary complex—in which the substrate and the inhibitor would simultaneously bind to the enzyme—cannot form. Accordingly, in the enzyme-inhibitor complex, the enzyme is completely inactive.

By popular—but quite misleading—terminology, these inhibitors are said to “displace” the substrate from the enzyme. While this term is aimed to be expressive, it is totally inadequate to explain the mechanism of this type of inhibition. The popular term suggests that the inhibitor would bind to the ES complex and would thus somehow actively force the substrate to dissociate. As already mentioned, no ternary complex is formed—not even temporarily. This inhibitory mechanism simply obeys a thermodynamic principle: two equilibria exist in parallel, one between the enzyme and the inhibitor and another between the enzyme and the substrate. More precisely, the latter one is a quasi-equilibrium because, during the measurement, the enzyme and the substrate are in a steady-state (as shown in Figure 9.4). The equilibrium concentrations of the free compounds and those of the complexes are dictated by the total concentrations of the individual compounds and the affinity of their interactions.

The equilibrium between the enzyme and the inhibitor is described by Equation 9.53 in which the  $K_I$  term represents a dissociation constant:

$$K_I = \frac{[E][I]}{[EI]} \quad (9.53)$$

The two equilibria are not independent as the complexes, ES and EI, equilibrate with the same free enzyme pool, E. Increasing EI concentration by increasing the inhibitor concentration can be achieved only through a decrease in ES concentration, and vice versa: an elevated substrate concentration can increase the concentration of the ES complex only at the expense of the EI complex.

This mechanism is illustrated in Figure 9.6.

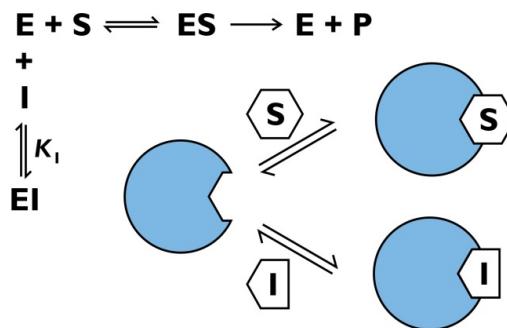


Figure 9.6. The scheme of competitive inhibition

When solving the Michaelis-Menten equation, we made use of the simple fact that the total enzyme concentration can be expressed as follows:  $[E]_T = [E] + [ES]$ . On the other hand, in the presence of a competitive inhibitor,  $[E]_T = [E] + [ES] + [EI]$ . Solving the Michaelis-Menten equation such that this difference is taken into consideration leads to Equation 9.54. (For brevity, the intermediate steps that yield this equation are not shown.)

$$V_0 = \frac{V_{\max}[S]}{(K_M + [S])} \quad (9.54)$$

The meaning of the term  $\alpha$  in Equation 9.54 is explained in Equation 9.55:

$$\alpha = 1 + \frac{[I]}{K_I} \quad (9.55)$$

It is readily apparent that, in the absence of inhibitor, the value of  $\alpha$  is one and, as expected, we get the original equation. In the presence of inhibitor, the value of  $\alpha$  exceeds one. The higher the concentration of the inhibitor compared to the value of the  $K_I$  dissociation constant, the higher the value of  $\alpha$ . Equation 9.54 clearly indicates that the measured  $V_{\max}$  will be invariant, regardless of the presence of the inhibitor. On the other hand, in the presence of a competitive inhibitor, the measured  $K_M$  will be higher than in the absence of the inhibitor. As in the case of this inhibitor type the substrate competes with the inhibitor, it is intuitively comprehensible that, at “infinitely” high substrate concentrations, the presence of the inhibitor should not affect the measurements, i.e. the maximal rate of the reaction should be unchanged. However, as in the presence of a competitive inhibitor higher than normal substrate concentration is needed to achieve a (half-) maximal rate, the value of  $K_M$  must be higher than in the uninhibited case. That is exactly what Equation 9.54 expresses.

When Equation 9.54 is rearranged according to the double reciprocal transformation, we get Equation 9.56, which is analogous to Equation 9.52 introduced previously for the uninhibited case:

$$\frac{1}{V_0} = \frac{\alpha K_M}{V_{\max}[S]} + \frac{1}{V_{\max}} \quad (9.56)$$

Equation 9.56 is graphically illustrated by the plots shown in Figure 9.7. The double reciprocal plots clearly show that, in the presence of a competitive inhibitor, the lines are steeper than in the uninhibited case; but the intercept on the y axis, which refers to the  $1/V_{\max}$  value, remains the same. The plot nicely illustrates the didactical strength of double reciprocal data analysis to demonstrate the mechanism of a reversible inhibitor.

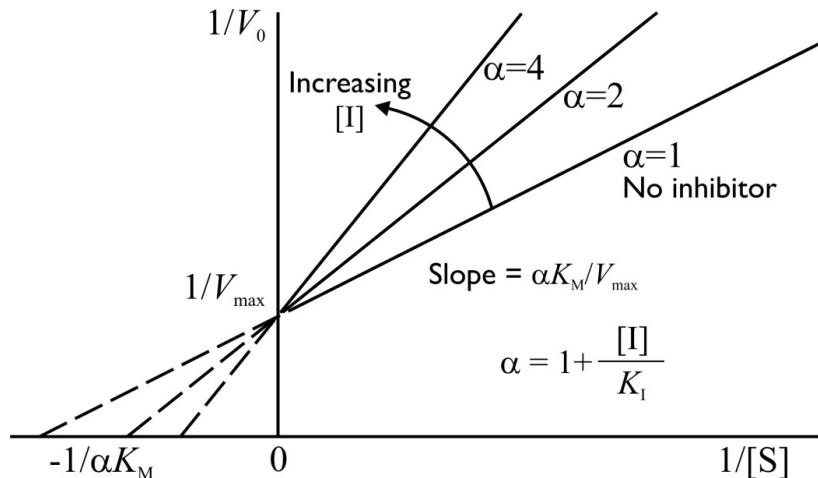


Figure 9.7. Double reciprocal Lineweaver-Burk analysis of competitive inhibition

As competitive inhibitors compete with the substrate for overlapping binding sites on the enzyme, it is not surprising that competitive inhibitors often resemble the substrate in terms of chemical structure, shape and polarity pattern. Due to this, competitive inhibitors are often used as useful reagents to study the substrate binding mechanism of enzymes. Comparative analysis of the structure of the substrate and that of a set of different competitive inhibitors can help in identifying the functionally most important parts of the substrate—those that provide the most binding

energy in the ES complex. Note that such indirect approaches are important because direct analysis of the short-lived ES complex is a demanding scientific challenge.

## 9.4.2. Uncompetitive inhibition

Some inhibitors bind only to the ES complex without binding to the free enzyme. This interaction scheme is illustrated in Figure 9.8.

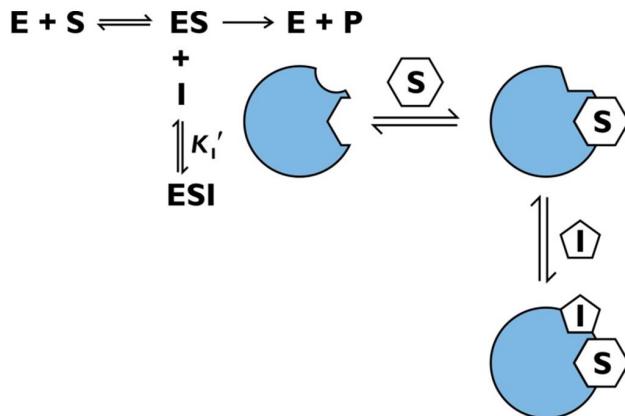


Figure 9.8. Uncompetitive inhibition

The kinetic equation of this type of inhibition can also be expressed as a modified version of the uninhibited case, as shown in Equation 9.57:

$$V_0 = \frac{V_{\max}[S]}{(K_M + \alpha'[S])} \quad (9.57)$$

The meaning of the  $K_{I'}$  term is shown in Equation 9.58:

$$K'_{I'} = \frac{[ES][I]}{[ESI]} \quad (9.58)$$

The meaning of the  $\alpha'$  term in Equation 9.57 is analogous to the meaning of  $\alpha$  introduced in Equation 9.55. However, as shown in Equation 9.59,  $\alpha'$  is associated with  $K_{I'}$  instead of  $K_I$ :

$$\alpha' = 1 + \frac{[I]}{K'_{I'}} \quad (9.59)$$

Double reciprocal transformation of Equation 9.57 results in Equation 9.60:

$$\frac{1}{V_0} = \frac{K_M}{V_{\max}[S]} + \frac{\alpha'}{V_{\max}} \quad (9.60)$$

A graphical illustration of Equation 9.60 is presented in Figure 9.9.

$$\frac{1}{V_0} = \left( \frac{K_M}{V_{\max}} \right) \frac{1}{[S]} + \frac{\alpha'}{V_{\max}}$$

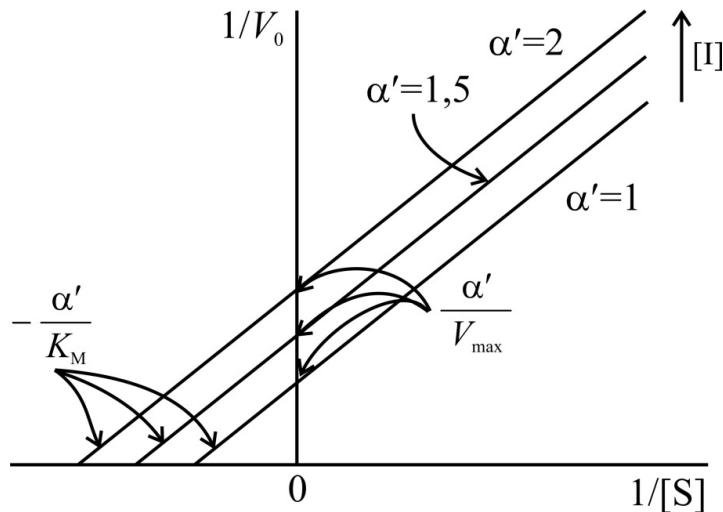


Figure 9.9. Double reciprocal Lineweaver-Burk analysis of uncompetitive inhibition

The plot clearly illustrates that both the  $K_M$  and  $V_{\max}$  values are divided by the value of  $\alpha'$  (i.e. their reciprocal value is multiplied by the value of  $\alpha'$ ). This means that, unlike in the case of competitive inhibitors, the presence of an uncompetitive inhibitor results in a decreased  $V_{\max}$  value. Moreover, the  $K_M$  (i.e. the substrate concentration at which the reaction rate reaches its half maximum) also decreases, and it does so to exactly the same extent as the  $V_{\max}$ . As both kinetic parameters decrease to the same degree, the slopes of the lines do not change. This type of inhibition is dramatically different from the competitive one. Namely, the effect of an uncompetitive inhibitor, although its binding is reversible, cannot be abolished by increasing substrate concentration.

### 9.4.3. Mixed inhibition

There are inhibitors that can bind both to the free enzyme as well as to the ES complex. These inhibitors represent a combination of the two already discussed types. The scheme of mixed inhibition is illustrated in Figure 9.10.

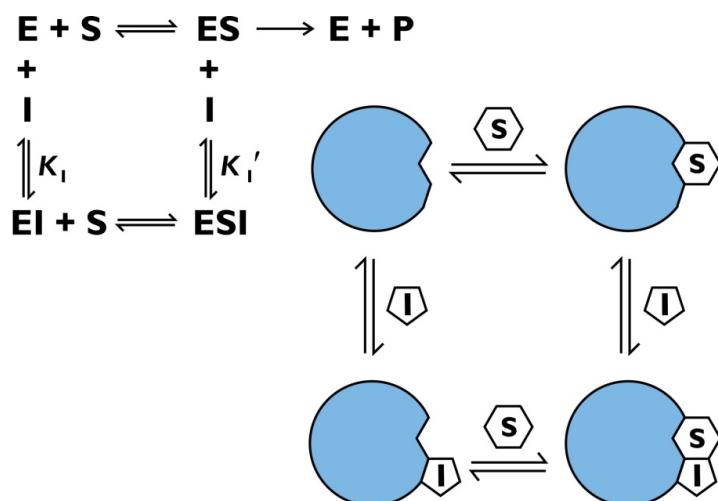


Figure 9.10. Mixed inhibition

In most cases of mixed inhibition, the inhibitor binds to the free enzyme and to the ES complex with different affinities—however the case where  $K_I$  is equal to  $K_I'$  cannot be excluded (see below).

As the scheme of mixed inhibition is the combination of the schemes of the previously mentioned inhibition types, the kinetic Equation 9.61 of mixed inhibition is also a combination of the previous equations:

$$V_0 = \frac{V_{\max}[S]}{(\alpha K_M + \alpha'[S])} \quad (9.61)$$

The double reciprocal transformation of Equation 9.61 results in Equation 9.62:

$$\frac{1}{V_0} = \frac{\alpha K_M}{V_{\max}[S]} + \frac{\alpha'}{V_{\max}} \quad (9.62)$$

The meaning of the terms  $\alpha$  and  $\alpha'$  is the same as was discussed for the competitive and uncompetitive cases, respectively. These represent the contribution of the competitive and uncompetitive components, respectively, to the observed inhibition.

A graphical illustration of the double reciprocal equation is shown in Figure 9.11.

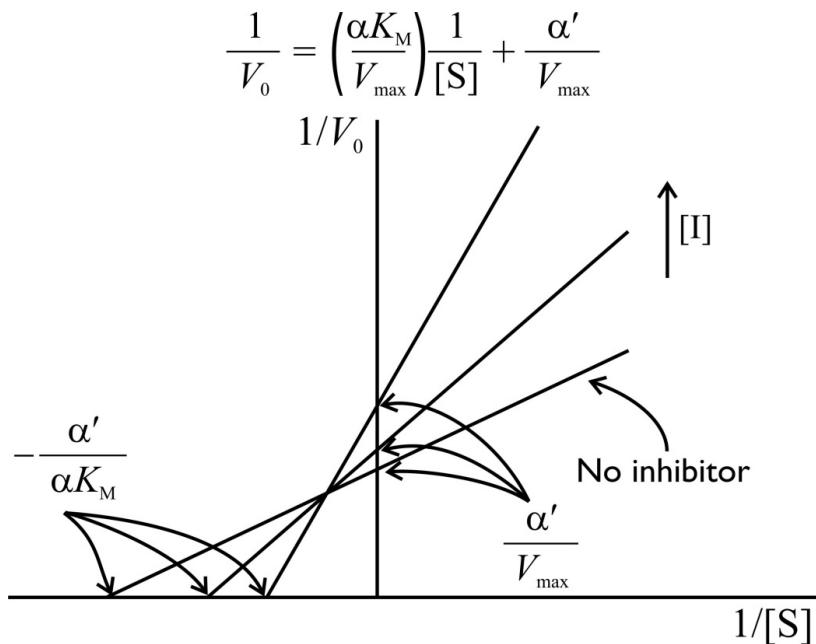


Figure 9.11. Double reciprocal Lineweaver-Burk analysis of mixed inhibition

In conclusion, the three different types of inhibition are accompanied by characteristically different double reciprocal plots. Therefore, these plots are of diagnostic value in the quick assessment of the type of the inhibitor.

It is worth noting that a certain (rather theoretical) variant of mixed inhibition is a type in which the inhibitor would bind to the free enzyme and the ES complex with exactly the same affinity. This would result only a single type of  $\alpha$  term with a single value. This would alter the plot in Figure 9.11 such that each line corresponding to a given inhibitor concentration would intercept the x axis in the same point (as the factor of  $-1/K_M$  would be  $\alpha/\alpha = 1$ ), while the y axis intercepts would depend on inhibitor concentration, having different  $\alpha/V_{\max}$  values.

In the case of all inhibitor types, the values of  $\alpha$  and/or  $\alpha'$  can be experimentally determined and, through these parameters, the corresponding values of the inhibitory constants can also be computed.

---

# Chapter 10. Recombinant DNA technology

by Mihály Kovács

## 10.1. Recombinant DNA techniques and molecular cloning

Recombinant DNA (molecular cloning) techniques enable the amplification, sequence determination, manipulation and functional investigation of DNA segments of interest (e.g. genes) as well as the production, manipulation and investigation of products (RNA, proteins) encoded by these DNA segments. The amplification of DNA molecules of interest can take place by two principal means:

- a. By creating recombinant DNA constructs that enable the *in vivo* amplification of a certain DNA segment in a host organism. During the process of cloning, one (or a small number of) DNA molecule(s) taken up by a single cell are amplified by the proliferation (division) of the host cell. Recombinant DNA constructs are created by linking the DNA segment to be cloned (the so-called insert) into a vector (carrier) DNA fragment. The vector DNA is necessary for DNA propagation in the host cell. Chapters 10.2-10.6 below will summarise the principles of the design, production, amplification and investigation of recombinant DNA constructs.
- b. Via *in vitro* amplification of the DNA segment of interest by applying polymerase chain reaction (PCR). PCR will be described in detail in Chapter 10.7.

Recombinant DNA technology is an essential and indispensable tool in practically all lines of modern biological and biomedical research as well as in biotechnology-based industry. The following list contains a few examples for the areas of utilisation of these techniques.

- a. Investigation, alteration and manipulation of the sequence and structure of DNA and RNA molecules, individual genes or entire genomes;
- b. Investigation, alteration and manipulation of the sequence and structure of protein molecules encoded by genes (protein engineering), or the metabolic products produced in reactions catalysed by these proteins;
- c. Investigation of spatial (cell-specific or tissue-specific) and temporal patterns of gene expression;
- d. Elucidation of the mechanisms of action of signal transduction pathways, cancerous (malignant) transformation, physiological and pathological processes, hereditary and infectious diseases, and diagnosis of these conditions;
- e. Research in the areas of developmental biology, evolution, ecology and environmental science;
- f. Applications in agriculture and power industry;
- g. Industrial-scale production of proteins, drug molecules, vaccines, hormones;
- h. Creation of transgenic organisms for research or industrial purposes;
- i. Gene therapy (replacement of missing gene functions);
- j. Legal disputes and criminal investigations; identification of samples, origins and individuals.

## 10.2. Plasmid vectors

Plasmids are extrachromosomal double-stranded circular DNA molecules that can be found in various bacteria and some eukaryotes. In their original form, the size of plasmids ranges between 1 and 200 kbp (kilo-base pairs).

Plasmids often contain genes encoding enzymes that confer a certain selective advantage to the host cell. Such selective advantages include resistance to certain antibiotics. In other cases, the conferred advantage can be the synthesis of antibiotics or various toxins. Elements of certain restriction-modification systems are also encoded in plasmid DNA.

The most commonly applied host cell in recombinant DNA techniques is the bacterium *Escherichia coli* (*E. coli*). Therefore, in the rest of this chapter we will focus on plasmids that occur and can be maintained in *E. coli*.

The replication of plasmids is governed by a subset of the enzymes involved in the replication of the bacterial chromosome (e.g. DNA polymerase I, DNA polymerase III). The starting point of replication is determined by the origin of replication. Sequences located in the vicinity of the origin of replication regulate the copy number of plasmids per cell. Together with the origin of replication, these regions form the so-called replicon. The replication of plasmids that are currently in laboratory use is independent of cell division, i.e. plasmid replication is under a so-called relaxed control. This condition can be asserted, for instance, by the pMB1 replicon located within the pMB1 plasmid. Plasmid vectors containing the pMB1 replicon with its original (native) sequence typically have a copy number of 15-20 per cell.

Several different replicon sequences were found in *E. coli* (e.g. ColE1, p15A, pSC101). Two plasmids harbouring replicons of different type and sequence are compatible, i.e. they can be stably maintained within the same host cell. However, plasmids carrying identical origins of replication are incompatible, i.e. they cannot be maintained in the same host cell.

As the replication of plasmids is independent of the expression of cellular proteins (the required enzymes have long half-lives in the cell), the inhibition of protein synthesis in the host cell (achieved, for instance, by chloramphenicol) can be used for the inhibition of the replication of the chromosomal DNA, with the concurrent maintenance of the replication of plasmid DNA. By this mechanism, the copy number of plasmids can be further increased.

In terms of utilisation, plasmids applied as vector (carrier) DNA can be classified into two groups. There exist cloning and expression vectors. Cloning vectors enable the investigation, alteration and manipulation of the carried foreign DNA segment of interest. Expression vectors also contain DNA sequences that enable transcription and translation based on the incorporated foreign DNA segment.

The first vectors used in practice, which were mainly cloning vectors, contained a multitude of selection markers in addition to the pMB1 replicon. One of such vectors is pBR322, which contains ampicillin and tetracycline resistance genes. Certain derivatives of this plasmid are still in use. During the construction of the plasmids, the manufacturers achieved the insertion of the recognition sequence of a variety of commercially-available restriction endonuclease enzymes in single copies (unique sites).

In the following, we list the most important developments in the use of plasmid vectors, and the most important functional DNA segments enabling and facilitating the efficient utilisation of plasmids (Figure 10.1).

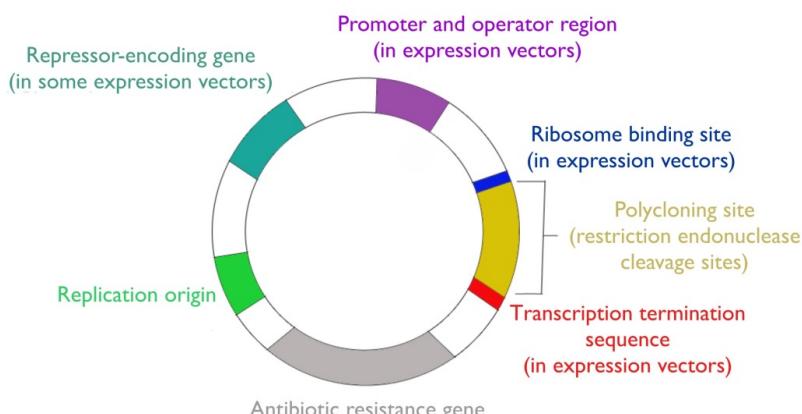


Figure 10.1. Map of a plasmid vector

- a. Decrease in the size of the plasmid. Smaller plasmids can be used to carry longer segments of foreign DNA. The decrease in size is usually associated with the elimination of certain non-essential genes (e.g. one of multiple antibiotic resistance genes).
- b. Increase in the number of plasmid copies per cell. By the modification of the sequence of the pMB1 replicon, the copy number of plasmids has increased to 500-700 per cell (e.g. in the pUC vector family).
- c. Increase in the number of restriction endonuclease recognition sites that can be utilised in order to incorporate the insert into the plasmid. By the incorporation of synthetic oligonucleotide segments (so-called polylinker segments or multiple cloning sites, MCS), the recognition sites of more than twenty restriction endonuclease enzymes were introduced into a short DNA segment of recently-developed plasmid families (e.g. pBluescript, pGEM).
- d. Many vectors contain viral RNA promoter sequences (e.g. ones originating from bacteriophages T3, T7 or SP6) flanking the two sides of the multiple cloning site. This arrangement enables the *in vitro* synthesis of RNA copies from the cloned DNA fragment of interest.
- e. Expression vectors contain the promoter, operator, repressor-binding, ribosome-binding and terminator sequences necessary for gene expression and protein production from the inserted DNA segment of interest.
- f. Several new vectors contain the origin of replication of filamentous phages (e.g. M13), which enables the production of *in vivo* single-stranded DNA copies from the plasmid. The produced single-stranded DNA can be used for sequencing or site-directed mutagenesis.

## 10.3. Creation of recombinant DNA constructs

The principal element in creating recombinant DNA constructs is the incorporation of the foreign DNA segment to be cloned (the so-called insert) into the vector DNA. The insert to be cloned is often a segment of the genomic DNA of an organism to be investigated, or a cDNA (complementary DNA) segment synthesised via reverse transcription from RNA molecules of the organism of interest. The genomic DNA fragments to be cloned are often produced by the fragmentation of the genome of the investigated organism by using restriction endonuclease enzymes, and the subsequent isolation of these restriction fragments. Restriction endonuclease-catalysed fragmentation is also often necessary during the cloning of cDNA products.

The vector DNA enables the replication of the recombinant DNA construct within the host cell. (It should be noted that, in addition to the plasmid vectors introduced in Chapter 10.2, vectors of other origin (e.g. viral vectors, artificial chromosomes) are also in use. These vector types will not be dealt with within the confines of the current chapter.)

The appropriate linking of the vector and insert molecules takes place most often by applying restriction endonuclease digestion followed by DNA ligase-catalysed ligation (Figure 10.2).

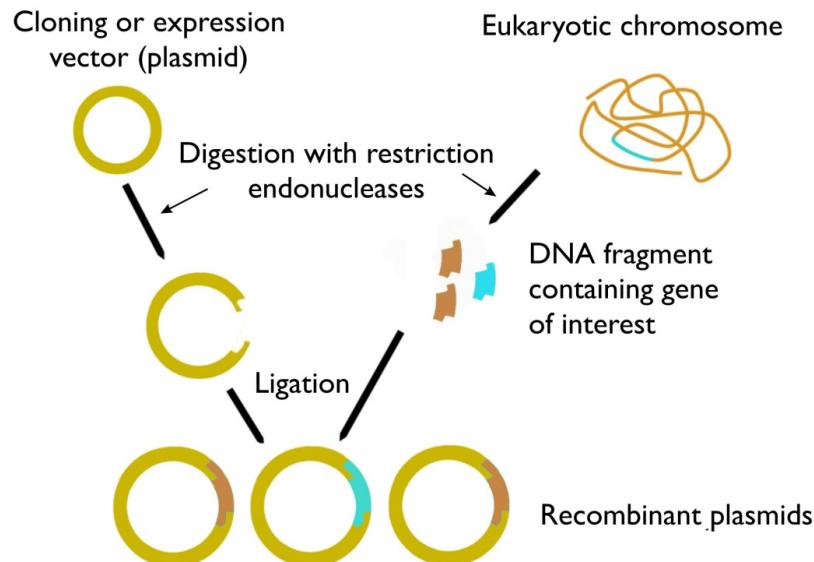


Figure 10.2. Incorporation of the insert into a plasmid vector

Among the known types of restriction endonuclease enzymes, type II enzymes are the most suitable for cloning procedures. These enzymes generally have short (4-8 base-pair (bp)) and specific recognition sequences, and cut the double-stranded DNA at specific sites within or in the vicinity of the recognition site. As it can be observed in Figure 10.3, some enzymes produce 5'-, while others produce 3'-single-stranded DNA overhangs, depending on the location of the cleaving site. Other enzymes, which cut the DNA strands just opposite each other, produce blunt DNA ends with no overhangs.

| Endonuclease | Cleavage products                            |            |           |  |
|--------------|--|------------|-----------|--|
| BamHI        | 5' G   GATCC 3'<br>3' CCTA GG 5'             | G 3'       | 5' GATCC  |  |
|              | Cleavage site                                | CCTAG 5'   | 3' G      |  |
| EcoRI        | 5' GAATT C 3'<br>3' CTTA AG 5'               | G 3'       | 5' AATT C |  |
|              | Palindromic sequence<br>(center of symmetry) | CTTAA 5'   | 3' G      |  |
| HaeIII       | 5' G G   C C 3'<br>3' C C   GG 5'            | GG 3'      | 5' CC     |  |
|              | Blunt end                                    | CC 5'      | 3' GG     |  |
| HhaI         | 5' G C G   C 3'<br>3' C   G C G 5'           | G CG 3'    | 5' C      |  |
|              |  | C 5'       | 3' GCG    |  |
| XhoI         | 5' C   T C G A G 3'<br>3' C A G C T   C 5'   | C 3'       | 5' TCGAG  |  |
|              |  | CA G CT 5' | 3' C      |  |

Figure 10.3. Recognition and cutting sites of some of the most commonly used restriction endonuclease enzymes

The ends of two DNA fragments containing complementary DNA overhangs (so-called „sticky ends”) can form base pairs with each other. Such overhangs are called compatible ends. Aligned compatible ends can be joined by using DNA ligase enzymes (Figure 10.2). The DNA ligase enzyme forms a covalent linkage between the 3'-OH

group of the deoxyribose moiety of one of the DNA ends and the terminal 5'-phosphate group of the adjacent DNA molecule. Similarly to sticky ends, two blunt DNA ends can be ligated too. Thus, two blunt ends are also considered as compatible ends.

