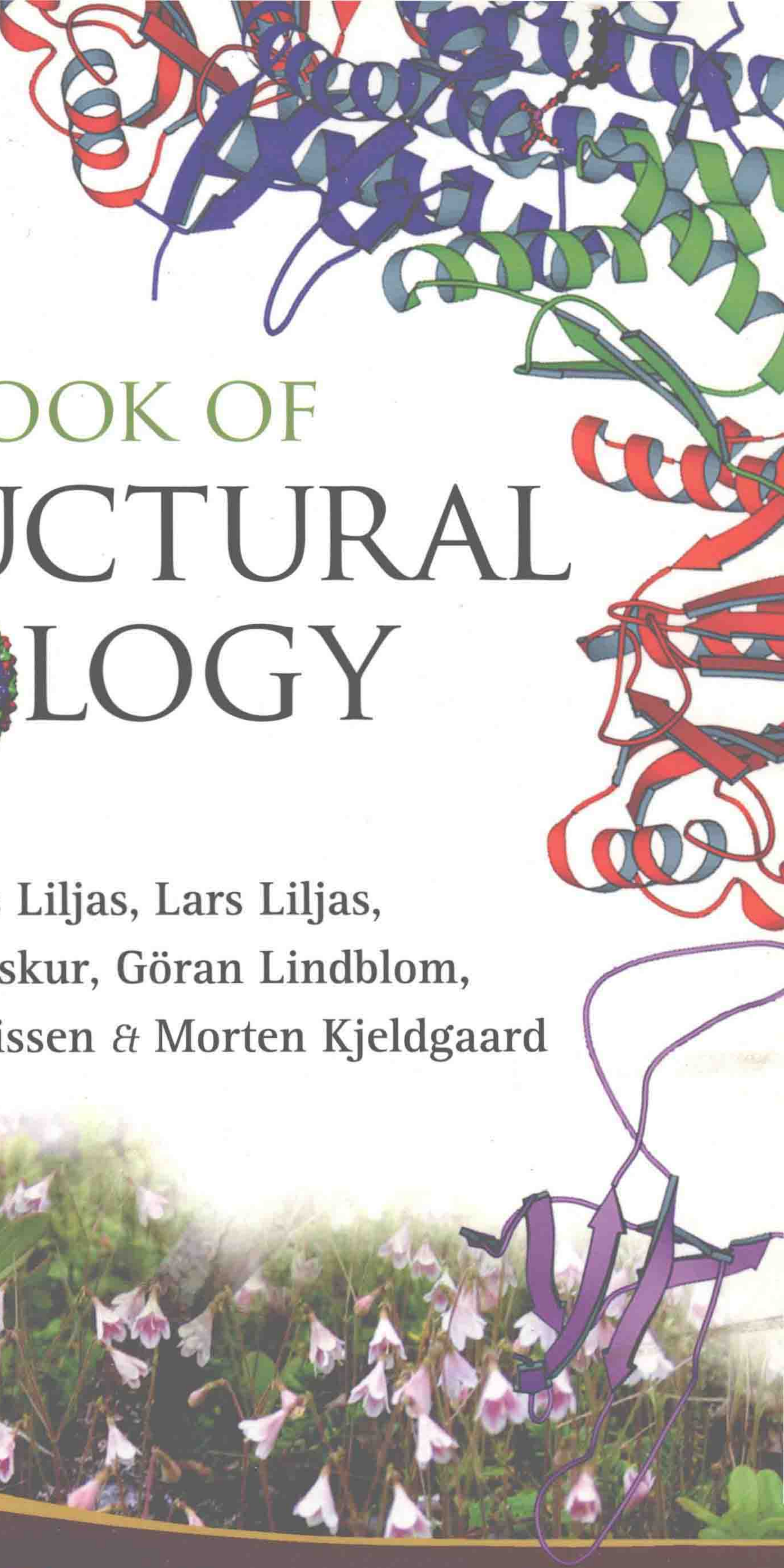

TEXTBOOK OF STRUCTURAL BIOLOGY

Anders Liljas, Lars Liljas,
Jure Piskur, Göran Lindblom,
Poul Nissen & Morten Kjeldgaard

