# APPLIED BIOPHYSICS

(Biophysics for Engineers through 15 Chapters)

Paata J. Kervalishvili
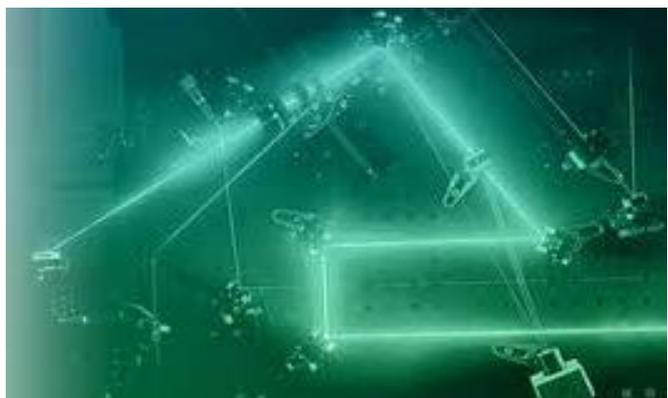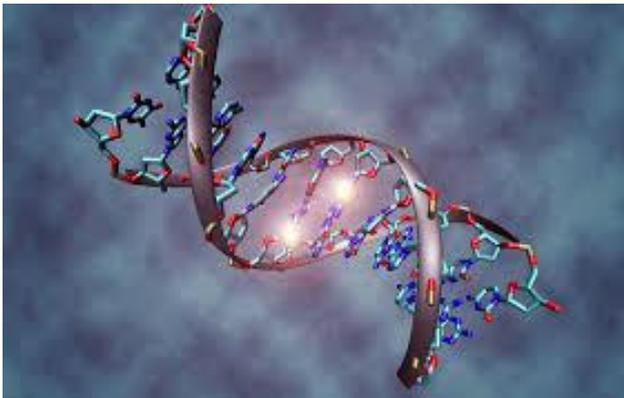
Georgian Technical University

Tbilisi 2021

Foreword

The lecture collection book Applied Biophysics is to present the fundamental and vital information about biophysics, via integration of physics and biology into biophysics, and is drawn on methods and research of this sciences.

This interdisciplinary, comprehensive, and comparative lecture collection is authored by distinguished scholar and scientist Professor Paata Kervalishvili. It offers expert scholarship on biophysics that is ideal for undergraduate/graduate students as well as for engineers and physicists working in applied sciences.

The book pays particular attention on the topic of unity in the sciences, which is explored through the following questions: Is there one privileged, most basic or fundamental concept or kind of thing, and if not, how are the different concepts or kinds of things in the universe related? How the various natural sciences (e. g. Physics, chemistry, biology) be unified into a single overarching theory, and can theories within a single science (e.g., physics, or models of evolution and development in biology) be unified?

Does the unification of these parts of physics and biology involve only matters of fact or are matters of value involved as well? What kinds of unity, not just units, in these sciences are there? What roles can unification play in scientific practices, their development, application and evaluation?

Views on matters of unity and unification make a difference in science and provide strong empirical or methodological guidance and even justification for hypotheses, projects, and specific goals. This sets a standard of what carries the authority and legitimacy of what it is to be scientific. As a result, they make a difference in scientific evaluation, management and application, especially in public domains such as healthcare and economic decision-making.

The lecture collection book Applied Biophysics point out that implicit assumptions about what unification can do have on science education. At the end of the day one should not lose sight of the larger context that sustains problems and projects in most disciplines and practices. The ideal of unity, frequently under the guise of harmony, has long been a standard of scientific advantage. Unities and unifications help scholars meet cognitive and practical demands upon the life.

Instead of introduction

This book is intended to be the basic and essential source of information about biophysics, not as an in-depth and exhaustive treatise of the field but useful for engineers and physicists working in applied sciences. The objective is to provide a basic knowledge of a broad range of topics so that even a newcomer can rapidly acquire the minimal necessary background for research and development and applications as well.

Although several monographs, textbooks and lecture-notes exist that cover selective aspects of biophysics, there is a clear necessity for a cluster of lectures that provides a unified synthesis of this subject. The need for such a book as this became apparent while teaching this topic as an interdisciplinary course available to students in many departments at the university level. The makeup of the registrants for these tutorial courses has been multidisciplinary. Over the years, participants in these courses have constantly emphasized the need for a comprehensive and multidisciplinary content, which could illuminate the main achievements in the field.

The book is written with the following readership in mind:

- Undergraduate and graduate students specialized in medical physics, biophysics and biomedical engineering as well.
- Researchers working in the area; it will provide useful information for them in areas outside their expertise and serve as a reference source.
- Educators who provide training and tutorial courses at universities as well as at various professional society meetings; it will serve them as a textbook that elucidates basic principles of existing knowledge and multidisciplinary approaches

The book encompasses the fundamentals and various applications involving the integration of physics and biology into biophysics. Each lecture begins with an introduction describing what a reader will find in it. Each lecture ends with highlights which consist additional information about subjects discussed in relevant lecture including figures, graphs, equations. In each of the lecture, a description of future directions of research and development is also provided, as well as a brief discussion of the current status, identifying some of areas of future opportunities. Some of the existing sources of instrumentation and supplies relevant to the content of many of the applications are also described.

Author intends that this book serve both as a textbook for education and training as well as a reference book that aids research and development of those areas integrating physics, physical chemistry and biology of different living organisms. Another aim of the book is to stimulate the interest of researchers and healthcare professionals and to foster collaboration through multidisciplinary programs.

Author hopes that this book will find wide relevant readership. At the same time he believes that fruitfulness of the book is absolutely personal object. Readers, have their favorite books, and this is an emotional statement, laden with context. Similarly, writers bring not just their knowledge and their technical skill to the creation of a book, but also their personalities. In writing something which might be used as a textbook author is giving to readers his view which strongly might be not a consensus of the domain field. This is advantage and at the same time disadvantage of free thinking, which is the main characteristic of creative minds.

And finally, Author would like to express his sincere appreciation the individuals whose broad-based support has been of paramount value in completing and editing this book: his teachers and colleagues from the past and present and mainly to his brilliant collaborators – editors of the book, professors Irina Gotsiridze, Lali Chakhvashvili and Tamar Berberashvili, for their uncountable help and support. Author is hugely grateful to all of them.

Chapter1:

The Scope and Topics of Biophysics

Introduction

When we are talking of a textbook about elementary particle physics, or condensed matter, or photonics there is often little doubt about what will be found inside the covers. There are questions, perhaps, about the level and style of presentation, or about the emphasis given to different subfields, but the overall topic is clear. The situation is very different for books or courses that attempt to bring the intellectual style of physics to bear on the phenomena of life.

"Biophysics" or "biological physics" has come to mean many different things simultaneously.

Here are always some questions. Where is the boundary between physics and biology? (Because physics - from "phusis" nature). Is biophysics really physics, or just the application of methods from physics to the problems of biology? In the interaction between physics and biology, what happens to chemistry?

Physicist's main task is to ask certain kinds of questions about Nature, and to seek certain kinds of answers. In physics, we (try to) teach principles and derive the predictions for particular examples. In biology, teaching proceeds (mostly) from example to example.

Following Galileo - the book of Nature written in the language of mathematics, but there is only one book, and we expect that if we really grasped its content it could be summarized in very few pages.

Physics is an exact science. Biology, on the other hand, can be classified as a descriptive science. However, in recent times, the latter is also being transformed into a more exact science with the advent of molecular biology and molecular biophysics. One aspect of this transformation originates from the applications of physics to physiology. Physical laws or concepts such as mechanics, hydrodynamics, optics, electrodynamics and thermodynamics are used to explain physiological observations like muscle contraction, neural communication, vision, etc. A second and more fundamental aspect of the transformation emerged from the search for universal principles governing the world around us. By the beginning of twentieth century it was realized that the laws of physics and chemistry, which were applied to non-living things, could equally well explain forces controlling biology, and that no new fundamentally different principles were necessary to explain the organisms and interactions which make up the living world. In the hundred years that have passed since then, this concept has become stronger and today nobody thinks it necessary to invoke any special physical or chemical forces or laws in the study of biology. This chapter will describe, in an elementary way, some of these basic concepts and laws of modern physics and chemistry, in particular those which are directly relevant to biology.

The giants of classical physics—Helmholtz, Maxwell, and Rayleigh, to name a few— often crossed borders among disciplines that we now distinguish as physics, chemistry, biology, and even psychology. Some of their forays into the phenomena of life were driven by a desire to test the universality of physical laws, such as the conservation and exchange of energy. A very different and primitive motivation was that our own view of the world is determined by what we can see and hear, and more subtly by what we can reliably infer from the data that our sense organs collect or even could not.

The reason is that it is no principal boundary between optics and vision, or between acoustics and hearing. The stimulus of our physical system to reach the nature of our perceptions and our ability to learn from the universe is driven by laws and rules nature.

The rise of modern physics motivated another wave of physicists to explore the phenomena of life. Fresh from the triumphs of quantum mechanics, they were emboldened to seek new challenges and brought new concepts.

For instant Erwin Schrodinger, in his influential series of lectures entitled *What is Life?*, seized upon the discovery that our precious genetic inheritance was stored in objects the size of single molecules, highlighting how surprising this is for a classical physicist, and contrasted the order and complexity of life with the ordering of crystals, outlining a strikingly modern view of how non linear and non equilibrium systems can generate structure out of disorder, continuously dissipating energy.

Theoretical physicist Max Delbruck did play a central role, not least because of his insistence that the community should focus on the simplest examples of crucial biological phenomena, reproduction and the transmission of genetic information. The goal of molecular biology to reduce these phenomena to interactions among a countable set of molecules surely echoed the physicists' search for the fundamental constituents of matter, and perhaps the greatest success of molecular biology

is the discovery that many of these basic molecules of life are universal, shared across organisms separated by hundreds of millions of years of evolutionary history.

Where classical biology emphasized the complexity and diversity of life, the first generation of molecular biologists emphasized the simplicity and universality of life's basic mechanisms, and it is not hard to see this as an influence of the physicists who came into the field at its start.

Perhaps inspired by the successes of their intellectual ancestors, each subsequent generation of physicists offered a few converts. The idea, for example, that the flow of information through the nervous system might be reducible to the behavior of ion channels and receptors inspired one group, armed with low noise amplifiers, intuition about the interactions of charges with protein structure, and the theoretical tools to translate this intuition into testable, quantitative predictions.

Understanding that the mechanical forces generated by a focused laser beam are on the same scale as the forces generated by individual biological molecules as they go about their business brought another generation of physicists to our subject. The sequencing of whole genomes, including our own, generated the sense that the phenomena of life could, at last, be explored comprehensively, and this inspired yet another group.

Through the many generations, some conventional views arose about the nature of science at the borders between physics and biology. First, there is a strong emphasis on technique. From X–ray diffraction to the manipulation of single molecules to functional imaging of the brain, it certainly is true that physics has developed experimental techniques that allow much more direct exploration of questions raised by biologists. Second, there is a sense that in some larger classification system, biophysics is a biological science.

At present, most questions about how things work in biological systems are viewed as questions that must be answered by experimental discovery. The situation in physics is very different, in that theory and experiment are more equal partners. In each area of physics we have a set of general theoretical principles, all interconnected, which define what is possible; the path to confidence in any of these principles is built on a series of beautiful, quantitative experiments that have extended the envelope of what we can measure and know about the world.

Beyond providing explanations for what has been seen, these principles provide a framework for exploring, sometimes playfully, what ought to be seen. In many cases these predictions are sufficiently startling that to observe the predicted phenomena (a new particle, a new phase of matter, fluctuations in the radiation left over from the big bang, ...) still constitutes a dramatic experimental discovery.

It is a remarkable thing that, pulling on the threads of one biological phenomenon, we can unravel so many general physics questions. In any one case, some problems will be presented in purer form than others, but in many ways everything is there.

The first problem, as noted above, is that there really is something different about being alive, and we'd like to know what this is—in the same way that we know what it is for a collection of atoms to be solid, for a collection of electrons to be superconducting, or for the vacuum to be confining (of quarks).

Looking around, we more or less immediately identify things which are alive, and the criteria that we use in making this discrimination between animate and inanimate matter surely have nothing to do with DNA or proteins.

Asking for the order parameters of the living state is a hard problem, and not terribly well posed. One way to make progress is to realize that as we make more quantitative models of particular biological systems.

If real biological systems occupy only a small region in the space of possible systems, we have to understand the dynamics by which systems find their way to these special parameters.

In order to survive in the world, organisms do indeed have to solve a wide variety of problems. Many of these are really physics problems: converting energy from one form to another, sensing weak signals from the environment, controlling complex dynamical systems, transmitting information reliably from one place to another, or across generations, controlling the rates of thermally activated processes, predicting the trajectory of multidimensional signals, and so on. it's obvious that everything which happens in living systems is determined by the laws of physics.

Identifying all the physics problems that organisms need to solve is not so easy. Thinking about how single celled organisms, with sizes on the scale of one micron,

manage to move through water, we quickly get to problems that have the look and feel of problems that we might find in fundamental book of Landau and Lifshitz. On the other hand, it really was a remarkable discovery that all cells have built Maxwell demons, and that our description of a wide variety of biochemical processes can be unified by this observation.

Beautiful description of problems at the interface of physics and biology is made in Werner Shroedinger's book "What is Life". To get a sense of the excitement and spirit of adventure that our intellectual ancestors brought to the subject, you should also look at the remarkable essays of Niels Bohr (1933) and Max Delbruck (A physicist looks at biology. M Delbruck, Trans Conn Acad Arts Sci 38, 173–190, 1949).

Generally, when we are searching for principles, we start by being fascinated with the phenomena of life.

Having introduced ourselves in some detail to one particular biological phenomenon, we proceed to explore three candidate principles: the importance of noise, the need for living systems to function without fine tuning of parameters, and the possibility that many of the different problems solved by living organisms are just different aspects of one big problem about the representation of information.

The past century has witnessed many technological breakthroughs, one of which is most advance part of optics - photonics. Photonics utilizes photons instead of electrons to transmit, process, and store information and thus provides a tremendous gain in capacity and speed in information technology. Photonics is an all-encompassing light-based optical technology that is being hailed as the dominant technology for this new millennium. The invention of lasers, a concentrated source of monochromatic andhighly directed light, has revolutionized photonics. Since the demonstration of the first laser in 1960, laser light has touched all aspects of our lives, from home entertainment, to high-capacity information storage, to fiber-optic telecommunications, thus opening up numerous opportunities for photonics. A

new extension of photonics is biophotonics, which involves a fusion of photonics and biology. Biophotonics deals with interaction between light and biological matter. A general introduction to biophotonics is illustrated in Figure 1.1.

The use of photonics for optical diagnostics, as well as for light-activated and light-guided therapy, will have a major impact on health care. This is not surprising since Nature has used biophotonics as a basic principle of life from the beginning. Harnessing photons to achieve photosynthesis and conversionof photons through a series of complex steps to create vision are the best examples of biophotonics at work. Conversely, biology is also advancing photonics, since biomaterials are showing promise as new photonic media for technological applications.

As an increasingly aging world population presents unique health problems, biophotonics offers great hope for the early detection of diseases and for new modalities of light-guided and light-activated therapies. Lasers have already made a significant impact on general, plastic, and cosmetic surgeries. Two popular examples of cosmetic surgeries utilizing lasers are skin resurfacing and hair removal. Laser technology also allows one to administer a burst of ultrashort laser pulses that have shown promise for use in tissue engineering. Furthermore, biophotonics may produce retinal implants for restoring vision by reverse engineering Nature's methods.

This book provides an introduction to the exciting new field of biophotonics and is intended for multidisciplinary readership. The book focuses on its potential benefits to medicine.

The focus of the book is on optical probing, diagnostics, and light activated therapies. However, biophotonics in a broad sense also includes the use of biology for photonics technology, such as biomaterials and development of bioinspired materials as photonic media.
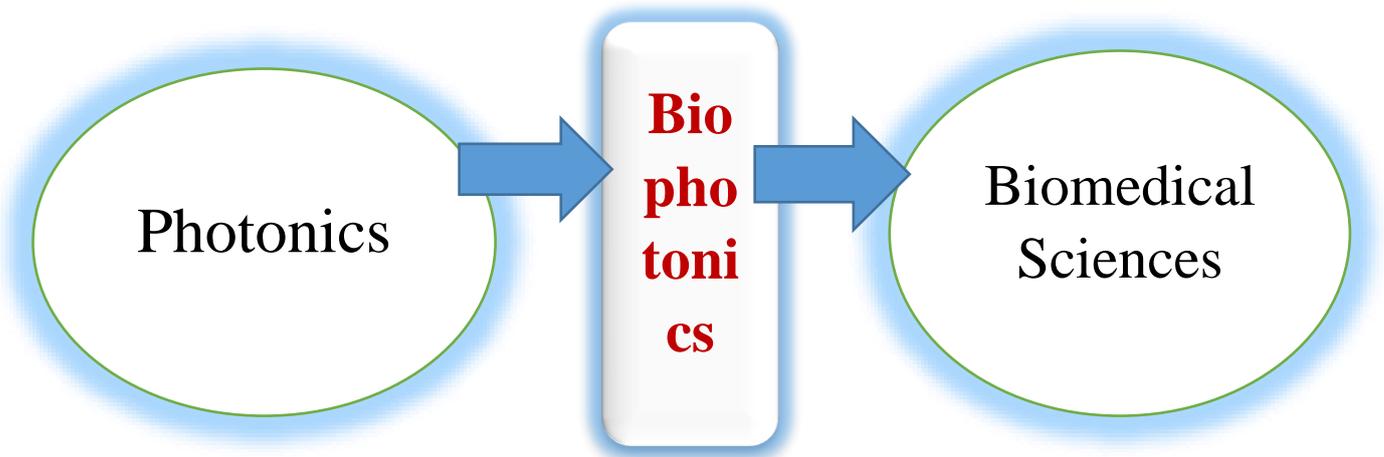

Figure 1.1. Photonics Biophotonics and Biomedicine

Photonics Technology for BioMedical Sciences includes:

Bioimaging/Biosensing

Optical diagnostics

Light based therapies

Tissue engineering

Light manipulations of cells

Laser media

Optical communication channels

Optical signal processing, etc.

In the 21st century, major technological breakthroughs are more likely to occur at the interfaces of disciplines. Biophotonics integrates four major technologies: lasers, photonics, nanotechnology, and biotechnology. Fusion of these technologies truly offers a new dimension for both diagnostics and therapy. Biophotonics creates many opportunities for chemists, physicists, engineers, physicians, dentists, health-care personnel, and biomedical researchers. The need for new materials and technologies to provide early detection of diseases, to produce more effective targeted therapies, and to restore impaired biological functions is constantly increasing. The world we live in has become more complex and increasingly dependent upon advanced technologies.

The benefits of lasers to health care are well recognized, even by the general population. Many light-based and spectroscopic techniques are already currently being used as optical probes in clinical laboratories as well as in medical and other health-care practices. Photodynamic therapy, which uses light to treat cancer and has a great potential for growth, is now being practiced.

Producing trained health-care personnel and new generations of researchers in biophysics is of the utmost importance to keep up with the increasing worldwide demands. Undergraduate and graduate research training programs are needed to develop a skilled workforce and a future generation of researchers respectively for a rapidly growing biotechnology industrial sector.

In the areas of research and development, many disciplines can contribute individually as well as collaboratively. Multidisciplinary interactions create unique opportunities that open new doors for the development and application of new technologies.

Biophisics and its subfields offer tremendous opportunities for both biotechnology development and fundamental research. From a technological perspective, biophysics integrates several major technologies: lasers, photonics, nanotechnology, biotechnology, etc. These technologies have already established themselves in the global marketplace, collectively generating hundreds of billions of dollars per year. Biophotonics also impacts a wide range of industries including biotechnology companies, health care organizations (hospitals, clinics, and medical diagnostic laboratories), medical instrument suppliers, and pharmaceutical manufacturers, as well as those dealing with information technology and

optical telecommunication. In the future, biophysics will have a major impact both in generating new technologies and in offering huge commercial rewards worldwide.

Biophysics offers challenging opportunities for researchers. For instant, a fundamental understanding of the light activation of biomolecules and bioassemblies, and the subsequent photoinduced processes, is a fundamental requirement in designing new probes and drug delivery systems. Also, an understanding of multiphoton processes utilizing ultra short laser pulses is a necessity both for developing new probes and creating new modalities of light-activated therapy. Some of the opportunities, categorized by discipline, are listed below:

Physical Chemistry:

- Development of new fluorescent tags
- Chemical probes for analyte detection and biosensing
- Nanoclinics for targeted therapy
- Nanochemistries for materials probes and nanodevices
- New structures for optical activation

Physics:

- Photoprocesses in biomolecules and bioassemblies
- New physical principles for imaging and biosensing• Single-molecule biophysics
- Nonlinear optical processes for diagnostics and therapy

Engineering:

- Efficient and compact integration of new generation lasers, delivery
- systems, detectors
- Device miniaturization, automation, and robotic control
- New approaches to noninvasive or minimally invasive light activation
- Optical engineering for *in vivo* imaging and optical biopsies
- Nanotechnologies for targeted detection and activation
- Optical BioMEMS (micro-electro-mechanical systems) and their
- nanoscale analogues.

Biomedical Research:

- Bioimaging to probe molecular, cellular, and tissue functions
- Optical signature for early detection of infectious diseases and cancers
- Dynamic imaging for physiological response to therapy and drug
- delivery
- Cellular mechanisms of drug action

- Toxicity of photo activatable materials
- Biocompatibility of implants and probes

    Clinical medicine:

- *In vivo* imaging studies using human subjects
- Development of optical *in vivo* probes for infections and cancers
- *In vivo* optical biopsy and optical mammography
- Tissue welding, contouring, and regeneration
- Real-time monitoring of drug delivery and action
- Long-term clinical studies of side effects

Additional material for the second lecture

BIOPHYSICS

Advanced interdisciplinary science which involves: physics, biology, chemistry, mathematics, computer science.

1892:Karl Pearson (missing link between biology and physics - name biophysics)

1943:Erwin Schrodinger (Nobel Prize, 1933)

Austrian-born physicist and theoretical biologist Erwin Schrödinger, one of the founders of quantum theory in physics, also became one of the first scientists to suggest a study of quantum biology in his 1944 book What Is Life?

Many biological processes involve the conversion of energy into forms that are usable for chemical transformations and are quantum mechanical in nature. Such processes include chemical reactions, light absorption, formation of excited electronic states, transfer of excitation energy, and the transfer of electrons and protons (hydrogen ions) in chemical processes such as photosynthesis and cellular respiration. Quantum biology uses computation to model biological interactions in light of quantum mechanical effects.

1946: Biophysics Research Unit, King's College,

London, hire physicists to work on questions of biological significance; Maurice Wilkins, Rosalind Franklin: Xray diffraction of DNA.

1953: Francis Crick (particle physicist turned into biophysicist at Cambridge) and James Watson (biologist): double helix structure of DNA

1957: The Biophysical Society founded.

TOPICS of BIOPHYSICS

Biophysical topics based on relative size of the subject: molecular and subcellular biophysics;

physiological and anatomical biophysics; environmental biophysics.

Biophysical techniques and applications: general; biophysical techniques; imaging biophysics;

Medical biophysics.

Molecular and Subcellular Biophysics

The Structure and Conformation of Biological Molecules: Structure Function Relationships; Conformational Transitions; Ligand Binding and Intermolecular Binding; Diffusion and Molecular Transport; Membrane Biophysics; DNA and Nucleic Acid Biophysics; Protein Biophysics; Energy Flow and Bioenergetics; Thermodynamics; Statistical Mechanics; Kinetics

Molecular Machines; Allosterics.

Biophysical Techniques and Applications:

Ultracentrifugation to separate molecules of different sizes based on the sedimentation principle, up to 10^6 g;

Electrophoresis to separate molecules of different molecular mass/size based on the sedimentation principle; electric field acts on the charged molecules; gel electrophoresis

Size Exclusion Chromatography (SEC) uses tightly packed gel beads and sedimentation based on gravity (and sometimes pressure) to trap small molecules and allow larger molecules to pass through the gel faster than small molecules;

Spectroscopy mostly with incident EM radiation and measuring the intensity/direction/polarization of the emitted radiation (originally only the visible spectrum 380-750nm was used; now also UV and IR); in addition to EM also electron and mass spectroscopy.

Absorption Spectroscopy to find e.g. the concentration of molecules in the solution by using EM of a particular λto shine on the sample and measure the intensity that comes out OR absorbance versus λ to identify the type of molecules;

Fluorescence Spectroscopy to characterize molecules and to follow conformational transitions; caused by absorption at a one wavelength and emission at a longer wavelength (electrons drop from their excited

energy state emitting light;

Mass Spectrometry to measure mass or molecular weight of molecules; molecules are ionized in a vacuum, then passed through a magnetic field;

X Ray Crystallography to determine the relative positions of atoms with ina crystal by using diffraction on a 3D crystal lattice; high resolution of structural details but the molecules need to be in a crystalline phase;

Nuclear Magnetic Resonance Spectroscopy (NMR) to obtain structural information about molecules of the highest resolution using EM of a radiofrequency, which interacts with nuclear spins of atoms in a large magnetic field, causing them to jump between the spin states and emit at different λ depending on the local structure around the atom;

Electron Microscopy to view objects 1,000 - 2,500smaller than those seen by light microscopes (electrons of a small wavelength are used instead of EM); transmission EM (TEM) and scanning electron microscopy (SEM);

Atomic Force Microscopy (AFM) with resolution similar to TEM, 3D

features like SEM; a mechanical probe (a tip) moves along the surface of

the scanned object to obtain 3D information;

Optical Tweezers to hold and manipulate microscopic particles even single molecules or atoms using focused laser beams to create forces of the order of $pN = 10^{12}N$ (0.1 nm to 10,000 nm size objects) and measure forces needed to bend or break DNA, for example;

Voltage Clamp is used in electrophysiology to determine electric currents in cells, in particular neurons; a fine microelectrode is inserted into the cell with another in contact with the surrounding fluid while the voltage is clamped (held constant) by a feedback that generates a counter current to that generated by the cell;

Current Clamp is analogous to voltage clamp; the current is clamped (held constant) and the voltage change induced by the cell measured;

Patch Clamp is alternative to voltage/current clamp; the electrode is placed inside a micropipette with electrolyte solution and the micropipette combined with a gentle suction electrically isolates a small patch on the membrane; enables to study a single ion channel within the membrane;

➔ Calorimetry measures CP or CV versus T: transitions or ligand and binding.

Chapter 2:

The Main Targets of Biophysics

Obviously, the main target of Biophysics is to bring a physical approach to the study of biological processes that range in scale from the sub-molecular (studying the interplay of the inter-atomic forces that give proteins their particular shape, motion and function) to the systems level (studying the concerted activity of neural and genetic circuits). Study in these research areas employs unique types of physical instrumentation and research methodologies such as X-ray diffraction, NMR, EPR and fluorescence spectroscopy, electronic and optical instrumentation, mathematical computer modeling and analysis. Study itself includes research areas of: Structural Biophysics and Protein Dynamics, Systems Neuroscience, Molecular Microscopy and Optical Probes, Cell Signaling and Cellular Physiology, Computational Biology and Genomics, Brain Imaging and Bioelectronics, Comparative Biomechanics, etc. With the development of new science technologies --most especially high resolution imaging-- there is an ongoing revolution in our understanding of intracellular architecture. The arrangement of proteins in three dimensional space reveals a great deal about how cellular processes occur. One example is neurotransmitter release at synapses, which requires a remarkable assemblage of different proteins to occur correctly. Ants are well known bio-engineers. Besides farming aphids and fungi and creating architecturally complex colonies, they are also capable of creating rafts that float on water -- and other structures-- by grasping each other in a complex three-dimensional assemblage and it could be the inspiration for self-assembling micro-robots. Nanotechnology has been around for at least 35 years when the first nano-fabrication laboratories were set up, but the technology has improved, leading to fascinating discoveries in basic science. Electrical conduction between bacterial cells is one example. Using the nanotech approach it was demonstrated that bacteria connect to each other with nanowires to share electricity. This one is more for the high energy particle physicists than it is for the biophysicists. It's an idea for obtaining evidence for dark matter in the universe, using long single stranded DNA molecules. The high energy particles would break the DNA, which could then be sequenced to obtain the lengths. The DNA molecules would be housed in a multi-array system to capture as many events as possible.

Biophysics as a recent science, forged in the heart of the molecular biological revolution. Many of the most efficacious treatments of cancer arise from developments in biophysics. Most notable are real-time imaging during radiation treatment of tumours, to maximize tumour irradiation and minimize damage to peripheral tissue. A biophotonic approach is even more selective. The state-of-the-art --tested in a mouse model-- is to use antibodies that bind specifically to tumour cells. The antibody has a molecule attached to it that absorbs in the infrared and infrared light can penetrate tissue quite deeply. Once photo-activated, the molecule selectively 'attacks' the cell to which it is attached.

All above mentioned move us to study and well understand the physical effects which occur in biological substances.

Four Classes of Macromolecules:

(A) DNA in a B form (B) Protein (hemoglobin) (C) Lipid molecule (phosphatidylcholine) (D) Branched complex carbohydrate


2 nm
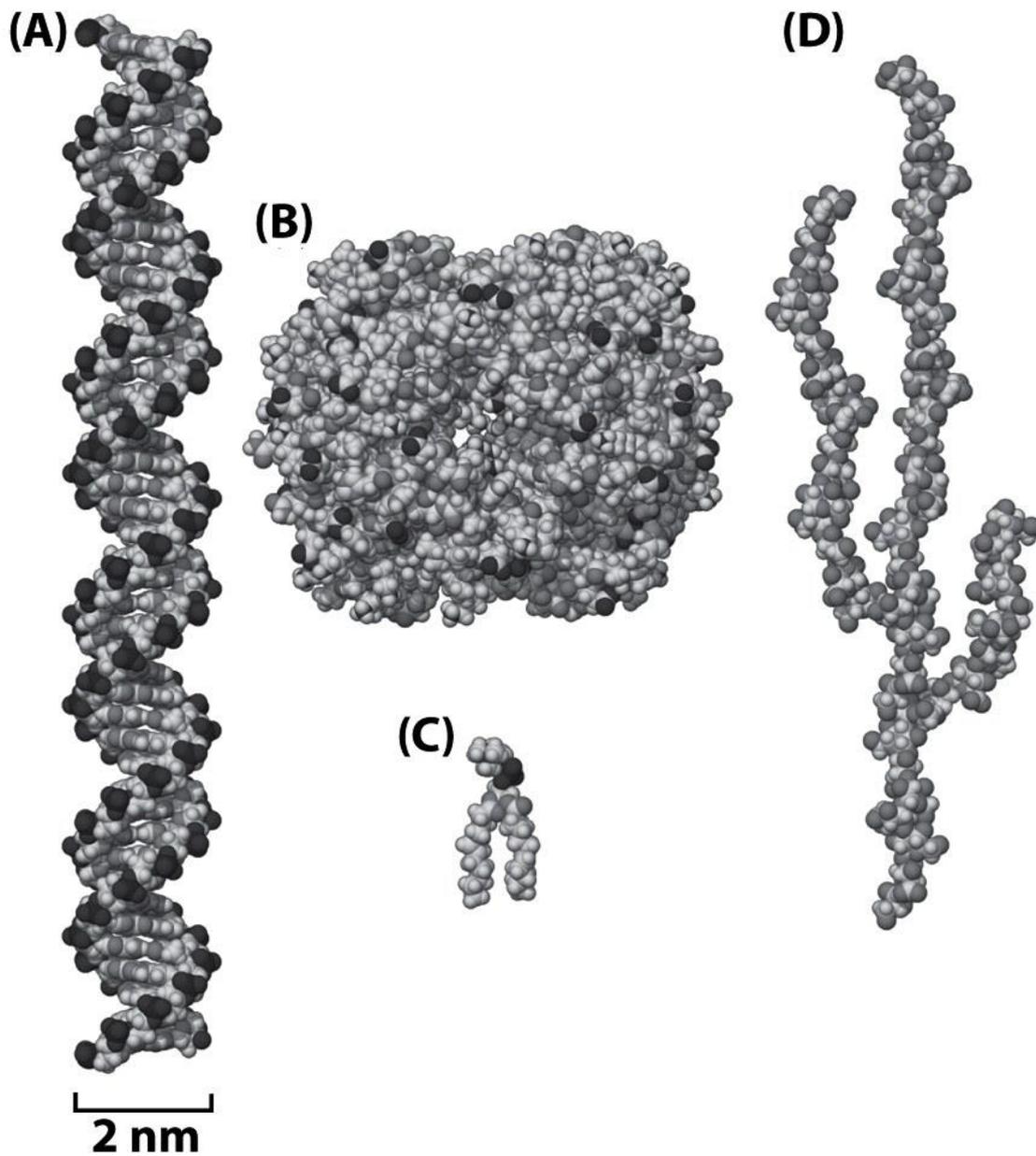
Fig. 2.1. Four Classes of Macromolecules. (Physical Biology of cell Garland Science 2009).

(1) DNA and RNA molecules: made of nucleic acids (2) proteins: made of amino acids
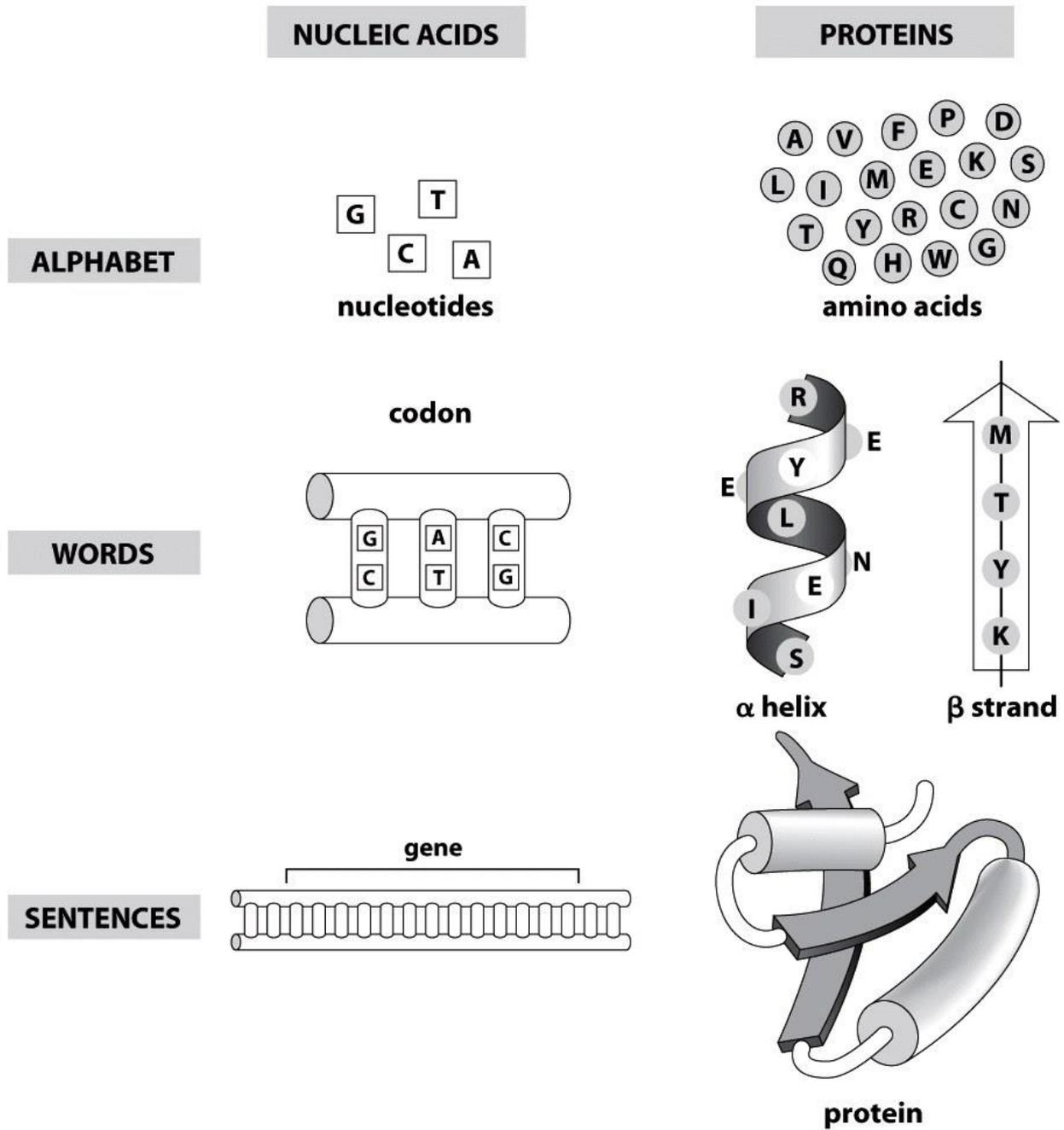
Fig. 2.2. Polymer Nature of Macromolecules. (Physical Biology of cell Garland Science 2009).
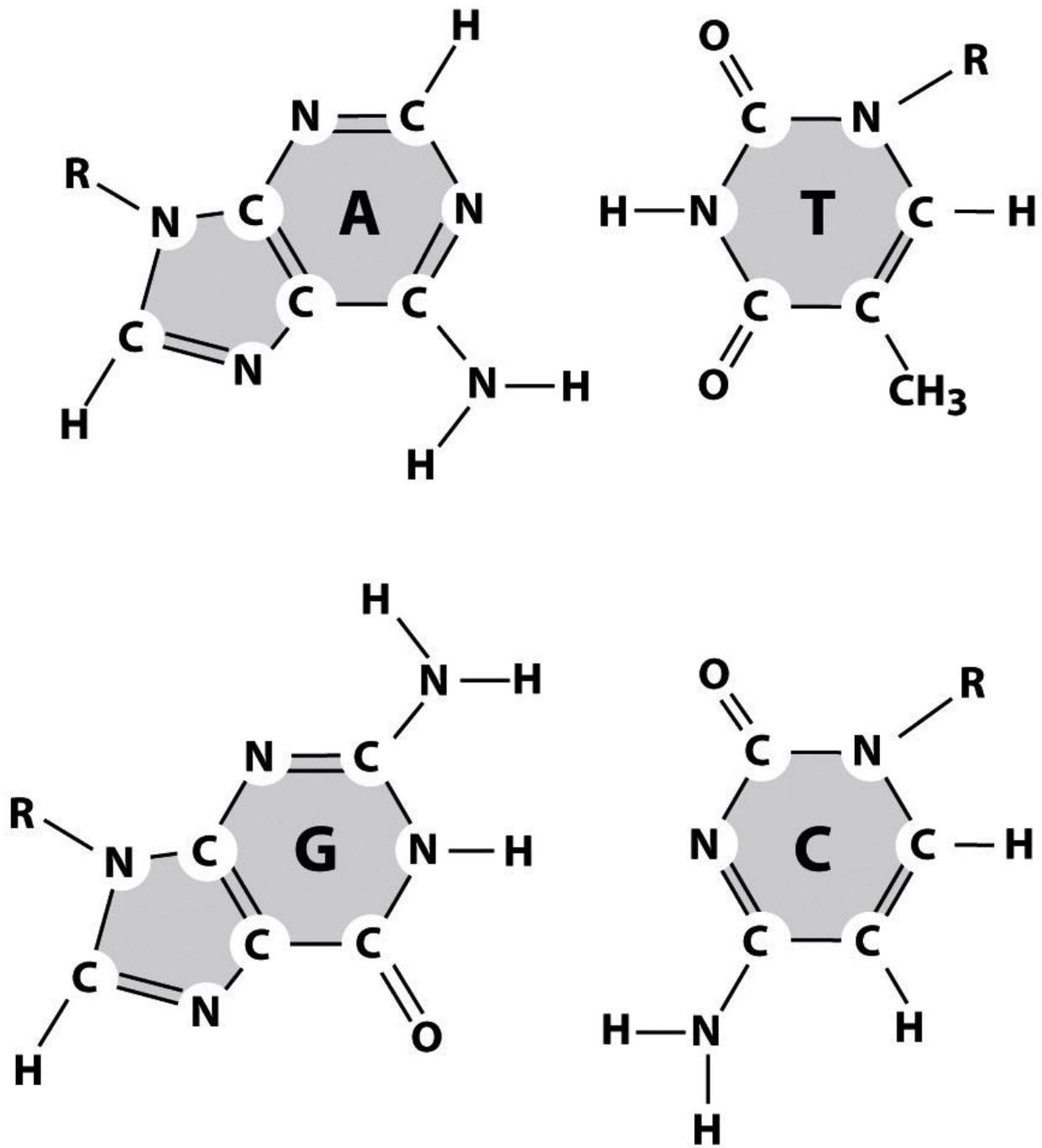
Fig.2.3. Each DNA molecule is polymers of four nucleic acids: A, T, G, C

Backbone groups: deoxyribose phosphate
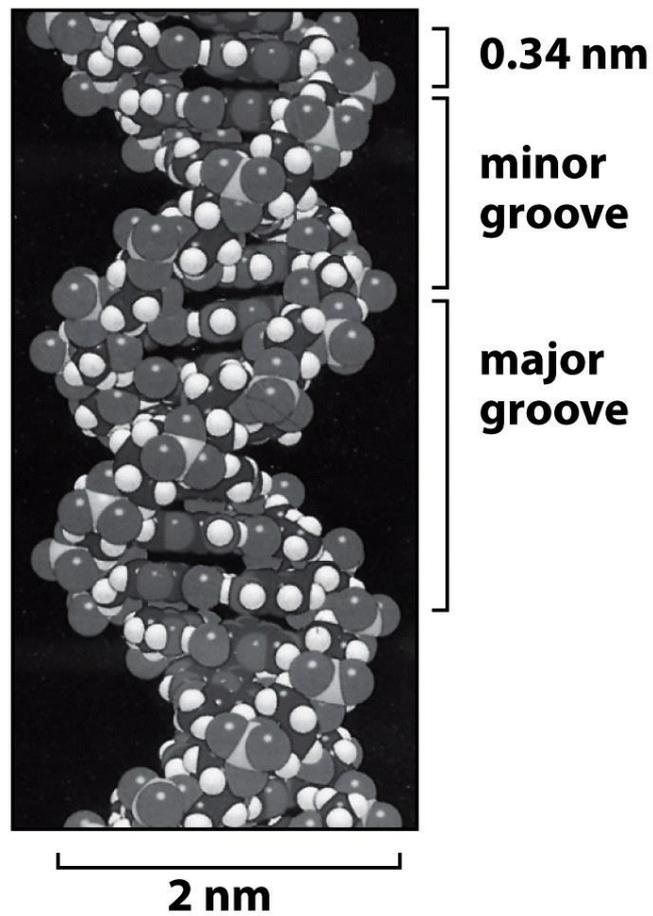
Fig.2.4. Nucleic acids within a doublestrand DNA:
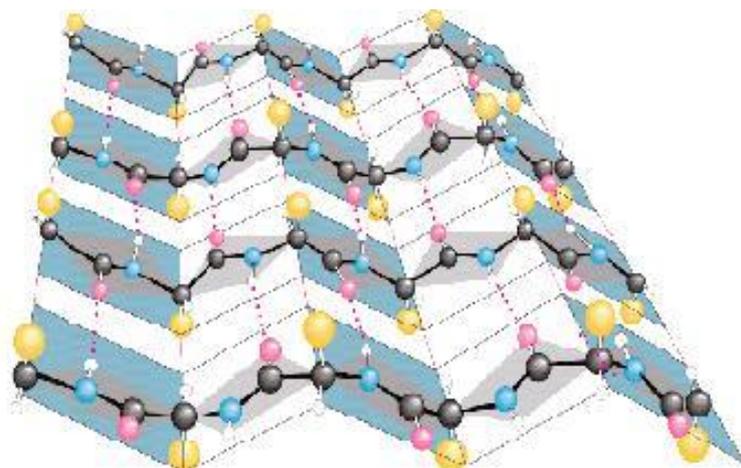
Fig. 2.5 Polypeptide bonds joined together by hydrogen bonds (red dot lines) in a β-pleated sheet. (http://www.people.virginia.edu
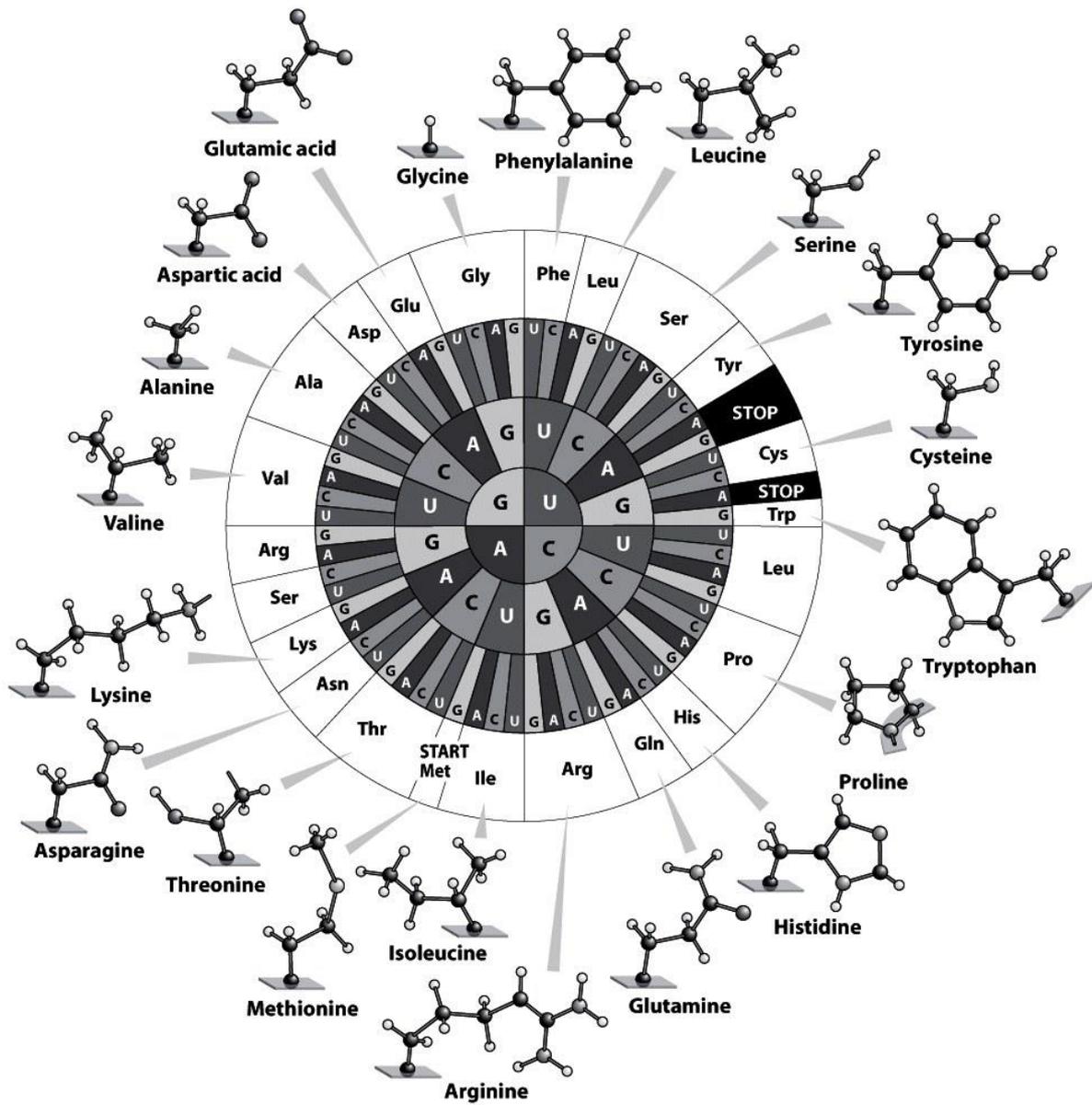
Fig. 2.6. mRNA encodes Proteins: nucleotide to amino acid sequences   Uracil (instead of Thymine).
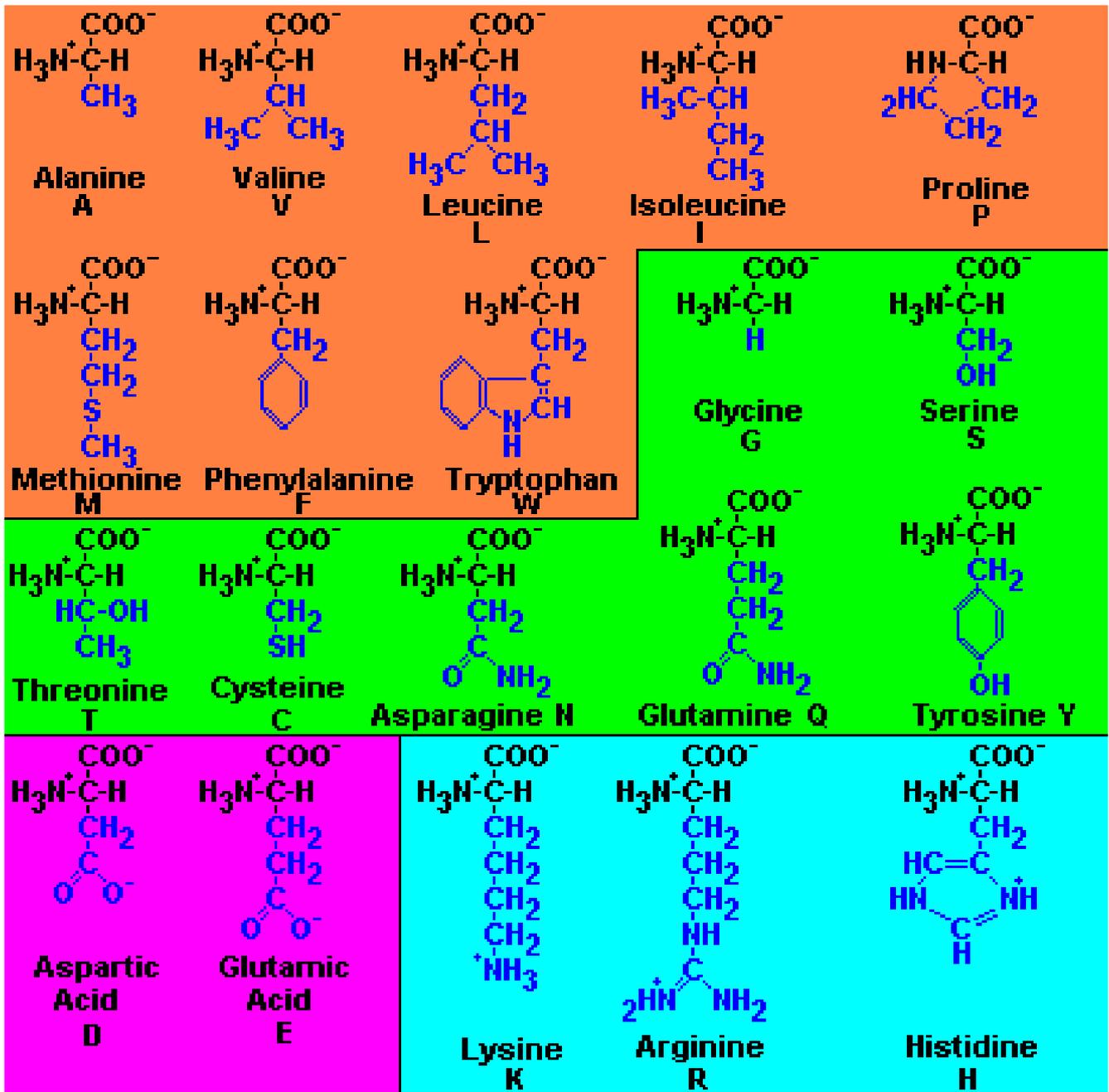
Fig.2.7. 20 natural amino acids; Up – hydrophobic ("hate" water); Down – hydrophilic ("love" water).
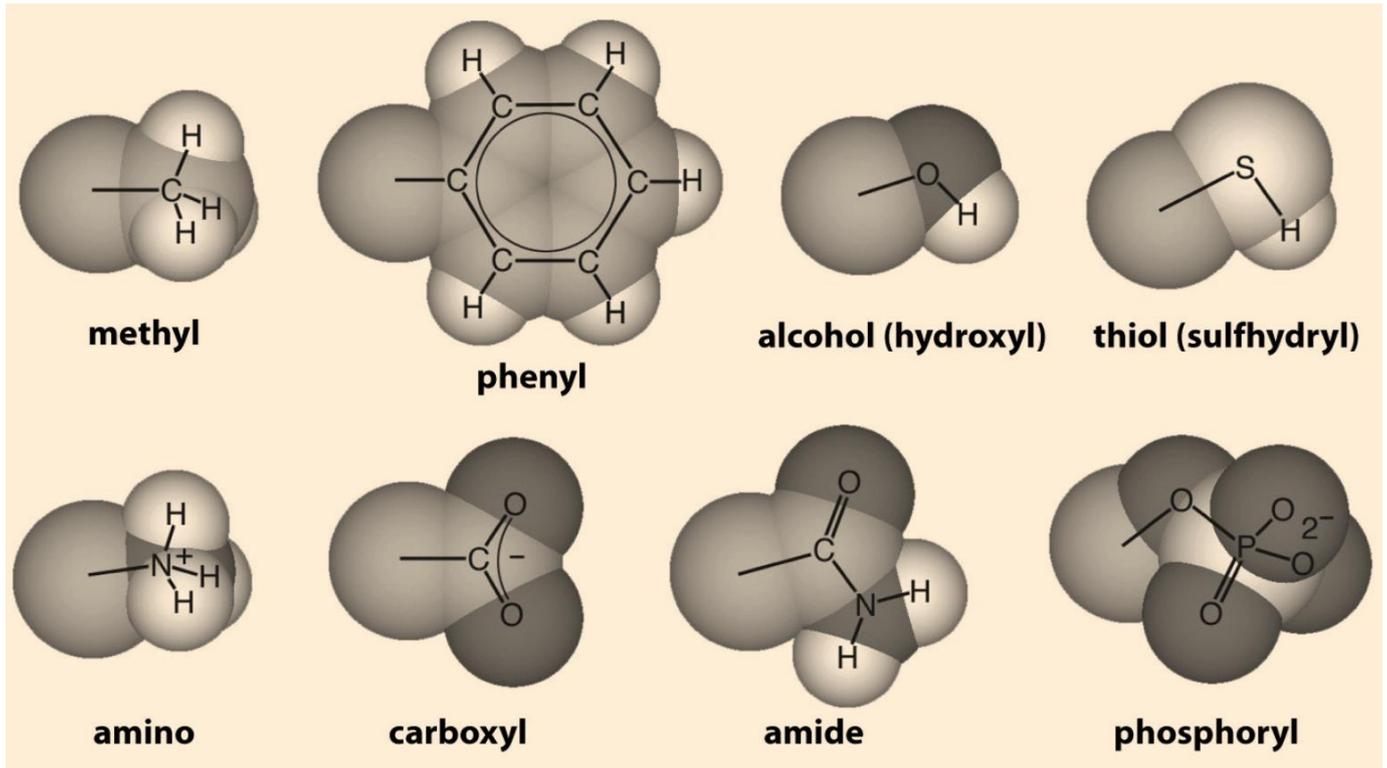
Fig.2.8. Most common chemical groups found in proteins: CHNOPS acronym: elements most commonly found in cells (accounts for about 98% of all atoms).
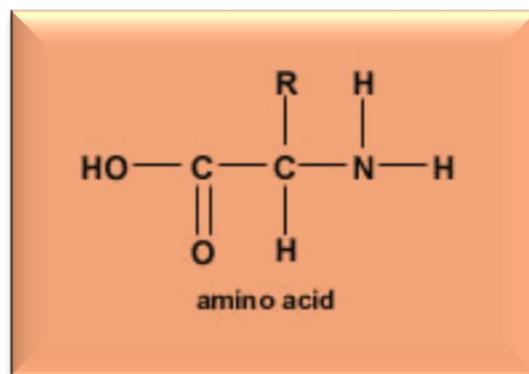
Fig.2.9. Amino acid: (2 of them could form a Peptide Bond)
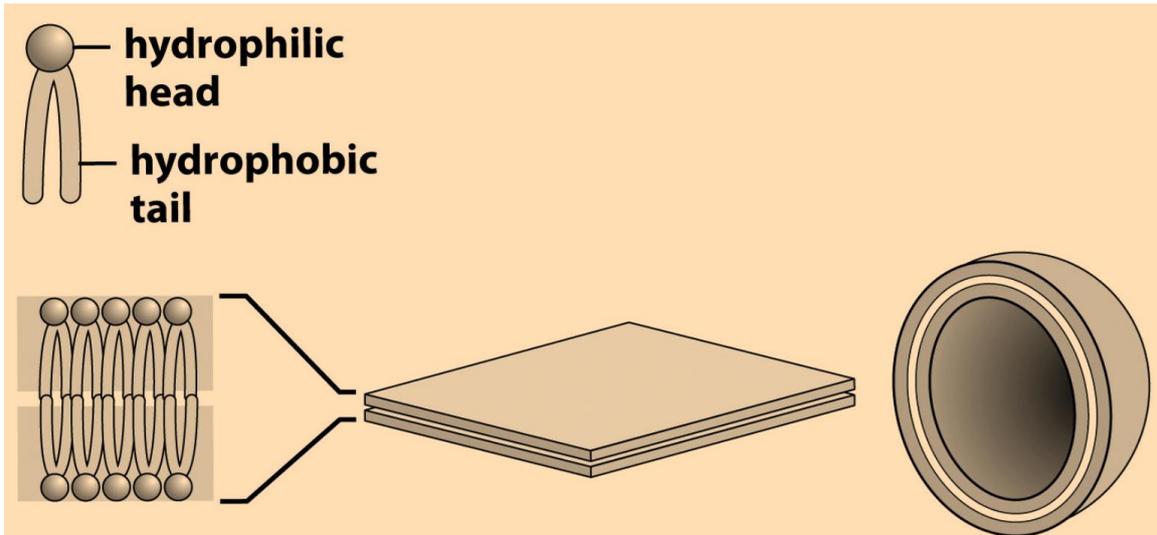
Fig.2.10. Membrane is a lipid bilayer.
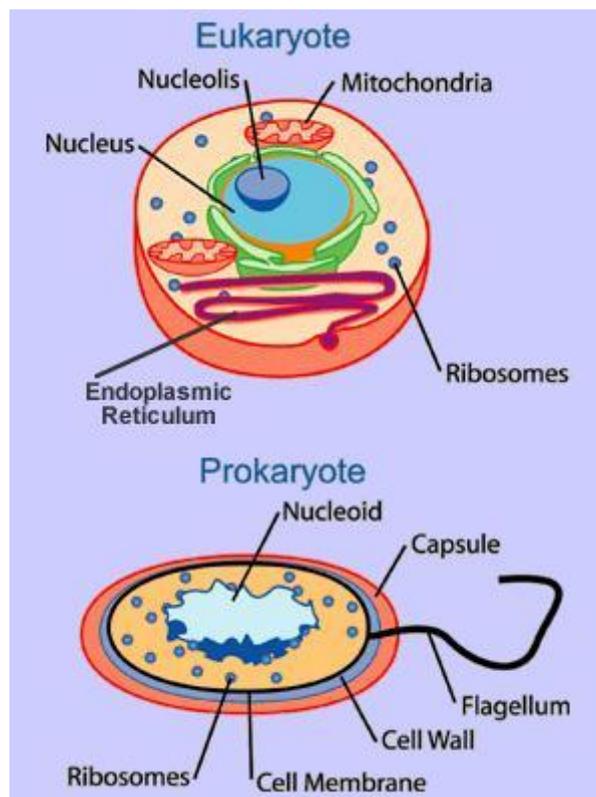
Fig.2.11. Structure of a Cell: bacteria, archaea (unicellular organisms) plants, animals (mostly multicelullar; also amoeba)
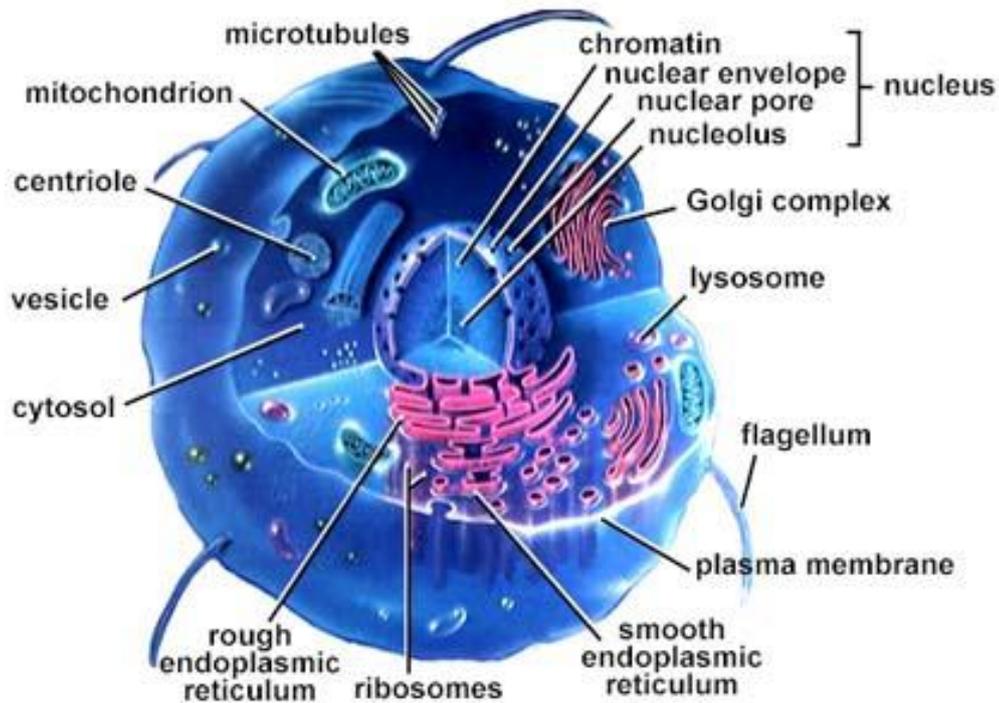
2.11. Cell structure

Major Cell Compartments Nucleus: approximately spherical membranebound organelle near the center of a cell; contains almost all of cell's genome; functions: gene expression (transcription of DNA to RNA to make proteins), DNA replication prior to cell division; surrounded by a double membrane called nuclear envelope (membrane); the outer membrane connected to rough ER.

Endoplasmic reticulum (ER): is a network of folded membranes with large surface to facilitate processes; rough ER hosts ribosomes, where the synthesis of proteins occurs; smooth ER contains lipid vesicles and is involved in lipid and steroid synthesis; ER also involved in adding carbohydrates to proteins, splicing and folding peptides, and packaging proteins into lipid vesicles for transport to other parts of the cell;

Golgi apparatus: similar to smooth ER (folded membrane); functions: processing and packaging of lipids and proteins, breakdown of carbohydrates and lipids;

Vesicles: small spherical bilayer containers, they fuse with or bud from the plasma membrane; lysosomes: vesicles with enzymes lysozymes to break down or digest larger molecules; peroxisomes: vesicles that break down long chain fatty acids;

Vacuoles: giant vesicles without a particular shape; functions: isolate harmfulobjects and waste products, help maintain correct hydrostatic pressure;

Ribosomes: large complex of proteins, enzymes, and ribosomal RNA (rRNA) found in both prokaryotes and eukaryotes; function: protein synthesis according to the sequence of messenger RNA (mRNA);

Mitohondria: membrane bound organelles, also contain DNA (mtDNA);

function: ATP (adenine triphosphate) synthesis, convert energy stored in food into high energy phosphate bonds of ATP;

Chloroplasts: organelles mostly found in plant cells (green parts), carry out photosynthesis (capture light and convert it into chemical bond energy of carbohydrates and ATP);

Cytoskeleton: interconnected tube or Ropelike fibrous structures made of proteins; function: to support, transmit, or apply forces, to preserve the shape of the cell and anchor various organelles in place; three types: microtubules, intermediate filaments, and microfilaments;

DNA: most significant structure inside the cell with genetic material organized in chromosomes: each chromosome is a single DNA molecule; sometimes DNA is organized into complexes with proteins; all chromosomes in a cell = cell's genome.
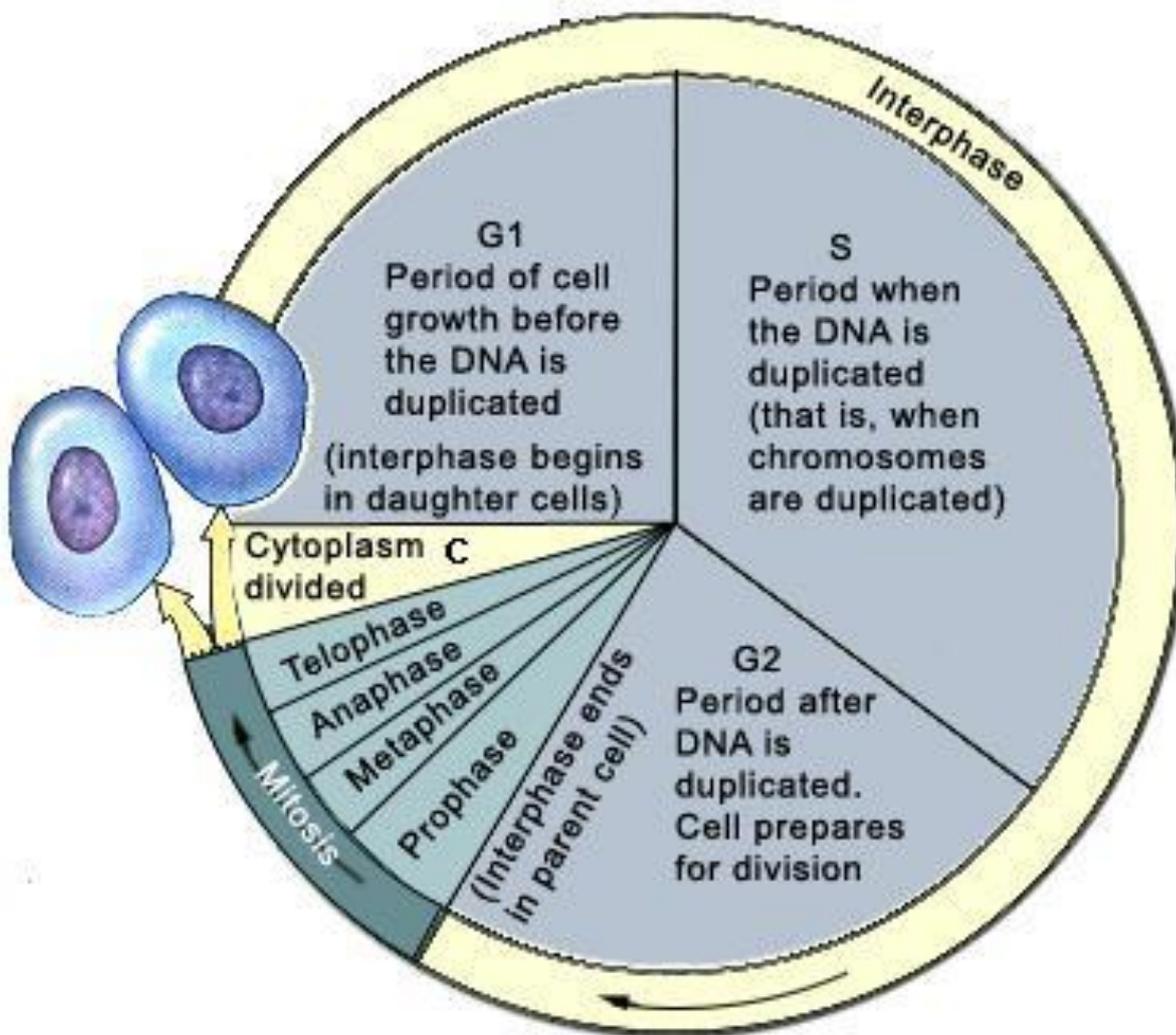
Fig.2.12. Life Cycle of an Eukaryotic Cell.

Main tasks of Biophysics:

- Create simplified models of biological systems

- Make quantitative predictions

- Experimentally test quantitative predictions.


Fig. 2.13. DNA models.

AMINO ACID SEQUENCE

...LFSIKREALSI...

HP MODEL

...HHPHPPPHHPH...

RIBBON DIAGRAM

COMPACT RANDOM WALK

RECEPTOR

TWO-STATE SYSTEM

Fig. 2.13. Protein models.

ARRAY OF SPRINGS

RANDOM SURFACE

RC CIRCUIT

SEMI-PERMEABLE BARRIER
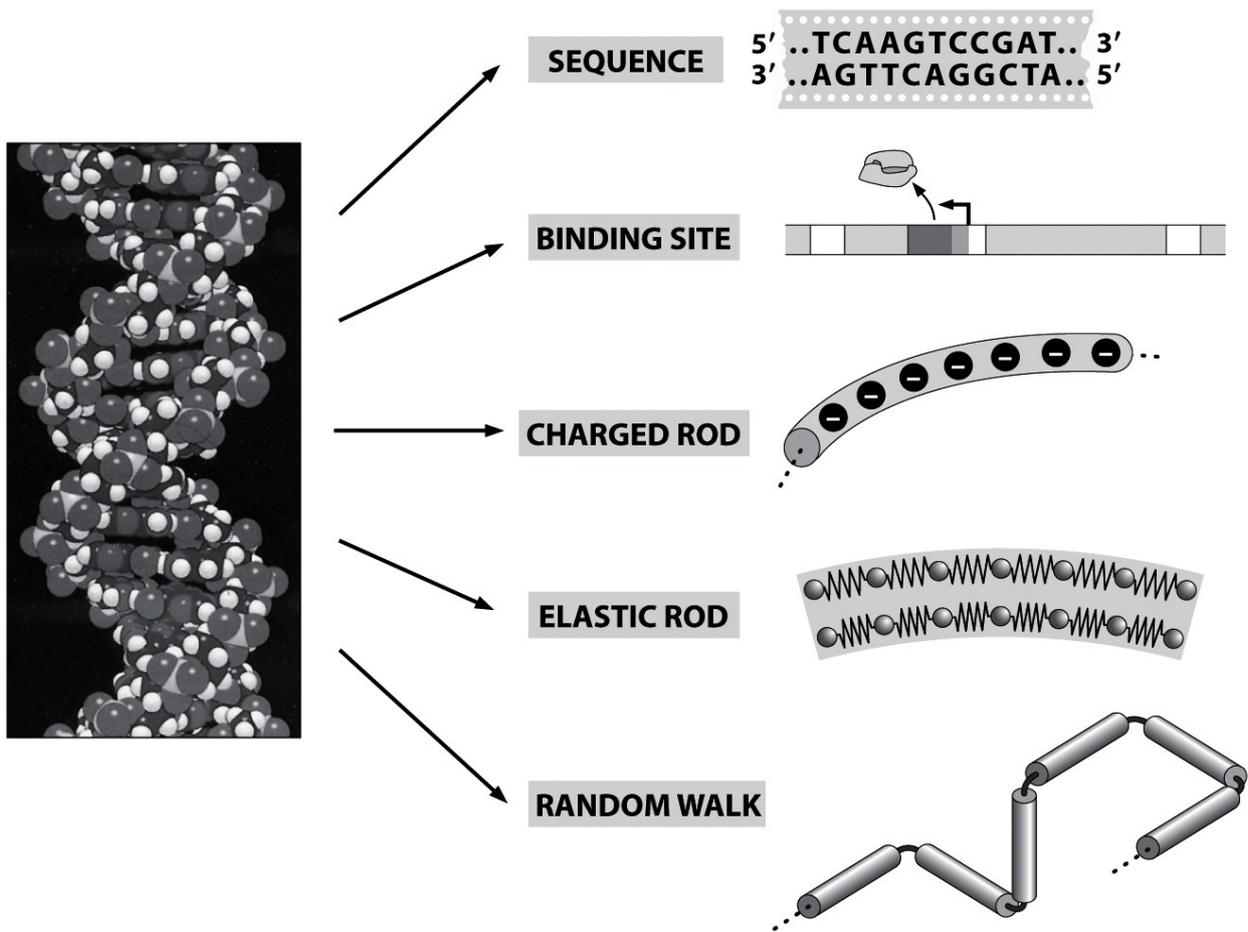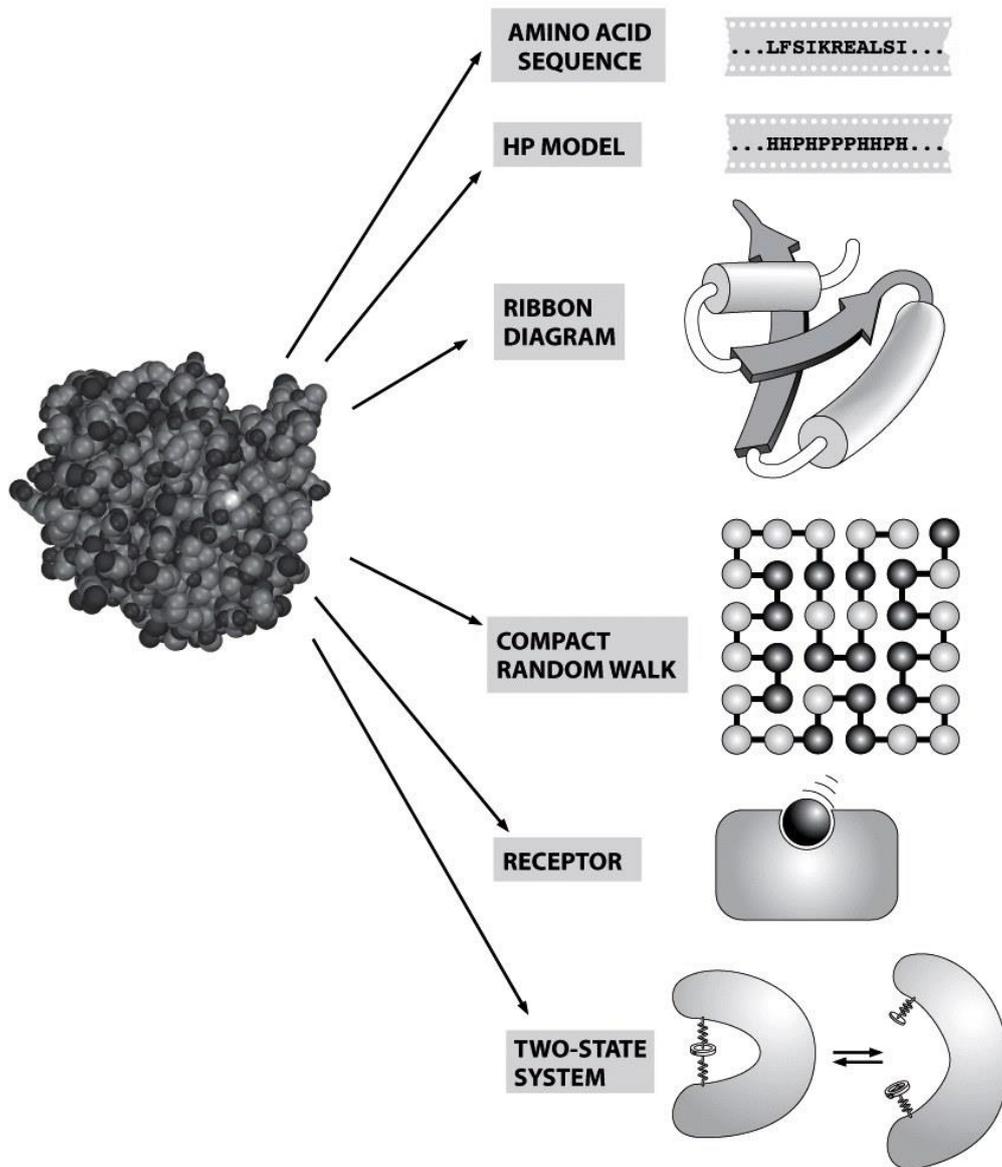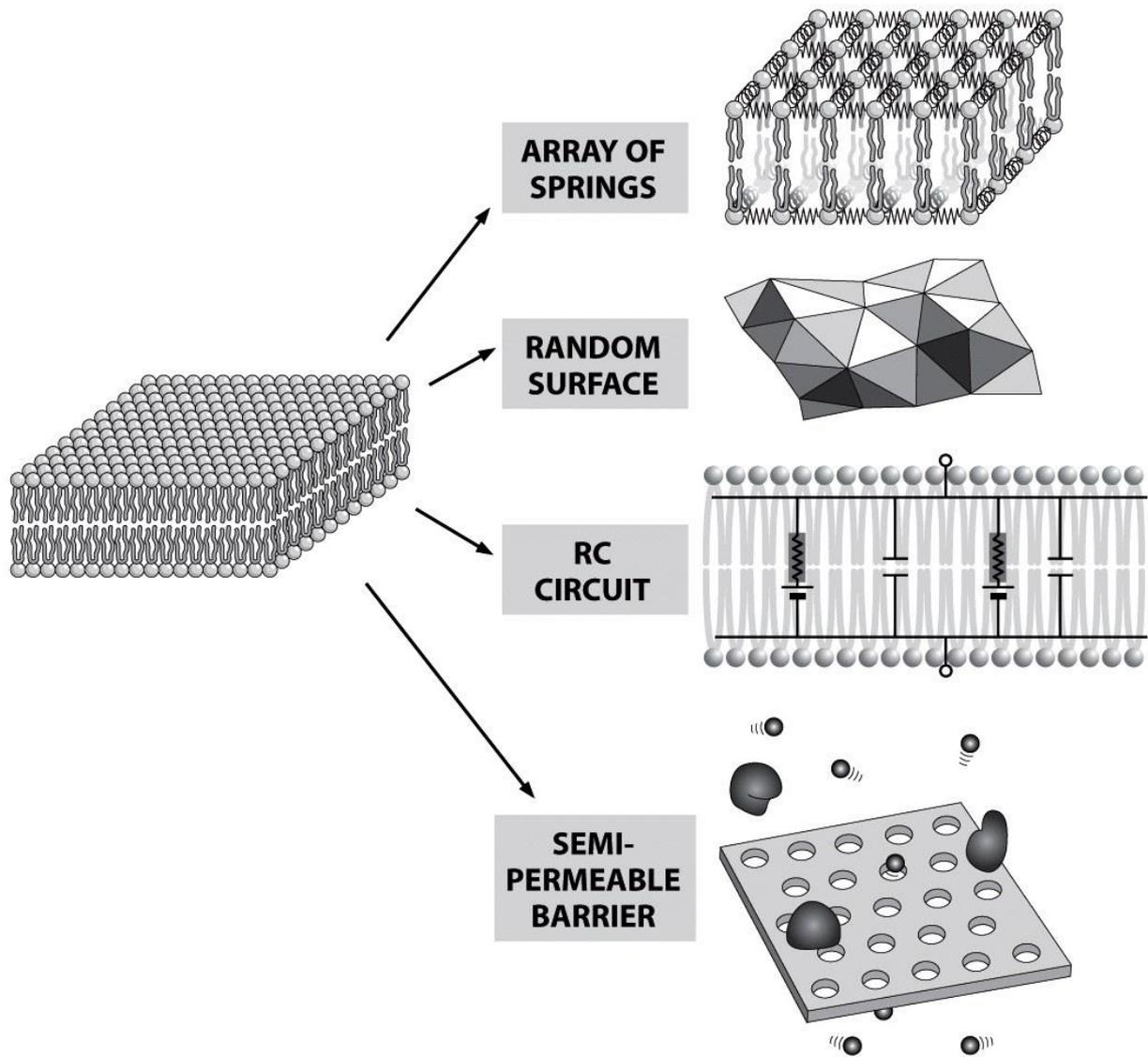
Fig. 2.14. Membrane model.

Chapter 3

Thermodynamics and Biosystems

The laws of thermodynamics are important unifying principles of biology. These principles govern the chemical processes (metabolism) in all biological organisms. The First Law of Thermodynamics, also known as the law of conservation of energy, states that energy can neither be created nor destroyed. It may change from one form to another, but the energy in a closed system remains constant. The Second Law of Thermodynamics states that when energy is transferred, there will be less energy available at the end of the transfer process than at the beginning. Due to entropy, which is the measure of disorder in a closed system, all of the available energy will not be useful to the organism. Entropy increases as energy is transferred. In addition to the laws of thermodynamics, the cell theory, gene theory, evolution, and homeostasis form the basic principles that are the foundation for the study of life.

First Law of Thermodynamics in Biological Systems

All biological organisms require energy to survive. In a closed system, such as the universe, this energy is not consumed but transformed from one form to another. Cells, for example, perform a number of important processes. These processes require energy. In photosynthesis, the energy is supplied by the sun. Light energy is absorbed by cells in plant leaves and converted to chemical energy. The chemical energy is stored in the form of glucose, which is used to form complex carbohydrates necessary to build plant mass. The energy stored in glucose can also be released through cellular respiration. This process allows plant and animal organisms to access the energy stored in carbohydrates, lipids, and other macromolecules through the production of ATP. This energy is needed to perform cell functions such as DNA replication, mitosis, meiosis, cell movement, endocytosis, exocytosis, and apoptosis.

Second Law of Thermodynamics in Biological Systems

As with other biological processes, the transfer of energy is not 100 percent efficient. In photosynthesis, for example, not all of the light energy is absorbed by the plant. Some energy is reflected and some is lost as heat. The loss of energy to the surrounding environment results in an increase of disorder or entropy. Unlike plants and other photosynthetic organisms, animals cannot generate energy directly from the sunlight. They must consume plants or other animal organisms for energy. The higher up an organism is on the food chain, the less available energy it receives from its food sources. Much of this energy is lost during metabolic processes performed by the producers and primary consumers that are eaten. Therefore,

much less energy is available for organisms at higher trophic levels. (Trophic levels are groups that help ecologists understand the specific role of all living things in the ecosystem.) The lower the available energy, the less number of organisms can be supported. This is why there are more producers than consumers in an ecosystem. Living systems require constant energy input to maintain their highly ordered state. Cells, for example, are highly ordered and have low entropy. In the process of maintaining this order, some energy is lost to the surroundings or transformed. While cells are ordered, the processes performed to maintain that order result in an increase in entropy in the cell's/organism's surroundings. The transfer of energy causes entropy in the universe to increase. Cells are open complex thermodynamic systems. Energy transformations, thermo-electro-chemical processes and transports occur across the cell's membranes. Different thermo-electro-biochemical behavior occurs between health and disease states. Moreover, living systems waste heat, the result of the internal irreversibility. This heat is dissipated into the environment. At the same time wasted heat represent a sort of information, which outflows from the cell toward its environment, completely accessible to any observer. Consequently, the analysis of irreversibility related to this wasted heat can represents a new approach to study the behavior of the cells. So, this approach allows us to consider the living systems as black boxes and analyze only the inflows and outflows and their changes in relation to the modification of the environment. Therefore, information on the systems can be obtained by analyzing the changes in the cell heat wasted in relation to external perturbations. In this lecture, a review of the recent state of the art by using this approach is proposed in order to highlight its thermodynamic fundamental: it could be the beginning of a new engineering science, the bioengineering thermodynamics.

Illustrations to the laws of thermodynamics in connection with biological substances

Laws of Thermodynamics

First Law: energy cannot be created or destroyed: $\Delta E = E_{in} - E_{out}$

$$\Delta U = Q - W$$

U … internal energy of a system, Q … heat the system receives from the environment, W … work done by the system.

Heat is treated separately from other forms of E: historically (heat engine: motor that converts heat into E). Results from kinetic energy and random motion of molecules.

Least organized form of energy: System is a part of the universe under study, the rest is surroundings

Units: calorie=energy needed to heat $1g\ H_2O$ from $14.5\ to\ 15.5\ C°$ at atmospheric pressure; $1\ cal = 4.184\ J$

Enthalpy H – Enthalpy (H) is the sum of the internal energy (U) and the product of pressure and volume (PV) given by the equation: $H = U + PV$. When a process occurs at constant pressure, the heat evolved (either released or absorbed) is equal to the change in enthalpy

Classic thermodynamic property of a system: $H = U + PV$

U … internal energy of a system, p … external pressure, V … volume of the system

Relevant changes: $\Delta H = \Delta(U + pV) = \Delta U + \Delta pV + p\Delta V$

or

$\Delta U = \Delta H - \Delta p V - p\Delta V = Q - W$, where $W = p\Box V$

(W … work done by the system: $W = \int F\, dx$) so $Q = \Delta H - V\Delta p$

The heat which is provided to the system is associated with the change in the enthalpy of a system.

The energy needed to change the pressure of the system: $\Delta H = Q + V\Delta p$

Most biological processes occur at $p = const.$, thus $\Delta H \approx Q$.

Entropy S is a thermodynamic property of all substances that is proportional to their degree of disorder. The greater the number of possible microstates for a system, the greater the disorder and the higher the **entropy**. Rudolf Clausius and heat machines to convert heat into work ($\sim 1\%$). Where is the remaining 99% of energy. If the change occurs at a constant T: $\Delta S = Q/T$ S … entropy of a system; T … absolute temperature of a system If the heat flows from $T1$ to $T2$ ($T1 > T2$): $\Delta S = Q/T2\ Q/T1$ S: trope [Greek] … transformation & energy $\rightarrow$ entropy S is a measure of disorder in a system and is associated with a number of different ways a system can be in one energy state:

- only a few ways to arrange molecules at a given temperature
- ordered system with a low S
- many possible ways to arrange molecules at a given temperature
- disordered system with a high S

Gibbs Free Energy G:

Gibbs used Clausius's definition of S to define available energy that can be converted into work: Free energy (as opposed to the energy lost through dissipation): $G = H - TS$ G … the energy that remains in the system after the energy losses due to dissipation are accounted for Gibbs energy change during a process occurring at constant T: $\Delta G = \Delta H - T\Delta S$

For a process that occurs spontaneously, $\Delta G < 0$, that is the final state has always a lower free energy than the initial state.

This principle is in a way analogous to $F = ma$ (Newton's 2nd law) as it drives all spontaneous processes, including chemical reactions.

The process that takes the system from state A to state B will occur spontaneously only if $\Delta G < 0$

The Second Law of Thermodynamics One way to state the second law is: In any process, when considering a system with its surroundings, the entropy will either remain constant or increase.

Life form:

1. Take energy from outside and create order

2. Propagate the order by having offspring

The Gibbs function principle $\Delta G < 0$ has the second law of TD build in.

- a system releases energy, $\Delta H < 0$, process enthalpy driven
- increasing disorder, $\Delta S > 0$, process entropy driven
- changing temperature, if $\Delta S > 0$, then $\Delta T > 0$ will result in $\Delta G < 0$;
- if $\Delta S < 0$, then $\Delta T < 0$ will result in $\Delta G < 0$

Example: When a certain protein binds to DNA the entropy decreases by $2\ kcal/K$. Atthe same time, the system releases $700\ kcal$ of enthalpy.
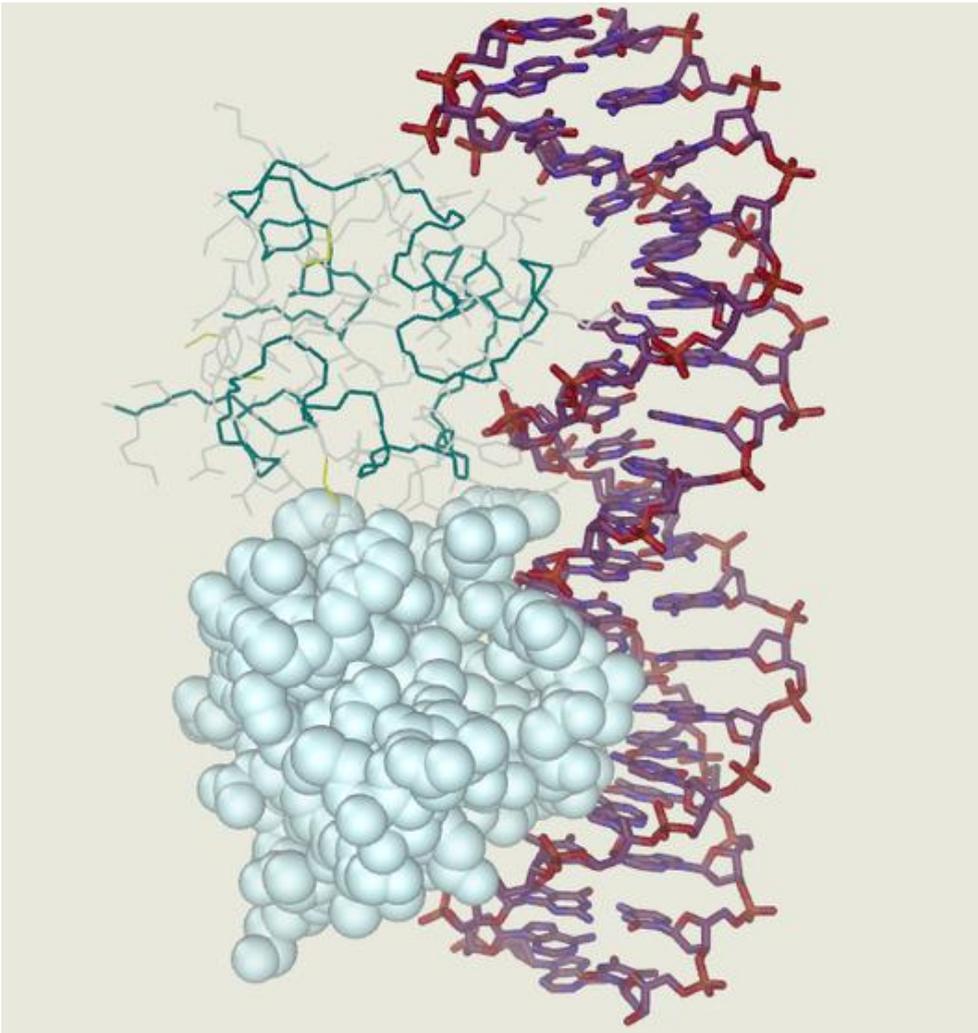
Fig.3.1. Process of protein binding to DNA.

Protein molecules binding to DNA initiate sequences of biochemical transitions that control and regulate all major processes in living cells. Contemporary theoretical views of protein-DNA interactions indicate that there are two main components of binding forces. One of them is purely electrostatic attraction between oppositely charged DNA and protein molecules that are mostly sequence-independent. It has been suggested that another contribution comes from particular DNA sequence motifs that strengthen the attraction of protein molecules

Chapter 4

Statistical Physics of Biosystems

Life is subject to the same laws of thermodynamics as other matter. Increasing entropy is commonly associated with decreasing order. Life takes place in regimes that are far from those studied by classical equilibrium thermodynamics, but also far from those captured by the past and current attempts to describe non equilibrium thermodynamics.

We should consider a way in which the statistical Entropy underpinning of the second law of thermodynamics can directly drive life and evolution towards order and complexity. We begin by recalling a number of potential confusions surrounding the relations between entropy, order, complexity, probability, and life, which often mislead the discussion on the statistical physics of life.

Formation of order and structure driven by entropy increase are ubiquitous in the universe: examples are cosmological formation of galaxies and stars from the initial uniform matter distribution. In these, and many other cases, elaborate structures form—not against statistics, but driven by the statistical logic of the second law. To explain this seemingly inexorable increase in order and complexity we point out some specific notions that can help in disentangling the complex relation between life and physics, in the regime far from thermodynamic equilibrium where life operates. Among these are the notions of macroscopic order providing a conduit for entropy to increase, metastable states, random motion within these states and channels among such states. This leads us to a perspective on the possible statistical underpinning of life, whereby life is not an improbable "fight against entropy"—as Erwin Schrödinger famously put it in his adventure into biology —but is rather a statistically favored process directly driven by entropy growth, in which movement of a system within a space of available states leads it to discover and traverse channels between metastable states. Here it is also necessary to underline the different roles that the notion of information plays in this context. The perspective we develop builds upon the numerous previous efforts towards understanding the statistical underpinning of life.

Much work has been done recently to analysis the dynamics of metabolic pathways, signaling networks, cell differentiation processes, etc. Using mathematical and computational models some trying to adopt chaos theory, concepts of attractors, order parameters etc. Transcriptomics has paved the way for a comprehensive understanding of how genes are expressed and interconnected. All these developments represent essential insights in the working of life and unravels concrete dynamical aspects of biological mechanisms.