As described above, recombinant DNA constructs are formed via the joining of the insert and vector fragments containing compatible ends (Figure 10.2). If the two ends of the vector fragment are compatible with each other (e.g. because they were cut by the same restriction endonuclease), this condition will also allow for the formation of „empty” (non-insert-containing) original circular vector molecules. The undesired formation of such molecules (i.e. the autoligation) can be prevented by treatment of the cut and isolated vector molecules with a phosphatase enzyme, which will lead to the elimination of their terminal 5'-phosphate groups. In the absence of these groups, DNA ligase will be unable to autoligate the two ends of the vector DNA molecule.

## 10.4. Introduction of recombinant DNA constructs into host cells and the identification of recombinant colonies

Under laboratory conditions, the isolated plasmids are introduced into bacterial host cells via a process called transformation (Figure 10.4). One of two principal means is generally applied to achieve an experimentally feasible efficiency of transformation:

- a. Prior to the transformation procedure, bacterial cells can be made „competent” via treatment with solutions containing bivalent cations. Plasmids are then introduced into host cells via a transformation procedure involving co-incubation and a subsequent heat shock step.
- b. Plasmids can also be introduced into the host cell via electroporation. During this procedure, cells co-incubated with plasmids are exposed to short pulses of electric shock.

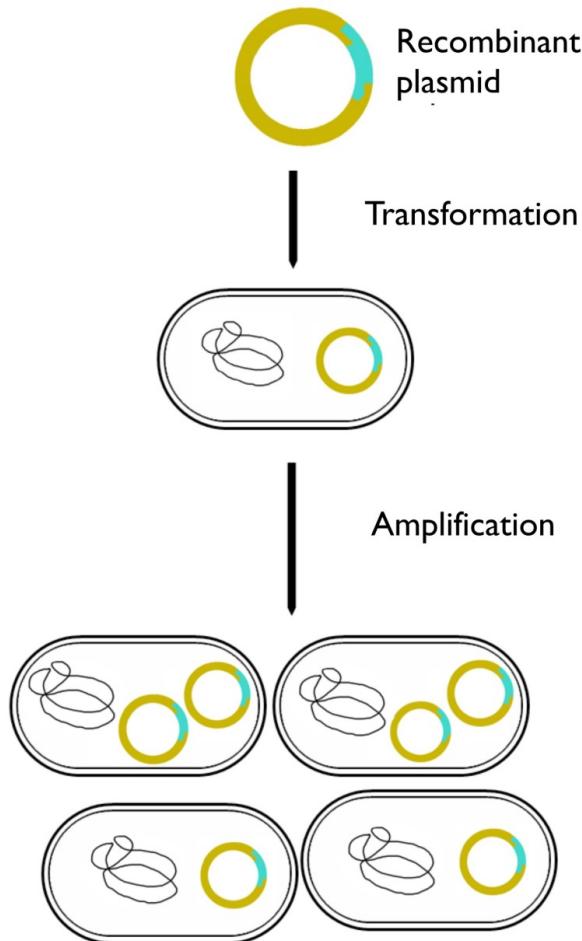


Figure 10.4. Introduction of plasmids into the host cell and their amplification therein

The extent of competence, which can be defined for a given set of experimental conditions, is specified as the number of plasmid-containing colonies grown after transformation (*cfu*, colony forming unit) extrapolated to units of mass of plasmids used for transformation (typically specified in *cfu/µg plasmid*).

Even with the maintenance of the above conditions facilitating the introduction of plasmids into the host cells, only a very small fraction of bacterial cells will stably take up a plasmid. The identification of transformed—i.e. plasmid-containing—cells is made possible by the utilisation of selection marker genes (Figure 10.5). The most commonly used selection markers confer resistance to certain antibiotics. The most commonly used antibiotics and the enzymes inactivating these antibiotics are listed below.

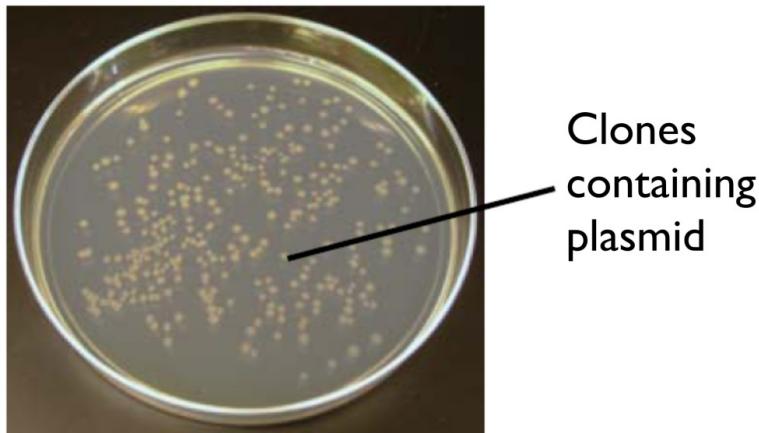


Figure 10.5. Appearance of plasmid-containing bacterial colonies on an antibiotic-containing agar nutrient plate

- Ampicillin, a penicillin derivative, inhibits one of the enzymes involved in the synthesis of the bacterial cell wall. Ampicillin resistance (*amp*<sup>r</sup>) is conferred by an enzyme called  $\beta$ -lactamase, which is located in the periplasm of *E. coli*. The enzyme hydrolyses the lactam ring of penicillin. The hydrolysed penicillin molecule is ineffective.
- Tetracycline binds to one of the proteins of the 30S ribosomal subunit. This molecule inhibits the translocation of the ribosome during translation. The tetracycline resistance gene (*ter*<sup>r</sup>) encodes a membrane protein that prevents the entry of the antibiotic into the host cell.
- Chloramphenicol binds to the 50S ribosomal subunit, thus inhibiting protein synthesis. The chloramphenicol resistance gene (*cm*<sup>r</sup>) encodes the enzyme chloramphenicol acetyl transferase (*cat*). This enzyme is localised in the cytoplasm, and acetylates chloramphenicol by using acetyl coenzyme A. The modified chloramphenicol molecule is unable to bind to the ribosome.

Following transformation, bacterial cells are spread onto a nutrient agar plate containing antibiotics according to the resistance gene contained in the plasmid. Only the plasmid-containing resistant bacteria will be able to grow on the antibiotic-containing agar plate. Ideally, the appearance of separate bacterial colonies can be observed on the agar nutrient plate (Figure 10.5). Theoretically, one colony contains the descendants (clones) of a single bacterial cell that had taken up the plasmid.

Following the selection of plasmid-containing bacterial cells, the next step is the identification of cells harbouring recombinant plasmids, i.e. the plasmids containing the insert of interest. The most widely applied methodologies for the identification of recombinant colonies are the following:

- Insertional inactivation: If the foreign DNA segment (the insert) is inserted within an antibiotic resistance gene, the successful insertion will result in the loss of the corresponding antibiotic resistance. By using this method, the identification of recombinant colonies is not straightforward, as it requires the application of replica techniques.
- $\alpha$ -complementation („blue/white selection“): Numerous vectors contain a segment of the *E. coli lac* operon containing the operator region and the region coding for the first ca. 100 amino acids of the  $\beta$ -galactosidase enzyme (the so-called  $\alpha$ -peptide). This part of the enzyme, whose synthesis can be induced by applying isopropyl-thio- $\beta$ -D-galactoside (IPTG), is able to exert intra-allelic complementation of the activity of the  $\beta$ -galactosidase variant encoded in the genome of the host cell—the latter lacking the first ca. 100 amino acids of the native enzyme. Upon IPTG induction, the bacterial cells harbouring non-insert-containing vector DNA will produce both fragments of the enzyme, resulting in the appearance of blue colonies in the presence of the chromogenic substrate 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside (X-gal). Vectors suitable for  $\alpha$ -complementation contain the multiple cloning site within the region coding for the  $\alpha$ -peptide. The insertion of a foreign DNA segment (insert) will thus disrupt continuity in the region coding for the  $\alpha$ -peptide, and thus abolish  $\alpha$ -complementation. Bacterial colonies containing recombinant (insert-containing) plasmids will thus produce white colonies, while the ones lacking the insert will produce blue colonies.

- c. Recombinant bacterial colonies can also be identified by applying a polymerase chain reaction (PCR, Chapter 10.7) using the bacterial colonies as templates.
- d. The presence of the insert in the plasmid constructs can also be assessed via restriction endonuclease digest of the isolated plasmids.
- e. The presence of the insert in the plasmid constructs can also be assessed via sequencing of the affected region.

## 10.5. Isolation of plasmid DNA

The various plasmid isolation techniques that are currently in use can be divided into three phases:

- a. Growth of bacterial cells;
- b. Harvesting and lysis of bacterial cells;
- c. Purification of plasmid DNA.

One of the most common bacterial host strains used for the *in vivo* amplification of plasmids is *E. coli* XL1-Blue, a derivative of the *E. coli* K12 strain. XL1-Blue cells can be readily transformed with plasmids. They are suitable for  $\alpha$ -complementation analysis, and can also be transduced with filamentous phages. Depending on the amount of plasmid DNA to be isolated, bacterial cells can be grown in various volumes of shaken culture in the presence of antibiotic(s) suitable for the desired selection. By applying the most commonly used, so-called „miniprep” plasmid isolation protocols, 1-10 µg of isolated plasmid can be prepared from 3-5 mL of cell culture volume.

Following bacterial growth, the cells can be pelleted by centrifugation in a microcentrifuge. After disposing of the supernatant, plasmid isolation from the pelleted cells can proceed principally in two different ways:

- a. By applying a „classical” method involving phenol-chloroform extraction and subsequent precipitation of the plasmid by using ethanol (see below in details).
- b. By using a plasmid isolation kit. In this case the isolation of the plasmid is performed using a miniaturised chromatographic column.

The „classical” method for plasmid isolation is composed of the steps described below.

- a. Resuspension of the bacterial pellet in an isotonic solution. In this solution, the lysis of the cells does not yet take place. The ethylene diamine tetraacetate (EDTA) contained in the resuspension solution inhibits the nuclease activity of cellular enzymes via complexation of Mg<sup>2+</sup> ions. Some earlier protocols also applied lysozyme to break down the bacterial cell wall. In addition to the above listed components, the resuspension solution can also contain RNase enzyme in order to break down ribonucleic acids that are later released into the lysate.
- b. Alkaline lysis of the cells by applying an alkaline solution of sodium dodecyl sulfate (SDS) that disintegrates the lipid structure of the cell membrane. In addition, this treatment will denature both proteins and DNA, and keep these molecules dissolved in their denatured form.
- c. Precipitation of dissolved proteins, membrane debris and associated genomic DNA by applying a solution of acidic potassium acetate. The potassium salt of dodecyl sulfate is also insoluble. Thus, this component will also precipitate. Plasmid DNA will still remain dissolved during this step.
- d. Sedimentation of the precipitated components in a microcentrifuge. The clear supernatant will contain the plasmid DNA.
- e. Purification of plasmid DNA using a mixture of phenol and chloroform (care should be taken when working with phenol as it is corrosive to human tissues). To remove the protein content of nucleic acid preparations, a mixture is often used that contains phenol and chloroform in a volume ratio of 1:1, isoamyl alcohol in a volume ratio of 1:24 with regard to the rest of the components, and is saturated with TE pH 8.0 (tris-hydroxymethyl-aminomethane buffer containing 1 mM EDTA (ethylene diamine tetraacetate)). Phenol denatures proteins, and chloroform readily dissolves phenol, which has limited water solubility. When the nucleic acid preparation is

shaken thoroughly with the described mixture and subsequently centrifuged, denatured proteins will be concentrated at the boundary of the upper aqueous and the lower phenol-chloroform phase of higher density. Isoamyl alcohol reduces the frothing associated with the separation procedure.

- f. The addition of ethanol to the aqueous phase containing the plasmid will result in the precipitation of the plasmid DNA, which can thus be subsequently sedimented by centrifugation.
- g. The precipitate containing the plasmid DNA is washed with 70 % ethanol in order to remove the salt content of the preparation.
- h. The plasmid DNA is sedimented repeatedly by centrifugation, and then dissolved in TE solution (see point (e) above for composition). The preparation can be stored on ice or in a freezer. The TE solution may also contain DNase-free RNase enzyme in order to eliminate ribonucleic acids.

During plasmid isolation using commercially-available kits (Figure 10.6), the steps of resuspension, alkaline lysis, precipitation with acidic potassium acetate and subsequent centrifugation (steps (a)-(d) above) are performed similarly to those described for the “classical” method, with the difference that the reagents supplied by the kit manufacturer are used for the procedure. The supernatant resulting from this series of steps will contain the plasmid DNA. This supernatant is loaded on top of a mini-column containing a silicate-based membrane. The mini-column is placed in an Eppendorf tube so that the flow-through can be collected upon centrifugation. Under the applied high ionic strength conditions, the column will bind DNA molecules in the size range of 100 base pairs to 10 kilobase pairs. The column is subsequently washed with a wash buffer and a solution with high ethanol content. The ethanol is then removed via repeated centrifugation. The plasmid DNA is then dissolved in TE or a similar low ionic strength solution.

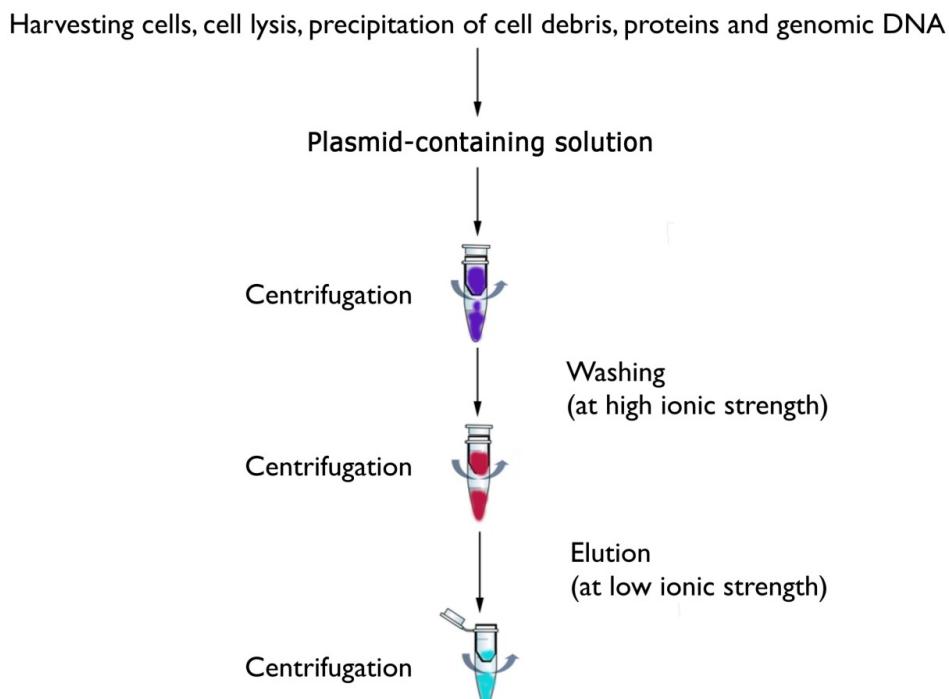


Figure 10.6. Isolation of plasmid DNA by using a mini-column containing a DNA-binding resin

## 10.6. Analysis of plasmid DNA by gel electrophoresis

Agarose gel electrophoresis (discussed also in Chapter 7) is the most commonly used method for the size- and shape-based separation of DNA molecules comprising several hundred or more base pairs, including plasmid DNA molecules (Figure 10.7). Agarose is one of the main components of agar extracted from the cell wall of red algae.

Agarose is a linear polysaccharide composed of galactose and anhydro-galactose units. The agarose gel possesses a number of features that make it especially advantageous for the purposes of gel electrophoresis. The gel is hydrophilic, chemically inert and stable. It does not bind the dye molecules used to visualise DNA molecules that are separated in the gel. The three-dimensional matrix of the agarose gel is brought about by non-covalent bonds formed between polysaccharide units.

As the structure of the agarose gel is held together by non-covalent bonds, the gel undergoes a phase transition at elevated temperature, and forms a sol state. The gel is prepared by mixing agarose powder into a running buffer, with subsequent formation of the sol state at high temperature, casting and subsequent cooling.

The pore size of the resulting gel depends on agarose concentration. The pore size determines the size range of DNA molecules that can be efficiently separated in the gel. For agarose gel electrophoresis, gels with agarose concentrations of 0.5-3 w/v % are generally used. The lower and higher ends of this concentration range are applicable in the case of larger and smaller DNA molecules, respectively.

DNA is a negatively charged molecule. Therefore, the electric field applied during electrophoresis will cause migration of DNA molecules towards the positive pole, i.e. the anode. Therefore, the gel is placed into the electrophoresis tank in an orientation that the sample loading wells will be towards the negative pole (i.e. the cathode) (Figure 10.7).

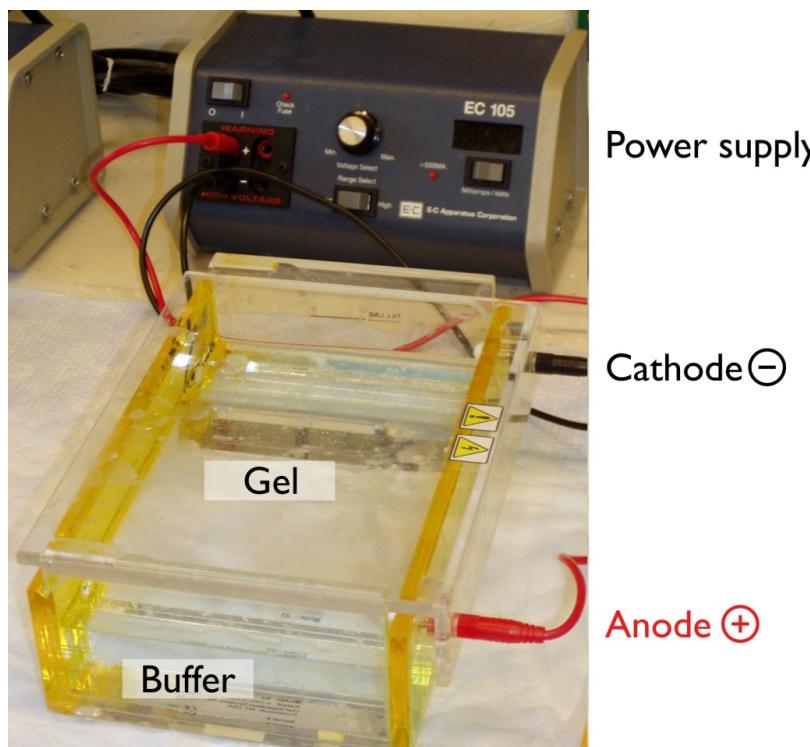


Figure 10.7. Gel electrophoresis apparatus

Before loading onto the gel, the plasmid samples are treated with a loading buffer solution. The loading buffer contains glycerol (or high molecular weight polysaccharides) in order to increase the density of the sample and thus facilitate efficient sample loading. This buffer also contains a loading dye (most often bromophenol blue), which has slightly higher mobility than the DNA molecules to be separated. Thus, by inspecting the progress of the blue spot, the proper termination time of the run can be determined.

Samples treated with the sample loading buffer are layered into the sample loading wells of the gel, below the surface of the loading buffer. The glycerol content of the loading buffer prevents dispersion by increasing the density of the sample (Figure 10.8). Besides the samples to be investigated, DNA molecular weight markers are usually applied (often at both sides, and also in the centre). These markers contain a series of linear DNA molecules of known length.

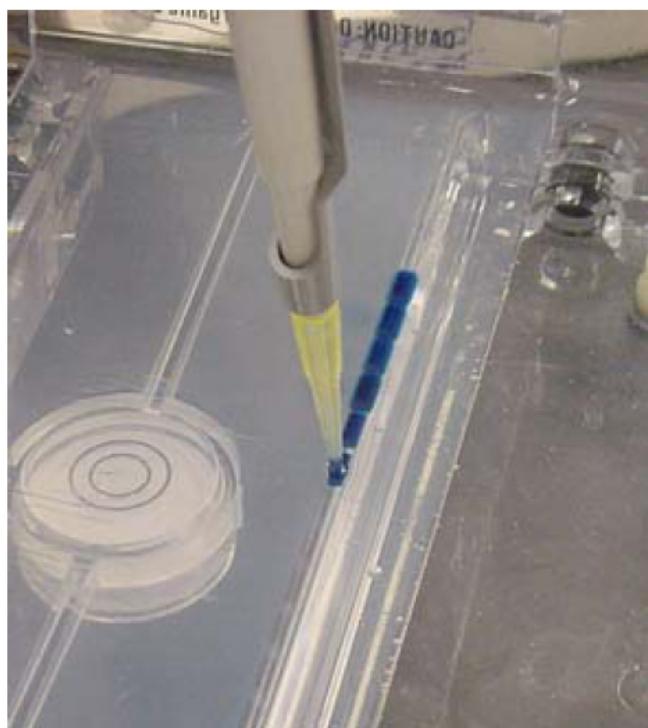


Figure 10.8. Loading samples on an agarose gel

Ethidium bromide (Figure 10.9), the DNA dye traditionally used for visualising DNA molecules within the gel, has a mutagenic effect. Ethidium bromide is a ring-containing compound that is able to intercalate between bases within the double helix of DNA. Therefore, the dye can cause insertions or deletions during replication. In contrast to ethidium bromide, the SYBR Safe dye is non-toxic. The DNA complexes of both dyes produce orange-coloured fluorescent light when illuminated by UV radiation. This effect serves as the basis of the in-gel detection of DNA molecules. The gel containing the stained DNA molecules is photographed in an UV-transillumination apparatus.

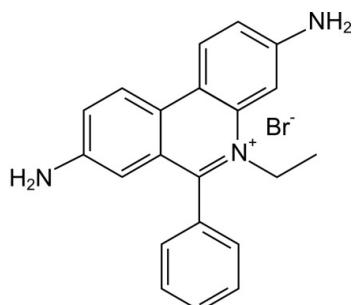


Figure 10.9. Ethidium bromide

Besides their size, the electrophoretic mobility of DNA molecules is also significantly affected by their shape. Superhelically packed circular plasmid DNA has a compact structure, and its hydrodynamic size is much smaller—and its electrophoretic mobility is therefore greater—than that of linear DNA molecules of the same size, as the latter form a freely moving entropic chain (Figure 10.10). When one of the strands of the superhelical plasmid DNA is cut, it will adopt a so-called relaxed circular form. The mobility of this form is smaller than even that of linear DNA. These phenomena are readily apparent in the electrophoretic image shown in Figure 10.10. From the intensity of the two bands appearing in the plasmid preparation run in the left lane, it can be inferred that the preparation largely contains the superhelical DNA form that is most abundant within the cells. A smaller amount of the plasmid DNA is in the relaxed circular form. In the other samples, treatment with a restriction endonuclease enzyme that cuts both DNA strands of the plasmid at a single recognition site results in the appearance of the linearised form of the plasmid, which has medium mobility.

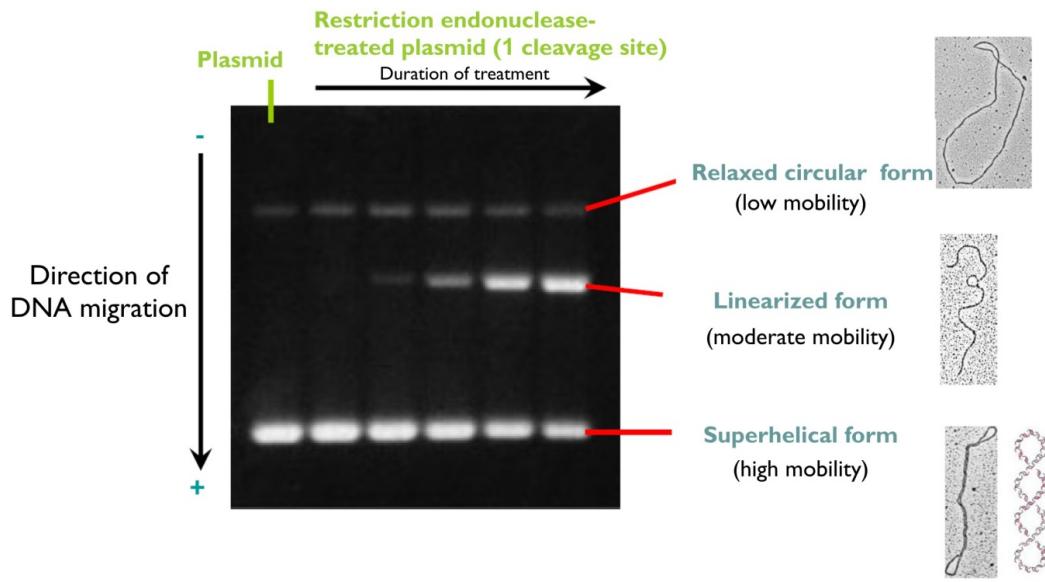


Figure 10.10. Gel electrophoretic image of plasmid DNA

## 10.7. Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) can be used for the selective amplification of a specific segment (target sequence or amplicon) of a DNA molecule. The DNA to be amplified can theoretically be present in a very small amount—even as a single molecule. The PCR reaction is carried out *in vitro* and, as such, it does not require a host organism. The size of the DNA region amplified during the PCR reaction typically falls within the range of 100 bp to 10 kbp.

PCR is based on the reaction scheme described as follows (see also Figure 10.11). First, a heat-induced denaturation of the target DNA sequence (template, panel *a*) is performed in order to separate the constituent complementary DNA strands. Then, short single-stranded DNA molecules (oligonucleotide primers) are added that are complementary to the flanking regions of the target sequence. Cooling of the sample allows annealing (panel *b*). Subsequently, the strand elongation activity of a DNA polymerase enzyme leads to the formation of new DNA strands (so-called primer extension products) starting from the 3' end of the annealed primers (panel *c*). After repeated heat denaturation and cooling, the primers will be able to anneal both to the original template molecules and to the primer extension products (panel *d*). In the latter case, the length of the nascent DNA strand will be limited by the primer extension product now serving as a template strand. This way, the resulting “end-product” strands will comprise the DNA segment defined by the template and the flanking primers (panel *e*). In further denaturation–annealing–synthesis cycles, the end-product strands will serve as templates for the synthesis of additional end-product strands. Therefore, the amount of these molecules will grow exponentially with the number of reaction cycles (panel *f*). Thus, the final result of the reaction will be a large amount of end-product molecules comprising the sequence flanked by the predefined primers. This highlights one of the crucial advantages of the PCR technique: via the design of primers, we can fully control which segment of the template DNA will be amplified—with only a few practical limitations.