World Scientific

TEXTBOOK OF STRUCTURAL BIOLOGY

Anders Liljas *(Lund University, Sweden)*

Lars Liljas *(Uppsala University, Sweden)*

Jure Piskur *(Lund University, Sweden)*

Göran Lindblom *(Umeå University, Sweden)*

Poul Nissen *(University of Aarhus, Denmark)*

Morten Kjeldgaard *(University of Aarhus, Denmark)*

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data
Textbook of structural biology / Anders Liljas ... [et al.].
p. ; cm.
Includes bibliographical references and index.
ISBN-13: 978-981-277-207-7 (hardcover : alk. paper)
ISBN-10: 981-277-207-3 (hardcover : alk. paper)
ISBN-13: 978-981-277-208-4 (pbk. : alk. paper)
ISBN-10: 981-277-208-1 (pbk. : alk. paper)
1. Proteins--Structure. 2. Ultrastructure (Biology) 3. Protein folding. 4. Biomolecules--Structure. I. Liljas, Anders.
[DNLM: 1. Proteins. 2. Cell Communication. 3. Cells--cytology. 4. Protein Folding. QU 55 T355 2008]
QP551.T49 2008
572'.633--dc22

2008035573

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore by Mainland Press Pte Ltd

TEXTBOOK OF STRUCTURAL BIOLOGY

PREFACE

The first textbook in structural biology may have been the one produced by Dickerson and Geis in 1969 entitled, *The Structure and Action of Proteins*. It was an excellent book covering most of what was known about protein structure. At that time, the first protein structure, myoglobin, had been known for ten years and only a few additional structures had been determined. The structure of DNA was derived from fiber diffraction data in 1953, but very little additional knowledge about nucleic acid structure was available. The exciting research of the initial phase is well covered by RE Dickerson in his recent book, "Present at the Flood. How Structural Molecular Biology came about." Now, almost forty years and numerous textbooks later the situation is very different. The very useful textbooks by Branden and Tooze, *Introduction to Protein Structure* (1991 and 1999) could naturally not cover the exciting developments that have taken place during the last ten years. The amount of structural information available from the Protein Data Bank (PDB), is now very extensive and is growing rapidly. The rate of the work has benefited very significantly from very dramatic technical developments, a key one being the use of synchrotrons as radiation sources. The scientific research in the field has since long ago not only been an academic effort, but very extensive work has also been extensively carried out in the pharmaceutical and biotechnological industries. In addition, the development of structural genomics laboratories has refined the laboratory protocols to the benefit of everybody and the rate of structure determination is steadily increasing.

In this flood of information a selection has to be made. Our selection has been made on the basis of our teaching experiences of what students would most benefit from. The central knowledge about protein, nucleic acid and membrane structure is described in relation to central systems in biology. Our research experience covers a broad range of structural biology and our personal research foci have had some influence on the selections. Moreover, since recent research also shows that membrane lipids play an important, active role in cell function, like in the regulation of membrane enzymes and channels, and formation of domains like rafts, a thorough chapter on the physical properties and functions of lipids has been included in this textbook.

In producing a textbook on structural biology, we have depended on the kindness of many colleagues, in reviewing parts of the text, providing expert opinions and permitting us to use their published or sometimes unpublished material. We cannot list them all without the risk of forgetting some. However, we are very grateful for all assistance. We are exceedingly grateful to Terese Bergfors for her very careful reading of the manuscripts and correction of both facts and language. Most illustrations of the book are produced using the programs Molscript (http://www.avatar.se/molscript/molscript.html, Kraulis, PJ (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* **24**, 946–950, Molray (Harris M, Jones TA (2001). Molray — a web interface between O and the POV-Ray ray tracer. *Acta Cryst* **D57**: 1201–1203) and PyMOL (http://pymol.sourceforge.net/). We are grateful for being given permission to use these programs.