Central to this discussion is the notion of order: that is, of correlations in time and space as a result of which particles or phenomena in the universe are not randomly dispersed, but rather have spatial or temporal structure. It is often assumed that any increase in order must imply a local decrease in entropy. The strict equivalence order = low entropy is a persistent prejudice that misleads efforts to understand the physics of

life. The key point is that macroscopic order can bring energy from macroscopic to microscopic variables, increasing entropy as it does so. This can happen not just when non equilibrium driving brings about novel patterns of complex organization in thermally fluctuating many body systems or 'active matter' mixtures, but also in simple isolated systems. There are familiar cases where an increase in entropy does amount to an increase in disorder. If system evolves freely the order is lost (increase of entropy generates disorder in some cases, but order in others).

There are very many examples in which entropy increase generates order and structure. They illustrate explicitly how an increase in entropy can lead to an increase in macroscopic order. This situation generalizes: for instance, it is entropy increase that separates air, water and rock on Earth, and is responsible for the existence of the sea, which stays above the rocks and below the atmosphere. Without dissipation, and hence entropy increase, water and air molecules would continue to mix. The common orderly arrangement on the Earth's surface, whereby the atmosphere is above the sea and the sea is above the rock, depends on the second law of thermodynamics. The conclusion is clear: entropy increase generates macroscopic disorder in some cases, but macroscopic order of various kinds in others. The relevance of this observation for biology is that the widespread idea that life is "a local fight against entropy", namely a trick to keep entropy locally low, is misleading. A classical reference for this misleading idea is Schrödinger famous observation: "The essential thing in metabolism is that the organism succeeds in freeing itself from all the entropy it cannot help producing while alive." In Schrödinger, E., "What is Life," 1944, chapter 6, Order, Disorder, and Entropy.

Life is driven directly by the second law, via a complex cascade through channels between metastable states opened by structure. The idea follows from the (correct) evidence of an important amount of order in life (life is a self-organizing process of increasing complexity) and from the prejudice that order and complexity necessarily imply low entropy. Since higher entropy is related to higher probability, this idea might lead to the misleading conclusion that life must be naturally improbable.

Since the second law of thermodynamics is the only fundamental law that distinguishes the past from the future, the physical basis of any irreversible phenomenon is this law, and nothing else currently known. Large-scale properties of the system, called "macroscopic variables", which can be numerous, but are still few in number compared to the number of degrees of freedom in the system. Entropy is a function of these variables that measures the number of states (or the volume of phase space—the space of possible states) where the variables have a given value. A "macroscopic state" is a set of values of these macroscopic variables, and it corresponds to an ensemble of "microscopic" states, whose size is measured by entropy. Notice that with this statistical definition (due to Boltzmann) the notion of entropy is defined for any value of the macroscopic variables, and not necessarily at equilibrium. The second law applies when in the past entropy is lower than its maximum. It states in particular that entropy increases in irreversible phenomena. The distinction between work and heat (two forms of energy), which lies at the foundation of classical

thermodynamics, is based on the macro/micro distinction as well: work is energy in macroscopic variables, while heat is kinetic energy in the microscopic variables. Importantly, macroscopic variables here are not just pressure, temperature and volume, but all the variables that enter in the description of a system such as those describing the position of distinguishable objects, the structure of an organism, the instantaneous chemical composition inside its cells, and so on. 'Macroscopic' variables can include, for instance, 'small' variables such as DNA nucleotides sequences, as we shall see below. Microscopic variables, on the other hand, are positions and velocities of all the individual molecules in the system. Energy available to do work is often called free energy, and is a measure of how much entropy the system can still gain.

All this drive towards maximal entropy is the 'reason' for all irreversible processes. This is what is meant by the metaphor of the system that "wants to increase its entropy". This is the logic underpinning the second law. It is not only relevant for the behavior of the systems at or near equilibrium states, which is quantitatively accounted for by equilibrium thermodynamics, but for any irreversible process in nature as well.

A biological system is typically an ensemble of atoms that include oxygen, hydrogen, carbon, and a few other elements. Naively, given so few elements one might think that the structure of the corresponding phase space should be relatively simple, but obviously this is not the case, in particular because of the extraordinary complexity of carbon chemistry, which arises in part from its ability to polymerize in all three dimensions. This opens the space for the extreme richness generated by combinatorics. The complexity of carbon chemistry is not a product of life: it is the aspect of the structure of the relevant physical state space that underpins it.

The dynamics of the interactions between organic chemicals is even more complex, as the interactions influence one another in all sort of manners. Therefore the structure of the relevant microscopic phase space is extremely complex: it is full of bubbles and channels between them. Life is a percolation among these channels. Not only is it far from equilibrium, it is also very far from having explored the full space of this complexity: for instance, only a minimal fraction of all possible proteins has been explored by life so far.

An important intuition arises from this model of phase space as a foam of quasi – independent bubbles. Consider that a complex system inhabits, in general, a larger bubble than a simple system, whereas a simple system inhabits a smaller bubble with fewer accessible states. It also follows that a larger bubble is harder to escape from the system. There is thus a time asymmetry in the progression through bubbles: a system will progress from smaller bubbles to larger ones more easily than the reverse. This means that, over time, a system will tend to become more complex rather than less. We can see the progression to increasing complexity as thus a consequence of the foamy structure of the phase space of complex systems, whereby ergodic wandering favours movement within a bubble more than movement between bubbles, and movement to larger bubbles more than movement to smaller ones.

Life is a process, not a state, and we can only hope to understand it as a system evolving in timerather than having a single state at a fixed time. As a state at a fixed time, it looks surprising, and it isonly by looking at it as a long term process that we can understand it.

This is a common pattern in science: the structure of the solar system or the structure of atoms were understood when instead of trying to figure out their instantaneous structure, scientists shifted their attention to their dynamics: the structure of the atom is understood by studying how its electrons move. Let us therefore think diachronically rather that synchronically; that is, let's consider the long term temporal evolution of the systems, rather than their instantaneous state.

We need two concepts. Instead of the phase space of the system, we should focus on the space of the motions of the system: namely the space of the possible histories (past and future) of the system. There is a strict relation between the two, because if we fix a time, each microstate at this time uniquely determines a motion and vice versa. We can extend the notion of coarse graining to the space of motions: macroscopic motions are families of ("microscopic") motions that are not distinguished by the macroscopic variables. Since we are using the second law of thermodynamics and assuming initial low entropy, we restrict our discussion to those motions that start in a suitable low-entropy region. We are interested in the properties of a generic macroscopic motion among these, with regard to order and structure formation.

The second notion we need is correlations in time rather than in space.

These can be defined as follows: Given a variable a that takes the value a(t), we say that there is order if there is a correlation in between $a(t_1)$ and $a(t_2)$, where t_1 and t_2 are different times. Life is then first and foremost characterized by a spectacular amount of correlation across time. Recalling that correlation is information, this can be expressed in informational terms: life is characterized by a spectacular amount of preservation of information across billions of years. One key holder of this information is of course the DNA molecule and the information it encodes. Using the precise definition of information - Shannon's "relative information" as physical correlation we can distinguish three distinct senses in which DNA molecules carry information: (i) Each single strand of a double-stranded DNA is the template of the other: given one, we can predict the other: hence it is correlated with it; hence it has information about it. The double strand has relative information because each strand has information about the other. This is key for reproduction. (ii) DNA encodes proteins and is therefore correlated with the proteins produced: in this sense, it has information about the proteins structuring the organism. (iii) What mostly concerns us here is the third sense in which DNA molecules carry information:

the entire molecule has information because it is reproduced across time - it has correlations exist in spite of the mortality-linked metastability of any single DNA molecule carrying that information, and are in fact to some extent a consequence of it. In order to understand correlation between DNA and the statistical underpinning of life We should recognize life as a phenomenon characterized by the following features: (i) Metabolism is a process that makes entropy grow, and as such it is directly entropically driven;

(ii) Metabolism would not happen if it were not for the biochemical structure of living matter: this structure provides channels in the complex phase space of carbon chemistry from metastable states to higher-entropy ones;

(iii) Inheritance is the process that allows the mechanics to be efficient in the long term, because total entropy continues to grow. This is possible thanks to long-interval correlations in time. These are given by the preservation of information, in particular in the DNA;

(iv) The structure that supports metabolism grows more complex over evolution because each complexity step opens new channels to higher entropy states. The densely packaged relative information in the DNA determines a structure which is produced by entropy growth, but is also capable of re-opening channels for further entropy growth. The main mechanism is of course the fact that the two strands of the DNA, which have information about each other, can separate, and an entropy-producing process can refurnish each strand with a new companion strand. The statistical physics of self-replication using the tools of non-equilibrium statistical mechanics says that this is an entropy-producing process that creates new structures capable of permitting a new entropy producing process by opening new channels for entropy to grow. As Francis Crick was once heard to say: "All life, sir, is DNA's way of making more DNA". Since the process can repeat, information is carried in time, and has been carried in time for some four billion years.

In this respect, a very simple model of life is a candle. The burning of a candle is a process with a certain amount of order and structure: the flame sustains itself via a self-regulatory (homeostatic) feed-back mechanism: if the flame is too strong it melts more wax, which suffocates it; if it fades, it consumes the melted wax, and liberates a new part of the wick, thus reinforcing itself. This is an elementary prototype of homeostasis. But the candle does not realize this by locally lowering entropy - it realizes this by being a channel for entropy to grow: a candle that is not lighted is in a metastable state and does not burn. Lighting the wick opens a channel that allows the (entropy growing) process which is burning to happen.

Statistical Mechanics provides a mathematical framework to explain TD quantities of a systemat the molecular level: For example: temperature is directly related to an average kinetic energy of atoms/molecules in the system. If we devise a model that defines all possible microscopic states of the system(e.g., 5 states of hemoglobin molecule: # of O atoms $0-4$); as sign energy differences between different microscopic states(preferably based on some experimental data); determine the number of different ways the molecules can distribute themselves among those energy states, which of the above distributions are most likely to occur(mathematics of probability and distribution) and knowing the distribution of energy and a probability of a molecule to be in each of the possible (allowed) energy states, we can derive allTDquantities.

EXAMPLE:

Our system has 4 molecules, each of which can have an energy that is a multiple of one energy unit E_o of 10^{-20}J (but not zero). Assume that the total energy of the system is $6\,E_o$
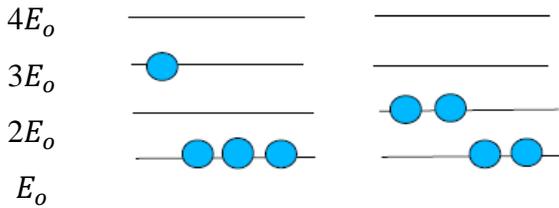
There are two states of the 4 molecule system with a total E of $6\,E_o$. Each of the two above distributions of molecules into the energy levels can be realized in several ways. How many permutations of molecules lead to each of the two distributions: left distribution: 4 permutations right distribution: 6 permutations. There are in total 10 permutations of molecules to be arranged into the energy levels, which are all equally likely to occur. Of the two distributions, the left one has thus 0.4 and the right one has 0.6 occurrence probability.

Consider the same example with a total energy of $7E_o$ (instead of $6E_o$).

How many distributions of 4 molecules into energy levels exist?
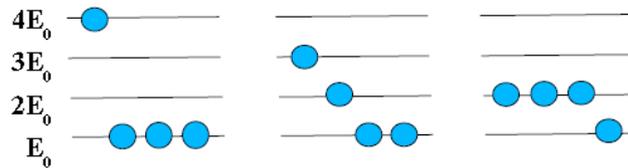
(i) (ii) (iii)

In how many ways (permutations) can molecules be arranged into each of

the three distributions: (i) 4; (ii) 12; (iii) 4

How do we calculate the total number of permutations, that is, different ways to arrange molecules in any given distribution:

$$W = \frac{N!}{n_1!\,n_2!\,n_3! \cdots n_L!}$$

N – total number of molecules

L – total number of energy levels

n – number of molecules on the energy level 1

$$N = \sum_{l-1}^{L} n_1$$

Remember: $0! = 1;\ 1! = 1;\ 2! = 2 \times 1 = 2;\ 3! = 3 \times 2 \times 1 = 6; \cdots$

What happens if we increase both the total energy and the number of Molecules: Consider 10 molecules with a total energy of $18E_o$:

(1) Determine the maximal number of energy levels L. A: 93

(2) Find all possible distributions of molecules into L energy levels. A: 22

(3) Calculate the number of permutations for each distribution.

9 1 1 1 1 1 1 1 1 1 (10) 5 2 2 2 2 1 1 1 1 1 (1260)

8 2 1 1 1 1 1 1 1 1 (90) 4 4 3 1 1 1 1 1 1 1 (360)

7 3 1 1 1 1 1 1 1 1 (90) 4 4 2 2 1 1 1 1 1 1 (1260)

7 2 2 1 1 1 1 1 1 1 (300) 4 3 3 2 1 1 1 1 1 1 (2520)

6 4 1 1 1 1 1 1 1 1 (90) 4 3 2 2 2 1 1 1 1 1 (5040)

6 3 2 1 1 1 1 1 1 1 (720) 4 2 2 2 2 2 1 1 1 1 (1260)

6 2 2 2 1 1 1 1 1 1 (840) 3 3 3 3 1 1 1 1 1 1 (210)

5 5 1 1 1 1 1 1 1 1 (45) 3 3 3 2 2 1 1 1 1 1 (2520)

5 4 2 1 1 1 1 1 1 1 (720) 3 3 2 2 2 2 1 1 1 1 (3150)

5 3 3 1 1 1 1 1 1 1 (360) 3 2 2 2 2 2 2 1 1 1 (840)

5 3 2 2 1 1 1 1 1 1 (2520) 2 2 2 2 2 2 2 2 1 1 (45)

All permutations are equally likely regardless of which distribution the ybelong to. If we have Adistribution with 100 permutations and distribution$B$ with 5 permutations, then it is 20times more likely that we will find thesystem in distribution A rather than in distribution B. The total number of permutations in our example is 24,310: (1) the least permutations (10) corresponds to the distribution9 1 1 1 1 1 1 1 1 1 (occurrence probability $p = 4.1 \times 10^4$) (2) the most permutations (5,040) corresponds to the distribution4 3 2 2 2 1 1 1 1 1 (occurrence probability $p = 0.21$)
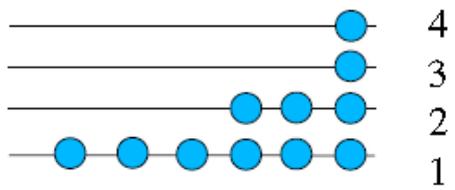
It could be shown that ignoring the 4 leastprobable distributions, the error is 1.5%. Ignoring 45% of all distributions, we would still capture the correctdistributions with 85% probability (15% error).

Ignoring 45% of all distributions, we would still capture the correct distributions with 85% probability (15% error).

# of Molecules [Total Energy]	# of Energy Levels	# of States (Distributions)	# of Permutations
4 [6 units]	3	2	10
4 [7 units]	4	3	20
10 [18 units]	9	22	24,310

Biophysical system: # of molecules at least 1,000 and up to 10^{20} !
For a fixed average energy per molecule, 1.8 E_0 per molecule, then:

# of Molecules	# of Levels	# of Distributions	% Ignored Dists [error < 5%]
5	5	9	14.0%
10	22	24,310	37.0%
20	231	4.1×10^9	71.0%
...			
100	15.8×10^6	1.7×10^{52}	99.9%

Boltzmann Distribution

In the limit $N \to \infty$, the most probable distribution is Boltzmanndistribution. We find it, if we can find ni for all energy levels i from 1 to L.

We assume that all energy levels between 1 and L are available and that L is the total number of available energy levels.

Mathematically, to find the most probable distribution means to maximize W.

The set of Ni that maximizes W is: $Ni = N/Z\ exp\ (\Delta Ei)$ and $Z = \Sigma\ exp\ (\Delta Ei)\ i = \{1, \ldots, L\}$

Things to remember about the Boltzmann distribution (derived in 1860s):

- the most probable distribution of N molecules over L energy levels
- for any reasonable # of molecules and total energy, the Boltzmann distribution overshadows all other distributions
- derivation of the functional form of the Boltzmann distribution uses Stirling's

approximation $[log(N!) \approx Nlog(N)]$, assuming N is a largenumber (at least 60).

Statistical Mechanical Calculations:

- The fraction of molecules, Fi, in a particular energy level: $Fi = Ni/N = exp\ (\beta Ei)/Z$ (is also a probability). The average energy: $<E> = \Sigma\ Ei\ Fi = 1/Z\ \Sigma\ Ei\ exp\ (\beta Ei)$.

Degeneracy of Energy Levels:

- atoms/molecules exist in different states (e.g. a particular conformation of a macromolecule, a mode of vibration, …)
- in many cases, each state (of a macromolecule) has a unique energy level
- sometimes, 2 or more distinct states (conformations, vibration modes)happen to have the

same energy level: how do we deal with such cases?

Distinct states occupying the same energy level are degenerate. Degeneracy of the E level = # of distinct states with this E level

Partition function Z can be then expressed as:$Z = \Sigma \acute{\omega} i \, exp(\beta \, Ei)$ with$\beta = 1/(kBT)$, where the sum of over all energy levels (not all states) and $\acute{\omega} i$ denotes thedegeneracy of the energy level i.

Reference State and Relative Energy:

- define a reference state E_{ref} (E_{ref}does not need to be known)
- define all energy levels relative to E_{ref} (in most cases the lowest energystate = ground state)
- example: double helix state of DNA as a reference state, partially orfully unwound states that are relevant to the situation under study defined with respect to this reference state

$$\Delta E_i = E_i - E_{ref}$$

$$Z = \omega_{ref} + \sum \omega_i exp(-\beta \Delta E_i)$$

Here we considered a possibility of a reference state to be degenerate.

Biophysics with Gibbs free energy:

$$Z = 1 + \sum \omega_i exp(-\beta \Delta G_i)$$

Chapter 5

Mechanical and Chemical Equilibrium in Biosystems

A system in dynamic equilibrium will have small changes that sum together to produce no net change. Many biological systems are in dynamic equilibrium, from the water inside cell to the dynamic equilibrium experienced by populations of predators and prey. Dynamic equilibrium is different from a static equilibrium, in which the parts do not move once they've reached equilibrium. Dynamic equilibrium has different meanings in each science sub-discipline, such as biochemistry or ecology. In chemistry, the equilibrium of a reaction is the point at which the products and reactants have the lowest free energy. The dynamic equilibrium, on the other hand, is the point at which products are being generated as fast as they are falling apart. This may not be the same as the chemical equilibrium, as enzymes force many reactions far past their natural equilibriums by making products faster than they break apart. Due to this, we often refer to dynamic equilibrium as a dynamic steady-state, to clearly differentiate between the two points in a reaction.

Ecologists and biologists will often refer to dynamic equilibrium when talking about populations of organisms. When studying the number of organisms in a population over time, many factors affect the growth of a population. Often, populations will go through periods of boom and bust. Ample resources cause high reproduction rates in all animals, leading to a much higher population. When the resources are distributed among this higher number, there are not nearly enough resources to go around. Thus, the population dies off. Ecologists see these cycles as a dynamic equilibrium that the population is stuck in, never really gaining or losing large numbers of individuals.

As one of the examples of Dynamic Equilibrium we can introduce Glucose in an Organism.

Throughout our entire lifetime, the glucose levels in our body remain relatively the same. Over the course of a day however, our bodies use enormous amounts of glucose and must replace it. Each cell in our body requires glucose to function. As the cells use this glucose, the liver and our digestive system work quickly to replace it. Glucose from the food you eat is moved from the stomach and intestines into the bloodstream. The liver stores glucose as glycogen, and must break this large molecule down to release glucose into the blood. In our body, glucose is in dynamic equilibrium. While glucose has periods of high and low concentration, it is relatively stable. If glucose levels in our body fall out of dynamic equilibrium, or we cannot replace the glucose we use, we would eventually die.

Ecologists often study the relationships between multiple species and their effects on each other. One relationship in nature that often shows dynamic equilibrium is the predatory-prey dynamic. Imagine a nature reserve that only contains rabbits and wolves. As the rabbit population increases, it provides more food for the wolf population. This sets both populations into dynamic equilibrium. The wolves, reaping the benefits of the increased rabbit population, also start to reproduce more. After a period of time, the wolf

population also starts to increase dramatically. As more wolves are born and eat the rabbits, their populations eventually level off. The wolves, still reproducing at high levels, eventually start decreasing the rabbit population, which cannot keep up. The rabbits decrease, and eventually the wolves are left without enough food to support a large population. This dynamic equilibrium of both populations is interesting because it shows a direct cause and effect relationship between different species in an ecosystem. The Static Equilibrium – When a system reaches a point of stability in which no parts are still moving; Equilibrium – A point in a reaction in which the lowest free energy exists on both sides of a chemical equation; Free Energy – The energy in a system capable of causing a reaction.

Our understanding of the evolutionary dynamics of living systems on all scales has lately developed from one based solely on gradualist Darwinian evolution through natural selection at the individual level to one embracing punctuated equilibrium, a scenario of large periods of stasis punctuated by episodic evolutionary change, with selection acting not only at the individual, but also at gene, species, and possibly higher levels. Stasis, once considered as an uninteresting triviality, now forms an important focus of evolutionary study at all levels of the hierarchy of life on Earth. In fact, paleontologists and ecologists, impressed by the ubiquity of stasis have argued for the search of an active force of stabilization. Stasis and punctuation of stasis is perhaps no better apparent than at the level of ecosystems. It is known that from their inception ecosystems go through a series of successional stages, each stage generally being more diverse, complex, and more stable than the previous one. We also know that the jump between successional stages occurs in a relatively short time span, and that most of the time, most ecosystems may be found in stasis, or in what is generally referred to as ecological steady states. In these states, species populations are either fixed or oscillate regularly, or perhaps even chaotically, but always about some fixed point in population space which is surprisingly stable to external perturbations. Every so often, however, rapid extinctions and speciation give rise to succession, instigated perhaps by either a critical change in the external conditions or by intrusion of a new species into the ecosystem. The lack of ''missing links'' between species, and between successive ecosystems, in the fossil record is an empirical fact, now taken as evidence of stasis punctuated by episodic change, prevalent at all levels of living systems and to the earliest times of life on Earth. At the species level, punctuated equilibrium may be described from within Darwinian theory of selection of the individual by allowing for Mayr's theory of allopatric speciation. Small populations of a particular species which become isolated geographically or otherwise from the main population are no longer subjected to dilution of their gene pool by the larger parent population and thus have an opportunity to evolve rapidly, perhaps forming new species. If such a new species becomes repatriated with the parent species and has some particular advantage over it in the same environment, then the new species may competitively cause the extinction of the parent species. Links between the two species are missing in the fossil record simply because rapid evolution occurred on a small population which was also geographically limited. Going up the hierarchy of living systems however, it becomes increasingly difficult to explain the

macroevolutionary dynamics of stasis and punctuation from within Darwinian theory. This because the individual units become further and further removed from the traditional Darwinian objects of selection and reproduction, their numbers dwindle so competition loose significance, and an appropriate target of selection becomes elusive. At these scales, the macro-evolutionary dynamics of living systems is thus an enigma, indicating a need for a more encompassing theory, one which might be effective at all levels of the hierarchy of living systems.

A more encompassing framework might be grounded in non-equilibrium thermodynamic theory for a number of reasons:

(1) Thermodynamic laws are the most universal of all laws and they work on all scales in similar ways, allowing for a unified hierarchical description.

(2) The study of the macroscopic behavior of whatever complex system benefits from a reduction in the number of variables to a smaller number of effective variables. Such a reduction is missing in traditional ecological theory and has led to an impasse in accounting for macro-evolutionary patterns. Thermodynamics, on the other hand, was developed in the physical sciences specifically out of this need to find a reduced number of relevant variables to describe macroscopic phenomena.

(3) Stasis and punctuation have intriguing analogues in the form of non-equilibrium thermodynamic stationary states and phase transitions.

(4) The problem of an elusive target of selection at higher than the species level, or, more specifically, the problem of the evolution of a system of a population of one, is solved because it can be reduced to a number of thermodynamic directives involving the entropy production. A shift in ecosystem analysis from a descriptive paradigm to one based on physical laws began with the seminal work concerning the flow of energy through an ecosystem. The possibility of framing ecology within a quantitative non-equilibrium thermodynamic paradigm, however, was first recognized by E. Schrodinger who pointed out that living systems were under the dictates of thermodynamic law and that biological structure and processes were maintained by a continual in-flow of negative entropy, at the expense of an entropy increase of the environment. Apart from developing the physical and mathematical ground work for the description of non-equilibrium phenomena, I. Prigogine has emphasized the remarkable similarity in characteristics that living systems share with thermodynamic stationary, non-equilibrium states.

Directives may be the basis of the active agent promoting stasis in ecosystems, which is nontrivial problem in the traditional ecological framework since a simple mathematical analysis shows that any complex interacting system, whether mechanical, chemical, or biological, will have little chance of being stable unless the interaction strengths between its component parts are very carefully chosen and continually maintained. A biological cause of such stabilization, for example through natural selection at the ecosystem level, however, remains elusive, leading to a stubborn complexity. The irreversible evolution toward the stationary state is an empirical fact for all abiotic systems under constant external constraints. It is thus

suggested that the ecological steady state, or periods of stasis in ecosystems, is a particular manifestation of the thermodynamic stationary state. Furthermore, it will be shown that evolution of the species interaction coefficients leading up to this period is driven necessarily in the direction of securing and maintaining global stability by thermodynamic restrictions on the entropy production.

These non-equilibrium ''phase transitions'', induced by changes through a critical point in the external conditions, may play an important part in episodic punctuation which, together with stasis, define the macro-evolutionary dynamics of the ecosystems toward larger, more complex and apparently more stable systems. The description of the stability and dynamics of ecosystems provides a convincing resolution of the complexity–stability paradox.

The traditional ecological framework is based on the empirical Lotka–Volterra-type equation

$$dp(t)/dt \,=\, Fg\,\{p_1(t) + p_2(t) + \cdots \ldots p_n(t)\} \tag{5.1}$$

where Fg is some empirically inspired, nonlinear function of the populations Pg of the n species The population dynamics and stability in the neighborhood of the fixed point can be determined by expanding Eq. (5.1). it is obvious that asymptotic stability near the steady state requires that the real parts of all the eigenvalues must be negative.

Consequently, the probability that a randomly constructed community will be stable decreases rapidly with the size of the ecosystem, becoming practically zero at an ecosystem size of only about 10 strongly interacting species. The most plausible mechanism from within the ecological framework thus far offered for tuning the parameters of the community matrix is natural selection. However, this explanation remains incomplete until the question of how a stable ecosystem could be the target of evolution through natural selection is addressed. In other words, it is a version of the conceptual problem of natural selection working on the evolution of a system of a population of one ecosystem. Before applying classical irreversible thermodynamic (CIT) formalism to ecosystems we acknowledge the general conditions under which CIT theory is valid and thus identify which ecosystems may be justifiably treated through CIT theory. The classical theory of irreversible thermodynamics is the non-equilibrium thermodynamic theory which has been the most empirically tested and universally accepted. The limitations of the theory have been discussed by I. Prigogine. Generally, the classical theory can be applied to any system for which it can be shown that the Gibbs relation holds locally. For transport processes it has been demonstrated, through a statistical-mechanics approach, that this corresponds to the requirement of linear phenomenological laws between the generalized forces and flows. However, the phenomenological coefficients may still be functions of the state variables. For chemical reactions, it is only required that the reaction rates are low enough to maintain a Maxwellian distribution of the velocities of each reacting component. Although these conditions may appear restrictive, they have, in fact, been shown, both experimentally and theoretically, to apply to a wide range of real phenomena, particularly to those to be considered here involving transport processes. At the hierarchical level at which the unit of entropy production and transport within the ecosystem is the

individual, the justification for the applicability of linear CIT is shown in the appendix to be obtained by limiting the analysis to interactions between individuals of the one- and two body form only. In this case, the relations between the generalized forces and the generalized flows (the species populations and the entropy flows, respectively, see below) are indeed linear.

Coupling of irreversible processes occurring within an ecosystem requires that the processes occur within the same ''macroscopic'' region. The scale of the macroscopic region being determined by the range of the forces of interaction between ecosystem components. Since, for example, metabolic rates of herbivores can be influenced by the mere sight or smell of a predator at distances of up to kilometers, or days after passage, it is reasonable to presume that the macroscopic space–time region available for coupling of irreversible processes within an ecosystem can be quite large. Stationary state coupling of irreversible processes is a further possibility. which shows that the change in entropy production due to changes in the generalized forces is always negative, independent of the systems closeness to equilibrium.

Quick Look on Mechanical and Chemical Equilibrium in Biosystems trough some examples

Chemical and Mechanical Aspect of Entropy-Exergy Relationship: the chemical potential is generated by electromagnetic interactions among atoms and molecules of a system and constitutes the internal potential energy as a component of the generalized internal energy and is defined as a form of distribution of energy among all elemental particles constituting the system, as kinetic energy with respect to thermal internal energy.

The maximum theoretical net useful work, according to the canonical definition of energy related to a reversible (internally and externally) chemical reaction process, is expressed by the Gibbs function $W = -\Delta G$.

The generalized state equation remains valid in the setting of the Stable-Equilibrium State Principle in turn overarching the Lowest-Generalized-Energy Principle and the Highest-Generalized Entropy Principle. Lowest-Energy Principle and Highest-Entropy Principle represent the paradigm underpinning reversible and irreversible contributions to be accounted for in the overall balance for the optimization of processes and the design of systems by means of extreme principles with global and local approaches.

This paradigm is obviously valid for processes occurring in isolated systems and in closed or open system experiencing irreversible processes. In isolated systems, energy remains constant and generalized entropy increases. In closed or open systems, entropy can remain constant under the condition that thermal, chemical or mechanical energy necessarily increases as it enters the system to compensate the entropy increase. Reduction of entropy determined by the increase of temperature or chemical potential implying increase of thermal energy or chemical energy or decrease of specific volume implying increase of mechanical energy (or increase of density).

The Central Dogma of molecular biology is an explanation of the flow of genetic information within a biological system. It is often stated as "DNA makes RNA, and RNA makes protein".
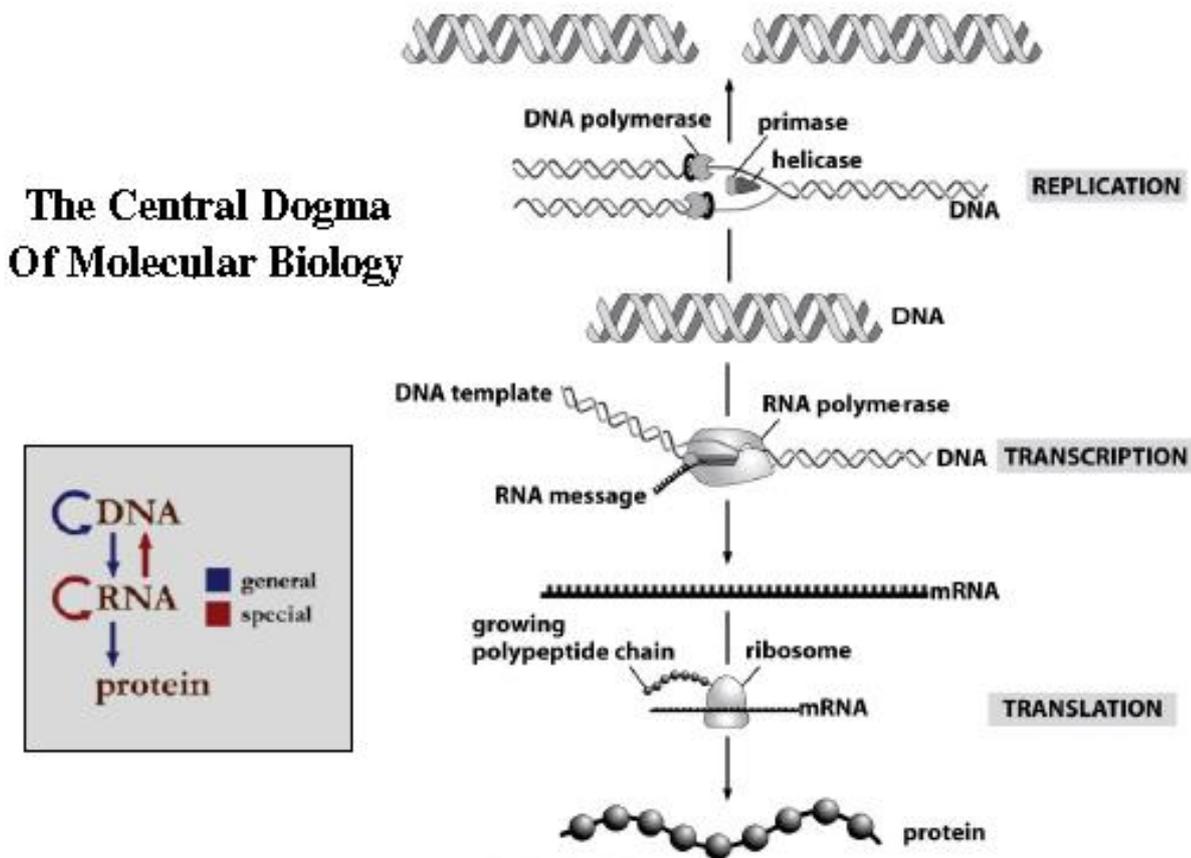
Fig. 5.1 The main scheme of molecular biology. Source: Physical Biology of cell. Garland Sciences 2009.
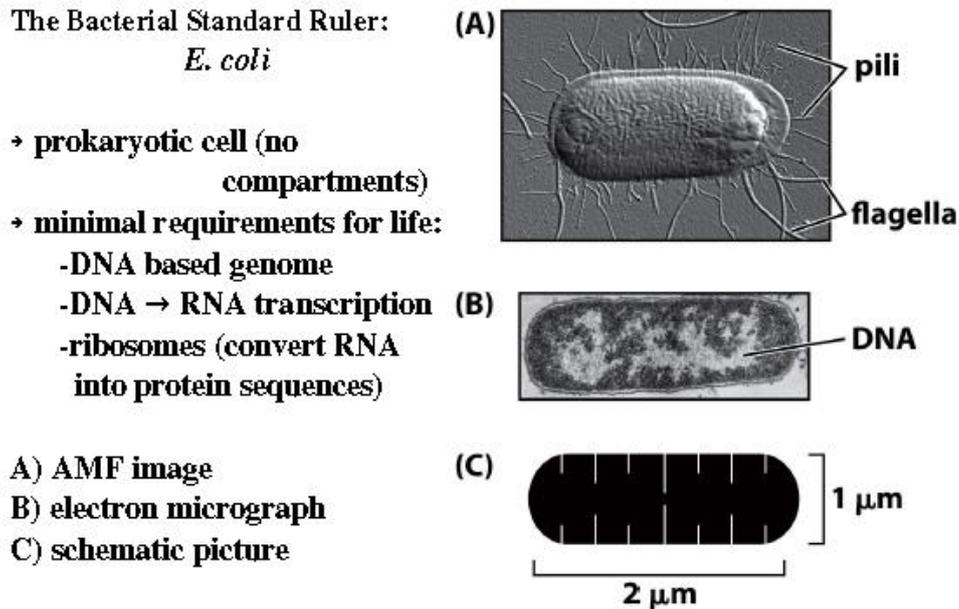
Fig. 5.2. The main actors of molecular biology. Source: Physical Biology of cell. Garland Sciences 2009.

What is the *E. coli's* intracellular environment like? crowded with many macromolecules

Figure 2.2 Physical Biology of the Cell (© Garland Science 2009)
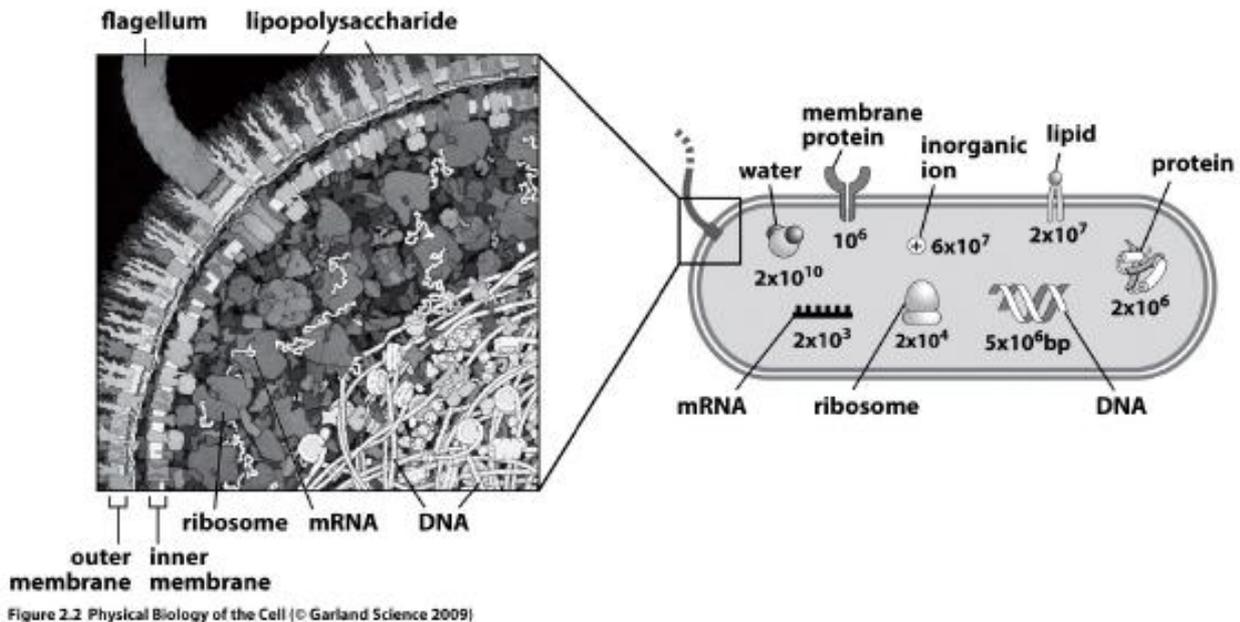
Fig. 5.3. The intracellular environment. Source: Physical Biology of cell. Garland Sciences 2009.

Molecular census on *E. coli* (has 4 lipid layers):

→ volume: $1 fL = 10 - 15 L$; mass: $1 pg = 10 - 12\ g$; density: $1 g/mL$ (H_2O)

→ dry weight of the cell: ~30% of its total ($0.30 pg$); half of dry weight is protein ($0.15\ pg$)

→ half of dry mass comes from the carbon content of E. coli, so there is ~10^{10} carbon atoms in a cell

→ number of proteins: assume each protein ~300 amino acids and each amino acid $100\ Da$ ($30\ kDa$; $1\ Da = 1.6 \times 1024 g$), thus: 3×106 proteins (1/3 of these within he membrane, 2/3 inside)

→ number of ribosomes: the mass of each ribosome $2.5\ Mda$; each ribosome consists of 1/3 of protein and 2/3 of RNA; 20% of all proteins in the cell resides in ribosomes: 20,000 ribosomes (Total ribosomal protein mass/protein mass inside on ribosome)

→ the diameter of ribosome is $20\ nm$: 10% cell volume

Coli cell: Macromolecular census

Table 5.1. Observed macromolecular census of an coli cell. (Source: F.C. Neidhardt et al., Physiology of Bacterial Cell. Sunderland Sinauer Associates inc. 1990 and M. Shchaechter et.al., Macrobes Washington DC. ASM Press, 2006.)

Substance	% of total dry weight	Number of molecules
Macromolecule		
Protein	55.0	2.4×10^6
RNA	20.4	
23S RNA	10.6	19,000
16S RNA	5.5	19,000
5S RNA	0.4	19,000
Transfer RNA (4S)	2.9	200,000
Messenger RNA	0.8	1,400
Phospholipid	9.1	22×10^6
Lipopolysaccharide	3.4	1.2×10^6
DNA	3.1	2
Murein	2.5	1
Glycogen	2.5	4,360
Total macromolecules	**96.1**	
Small molecules		
Metabolites, building blocks, etc.	2.9	
Inorganic ions	1.0	
Total small molecules	**3.9**	

(A) the protist *Giardia lamblia*
(B) a plant cell
(C) yeast cell (with a bud)
(D) red blood cell
(E) a fibroblast cell
(F) a nerve cell
(G) retinal rod cell
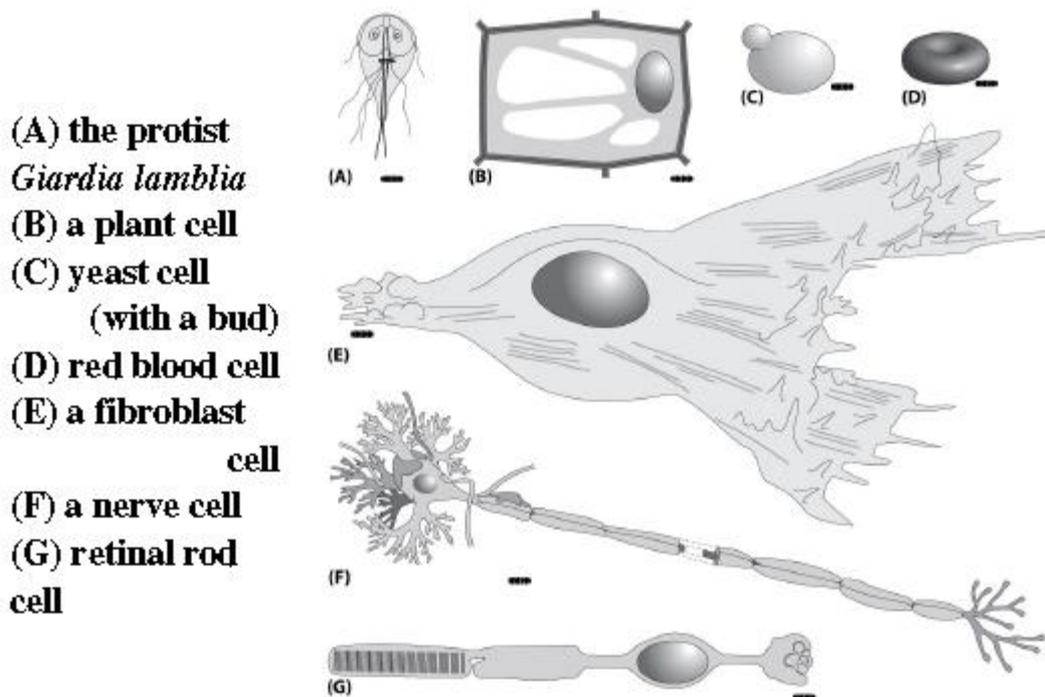
Fig. 5.4. The well- known cells. Source: Physical Biology of cell. Garland Sciences 2009.

Molecular census on a yeast cell:

$\rightarrow$ $E.\ coli$ volume: $VEColi = 1.0\ \mu m3 (1 \times 2\ \mu m)$

$\rightarrow$ Y_{east}: a sphere of diameter $5\ \mu m$: $V_{east} = 65\ \mu m3 \sim 60\ V_E.Coli$;surface area $A_{yeast} \sim 80\ \mu m2$; yeast nucleus a sphere of diameter $2\ \mu m$ and volume $\sim 4 \mu m3\ with\ 1.2 \times 107$ base pairs (bp) of yeastgenome (16 chromosomes).

$\rightarrow$ DNA packed into nucleosomes (histone DNA complexes): 150 bpwrapped around a cylindrical core, histone octamer (radius $3.5\ nm$, height $6\ nm$, volume $230\ nm3$), with $50\ bp$ spacers:$N_{nucleosomes} \sim 60{,}000$. (Exp. 80,000 nucleosomes with a mean spacing of $\sim 170\ bp$)

$\rightarrow$ volume of one $bp \sim 1\ nm3$ & volume of all histones: $14 \times 106\ nm3$

$\rightarrow$ the genomic DNA packing fraction$\rho_{pack} \sim 3 \times 10^{-3}$.

Chemical, mechanical, electromagnetic, thermal versus length scale
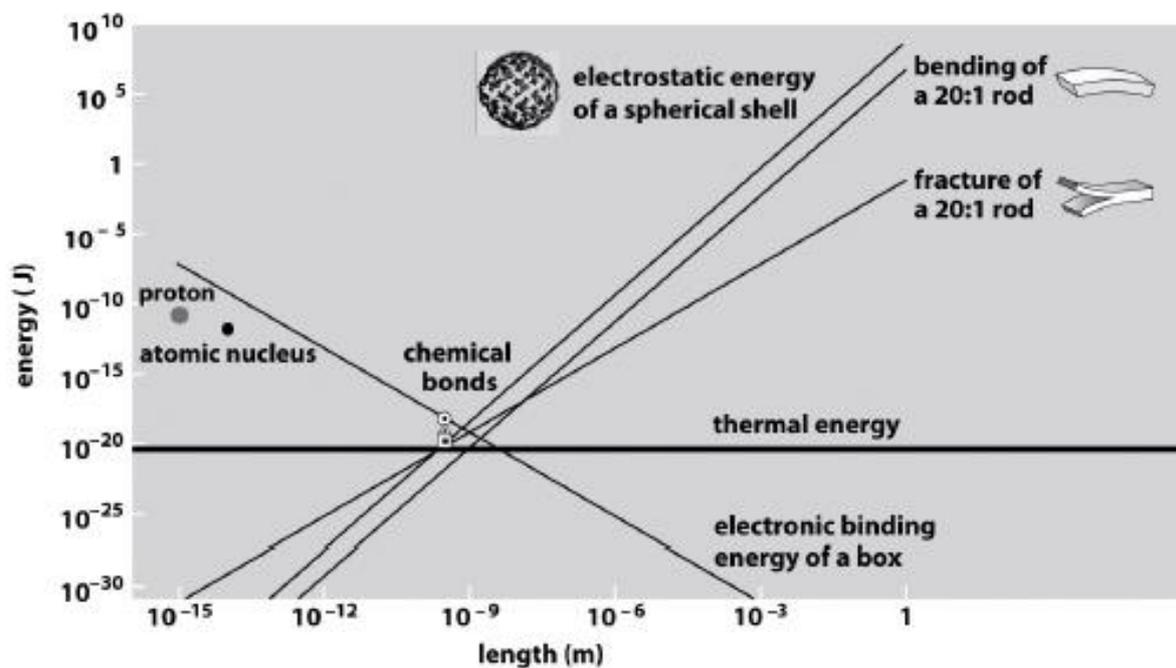
Fig. 5.5. Correlation between energy and length. Source: Physical Biology of cell. Garland Sciences 2009.

Thermal energy at room temperature $T = 300\ K$: $kBT = 1.38 \times 10^{-23}$
$J/K \times 300\ K = 4.1\ pN\ nm = 25\ meV = 2.5\ kJ/mol = 0.6\ kcal/mol$
(Avogadro's number: $N_A = 6.022 \times 10^{23}$)
Brownian (thermal) motion: important for nm to μm length scales. For macromolecules (DNA, proteins, lipids, and carbohydrates) $\sim nm$ Scale: thermal energy $\sim$ energy needed for intramolecular rearrangement.

Metabolic breakdown of Glucose in a glycolysis pathway

metabolism ... cellular transformation of one molecule into another

needs input energy (ATP)

produces more energy (ATP, NADH)

end product: two molecules of pyruvate
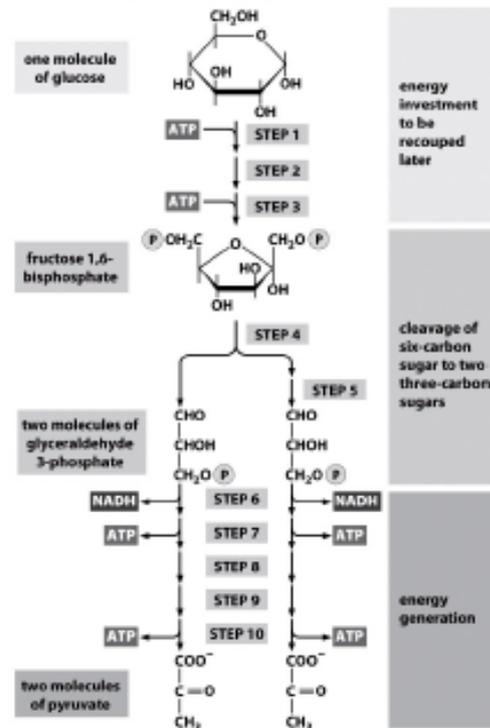
Fig. 5.6. The scheme of metabolic breakdown of glucose. Source: Physical Biology of cell. Garland Sciences 2009.

ATP conversion to ADP (releases – 20 KbT energy: unit of energy in the cell process
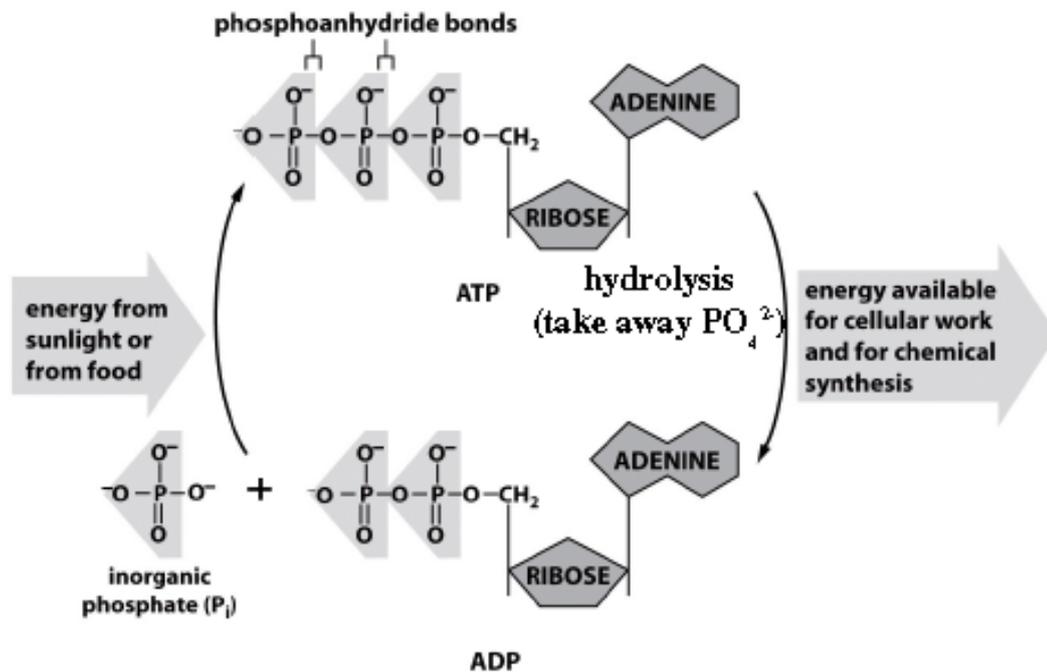
Fig. 5.7. ATP –ADP transformation mechanism. Source: Physical Biology of cell. Garland Sciences 2009.

Transferable electrons on NADP and NADPH

NADPH gives up its hydrate ion and liberates energy
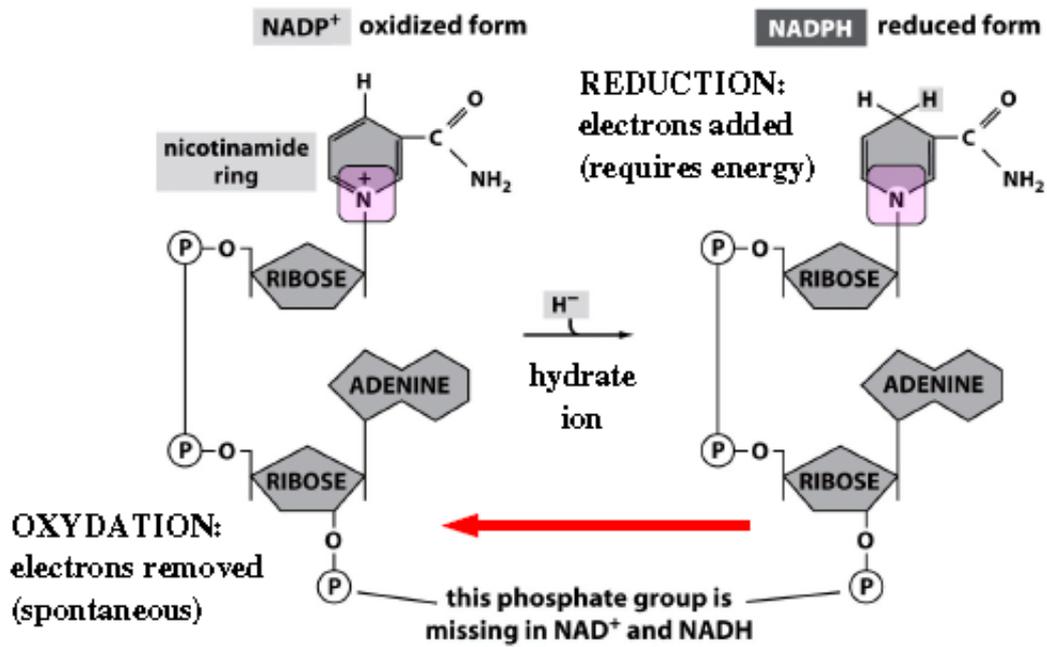
Fig. 5.8. The electrons transferring scheme in the main biomolecular systems. Source: Physical Biology of cell. Garland Sciences 2009.

Create H+ gradients across membrane (can be converted to ATP or NADH energy)
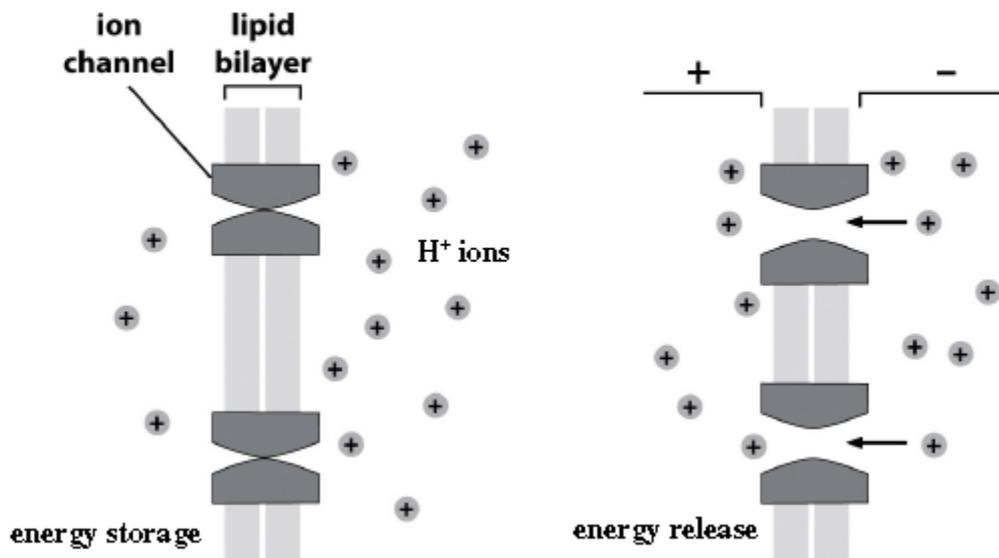
Fig. 5.9. The charge gradient in energy conversion process.. Source: Physical Biology of cell. Garland Sciences 2009.
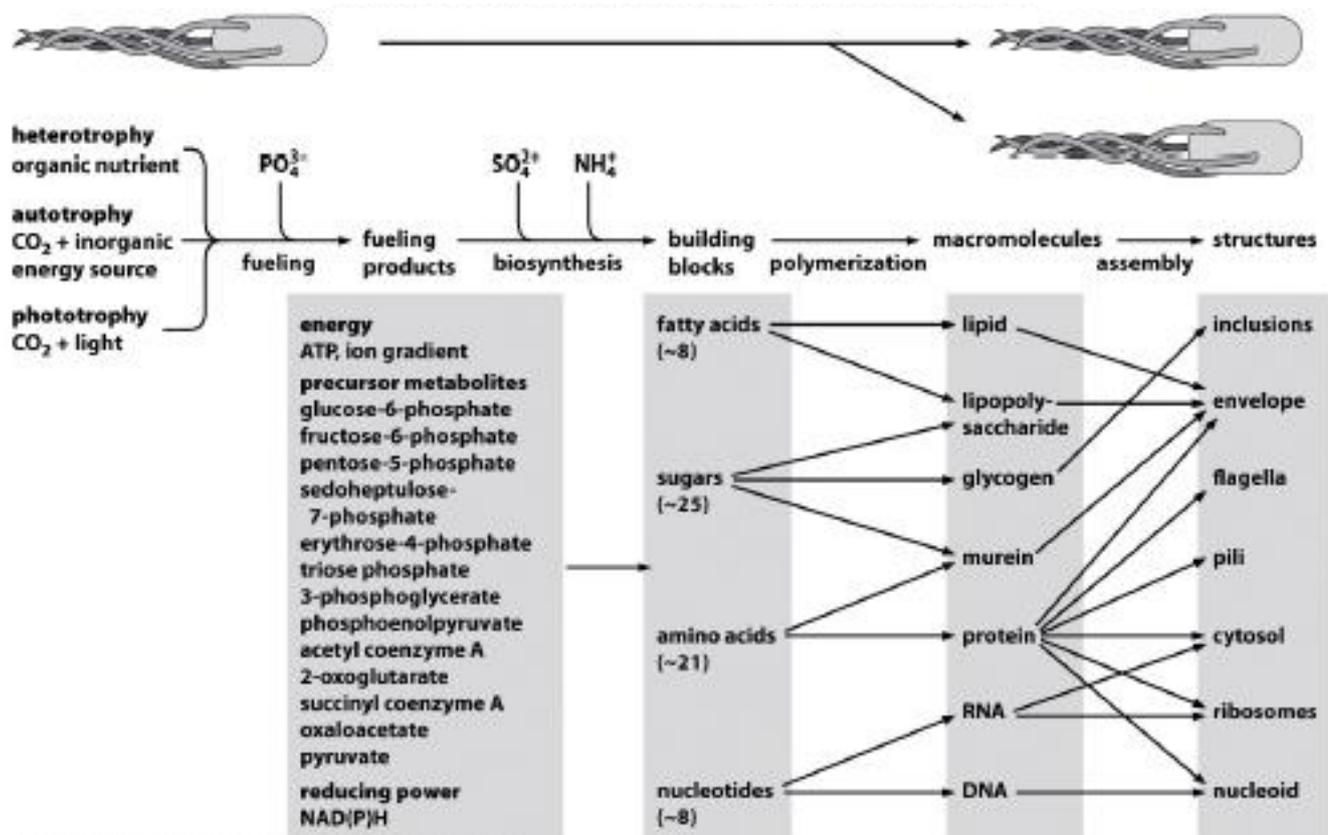
Fig. 5.10. Schematic diagram of process of biosynthesis. Source: Physical Biology of cell. Garland Sciences 2009.

Biosynthesis of proteins

→ glucose as a sole carbon source: 10^{10} Catomsin $E.coli$cell (6 Catomsper glucose, nee2×10^9d glucose molecules just to construct a cell)

→ metabolic path ways for synthesis of 20 amino acids known butcomplex; connected to glycolytic pathway:(alanine from pyruvate in a single step by a single enzyme)

→ an average energetic cost to synthesize an amino acid is:1.2 ATP equivalents aerobically, 4.7 ATP equivalents anaerobically

→ Build a protein from amino acids: 4 ATP equivalents form peptide bonds attach amino acids to tRNA (carries one codon, 7393 nucleotides)power the movement of ribosome. In total, to build a protein: 5.2 ATP equivalents per amino acid: For the entire $E.coli$cell: 5.2 $ATP \times 300 \times 3 \times 10^6 = 4.5 \times 10^9 ATP$

How about DNA/RNA building: To synthesize one nucleotide: 10^{20} ATPcost of assembling nucleotides into polymers is small (10%).

How much energy can one glucose generate: Under ideal growing conditions: 30 ATP (CO_2 waste product) Total cost of$E.coli$cell building: 10^{10} ATP or 6×10^8glucosemolecules (1/3 of the material required under ideal conditionsand up to 10foldin less efficient growth conditions)

Table 5.2. Parameters of the process of protein synthesis

Amino acid	Abundance (molecules per cell)	Glucose equivalents	ATP equivalents (aerobic)	ATP equivalents (anaerobic)
Alanine (A)	2.9×10^8	0.5	−1	1
Arginine (R)	1.7×10^8	0.5	5	13
Asparagine (N)	1.4×10^8	0.5	3	5
Aspartate (D)	1.4×10^8	0.5	0	2
Cysteine (C)	5.2×10^7	0.5	11	15
Glutamate (E)	1.5×10^8	0.5	−7	−1
Glutamine (Q)	1.5×10^8	0.5	−6	0
Glycine (G)	3.5×10^8	0.5	−2	2
Histidine (H)	5.4×10^7	1	1	7
Isoleucine (I)	1.7×10^8	1	7	11
Leucine (L)	2.6×10^8	1.5	−9	1
Lysine (K)	2.0×10^8	1	5	9
Methionine (M)	8.8×10^7	1	21	23
Phenylalanine (F)	1.1×10^8	2	−6	2
Proline (P)	1.3×10^8	0.5	−2	4
Serine (S)	1.2×10^8	0.5	−2	2
Threonine (T)	1.5×10^8	0.5	6	8
Tryptophan (W)	3.3×10^7	2.5	−7	7
Tyrosine (Y)	7.9×10^7	2	−8	2
Valine (V)	2.4×10^8	1	−2	2

Table 5.3. Biosynthetic cost in ATP equivalents to synthesize the macromolecules of a single cell.

Class	Biosynthetic cost (aerobic) – ATP equiv.
Protein	4.5×10^9
DNA	3.5×10^8
RNA	1.6×10^9
Phospholipid	3.2×10^9
Lipopolysaccharide	3.8×10^8
Peptidoglycan	1.7×10^8
Glycogen	3.1×10^7

Biological Systems as Minimizers

→ mechanical and chemical equilibrium: minimization problems

→ mechanical/chemical equilibrium: short time scales

Example:

$$A \underset{k-}{\overset{k+}{\rightleftarrows}} B \xrightarrow{r} C$$
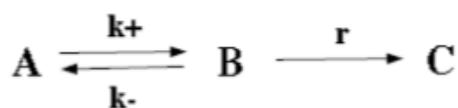
If $k+$ and $k- >> r$ (faster reactions), then A and B can be treated as if in chemical equilibrium.
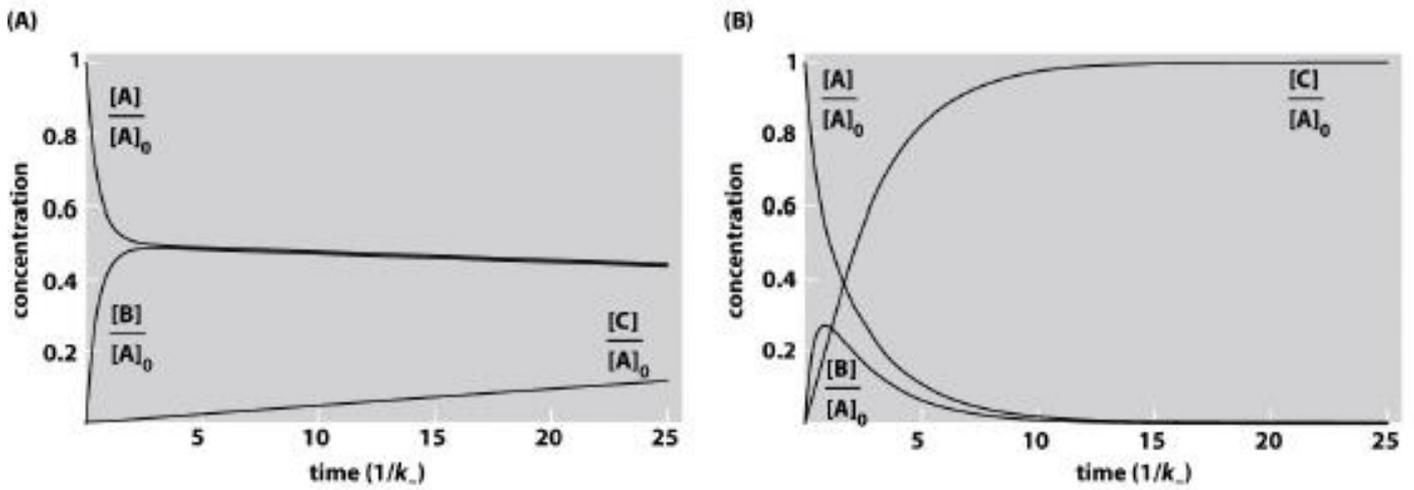
Fig. 5.11. Time dependence of concentration changes in biomolecules.. Source: Physical Biology of cell. Garland Sciences 2009.

Free energy minimization principles: Protein folding
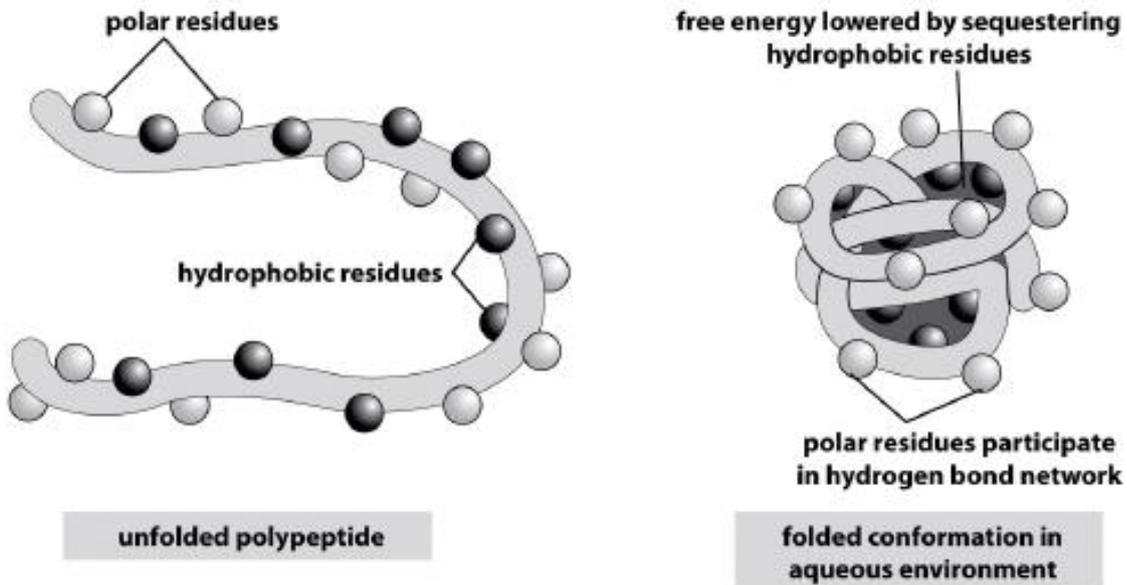
Fig. 5.12. Some examples of free energy minimization principles realization in biomolecules. Source: Physical Biology of cell. Garland Sciences 2009.

Mechanical equilibrium: Potential energy minimization.

$$U(x) = k\,(x-x_0)^2/2 - mg\,(x - x_0)$$

(A)

(B)
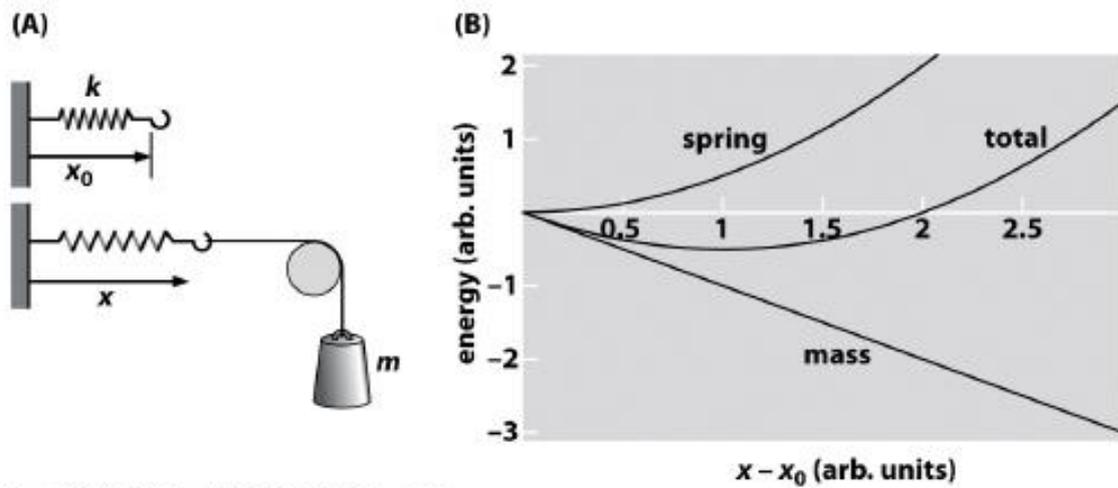
Fig. 5.13 .Mechanism of potential energy minimization and equilibrium. Source: Physical Biology of cell. Garland Sciences 2009.

Optical trap as a mass-spring system: Bead in an optical3 trap with DNA tether exerting force

$$U(x) = k_{trap}\,x^2/2 - Fx$$

(A) laser beam

DNA tether

optical bead

(B) k_{trap}

(C)
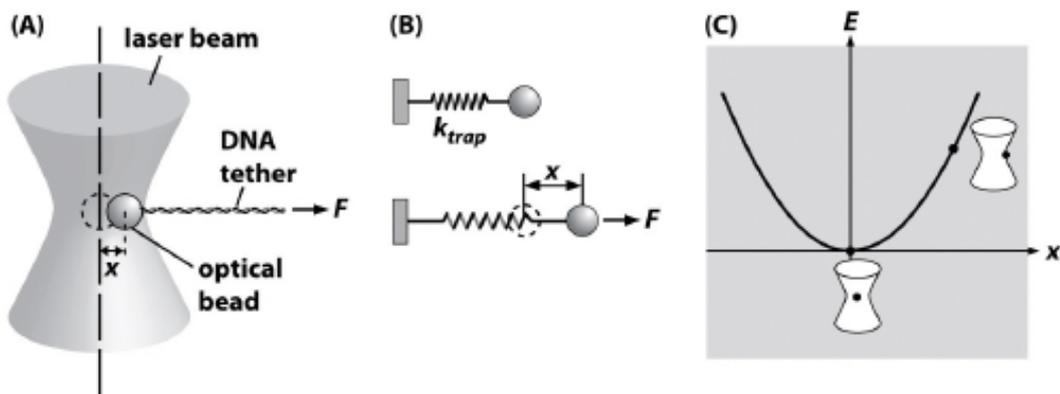
Fig.5.14.Trapping of optical forces in biosystems. Source: Physical Biology of cell. Garland Sciences 2009.

Chapter 6

Biological Processes and Entropy

Energy and Order in Biological Systems: The concept of entropy and the second law of thermodynamics suggests that systems naturally progress from order to disorder. ... Using chlorophyll in the process called photosynthesis, they convert the sun's energy into storable form in ordered sugar molecules.

Processes leading to randomness, disorder, chaos or loss of information would appear to be out of place in the world of biology. Living things are characterized by a very high degree of structure and assembly, whether at the level of molecules, genetic information, cells, tissues, organs, organisms or populations of organisms. On the other hand, the second law of thermodynamics implies the principle that the total entropy, which is a measure of disorder, must increase steadily. Even though thermodynamics itself does not describe processes as a function of time, the second law defines a unique direction of time as the direction in which total entropy increases. Nevertheless, thermodynamics does not exclude local exceptions. Living things are local exceptions. Every isolated system moves towards a state of maximum entropy. The universe is presumably an isolated system, whose entropy can never decrease. Thus, it is clear that the decrease in entropy that accompanies growth of living structure, must always be accompanied by an increase in entropy in the sustaining physical environment. In his classic book, "What is life?" Erwin Schroedinger discussed how life is a state of very low probability because so much energy is needed to create and sustain it. The 'vital force' that maintains life is energy. Living things preserve their low levels of entropy throughout time, because they receive energy from their surroundings in the form of food. They gain their order at the expense of disordering the nutrients they consume. For example, the development. Progress in nature is nearly always achieved by the development of more complicated biological structures, usually when these structures are better adapted to the environment. This development occurs by accumulation of small random changes in the DNA master plan, not by starting again from scratch. In other words, the mechanism of evolution makes the whole process irreversible. Thus, the analogy between the theory of the evolution of life and the laws of statistical physics and entropy, both of which apply to large populations (of organisms and molecules, respectively), extends to the concept of the irreversibility of time. It is interesting that Darwin developed his ideas on the evolution of life forms at about the same time as his contemporaries, the physicists Clausius, Boltzmann, Maxwell and Gibbs, formulated the laws of statistical physics and entropy. It was Boltzmann who remarked that the nineteenth century would be known as the century of Darwin.

Entropy and Free Energy

According to the second law of thermodynamics, spontaneous chemical reactions, which in biological processes typically occur at constant temperature and pressure, are always accompanied by decreases in free energy. Free energy is energy that is capable of doing work. All reactions proceed towards a state of equilibrium, the state of maximum probability where nothing further happens, and hence, no further decrease in free energy occurs. The free energy lost appears either as heat (enthalpy) or is used to increase the entropy. Thus, the spontaneous chemical reactions, on which life depends, may proceed without release of heat, and may, in fact, proceed with take-up of heat, but only at the cost of expending free energy. For example, the unfolding of a protein consumes a large amount of heat, and as described below, is the ultimate example of a reaction that can be driven by an increase in entropy. The original definition of entropy, which was proposed by Rudolf Clausius in 1864, when applied to biological processes that occur typically at constant temperature, is that the change in entropy is equal to the heat supplied divided by the temperature. Since for a protein unfolding reaction, heat supplied is positive because heat is taken up, the change in entropy is always positive. Entropy as a driving force for protein unfolding. One manifestation of the high order and low entropy characteristic of living systems is evident in the intricate and hierarchical, but yet very precise structures of proteins, into which their polypeptide chains comprising amino acid residues fold. It is the structure that determines the function of any protein, whether in catalysis, structural support, locomotion, transport or communication. The functional or native state of a protein has very low entropy because its conformation is highly restricted. On the other hand, the unfolded form can exist in many different conformations: even if each amino acid can adopt only 3 positions, a polypeptide chain of 100 amino acid residues can adopt 3100 or 1047 different conformations. Since the product of the protein unfolding reaction can exist in a larger number of equivalent states, compared with the native state, entropy increases during the unfolding reaction. From the simple formula defined by Ludwig Boltzmann in 1877, the entropy of the unfolded state, which is equal to k W (where k is the Boltzmann constant, and W is the number of accessible states) can be calculated to be - 250 cal K-l mol-I. This is more than 10 times higher than the entropy changes of 20 cal K-I mol-I, that are usually observed for the reactions of small molecules. Thus, the entropy loss during protein folding plays a much larger role in determining the shape of the free energy reaction landscape than it does in most small molecule reactions. For a protein to fold, the loss of entropy must be balanced by the gain in enthalpy for the free energy to favor folding. Strong noncovalent forces from hydrogen bonding and other physical interactions compensate for the low entropy of the native state, but just barely. Changes in the entropy of the solvent, water, play an important role in compensating for the loss in conformational entropy. In the native state, many non-polar amino acid residues are packed into the interior of the protein, and hence, sequestered away from water. In the unfolded form, these residues are exposed to water molecules, which consequently, get arranged around the non-polar residues, into cage-like structures: the normal hydrogen bonding network present in water is reorganized so that the number of

hydrogen bonds is nearly preserved. The additional hydrogen bond donors and acceptors of the polar polypeptide backbone of the protein, which get exposed upon unfolding, probably also playa role in restricting the freedom of movement of more water molecules. This ordering of water molecules lowers the entropy of water. When the protein folds, these water molecules are released as the non-polar residues get secluded from water. The resultant regain of entropy by water, is thought to be a dominant force in protein folding, and the effect is commonly referred to as the hydrophobic effect. According to the original definition by Clausius, the change in entropy is larger at lower temperatures: a fixed amount of heat has a greater disordering effect at a lower temperature than at a higher temperature. The entropy of protein unfolding would therefore be expected to become smaller with increasing temperature. Nevertheless, it is commonly observed that the entropy of unfolding becomes higher with increasing temperature, because the cage-like structures of water molecules melt at higher temperatures, and so fewer water molecules remain restricted' in motion. The solvent can therefore have profound effects on biological reactions. Entropy as a driving force for catalysis by enzymes. Although life is driven by spontaneous reactions, it is only by the intervention of large, very elaborately structured molecules in cells, usually proteins called enzymes, that the rates of these reactions are brought into biologically relevant time scales. Entropy plays a big role in enzyme catalysis. Reactions in solution are usually slow because of the entropic cost in bringing the reactants or reactant and catalyst together. Two or more molecules associating to form one involves considerable loss of entropy. On the other hand, when an enzyme binds its substrate, the binding energy released is used to compensate for the loss in translational and rotational entropies which occurs on formation of the enzyme-substrate complex, which is a state of very low probability because catalytic groups have to be oriented very precisely, within fractions of an angstrom. This occurs at the expense of an increase in the dissociation constant of the enzyme-substrate complex. Very little entropy loss occurs during the actual chemical reaction steps, because the catalytic groups are already properly oriented on the same enzyme-substrate complex, and hence, their effective concentrations are very high compared to the corresponding bimolecular reactions that occur free in solution. Thus, the ability of enzymes to accelerate chemical reactions by factors as large as 1015, resides in their ability to reduce the amount of entropy that must be lost for the reaction to occur. Entropy and Single Molecules in Biology One of the most exciting developments in biology today is the application of physical methods to study the behavior of single protein or DNA molecules. For instance, the use of optical tweezers and atomic force microscopes now makes it possible to take hold of a single multi-domain protein molecule and mechanically stretch it, thereby unfolding it. A fully stretched protein molecule has near zero entropy; When it is freed from mechanical restraint, it first adopts a random coil unfolded form of high entropy, and then folds to the native state which again has near-zero entropy. Measurement of the entropic restoring force in a single stretched protein molecule allows direct determination of the mechanical force and energy that can unfold a single protein domain. The methods of single molecule biology are now being applied to many diverse mechanical

systems in biology, including molecular motors. Molecular motors carry out the basic functions that keep cells living: these protein molecules build or destroy other proteins, transport materials, pump ions, or propel the cell. They do their jobs by converting chemical energy into mechanical energy with near 100% efficiency. The directed motion of molecular motors, overcoming entropy, is crucial for processing information in biology. It appears that biological nanomotors might be able· to perform their tasks because they use the energy from biased Brownian motion to convert the molecular disorder around them into order. Nothing escapes the long arm of the second law.

Changes in configurational entropy represent one of the major contributions to the thermodynamics of folding, binding, and oligomerization. Methods have been developed to estimate changes in the entropy of the backbone and side chains, and for the loss of translational entropy. These methods have been used in combination with empirical methods that provide estimates of the changes in entropy of solvation as well as estimates of the changes of enthalpy. The results of such calculations are in excellent agreement with experimentally observed values. The laws of thermodynamics are important unifying principles of biology. These principles govern the chemical processes (metabolism) in all biological organisms. The First Law of Thermodynamics, also known as the law of conservation of energy, states that energy can neither be created nor destroyed. It may change from one form to another, but the energy in a closed system remains constant. The Second Law of Thermodynamics states that when energy is transferred, there will be less energy available at the end of the transfer process than at the beginning. Due to entropy, which is the measure of disorder in a closed system, all of the available energy will not be useful to the organism. Entropy increases as energy is transferred. In addition to the laws of thermodynamics, the cell theory, gene theory, evolution, and homeostasis form the basic principles that are the foundation for the study of life.

First Law of Thermodynamics in Biological Systems

All biological organisms require energy to survive. In a closed system, such as the universe, this energy is not consumed but transformed from one form to another. Cells, for example, perform a number of important processes. These processes require energy. In photosynthesis, the energy is supplied by the sun. Light energy is absorbed by cells in plant leaves and converted to chemical energy. The chemical energy is stored in the form of glucose, which is used to form complex carbohydrates necessary to build plant mass.The energy stored in glucose can also be released through cellular respiration. This process allows plant and animal organisms to access the energy stored in carbohydrates, lipids, and other macromolecules through the production of ATP. This energy is needed to perform cell functions such as DNA replication, mitosis, meiosis, cell movement, endocytosis, exocytosis, and apoptosis.

Second Law of Thermodynamics in Biological Systems

As with other biological processes, the transfer of energy is not 100 percent efficient. In photosynthesis, for example, not all of the light energy is absorbed by the plant. Some energy is reflected and some is lost as heat. The loss of energy to the surrounding environment results in an increase of disorder or entropy. Unlike plants and other photosynthetic organisms, animals cannot generate energy directly from the sunlight. They must consume plants or other animal organisms for energy. The higher up an organism is on the food chain, the less available energy it receives from its food sources. Much of this energy is lost during metabolic processes performed by the producers and primary consumers that are eaten. Therefore, much less energy is available for organisms at higher trophic levels. (Trophic levels are groups that help ecologists understand the specific role of all living things in the ecosystem.) The lower the available energy, a smaller number of organisms can be supported. This is why there are more producers than consumers in an ecosystem. Living systems require constant energy input to maintain their highly ordered state. Cells, for example, are highly ordered and have low entropy. In the process of maintaining this order, some energy is lost to the surroundings or transformed. While cells are ordered, the processes performed to maintain that order result in an increase in entropy in the cell's/organism's surroundings. The transfer of energy causes entropy in the universe to increase.

Some necessary additional examples Information theory: Based on the constraints (for example the total energy) *derive the least biased probability distribution* Shannon entropy:

$$S(P_1, P_2, P_3 \ldots, P_n) = - \sum_{i=1}^{N} P_i \, ln \, P_i$$

where P_i is the probability the system to be in the i-th microstate.

For example, if nothing is known about the system that there is N microstates, then for all $P_i = 1/N$ and $S = lnN$ (the maximal value).

Maximize Shannon entropy S' using Lagrange multiplier method for each constraint:

$$S' = - \sum_{i=1}^{N} P_i \, ln \, P_i - \lambda \left[\sum_{i=1}^{N} P_i - 1 \right]$$

Maximize equations:

$$\frac{\partial S'}{\partial \lambda} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} P_i + 1 = 0$$

$$\frac{\partial S'}{\partial P_i} = 0 \quad \rightarrow \quad -lnP_i - 1 - \lambda = 0$$

$$P_i = exp(-1 - \lambda)$$

Note that the probabilities P_i do not depend on i!

$$\sum_{i=1}^{N} P_i = 1 \;\rightarrow\; \sum_{i=1}^{N} exp(-1-\lambda) = 1 \;\rightarrow\; exp(-1-\lambda) = 1/N$$

$$P_i = 1/N$$

Boltzmann distribution is a maximum entropy distribution with a fixed average energy:

$$S' = -\sum_i P_i \, lnP_i - \lambda \left[\sum_i P_i - 1 \right] - \beta \left[\sum_i P_i E_i - \langle E \rangle \right]$$

$$-lnP_i - 1 - \lambda - \beta E_i \;\rightarrow\; P_i = exp(-1-\lambda)exp(-\beta E_i)$$

$$\sum_i P_i = 1 \;\rightarrow\; exp(1+\lambda) = \sum_i exp(-\beta E_i) = Z$$

$$P_i = \left. exp(-\beta E_i) \middle/ \sum_i exp(-\beta E_i) \right.$$

Gases are complicated. They're full of billions and billions of energetic gas molecules that can collide and possibly interact with each other. Since it's hard to exactly describe a real gas, people created the concept of an Ideal gas as an approximation that helps us model and predict the behavior of real gases. The term ideal gas refers to a hypothetical gas composed of molecules which follow a few rules: Ideal gas molecules do not attract or repel each other. The only interaction between ideal gas molecules would be an elastic collision upon impact with each other or an elastic collision with the walls of the container. Ideal gas molecules themselves take up no volume. The gas takes up volume since the molecules expand into a large region of space, but the Ideal gas molecules are approximated as point particles that have no volume in and of themselves. Ideal Gas Law is an Approximation. The ideal gas law describes how gases behave, but does not account for molecular size or intermolecular forces. Since molecules and atoms in all real gases have size and exert force on each other, the ideal gas law is only an approximation, albeit a very good one for many real gases.
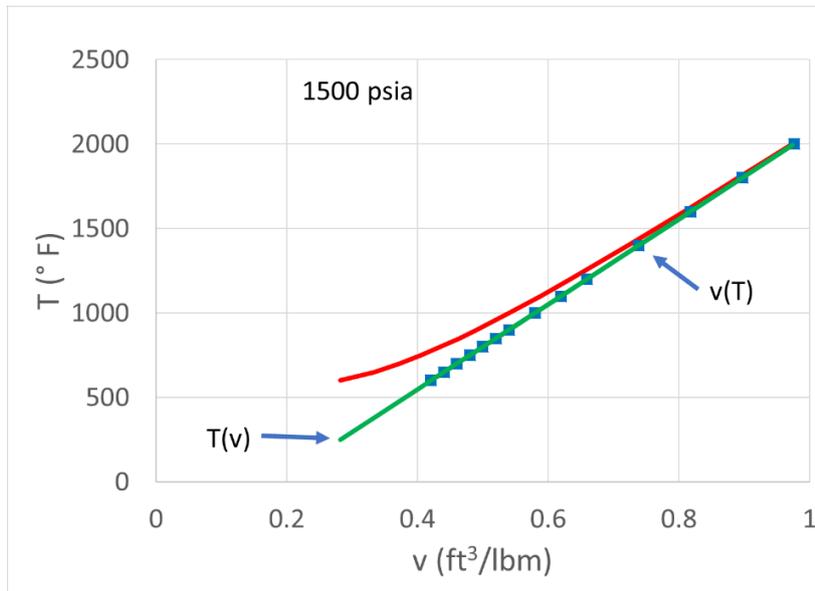
Figure 6.1. Steam table data with ideal gas approximation calculations for 4000 psia. The red line represents the actual $v - T$ data from the superheated tables for water, while the green line represents temperatures calculated from tabulated pressure and specific volume data using the ideal gas approximation, and the blue symbols represent specific volumes calculated from tabulated pressure and temperature data.

Figure 6.2. The theories and gas owication.
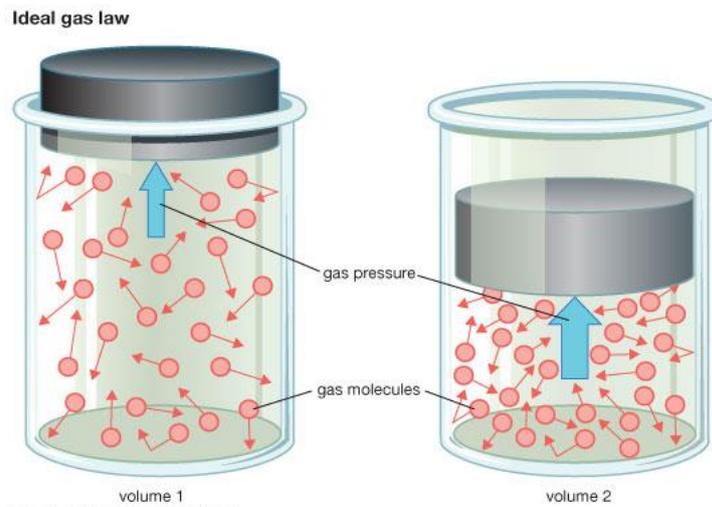
Figure 6.3. The scheme of ideal gas law. Source: Encyclopedia Britannica

Ideal gas approximation: the interaction range is small in comparison to the mean spacing between molecules

For a system with N independent variables $(x_1, x_2, \ldots, x_N)$ the probability distribution can factorize:

$$P(x_1, x_2, \ldots x_N) = P(x_1) \, P(x_2) \ldots P(x_N)$$

Probability distribution for a system with average energy is:

$$P_{(p_x)} = \frac{\exp\left[-\beta p_x^2 / (2m)\right]}{\sum_{states} \exp\left[-\beta p_x^2 / (2m)\right]}$$

Instead of a sum, we use an integral over a continuous p_x:

$$\sum_{states} \rightarrow \int_{-\infty}^{\infty} dp_x$$

Where of a sum, we use an integral:

$$\int_{-\infty}^{\infty} \exp(-\alpha p_x^2) dp_x = \sqrt{\frac{\pi}{\alpha}}$$

According to the equipartition, each degree of freedom is associated with:

$$\langle E \rangle = \frac{1}{2} k_B T$$

Calculate $\langle E \rangle$ using the Boltzmann distribution:

$$Z = \int_{-\infty}^{\infty} exp\left[-\beta p_x^2 \middle/ (2m)\right] dp_x = \sqrt{2m\frac{\pi}{\beta}}$$

$$\langle E \rangle = \int_{-\infty}^{\infty} \frac{p_x^2}{(2m)} exp\left[-\beta p_x^2 \middle/ (2m)\right] dp_x$$

We use the following trick:

$$\langle E \rangle = -Z^{-1}\frac{\partial Z}{\partial \beta} = \frac{1}{2}\beta^{-1}$$

So we showed that the Lagrange multiplier $\beta = (k_B T)^{-1}$.

Free energy and chemical potential of a dilute solution

Application of a lattice model and ideal gas approximation
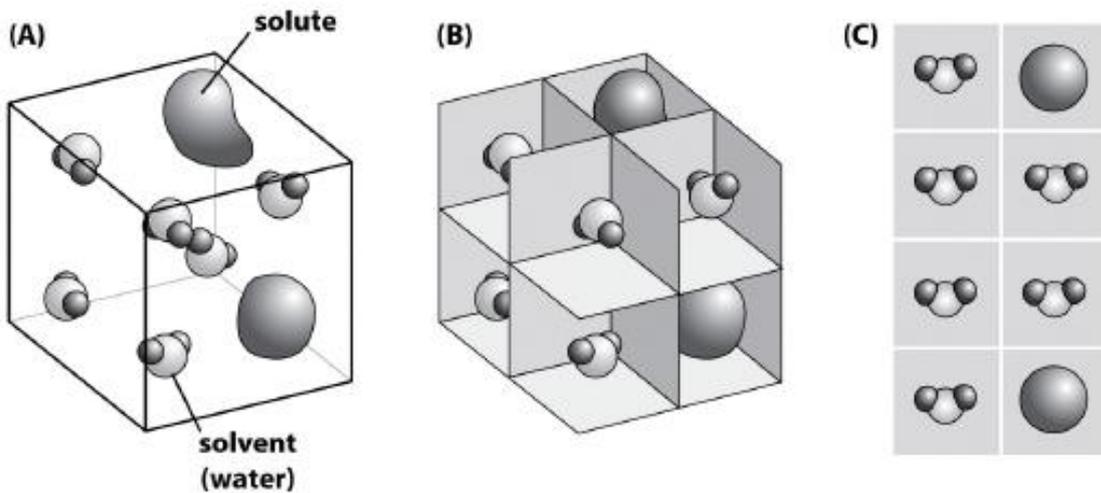
Figure 6.4. Example of lattice model in connection to free energy an chemical potential of a dilute solution.