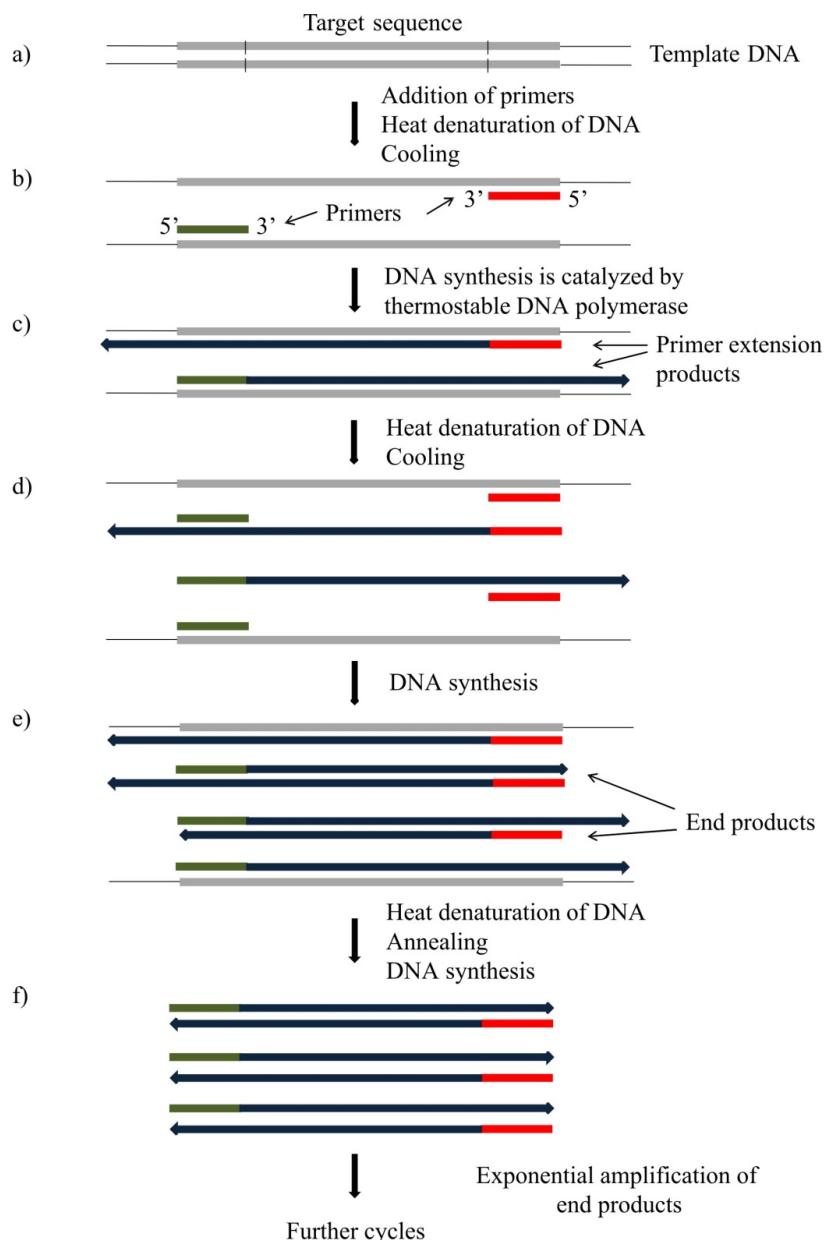


Figure 10.11. Cycles of the polymerase chain reaction (PCR)

In order to successfully perform the reaction described above, the following solution components are necessary:

- DNA molecules that serve as template for the reaction. The amount of the template can be very low—in principle, the reaction can start even from a single template molecule. Another advantage of PCR is that the selective amplification of the desired DNA segment can be accomplished even using a heterogeneous DNA sample as template.
- A pair of oligonucleotides serving as primers. The 3' ends of the oligonucleotides must be able to anneal to the corresponding strands of the template. A further advantage of PCR is that the 5' end of the applied primers may contain segments that do not anneal to the original template. These regions of the primers may be specific engineered sequences or even contain labelling or other modifications, which will be present in the end product and thus facilitate its further analysis and/or processing. (For instance, recognition sites of restriction endonucleases can be incorporated in order to facilitate the subsequent cloning of the PCR product.)
- The DNA polymerase enzyme catalysing DNA synthesis. As the heat-induced denaturation of the template is required during each cycle, heat stable polymerases are usually applied that originate from thermophilic organisms (e.g. *Thermus aquaticus* (Taq) or *Pyrococcus furiosus* (Pfu) DNA polymerase).

- d. Deoxyribonucleoside triphosphate (dNTP) molecules that serve as building blocks for the DNA strands to be synthesised. These include dATP (deoxyadenosine triphosphate), dGTP (deoxyguanosine triphosphate), dTTP (deoxythymidine triphosphate), and dCTP (deoxycytidine triphosphate).
- e. A buffer providing optimal reaction conditions for the activity of DNA polymerase. Among other components, PCR buffers contain bivalent cations (e.g.  $Mg^{2+}$  or  $Mn^{2+}$ ).

For an effective polymerase chain reaction, it is necessary to change the temperature of the solution rapidly, cyclically and in a wide range (see below). This can be achieved by using a programmable instrument containing a thermoblock equipped with a Peltier cell. To achieve effective heat exchange, PCR reactions are performed in thin-walled plastic tubes in small reaction volumes (typically, in the order of 10-200  $\mu$ l). The caps of the PCR tubes are constantly held at high temperature by heating the lid of the thermoblock, in order to prevent condensation of the reaction mixture in the upper part of the tubes. In the absence of a heated lid, oil or wax can be layered on top of the aqueous PCR samples in order to prevent evaporation.

The programmed heat profile of a PCR reaction generally consists of the following steps:

- a. Initial denaturation of the template, performed at high temperature (typically, around 95°C).
- b. Denaturation: Heat-induced separation of the strands of double-stranded DNA molecules at high temperature (typically, around 95°C).
- c. Annealing: Cooling of the reaction mixture to a temperature around 45-65°C in order to facilitate the annealing of the oligonucleotide primers to complementary stretches on template DNA molecules.
- d. DNA synthesis: This step takes place at a temperature around the optimum of the heat-stable DNA polymerase (typically, 72°C), for a time period dictated by the length of the DNA segment to be amplified (typically, 1 minute per kilo-base pair).
- a. Steps (b)-(d) are repeated typically 20-35 times, depending on the application.
- a. Final DNA synthesis step: After completion of the cycles consisting of steps (b)-(d), this step is performed at a temperature identical to that during step (d) (72°C), in order to produce complementary strands for all remaining single-stranded DNA molecules.

PCR reactions are widely applied in diverse areas of biology and medical science. In the following, we list a few examples.

- a. Molecular cloning, production of recombinant DNA constructs and hybridisation probes, mutagenesis, sequencing;
- b. Investigation of gene function and expression;
- c. Medical diagnostics: identification of genotypes, hereditary disorders, pathogenic agents;
- d. Forensic investigations: identification of samples and individuals based on DNA fingerprints (unique individual DNA sequence patterns);
- e. Evolutionary biology, molecular evolution, phylogenetic investigations, analysis of fossils.

## 10.8. Site directed *in vitro* mutagenesis

Various genetic, biochemical and molecular cell biological investigations often require the designed alteration of specific DNA segments. The process is called mutagenesis. These alterations are most often targeted at regulatory or protein-coding regions of genes. The aim of mutagenesis is the alteration of the functioning of target genes, or the alteration of the sequence, structure and function of their encoded RNA and protein products. Modifications introduced into DNA molecules can be classified into three groups:

- a. Deletion: the elimination of a given number of nucleotides from a segment of DNA;
- b. Insertion: the addition of a given number of nucleotides to a segment of DNA;

- c. Substitution: the exchange of certain nucleotides with others without changing the length of the affected segment of DNA.

Many different techniques are available for performing *in vitro* mutagenesis. Below we introduce two of the most often used methodologies.

- a. Site-directed mutagenesis via PCR (Figure 10.12). With the help of this method, linear DNA segments carrying the desired mutation can be created. Mutagenesis is achieved via the application of the PCR technique. Primers are designed in a way that one of them will anneal to one of the flanking regions of the DNA segment to be amplified and mutagenised, whereas the other primer—the one that contains the mutant sequence—will hybridise to the region to be mutated. The product of this PCR reaction will contain the sequence of the primers, i.e. also the mutation that was present in one of the primers. However, the product of this first PCR reaction will comprise only the region of the original DNA that was flanked by the primers, i.e. the region ranging from the site of the mutation to one end of the DNA segment of interest. If necessary, a second PCR reaction can be performed in which the product of the first PCR reaction will serve as one primer, whereas the other will anneal to the remaining end of the DNA segment to be amplified. The resulting mutant DNA product is inserted into a plasmid vector and amplified in bacteria. Bacterial colonies containing the plasmid are identified via selection procedures. Plasmids are isolated from multiple colonies, and DNA sequencing is used to identify the ones carrying the desired mutation.

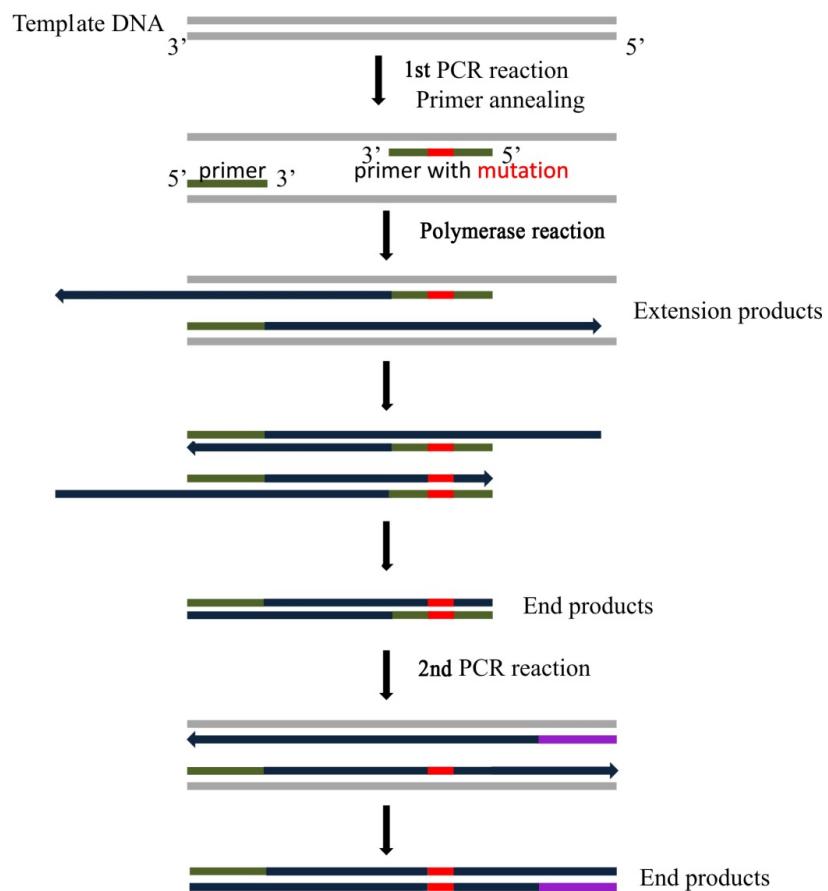


Figure 10.12. PCR-based site-directed mutagenesis method (see text for details)

- b. Selective template degradation method (Figure 10.13). In this method, mutagenesis is performed in a DNA segment that has already been incorporated into a plasmid. A prerequisite for this procedure is that the guanine bases of the plasmid should be methylated. In order to achieve this, plasmids are amplified in a bacterial strain that methylates its DNA content. Mutagenesis is performed via PCR reaction. However, in contrast to the classical PCR technique, both primers applied in this method hybridise to the same region of the DNA template, and both of them contain the desired mutation. (Thus, the primers are complementary to each other.) The PCR reaction using these primers is set up so that the entire plasmid will be copied. The reaction will thus result in a mixture containing the original template plasmid and also the newly synthesised plasmids containing the

mutation. Subsequently, a special endonuclease enzyme (DpnI) is added to the reaction mixture. DpnI cleaves only the methylated DNA molecules. Due to the activity of this endonuclease, the original plasmid will be digested into fragments. However, the mutation-carrying plasmids synthesised during the PCR reaction will remain intact. Mutant plasmids are subsequently transformed into a bacterial strain (most often, *E. coli*). Amplified plasmids are then isolated and analysed using methods described in Chapters 10.4-10.6.

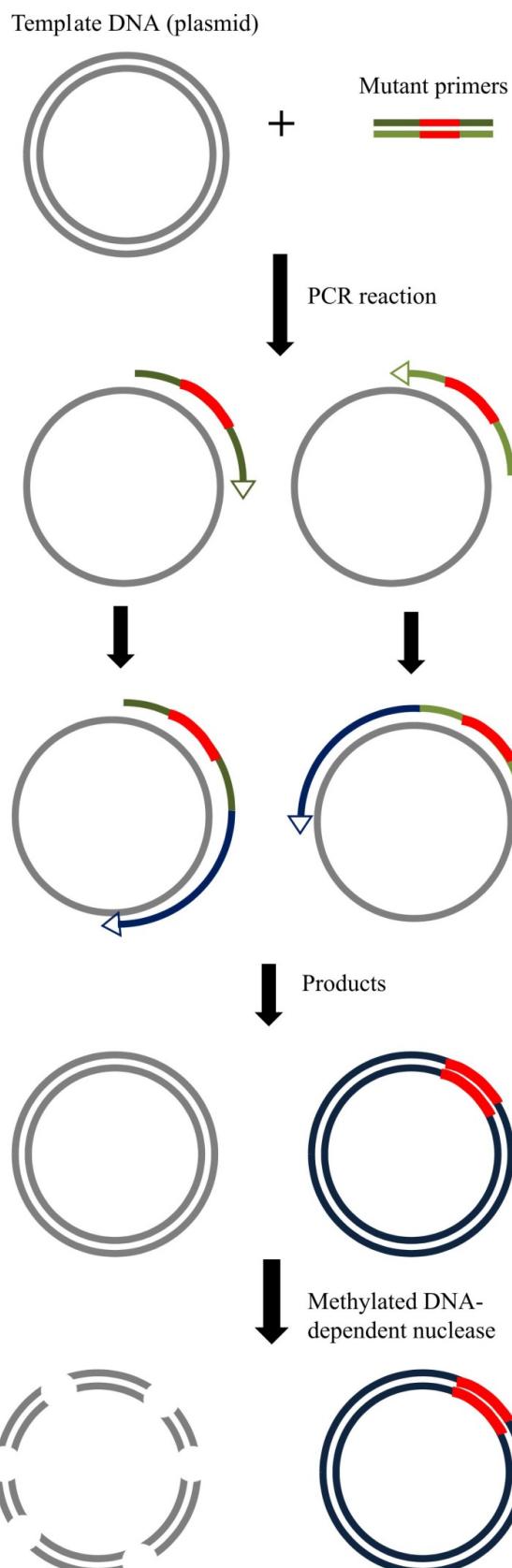


Figure 10.13. Mutagenesis method applying selective degradation of template

## 10.9. DNA sequencing

Determination of the base sequence of DNA molecules (DNA sequencing) is necessary in various cases. Below we provide an overview of DNA sequencing methodologies.

### *Principle of chain-termination DNA sequencing*

The principles of the Sanger-Coulton DNA sequencing method were laid down in the late 1970s in order to provide a readily executable, rapid, and relatively cheap method to determine the nucleotide sequence of DNA molecules. During the procedure, the two strands of the DNA to be sequenced are separated via high-temperature treatment, and then a short single-stranded DNA molecule (a primer oligonucleotide), is hybridised upstream (5') of the region of interest within the strand to be sequenced. Four different sequencing reaction mixtures are assembled. All reaction mixtures contain the DNA template, the sequencing primer, the DNA polymerase enzyme, and equal amounts of the four deoxynucleoside triphosphate (dNTP) building blocks. In addition, each of the four parallel reactions contains one of four dideoxynucleoside triphosphate molecules (ddNTPs). In the ribose ring of ddNTPs, the 3' carbon atom has an attached H atom instead of the -OH group present in “normal” dNTPs (Figure 10.14).

As an example, let us consider the reaction that contains ddATP. The polymerase enzyme will start the synthesis of the complementary DNA strand from the 3' end of the oligonucleotide primer, using both the dNTP and ddATP molecules. Juxtaposed to thymine residues present in the template DNA strand, the nascent DNA strand will have either dATP or ddATP incorporated—with probabilities depending on the ratio of concentrations of the two corresponding nucleotide blocks. If dATP is incorporated, strand synthesis will continue. However, if ddATP is incorporated, the synthesis of the new strand will halt due to the absence of a reactive 3'-OH group at the strand terminus. This way, new DNA strands of various length will be synthesised during the reaction. The lengths of these strands will reflect the position and distance of thymine bases of the template relative to the 5' end of the primer (Figure 10.15). In the case of parallel reactions performed in the presence of ddGTP, ddCTP or ddTTP, the lengths of the synthesised new strands will reflect the positions of cytosine, guanine and adenine bases in the template strand, respectively.

### Dideoxynucleoside triphosphates

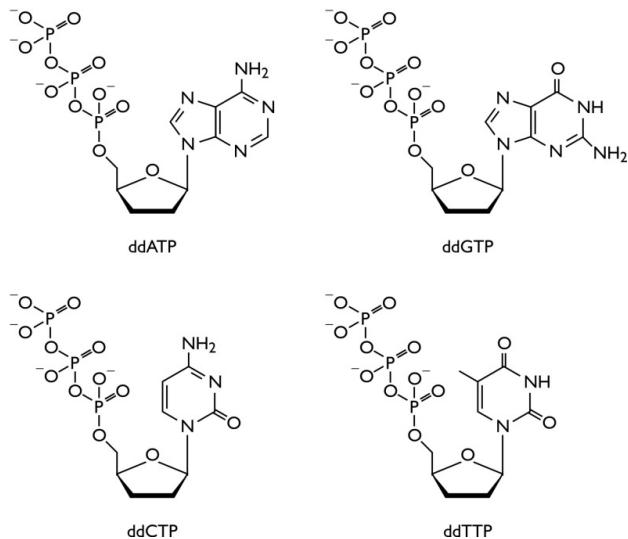


Figure 10.14. Structure of dideoxy nucleoside triphosphates

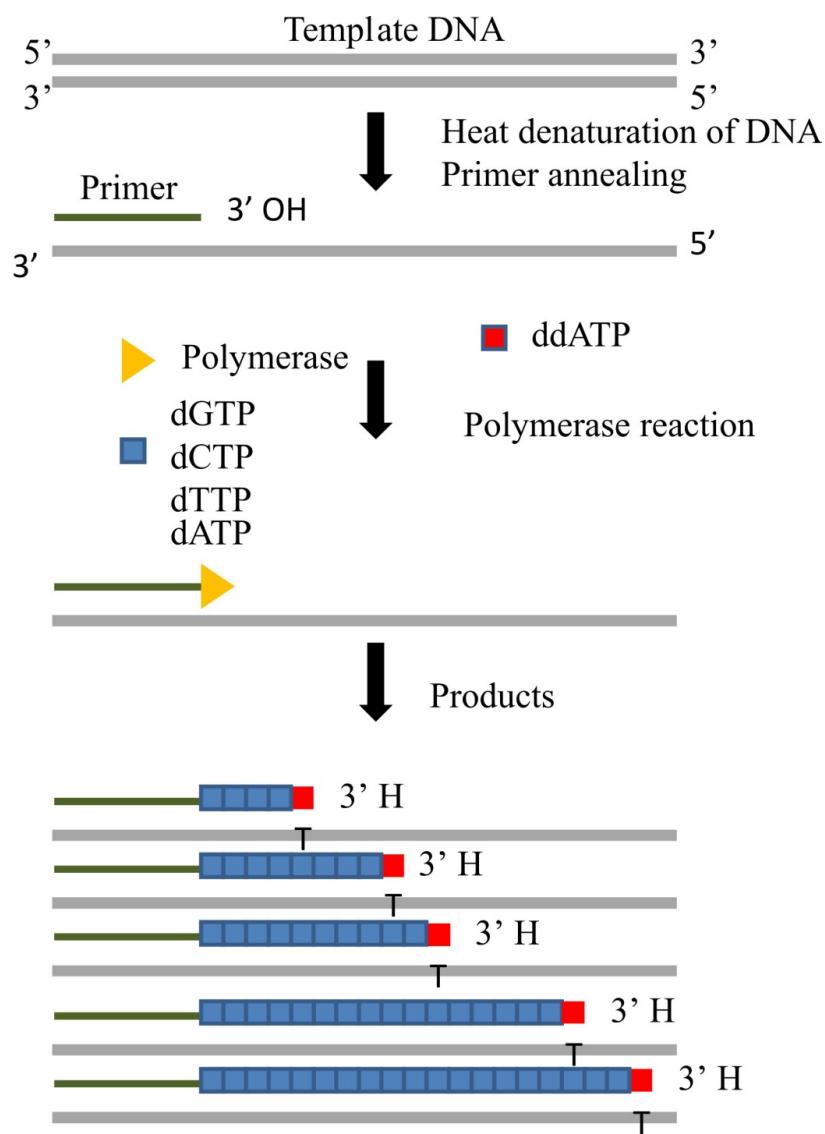


Figure 10.15. DNA sequencing with the chain termination method (one of four reactions is shown; see text for details)

#### *Size-based separation of DNA strands in a polyacrylamide gel*

The knowledge of the sizes of the DNA strands synthesised in the four different sequencing reactions enables the determination of the base sequence of the template DNA molecule. Agarose gel electrophoresis (Chapter 10.6) is suitable only for the separation of large (mostly double-stranded) DNA chains. This technique can detect a difference of at least 100 base pairs. High-resolution separation of shorter DNA strands can be achieved via polyacrylamide gel electrophoresis (PAGE) (see also in Chapter 7) combined with urea denaturation. Similarly to the situation during agarose gel electrophoresis, in PAGE the negatively charged DNA molecules are separated in an electric field while they are moving towards the anode. However, in the polyacrylamide gel that has a much smaller pore size than the agarose gel, it is possible to resolve much smaller size differences (down to a single nucleotide unit). Polyacrylamide gels are created using synthetic acrylamide and bisacrylamide ( $N,N'$ -methylenebisacrylamide) mixtures. (Acrylamide is harmful by inhalation or skin contact, and thus it should be handled with care.) Acrylamide undergoes radical polymerisation in an aqueous environment, resulting in polymer chains of various length. Bisacrylamide is a crosslinking reagent that is able to link individual acrylamide chains. The crosslinking results in a three-dimensional gel matrix. The rate of polymerisation and crosslink formation is greatly enhanced by the presence of free radicals. Therefore, during the creation of the gel, ammonium persulfate (APS) and tetramethyl-ethylene-diamine (TEMED) are used to form and stabilise free radicals, respectively. The pore size of polyacrylamide gels—and thus, the resolving capacity of the gels—will be determined by the ratio of concentrations of acrylamide and bisacrylamide. In the case of sequencing gels, acrylamide is applied at a concentration of 4–8 w/v

%, whereas bisacrylamide is applied at a concentration of 0.2 w/v %. In this concentration range, DNA molecules with lengths of up to 600 bases can be resolved with single-base resolution. As single-stranded DNA molecules may possibly form different secondary structures with themselves or with other DNA molecules of complementary sequence, 7 M urea is used in the gel in order to avoid this possibility. Urea is a DNA denaturing agent that disrupts base pairs, thus ensuring that truly single-stranded DNA molecules will be present in the gel.

Sequencing gels are usually run in a vertical arrangement: the lower part of the gel is oriented towards the positively charged anode, whereas its upper part is oriented towards the negatively charged cathode (Figure 10.16). The upper end of the gel contains the sample loading wells. The samples produced in the sequencing reactions are treated with a loading solution before loading onto the gel. The loading solution contains tracking dyes (bromophenol blue and xylene cyanol) as well as formamide. Formamide is an organic solvent that denatures the polymerase enzyme used in the sequencing reactions, thus stopping the reactions. As the newly synthesised DNA strands are base paired to the template, these strands must be separated prior to the run. The samples to be run are therefore heat-denatured at 95°C for several minutes. The presence of formamide significantly decreases the heat stability of double-stranded DNA, thereby aiding denaturation and preventing the re-formation of base pairs at the lower temperature applied subsequently during the run. The treated samples that originate from the four sequencing reactions containing different ddNTP reagents are loaded into neighbouring wells of the sequencing gel. Following loading, the gel is run in TBE (TRIS (tris-hydroxymethyl-aminomethane), borate, EDTA (ethylene diamine tetraacetate)) buffer at 50°C in a high electric field (5-20 V/cm). In the gel, the DNA strands of smaller size will run faster, whereas increasing size will be associated with reduced electrophoretic mobility. The run is performed until the bromophenol blue tracking dye reaches the lower end of the gel.

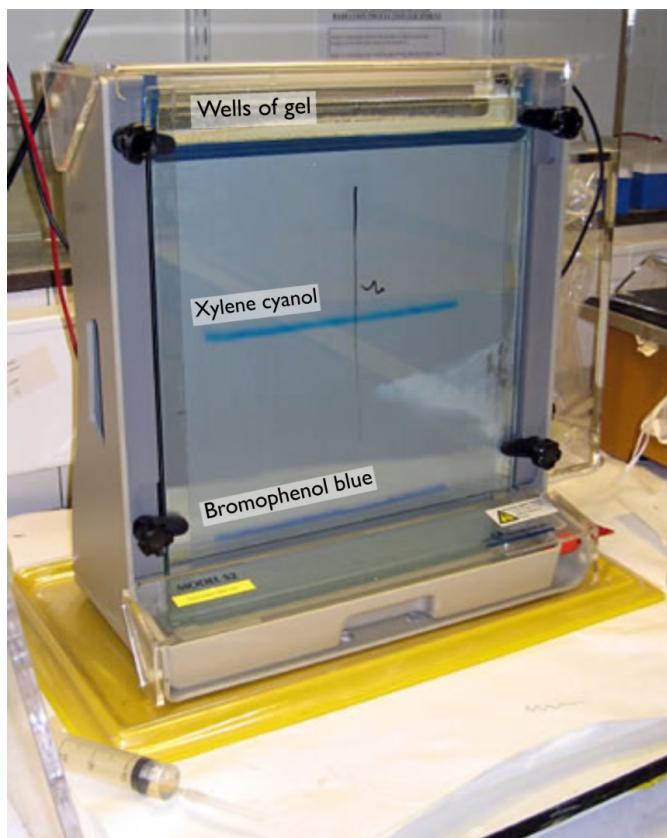


Figure 10.16. Polyacrylamide gel electrophoresis

#### *Detection of DNA and determination of the base sequence*

In the sequencing gel, the detection of DNA is achieved by applying a sensitive radiometric method. During the sequencing reaction, the DNA strands are labelled by using  $\alpha$ - $^{32}\text{P}$ -dATP ( $\alpha$ - $^{32}\text{P}$ -deoxyadenosine triphosphate). Following the run, the gel is dried and placed onto a radiosensitive photo paper.  $^{32}\text{P}$  is a  $\beta$ -emitting radioactive isotope. During  $\beta$ -decay, an electron leaves the nucleus of the isotope. As the electrons interact with the photo paper, the film will darken: thus, the “fingerprint” of DNA bands will become visible (Figure 10.17). Recently,

photo papers have mostly been replaced by the usage of *phosphorimager* equipment, which is reusable and offers enhanced sensitivity. Radioisotope techniques have the advantages of low cost and very high sensitivity. However, their application presents a considerable health hazard (due to carcinogenicity), and thus requires special care and attention.

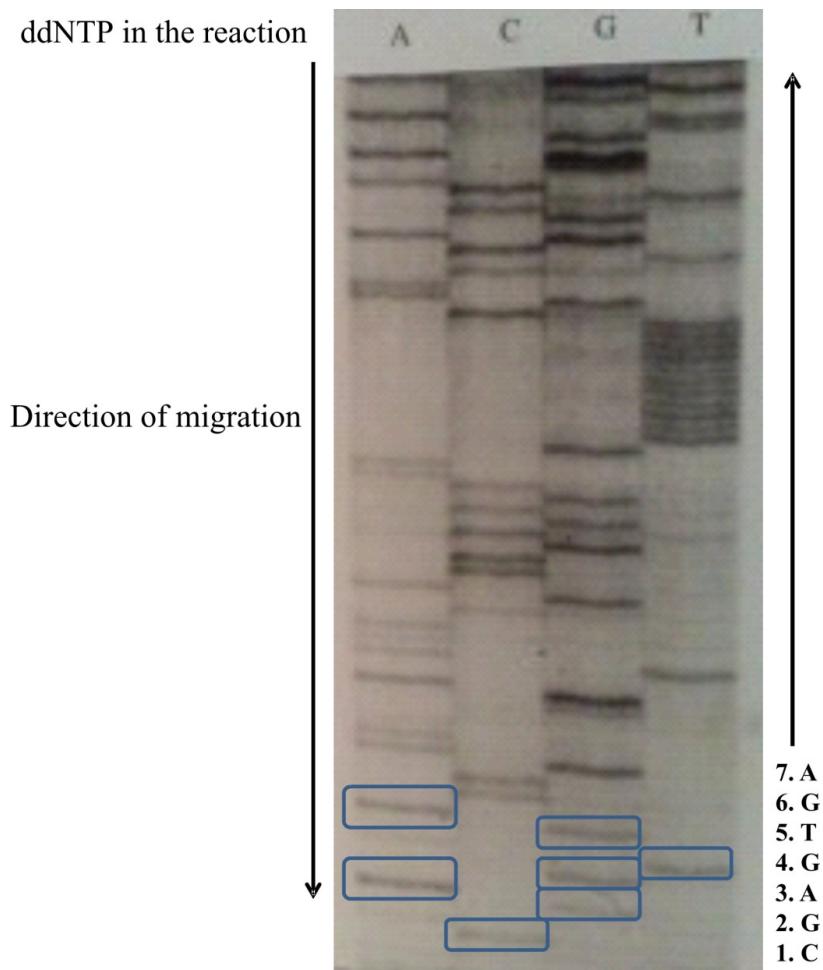


Figure 10.17. Reading of the base sequence of DNA. The pattern of product bands, read from bottom to top, reflects the base sequence of the DNA strand synthesised during the sequencing reaction (which is complementary to that of the template strand).

