1 Biochemical and Molecular Biological Methods in Life Sciences Studies

SAMUEL CLOKIE AND ANDREAS HOFMANN

1.1 FROM BIOCHEMISTRY AND MOLECULAR BIOLOGY TO THE LIFE SCIENCES

Biochemistry is a discipline in the natural sciences that is chiefly concerned with the chemical processes that take place in living organisms. Starting in the 1950s, a new stream evolved from traditional biochemistry, which, until then, mainly investigated bulk behaviour and macroscopic phenomena. This new stream focussed on the molecular basis of biological processes and since it put the biologically important molecules into the spotlight, the term molecular biology was coined. Importantly, molecular biology goes beyond the mere characterisation of molecules. It includes the study of interactions between biologically relevant molecules with the clear goal to reveal insights into functions and processes, such as replication, transcription and translation of genetic material.

Major technological and methodological advances made during the 1980s enabled the development and establishment of several specialised areas in the life sciences. These include structural biology (in particular the determination of three-dimensional structures), genetics (for example DNA sequencing) and proteomics. The refinement and improvement of methodologies, as well as the development of more efficient software (in line with more powerful computing resources), contributed substantially to specialist techniques becoming more accessible to researchers in neighbouring disciplines. What once had been the task of a highly specialised scientist who had been extensively trained in that particular area has consistently been transformed into a routinely applied methodology. These tendencies have pushed the feasibility of cross-disciplinary studies into an entirely new realm, and in many contemporary laboratories and research groups, methods originally at home in different basic disciplines are frequently used next to each other.

It is thus not surprising, that in the past 15–20 years, the term 'life sciences' has been used to describe the general nature of studies and research areas of a scientist working on studies related to living organisms.

1.2 THE EDUCATION OF LIFE SCIENTISTS

The life sciences embrace different fields of the natural and health sciences, all of which involve the study of living organisms, from microorganisms, plants and animals to human beings. Importantly, even satellite areas that are methodologically rooted outside natural or health sciences, for example bioethics, have become a part of the life sciences. And with neuroscience and artificial intelligence being two major current areas of interest, one might expect further subjects to be included under the life sciences umbrella.

The growing complexity of scientific studies requires ever-increasing cross-disciplinarity when it comes to particular methodologies utilised in the quest to reach the goals of these studies. This poses entirely new problems when it comes to the education of future scientists and researchers. At the core, this requires an appreciation or even acquisition of mindsets from other disciplines, as opposed to a mere concatenation of methods specific to one discipline. Therefore, the achievement of true cross-disciplinarity poses particular challenges since, according to Simon Penny, it requires deep professional humility, intellectual rigour and courage. At the same time, due to time and practical constraints, educational programmes that teach life science studies often focus on a select group of individual disciplines, rather than attempting to include every subject that could be included under the global term 'life sciences'.

In order to successfully embark on a contemporary life science study or contribute to large cross-disciplinary teams, scientists also need to possess knowledge and skills in a diverse range of areas. For example, in order to screen a small-molecule compound library against a target protein, the chemist needs to understand the nature and behaviour of proteins. Vice versa, if a cell biologist wants to screen a small-molecule library to identify novel effectors for a pathway of interest, they need to deal with the logistics and characteristics of small molecules.

One relatively new aspect that the life scientist is faced with is the large amount of data generated by improved instrumentation and methodologies. The term 'bioinformatics' has been coined that loosely describes the processing of biological data. It is now considered a stand-alone discipline that includes methods on processing 'big data' that are commonplace in the field of genomics. Such data-processing techniques are ubiquitous in the life sciences, and bioinformatics serves as an example that spans almost all the life science disciplines.

Life scientists work in many diverse areas, including hospitals, academic teaching and research, drug discovery and development, agriculture, food institutes, general education, cosmetics and forensics. Aside from their specialist knowledge of the interfacing of core disciplines, they also require a solid basis of transferable skills, such as analytical and problem-solving capabilities, and written and verbal communication, as well as planning, research, observation and numerical skills.

1.3 AIMS OF LIFE SCIENCE STUDIES

Studies in the life sciences ultimately aspire to an advanced understanding of the nature of life in molecular and mechanistic terms. **Biochemistry** still constitutes the core discipline of the experimental life sciences, and involves the study of the chemical processes that occur in living organisms. Such studies rely on the application of appropriate techniques to advance our understanding of the nature, and relationships between, biological molecules, especially proteins and nucleic acids in the context of cellular function.