Configurational entropy of N objects placed into W available spots:

$$W(N,\Omega) = \frac{\Omega!}{N!\,(\Omega - N)!} \qquad\qquad S = k_B \, lnW$$

How to calculate the chemical potential of a dilute solutions?

$$\mu_{SOLUTE} = \left(\frac{\partial G_{TOT}}{\partial N_S}\right)_{Tp}$$

$$G_{TOT} = N_W \mu_W^0 + N_S \varepsilon_S - T S_{MX}$$

65

Water G+ solute energy + mixing entropy contribution

Mixing entropy contribution:

- Independent solute molecules (ideal gas)
- Lattice model: $N_W + N_S$ is a total number of lattice sites

$$W\,(N_W, N_S) = \frac{(N_W + N_S)!}{(N_W!\,N_S!)}$$

$$S = k_B lnW = ln\frac{(N_W + N_S)!}{(N_W!\,N_S!)} \approx -k_B \left\{ N_W ln\left[\frac{N_W}{(N_W + N_S)}\right] + N_S ln\left[\frac{N_S}{(N_W + N_S)}\right]\right\}$$

$$= -k_B \left[N_W ln\left(\frac{1}{\frac{N_S}{N_W}}\right) + N_S ln\left(\frac{1}{\frac{N_S}{N_W}}\right)\right]$$

Taking into account the Taylor expansion of in $ln(1 + x) \approx x$, we get:

$$S_{MX} = -k_B \left[N_S ln\left(\frac{N_S}{N_W} - N_S\right)\right]$$

$$G_{TOT}(T, p, N_W, N_S) = N_W \mu_W^0 + N_S \varepsilon_S(T, p) + k_B T\left(N_S ln\left(\frac{N_S}{N_W}\right) - N_S\right)$$

$$\mu_S(T, p) = \varepsilon_S(T, p) + k_B T ln\left(\frac{c}{c_0}\right)$$

Or in a general form expressed in concentration $c = \frac{N}{V}$:

$$\mu_i = \mu_{i0} + k_B T ln\left(\frac{c_i}{c_{i0}}\right)$$

Osmotic pressure Is an Entropic Effect

> → consider a cell in an aqueous environment exchanging material with a solution (intake of food, excretion of waste)

> → chemical potential difference proportional to ΔG:

> $\Delta G = (\mu_1 - \mu_2)dN \leq 0$ (spontaneous)

> → cell with a crowded environment of biomolecules: tendency of almost all components to move out causes a mechanical pressure called *osmotic pressure*

> → lipid membranes with ion channels to regulate ion concentration

> → calculate osmotic pressure due to a dilute solution of N_S molecules

Osmotic pressure on a semipermeable membrane, which only allows water molecules through
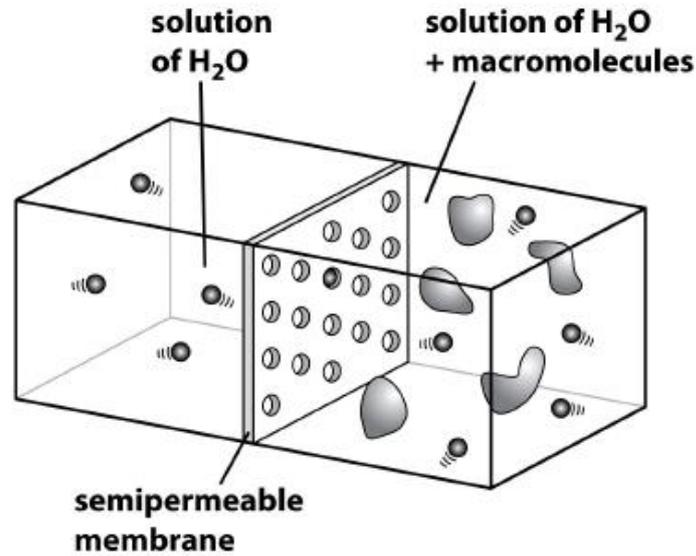
Figure 6.5. Scheme of osmotic pressure for water and water+ macromolecules.

$$\mu_W(T,p) = \left(\frac{\partial G_{TOT}}{\partial N_S}\right)_{Tp} = \mu_W^0(T,p) - k_B T \frac{N_S}{N_W}$$

For both sides of the membrane in equilibrium:

$$\mu_W^0(T,p_1) = \mu_W^0(T,p_2) - k_B T \frac{N_S}{N_W}$$

Expand the chemical potential at p_2 around the p_1 value:

$$\mu_W^0(T,p_2) \approx \mu_W^0(T,p_1) + \left(\frac{\partial \mu_W^0}{\partial p}\right)(p_2 - p_1)$$

And consider that $\left(\frac{\partial \mu_W^0}{\partial p}\right) = v = \frac{N_W}{V}$ is volume per water molecule so that

$$(p_2 - p_1) = k_B T \frac{N_S}{V}$$

Measuring inter strand interactions in DNA using osmotic pressure
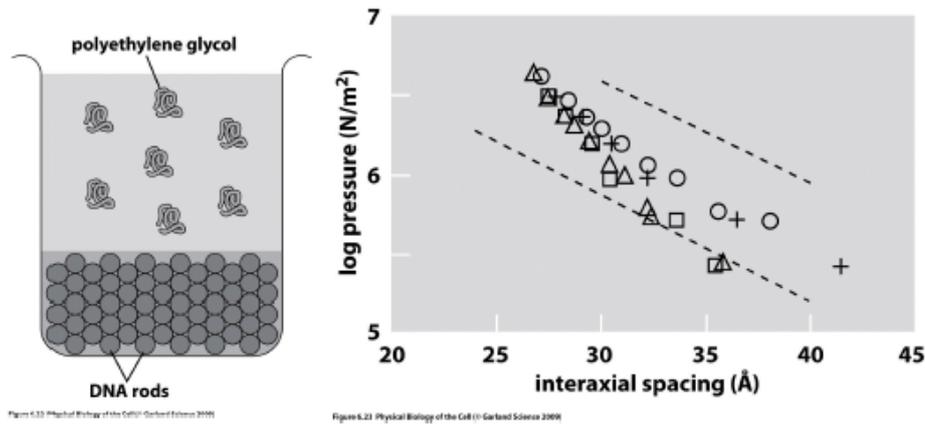
pressure: $p(d_s) = F_0 \exp(-d_s/c)$

d_s ... interstrand spacing; F_0 depends on the ionic solution

Figure 6.6. Example of measurement of inter strand interaction by osmotic pressure.

Law of Mass Action and Equilibrium Constants (Chemical Reactions)

→ chemical equilibrium between A,B and their complex AB:

$$A + B \leftrightarrow AB$$
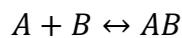
→ final equilibrium independent of whether we start with only A and B, or with a high concentration of A or B

→ N_A, N_B, N_{AB} ... number of $A, B,$ and AB molecules

→ In equilibrium: $dG = 0$

$$0 = \left(\frac{\partial G}{\partial N_A}\right) dN_A + \left(\frac{\partial G}{\partial N_B}\right) dN_B + \left(\frac{\partial G}{\partial N_{AB}}\right) dN_{AB}$$

→ A more convenient and general expression:

$$\sum_{i=1}^{N} \mu_i \, dN_i = 0$$

Stoichiometric coefficients for each of the reactants are defined as:

$$V_i = \pm 1$$

Depending on whether the number of particles of the i-th type increases or decreases the reaction:

$$\sum_{i=1}^{N} \mu_i V_i = 0$$

or

$$\sum_{i=1}^{N} \mu_{i0} V_i = -k_B T \sum_{i=1}^{N} \ln\left(\frac{c_i}{c_{i0}}\right)^{Vi} - \beta \sum_{i=1}^{N} \mu_{i0} V_i = \ln\left[\prod_{i=1}^{N}\left(\frac{c_i}{c_{i0}}\right)^{Vi}\right]$$

or

$$\prod_{i=1}^{N}(c_i)^{Vi} = \left(\prod_{i=1}^{N}(c_{i0})^{Vi} \, exp\beta \sum_{i=1}^{N} \mu_{i0} V_i\right)$$

Where we define the equilibrium constant K_{eq}:

$$K_{eq} = \left(\prod_{i=1}^{N}(c_{i0})^{Vi}\right) exp\left(-\beta \sum_{i=1}^{N} \mu_{i0} V_i\right)$$

$K_d = \frac{1}{K_{eq}}$ ... dissociation constant

In our case of the reaction $A + B \leftrightarrow AB$, we can express K_d as:

$$K_d = \prod_{i=1}^{N} c_i^{Vi} = \frac{c_A c_B}{c_{AB}}$$

Example: total concentration: 50 µM

$K_d = 2.5\ \mu M$ $K_d = 25\ \mu M$

80% bound   50% bound
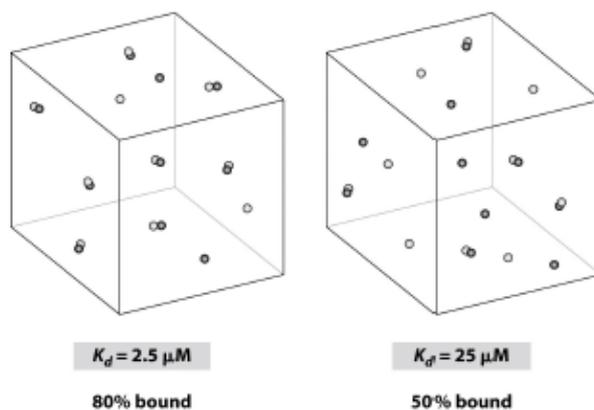
Application to Ligand-Receptor Binding:
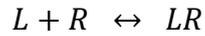
$$L + R \ \leftrightarrow \ LR$$

$$K_d = \frac{[L][R]}{[LR]} \quad or \ [LR] = \frac{[L][R]}{K_d}$$

Binding probability:

$$p_{boad} = \frac{[LR]}{[LR] + [R]} = \frac{\frac{[L]}{K_d}}{1 + \frac{[L]}{K_d}}$$

A natural interpretation of K_d: K_d is the concentration at which the receptor has a probability of $\frac{1}{2}$ of being occupied by a ligand. Based on out prior result, we can express it in terms of lattice model parameters as: $K_d = V^{-1}exp(\beta\Delta\varepsilon)$

Important: K_d depends on the concentration of *free* ligands not their *total* concentration!
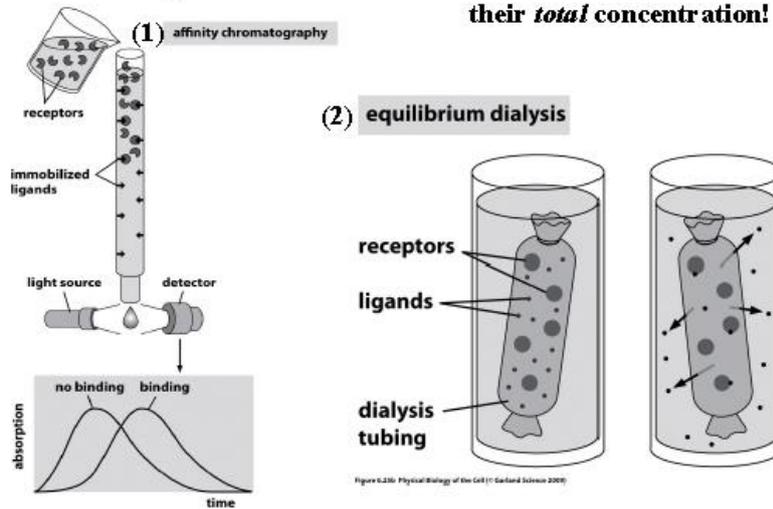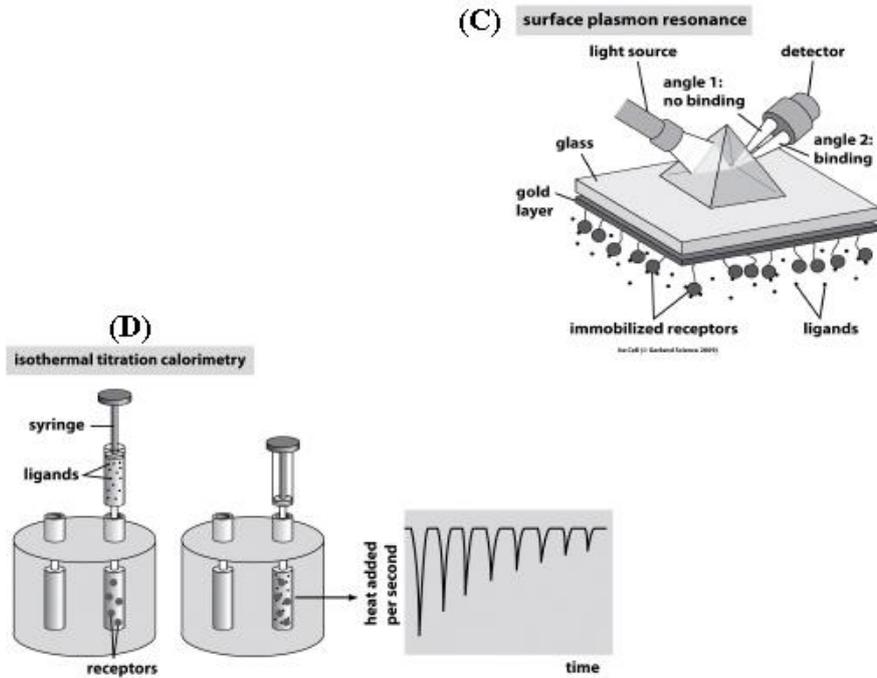
Figure 6.7. Examples of processes of formation of dissociation constant dependence.

Cooperative Ligand-Receptor Binding: The Hill Function

Biological function: an on off switch behavior triggered by Binding of a ligand to a receptor involves a cooperative (all or none) mechanism:

$$L + L + R \; \leftrightarrow \; L_2R$$

$$K_d^2 = \frac{[L]^2[R]}{[L_2R]}$$

$$p_{boad} = \frac{\left(\frac{[L]}{K_d}\right)^2}{1 + \left(\frac{[L]}{K_d}\right)^2} = \frac{\left(\frac{[L]}{K_d}\right)^n}{1 + \left(\frac{[L]}{K_d}\right)^n}$$

Figure 6.8. Graphical picture of dependence of bounding strength from ligand concentration.

Lecture 7

Ions and Binding

Ionic bond, also called **electrovalent bond**, type of linkage formed from the electrostatic attraction between oppositely charge ions in a chemical compound. Such a bond forms when the valence electrons of one atom are transferred permanently to another atom. The atom that loses the electrons becomes a positively charged ion (cation), while the one that gains them becomes a negatively charged ion (anion).
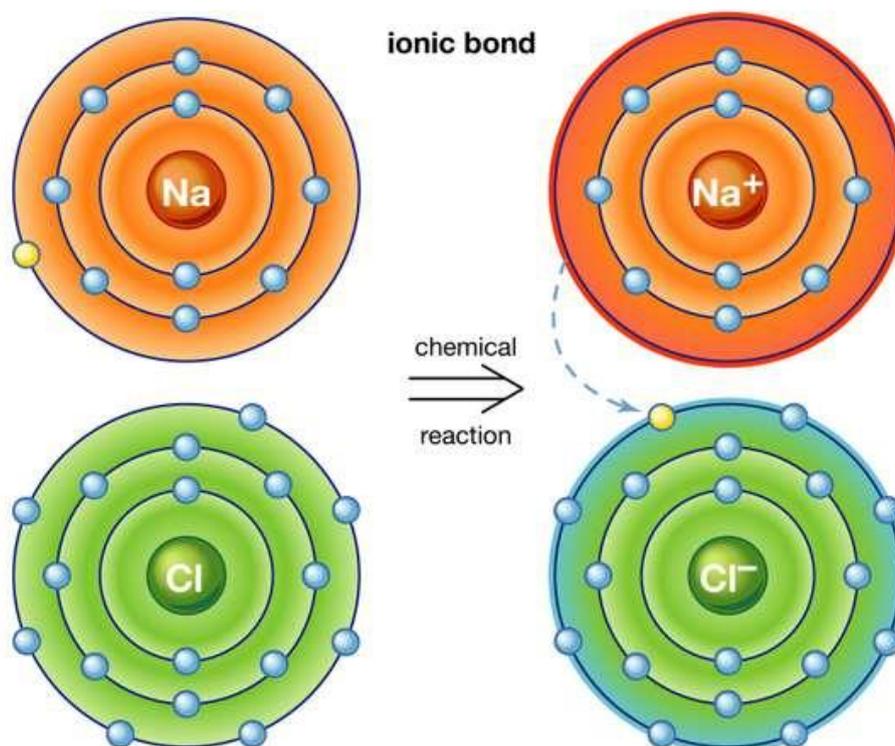
Fig. 7.1 Examples of ionic bonds.Ionic bond: sodium chloride, or table saltIonic bonding in sodium chloride.
Source: Encyclopedia Britannica, 2012.

An atom of sodium (Na) donates one of its electrons to an atom of chlorine (Cl) in a chemical reaction, and the resulting positive ion (Na^+) and negative ion (Cl^-) form a stable ionic compound (sodium chloride; common table salt) based on this ionic bond.
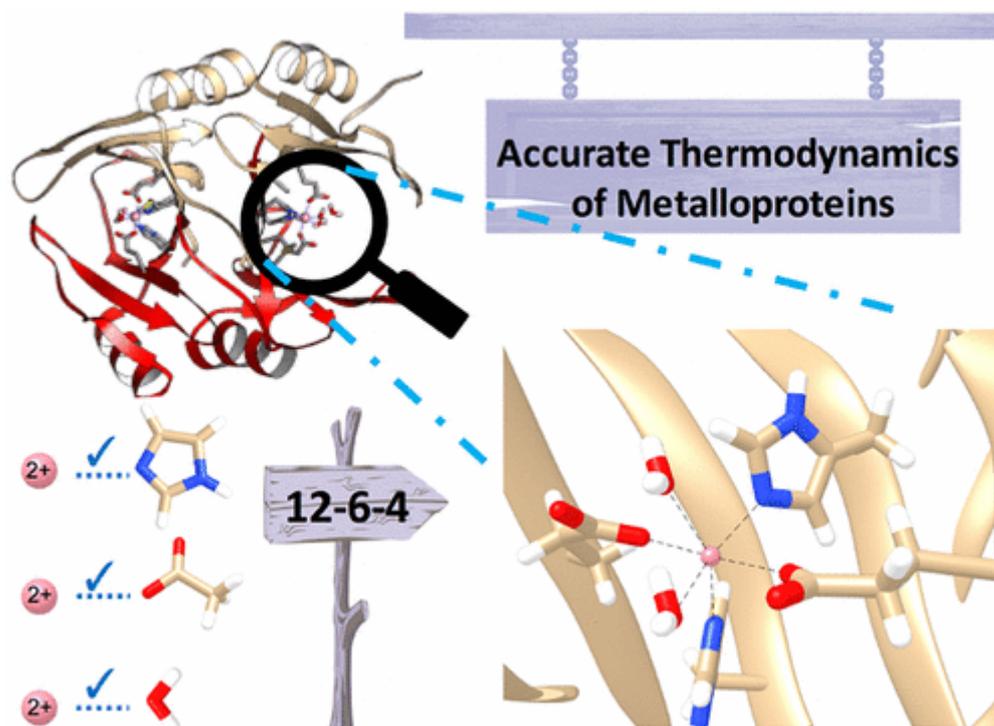
Fig.7.2. Schematic picture of thermodynamics of metal proteins.

Modeling the thermodynamics of a transition metal (TM) ion assembly be it in proteins or in coordination complexes affords us a better understanding of the assembly and function of metalloclusters in diverse application areas including metal organic framework design, TM-based catalyst design, the trafficking of TM ions in biological systems, and drug design in metalloprotein platforms. While the structural details of TM ions bound to metalloproteins are generally well understood via experimental and computational approaches, accurate studies describing the thermodynamics of TM ion binding are rare. It is is possible to obtain accurate structural and absolute binding free energies of $Co2+$ and $Ni2+$ to the enzyme glyoxalase I using an optimized 12–6–4 (m12–6–4) potential. Critically, this model simultaneously reproduces the solvation free energy of the individual TM ions and reproduces the thermodynamics of TM ion–ligand coordination as well as the thermodynamics of TM ion binding to a protein active site unlike extant models. We find the incorporation of the thermodynamics associated with protonation state changes for the TM ion (un)binding to be crucial. The high accuracy of m12–6–4 potential in this study presents an accurate route to explore more complicated processes associated with TM cluster assembly and TM ion transport.

Examples of two state systems, ion channels and state variables of binding

Examples of two-state systems: state variable σ
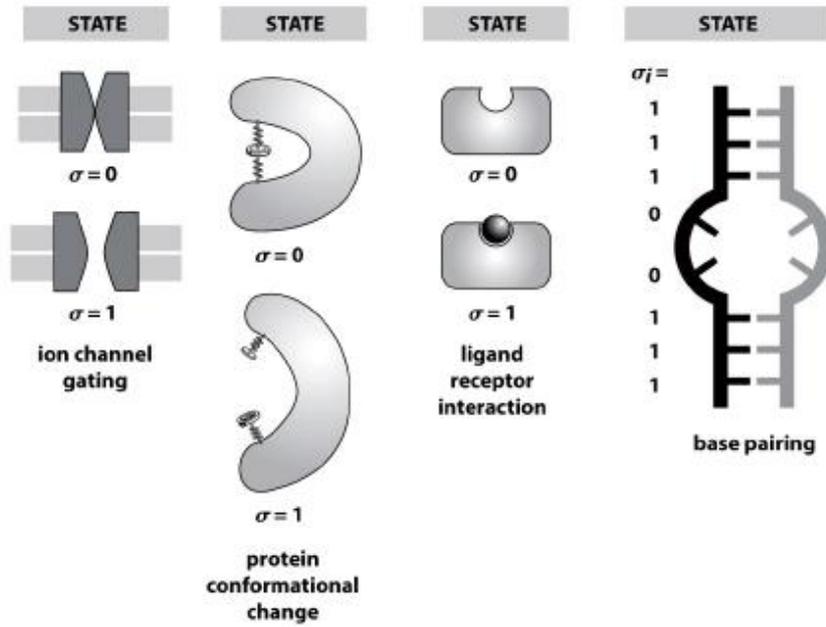
Fig.7.3. Examples of 2-state systems: state variable σ.

Current trace and its two state idealization (transition time or "dwell time")
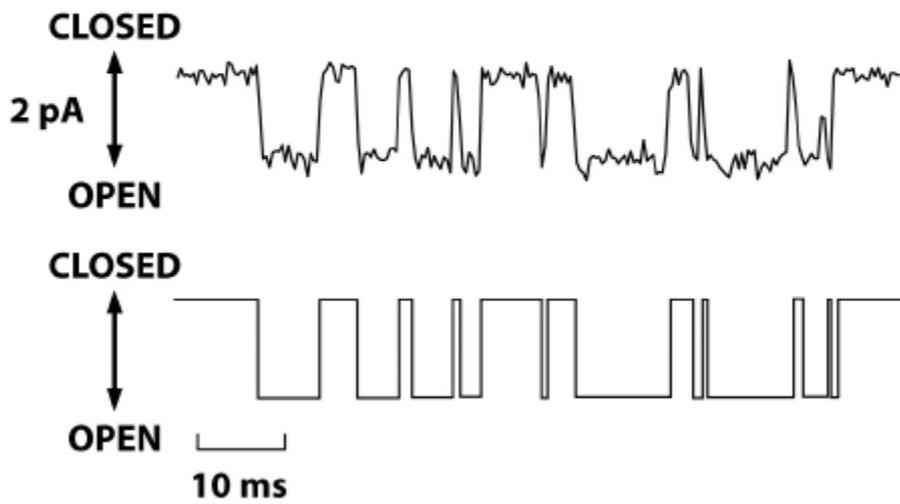
Fig.7.4 Schematic picture of a transition time

Ion channels within a lipid Membrane:

Different voltages result in different currents across the membrane

	applied voltage (mV)	$\varepsilon_{open} - \varepsilon_{closed}$ (k_BT)
	−125	3.24
	−105	1.14
	−95	0.05
	−85	−1.27
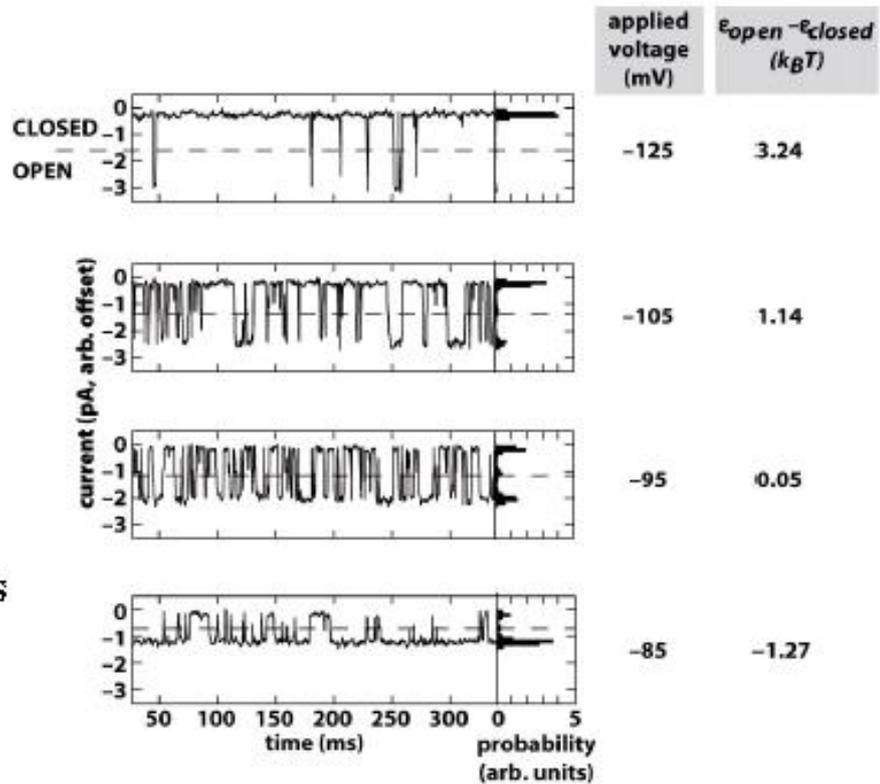
Fig.7.5. Current in ion channels.

Current proportional to the probability of a channel to be open
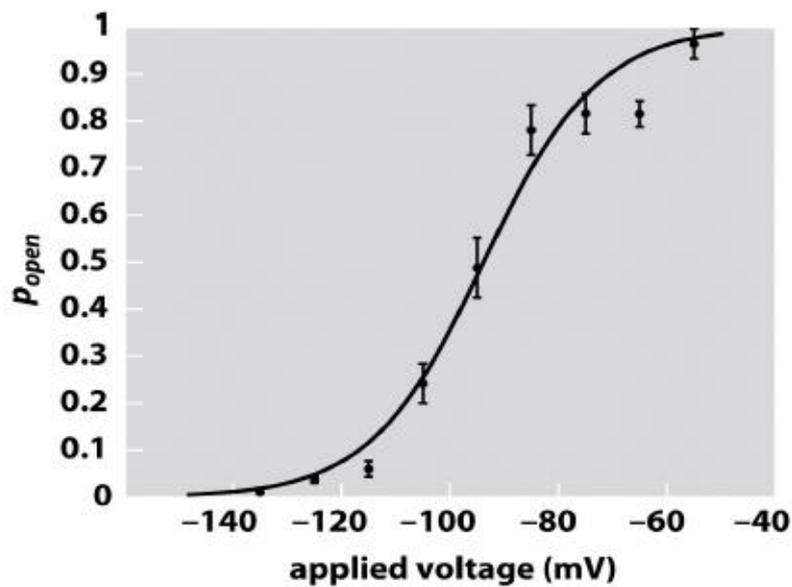
Fig.7.6. Voltage of channel necessary to be open.
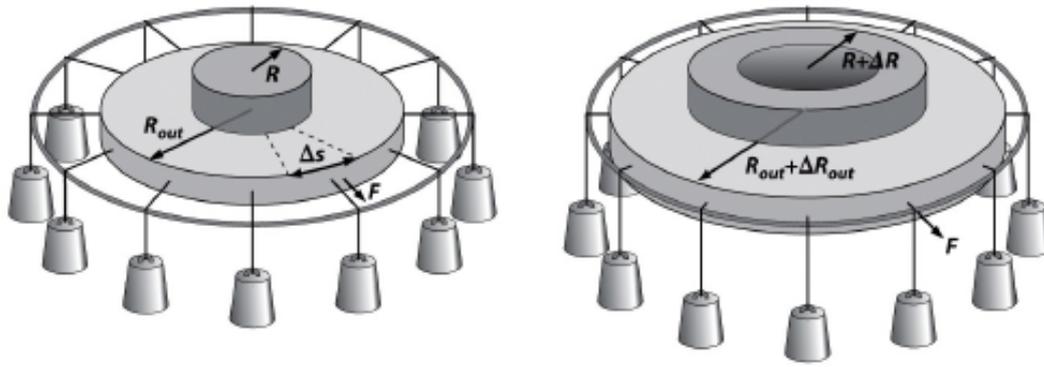
Fig.7.7. Schematic picture of ion channels opening.

When the ion channel opens, the energy of the loading device (weight in the figure 7.6. decreases (potential energy of weights decreases as the weights are lowered.

How does the external tension couple to the energy?

$$\Delta G_{TENSION} = -\tau \Delta s \times \frac{R \Delta R}{R_{OUT}} \times \frac{2\pi R_{OUT}}{\Delta s}$$

$\tau \Delta s$ ... force on the arc

$\frac{R \Delta R}{R_{OUT}}$ ... the outer radius change (parch displacement)

$\frac{2\pi R_{OUT}}{\Delta s}$ ... the number of parches

The outer radius change R_{OUT} follows from the condition that

The membrane area is constant: $\pi R_{OUT}^2 - \pi R^2 = const.$

$$R_{OUT} \Delta R_{OUT} - R \Delta R = 0$$

$$\Delta G_{TENSION} = -\tau 2\pi R \Delta R = -\tau \Delta A$$

Energy as a function of the internal state variable σ

$$(\sigma = 1: open \ \sigma = 0: closed)$$

No external driving force:

$$E(\sigma) = \sigma \varepsilon_{OPEN} + (1 - \sigma)\varepsilon_{CLOSED}$$

With an external driving force:

$$E(\sigma) = \sigma \varepsilon_{OPEN} + (1 - \sigma)\varepsilon_{CLOSED} - \sigma \tau \Delta A$$

Two states with the corresponding weights
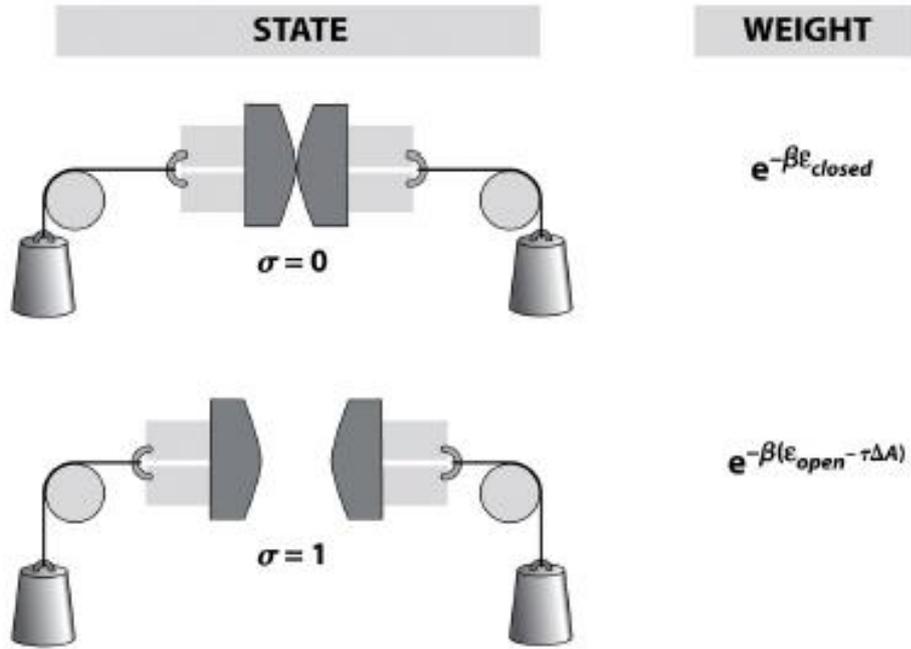
Fig.7.8. Schematic picture of corresponding weights distribution.

Energy landscape for an ion channel as a function of radius R for different forces that promote opening.
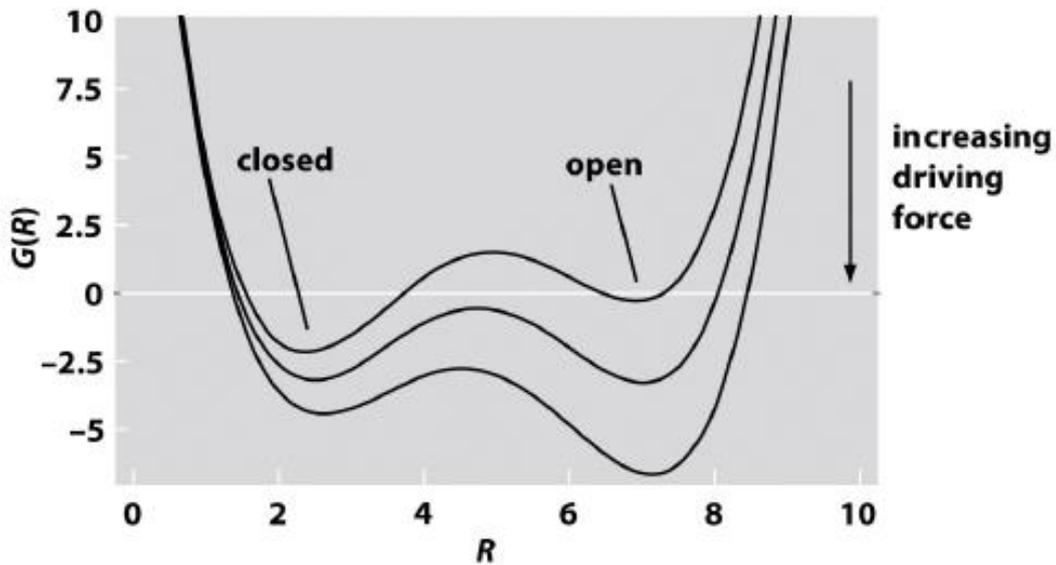
Fig.7.9. Dependence of landscape of ion channel from radius R.

$$Z = \sum_{\sigma=\{0,1\}} exp[-\beta E(\sigma)] = exp[-\beta \varepsilon_{CLOSED}] + exp[-\beta \varepsilon_{Open} + \beta \tau \Delta A]$$

$$\langle \sigma \rangle = p_{Open} = p(1) = \frac{exp[-\beta \varepsilon_{Open} + \beta \tau \Delta A]}{Z}$$

Ion channel opening probability as a function of driving force for gating (membrane tension τ)
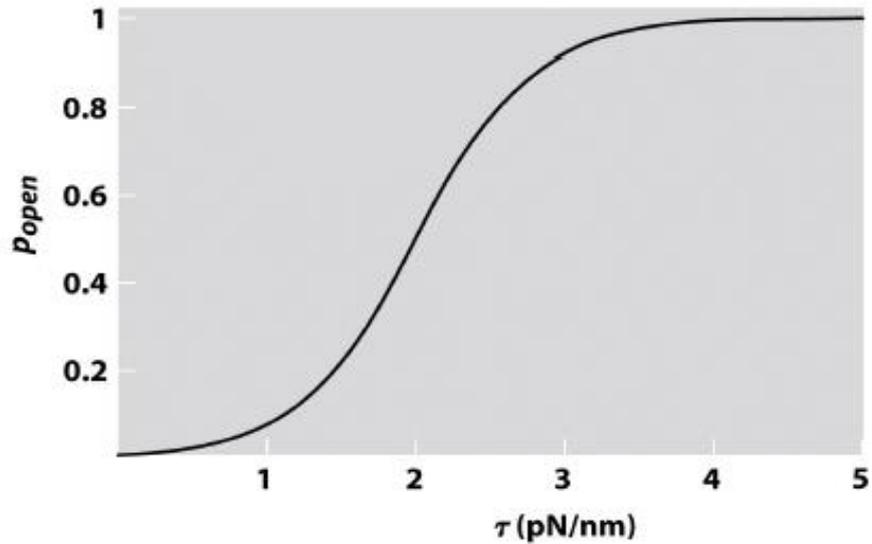
Fig.7.10. Dependence of opening probability from membrane tension.

State variable description of binding: **Grand partition function and Gibbs distribution**

Open system exchanges both energy and particles with the reservoir.
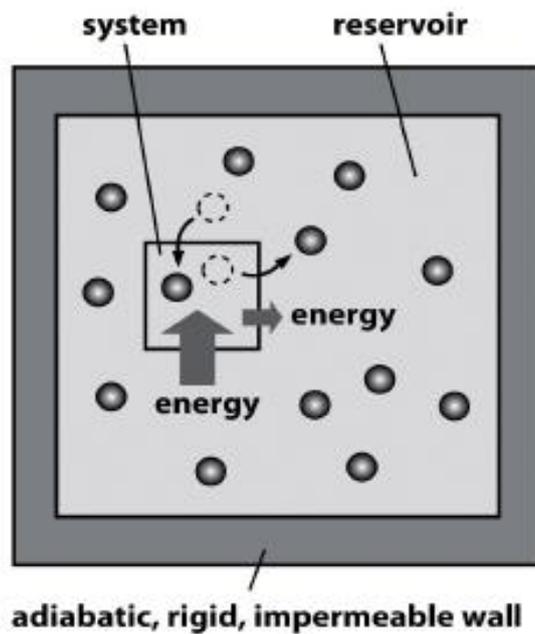
Fig.7.11. Scheme of open system energy and mass exchanges.

We apply the same principle of maximal entropy to the isolated "universe" of the system and the reservoir:

$$W_{TOT}\big(E_{TOT} - E_{S}, N_{TOT} - E_{S}\big) = 1 \times W_{R}\big(E_{TOT} - E_{S}, N_{TOT} - N_{S}\big)$$

Such that the ratio of the probabilities to find system in microstates I and II can be expressed as:

$$\frac{p(E_S^{,I}, N_S^I)}{p(E_S^{,II}, N_S^{II})} = \frac{W_R(E_{TOT} - E_S^{,I}, N_{TOT} - N_S^I)}{W_R(E_{TOT} - E_S^{,II}, N_{TOT} - N_S^{II})}$$

Because $S = k_B lnW$, the above ratio can be expressed as:

$$\frac{W_R(E_{TOT} - E_S^{,I}, N_{TOT} - N_S^I)}{W_R(E_{TOT} - E_S^{,II}, N_{TOT} - N_S^{II})} = \frac{exp\left[\frac{S_R(E_{TOT} - E_S^{,I}, N_{TOT} - N_S^I)}{k_B}\right]}{exp\left[\frac{S_R(E_{TOT} - E_S^{,II}, N_{TOT} - N_S^{II})}{k_B}\right]}$$

$$S_R(E_{TOT} - E_S, N_{TOT} - E_S) = S_R(E_{TOT}) - \frac{\partial S_R}{\partial E} E_S - \frac{\partial S_R}{\partial N} N_S = S_R(E_{TOT}) - \frac{E_S}{T} + \mu \frac{N_S}{T}$$

We derived the Gibbs distribution:

$$p(E_S^{,I}, N_S^I) - \frac{exp[-\beta(E_S^{,I} - \mu N_S^I)]}{Z}$$

$$Z = \sum exp\left[-\beta(E_S^{,I} - \mu N_S^I)\right] \text{ ... grand partition function.}$$

Simple Ligand-Receptor Binding through Gibbs distribution

$$Z = 1 + exp[-\beta(\varepsilon_b - \mu)] \text{ ...}$$

$$p_B = \langle N \rangle = \beta^{-1} \frac{\partial lnZ}{\partial \mu} = \frac{exp[-\beta(\varepsilon_b - \mu)]}{Z}$$

$$\mu = \mu_0 + \beta^{-1} ln \frac{c}{c_0}$$
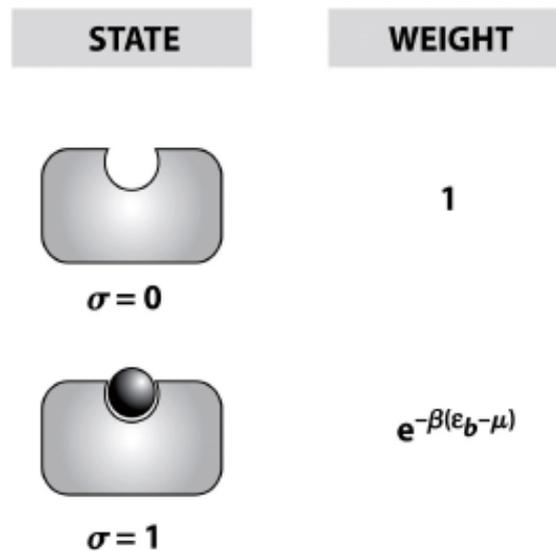
Fig.7.12. Schematic picture of ligand – receptor binding.

Phosphorylation

(covalent post-translational modification)

Attachment of phosphate groups to specific substrates:

- In proteins within eukaryotic cells amino acids with hydroxyl groups (serine, threonine and tyrosine), added a phosphate group with two negative charges
- ATP
- Kinase enzymes assist in phosphorylation
- Reversed process facilitated by another type of enzymes: phosphatases (use water to hydrolyze the bonds)
- Controls protein activity (in response to environmental changes)

Example of two state distribution
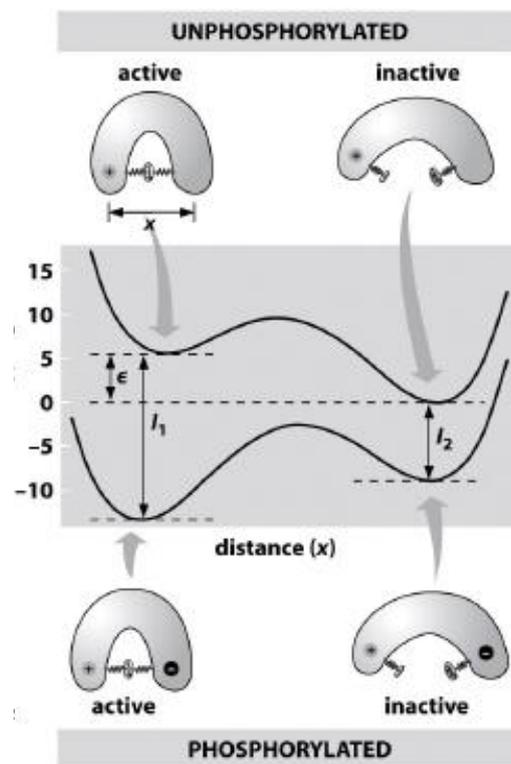
Fig.7.13. Dependence of free energy from distance.

$\sigma_S = 1 \ldots$ active

$\sigma_S = 0 \ldots$ inactive

$\sigma_p = 1 \ldots$ phosphorylated

$\sigma_p = 0 \ldots$ unphosphorylated

Phosphorylation lowers the energy of both states but more the active state.

$$G(\sigma_p, \sigma_S) = (1 - \sigma_p)[(1 - \sigma_S)0 + \varepsilon\sigma_S] + \sigma_p[(1 - \sigma_S)(-I_2) + \sigma_S(\varepsilon - I_1)]$$
$$= \varepsilon\sigma_S - I_2\sigma_p + (I_2 - I_1)\sigma_S\sigma_p$$

$$-p_A = \frac{exp(-\beta\varepsilon)}{1 + exp(-\beta\varepsilon)}$$

For unphosphorylated case

$$-p_A^* = \frac{B}{B + C} \gg p_A$$

For phosphorylated case

$$B = exp[-\beta(\varepsilon - I_1)] \& C = exp(\beta I_2)$$

Factor of 2 up to 1,000 fold increase in activity upon phosphorylation

Free energy of the protein
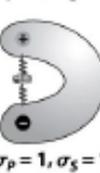
Fig.7.14. Free energy distribution scheme in protein.

Oxygen binding in hemoglobin

- 4 *oxygen binding sites* in each hemoglobin molecule

- *Cooperativity*: binding energy of each oxygen depends on the number of oxygen that are already bound to hemoglobin
- Biophysical basis: each bound O: conformational change that makes is easier for other O-es to bind
- Toy model of dimeric hemoglobin (O_2) dimoglobin binding

$$E = \varepsilon(\sigma_1 + \sigma_2) + J\sigma_1\sigma_2$$
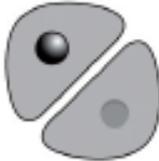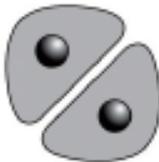
STATE	WEIGHT
	1
	$e^{-\beta(\varepsilon-\mu)}$
	$e^{-\beta(\varepsilon-\mu)}$
	$e^{-\beta(2\varepsilon+J-2\mu)}$

Fig.7.15. Oxygen binding scheme in hemoglobin.

Grand partition function

$$Z = 1 + 2exp[-\beta(\varepsilon - \mu)] + exp\,[-\beta(2\varepsilon - 2\mu + J)]\ldots(p_0, p_1, p_2)$$

$$p_0 = \frac{1}{Z}; \quad p_1 = \frac{2exp[-\beta(\varepsilon - \mu)]}{Z}; \quad p_2 = \frac{exp\,[-\beta(2\varepsilon - 2\mu + J)]}{Z}$$
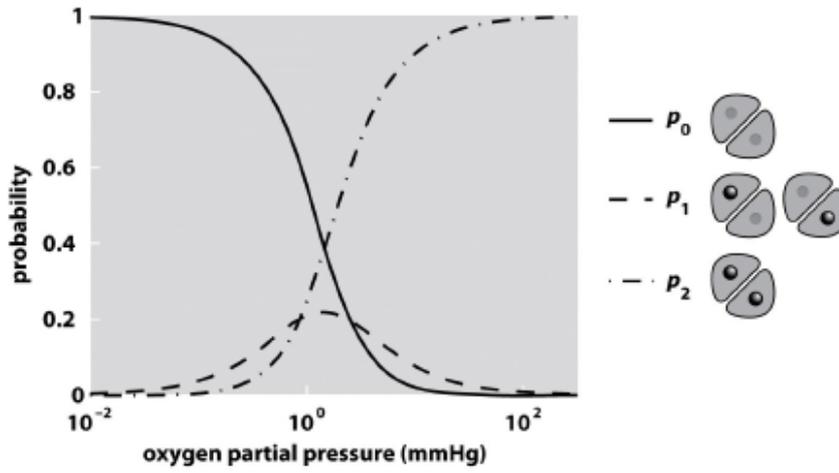
Fig.7.16. Dependence of probability from Oxygen partial pressure.

$$\mu = \mu_0 + k_B T ln\frac{c}{c_0}; \quad \Delta\varepsilon = \varepsilon - \mu_0; \quad x = \frac{c}{c_0};$$

$$\langle N \rangle = \frac{A}{1+A}; \quad A = 2x exp(-\beta\Delta\varepsilon) + 2x^2 exp[-\beta(2\Delta\varepsilon + J)]$$
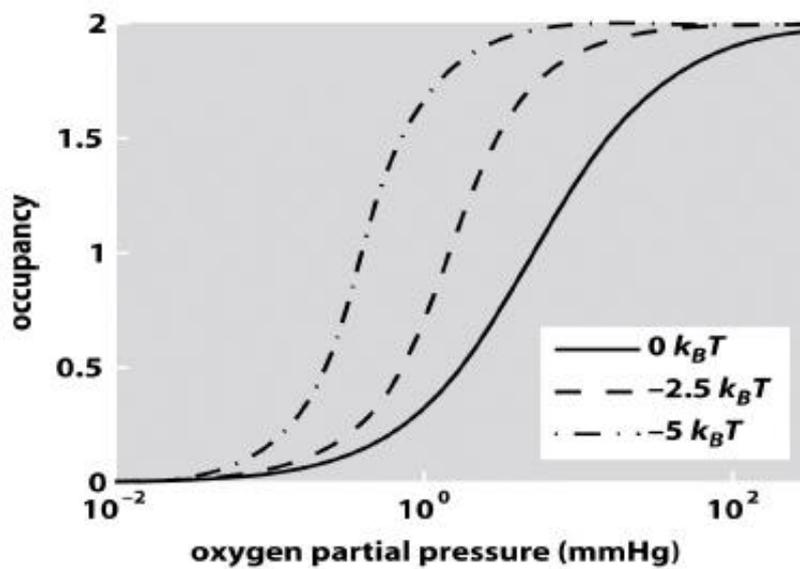
Fig.7.17. Dependence of Oxygen occupancy from its partial pressure.

The Monod-Wyman-Changeux (MWC) Model of Cooperative binding
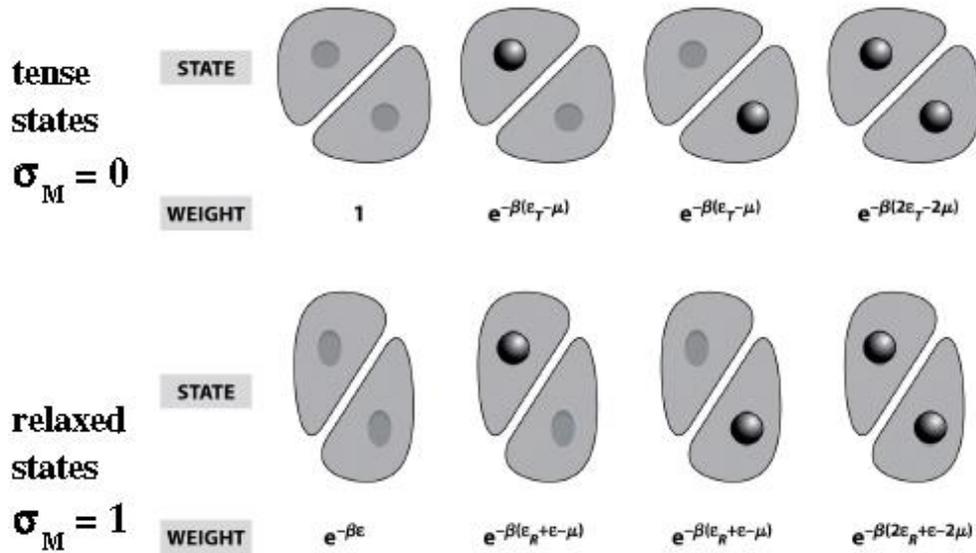
Fig.7.18. Schematic picture of MWC model of cooperative binding.

Description of MWC model of Oxygen binding in dimoglobin

- A classic two-state model for binding

- Protein can exist in two distinct states: T (tense) and R (relaxed)

- T is favored over R: $\varepsilon_T < \varepsilon_R$ so that $\varepsilon = \varepsilon_T - \varepsilon_R > 0$

- The first two-state variable: $\sigma_M = 0$ (Tense), $\sigma_M = 1$ (Relaxed)

- The second two-state variable: $\sigma_i = 0$ (empty), $\sigma_i = 1$ (occupied)

- The variable $i \in \{1,2\}$ corresponds to the receptor place in dimoglobin (two sites of binding)

- Complete energy E:

$$E = (1 - \sigma_M)\varepsilon_T \sum_I \sigma_i + \sigma_M(\varepsilon + \varepsilon_R \sum_I \sigma_i)$$

Tense Part+Relaxed Part

$$Z = 1 + exp[-\beta(\varepsilon_T - \mu)] + exp[-\beta(2\varepsilon_T - 2\mu)]$$
$$+ exp(-\beta\varepsilon) \times \{1 + exp[-\beta(\varepsilon_R - \mu)] + exp[-\beta(2\varepsilon_R - 2\mu)]\}$$

$$\langle N \rangle = \beta^{-1}\left(\frac{\partial}{\partial\mu}\right) lnZ$$

$$\langle N \rangle = \frac{2}{Z} \times [x + x^2 + exp(-\beta\varepsilon)(y + y^2)]$$

$$x = \frac{c}{c_0} exp[-\beta(\varepsilon_T - \mu_0)]$$

$$y = \frac{c}{c_0} exp[-\beta(\varepsilon_R - \mu_0)]$$
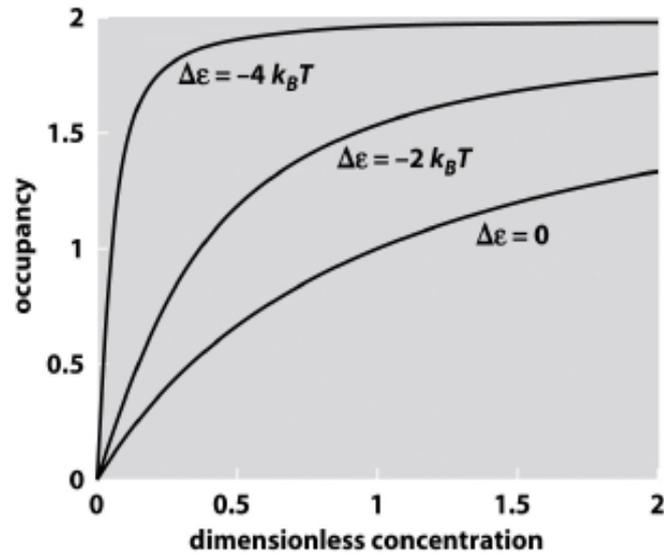
Fig.7.19. Dependence of occupancy from dimensionless concentration.

Increasingly Complex Binding Models for Hemoglobin

(1) Noncooperative Model:

- Binding on the 4 sites of hemoglobin independent

- Energy: $E = \varepsilon \sum_i \sigma_i \quad i \in \{1, 2, 3, 4\}$

(2) Pauling Model:

- Four sites as vertices of a tetrahedron

- Binding to the neighboring sites is favorable

- Energy: $E = \varepsilon \sum_i \sigma_i + \frac{J}{2} \sum_{ij} \sigma_i \sigma_j \quad ij \in \{1, 2, 3, 4\}$

(3) Adair Model:

- In addition to cooperativity of the Pauling model accounts for three- and four-body interactions

- Energy: $E = \varepsilon \sum_i \sigma_i + \frac{J}{2!} \sum_{ij} \sigma_i \sigma_j + \frac{K}{3!} \sum_{ijk} \sigma_i \sigma_j \sigma_k + \frac{L}{4!} \sum_{ijkl} \sigma_i \sigma_j \sigma_k \sigma_l$

non-interacting model

weights	1	$4e^{-\beta(\varepsilon-\mu)}$	$6e^{-2\beta(\varepsilon-\mu)}$	$4e^{-3\beta(\varepsilon-\mu)}$	$e^{-4\beta(\varepsilon-\mu)}$

Pauling model

weights	1	$4e^{-\beta(\varepsilon-\mu)}$	$6e^{-2\beta(\varepsilon-\mu)-\beta J}$	$4e^{-3\beta(\varepsilon-\mu)-3\beta J}$	$e^{-4\beta(\varepsilon-\mu)-6\beta J}$

Adair model

weights	1	$4e^{-\beta(\varepsilon-\mu)}$	$6e^{-2\beta(\varepsilon-\mu)-\beta J}$	$4e^{-3\beta(\varepsilon-\mu)-3\beta J-\beta K}$	$e^{-4\beta(\varepsilon-\mu)-6\beta J-4\beta K-\beta L}$
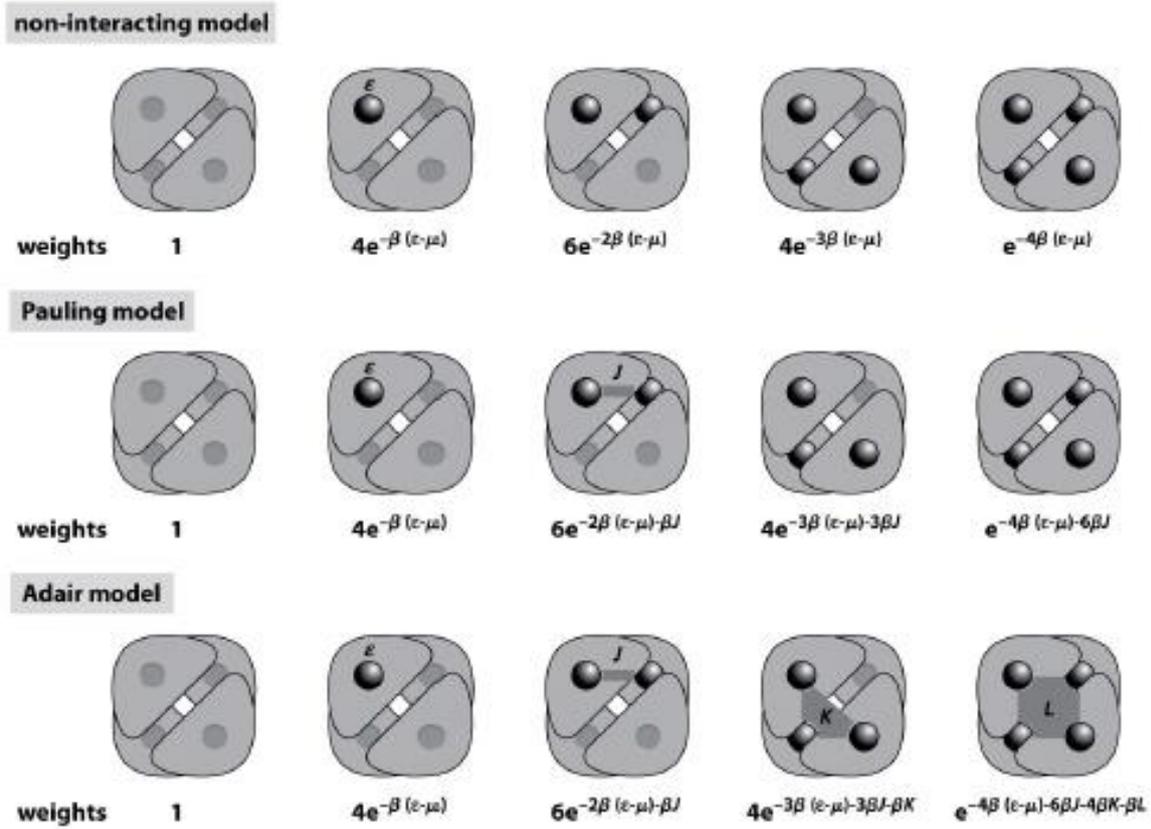
Fig.7.20. Complex binding models for hemoglobin.

Cooperative binding associated with a sharper transition
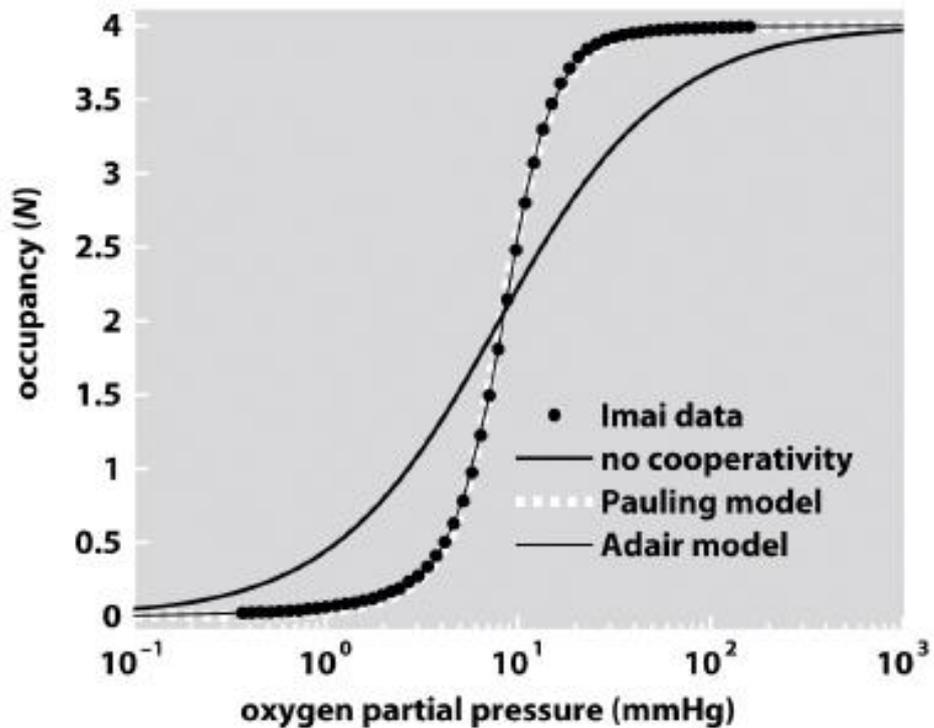
Fig.7.21. Dependence of occupancy from oxygen partial pressure.

Cooperative binding eliminates the intermediate states:(1,2, or 3 oxygens bound to hemoglobin)
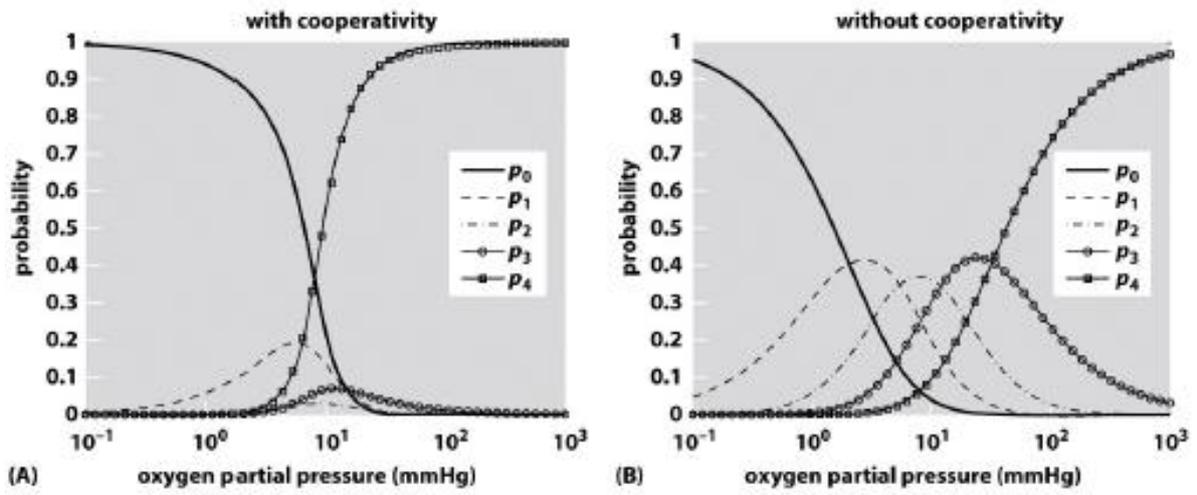
Fig.7.22. Dependence of probability from oxygen partial pressure with and without cooperativity.

Chapter 8

Molecular Structures: The Shape of Molecules

The three dimensional shape or configuration of a molecule is an important characteristic. This shape is dependent on the preferred spatial orientation of covalent bonds to atoms having two or more bonding partners. Three dimensional configurations are best viewed with the aid of models. In order to represent such configurations on a two-dimensional surface (paper, blackboard or screen), we often use perspective drawings in which the direction of a bond is specified by the line connecting the bonded atoms. In most cases the focus of configuration is a carbon atom so the lines specifying bond directions will originate there. As defined in the diagram on the right, a simple straight line represents a bond lying approximately in the surface plane. The two bonds to substituents A in the structure on the left are of this kind. A wedge shaped bond is directed in front of this plane (thick end toward the viewer), as shown by the bond to substituent B; and a hatched bond is directed in back of the plane (away from the viewer), as shown by the bond to substituent D. Some texts and other sources may use a dashed bond in the same manner as we have defined the hatched bond, but this can be confusing because the dashed bond is often used to represent a partial bond (i.e. a covalent bond that is partially formed or partially broken). The following examples make use of this notation, and also illustrate the importance of including non-bonding valence shell electron pairswhen viewing such configurations .
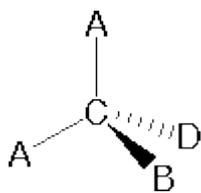
| Methane | Ammonia | Water |
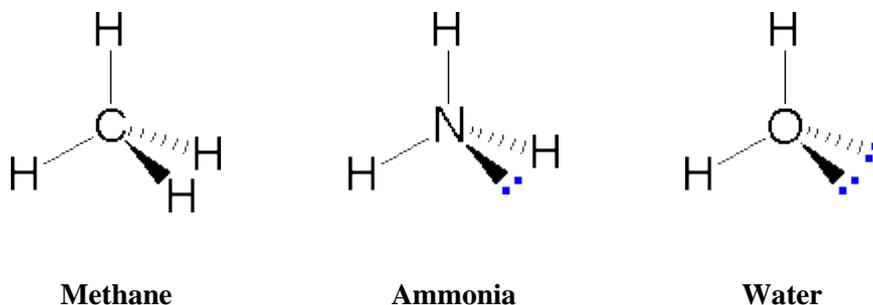
Fig. 8.1. Configuration of valence shell electron pairs.

Bonding configurations are readily predicted by valence-shell electron-pair repulsion theory, commonly referred to as VSEPR in most introductory biophysical texts. This simple model is based on the fact that electrons repel each other, and that it is reasonable to expect that the bonds and non-bonding valence electron pairs associated with a given atom will prefer to be as far apart as possible. The bonding configurations of carbon are easy to remember, since there are only three categories.

In the three examples shown above, the central atom (carbon) does not have any non-bonding valence electrons; consequently the configuration may be estimated from the number of bonding partners alone. For molecules of water and ammonia, however, the non-bonding electrons must be included in the

calculation. In each case there are four regions of electron density associated with the valence shell so that a tetrahedral bond angle is expected. The measured bond angles of these compounds (H_2O 104.5º & NH_3 107.3º) show that they are closer to being tetrahedral than trigonal or linear. Of course, it is the configuration of atoms (not electrons) that defines the the shape of a molecule, and in this sense ammonia is said to be pyramidal (not tetrahedral). The compound boron trifluoride, BF_3, does not have non-bonding valence electrons and the configuration of its atoms istrigonal.

The best way to study the three-dimensional shapes of molecules is by using molecular models. One way in which the shapes of molecules manifest themselves experimentally is through molecular dipole moments. A molecule which has one or more polar covalent bonds may have a dipole moment as a result of the accumulated bond dipoles. In the case of water, we know that the O-H covalent bond is polar, due to the different electronegativities of hydrogen and oxygen. Since there are two O-H bonds in water, their bond dipoles will interact and may result in a molecular dipole which can be measured. The following diagram shows four possible orientations of the O-H bonds.
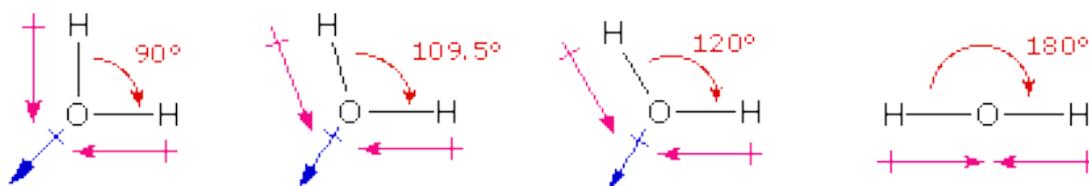
Fig. 8.2. Diagram of orientation of O - H bonds.

The bond dipoles are colored magenta and the resulting molecular dipole is colored blue. In the linear configuration (bond angle 180º) the bond dipoles cancel, and the molecular dipole is zero. For other bond angles (120 to 90º) the molecular dipole would vary in size, being largest for the 90º configuration. In a similar manner the configurations of methane (CH_4) and carbon dioxide (CO_2) may be deduced from their zero molecular dipole moments. Since the bond dipoles have canceled, the configurations of these molecules must be tetrahedral (or square-planar) and linear respectively.

Substitution of one hydrogen by a chlorine atom gives a CH_3Cl compound. Since the tetrahedral, square-planar and square-pyramidal configurations have structurally equivalent hydrogen atoms, they would each give a single substitution product. However, in the trigonal-pyramidal configuration one hydrogen (the apex) is structurally different from the other three (the pyramid base). Substitution in this case should give two different CH_3Cl compounds if all the hydrogens react. In the case of disubstitution, the tetrahedral configuration of methane would lead to a single CH_2Cl_2 product, but the other configurations would give two different CH_2Cl_2 compounds.

Isomers

It is necessary to draw structural formulas for organic compounds because in most cases a molecular formula does not uniquely represent a single compound. Different compounds having the same molecular formula are called isomers, and the prevalence of organic isomers reflects the extraordinary versatility of carbon in forming strong bonds to itself and to other elements.

When the group of atoms that make up the molecules of different isomers are bonded together in fundamentally different ways, we refer to such compounds as constitutional isomers. There are seven constitutional isomers of $C_4H_{10}O$, and structural formulas for these are drawn in the following table. These formulas (Kekulé Formula)represent all known and possible $C_4H_{10}O$ compounds, and display a common structural feature.

Simplification of structural formulas may be achieved without any loss of the information they convey. In condensed structural formulas the bonds to each carbon are omitted, but each distinct structural unit (group) is written with subscript numbers designating multiple substituents, including the hydrogens. Shorthand (line) formulas omit the symbols for carbon and hydrogen entirely. Each straight line segment represents a bond, the ends and intersections of the lines are carbon atoms, and the correct number of hydrogens is calculated from the tetravalency of carbon. Non-bonding valence shell electrons are omitted in these formulas.   Developing the ability to visualize a three-dimensional structure from two-dimensional formulas requires practice, and in most cases the aid of molecular models.

Distinguishing Carbon Atoms

When discussing structural formulas, it is often useful to distinguish different groups of carbon atoms by their structural characteristics. A primary carbon (1º) is one that is bonded to no more than one other carbon atom. A secondary carbon (2º) is bonded to two other carbon atoms, and tertiary (3º) and quaternary (4º) carbon atoms are bonded respectively to three and four other carbons. The three C_5H_{12} isomers shown below illustrate these terms.
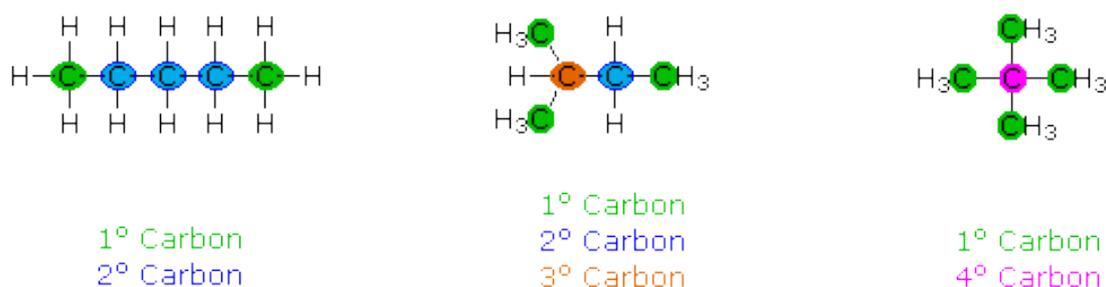
Fig. 8.3. Scheme of illustration of carbon atoms' bonds.

90

Structural differences may occur within these four groups, depending on the molecular constitution. In the formula on the right all four 1º-carbons are structurally equivalent (remember the tetrahedral configuration of tetravalent carbon); however the central formula has two equivalent 1º-carbons (bonded to the 3º carbon on the left end) and a single, structurally different 1º-carbon (bonded to the 2º-carbon) at the right end. Similarly, the left-most formula has two structurally equivalent 2º-carbons (next to the ends of the chain), and a structurally different 2º-carbon in the middle of the chain. A consideration of molecular symmetry helps to distinguish structurally equivalent from nonequivalent atoms and groups.

Resonance

Kekulé structural formulas are essential tools for understanding behavior of molecules. . However, the structures of some compounds and ions cannot be represented by a single formula. For example, sulfur dioxide (SO_2) and nitric acid (HNO_3) may each be described by two equivalent formulas. For clarity the two ambiguous bonds to oxygen are given different colors in these formulas.

1) sulfur dioxide

2) nitric acid
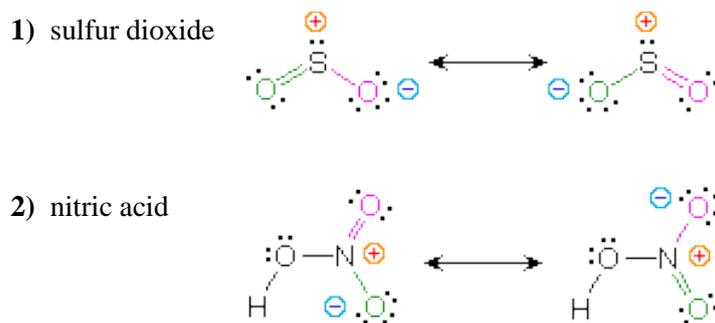
If only one formula for sulfur dioxide was correct and accurate, then the double bond to oxygen would be shorter and stronger than the single bond. Since experimental evidence indicates that this molecule is bent (bond angle 120º) and has equal length sulfur : oxygen bonds (1.432 Å), a single formula is inadequate, and the actual structure resembles an average of the two formulas. This averaging of electron distribution over two or more hypothetical contributing structures to produce a hybrid electronic structure is called resonance. Likewise, the structure of nitric acid is best described as a resonance hybrid of two structures, the double headed arrow being the unique symbol for resonance.

The above examples represent one extreme in the application of resonance. Here, two structurally and energetically equivalent electronic structures for a stable compound can be written, but no single structure provides an accurate or even an adequate representation of the true molecule. In cases such as these, the electron delocalization described by resonance enhances the stability of the molecules, and compounds or ions composed of such molecules often show exceptional stability.

3) formaldehyde
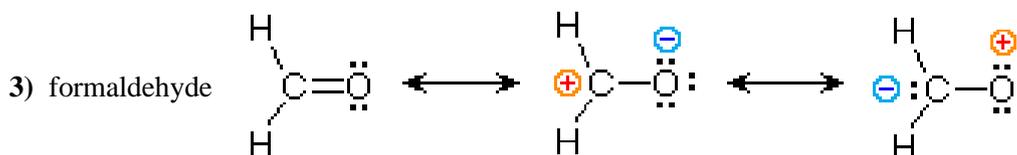
The electronic structures of most covalent compounds do not suffer the inadequacy noted above. Thus, completely satisfactory Kekulé formulas may be drawn for water (H_2O), methane (CH_4) and acetylene C_2H_2). Nevertheless, the principles of resonance are very useful in rationalizing the chemical behavior of many such compounds. For example, the carbonyl group of formaldehyde (the carbon-oxygen double bond) reacts readily to give addition products. The course of these reactions can be explained by a small contribution of a dipolar resonance contributor, as shown in equation 3. Here, the first contributor (on the left) is clearly the best representation of this molecular unit, since there is no charge separation and both the carbon and oxygen atoms have achieved valence shell neon-like configurations by covalent electron sharing. If the double bond is broken heterolytically, formal charge pairs result, as shown in the other two structures. The preferred charge distribution will have the positive charge on the less electronegative atom (carbon) and the negative charge on the more electronegative atom (oxygen). Therefore the middle formula represents a more reasonable and stable structure than the one on the right. The application of resonance to this case requires a weighted averaging of these canonical structures. The double bonded structure is regarded as the major contributor, the middle structure a minor contributor and the right hand structure a non-contributor. Since the middle, charge-separated contributor has an electron deficient carbon atom, this explains the tendency of electron donors (nucleophiles) to bond at this site.

The basic principles of the resonance method may now be summarized:

- The number of covalent bonds in a structure. (The greater the bonding, the more important and stable the contributing structure.)

- Formal charge separation. (Other factors aside, charge separation decreases the stability and importance of the contributing structure.)

- Electronegativity of charge bearing atoms and charge density. (High charge density is destabilizing. Positive charge is best accommodated on atoms of low electronegativity, and negative charge on high electronegative atoms.)

- The stability of a resonance hybrid is always greater than the stability of any canonical contributor. Consequently, if one canonical form has a much greater stability than all others, the hybrid will closely resemble it electronically and energetically. This is the case for the carbonyl group.

Molecular Orbitals

Just as the valence electrons of atoms occupy atomic orbitals (AO), the shared electron pairs of covalently bonded atoms may be thought of as occupying molecular orbitals (MO). It is convenient to approximate molecular orbitals by combining or mixing two or more atomic orbitals. In general, this mixing of n atomic orbitals always generates n molecular orbitals. The hydrogen molecule provides a simple example of MO formation. In the following diagram, two 1s atomic orbitals combine to give a sigma (σ) bonding (low energy) molecular orbital and a second higher energy MO referred to as an antibonding orbital. The bonding MO is occupied by two electrons of opposite spin, the result being a covalent bond.
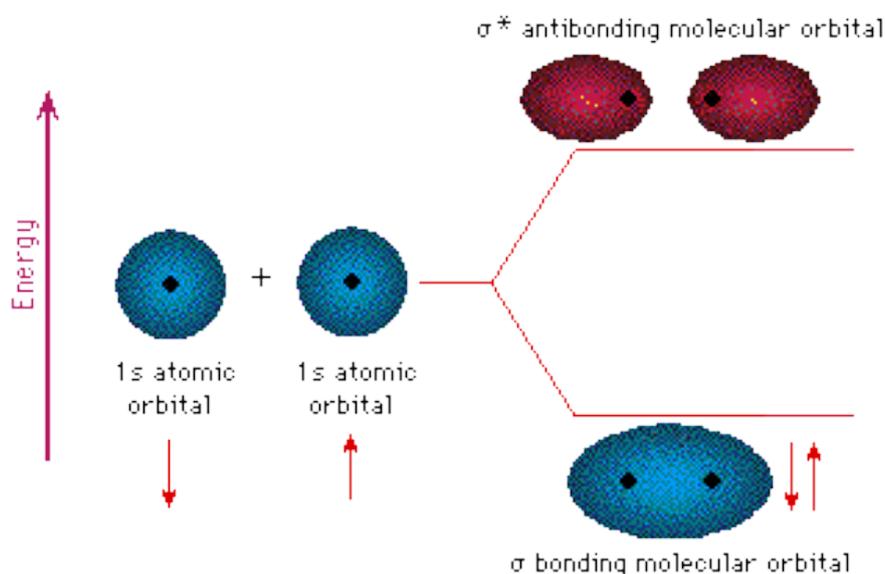
Fig. 8.4. Scheme of building of molecular orbitals.

The notation used for molecular orbitals parallels that used for atomic orbitals. Thus, s-orbitals have a spherical symmetry surrounding a single nucleus, whereas σ-orbitals have a cylindrical symmetry and encompass two (or more) nuclei. In the case of bonds between second period elements, p-orbitals or hybrid atomic orbitals having p-orbital character are used to form molecular orbitals.
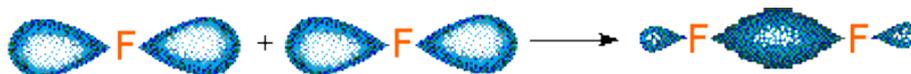
Fig. 8.5. σ-orbital formation scheme.

Another type of MO (the π orbital) may be formed from two p-orbitals by a lateral overlap, as shown in the following diagram. Since bonds consisting of occupied π-orbitals (pi-bonds) are weaker than sigma bonds, pi-bonding between two atoms occurs only when a sigma bond has already been established.
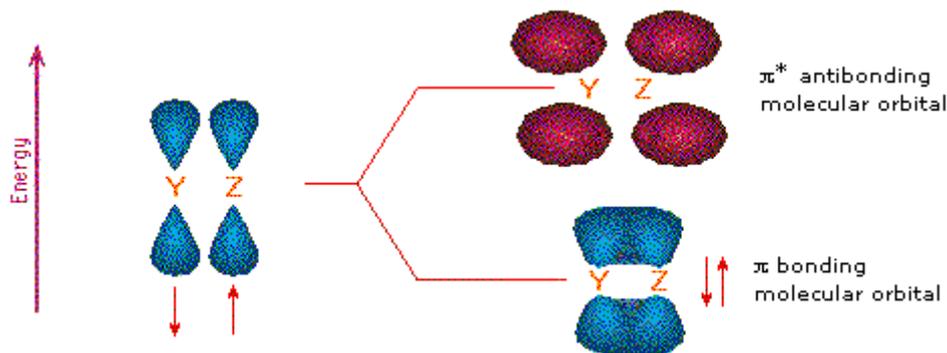
Fig. 8.6. π – orbital formation from 2 p-orbitals.

The manner in which atomic orbitals overlap to form molecular orbitals is actually more complex than the localized examples given above. These are useful models for explaining the structure and reactivity of many organic compounds, but modern molecular orbital theory involves the creation of an orbital correlation diagram. Two examples of such diagrams for the simple diatomic elements F_2 and N_2 will be drawn above when the appropriate button is clicked. The 1s and 2s atomic orbitals do not provide any overall bonding, since orbital overlap is minimal, and the resulting sigma bonding and antibonding components would cancel. In both these cases three 2p atomic orbitals combine to form a sigma and two pi-molecular orbitals, each as a bonding and antibonding pair. The overall bonding order depends on the number of antibonding orbitals that are occupied.

Molecular structure as a random walk

The basis of random walk theory can be traced back to the irregular motion of individual pollen particles, famously studied by the botanist Brown (1828), now known as Brownian motion. Classical works on probability have been in existence for centuries, so it is somewhat surprising that it was only at the beginning of the twentieth century that a random walk was described in the literature, when the journal *Nature* published a discussion between Pearson (1905) and Rayleigh (1905). Physicists, such as Einstein (1905, 1906) and Smoluchowski (1916), were then drawn to the subject and many important fields, such as random processes, random noise, spectral analysis and stochastic equations, were developed during the course of research on random walks. Random walk theory was further developed with the mean-reversion process.

The first simple models of movement using random walks are uncorrelated and unbiased. In this context, uncorrelated means the direction of movement is completely independent of the previous directions moved: the location after each step taken in the random walk is dependent only on the location in the previous step and the process is Markovian with regard to the location. Unbiased means there is no preferred direction: the direction moved at each step is completely random. Assuming that movement in any direction is

allowed, this process is essentially Brownian motion and such models can be shown to produce the standard diffusion (or heat) equation.

Correlated random walks (CRWs) involve a correlation between successive step orientations, which is termed 'persistence'. This produces a local directional bias: each step tends to point in the same direction as the previous one, although the influence of the initial direction of motion progressively diminishes over time and step orientations are uniformly distributed in the long term

The simple isotropic random walk model (SRW) is the basis of most of the theory of diffusive processes. The walk is isotropic, or unbiased, meaning that the walker is equally likely to move in each possible direction and uncorrelated in direction, meaning that the direction taken at a given time is independent of the direction at all preceding times.

Consider a walker moving on an infinite one-dimensional uniform lattice (i.e. a line split into discrete points). Suppose the walker starts at the origin ($x=0$) and then moves a short distance δ either left or right in a short time τ. The motion is assumed to be completely random, so the probabilities of moving both left and right are 1/2. After one time step, the walker can either be at a distance δ to the left or right of the origin, with probability 1/2 each. After the next time step, the walker will either be at a distance 2δ to the left or right of the origin (with probability 1/4 each) or will have returned to the origin (with probability 1/2). Note that, after an even (odd) number of steps, the walker can only be at an even (odd) distance away from the origin. Continuing in this way, the probability that a walker will be at a distance $m\delta$ to the right of the origin after n time steps (where m and n are even) is given by

$$p(m,n) = \left(\frac{1}{2}\right)^n \left(n \, \frac{n}{2} \, m\right) = \frac{n!}{2^n \left(\frac{(n+m)}{2}!\right)\left(\frac{(n-m)}{2}\right)!} \tag{8.1}$$

This is a form of the binomial distribution, with mean 0 and variance n. For large n, this converges to a normal (or Gaussian) distribution so, after a sufficiently large amount of time $t=n\tau$, the location $x=m\delta$ of the walker is normally distributed with mean 0 and variance $\delta^2 t/\tau$. Taking the limit $\delta, \tau \rightarrow 0$ such that $\delta^2/\tau = 2D$, where D is a constant known as the diffusion coefficient, gives the PDF for the location of the walker after time t

$$p(x,t) = \frac{1}{\sqrt{2\pi Dt}} exp\left(\frac{-x^2}{4Dt}\right). \tag{8.2}$$

Note that this is the fundamental solution of the *diffusion equation*.

Useful time-dependent statistics of this process are the mean location $E(X_t)$ and the mean squared displacement (MSD) $E(X_t^2)$, defined as

$$\begin{cases} E(X_t) = \displaystyle\int_{-\infty}^{\infty} x p(x,t)dx, \\ E(X_t^2) = \displaystyle\int_{-\infty}^{\infty} x^2 p(x,t)dx. \end{cases} \tag{8.3}$$

Note that here, and throughout the paper, we adopt the convention that random variables are denoted by upper case letters (e.g. X, Θ), while the possible numerical values of these variables are denoted by lower case letters. For the one-dimensional solution (8.2), it is easy to show that $E(X_t)=0$ and

$$E(X_t^2) = 2Dt$$

The first result illustrates the absence of a preferred direction or bias, while the second result illustrates the standard property of a diffusive process—that MSD increases linearly with time. This contrasts with a system or process where the signal propagates as a wave (ballistic movement), in which MSD increases. This relationship between MSD and D, and its equivalence in higher dimensions, is extremely important as it provides a means of estimating from empirical data the diffusion coefficient D that is used in many mathematical models of spatial population dynamics and diffusive processes.

A similar derivation can be completed using an N-dimensional lattice to give the standard drift–diffusion equation

$$\frac{\partial p}{\partial t} = -u \cdot \nabla p + D\nabla^2 p \tag{8.4}$$

where $\boldsymbol{u}$ is the average drift velocity (now an N-dimensional vector); ∇ is the gradient operator; and ∇^2 is the Laplacian.
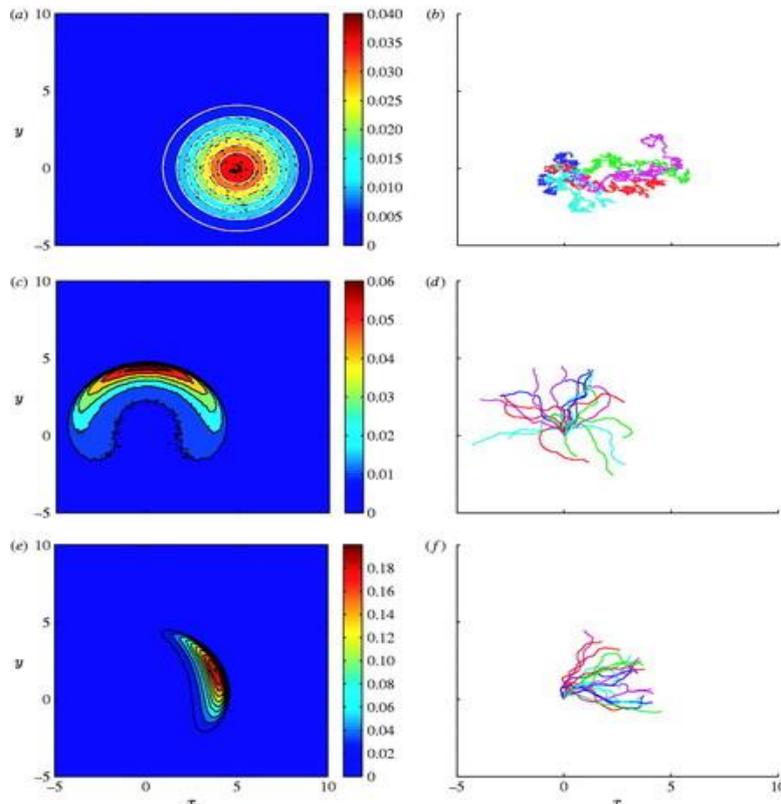
Fig.8.7. (*a,c,e*) PDFs and (*b,d,f*) sample paths of different random walks.

At this figure (*a,b*) A lattice BRW with probabilities of moving a distance δ right or left of $\tau(D/\delta^2 \pm u/(2\delta))$ and up or down of $\tau D/\delta^2$. (*c,d*) A non-lattice CRW with probabilities of turning an angle δ_θ clockwise or anticlockwise of $\frac{\tau\sigma_0^2}{(2\delta_\theta^2)}$. (*e,f*) A non-lattice BCRW with probabilities of turning clockwise or anticlockwise of $\tau\left(\frac{\sigma_0^2}{(2\delta_\theta^2)} \pm \frac{\theta}{(2B\delta_\theta)}\right)$. In the BRW and BCRW, the global preferred direction is $\theta_0=0$; in the CRW and BCRW, the initial direction is $\theta=\pi/2$ and the walker moves with constant speed v. In all cases, the walker starts at $(x,y)=(0,0)$ at $t=0$ and is allowed to move until $t=10$. The PDFs $p(x, y, t=10)$ were calculated from 10^6 realizations of the walk. In (*a*), the white lines show the contours of the corresponding theoretical PDF **(2.12)**. In the sample paths for the BRW, at each step the walker either stays still or moves right, left, up or down by a distance δ. In the CRW and BCRW, at each step the walker's direction of motion θ either stays the same or turns clockwise or anticlockwise by an angle δ_θ, and the walker's movement is given by the vector $v\tau(\cos\theta, \sin\theta)$.

Random walks as models of cell movement

Cell movements are often characterized by some directional correlation (persistence) and, unfortunately, with a CRW, it is not usually possible to calculate $p(\pmb{x},t)$ directly, or even to derive a system of differential equations for $p(\pmb{x},t)$. Variations of the telegraph equation can be used to model a one-dimensional CRW and $p(\pmb{x},t)$ (and associated moments) can be found. However, it is still a non-trivial problem to derive a solution for $p(\pmb{x},t)$ for a CRW in higher dimensions but it is, in many cases, still possible to calculate statistics of the CRW directly through the analysis of paths.

The situation is slightly more complex when movement is both correlated and biased in a global preferred direction (i.e. a BCRW). Simple extensions of the BRW model is possible to introduce as generalized mass-balance equation (the transport equation) that describes hyperbolic movement and discuss how this can be used as a general framework for modelling BCRW.

In general, for organisms moving in an environment that is varying spatially and/or temporally, transition probabilities will depend explicitly on the time t and walker's location $\pmb{x}$. Typically, this dependence is via some 'control signal', such as a chemical substance, light, heat, humidity or odour. A control signal can stimulate the organism in four main ways: the stimulus may be an attractant (or repellent), providing a directional bias that stimulates the organisms to migrate up (or down) a concentration gradient field, or may be an inducer (or inhibitor), causing the rate of diffusive unbiased movement to increase (or decrease).

As a further complication, it is common for migrating cells to modify their own bio environment by producing or degrading the control substance. For example, the slime mould *Dictyostelium discoideum* secretes cyclic adenosine monophosphate (cAMP), which acts as a chemoattractant, leading to

the aggregation of cells from a wide area; certain types of bacteria secrete slime trails, which provide directional guidance for other cells. The basic theory of reinforced random walks (RRWs) is introduced, together with some models for the transition probabilities which lie at the heart of the RRW description.

Some examples of Random Walks in a molecular biosystems

Description of the macromolecular structure as random walks:

- Motivation: calculate entropic cost of DNA packed into a cell; description of DNA stretching by optical tweezers
- Atomic coordinates deposited on databases such as Protein Data Bank (pdb format of an ascii file with x,y,z coordinates of know protein structures) $(r_1, r_2 \ldots r_n)$ $with$ $r_i = (x_1, y_1, z_1)$
- Data derived based on X-ray crystallography or NMR
- Statistical measure of the size of a macromolecule: Radius of Gyration R_G
- Example of DNA characterization $r(S)$ where s is the distance along the contour of the molecule.

Random walk model of a polymer
Series of rigid rods – Kuhn segments connected by flexible hinges
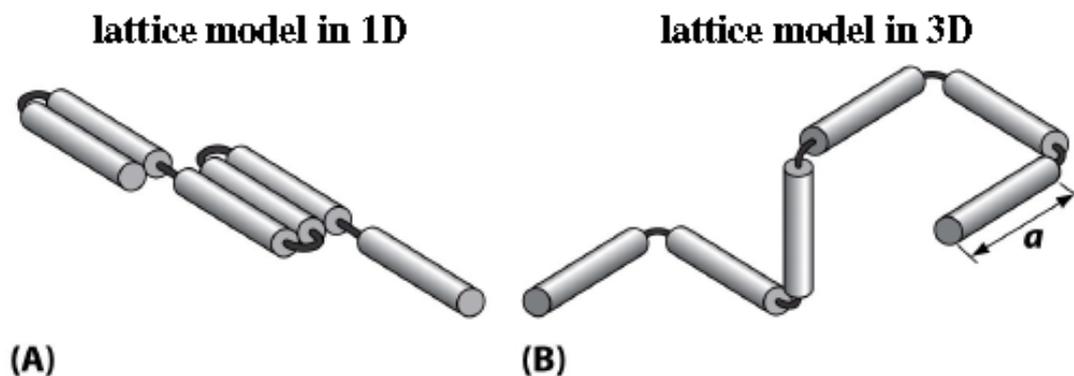
Fig.8.8. Random walks of polymer in one and three dimensions.

DNA molecule as a random walk
Every macromolecular configuration ai equally probable
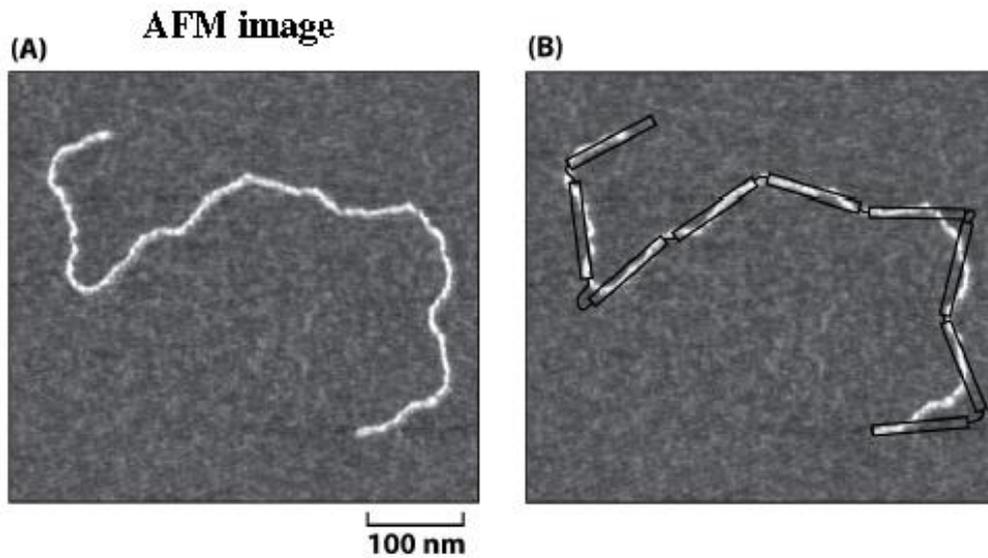
Fig.8.9. Atom Force Microscope images of DNA molecule movement.