Another opportunity for the detection of DNA molecules is provided by fluorescent tracers linked to the 5' end of the primer used in the sequencing reaction. When illuminated at a specific wavelength, fluorescent dyes (fluorophores) enter an excited state. The molecules then return to the ground state, accompanied by light emission. This is the principle of the phenomenon of fluorescence. The wavelength of the emitted light is higher—i.e. its energy is lower—than that of the absorbed light. If the gel is illuminated with light absorbed by the applied dye molecules, the fluorescent signals can be detected at the locations of the gel that contain labelled DNA molecules.

#### *Automated DNA sequencing*

In the currently most widely applied method, the fluorescent dyes are covalently linked to the ddNTP molecules (Figure 10.18). The fluorophores are attached to ddNTPs in a way that they do not interfere with the enzymatic incorporation of the given ddNTP molecule into the nascent DNA strand. The four different ddNTP molecules have four different fluorophores attached, characterised by different absorption and emission wavelengths. This setup does not need four different sequencing reactions: the four different labelled ddNTP molecules are added to a single sample at the same time. After performing the sequencing reaction using the DNA polymerase enzyme, samples are loaded in a single well of the gel. Following gel electrophoresis, the gel is placed into a reader that is capable of separately detecting the signals of all four fluorescent ddNTP conjugates. By reading in the direction towards the top of the gel, the base sequence of the template DNA strand can be determined.

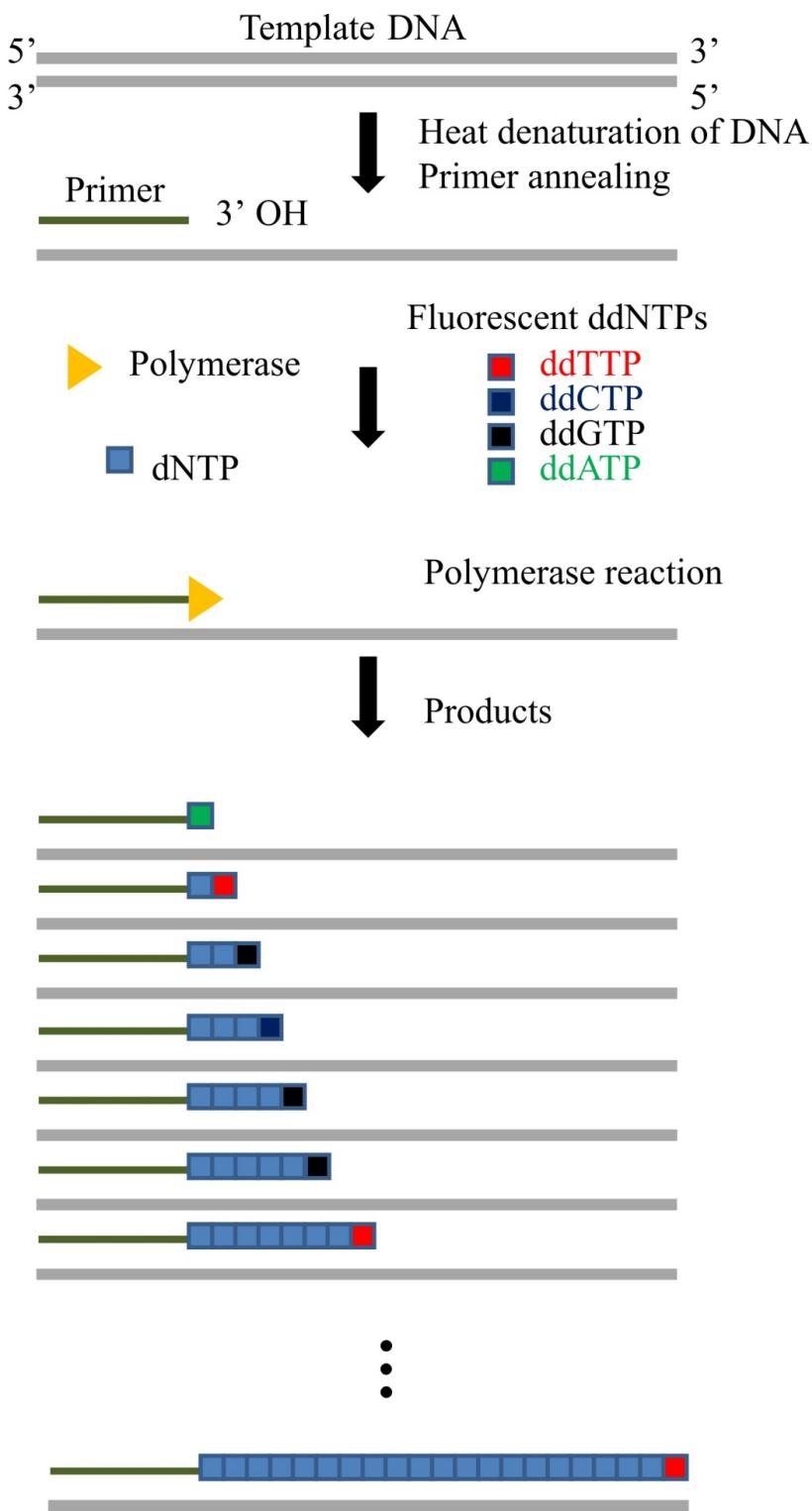


Figure 10.18. Automated sequencing of DNA using fluorescently-labelled ddNTP deoxyribonucleotide analogues. A key advantage of the method is that all four (differently labelled) ddNTPs are present in a single reaction mixture, enabling the determination of the base sequence of DNA.

The use of fluorescent dNTPs has enabled the full automation of the sequencing procedure. Instead of slab gel electrophoresis, the DNA strands synthesised during the sequencing reaction are separated via capillary gel electrophoresis (Figure 10.19). During capillary gel electrophoresis, the gel matrix is contained in a capillary with a length of 50–70 cm and an inner diameter of 50–100 µm. The sample is loaded into the capillary, and then an electric voltage is applied between the ends of the capillary. Thus, the DNA molecules will start to migrate towards

the positively charged anode. During migration, the gel matrix will exert a molecular sieving effect, and thus the molecules of smaller sizes will migrate faster. At the anode-oriented end of the capillary, the fluorescent signals linked to the four different ddNTP molecules are continuously monitored and recorded via a computer. The result of the continuous detection will be a chromatogram on which the time-dependent passage of different fluorophores is tracked. This way, the sequence of the template DNA molecule can be determined (Figure 10.20).

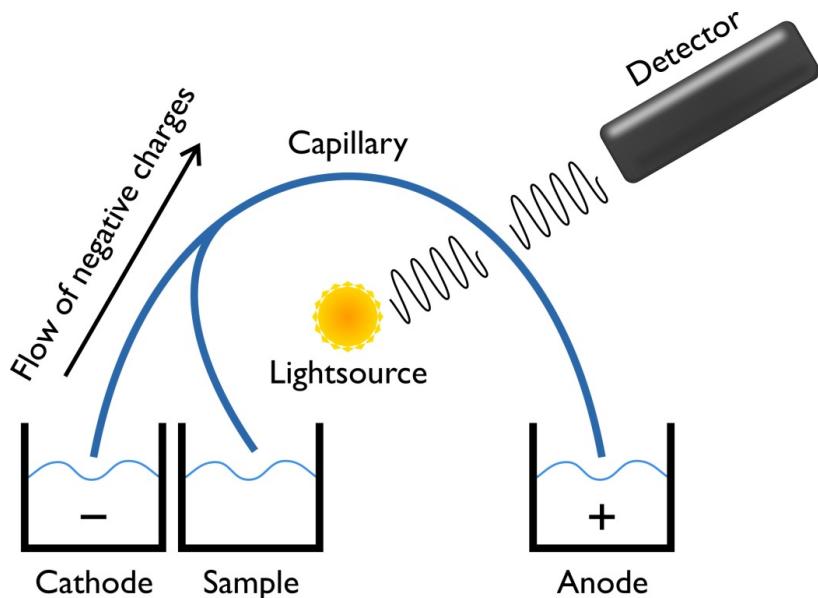


Figure 10.19. Capillary gel electrophoresis

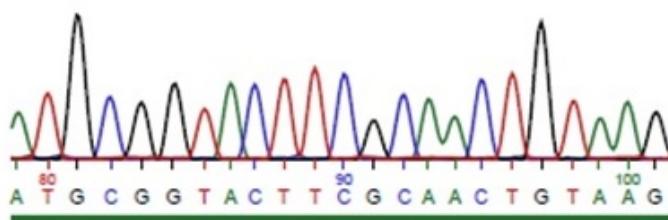


Figure 10.20. Chromatogram produced via a capillary gel electrophoretic method

#### *Second-generation sequencing methods*

With the help of a fully automated procedure, a DNA stretch of about 1000 bases can be read. With an automated apparatus working at full capacity (which can typically run 96 samples at a time), ca. 5 million bases can be read in one day. The size of the human genome is ca. 3.2 billion base pairs. In an ideal case, considering the operating time of one sequencing apparatus, it would take around 3 years to read the entire human genome. In recent years, so-called second-generation sequencing techniques have emerged. These techniques represented a major breakthrough in the field of DNA sequencing by enabling the parallel reading of as many as  $10^5$ - $10^6$  different DNA samples in a single experiment. These procedures do not require the laborious and time-consuming size-based separation of DNA molecules. Production and development of second-generation DNA sequencing apparatuses has become a remarkably big line of industry. Many different biotechnological companies provide second-generation DNA sequencers applying various methodological principles.

---

# Chapter 11. Bioinformatics

by László Nyitrai

## 11.1. Introduction

In the last decade, more biology-related information has accumulated than in the preceding two and a half thousand years of the history of science. This new surge of information mostly consists of nucleic acid and protein sequences, primarily due to the fact that DNA sequencing has become a routine technique after the recombinant DNA revolution. Bioinformatics has emerged as a new field at the interface of informatics and molecular biology in the mid-1980s, with the aim of storing and analysing the huge amount of data provided by DNA sequencing. **Bioinformatics combines mathematical algorithms, computer sciences and statistics (i.e. informatics methods) to derive knowledge from computational analysis of experimental biological data.** From a molecular biological perspective, bioinformatics mostly deals with the storage, retrieval, and analysis of nucleic acid and nucleic acid-derived amino acid sequences of proteins. It has several specific subfields. For instance, structural bioinformatics deals with the *in silico* analysis of the three-dimensional structure of macromolecules. Beyond sequencing, a massive amount of data is produced by many other so-called high-throughput (HTP) methods that can be managed only by bioinformatics. These HTP methods include, just to mention a few, gene expression analysis, electrophoresis and mass spectrometry that generate data to establish genetic, metabolic, signal transduction, protein-protein and other interaction pathways and networks.

Bioinformatics provides the core toolbox for the emerging new field of systems biology. Systems biology aims to understand biology by a holistic approach and is based, among other things, on the enormous datasets supplied by HTP methods. These approaches expand the traditional reductionist approach of molecular biology. The „omics” fields are part of system biology that started with **genomics** (the genome is the full complement of genetic material within an organism) followed by other fields of study, named using language neologisms as **proteomics** (large-scale study of the **proteome**, the full complement of proteins within an organism), **transcriptomics** (the **transcriptome** is a full complement of transcribed RNA within an organism or cell type or a physiological state of a particular cell), **interactomics** (study of the  **interactome**, protein-protein interactions within an organism or cell). One could continue with an „omics” list to study the complete set of small-molecule metabolites (metabolome), the complete set of lipids (lipidome), the entire complement of carbohydrates (glycome), the full set of protein kinase enzymes (kinome) and so on.

In this chapter, we will describe the so-called primary databases that contain nucleic acid and protein sequences as well as three-dimensional structures of macromolecules and their complexes. Moreover, we will give an introduction to *in silico* sequence (and structure) analysis. The role of bioinformatics in molecular cloning experiments (such as restriction mapping of DNA constructs, design of oligonucleotide primers) will be covered only briefly. (More details of recombinant DNA technological methods can be found in Chapter 10.) The first steps in sequence analysis are **similarity searches** and **sequence alignments**; programs to perform these analyses will be described. Data from sequence alignments can be used to construct phylogenetic trees and to infer evolutionary relationships among sequences (and among species). Principles of molecular evolution are not covered in this e-book. *In silico* methods will be discussed that are used to predict structural and functional motifs within nucleic acid and protein sequences. **We must keep in mind that most of the sequence analysis data are predictions, and laboratory experiments should be conducted to validate them.**

Although the three-dimensional structure (conformation) of a protein is determined by its amino acid sequence (recall the conclusive **Anfinsen experiment** proving that the polypeptide chain spontaneously folds into its native three-dimensional shape), currently this information can only be partially inferred from the sequence. *Ab initio* protein structure prediction is still in its infancy. On the contrary, the visualisation of protein conformation is a relatively simple task. If the atomic-resolution structure of a protein (or a nucleic acid or their complexes) has previously been determined (by X-ray crystallography, nuclear magnetic resonance spectroscopy or homology modelling), for its visualisation one needs a file containing the atomic coordinates of the structure and any of the several freely available **molecular graphics programs** to handle these coordinates. A few of the most popular such programs will be described at the end of this chapter. Practical problems and exercises in bioinformatics can be found in Chapter 12.

## 11.2. Primary sequence and three-dimensional structure databases

The most common use of bioinformatics for a biologist (and for a student of biology) is a search within primary molecular biological databases. These mostly include sequence and three-dimensional structure databases, and also experimental datasets provided by „omics” HTP methods (protein-protein interactions, large-scale mass spectrometry analysis and identification of lipids, sugars or small-molecule metabolites). The best known nucleotide sequence database is called **GenBank**, which is part of the **Entrez** bioinformatics web portal.

The most familiar protein sequence database is called **UniProt**, which is part of the **ExPASy** portal. The vast majority of amino acid sequences of polypeptides has been determined as nucleotide sequence and subsequently translated *in silico* (by bioinformatics tools) using the genetic code table. (Note that protein sequences determined on amino acid level are more relevant, since quite a few different functional proteins may originate from a single gene, due to e.g. alternative splicing and/or post-translational modifications. Actually there are many more proteins than genes!) Experimentally-determined three-dimensional structures of macromolecules (proteins, nucleic acids as well as protein-protein and protein-nucleic acid complexes) are stored in the **Protein Data Bank** (PDB). Secondary databases contain data from analysis of sequences and structures, and will be mentioned briefly later in the text.

The relationship between the informational macromolecules and the primary bioinformatics databases is summarised in Figure 11.1.

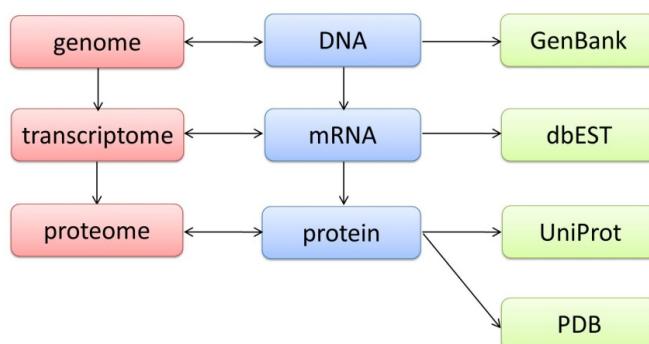


Figure 11.1. Relationship between informational macromolecules and primary bioinformatics databases

### 11.2.1. GenBank

**GenBank** ([ncbi.nlm.nih.gov/genbank](http://ncbi.nlm.nih.gov/genbank)) is a DNA (nucleotide) sequence database maintained by the NCBI (National Center for Biotechnology Information), a US-government sponsored resource for bioinformatics information, which is part NIH (National Institutes of Health).

GenBank currently (late 2012) contains ~150 Gbp (150 billion bp) of information in 160 million sequence files. Only original, experimentally-derived sequences can be submitted to GenBank. It is a redundant database, meaning that a particular sequence can be determined by independent research projects (cloning of a single gene or by genome sequencing projects). GenBank continues to grow at an exponential rate, doubling every 18 months. Presently, the major sources of submitted sequences are **genome projects** (complete sequencing of the full genetic material of an organism). Up to now, more than a thousand genomes have been sequenced, including our own genome. The **Human Genome Project** ([www.ornl.gov/Human\\_Genome](http://www.ornl.gov/Human_Genome)), i.e. the sequencing of the 3.2-Gbp human haploid genome (the 23 chromosomes) was finished in 2003. More precisely, only the gene-rich euchromatin region of the chromosomes (~90%) were sequenced because the highly repetitive so-called constitutive heterochromatin (around the centromere and the telomeres of the chromosomes) cannot be cloned. The human genome sequence, and in fact most of the genome sequences, are freely available in GenBank and in other databases (e.g. Ensemble, GenCard). GenBank is an **annotated database**, i.e. the sequences are supplemented with explanations or commentaries on its information content (including the coding region, the source of the sequence, and related publications). Nucleic acid sequences and any analysis derived from those sequences can be published only after they have been

deposited in a freely accessible database. The main page of NCBI is shown in Figure 11.2, while a sequence entry is shown in Figure 11.3. An online example of a sequence record (that of the human hemoglobin beta chain) is accessible here.

Newly determined nucleotide sequences can be identified by and compared using the GenBank database (using the **BLAST** program described in Chapter 11.3.2.1), and the results of this analysis are GenBank files identified by an accession code (e.g. D32013 in Figure 11.3).

Figure 11.2. The NCBI homepage (<http://www.ncbi.nlm.nih.gov/>). A few databases that are mentioned in the text are marked.

**Thermus aquaticus gene for DNA polymerase, complete cds**

GenBank: D32013.1

LOCUS TTHDNP 3026 bp DNA linear BCT 26-JAN-2008

DEFINITION Thermus aquaticus gene for DNA polymerase, complete cds.

ACCESSION D32013

VERSION D32013.1 GI:507890

KEYWORDS .

SOURCE Thermus aquaticus

ORGANISM Thermus aquaticus

Bacteria; Deinococcus-Thermus; Deinococci; Thermoales; Thermaceae; Thermus.

REFERENCE 1

AUTHORS Ishino,Y., Ueno,T., Miyagi,M., Uemori,T., Imamura,M., Tsunasawa,S. and Kato,I.

TITLE Overproduction of Thermus aquaticus DNA polymerase and its structural analysis by ion-spray mass spectrometry

JOURNAL J. Biochem. 116 (5), 1019-1024 (1994)

PUBMED 7896728

CDS 1..2499  
 /EC\_number="2.7.7.7"  
 /db\_xref="GI:507891"  
 /translation="MRGMPLFEPKGRVLLVDGHHLAYRTFHAKGLTTSRGEVQAV  
 YGFAKSLLKALKEDGDAIVVVFDAKAPSFRHEAYGGYKAGRPTPEDFPRQLALIKEL  
 VDLIGLARLEVPGYEADDVLASLAKKAKEGYEVRLTADKDLYQLLSDRIHALHPEG  
 YLITPAWILWEKYGLRDQWADYRALTGDESDNLPGVKIGEKTARKLLEEWGSLEALL  
 KNLDRLKPAIREKILAHMDDKLWSWLAKVRTDLPLEVDFAKRREPDRERLRAFLERL  
 EFGSLLHEFGLLSPKALEEAAPWPPEGAFFGVLSRKEMWADLLALAAARGGRVHR  
 APEPYKALRDLKEARGLLAKDSLVALEREGLGLPPGDDMILLAYLDDPSNITPEGVAR  
 RYGGGEWTEAAGERAALSERLFANLWGRLEGERLILWLYREVERPLSAVLAHMEATGVR  
 LDVAYLRLSLEVAEEIALEAEVFRLAGHPFNLSRDQLERVLFDELGLPAIGKTEK  
 TGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLTHPRTGRLHTRFNQ  
 TATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSG  
 DENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRRAAKTINFVLYGMSAHRQLSQEL  
 AIPYEEQAIFIERYFGSFVKVRAWIEXTLEGGRRGYVETLEGRRRYVPDLEARVKS  
 REAAERMAFNMPVQGTAIDLMLAMVLFPRLEEMGARMLLQVHDELVLEAPKERAEE  
 VARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE"

ORIGIN

```

1 atgagggggg tgctggccct ctttggcccc aaggggccgg tcctcctggt ggacggccac
61 cacctggccct accgcacccctt ccacggccctt aaggggccctt ccaccagccg gggggagccg
121 gtgcaggccgg tctacggctt cgccaagggc ctccctaagg ccctcaagg ggacggggac
181 gccgtatcg tggctttttt ccacggccggcc cccctccccc ccacggccggc ctacgggggg
241 tacaaggccgg gccggggcccc cacggccggag gactttcccc ggcaactcgc cctcatcaag
301 gagctgggtgg accttcgtgg gctggccggc ctggaggcc cgggctacga ggcggacgac

```

Figure 11.3. An example GenBank file (DNA polymerase from *Thermus aquaticus*; accession code D32013)

GenBank is part of the **Entrez** web portal ([www.ncbi.nlm.nih.gov/sites/gquery](http://www.ncbi.nlm.nih.gov/sites/gquery)), which is a powerful web tool to search for a large number of bioinformatics databases maintained by NCBI. **PubMed** ([ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)) is a bibliography database of life sciences and biomedical topics (covering practically all scientific journals in biochemistry and molecular biology). It contains more than 20 million bibliographical records of biomedical publications including free abstracts. More and more open access articles are freely available on the original journal websites in html format or downloadable as pdf file (directly accessed via PubMed). The **Bookshelf** online library contains many university textbooks in a fully searchable format (e.g. Stryer: Biochemistry, Lodish et al.: Molecular Cell Biology, Alberts et al.: Molecular Biology of the Cell). Part of the search page of the Entrez portal is shown in Figure 11.4.

Figure 11.4. Search page of the Entrez portal (databases mentioned in the text are highlighted) ([www.ncbi.nlm.nih.gov/sites/gquery](http://www.ncbi.nlm.nih.gov/sites/gquery))

Among the databases accessible via the Entrez search engine, only the most important ones are mentioned here. **Genome** provides views for a variety of *genomes*, complete chromosomes, *sequence* maps from many organisms whose genome has been fully sequenced; **dbEST** (Expressed Sequence Tag) contains **cDNA (complementary DNA) sequences** that were reverse-transcribed from mRNA sequences (transcripts); **OMIM** (Online Mendelian Inheritance in Man) contains detailed, full-text, referenced overviews of all human Mendelian disorders (> 12,000 genes). The **Ensembl** (<http://www.ensembl.org>) database also contains genome sequences. It is maintained as a joint project by the European Bioinformatics Institute (EBI, <http://www.ebi.ac.uk/>), the European Molecular Biology Laboratories (EMBL) and the British non-profit Wellcome Trust Sanger Institute (named after the double Nobel prize laureate British scientist Frederick Sanger who developed protein as well as DNA sequencing methods). EMBL also maintains a nucleotide database (ENA: European Nucleotide Archive), which contains the same information as GenBank.

## 11.2.2. UniProt

**UniProt** ([uniprot.org](http://uniprot.org)) is an annotated, non-redundant amino acid sequence database that actually consists of two sub-databases. The **Swissprot** division contains only experimentally validated and manually curated (annotated) protein sequences together with references to scientific publications (currently it contains more than 200 million amino acid residues in more than 500,000 annotated sequence files), while the **TrEMBL** division contains automatically translated sequences (currently more than 8 billion amino acid residues in approximately 24 million sequence files) from the EMBL nucleic acid database. Annotations of UniProt files include alternative versions of the particular sequence (alternatively spliced isoforms), other sequence variations (polymorphisms, mutations, sequence conflicts), information on the protein family to which the sequence belongs, structural and functional elements (motifs) of the polypeptide sequence, posttranslational modifications, cross-references to other databases (nucleotide sequence, structural and secondary databases) and, finally, literature references. An important part of the annotation is the so-called Gene Ontology (GO), a standardised vocabulary of the gene product across species and databases. It covers three attributes of the protein: the **cellular component** is the biological localisation of the protein (the parts of a cell or extracellular environment); the **molecular function** describes the elementary activities at the molecular level, e.g. binding or catalysis; and finally the **biological process**, functions in integrated living units: cells, tissues, organs, and organisms. An example UniProt file is shown in Figure 11.5 and Figure 11.6 (human skeletal muscle  $\alpha$ -actin with accession code P68133). Protein sequences are referred to using their accession

number (six alphanumeric characters) in research publications. The reader is encouraged to read the following short tutorial about the use of the UniProt database.

Figure 11.5. Example of a UniProt record (name, general annotation) (<http://www.uniprot.org/uniprot/P68133>)

Figure 11.6. Example of a UniProt record (secondary structure and amino acid sequence) (<http://www.uniprot.org/uniprot/P68133>)

The UniProt database is part of the **ExPASy** (Expert Protein Analysis System; [expasy.org/](http://expasy.org/)) bioinformatics resource portal, which provides access to scientific databases and software tools to different areas of life sciences including proteomics, genomics, transcriptomics and systems biology. Moreover, it is an entry point to many other secondary databases. For instance, proteomics tools include online programs of DNA-to-protein translation, calculation of the molecular mass and isoelectric point of proteins, prediction of structural and functional motifs, posttranslational modifications and three-dimensional structure. A screenshot of the portal is shown in Figure 11.7 (highlighted are databases and tools described in the text).

Figure 11.7. Screenshot of the ExPASy bioinformatics resource portal ([www.expasy.org](http://www.expasy.org))

### 11.2.3. Protein Data Bank (PDB)

**PDB** ([www.rcsb.org/pdb](http://www.rcsb.org/pdb)) is a database of experimentally-determined three-dimensional structures of proteins, nucleic acids and their complexes. Currently it stores nearly 80,000 structures determined by X-ray diffraction and approximately 10,000 structures determined by nuclear magnetic resonance (NMR) spectroscopy. (These two methods can be used to determine atomic-resolution structures of biological macromolecules.) The annotated PDB files contain additional useful information beyond the Cartesian atomic coordinates of the three-dimensional structures. The main page of the PDB website and details of an entry are shown in Figure 11.8 and Figure 11.9. PDB entries have a unique identification code consisting of a number and three letters (e.g. 1GFL is the PDB code of a Green Fluorescent Protein structure shown in Figure 11.9). Three-dimensional structures can be visualised online by using the **Jmol** applet (integrated into web browsers; see in Chapter 11.4.3). Alternatively, the PDB file can be downloaded and utilised by any of the freely available (open-source) molecular graphics programs (see Chapter 11.4).

Molecule of the Month regularly describes the structure and function of an interesting or important molecule. It is part of the PDB-101 interface, an educational resource for exploring a structural view of biology. It is highly recommended to download and study the poster “Molecular Machinery: Tour of the Protein Data Bank” that illustrates 80 PDB entries (enzymes, membrane proteins, motor proteins, DNA-binding proteins, protein complexes such as ribosomes) alongside water and ATP at a scale of one to three million.

Figure 11.8. Web page of PDB (<http://www.rcsb.org>), a repository for three-dimensional structural data of large biological molecules such as proteins and nucleic acids

The screenshot shows the PDB homepage with a search bar and navigation links. The main content area displays the details for the protein 1GFL. It includes the title 'STRUCTURE OF GREEN FLUORESCENT PROTEIN', the author 'Yang, F., Moss, L.G., Phillips Jr., G.N.', and the journal 'J. Biomol. Struct. Funct.' The page also shows the molecular weight (53783.00), length (238), and various deposition and validation information. On the right, there is a 3D ribbon model of the protein structure.

Figure 11.9. Details of a PDB entry (Green Fluorescent Protein) (<http://www.rcsb.org>)

## 11.3. Introduction to bioinformatics analysis of sequences

### 11.3.1. Bioinformatics tasks during molecular cloning

When we design a recombinant DNA construct (see Chapter 10), it is important to know the potential restriction endonuclease recognition sites both in the vector and the insert. Finding these sites is a relatively simple bioinformatics task using online programs such as NEBCutter or RestrictionMapper. It is also useful to draw a map of the recombinant DNA we plan to construct with the help of downloadable or online programs such as pDRAW, BioEdit or SnapGene.

Another typical bioinformatics task is the design of oligonucleotide primers for sequencing, polymerase chain reaction (PCR) or site-specific *in vitro* mutagenesis. This can also be achieved by online programs such as Primer3, Oligo or OligoCalc.

### 11.3.2. Sequence similarity search and sequence alignment

We can do a similarity search to learn if our sequenced DNA can be found in a public nucleotide database (i.e. it has already been cloned by others) and/or whether it is evolutionarily related (i.e. homologous) to other sequences. In a simple similarity search, one can compare a sequence with sequences found in an entire nucleotide database (see later the BLAST program), while for a homology search the method of choice is multiple sequence alignment by the ClustalW program. By comparing either nucleotide or amino acid sequences we can find homologs. If these are from different species (that had a common ancestor) but have identical or similar functions they are called **orthologs**; while those homologs that are found in the same organism and originate from a gene duplication event followed by divergent evolution within the species are called **paralogs**. We will not cover the construction of evolutionary trees in this e-book—one can learn about these in bioinformatics or evolutionary biology courses.