The huge advances made in the past 10–15 years, with the **Human Genome Project** being a particular milestone, have stimulated major developments in our understanding of many diseases and led to identification of strategies that might be used to combat these diseases. Such progress was accompanied – and enabled – by substantial developments in technologies, data acquisition and data mining. For example, the genome of any living organism includes coding regions that are transcribed into messenger RNAs (mRNAs), which are subsequently translated into proteins. In vitro, mRNAs are reverse-transcribed (Section 4.10.4), resulting in stable complementary DNA, which is traditionally sequenced using a DNA polymerase, an oligonucleotide primer and four deoxyribonucleotide triphosphates (dNTPs) to synthesise the complementary strand to the template sequence (see Section 20.2.1). The development of high-throughput sequencing ('next-generation sequencing') technologies that can produce millions or billions of sequences concurrently, has made the sequencing of entire genomes orders of magnitude faster and, at the same time, less expensive. This particular development has made it possible to sequence many thousands of human genomes, making it possible to truly understand **population genetics**. Such information can be used to aid and improve genetic diagnosis of common and rare human diseases.

The combination of molecular biology and genomics applied to the benefit of humankind can be best illustrated by the invention and recent improvement of genome editing techniques, such as the CRISPR/Cas method (see Section 4.17.6). The potential impact on **health economics** is substantial; individually unpleasant and costly conditions, genetically or environmentally acquired, can be addressed and potentially be eradicated.

Similar developments have occurred in many other disciplines. In structural biology, robotics, especially at synchrotron facilities, have drastically reduced the time required for handling individual samples. Plate readers that perform particular spectroscopic applications (see Chapter 13) in a multi-well format are now ubiquitously present in laboratories and enable medium and **high throughput** for many standard assays (see also Section 24.5.3).

All these developments are accompanied by a massive increase in (digital) data generated, which opens an avenue for entirely new types of studies that are more or less exclusively concerned with data mining and analysis.

This text aims to cover the principles and methodologies underpinning life science studies and thus address the requirements of today's students and scientists who operate in a much broader area than the typical biochemists one or two decades ago.

The contents of this text are therefore structured around six different disciplines or methodologies, all of which play pivotal roles in current research:

- Basic Principles (Chapter 2)
- Biochemistry and Molecular Biology (Chapters 3–10)
- Biophysics (Chapters 11–15)
- Information Technology (Chapters 16–19)
- 'omics Methods (Chapters 20–22)
- Chemical Biology (Chapters 23–24)

The Basic Principles chapter introduces some important general concepts surrounding biologically relevant molecules, as well as their handling in aqueous solution. It further highlights fundamental considerations when designing and conducting experimental research.

Information technology has become an integral part of scientific research. The ability to handle, analyse and visualise data has always been a core skill of science and gained even more importance with the advent of 'big data' on the one hand, and the fact that data acquisition, processing and communication is done entirely in digital format on the other. Furthermore, in the areas of bio- and chemoinformatics, standardisation of data formats has resulted in the availability of unprecedented volumes of information in databases that require an appropriate understanding in order to fully utilise the data resource.

Among the core methodologies of biochemistry and molecular biology are techniques to culture living cells and microorganisms, and the preparation and handling of DNA, as well as the production and purification of proteins. Such experimental work is frequently accompanied by analytical or preparative gel electrophoresis, the use of antibodies (immunochemistry) or radio isotopes. In the medical sciences, a diagnostic test requires the application of biochemical or molecular techniques in a regulated environment, comprising the area of clinical biochemistry (discussed in Chapter 10).

The characterisation of cells and molecules, as well as their interactions and processes, involves an array of biophysical techniques. Such techniques apply gravitational (centrifugation) or electrical forces (mass spectrometry), as well as interactions with light over a broad range of energy.

The neologism 'omics is frequently being used when referring to methodologies that characterise large pools of biologically relevant molecules, such as DNA (genome), mRNA (transcriptome), proteins (proteome) or metabolic products (metabolome). The first application of these methodologies were mainly concerned with the acquisition and collection of large datasets specific to one experiment or biological question. However, several areas of the life sciences now leave behind the phase of pure observation and are increasingly applied to study the dynamics of an organism, a so-called systems biology approach.