Derivation based on probability theory

- a macroscopic state: many microscopic states

- N steps in 1D random walk: $N = n_R + n_L$, each with a probability of ½

- A probability associated with each microstate of N steps is $\left(\frac{1}{2}\right)^N$

- Multiplicity W associated with n_R steps is:

$$W = \frac{N!}{[n_R! \, (N - n_R)!]}$$
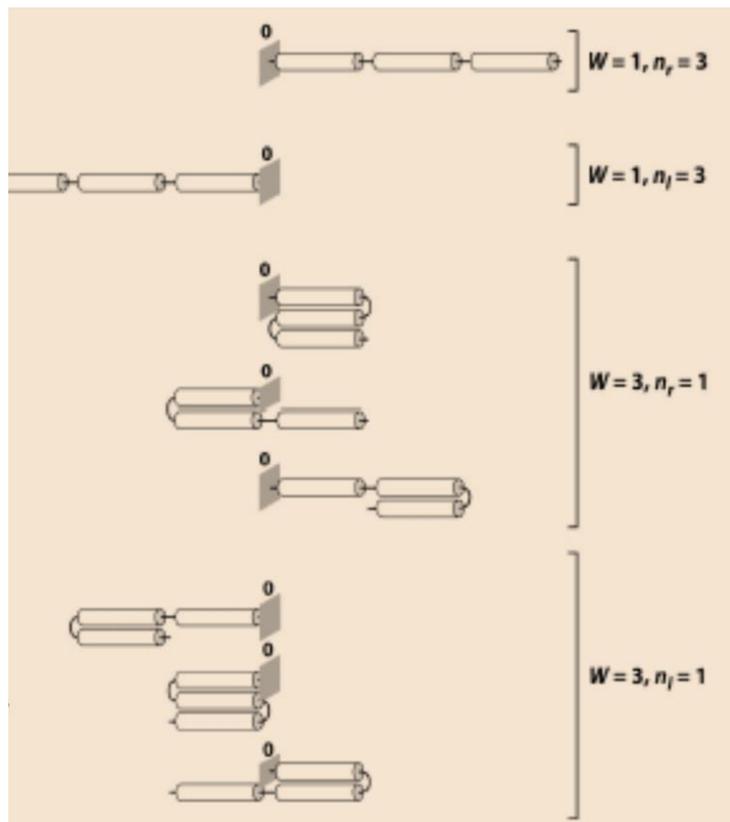
Probability of an overall departure n_R from the origin:

$$p(n_R; N) = \frac{N!}{[n_R!\,(N - n_R)!]}\left(\frac{1}{2}\right)^N$$

This distribution is normalized:

$$\sum_{nR} p(n_R; N) = 1, \quad n_R \in \{0, N\}$$

Probability distribution for the end-to-end distance:

$$R = a(n_R - n_L) = a(2n_R - N); \quad n_R = \frac{N}{2} + \frac{R}{(2a)}$$

Such that

$$p(R; N) = \frac{\left(\frac{1}{2}\right)^N N!}{\left\{\left[\frac{N}{2} + \frac{R}{(2a)}\right]!\,\left[\frac{N}{2} - \frac{R}{(2a)}\right]!\right\}}$$

Which takes the form of a Gaussian distribution for $R \ll Na$:

$$p(R; N) = \frac{2}{\sqrt{2\pi N}} \; exp\left[-\frac{R^2}{(2Na^2)}\right]$$

To obtain the probability distribution function for the end-to-end distance of a freely jointed chain $P(R; N)$, $p(R; N)$ needs to be divided by the number of integer R values per unit length $(= 2a)$:

$$P(R; N) = \frac{2}{\sqrt{2\pi Na^2}} \; exp\left[-\frac{R^2}{(2Na^2)}\right]$$

Note that $\langle R \rangle = 0$ and $\langle R^2 \rangle = Na^2$, independent of the space Dimension. This can be used to derive $P(R; N)$ in 3D space:

$$P(R; N) = A \; exp(-\kappa R^2);$$

$$A = \left[\frac{3}{(2\pi Na^2)}\right]^{\frac{3}{2}} ; \quad \kappa = \frac{3}{(2Na^2)}$$

Use normalization condition in 3D and calculate the variance to determine A and κ (integration in 3D).

End – to end distance probability distribution for 1D random walk

Binomial versus of Gaussian distribution for N=100
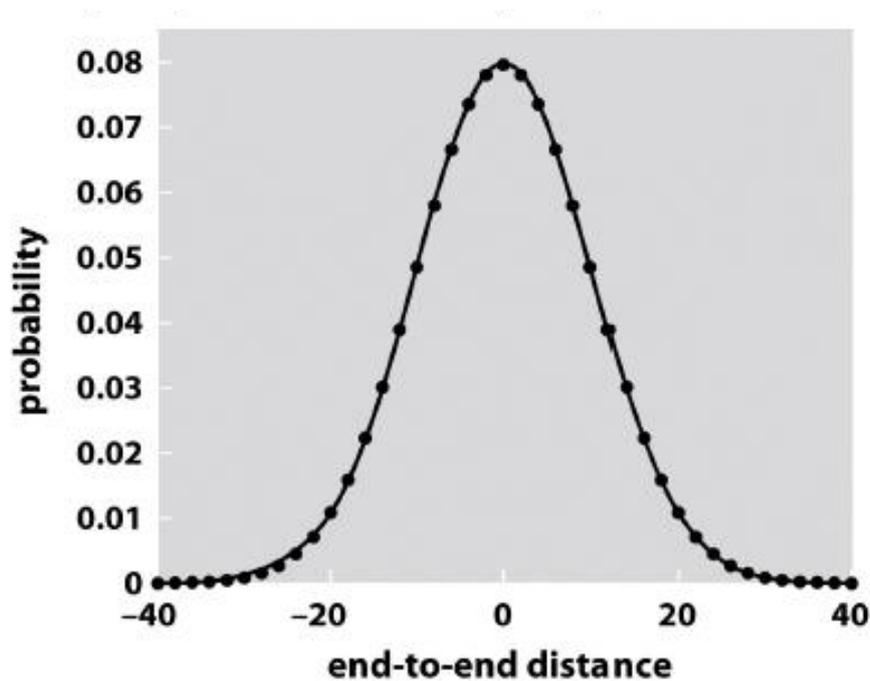
Fig.8.10. Gaussian like distribution of dependence of probability from end-to-end distance.

Persistence Length versus the Kuhn Length

- Persistence length: the length scale ξ_P over which the polymer remains approximately straight
- The Kuhn length: the length of a step a in the random walk model

$$\langle R^2 \rangle = La$$

- persistence length ξ_P can be calculated from the unit tangent vector $t(s)$ where s is a distance along the polymer:

$$\langle t(s)t(u) \rangle = exp\left(-Is - \frac{uI}{\xi_P}\right)$$

- To find the relationship between a and ξ_P we use $R = \int_0^L ds\, t(s)$:

$$\langle R^2 \rangle = \langle \int_0^L ds\, t(s) \int_0^L du\, t(u) \rangle = \int_0^L \int_0^L ds\, du\, \langle d(s)\, t(u) \rangle = 2\int_0^L ds \int_s^L du\, exp\left[-\frac{(s-u)}{\xi_P}\right] =$$

$$= 2\int_0^L ds \int_s^\infty dx\, exp\left[-\frac{x}{\xi_P}\right] = 2L\xi_P$$

$$a = 2\xi_P$$

Size of Genomic DNA in solution

$$\langle R^2 \rangle = 2L\xi_P$$

$$\langle R_G^2 \rangle > \frac{1}{3L\xi_P}$$

(use definition)

$$L = 0,34nm\, N_{BP}$$

$$N_{BP\ ...}\ \#\ of\ basepairs$$

$$R_G = \frac{1}{3}\sqrt{N_{BP}\xi_P}nm$$
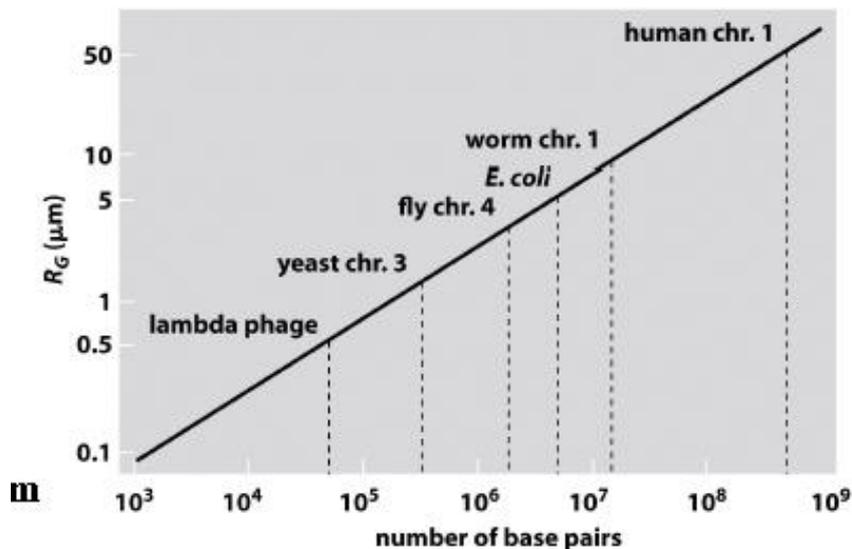
Fig.8.11. Distribution of size of Genomic DNA in solution

Chapter 9

Structure of Biosystems

One of our goals is to understand the molecular mechanism of working macro biosystems such as DNA, lipid, and protein. A simple way to do this is study the biomolecules in action in real time. For investigation of biomolecules we exploit the merit of different techniques, such as optical spectroscopy, scan-probe and transmission electron microscopy, atom force microscopy, Raman spectroscopy, resonance laser spectrometry and many others. The labeled fluorescent method dyes on the target molecules would give a variety information including high precision temporal and spatial movement and intra molecular motion of the molecules. With this simple approach, we attempt to give answers on the open questions in biology such as the conformational change of proteins, the biological role of nucleic acid structure, and microscopic domain formation of lipids bilayer.

There are two most interesting examples of biosystems (biomolecules) and their main properties: one is RecA filament dynamics and second – DNA bubble formation.

RecA is a protein involved in several DNA repair pathways s.a. homologous recombination and SOS response. RecA forms a right handed helical filament on a single-strand DNA, so that the DNA can be stretched out allowing the homologous sequence searching.
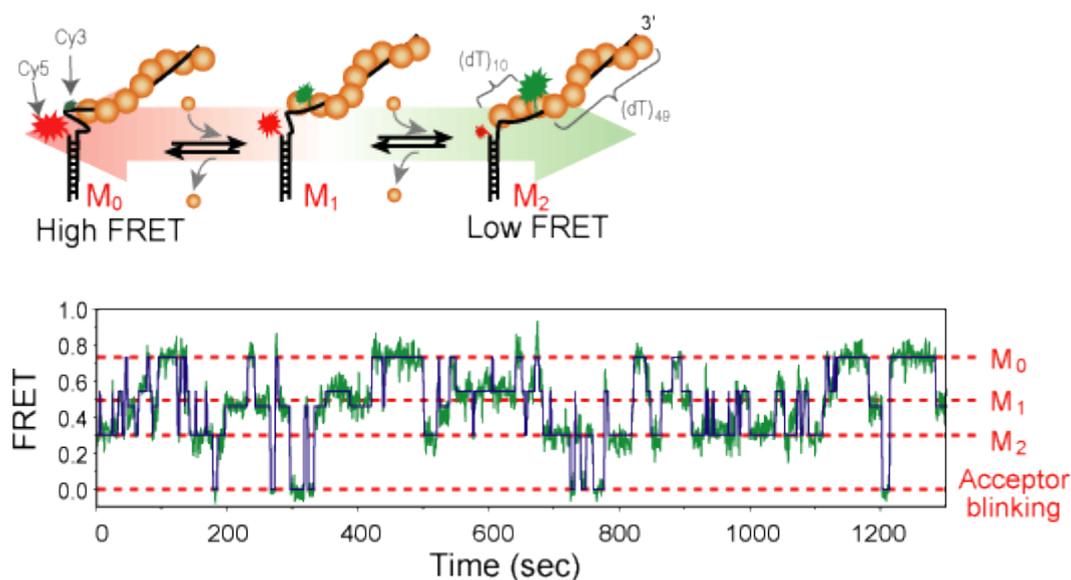
Fig.9.1. A model for binding and dissociation of RecA monomer from a RecA filament. Distance change of the two dye molecules by monomer binding.

DNA has a double-helical structure with two complimentary strand. In the aqueous solution, however, partial unwinding of double-strand DNA happens spontaneously by thermal (or quantum) fluctuation, which is called DNA bubble and fork formation. As FRET is sensitive to the distance change in nanometer scale, we study the DNA bubble and fork formation in real time.

Bubble   Fork
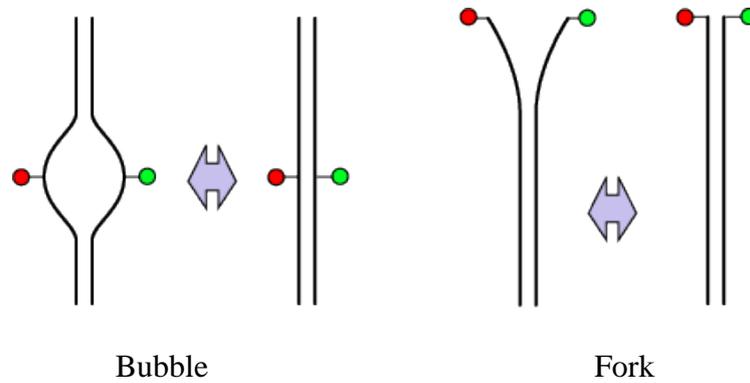
Fig. 9.2. DNA bubble and fork formation monitored by FRET.

The macro-biomolecule level includes the complexes of proteins nucleic acids and their small ligands. The typical sizes are on the scale of 10nm to fractions of micrometers. Even systems as large as a single protein are included in this range, as is the case of complexes between proteins and small ligands, when one wants to consider the process of association-dissociation or the very slow conformational modification related to the binding and subsequent events, since these can occur on the micro-millisecond time scale. The biological properties of a protein molecule depend on its physical interaction with other molecules. Thus, antibodies attach to viruses or bacteria to mark them for destruction, the enzyme hexokinase binds glucose and ATP so as to catalyze a reaction between them, actin molecules bind to each other to assemble into actin filaments, and so on. Indeed, all proteins stick, or bind, to other molecules. In some cases, this binding is very tight; in others, it is weak and short-lived. But the binding always shows great specificity, in the sense that each protein molecule can usually bind just one or a few molecules out of the many thousands of different types it encounters. The substance that is bound by the protein—no matter whether it is an ion, a small molecule, or a macromolecule— is referred to as a ligand for that protein (from the Latin word ligare, meaning "to bind").
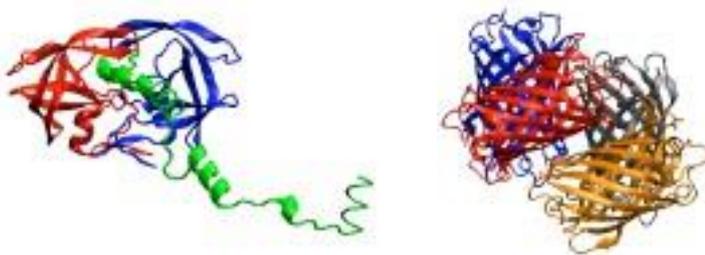
Fig.9.3. (a) HIV-1 protease with docked substrate (b) The tetramer of DsRed (c) The S70 Bacterial ribosome (~50 between proteins and RNA chains) (d) The cucumber mosaic virus capsid (~200 proteins).

(a)   (b) (c) (d) at Fig 9.3:(a) HIV-1 protease with docked substrate (b) The tetramer of DsRed (c) The S70 Bacterial ribosome (~50 between proteins and RNA chains) (d) The cucumber mosaic virus capsid

(~200 proteins). The experimental technique is mainly based on X-ray diffraction where the data post processing needs much work, because a much large molecular model must be fitted into experimental data (larger complexes are more difficult to crystallize, and due to the large amount of atoms within the crystal unit cell, the resolution turns out to be smaller).Thus only in particular conditions (presence of some peculiar symmetry) it is possible to apply this technique to large macromolecular systems. One example are certain virus capsids. A technique that is typically used at this level is the electron cryo-microscopy thatcan be complemented by X-ray. Both the experimental and modeling techniques used at this level are much less standard than those used both at the lower and at the higher levels. This is because this is a mesoscopic domain, bridging between an "atomic" description with a "continuum like" description of the system. Thus, especially at this level, in many cases multi-scale approaches are necessary to obtain a realistic description. Bridging macro to micro and finding reliable models for this level is one of the challenging topics in the field of bio-world description. Typical bead models for nucleic acids (a) one bead (b) two beads (c) multiple beads models for nucleic acids are even less standard than those for proteins.
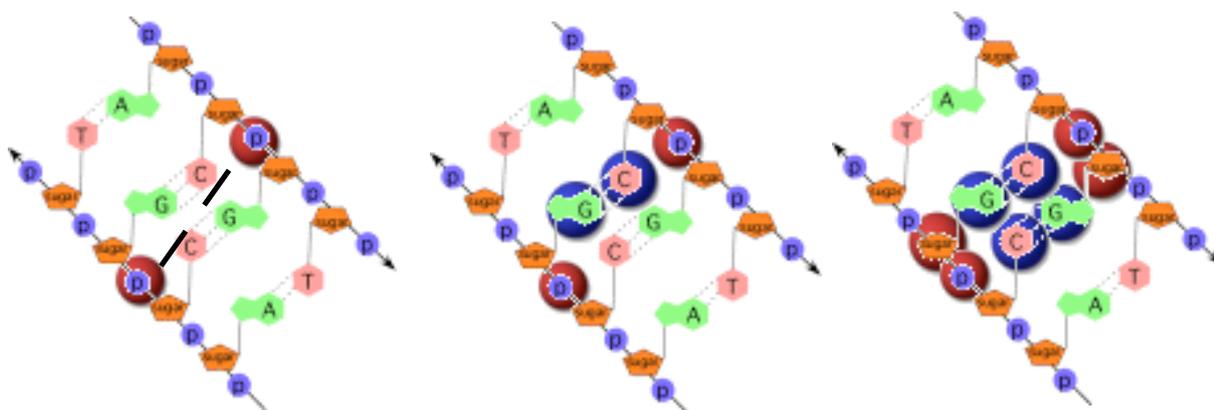
Fig. 9.4. Typical bead models for nucleic acids (a) one bead (b) two beads (c) multiple beads.

One bead model has the bead usually placed on the Phosphate atom, two beads have an additional bead on the base and can more properly represent the directionality of the base-base interaction. But also, multiple beads models with many beads placed between the two strands are available.

Fig 9.5 represents classification of CG models for proteins. The CG models for proteins can be classified based on the level of coarse graining (beads per amino acid #) on the complexity of the parameterization (parameters and/or FF terms #) and on the kind of parameterization. Fig. 9.5 can give an idea of the landscape of the FF and of their potentialities. In general, coarser models are preferable because give a

larger advantage in terms of computational cost, but their parameterization is more difficult because many complex interactions effect must be included in a small number of interaction terms and parameters.
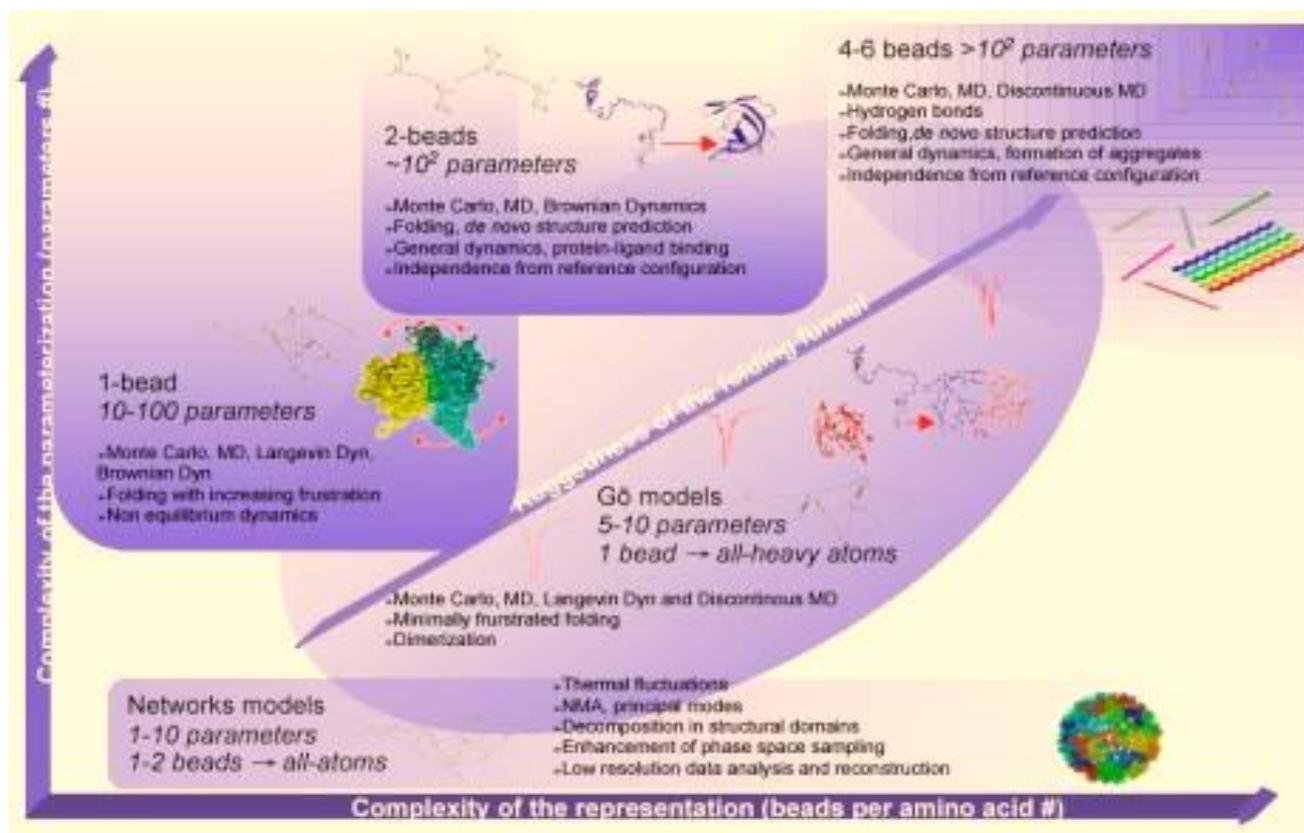
Fig. 9.5 Classification of CG models for proteins

The parameterization is usually based on the following methodologies:

- Force matching methods: the parameters of the CG Force Field are fitted on the forces calculated on trajectories obtained with all-atom simulations.
- "Statistical" methods: the parameters of the CG FF are fitted on the "Boltzmann inverse" of the probability distributions obtained from a set of experimental structural data;
- Knowledge-based methods: a priori knowledge of the physics-chemical properties of the system are included (local structure, secondary structure propensity, hydrophobicity, effective electrostatic effective charges)

Usually, a combination of these three approaches is used to parameterize a single CG-FF, and, given a level of coarse graining, different FF differ for the relative weight of given to them. In any case, especially 1. and 2., the thermodynamic consistency with the all-atom representation would be a requirement to fulfill, that can be expressed by the equation:

$$Z = exp(-\beta F) = \int dq \, exp(-\beta V(q)) = \int dQ \, exp(-\beta U(Q)) \qquad (9.1)$$

where Z is the partition function, F is the free energy, q and Q are the all-atom and CG internal coordinates, V(q) and U(Q) are the corresponding potentials and $\beta=1/KT$. In 1. the mechanical consistency is considered, by matching the forces on the CG sites with the sum of the forces over all the atomic sites included in each CG site, averaged over a set of trajectories obtained from all atom simulations.

The simplest one-bead models are the network. The procedure to build such models is very simple and is illustrated in Fig.9.6. Starting from a known reference structure (usually a crystallographic one), only the atoms where the bead will be located are selected (usually the Cαs) and a network of bonds that connect all the beads located within certain cutoff between each other is build.

a   b   c
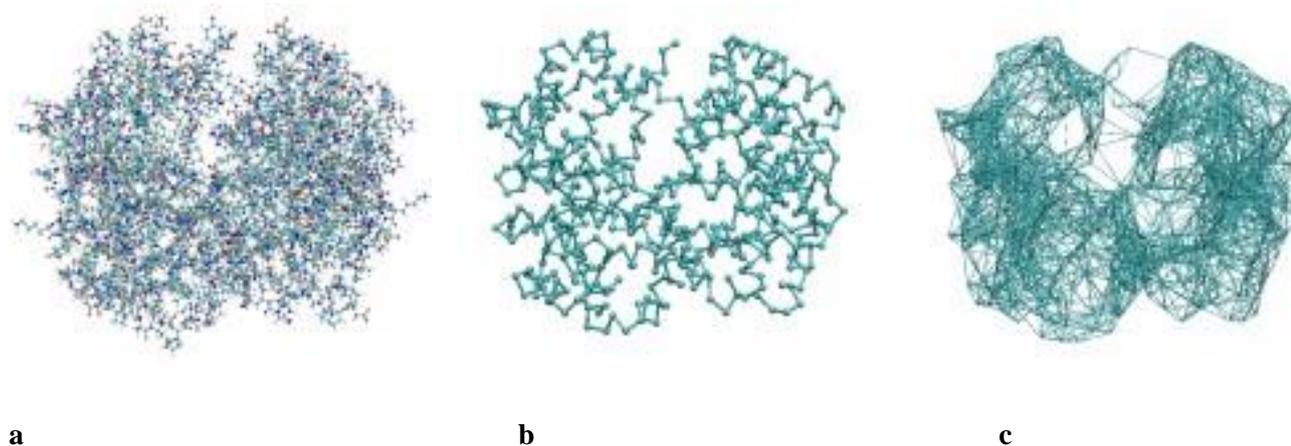
Fig. 9.6. Steps to build a network model for hemoglobin: starting from the allatom structure (a) only the Cα are selected (b) and a network of bonds for the beads within a certain cutoff are build (c).

The interactions are treated with harmonic or anharmonic (usually Morse) potentials

$$U(rij) = 1/2 \, k \, (rij - rij, 0)2 \qquad (9.2)$$

$$U(rij) = E(rij, 0)[exp(\alpha(rij - rij, 0)) - 1]2 \qquad (9.3)$$

The predictivity and transferability of the network models is very low, because a high level of a priori knowledge of the system is necessary. However, they are still very useful for special purposes. One is the principal components analysis. Principal modes are defined as the eigenvectors of the covariance matrix $Cij = < (ri-< ri >)(rj-< rj >) >$

They coincide with the normal modes in a purely harmonic system, and their eigenvalues are the squared amplitudes of modes. Thus, when ordered by eigenvalues, the first principal modes correspond to those that exhaust the dynamics of the system, i.e. they define the "Essential Dynamics". Additionally, assuming that the equipartition theorem applies, one has $<\Delta r2>= KT/m\omega2$, thus are also the slowest. It was shown that the first ~10 modes usually exhaust the relevant dynamics of a biological system and correspond to motions

related to some biological function. An example is the PCA analysis of the S70 bacterial ribosome, that reveals a few modes related to the function of translocation of the mRNA in the process of translation.
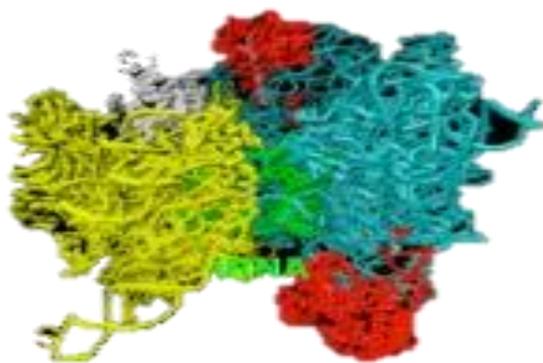
Fig.9.7. Example of PCA. Representation of the first two principal modes of the S70 bacterial ribosome.

The power spectrum of these vibrations can be evaluated, and turns out to be on the scale of 1cm-1, corresponding approximately to a period of 0.1-0.2 nsec. It should be kept in mind that in coarse grained models the time constants obtained from the simulation cannot be taken as they are. As already mentioned, there is an intrinsic unphysical acceleration of the dynamics due to the fact that the potential energy surfaces are artificially smoothened, implying a faster sampling of the phase space. Elastic and inelastic network models can be conveniently used to fit models into the low resolution electron maps obtained from cryo-EM. In particular, it is possible to obtain models for different conformations that might be present in the EM data, and to flexibly fit high resolution data into low resolution cryo-EM data, to obtain atomic level resolution even for non crystallizable large complexes.

Domain V   Domain I

Domain IV

Domain III   Domain II
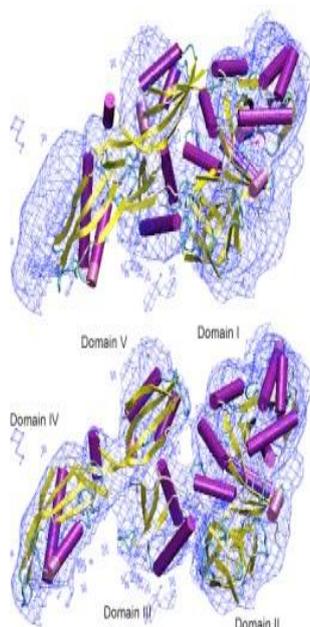
a

Fig 9.8. (a) The cry-EM map of two conformations of a part of the ribosome obtained with the aid of a network model (b) flexibly fitting the high resolution structures into a low resolution map with the aid of network models.

For one-bead models especially designed for folding the potential energy is written as the sum of bonds, bond angles, dihedrals and non bonded interactions, as in the all atoms approaches

$$U = (U_b) + U_\theta + U\alpha + Unb \tag{9.4}$$

where U_b is harmonic or substituted with a constraint, $U_\theta = \frac{1}{2}k_\theta(\theta - \theta_0)^2$, $U_a = \frac{1}{2}k_a(a - a_0)^2$ and $Un = \varepsilon\left[6(r_{0,ij/rij})12 - 12(r_{0,ij/rij})^6\right]$ is a Lennard Jones term. All the terms of the potential are biased towards a known structure in the other words the equilibrium values for the angle, dihedral and all the $r_{0,ij}$ are taken from it. Of course, as in the case of the network models, there is a large amount of a priori knowledge included in the model, and the native structure is a minimum of the free energy surface. In spite of their strong bias and extreme simplicity, this model already can give a lot of interesting information: the functional motions of a system, the pathways to folding. Since the main input is the (topology of the) native structure, this indicates that the fundamental properties of a system depend mainly on the topology and structure of the folded protein. It is true that the primary sequence determines the folded structure (in the natural proteins), however, sequences with remarkably low identity (down to 20%) can fold in the same tertiary structure. In order to build a really predictive model, the bias towards a reference structure should be eliminated. A possibility is to parameterize each term of the potential energy based on the Boltzmann inversion. A first approximation for the energy terms can be obtained taking the logarithm of the probability distribution function evaluated on a statistical set of structures. The generality/transferability of the model will also depend on how extended is the phase space sampled by the statistical set.protein-protein or protein-ligand docking methods can be seen as another example of extreme coarse graining. In these methods the whole protein (or macromolecule) is represented usually as a rigid body defined by its solvent exposed surfaces, that is obtained rolling a probe sphere (representing the solvent molecule) over the VanDerWaals spheres of the surface atoms of the protein. It is the first requisite to satisfy for docking is the surface complementarity. Then the interaction energy is evaluated through "scoring functions": several properties are mapped onto the surfaces (electrostatics, hydrophobicity and others) and based on these a score is given indicating how well these properties match. The relative conformational space of the two molecules to be docked is explored with several methods (systematic, monte - carlo), but, obviously, the timestep has only the meaning of progression in the exploration, no dynamical information can be extracted. Recently, flexible docking methods were also considered, that allow partial fluctuation of the structure to better dock the two molecules.
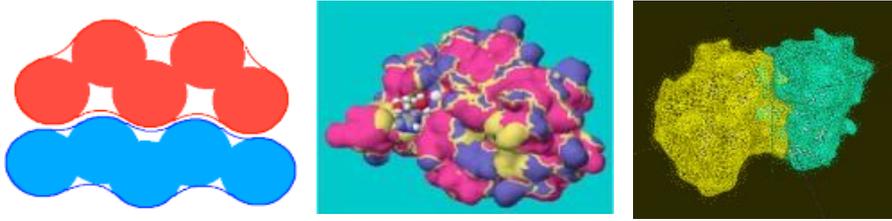
Fig 9.9 (a) A schematic representation of the shape complementarity between molecules (b) a protein surface colored according to hydrophobicity with a docked ligand (c) a protein-protein docking

Some examples of Biosystems organization
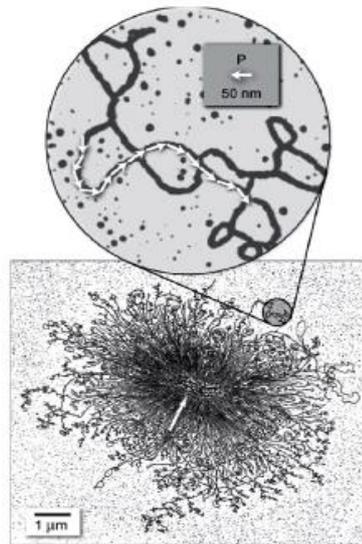
Fig.9.10. Bacterial genome and bacterial cell.

DNA organization:

- packed into chromatin fibers;

- unit of chromatine (nucleosome);

- characterized with various packing densities;

- linear density of chromatin
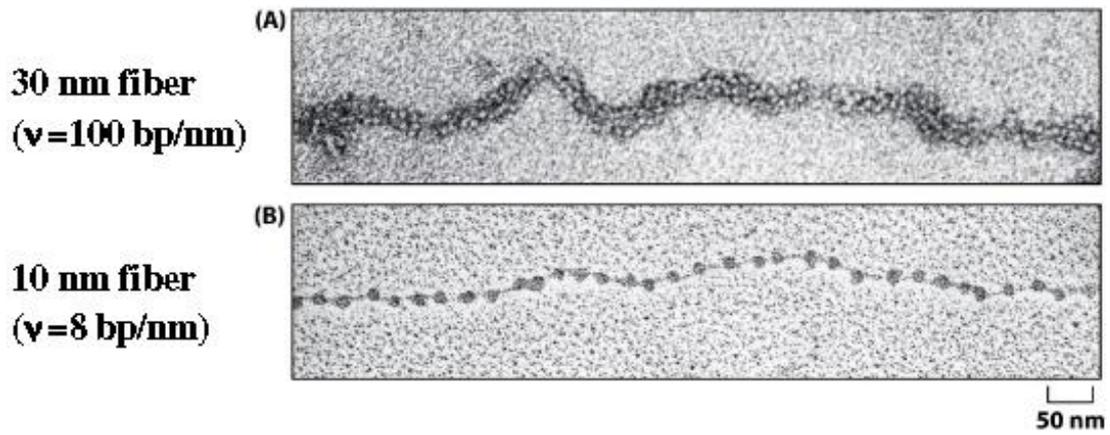
Fig 9.11. A schematic representation of the shape of DNA: with 30 fiber (A), and 10nm fiber (B).

Chromosome 18 and 19 in a nucleus of a human cell

(observed with fluorescence spectroscopy)

- existence of chromosome territories
- chromatin density: a dense polymer system
- free polymers in a dense system inter-penetrate
- why is chromatin packed and localized?
- existence of *tethering*
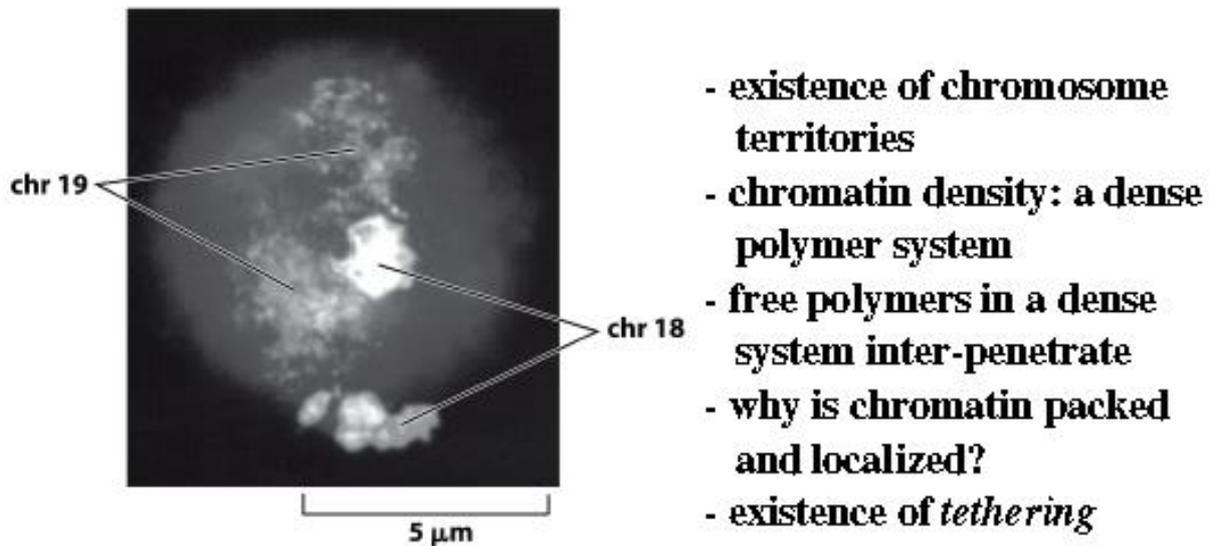
Fig 9.12. Scan probe electron microscopy picture of chromosome.
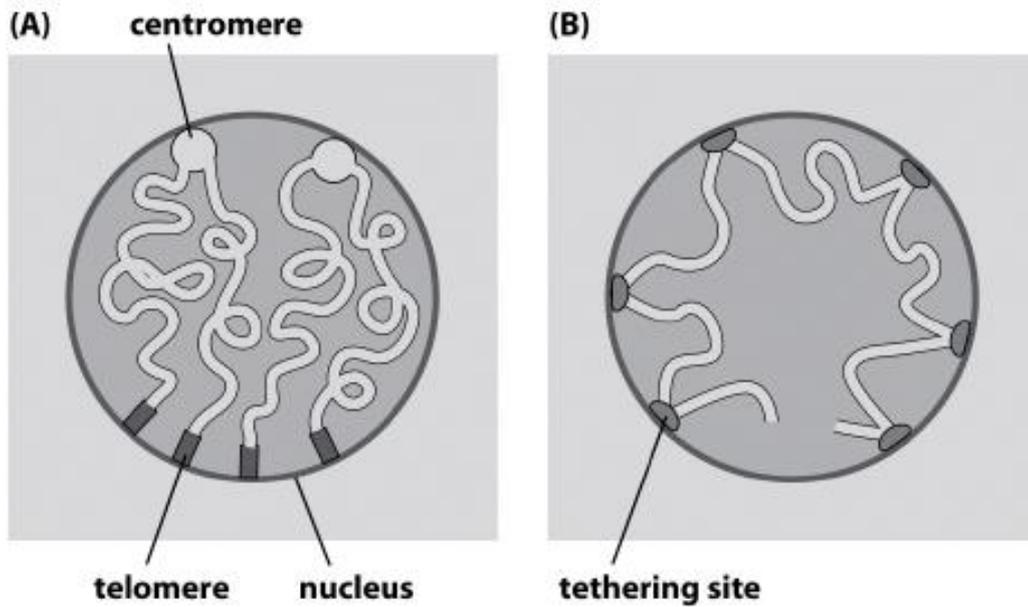
Fig 9.13. Chromosomes are tethered at different locations inside the nucleus.
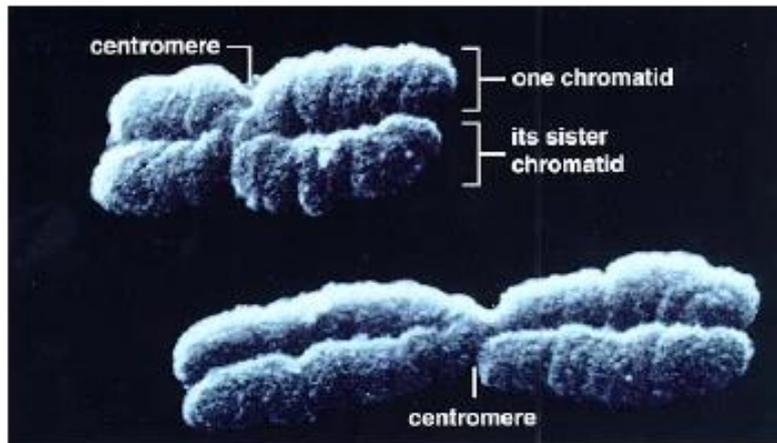
Fig 9.14. Chromosomes: centromeres and chromatids.

None or only one tethering site, a random walk model of chromatin predicts *a Gaussian distribution*:

$$P(r) = C \exp\left(-\frac{r^2}{\sigma^2}\right)$$

$$C' = \left[\frac{3}{(2\pi Na^2)}\right]^{\frac{3}{2}} \& \sigma^2 = \frac{2Na^2}{3} \& a = 2\xi_P$$

N ... total number of Kuhn's segments (of length a); $L = Na$ ...length of the polymer; r ... a 3D vector.

For two tethering sites (and two fluorescent markers) leads to *a displaced Gaussian distribution*:

$$P(r) = C' \exp\left(-\frac{(r-R)^2}{\sigma'^2}\right)$$

$$C' = \left[\frac{3}{(2\pi N'a^2)}\right]^{\frac{3}{2}} \& \sigma'^2 = \frac{2N'a^2}{3} \& \ a = 2\xi_P$$

N ... total number of Kuhn's segments between 2nd tether and 2nd fluorescent marker, R and r ... 3D vector.

Flory theorem

For dense polymer systems distribution of distances between monomers are described by random-walk statistics. $Na = N_{gd}/v$; N_{gd} ... genomic distance; v... linear packing density
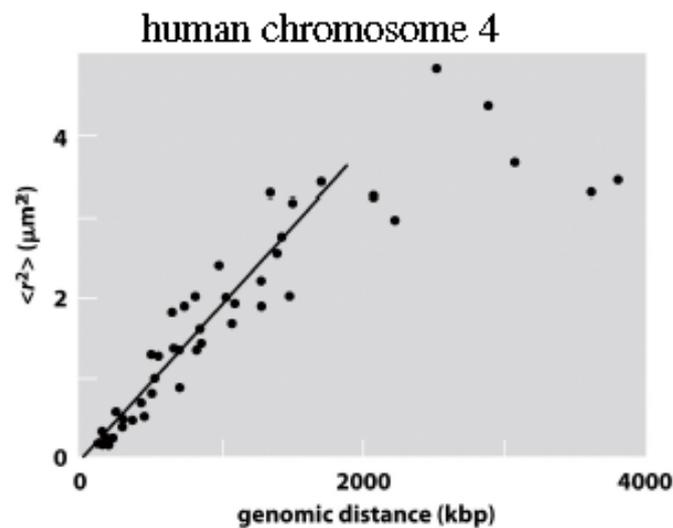
Fig 9.15. distribution of radius-vectors and their dependence from demonic distance.

Model of polymer confinement and tethering
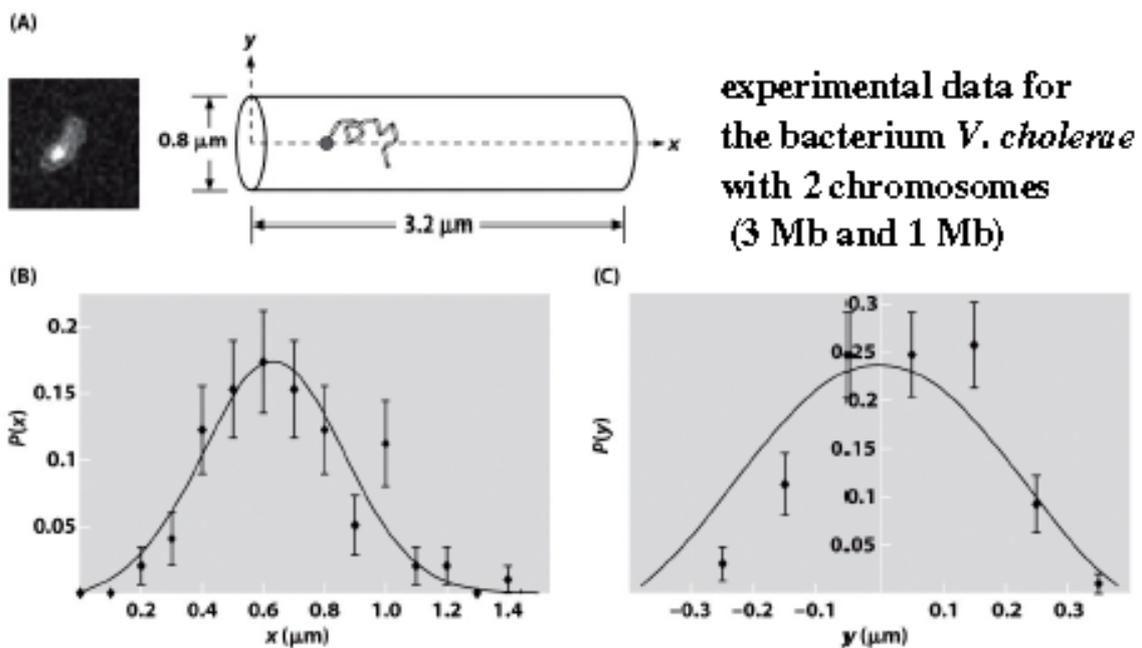
Fig. 9.16. Gaussian distribution for distance along x, but confined random walk distribution along y.

DNA Structures
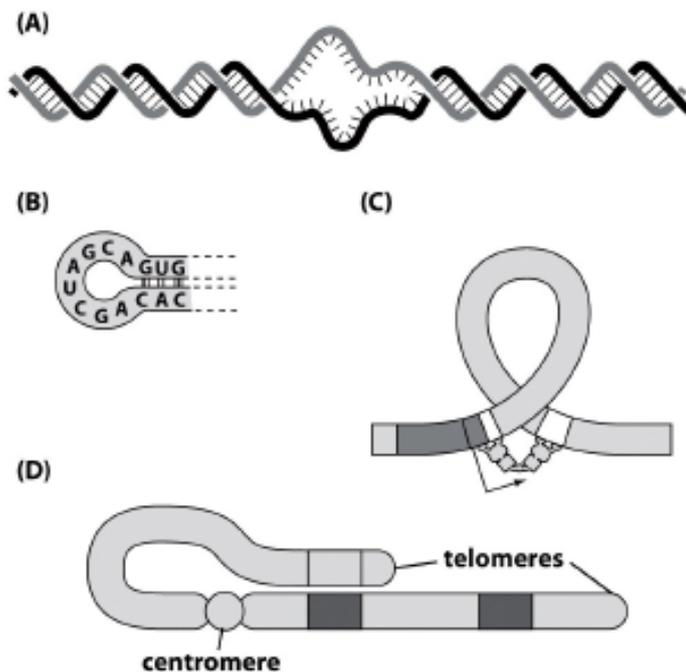
Fig. 9.17. A – DNA bubbles, B – RNA hairpins, DNA loop, D- long-range looping of chromosomal DNA
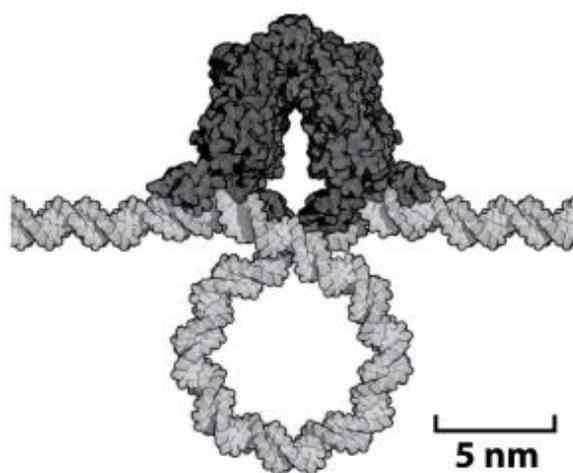
Fig. 9.18. Schematic model of DNA loop formation

Entropic cost of loop formation

(1) in 1D random walk model, calculate the fraction of conformations which close on themselves:

$$p_0 = \#\ of\ looped\ configs/total\ \#\ of\ configs = \frac{\left\{\frac{N!}{\left[\left(\frac{N}{2}\right)!\left(\frac{N}{2}\right)!\right]}\right\}}{2^N}$$

$$\sim \left[\frac{2}{(\pi N)}\right]^{\frac{1}{2}}$$

Or use the probability distribution 1D or 3G

$$P(R;N) \sim (2\pi N a^2)^{-\frac{1}{2}} \quad or \quad P(R;N) \sim = \left[\frac{3}{(2\pi N a^2)}\right]^{\frac{3}{2}}$$

To calculate $p_0 = \int_{-\sigma}^{+\sigma} P(R;N)\,dR$ and put $\sigma = a$

(2) in 3D random walk model,

$$p_0 = \int_0^{+\sigma} 4\pi R^2\, P(R;N)\,dR = \left[\frac{6}{(\pi N^3)}\right]^{\frac{1}{2}}$$

Calculate that the probability of a bubble of length n base pairs:

$$p_1(n) = Z^{-1} \exp\left[-\frac{G_1(n)}{k_B T}\right]$$

$G_1(n)$ is the free energy for formation of a bubble of length n:

$$G_1(n) = E_{IN} + n E_{EL} - k_B T \ln[\Omega_0(n)(N - n + 1)]$$

E_{IN} ... energy of bubble initiation

E_{EL} ... energy of bubble elongation by one base pair

$(N - n + 1)$ ... number of choices for the bubble location

$\Omega_0(n)$ ... the number of ways to make a bubble of two strands each n base pairs long (1D)

$$\Omega_0(n) = 2^{2n} p_0(2n) \sim \frac{2^{2n}}{(\pi N)^{\frac{1}{2}}} \quad for \ N \gg 1$$

$$\frac{\Delta G_1(n)}{k_B T} = n(\varepsilon_{EL} - 2\ln 2) + \frac{1}{2}\ln n - \ln(N - n + 1)$$

$$where \quad \varepsilon_{EL} = \frac{E_{EL}}{k_B T}$$

Minimize $\frac{\Delta G_1(n)}{k_B T}$ with respect to n:

$$\varepsilon_{EL} - 2\ln 2 + \frac{1}{(2n)} + \frac{1}{(N - n + 1)} = 0$$

Two possible situations:

(A) $\varepsilon_{EL} - 2\ln 2 > 0$: no real solution for n, bubbles small

(B) $\varepsilon_{EL} - 2\ln 2 < 0$: two solutions solution for n (small n local maximum and $n \sim N$ local minimum)

(A) low temperatures; (B) high temperatures
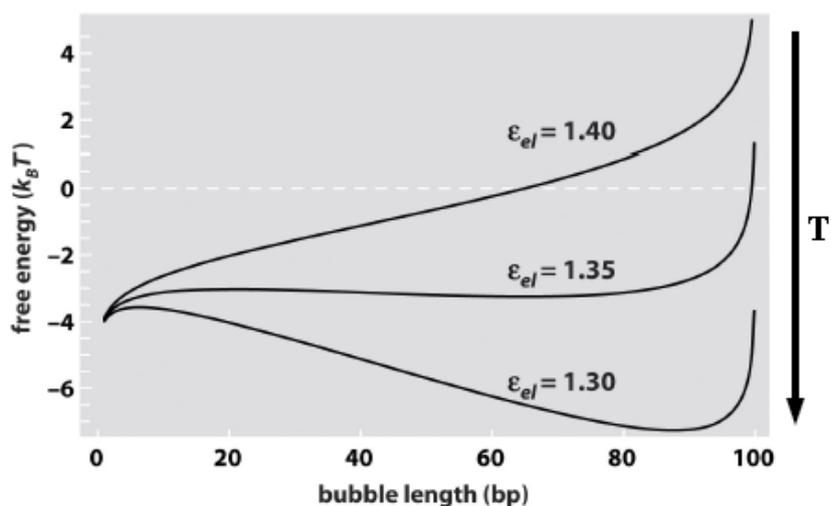
Fig. 9.18. Interplay between energy and bubble elongation.
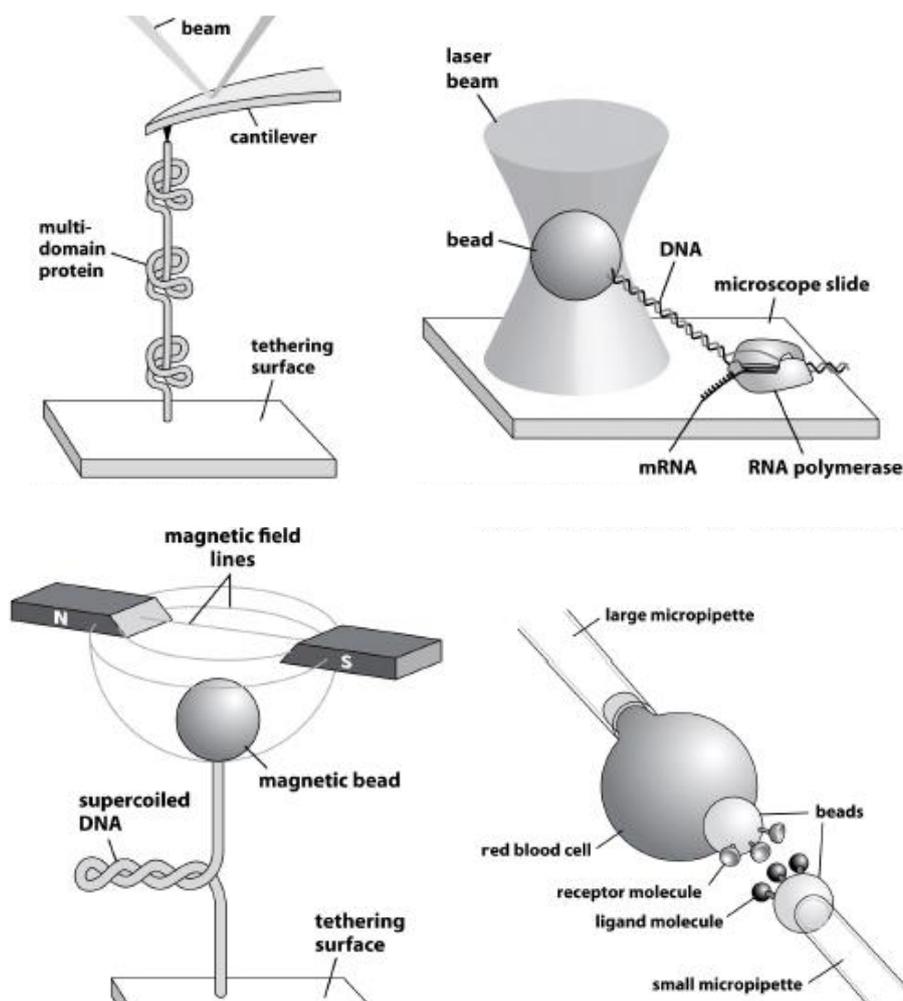
Fig. 9.19. A - Scheme of Atom Force Microscope work; B – Optical tweezers; C - magnetic tweezers; D - pipette based force instruments

The free energy can be expressed as:

$$G(n_R) = -2fn_R a - k_B T \left[n_R \ln(n_R) + (N - n_R) \ln(N - n_R) \right]$$

$$\frac{\partial G(n_R)}{\partial n_R} = 0 \implies \frac{n_R}{n_L} = \exp\left[\frac{2fa}{(k_B T)} \right]$$

The 1D random walk model predicts the expansion z:

$$z = \frac{\langle L \rangle}{L_{TOT}} = \frac{(n_R - n_L)}{(n_R + n_L)} = \tanh\left(\frac{fa}{k_B T} \right)$$

For $fa \ll k_B T$ we obtain a linear relationship

$$\langle L \rangle = L_{TOT} \frac{fa}{k_B T}$$

Hooke's law: $f = kx$, thus the stiffness' constant:

$k = \frac{k_B T}{(a L_{TOT})}$ . . .entropic origin

The free energy of stretching with a force f can be expressed as:

$$G(L) = -fL - k_B T \ln W(L; L_{TOT})$$

L . . . end-to-end distance of the macromolecule

L_{TOT} . . . total length of the macromolecule

$W(L; L_{TOT})$ . . . the number of microstates (realizations, permutations) corresponding to L

f . . . pulling force

In a 1D random walk model:

$$L = (n_R - n_L)\, a \ \ \& L_{TOT} = (n_R + n_L)\, a = Na$$

$$W(n_R; N) = \frac{N!}{[n_R! (N - n_R)!]}$$

Leventhal' Paradox

The Levinthal paradox dominated ideas about protein folding until very recently. The essential concept introduced by Levinthal is that the appropriate point of reference for protein folding is a random search problem. This means that all conformations of the polypeptide chain are equally probable, so that the native state can be found only by an unbiased random search. This has been referred to as the 'golf course' model

of the protein potential energy surface. For such a surface, the time to find the native state is given by the number of configurations of the polypeptide chain multiplied by the time required to find one configuration. Since proteins generally fold in times on the order of milliseconds to seconds (except for special factors that slow the folding, such as proline isomerization, there was indeed a paradox. An important point that emerged from these studies is that an essential element of the complexity is the presence of long-range interactions, which lead to the well-known cooperative character of the folding transition. Levinthal's solution to the protein folding problem was that there were well-defined pathways to the native state, so that protein folding was under 'kinetic' control; Most of these models are descriptive in character and do not provide a means for estimating the folding time, which is clearly an important element in a resolution of the Levinthal paradox. The diffusion-collision model is an exception that has been used to relate the folding time to certain system parameters. In 1990, Baldwin wrote: "After many years of skepticism, it needs no longer be doubted that proteins go through a series of identifiable intermediate changes as they fold up to assume their native conformations…the ghost of the 'jigsaw-puzzle' model, which postulates a random collection of folding intermediates, has now been laid to rest. The impact of lattice-based folding simulations has continued to grow with a veritable deluge of papers published recently. They have demonstrated the versatility of lattice simulations and their ability to mimic a wide range of folding behavior. An important element in the change in the perception of the protein folding problem has been the felicitous introduction of the 'folding funnel' as an idealized construct for the free energy surface of the polypeptide chain. Two different models for the interaction give very similar surfaces helps to establish the generality of the original result. The 'folding funnel' represents the energy and entropy of the polypeptide chain as a function of the progress variable. Generally, the protein folding (Levinthal's) paradox states that it would not be possible in a physically meaningful time to a protein to reach the native (functional) conformation by a random search of the enormously large number of possible structures. This paradox has been solved: it was shown that small biases toward the native conformation result in realistic folding times of realistic-length sequences. This solution of the paradox is, however, not amenable to most chemistry or biology students due to the demanding mathematics. Here, a simplification of the study of the paradox and its solution is provided so that it is accessible to chemists and biologists at an undergraduate or graduate level. Despite its simplicity the model captures some fundamental aspects of the protein folding mechanism and allows students to grasp the actual significance of the popular folding funnel representation of the protein energy landscape. The analysis of the folding model provides a rich basis for a discussion of the relationships between kinetics and thermodynamics on a fundamental level and on student perception of the time-scales of molecular phenomena.
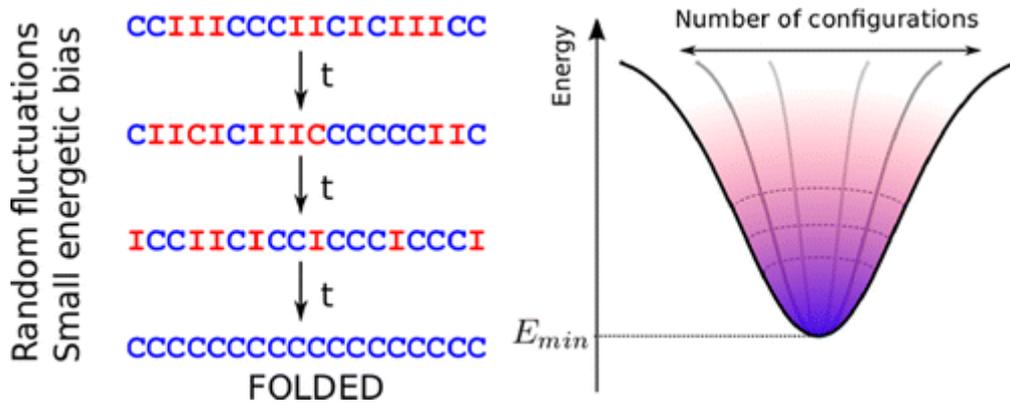
Fig. 9.20. Levinthal's Protein Folding Paradox and Its Solutions.

Lattice Models: The Simplest Protein Model

$\{H, P\}$; $H/P = hydrophobic/polar$
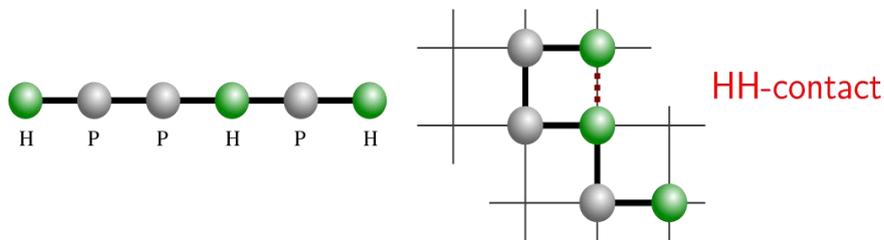
Fig. 9.21. Scheme of HP-Model Definition

The HP-model is a protein model, where

Sequence Structure $s\omega \in \{: [1H..n,]P \rightarrow\}n\ L\ (e.g.\ L = Z2, L = Z3)$,

1. $for\ all\ 1 \leq (ii + 1)) = < n : dmin(L)\quad [dmin(Z2) = 1]\ d(\omega(i), \omega$

2. $for\ all\ 1 \leq i < j \leq n : \omega(i)\ 6 = \omega(j)$

- $Energy\ function\ E(s, \omega) = P1 \leq i < j \leq n\ Esi, sj\ \Delta(\omega(i), \omega(j))$,

where E $=$

	H	P
H	−1 0	0
P	0 0	0

$and\ \Delta(p, q) = 1 if\ d(p, q) = dmin(L)$

Chapter 10

The Building Blocks and Architecture of Cells

The basic properties of the simple biological molecules, lipids and microorganisms are very depending of rich variety of their structures that life presents, and some respect for the extreme complexity of the biological molecules that operates in a wide range of cellular processes.

Proteins

Polymers consist of a large number of sub-units (monomers) connected together with covalent bonds. A protein is a special type of polymer. In a protein there are up to twenty different amino acids (Figure 10.1) that can function as monomers, and all the monomers are connected together with identical peptide linkages (C–N bonds, Figure 10.2). The twenty amino acids can be placed in different family dependent on the chemistry of their different side groups. Five of the amino acids form a group with lipophilic (fat-liking) side-chains: glycine, alanine, valine, leucine, and isoleucine. Pro line is a unique circular amino acid that is given its own separate classification. There are three amino acids with aromatic side-chains: phenylalanine, tryptophan, and tyrosine. Sulfur is in the side-chains of two amino acids: cysteine and methionine. Two amino acids have hydroxyl (neutral) groups that make them water loving: serine and threonine. Three amino acids have very polar positive side-chains: lysine, arginine and histidine. Two amino acids form a family with acidic. Linkages between amino acids all have the same chemistry and basic geometry (Figure 10.2). The peptide linkage that connects all amino acids together consists of a carbon atom attached to a nitrogen atom through a single covalent bond. Although the chemistry of peptide linkages is fairly simple, to relate the primary sequence of amino acids to the resultant three dimensional structure in a protein is a daunting task and predominantly remains an unsolved problem. To describe protein structure in more detail it is useful to consider the motifs of secondary structure that occur in their morphology. The motifs include alpha helices, beta sheets and beta barrels (Figure 10.3). The full three dimensional tertiary structure of a protein typically takes the form of a compact globular morphology (the globular proteins) or a long extended conformation (fibrous proteins, Figures 1.4 and 1.5). Globular morphologies usually consist of a number of secondary motifs combined with more disordered regions of peptide. Charge interactions are very important in determining of the conformation of biological polymers. The degree of charge on a poly acid or poly base (e.g. proteins, nucleic acids etc.) is determined by the pH of a solution, i.e. the concentration of hydrogen ions. Water has the ability to dissociate into oppositely charged ions; this process depends on temperature

$$H_2O \rightleftarrows H^+ + OH \qquad (10.1)$$

Aliphatic amino acids

H_3N^+—C—COO^- (H above, H below) — **Glycine**

H_3N^+—C—COO^- (CH_3 above, H below) — **Alanine**

H_3N^+—C—COO^- (H_3C, CH_3 above; H below) — **Valine**

H_3N^+—C—COO^- (H_3C, CH_3–CH–CH_2 above; H below) — **Leucine**

H_3N^+—C—COO^- (CH_3–CH_2; H_3C–CH_2; CH_3) — **Isoleucine**
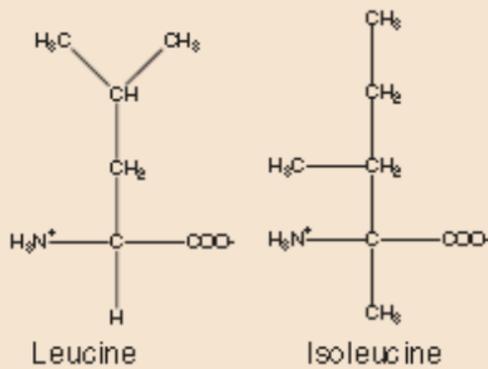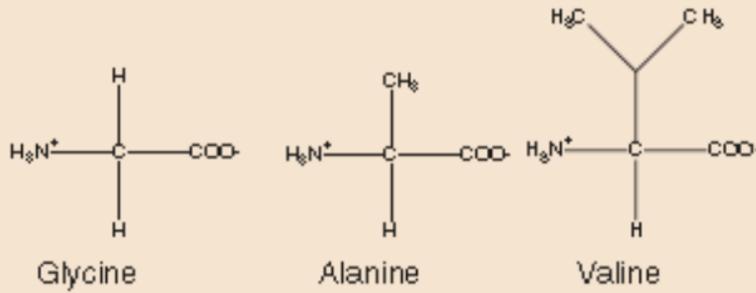
Amino acids with hydroxyl or sulfur containing groups

H_3N^+—C—COO^- (OH–CH_2 above; CH_3 below) — **Serine**

H_3N^+—C—COO^- (SH–CH_2 above; H below) — **Cysteine**

H_3N^+—C—COO^- (CH_3–HCOH above; H below) — **Threonine**

H_3N^+—C—COO^- (CH_3–S–CH_2–CH_2 above; H below) — **Methionine**

Aromatic amino acids

H_3N^+—C—COO^- (benzene ring–CH_2 above; H below) — **Phenylalanine**

H_3N^+—C—COO^- (OH–benzene ring–CH_2 above; H below) — **Tyrosine**

H_3N^+—C—COO^- (indole ring–CH_2 above; H below) — **Tryptophan**
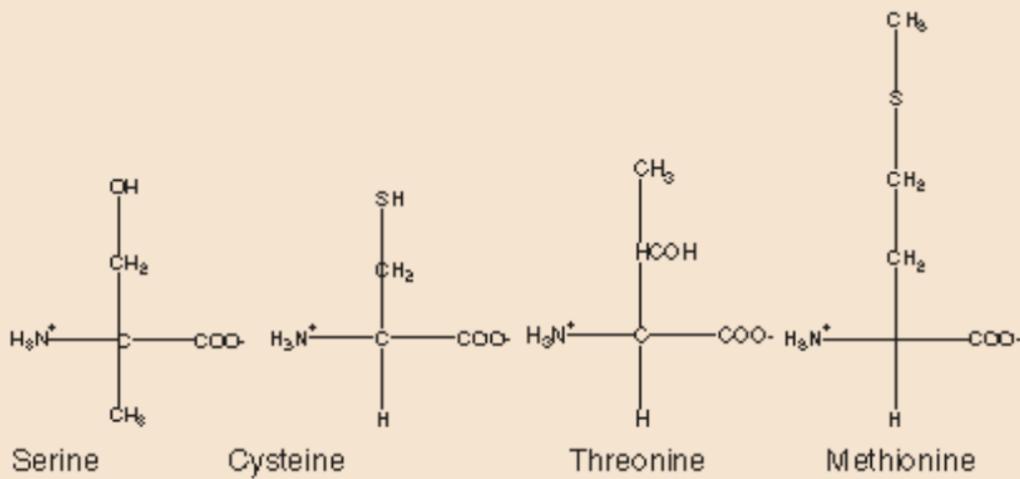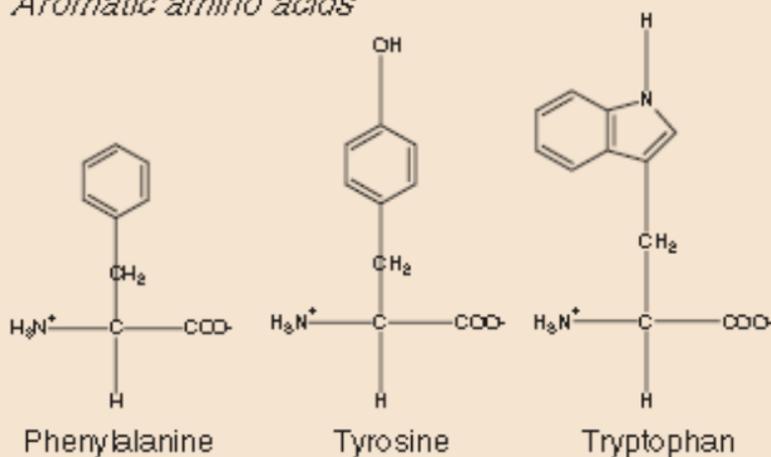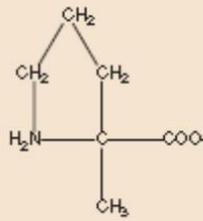
Cyclic amino acid

Proline

Basic amino acids
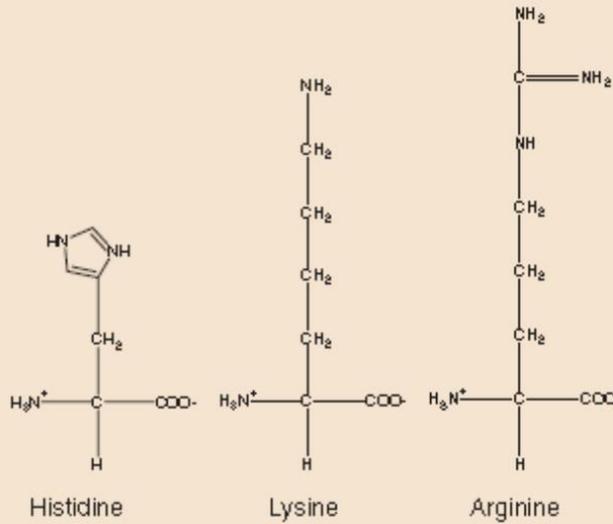
Histidine   Lysine   Arginine

Acidic amino acids and amides
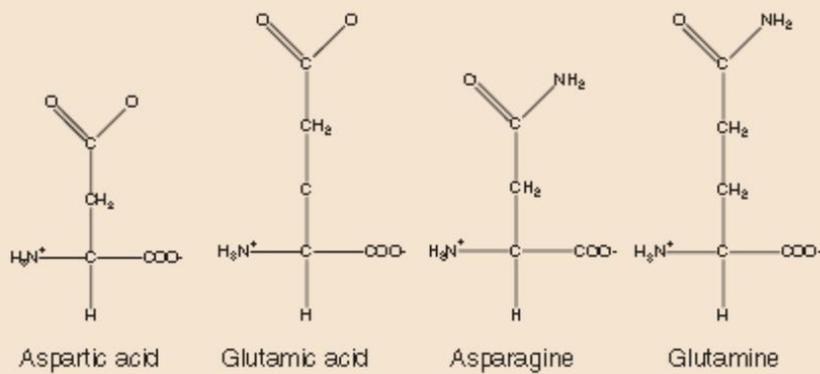
Aspartic acid   Glutamic acid   Asparagine   Glutamine

Figure 10.1. The chemical structure of the twenty amino acids found in nature

122
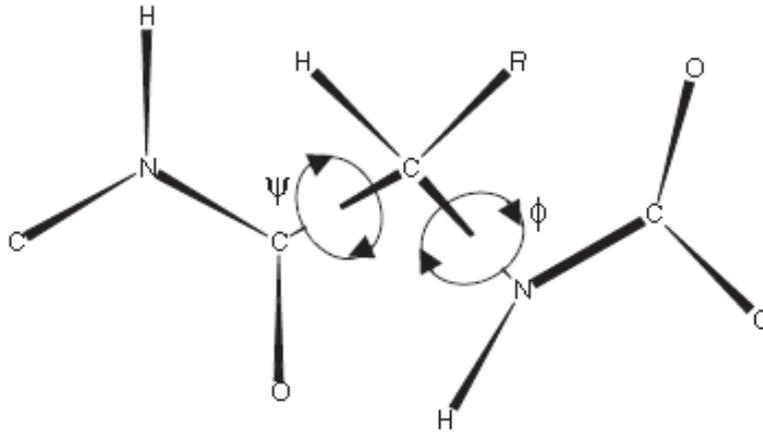
Figure 10.2. All amino acids have the same primitive structure and are connected with the same peptide linkage through $C - C - N$ bonds (O, N, C, H indicate oxygen, nitrogen, carbon and hydrogen atoms respectively. R is a pendant side-group which provides the amino acid with its identity, i.e. pro line, glycine etc.)

The product of the hydrogen and hydroxyl ion concentrations formed from the dissociation of water is a constant at equilibrium and at a fixed temperature (37°C)

$$C_{H+}C_{OH} = 1 \times 10^{14}M^2 = K_\omega \qquad (10.2)$$

where C_{H+} and C_{OH} are the concentrations of hydrogen and hydroxyl ions respectively. Addition of acids and bases to a solution perturbs the equilibrium dissociation process of water, and the acid/base equilibrium phenomena involved are a corner stone of the physical chemistry of solutions.
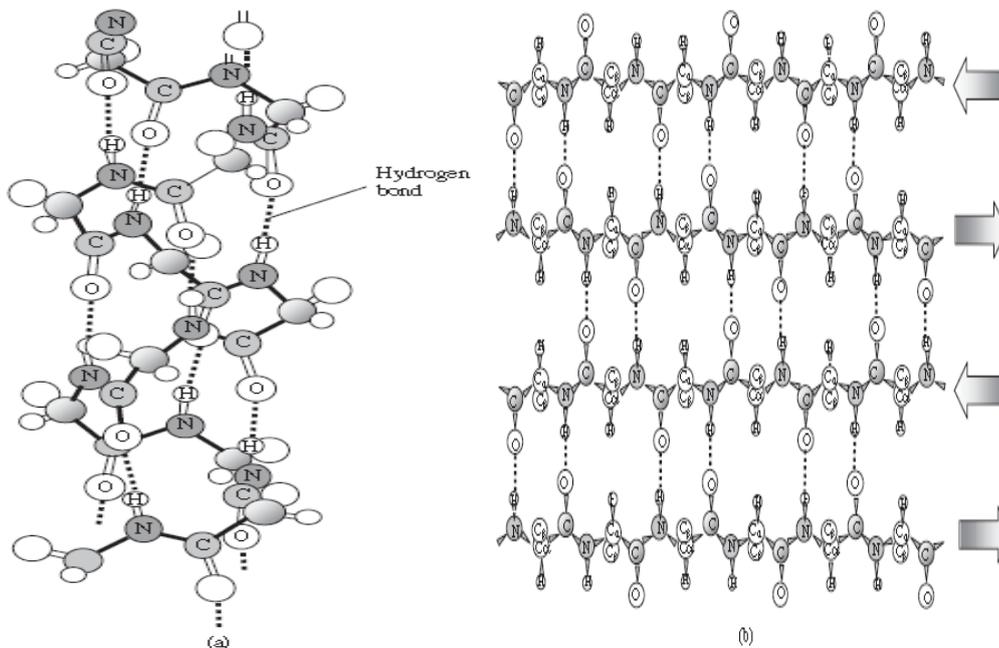
Figure 10.3. Simplified secondary structures of (a) an a-helix and (b) a b-sheet that commonly occur in proteins (Hydrogen bonds are indicated by dotted lines.)

Due to the vast range of possible hydrogen ion $(H+)$ concentrations typically encountered in aqueous solutions, it is normal to use a logarithmic scale (pH) to quantify them. The pH is defined as thenegative logarithm (base 10!) of the hydrogen ion concentrations.

$$PH = \log C_{H+} \tag{10.3}$$
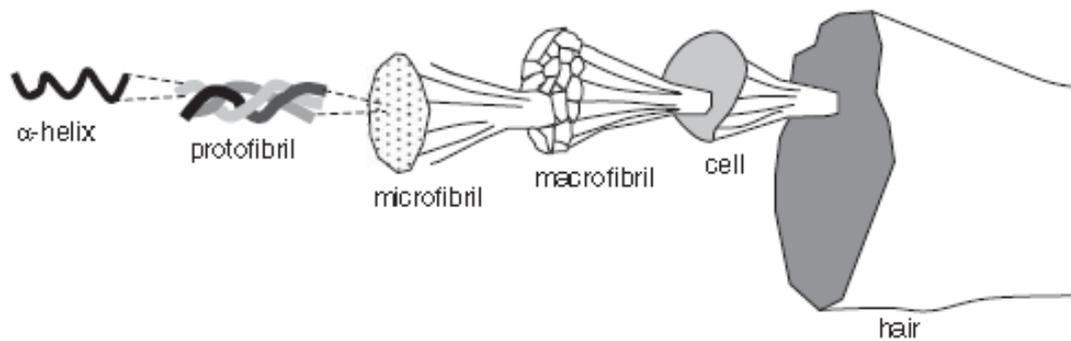
Figure 10.4. The complex hierarchical structures found in the keratins of hair (a-helices are combined in to proto fibrils, then into micro fibrils, macro fibrils, cells and finally in to a single hair fibre [Reprinted with permission from J.Vincent, Structural Biomaterial, Copyright (1990) Princeton University Press])
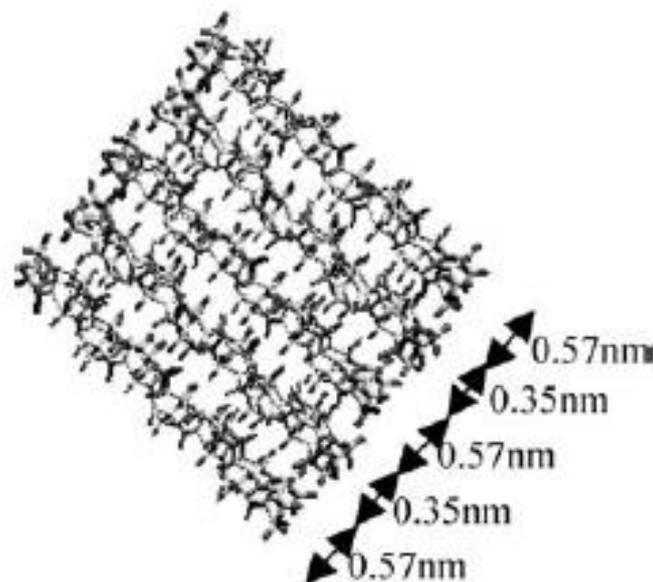
Figure 10.5. The packing of anti-parallel beta sheets found in silk proteins (Distances between the adjacent sheets are shown.)

Typical values of pH range from 6.5 to 8 in physiological cellular conditions. Strong acids have a pH in the range 1–2 and strong bases have a pH in the range 12–13.

When an acid (HA) dissociates in solution it is possible to define an equilibrium constant (Ka) for the dissociation of its hydrogen ions ($H+$)

$$HA \rightleftarrows H^+ + A \qquad\qquad K_a = \frac{C_{H+}C_A}{C_{HA}} \qquad\qquad (10.4)$$
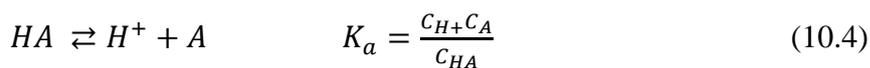
where C_{H+}, C_A and C_{HA} are the concentrations of the hydrogen ions, acid ions, and acid molecules respectively. Since the hydrogen ion concentration follows a logarithmic scale, it is natural to also define the dissociation constant on a logarithmic scale pK_a

$$pK_a = \log K_a \qquad\qquad (10.5)$$

The logarithm of both sides of equation (10.4) can be taken. That gives a relationship between the pH and the pK_a value:

$$PH = pK_a + \log\left\{\frac{C_{conjugate-base}}{C_{acid}}\right\} \qquad\qquad (10.6)$$

where C conjugate-base and C acid are the concentrations of the conjugate base (e.g. A⁻) and acid (e.g. HA) respectively. This equation enables the degree of dissociation of an acid (or base) to be calculated, and it is named after its inventors Henderson and Hassel Balch. Thus a knowledge of the pH of a solution and the pK_a value of an acidic or basic group allows the charge fraction on the molecular group to be calculated to a first approximation.

The propensity of the amino acids to dissociate in water is illustrated in Table 1.1. In contradiction to what their name might imply, only amino acids with acidic or basic side groups are charged when incorporated into proteins. These charged amino acids are arginine, aspartic acid, cysteine, glutamic acid, histidine, lysine and tyrosine.

Another important interaction between amino acids, in addition to charge interactions, is their ability to form hydrogen bonds with surrounding water molecules; the degree to which this occurs varies. This amino acid hydrophobicity (the amount they dislike water) is an important driving force for the conformation of proteins. Crucially it leads to the compact conformation of globular proteins (most enzymes) as the hydrophobic groups are buried in the centre of the globules to avoid contact with the surrounding water. Covalent interactions are possible between adjacent amino acids and can produce solid protein aggregates (Figures 1.4 and 1.6). For example, disulfide linkages are possible in proteins that contain cysteine, and these form the strong inter-protein linkages found in many fibrous proteins e.g. keratins in hair.

Table 10.1 Fundamental physical properties of amino acids found in protein. Ref.: Data adapted from C.K. Mathews and K.E. Van Holde, Biochemistry, 137].

Name	pK$_a$ value of side chain	Mass of residue	Occurrence in natural proteins (%mol)
Alanine	—	71	9.0
Arginine	12.5	156	4.7
Asparagine	—	114	4.4
Apartic acid	3.9	115	5.5
Cysteine	8.3	103	2.8
Glutamine	—	128	3.9
Glutamic acid	4.2	129	6.2
Glycine	—	57	7.5
Histidine	6.0	137	2.1
Isoleucine	—	113	4.6
Leucine	—	113	7.5
Lysine	10.0	128	7.0
Methionine	—	131	1.7
Phenylalanine	—	147	3.5
Proline	—	97	4.6
Serine	—	87	7.1
Threonine	—	101	6.0
Tyrptophan	—	186	1.1
Tyrosine	10.1	163	3.5
Valine	—	99	6.9
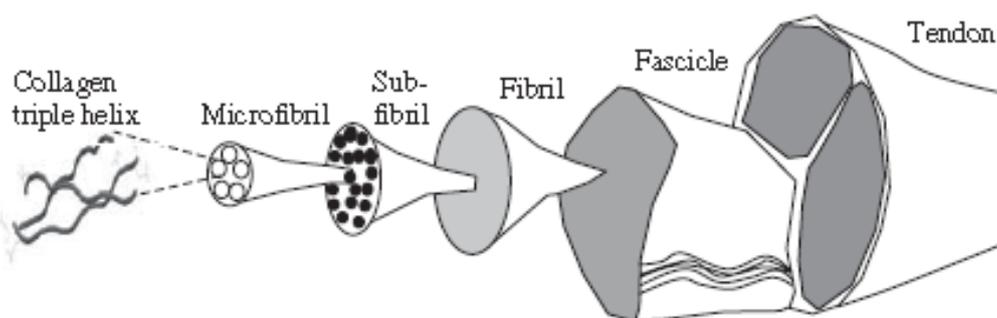
Figure 10.6. Hierarchical structure for the collagen triple helices in tendons (Collagen helices are combined into micro fibrils, then into sub-fibrils, fibrils, fascicles and finally into tendons.)

The internal secondary structures of protein chains (a helices and b sheets) are stabilized by hydrogen bonds between adjacent atoms in the peptide groups along the main chain. The important structural proteins such as keratins (Figure 10.4), collagens (Figure 10.6), silks (Figure 10.5), anthropod cuticle matrices, elastins (Figure 10.7), resilin and abductin are formed from a combination of intermolecular disulfide and hydrogen bonds.
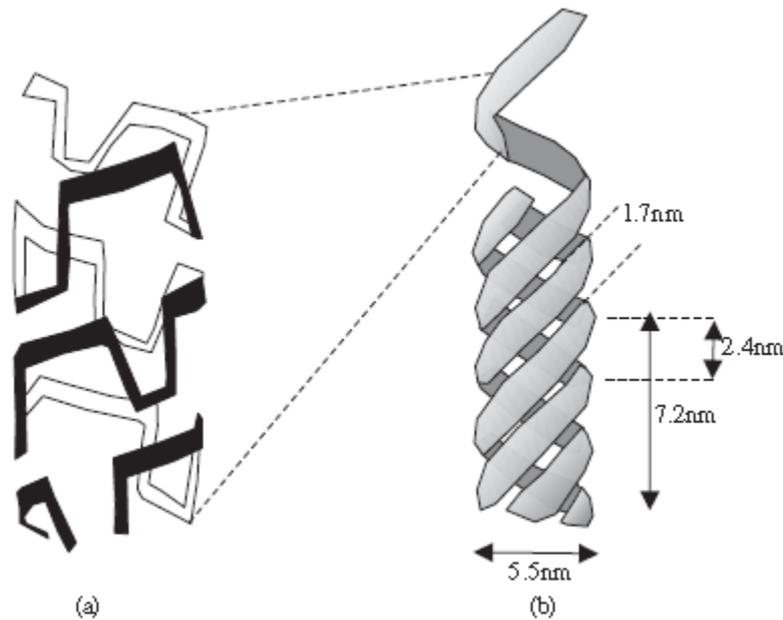
Figure 10.7. The b turns in elastin (a) form a secondary elastic helix which is subsequently assembled into a super helical fibrous structure (b).

Some examples of the globular structures adopted by proteins are shown in Figure 10.8. Globular proteins can be denatured in a folding/unfolding transition through a number of mechanisms, e. g. an increase in the temperature, a change of pH, and the introduction of hydrogen bond breaking chaotropic solvents. Typically, the complete denaturation transition is a first order thermodynamic phase change with an associated latent heat (the thermal energy absorbed during the transition). The unfolding process involves an extremely complex sequence of molecular origami transitions. There are a vast number of possible molecular configurations (~10N for an N residue protein) that occur in the reverse process of protein folding, when the globular protein is constructed from its primary sequence by the cell, and thus frustrated structures could easily be formed during this process. Indeed, at first sight it appears a certainty that protein molecules will become trapped in an intermediate state and never reach their correctly folded form. This is called Levinthal's paradox, the process by which natural globular proteins manage to find their native state among the billions of possibilities in a finite time. The current explanation of protein folding that provides a resolution to this paradox, is that there is a funnel of energy states that guide the kinetics of folding across the complex energy landscape to the perfectly folded state (Figure 10.9).

There are two main types of inter-chain interaction between different proteins in solution; those in which the native state remains largely unperturbed in processes such as protein crystallization and the formation of filaments in sheets and tapes, and those interactions that lead to a loss of conformation e.g. heat set gels (e. g. table jelly and boiled eggs) and amyloid fibers (e.g. Alzheimer's disease and Bovine Spongiform Encephalopathy).
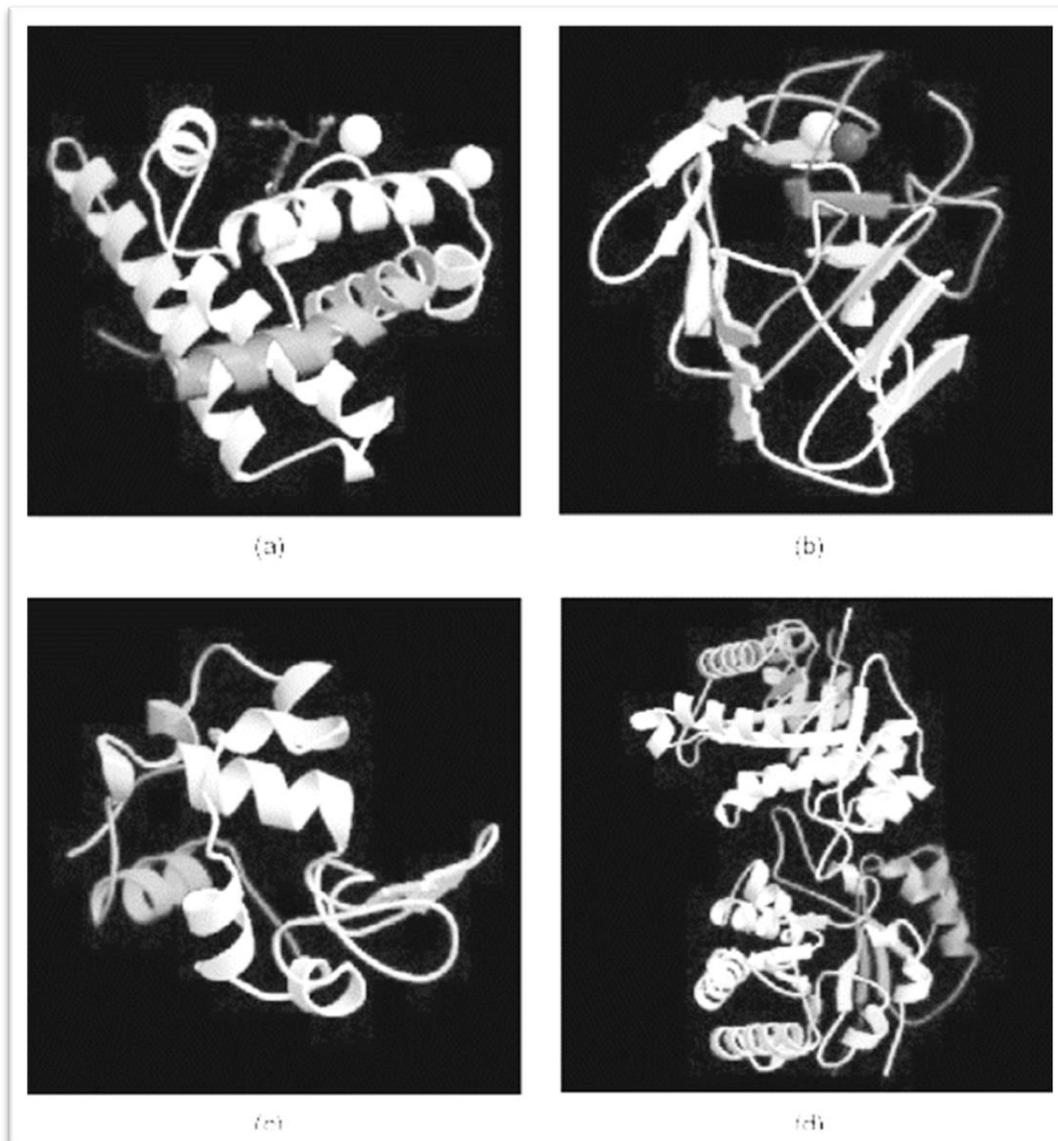
Figure 10.8. Ribbon diagram for four typical protein structures in different structural classes (a) all-α (4MBN), (b) all-β (3CNA), (c) α+β (4LYZ), and (d) α/β (1TIM). Figure was adapted from Gromiha and Selvaraj (2004).
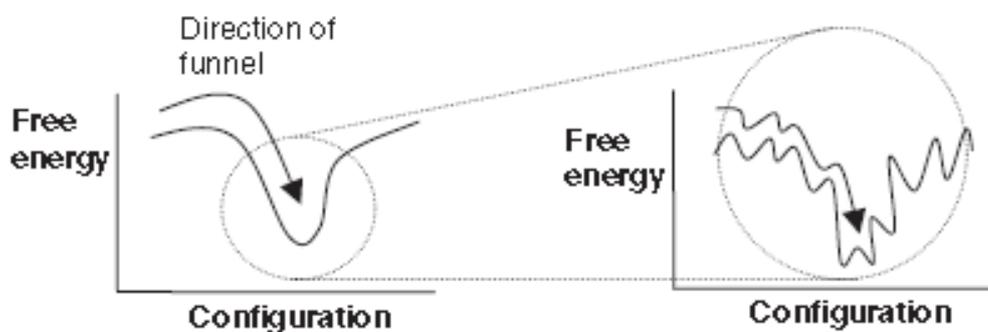
Figure 10.9. Schematic diagram indicating the funnel that guides the process of protein folding through the complex configuration space that contains many local minima. The funnel avoids the frustrated misfolded protein structures described in Levinthal's paradox.