#### 11.3.2.1. The BLAST program

If we sequence a DNA clone, the first bioinformatics analysis is a similarity search against a nucleotide database. The most widely used similarity search program accessible on the internet is **BLAST** (Basic Local Alignment Search Tool), which will be described here and will be used by the students during the laboratory practice. The

BLAST program is available online at several servers including the one at NCBI: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

BLAST uses a heuristic algorithm that makes it possible to search a huge database in a very short period of time by using a **query sequence**. The high speed of the algorithm stems from the fact that the query sequence is divided into short „words” that are used, instead of the full-length sequence, during the alignment process. These words are searched in the database first (called „seeding”, i.e. finding the best local alignments). The most relevant hits are then scored with the help of a scoring matrix, extended to neighbouring words, and finally assembled and compiled into a final list of similarity hits. It is important that the query sequences must be in the so-called **FASTA format** (FASTA was a previously popular but much slower similarity search program). The FASTA format is shown in Figure 11.10.

```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
(the first line can be omitted)
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPIYGTNLV
EWIWGGSFVDKATLNRRFFAHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFH PYYTIKDFLG
LLILLLLLLALLSPDMLGDPDNHMPADPLNTPHLIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLMDLLTLTWIGSQPV EYPYTIIGQMASILYFSIILAFLPIAGX
IENY
```

Figure 11.10. The FASTA sequence format

If we want to search using a nucleotide query sequence within a nucleotide database, we can use the **BLASTN** version of the program. If we have an amino acid sequence, we can search a protein database by the **BLASTP** version of the program. The **BLASTX** version of the program translates a nucleotide sequence in all six reading frames (three on each strand) and allows searching a protein database. Finally, with the **TBLAST** subprogram, we can search against a translated nucleotide database using either a protein (TBLASTN) or a nucleotide (TBLASTX) query sequence. These similarity search options are summarised in Figure 11.11.

| Program        | Query sequence                       | Target database                      |
|----------------|--------------------------------------|--------------------------------------|
| <b>BLASTN</b>  | <b>Nucleotide</b>                    | <b>Nucleotide</b>                    |
| <b>BLASTP</b>  | <b>Protein</b>                       | <b>Protein</b>                       |
| <b>BLASTX</b>  | <b>Nucleotide in 6 reading frame</b> | <b>Protein</b>                       |
| <b>TBLASTN</b> | <b>Protein</b>                       | <b>Nucleotide in 6 reading frame</b> |
| <b>TBLASTX</b> | <b>Nucleotide in 6 reading frame</b> | <b>Nucleotide in 6 reading frame</b> |

Figure 11.11. Search possibilities in the BLAST program

The result of a BLAST analysis is a list of sequences from the searched database that show significant similarity to the query sequence. Besides the sequence identifiers of the similar sequence hits in the database, the final list of alignments contains a score number and a statistical significance number, the **E-value**. The *E*-value is a parameter that describes the number of hits one can expect to see by chance when searching a database of a particular size. It decreases exponentially as the score (S) of the match increases. Essentially, the *E*-value describes the random background noise. The lower the *E*-value, or the closer it is to zero, the more "significant" the match ( $E > 0.01$  is usually considered to reflect a homologous, i.e. evolutionarily-related sequence). The score value is calculated based on the alignment, taking into account the gaps and the similarity of the amino acids at the aligned positions. The most often used similarity matrix (an amino acid substitution matrix) is the **BLOSUM** (BLOcks SUbstitution Matrix) matrix. The numbers within a BLOSUM are "log-odds" scores that measure, in an alignment, the logarithm of the ratio of the likelihood of two amino acids appearing with a biological sense and the likelihood of the same amino acids appearing by chance.

The similarity hits can be found and downloaded from the database using their accession number (identifier). BLAST hits are usually hyperlinked directly to the corresponding entries in the GenBank database where we can learn much more about the related sequences, the gene, cDNA and/or the coded protein. As we have already mentioned, the most comprehensive information on a given protein can be found in the UniProt database. In Figure 11.12, a detail of a BLAST run is shown in which the BLASTP program was used to search the UniProt database using a human skeletal actin query sequence.

| Alignments | Entry  | Entry name   | Status | Protein names                | Organism  | Length | Identity | Score* | E-value | Gene names   |
|------------|--------|--------------|--------|------------------------------|---|--------|----------|--------|---------|--------------|
| 1          | M7AKZ6 | M7AKZ6_CHEMY | ★      | Actin, alpha skeletal muscle | Chelonia mydas (Green sea-turtle) (Chelonia agassizi)                                       | 377    | 100.0%   | 1,975  | 0.0     | UY3_17810    |
| 2          | L7MZF2 | L7MZF2_ANOCA | ★      | Uncharacterized protein      | Anolis carolinensis (Green anole) (American chameleon)                                      | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |
| 3          | K7F4I9 | K7F4I9_PELSI | ★      | Uncharacterized protein      | Pelodiscus sinensis (Chinese softshell turtle) (Trionyx sinensis)                           | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |
| 4          | J3SBR4 | J3SBR4_CROAD | ★      | Actin, alpha skeletal muscle | Crotalus adamanteus (Eastern diamondback rattlesnake)                                       | 377    | 100.0%   | 1,975  | 0.0     |              |
| 5          | I3N266 | I3N266_SPETR | ★      | Uncharacterized protein      | Spermophilus tridecemlineatus (Thirteen-lined ground squirrel) (Ictidomys tridecemlineatus) | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |
| 6          | H0VGS3 | H0VGS3_CAVPO | ★      | Uncharacterized protein      | Cavia porcellus (Guinea pig)  | 377    | 100.0%   | 1,975  | 0.0     | LOC100716933 |
| 7          | H0XEG0 | H0XEG0_OTOGA | ★      | Uncharacterized protein      | Otolemur garnettii (Small-eared galago) (Garnett's greater bushbaby)                        | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |
| 8          | G3WM41 | G3WM41_SARHA | ★      | Uncharacterized protein      | Sarcophilus harrisii (Tasmanian devil) (Sarcophilus laniarius)                              | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |
| 9          | G3SX7  | G3SX7_LOXAF  | ★      | Uncharacterized protein      | Loxodonta africana (African elephant)   | 377    | 100.0%   | 1,975  | 0.0     | LOC100661558 |
| 10         | F7HM15 | F7HM15_MACMU | ★      | Uncharacterized protein      | Macaca mulatta (Rhesus macaque)   | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |
| 11         | F7CZ92 | F7CZ92_HORSE | ★      | Uncharacterized protein      | Equus caballus (Horse)  | 377    | 100.0%   | 1,975  | 0.0     | ACTA1        |

Figure 11.12. Result of a sequence similarity search by the BLAST program (human skeletal muscle actin was used as a query sequence against the UniProt database)

It is important to note that, since 3-D structure is more conserved than primary structure, it is easier to recognise two related proteins by comparing their three-dimensional structure than their amino acid sequence. Obviously, it is more convenient to compare primary sequences, since they are available for much more proteins than the atomic-resolution structures. Similarity searches and protein structure comparisons are dealt with in more detail in bioinformatics (or structural bioinformatics) courses.

### 11.3.2.2. Multiple sequence alignment

More detailed sequence similarity analysis can be performed by creating multiple sequence alignments. By this method, three or more biological sequences (those of proteins or nucleic acids) of similar length are aligned to minimise gaps (insertions or deletions in one sequence compared to the others) and maximise the occurrence of identical or similar residues at the aligned positions.

From the output alignments, homology—i.e. the evolutionary relationships—between the sequences can be inferred. Moreover, the presence of conserved regions indicates conserved structural and/or functional elements (motifs) within the sequence. The most often used program for multiple sequence alignment is **ClustalW** that can be reached at the ExPASy portal ([embnet.vital-it.ch/software/ClustalW.html](http://embnet.vital-it.ch/software/ClustalW.html)) or via the web page of the European Bioinformatics Institute ([ebi.ac.uk/Tools/msa/clustalw2/](http://ebi.ac.uk/Tools/msa/clustalw2/)).

In Figure 11.13, human alpha and beta hemoglobin as well as the myoglobin sequences are aligned by the ClustalW program. The asterisks at the bottom line of the alignment indicate identical (fully conserved, i.e. invariant) residues in a given sequence position, while single and double dots refer to highly and moderately conserved (chemically similar) residues, respectively. Within the aligned sequences, the dashes indicate the „gaps” that were inserted in order to optimise the alignment.

|                    |   |     |
|--------------------|---|-----|
| P69905 (HBB_HUMAN) | MV-LSPADKTNVKAAGWKVGAGHEYGAELERMFLSFPPTTKTYFPHF-DLSH-----GS               | 53  |
| P68871 (HBB_HUMAN) | MVHLTPEEKSATVALWGKV--NVDEVGGEALGRLLVVYPWTQRFESFGDLSTPDAVMGN               | 58  |
| P02144 (MYG_HUMAN) | -MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFHKLKSDEDMKAS              | 59  |
|                    | : * : * **** * * * : * * * : * * * :                                      |     |
| P69905 (HBB_HUMAN) | AQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAKLDRVDPVNFKLLSHCLLVTLAAH              | 113 |
| P68871 (HBB_HUMAN) | PKVKAHGKKVLGAFSDGLAHLDNLKGTFTALSELHCDDKLHVDPENFRLLGNVLVCVLAHH             | 118 |
| P02144 (MYG_HUMAN) | EDLKKHGATVLTA <del>LGGI</del> KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSK  | 119 |
|                    | : * *** * * : : : * : * * : : : : : : * : :                               |     |
| P69905 (HBB_HUMAN) | LPAEFTP <del>AV</del> HASLDKFLASVSTVLTSKYR-----                           | 142 |
| P68871 (HBB_HUMAN) | FGKEFTP <del>PPV</del> QAA <del>Y</del> KV <del>V</del> AGVANALAHKYH----- | 147 |
| P02144 (MYG_HUMAN) | HPGDFGADAQGAMNKALELFRKDMSA <del>N</del> YKELGFGQ                          | 154 |
|                    | : * : : * : : * :   |     |

Figure 11.13. Three polypeptide chains from the globin family aligned by the ClustalW program and shown together with their UniProt accession code (and short name in the database)

### 11.3.3. Bioinformatics analysis of protein sequences

The wide range of *in silico* analysis possibilities of protein sequences is summarised in Figure 11.14. Note that many of these analyses can be performed also with nucleic acid sequences. Sequences can be compared to each other and to full databases. The physical and structural/functional properties of polypeptide chains can be predicted via this analysis. Sequence comparisons (alignments) were described in the previous section (BLAST and ClustalW programs). During the so-called profile analysis, the analysed sequences are compared to secondary databases that contain information about protein structural families, structural and functional domains, modules, phosphorylation, glycosylation and other posttranslational modification consensus sequences. Many online programs are available on the internet that can search secondary databases. For instance, the InterProScan profile analysis program can be used to search the InterPro secondary database (in fact it is a „superdatabase” of several individual derived databases) maintained by the EBI. Another example is the PhosSitePlus database that can be searched by any query sequence to predict phosphorylation or other posttranslational modification sites.

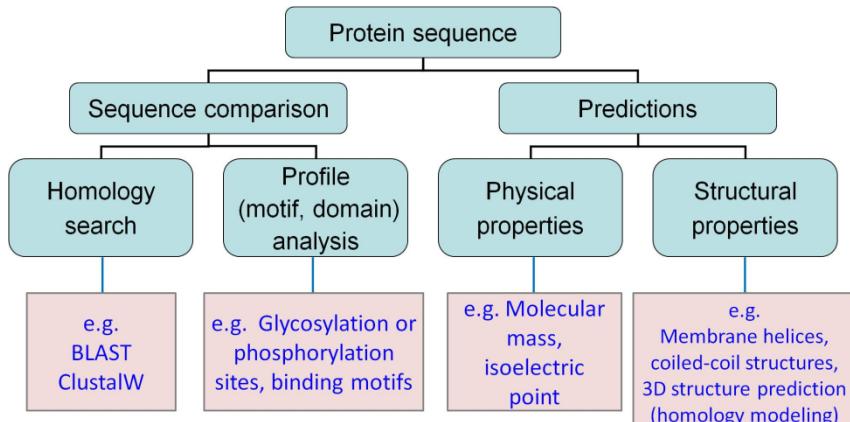


Figure 11.14. The wide range of *in silico* analysis possibilities of protein sequences. (Most of these options are also available for nucleic acid sequences.)

## **11.4. Visualisation of protein structures by molecular graphics programs**

### 11.4.1. RasMol

**RasMol** is a free, open-source and stand-alone molecular graphics program. Three-dimensional structures can be constructed and displayed if the atomic-resolution coordinates of a biomolecule or its complex are available. The program consists of two windows: one for the command line, another for providing the graphics. The input file

can be in PDB format and can be downloaded from the PDB structure database. The initial image is shown as a „wire” model. However, from the Display menu one can choose other visualisation styles such as „spacefill”, „stick”, „ball and stick” as well as the visually most attractive „ribbon” and „cartoon” models. In the last two styles, alpha helices are rendered as helical ribbons and beta structures as flat arrows pointing in the direction of the polypeptide chain. The atoms of the model can be coloured by the standard CPK (named after Corey, Pauling and Koltun) scheme in which white is hydrogen, black is carbon, blue is nitrogen and red is oxygen. The protein can be coloured based on polypeptide chains („chain” command), the chemical property of the amino acids („shapely” command), or the inner mobility of the atoms (as determined by the crystallographic so-called B-factor; „temperature” command). The structure can be cut in the z-dimension by the „slab” command, can be rendered in a stereo view, and the otherwise invisible hydrogen atoms and hydrogen bonds can also be visualised.

The left and right mouse buttons can be used to rotate the protein along the „x” and „y” axes. Shift + left-click will shrink and enlarge the rendered structure, whereas shift + right click will translate it along the „z” axis. By clicking any part of the structure, the residue number of the given chain and the particular atom will be shown in the command window. One can select a chain, a particular residue or segment of the chain by using the „select” command. „Select 25A” in the command window means that the 25th residue of chain A will be selected. The name of the residue can also be used. A segment of a chain can also be selected: for example, „select 1-33” means that the first 33 residues of the chain will be selected. If the structure contains a ligand (coenzyme, substrate, metal ion etc.) besides the polypeptide chain, it can be selected by the „hetero” command or by its name (e.g. „ca” refers to a  $\text{Ca}^{2+}$  ion). By clicking the ligand, we can learn its name. **Commands in the menu or in the command line will always be executed on the last selection.** Parts of the molecule (chains, segments, residues, ligands) can be changed by the „color” command followed by the name of a particular colour. The background of the image can be set by the „background color” syntax. If we want to remove part of the structure, it can be done by using the „restrict” command (e.g. „restrict 1-56” will remove the rendering of the chain from residue 57 to the C-terminal end). One can save the modified structure (e.g. for later manipulation) by the „write script” command and a file name. To continue working on a particular structure view, the „script” command together with a file name will reload the previously saved script file. The finished structure can be saved in common graphics file formats (gif, jpeg, etc.). The „help” menu can explain many additional commands that can be used to manipulate the structure. RasMol is a powerful program, and it takes time to learn all of its features. Further help for using RasMol can be found at <http://www.openrasmol.org/doc/>.

It should be noted that RasMol is not suitable for homology modelling, to study the structural effect of mutations, to energy-minimise structures or for molecular dynamics simulations. If the reader is interested in learning such structural modelling software, he/she should search the internet for such programs (many free and commercial programs are available for that purpose) and/or attend a structural bioinformatics course.

In Figure 11.15 and Figure 11.16 we illustrate two protein structures rendered by RasMol. In the first one, an EF-hand  $\text{Ca}^{2+}$ -binding domain of the ubiquitous eukaryotic  $\text{Ca}^{2+}$ -modulator protein calmodulin is shown ( $\alpha$ -helices are magenta, a short  $\beta$ -sheet at the beginning of the  $\text{Ca}^{2+}$ -binding loop is shown in yellow,  $\text{Ca}^{2+}$  is blue; and the residues participating in the coordination of the divalent ion are shown in a ball-and-stick style). In the second molecular graphics image, the three-dimensional structure of myoglobin is shown in a cartoon model ( $\alpha$ -helices are red, the loops are yellow, the ball-and-stick hem atoms are blue, and the  $\text{Fe}^{2+}$  ion is shown in magenta in spacefill style).

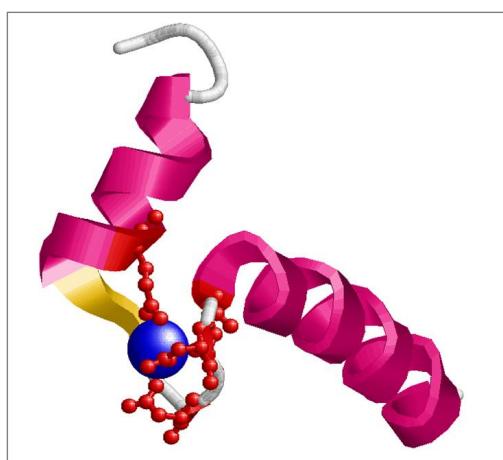


Figure 11.15. An EF-hand of calmodulin rendered by the RasMol program

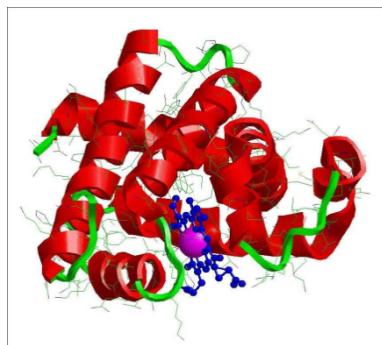


Figure 11.16. Myoglobin structure visualised by RasMol

## 11.4.2. PyMOL

**PyMOL** is another open-source molecular graphics program. It is probably the one most often used by molecular biologists to visualise three-dimensional structures of macromolecules. The program is free for academic users and students (after registration), except for its molecular modelling modules. Its advantage over RasMol is that it has superb publication-quality graphics outputs. In Figure 11.17, a „publication-ready” structural model is shown: that of the vertebrate-specific  $\text{Ca}^{2+}$ -binding protein S100A4 in complex with a non-muscle myosin 2a peptide. In Figure 11.18, the three-dimensional structure of hemoglobin is modelled by PyMol in three styles; the polypeptide chains in stick and atoms of the hem prosthetic group in spacefill representation (left), the four chains in four coloured ribbon models (middle) and a van der Waals-surface representation with the hem in wireframe (right). The detailed description of the program is given in M.Sc. programme courses and is also available online (<http://pymol.sourceforge.net/newman/user/toc.html>).

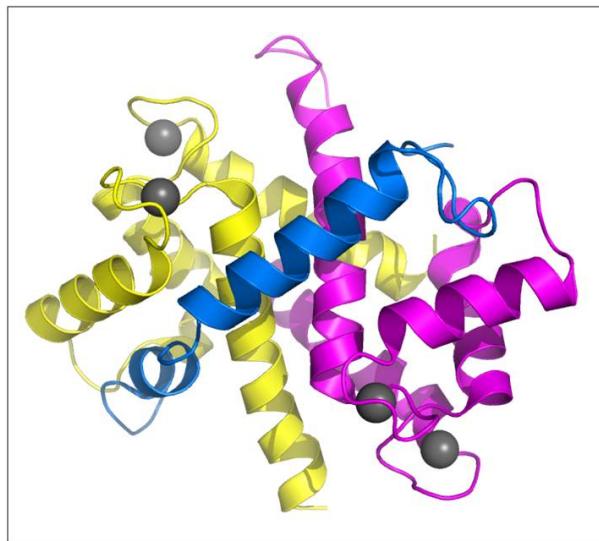


Figure 11.17. The Complex of the  $\text{Ca}^{2+}$ -binding protein S100A4 with a non-muscle myosin 2A peptide (figure made using the PyMOL program)

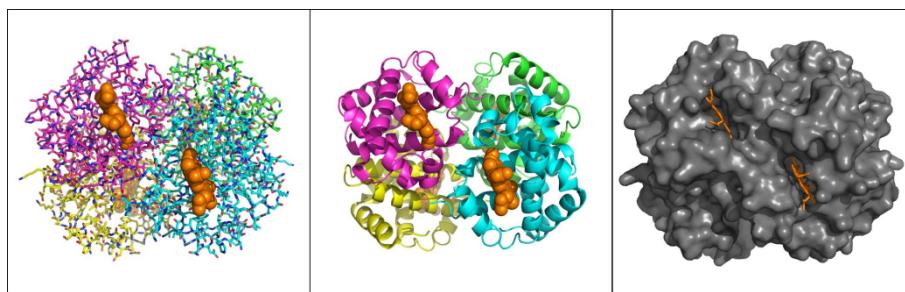


Figure 11.18. Three-dimensional structure of deoxyhemoglobin, visualised by PyMOL as „sticks” (left), as a „cartoon” model (middle) and displaying the surface of the protein (right)

### 11.4.3. Jmol

**Jmol** is an open-source molecular graphics application written in Java programming language that can be used either as a browser applet or a downloadable stand-alone program. Its input file can be in PDB file format or many other structural biology or organic chemistry file formats that contain atomic coordinates of a structure. The commands of Jmol are similar to (actually derived from) those of RasMol, and they follow the same logic. Jmol can be used to perform simple molecular modelling tasks such as adding or deleting atoms or residues. The command window can be opened by clicking the right mouse button. Jmol has versions in many languages (including Hungarian). Further information can be read on the Jmol Wiki page.

Jmol is often used for educational purposes as part of e-learning pages or other applications. Any file from the PDB database can be interactively viewed online by using the Jmol applet. In Figure 11.19, a static view of the hemoglobin mutant causing sickle-cell anemia is shown from the interactive Jmol Hemoglobin Tutorial. Other examples of structural biology Jmol tutorials include DNA structure, Antibody, Lipid Bilayers and Membrane Channel, Collagen, Water, and Lac Repressor. In Figure 11.20, we illustrate B-DNA structure (left) and the three-chain collagen helix (right) captured by Jmol.

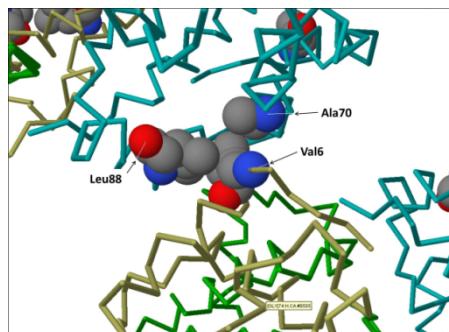


Figure 11.19. Detail of a Jmol tutorial explaining the structural background of the hemoglobin mutation causing sickle-cell anemia. The close-up view of a subunit interface is shown. (The mutant valine 6 of the beta chain binds to alanine 70 and leucine 88 of the beta chain of a second hemoglobin tetramer.)

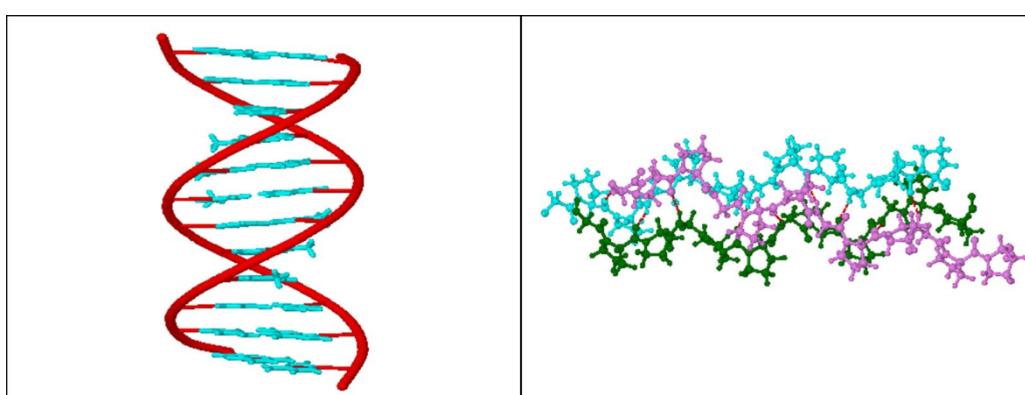


Figure 11.20. Structure of B-DNA (left) and that of the collagen triple helix (right), visualised by the Jmol program

---

# Chapter 12. Calculations and problem solving exercises

by László Radnai

The structure of this chapter has been adjusted to that of the whole e-book. After each chapter title, several examples of exercises are shown. Those together with their detailed solutions are marked by the tag “(SAMPLE EXERCISE)”. Solutions of exercises that are to be calculated (marked by “(\*)”) can be found in Chapter 12.3. Some examples do not have any kind of marks. These are meant to aid revisiting the curriculum of the given chapter and deepening the acquired knowledge. Consequently, answers to these problems are not provided; these are to be developed by the reader.

## 12.1. Useful preliminary information

i. Relative molecular mass ( $M_r$ ) and molar mass ( $M$ ):

Relative molecular mass ( $M_r$ ) is a dimensionless numerical value that shows how many times the 1/12th of the mass of the  $^{12}\text{C}$  isotope equals the mass of the given molecule. In biochemistry, it is expressed in units of dalton (Da) or kilodalton (kDa).

Molar mass is the mass of the substance divided by the molar amount of the substance, with the dimension of mass/mol, usually expressed in g/mol.

The numerical value of the relative molecular mass of a given molecule in daltons equals the molar mass of the same chemical compound expressed in g/mol.

ii. Dimensions of concentration—percentages:

Hybrid percentage (w/v %): expresses how many grams of solute are present in 100 millilitres of final solution volume.

Volume percentage (v/v %): expresses how many millilitres of a given liquid have been measured into 100 millilitres of final solution volume.

Weight percentage (w/w %): expresses how many grams of solute are present in 100 grams of solution.

iii. Dimensions of concentration—molarity: expresses how many moles of solute are present in 1 litre of solution (expressed in mol/dm<sup>3</sup>, or M).

iv. Dimensions of concentration—mg/mL: an „arbitrary” concentration unit that is frequently used during laboratory work. Note that this is identical to the concentration in g/L as 1 g/L = 1000 mg/1000 mL = 1 mg/mL. Furthermore, this is also identical to the unit kg/m<sup>3</sup> included in the SI system.

v. Frequently used orders of magnitude in biochemistry:

- milli: one thousandth (symbol: m), e.g. 1 mM =  $10^{-3}$  M.
- micro: one millionth (symbol:  $\mu$ ), e.g. 1  $\mu\text{L} = 1/10^6 \text{ L} = 10^{-6} \text{ L}$ .
- nano: one billionth (symbol: n), e.g. 1 nm =  $10^{-9} \text{ m}$ .

## 12.2. Problems and exercises

### 12.2.1 Units of measure, solutions

(1) (SAMPLE EXERCISE) How many moles and how many milligrams of sodium chloride are present in 15  $\mu\text{l}$  of a 0.5-M sodium chloride solution? ( $M_{\text{NaCl}} = 58 \text{ g/mol}$ )

|      |                       |                |
|------|-----------------------|----------------|
| When | 1 L solution contains | 0,5 mol NaCl,  |
| then | 1 mL contains         | 0,5 mmol NaCl, |
| and  | 1 µL contains         | 0,5 µmol NaCl. |

When 1 µL solution contains 0,5 µmol, then 15 µL contains 15 times more, which is 7,5 µmol.

This equals  $\underline{7,5 \cdot 10^{-6} \text{ mol}}$ .

The calculation can be solved by using the rule of three:

$$1 : 0,5 = 1,5 \cdot 10^{-5} : x \quad x = \underline{7,5 \cdot 10^{-6} \text{ mol}}$$

This way, we obtain the amount of NaCl in grams by multiplying x by the molar mass. Then we convert the result to milligrams:

$$7,5 \cdot 10^{-6} \text{ mol} \cdot 58 \text{ g/mol} = 4,35 \cdot 10^{-4} \text{ g} = \underline{0,435 \text{ mg}}$$

(2) (SAMPLE EXERCISE) A given protein solution has a concentration of 5 mg/mL, and the molecular mass of the protein is 25000 Da. (A) What is the molarity [M] of the solution? (B) What is the hybrid percentage concentration of the solution?