A hallmark of chemical biology is the use of small molecules, either purified or synthetically derived natural products or purpose-designed chemicals, to study the modulation of biological systems. The use of small-molecule compounds can be either

exploratory in nature (probes) or geared towards therapeutic use where the desired activities of the compounds are to either activate or inhibit a target protein, which is typically, but not necessarily, an enzyme. Due to the specific role of a target protein in a given cellular pathway, the molecular interaction will ideally lead to modulation of processes involved in pathogenic situations. Even if a small molecule has a great number of side effects, it can be used as a probe and be useful to delineate molecular pathways in vitro.

Given the breadth of topics, methodologies and applications, selections as to the individual contents presented in this text had to be made. The topics selected for this text have been carefully chosen to provide undergraduate students and non-specialist researchers with a solid overview of what we feel are the most relevant and fundamental techniques.

Methods and techniques form the tool set of an experimental scientist. They are applied in the context of studies which, in the life sciences, address questions of the following nature:

- the structure and function of the total protein component of the cell (proteomics) and of all the small molecules in the cell (metabolomics)
- the mechanisms involved in the control of gene such expression
- the identification of genes associated with a wide range of diseases and the development of gene therapy strategies for the treatment of diseases
- the characterisation of the large number of 'orphan' receptors, whose physiological role and natural agonist are currently unknown, present in the host and pathogen genomes and their exploitation for the development of new therapeutic agents
- the identification of novel disease-specific markers for the improvement of clinical diagnosis
- the engineering of cells, especially stem cells, to treat human diseases
- the understanding of the functioning of the immune system in order to develop strategies for protection against invading pathogens
- the development of our knowledge of the molecular biology of plants in order to engineer crop improvements, pathogen resistance and stress tolerance
- the discovery of novel therapeutics (drugs and vaccines) to the nature and treatment of bacterial, fungal and viral diseases.

1.4 PERSONAL QUALITIES AND SCIENTIFIC CONDUCT

The type of tasks in scientific research and the often long-term goals pursued by science require, very much like any other profession, particular attributes of people working in this area:

- Quite obviously, a substantial level of intelligence is required to grasp the scientific concepts in the area of study. The ability to think with clarity and logically, and to transform particular observations (low level of abstraction) into concepts (high level of abstraction) is also required, as is a solid knowledge of the basic mathematical syllabus for any area of the life sciences.

- For pursuing longer-term goals it is also necessary to possess stamina and persistence. Hurdles and problems need to be overcome, and failures need to be coped with. Often, experimental series can become repetitive and it is important to not fall into the trap of boredom (and then become negligent).
- A frequently underestimated quality is attention to detail. Science and research is about getting things right. At the time when findings from research projects are written up for publication, or knowledge is summarised for text books such as this one, the readership and the public expect that all details are correct. Of course, mistakes can and do happen, but they should not happen commonly, and practices need to be in place that prevent mistakes being carried over to the next step. Critical self-appraisal and constant attention to detail is probably the most important element in this process.
- Communication skills and the ability to describe and visualise fairly specialist concepts to non-experts is of great importance as well. The best set of data is not put to good use if it is not presented to the right forum in the right fashion. Likewise, any set of knowledge acquired cannot be taught effectively to others without such skills.
- Lastly, curiosity and the willingness to explore are a requirement if an independent career in the life sciences is being sought. Just having excellent marks in science subjects does not automatically make a scientist if one needs to be told every single next step through a research project.

Since science does not happen in an isolated situation but a community (the scientific community on the one hand, but also society as a whole on the other), norms need to be put in place that define and guide what is acceptable and unacceptable behaviour. Beset by the occasional fraudulent study and scientific misconduct case, more attention has been paid to ethical conduct in science in the recent decade (see also Section 2.8.2).

Despite ethical frameworks and policies put in place to varying degrees, ultimately, the responsibility of ethical conduct rests with the individual. And while many of the ethical norms are geared towards the interactions of scientists when it comes to specific scientific tasks or procedures, there are certainly elements that apply to any (scientific or non-scientific) situation:

- Critically assess potential conflicts of interest. In a surprisingly large number of situations, any individual might find themselves playing multiple roles, and the objectives of each of the roles may be in conflict with each other. Where conflicts of interest arise, the individual should withdraw from the decision-making process.
- Respect confidentiality and privacy. This not only applies to ongoing research work, results, etc., but also to conversations and advice. When approached for advice or with a personal conversation, most parties expect this to be treated in confidence and not made available to the public.
- Follow informed consent rules and discuss intellectual property frankly. In order to avoid disagreements about who should get credit and for which aspects, talk about these issues at the beginning of a working relationship.
- Engage in a sharing culture. Resources, methodologies, knowledge and skills that have been established should be shared where reasonably possible. This fosters positive

interactions, contributes to transparency and frequently leads to new collaborations, all of which advance science.