CONTENTS

1

INTRODUCTION

1.1 Life

We are surrounded by microbes, plants and animals that we can immediately recognize as living things (Fig. 1.1). However, it is still difficult to provide a concise definition of what life is. Perhaps the most useful definition for the purpose of our book is that life is a unit capable of chemical activities, and which can reproduce and evolve.

Chemical activities, which involve conversion of energy and matter, are termed *metabolism*. These activities capture energy and chemical matter in different forms. Thousands of chemical activities can take place simultaneously in a living organism and they must be well coordinated to maintain the stability of the living unit.

Reproduction of the unit (generating new units) provides both the continuity and the variation that is also an important characteristic of life. The combination of reproduction, horizontal transfer of information and "erroneous" duplicates provides the basis of evolution. In other words, the composition of the unit should be able to change over time to better adapt to the changing environmental conditions. The living organisms appear in very different forms and follow very different lifestyles. However, the basic characteristics of life, including metabolism, reproduction and evolution are provided and governed by very similar sub-structures, biological macromolecules and cells.

1.2 Levels of Organization of Life

The living world has several hierarchical levels, ordered from the smallest to the largest. At the bottom are molecules, a mix of inorganic and organic compounds

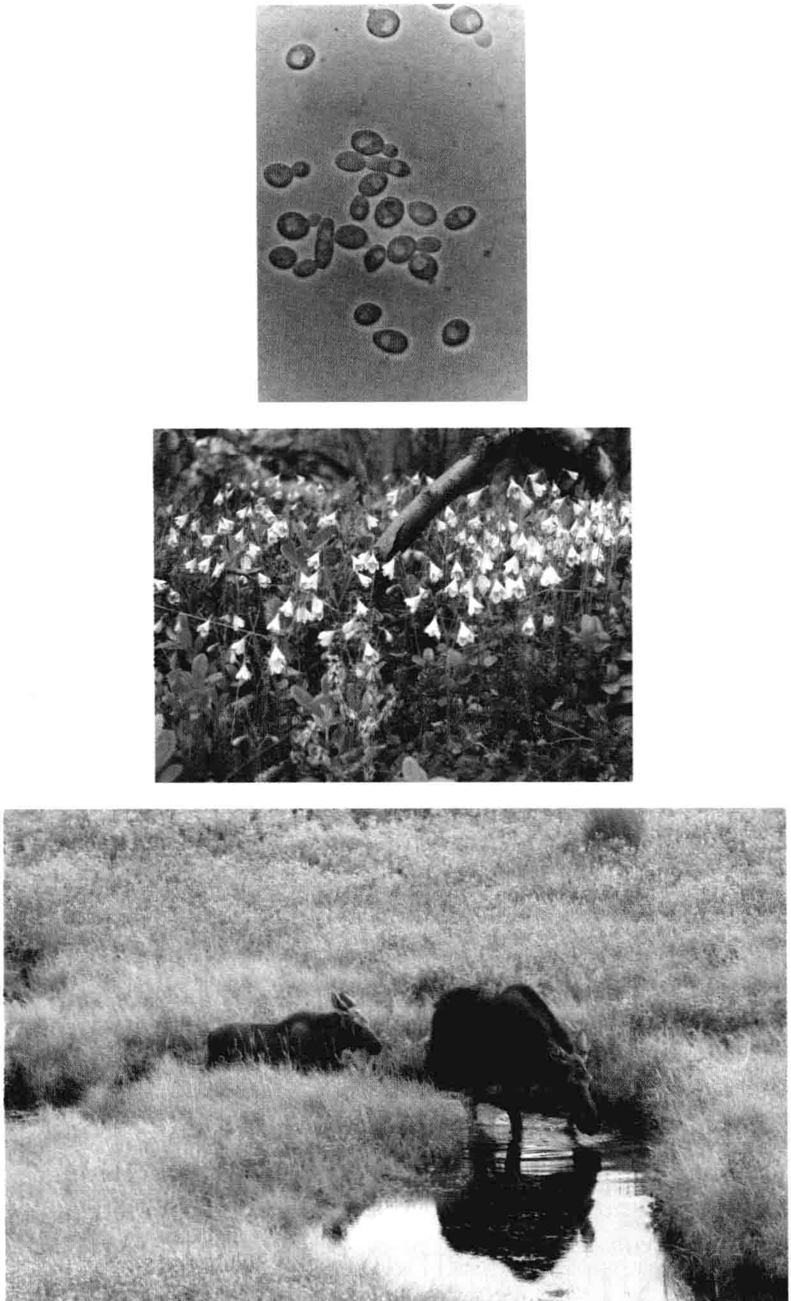