It is easy to see that  $5 \text{ mg/mL} = 5 \text{ g/L}$  because:  $1 \text{ g/L} = 1000 \text{ mg}/1000 \text{ mL} = 1 \text{ mg/mL}$ . The molar amount of 5 g protein in 1 L ( $1 \text{ dm}^3$ ) of solution is:

$$5/25000 = 2 \cdot 10^{-4} \text{ mol}. \text{ Therefore, the concentration is } \underline{2 \cdot 10^{-4} \text{ mol/dm}^3} (= 2 \cdot 10^{-4} \text{ M}).$$

By a reverse approach: the molar mass is 25000 g/mol, thus the concentration of the 1-M solution would be 25000 g/L, which is 25000 mg/mL. Then what is the molarity of a solution with 5 mg/mL concentration?

Calculating with a simple proportion:

$$\begin{array}{ll} 1 \text{ M} & 25000 \text{ mg/mL} \\ x \text{ M} & 5 \text{ mg/mL} \end{array}$$

$$x = (5 \text{ mg/mL} \cdot 1 \text{ M}) / 25,000 \text{ mg/mL} = \underline{2 \cdot 10^{-4} \text{ M}}$$

(B) When 1 g solute is present in 100 mL of solution, then the solution has 1 % hybrid concentration. In the specified case, the solution has a concentration of 5 mg/mL (= 5 g/L), that is, 5 g of solute is present in 1000 mL of solution. Thus, 100 mL of solution contains 1/10 of this; that is, 0,5 g. Therefore, the solution has a hybrid percentage (w/v) concentration of 0,5 %.

(3) (\*) The diameter of a eukaryotic cell, which can be considered as a sphere, is 50 µm. The diameter of an average-sized globular protein (with a molecular mass of ca. 40 kDa) is  $\approx 3,6 \text{ nm}$ . (A) Supposing that there are no other cellular components present and the available space is filled up by small, spherical protein molecules, how many molecules of protein can be fitted into this cell? (In this case, approx. 74 % of the space is filled with protein and approx. 26 % of „empty” space remains in between.) (B) The glucose concentration within this theoretical cell is 1 mM (mmol/dm<sup>3</sup>). Calculate the number of glucose molecules within the cell. (C) The concentration of hexokinase, a key enzyme in glucose metabolism, is 20 µM in this given cell. Calculate how many glucose molecules are available for one hexokinase molecule. (The volume of a sphere can be calculated as  $V = 4r^3\pi/3$ .)

(4) (\*) An *E. coli* cell can be considered as a cylinder with a diameter of 0,8 µm and a depth of 2 µm. (A) The average density of an *E. coli* cell is  $1,1 \cdot 10^3 \text{ g/L}$ . What is the mass of a single *E. coli* cell? (B) The cell wall of an *E. coli* cell is 10 nm thick. What percentage of the total cell volume is made up by the cell wall? (C) Intensive protein synthesis in *E. coli* cells is provided by the ~15000 ribosomes present in each cell. Ribosomes can be considered as spherical cellular compartments with a diameter of 18 nm. What percentage of the total cell volume is occupied by ribosomes? (D) Regarding the volume values given above, let us assume that the remaining cell volume is an aqueous solution with a pH of 7. How many  $\text{H}^+$  ( $\text{H}_3\text{O}^+$ ) ions are present in a single *E. coli* cell? (E) Assuming that there are 2 molecules of a given protein in an *E. coli* cell, what is the molar concentration (mol/dm<sup>3</sup>) of this protein?

(F) How many *E. coli* cells can be fitted on the point of a needle? (Let us assume that the diameter of the point of the needle is 0.3 mm, and it is flat and circle-shaped. Cells are put onto the surface vertically so their projection will be a circle. A plane can optimally be filled with identical circles in a hexagonal arrangement. In this case, the circles will cover 90.69 % of the plane.)

(5) (\*) Inside an *E. coli* cell, there is a single circular double-stranded DNA molecule with a relative molecular mass of  $3.1 \cdot 10^9$  Da. The average molecular mass of a single pair of nucleotides in double-stranded DNA is 660 Da, and each nucleotide pair adds 0.34 nm to the length of the DNA molecule. (A) Calculate the length of *E. coli* DNA. Collate the length of the DNA with the dimensions of the *E. coli* cell. How can the DNA molecule be fitted into the cell? (B) Let us assume that proteins in *E. coli* cells consist of 400 amino acids on average. How many protein molecules can be coded for by the *E. coli* DNA if, for simplicity, we neglect the regulatory regions and consider the whole DNA strand as a coding region?

(6) (\*) A mitochondrion can be considered as a cylinder with a length of 1.5  $\mu\text{m}$  and a diameter of 0.6  $\mu\text{m}$ . The membrane system makes up 35 % of the total volume of the mitochondrion. The remaining volume can be regarded as an aqueous solution. The concentration of oxaloacetate inside the mitochondrion is 0.03  $\mu\text{M}$ . How many oxaloacetate molecules are present in a single mitochondrion? (Avogadro's number =  $6.022 \cdot 10^{23}$ .)

(7) (\*) Tropomyosin is a dimeric muscle protein with a total molecular mass of 70 kDa. It has a coiled-coil structure: it consists of two  $\alpha$ -helical chains spiralled onto each other. Using the dimensions of the  $\alpha$ -helix (3.6 amino acids/turn; pitch = 5.4 Å), calculate the length of the tropomyosin molecule. (The average mass of an amino acid is 110 Da.)

(8) (\*) 1 mL of a 0.35 w/v % solution of tropomyosin contains  $10^{15}$  latex beads. A small volume of the solution is dried and examined under an electron microscope. 300 protein molecules and 10 latex beads can be counted in a given field. What is the molecular mass of tropomyosin?

(9) (\*) 1 mL of a 0.1 w/v % solution of myosin contains  $10^{14}$  latex beads. A small volume of the solution is dried and examined under an electron microscope. 120 protein molecules and 10 latex beads can be counted on a given area. What is the molecular mass of myosin?

(10) (\*) The growth speed of hair is 150-200 mm/year. The structure of hair is built up from  $\alpha$ -keratin, which is synthesised in the epidermal cells of hair bulbs.  $\alpha$ -keratin has an  $\alpha$ -helical structure in which 3.6 amino acids constitute a full turn and the vertical pitch is 5.4 Å. Assuming that the rate-limiting step of hair growth is the biosynthesis of the  $\alpha$ -keratin peptide chain, calculate the rate of peptide synthesis during hair growth.

(11) (\*) A copper-containing protein contains 320 µg copper per gram of total mass. What is the molecular mass of the protein if one protein molecule contains one  $\text{Cu}^{2+}$ ? (The atomic mass of copper is 64 Da.)

(12) (\*) According to amino acid analysis, ribonuclease contains 2.47 w/w % isoleucine and 1.65 w/w % leucine. The molecular mass of both amino acids is 131 Da. What is the (minimal) molecular mass of the ribonuclease molecule?

(13) (\*) Cytochrome C contains 0.423 w/w % iron. The atomic mass of iron is 55.84 Da. What is the molecular mass of cytochrome C? (One cytochrome C molecule contains a single iron bound to a heme group.)

(14) (\*) How many femtomoles ( $\text{fmol} = 10^{-15}$  mol) of alanine are present in a liver cell if the volume of a liver cell is  $10^{-8}$  mL and the concentration of alanine is 1 mM? Calculate the w/w percentage concentration of alanine in the liver if 1 g tissue contains  $10^9$  liver cells. (The molecular mass of alanine is 89 Da.)

(15) (\*) The human body contains 3 g iron, 75 % of which is located in hemoglobin. (A) Calculate the amount of hemoglobin in the human body by using the information given in brackets. (B) How many moles of oxygen can be bound by this amount of hemoglobin? (C) How many litres of 25°C air (at a pressure of 1 atm) contain this amount of oxygen? (Hemoglobin is a tetramer with a molecular mass of 68,000 Da. It contains one heme-bound iron per subunit and can bind 1 oxygen molecule per subunit. The atomic mass of iron is 55.8. Air contains 20 v/v %  $\text{O}_2$ . The volume of 1 mol of an ideal gas at 1 atm pressure and 0°C is 22.4 litres.)

(16) (\*) 4.9 mL water is added to 100 µL of a solution of unknown concentration. The determined concentration of the diluted solution is 20 µM. What is the concentration of the original unknown solution, and how many µmol of solute does it contain per millilitre?

(17) (\*) One pill of vitamin C contains 350 mg ascorbic acid. The molecular mass of ascorbic acid is 175 Da. We wish to prepare a 10-mM solution of ascorbic acid. What should be the volume of the solution if we dissolve one pill?

(18) (\*) Suppose you throw two lumps of sugar into a 200-mL mug, and then fill it up with tea. A 500-g box of sugar contains 250 lumps. The molecular mass of sucrose is 342 Da. (A) What will be the sugar concentration in the tea? (B) The well-known sweetener aspartam is the methyl ester derivative of the aspartyl-phenylalanine dipeptide. Draw the structure of the dipeptide. (C) One weight unit of aspartam is 180 times as sweet as one weight unit of sugar. What is the mM concentration of aspartam with a sweetness identical to that of the sugar concentration of the tea described above? (The molecular mass of aspartam is 294 Da.)

(19) (\*) The molecular mass of a protein is 25000 Da. (A) What is the molarity of its 1 mg/mL solution? (B) What is the hybrid percentage concentration of protein in this solution?

(20) (\*) The molecular mass of NADH is 709.43 Da. How many millilitres of a 50 µM solution of NADH can be prepared by dissolving 1 mg NADH?

(21) (\*) The molecular mass of a protein is 50000 Da. (A) What is the mg/mL concentration of a 40-µM solution of the protein? (B) What is the hybrid percentage concentration of this solution?

(22) (\*) How would you prepare 100 mL solution with the following final concentrations: 20 mM potassium phosphate buffer, 150 mM NaCl, 50 µM serum albumin? The available materials are 1 M potassium phosphate buffer, solid NaCl (molar mass = 58.44 g/mol), 13 mg/mL serum albumin solution (molecular mass = 65000 Da) and distilled water.

(23) (\*) You wish to determine the activity of the enzyme lactate dehydrogenase. For this measurement, a 4-mL reaction mixture is needed in which the concentrations of the components are the following: 5 mM lactate, 20 mM phosphate buffer, 0.5 mM NAD<sup>+</sup> and 10 pM lactate dehydrogenase. The following stock solutions are available: 0.1 M lactate, 0.1 M phosphate buffer, 0.1 M NAD<sup>+</sup>. The stock solution of lactate dehydrogenase (146 kDa) has a concentration of 0.544 µg/mL. How many millilitres or microlitres of the individual components need to be taken in order to prepare a solution with the specified volume and composition?

## 12.2.2. Ionisation equilibria

(24) (SAMPLE EXERCISE) What is the pH of the following solutions: (A) 0.35 M HCl; (B) 200 mL distilled water + 50 mL 1 mM HCl; (C) 0.35 M acetic acid; (D) 0.035 M acetic acid? (Keep in mind that HCl is a strong acid, and thus in aqueous solution its dissociation is practically complete; whereas acetic acid is a weak acid with a dissociation constant of K<sub>a</sub> = 1.74·10<sup>-5</sup>.)

(A)

$$[\text{H}^+] = 0.35 \text{ mol/dm}^3$$

$$\text{pH} = -\log 0.35 = 0.46$$

(B)

$$250/50 = 5\text{-fold dilution}$$

$$[\text{H}^+] = 10^{-3}/5 = 2 \cdot 10^{-4} \text{ M}$$

$$\text{pH} = -\log (2 \cdot 10^{-4}) = 3.7$$

(C)

$$K_a = 1.74 \cdot 10^{-5} = [\text{H}^+][\text{Ac}^-]/[\text{HAc}] = [\text{H}^+]^2/(0.35 - [\text{H}^+]) \quad (\text{as in this case } [\text{H}^+] = [\text{Ac}^-])$$

$$1.74 \cdot 10^{-5} = [\text{H}^+]^2/(0.35 - [\text{H}^+])$$

$$[\text{H}^+]^2 + 1.74 \cdot 10^{-5}[\text{H}^+] - 6.09 \cdot 10^{-6} = 0$$

This is a quadratic equation. One of its roots is negative, the other is:

$$[\text{H}^+] = 2.46 \cdot 10^{-3} \text{ mol/dm}^3$$

$$\text{pH} = -\log(2.46 \cdot 10^{-3}) = \underline{2.61}$$

(D)

Following the above logic:

$$[\text{H}^+] = 7.72 \cdot 10^{-4} \text{ mol/dm}^3$$

$$\text{pH} = -\log(7.72 \cdot 10^{-4}) = \underline{3.11}$$

(25) (SAMPLE EXERCISE) 50 mL 2.0 M  $\text{K}_2\text{HPO}_4$  and 25 mL 2.0 M  $\text{KH}_2\text{PO}_4$  solutions are mixed and then filled up to 200 mL with distilled water. What will be the pH of this solution? ( $\text{pK}_2 = 7.2$ . Knowledge of  $\text{pK}_1$  and  $\text{pK}_3$  values is not necessary, as the concentrations of fully protonated and fully deprotonated forms are negligible in this case.)

$\text{K}_2\text{HPO}_4$ : 200/50 = 4-fold dilution, final concentration:  $2/4 = 0.5 \text{ M}$

$\text{KH}_2\text{PO}_4$ : 200/25 = 8-fold dilution, final concentration:  $2/8 = 0.25 \text{ M}$

$$\text{pH} = \text{pK}_2 + \lg([\text{HPO}_4^{2-}] / [\text{H}_2\text{PO}_4^-]) = 7.2 + \lg(0.5/0.25) = 7.2 + \lg 2 = \underline{7.5}$$

(26) (\*) What is the pH value of the following buffer solutions: (A) 0.5 M acetic acid + 0.1 M sodium-acetate; (B) 0.5 M acetic acid + 0.2 M sodium-acetate; (C) 0.5 M acetic acid + 0.4 M sodium-acetate? ( $K_a = 1.74 \cdot 10^{-5}$ )

(27) (\*) How many grams of Na-H-succinate (molecular mass = 140 Da) and of  $\text{Na}_2\text{-succinate}$  (molecular mass = 162 Da) are needed for 1 litre of solution with a succinate concentration of 50 mM and a pH of 6.0? The second dissociation constant of succinic acid is  $K_{d2} = 2.3 \times 10^{-6}$ . (Knowledge of  $K_{d1}$  is not necessary, as the concentration of the fully protonated form of succinic acid is negligible in this case.)

(28) (\*) How many millilitres of a 5-M NaOH solution must be added to 100 mL of 0.1 M phosphoric acid in order to increase the pH value of the solution from 4 to 9? (Phosphoric acid  $\text{pK}_1 = 2.15$ ,  $\text{pK}_2 = 7.2$ ,  $\text{pK}_3 = 12.37$ .)

(29) (\*) A 0.02-M solution of a certain weak acid has a pH of 4.6. (A) What is the  $[\text{H}^+]$  in the solution? Also calculate the following: (B) the dissociation constant and (C) the pK.

(30) (\*) The acid dissociation constant of acetic acid is  $K_a = 1.74 \cdot 10^{-5}$ . (A) How much does the pH value of 10 mL 0.05 M acetic acid solution change when we add 4 mL of a 0.05-M NaOH solution to it? (B) How many millilitres of a 0.5-M Na-acetate solution must be added to the acetic acid solution described above in order to change the pH value by exactly 1 unit?

(31) (\*) The acid dissociation constant of acetic acid is  $K_a = 1.74 \cdot 10^{-5}$ . (A) How much does the pH value of 10 mL 0.05 M acetic acid solution change if we add 40 mL  $1.25 \cdot 10^{-2}$  M Naacetate solution to it? (B) How many millilitres of a 0.05-M NaOH solution must be added to the acetic acid solution above in order to change the pH value by exactly 1 unit?

(32) (\*) (A) What percentage of arginine side chains are protonated at a pH value of 11.5? (B) What is the average charge of the side chain of arginine at this pH? ( $\text{pK}_a = 12$ )

(33) (\*) (A) What is the protonation percentage of the side chain of tyrosine at a pH value of 10.5? (B) What is the average charge of the side chain of tyrosine at this pH? ( $\text{pK}_a = 10$ )

(34) (\*) The side chain of an amino acid is 14 % protonated at pH = 6.8. (A) Calculate the  $\text{pK}_a$  value of the side chain. (B) Which proteinogenic amino acid is this?

(35) (\*) The side chain of the amino acid selenocysteine has a charge of -0.91 at pH = 6.5. Calculate the  $\text{pK}_a$  value of the side chain.

(36) (\*) How many millilitres of 1.0 M NaOH are needed to change the pH of 50 mL of a 50-mM histidine solution from 2 to 9? (Carboxyl group  $pK_a = 4$ , imidazole group  $pK_a = 6$ , amino group  $pK_a = 9$ .)

(37) Prove the following statement: the isoelectric point of glycine is exactly the arithmetic mean of the  $pK_a$  values of its amino and carboxyl functional groups.

## 12.2.3. Spectrophotometry of biomolecules

(38) (SAMPLE EXERCISE) The molecular mass of the LDH (lactate dehydrogenase) enzyme is 140 kDa. Its molar extinction coefficient is  $\epsilon_{280} = 1.6 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ . The optical absorption of a given LDH solution at 280 nm is  $E_{280} = 0.8$ . (A) Give the molar concentration of the solution (in  $\mu\text{M}$ ). (B) Calculate the hybrid (w/v) percentage concentration and (C) the mg/mL concentration.

(A)

$$E = \epsilon cl \quad c = E/\epsilon = 0.8/(1.6 \cdot 10^5) \quad c = 5 \cdot 10^{-6} \text{ M} = 5 \mu\text{M}$$

(B)

$$\text{in } 100 \text{ mL: } 0.1 \cdot 5 \cdot 10^{-6} = 5 \cdot 10^{-7} \text{ mol LDH}$$

$$\text{This is } 140000 \cdot 5 \cdot 10^{-7} = 0.07 \text{ g}$$

$$0.07 \text{ w/v \%}$$

(C)

$$\text{in 1 litre: } 10 \cdot 0.07 = 0.7 \text{ g}$$

$$\text{g/L} = \text{mg/mL}$$

$$0.7 \text{ mg/mL}$$

(39) (SAMPLE EXERCISE) What percentage of  $\text{NAD}^+$  is reduced in a solution that had an absorbance of 0.9 at 260 nm before reduction, and an absorbance of 0.11 at 340 nm after the reaction? (The molar extinction coefficients of  $\text{NAD}^+$  are  $E_{260} = 18000$ ,  $E_{340} = \text{negligible}$ . The molar extinction coefficient of NADH are  $E_{260} = 18000$ ,  $E_{340} = 6200$ .)

260 nm:

$$c = E/\epsilon = 0.9/18000 = 5 \cdot 10^{-5} \text{ M} \text{ (total amount of } \text{NAD}^+ \text{)}$$

340 nm:

$$c = E/\epsilon = 0.11/6200 = 1.77 \cdot 10^{-5} \text{ M} \text{ (the amount of } \text{NAD}^+ \text{ that became NADH)}$$

$$(1.77 \cdot 10^{-5}/5 \cdot 10^{-5}) \cdot 100 = 35.5\%$$

(40) Write the Lambert-Beer equation. What is determined during a spectrophotometric analysis when one records the spectrum of a substance? What can this spectrum be used for?

(41) Give the definition of extinction and transmittance. How can you detect and quantify protein contamination in a DNA sample? What type of cuvette should be used for this measurement and why?

(42) (\*) The number of SH groups in skeletal muscle myosin can be determined by using Ellmann's reagent (dithionitrobenzoic acid, DTNB). The molar extinction coefficient of the thionitrobenzoate that is released during the reaction of the SH group is  $\epsilon^M = 12000 \text{ M}^1\text{cm}^1$  at 412 nm. 0.3 mL myosin solution, 2.6 mL phosphate buffer and 0.1 mL Ellmann's reagent are measured into the cuvette. The measured  $E_{412\text{nm}}$  is 0.36. The concentration of the myosin stock solution is determined at 100-fold dilution by UV-photometry:  $E_{280\text{nm}} = 0.2$ ;  $\epsilon\% = 8$  (hybrid (w/v) percentage extinction coefficient). How many reactive cysteines are present per enzyme? (The enzyme has a molecular mass of  $5 \cdot 10^5 \text{ Da}$ .)

(43) (\*) The reactive Lys side chains of a protein, which has a molecular mass of 50 kDa, are modified by dansyl chloride. The reaction is performed in 1-mL volume with 0.1 mL protein stock solution. The absorbance of the modified lysine derivate measured in a 1-cm optical path cuvette at 335 nm is  $E_{335\text{nm}} = 0.48$  ( $l = 1 \text{ cm}$ ). The molar extinction coefficient of the lysine derivative (dansyl lysine) is  $\epsilon^M = 3000 \text{ M}^{-1}\text{cm}^{-1}$ . The concentration of the protein solution is determined by applying 0.2 mL stock solution in a 2-mL biuret reaction in which  $E_{540\text{nm}} = 0.25$ . 1 E is equivalent to 4 mg/mL protein concentration. What is the extent of protein labelling (mol Lys/mol protein)?

(44) (\*) The molar absorption coefficient of myoglobin is  $\epsilon_{580\text{nm}} = 1.5 \cdot 10^4 \text{ M}^{-1}\text{cm}^{-1}$ . Its iron content is 0.33 %. Calculate the optical absorption of a myoglobin solution with a concentration of 0.5 mg/mL in a 1-cm cuvette at the wavelength of 580 nm. (The atomic mass of iron is 55.8.)

(45) (\*) The concentration of a protein solution is measured by biuret reaction. 0.05 mL protein solution is mixed into a final test volume of 3.00 mL. The measured  $\Delta E$  is 0.4. The equation of the calibration line is  $y = 2x$  (mg/test). (A) Calculate the concentration of the stock solution. (B) The stock solution is diluted fourfold with distilled water. The absorbance of the diluted solution is 0.5 (measured at 280 nm). It is known from the sequence of the protein that it contains two tryptophans and no tyrosine or phenylalanine. What is the molecular mass of the protein if the molar extinction coefficient of tryptophan is  $5000 \text{ M}^{-1}\text{cm}^{-1}$ ?

(46) (\*) You are to determine the concentration of a protein solution using the Folin reaction. The final volume of the unknown protein solution is 0.2 mL, while the final volume of the test reaction is 1.0 mL. The absorbance of the reagent control is  $E = 0.1$ . The absorbance of the unknown sample is  $E = 0.54$ . The equation of the calibration line for the reagent, corrected with the control, is  $y (A) = 0.5x$  (mg). (A) What is the concentration of the unknown protein solution? (B) After determination of the concentration, the solution is diluted fourfold with distilled water. This diluted solution has an absorbance value (E) of 0.25 at 280 nm. We know that there are two tryptophan residues, but no tyrosine or phenylalanine residues within the protein. The molar extinction coefficient of tryptophan is  $5000 \text{ M}^{-1}\text{cm}^{-1}$ . We suppose that this value does not change when tryptophan is incorporated into the protein. Calculate the molecular mass of the protein.

(47) (\*) The amino acid sequence of the peptide hormone vasopressin is the following: AspTyrPheGluAsnCysPro-LysGly. (A) Calculate the approximate molecular mass of the peptide. (B) The optical absorbance of the peptide at 280 nm is the result of the absorbance of its aromatic amino acids. What is the molar extinction coefficient of vasopressin ( $\epsilon_{280}$ ) if the molar extinction coefficient of tyrosine is  $1.2 \cdot 10^3$ , and that of phenylalanine is  $1.1 \cdot 10^2$ ? (C) When measuring the concentration of a vasopressin solution in a 1-cm cuvette at 280 nm, we record an absorbance of 1.3 A. What is the concentration of the solution, expressed in mg/mL?

(48) (\*) The molecular mass of a protein is 40000 Da. According to the amino acid analysis, it contains 16 mol tyrosine/mol protein. As the protein does not contain any tryptophans, its absorbance measured at 280 nm is solely due to the absorbance of tyrosine side chains, aside from the negligible absorbance of other chromophores. The molar absorptivity (extinction coefficient) of the amino acid tyrosine is  $\epsilon_{280} = 1.5 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$ . (A) What is the molar absorptivity of the protein ( $\epsilon_{280}$ )? (B) A dilute solution of the protein is measured at 280 nm. The resulting absorbance (extinction) is 0.6. Calculate the concentration of the protein solution, expressed in mg/mL.

(49) (\*) The methyl red test is utilised by a microbiologist to determine whether a certain bacterium is weakly or strongly acidogenic. During the test, a glucose-containing liquid growth substrate is inoculated with a pure culture of the bacteria and, after the proper incubation time, a few drops of methyl red are added to the growth substrate. If the bacterium is strongly acidogenic, the indicator will show a red colour. Methyl red (MR) is an organic acid with a  $pK_a$  value of 5.1. The molar mass of its acidic form (HMR) is 269 g/mol, while the sodium salt of methyl red (NaMR) has a molar mass of 291 g/mol. Methyl red is a coloured substance that can be used as a pH indicator because its colour (absorption spectrum) depends on the extent of protonation. (The extent of protonation depends on the  $[\text{H}^+]$ .) The maximal absorbance of the protonated form occurs at 520 nm (the solution is red), while the maximal absorbance of the deprotonated form occurs at 425 nm (the solution is yellow). At 500 nm, both forms show absorbance. The molar extinction coefficients of the protonated (HMR) and deprotonated ( $\text{MR}^-$ ) forms are determined at this wavelength, with the result  $\epsilon_{\text{HMR},500\text{nm}} = 260000 \text{ M}^{-1}\text{cm}^{-1}$  and  $\epsilon_{\text{MR}^-,500\text{nm}} = 65000 \text{ M}^{-1}\text{cm}^{-1}$ . (A) What is the protonation percentage of methyl red at pH = 5.8? (B) What would be the absorbance of a methyl red solution with 10  $\mu\text{M}$  concentration, measured at 500 nm and a pH of 5.8 in a 1-cm cuvette? (C) The following substances are available: solid powder of methyl red (acid form, HMR), sodium salt of methyl red (NaMR) and distilled water. A solution is prepared from these substances with the following parameters:  $c_{\text{MV}} = 10 \mu\text{M}$ , pH = 5.8. How many micrograms of HMR and NaMR are contained in 100 mL of this solution?

## 12.2.4. Cell disruption, cell fractionation and protein isolation

(50) (SAMPLE EXERCISE) During the preparation of an enzyme, the enzyme activity is first measured from the initial tissue homogenate. The result is that 0.2 mL of the homogenate transforms 0.15 µmol of substrate in 5 minutes. The total volume of the tissue homogenate is 200 mL and it has a total protein content of 6 mg/mL. (Note that this is not the amount of the enzyme examined.) After several steps of purification, the volume of the solution containing the isolated enzyme in homogenous form is 0.5 mL with an enzyme concentration of 2 mg/mL. 2 µL of purified enzyme transforms 0.18 µmol of substrate in 5 minutes. What is the degree of purification and the yield (%)?

Tissue homogenate:

200 mL, 6 mg/mL → 1200 mg protein

0.2 mL, 6 mg/mL → 1.2 mg protein

In 5 minutes, 1.2 mg protein transforms 0.15 µmol substrate:

Specific activity:  $0.15/(1.2 \cdot 5) = 0.025 \text{ µmol S/mg protein/min}$

Total activity (1200 mg protein):  $1200 \cdot 0.025 = 30 \text{ µmol S/min}$

Pure enzyme:

2 µL, 2 mg/mL → 0.004 mg protein

0.5 mL, 2 mg/mL → 1 mg protein

In 5 minutes, 0.004 mg protein transforms 0.18 µmol substrate:

Specific activity:  $0.18/(0.004 \cdot 5) = 9 \text{ µmol S/mg protein/min}$

Total activity (1 mg protein):  $1 \cdot 9 = 9 \text{ µmol S/min}$

The specific activity increased 360-fold, therefore the degree of purification is 360-fold.

Yield:  $100 \cdot (9/30) = 30\%$  (70 % of total enzyme activity is lost.)

(51) What is the specific activity of an enzyme?

(52) What is the theoretical base of salting out? What other methods can be used to fractionate the samples?