- Lead by example. Not just in a teaching situation, but in most other settings, too, general principles and cultural norms are only effectively imprinted into the environment if the rules and guidelines are lived. In many instances, rules and protocols are 'implemented' by institutions and exist mainly as a ticking-the-box exercise. Such policies do not address the real point as they should and are largely ineffective.

1.5 SUGGESTIONS FOR FURTHER READING

1.5.1 Experimental Protocols

Holmes D., Moody P. and Dine D. (2010) *Research Methods for the Biosciences, 2nd Edn.*, Oxford University Press, Oxford, UK.

1.5.2 General Texts

Smith D. (2003) Five principles for research ethics. *Monitor on Psychology* **34**, 56.

1.5.3 Review Articles

Duke C.S. and Porter J.H. (2013) The ethics of data sharing and reuse in biology. *BioScience* **63**, 483–489.

Puniewska M. (2014) Scientists have a sharing problem. *The Atlantic*, 15 Dec 2014, www.theatlantic.com/health/archive/2014/12/scientists-have-a-sharing-problem/383061/ (accessed April 2017).

Taylor P.L. (2007) Research sharing, ethics and public benefit. *Nature Biotechnology* **25**, 398–401.

1.5.4 Websites

Mike Brotherton's Blog: Five qualities required to be a scientist
www.mikebrotherton.com/2007/11/05/five-qualities-required-to-be-a-scientist/ (accessed April 2017)

Simon Penny: Rigorous Interdisciplinary Pedagogy
simonpenny.net/texts/rip.html (accessed April 2017)

Andy Polaine's Blog: Interdisciplinarity vs Cross-Disciplinarity
www.polaine.com/2010/06/interdisciplinarity-vs-cross-disciplinarity/ (accessed April 2017)

2 Basic Principles

PARISA AMANI AND ANDREAS HOFMANN

2.1 BIOLOGICALLY IMPORTANT MOLECULES

Molecules of biological interest can be classified into ions, small molecules and macromolecules. Typical organic small molecules include the ligands of enzymes, substrates such as adenosine triphosphate (ATP) and effector molecules (inhibitors, drugs). Ions such as Ca^{2+} play a key role in signalling events. Biological macromolecules are polymers which, by definition, consist of covalently linked monomers, the building blocks. The four types of biologically relevant polymers are summarised in Table 2.1.

2.1.1 Proteins

Proteins are formed by a condensation reaction of the α-amino group of one amino acid (or the imino group of proline) with the α-carboxyl group of another. Concomitantly, a water molecule is lost and a peptide bond is formed. The peptide bond possesses partial double-bond character and thus restricts rotation around the C–N bond. The progressive condensation of many amino acids gives rise to an unbranched polypeptide chain. Since biosynthesis of proteins proceeds from the N- to the C-terminal amino acid, the N-terminal amino acid is taken as the beginning of the chain and the C-terminal amino acid as the end. Generally, chains of amino acids containing fewer than 50 residues are referred to as peptides, and those with more than 50 are referred to as proteins. Most proteins contain many hundreds of amino acids; ribonuclease, for example, is considered an extremely small protein with only 103 amino-acid residues. Many biologically active peptides contain 20 or fewer amino

Table 2.1 **The four types of biologically relevant polymers**

Polymer	Monomers	Monomer details
Ribonucleic acid (RNA)	4 Bases	Adenine (A), uracil (U), cytosine (C), guanine (G)
Deoxyribonucleic acid (DNA)	4 Bases	Adenine (A), thymine (T), cytosine (C), guanine (G)
Protein	20 Amino acids	Ala, Cys, Asp, Glu, Phe, Gly, His, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Ser, Thr, Val, Trp, Tyr
Polysaccharide	Monosaccharides	Trioses, tetroses, pentoses, hexoses, heptoses

acids, such as the mammalian hormone oxytocin (nine amino-acid residues) which is clinically used to induce labour since it causes contraction of the uterus, and the neurotoxin apamin (18 amino-acid residues) found in bee venom.