FIGURE 1.1 ■ Living organisms are found in numerous different forms. *Top*: A microscope picture of baker's yeast cells (courtesy of Concetta Compagno); *middle*: linneas (*Linnea borealis*) covering vast areas of Lapland (courtesy of Bernarda Rotar); and *bottom*: a moose, the largest land animal in Scandinavia (courtesy of Aca.Pixus.dk). Within these different macro-forms, very similar molecular structures can be found, which determine the form and life style of the carrier organisms.

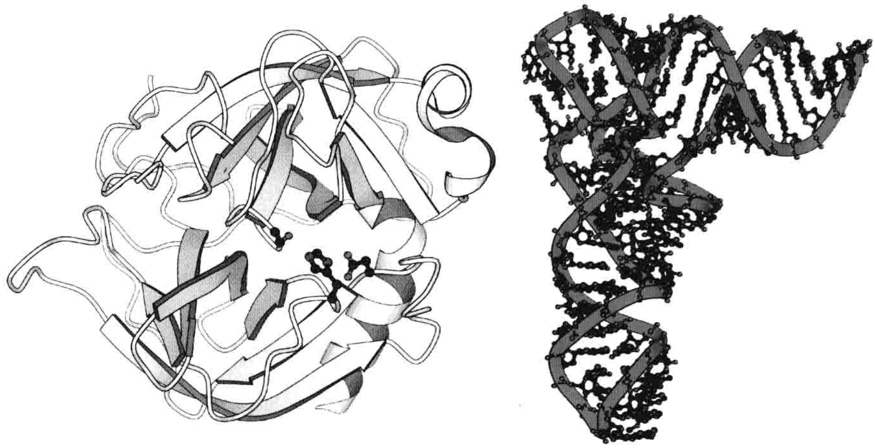

FIGURE 1.2 ■ A simplified picture of two bio-macromolecules, which are the focus of our further chapters. *Left*: The structure of a well-known protein (chymotrypsin, PDB: 4CHA). *Right*: A nucleic acid molecule (yeast tRNAPhe, PDB: 1EHZ).

and macromolecules, followed by sub-cellular structures, cells, tissues, organs, organisms, populations, communities and the biosphere, which encompass all biological communities on the Earth.

Macromolecules are central molecules of all living organisms. They are giant polymers consisting of monomeric units (Fig. 1.2), which may or may not be identical, and are connected by covalent or non-covalent bonds. Macromolecules perform several functions, which are the basis of metabolism, reproduction and evolution, such as energy or information storage, reaction catalysis, coordination and regulation, communication, structural support, defense, movement and transport. Macromolecules, on the basis of chemical composition, can be categorized into four different types: (i) peptides and proteins which are polymers of amino acid residues; (ii) nucleic acids, which are polymers of nucleotides; (iii) carbohydrates, which are polymers of sugars; and (iv) membranes which are structural aggregates of lipids. In the following chapters we will discuss the structures of bio-macromolecules and link them to their functions and the higher levels of the living world.