Lipids

Cells are divided into a series of subsections or compartments by membranes which are formed predominantly from lipids. The other main role of lipids is as energy storage compounds, although the molecules play a role in countless other physiological processes. Lipids are amphiphilic, the head groups like water (and hate fat) and the tails like fat (and hate water). This amphiphilicity drives the spontaneous self-assembly of the molecules into membranous morphologies. There are four principal families of lipids: fatty acids with one or two tails (including carboxylic acids of the form RCOOH where R is a long hydrocarbon chain), and steroids and phospholipids where two fatty acids are linked to a glycerol backbone (Figure 10.10). The type of polar head group differentiates the particular species of naturally occurring lipid. Cholesterol is a member of the steroid family and these compounds are often found in membrane structures. Glycolipids also occur in membranes and in these molecules the phosphate group on a phospholipid is replaced by a sugar residue. Glycolipids have important roles in cell signaling and the immune system. For example, these molecules are an important factor in determining the compatibility of blood cells after a blood transfusion, i.e. blood types A, B, O, etc.
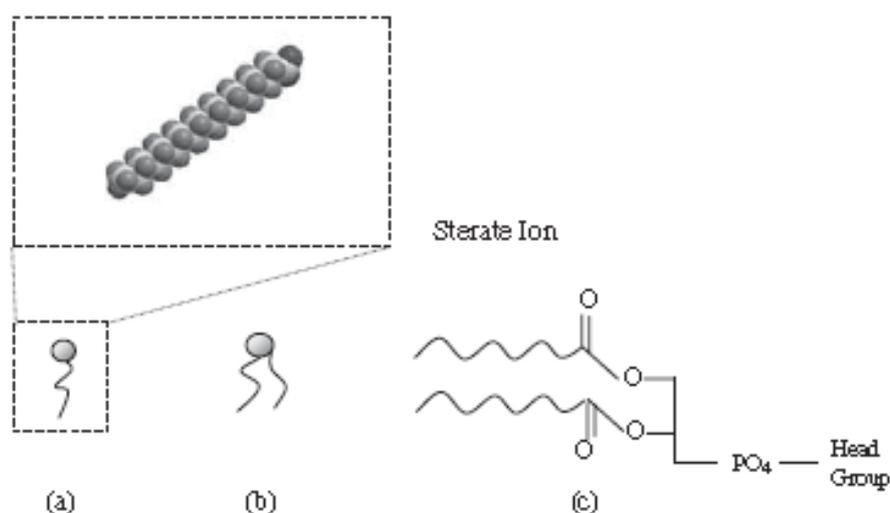
Figure 10.10. Range of lipid molecules typically encountered in biology (a) fatty acids with one tail; (b) steroids and fatty acids with two tails; (c) phospholipids.

Nucleic Acids

The 'central dogma of biochemistry' according to F.C. Crick is illustrated in Figure 10.11. DNA contains the basic blueprint for life that guides the construction of the vast majority of living organisms. To implement this blue print cells need to transcribe DNA to RNA, and this structural information is subsequently translated into proteins using specialized protein factories (the ribosomes). The resultant

proteins can then be used to catalyze specific chemical reactions or be used as building materials to construct new cells.

This simple biochemical scheme for transferring information has powerful implications. DNA can now be altered systematically using recombinant DNA technology and then placed inside a living cell. The foreign DNA hijacks the cell's mechanisms for translation and the proteins that are subsequently formed can be tailor-made by the genetic engineer to fulfil a specific function, e.g. bacteria can be used to form biodegradable plastics from the fibrous proteins that are expressed.

Figure 10.11. The central dogma of molecular biology considers the duplication and translation of DNA. DNA is duplicated from a DNA template. DNA is transcribed to form RNA chain, and this information is translated into a protein sequence.

The monomers of DNA are made of a sugar, an organic base and a phosphate group (Figure 10.12). There are only four organic bases that naturally occur in DNA, and these are thymine, cytosine, adenine and guanine (T, C, A, G). The sequence of bases in each strand along the backbone contains the genetic code. The base pairs in each strand of the double helical DNA are complementary, A has an affinity for T (they form two hydrogen bonds) and G for C (they form three hydrogen bonds). The interaction between the base pairs is driven by the geometry of the hydrogen bonding sites. Thus, each strand of the DNA helix contains an identical copy of the genetic information to its complementary strand, and replication can occur by separation of the double helix and resynthesis of two additional chains on each of the two original double helical strands. The formation of helical secondary structures in DNA drastically increases the persistence length of each separate chain and is called a helix-coil transition.
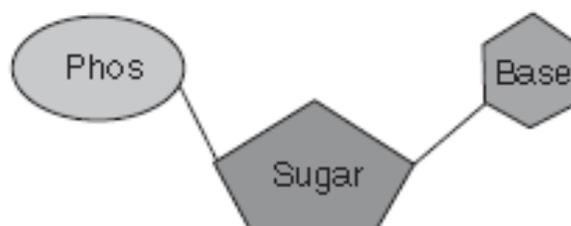
Figure 10.12. The chemical structure of the base of a nucleic acid consists of a phosphate group, a sugar and a base.

There is a major groove and a minor groove on the biologically active A and B forms of the DNA double helix. The individual polynucleotide DNA chains have a sense of direction, in addition to their individuality (a complex nucleotide sequence). DNA replication in vivo is conducted by a combination of the DNA

polymerases (I, II and III). DNA in its double helical form can store torsional energy, since the monomers are not free to rotate (like a telephone cable). The ends of a DNA molecule can be joined together to form a compact super coiled structure that often occurs in vivo in bacteria; this type of molecule presents a series of fascinating questions with regard to its statistical mechanics and topological analysis.

DNA has a wide variety of structural possibilities (Table 10.2, Figure 10.13). There are 3 standard types of averaged double helical structure labelled A, B and Z, which occur ex vivo in the solid fibers used for X-ray structural determination. Typically, DNA in solution has a structure that is intermediate between A and B, dependent on the chain sequence and the aqueous environment. An increase in the level of in the level of hydration tends to increase the number of B type base pairs in a double helix. Z-type DNA is favored in some extreme non-physiological conditions.

Table 10.2. Structural parameters of polynucleotide helices

Property	A form	B form	Z-form
Direction of helix rotation	Right	Right	Left
Number of residues per turn	11	10	12
Rotation per residue	33°	36°	30°
Rise in helix per residue	0.255 nm	0.34 nm	0.37 nm
Pitch of helix	2.8 nm	3.4 nm	4.5 nm
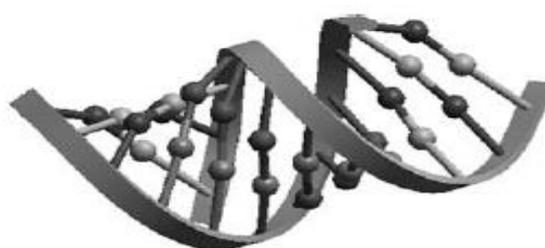
A-DNA
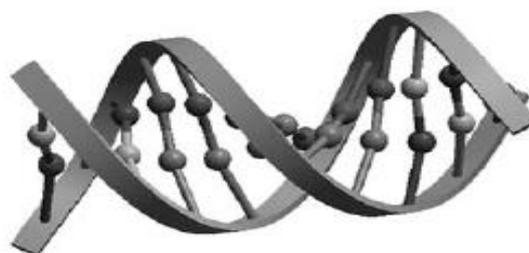
B-DNA

Z-DNA

Figure 10.13. Molecular models of A, B and Z type double helical structures of DNA (A and B type helical structures, and their intermediates typically occur in biological systems. Z-DNA helical structures crystallize under extreme non-physiological conditions.).

There are a number of local structural modifications to the helical structure that are dependent on the specific chemistry of the individual DNA strands, and are in addition to the globally averaged A, B and Z classifications. The kink is a sudden bend in the axis of the double helix which is important for complexation in the nucleosome. The loop contains a rupture of hydrogen bonds over several base pairs, and the separation of two nucleotide chains produces loops of various sizes. In the process of DNA transcription RNA polymerase is bound to DNA to form a loop structure. In the process of breathing of a double helix, hydrogen bonds are temporarily broken by a rapid partial rotation of one base pair. The hydrogen atoms in the NH groups are therefore accessible and can be exchanged with neighboring protons in the presence of a catalyst. The cruciform structure is formed in the presence of self -complementary palindromic sequences separated by several base pairs.

Hydrophobic molecules (e.g. DNA active drugs) can be intercalated into the DNA structure, i.e. slipped between two base pairs. Helices that contain three or four nucleic acid strands are also possible with DNA, but do not occur naturally. DNA has a number of interesting features with respect to its polymer physics. The persistence length (lp) of DNA is in the order of 50 nm for E. coli (which depends on ionic strength), it can have millions of monomers in its sequence and a correspondingly gigantic contour length (L) (for humans L is ~1.5 m). The large size of DNA has a number of important consequences; single fluorescently labelled DNA molecules are visible under an optical microscope, which proves very useful for high resolution experiments, and the cell has to solve a tricky packaging problem in vivo of how to fit the DNA inside the nucleus of a cell which is, at most, a few microns in diameter (it uses chromosomes).

Carbohydrates

Historically, advances in carbohydrate research have been overshadowed by developments in protein science. This has in part been due to the difficulty of analyzing of the structure of carbohydrates, and the extremely large variety of chemical structures that occur naturally.

Carbohydrates play a vital role in a vast range of cellular processes that are still only partly understood. There are two important glucose polymers which occur in plants that are differentiated by the linkage between the monomers: cellulose and amylopectin. Cellulose is a very rigid polymer, and has both nematic and semi-crystalline phases. It is used widely in plants as a structural material. The straight chain formed by the β (1-4) linkage between glucose molecules is optimal for the construction of fibers, since it gives them a high tensile strength in the chain direction (Figures 10.14 and 10.15), and reasonable strength

perpendicular to the chain due to the substantial intrachain hydrogen bonding in sheet-like structures. Amylose and its branched form, amylopectin (starch), are used in plants to store energy, and often amylopectin adopts smectic liquid crystalline phases (Figure 10.16).
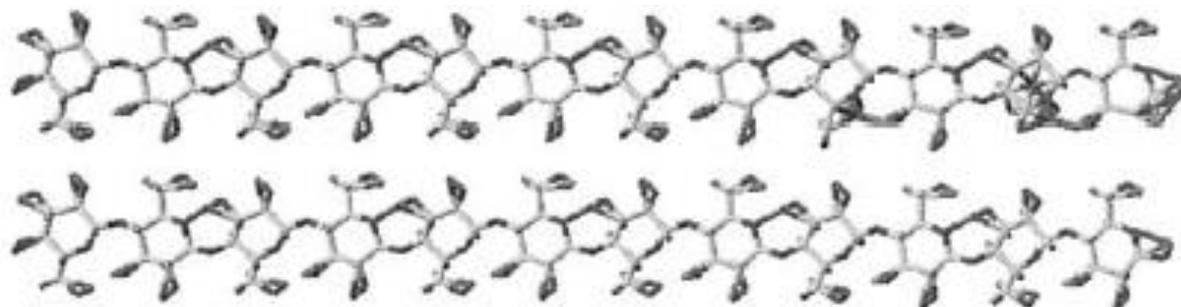

Figure 10.14 Sheet-like structures formed in cellulosic materials. (The β (1-4) Þ linkages between glucose monomers induce extended structures, and the cellulose chains are linked together with hydrogen bonds.)
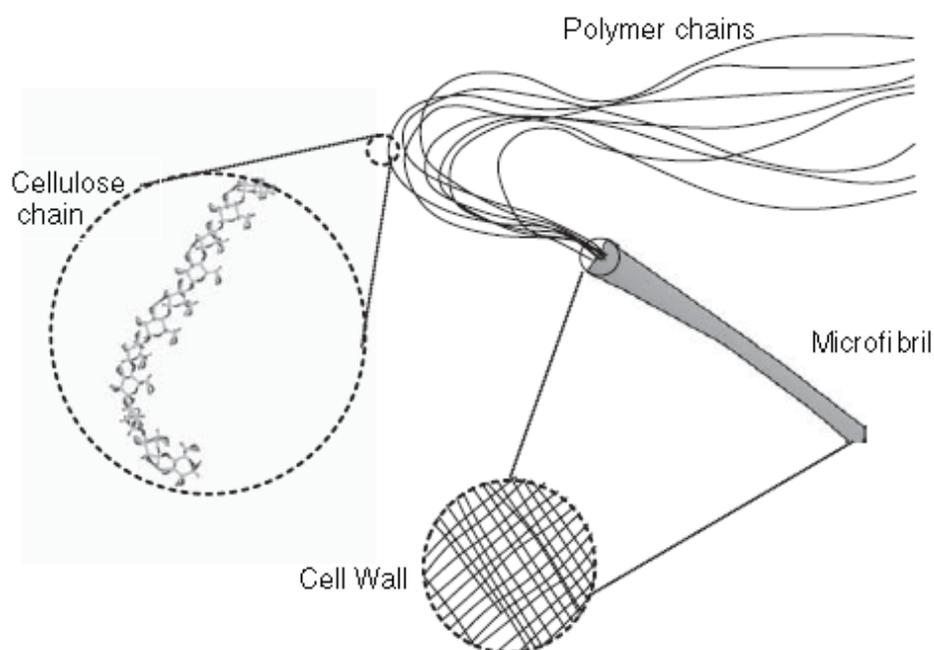

Figure 10.15. The hierarchical structure of cellulose found in plant cell walls (Cellulose chains are combined into microfibrils that form the walls of plant cells [Ref.: adapted from C.K. Mathews and K.E. Van Holde, Biochemistry, Benjamin Cummings]).
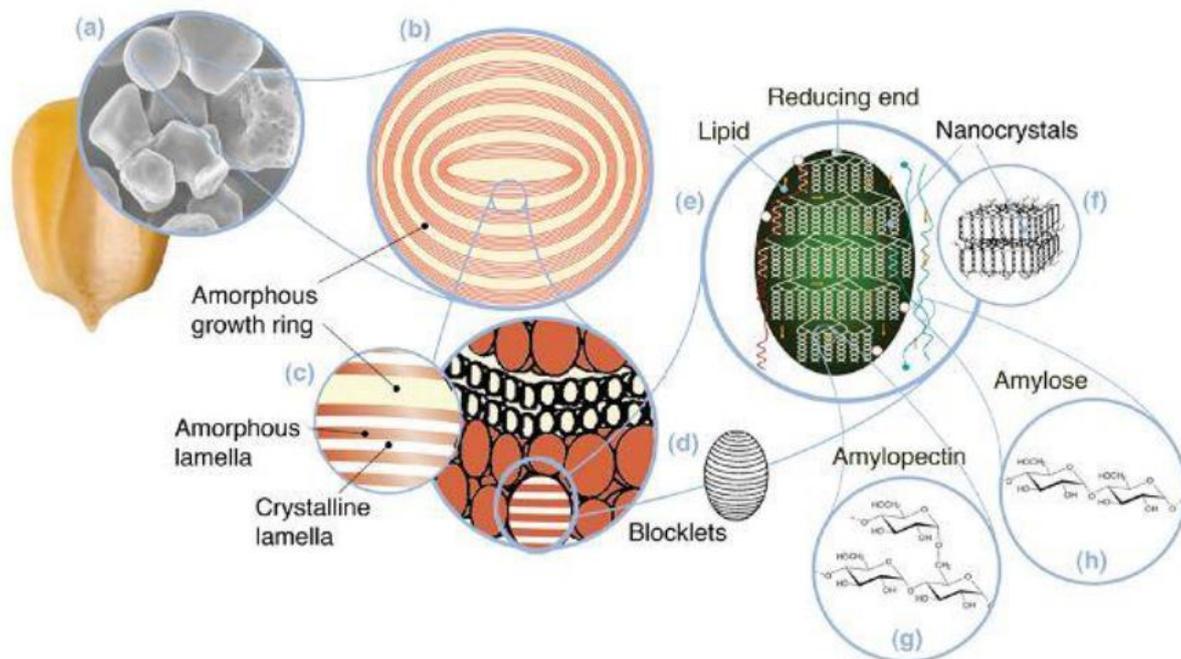
Figure 10.16. Multi-scale structure of starch granules. [Bin Zhang. Molecular Organisation, Physical and Digestion Properties of Less-Ordered Starch Matrices, PhD Thesis, The University of Queensland, Australia, 2015].

Starch, an amylose/amylopectin composite, forms the principal component of mankind's food source. In amylose the glucose molecules are connected together with an β (1-4) linkage. α-linkages between the glucose molecules are well suited to the formation of an accessible sugar store, since they are flexible and can be easily degraded by enzymes. Amylopectins are formed from amyloses with additional branched α(1-6) flexible linkages between glucose molecules Glycogen is an amorphous hyperbranched glucose polymer analogous to amylopectin, and is used inside animal cells as an energy store.

Chitin is another structural polysaccharide; it forms the exoskeleton of crustaceans and insects. It is similar in its functionality to cellulose it is a very rigid polymer and has a cholesteric liquid crystalline phase. It must be emphasized that the increased complexity of linkages between sugar molecules, compared with nucleic acids or proteins, provides high density mechanism for encoding information. A sugar molecule can be polymerized in a large number of ways, e.g. the six corners of a glucose molecule can each be polymerized to provide an additional N^6 arrangements for a carbohydrate compared with a protein of equivalent length (N). In proteins there is only one possible mechanism to connect amino acids, the peptide linkage. These additional possibilities for information storage with carbohydrates are used naturally in a range of immune response mechanisms.

Pectins are extra cellular plant polysaccharides forming gums (used in jams), and similarly algins can be extracted from sea weed. Both are widely used in the food industry. Hyaluronic acid is a long negatively charged semi-flexible polyelectrolyte and occurs in a number of roles in animals. For example, it is found as a component of cartilage (a biological shock absorber) and as a lubricant in synovial joints.

Fresh granule ghosts from maize and potato starches were isolated by centrifugation. This method ensures full gelatinization and maximum swelling (for most normal starches) when starch is heated at 95 °C for 30 min in excess water. The risks of retrogradation from leached amylose and shear degradation are minimized by the low starch concentration (< 1% w/v) and low stirring rate during ghost preparation respectively. Light microscopy and particularly confocal microscopy suggest that ghost particles are essentially hollow with a surface 'skin' composed of apparently densely packed starch (Figure 10.17 A1 and B1). Granule ghosts have characteristic microscopic features depending on their botanical and genetic origins.

The morphology of granule ghosts was also investigated using CLSM after being stained by Nile blue, a water-soluble basic oxazine dye, which is one of the most suitable fluorescent staining agents for both granular and gelatinized starches in the absence of proteins. CLSM micrographs of granule ghosts show isolated balloon-like structures, lacking their original contents (Figure 10.17 A1 and B1 insets). The skins of maize ghosts appear to be thicker and more substantial and do not show much deterioration or collapse, compared with potato ghosts.
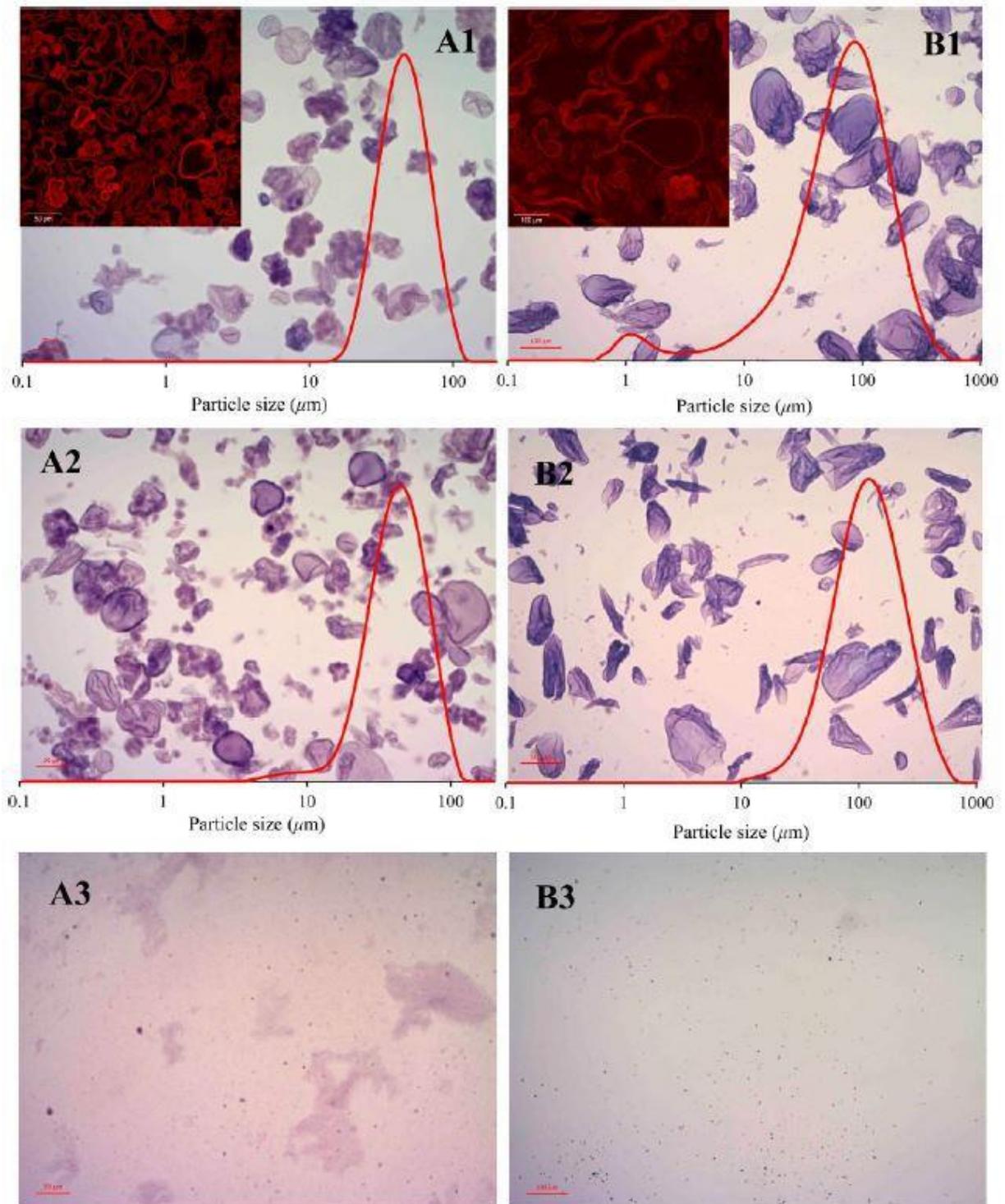
Figure 10.17. Light micrographs and particle size distribution of fresh granule ghosts (A, maize, scale bar = 20 μm; B, potato, scale bar = 100 μm) with/without SDS pre-treatment or high shear force (1, without any treatment; 2, with SDS pre-treatment; 3, with high shear force treatment after ghost formation). Insets in panels A1 and B1 are confocal microscopy images after staining with Nile Blue [Atkin, N. J.; Abeysekera, R. M.; Robards, A. W., The events leading to the formation of ghost remnants from the starch granule surface and the contribution of the granule surface to the gelatinization endotherm. Carbohydr Polym 1998, 36, 193-204].

Water Molecules

Water is a unique polar solvent and its properties have a vast impact on the behavior of biological molecules (Figure 10.18). Water has a high dipole moment (P) of $6,11 \times 10^{30}$ Cm, a quadrupole moment of $1,87 \times 10^{39}$ Cm^2 and a mean polarizability of $1,44 \times 10^{30}$ m3.

Water exists in a series of crystalline states at subzero temperature or elevated pressures. The structure of ice formed in ambient conditions has unusual cavities in its structure due to the directional nature of hydrogen bonds, and it is consequently less dense than liquid water at its freezing point. The polarity of the O–H bonds formed in water allows it to associate into dimers, trimers etc. (Figure 10.19), and produces a complex many body problem for the statistical description of water in both liquid and solid condensed phases.
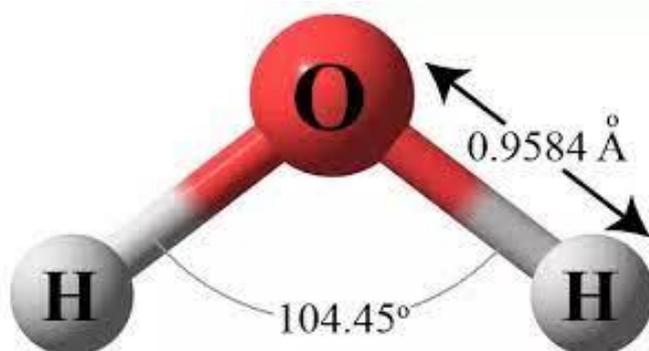
Figure 10.18. The geometry of a single water molecule. (The molecule tends to form a tetrahedral structure once hydrogen bonded in ice crystals).
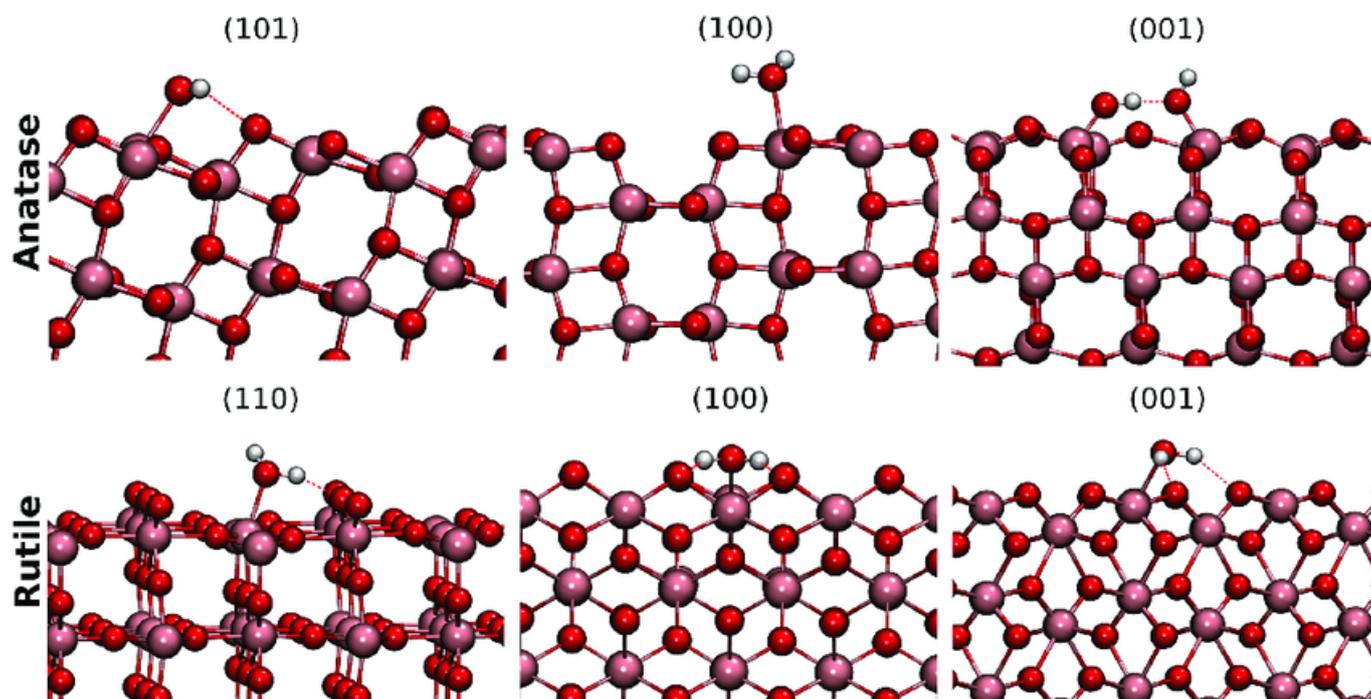
Figure 10.19. Schematic diagram of the network structure formed by water molecules. (Lines indicate hydrogen bonds. Such chains of hydrogen bonded water molecules occur over a wide range of angles for liquid water.)

Antifreeze proteins have been designed through evolution to impair the ability of the water that surrounds them in solution to crystallize at low temperatures. They have an alpha helical dipole moment that disrupts the hydrogen bonded network structure of water. These antifreeze molecules have a wide range of applications for organisms that exist in subzero temperatures e.g. arctic fish and plants.

The imaging of biological processes is possible in vivo using the technique of nuclear magnetic resonance, which depends on the mobility of water to create the image. This powerful non- invasive method allows water to be viewed in a range of biological processes, e.g. cerebral activity.

Even at very low volume fractions water can act as a plasticizer that can switch solid biopolymers between glassy and non-glassy states. The ingress of water can act as a switch that will trigger cellular activity in plant seeds, and such dehydrated cellular organisms can remain dormant for many thousands of years before being reactivated by the addition of water.

A wide range of time scales (10^{-18}–10^{3}s) of water are important to understand its biological function (Figure 10.20). The range of time scales includes such features as the elastic collisions of water at ultrafast times ($\sim 10^{-15}$ seconds) to the macroscopic hydrodynamic processes observed in blood flow at much slower times ($\sim$seconds).
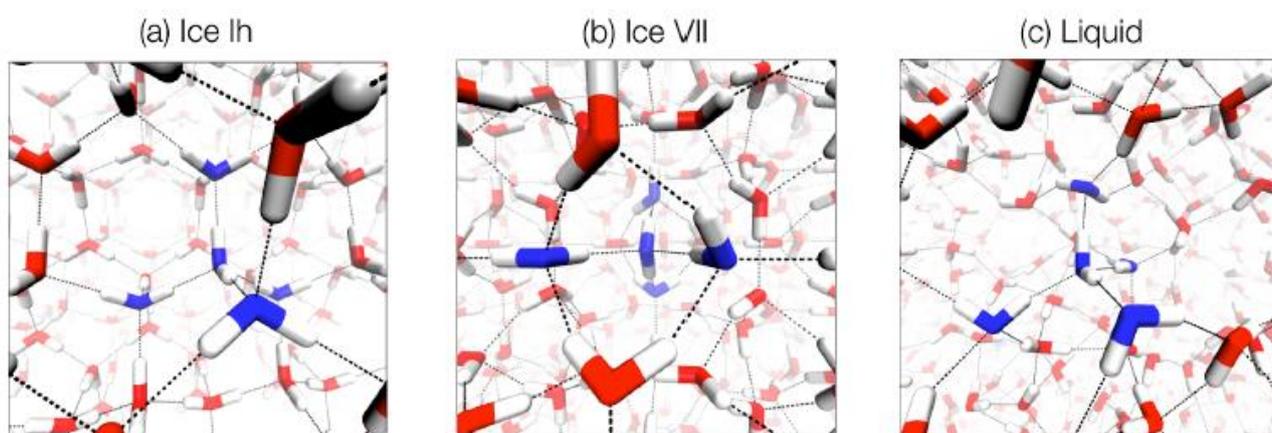
Figure 10.20. Typical configurations of liquid and solid water taken from molecular simulations. Hydrogen bonds are drawn in black and a representative tetrahedrally coordinated molecule and its neighbors are highlighted in blue.

Proteoglycans and glycoproteins

Proteoglycans (long carbohydrate molecules attached to short proteins) and glycoproteins (short carbohydrate molecules attached to relatively long proteins) are constructed from a mixture of protein and carbohydrate molecules (glycosaminoglycans). In common with carbohydrates, proteoglycans/glycop-roteins exhibit extreme structural and chemical heterogeneity. Furthermore, the challenges presented to

crystallography by their non-crystallinity means that a full picture of the biological function of these molecules is still not complete.

Many proteoglycans and glycoproteins used in the extracellular matrix have a bottle brush morphology (Figures 10.21 and 10.22).

Proteoglycan Turnover

- Shedding by exoproteolytic activity, MMP-7 for one

- Endosulfatases remove sulfate groups on proteoglycans at cell surface: remodeling

- Heparanase (endohexosaminidase) clips at certain sites in the chain. Outside cells, it plays a role in cell invasion processes

- Inside cells it's the first step towards complete degradation in lysosomes by exoglycosidases and sulfatases
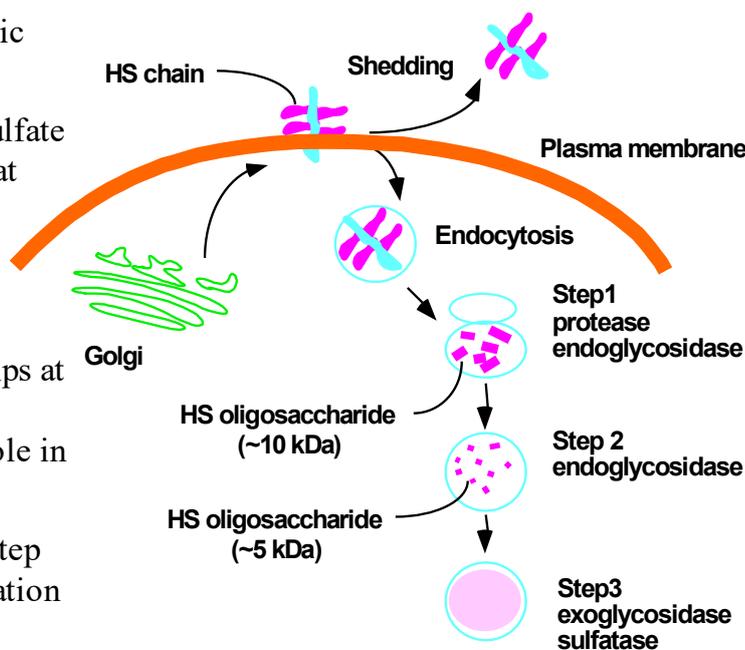
Mucopolysaccharidoses

After Esko, J

Figure 10.21. The Scheme of turnover of proteoglycans
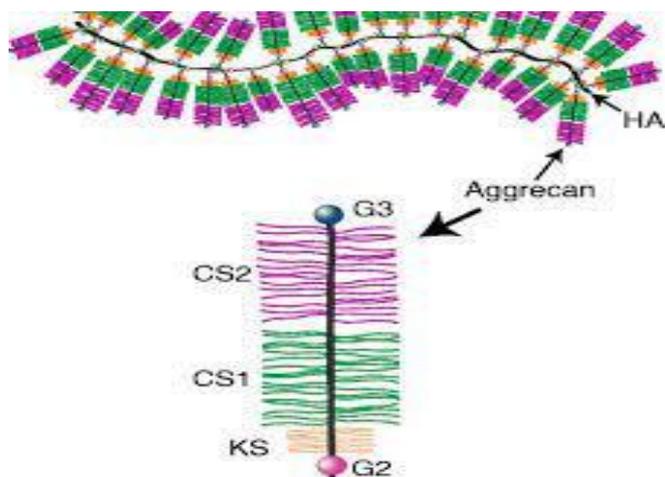
Figure 10.22. The structure of proteoglycan aggregates. Researchgate net.

An example of a sophisticated proteoglycan is aggrecan, a giant polymeric molecule that consists of a bottle-brush of bottle-brushes (Figure 10.21). These materials have a very large viscosity in solution, and are used to dissipate energy in collagenous cartilage composites and to reduce friction in synovial joints as boundary lubricants. An example of an extracellular glycoprotein is the mucins found in the stomach of mammals. These molecules experience telechelic (either end) associations to form thick viscoelastic gels that protect the stomach lining from auto digestion (Figure 1.22). Other examples of glycoproteins occur in enzymes (Ribonuclease B), storage protein (egg white), blood clots (fibrin) and antibodies (Human IgG).

Different construction of cells

Cells act co-operatively in multicellular organisms and are hierarchically arranged into tissues, organs and organ systems. Tissues contain both cells and other materials such as the extracellular matrix. There are four distinct forms of mammalian muscle cells: skeletal and cardiac (which both forms striated muscular tissues), smooth muscle (found in blood vessels and intestines) and myoepithelial cells (again present in intestines). Nerve cells are used to send and receive signals. They are highly branched and this structure allows them to react to up to one hundred thousand inputs from other cells. The electrochemistry of nerve cells is a fascinating area; the efficiency and time response of these electrical circuits has been carefully optimized bio evolution.

Blood cells have a squashed donut shape (Figure 10.23) which is related to the differential geometry of their cytoskeleton. Red blood cells carry oxygen and carbon dioxide, towards and away from the lungs. White blood cells play a role in the fight to remove infections from an organism. Fibroblast cells are largely responsible for the secretion and regulation of the extracellular matrix, e.g. the production of molecules such as the collagens. Epithelial cells control the passage of material across the boundary of organs, e.g. in the interior of the intestinal tract.
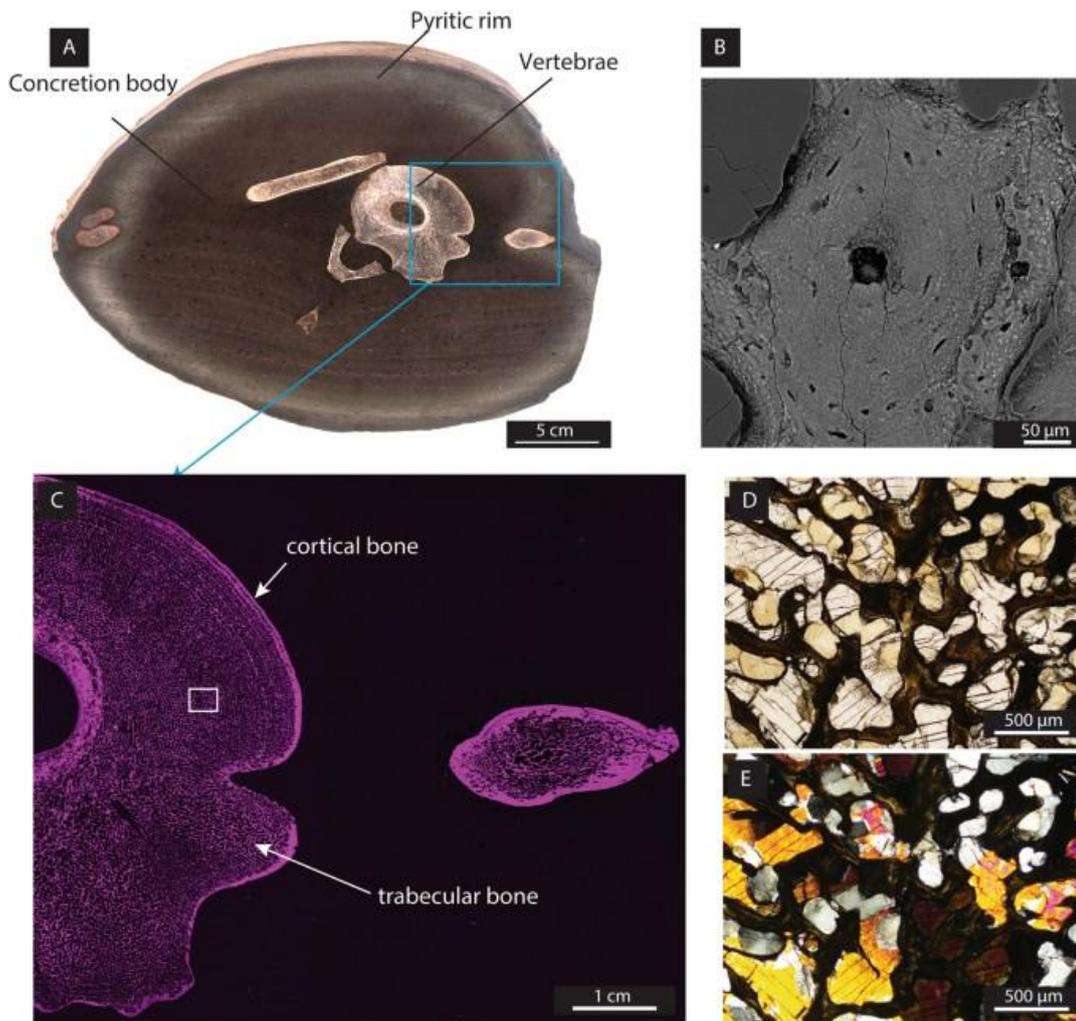
Figure 10.23. White blood cell-like structures, Source: Nature.com.

Viruses as biomolecules

Viruses are intra-cellular parasites, biological entities that multiply through the invasion of cellular organisms. In addition to aspects related to their biological role in disease, viruses have attracted a great deal of attention from biophysicists for their physical properties. Viruses self - assemble into well-defined monodisperse geometrical shapes (rods and polyhedra) (Figure 10.24, 10 25.) from their constituent components. Such materials have proven ideal model systems for the examination of the phase behaviour of charged colloids and lyotropic liquid crystals and allow the processes involved in their self-assembly to be investigated in detail.

To fully appreciate the powerful use of virus-like (VLPs) nano-particles one could understand the surface chemistry and surface physics of particles, physico-mechanical, chemical, bio-chemical properties and related mechanisms of viruses and virus-like particles. Generally, biological objects are working as mechanical machines, but fundamentally they are functioning on a different principals because of their nano-scale size. The common and unique properties of nanobio particles (virus-like particles - VLPs) and field of application may be briefly presented by scheme (Figure 10.26).
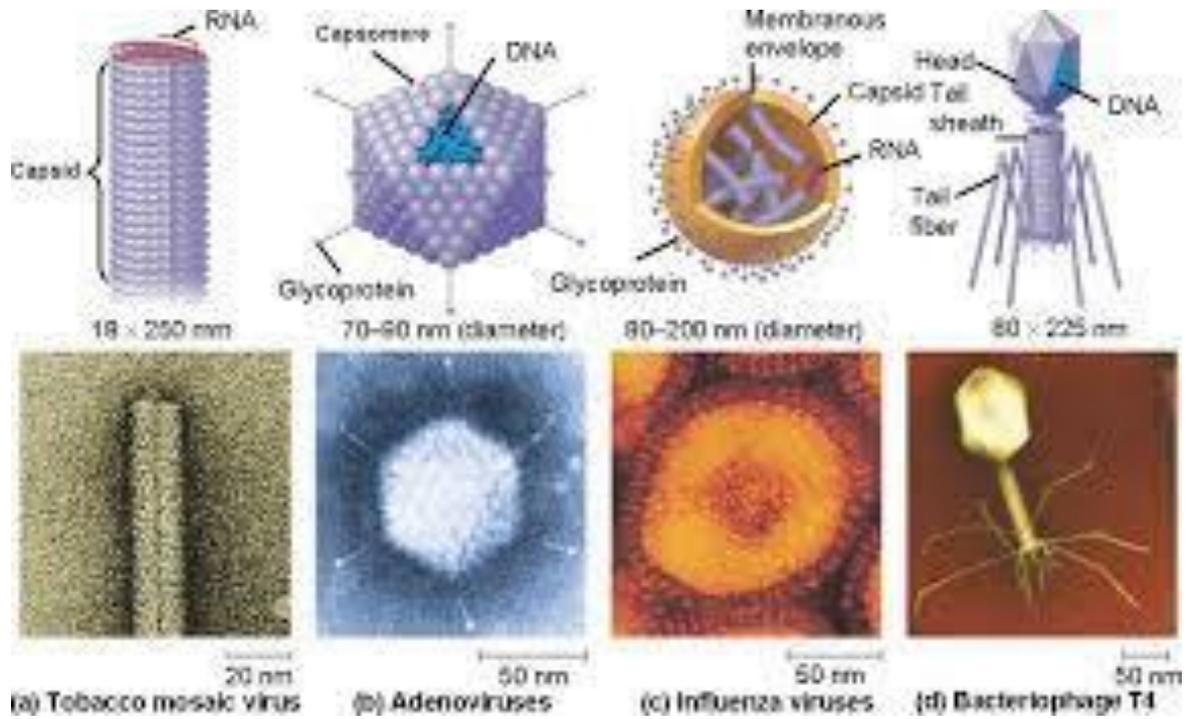
Figure 10.24. Virus Classification on the basis of morphology and replication. Source: January 18, 2020 by Somak Banerjee
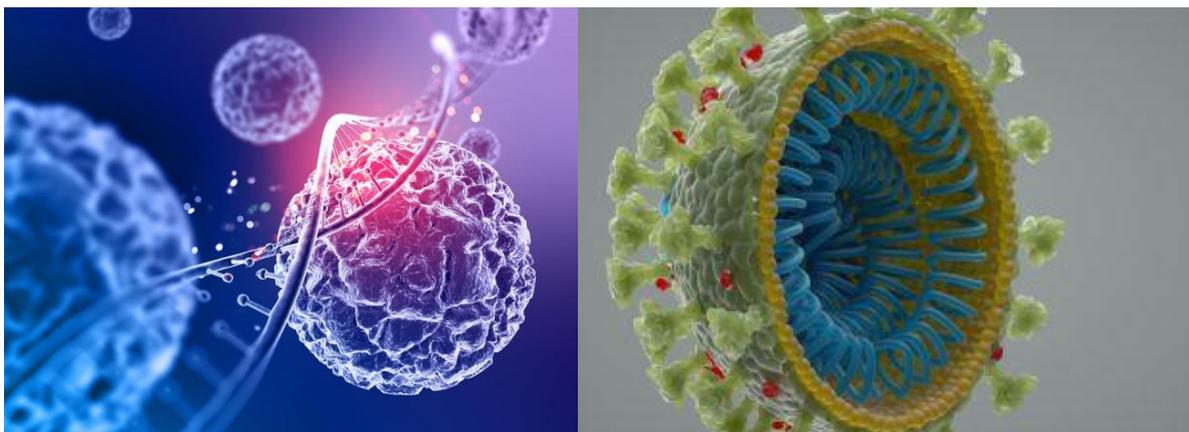
a   b

Figure 10.25.a) Range of virus structures; b) Corona virus

Study the common and specific properties of viral particles possess the great potential for developing the new concepts of detection and sensing bio-molecules, monitoring macro/micro environment, medical diagnosis and treatment. The ability to detect rapidly, directly, and selectively individual virus particles can significantly impact healthcare, improve clinical decisions, lead to better health outcomes.
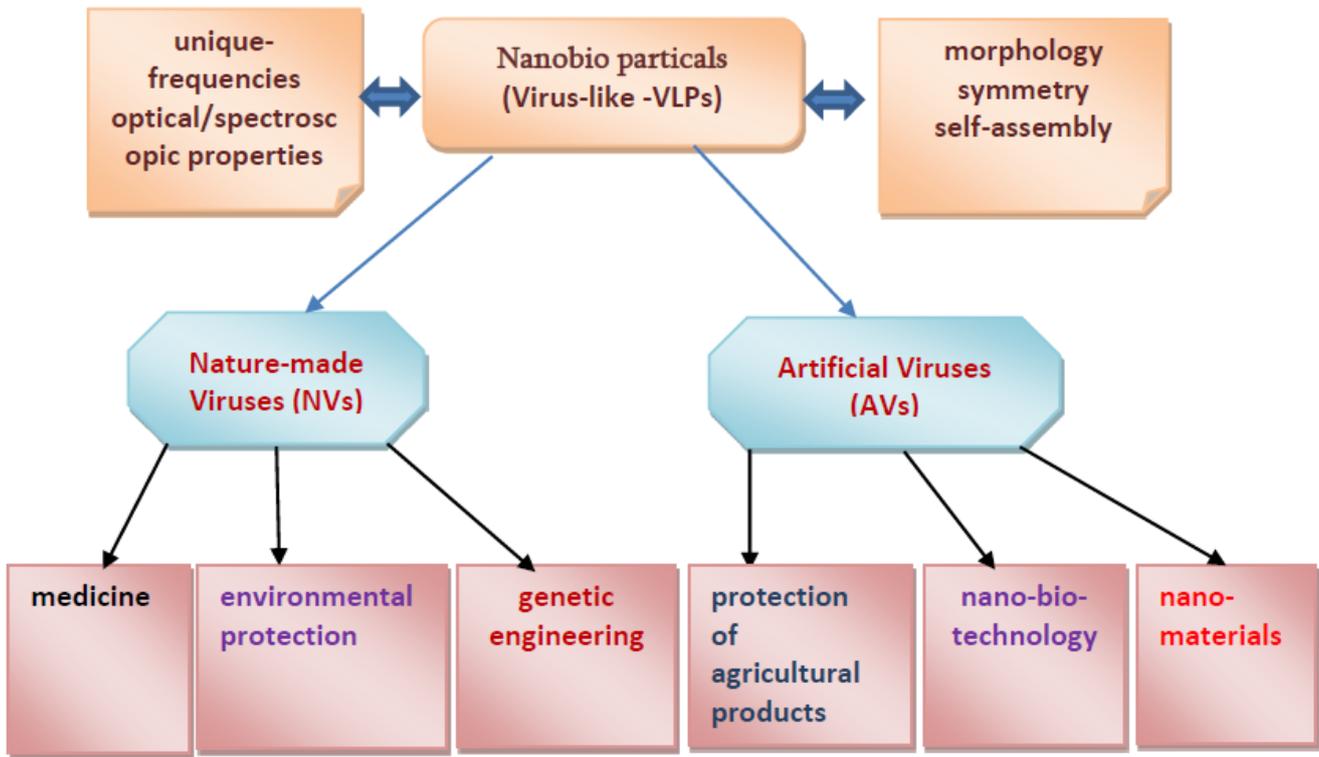
Figure 10.26. The scheme of main properties and applications of nanobioparticles (Virus-like particles - VLPs).

Behavior of nano-micro sized pathogenic microorganisms such as bacteria, virus, organic and non-organic agents is selectively sensitive toward the electromagnetic (EM) field excitation. Estimation of vibration frequencies of microorganisms is the scope of investigations revealing the Raman-active low frequency vibrational modes, and inactivation mechanisms for excitation of mechanical modes of viral capsids, resulted in mechanical resonances, also it is the object of spectroscopy methods or atomistic molecular dynamics (MD) simulation studies giving opportunity to ap-preciate the wavenumbers corresponding to specific vibration bonds of functional groups and structures.

Method of estimation of spectral response on EM field & particle interaction is based on solutions of electrodynamics two (2D) or three (3D) dimensional boundary tasks. Analytical expressions of EM fields are derived from rigorous solutions of Maxwell's and Helmholtz's equations and defined through the dimensionless parameters, diameters (d) over an excitation wave-length (λ). It makes possible to apply the classical well-known approach to sub-micro particles characterization.

Proposed method today is using for investigation of viral particle's physical properties. Namely, EM near and far fields distribution, EM spectrum studied for viruses, having rod-like, prolate un-enveloped virions (e.g. Tobacco Mosaic Virus (TMV), bacteriophage M13). Electrodynamics 2D solutions are applicable for TMV particles, the length of which is approximately 16 times greater than its diameter. Helmholtz's equation in cylindrical coordinates is converted into three differential equations. Radial equation with respect to radial function gets the form:

$$\xi^2 R''(\xi) + \xi R'(\xi) + \left(\xi^2 - m^2\right) R(\xi) = \frac{k^2}{k_q^2} f\left(\xi / k_q\right) \xi^2 R(\xi)$$

(10.1)

where, $\xi = k_q r_q$ "/" - denotes the derivative with respect to an argument. is the function of inhomogeneity (through the radius of cylinder) of dielectric permittivity. The field in different areas is written as the sums of multipole waves, the radial parts of that are represented by the solutions of equation (10.1).

Bio-particles characterization is based on the concept considering the bio-object in physical point of view. Virions, the extracellular infective forms of viruses are modelled by the particles of cylindrical shape of different structures, considered, as homogeneous dielectric particles or inhomogeneous ones through the radius, also as particles of core-shell structure reflecting the properties of ribonucleic acids (DNA or RNA) of viruses and capsid's proteins. Shape, structure and the set of geometrical, magnetic and electrical characteristics are proposed as the main parameters defining the particles' EM spectral properties. Advantage of simulation study of complex molecular systems such as virions in contrast of measuring experiments associated with weak signals detection is noteworthy. Computer simulation (based on MatLabR2013b software) is currently doing for TMV particles characterization.

Bacteria and energy biomolecules

Bacteria are small structurally simple cellular organisms. Only a minority of bacterial species have developed the ability to cause disease in humans. Bacteria take the form of spheres, rods and spirals. They are mostly encountered in terms of their mechanisms of molecular motility (Figure 10.27)

Bacteria (singular: bacterium) are classified as prokaryotes, which are single-celled organisms with a simple internal structure that lacks a nucleus, and contains DNA that either floats freely in a twisted, thread-like mass called the nucleoid, or in separate, circular pieces called plasmids. Ribosomes are the spherical units in the bacterial cell where proteins are assembled from individual amino acids using the information encoded in ribosomal RNA.

Bacterial cells are generally surrounded by two protective coverings: an outer cell wall and an inner cell membrane. Certain bacteria, like the mycoplasmas, do not have a cell wall at all. Some bacteria may even have a third, outermost protective layer called the capsule. Whip-like extensions often cover the surfaces of bacteria — long ones called flagella or short ones called pili — that help bacteria to move around and attach to a host. A few different criteria are used to classify bacteria. The organisms can be distinguished by the nature of their cell walls, by their shape, or by differences in their genetic makeup.

The Gram stain is a test used to identify bacteria by the composition of their cell walls, named for Hans Christian Gram, who developed the technique in 1884. The test stains Gram-positive bacteria, or bacteria that do not have an outer membrane. Gram-negative bacteria don't pick up the stain.

For example, Streptococcus pneumoniae (S. pneumoniae), which causes pneumonia, is a Gram-positive bacterium, but Escherichia coli (E. coli) and Vibrio cholerae, which causes cholera, are Gram-negative bacteria.

There are three basic bacterial shapes: Round bacteria called cocci (singular: coccus), cylindrical, capsule-shaped ones known as bacilli (singular: bacillus); and spiral bacteria, aptly called spirilla (singular: spirillum). The shapes and configurations of bacteria are often reflected in their names. For example, the milk-curdling Lactobacillus acidophilus are bacilli, and pneumonia-causing S. pneumoniae are a chain of cocci. Some bacteria take other shapes, such as stalked, square or star.
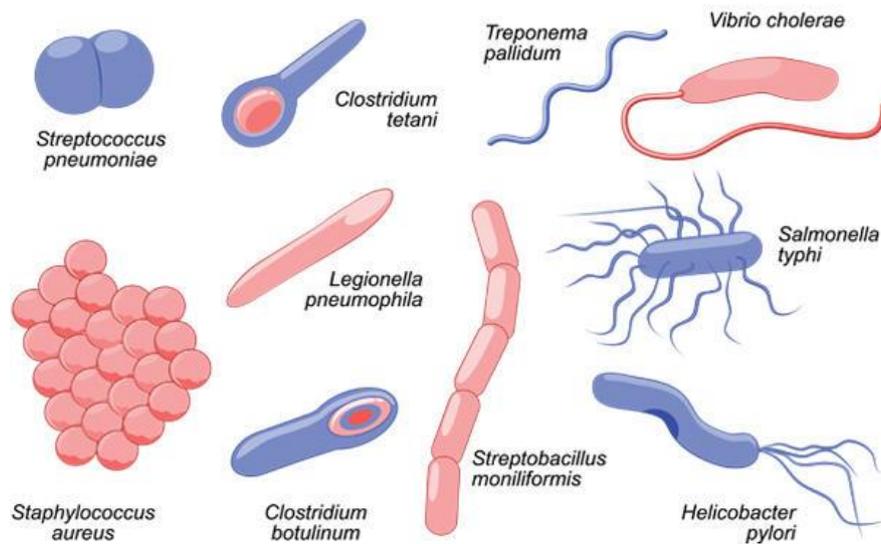
Figure 10.27. Different Range of baceria.
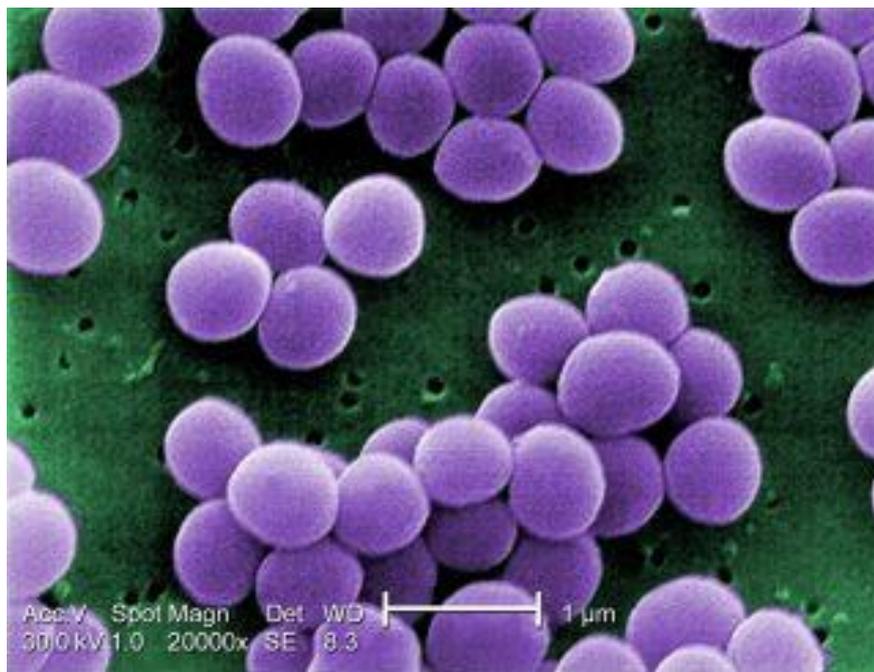
Figure 10.28. A scanning electron microscope image of resistant Staphylococcus aureus bacteria, with false color added.(Image: © Centers for Disease Control).

ADP and ATP are the 'currency of energy' in many biochemical processes. Energy is stored by the addition of the extra strongly charged phosphate group in the ATP molecule and can be released when it is metabolized into ADP. ATP (Adenosine tri-phosphate) is an important molecule found in all living things. Think of it as the "energy currency" of the cell. If a cell needs to spend energy to accomplish a task, the ATP molecule splits off one of its three phosphates, becoming ADP (Adenosine di-phosphate) + phosphate. The energy holding that phosphate molecule is now released and available to do work for the cell. When the cell has extra energy (gained from breaking down food that has been consumed or, in the

case of plants, made via photosynthesis), it stores that energy by reattaching a free phosphate molecule to ADP, turning it back into ATP. The ATP molecule is just like a rechargeable battery. When it's fully charged, it's ATP. When it's run down, it's ADP. However, the battery doesn't get thrown away when it's run down–it just gets charged up again.

ATP (Adenosine tri-phosphate) is an important molecule found in all living things. Think of it as the "energy currency" of the cell. If a cell needs to spend energy to accomplish a task, the ATP molecule splits off one of its three phosphates, becoming ADP (Adenosine di-phosphate) + phosphate. The energy holding that phosphate molecule is now released and available to do work for the cell. When the cell has extra energy (gained from breaking down food that has been consumed or, in the case of plants, made via photosynthesis), it stores that energy by reattaching a free phosphate molecule to ADP, turning it back into ATP. The ATP molecule is just like a rechargeable battery. When it's fully charged, it's ATP. When it's run down, it's ADP. However, the battery doesn't get thrown away when it's run down–it just gets charged up again. There are times when the cell needs even more energy, and it splits off another phosphate, so it goes from ADP, adenoside di-phosphate, to AMP, adenosine mono-phosphate.
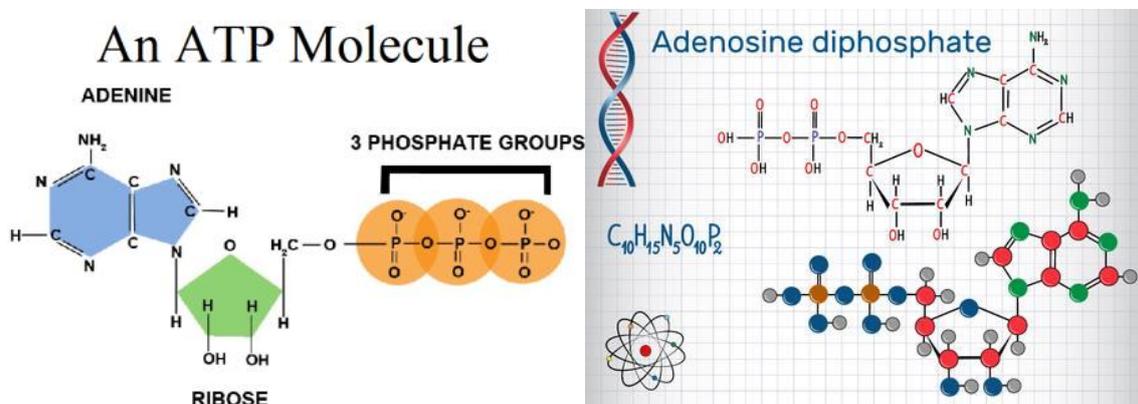
Figure 10.29. ATP and ADP molecules..
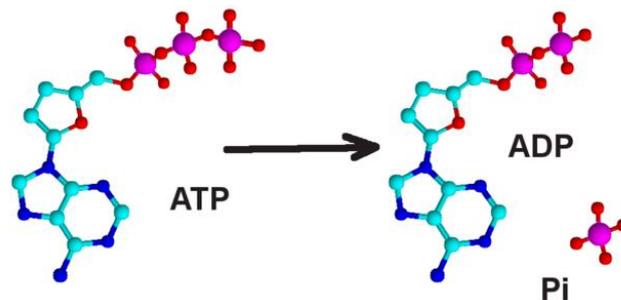
Figure 10.30. ATP – ATP circle.

Some physical approaches to investigation of biomolecules: View from applied physics

Beam Theory

Macromolecules can be viewed as networks of one-dimensional elements (chains of aminoacids, chains of base pairs, etc.). Beam theory addresses mechanisms of those networks.

A bead treated as a rod with elastic properties.
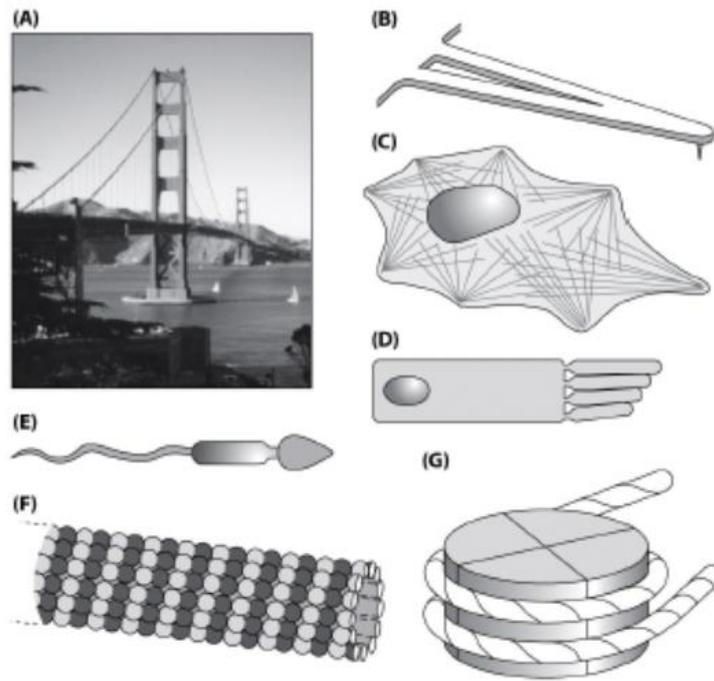
Figure 10.31. Illustrations to beam theory.
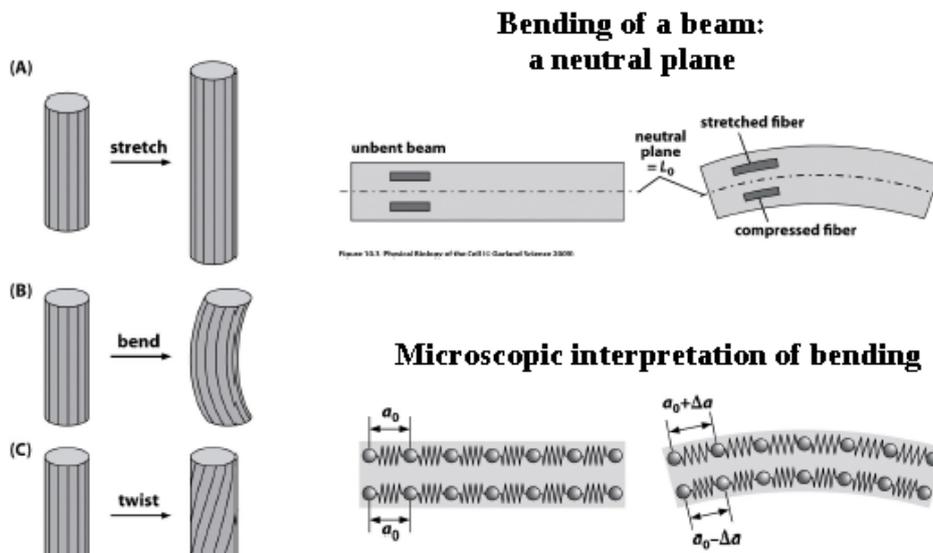

Figure 10.32. Three types of deformations.
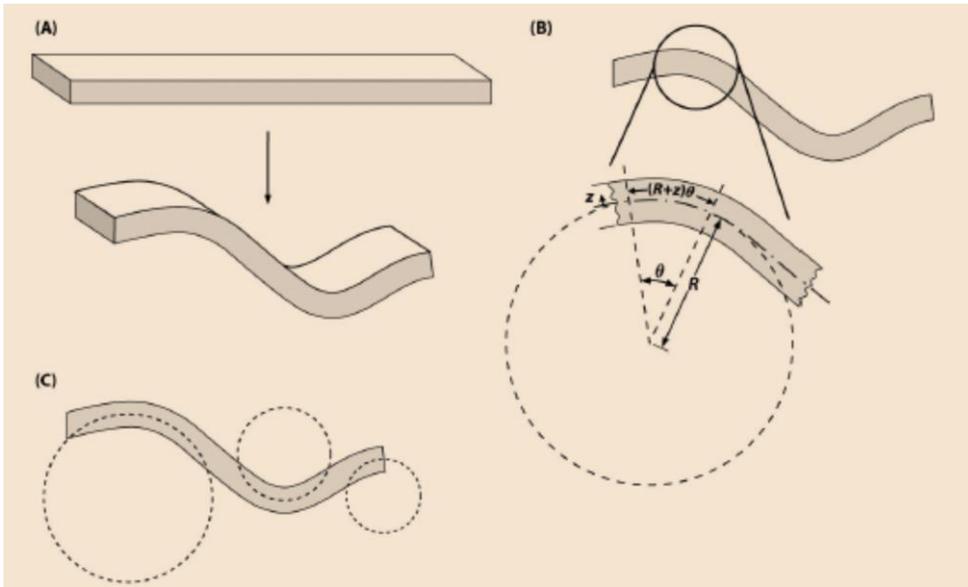
Figure 10.33. Imagine a beam split up in segments each with own curvature.

Note that the extent of extension or compression is a linear function of the distance from the neutral plane z:

$$\theta = \frac{L_0}{R} \quad L(z) = (R + z)\theta = \frac{(R + z)L_0}{R}$$

$$\theta = \frac{L_0}{R} \quad \Delta L(z) = L(z) - L_0 = (R + z)\frac{L_0}{R} - L_0\theta = z\frac{L_0}{R}$$

Extensional strain $\varepsilon(z)$ at a distance z is defined as:

$$\varepsilon(z) = \frac{\Delta L}{L_0} = \frac{z}{R}$$

Such that material above the neutral axis is stretched, $\varepsilon(z) > 0$, and material below the neutral axis is compressed, $\varepsilon(z) < 0$.

The energy cost is a quadratic function of the strain (Hooke's law):

$$w(\varepsilon) = \frac{1}{2}E\varepsilon^2 = \frac{1}{2}E\frac{\Delta L}{L_0}$$

E . . . the Young modulus

$$E_{bend} = L_0 \int_{\theta\Omega} dA \frac{E}{2R^2}z^2 = \frac{E\,|\,\mathrm{L}}{2R^2}$$

148

$$I = \int_{\partial\Omega} z^2 \, dA \qquad I \ldots \text{geometric moment}$$
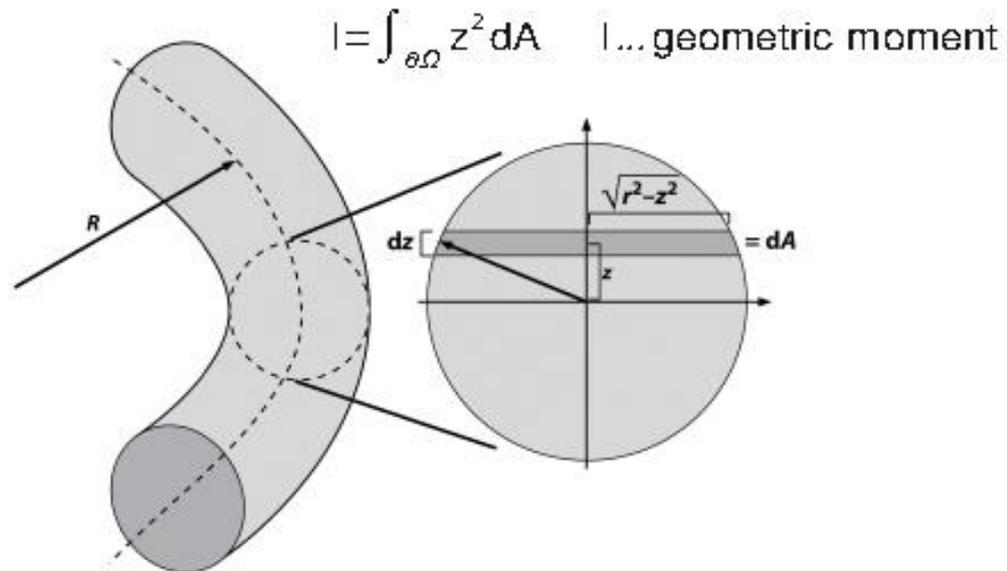
Figure 10.34. Precise calculation of the total strain energy of the beam.

$$E_{bend} = \frac{E|L}{2R^2} \qquad For \ L = 2\pi R, \ we \ get: \ E_{loop} = \frac{E|}{R}$$

We can introduce flexural rigidity: $K_{off} = E|$

Then the energy of bending is: $E_{bend} = \frac{K_{off}}{2} \int_0^L ds \, \frac{1}{R(s)^2}$

In general, we can express the bending energy in terms of a curvature: $k = \frac{1}{R(s)}$

And we finally get:

$$E_{bend} = \frac{K_{off}}{2} \int_0^L \left|\frac{dt}{ds}\right|^2 ds \left|\frac{dt}{ds}\right| \ldots \textit{derivative the tangent vector}$$

Persistence Length and Stiffness

$\rightarrow$ Persistence length is a measure of the competition between the entropy ("randomizer") and energy cost of bending

$\rightarrow$ equate the deterministic energy cost to thermal energy:

$$k_B T \approx \frac{E|L}{2R^2}$$

$\rightarrow$ persistence length is the length of a polymer for which the radius of curvature is equal to the length of polymer itself: $\xi_P \approx R \approx L$

$\rightarrow$ estimate of the persistence length: $\xi_P \approx \frac{E|}{k_B T}$

149
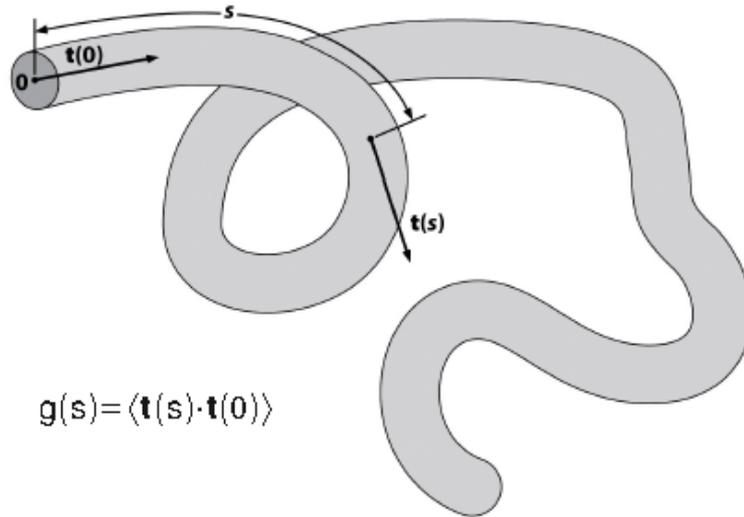
Figure 10.35. Persistence length which characterizes the correlation in the tangent vectors of different positions along the polymer.

General properties of the tangent-tangent correlation function:

$$g(s = 0) = 1 \qquad g(s \to \infty) = 0$$

$$g(s) = \langle t(s) \cdot t(0) \rangle = e^{-\frac{g}{\xi_P}}$$

Relationship between flexural rigidity and persistence length:

→ short beam of length: $L \approx s \ll \xi_P$

→ short beam is only slightly bent: $E_{bend} = \frac{E|L}{2R^2} \approx \frac{E|L}{2s} \theta^2 (s = R\theta)$

→ if a tangent at $s = 0$ points in the z-direction, then:

$$g(s) = \langle cos\theta(s) \rangle \quad \theta \ll 1 \quad \Longrightarrow \quad cos\theta \approx 1 - \frac{1}{2}\theta^2$$

$$g(s) \approx 1 - \frac{1}{2}\langle \theta^2 \rangle$$

Calculate:

$$\langle \theta(s)^2 \rangle = \frac{1}{Z} \int_0^{2\pi} d\Phi \int_0^{\pi} d\theta \sin\theta \, \theta^2 e^{\left(\frac{E|}{2k_B T}\right)\theta^2}$$

$$Z = \int_0^{2\pi} d\Phi \int_0^{\pi} d\theta \sin\theta \, e^{\left(\frac{E|}{2k_B T}\right)\theta^2}$$

We apply the same trickas before:

$$\langle \theta(s)^2 \rangle = \frac{1}{Z} \frac{2k_B T s}{|} \frac{\partial Z}{\partial E}$$

Thus, the only integral we really need to calculate is:

$$Z = \int_0^{2\pi} d\Phi \int_0^{\pi} d\theta \, \sin\theta \, e^{\left(\frac{E\!\!\mid}{2k_BT}\right)\theta^2} \approx \int_0^{2\pi} d\Phi \int_0^{\pi} d\theta\, \theta \, e^{\left(\frac{E\!\!\mid}{2k_BT}\right)\theta^2}$$

Where we used the Taylor expansion for $\sin\theta \approx \theta$

$$Z = \frac{2\pi k_B T s}{E\!\!\mid} \int_0^{\infty} du \, e^{-u} = \frac{2\pi k_B T s}{E\!\!\mid} \frac{1}{Z}\frac{\partial Z}{\partial E} = -\frac{1}{E}$$

$$\langle \theta(s)^2 \rangle = \frac{2k_B T s}{E\!\!\mid} \quad \langle g(s) \rangle \approx 1 - \frac{k_B T}{E\!\!\mid} s = 1 - \frac{s}{\xi_P} \quad \xi_P = \frac{E\!\!\mid}{k_B T} = \frac{k_{off}}{k_B T}$$

The Worm – Like Chain Model

Contains both the entropic and elastic bending contribution to free energy

(A)
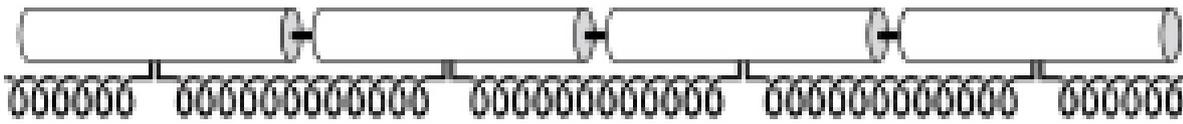
(B)
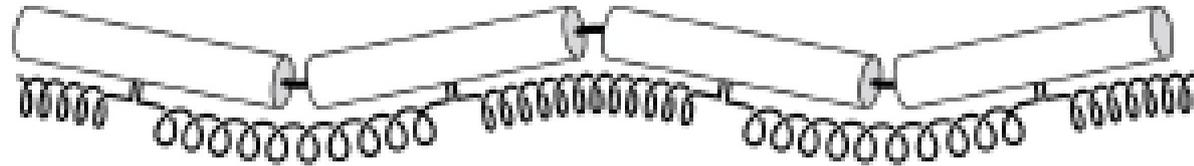
Figure 10.36. Scheme of concept that chains are cylinders connected by flexible links. A – no bending (optimal E, non - optimal S); B – bending (non - optimal E, optimal S).

$$Z = \int Dt(s) \exp\left(-\frac{\xi_P}{2}\int_0^L \left|\frac{dt}{ds}\right|^2 ds\right)$$

Sum over all possible $t(s)$ curves of length L: Feynman Path Integral

To calculate the force-extension curve, we need to include the work done by the external force F (along the z-direction), which adds:

$$-F\int_0^L t_z ds$$

Such that an average change extension can be calculated as:

$$\langle z \rangle = \frac{1}{Z(f)} \int Dt(s) z \exp\left(-\frac{\xi_P}{2} \int_0^L \left|\frac{dt(s)}{ds}\right|^2 ds + \int_0^L t_z ds\right)$$

$Z(f) \ldots$ partition function in the presence of the force $F = k_B T f$ (f has a unit of inverse length)

Again, the same trick can be used to avoid a calculation of two path integrals:

$$\langle z \rangle = \frac{d\ln Z(f)}{df} \rightarrow approx. \, solution: \; f\xi_P = \frac{Z}{L} + \frac{1}{4\left(\frac{1-z}{L}\right)^2} - \frac{1}{4}$$
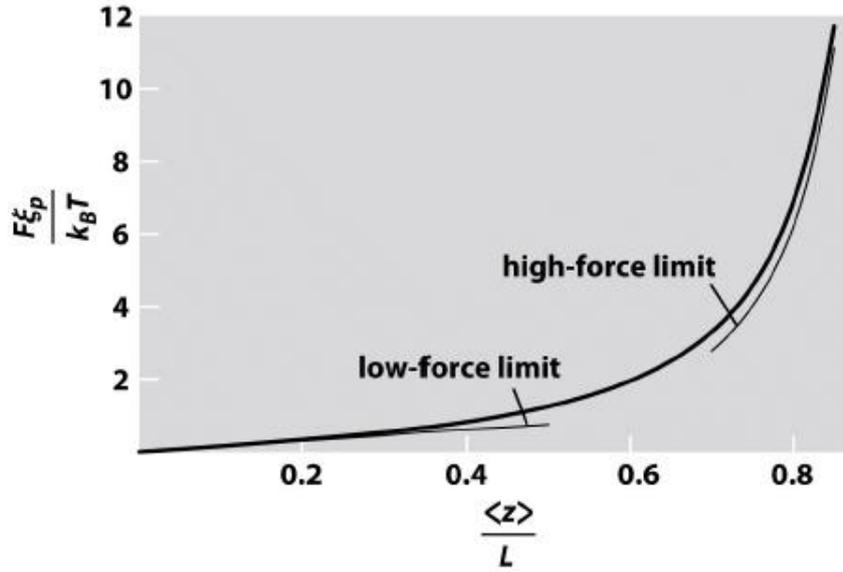
Figure 10.37. Force extension curve form a Warm-Like Chain model.

Energetics of DNA looping

Bending energy of a loop with a radius R was calculated already:

$$E_{loop} = \frac{\xi_P \pi k_B T}{R}$$

For a DNA loop of length L, we take into account the following:

$$L = 2\pi R \quad \rightarrow \quad R = \frac{L}{2\pi} = \frac{\delta N_{bp}}{2\pi} \qquad where \;\; \delta = 0,34 \, nm$$

To obtain a more convenient bending energy of a DNA loop:

$$\frac{E_{loop}}{k_B T} = \frac{2\pi^2}{N_{bp}} \left(\frac{\xi_P}{\delta}\right) \approx \frac{3000}{N_{bp}}$$

Expressed in terms of the number of base pairs. We used the estimate for the persistence length:

$$\xi_P = 50 \, nm$$
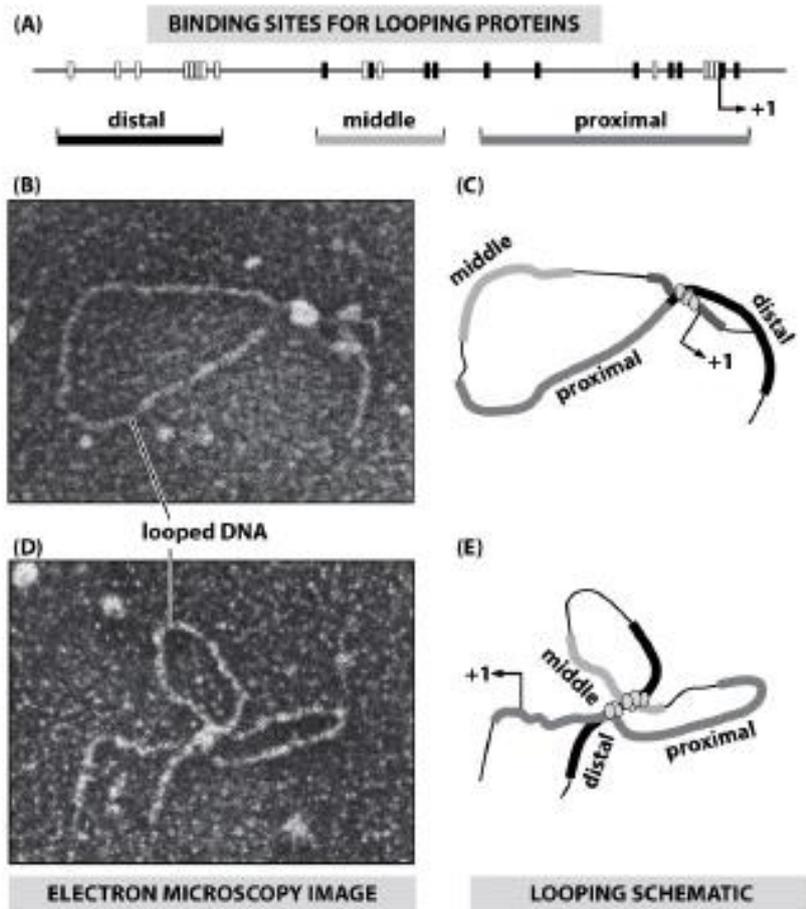
Figure 10.38. Scheme of concept that chains are cylinders connected by flexible links. A – no bending (optimal E, non optimal S); B – bending (non optimal E, optimal S).


Figure 10.39. The bending energy of DNA loops as a function of the number of base pairs in the DNA.

Free Energy Estimate for DNA Looping: Elastic Energy Combined With Entropic Contribution

→ Long DNA fragments: entropic contribution unfavorable

→ short DNA fragment: bending contribution unfavorable

→ elastic and entropic contributions:

$$\Delta E_{loop} = \frac{3000 k_B T}{N_{bp}}$$

$$p_0 \, ^{\circ}\text{C} \frac{1}{\sqrt{N_{bp}^3}} \rightarrow \Delta S_{loop} = k_B \ln p_0 = -k_B \left(\frac{3}{2} \ln N_{bp} + const\right)$$

→ result in the total free energy change upon DNA looping:

$$\Delta G_{loop} = \Delta E_{loop} - TS_{loop} \approx k_B T \left(\frac{3000}{N_{bp}} + \frac{3}{2} \ln N_{bp} + const\right)$$

Cyclization experiments report a quantity J [in units of Concentration]:

$$J \, ^{\circ}\text{C} e^{-\beta \Delta G_{loop}}$$

So that the minimum of the free energy of looping is equivalent to the maximum of J.
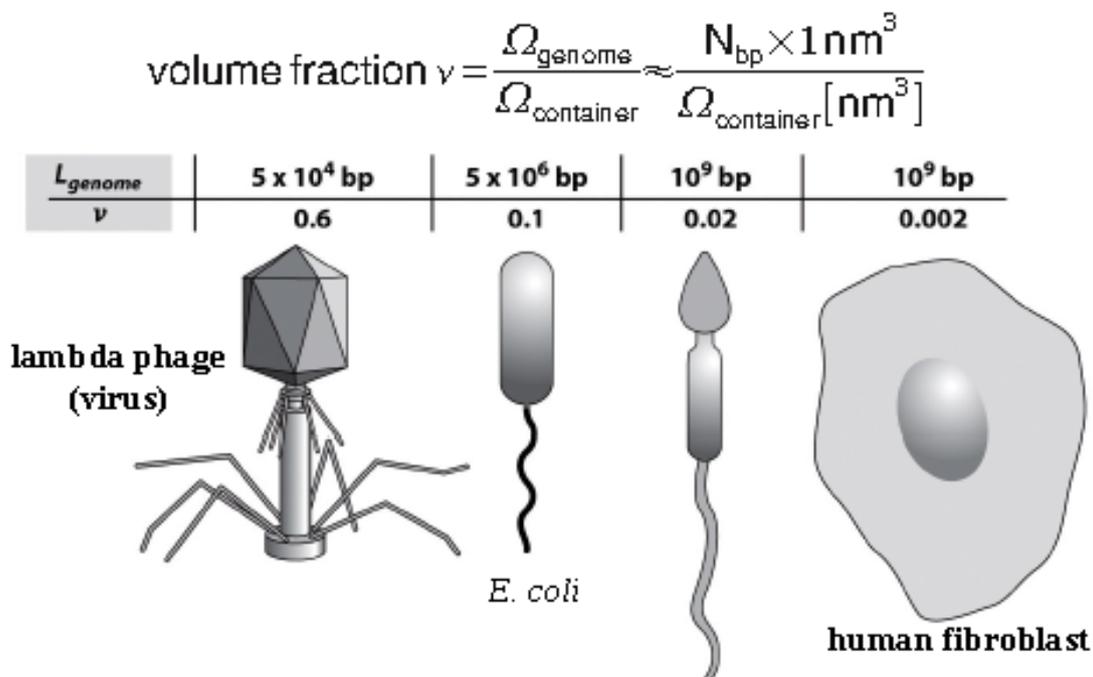
Figure 10.40. DNA packing – form viruses to eukaryotes.

The DNA Packing Compaction ratio

a) Lambda phage with spherical capsid of radius of 27 nm holding 48,500 base pairs:

$$\nu = \frac{N_{bp} nm^3}{\frac{4}{3}\pi R^3 nm^3} \approx \frac{5 \times 10^4}{4 \times 27^3} \approx 0,6$$

b) Bacterial nucleoid: a sphere of radius 0,25 μm holding 5×10^6 base pairs:

$$\nu \approx \frac{5 \times 10^6}{4 \times 250^3} \approx 0,1$$

c) human sperm cell with a spherical nucleon of radius of 0,25 μm holding 10^9 base pairs:

$$\nu \approx \frac{10^9}{4 \times 2500^3} \approx 0,02$$

d) human fibroblast nucleus of radius of 5 μm holding 10^9 base pairs:

$$\nu \approx \frac{10^9}{4 \times 5000^3} \approx 0,002$$
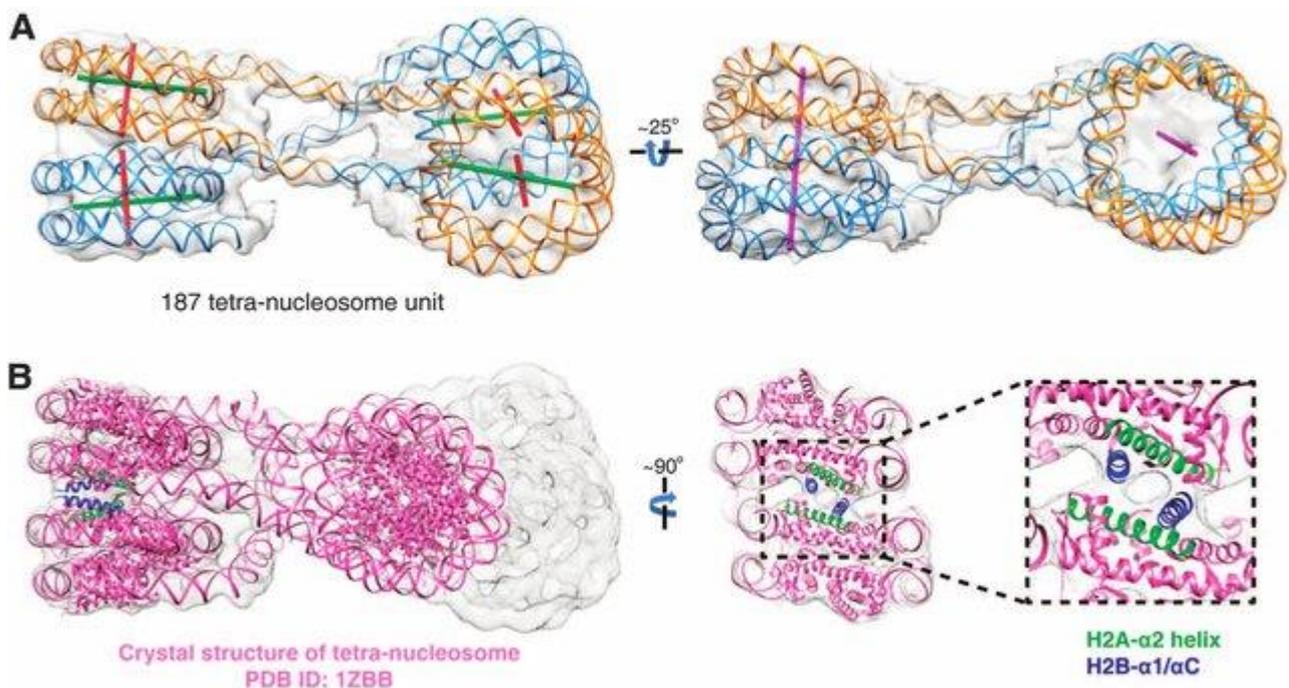
Figure 10.41. Cryo-electron microscopy images of packed DNA. Fig. 2. The structure of a tetranucleosomal unit with the 187-bp NRL. (A) The segmented density map for the tetranucleosomal unit in the 30-nm chromatin fibers reconstituted on 12 × 187 bp DNA, shown with the atomic structure of DNA from a docked mononucleosome crystal structure (PDB 1AOI) and modeled presumptive linker DNA and viewed from two angles. Different axes are highlighted by colors, including nucleosome core dyad (green), nucleosome superhelix axes (red), and stack axes (pink). (B) A comparison of the 3D cryo-EM map (gray) with the x-ray structure (PDB 1ZBB, pink) of the

tetranucleosome (16). The strong density where the adjacent H2A-H2B dimer meets is magnified and highlighted in the interface between the nucleosome cores within each stack. Source: Feng Song, Ping Chen, Guohong Li. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranuc leosomal Units. April 2014Science 344(6182):376-80 F.
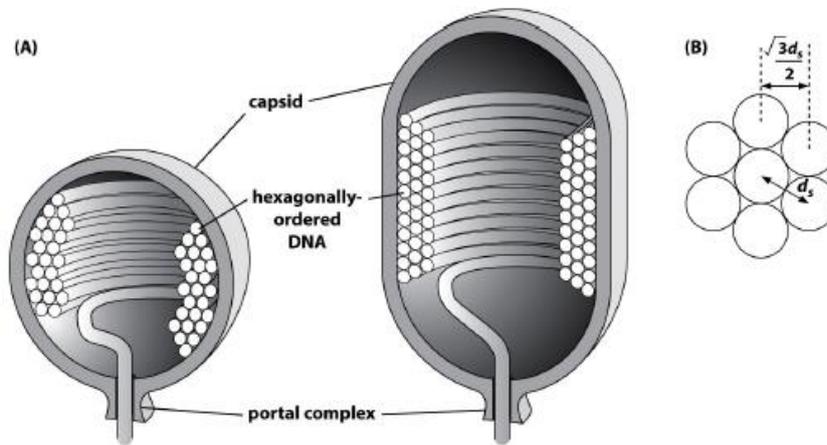

Figure 10.42. Capsid volume: $V_{capsid} = L_{dna}DNA\ d^2 \quad \rightarrow d_s\ 1/\ L_{DNA}^{1/2}$

DNA packing in viruses requeres energy

→ the entropic spring effect (DNA tends to spread out)

→ elastic bending on the scale smaller than the persistence length of 50 nm

→ strongly negatively charged DNA (tends to avoid itself)

→ the entropic contribution is about 10-tim smaller than the bending and electrostatic free energy associated with DNA packing (in viruses):

$$G_{tot}(d_s, L) \approx G_{bend}(d_s, L) + G_{charge}(d_s, L)$$

→ the force that resists the packing is then:

$$F = \frac{-dG_{tot}(d_s, L)}{dL}$$

Elastic Bending Contribution to the Free Energy:

Circular Hoops:

$$E_{boop} = \frac{\pi E|}{R} = \frac{\pi \xi_P k_B T}{R} \rightarrow G_{bend} = \pi \xi_P k_B T \sum_i \frac{N(R_i)}{R_i}$$

$$\sum_i \rightarrow \frac{2}{\sqrt{3}d_3} \int dR' \frac{\sqrt{3}}{2} d_3 \dots distance\ between\ DNA\ strans$$

$$G_{bend} = \frac{2\pi\xi_P k_B T}{\sqrt{3}d_3} \int_R^{R_{out}} \frac{N(R')}{R'} dR'$$

$$DNA\ length:\ L = \frac{2}{\sqrt{3}d_3} \int_R^{R_{out}} 2\pi R' N(R')\, dR'$$

Cylindrical capsid of height z, radius R_{out}:

$$G_{bend}(R) = \frac{2\pi\xi_P k_B T z}{\sqrt{3}d_s^2} \ln\left(\frac{R_{out}}{R}\right) \quad L(R) = \frac{2\pi z}{\sqrt{3}d_s^2}(R_{out}^2 - R^2)$$

The second equation for L(R) can be used to express R:

$$R = R_{out}\sqrt{1 - \frac{\sqrt{3}d_s^2 L}{2\pi z R_{out}^2}}$$

And the final expression for the free energy of bending:

$$G_{bend} = \frac{-\pi\xi_P k_B T z}{\sqrt{3}d_s^2} \ln\left(1 - \frac{\sqrt{3}d_s^2 L}{2\pi z R_{out}^2}\right)$$

And the force associated with ac cumulated bending energy:

$$f(L) = \frac{-dG_{bend}}{dL} = \frac{-\dfrac{\xi_P k_B T}{2R_{out}^2}}{1 - \dfrac{\sqrt{3}d_s^2 L}{2\pi z R_{out}^2}}$$

polyethylene glycol

DNA rods
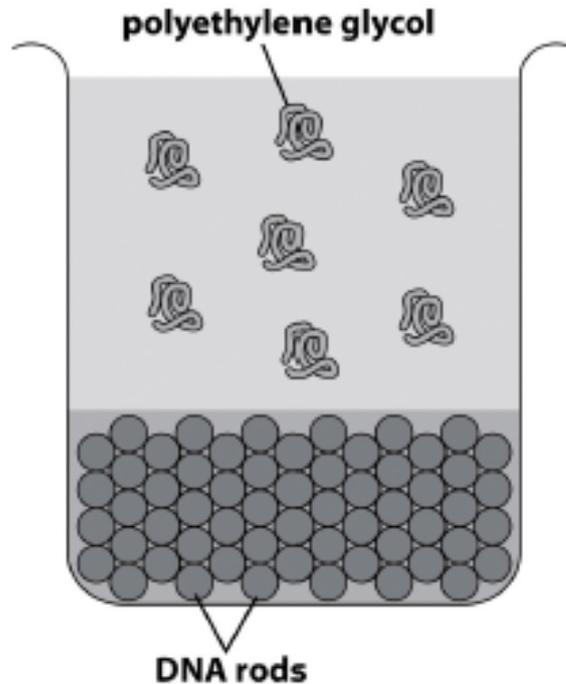
Figure 10.43. Electrostatic contribution to the free energy of DNA packing in viruses.