(53) (\*) The protein content and enzyme activity of a preparation changes during purification according to the following table. Calculate the changes in specific activity (which reflects the extent of purification) and the yield during the purification. Fill out the table with the missing data.

| Purification step                           | Protein (mg) | Enzyme units (calculated based on activity) | Specific activity (unit/mg protein) | Yield (%) |
|---|--------------|---|-------------------------------------|-----------|
| Tissue extract                              | 20000        | 4000000                                     |                                     | 100 %     |
| NH <sub>4</sub> SO <sub>4</sub> precipitate | 5000         | 3000000                                     |                                     |           |
| Ion exchange chromatography                 | 1500         | 1000000                                     |                                     |           |
| Gel filtration                              | 500          | 750000                                      |                                     |           |
| Affinity chromatography                     | 45           | 675000                                      |                                     |           |

(54) (\*) The enzyme activity of an initial tissue homogenate is measured during enzyme preparation. 0.25 mL of homogenate transforms 0.060 µmol substrate in 10 minutes. The total volume of the sample is 100 mL. The protein

concentration of the homogenate is measured using a 0.2-mL sample in a 2-mL Bradford reaction. The result is  $E = 0.2$ . (According to the calibration curve,  $E = 1$  is equivalent to 2 mg/mL protein concentration.) At the end of the purification, 0.5 mL enzyme solution with a concentration of 2 mg/mL is obtained, of which 30  $\mu\text{L}$  is used to measure the activity. 0.16  $\mu\text{mol}$  substrate is transformed in 3 minutes. Calculate the specific activity of the initial homogenate and the purified enzyme, the degree of purification and the yield (%).

## 12.2.5. Peptides and proteins

(55) (SAMPLE EXERCISE) What is the approximate molecular mass of a protein that consists of a single polypeptide chain with 682 amino acids?

The average molecular mass of an amino acid is 110 Da.

$$110 \cdot 682 = 75020 \text{ Da}$$

The approximate molecular mass of the protein is 75 kDa.

(56) (SAMPLE EXERCISE) The following peptides are isolated after trypsin digestion of a peptide: PheArg, SerGlyVal, ProGluAlaLys, and AlaGlyTyrLys. After chymotrypsin digestion of the original peptide, the isolated peptides are the following: ArgAlaGlyTyr, LysSerGlyVal, and ProGluAlaLysPhe. What is the full sequence of the peptide?

Trypsin cleaves this peptide after Arg and Lys, while chymotrypsin does after Phe and Tyr.

By looking for overlaps between the fragments, the full-length sequence can be reconstructed:

|                |  |
|----------------|--|
| Trypsin:       | Pro-Glu-Ala-Lys/Phe-Arg/Ala-Gly-Tyr-Lys/Ser-Gly-Val  |
| Chymotrypsin:  | Pro-Glu-Ala-Lys-Phe/Arg-Ala-Gly-Tyr/Lys-Ser-Gly-Val  |
| Full sequence: | Pro-Glu-Ala-Lys-Phe-Arg-Ala-Gly-Tyr-Lys-Ser-Gly-Val. |

(57) (\*) The amino acid tyrosine is poorly soluble in water under neutral conditions. (A) Is the solubility of a (Tyr)<sub>5</sub> pentapeptide better or worse than the solubility of tyrosine? (B) Can the solubility be improved by changing the pH?

(58) The inactive, denatured and reduced form of the ribonuclease enzyme is rapidly transformed to the active form by the protein disulfide isomerase (PDI) enzyme, whereas insulin is inactivated by PDI in an irreversible manner. What are the structural features that can be in the background of the described effect?

(59) (\*) The six amino acid differences between the sequences of human and duck insulin are shown below. Which insulin has a higher isoelectric point?

| AA position | A8  | A9  | A10 | B1  | B2  | B27 |
|-------------|-----|-----|-----|-----|-----|-----|
| Human       | Thr | Ser | Ile | Phe | Val | Thr |
| Duck        | Glu | Asn | Pro | Ala | Ala | Ser |

(60) For what purpose are the following reagents used during protein sequence determination? (A) CNBr, (B) urea, (C)  $\beta$ -mercaptoethanol, (D) trypsin, (E) performic acid, (F) dansyl chloride, (G) 6 N HCl, (H) ninhydrin, (I) phenyl-isothiocyanate, (J) chymotrypsin.

(61) (\*) The amino acid constitution of a hexapeptide is the following: 2 Arg, Ala, Ser, Val, Tyr. Ala can be detected by N-terminus determination. Two peptides can be isolated by trypsin digestion, the amino acid constitution of which is Arg, Ala, Val and Arg, Ser, Tyr. Similarly, two peptides can be isolated by chymotrypsin digestion, with amino acid constitutions of Arg, Ala, Tyr, Val and Arg, Ser. No amino acids are released after adding carboxypeptidase-A. What is the sequence of the hexapeptide?

(62) (\*) The sequence of human melanotropin is the following: Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val. The  $pK_a$  values of the dissociable groups are: Asp, Glu and the C-terminal carboxyl group:  $pK_a \approx 4$ ; Lys and the N-terminal amino group:  $pK_a \approx 9$ ; Arg:  $pK_a \approx 12$ ; His:  $pK_a \approx 6$ ; Tyr:  $pK_a \approx 11$ . (A) What is the approximate

net charge of the molecule (values smaller than 0.1 can be disregarded) at pH 7 and 9? (B) Is the isoelectric point of the molecule below or above pH = 7? (C) How many peptides will result from CNBr treatment of the molecule? (D) How many peptides will result from trypsin digestion of the molecule? (E) How many peptides will result from chymotrypsin digestion of the molecule?

(63) (\*) A polypeptide yields two peptides after 2-mercaptoethanol treatment:

Chain A: Ala-Val-Cys-Arg-Thr-Gly-Cys-Lys-Asn-Phe-Leu

Chain B: Tyr-Lys-Cys-Phe-Arg-His-Thr-Lys-Cys-Ser

Peptides with the following amino acid content are obtained after trypsin digestion of the original, non-reduced peptide:

1. Lys, Tyr
2. His, Lys, Thr
3. Asn, Leu, Phe
4. Ala, Arg, Cys<sub>2</sub>, Ser, Val
5. Arg, Cys<sub>2</sub>, Gly, Lys, Thr, Phe

Mark the disulfide bonds in the intact polypeptide.

(64) Due to a genetic mutation of a protein that contains a single disulfide bond, a third cysteine appears in the primary structure. What kind of experiment would you perform in order to determine whether the original or a new disulfide bond is present in the mutant protein?

## 12.2.6. Chromatographic methods

(65) (SAMPLE EXERCISE) The molecular mass of a purified protein is 60 kDa, according to gel filtration chromatography. When performing the chromatography in the presence of 6 M urea, the resulting molecular mass is 30 kDa. When 6 M urea and 10 mM mercaptoethanol are present, a single product with a molecular mass of 15 kDa can be detected. What is the structure of the protein like?

Urea denatures proteins, leading to their dissociation into their subunits.

Mercaptoethanol reduces disulfide bonds.

The original protein is a dimer of 2·15 kDa dimers with disulfide bonds.

(66) (SAMPLE EXERCISE) The following protein mixture is separated using a DEAE chromatographic column: fibronectin ( $pI = 5.8$ ), hemoglobin ( $pI = 7.1$ ), lysozyme ( $pI = 11$ ), and RNase ( $pI = 7.8$ ). The pH of the sample is set to 8.5 and the column is also equilibrated with the same buffer solution. (A) What is the chemical nature of the DEAE stationary phase? (B) What happens during the loading of the sample? (C) What principles can be applied to elute the proteins bound to the column? (D) What is the expected order of elution by using the different elution methods?

DEAE is an anion exchanger, i.e. it has positive charges, thus it is able to bind anions.

At  $pH = 8.5$ , proteins with a  $pI$  value below 8.5 will bind to the column, while the ones with a higher  $pI$  will flow through. Therefore, lysozyme will be contained in the flow-through.

Decreasing pH (pH gradient) or increasing ionic strength („salt gradient”) can be used.

When applying a pH gradient, proteins elute in the decreasing order of their  $pI$  values. Thus, RNase will be the first and fibronectin the last. In a salt gradient, the order of elution is less predictable because the binding to the column depends on the number and distribution of charges on the proteins.

(67) How is gel filtration chromatography affected by (A) column length, (B) column diameter, (C) bead size of the stationary phase, (D) pore size, (E) sample volume, and (F) the flow rate of the eluent?

(68) (\*) Three peptides with the following compositions are present in a peptide mixture:

- 25 % Ala, 10 % Gly, 10 % Ser, 10 % Ile, 10 % Leu, 5 % Glu, 5 % Gln, 10 % Lys, 5 % Phe, 10 % Val
- 20 % Ala, 30 % Gly, 10 % Ser, 10 % Asn, 5 % Glu, 10 % Gln, 10 % Lys, 5 % Val
- 20 % Ala, 20 % Gly, 15 % Ser, 10 % Asn, 5 % Asp, 15 % Gln, 10 % Lys, 5 % Val

All three peptides have approximately the same size and isoelectric point. Which separation method would you apply to separate the first peptide from the other two?

(69) (\*) According to quantitative amino acid analysis, bovine serum albumin contains 0.58 % tryptophan. The molecular mass of serum albumin is 70000, based on gel filtration chromatography and gel electrophoresis data. How many tryptophans does serum albumin contain? (The molecular mass of tryptophan is 204 Da.)

(70) (\*) Trypsin digestion is performed on a decapeptide with the following sequence: Gly-Ala-Val-Gly-Tyr-Arg-Val-Lys-Ser-Ile. (A) Which method would you choose to separate the products: ion exchange chromatography or gel filtration? (B) Which chromatographic method should be applied if the digestion was performed with chymotrypsin instead of trypsin?

(71) (\*) (A) What would be the order of elution of the amino acids Met, Phe and Ser, if they were separated on a C18 reverse-phase HPLC column? (B) Why?

(72) (\*) You wish to separate BSA ( $pI = 4.9$ ) and myoglobin ( $pI = 6.0$ ) using a carboxymethyl cellulose column. Is a pH = 8.0 buffer solution suitable for the separation? Why?

(73) (\*) What would be the result of loading a lactate dehydrogenase isoenzyme mixture onto a Sephadex gel filtration column with a pore size fitted to the size of the enzyme?

(74) (\*) What is the order of elution of a mixture of proteins with molecular masses of 8, 15, 54, and 62 kDa in a gel filtration column that separates between 5-20 kDa?

(75) (\*) A peptide mixture contains the following three peptides:

- Val-Glu-Glu-Ser-His-Lys-Gly-Trp-Thr
- Val-Gln-Gln-Ser-His-Lys-Gly-Trp-Thr
- Glu-Glu-Glu-His-Lys

Draw the structure of the amino acids Glu and Gln at neutral pH. Do the masses of the first and second peptides differ? If so, to what extent do they differ? (Carbon: 12 Da, oxygen: 16 Da, sulphur: 32 Da, nitrogen: 14 Da, hydrogen: 1 Da.)

A solution of the peptide mixture at a pH value of 6.0 is prepared. Estimate the charges of the peptides in the solution. ( $pK$  values: N-terminus: 9, Cterminus: 2, lysine: 10, arginine 12, aspartic acid: 3, glutamic acid: 4, histidine: 6, cysteine: 8, tyrosine: 10.)

What would be the result if the peptides were separated by using an anion exchange column at a pH of 6.0 and a linear salt concentration gradient for elution? What would happen if a cation exchange column was used? What would be the result if a gel filtration column was used at the same pH value? (Which peptide would bind to the column? Which peptide would flow through? What would be the order of elution? Give explanations.)

Propose a chromatographic method to separate all three peptides.

(76) (\*) A peptide mixture contains the following four peptides:

- Lys-Glu-His-Leu-Asn-His-Tyr-Gly-Trp-Thr-Ala
- Val-Gln-Ala-Gln-Ser-His-Lys-Gly-Trp-Thr-Lys-Arg-Arg
- Asn-Gln-Gly-His-Gln-Ser-Lys-Lys-Ala-Thr-Ser
- Ala-Val-Tyr-Glu-Glu-His-Lys-Asp-Asp-Glu

The pH of the aqueous solution of the peptides is set to 8.0. Estimate the charge of each peptide in this solution. (pK values: N-terminus: 9, Cterminus: 2, lysine: 10, arginine 12, aspartic acid: 3, glutamic acid: 4, histidine: 6, cysteine: 8, tyrosine: 10.)

What would be the result if the peptide mixture was flown through a cation exchange column at pH = 8?

Calculate the isoelectric point of the peptides that bind to the column.

You wish to perform linear pH gradient elution. Propose a gradient and guess the order of elution.

## 12.2.7. Electrophoretic methods

(77) (SAMPLE EXERCISE) The relative mobility values of a 30-kDa and a 92-kDa protein during SDS polyacrylamide gel electrophoresis are 0.80 and 0.41, respectively. What is the molecular mass of a protein that has a relative mobility of 0.62 in this gel?

A linear relation exists between relative mobility and the logarithm of molecular mass.

The slope of a line ( $y = mx+b$ ) fitted to two points, ( $x_1 = \log 30000$ ,  $y_1 = 0.80$ ) and ( $x_2 = \log 92000$ ,  $y_2 = 0.41$ ), will be:

$$m = (y_2 - y_1) / (x_2 - x_1) = (0.41 - 0.80) / (4.964 - 4.477) = -0.801$$

Its intercept will be:

$$b = y_1 - m \cdot x_1 = 4.39$$

Consequently, the equation of the line will be  $y = -0.801x + 4.39$

Substituting the relative mobility value of the unknown protein:

$$(0.62 - b) / m = \log(M_r) = 4.702$$

$$M_r = 10^{4.702} = 50320 \text{ Da}$$

(78) Tropomyosin is eluted earlier than actin during gel filtration separation. Which protein would have a greater mobility in SDS-PAGE and what is the explanation of this apparent contradiction?

(79) (A) What kind of advantages does the polyacrylamide gel have that makes it especially suitable for electrophoretic uses? (B) Which anions conduct electricity in the acrylamide gel used in the laboratory for native electrophoresis (pH = 8.8)? Write their formulae at pH = 8.8.

(80) (A) Which parameters influence the mobility of a protein during gel electrophoresis? (B) How does the presence of urea (carbamide) and Na dodecyl sulphate (SDS) influence the mobility of the sample?

(81) The relative mobility of a given protein is found to be significantly higher during SDS-PAGE in the presence of mercaptoethanol than in its absence. How can this effect be explained?

(82) (\*) Proteins are separated using SDS polyacrylamide gel electrophoresis. Electrophoresis is performed at 150 V for 90 minutes. The current intensity is 100 mA during the experiment. The length of the gel is 10 cm, and the migration distance of the front is 9.1 cm. The relative mobility values of two proteins with molecular masses of 25 kDa and 85 kDa are 0.85 and 0.45, respectively. (A) What is the molecular mass of a protein that has a relative mobility of 0.7 in this particular gel? (B) The molecular mass of a dimeric protein is 60 kDa. It has two polypeptide chains with identical masses, linked by one disulfide bond. What mobility would this protein show in the same gel? (C) What would be the mobility of this protein in the presence of mercaptoethanol?

(83) When the enzyme enterokinase produces active trypsin from inactive trypsinogen, the following pentapeptide is released from the N-terminus of trypsinogen: AspAspAspAspLys. The cleavage of the pentapeptide does not cause a significant change in the size of the protein. Can the activity of enterokinase be demonstrated by combining native gel electrophoresis with Coomassie staining? Give an explanation of your answer.

(84) Can a point mutant enzyme containing a single Arg-to-His substitution be separated from its wild-type counterpart by native gel electrophoresis at pH = 8.8? (Arg: pK<sub>a</sub> = 12.0, His: pK<sub>a</sub> = 6.5.)

## 12.2.8. Protein-ligand interactions

(85) (SAMPLE EXERCISE) What percentage of receptor molecules will be in complex if the concentration of the free ligand is exactly three times the value of the dissociation constant?

$$K_d = [R][L]/[RL]$$

$$K_d = [R]3K_d/[RL]$$

$$1/3 = [R]/[RL]$$

$$[RL] = 3[R]$$

The fraction of receptor molecules in complex will be:

$$[RL]/[R]_{\text{total}} = [RL]/([RL]+[R]) = 3[R]/4[R] = 3/4 = 0.75$$

Therefore, 75% of receptor molecules will be in complex.

(86) (\*) What is the dissociation constant if 67 % of receptor molecules are in complex and the free ligand concentration is 5 μM?

(87) (\*) The dissociation constant of a simple protein-ligand interaction is  $3 \cdot 10^{-6}$  M. (A) What is the free protein:bound protein ratio when the system is in equilibrium and the ligand concentration is 600 nM? (B) What percentage of protein molecules are ligand-bound in this case?

(88) (\*) (A) In a solution, the total receptor concentration is 5 μM. The free ligand concentration is 10 μM in equilibrium. Calculate the dissociation constant if 33 % of receptor molecules are in complex. (B) Calculate the total ligand concentration. (C) What percentage of ligand molecules are in complex?

(89) (\*) A given protein forms a complex with a small ligand molecule. The complex contains one molecule of protein and one molecule of ligand. The dissociation constant is 2 μM. One millilitre of solution of the complex is dialysed against 99 mL phosphate buffer. The total concentration of the protein is 300 μM within the dialysis bag. The dialysis membrane is impermeable for the protein, but is permeable for the ligand. After the equilibrium has been reached, the ligand concentration in the external solution is 5 μM. (A) What percentage of protein is in complex? (B) What was the (molar) ligand/protein ratio within the membrane before the dialysis?

(90) (\*) The following solutions are measured into an Eppendorf tube: 10 μl receptor (from a 100-μM stock solution), 20 μl ligand (from a 200-μM stock solution) and 70 μl phosphate buffer (pH = 7.5). The receptor is a monomer. It can bind only one ligand molecule. The dissociation constant ( $K_d$ ) is 20 μM. Calculate the equilibrium concentrations of the receptor, the ligand and the complex.

(91) (\*) A protein forms a complex at a 1:1 ratio with a peptide. The dissociation constant is 30 μM. What percentage of the protein will be peptidebound if 10 μl of a 2.2-mM peptide solution is mixed with 10 μl of a 2-mM protein solution?

## 12.2.9. Enzyme kinetics

(92) (SAMPLE EXERCISE) At what substrate concentration can you measure the quarter of the maximal velocity of the enzyme if  $K_M = 5 \cdot 10^{-3}$ ?

$$v_0 = v_{\max}/4$$

Substituting this into the Michaelis-Menten equation:

$$v_0 = v_{\max}[S]/(K_M+[S])$$

$$v_{\max}/4 = v_{\max}[S]/(K_M + [S])$$

$$1/4 = [S]/(5 \cdot 10^{-3} + [S])$$

$$[S] = 1.67 \text{ mM}$$

(93) (SAMPLE EXERCISE) We measure the velocity of an enzyme reaction by determining light absorbance of the product by photometry. The volume of the reaction mixture is  $2 \text{ cm}^3$ , which contains  $5 \mu\text{g}$  enzyme. The change in light absorbance in 1 minute during the linear phase of the reaction is  $\Delta E = 0.15$  (measured in a 1-cm cuvette). The molar extinction coefficient of the product is  $\epsilon = 3 \cdot 10^4 \text{ M}^{-1}\text{cm}^{-1}$ . The molar mass of the enzyme is 50000 g/mol. How many moles of substrate are transformed by 1 mol enzyme in 1 minute?

According to the Lambert-Beer equation:  $E = \epsilon cl$

$$\Delta E = \Delta c \epsilon l$$

$$\Delta c_{\text{product}} = \Delta E / (\epsilon \cdot l) = 0.15 / (3 \cdot 10^4 \text{ M}^{-1}\text{cm}^{-1} \cdot 1 \text{ cm}) = 5 \cdot 10^{-6} \text{ M}$$

$$c_{\text{enzyme}} = m_{\text{enzyme}} / M_{\text{enzyme}} / V = 5 \cdot 10^{-6} \text{ g} / 50000 \text{ gmol}^{-1} / 0,002 \text{ dm}^3 = 5 \cdot 10^{-8} \text{ M}$$

$$\text{Turnover number} = \Delta c_{\text{product}} / (c_{\text{enzyme}} \cdot t) = 100 \text{ min}^{-1}$$

1 mol enzyme transforms 100 mol substrate in 1 minute.

(94) (A) Draw the kinetic scheme, the progress curve and the substrate saturation curve of an enzyme reaction with one substrate. (B) What is the Michaelis constant ( $K_M$ )?

(95) Enzyme reaction velocities are usually determined by measuring the amount of product, and not the decrease in the amount of substrate. Why?

(96) (\*) Specify the dimensions of  $K_M$ ,  $v_{\max}$  and  $k_{\text{cat}}$ .

(97) (\*) Do we need to know the value of  $[E]_T$  when measuring the kinetic parameters of an enzyme in order to determine  $K_M$ ,  $v_{\max}$  and  $k_{\text{cat}}$ ?

(98) (\*) The following reaction mixture needs to be prepared for determining the reaction velocity of an enzyme: 50 mM phosphate buffer, 2 mM NADH, 8 mM Na-piruvate, and  $5 \mu\text{g}/\text{mL}$  lactate dehydrogenase. (A) What kind of stock solutions would you prepare? How much of these stock solutions would you measure into a final reaction volume of 2 mL to obtain the mixture described above? In order to maintain the desired accuracy of the measurement, use volumes of reagents at least 10  $\mu\text{l}$ . (Solubility values: NADH: ~140 mM; Na-piruvate: ~910 mM; phosphate buffer: ~1 M; lactate dehydrogenase: ~10 mg/mL.) (B) What is the simplest method to determine the activity of the enzyme? (Keep in mind the spectra of  $\text{NAD}^+$  and  $\text{NADH}$ .)

(99) (\*) (A) What will be the  $v_0/v_{\max}$  ratio if the value of the substrate concentration equals three times the value of  $K_M$  in an enzyme-catalysed reaction? (B) What percentage of the enzyme measured into the reaction will be substrate-bound in this case?

(100) (\*) The maximal velocity of an enzyme is  $100 \mu\text{mol} \cdot \text{dm}^{-3} \cdot \text{s}^{-1}$ . The  $K_M$  is 2 mM. What will be the reaction velocity if the substrate concentration is 20 mM?

(101) (\*) How many moles of p-nitrophenyl- $\beta$ -D-glucoside are transformed in 1 second by 1 mol glucosidase enzyme (molar activity or turnover number) if a reaction mixture (2 mL final volume) containing  $5 \mu\text{g}$  enzyme shows an extinction change of 0.63 in 2 minutes at 410 nm? The molar extinction coefficient of p-nitrophenol is  $\epsilon_{410} = 1.75 \cdot 10^4 \text{ M}^{-1}\text{cm}^{-1}$ . The molecular mass of glucosidase is 350000 Da.

(102) (\*) The activity of lactate dehydrogenase is measured using a spectrophotometer. The 3-mL test contains 0.1 mL of an LDH stock solution, which has a concentration of 1  $\mu\text{M}$ , and the adequate amount of substrates. The decrease in the amount of one substrate is measured at 340 nm during the reaction. The change in extinction is 0.6 in 5 minutes. The molar extinction coefficient of the substrate is  $6 \cdot 10^3 \text{ M}^{-1}\text{cm}^{-1}$ . The molecular mass of the enzyme is  $10^5$  Da. (A) What is the specific activity of LDH, expressed in  $\mu\text{mol S/mg E/min}$ ? (B) What is the turnover number of the enzyme ( $k_{\text{cat}}$ ), expressed in  $\text{s}^{-1}$ ? (The enzyme is a tetramer.) (C) What are the substrates of the reaction?

(103) (\*) Carbonic anhydrase is an enzyme with one of the highest turnover numbers. It catalyses the reversible hydration of  $\text{CO}_2$ , which is a key step during  $\text{CO}_2$  transport between the lungs and other tissues. The molecular mass of erythrocyte carbonic anhydrase is  $3 \cdot 10^4$  Da. 10 µg pure carbonic anhydrase catalyses the hydration of 0.7 g  $\text{CO}_2$  in one minute at 37°C. Calculate the turnover number of the enzyme ( $k_{\text{cat}}$ ).

(104) (\*) The reaction velocity of an enzyme is determined by following the increase in light absorbance of the product. The volume of the reaction mixture is 1 cm<sup>3</sup>, and it contains 1 µg enzyme. The change in light absorbance in one minute is 0.3 E, measured in a 1-cm cuvette. The molar extinction coefficient of the product is  $\epsilon = 6 \cdot 10^4$  M<sup>-1</sup>cm<sup>-1</sup>. The molar mass of the enzyme is 25000 g/mol. How many moles of substrate are transformed by 1 mol enzyme in 1 second?

(105) (\*) We follow the activity of trypsin by using BAPNA (N-benzoyl-arginine p-nitroanilide) substrate. (A) What is the name and structural formula of the coloured product? (B) The concentration of the trypsin stock solution is 0.02 mg/mL. The molecular mass of the enzyme is 25000 Da. With increasing substrate concentration in reaction mixtures having a final volume of 2 mL, a maximal extinction change ( $\Delta E$ ) of  $\sim 0.4 \text{ min}^{-1}$  can be detected. All tests are performed in a 0.5 cm path length cuvette. Twenty microlitres of the trypsin stock solution are added to each mixture. The molar extinction coefficient of the product is  $\epsilon_{405} = 8000 \text{ M}^{-1}\text{cm}^{-1}$ . Calculate the maximal number of BAPNA molecules transformed by a single trypsin molecule in 1 second. What is the name of this value?

(106) (\*) Hydrolysis of pyrophosphate resulting from some synthetase reactions is an important thermodynamic driving force of biosynthetic reactions. The pyrophosphatase enzyme of *E. coli* contains six identical subunits and has a molecular mass of 120 kDa. Its activity under certain conditions is defined as the following: one unit of the enzyme hydrolyses 10 µmol pyrophosphate in 15 minutes. The activity of the purified, homogenous enzyme is  $v_{\text{max}} = 2800 \text{ unit/mg enzyme}$ . (A) How many moles of substrate are hydrolysed per second by 1 mg enzyme? (B) How many moles of substrate are hydrolysed per second by 1 mol enzyme? (C) How many moles of active sites does 1 mg enzyme contain? (D) What is the turnover number of the enzyme?

(107) We wish to determine the  $K_M$  of an enzyme. For this purpose, initial reaction velocities are measured as the function of increasing substrate concentration. At the end of the measurement it turns out that, above a certain concentration value, the substrate precipitates under the reaction conditions. How does this undesired effect influence the results of the measurement?

(108) We measure the reaction velocity of an enzyme as the function of substrate concentration. The  $K_M$  of the enzyme is 2 µM. The enzyme concentration during the measurement is 200 nM, and the substrate concentration is varied within the 0.1 µM-10 µM interval. What is wrong with this assay and how can it be improved?

(109) (\*) Penicillin-resistant bacteria contain an enzyme named penicillinase (or  $\beta$ -lactamase) that hydrolyses penicillin. The molecular mass of *Staphylococcus aureus* penicillinase is 19.6 kDa. Its activity is measured using  $10^{-9}$  g purified penicillinase in a test solution of 10 mL volume. We assume that substrate concentration does not change significantly during the experiment. The activity of the enzyme (the amount of hydrolysed penicillin) is measured for one minute as the function of substrate concentration. The results are shown below. (A) Does penicillinase follow MichaelisMenten kinetics? (B) If yes, what are the  $K_M$  and  $v_{\text{max}}$  values of the enzyme? (C) Under the given conditions, what is the turnover number of the enzyme?

| [penicillin] (µM) | amount of hydrolysed penicillin (nmol) |
|-------------------|--|
| 1                 | 0.11                                   |
| 3                 | 0.25                                   |
| 5                 | 0.34                                   |
| 10                | 0.45                                   |
| 30                | 0.58                                   |
| 50                | 0.61                                   |

(110) (\*) The catalytic rate constant of an enzyme ( $k_{\text{cat}}$ ) is  $50 \text{ s}^{-1}$ . The extinction change measured in 3 minutes in a test having a final volume of 1 mL is  $\Delta E = 0.36$ . The molar extinction coefficient of the product is  $8000 \text{ M}^{-1}\text{cm}^{-1}$ . (A) Calculate the enzyme concentration. (B) What amount of substrate needs to be measured into the test described above in order to meet the requirements of determining the initial reaction velocity?