The basic unit of life is a *cell*. Cells are surrounded by a plasma membrane, which separates each cell from the external environment and creates a segregated compartment with relatively constant internal environment. The cells show two organizational patterns: (i) prokaryotic characteristic for bacteria and archaea, and (ii) eukaryotic characteristic for eukarya. Prokaryotic cells are usually smaller than eukaryotic cells, on average 1 μm, and they usually exist as single cells. All prokaryotic cells have the basic structure with a plasma membrane, an intracellular nucleoid containing DNA and the rest of the intracellular material called cytosol, where ribosomes and many enzymes and cytoskeletal elements can be found. Eukaryotic cells are usually at least 10 times larger than prokaryotic cells and more complex. In eukaryotic cells, the basic prokaryotic cell structure with the plasma membrane and cytoplasm is upgraded with compartments, also called organelles. In the cytoplasm, additional distinctive structures can be found. In some cases their interior is segregated from each other by a membrane. The most common organelles are: (i) the nucleus storing the cell genetic material and where replication and gene expression takes place; (ii) the cytosol, where protein synthesis and many essential biochemical reactions take place; (iii) the mitochondrion, a power plant and energy storage compartment; (iv) the endoplasmatic reticulum and Golgi apparatus, where proteins are packaged and sent to further locations; (v) the lysosomes or vacuoles, where polymeric macromolecules, such as proteins, are degraded into usable monomers (Fig. 1.3).

It is believed that all organisms on Earth originate from a single kind of unicellular organism. Today many millions of different kinds of organisms that do not interbreed with one another can be found and we call them species. They are all successfully adapted to their different environments and in this sense, perfect. However, some of them may not be perfect tomorrow and can thereby become

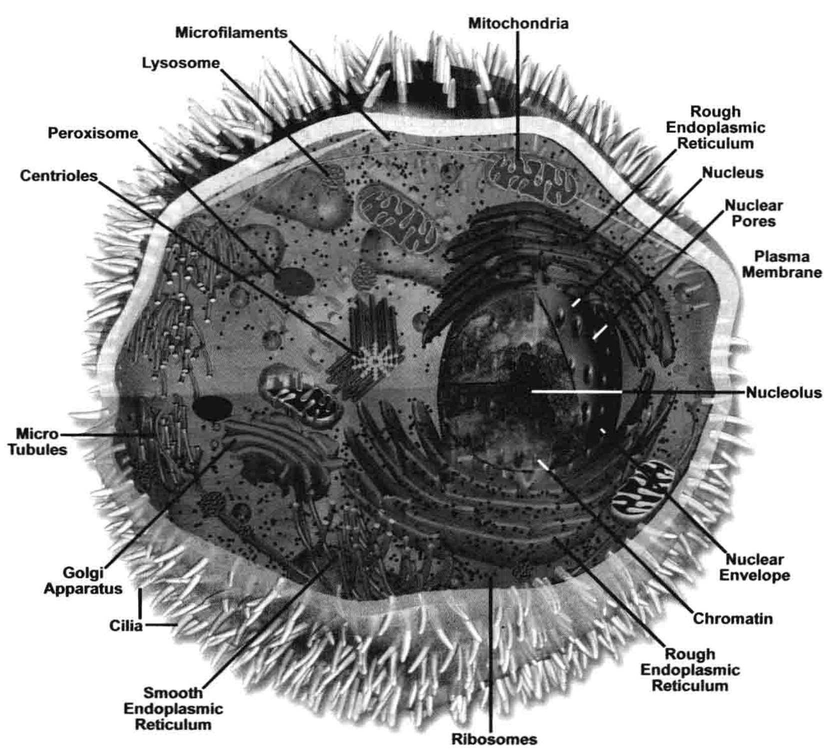

FIGURE 1.3 ■ A schematic picture of an animal cell showing sub-cellular structures, such as nucleus, membrane systems (ER), mitochondrion, etc. (Made by Michael W. Davidson, Florida State University.)

extinct, like so many other species that have previously populated the Earth. In short, diversity among organisms is a very dynamic process and changes tremendously with time and space.