Consider N parallel strands of length I, each packed into a hexagonal array with a spacing d_s:

$$G_{bend} = 3N|v(d_s)$$

(a factor of $3 = \frac{1}{2} \times 6$ for 6 nearest neighbors in the array)

The volume of the assembly is:

$$V = \frac{\sqrt{3}}{2}Nd_s^2|$$

The pressure can be calculated as:

$$p(d_s) = \frac{-dG_{charge}}{dV}, \quad dV = N|\sqrt{3}d_s dd_s$$

So that the force is:

$$f(d_s) = \frac{dv(d_s)}{dd_s} = \frac{1}{\sqrt{3}}p(d_s)d_s$$

Which leads us to the final expressions:

$$v(d_s) = \frac{1}{\sqrt{3}}F_0(c^2 + cd_s)e^{\frac{-d_s}{c}}$$

$$G_{charge} = \sqrt{3}F_0(c^2 + cd_s)Le^{\frac{-d_s}{c}}, \quad L = N|\ldots \text{ total DNA length}$$

This electrostatic contribution can be experimentally controlled Through ionic concentration, which affects the energy through F_0 whereas $c \approx 0{,}27\ nm$ is more or less constant over a wide range of salt conditions.

The total energy associated with DNA packing in viruses is:

$$G_{tot} = G_{bend} + G_{charge}$$

$$G_{tot} = \frac{-\pi\xi_P k_B Tz}{\sqrt{3}d_s^2}\ln\left(1 - \frac{\sqrt{3}d_s^2 L}{2\pi z R_{out}^2}\right) + G_{charge} = \sqrt{3}F_0(c^2 + cd_s)Le^{\frac{-d_s}{c}}$$

Chapter 11

Biomechanical Systems and Muscles

The nervous system 'communicates' with muscle via neuromuscular (also called myoneural) junctions. These junctions work very much like a synapse between neurons In other words: the impulse arrives at the end bulb, chemical transmitter is released from vesicles (each of which contains 5,000 - 10,000 molecules of acetylcholine) and diffuses across the neuromuscular cleft, the transmitter molecules fill receptor sites in the membrane of the muscle & increase membrane permeability to sodium, it then diffuses in & the membrane potential becomes less negative, and, if the threshold potential is reached, an action potential occurs, an impulse travels along the muscle cell membrane, and the muscle contracts. Some muscles (skeletal muscles) will not contract unless stimulated by neurons; other muscles (smooth & cardiac) will contract without nervous stimulation but their contraction can be influenced by the nervous system. Thus, the nervous and muscle systems are closely interconnected.

Muscles, their structure and working mechanisms
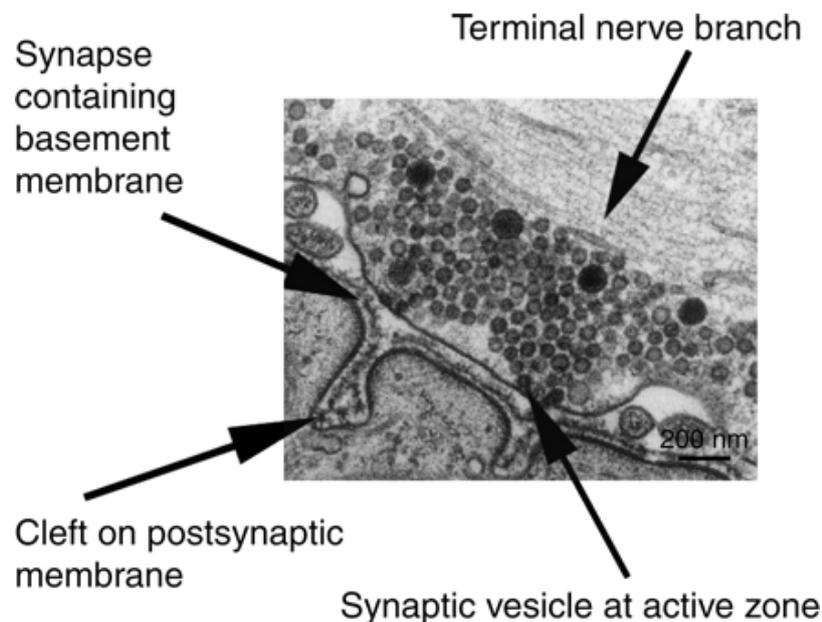
Fig.11.1. Essential structure of a Muscle.

Characteristics of muscle:

- excitability - responds to stimuli (e.g., nervous impulses)

- contractility - able to shorten in length

- extensibility - stretches when pulled

- elasticity - tends to return to original shape & length after contraction or extension

Functions of muscle:

- motion

- maintenance of posture

- heat production

Types of muscle:

- skeletal:

    o   attached to bones & moves skeleton

    o   striated muscle

    o   voluntary muscle

- smooth

    o   involuntary muscle

    o   muscle of the viscera (e.g., in walls of blood vessels, intestine, & other 'hollow' structures and organs in the body)

- cardiac:

    o   muscle of the heart

    o   involuntary
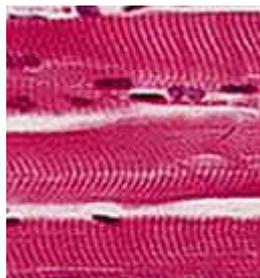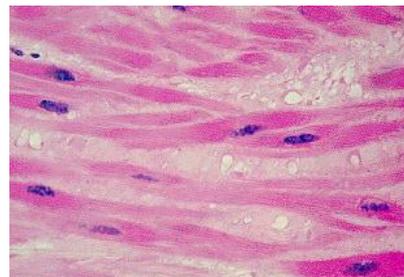
a   b

Fig.11.2. Electron microscope photographs of a – striated and b – smooth muscles.

Structure of Skeletal Muscles

Skeletal muscles are usually attached to bone by tendons composed of connective tissue. This connective tissue also sheaths the entire muscle & is called epimysium. Skeletal muscles consist of numerous subunits or bundles called fasicles (or fascicles). Fascicles are also surrounded by connective tissue (called the perimysium) and each fascicle is composed of numerous muscle fibers (or muscle cells). Muscle cells,

ensheathed by endomysium, consist of many fibrils (or myofibrils), and these myofibrils are made up of long protein molecules called myofilaments. There are two types of myofilaments in myofibrils: thick myofilaments and thin myofilaments.
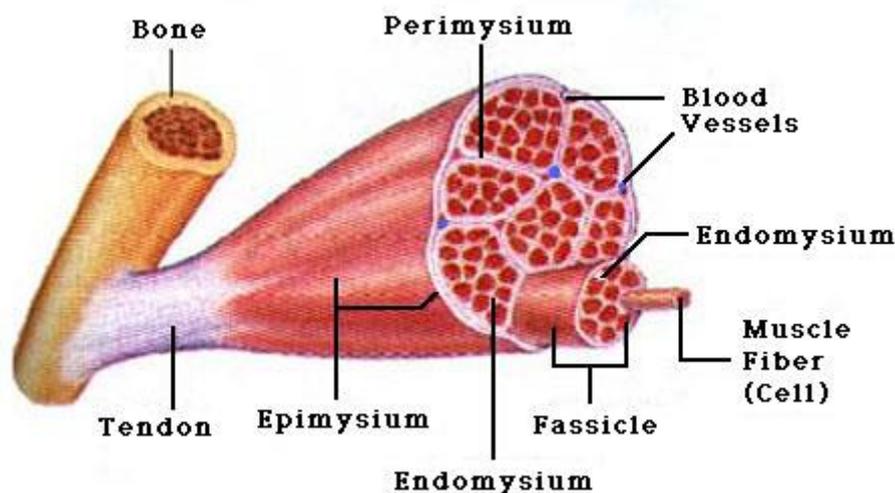

Fig.11.3. Skeletal muscle structure.

Skeletal muscles vary considerably in size, shape, and arrangement of fibers. They range from extremely tiny strands such as the stapedium muscle of the middle ear to large masses such as the muscles of the thigh. Skeletal muscles may be made up of hundreds, or even thousands, of muscle fibers bundled together and wrapped in a connective tissue covering. Each muscle is surrounded by a connective tissue sheath called the epimysium. Fascia, connective tissue outside the epimysium, surrounds and separates the muscles. Portions of the epimysium project inward to divide the muscle into compartments. Each compartment contains a bundle of muscle fibers. Each bundle of muscle fiber is called a fasciculus and is surrounded by a layer of connective tissue called the perimysium. Within the fasciculus, each individual muscle cell, called a muscle fiber, is surrounded by connective tissue called the endomysium. Skeletal muscles have an abundant supply of blood vessels and nerves. Before a skeletal muscle fiber can contract, it has to receive an impulse from a neuron. Generally, an artery and at least one vein accompany each nerve that penetrates the epimysium of a skeletal muscle. Branches of the nerve and blood vessels follow the connective tissue components of the muscle of a nerve cell and with one or more minute blood vessels called capillaries. When muscle cells are viewed under the microscope, one can see that they contain a striped pattern (striations). This pattern is formed by a series of basic units called sarcomeres that are arranged in a stacked pattern throughout muscle tissue. There can be thousands of sarcomeres within a single muscle cell. Sarcomeres are highly stereotyped and are repeated throughout muscle cells, and the proteins within them can change in length, which causes the overall length of a muscle to change. An individual sarcomere contains many parallel actin (thin) and myosin (thick) filaments. The interaction of myosin and actin proteins is at the core of our current understanding of sarcomere shortening. How does this shortening happen? It has something to do with a sliding interaction between actin and myosin.

161
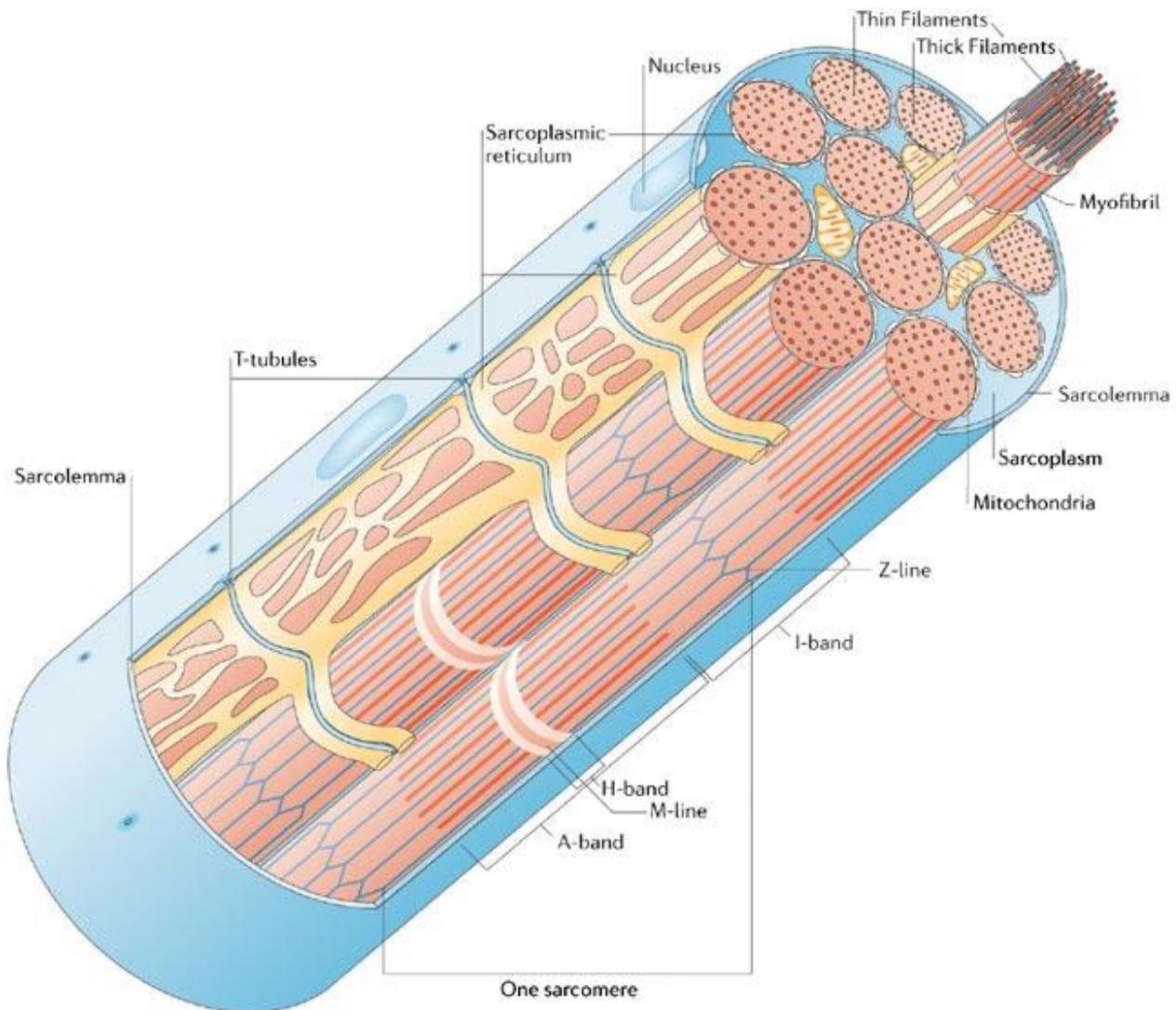
Fig.11.4.   Schematic   structure of a skeletal muscle. Source: Molecular cell biology, 2006.

Skeletal muscle is the muscle attached to the skeleton. Hundreds or thousands of muscle fibers (cells) bundle together to make up an individual skeletal muscle. Muscle cells are long, cylindrical structures that are bound by a plasma membrane (the sarcolemma) and an overlying basal lamina and when grouped into bundles (fascicles) they make up muscle. The sarcolemma forms a physical barrier against the external environment and also mediates signals between the exterior and the muscle cell. The sarcoplasm is the specialized cytoplasm of a muscle cell that contains the usual subcellular elements along with the Golgi apparatus, abundant myofibrils, a modified endoplasmic reticulum known as the sarcoplasmic reticulum (SR), myoglobin and mitochondria. Transverse (T)-tubules invaginate the sarcolemma, allowing impulses to penetrate the cell and activate the SR. As shown in the figure, the SR forms a network around the myofibrils, storing and providing the Ca^{2+} that is required for muscle contraction. Myofibrils are contractile units that consist of an ordered arrangement of longitudinal myofilaments. Myofilaments can be either thick filaments (comprised of myosin) or thin filaments (comprised primarily of actin). The characteristic

'striations' of skeletal and cardiac muscle are readily observable by light microscopy as alternating light and dark bands on longitudinal sections. The light band, (known as the I-band) is made up of thin filaments, whereas the dark band (known as the A-band) is made up of thick filaments. The Z-line (also known as the Z-disk or Z-band) defines the lateral boundary of each sarcomeric unit. Contraction of the sarcomere occurs when the Z-lines move closer together, making the myofibrils contract, and therefore the whole muscle cell and then the entire muscle contracts.The SARCOLEMMA has a unique feature: it has holes in it. These "holes" lead into tubes called Transverse Tubules or T Tubules for short. These tubules pass down into the muscle cell and go around the MYOFIBRILS. However, these tubules DO NOT open into the interior of the muscle cell; they pass completely through and open somewhere else on the sarcolemma (i.e., these tubules are not used to get things into and out of the muscle cell). The function of T-TUBULES is to conduct impulses from the surface of the cell (SARCOLEMMA) down into the cell and, specifically, to another structure in the cell called the SARCOPLASMIC RETICULUM.
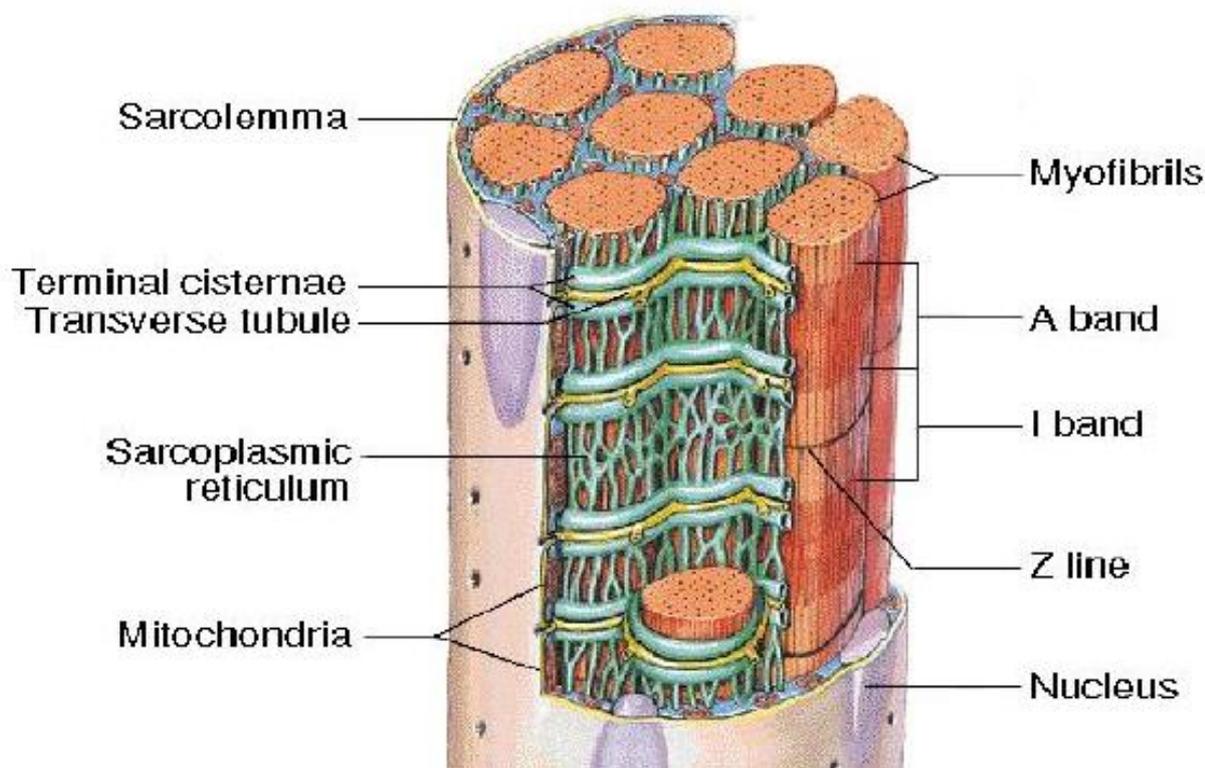

Fig.11.5. Main working parts of a muscle.

The SARCOPLASMIC RETICULUM (SR) is a bit like the endoplasmic reticulum of other cells, e.g., it's hollow. But the primary function of the SARCOPLASMIC RETICULUM is to STORE CALCIUM IONS. Sarcoplasmic reticulum is very abundant in skeletal muscle cells and is closely associated with the MYOFIBRILS (and, therefore, the MYOFILAMENTS). The membrane of the SR is well-equipped to handle calcium: there are "pumps" (active transport) for calcium so that calcium is constantly being "pumped" into the SR from the cytoplasm of the muscle cell (called the SARCOPLASM). As a result, in a

relaxed muscle, there is a very high concentration of calcium in the SR and a very low concentration in the sarcoplasm (and, therefore, among the myofibrils & myofilaments). In addition, the membrane has special openings, or "gates", for calcium. In a relaxed muscle, these gates are closed and calcium cannot pass through the membrane. So, the calcium remains in the SR. However, if an impulse travels along the membrane of the SR, the calcium "gates" open &, therefore, calcium diffuses rapidly out of the SR & into the sarcoplasm where the myofibrils & myofilaments are located. This, as you will see, is a key step in muscle contraction. Sarcoplasmic reticulum (SR) membranes are in close proximity to a T-tubule. 'RyR' are proteins the aid in the release of calcium from the SR, 'SERCA2' are proteins that aid in the transport of calcium into the SR.
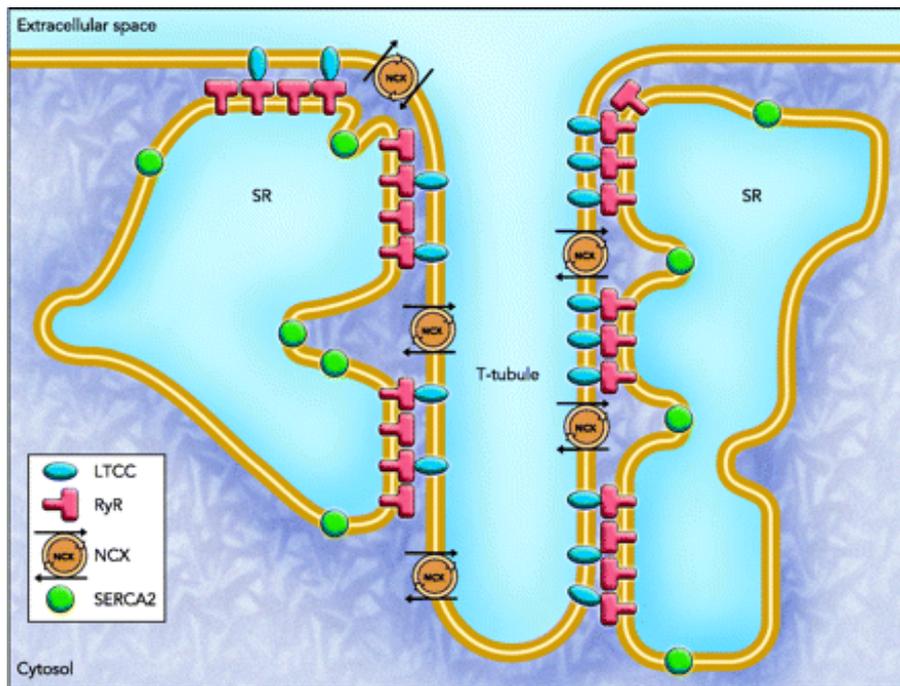
Fig.11.6. Scheme of Sarcoplasmic membranes framework.
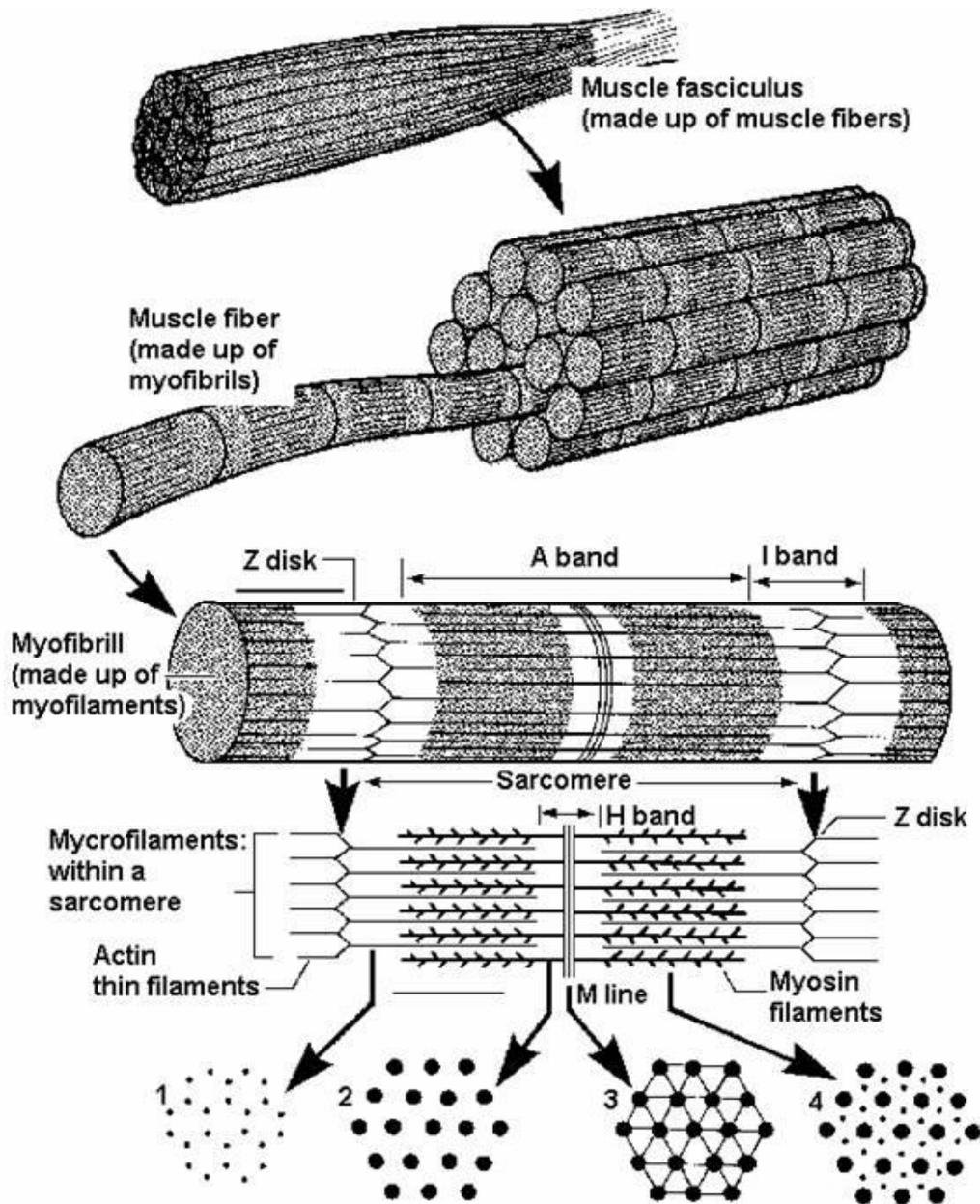
Fig.11.7. Structure of muscle fibers.

Each myofibril is composed of many subunits lined up end-to-end. These subunits are, of course, composed of myofilaments and are called SARCOMERES. The drawings above & below show just a very small section of the entire length of a myofibril and so you can only see one complete SARCOMERE.
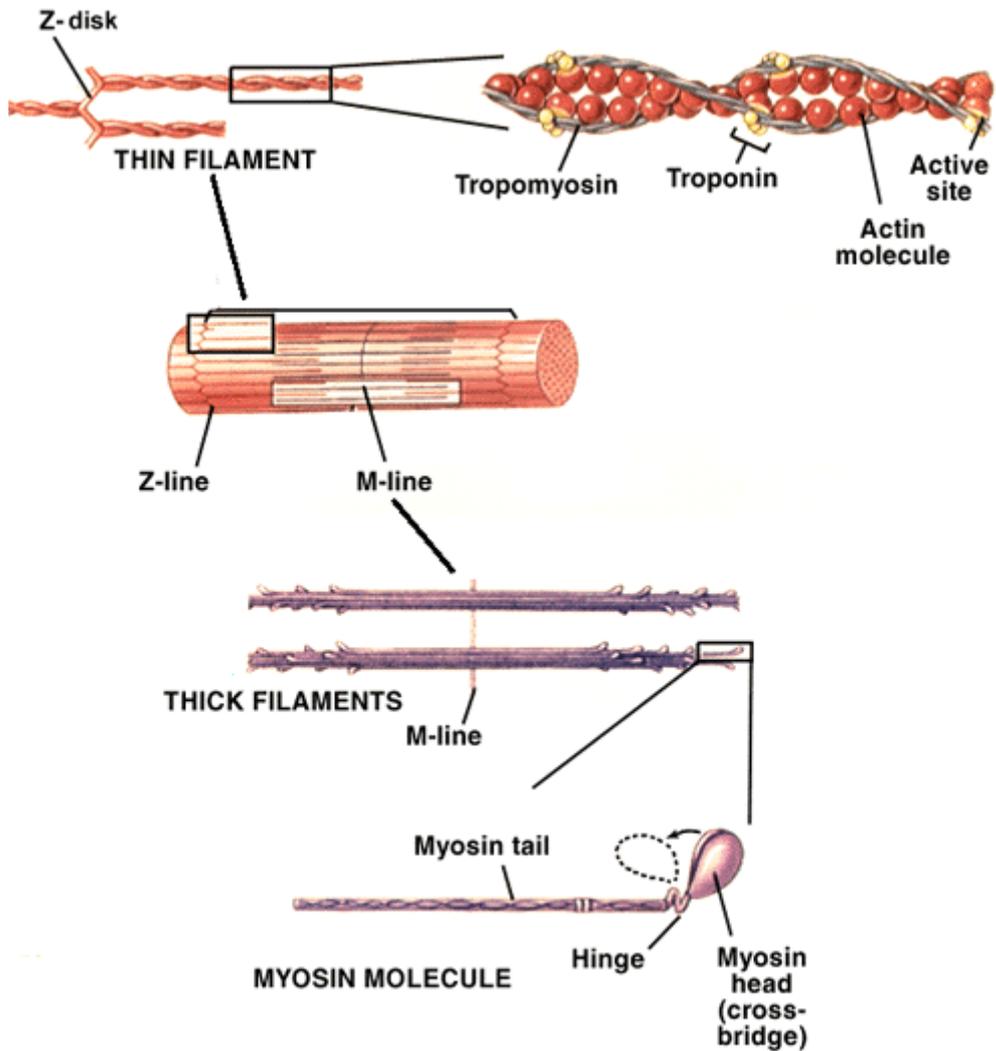
Fig.11.8. Muscle filaments and their interconnections.
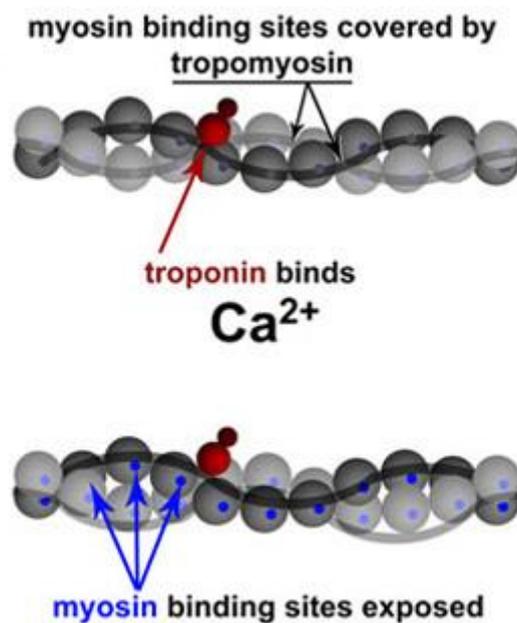
Fig.11.9. Troponin and tropomyosin regulate contraction via calcium binding

In each sarcomere, thin myofilaments extend in from each end. Thick myofilaments are found in the middle of the sarcomere and do not extend to the ends. Because of this arrangement, when skeletal muscle is viewed with a microscope, the ends of a sarcomere (where only thin myofilaments are found) appear lighter than the central section (which is dark because of the presence of the thick myofilaments). Thus, a myofibril has alternating light and dark areas because each consists of many sarcomeres lined up end-to-end. This is why skeletal muscle is called STRIATED MUSCLE (i.e., the alternating light and dark areas look like stripes or striations). The light areas are called the I-BANDS and the darker areas the A-BANDS. Near the center of each I-BAND is a thin dark line called the Z-LINE (or Z-membrane in the drawing below). The Z-LINE is where adjacent sarcomeres come together and the thin myofilaments of adjacent sarcomeres overlap slightly. Thus, a sarcomere can be defined as the area between Z-lines.

Thick myofilaments are composed of a protein called MYOSIN. Each MYOSIN molecule has a tail which forms the core of the thick myofilament plus a head that projects out from the core of the filament. These MYOSIN heads are also commonly referred to as CROSS-BRIDGES.

The MYOSIN HEAD has several important characteristics:

- it has ATP-binding sites into which fit molecules of ATP. ATP represents potential energy.

- it has ACTIN-binding sites into which fit molecules of ACTIN. Actin is part of the thin myofilament and will be discussed in more detail shortly.

- it has a "hinge" at the point where it leaves the core of the thick myofilament. This allows the head to swivel back and forth, and the "swivelling" is, as will be described shortly, what actually causes muscle contraction.
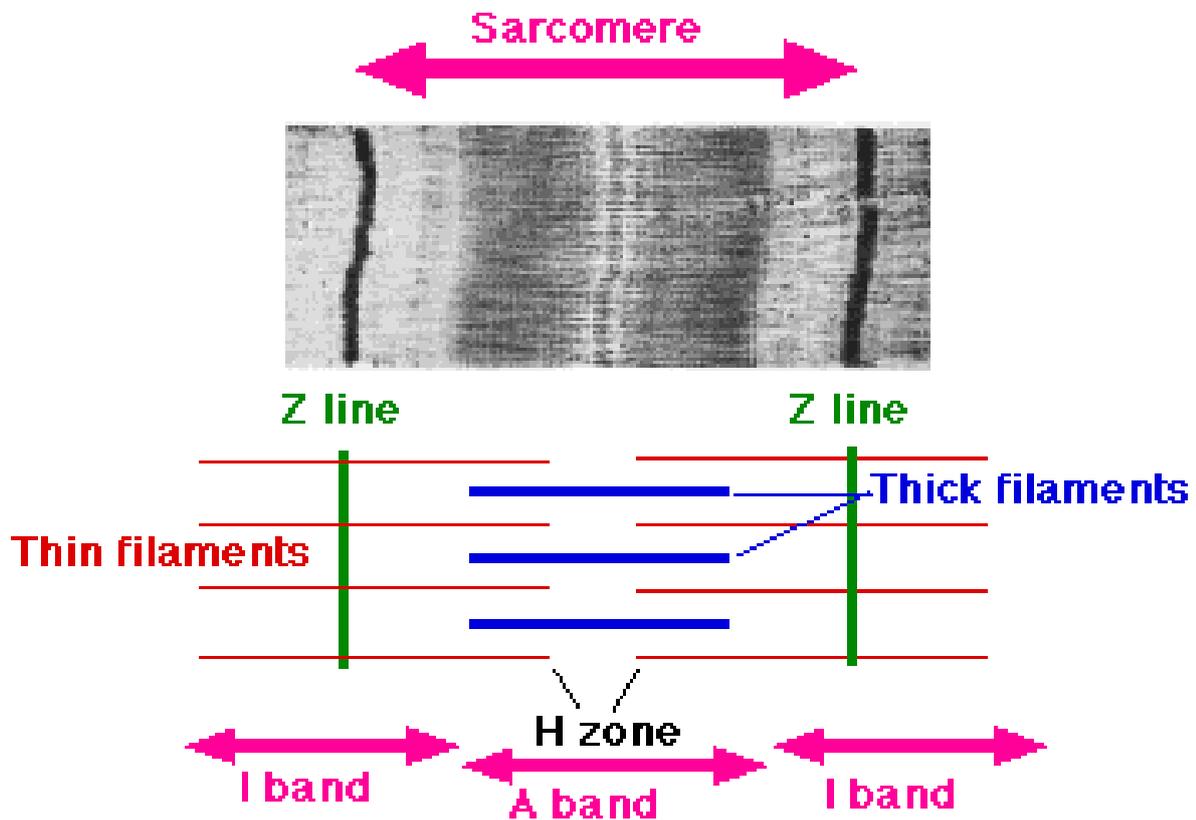
Fig.11.10. Distribution of sarcomere and filaments in a Muscle.

Thin myofilaments are composed of 3 types of protein: ACTIN, TROPONIN, and TROPOMYOSIN.
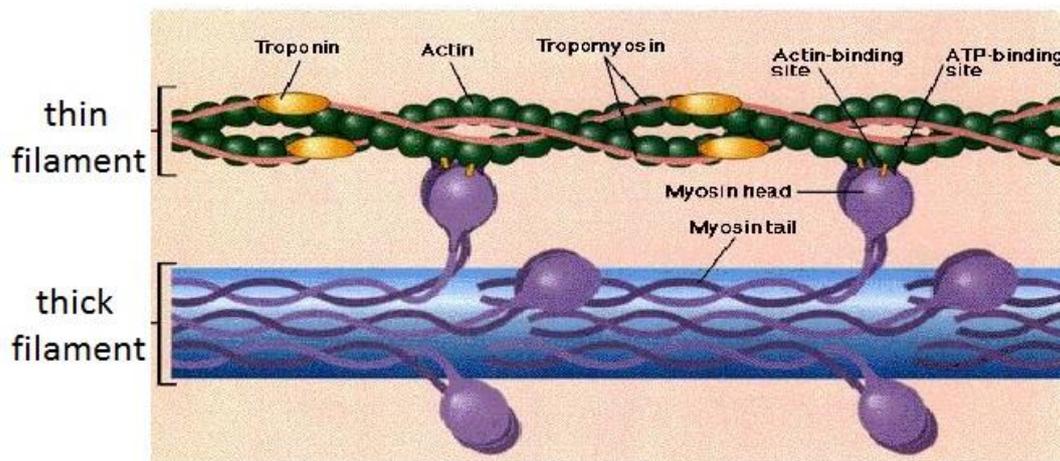
Fig.11.11. Structure of thick and thin filaments.

The actin molecules (or G-actin as above) are spherical and form long chains. Each thin myofilament contains two such chains that coil around each other. TROPOMYOSIN molecules are lone, thin molecules that wrap around the chain of ACTIN. At the end of each tropomyosin is an TROPONIN molecule. The TROPOMYOSIN and TROPONIN molecules are connected to each other.

- ACTIN - when actin combines with MYOSIN HEAD the ATP associated with the head breaks down into ADP. This reaction released energy that causes the MYOSIN HEAD to SWIVEL.

- TROPOMYOSIN - In a relaxed muscle, the MYOSIN HEADS of the thick myofilament lie against TROPOMYOSIN molecules of the thin myofilament. As long as the MYOSIN HEADS remain in contact with TROPOMYOSIN nothing happens (i.e., a muscle remains relaxed).

- TROPONIN - Troponin molecules have binding sites for calcium ions. When a calcium ion fills this site it causes a change in the shape and position of TROPONIN. And, when TROPONIN shifts, it pulls the TROPOMYOSIN to which it is attached. When TROPOMYOSIN is moved, the MYOSIN HEAD that was touching the tropomyosin now comes in contact with an underlying ACTIN molecule.

And finally, the standard model of muscle contraction includes only two main players, the filaments actin and myosin, which slide past each other, causing the sarcomere - the basic unit of muscle tissue - to shorten. Driven by ATP-powered myosin motors, sarcomere contraction causes whole muscles to shorten. In the past 40 years, researchers have explored the role of another filament: titin. Its emerging role in muscle function, unwinding during muscle relaxation (above) and folding during contraction (below), suggests that the ATP-driven motors also act as latches allowing titin to fold, providing a powerful boost to muscle contraction. One of the most important observation is that proteins can readily fold against a pulling force, delivering a large amount of mechanical work that surpasses that generated by ATP-fueled motors. From this, it is evident how titin might operate in intact muscle tissues. Titin modules are arranged in tandem and can number up to 100 in the elastic I band region, which spans sarcomeres. Passive stretching, which occurs when muscle is relaxed and elongated during activities such as yoga, will unfold and extend titin modules under a wide range of forces and time scales. Holding a yoga pose for long periods of time results in storage of large amounts of potential energy in the stretched muscle through titin unfolding. It is now clear that protein unfolding and folding under force is prevalent in biology and plays crucial roles in processes from protein translation to protein degradation and most things in between. By contrast, titin folding occurs over only a few pico-newtons in the physiological force range. Strikingly, this range matches the range of forces produced by active ATP-driven myosin motors. Given how these molecules are organized in the sarcomere, it seems likely that the activation of the ATP-driven motors relieves the force on titin, triggering spontaneous titin folding. This partnership suggests that the motors act as release latches for the elastic energy stored in titin during the human motion. Our most recent data show that in the physiological force range, the power output of disulfide-bonded titin domains matches the power output of the myosin thick filament, suggesting that these two sources of energy combine to deliver muscle power.
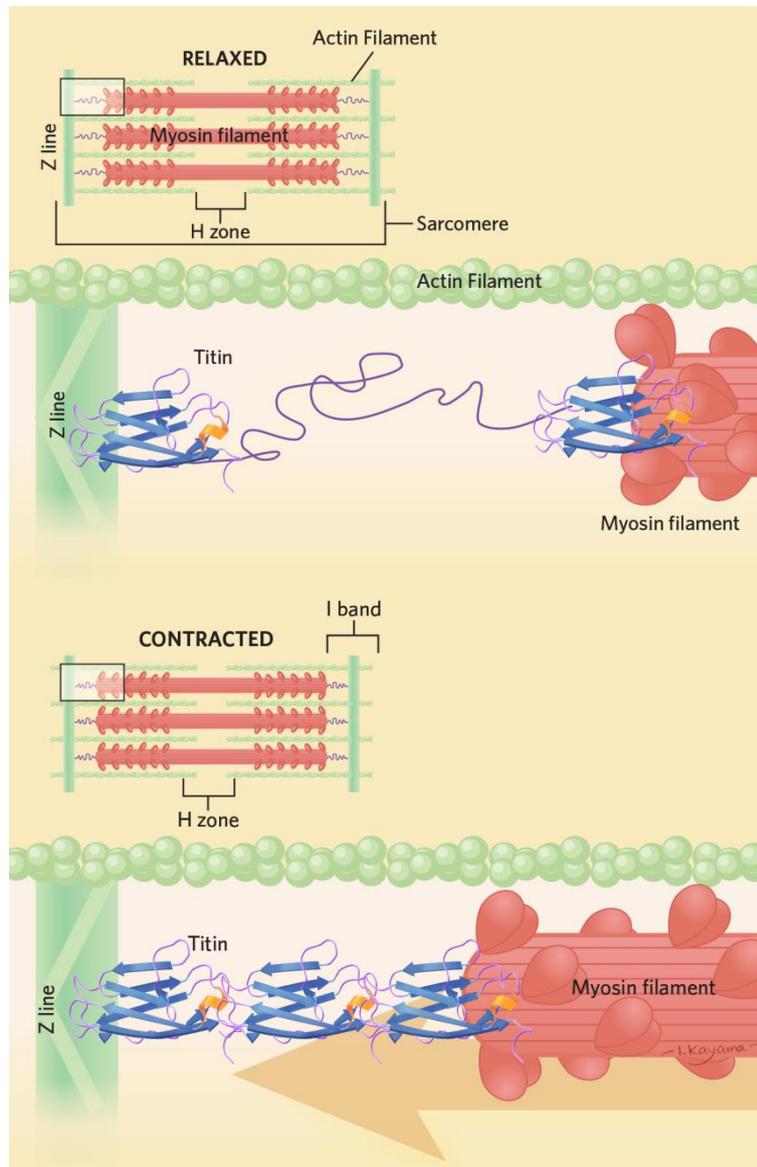
Fig.11.12. Scheme of interaction of actin, titin and myosin in muscle filaments.

Chapter 12

Mesoscopic Forces

The reader should be familiar with some simple manifestations of the fundamental forces that drive the interactions between matter, such as electro statics, gravity and magnetism. However, nature has used a subtle mixture of these forces in combination with geometric and dynamic effects to determine the interactions of biological molecules. The mesoscopic forces are not fundamental, but separation into the different contributions of the elementary components would be very time consuming and require extensive molecular dynamic simulations. Therefore, in this lecture a whole series of simple models for mesoscopic forces is introduced and some generic methods to measure the forces experimentally reviewed. There is a rich variety of mesoscopic forces that have been identified. These include Van der Waals, hydrogen bonding, electrostatics, depletion forces and hydrodynamic interactions.

Cohesive forces

The predominant force of cohesion between matter is the Van der Waals interaction. Objects made of the same material always attract each other due to induced dipoles. The strength of Van der Waals bonds is relatively weak, with energies of the order of 1 kJ mol, but the forces act between all types of atom and molecule (even neutral ones).

A fundamental definition of the Van der Waals interaction is an attractive force of quantum mechanical origin that operates between any two molecules, and arises from the interaction between oscillating dipoles. Without overburdening the description with the detailed quantum mechanics, the potential V_{12} (r), which gives rise to the dispersive Van der Waals force between molecules 1 and 2 can be defined as:

$$V_{12}(\gamma) = \frac{1}{24(\pi\varepsilon_0)^2\gamma^6} \sum_{n,k} \frac{|(n|\vec{m}|o)_1|^2|(k|\vec{m}|o)_2|^2}{(E_1^n E_1^o) + (E_2^k E_2^o)} \frac{A_{12}}{\gamma^6} \tag{12.1}$$

where nm, k mare the transition dipole moment from the quantized state to 0 for molecule 1,r is the distance between the two molecules, andA_{12} is a constant. $E^n_1$and $E^k_1$are the quantum energies of state n and k for molecules 1 and 2 respectively. There is thus a characteristic $1=r^6$ decay of the potential between point-like molecules, with a single characteristic constant of proportionality (A_{12}, the Hamaker constant) that is dependent of the variety of molecule considered. Van der Waals forces are sometimes called a dispersion interaction, because the same quantities determine both the optical properties of the molecules (dispersion of light) and the forces between them. It is therefore possible to observe the effects of Van der Waals forces optically with micron sized colloidal particles in solution. If a material contains permanent dipoles, they can induce temporary dipoles in another material giving rise to further Vander Waals type interactions (Keesom or Debye forces).

The analysis of Van der Waals forces tends to be more complicated in practice than many of the fundamental interactions that may have been encountered previously in foundation physics courses. Van der Waals forces are long range and can be effective from large distances (> 10 nm) down to interatomic spacings (< 0.1 nm). The forces may be repulsive or attractive, and crucially in general, they do not follow a simple power law, as is illustrated in Figure 12.1 for four separate possible geometries.

Van der Waals forces tend to both bring molecules together and mutually align or orientate them. Unlike gravitational and Coulomb forces, Vander Waals forces are not generally additive. At larger separations (> 10 nm) the effect of the finite speed of propagation (the speed of light, c) of the interaction also becomes important.

The force laws illustrated in Figure 12.1 can be proved by careful summation of the contributions in equation (12.1) over an extended body.

(a)
$$w = -\frac{C}{r^6}$$

(b)
$$w = \frac{-A}{6D}\frac{R_1 R_2}{(R_1 + R_2)}$$

(c)
$$w = -\frac{\pi C\rho}{6D^3}$$

(d)
$$w = -\frac{A}{12\pi D^2}$$
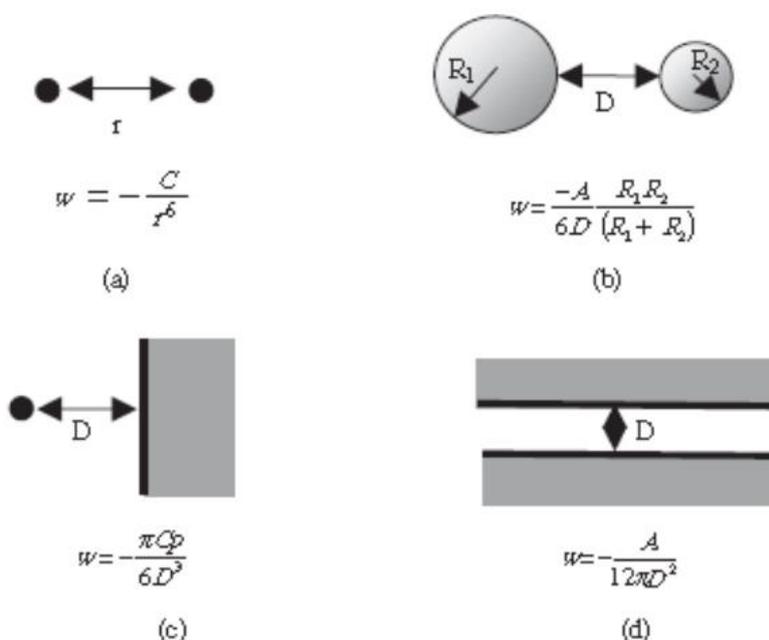
Figure 12.1 The leading term in the energy (w) of the Van der Waals interaction between surfaces depends on the geometry. (Four geometries are shown: (a) point atoms, (b) two spheres, (c) atom/plane, and (d)two plane surfaces.).

Hydrogen bonding

Water exhibits an unusually strong interaction between adjacent molecules, which persists into the solid state (Figure 12.2). This unusual interaction is given a special name, hydrogen bonding, and is an important effect in a wide range of hydrogenated polar molecules and determines their different molecular geometries (Figure 12.3), e.g. chain structures, crystals, bifurcated associations and intramolecular bonds.

Hydrogen bonds are typically stronger than Van der Waals forces and have energies in the range 10–40 kJmol1, but are still weaker than ionic or covalent interactions by an order of magnitude. Hydrogen bonding

plays a central role in molecular self-assembly processes such as micelle formation, biological membrane structure and the determination of protein conformation.
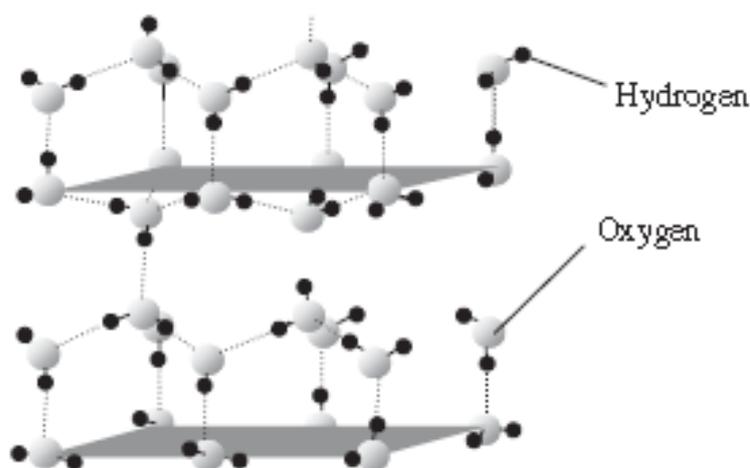

Figure 12.2 The molecular structure of crystalline water (ice I)(The hydrogen bonds are indicated by dotted lines, and the covalent bonds by continuous lines [Reprinted with permission from L. Pauling, Nature of Chemical Bond, Copyright (1960), Cornell University Press.
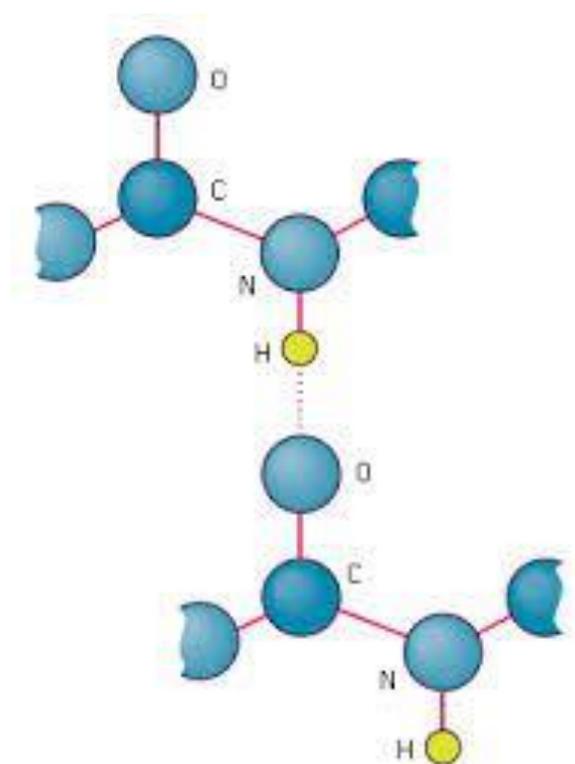

Figure 12.3 Examples of the range of possible geometries of hydrogen bonds encountered in organic molecules ((a) chain structure, (b) three dimensional structure, (c) bifurcated structure and (d)intramolecular bond). Source: Britannica.com

Hydrogen bonds occur between a proton donor group (D), which is the strongly polar group in a molecule such as FH, OH, NH, SH, and a proton acceptor atom (A), which is a slightly electronegative atom such as fluorine, oxygen, nitrogen and sulfur. Hydrogen bonding also has important consequences for a polar (nonpolar) biomolecules in aqueous solutions.

Water molecules arrange themselves in clathrate structures around hydrophobic compounds, e.g. the hydrophobic tails of lipids (Figure 12.4). The clathrates are labile (the water molecules can exchange position with their neighbours), but the water molecules are more ordered in the cages. Thus for a polar biomolecules (e.g. hydrocarbons) the free energy of transfer into an aqueous environment is proportional to the surface area of the molecules, since the entropy change is proportional to the area of the clathrate. This hydrophobic interaction is sometimes given the status of a separate mesoscopic force, since the reduction in free energy causes hydrophobic molecules to be driven together in aqueous solutions.
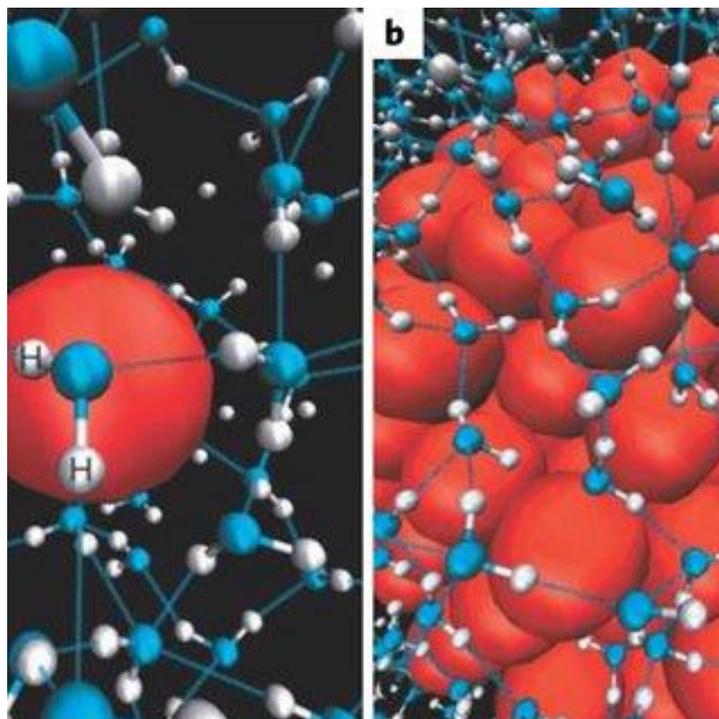
Figure 12.4 Water molecules around a hydrophobic Compound. ResearchGate - Configurations of liquid water molecules

The entropy of the associated water molecules plays a critical role in this case. However, to model such interactions care must be taken not to double count a queues mesoscopic forces under both the 'hydrogen bonding' and 'hydrophobic' banners. Surface force apparatus experiments have provided evidence on the long range nature of the hydrophobic effect. It is still an active area of study, but the energy of repulsion (W) is thought to have the basic form:

$$W = W_0 e^{\frac{\gamma}{\lambda}} (12.2)$$

where λ is the decay length and is typically in the order of nanometers, W_0 is a constant (with units of energy) and r is the distance between the surfaces.

Ab initio computational methods to quantify the strength of hydrogen bonds remain at a rudimentary level. One stumbling block to the analysis is the ability of hydrogen bonds to bifurcate (e.g. a single oxygen atom

can interact with two hydrogen molecules simultaneously); this leaves a would be model with a tricky multi-body problem. Another challenge is the wide spectrum of dynamic phenomena possible in hydrogen bonded solutions and care must be taken to determine the critical time window for the biological phenomena that need to be modelled. A series of important experimental advances in the dynamics of hydrogen bonds have been recently made using pulsed

femtosecond lasers. The lifetime of water molecules around solution state ions has been directly measured to be of the order of 10 picoseconds. It is hoped that such detailed experiments on the structure and dynamics of hydrogen bonds will allow tractable potentials to be refined.

Electrostatics

Unscreened Electrostatic Interactions

In principle, the electrostatic interaction between biomolecules can be calculated explicitly in a molecular dynamics simulation. Coulombs law ion–dipole and dipole–dipole interactions need to be treated, but rigorous quantitative potentials exist that can be calculated if the number of charged entities is sufficiently small (rarely the case with biological molecules, Figure 12.5). The direction of the interaction is paramount in the calculation of dipolar interactions, and also becomes important when Coulombic forces are experienced by extended objects, e.g. the parallel alignment of charged rods.
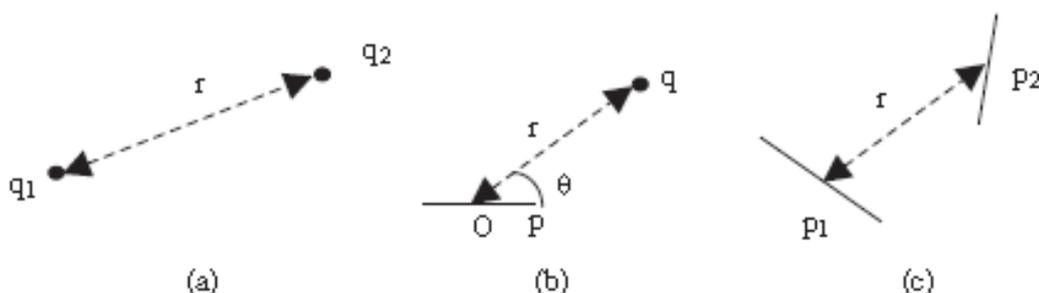

Figure 12.5 Geometry for the interaction between electrostatically charged ions and dipoles: (a) ion–ion ($q1, q2$), (b) ion–dipole (q, p), and (c) dipole–dipole ($p1, p2$))

Coulomb's law for the interaction energy (Ec) between two point charges is given by:

$$E_C = \frac{q_1 q_2}{4\pi\varepsilon\varepsilon_0 \gamma},$$

(12.3)

where e is the relative dielectric permittivity, e0 is the permittivity of free space, q_1 and q_2 are the magnitude of the two charges, and r is the distance between the charges. The next most important electro static interactions experienced by charged molecules are those between ions and dipoles. The energy of interaction (Ep) between a dipole (p) and a point charge (q) is given by:

$$E_P = \frac{p^2 q^2}{(4\pi\varepsilon_0)3kT\gamma^2} \quad (12.4)$$

where kT is the thermal energy. similarly there is an interaction energy

between two separate electric dipoles ($\eth Epp$) which is given by:

$$E_{PP} = \frac{p_1 p_2 K}{4\pi\varepsilon_0 \gamma^3} \qquad (12.5)$$

where K is a constant. Higher order electrostatic interactions (quadrupolar) are also possible, but provide progressively smaller contributions to the force of interaction in most biological scenarios.


(a) DNA chains   (b) Protein/DNA interaction   (c) Protein/Protein interaction   (d) Proteoglycans
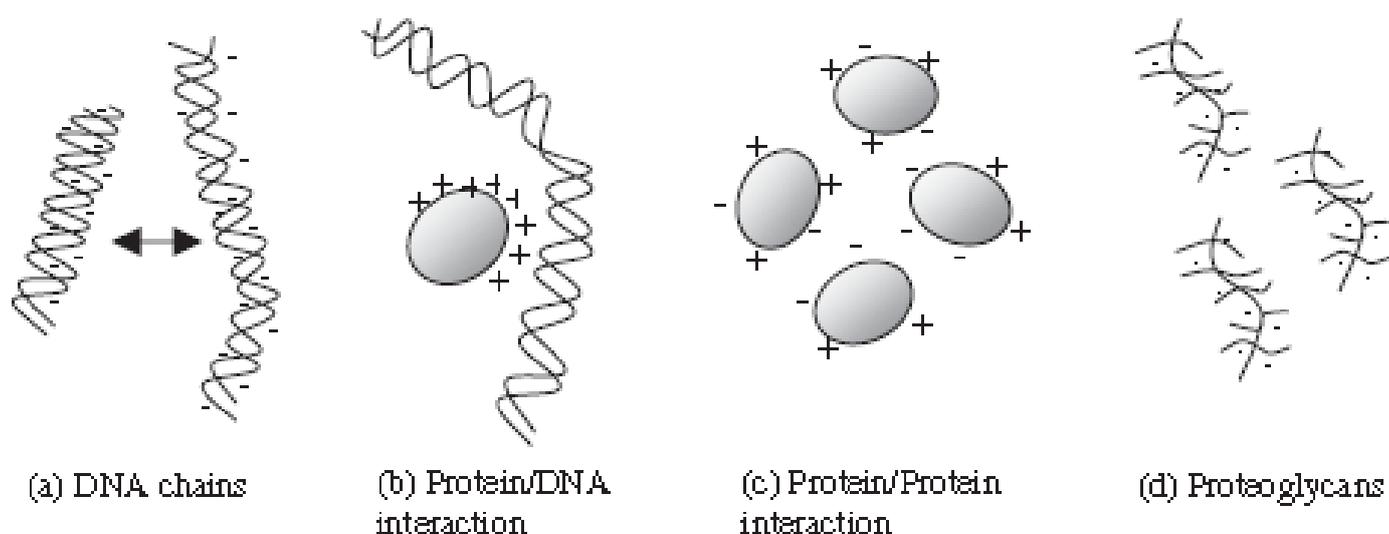
Figure 12.6 Schematic diagram of some molecular systems in which electrostatic interaction dominates the intermolecular forces((a) nucleic acids, (b) nucleic acids and proteins, (c) the aggregation of proteins and (d) proteoglycans.

Ionic bonds between molecules typically have a strength of the order of 500 kJmol. For a large range of biological molecules, electro static forces are vitally important for their correct functioning (Figure 2.6) and provide the dominant long range interaction.

Screened Electrostatic Interactions

An electric 'double layer' forms around charged groups in aqueous solution(Figure 12.7) and the process by which the double layer screens the Coulombic interaction is important in determining the resultant electrostatic forces.
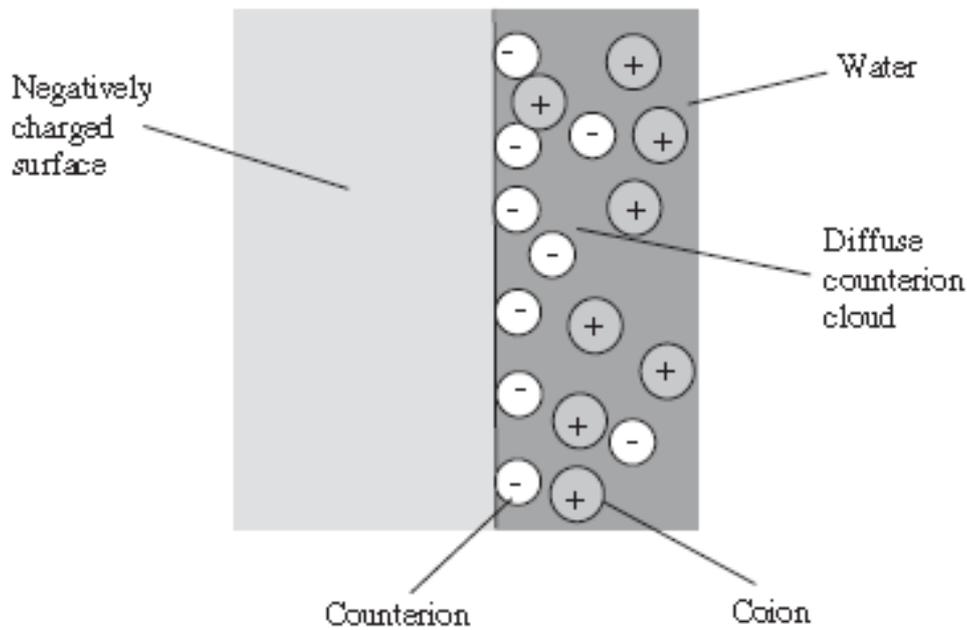
Figure 12.7. The distribution of counter ions in an electric double layer around a negatively charged surface.

The concept of screening allows a complex many body problem involving two strongly charged objects immersed in an electrolyte of many billions of simple ions to be reduced to a simple two body problem with a modified potential between the two strongly interacting objects.

The charging of a surface in a liquid can happen in two ways: by the dissociation of surface groups or by the adsorption of ions onto the surface.

For example, surface carboxylic groups can be charged by dissociation:

$$COOH \rightarrow COO + H^+, \tag{12.6}$$

which leaves behind a negatively charged surface and liberates approximately

14 kT of energy. The adsorption of an ion from solution onto apreviously uncharged or oppositely charged surface (e.g. binding Ca2þonto a negatively charged protein) could charge the surface positively. The chemical potential (m, total free energy per molecule) for the electric double layer that surrounds a charged aqueous system is the sum of two terms:

$$\mu = ze\Psi + kT log \, \rho, \tag{12.7}$$

where c is the electric potential, r is the number density of counter ions, kT is the thermal energy, z is the valence of the charged groups on themolecule, and e is the electronic charge. The first term is due to the electrostatic energy and the second is the contribution of the entropy of the counter ions. The form of the chemical potential (equation) is consistent with the Boltzmann distribution for the density of the counter ions (r) and can be reexpressed as:

$$\rho = \rho_0 e^{\frac{ze\Psi}{kT}} \tag{12.8}$$

where r_0 is related to the chemical potential:

$$\rho_0 = e^{\frac{\mu}{kT}} \tag{12.9}$$

A fundamental formula from electromagnetic theory is the Poisson equation for electrostatics. It relates the potential (c) to the free ion concentration immersed in a dielectric at a distance x from a charged surface:

$$\varepsilon, \varepsilon_0 \frac{d^2\psi}{dx^2} = \rho_{freeion} \tag{12.10}$$

where ε_0 is the permittivity of free space and ε_γ is the relative permittivity of the dielectic (e.g. water) in which the ions are embedded. The onedimensional version of the Poisson equation, dependent only on the perpendicular distance from the surface (x), is quoted for simplicity.

The Poisson equation for electrostatics can be combined with the Boltzmann distribution for the thermal distribution of ion energies and gives the Poisson–Boltzmann (PB) equation:

$$\frac{d^2\psi}{dx^2} = \frac{ze\rho}{\varepsilon_\gamma\varepsilon_0} e^{\frac{ze\Psi}{kT}} \tag{12.11}$$

The PB equation can be solved to give the potential (c), the electric field (E ¼ @c = @x) and the counter ion density (r) at any point in the gap between two planar surfaces. The density of counter ions and co ions from a planar surface can therefore be calculated as shown schematically in Figure 12.8.

There are some limitations on the validity of the PB equation at short separations which include: ion correlation effects (electronic orbitals become correlated), finite ion effects (ions are not point-like), image forces (sharp boundaries between dielectrics affect the solutions of the electromagnetism equations), discreteness of surface charges (the surface charge is not smeared out smoothly), and solvation forces (interaction of water molecules with the charges). The pressure (P) between two charged surfaces in water can often be calculated using the contact value theorem. It relates the force between two surfaces to the density of contacts (or ions in this case) at the midpoint:

$$P(\gamma) = kT[\rho_s(\gamma)\rho_s(\infty)] \tag{12.12}$$

where ρ_s is the ion concentration at infinity (e.g. bulk salt concentration), kT is the thermal energy and r is the separation between the surfaces.

With screened electrostatic interactions the inter surface pressure is given by the increase in the ion concentration at the surfaces as they approach each other. The theorem is valid as long as there is no specific interaction between the counter ions and the surfaces. The contact value theorem also functions well in other calculations of mesoscopic forces such as those of solvation interactions, polymer associated steric and depletion interactions, undulation and protrusion forces.

Depletion forces

Depletion forces are another mesoscopic interaction that are formed by a subtle range of more fundamental forces. An illustrative example is when colloidal spheres and polymers are mixed in aqueous solution. The colloids can experience an effective attractive interaction when the polymers are excluded from the volume between the approaching colloidal spheres (Figure 12.16). Such phenomena were originally verified

through macroscopic measurements on the phase separation of polymer/colloid mixtures. Recent experiments with dual trap optical tweezers have provided direct evidence for the depletion potential between two colloidal probes in DNA solutions (Figure 12.17).

The depletion force can be understood from an analysis of the thermodynamics. The addition of the polymer lowers the solvent's chemical potential, and creates a depletion force that drives the colloidal surfaces together.
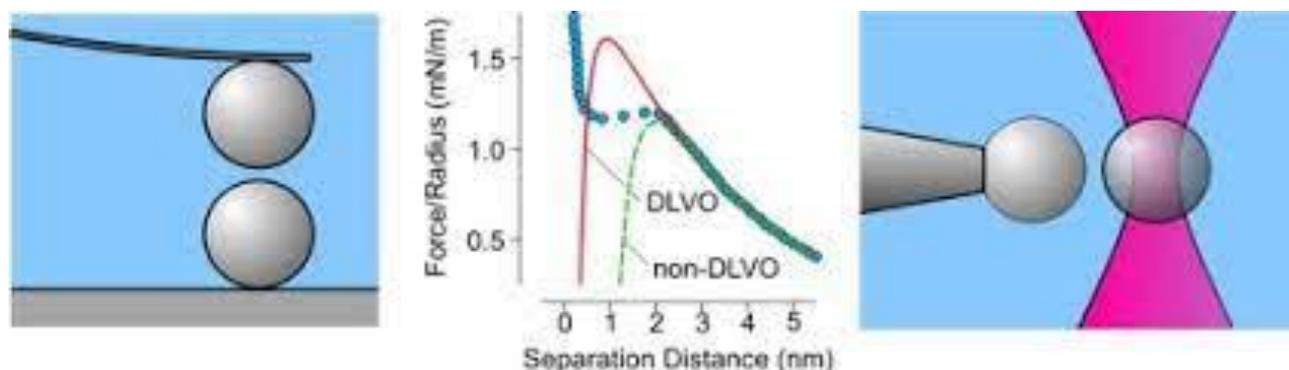

Figure 12.16 Depletion forces between two colloids in a solution of water.
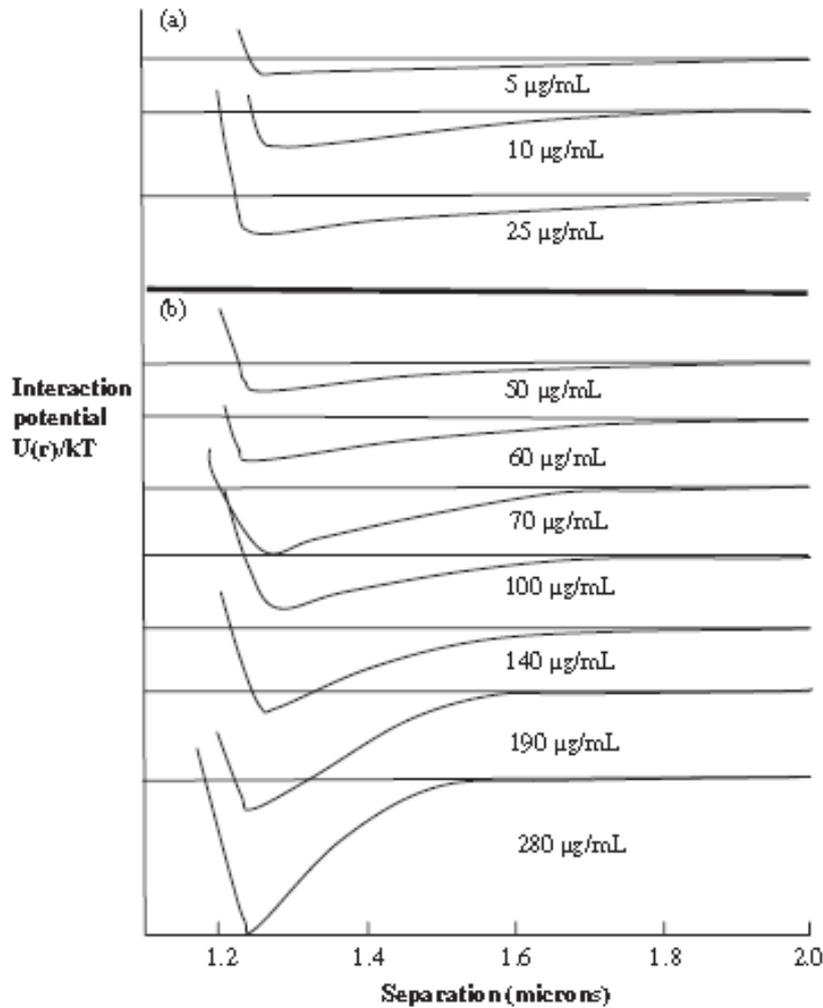
Figure 12.17 Depletion potential between two 1.25 *m*m silica spheres as a function of DNA concentration in (a) dilute and (b) semi-dilute DNA solutions measured with dual trap optical tweezers[Reprinted with permission from R. Verma, J.C. Crocker, T. C. Lubensky and A.G.Yodh, Macromolecules, 33, 177–186, Copyright (2000) American Chemical Society].

For a dilute concentration of polymers in a colloidal solution the osmotic pressure (P) is proportional to the number density of polymer chains (N=V), with each chain contributing kT to the osmoticpressure:

$$\Pi = \frac{N}{V}kT \tag{12.13}$$

This pressure is analogous to that of an ideal gas, that gives rise to the Van der Waals equation (P ¼ NRT=V, where R is the ideal gas constant).

The depletion force (F_{dep}) is approximately the product of theosmotic pressure with the volume from which the chains are depleted (V_{dep}):

$$F_{dep} = \Pi V_{dep}\Pi\frac{4}{3}\pi R_g^3 \tag{12.14}$$

where R_g is the radius of gyration of the polymer molecules. Thus a high molecular weight (large R_g) and a high polymer concentration (large P) are required for a strong depletion force. In ageneral biological

180

context the naturally occurring intracellular environment is extremely crowded and an intricate hierarchy of excluded volume depletion interactions can occur.

Hydrodynamic interactions

Each of the mesoscopic forces discussed in the previous section have a time scale associated with their action, since they can not occur stantaneously. Therefore the dynamics of the components of each system(e.g. solvents, counter ion clouds and tethered polymers) needs to be under stood to realistically gauge the strength of the interaction potentials. More advanced treatments of mesoscopic forces often consider how to evaluate the dependence of the forces on time.

Direct experimental measurements of intermolecular and surface forces

There are a large number of experimental probes for intermolecular forces, which operate using a small range of physical principles. (Figure 2.18). Some of the most important principles are briefly reviewed here. The thermodynamic properties of gases, liquids and solids (pressure–volume- temperature phase diagrams, boiling points, latent heats of and lattice energies) provide important information on short range inter particle forces. Similarly, adsorption isotherms provide information on the interactions of molecules with surfaces. A range of direct physical techniques on gases, liquids and solids (e.g. molecular beam scattering, viscosity, diffusion, compressibility, NMR, x-ray and neutron scattering experiments) can provide information on short range interactions of molecules, with particular emphasis on their repulsive forces. A sensitive method for the characterization of hydrated molecules is to measure the separation of the molecules (using X-ray scattering) in parallel with osmotic pressure measurements. This provides a non-invasive pico newton measurement of intermolecular forces (Figure 12.18).
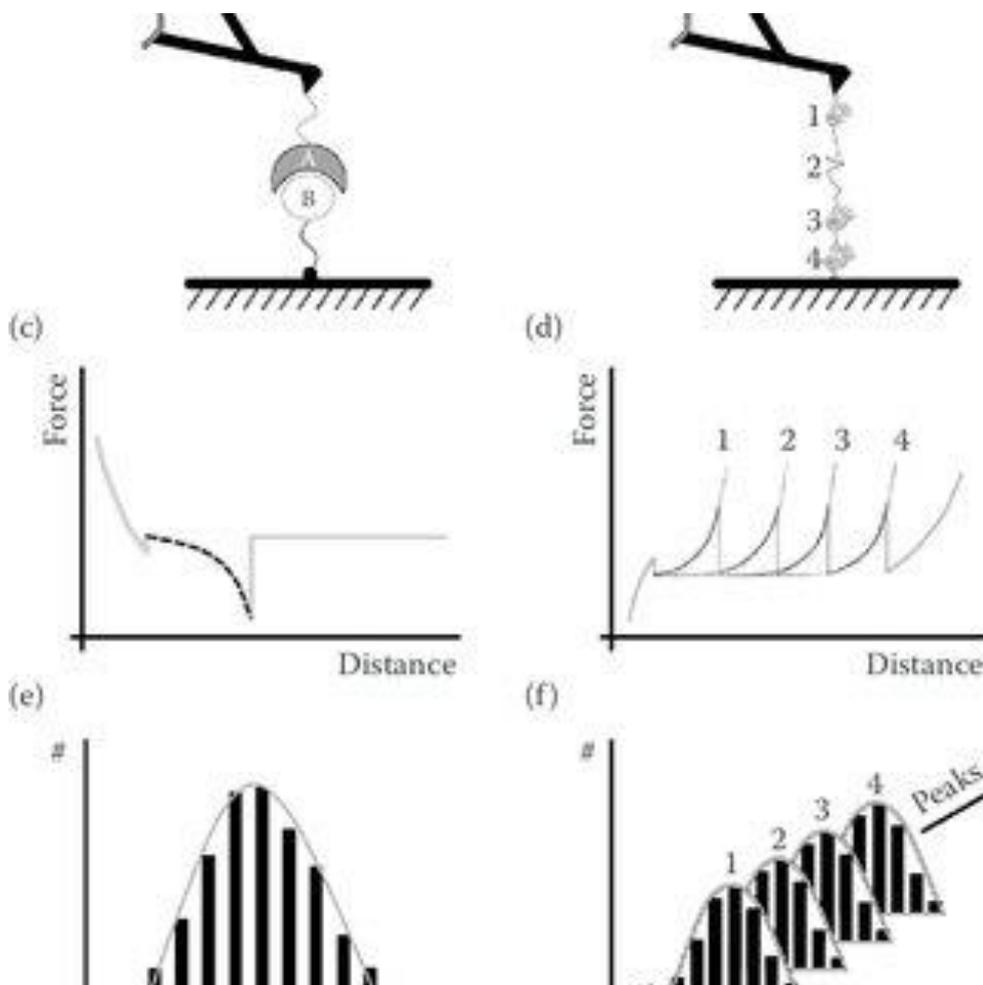
Figure 12.18 Range of techniques for the measurement of intermolecular forces.

Thermodynamic data on solutions (Phase stability)
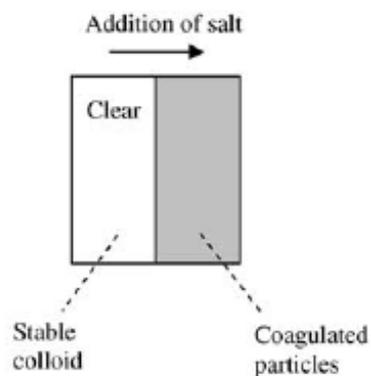

Adhesion experiments
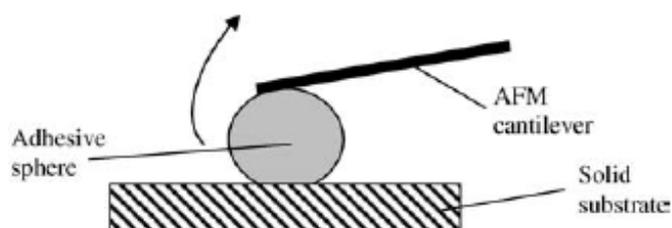
Direct force measurement (Optical tweezers)
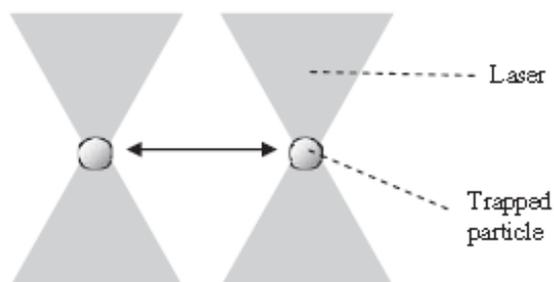

Surface studies
Reflectivity
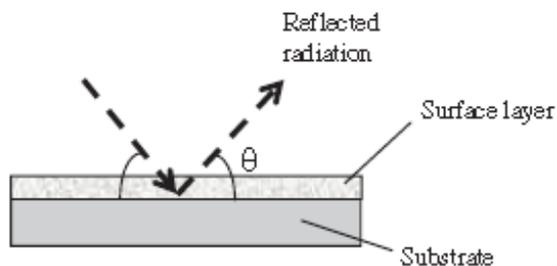
Contact angle
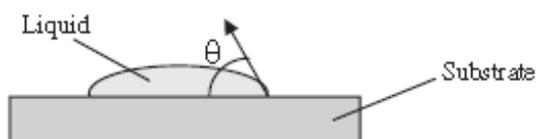
Figure.12.18 (Continued).
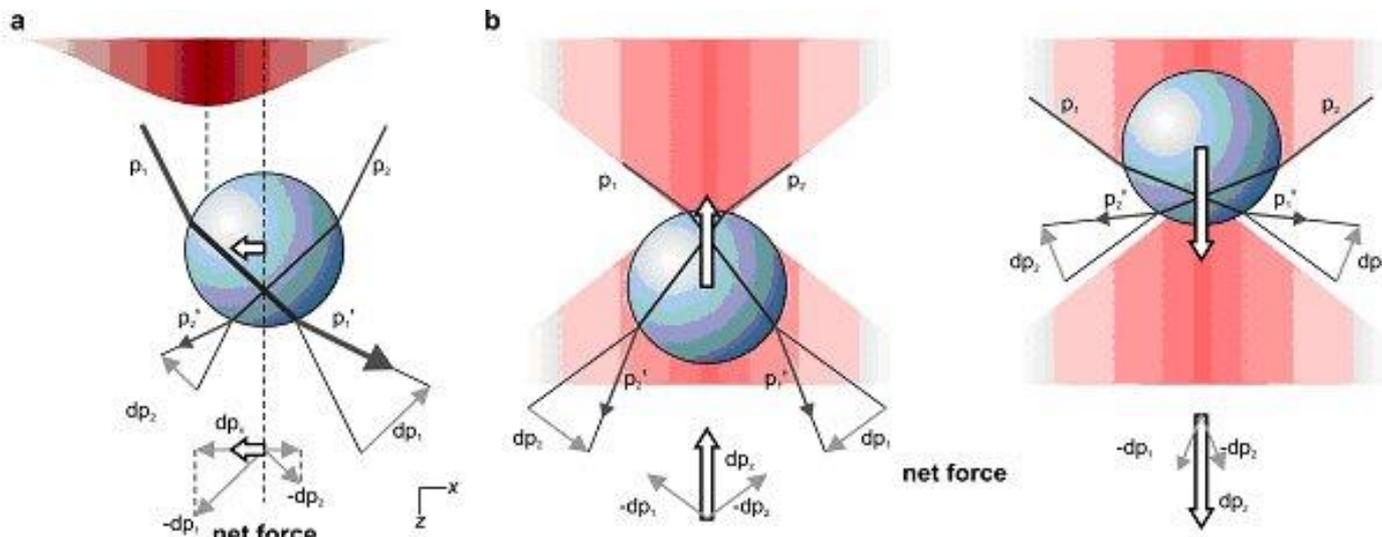
Figure.12.19. Examples of Optical Tweezers.

Thermodynamic data on solutions (phase diagrams, solubility, partitioning, miscibility and osmotic pressure) provide information on short range solute–solvent and solute–solute interactions. With colloidal dispersions coagulation studies as a function of the salt concentration, pH or temperature yield useful information on inter particle forces. Adhesion experiments provide information on particle/surface adhesion forces and the adhesion energies of solid surfaces in contact. Direct force measurement between two macroscopic/microscopic surfaces as a function of surface separation can characterise forces of interaction in great detail, e.g. surface force apparatus (SFA), optical/magnetic tweezers and atomic force microscopes (AFM).

Surface studies such as surface tension and contact angle measurements can elucidate liquid–liquid and solid–liquid adhesion energies. Similarly, the reflectivity of radiation (neutrons, X-rays and light) from surfaces can provide invaluable information on surface energies. With film balances, the thicknesses of free soap films and liquid films absorbed on surfaces can be measured using optical interferometry as a function of salt concentration or vapour pressure, and this again provides a direct measurement of inter surface potentials. Hydrodynamic studies of liquids can be made using nuclear magnetic resonance spectroscopy (NMR), and elastic/inelastic scattering of light, X-rays and neutrons. These methods are particularly useful for the measurement of hydrodynamic effects that lead to time dependent mesoscale forces.

Chapter 13

Biological Membranes

Structure and organization of membranes

Biological membranes They form cells and enable separation between the inside and outside of an organism, controlling by means of their selective permeability which substances enter and leave. By allowing gradients of ions to be created across them, membranes also enable living organisms to generate energy. In addition, they control the flow of messages between cells by sending, receiving and processing information in the form of chemical and electrical signals. This essay summarizes the structure and function of membranes and the proteins within them, and describes their role in trafficking and transport, and their involvement in health and disease. Techniques for studying membranes are also discussed.
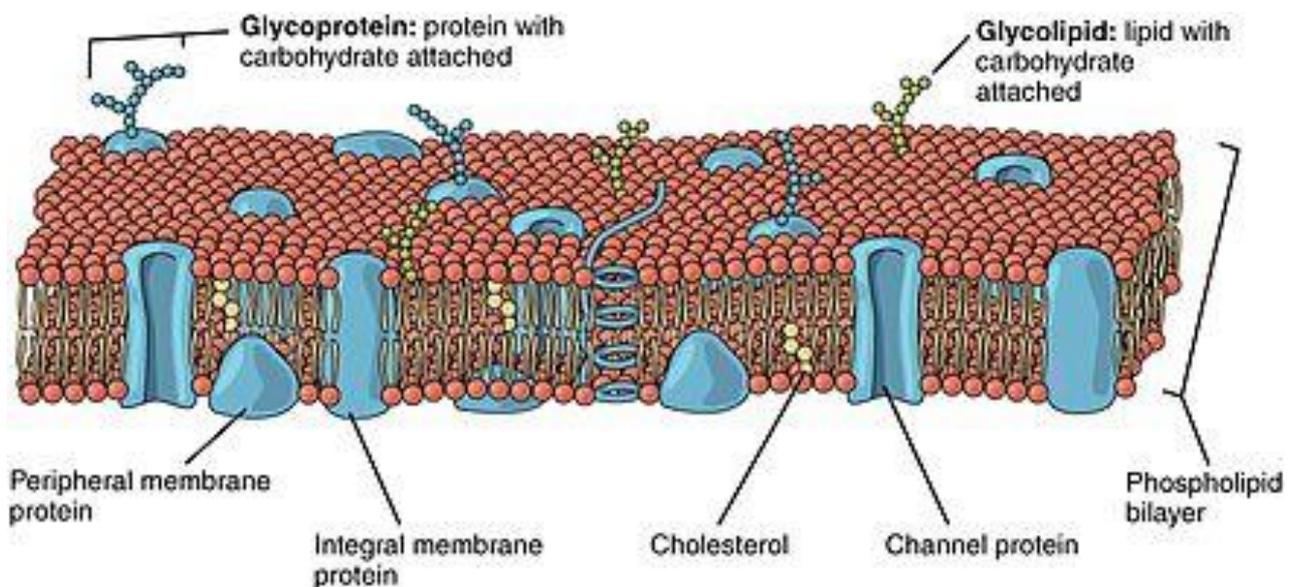
Fig.13.1 Biological membrane.(Wikipedia en.wikipedia.org)

Membranes - composition of lipids, proteins and sugars

Biological membranes consist of a double sheet (known as a bilayer) of lipid molecules. This structure is generally referred to as the phospholipid bilayer. In addition to the various types of lipids that occur in biological membranes, membrane proteins and sugars are also key components of the structure. Membrane proteins play a vital role in biological membranes, as they help to maintain the structural integrity, organization and flow of material through membranes. Sugars are found on one side of the bilayer only, and are attached by covalent bonds to some lipids and proteins.

Three types of lipid are found in biological membranes, namely phospholipids, glycolipids and sterols. Phospholipids consist of two fatty acid chains linked to glycerol and a phosphate group. Phospholipids containing glycerol are referred to as glycerophospholipids. An example of a glycerophospholipid that is commonly found in biological membranes is phosphatidylcholine (PC) (Figure 13.1), which has a choline molecule attached to the phosphate group. Serine and ethanolamine can replace the choline in this position, and these lipids are called phosphatidylserine (PS) and phosphatidylethanolamine (PE), respectively. Phospholipids can also be sphingophospholipids (based on sphingosine), such as sphingomyelin. Glycolipids can contain either glycerol or sphingosine, and always have a sugar such as glucose in place of the phosphate head found in phospholipids. Sterols are absent from most bacterial membranes, but are an important component of animal (typically cholesterol) and plant (mainly stigmasterol) membranes. Cholesterol has a quite different structure to that of the phospholipids and glycolipids. It consists of a hydroxyl group (which is the hydrophilic 'head' region), a four-ring steroid structure and a short hydrocarbon side chain.