2.2 THE IMPORTANCE OF STRUCTURE

Three main factors determine the three-dimensional structure of a macromolecule:

- allowable backbone angles
- interactions between the monomeric building blocks
- interactions between solvent and macromolecule.

The solvent interactions can be categorised into two types: binding of solvent molecules (solvation) and hydrophobic interactions. The latter arise from the inability or reluctance of parts of the macromolecule to interact with solvent molecules (hydrophobic effect), which, as a consequence, leads to exclusive solvent–solvent interactions. Phenomenologically, a collection of molecules that cannot be solvated will stick close to one another and minimise solvent contact.

The interactions between the building blocks of the macromolecule comprise negative interactions (by avoiding atomic clashes) and positive interactions, which may be provided by hydrogen bonds, electrostatic interactions and van der Waals interactions (see also Section 17.4). Hydrogen bonds are the main constitutive force of backbone interactions in proteins, but can also be observed between residue side chains. Electrostatic interactions in proteins occur between residue side chains that possess opposite charges (arginine, lysine and aspartate, glutamate). The van der Waals attraction is a weak short-range force that occurs between all molecules. It becomes particularly important if two molecules possess highly complementary shapes. Thus, van der Waals interactions are responsible for producing complementary surfaces in appropriate regions of macromolecules.

The allowable backbone angles provide a framework of geometric constraints and balance attractive interactions and geometrical/steric tension within the macromolecule.

2.2.1 Conformation

The structural arrangement of groups of atoms is called conformation (see also Section 17.1.1) and the **conformational isomerism** of molecules describes isomers that can be inter converted exclusively by rotations about formally single bonds. The rotation about a single bond is restricted by a rotational energy barrier that must be overcome; the individual isomers are called **rotamers**. Conformational isomerism arises when the energy barrier is small enough for the interconversion to occur. The angle describing the rotation around a bond between two atoms is called the **dihedral** (or torsion) angle.

The protein **backbone** (Figure 2.1) is geometrically defined by three dihedral angles, namely Φ (N-Cα), Ψ (Cα-C) and ω (C-N); the last angle can take only two values, 0° and 180°, due to the partial double-bond character of the peptide bond. The conformation of the protein backbone is therefore determined by the Φ and Ψ **torsion angles**. The values these angles assume determine which type of secondary structure (see below) a certain consecutive region in the protein will adopt.

Many organic small molecules possess cyclic aromatic structures and are thus planar. However, there are also many non-aromatic cyclic structures, in particular **carbohydrates** (**sugars**), which are of great importance in biochemical processes. Notably, many sugars exist in aqueous solution as both open-chain and cyclic forms. Figure 2.2 illustrates this using the example of *D*-**glucose**. It is obvious from the open-chain form, that there are a number of stereogenic centres in *D*-glucose. If the positions of the hydroxyl groups are changed to the opposite enantiomer for each stereogenic carbon, the resulting molecule is *L*-glucose. Upon ring closure of the **open-chain** form, atoms or groups bonded to tetrahedral ring carbons are either pointing up or down, as indicated by the use of dashed or solid wedges when drawing the two-dimensional structures. If two neighbouring hetero-atom substituents (e.g. hydroxyl groups) on the ring are both pointing in the same direction, this conformation is called ***cis***; if they point in opposite directions, they are said to be in the ***trans*** conformation. The open-chain form is characterised by an aldehyde function which, upon ring closure, is converted to a **hemi-acetal** (comprising $R_1C(OR_2)(OH)H$; see the carbon

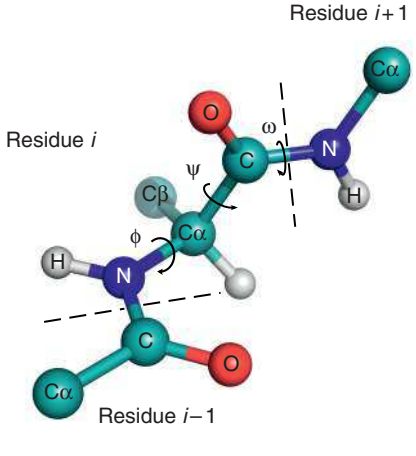

Figure 2.1 Polypeptide chain comprising three amino acids (numbered *i*−1, *i*, *i*+1). The limits of a single residue (*i*) are indicated by the dashed lines. The torsion angles Φ, Ψ and ω describe the bond rotations around the N-Cα, Cα-C and C-N bonds, respectively.