The unfolding of events leading to the present diversity can be expressed as an evolutionary tree showing the order in which a species split and evolved into new species. This tree traces the descendants from ancestors that lived at different times in the past. In other words, the evolutionary tree shows the evolutionary relationship among modern and ancient species. It is important to understand the evolutionary relationship between organisms when one compares the structure of macromolecules originating from these organisms and being involved in similar processes. Some of the earlier branching is difficult to reconstruct because there are no available fossils. However, based on the molecular evidence in the modern organisms we can separate all living organisms into three domains, which have been evolving separately for more than one billion years: (i) Archaea, (ii) Bacteria, and (iii) Eukarya (Fig. 1.4). Even if, superficially, they looked similar, archaea and bacteria separated into distinct lineages very early during the evolutionary history.

Archaea were discovered as inhabitants of extreme environments, such as hot and acidic springs, sea depths, salt brines, but can also be found in more "normal" environments. Their replication, transcription and translation machinery

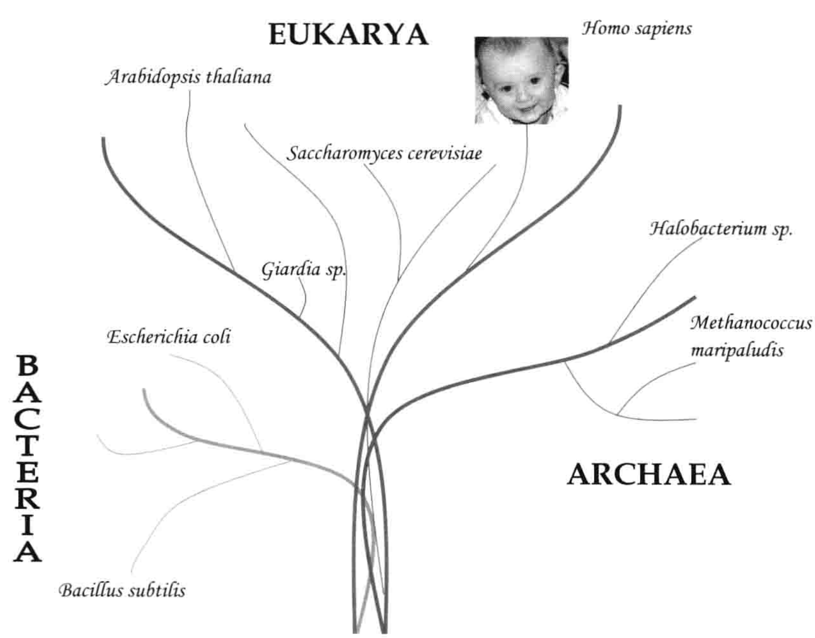

FIGURE 1.4 ■ A simplified tree of life. The common progenitor originated appr. four billion years ago. The position of the first branchings, occurring between the progenitor of bacteria, archaea and eukarya, are still unclear.

resembles the eukaryote machinery, while the Archaea metabolism and energy conversion resemble those of the bacterial ones.

Bacteria consist of more than a dozen sub-groups, also called clades, but the most important are: Protobacteria, Cyanobacteria, Spirochetes, Chlamydias and Firmicutes. The Protobacteria form the largest group. This is a very diverse group which includes one of the best-studied organisms, *Escherichia coli*. Sometimes bacteria are divided, on the basis of their cell wall composition, into Gram-positive, including *Bacillus subtilis*, and Gram-negative, including *E. coli*. Bacteria exhibit the greatest biochemical diversity.

Eukarya can be divided into four groups: Protista, Plantae, Fungi and Animalia. The Protista contain mostly single-celled organisms and have a polyphyletic origin, meaning that some represent very primitive eukaryotes, such as *Giardia*, while others are closely related to animals, such as *Dictyostelium*, or plants, such as red algae. Phagotrophy, a feeding mode to form a pocket in the plasma membrane and enclose the "food," is a hallmark of Eukarya.