The sugars attached to lipids and proteins can act as markers due to the structural diversity of sugar chains. For example, antigens composed of sugar chains on the surface of red blood cells determine an individual's blood group. These antigens are recognized by antibodies to cause an immune response, which is why matching blood groups must be used in blood transfusions. Other carbohydrate markers are present in disease (e.g. specific carbohydrates on the surface of cancer cells), and can be used by doctors and researchers to diagnose and treat various conditions.

Amphipathic lipids – bilayers formation

All membrane lipids are amphipathic—that is, they contain both a hydrophilic (water-loving) region and a hydrophobic (water-hating) region. The most favorable environment for the hydrophilic head is an aqueous one, whereas the hydrophobic tail is more stable in a lipid environment. The amphipathic nature of membrane lipids means that they naturally form bilayers in which the hydrophilic heads point outward towards the aqueous environment and the hydrophobic tails point inward towards each other. When placed in water, membrane lipids will spontaneously form liposomes, which are spheres formed of a bilayer with

water inside and outside, resembling a tiny cell. This is the most favourable configuration for these lipids, as it means that all of the hydrophilic heads are in contact with water and all of the hydrophobic tails are in a lipid environment.
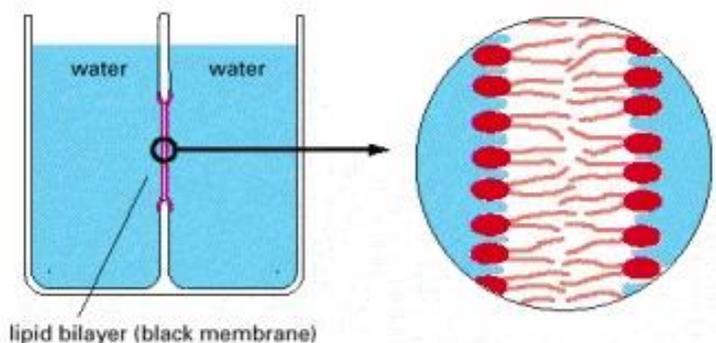

Figure 13.2. A cross-sectional view of a black membrane, a synthetic lipid belayers.

Early experiments by E. Gorter and F. Grendel in 1925 were the first to demonstrate that biological membranes are bilayers. These researchers extracted the lipids from red blood cells and found that they occupied a space that was twice the surface area of the cell. Red blood cells contain no internal membranes, so they deduced that the plasma membrane must be composed of two layers of lipids.

Biological membranes and the fluid mosaic model

The fluid mosaic model proposed by Jonathan Singer and Garth Nicolson in 1972 describes the dynamic and fluid nature of biological membranes. Lipids and proteins can diffuse laterally through the membrane. Phospholipids can diffuse relatively quickly in the leaflet of the bilayer in which they are located. A phospholipid can travel around the perimeter of a red blood cell in around 12 s, or move the length of a bacterial cell within 1 s. Phospholipids can also spin around on their head-to-tail axis, and their lipid tails are very flexible. These different types of movements create a dynamic, fluid membrane which surrounds cells and organelles. Membrane proteins can also move laterally in the bilayer, but their rates of movement vary and are generally slower than those of lipids. In some cases, membrane proteins are held in particular areas of the membrane in order to polarize the cell and enable different ends of the cell to have different functions. One example of this is the attachment of a glycosyl-phosphatidylinositol (GPI) anchor to proteins to target them to the apical membrane of epithelial cells and exclude them from the basolateral membrane. Fluorescence photobleaching is one experimental method that is used by scientists to demonstrate visually the motility of proteins and lipids in a bilayer (Figure 13.3). A lipid or membrane protein located on the surface of a cell is tagged with a fluorescent marker such as green fluorescent protein (GFP). A beam of laser light is then focused on to a small area of the cell surface using a fluorescence microscope in order to bleach the fluorescent tags in this area so that they no longer emit a fluorescence signal. This small area of membrane is observed over time and gradually the fluorescence increases again, indicating that other tagged proteins or lipids are diffusing into this region from elsewhere in the membrane. This demonstrates that the

lipid bilayer surrounding cells is fluid in nature and allows lateral diffusion of both lipids and membrane proteins.
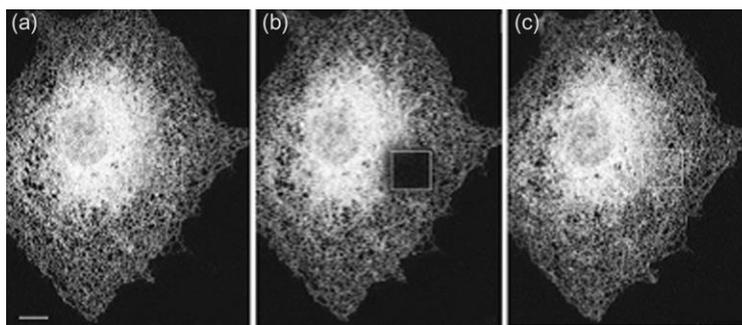

Fig. 13.3. Cells expressing a GFP-labelled protein in the endoplasmic reticulum were subjected to photobleaching. (a) A cell before bleaching. (b) The same cell immediately after bleaching of the square section shown. (c) The same cell 5 min after photobleaching. Source: Lippincott-Schwartz, J., Snapp, E. and Kenworthy, A. (2001) Studying protein dynamics in living cells. Nat. Rev. Mol. Cell. Biol. 2, 444–456.

Despite all this movement of lipids and proteins in the bilayer, vertical movement, or 'flip-flop', of lipids and proteins from one leaflet to another occurs at an extremely low rate. This is due to the energetic barrier encountered when forcing the hydrophilic head (in the case of lipids) or hydrophilic regions (in the case of proteins) through the hydrophobic environment of the inside of the membrane. This near absence of vertical movement allows the inner and outer leaflets of the bilayer to maintain different lipid compositions, and enables membrane proteins to be inserted in the correct orientation for them to function. However, some enzymes facilitate the process of lipid flip-flop from one leaflet to another. These flippases, or phospholipid translocators, use ATP to move lipids across the bilayer to the other leaflet. In eukaryotic cells, flip passes are located in various organelles, including the endoplasmic reticulum (ER), where they flip-flop newly synthesized lipids.

Formation of membranes

Biological membranes are formed by adding to a pre-existing membrane. In prokaryotes this occurs on the inner leaflet of the plasma membrane, facing the cytoplasm. In eukaryotes, membrane synthesis takes place at the ER on the cytoplasmic leaflet of the ER membrane (termed the 'inside' of the cell). Lipids then leave the ER and travel through the secretory pathway for distribution to various subcellular compartments or the plasma membrane.

In eukaryotic cells, enzymes that span the ER catalyze the formation of membrane lipids. In the cytoplasmic leaflet of the ER membrane, two fatty acids are bound, one by one, to glycerol phosphate from the cytoplasm. This newly formed diacylglycerol phosphate is anchored in the ER membrane by its fatty acid chains. The phosphate is then replaced by the head group (e.g. phosphate and choline). Flip passes in the ER membrane can then move some of these newly formed lipids to the luminal side of the ER membrane. Similarly, flippases in prokaryotes can transfer new lipids from the inner leaflet of the plasma membrane

to the outer leaflet. These flippases are responsible for adjusting the lipid composition of each layer of the membrane. In eukaryotes, lipids must then be distributed to the various intracellular membranes. The traffic of vesicles between organelles in combination with signals that direct particular lipids to specific locations is required to create the correct lipid composition in all of the cellular membranes. Vesicles bud from the ER and travel via the ER–Golgi intermediate compartment (ERGIC) to join with the Golgi, where sorting of lipids takes place. The Golgi then sends lipids in vesicles to various destinations, including the plasma membrane and lysosomes. Lipids and proteins are internalized from the plasma membrane into endosomes. Organelles, such as mitochondria, acquire lipids from the ER by a different mechanism. Water-soluble proteins called phospholipid-exchange proteins remove phospholipids from the ER membrane and deposit them in the membranes of the appropriate organelles.

Distribution of lipids

The inner and outer leaflets of bilayers differ in their lipid composition. In mammalian cells, the outer leaflet of the plasma membrane contains predominantly PC and sphingomyelin, whereas PS and PE are found on the inner leaflet. During programmed cell death (apoptosis), PS is no longer restricted to the inner leaflet of the plasma membrane. It is exposed on the outer leaflet by the action of an enzyme called scramblase which is a type of flippase enzyme. PS is negatively charged, unlike PC, which has no net charge. The movement of PS into the outer leaflet therefore changes the charge of the plasma membrane as viewed from the outside of the cell. This change in surface charge labels the apoptotic cell for phagocytosis by phagocytic cells such as macrophages.

Lipid composition also varies between the organelles within eukaryotic cells. Cholesterol is synthesized in the ER, but the ER membrane has a relatively low cholesterol content, as much of the cholesterol is transported to other cellular membranes. The prevalence of cholesterol in membranes increases through the secretory pathway, with more in the Golgi than in the ER (the *trans*-Golgi network is richer in cholesterol than the *cis*-Golgi), and most in the plasma membrane. This increase in cholesterol through the secretory pathway results in slightly thicker membranes in the late Golgi and plasma membrane compared with the ER, and is thought to be a contributing factor to protein sorting through the pathway, as membrane proteins in the plasma membrane generally have longer hydrophobic transmembrane domains than membrane proteins that reside in the ER.

Membrane proteins and its synthesis

Membrane proteins are the nanomachines that enable membranes to send and receive messages and to transport molecules into and out of cells and compartments. Without membrane proteins the phospholipid membrane would present an impenetrable barrier and cells would be unable to communicate with their

neighbours, transport nutrients into the cell or waste products out of it, or respond to external stimuli. Both unicellular and multicellular organisms need membrane proteins in order to live. The membrane proteins that are present in a particular membrane determine the substances to which it will be permeable and what signal molecules it can recognize.

In eukaryotic cells, the synthesis of membrane proteins destined for the plasma membrane, ER or any other membrane-bound compartment begins on cytosolic ribosomes. After a short segment of protein has been synthesized, the ribosome, mRNA and nascent protein chain associate with the ER, where the rest of the protein is made and simultaneously inserted into the membrane. This phenomenon was first explained by Günter Blobel, David Sabatini and Bernhard Dobberstein in the 1970s. These scientists proposed that there is a 'binding factor' which recognizes the emerging protein chain and can dock the ribosome at the ER membrane. We now know that there is an N-terminal signal sequence within membrane proteins. These signal sequences are not identical but share a common motif, namely a hydrophobic stretch of 20–30 amino acids, a basic region at the N-terminus and a polar domain at the C-terminus of the signal. These N-terminal signal sequences are recognized by the signal recognition particle (SRP), which has binding sites for the signal sequence, ribosome and the SRP receptor which is embedded in the ER membrane. Upon binding the SRP, the ribosome pauses protein synthesis. The SRP binds to the SRP receptor, adjacent to a translocon pore in the ER membrane. The translocon is a protein pore through which membrane protein chains can be threaded into the membrane. It has a laterally opening gate to allow newly synthesized proteins into the ER membrane. Once the ribosome is at the translocon, the SRP dissociates and protein synthesis resumes.

The key steps of ER targeting are summarized. Each component is labelled and the ER membrane is represented by double blue lines. The signal sequence (shown in black) becomes the first transmembrane domain of the protein in this example. Co-translational targeting is the dominant mechanism for protein delivery to the ER in higher eukaryotes, whereas yeast and prokaryotes favor post-translational targeting, whereby proteins are delivered to the ER after completion of synthesis. Post-translational targeting also occurs in higher eukaryotes, often when a membrane protein is so small that the signal sequence does not emerge until the whole protein has been synthesized. Post-translational targeting can be carried out both by SRP-dependent and by SRP-independent mechanisms.

Membrane-spanning proteins are diverse in structure and function. They can be constructed of α-helices or from β-barrels. The β-barrel membrane proteins often function as pores, with hydrophobic amino acids facing out into the bilayer. In addition, there are other non-spanning proteins which associate with the bilayer, often using a hydrophobic anchor. Here we shall focus on the α-helical membrane proteins. These proteins have at least one α-helical hydrophobic stretch of amino acids, around 20 residues in length, which corresponds to around 30 Å (the thickness of an average phospholipid bilayer). If an α-helical membrane protein spans the membrane more than once, it will have more than one of these hydrophobic sections. For

189

example, the Ca^{2+}-ATPase of the ER and sarcoplasmic reticulum (SR) spans the membrane 10 times, so it has 10 hydrophobic stretches of around 20 amino acids each.

A vital class of membrane proteins are those involved in active or passive transport of materials across the cell membrane or other subcellular membranes surrounding organelles. For a cell or an organism to survive, it is crucial that the right substances enter cells (e.g. nutrients) and the right substances are transported out of them (e.g. toxins). Molecules can cross biological membranes in several different ways depending on their concentration on either side of the membrane, their size and their charge. Some molecules, including water, can simply diffuse through the membrane without assistance. However, large molecules or charged molecules cannot cross membranes by simple diffusion. Charged molecules such as ions can move through channels passively, down electrochemical gradients. This movement is described as 'downhill', as the ions or molecules travel from an area of high concentration to an area of low concentration. This requires channel proteins but no energy input. Passive transport can also be mediated by carrier proteins that carry specific molecules such as amino acids down concentration gradients, again without any requirement for energy. Active transport moves species against concentration gradients and requires energy, which is obtained from ATP, from light, or from the downhill movement of a second type of molecule or ion within the same transporter (Figure 13.4).
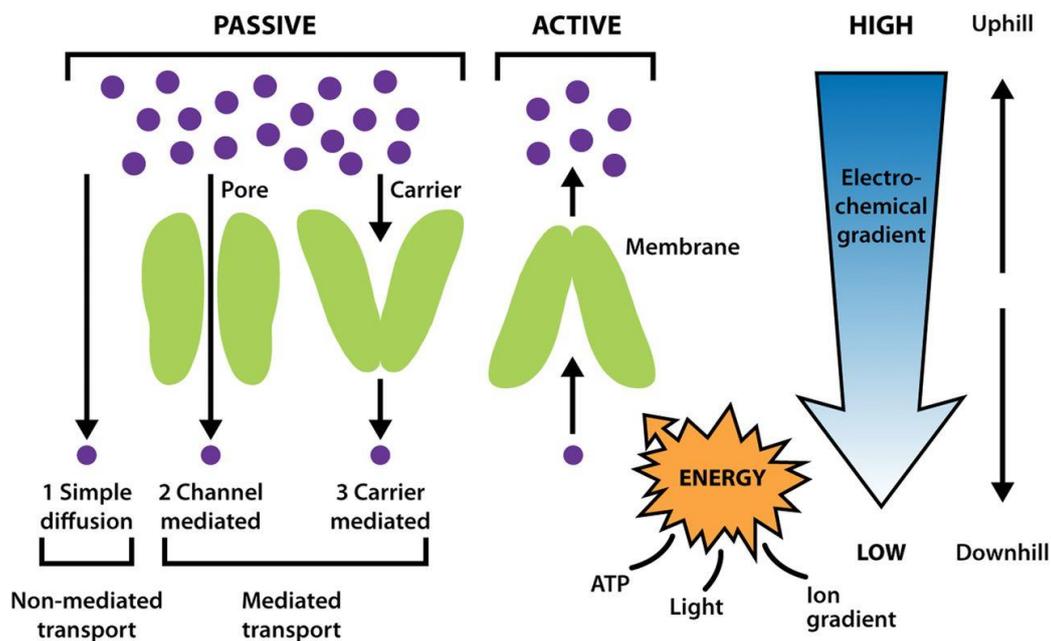
Fig.13.4. Passive and active transport. The different types of membrane proteins involved in passive and active transport are shown.

Passive transport is the movement of molecules across biological membranes down concentration gradients. This type of transport does not require energy. Channels form water-filled pores and thus create a hydrophilic path that enables ions to travel through the hydrophobic membrane. These channels allow downhill movement of ions, down an electrochemical gradient. Both the size and charge of the channel

pore determine its selectivity. Different channels have pores of different diameters to allow the selection of ions on the basis of size. The amino acids that line the pore will be hydrophilic, and their charge will determine whether positive or negative ions travel through it. For example, Ca^{2+} is positively charged, so the amino acids lining the pores of Ca^{2+} channels are generally basic (i.e. they carry a negative charge).

Channels are not always open. They can be gated by ligands which bind to some part of the protein, either by a change in membrane potential (voltage gated) or by mechanical stress (mechanosensitive). The nicotinic acetylcholine receptor is an example of a ligand-gated ion channel which opens upon binding the neurotransmitter acetylcholine. The nicotinic acetylcholine receptor is a pentameric membrane protein composed of five subunits arranged in a ring, with a pore through the centre. In the closed state, the pore is blocked by large hydrophobic amino acid side chains which rotate out of the way upon acetylcholine binding to make way for smaller hydrophilic side chains, allowing the passage of ions through the pore. Opening of the nicotinic acetylcholine receptor allows rapid movement of Na^+ ions into the cell and slower movement of K^+ ions out of the cell, in both cases down the electrochemical gradient of the ion. The difference in gradients between Na^+ and K^+ across the membrane means that more Na^+ enters the cell than K^+ leaves it. This creates a net movement of positive charges into the cell, resulting in a change in membrane potential. Acetylcholine released by motor neurons at the neuromuscular junction travels across the synapse and binds to nicotinic acetylcholine receptors in the plasma membrane of the muscle cells, causing membrane depolarization. This depolarization of the muscle cells triggers Ca^{2+} release and muscle contraction. There are two types of co-transport. In order to transport glucose into cells, the Na^+–glucose symporter uses the electrochemical gradient of Na^+ across the plasma membrane. The concentration of Na^+ is much higher outside the cell, and the inside of the cell is negatively charged relative to the outside, so by allowing Na^+ to travel down its electrochemical gradient, these transporters can move glucose uphill, into the cell and against its concentration gradient. This is referred to as symport, as both Na^+ and glucose travel in the same direction—in this case into the cell. In order for this symport to be sustainable, the Na^+ gradient must be maintained. This is done by the Na^+/K^+-ATPase, which uses ATP to pump the Na^+ back into the extracellular space, thus maintaining a low intracellular Na^+ concentration. Both Na^+ and Ca^{2+} are present at much higher concentrations outside the cell than inside it. Like the Na^+–glucose symporter, the Na^+–Ca^{2+} exchanger uses the electrochemical gradient of Na^+ across the plasma membrane to move a second species (Ca^{2+}) against its electrochemical gradient. However, in this case the transporter is an antiporter, as it uses the concentration gradient of one substance moving in (Na^+) to move another (Ca^{2+}) out of the cell. This antiporter has an exchange rate of three Na^+ ions in to two Ca^{2+} ions out. It moves Ca^{2+} out of the cell faster than the plasma membrane equivalents of SERCA, but has a lower affinity for Ca^{2+} than these P-type ATPases. Again this transporter relies on the Na^+/K^+-ATPase to maintain the low intracellular Na^+ concentration.

Investigations of the structure of membrane proteins

In order to understand more fully the mechanisms of action of membrane proteins such as the transporters described here, we can determine their three-dimensional protein structures. As a result of huge advances in structural biology in the last 50 years, we now have access to many thousands of protein structures in online databases. This enables researchers to visualize the structure of their protein of interest, and thus gain insight into its mechanism.

The structure of whale myoglobin was solved in 1958 using X-ray crystallography, earning John C. Kendrew and Max Perutz the Nobel Prize in Chemistry. This was the first protein structure to be solved using this technique, and since then thousands of proteins have been solved using this method. X-ray crystallography works by firing a beam of X-rays at a crystalline structure and measuring the diffraction of the X-rays after they have passed through the structure of interest. This generates an electron density map, showing where different atoms in the structure are located. For regular crystalline solids such as salts this is relatively straightforward, but for large irregular molecules such as proteins it can present many technical challenges. Before a protein is subjected to X-ray beams, it must first be purified and crystallized. In nature, proteins exist in the busy milieu of a cell, surrounded by thousands of other types of proteins, as well as lipids and other molecules. A common method of obtaining enough of the protein of interest involves expressing the relevant gene in a system such as bacteria. The gene is tagged with a small protein tag which can be used to isolate the protein of interest. Bacterial systems allow large amounts of protein to be produced cheaply and quickly. However, if the protein of interest is from a species that is only distantly related to that in which it is normally expressed (e.g. a human protein produced in *Escherichia coli* (*E. coli*)), the lack of correct glycosylation enzymes and the differences in protein folding and assembly may prevent the production of a biologically active protein. In addition, the expression of membrane proteins that make pores or channels can kill the host organism.

A pure protein sample is then crystallized by allowing water to evaporate away, in exactly the same way as a solution of salt will form crystals naturally when left to dry. Optimum conditions for this must be determined, and crystallization conditions are not always straightforward, as they differ from one protein to another. For soluble proteins such as myoglobin this is easier than for insoluble membrane proteins. Membrane proteins have lipid-soluble domains that will not dissolve in an aqueous medium. This significantly decreases the ease with which membrane protein structures can be solved using X-ray diffraction. However, there are ways in which scientists can overcome this difficulty. Generally, membrane proteins are removed from the membrane in which they were made and placed in an environment of lipids and detergents for crystallization. Sometimes the lipids associated with the protein are apparent in the crystal structure.

The number of solved crystal structures of proteins is constantly growing as technology improves and expertise is shared among scientists to help to optimize conditions for crystal production. The Protein Data Bank (PDB) is an online archive of protein structures which can be freely accessed by scientists worldwide. At the time of writing, 88% of the structures in the PDB have been solved by X-ray crystallography, and there are currently just under 70 000 X-ray crystal structures in the database. The number of membrane protein structures in the PDB is increasing rapidly with the refinement of crystallization techniques.
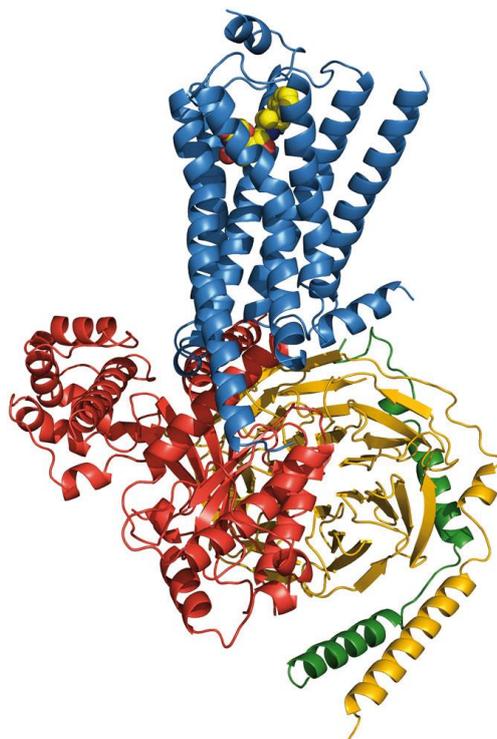
Fig.13.5. Passive and active transport. A structural model of a GPCR (blue) with signalling molecule bound (yellow spheres), activating a G protein (red, gold, and green).

The lipids that surround membrane proteins in biological membranes play an important role in the activity of these proteins. As was mentioned earlier, some membrane protein crystal structures include lipids bound to the outside surface of the transmembrane domains of the proteins. It is thought that these lipids bind tightly to the protein, and have a long-lived interaction with the transmembrane region. In other cases, lipids are thought to interact briefly with membrane proteins, rapidly moving away and being replaced by other membrane lipids. The activity of membrane proteins is considered to be dependent to some extent on the lipids that surround them in the membrane. Certain types of K^+ channel are thought to bind to negatively charged membrane lipids, as the activity of these channels increases at higher anionic lipid concentrations. These types of interaction can be studied by placing a purified form of the protein of interest in an artificial bilayer and measuring its activity. By altering the types of lipid present in the artificial bilayer, deductions can be made about the lipids that the protein requires in order to be active. Fluorescence spectroscopy and electron spin resonance are two techniques that are used to measure how strongly membrane proteins interact with specific lipids around them.

Molecular dynamics simulations use computer algorithms to work through theoretical problems. These simulated experiments are useful for investigating interactions between membrane proteins and lipids, as in real membranes these interactions are often so fleeting that they are very difficult to measure. Molecular dynamics simulations have predicted that in the case of the nicotinic acetylcholine receptor, the negatively charged lipid, phosphatidic acid, is required for activity. These simulations have also shown that cholesterol stabilizes the receptor and that the phosphatidic acid forms a shell around the protein which is more long-lasting than the interactions with other membrane lipids. Although molecular dynamics simulations are extremely useful, they are limited by the assumptions and approximations on which they are based. As in many areas of biology, a combination of experimental and computational research is required if real progress is to be made in understanding the complexity of biological membranes.

Inside the plasma membrane that surrounds eukaryotic cells lie many other membranes which define the intracellular compartments, or organelles. Each of these organelles has distinct functions and contains specific complements of proteins adapted for these roles. With the exception of a few proteins that are coded for by the mitochondrial genome, synthesis of all of the proteins that are required in these organelles begins on ribosomes in the cytoplasm, and therefore the proteins must be directed to the correct destination. We have seen earlier how this is achieved with membrane proteins, and most organelles have some kind of signal sequence that can be recognized by various receptors and which ensures that the protein arrives at the correct organelle.

All these gene to cells, by harnessing their ability to inject cells with foreign DNA or RNA. Patients may also be able to be given liposomes containing the functional gene, which fuse with cell membranes and deliver the therapeutic gene. Gene therapy is a growing and important area of research, and it is hoped that many diseases, including some cancers, will eventually be able to be treated using DNA.

Membrane proteins and viruses

Viruses that attack the human body can use the body's own membrane proteins to recognize their target cells. HIV attacks cells of the immune system. A protein on the surface of HIV called gp120 binds to CD4 protein molecules (CD4 is a membrane glycoprotein - molecular weight $55\ kDa$ expressed on helper T lymphocytes) on the surface of cells that are involved in immunoregulation, and allows fusion of the virus with the host cell. Once the contents of the virus have entered the CD4-positive cell, the HIV genome is integrated with the host genome and uses the host machinery to make new copies of the virus. Over time, the numbers of CD4 T-cells are reduced by the virus, and the patient's immune system eventually becomes so compromised that they are unable to fight invading pathogens. Many therapeutic agents have been created to help to fight HIV, and the interaction between CD4 and gp120 is just one of the points at which drugs can be used to stop the progression of the virus.

Various toxins interfere with the transmission of messages across biological membranes. Tetanus neurotoxin (TeNT) and botulinum neurotoxin (BoNT) are both protein toxins that affect nerve impulse

transmission between nerves and muscles. TeNT is produced by a soil bacterium and causes the skeletal muscle spasms that characterize tetanus infection. TeNT-producing bacteria generally enter the body through wounds, and TeNT binds glycolipids enriched at presynaptic membranes of motor neurons (Figure 14). TeNT then undergoes endocytosis and moves up the axon to the dendrites that connect the motor neuron to an inhibitory interneuron. TeNT is released into the synapse between these two cells and is endocytosed into the inhibitory interneuron. Acidification of vesicles containing TeNT causes the protein toxin to break apart into two domains. One of these, the L domain, is translocated into the cytoplasm of the interneuron, where it uses its proteolytic activity to cleave vesicle-associated membrane protein (VAMP). Under normal circumstances, VAMP is part of the protein complex that allows synaptic vesicles to fuse with the presynaptic membrane and release inhibitory neurotransmitters.

Membrane proteins are important drug targets. As our structural and functional knowledge of membrane proteins expands, it is becoming possible to design more effective medicines. Computational tools are becoming an increasingly important part of the process. One important class of drug targets are pore-forming membrane proteins encoded by viruses. HIV, influenza and polio, among other viruses, encode membrane proteins that form pores in the host cell membranes in order to cause leakage and promote infection. One of these pore-forming proteins was formerly used as a drug target in the treatment of influenza. NMR studies have provided structural information about the pore-forming Vpu protein from HIV-1. Using these data together with structural information about pores with similar sequences, computational models of the structure of the channel in the host membrane can be produced. These models, combined with advanced biophysical techniques, are invaluable for predicting sites for potential drug molecule binding which can then be tested both computationally and experimentally.

Using different drug molecules to stabilize different conformations in different signaling pathways may be the best approach to finding more effective medicines in the future. There is now much pressure on researchers to replace, refine and reduce the use of animals in drug discovery (an approach referred to as the 'three Rs'). By using computers in the early stages of the process to model drug–target interactions, researchers can produce much more promising compounds to test in experiments and drug trials.
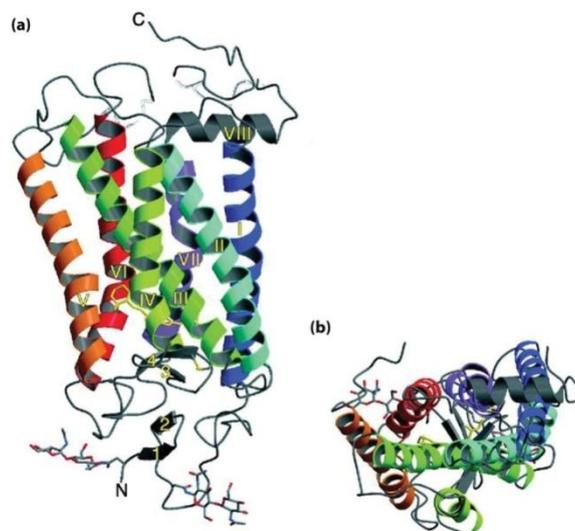
Fig.13.6. The crystal structure of rhodopsin. A ribbon representation of the first crystal structure of rhodopsin is shown in the plane of the membrane (a) and from the cytoplasmic side (b). The N- and C-termini are labelled, as are the seven transmembrane helices (I-VII). Source: Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E. et al. (2000) Crystal structure of rhodopsin: a G-protein-coupled receptor. Science 289, 739–745.

Biological membranes allow life to exist. From simple unicellular prokaryotes to complex multicellular eukaryotes such as humans, the properties of the membranes that surround cells are remarkably similar.

Our understanding of the structure of these lipid bilayers is now expanding rapidly as a result of significant advances in biophysical techniques and the huge computational power now available to researchers. The proteins that inhabit these membranes allow messages to be sent and received so that the cell can communicate with the external environment. Many messages are relayed by hydrophilic molecules that require receptors to transmit information across the bilayer. It is this step that is targeted by the majority of drugs which are on the market today, as it enables us to modify the message before it enters the cell. An understanding of how membrane proteins work, how they reach the correct destinations and how we can alter their functions is key to the fight against human disease.

Cells and Cellular Compartments Bound by Membranes

→ protect the content of the cell/compartment

→ must enable passage of the critical nutrients in and waste out

→ must be flexible (to allow cell growth/divison)
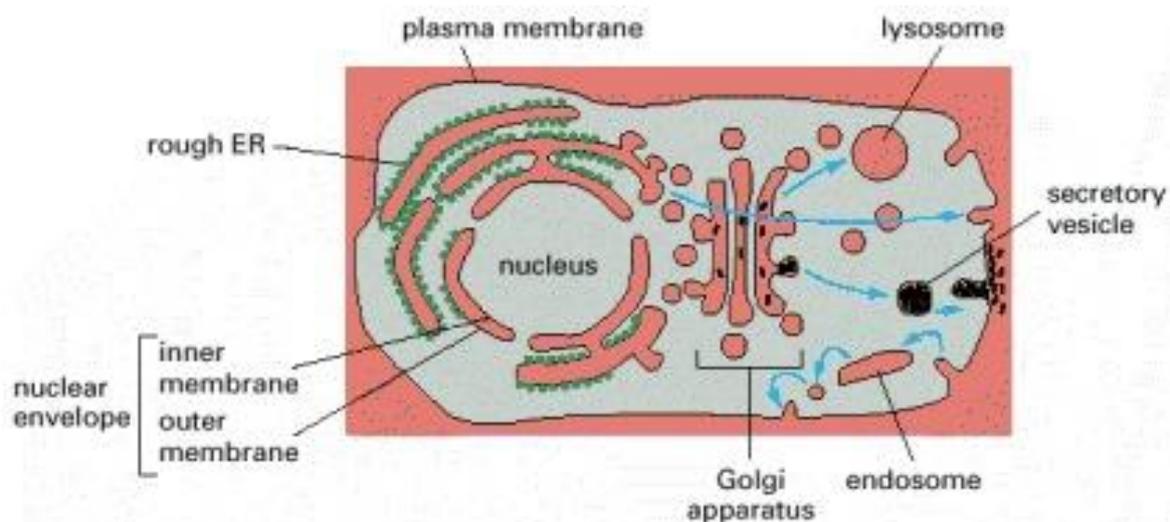
Fig.13.6. Topological relationships between compartments of the secretory and endocytic pathways in a eucaryotic cell Passive and active transport. Source: Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.

Within a membrane bilayer:

● individual lipid molecules diffuse laterally (left) within each of

the two *leaflets* as in a 2D liquid;

→ membrane proteins also diffuse laterally (center);

→ individual lipid molecule may flip over from one leaflet to the other at a very slow rate (right); can be sped up by *flippases*
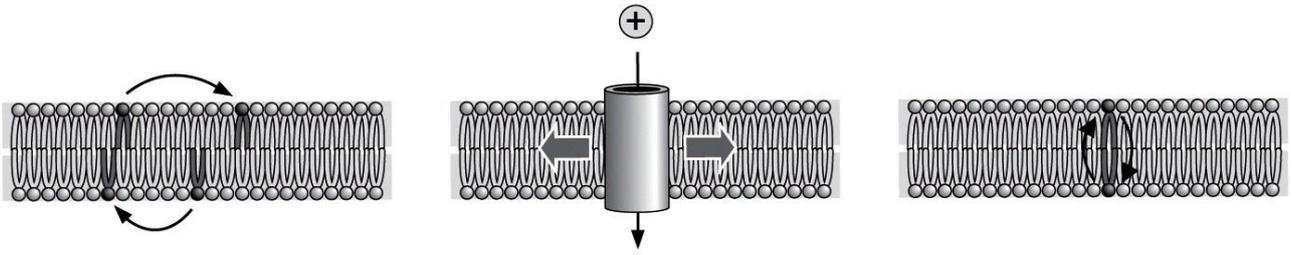
Fig.13.6. Processes occur in membranes Scheme of membrane shape change.

Membranes change shape due to (*i*) thermal motion;

(*ii*) external force; (*iii*) fusion and fission
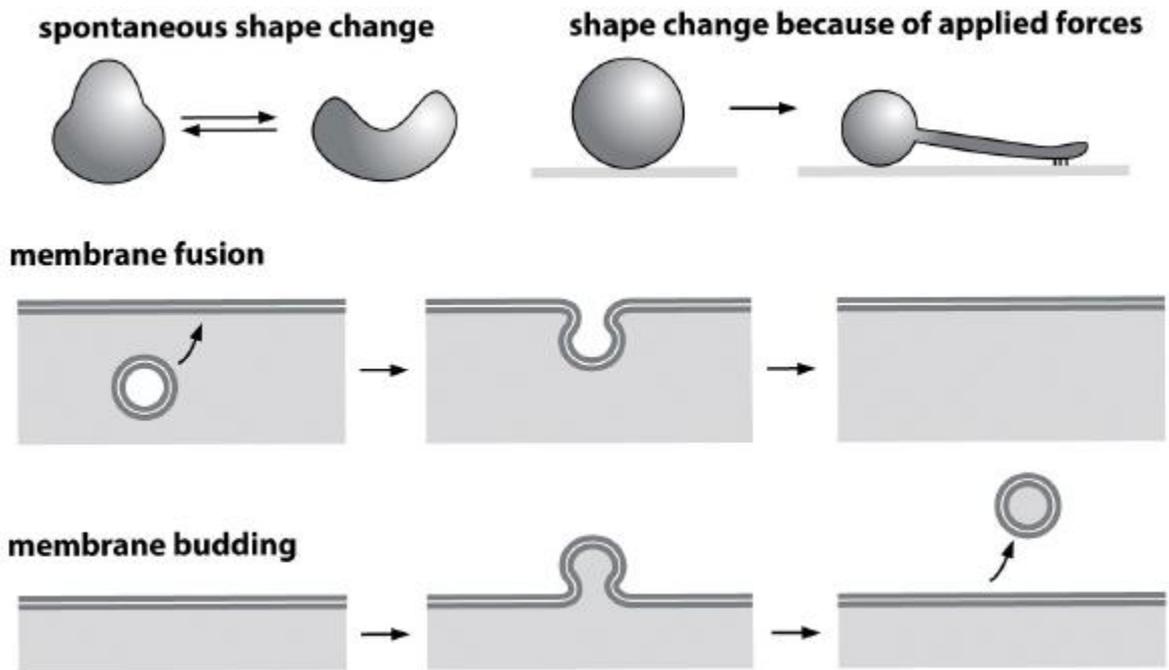
Fig.13.7. Scheme of membrane shape change.

Electron microscopy images of a variety of membranes:

A) *C. crescent us*: membrane layers and cell wall

B) intestinal epithelial cells with dense membrane folds

C) stacks of membranes with photoreceptors in a rod cell

D) mitochondrion surrounded by rough ER
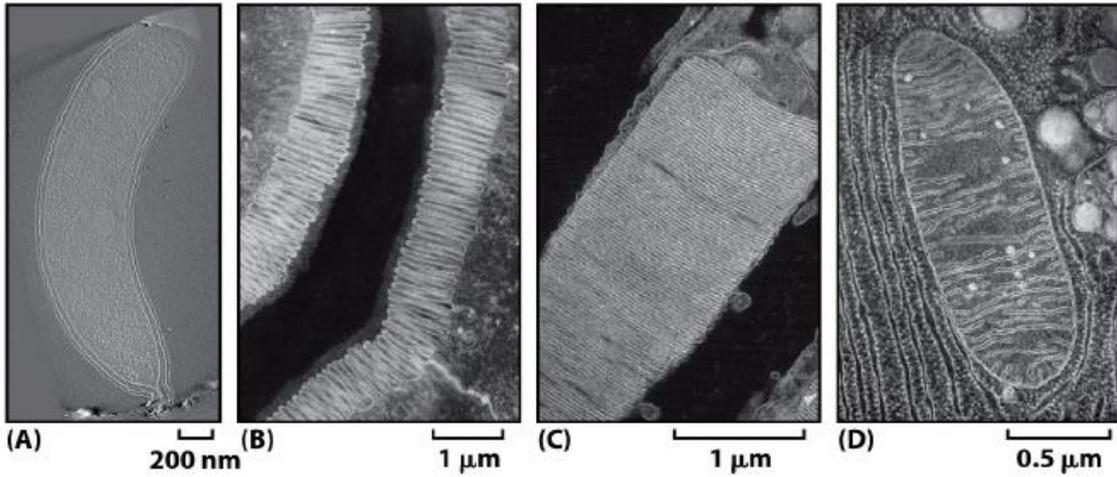
Fig.13.8. Electron microscopy of membranes (scan probe).

Lipids are one of the four basic building blocks of cells. Generic structure of a lipid molecule (*amphipathic*):

→ hydrophilic head (can form HBs with water)

→ elongated hydrophobic domain (fatty acids) Schematic Models of Membranes (right)
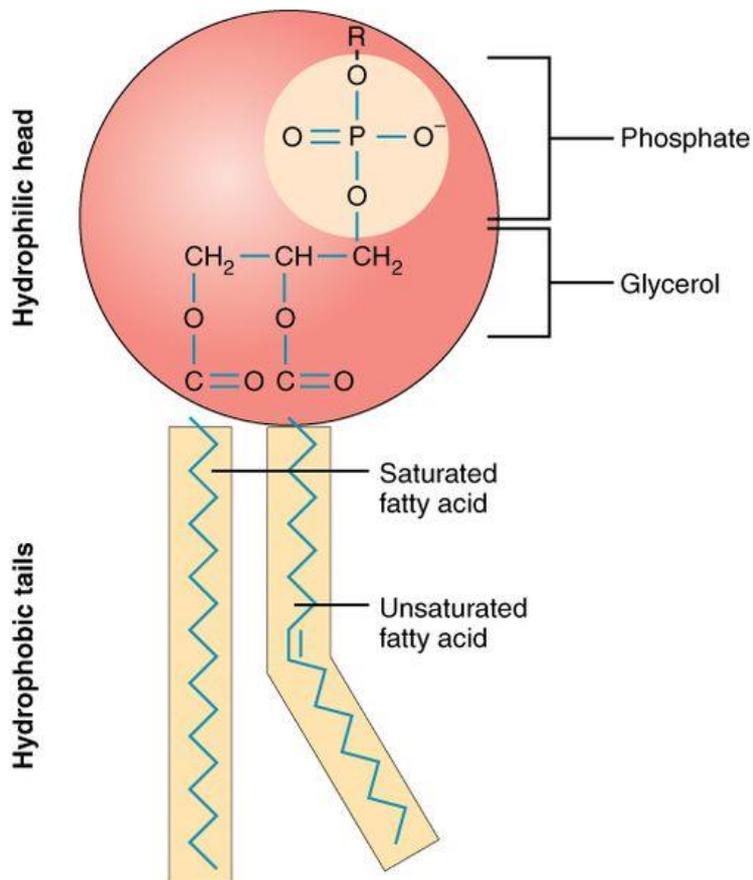
Fig.13.8. Chemical Structure of Lipids.

Fig.13.9. Key Molecules of the Cell Membrane

A phospholipid is built around a *glycerol* (3 carbons and 3 hydroxyl groups).

Fig.13.10. A phospholipid is built around a glycerol (3 carbons and 3 hydroxyl groups).

Fig.13.11. A Sphingolipids.
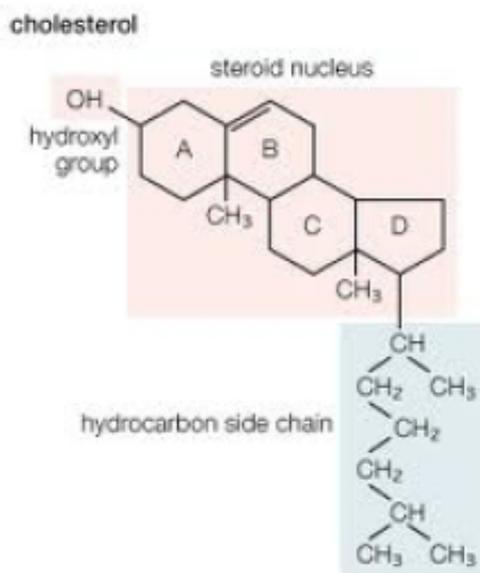
Fig.13.12. A structure of cholesterol.
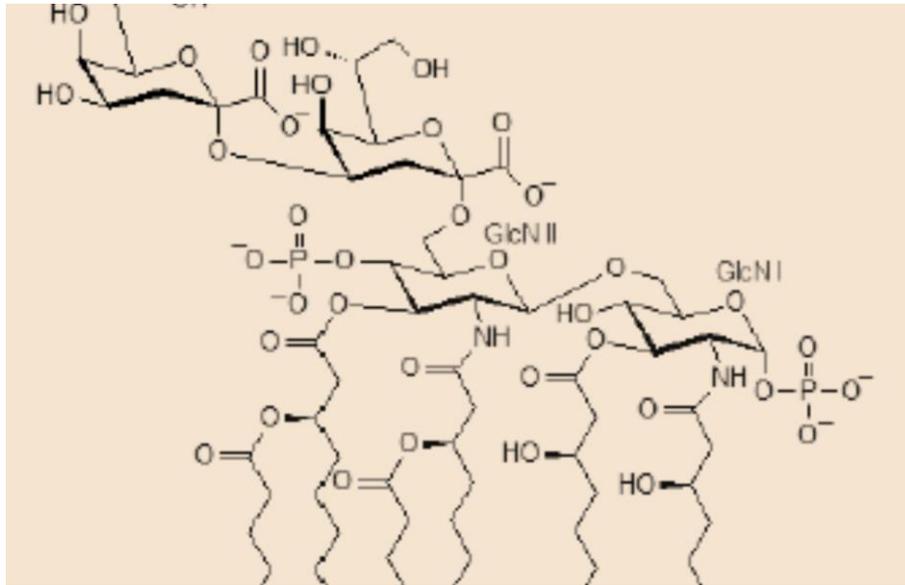
Fig.13.13. The basic Lipopolysaccharide (*E.coli*), incorporating.

The geometric properties of lipid molecules (shape) can induce spontaneous curvature and result in various assemblies

bilayer micelle H$_{II}$-phase
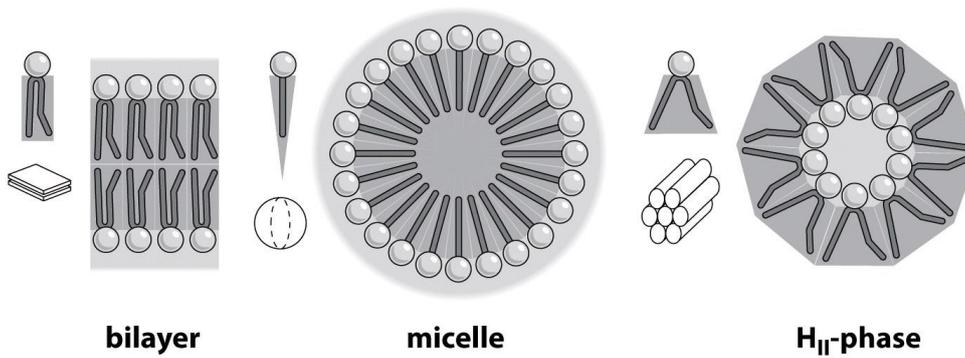
Fig.13.14. The geometric properties of lipid molecules (shape).

5 μm
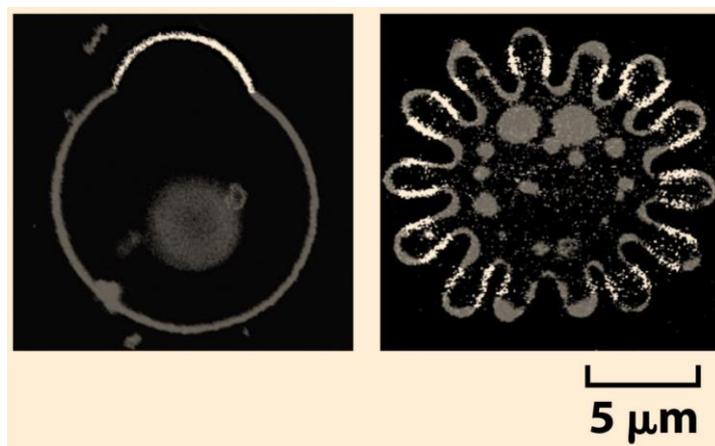
Fig.13.15. Structures of multicomponent vesicles at low and high temperature.

Membrane Heterogeneity:

→in *E. coli*: 1/3 of all proteins are membrane proteins: 10^6 membrane proteins per cell

→in *E. coli*: 2 membranes, each with 500,000 proteins

→membrane area: 6 square micrometer

→area per protein: 12 square nanometer

→average spacing between proteins in a membrane: 3,5 nanometer

→mitochondrial membranes: proteins ~ 70% of the total mass
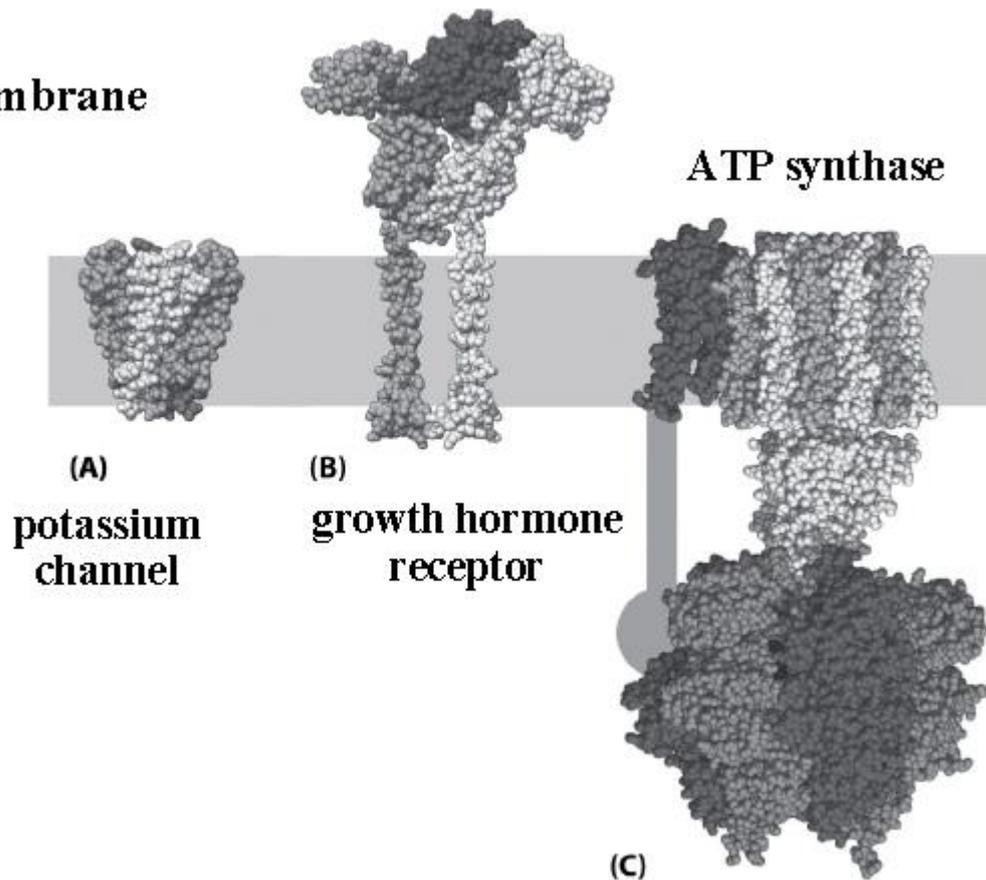
Fig.13.16. Heterogeneity of various membranes.
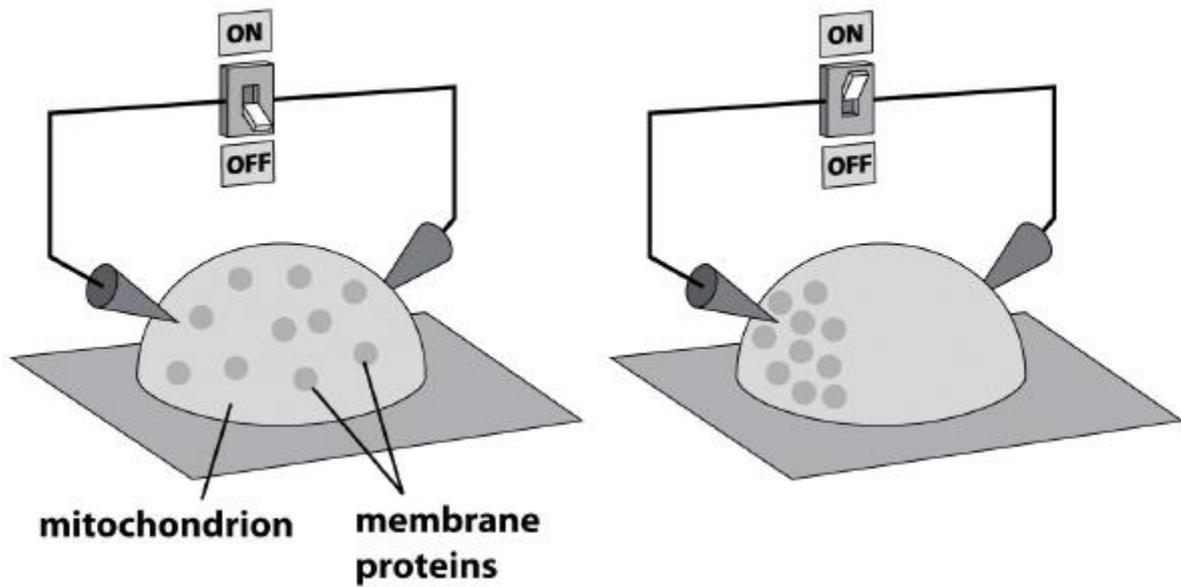
Fig.13.17. Proteins in the Mitochondrial Membrane.
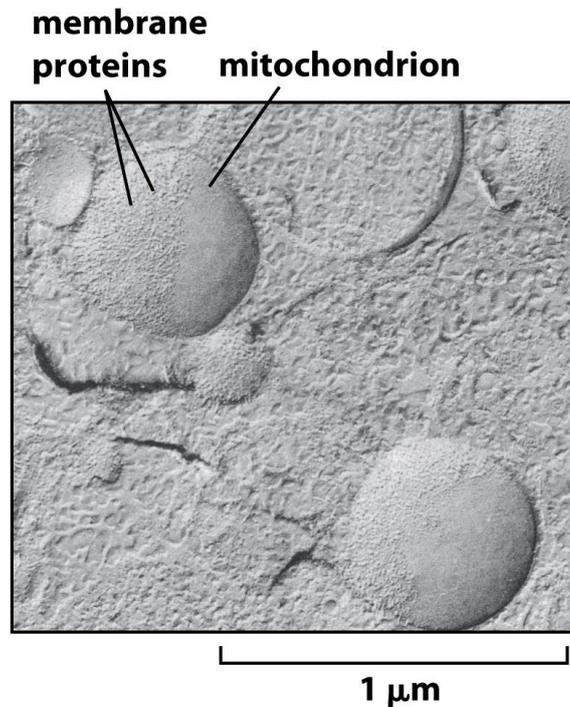
Fig.13.18. The geometric properties of lipid molecules (shape). Freeze Fracture Electron Microscopy.

Membrane proteins help transport mass across the membrane:

→membrane *permeability coefficient P* is defined as following:

$$flux = P(C_{INSIDE} + C_{OUTSIDE})$$

$$FLux \ldots [m^2 s^1] \quad P[m/s]$$

$$C_{INSIDE} \ldots ion\ conc.'inside$$

$$C_{OUTSIDE} \ldots ion\ conc.'inside$$

→permeability depends on the molecule crossing the membrane (right)

-water has the highest permeability

-ions have a rather low permeability

→Transporter proteins act as channels that selectively increase P for the particular ion/molecule
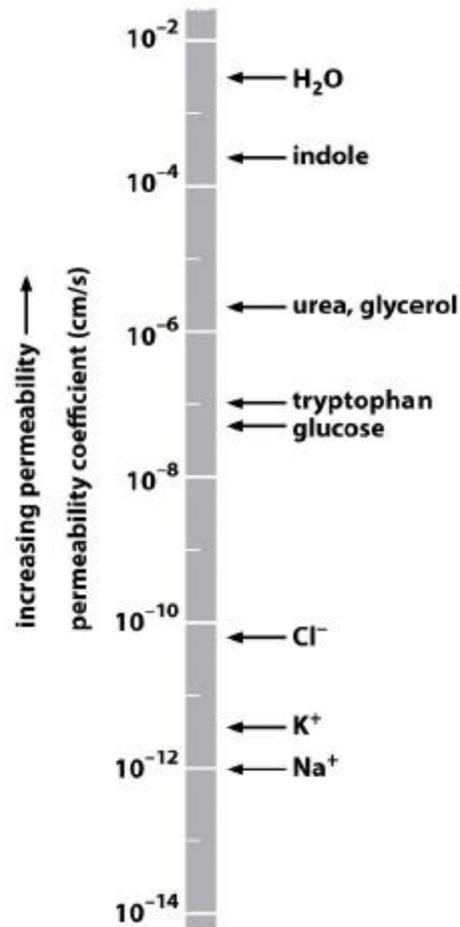
Fig.13.19. The Permeability of various biomolecules.

Growth factor attaches to the receptor extracellularly and initiates microtubule formation inside the neuron (dendrites, axon growth)
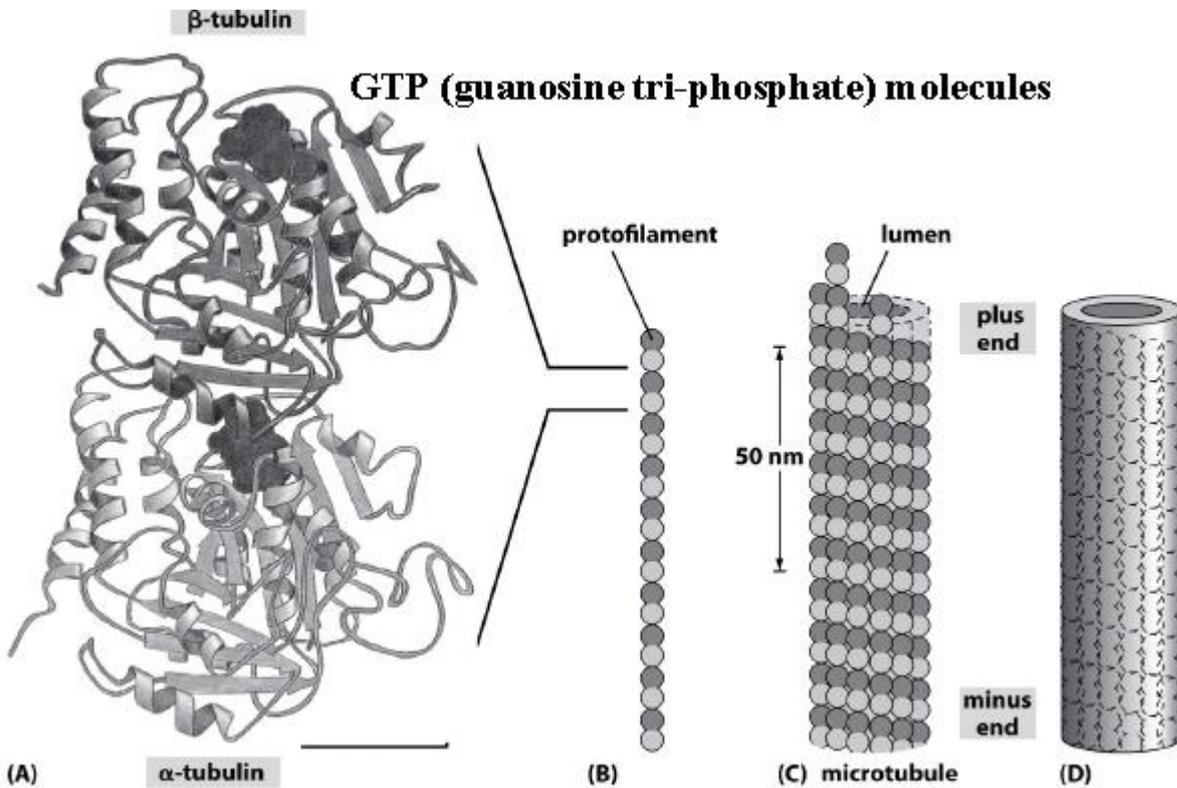
Fig.13.20. Picture of guanosine tri-phosphate biosystems.

Examples of Important Membrane Proteins ATP synthase: a molecular machine that converts a transmembrane H+ gradient into ATP using ADP and inorganic phosphate. Bacteriorhodopsin: pumps protons across the membrane in response to light.

When ATP synthase and bacteriorhodopsin are reconstructed in artificial phospholipid vesicles, they adopt a symbiotic relationship:

- bacteriorhodopsin creates proton gradient using light

- ATP synthase uses the proton gradient to create ATPs

Together, they form a minimalistic green plant that converts light into ATP chemical energy!
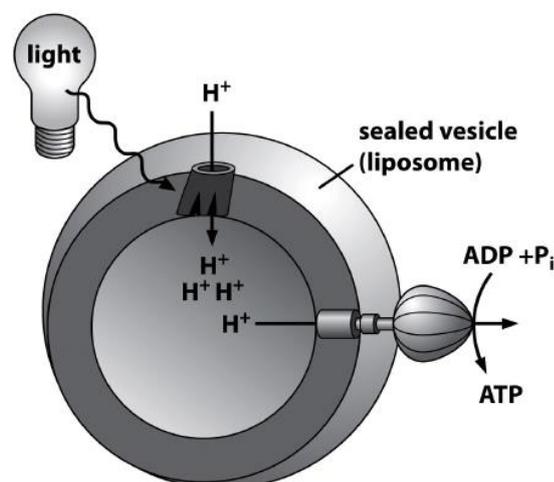
Fig.13.21. Bacteriorhodopsin and ATP synthase in an artificial vesicle.

Four types of membrane deformations:

→ Stretch

→ Bend

→ Thickness Change

→ Shear
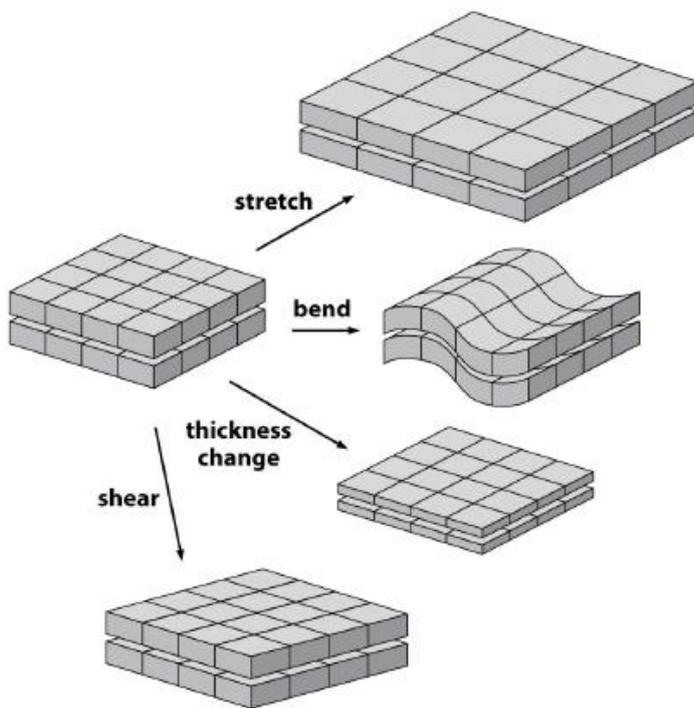
Fig. 13.22. Scheme of four type of membranes deformation.

Mathematical description of stretching, bending, compression, and shear of the membrane

Consider a patch of membrane described in the (x, y) plane:

→ Stretching:

→ Bending:

→ Compression:

→ Shear:

$\Delta a(x, y)$... change of the local area

$H(x, y)$ ...eight of the local area

$W(x, y)$... thickness of the local area

$\Theta(x, y)$... shear angle of the local area
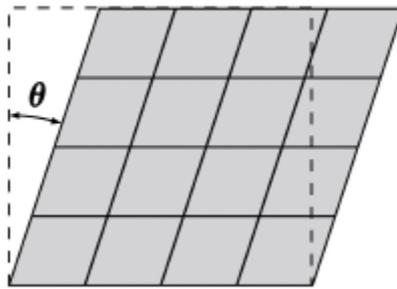
Fig.13. 23. Scheme of mathematical description of membrane's deformation.

## Description of bending geometry by a height function

$$h(x,y)$$
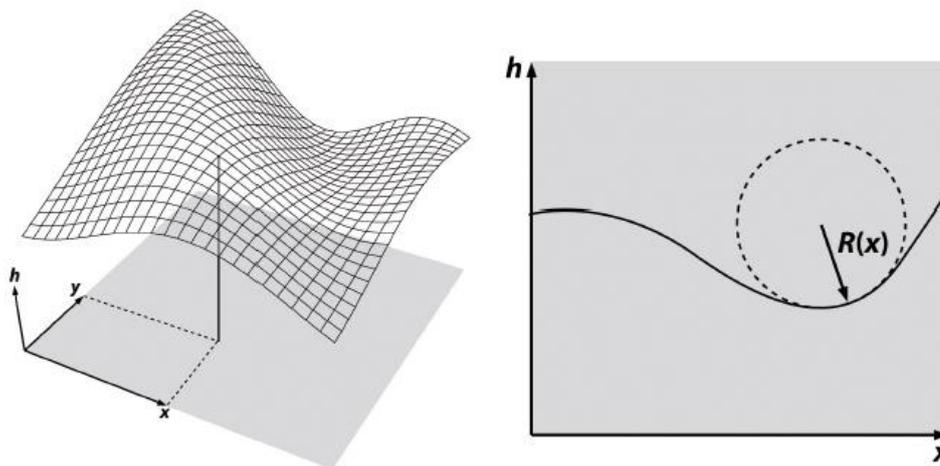
Fig.13. 25. Scheme of mathematical description of membrane's bending geometry.

Local curvature depends on the angle of intersection between the plane and the membrane
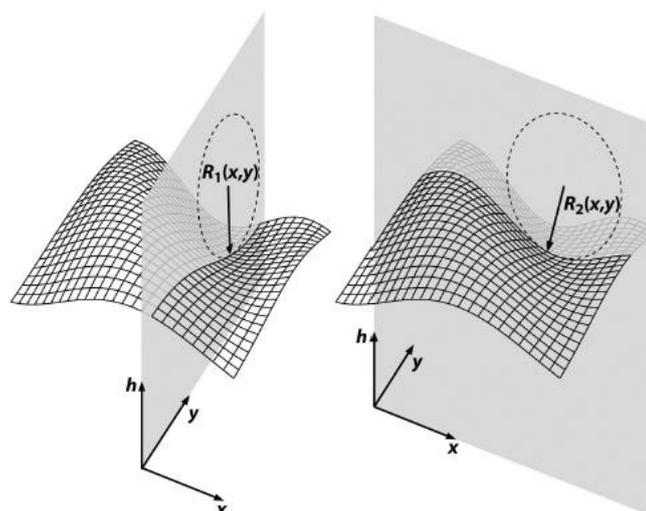
Fig.13.26. Scheme of local curvature depends on the angle of intersection between the plane and the membrane.

How do we calculate a curvature of a membrane at (x, y)?

→ Construct the tangent plane at the point (x, y)

→ Expand the height h in powers of x and y around the origin

$$h(x, y) = k_{11}x^2 + k_{21}yx + k_{22}y^2$$

$$k = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}$$

$$k_{11} = \frac{\partial^2 h}{\partial x \partial x} \quad k_{12} = \frac{\partial^2 h}{\partial x \partial y}$$

$$k_{21} = \frac{\partial^2 h}{\partial y \partial x} \quad k_{22} = \frac{\partial^2 h}{\partial y \partial y}$$

→ the eigenvalues of this matrix are the two principal curvatures and the eigenvectors correspond to the axes of the largest and smallest curvatures (principal values)
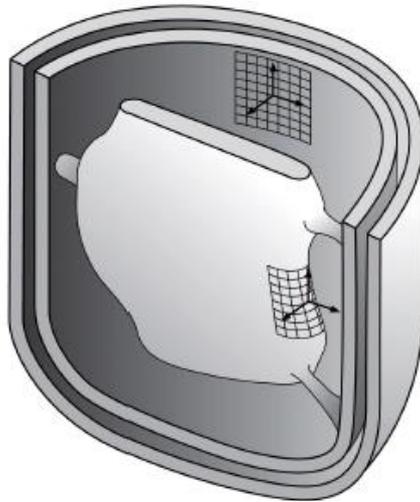

Fig.13.27. Two principal curvatures and the eigenvectors correspond to the axes of the largest and smallest curvatures.

Follow Taylor expansion of $h(x, y)$ around (x_0, y_0)

$$h(x, y) = h(x_0, y_0) + \frac{\partial h}{\partial x} \Delta x + \frac{\partial h}{\partial y} \Delta y$$

$$+ \frac{1}{2} \left(\frac{\partial^2 h}{\partial x^2} \Delta x^2 + 2 \frac{\partial^2 h}{\partial x \partial y} \Delta x \Delta y + \frac{\partial^2 h}{\partial y^2} \Delta y^2 \right)$$

translate
$h(x_0, y_0)$

rotate
$\frac{\partial h}{\partial x} \Delta x + \frac{\partial h}{\partial y} \Delta y$

bend
$\frac{1}{2} \frac{\partial^2 h}{\partial x^2} \Delta x^2 + \frac{\partial^2 h}{\partial x \partial y} \Delta x \Delta y + \frac{1}{2} \frac{\partial^2 h}{\partial y^2} \Delta y^2$
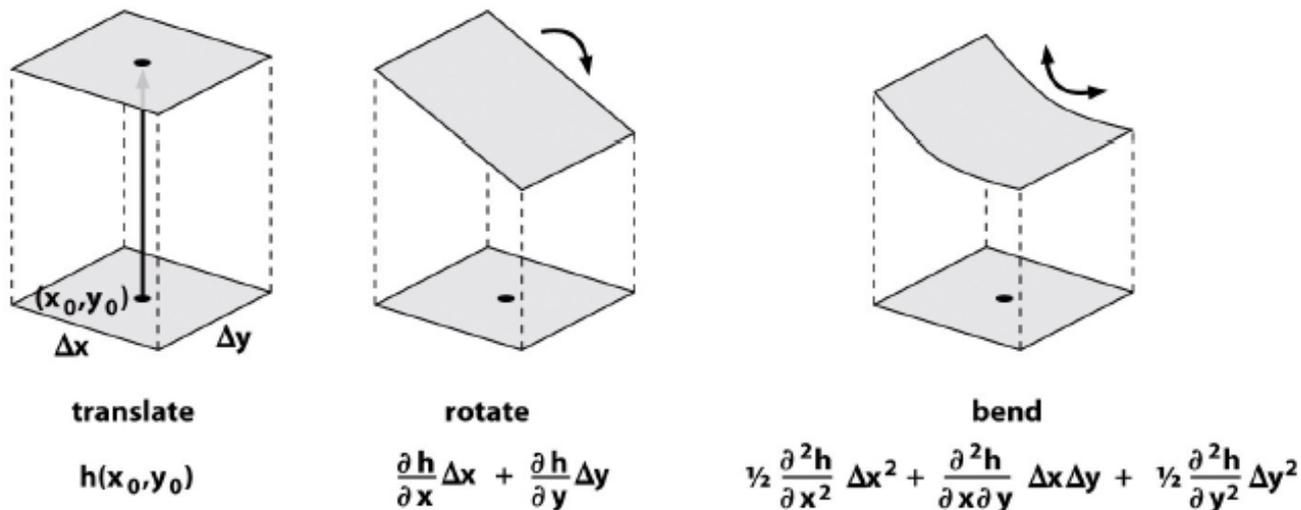
Fig.13.28. Three versions of mathematical modeling of membrane's bending.

Free Energy Penalty due to Area Change

$$G_{stretch} = \frac{K_a}{2} \int \left(\frac{\Delta a}{a_0}\right)^2 da$$

$$K_a \ldots \text{area stretch modulus } \lfloor 55 - 70 \, k_B T/m^2 \rfloor$$
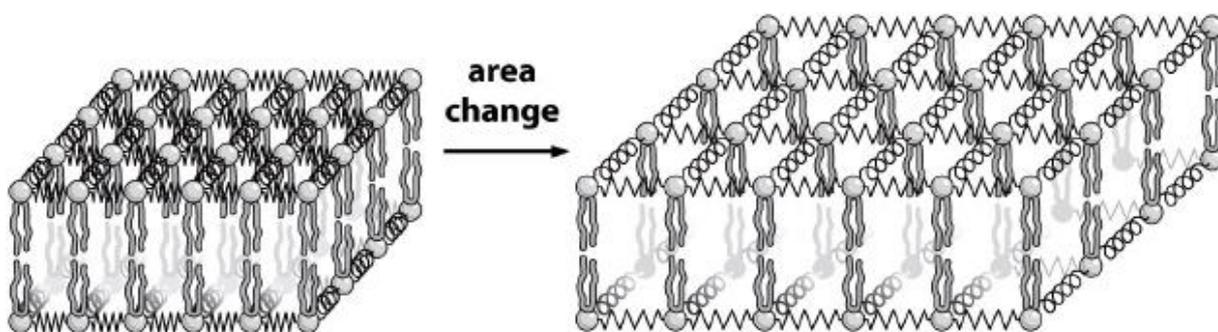

Fig.13.29. Free Energy dependence of area change.

Free Energy Penalty due to Bilayer Bending: Helfrich-Canham-Evans Free Energy

$$G_{bend} = \frac{K_b}{2} \int da \, |\kappa_1 + \kappa_2|^2$$

K_b...bending rigidity $\lfloor 10 - 20 \, k_B T \rfloor$
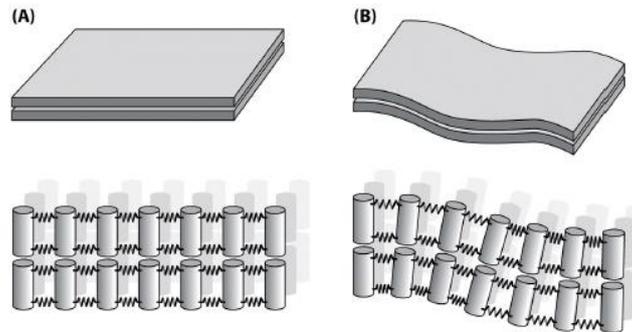
Fig.13.30. The scheme of free energy dependence mechanism from layer's bending.

Free Energy Penalty due to Bilayer Thickness Change

$$G_{compression} = \frac{K_t}{2} \int da \left(\frac{w(x,y) - w_0}{w_0} \right)^2$$

K_t...stiffness $\lfloor 60 \, k_B T / nm^2 \rfloor$
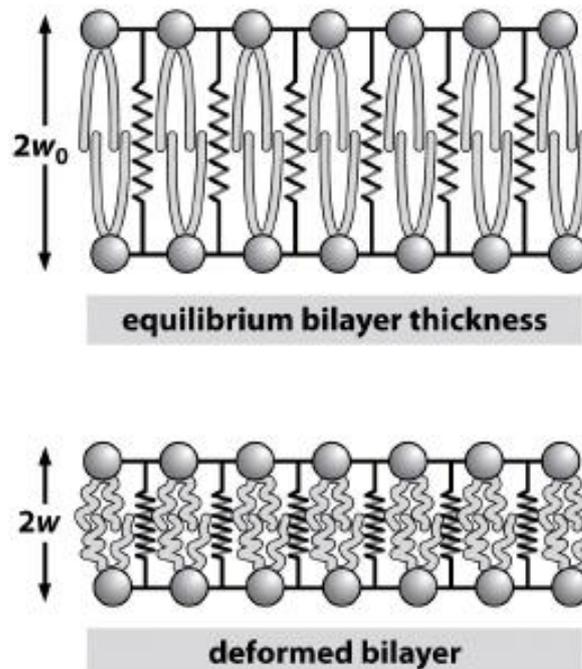
Fig.13.31. The scheme of free energy dependence mechanism from layer's thickness.

Chapter 14

Self-Assembly of Biostructures

Biological complexes are often extremely complicated and it was an important advance when many were found to self-assemble on a molecular level from their 'raw ingredients'. The molecules arrange themselves spontaneously into aggregates without any outside assistance. If the components, the solvent, the pH and the temperature are correctly chosen, the system will minimize its free energy and organize itself in the correct manner. Such strategies for self-assembly have been invented countless times during biological evolution and appear intimately connected with life itself.

There are a diverse range of examples of self-assembling biological systems. In the construction of the tobacco mosaic virus, RNA attaches itself onto 'pie-shaped' coat proteins to produce a rod-like helical virus which is pathogenic to tobacco plants. Similarly, many globular enzymes can self-assemble from their primary structure into fully functioning chemical factories. This is the extremely complicated Levinthal problem referred to the globule-coil transition.

Actin, tubulin and flagellin can self-assemble to provide a force for cellular locomotion. The gelation of hemoglobin in the interior of red blood cells can disrupt the functioning of the cells and gives them a characteristic sickle shape (Figure 14.1); a first example of a self-assembling disease, sickle cell anemia. A further medical condition which is currently the subject of intense research is amyloidosis in prion diseases. Self-assembled beta-sheet amyloid plaques are implicated in a large range of diseases including Alzheimer's, bovine spongiform encephalopathy (mad cow disease) and Parkinson's disease (Figure 14.2).
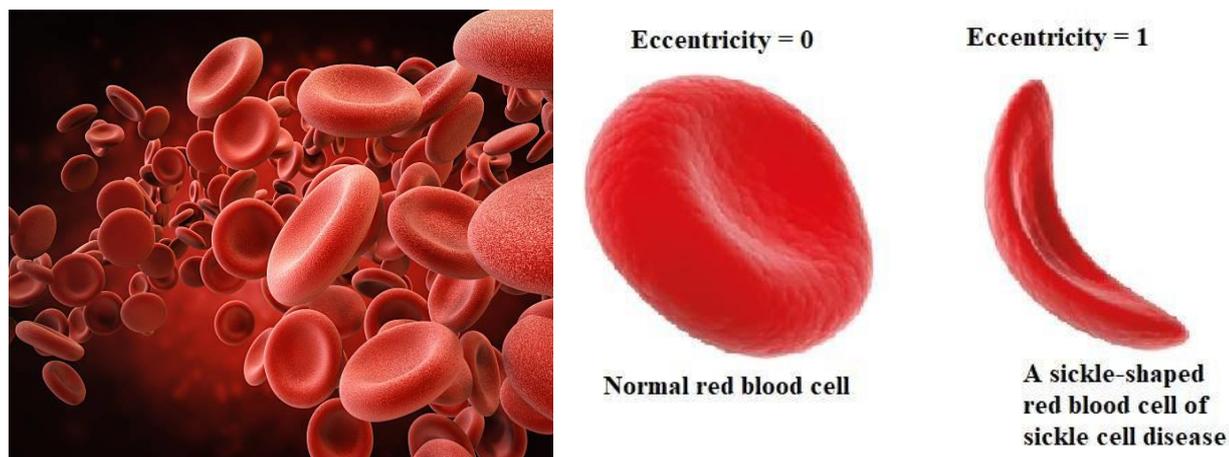

Figure 14.1 Sickle shaped cells are formed when fibrous aggregates of misfolded hemoglobin molecules self-assemble within cells and produce elongated structures on the membrane surfaces. The flow of sickled cells is impaired in the circulation system and this reduces the transport of oxygen in the body (an anemia).

Cell membranes are found to self-assemble from their raw components. Bilayers are easily created synthetically from a range of lipid molecules and spontaneously arrange themselves into vesicles. Naturally occurring cell membranes follow more complicated schemes of construction (their structures include

intramembrane proteins and scaffolding, Figure 14.3); however, the underlying scheme of amphiphilic selfassembly is still thought to hold. Carbohydrates also experience a processof self-assembly; the double helices of starch in plant storage organs are expelled into smectic-layered structures when the carbohydrate is hydrated. A distinction is made between examples of aggregating self-assembly (e.g. micellisation of lipids) and non-aggregating self-assembly (e.g. folding of globular proteins). Aggregating self-assembly has some conceptually sophisticated universal thermodynamic features (e.g. a critical micelle concentration) which are considered in detail in this chapter.

Non-aggregating self-assembly usually describes the behavior of a system that moves between some hidden free energy minima, e.g. the subtle molecular origami involved in the folding of globular proteins.
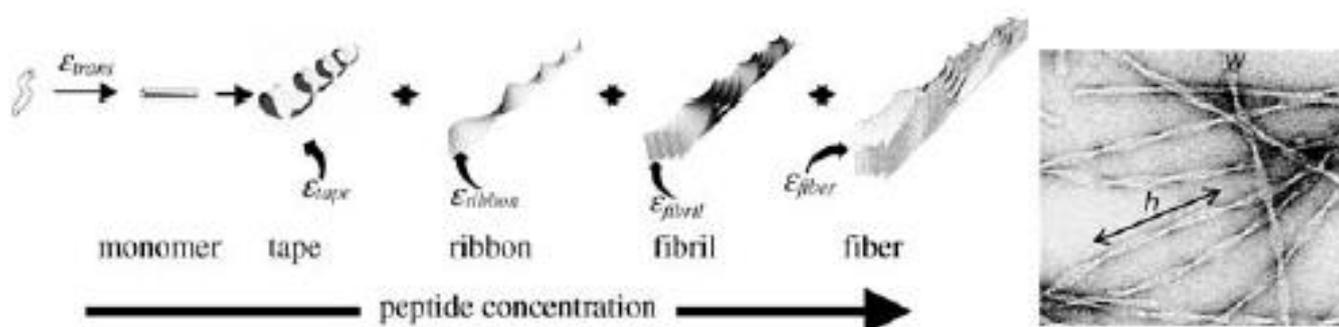

Figure 14.2 Giant self-assembled beta sheet aggregates are thought to be responsible for a range of prion diseases. A hierarchy of structures are found with model peptide systems as a function of peptide concentration. The diameters of the resultant fibers are controlled by the chirality of the peptide monomers. h is the pitch of the self-assembled fiber and w is the width of a fiber [Reprinted with permission from A. Aggeli, I.A. Nyrkova,M.Bell et al., Proceedings of the National Academy of Sciences, 98, 21, 11857–11862, Copyright (2001) National Academy of Sciences]

Other more general examples of self-assembly exist in soft-condensed matter physics, such as the morphologies produced in the phase separation of liquids, liquid-crystals, polymers and block copolymers. All of these have analogues in molecular biophysics. Self-organization is another closely related field of pattern formation in molecular biology and typically is used to describe the results of non-equilibrium thermodynamic processes, e.g. morphogenesis during cell division. The only examples of driven non-equilibrium self-assembly that will be considered are the self-assembly of motor proteins.

To use thermodynamics to describe processes of aggregating self-assembly, the change in free energy (G) of a system due to the exchange of one of its components needs to be considered. The partial molar Gibbs free energy of a biomolecular system with a number of components is given by the symbol μ (the chemical potential). The chemical potential μ_i with respect to one of the species is defined as:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T, P_{\mu i/i}}$$
(14.1)

where n_i is the number of species of type i in the system; the subscripts T; P and n_i on the differential indicate that the temperature, pressure and number of other species are held constant. The total Gibbs free energy (G) of a biomolecular system is the sum of the partial free energies of each of its components:

$$G = \sum_{i=1}^{N} n_i \mu_1 \tag{14.2}$$

where μ_1 are the potentials that drive chemical reactions or diffusion in which changes in the amounts of chemical substances occurred.

Generally the processes of aggregating self-assembly in molecular biophysics have a number of common themes; there exists a critical micellar concentration – a value of the concentration of subunits above which self-assembly occurs (the free monomer concentration is pinned at a single value above this concentration), the entropy change is positive on assembly as the aggregate becomes more ordered (globally the entropy is still maximized due to the increased randomization of associated solvent molecules), hydrogen bonding and hydrophobicity are often an important driving factor, and the surface free energy is minimized as the self-assembly proceeds.

The general features of self-assembly also depend on the dimensionality of the system. Self-assembly in one dimension produces highly polydisperse polymeric aggregates. In two dimensions self-assembly tends to form an aggregate consisting of a single raft and in three dimensions the aggregate is a single micelle or crystal. Self-assembly is driven by the minimization of the surface free energy. In one dimension the reduction in free energy is independent of polymer length, and thus polydisperse aggregates are formed in the self-assembly of single-stranded fibrous proteins and linear surfactant aggregates. Fusion of two surface rafts in two dimensions reduces the surface area and drives the process of coarsening of raft morphologies, eventually leading to the formation of a single giant raft. Similarly, in three dimensions the process of Ostwald ripening causes a gradual increase in aggregate size as small micelles are subsumed by their larger neighbors as they minimize their surface free energy, eventually forming a single giant crystal.

Surfactants

The essential framework of biological membranes in cells is provided by lipid amphiphiles that spontaneously aggregate to form bilayer vesicles (Figure 14.3). The bilayer encapsulates an internal cavity in which the environment for a living cell is maintained (its osmotic pressure, salt concentration and pH). Amphiphilic molecules such as surfactants, lipids, copolymers and proteins can spontaneously associate into a wide variety of structures in aqueous solutions. With naturally occurring lipids the critical micelle concentrations (CMCs) occur at extremely low concentrations and allows stable bilayers to be formed from globally low concentrations of sub-units. Critical micelle concentrations are typically in the range 10^{-2}-10_5 M and 10^{-2}-10^{-9}M for single and double chained phospholipids respectively.
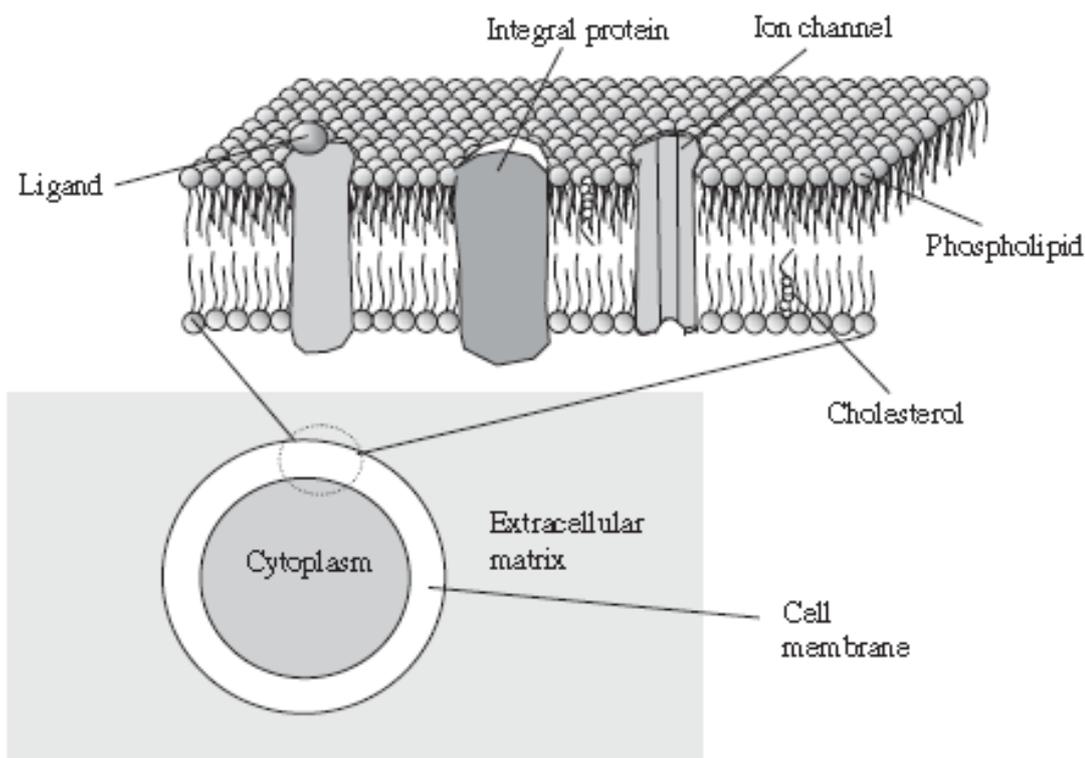
Figure 14.3 Schematic diagram of a cell membrane showing phospholipids, ligands, proteins, cholesterol and ion channels. The cell membrane separates the cytoplasm from the extracellular matrix.

A graphical illustration of surfactant self-assembly is shown in Figure 14.4. Surfactants are partitioned between micelles and unassociated sub-units above the CMC (Figure 14.5). The unusual phenomenon above the CMC is that the monomers are pinned at a fixed concentration due to the thermodynamics of the assembly process and this process will be motivated theoretically in the following section.

Normally surfactants are in dynamic equilibrium with their aggregates during the process of micellar assembly. There is a constant interchange between lipids in micelles and those free in solution (Figure 6.4). The morphology of the aggregates is determined by the geometry of the amphiphilic molecules (the head group area and the length of the tails etc), and the hydrophobicity/hydrophilicity of both the head and tail groups.

There is a strong similarity between Figure 14.5, which indicates the CMC for lipid amphiphiles.

This indicates that a universal thermodynamic process is in action. In equilibrium, thermodynamics requires that the chemical potential of all identical molecules in the different sized self-assembled aggregates are equal. The chemical potential for monomers, dimers, and trimers can thus be equated:
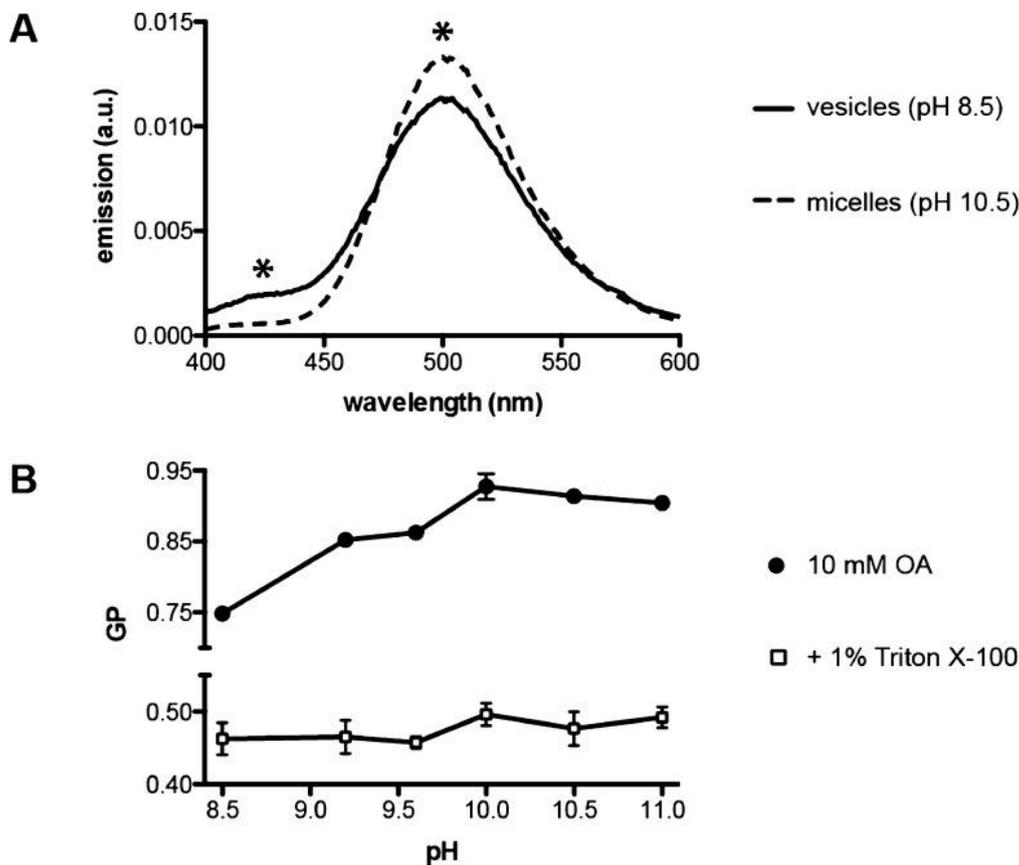
Figure 14.5 Aggregate dependence of Laurdan emission. (A) Emission spectrum for 25 μM Laurdan (excitation 364 nm) in 10 mM oleate at pH 8.5 (vesicles) or pH 10.5 (micelles). Asterisks indicate peaks whose emission intensities are used to calculate GP. (B) Dependence of Laurdan GP on pH in 10 mM oleate with (open squares) or without (closed circles) 1% v/v Triton X100, which disrupts fatty acid aggregates. Error bars indicate SD (n = 3). Source: Itay Budin, Anik Debnath, and Jack W. Szostak. Concentration-Driven Growth of Model Protocell Membranes. J Am Chem Soc. 2012 Dec 26; 134(51): 20812–20819. Published online 2012 Nov 30. doi: 10.1021/ja310382d

$$\mu = \mu_1 + kT log\ c_1 = \mu_2 + \frac{1}{2}kT log\ c_2 = \ldots \ldots \qquad (14.3)$$

and more generally for a N-mer aggregate:

$$\mu = \mu_N^0 + \frac{kT}{N}\log\left(\frac{c_n}{N}\right) = const \qquad (14.4)$$

where N is the aggregation number, kT is the thermal energy, c_N is the concentration of the micellar species of aggregation number N and $m^0 N$ is the standard part of the chemical potential. The rate of association is:

$$k_{1_c}N_1 \qquad (14.5)$$

and the rate of dissociation is $kN\partial cN = NP$: kN is the reaction rate for the Nth order association process, and K is the dissociation constant for the equilibrium process. By definition the dissociation constant is given by:

$$K = \frac{k_1}{k_N} = \exp\left[\frac{N(\mu_N^0 N_1^0)}{rT}\right] \qquad (14.6)$$

Solute molecules self-assemble in solution to form clusters of aggregation number (N) per cluster.