(111) (\*) We find an Eppendorf tube with the title „hu 4, 03.07.28.” in the -20°C refrigerator. This can only be trypsin in the given laboratory, but no further information can be found in any of the records. The concentration of the trypsin solution is determined by spectrophotometry. For this, the stock solution is diluted 10-fold, and then the extinction is measured at 280 nm. The result is  $E_{280} = 0.2$ . The enzyme activity of the trypsin solution is measured as well. 10 µl of a 50-times diluted enzyme solution and 10 µl of 15 mM ZGlyProArgpnitroaniline are measured into a final volume of 1 mL. What percentage of the enzyme amount is active if the slope of the linear phase of the time curve is  $0.72 \text{ min}^{-1}$ ? (The molar extinction coefficient of trypsin is  $\epsilon_{280} = 40000 \text{ M}^{-1}\text{cm}^{-1}$ . The molar extinction coefficient of p-nitroaniline is  $\epsilon_{405} = 8000 \text{ M}^{-1}\text{cm}^{-1}$ . The light path is 1 cm in both cases. The kinetic constants of human trypsin 4 at 37°C are approximately  $k_{\text{cat}} = 300 \text{ s}^{-1}$  and  $K_M = 5 \cdot 10^{-5} \text{ M}$ .)

(112) (\*) The  $K_M$  value of an enzyme reaction is expected to be in the 0.1-1 mM range. What volumes would you measure into 6 different tubes from the stock solutions of the enzyme and the substrate to determine  $K_M$  if the concentration of the enzyme stock solution was 2 mg/mL, while there were substrate solutions in 100 mM and 10 mM concentrations available, and the final volume of the tests should be 2 mL? (The final concentration of the enzyme must be 10 µg/mL.)

(113) (A) How would you measure the activity of trypsin? (B) How does the value of  $v_{\text{max}}$  change in the presence of a competitive inhibitor? (C) How do  $K_M$  and  $v_{\text{max}}$  values, respectively, change if you irreversibly inhibit one third of the enzyme molecules?

(114) The reaction velocity of an enzyme is measured at different substrate concentrations, in the presence or absence of an inhibitor. When the inhibitor is present, the resulting reaction velocity values at each substrate concentration are smaller than those in the absence of the inhibitor. However, the extrapolated value of  $v_{\text{max}}$  is the same in both cases. Interpret this finding.

## 12.2.10. Recombinant DNA technology

(115) (SAMPLE EXERCISE) 7 guanines are present in a DNA segment of 20 base pairs. How many adenines are there in the segment?

20 base pairs → 40 bases

C = G → 7 pieces of G and 7 pieces of C are there in the segment

A+T = 40-(C+G) = 40-(7+7) = 26 bases (A and T together)

A = T → 26/2 = 13 adenines

(116) (\*) (A) Write the mRNA sequence transcribed from the DNA sequence shown below. (B) Presuming that the reading frame starts at the beginning of this mRNA sequence, write the corresponding amino acid sequence.

5'-CATCCAATCACTAGGCCAA-3'

(117) (\*) We add a poly(UUAC) template to a cellfree *in vitro* protein synthesis system. What will be the sequence of the resulting polypeptide chain?

(118) (\*) The size of the haploid genome of a diploid organism is 45000 kb. The genome contains 21 % guanine base. Calculate the number of the four different bases in the diploid genome.

(119) (\*) Design two primer sequences with a total length of 10 bases each that can be used to amplify the following DNA segment by PCR:

ATAGGCATAGGCCATATGGCATCTTATAATATGCGATAGCGCTGGTCA

## 12.2.11. Bioinformatics

(120) (A) Identify the nucleotide sequence below by using the *BLAST* sequence alignment algorithm (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). (B) Translate the sequence using the *Translate* software (<http://web.expasy.org/translate/>), and then identify the protein sequence with *BLAST*. (C) Find the protein in the

*UniProt* database (<http://www.uniprot.org/>). (D) Determine the isoelectric point and the molar extinction coefficient of the protein using the *ProtParam* software (<http://web.expasy.org/protparam/>).

TACGCAATCGTATCATTCTGCTGGCGCTCCGGCGCAGGTAAAGGTACTCAGGCTCAATTCAT-CATGGAGAAAATACGGCATTCCGCAAATCTACTGGTGACATGTTGCGCGCCGCTGAAAAGCAG-GTTCTGAGTTAGGTCTGAAAGCAAAAGAAATTATGGATGCGGGCAAGTTGGTACTGATGAGTTAGTTATCGCATTACTCAAAGAACGTATCACACAGGAAGAGATTGCCGCATGGTTCTGTTAGACGG-GTTCCCGCGTACCATTCCTCAGGCAGATGCCATGAAAGAAGCCGGTACAAAGTTGATTATGTGCT-GGAGTTGATGTTCCAGACGAGCTGATTGAGCGCATTGTCGGCGTCGGTACATGCTGCTTCAG-GCCGTGTTATCACGTTAAATTCAACCCACCTAAAGTTGAAGATAAAAGATGATGTTACCGGTGAA-GAGCTGACTATTCTGAAAGATGATCAGGAAGCGACTGTCCGTAAGCGTCTTATCGAATATCAT-CAACAAACTGCACCATTGGTTCTACTATCATAAAGAAGCAGGATGCAGGTAAATACGCAATA-TTTAAACTGGACGGAACCGTAATGTAGCAGAAGTCAGTGTGAACGGCGACTATTCTCGGT-TAATTCTGGATGGCCTTATAGCTAAGGCGTTAAGGCCCTAGCTATTCAAGTAAGAAGG-GCGTAGTACCTACAAAAGGAGATTGGCATGATGCAAAGCAAACCCGGGTATTAATGGTTAATTGGG-GACACCAGATGCTCAACGTCGAAAGCTATCAAGCGTTATTAGCTGAGTTTGAGTGACCGCCGG-GTAGTTGATACTCCCCATTGCTATGGTGCCATTGCTGCATGGTGTATTACCGCTTCGGTCAC-CACGTGTAGCAAAACTTATCAATCCGTTGGATGGAAGAGGGCTCCTTATTGGTTATAGCCGC-CGCCAGCAGAAAGCACTGGCAGCAAGAATGCCTGATATTCCGTAGAATTAGGCATGAGCTATG-GTTCAC

(121) Identify the following protein sequence using the sequence similarity search software *BLAST* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Does it have any orthologs?

MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVVMGMQKDSYVGDEAQSKR-GILTLKYPIEHHIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREKMTQIMFETFN-VPAMYVAIQAVLSLYASGRTTGIVLDSDGDGVTHNVIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYS-FVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEKSYELPDGVITIGNERFRCPETLFQPSFIGMESAGI-HETTYNSIMKCDIDIRKDLYANNVMSGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGG-SILASLSTFQQMWITKQEYDEAGPSIVHRKCF

(122) Identify the following protein sequence using the sequence similarity search software *BLAST* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Identify a human, an archaeal and a plant ortholog of the protein. What are the extents of their similarity?

APSRKFFVGGNWKMNGRKKNLGELITTLNAAKVPADTEVVVCAPPTAYIDFARQKLDPKIAVAAQNCYK-VTNGAFTGEISPGMIKDCGATWVVLGHSERRHVFGESDELIGQKVAHALSEGLVIACIGEKLDEREAGIT-EKVVFEQTKVIAADNVKDWSKVVLAYEPVWAIGTGKTATPQQAQEVHEKLRGWLKSNVS-DAVAQSTRIIYGGSVTGATCKELASQPDVDGFLVGGASLKPEFVDIINAKQ

## 12.3. Solutions

### 12.3.1. Units of measure, solutions

(1) (SAMPLE EXERCISE)

(2) (SAMPLE EXERCISE)

(3) (A)  $1.98 \cdot 10^{12}$ ; (B)  $3.94 \cdot 10^{10}$ ; (C) 50.

(4) (A)  $1.1 \cdot 10^{-12}$  g; (B) 5.36 %; (C) 4.56 %; 54.54 pcs;  $3.32 \cdot 10^{-9}$  mol/dm<sup>3</sup>; 127470.

(5) (A) 1.6 mm; the *E. coli* cell is approximately 1  $\mu$ m, therefore the DNA is much longer => it is coiled up. (B) 3914 pcs.

(6) 5.

(7) 95.5 nm.

- (8) 70257 Da.
- (9) 501833 Da.
- (10) 32-42 peptide bonds/s.
- (11) 200000 Da.
- (12) 15879 Da.
- (13) 13201 Da.
- (14) 0.089 %.
- (15) (A)  $1.01 \cdot 10^{-2}$  mol / 685.48 g; (B)  $4.03 \cdot 10^{-2}$  mol; (C) 4.93 L.
- (16) 1000  $\mu\text{mol}/\text{dm}^3$ , 1  $\mu\text{mol}/\text{mL}$ .
- (17) 0.2  $\text{dm}^3$ .
- (18) (A)  $5.85 \cdot 10^{-2}$  mol/ $\text{dm}^3$ ; (B) Not provided; (C)  $3.78 \cdot 10^{-1}$  mM.
- (19) (A)  $4 \cdot 10^{-5}$  mol/ $\text{dm}^3$ ; (B) 0.1 w/v %.
- (20) 28.2 mL.
- (21) (A) 2 mg/mL; (B) 0.2 w/v %.
- (22) Measure 2 mL potassium phosphate buffer + 25 mL serum albumin + 0.877 g NaCl, then add distilled water to a volume of 100 mL.
- (23) Lactate: 0.2 mL (200  $\mu\text{l}$ ); phosphate buffer: 0.8 mL (800  $\mu\text{l}$ ); NAD: 0.02 mL (20  $\mu\text{l}$ ); lactate dehydrogenase: 0.011 mL (11  $\mu\text{l}$ ).

## 12.3.2. Ionisation equilibria

- (24) (SAMPLE EXERCISE)
- (25) (SAMPLE EXERCISE)
- (26) (A) 4.06; (B) 4.36; (C) 4.66.
- (27)  $m_{\text{Na-H-succinate}} = 2.12 \text{ g}$ ;  $m_{\text{Na}_2\text{-succinate}} = 5.65 \text{ g}$ .
- (28) 1.97 mL.
- (29) (A)  $2.51 \cdot 10^{-5}$  mol/ $\text{dm}^3$ ; (B)  $3.16 \cdot 10^{-8}$  M; (C) 7.5.
- (30) (A) 1.55; (B) 0.19 mL.
- (31) (A) 1.73; (B) 1.58 mL.
- (32) (A) 76 %; (B) +0.76.
- (33) (A) 24 %; (B) -0.76.
- (34) (A) 6; (B) His.
- (35) 5.5.
- (36) 6.25 mL.

(37) Not provided.

### 12.3.3. Spectrophotometry of biomolecules

(38) (SAMPLE EXERCISE)

(39) (SAMPLE EXERCISE)

(40) Not provided.

(41) Not provided.

(42) 6.

(43) 8 mol Lys/mol protein.

(44) 0.44.

(45) (A) 4mg/mL; (B) 20000 Da.

(46) (A) 4.4 mg/mL; (B) 44000 Da.

(47) (A) 990 Da; (B)  $1.31 \cdot 10^3 \text{ M}^{-1}\text{cm}^{-1}$ ; (C) 0.982 mg/mL.

(48) (A)  $2.4 \cdot 10^4 \text{ M}^{-1}\text{cm}^{-1}$ ; (B) 1 mg/mL.

(49) (A) 16.6 %; (B) 0.97; (C) 44.7  $\mu\text{g}$  HMV and 242.6  $\mu\text{g}$  NaMV.

### 12.3.4. Cell disruption, cell fractionation and protein isolation

(50) (SAMPLE EXERCISE)

(51) Not provided.

(52) Not provided.

(53) Specific activity (unit/mg protein): 200, 600, 666.6, 1500, 15000; Yield: 100 %, 75 %, 25 %, 18.7 %, 16.8 %.

(54) Specific activity (homogenate): 6 nmol/min/mg protein; specific activity (purified enzyme): 0.88  $\mu\text{mol}/\text{min}/\text{mg}$  protein; degree of purification: 148-fold; yield: 37 %.

### 12.3.5. Peptides and proteins

(55) (SAMPLE EXERCISE)

(56) (SAMPLE EXERCISE)

(57) (A) The pentapeptide is less soluble. (B) Yes.

(58) Not provided.

(59) Human.

(60) Not provided.

(61) Ala-Val-Arg-Tyr-Ser-Arg.

(62) (A) at pH = 7: +1; at pH = 9: 0. (B) pI = 9.

(C) Results in 2 peptides; results in 3 peptides; results in 4 peptides.

(63) A chain: Ala-Val-Cys-Arg-Thr-Gly-Cys-Lys-Asn-Phe-Leu; B chain: Tyr-Lys-Cys-Phe-Arg-His-Thr-Lys-Cys-Ser; Disulfide bonds: (A)Cys3-(B)Cys9 and (A)Cys7-(B)Cys3.

(64) Not provided.

### 12.3.6. Chromatographic methods

(65) (SAMPLE EXERCISE)

(66) (SAMPLE EXERCISE)

(67) Not provided.

(68) Hydrophobic chromatography.

(69) 2.

(70) (A) Both are needed in the case of trypsin treatment. (B) In the case of chymotrypsin treatment, ion exchange should be used.

(71) (A) Order of elution: Ser, Met, Phe. (B) Polar molecules bind weakly, whereas apolar molecules bind strongly to the reverse-phase beads.

(72) No, because at pH = 8.0 both proteins have a negative charge and do not bind to carboxymethyl cellulose, which also has a negative charge.

(73) The isoenzyme mixture cannot be separated by gel filtration chromatography, as the sizes of the tetramers with four subunits are nearly identical. (They differ only in their charges.)

(74) The 54-kDa and 62-kDa proteins elute together in the exclusion volume; then the 15-kDa, and finally the 8-kDa protein elutes, assuming that their shape is not significantly asymmetrical.

(75) (A) There is no difference between the masses. (B) 1st: -1.5; 2nd: +1.5; 3rd: -1.5. (C) Anion exchange: the 1st and the 3rd bind, the 2nd flows through (positive charge); during elution, the 1st and the 3rd give a single peak as their charge is the same. Cation exchange: the 2nd binds, the 1st and the 3rd flows through (negative charge); during elution, the 2nd gives a single peak. Gel filtration: there is no difference between the size of the 1st and the 2nd, thus they produce a single peak; the 3rd has a smaller size, therefore it gives a separate peak at a larger elution volume. (D) E.g. gel filtration, followed by anion exchange chromatography for 1st + 2nd.

(76) (A) 1st: 0; 2nd: +4; 3rd: +2; 4th: -4. (B) The 2nd and 3rd peptides bind, the others flow through. (C) 2nd: pI = 12; 3rd: pI = 10. (D) Increasing pH gradient: pH = 8 => pH = 13. Order: The 3rd, then the 2nd peptide.

### 12.3.7. Electrophoretic methods

(77) (SAMPLE EXERCISE)

(78) Not provided.

(79) Not provided.

(80) Not provided.

(81) Not provided.

(82) (A) 39559 Da; (B) 0.56; (C) 0.79.

(83) Not provided.

(84) Not provided.

## 12.3.8. Protein-ligand interactions

(85) (SAMPLE EXERCISE)

(86)  $2.5 \cdot 10^{-6}$  M.

(87) (A)  $[P]/[PL] = 5$ ; (B) 16.67 %.

(88) (A)  $2 \cdot 10^{-5}$  M; (B)  $1.165 \cdot 10^{-5}$  M; (C) 14.16 %.

(89) (A) 71.43 %; (B) 2.38.

(90)  $[RL] = 6.28 \cdot 10^{-6}$  M;  $[R] = 3.72 \cdot 10^{-6}$  M;  $[L] = 3.372 \cdot 10^{-5}$  M.

(91) 88 %.

## 12.3.9. Enzyme kinetics

(92) (SAMPLE EXERCISE)

(93) (SAMPLE EXERCISE)

(94) Not provided.

(95) Not provided.

(96)  $[K_M] = \text{mol} \cdot \text{dm}^{-3}$ ;  $[v_{\max}] = \text{mol} \cdot \text{dm}^{-3} \cdot \text{s}^{-1}$ ;  $[k_{\text{cat}}] = \text{s}^{-1}$ .

(97)  $K_M$ : no;  $v_{\max}$ : no;  $k_{\text{cat}}$ : yes.

(98) (A) Lactate dehydrogenase: 20  $\mu\text{l}$  of 0.5 mg/mL stock solution; Na pyruvate: 40  $\mu\text{l}$  of 400 mM stock solution; NADH: 40  $\mu\text{l}$  of 100 mM stock solution; phosphate buffer: 200  $\mu\text{l}$  of 500 mM stock solution. (Several correct answers are possible.) (B) Following absorbance at 340 nm.

(99) (A) 3/4; (B) 75 %.

(100)  $90.9 \mu\text{mol} \cdot \text{dm}^{-3} \cdot \text{s}^{-1}$ .

(101) 42 mol p-nitrophenyl- $\beta$ -D-glucoside is transformed by 1 mol glucosidase enzyme per 1 second.

(102) (A) 6  $\mu\text{mol S/mg E/min}$ ; (B)  $2.5 \text{ s}^{-1}$ ; (C) pyruvate and NADH, or lactate and  $\text{NAD}^+$  (the reaction is reversible).

(103)  $795000 \text{ s}^{-1}$ .

(104) 1 mol enzyme transforms 2.08 mol substrate per 1 s.

(105) (A) Para-nitroaniline; (B) One trypsin molecule transforms a maximum of 208 BAPNA molecules per second (turnover number).

(106) (A)  $3.11 \cdot 10^{-5} \text{ mol PP}_i / \text{s/mg enzyme}$ ; (B) 1 mol enzyme hydrolyses 3732 mol substrate per second; (C)  $5 \cdot 10^{-8} \text{ mol}$ ; (D)  $622 \text{ s}^{-1}$ .

(107) Not provided.

(108) Not provided.

(109) (A) It follows Michaelis-Menten kinetics. (B)  $K_M = 5.22 \cdot 10^{-6} \text{ M}$ ;  $v_{\max} = 1.14 \cdot 10^{-9} \text{ M/s}$ ;

(C)  $k_{\text{cat}} = 224 \text{ s}^{-1}$ .

(110) (A)  $5 \cdot 10^{-9}$  mol/dm<sup>3</sup>; (B)  $9 \cdot 10^{-4}$  M (minimal substrate concentration measured).

(111) 66.7 %

(112) Enzyme: 10 µl; 100 mM substrate: 200 µl, 100 µl, 20 µl; 10 mM substrate: 100 µl, 20 µl, 10 µl.

(113) Not provided.

(114) Not provided.

## 12.3.10. Recombinant DNA technology

(115) (SAMPLE EXERCISE)

(116) (A) 5'-UUGCCUAGUGAUUGGAUG-3'; (B) Leu-Pro-Ser-Asp-Trp-Met.

(117) [Leu-Leu-Thr-Tyr]<sub>n</sub>.

(118) A, T =  $5.22 \cdot 10^7$ ; C, G =  $3.78 \cdot 10^7$ .

(119) Forward primer: 5'-ATAGGCATAG-3'; Reverse primer: 5'-TGACCAGCGC-3'.

## 12.3.11. Bioinformatics

(120) Not provided.

(121) Not provided.

(122) Not provided.

---

# Chapter 13. Epilogue

by Attila Reményi

The real test of this e-book will be if students will successfully employ its information in laboratory practice to solve real-life problems. The aim of this overview is to support students in this complex learning exercise. Below we will collate and summarise the study material that was dealt with in more detail during the course. Based on the extensive experience of the author team, we strongly believe that this knowledge will be useful for future researchers in biology. It will enable them to perform biochemical measurements and to design and execute their own experiments. The course material will be equally useful for researchers working in academia and for students who wish to pursue a non-academic career in the field of applied biological sciences.

After completing this course, one should be able to deal with some real-life tasks in the biochemical/molecular biological laboratory as follows:

- using biochemical units correctly;
- storing and handling biological samples properly;
- preparing solutions for basic biochemical experiments;
- determining the charge of ionisable molecules depending on pH;
- determining the concentration of proteins and nucleic acids;
- designing expression and purification protocols for proteins;
- characterising interactions between proteins and their ligands;
- measuring the activity of enzymes;
- designing expression plasmids for recombinant protein production;
- using public databases to acquire information about the physical, chemical and biological properties of macromolecules.

The short recapitulation below will help you in revisiting the course material.

#### *Biochemical and molecular biological instruments*

Commonly used plastic and glassware; pipettes; measuring mass (weight) and pH; instruments in the biochemical laboratory practice: water softening systems, centrifuges, spectrophotometers, shakers and incubators; important instruments in the molecular biological practice: PCR apparatus, electroporators, homogenisers, gel documentation apparatus; equipment for sterile work; storage of biological samples.

#### *Biochemical units and working with solutions*

Preparation of commonly used solutions and buffers; concentrations and their conversions (mg/mL, %, molarity); the molecular mass of macromolecules; different ways of expressing quantities in numbers (exponential terms, prefixes, rules of rounding).

#### *Ionisation equilibria*

Acid-base reactions in aqueous solutions; the Henderson-Hasselbalch equation;  $pK_a$ ; weak acids and bases; buffers; conditions influencing the degree of ionisation; estimation of the charge; the  $pI$  value.

#### *Spectrophotometry of biomolecules*

Transmittance and absorbance; the Lambert-Beer equation; the principle of a spectrophotometer; the spectra of ATP,  $NAD^+$ , NADH, some amino acids; spectral characteristics of proteins, nucleotides and polynucleotides; the practice of determining protein and DNA concentration; use of calibration curves for concentration determination.

#### *Isolation of proteins from cells and tissues*

Preparation of biological samples and the lysis of cells for the purification of macromolecules; steps of protein purification; factors influencing enzyme stability; cell lysis; differential and density gradient centrifugation; the

solubility of proteins; the significance of the isoelectric point in protein solubility and stability; fractionation methods: salting-out and salting-in, irreversible precipitation; dialysis; ultrafiltration; lyophilisation.

### *Chromatographic techniques*

Gel filtration, ion exchange, hydrophobic interaction and affinity chromatography; the setup of a chromatographic workstation; separation columns; the concept of stationary and mobile phases; column resolution and capacity; design of purification schemes; FPLC and HPLC.

### *Gel electrophoresis*

Electrophoresis; different gel electrophoresis techniques: native and SDS-polyacrylamide gel electrophoresis (PAGE), isoelectric focusing, denaturing PAGE, 1D and 2D PAGE, agarose gel electrophoresis; detection techniques: staining and blotting; Western-blot; zymography.

### *Protein-ligand interactions*

The dissociation constant ( $K_d$ );  $\Delta G$  and  $K_d$ ; experimental determination of the  $K_d$ ; the Scatchard plot; affinity and specificity; the role of H-bonding, van der Waals and electrostatic interactions in molecular recognition; study of interacting partners; techniques for the determination of binding affinity.

### *Enzyme kinetics*

Physico-chemical background of enzyme action; the substrate saturation curve; determination of initial velocity; the Michaelis-Menten equation; the meaning and determination of  $k_{cat}$ ,  $V_{max}$ ,  $K_M$  and  $k_{cat}/K_M$ .

### *Recombinant DNA technology*

Recombinant DNA and molecular cloning; the polymerase chain reaction (PCR); plasmids as vectors; restriction endonucleases; DNA ligases; introduction of recombinant DNA into cells (transformation, transfection, infection); DNA miniprep and DNA agarose gel electrophoresis.

### *Bioinformatics*

Public databases for DNA and protein sequences and structures (GenBank, UniProt, PDB); sequence analysis (BLAST, ClustalW); use of public databases in the biochemical/molecular biological practice; orthologous and paralogous genes and proteins; three-dimensional structures of macromolecules and their visualisation by molecular graphics.

## **Biochemical calculations and problem solving**

The following problem set will help you in putting your practical biochemical knowledge to a test. We advise you to discuss these problems with your fellow students. If you get stuck, do not hesitate to ask someone who might know the answers better. This way you find yourself in a real lab situation. Scientific research is done in research groups, which normally combine efforts of people with vastly varying degrees of experience. Note that science is done in groups and rarely in isolation. It is a joint enterprise in which questions stimulate people and drive the work forward.

### *1. Design a complete experimental plan that would enable you to characterise the enzymatic activity of a newly isolated protease.*

What reagents are needed? What experiments will need to be done? How would you be able to monitor substrate consumption during the enzymatic reaction? After employing several purification steps on a biological sample in which this protease is highly abundant, you managed to prepare a homogeneous protease solution. The concentration of this protease solution was determined to be 2 mg/mL by using a BSA calibration curve with the Biuret method. Let us assume that the  $K_M$  of the reaction of your enzyme with a substrate is between 0.1 and 1 mM. How many measurement points are minimally needed in order to determine the  $K_M$  with reasonable accuracy? Draw the expected substrate saturation curve. How will you be able to determine the  $K_M$  based on this curve? How much would you put into the test tube from the purified protease sample, from the 100 mM or 10 mM substrate stock solutions, and from the 1 M buffer stock if the total test volume needs to be 2 mL? (Naturally, you need to decide on what

concentrations you will use for the enzyme and the substrate, respectively, in each different test. The concentration of the buffer in the reaction should be 20 mM. Make a table for the test solutions and indicate the amount of the required stock solutions to prepare them.)

*2. Design an experiment that will be suitable for the determination of the affinity of a DNA-binding protein (e.g. a transcription factor) to its DNA binding site.*

What reagents would you need? What type of experiments would you carry out? How would you be able to monitor DNA binding to the protein? Let us assume that the protein-DNA dissociation constant ( $K_d$ ) is between 10 and 100 nM. How many measurement points would be minimally required to determine the  $K_d$  with reasonable accuracy? How would you prepare the required solutions? Draw the expected ligand binding curve. In an additional experiment, you would like to find out how mutations at the DNA binding surface of this protein affect its binding affinity. By using recombinant DNA-based techniques, you produce two different point mutants (m1 and m2) and one double mutant (m3) (m1: T104A, m2: R321A, m3: T104A, R321A). These are expressed and purified, and are available for experiments. Previously you determined the binding affinity of these mutants by the assay discussed above (wild type: 20 nM, m1: 40 nM, m2: 200 nM, m3: 1  $\mu$ M). Draw the  $\Delta G$  values corresponding to the different proteins in an energy diagram. What are the  $\Delta\Delta G$  values caused by the different mutations? Speculate on the reason for the apparently lower binding affinity of the different mutants. What can be the background of the “unusual” binding affinity of m3?

*3. The goal is to produce the human growth hormone in milligram amounts. How would you do it?*

*4. The results of your recent biochemical experiments suggest the existence of an undiscovered enzyme that seems to play an important role in remodelling connective tissue. Your task is to design a research plan to characterise this new protein.*

The goals are the following: characterisation of the enzymatic properties of this novel protein, its identification by sequence, and its evolutionary analysis in case there are other similar known enzymes. You also want to create a possibility for studying this enzyme by X-ray crystallography, and for structure-function studies (e.g. by directed mutagenesis). Outline a reasonable experimental plan that could be executed by a researcher trained in biochemistry and molecular biology. Put this plan together by outlining the different steps in a logical order, constantly bearing in mind the original goals of the project.

*5. You want to determine the substrate saturation curve of an enzyme.*

The substrate, which is a weak base, is used in the 2–30  $\mu$ M range. Its molecular weight is 310 and its  $pK_a$  value is 6.8. Its solubility is 10  $\mu$ g/mL at pH 8.0 where the enzyme has its pH optimum. The reaction volume is 1 mL, and the maximum volume of the added substrate stock may be 30  $\mu$ L. What could be the highest concentration of the substrate stock, and how could this be accomplished?

*6. How would you be able to determine to what degree your plasmid DNA miniprep sample is contaminated by protein?*

*7. By accident, you mixed up two plasmid DNA samples in the lab.*

One of the samples contains the wild-type sequence of protein X, while the other is a mutant version of protein Y. How would you be able to set the situation to rights and determine which sample contains which DNA plasmid?

*8. An inhibitor displays competitive inhibition of an enzyme. How would you be able to determine the binding affinity of this inhibitor to the enzyme? Where does most probably the inhibitor bind?*

The author team of the Department of Biochemistry at Eötvös Loránd University wish you a pleasant and fruitful problem solving experience!

---

# Bibliography

Berg, J.M., Tymoczko, J.L., Stryer, L.: Biochemistry (2012) 7th edition, W. H. Freeman and Company, New York; ISBN-13: 9781429229364

Nelson, D.L., Cox, M.M.: Principles of Biochemistry (2008) 5th edition, W. H. Freeman and Company, New York; ISBN-13: 9780716771081

Voet, D., Voet, J.G., Prat, C.W.: Fundamentals of Biochemistry (2009) 3d edition, John Wiley et Sons, ISBN: 978-0-470-22842-5

Voet, D., Voet, J.G.,: Biochemistry (2011) 4th edition, John Wiley et Sons, ISBN-13: 978-0470570951