1.3 A Short History of Life on Earth

The theory of chemical evolution holds that conditions on the primitive Earth, at least four billion years ago, led to the emergence of the first biological molecules. Oparin and Haldane independently suggested in the 1920s, that if the Earth's first atmosphere was reducing, and if there was a supply of external energy, then a range of organic compounds might be synthesized. In the 1950s Stanley Miller and Harold Urey mimicked these conditions in the lab. Water vapor, hydrogen gas, ammonia and methane gas were exposed to sparks, and after a few days the system was observed to contain several complex molecules, such as amino acids and nucleic acid bases, the building blocks of today's life. When the monomeric units were present, it was not so difficult to achieve polymerization even under abiotic conditions, but how could the first peptides and nucleic acids become "alive"? In other words, how could they start to reproduce and evolve?

The term *replicator* means a structure that can arise only if there is a preexisting structure of the same kind in the vicinity. For example, a supersaturated solution crystallizes if a small seeding crystal is added. However, this represents a simple replicator relying on a single structure. More sophisticated replicators should be able to exist in several forms and thereby exhibit a simple variation in heredity. For sustained evolution an indefinite number of forms and indefinite variation in heredity are necessary. The first artificial replicator, a simple hexadeoxynucleotide not needing enzymes for its replication (polymerization from the mono-units present) was synthesized by von Kiedrowski in 1986. Now it is believed that the first short RNA molecules had the ability to catalyze the polymerization of offspring molecules. The first living RNA molecules competed successfully with their own erroneous copies and with other less-efficient systems for the monomers needed for their replication. Even if self-replicating RNA molecules fulfill the above criteria for life, the path to the first cells was still more sophisticated. One of the main steps that followed was to include proteins and establish the nucleic acids–the protein world, followed by the introduction of membrane systems and thereby a segregated environment of a primitive cell.

The origin of the first cell, the common progenitor of all living organisms, could go as far back as approximately four billion years ago and at that time, replication, involving translation machineries already existed. One hypothesis suggests that during the following two billion years, the unicellular system evolved to represent a fine net of metabolic reactions connected to more and more sophisticated machineries for nucleic acids replication and RNA to protein translation, as well as to maintain a large plasticity to enable the cell to respond to demands from the ever-changing environment. During this period, the first living cells were still dependent on organic compounds, which were the primary source of energy, and had an abiotic origin. Later, approximately 2.5 billion years ago, one of the major steps was the evolution of the ability of cells to use the energy of sunlight to power metabolism. Photosynthesis provided energetic independence and soon resulted in vast quantities of organic materials and oxygen. The introduction of aerobic metabolism significantly changed the cell biochemistry. Many enzymatic reactions became dependent, directly or indirectly, on the presence of oxygen. Aerobic metabolism allowed cells to grow larger. Some of the further major transitions include the origin of sex, the origin of multicellular organisms and the origin of social groups. Behind all these events stood proteins, nucleic acids, carbohydrates and lipids, with their evolving structures and functions.

1.4 What is Structural Biology and When Did It Start?

The field of structural biology focuses on a classical insight: In order to understand, we need to see. There is also a related statement: "Seeing is believing." This is true whether we deal with large objects like stars and galaxies, medium-sized objects like cars or watches, or very small objects like biological systems or particles. Structural biology is the science trying to make the biological subcellular and molecular objects visible and understood.

In a review in 1964, one of the pioneers in structural biology, James Watson expressed: "Unfortunately, we cannot accurately describe at the chemical level how a molecule functions unless we first know its structure." This describes the situation in a nutshell and it is manifested in one field of biology after another. The insight that he provided in generating, about 10 years earlier, the structure of the double helical DNA, is a classical example of this. The central steps in molecular biology, where nucleic acids took part, were not understood until Francis Crick and James Watson, using diffraction data obtained by Rosalind Franklin and Maurice Wilkins, deduced a model for DNA. Instantaneously, the model provided an extensive insight into the replication of DNA, the transcription of DNA to RNA and also the key steps of translation.

It is difficult to pinpoint the very beginning of structural biology. One important step is the purification of the molecular components. Friedrich Miescher discovered and isolated DNA in 1869. An understanding of the biological role of DNA was gained in 1944 when Avery, MacLeod, and McCarty showed that DNA is the genetic material. The elucidation of the structure of DNA in 1953 was a major milestone in structural biology.

Proteins have long been known, but their molecular nature was poorly understood. Proteins were classified as colloids without defined structures and shapes. Enzymes were believed to be of a different nature than lipids, carbohydrates or proteins and present in only very low concentrations in plants or animals. Theodor (The) Svedberg showed, using his ultracentrifuge, that proteins have unique molecular weights, and this finding constitutes an important step towards understanding the nature of proteins. J.B. Sumner and subsequently J.H. Northrop showed that proteins can have unique structures, as the enzymes urease and pepsin could be purified and crystallized. During the same period, W.M. Stanley crystallized a number of viruses. The perfection of crystallization and the crystallographic analysis of the crystals took several decades until it matured in the well resolved crystallographic structures of myoglobin and hemoglobin in 1959 and 1968, respectively.

Structural biology includes a number of methods in addition to diffraction and scattering. Electron microscopy, already in an early phase, provided an important technique to get an insight into the organization of biological systems and macromolecules. One major advance was the analysis of virus particles by Caspar and Klug at the end of the 1950s and the beginning of the 1960s. The symmetry principles could be deduced and for the larger viruses, different functional components were identified. Another development in the field of electron microscopy was the studies of 2D crystals of bacteriorhodopsin carried out by R. Henderson and N. Unwin using electron diffraction. Even though this opened new possibilities for structural studies, only a limited range of objects have yielded material good enough for structural studies. Subsequent to these developments, the single particle reconstruction studies of large molecular complexes and tomography have very significantly extended the capabilities of electron microscopy to contribute entirely new insights of structural biology at low and medium resolution. The single particle reconstruction, particularly, has the capacity to identify several different conformations in a sample to add to the dynamic picture of functional molecules.

Structural biology has moved numerous systems from a level where the molecules are represented as blobs to a level where the atomic coordinates as well as details of molecular interactions are known.

A limitation of crystallography is that, in principle, it gives still pictures of the molecular systems studied. In fortunate cases, a number of states can be crystallized and characterized down to atomic resolution. Usually, this limits the possible hypotheses of functional activities down to a few or sometimes only one. However, in many cases one would need insights into states that are not accessible to crystallization (perhaps because they are too short-lived), or to understand the dynamics of the system. Here NMR studies can sometimes provide the information that is missing.

Generally, NMR spectroscopy can yield structures. They are particularly valuable when crystals cannot be obtained, but when both NMR and crystallographic structures are available, the quality of the crystallographic information is normally better. However, the unique contributions of NMR spectroscopy come from dynamic studies of systems where the structures are already known. Here the mobility and details of transient interactions can be characterized.

To obtain optimal information several methods should be employed. Wrong or partial information can be corrected or extended. In the best-understood systems, physical and theoretical chemists have contributed their experimental or computational methods to achieve additional angles on the understanding of the systems.

1.5 A Short Summary of the Book

This book is a textbook of structural biology for undergraduate and graduate students. The focus is on the central and most interesting aspects of structure, combined with a focus on biology. The book makes no attempt to cover all the fields of biology or molecular biology. The emphasis is on systems where we know the structures reasonably well.

We cover the basic knowledge in the fields of proteins, nucleic acids and lipids and membranes, as well as include observations which are sometimes hidden in the original literature and that are important for scientists working in structural biology or related fields. For students who would need a reminder of some basic chemistry, the appendices are recommended. The appendices also include some tools necessary for using the databases of deposited sequence and structural data as well as a brief summary of Nobel prizes related to structural biology.