The chemical equation for the association of monomers (A) into aggregates (B) in solution can be expressed as:

$$A + A + A + \ldots = B \tag{14.7}$$

c_A and c_B are defined to concentrations of A and B in mole fraction unitsrespectively, and c is the total concentration of the solute molecules. A general relationship can be obtained between K, N, c and c_A. For largevalues of the dissociation constant $(K > 1)$ and large micellar aggregates $(N > 1)$ it can be shown that the concentration of monomers$c_A$can never exceed $(NK)^{\frac{1}{N}}$. By definition the equilibrium dissociation constant is:

$$K = \frac{c_B}{c_A^N} \tag{14.8}$$

The total concentration of species equals the sum of the concentration of the components $\quad c = c_A + Nc_B$

This allows equation (14.8) to be reexpressed as:

$$K = \frac{(c - c_A)}{Nc_A^N} = const \tag{14.9}$$

Equation (14.9) can be rearranged to give:

$$c_A = \left[\frac{c - c_A}{NK}\right]^{\frac{1}{N}} \tag{14.10}$$

The maximum possible value of$c$, $c_A$is 1, since the calculation is in fractional molar units, e.g. when $c_A \sim 0$; $c_A \sim 1$. Therefore c_A cannot exceed $(NK)^{\frac{1}{N}}$. This was shown graphically on Figure 14.5. If a large value is taken for the dissociation constant $(K = 10^{80}$ and a reasonable value for the aggregation number $(N = 20)$ is chosen the critical concentrationc_A is $0,86 \times 10^4$. Substitution in equation (14.10) gives:

$$c_A = 10^4 \left[\frac{c - c_A}{20}\right]^{\frac{1}{20}} \tag{14.11}$$

Detailed analysis of this equation shows that for c $<10^4 c_A \sim$ c, whereas for c $\sim 10^4 c_A \sim Nc_B$, and there is an equal partition between micelles and unimers. Thus $(NK)^{\frac{1}{N}}$ is the critical micelle concentrationfor this process of self-assembly. The process of one-dimensional aggregation can now be considered in detail.αkT is defined to be the monomer–monomer 'bond' energy of the linear aggregate relative to isolated monomers in solution. The total interaction free energy $(N\mu_N)$ of an aggregate of N monomers is therefore (terminal monomers are unbonded):

$$N\mu_N = (N - 1)\alpha kT \tag{14.12}$$

This can be rearranged as:

$$\mu_N = \left(1 \quad \frac{1}{N}\right)\alpha kT \qquad (14.13)$$

and can be written in the equivalent form:

$$\mu_N = \mu_\infty + \frac{\alpha kT}{N} \qquad (14.14)$$

Thus αkT decreases asymptotically towards μ_∞ the bulk energy of an extremely large aggregate ($N \to \infty$). In two - dimensional aggregation the number N of molecules per disc like aggregate is proportional to the area πR^2. The number of unbonded molecules in the rim of the disc is proportional to the circumference $2\pi R$ and hence $N^{\frac{1}{2}}$

This implies that the free energy of an aggregate is therefore:

$$N\mu_N = \left(N \quad N^{\frac{1}{2}}\right)\alpha kT \qquad (14.15)$$

and the free energy per molecule in an aggregate is:

$$\mu_N = \mu_\infty + \frac{\alpha kT}{N^{\frac{1}{2}}} \qquad (14.16)$$

where $(\mu_\infty = \alpha kT)$ is again the free energy per particle of an infinitely large aggregate.

For spherical, three dimensional aggregates, N is proportional to the volume $4/3\pi R^3$ and the number of unbonded molecules is proportional to the area $4\pi R^2$, and hence $N^{\frac{2}{3}}$. Therefore the total free energy of an aggregate is:

$$N\mu_N = \left(N \quad N^{\frac{2}{3}}\right)\alpha kT \qquad (14.17)$$

This can be rearranged and the free energy per particle is:

$$\mu_N = \mu_\infty + \frac{\alpha kT}{N^{\frac{1}{3}}} \qquad (14.18)$$

where $\mu_\infty = \alpha kT$ is the free energy per particle of an infinitely large aggregate. It can be shown that for a spherical micelle the proportionality constant (α) is related to the surface tension (γ) and the size of the aggregate (R):

$$\alpha = \frac{4\pi R^2}{kT} \qquad (14.18)$$

In general the CMC for the aggregation of surfactants is given by the exponential of the difference in chemical potentials for a monomer and an aggregate:

$$CMC \approx e^{\frac{(\mu_1 \mu_N)}{kT}} \qquad (14.20)$$

Therefore, for three dimensional aggregates the CMC is:

$$CMC \approx e^{\frac{\alpha}{N^{\frac{1}{3}}}}$$
(14.21)

where α is given by above mentioned equation (14.19.)

Viruses

The self-assembled geometrical structure of heptatitis B determined by X-ray crystallography measurements is shown in Figure 14.6 a. This virus is pathogenic to humans, and the self-assembly (reproduction) of such parasites is of vital importance to medical science. The general process of self - assembly in viruses is thought to originate from an interplay between shorter range hydrophobic and longer range electrostatic forces. However, the details of the mechanism of self-assembly can be very complicated and are specific to the particular variety of virus that is considered.
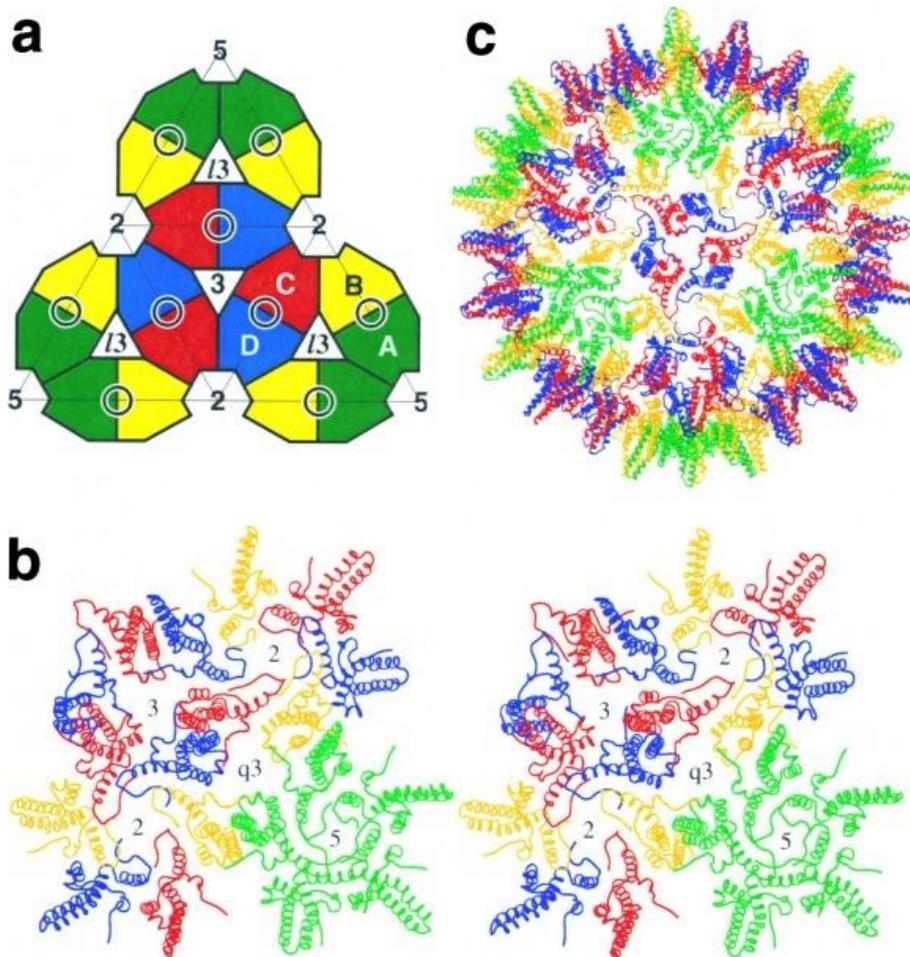
Fig. 14.6: Packing of Dimers in the T = 4 Capsid

(a) A schematic representation of the arrangement of the four independent subunits in the T = 4 capsid, showing one icosahedral face. The subunit nomenclature follows Zlotnick et al. 1996. The positions of the four-helix bundles are shown by black circles.

(b) Stereo view of the packing of the dimers in the T = 4 capsid in the region of a quasi three-fold axis. Subunits A, B, C, and D are shown in green, yellow, red, and blue as in Böttcher et al. 1997. The positions of the icosahedral five-fold (5), three-fold (3), quasi three-fold (q3), and two-fold (2) axes are shown.

(c) The capsid viewed down an icosahedral three-fold axis. a- Graphical representation of a hepatitis B virus based on X-ray crystallography.

Source: S.A Wynne, R.A Crowthe, A.G.W Leslie, The Crystal Structure of the Human Hepatitis B Virus Capsid. Molecular Cell Volume 3, Issue 6, June 1999, Pages 771-780

Tobacco mosaic virus (TMV, Figure 14.7) is one of the simplest helical viruses known and consists of a single strand of RNA surrounded by_2000 identical pie shaped protein subunits. TMV has been a favorite topic of research for physicists, since it is not pathogenic to man and presents an ideal monodisperse rod-like system for studying liquid crystalline phases.

Assembly of TMV in vitro can occur with and without a chain of RNA. Without RNA molecules, the protein monomers of TMV first form double disks of 17 monomer units. The disks contain holes at their centres. If the pH is changed appropriately, it modulates the electrostatic interactions of the disks and they slip with respect to each other and aggregate. The protein disk like sub-aggregate units have a' lock washer' morphology and slowly stack upon each other to form rods with a high polydispersity in their length. With RNA molecules the nucleic acid chain directs the growth of disk aggregation, a monodisperse virus is formed, since the RNA dictates a well defined length for the helical virus.
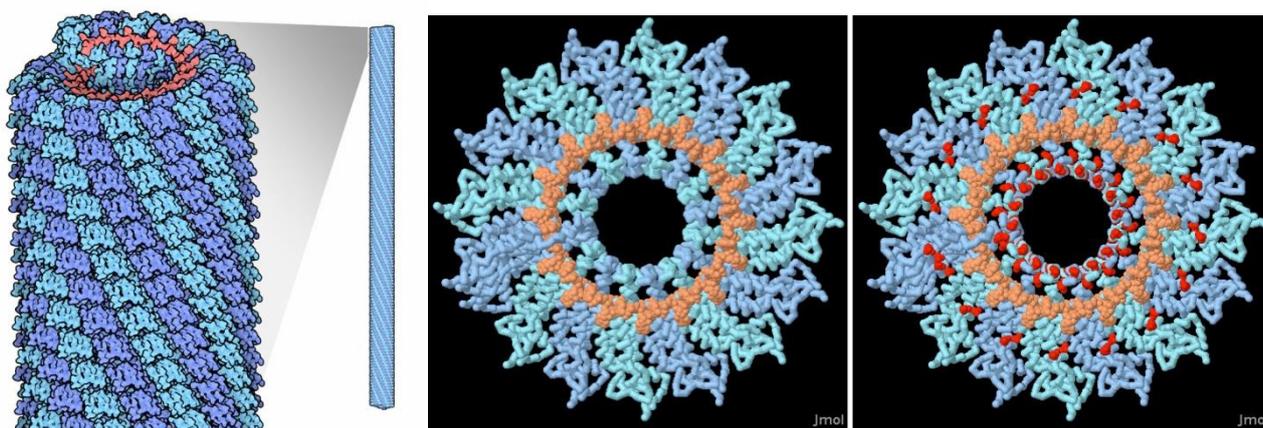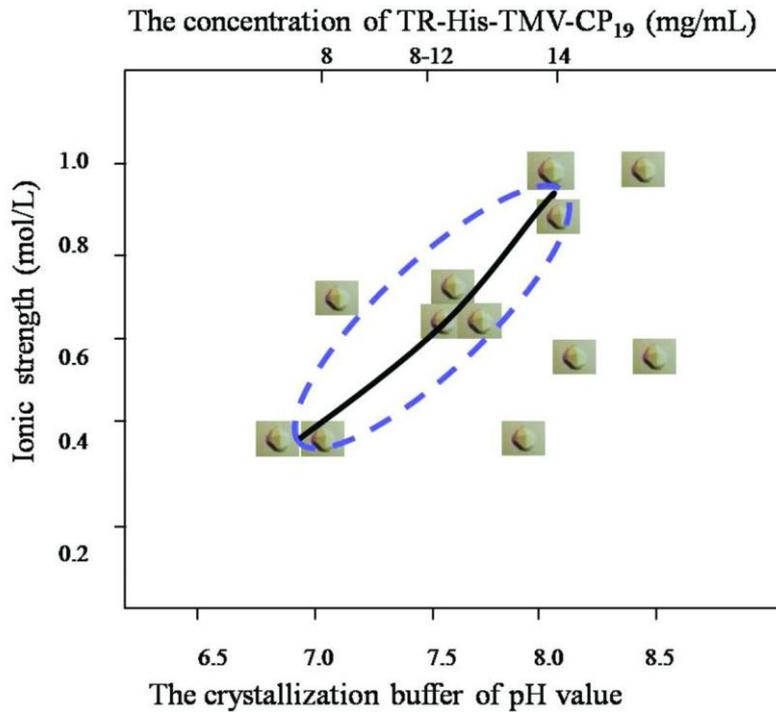
Figure 14.7 (a, b) Tobacco mosaic virus self-assembles from 'lock washer' units that become attached to a central RNA core. (c) The ionic strength and pH can switch the process of assembly on and off. The self-assembly of a complete TMV virus is favoured at low pH values.

Many other viruses consist of a nucleic acid core surrounded by asymmetrical shell that is assembled from identical protein molecules (icosohedral viruses such as hepatitis B, Figure 14.6). There are geometrical selection rules for the symmetry of the arrangement of the identical coat proteins. The process of self-assembly in these cases is often much more complicated than with TMV. The self-assembly is sometimes directed by chaperone proteins that guide the process. The complexity of the steps involved in the self-assembly and life cycle of the P4 icosohedral virus is illustrated in Figure 14.8.
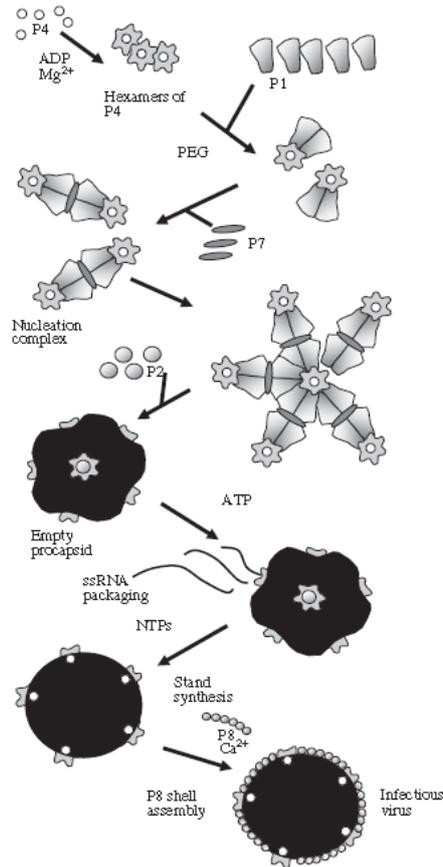
Figure 14.8. Schematic diagram of the sophisticated process of self-assembly found in the P4 virus [Ref.: D.H. Bam ford, Phil. Trans. R. Soc. Lond. A, 2003, 361, 1187–1203]

The T4 bacteriophage is one of several DNA containing viruses that can infect E. Coli bacteria (Figure 14.9) and the process of self-assembly is again slightly more sophisticated than TMV. Separate sections of this virus have been observed to self-assemble from their constituent components.

The tail tube forms spontaneously from the core proteins and purified base plates (Figure 14.10). Starting only with purified base plates and core protein monomers the tail tube self-assembles in vitro to a length of ~100 nm.
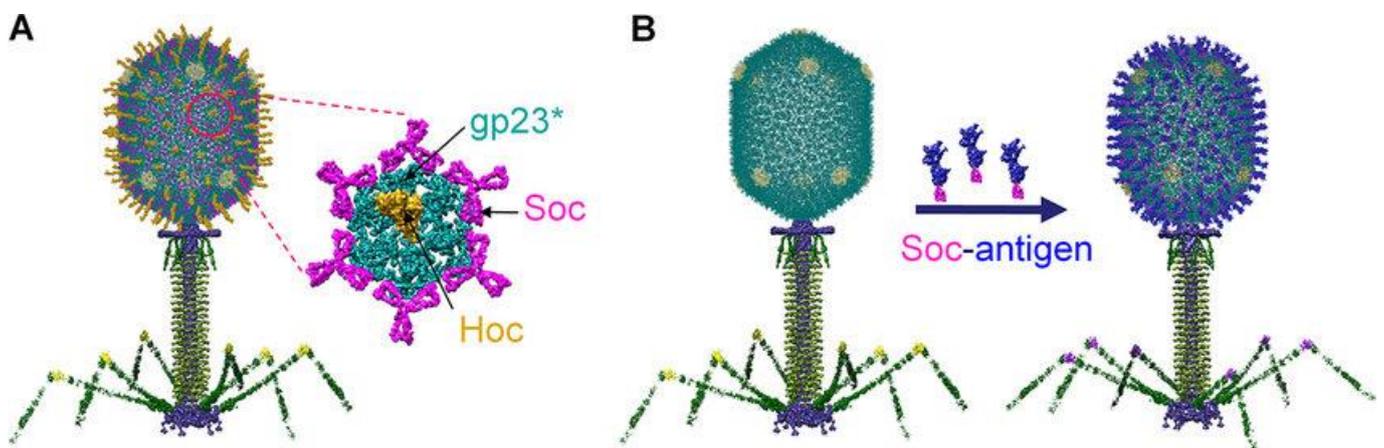

Figure 14.9 Schematic diagram of the self-assembled structure of T4 bacteriophage (The tail tube is found to spontaneously self-assemble in vitro (Figure 14.10).)

221
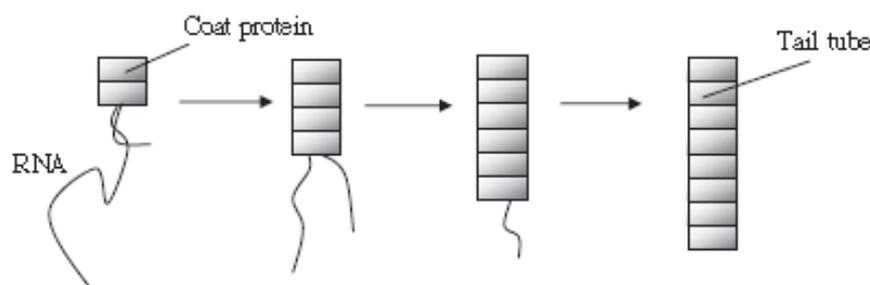
Figure 14.10 Self assembly of the coat proteins of T4 bacteriophage on a single RNA chain template

Self-assembly of proteins

Two important examples of aggregating protein self-assembly in medical conditions are the aggregation of proteins in amyloid diseases (Alzheimer's, Bovine Spongiform Encephalopathy etc., (Figure 14.2) and the aggregation of hemoglobin molecules in sickle cell anemia (Figure 14.1). There are a series of unusual features found during the aggregation of filamentous proteins in vitro which are indicative of a process of self assembles opposed to a conventional chemical reaction. The temperature dependence of the process of polymerization is very different to that of small inorganic species. At low temperatures there is no polymerization and at high temperatures polymerization occurs; normally with synthetic polymerization reactions the reverse is true. A rise in pressure causes depolymerization; behavior opposite to that normally found with covalent bonds, e.g. polymerization of polyethylene. Polymerization of self-assembling proteins only occurs above a critical initial monomer concentration, the CMC of the self-assembly process, as in the case of surfactant self-assembly. The kinetics of protein polymerization are characterized by a long lag period followed by rapid formation of polymers. This is of particular concern in prion diseases, since the resultant self-assembled amyloid aggregates are implicated in these fatal conditions.

The change in Gibbs free energy (ΔG) during polymerisation is givenby the standard thermodynamic equation:

$$\Delta G = \Delta U + P\Delta V \qquad T\Delta S \qquad (14.22)$$

where ΔU is the change in free energy, P is the pressure, ΔV is the changein volume, T is the temperature and ΔS is the change of entropy. The polymerization reaction is found to proceed move favourably at high temperatures. From equation (14.22) this signifies that the entropy changes that occur upon polymerization is positive. This is allowed thermodynamically since the entropy of the associated solvent molecules increases to compensate for the entropy of the protein aggregates which decreases as they self-assemble into ordered structures.

The mechanism of non-aggregating self-assembly that forms the internal structure of globular proteins intimately associated with the globule–coil, helix–coil and beta sheet–coil phase transitions. With natural

proteins there is the additional complication of frustration during folding which exists due to the large number of closely spaced local minima explored by the protein on its free energy landscape (Levinthal's paradox). This frustration can lead to misfolded proteins if they are improperly chaperoned to their final active states or unfolded by chemical/physical denaturants. The misfolding of a range of proteins is thought to be the nucleation step in a number of amyloid diseases and constitutes a rate limiting step in the development of such conditions.

Polymerisation of cytoskeletal filaments (motility)

The polymerisation of cytoskeletal polymers is an important example of self-assembly, since it can lead to motility with actin polymerisation and has thus been intensively researched. Single stranded polymerisation is found to be an unlikely mechanism for the construction of long fibres dueto the surface free energy effects described in the introduction to this chapter: 1-D aggregates are short and polydisperse). Actin circumvents this problem by using two interacting (double helical) strands. Similar schemes hold for a range of other helical cytoskeletal filaments, e.g. tubulins.

The addition of a monomer unit to an actin fibre is a process of selfassembly and the equilibrium morphology adopted is a balance between the rate of addition k_{on} and dissociation k_{off} of a globular proteinsubunit:
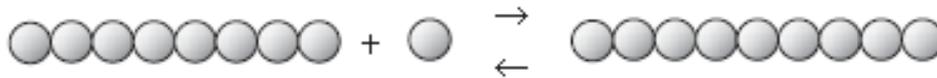

Figure 14.11 Self-assembly of filamentous proteins. (The equilibrium constant is the same for each unit added to the polymer for a single filament that consists of symmetric subunits.)

$$A_n + A_1 \underset{k_{off}}{\overset{k_{on}}{\leftrightarrow}} A_{n+1} \tag{14.23}$$

where A_n is an n-mer aggregate and A_1 is a monomer. There is a dissociation constant (K) attributed to this process of monomer addition on to a n-mer aggregate:

$$\frac{c_n c_1}{c_{n+1}} = K = \frac{k_{off}}{k_{on}} \qquad n > 1 \tag{14.24}$$

where $c_{n,} c_1$ and c_{n+1} are the molar concentrations of monomer, n-mer and $n+1$-mer aggregates respectively. It is assumed that all the individual subunit addition reactions for a length of n-mer have the same dissociation constant (K), dissociation rate constant (k_{off}) and second order association constant (k_{on}) However, the dissociation constants (K) are not held fixed and change with the length of the

polymeric aggregate. The equilibrium dissociation constant is given by the standard thermodynamic relationship:

$$K = exp\left(\frac{\Delta G_0}{kT}\right) \tag{14.25}$$

The dissociation constant (K) is thus associated with the standard free energy change ΔG_0 of the reaction. G_0 is the sum of the potential energy (negative) associated with the formation of a monomer–protein bond and an entropy change (positive) associated with the loss of translational and rotational entropy as the subunits are transferred from the standard 1M concentration in solution to a bound state in a polymeric aggregate {equation (14.22)), at constant volume and pressure, $\Delta V = 0$}.

As argued qualitatively in the introduction it is found that single stranded self-assembled filaments are polydisperse and short. This result can be deduced more formally by calculating the dependence of the average length of the filamentous aggregates on the total concentration of subunits. The dissociation constant for the reaction is given by equation (14.24). The concentration of different lengths of polymers (c_n) are assumed to follow an exponential distribution given by:

$$c_n = K exp\left(\frac{n}{n_0}\right) \tag{14.26}$$

where K is the equilibrium constant as defined by equation (14.25), n is the length of the polymer and n_0 is defined as:

$$n_0 = \frac{1}{ln\alpha_1} \tag{14.27}$$

where α_1 is given by:

$$\alpha_1 = \frac{c_1}{K} \tag{14.28}$$

The exponential distribution can be proven by substitution in equation (14.24), but to calculate the average number of monomers in a filament (n_{av}) requires more effort. It is:

$$n_{av} \approx \sqrt{\frac{c_t}{K}} \tag{14.29}$$

Where c_t is the total concentration of monomers and K is the equilibrium constant. This result can be derived from the exponential distribution of the n-mer aggregate concentration. The contribution from the monomer lengths is discounted and the definition of an independent probabilistic average gives:

$$n_{av} = \sum_2^\infty np_n = \sum_2^\infty n\frac{\alpha_n}{\sum_2^\infty \alpha_n} = 1 + \frac{1}{1}\frac{}{\alpha_1} \tag{14.30}$$

where an are the statistical weights given by $\alpha_n = \frac{c_n}{K}$, and from equation (14.26) these are given by:

$$\alpha_n = \alpha_1^n = e^{n/n_0} \tag{14.31}$$

The total number of subunits (α_C) is the algebraic sum:

$$\alpha_t = \sum_2^\infty n\alpha_n = \frac{\alpha_1}{(1 \quad \alpha_1)^2}$$

$$\alpha_1 = 1 + \frac{1}{2\alpha_t}\sqrt{\frac{1}{\alpha_t} + \frac{1}{4\alpha_t^2}} \tag{14.32}$$

When $\alpha_C > 1$, the expression for a1 simplifies to:

$$\alpha_1 \approx 1 \quad \alpha_t^{1/2} \tag{14.33}$$

From equation (6.27) and using the Taylor expansion $ln(1+x) = x_1 + \frac{x^2}{2} + \frac{x^3}{3}\ldots$ gives:

$$n_0 \approx \sqrt{\alpha_t} \tag{14.34}$$

Finally, using equation (14.30) gives the expression:

$$n_{av} \approx 1 + \sqrt{\alpha_t} \tag{14.35}$$

which is equivalent to equation (14.29).

A similar type of analysis shows that multi-stranded filaments tend to be very long. The problem is slightly more complicated because the geometry of the fibre imposes three separate dissociation constants K; K_1 and K_2 on the process of assembly (Figure 14.12). The average length of multi-stranded filaments is found to be given by:

$$n_{av} \approx \sqrt{\frac{K_1}{K}}\sqrt{\frac{c_t}{K_2}} \tag{14.36}$$

where c_t is the total monomer concentration. This average length is typically much bigger than found with single stranded filaments. Filaments of actin and microtubules can polymerise and are sufficiently long for their roles in motility because they are multi-stranded. Calculations show that the lengths of the polymeric aggregates are again distributed exponentially, but the average filament length is much greater than predicted by equation (14.36).
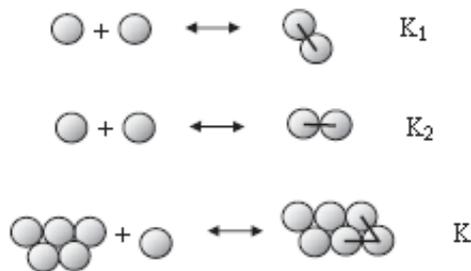
Figure 14.12 Three association constants are required to describe the self-assembly of a two stranded filament formed from symmetric subunits (K; K_1 and K_2)

This is due to the active nature of the motility process with actin, i.e. ATP (fuel) drives the reaction and the process is not in thermodynamic equilibrium. The rate of elongation (dn/dt) for filamentous self-assembly is given by the Oosawa equation:

$$\frac{dn}{dt} = k_{on}c_1 \qquad k_{off} \qquad (14.37)$$

The graphical solution is shown in Figure 14.13(a). The partitioning of the monomers between those free in solution and those in the filamentous state that defines the CMC for one dimensional self-assembly of the cytoskeletal fibers is shown in Figure 14.13(b).
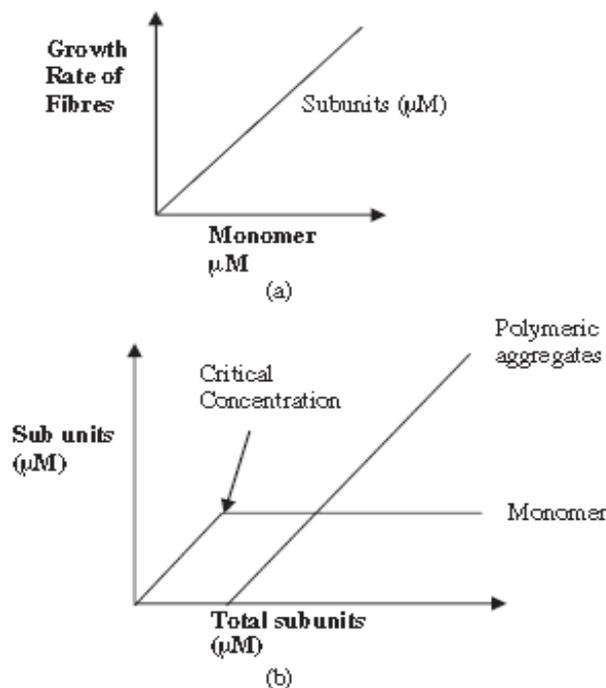
Figure 14.13 The generic self-assembly of filamentous protein aggregates: (a) shows the growth rate of aggregates as a function of the monomer concentration and (b) indicates how the concentration of subunits is partitioned between monomers free in solution and those in fibrous aggregates.
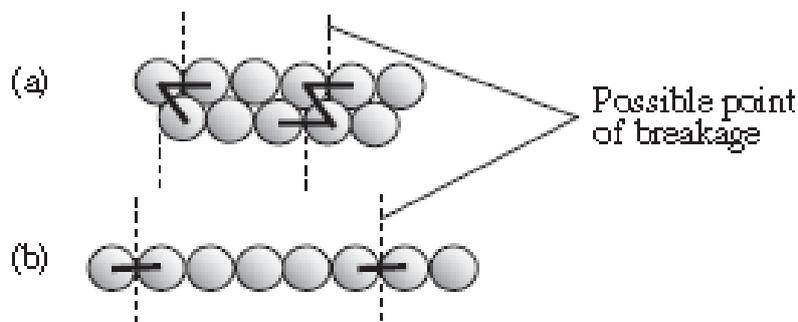

Figure 14.14 Multi stand filaments (a) are stable against breakage whereas single stranded filaments (b) continually break and reform along their length

226

There is a close resemblance to the equivalent diagram for surfactant self-assembly (Figure 14.5) and this emphasizes the general features involved in both processes. Extensive theoretical development of the assembly of multi-stranded fibres shows that multi-stranded filaments grow and shrink at both their ends, and there again exists a critical concentration for self-assembly.

Multi-stranded filaments are stable to breakage whereas single stranded filaments continually break and reform (Figure 14.14). Other biological examples which follow the same trends in their self-assembly as actin (Figure 14.13) are sickle cell hemoglobin aggregates (double helical fibres) and amyloid aggregates (they often are twisted chiral multi-tape fibres). The critical concentration (c_c) for self-assembly is given by the minimum in the rate of addition ($\frac{dn}{dt} = 0$)and substitution in the Oosawa equation (14.37) gives:

$$c_c = \frac{k_{off}}{k_{on}} = K \qquad\qquad (14.38)$$

This equation provides a method of determining the dissociation constant experimentally from a plot of the monomer concentration against the total concentration of sub-units. It is found that the concentration of nuclei are very small during the self-assembly of multistranded fibres, the ends of the filaments are blunt due to the extra stability of a snug geometrical fit in these systems, and the mean lengths of the filaments increase very steeply above the critical concentration. Thus slight changes in free monomer concentration give a large change in polymer length for multi-stranded filaments in equilibrium (Figure 14.15).

Many fibrous biopolymers are self-assembled through the intermediate step of protofilaments, in contrast to the mechanism of the direct addition of globular protein sub-units to the end of a fibrous aggregate seen with actin and microtubules. Examples of protofilament assembly are shown in Figure 14.16. Such mechanisms are found to be important in a range of proteins, including the collagens and the intermediate filaments such as the keratins and desmins. Protofilament assembly provide another mechanism for these systems to circumvent the problems with one dimensional self-assembly, such as single filament polydispersity and low size, and enables the construction of giant fibrous networks.


Mean Polymer Length (subunits)

Total Subunits (μM)
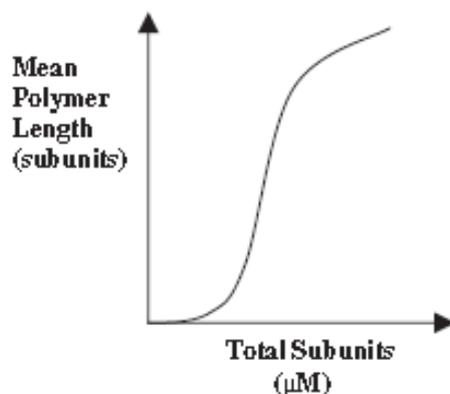
Figure 14.15 The mean length of a self-assembled two stranded filamentous aggregate chains is an S-shaped function of the number of subunits


(a) Collagen
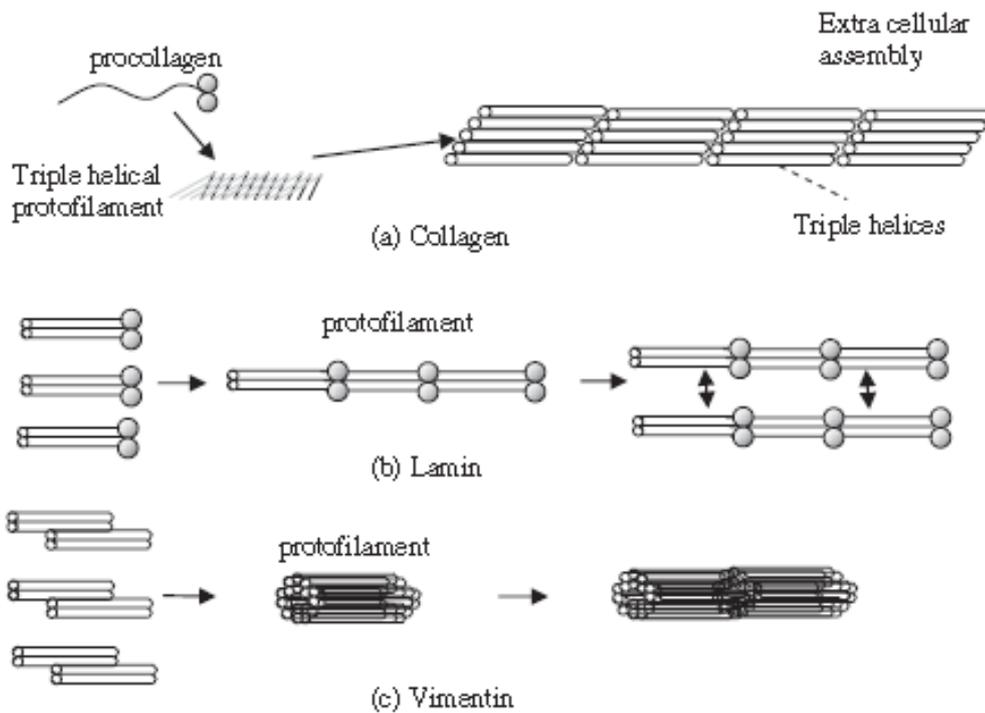
(b) Lamin

(c) Vimentin

Figure 14.16. Collagen, lamin and vimentin fibres form from extended protofilament units, in contrast to actin and tubulin fibres that self-assemble directly from small spherical subunits [Ref.: H. Herrmann, U. Aebi, Annu. Rev. Biochem., 2004, 73, 749–789]
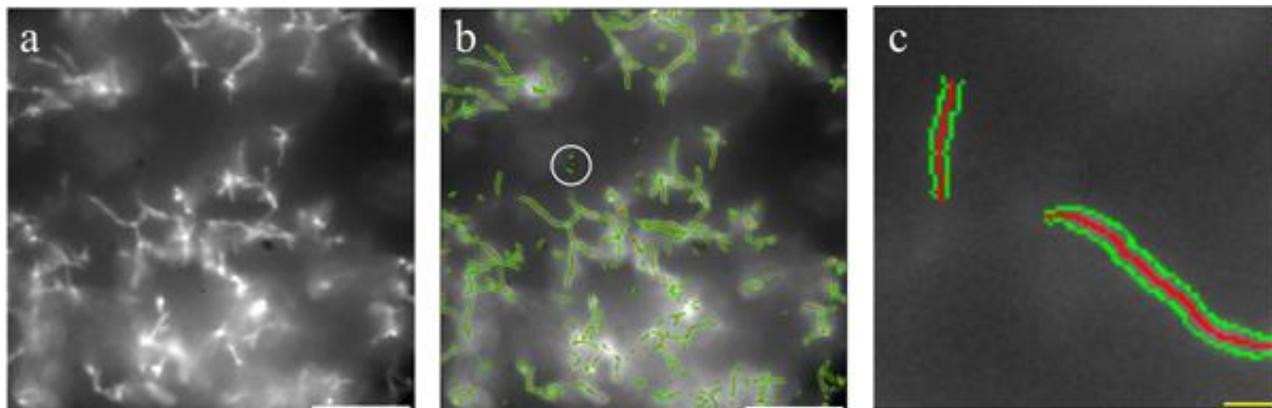

Figure 14.17.. Ridge detection algorithm used to determine bundle width. (a) Gray scale image of 6HT at 16 μM DNA concentration and 5% PEG content. (b) Analysis of the gray scale image. Source: Martin Glaser Jörg Schnauß, Teresa Tschirner. Self-assembly of hierarchically ordered structures in DNA nanotube systems. Published 29 April 2016 • © 2016 IOP Publishing Ltd and Deutsche Physikalische Gesellschaft.

Chapter 15

Experimental Methods and Techniques

A vast range of methods experimental techniques is used to analyze the structure and dynamics of biomolecules. A subset of methods that emphasize the physical behavior of biological molecules is examined here and reference should be made to more specialized texts for detailed descriptions of analytical biochemical methods. The discussion of nuclear magnetic resonance, terahertz, ultraviolet, infra-red, mass and Raman spectroscopies is avoided, since they require too much space to be covered satisfactorily. Only the mechanical form of spectroscopy (rheology)–where the sample is hit with a mechanical perturbation and its response in time is observed–is examined in detail here.

Historically, the primary methods for structural determination of biological molecules have been high resolution probes such as scattering (neutrons, X-ray, light, elastic/inelastic) and microscopy (light and electron).

There are many good accounts of standard scattering techniques in the biophysical literature and microscopy is well described in undergraduate optics textbooks. It is hard to beat the discussion in Cantor and Schimmel for detail on the process of scattering from biomolecules, so instead, after a brief introduction, some modern developments in the field of scattering, such as quasi-elastic scattering, micro focus scattering and coherent diffraction microscopy, are covered. In addition, methods of single molecule force measurement, osmometry, sedimentation, tribology, solid mechanics and electrophoresis are explained to give a modern emphasis to the subject.

Scattering Techniques

The field of scattering encompasses a vast range of fundamental physical processes and techniques. The basic geometry of a scattering experiment is shown in Figure 15.1. Incident radiation or particles interact with the sample and are deflected through an angle (θ). The momentum transfer (q) of the scattering process is simply related to the reciprocal of the length scale (d) probed:

$$q = \frac{2\pi}{d} \tag{15.1}$$

For a particular form of radiation the momentum transfer for an elastic (energy is conserved) scattering process can be calculated from:

$$q = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right) \tag{15.2}$$

where λ is the wavelength. The use of momentum transfers rather than scattering angles allows the results of experiments with a range of different forms of radiation (X-rays, light, neutrons etc.) to be compared easily. The varieties of radiation typically used in biological scattering experiments are shown in Table 15.1, which includes the wavelength of the radiation and the length scales in the sample that can be typically probed. Specialised detectors and optics are required for each different form of radiation.

$$|q_i| = |q_s| = \frac{2\pi}{\lambda}$$

$$q = \frac{4\pi}{\lambda}\sin\left(\frac{\theta}{2}\right)$$
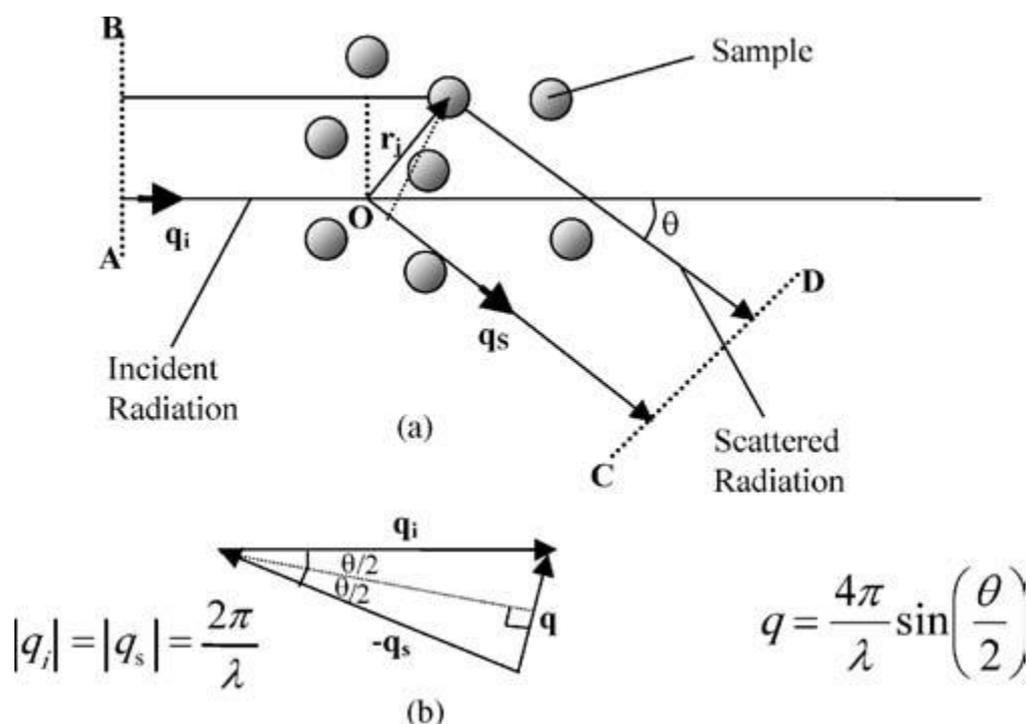
Figure 15.1 (a) The elastic scattering processes involved with a wide range of radiation (X-rays, light, neutrons and electrons) can all be understood using the same schematic diagram. (b) The momentum transfer can be calculated from the wavelength and scattering angle using a simple geometrical construction

Technique	Typical wavelength (nm)	q values (nm^{-1})	Real space distances (nm)	Sample contrast	Comments
Small angle X-ray scattering	0.15	0.009–6.284	1–700	Electron density	Beam damage can be a problem
Small angle neutron scattering	0.4	0.0031–6.284	1–200	Scattering length density	Samples can be labelled with deuterium
Light scattering	450	0.0003–0.126	50–2000	Refractive index	Multiple scattering is a problem at high sample concentrations (>2%)
Wide angle X-ray scattering	0.15	6.284–62.84	0.1–1	Electron density	Beam damage can be a problem
Wide angle neutron scattering	0.4	6.284–62.84	0.1–1	Scattering length density	Useful to measure hydrogen bonds
Electron scattering	0.0037	0.006284–6.284	1–1000	Electron density	Very thin sections are required due to strong Coulombic interaction with samples

Table 15.1. Comparison of scattering techniques applied to biological systems.

There are a number of novel techniques that utilize the coherence of electromagnetic radiation and in particular that of X-rays. Two promising new coherent methods are X-ray diffraction imaging and quasi-elastic x-ray scattering. Images of completely aperiodic magnesia stained bacteria have now been

reconstructed using coherent X-ray diffraction with 100 nm resolution and quasi-elastic X-ray scattering offers dynamic measurements from soft matter systems with unprecedented sensitivity to length scale.

A further modern advance with X-ray and neutron scattering is the introduction of effective focusing techniques. Focusing of X-rays is routinely made to submicron levels at third generation synchrotron sources and micron sized beams can be made with laboratory based microfocus sources. Such beams can be rastered across heterogeneous biological materials and the molecular structure probed as a function of position on the sample (Figure 15.2). New focusing devices have now been made to create micron sized beams that include capillaries, Fresnel lens, mirrors and even simple compound lens (e.g. lenticular holes in a block of aluminum). Such scanning X-ray microdiffraction techniques are helping to revolutionize the field of fibrous carbohydrates and proteins.
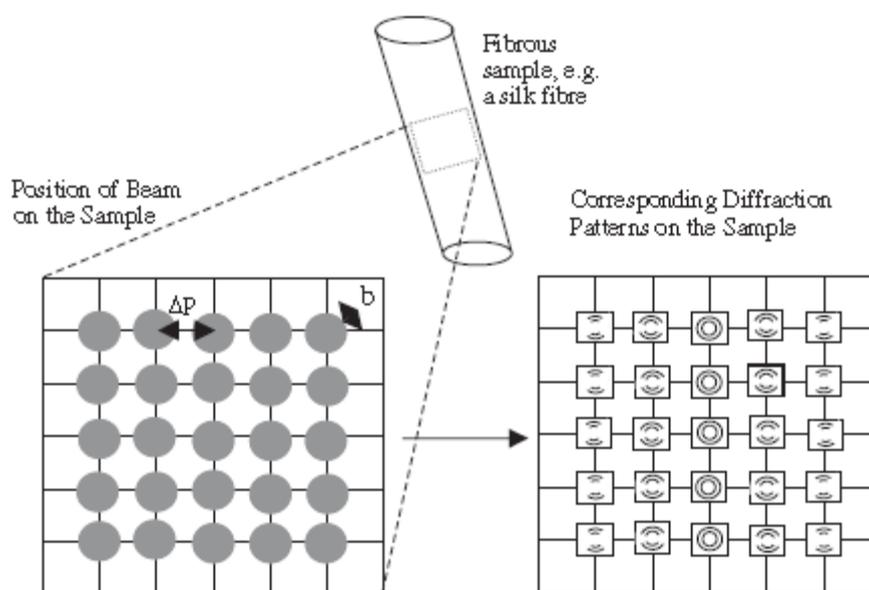
Figure 15.2 Scanning X-ray microdiffraction across a semi-crystalline anisotropic Material (b is the diameter of the X-ray beam. Two dimensional diffraction images are recorded at micron spaced steps (DP) across the sample and provide detailed information on the molecular structure of the fibres.

The contrast that is measured during a scattering process can be varied in both neutron and X-ray scattering experiments. With neutrons isotopic substitution can be used to label biomolecules through the replacement of hydrogen atoms with chemically equivalent deuterium atoms. This labelling scheme is particularly attractive when electronically light atoms need to be located in a crystalline structure, e.g. the elucidation of the structure of hydrogen bonds. With X-rays the wavelength of the radiation can be matched to the absorption edge of a heavy atom that exists in a biological structure, and the contrast can be varied to elucidate both the crystalline and solution state structures with much improved resolution, e.g. through the method of anomalous small-angle X-ray diffraction.

Dynamic Scattering Techniques

Once the structure of a biological sample is well understood, quantifying the dynamics of the components of the material poses some important questions. A challenge is to examine the dynamics of the material without perturbing the sample morphology. With soft biological materials the dynamics can be studied quantitatively with scattering methods by observing the time decay of stimulated emission (fluorescence techniques) or by measuring the change in energy of the scattered particles (quasi elastically or inelastically scattered).

Fluorescence intensity correlation spectroscopy is a modern example of a scattering technique that has been adapted to single molecule experiments. Fluorescent probes are added to the biological molecules whose dynamics are of interest (there is a vast range of possible ways to do this and large commercial catalogues exist of the available fluorescent probe molecules) and made to fluoresce using a tightly focused laser beam under an optical microscope (Figure 15.3).
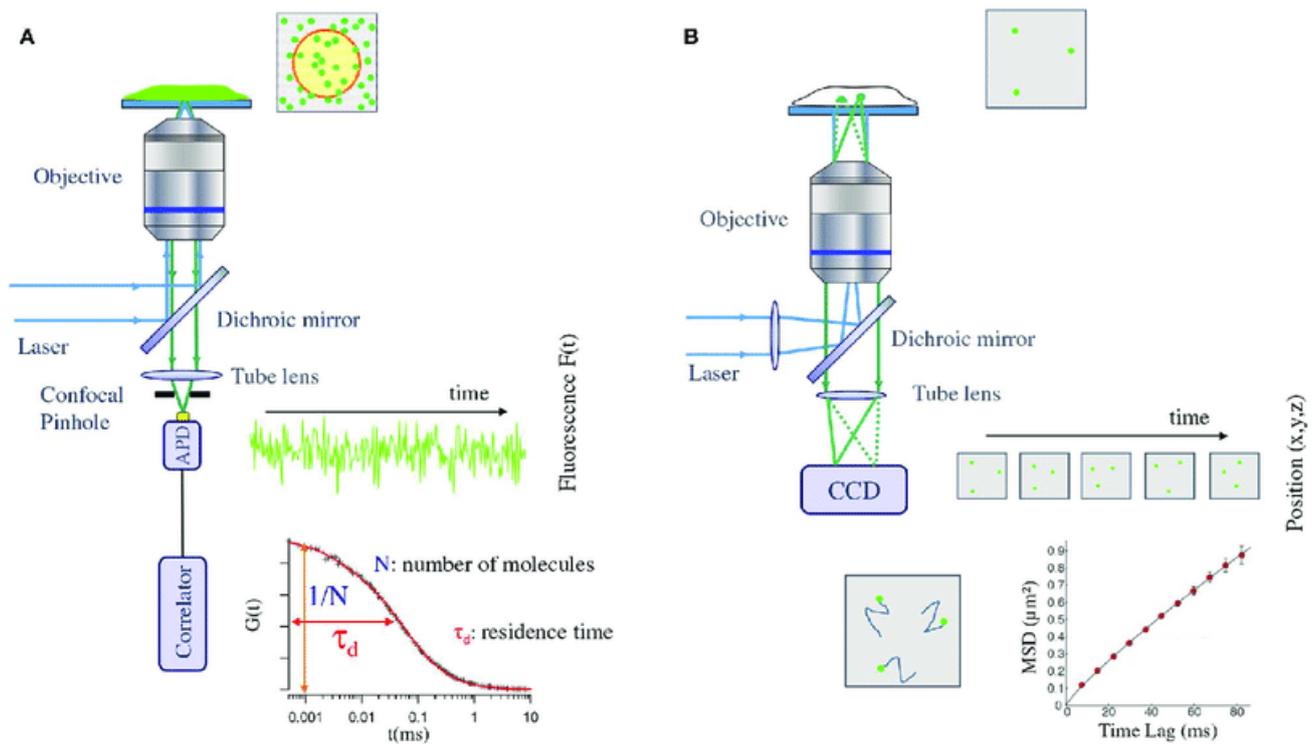
Figure 15.3 Schematic view of the typical setup used in fluorescence correlation spectroscopy (A) and single/multiple particle tracking (B) experiments. (A) A laser is focused on the fluorescently labeled sample by the objective of a microscope. The fluorescence is then collected by the objective and focused in a confocal way (using a pinhole) on a single photon. Source: Anomalous Sub diffusion in Living Cells: Bridging the Gap Between Experiments and Realistic Models Through Collaborative Challenges, Article, May 2020.

The intensity of the fluorescently emitted radiation $\{If(t)\}$ is proportional to the concentration of fluorescent molecules in the scattering volume $\{c\,(\gamma, t)\}$, the fluorescent yield (Q) and the intensity of the incident laser beam (I):

$$If(t) = Q_\varepsilon \int I(\gamma)\, c(\gamma_1 t) d^3\gamma \tag{15.3}$$

where ε is the extinction coefficient of the molecular species and the integral over $d^3\gamma$ is over the complete scattering volume. The dynamic information is contained in the fluctuations of the emitted fluorescent intensity with time, which are directly related to the fluctuations of the concentration of the fluorescent molecules $\big(\delta c(\gamma_1 t)\delta c(\gamma'_1 0)\big)$:

$$(\delta If(t)\delta If(0)) = (\varepsilon Q)^2 \iint I(\gamma)I(\zeta)\big(\delta c(\gamma_1 t)\delta c(\zeta_1 0)\big)d^3\gamma d^3\zeta \tag{15.4}$$

where $\{\zeta\}$ denotes a time average and δ is a small fluctuation in a quantity.

For translational diffusion of the fluorescent probes the intensity autocorrelation function has a simple exponential form. Thus, an exponential can be fitted to the auto correlated fluorescent signal from a fluorescently tagged biomolecule, which gives the characteristic time constant (t) to diffuse out of the scattering volume (e.g. a cube of side b), and hence the diffusion coefficient (D) of the molecules (D = $6b^2t$). Fluctuations in the intensity of the fluorescent light emitted as particles move across the irradiated volume can thus be related to the diffusion coefficients of the fluorescent species. Unfortunately, the information that relates to the momentum transfer (equation (15.2)) is lost in this inelastic scattering method and, for larger scattering volumes of biological materials, data from quasi-elastic scattering are much more rich in information. Fluorescence depolarization experiments are a further powerful tool. A pulsed laser excites fluorescent probes attached to biological molecules, whose motion can be detected by the change in polarization of the reemitted photon. If the molecule reorientates a considerable amount over the picosecond time scale of fluorescent emission, the polarization state of the emitted photon is changed (15.3).
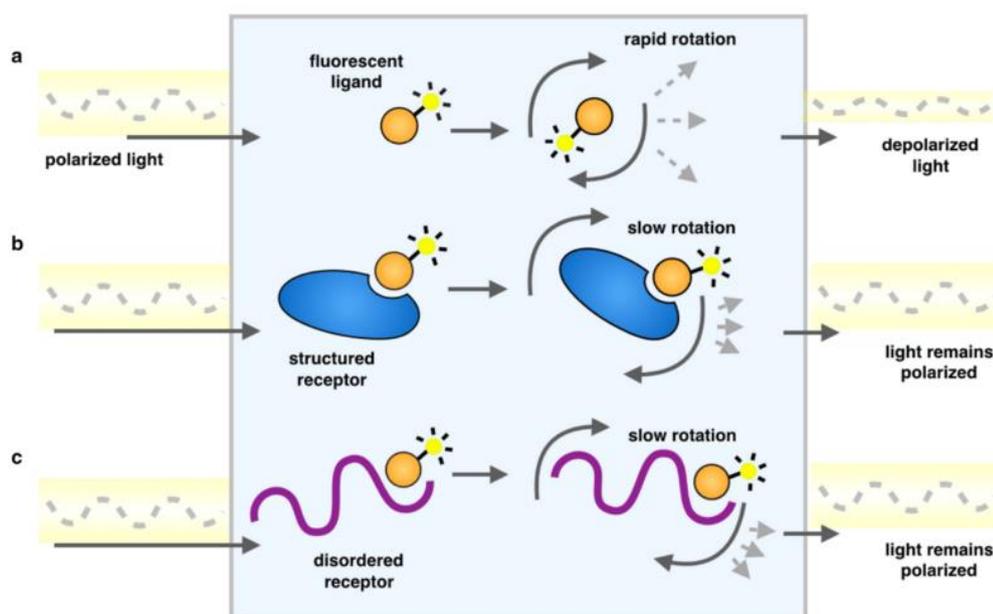

Fig. 15.3. Schematic representation of a fluorescence polarization experiments. Source: ResearchGate - Methods of probing the interactions between small molecules and disordered proteins

The utility of fluorescence depolarization stems from the fact that the technique can probe dynamics in the ultra -fast picosecond time regime due to the availability of intense ultra-fast pulsed laser sources (Figure 15.4)


.
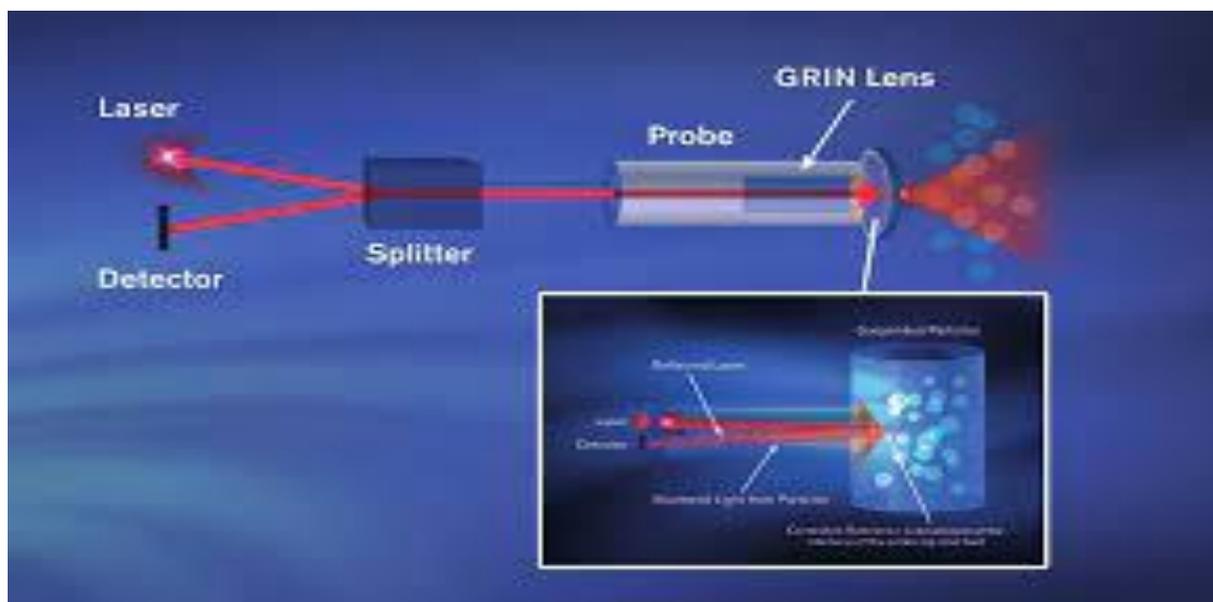
Fig. 15.4. Scheme of dynamic light scattering.

Futhermore, the high yield of fluorescent reemission makes the experimental measurement correlation functions over ultra-fast time scales and with single molecules feasible. (Fig. 15.5).
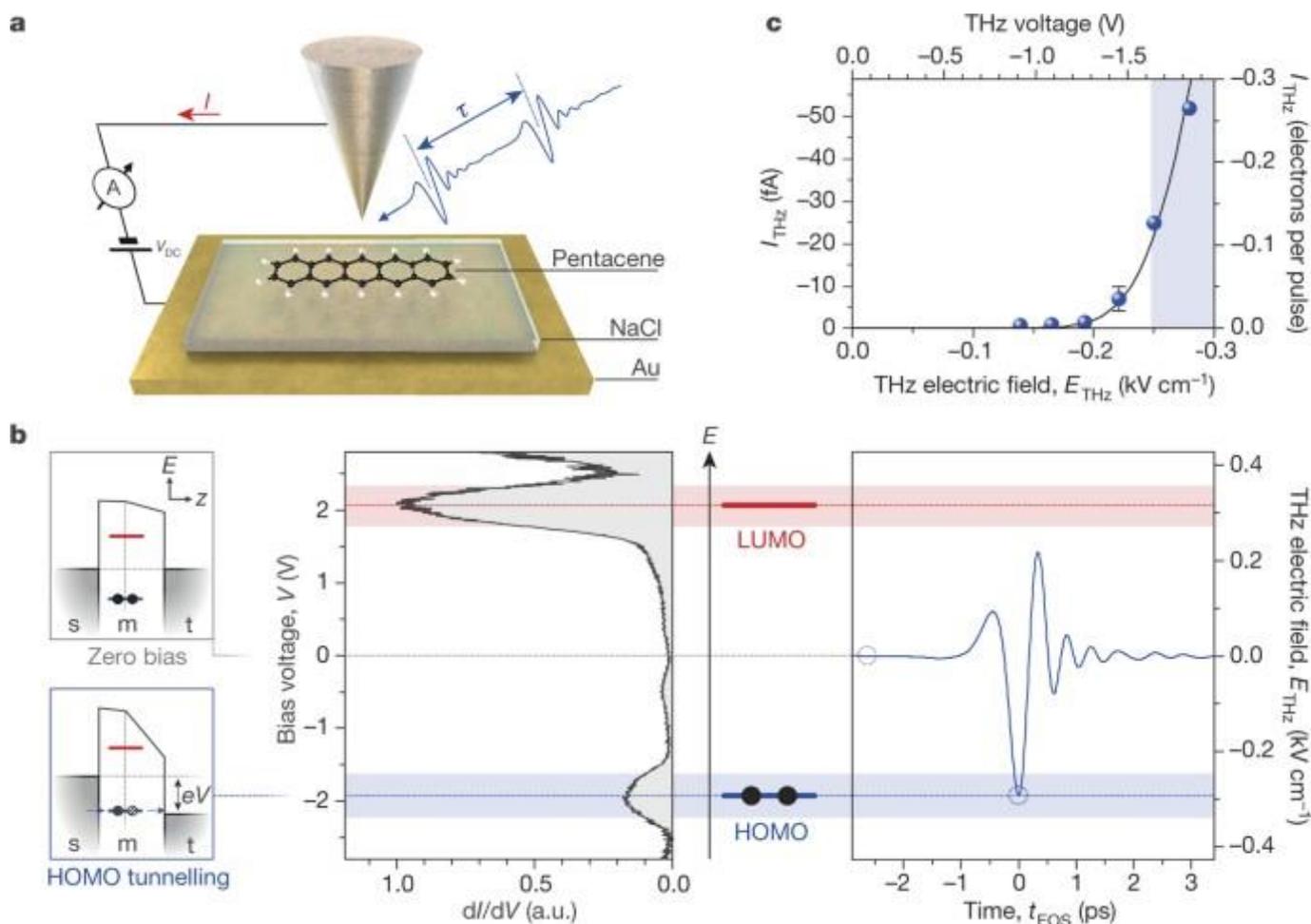

Fig.15.4. Tracking the ultrafast motion of a single molecule. Tyler L. Cocker, Dominik Peller, Ping Yu, Jascha Repp & Rupert Huber. Tracking the ultrafast motion of a single molecule by femtosecond orbital imaging. Published: 09 November 2016. Nature volume 539, pages263–267(2016).

Quasi-elastic scattering experiments cover a wide range of techniques that monitor small energy changes in scattered radiation due to the motion of a sample, i.e. a Doppler shift of the energy of the scattered particle. Typically, it is the normalised intermediate scattering function ($F(q, t)$ that is measured in a quasi-elastic scattering experiment.
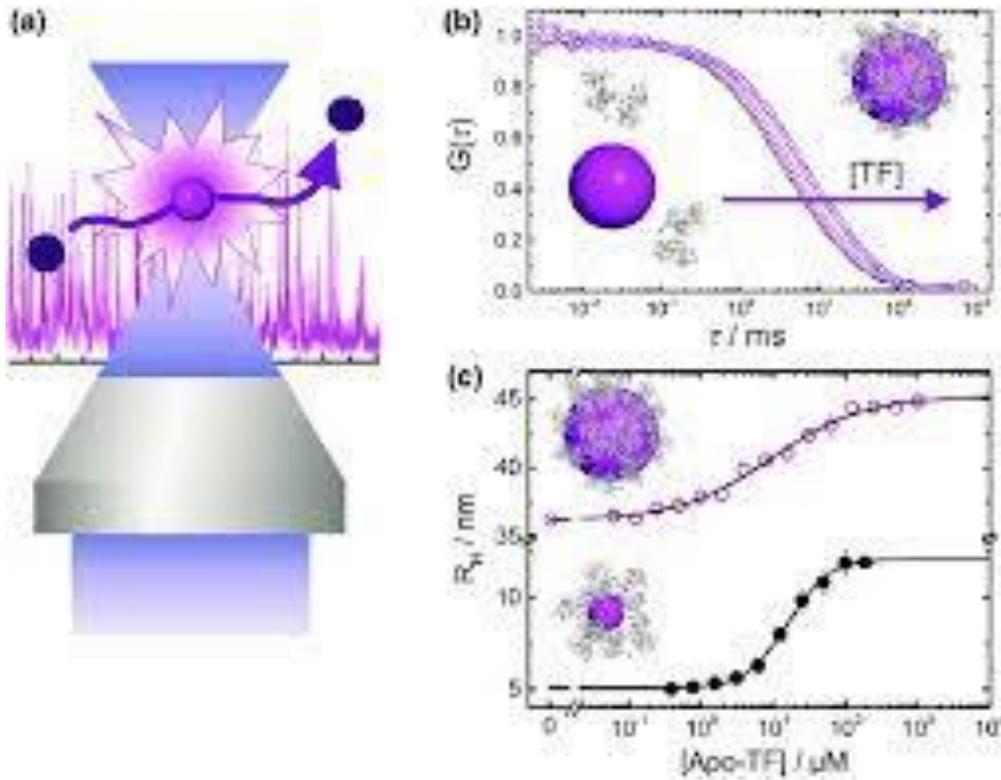
The intermediate scattering function is a useful general tool, which is amenable to accurate quantitative theoretical analysis. For the scattering of coherent light (called equivalently quasi-elastic light scattering, photon correlation spectroscopy or dynamic light scattering) the quantity examined experimentally $\{g(\tau)\}$, the field correlation function} relates to the correlation function of the scattered electric field. In terms of the scattered electric field (E) at time (t) the field correlation function measured at a certain angle is defined as:

$$g_1(\tau) = \frac{\left(E^*(0)E(\tau)\right)}{(I)} \tag{15.5}$$

where $\{E(0)\,E(\tau)\}$ is the ensemble averaged quantity over the array of scatterers, and I is the intensity measured on the detector. The electric field strength (E) scattered by the collection of diffusing particles in the sample is given by:

$$E = \sum_{j-1}^{N} A_j e^{iq\gamma} E_0 e^{i\omega t} \tag{15.6}$$

where $E(0)$ is the magnitude of the incident wave, A_j is the scattered wave amplitude of the jth particle, rj is the position of the jth particle, q is the momentum transfer, t is the time and v_0 is the frequency of the incident radiation. The sum from $j = 1 \ldots N$ is over all the particles that scatter

236

radiation in the sample. The form of the electric field can be substituted in equation (15.5) and the correlation function can therefore be written:

$$g_1(\tau) = \left(e^{iq\gamma}\right) \sum_{j-1}^{N} \left(e^{iq(\gamma_1(\tau)\gamma_1(0))}\right) e^{i\omega_0\tau} \qquad (15.7)$$

Fick's second law of diffusion in three dimensions is:

$$\frac{\partial c(\gamma_1\ t)}{\partial t} = D\nabla^2\left(\gamma_1\ t\right) \qquad (15.8)$$

where $c(r,t)$ is the concentration of molecules in the scattering volume and D is the translational diffusion coefficient. Let $P(O/r,t)$ be the conditional probability that a particle can be found in volume element $d^3\gamma$ at time t. For low particle concentrations the conditional probability $P(O/r,t)$ also obeys the diffusion equation:

$$\frac{\partial P\left(O\,|\,\underline{\gamma_1}\ t\right)}{\partial t} = D\nabla^2 P\left(O\,|\,\underline{\gamma_1}\ t\right) \qquad (15.9)$$

A Fourier transform can be taken of either side of (15.9) and allows the equation to be solved:

$$\int_0^\infty e^{iq.\gamma} \frac{\partial P\left(O\,|\,\underline{\gamma_1}\ t\right)}{\partial t} d^3\gamma = D \int_0^\infty e^{iq.\gamma} \nabla^2 P\left(O\,|\,\underline{\gamma_1}\ t\right) d^3\gamma \qquad (15.10)$$

A general property of Fourier transforms is that the Fourier transform of n^{dr} order differential is equal to $(iq)^n$ times the Fourier transform of the argument of the differential:

$$\int_0^\infty e^{iq.\gamma} \frac{\partial^n}{\partial\gamma^n} P(\gamma)d\gamma = (iq)^n \int_\infty^\infty e^{iq\gamma} P(\gamma)d\gamma \qquad (15.11)$$

The intermediate scattering function $\{F(q;t)\}$ is equal to the Fourier transform of the probability distribution $(P(O\,|\,r,t)$:

$$P_s(q_1\ t) = \int_0^\infty P\left(O\,|\,\gamma_1\ t\right) e^{iq.\gamma} d^3\gamma \qquad (15.12)$$

This definition and the Fourier transform identity (equation (15.11) allows equation (15.10) to be simplified in terms of the intermediate scattering function:

$$\frac{\partial F_s(q_1\ t)}{\partial t} = Dq^2 F_s(q_1\ t) \qquad (15.13)$$

This is a simple variables separable differential equation which has the solution:

$$F_s(q_1\ t) = F_s(q_1\ t) e^{Dq^2 t} \qquad (15.14)$$

where the initial condition is given by Fs(q,0)=1. From equation (15.7) the field correlation function can therefore be written as:

$$g_1(\tau) = F_s(q_1\ t)e^{i\omega 0\tau} = e^{Dq^2\ t}e^{i\omega 0\tau} \qquad (15.15)$$

The intensity correlation function $g_2(\tau)$ is measured by time correlation of the signal on a photomultiplier tube due to the scattered radiation and is related to the electric field correlation function $g_1(\tau)$ by:

$$g_2(\tau) = 1 + |g_1(\tau)|^2 \qquad (15.6$$

By substitution of equation (15.15) in (15.16) it is seen that diffusional process introduces a $\exp(Dg_2(\tau))$ term in the intensity correlation function, $g_2(\tau)$. A typical correlation function for a fibrous protein that experiences translational diffusion is shown in Figure 15.6.A qualitative understanding of the form of a correlation function can be achieved. At short times the particles have insufficient time to move anywhere, they do not diphase the scattered light (equivalently the scattered speckle pattern does not move), and the correlation is nearly perfect $g_1(\tau)\sim 1$. At longer times the motion of the particle's decor relates the phase of the scattered radiation and reduces the value of $g_1(\tau)$. Eventually at very long times the correlation function reduces to zero due to a complete random phasing of the scattered photons.
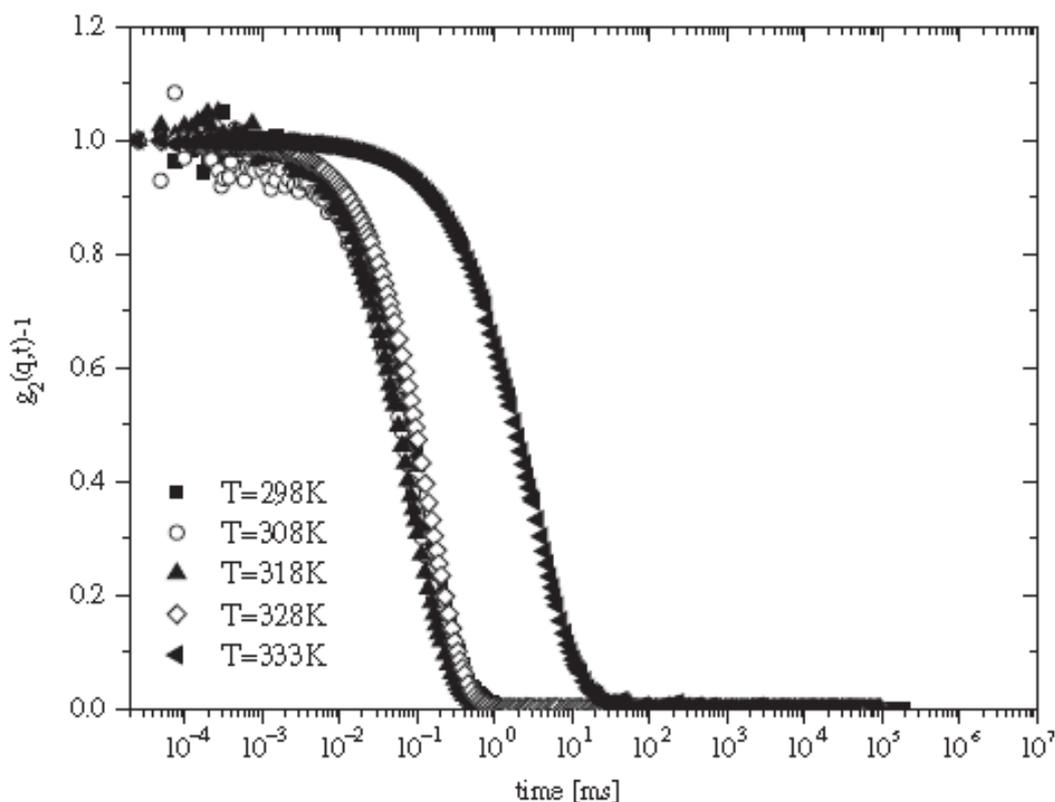

Figure 13.7. Intermediate scattering (intensity correlation, g_2(q,t) functions from a quasi-elastic light scattering experiment on titin, a giant protein from skeletal muscle (The dynamics slow down as the protein unfolds with increasing temperature (T) due to the increase in protein length [E. Di Cola, T.A. Waigh, J. Trinick, et al., Biophysical Journal, 88, 4095–4106, Copyright (2005) Biophysical Society])

A wide range of other radiation can be used to probe dynamic processes in mesoscopic biological materials with quasi-elastic scattering experiments that includes X-rays (time scales 10^7 -1000 s) and neutrons (time scales 0,1-100 ns). X-ray quasi-elastic techniques are similar in conception to dynamic light scattering, whereas neutron spin echo measurements use the spin of scattered neutrons to clock the dynamics of the scattering process

Fig. 15. 8. Diode-Pumped Solid-State Laser Kit for Education & Research. Features 5 modes of operation, incl. CW & Q-switched.

Osmotic Pressure

The phenomena associated with osmotic pressure are important in a series of biological processes: how cell metabolism is regulated (animal cells walls are ruptured if the external osmotic pressure is too high or low), how intermolecular forces are mediated by solvent molecules and the molecular crowding of the intracellular environment. Consider an idealized experiment with a semi-permeable membrane that separates two polymer solutions (Figure 15.7). This apparatus is an example of a membrane osmometer, a device for the measurement of osmotic pressure.

From standard thermodynamic theory the partial differential of the Gibbs free energy G = F TS + PV) with respect to the pressure (P) is equal to the volume (V) at constant temperature (T):

$$\left(\frac{\partial G}{\partial P}\right)_T = V \tag{15.17}$$
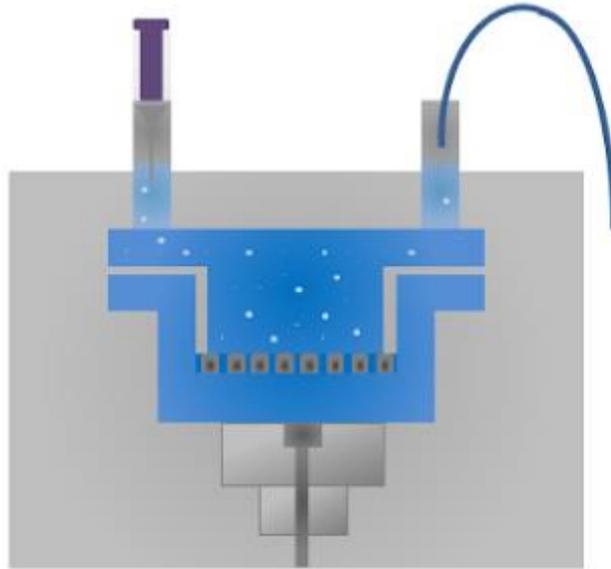
Figure 15.9. A membrane osmometer can be used to measure the difference in osmotic pressure between two solutions, a and b, which have P_a and P_b external applied pressures respectively. The difference in fluid heights (the capillary rise h) provides the osmotic pressure ($\pi = hrg$), where r is the fluid density and g is the acceleration due to gravity. WEE-Solve GmbH.

The chemical potential (μ, the Gibbs free energy per particle) with respect to one of the components in the solution (μ_1) is:

$$\left(\frac{\partial \mu_1}{\partial P}\right)_T = \bar{V}_1 \qquad (15.18)$$

where V_1 is the partial molar volume of the particles. Consider the chemical potentials on either side of the membrane in sections a and b (Figure 15.7). Compartment b only contains solvent molecules and the chemical potential is that of the solvent (μ_1^0):

$$\mu_1^b = \mu_1^0 \qquad (15.19)$$

In compartment as the chemical potential has two additional effects, that due to the solute and that due to the reservoir used to measure the pressure:

$$\mu_1^b = \mu_1^0 (solute - effect) + (pressure - effect) \qquad (15.20)$$

The effect of the solute on the chemical potential is given by a van der Waals type expansion in the component concentration:

$$\mu_1^b = \mu_1^0 \quad RTV_1^0 \left(\frac{c}{M} + Bc^3 + \dots\right) + \int_{P_0}^{P_0 + \pi} \bar{V}_1 \, dP \qquad (15.21)$$

240

where c is the concentration of the species too large to permeate the membrane (the solute), M is the molecular weight of the solute and B is the second virial coefficient of the solute. $V_1^0 = V_1$ is the molecular volume of solvent at one atmosphere pressure, which allows the last term in equation (13.21) to be evaluated. In thermal equilibrium the chemical potentials on each side of the membrane are equal ($\mu_1^6 = \mu_1^a$).

$$\mu_1^0 = \mu_1^0 \quad RTV_1^0 \left(\frac{c}{M} + Bc^3 + \ldots\right) + V_1^0 \pi \qquad (15.22)$$

This expression can be solved for the osmotic pressure (π) of the solution:

$$\pi = RT \left(\frac{c}{M} + Bc^2 + \ldots\right) \qquad (15.23)$$

For a dilute solution the second virial coefficient is very small (B= 0), so to a good approximation:

$$\pi = \frac{RT_c}{M} = nkT \qquad (15.24)$$

The osmotic pressure of a solution of non-interacting particles is directly proportional to the number of solute molecules (n= c/N^AM), where N_A is Avogadro's number) and, if the concentration is known, an accurate determination of the molecular weight of the particles can be made. An example of the osmotic pressure of a protein solutions shown in Figure 15.10 with the simple osmometer shown in Figure 15.9, the osmotic pressure (π). The osmotic pressure of a solution of non-interacting particles is directly proportional to the number of solute molecules (n= c/N^AM), where N_A is Avogadro's number) and, if the concentration is known, an accurate determination of the molecular weight of the particles can be made. An example of the osmotic pressure of a protein solution is shown in Figure 15.8. With the simple osmometer shown in Figure 15.6, the osmotic pressure (π) is given by:

$$\pi = hg\rho \qquad (15.25)$$

where h is the difference in fluid height, g is the acceleration due to gravity and r is the fluid density. Thus a simple measurement of the height of the fluid in the capillary leads to a direct calculation of the solution's osmotic pressure.
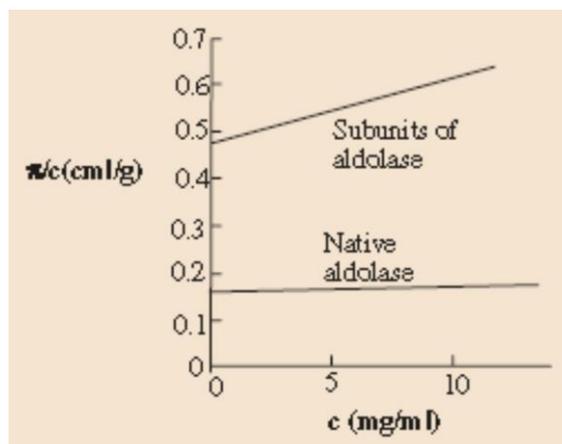
Figure 15.10. The osmotic pressure of a solution of globular proteins (A number of al do lase subunits assemble into the native structure which causes a decrease in the osmotic pressure of the solution. The osmotic pressure per gram of protein (p=c) is plotted as a function of protein concentration {Reprinted with permission from F.J. Castellino and O.R. Baker, Biochemistry, 7, 2207–2217, Copyright (1968) American Chemical Society}).

The effects of osmotic pressure are extremely important in determining the physical state and morphology of a biological material. For example, a charged polyelectrolyte gel placed in a dilute solution expands many times in volume due to the pressure exerted by the counterions associated with the polyelectrolyte chains. Polyelectrolyte gels can exhibit significant elasticity at very low volume fractions, e.g. 'wobbly solid' gelatine gels are 2–3% polymer and 97–98% water. Simple macroscopic measurements of gel sizes and concentrations thus provide an important tool to understand their molecular structuring (Figure 15.11).
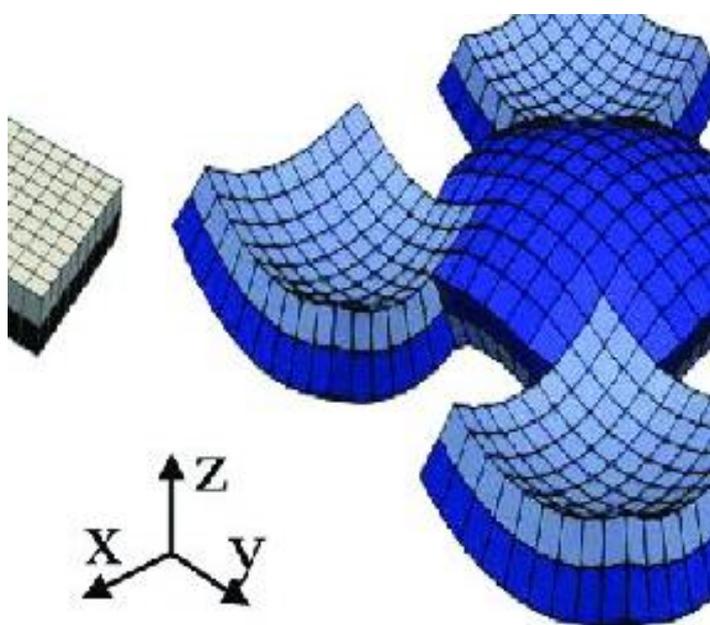
Figure 15.11 The swelling of polymer gels is intimately related to the osmotic pressure of the network and the swelling is resisted by the elasticity of the polymer chains. "Patterning with Loops" to Dynamically Reconfigure Polymer Gels.Source: Santidan Biswas, Victor V. Yashin, Anna Balazs. "Patterning with Loops" to Dynamically Reconfigure Polymer Gels.April 2018, Soft Matter 14(17). DOI: 10.1039/C8SM00270C.

There are a number of other methods typically used to measure osmotic pressure. These include the vapour pressure osmometer (it examines the depression of the boiling point by the osmotic effect) andoptical tweezers (a direct measurement of the piconewt on pressures on colloidal particles is made).

Force Measurement

There is a wide range of techniques for the measurement of mesoscopic forces. Some of the more modern developments in the field of force measurement include atomic force microscopy (AFM), glass fibers,

surface force apparatus (SFA) and magnetic/optical tweezers. The range of forces that can typically be measured with each of the techniques is compared in Figure 15.12. In general terms AFM and SFA offer the largest forces, and magnetic/optical tweezers offer the greatest sensitivity.
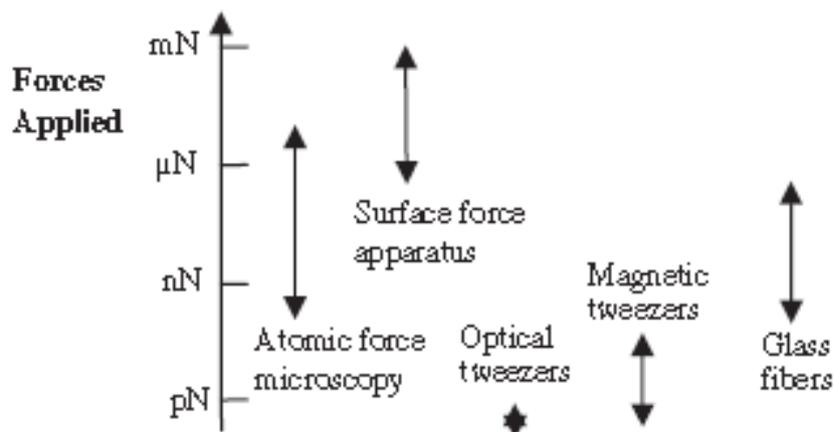

Figure 15.12. Forces that can typically be measured by the different techniques.
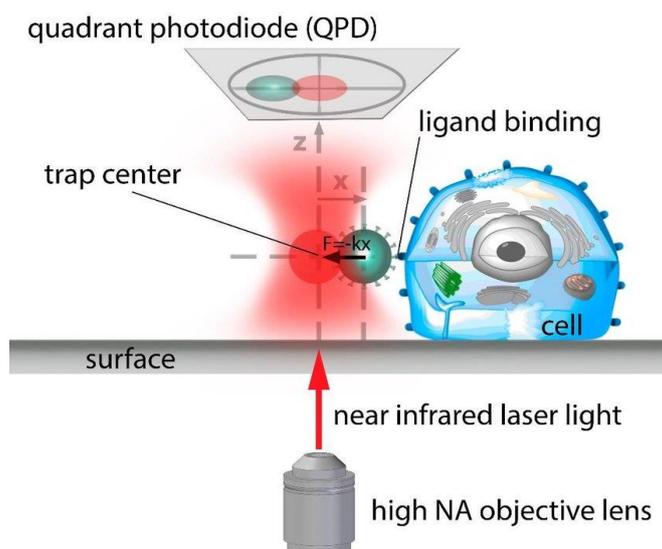
Figure 15.13. An optical trap is created by focusing a near infrared laser light to a diffraction-limited spot with a high numerical aperture (NA) microscope objective. The laser beam used for trapping has a 2-D Gaussian intensity distribution with an intensity gradient in the x-and y-plane (perpendicular to the beam axis). Source: A practical review on the measurement tools for cellular adhesion force, Article in Advances in Colloid and Interface Science, May 2019.

Optical and magnetic tweezers are both similar in application, a feedback mechanism is used to clamp the position of a colloidal particle in three dimensions under an optical microscope, and this force is subsequently used to manipulate biological molecules. However, the physical processes which control the mode of action of the two types of tweezer are radically different. Optical tweezers focus laser light to trap a dielectric particle using the pressure of photons in the incident laser beam (Figure 15.14). Magnetic

tweezers use the magnetic force on super paramagnetic or ferromagnetic beads exerted by gradients in an applied magnetic field (Figure 15.14).

For optical tweezers, the electric dipole (p) induced in a trapped particle is given by:

$$\underline{p} = \underline{\alpha} \cdot \underline{E} \tag{15.25}$$

where α is the polarizability of the irradiated material and E is the electric field of the incident laser. The induction of the optical dipole moment in the trapped particle by the laser beam provides a force (F_{light}), proportional to the Laplacian of the electric field, or equivalently the gradient of the intensity of the incident light (∇I):

$$\underline{F_{light}} = \alpha \underline{\nabla^2} E = \alpha \underline{\nabla} I \tag{15.27}$$
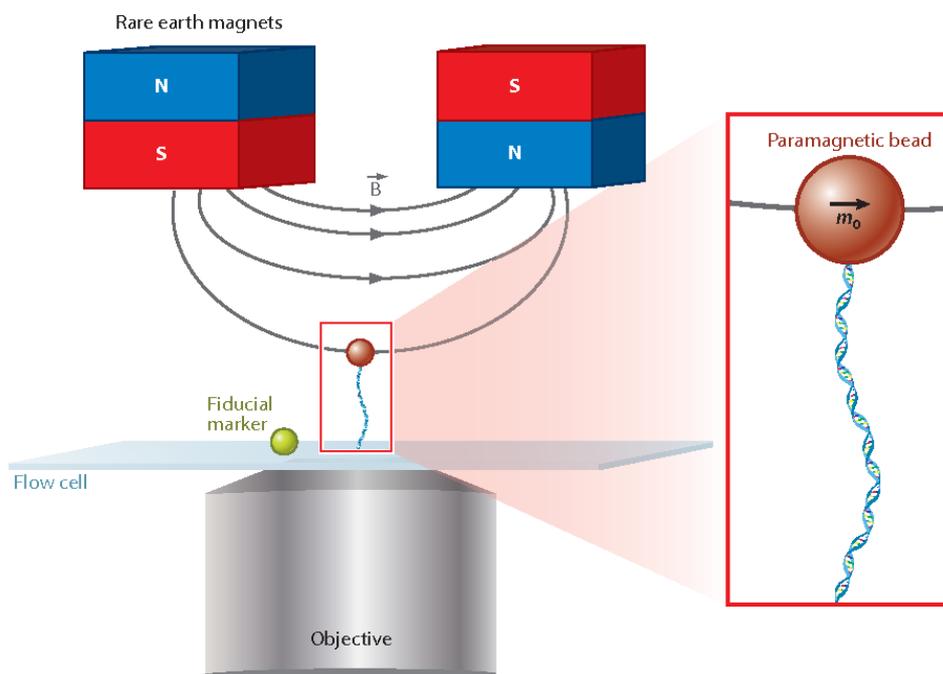
Figure 15.14. Schematic of basic implementation of magnetic tweezers. A molecule is tethered between the surface of a flow cell and a paramagnetic bead. The magnetic field generated by a pair of magnets induces a magnetic moment m0 in the paramagnetic bead. The bead experiences a force proportional to the gradient of the field. The molecule can be coiled by rotating the external magnet. Abbreviations: N, magnetic north pole; S, magnetic south pole.

A single well focused laser beam can trap dielectric particles $(0,5 \mu m)$ in three dimensions using the pN forces induced from the optical dipolar force. For a specific laser, microscope and optical set up, the tweezerforce (F_{light}) is related to the incident laser power by the equation:

$$\underline{F_{light}} = \frac{Q n_m P}{c} \tag{15.28}$$

where Q is the efficiency of the trap, n_m is the index of refraction of the particle, c is the speed of light and P is the incident laser power. Double optical traps can be used to extend single molecules that are attached

at either end to colloidal probes. The correct choice of the laser in a trapping experiment is important to reduce the damage on fragile biological molecules. Often, with delicate biological materials, infrared lasers are used to minimize this damage.

In contrast to optical tweezers, with magnetic tweezers the potential energy (U) of a magnetic dipole (m) placed in a magnetic field (B) is given by the scalar product:

$$U = \underline{m} \cdot \underline{B} \tag{15.29}$$

Thus a free permanent magnetic dipole experiences a torque as it minimizes its energy through the alignment of the dipole with the applied magnetic field. The magnetic forces experienced by a probe particle depend sensitively on its type of magnetism. The colloidal probes used in magnetic tweezer experiments are typically either ferromagnetic or superparamagnetic. The corresponding magnetic force (Fmag) is the gradient of the potential (∇U):

$$\underline{F_{mag}} = \underline{\nabla}(\underline{m} \cdot \underline{B}) \tag{15.30}$$

For superparamagnetic spheres the magnetization is approximately equal to the saturated value (Mmax), m ~ Mmax, and the magnetic field gradient occurs parallel to the x-axis, so:

$$\underline{F_{mag}} \approx M_{max}V\frac{dB}{dx} \tag{15.31}$$

where V is the particle volume. With superparamagnetic particles the application of a magnetic field gradient can provide forces on the order of$100\ pN$. The torque on a large ferromagnetic particle ($4\ \mu$m) can be quite considerable ($\sim 1000\ pN\mu m$) and magnetic tweezer cytometry has found applications for determining the elasticity of cells adhered to magnetic beads. Hysteresis effects in the magnetism curves of the probe particles and pole pieces pose a number of technical challenges for accurate quantitative analysis of magnetic forces, particularly with ferromagnetic beads.

The hydrodynamic drag force (Fdrag) experienced by a trapped particle moved (velocity, v) through its surrounding solvent with optical or magnetic tweezers is given by Stoke's law:

$$F_{drag} = 6\pi\eta\alpha v \tag{15.32}$$

This equation provides a method for calibrating both optical and magnetic traps. The trapped colloidal probe is held at rest with respect to the laboratory and the solvent is given a constant velocity using flow cell. The critical velocity at which the trapped bead becomes dislodged is measured, which allows the force applied to the tweezers to be calculated using equation (15.32).

A more accurate method for calibrating traps uses an analysis of the thermal fluctuations of the trapped particle and is based on Langevin's equation for the particle's motion:

$$m\frac{dv}{dt} = F_{thermal}(t)\ \ \gamma v\ \ kx \tag{15.33}$$

This is just Newton's second law, a balance of the inertial force (mdv/dt), the thermal force Fthermal(t), the drag force γv and the elastic trap force (the effective lateral trap spring constant is k and x is the particle displacement). To solve equation (15.33), the thermal force (F thermal) is assumed to be completely random over time (t); mathematically this is equivalent to:

$$\left(F_{thermal}(t)F_{thermal}(t\ \tau)\right) = 2kT\gamma\delta(\tau) \qquad (15.34)$$

where δt is the dirac delta function, kT is the thermal energy and γ is the frictional coefficient. In the low Reynolds number regim the inertial term can be neglected ($mdv/dt\ =\ 0$); this greatly simplifies equation (15.33) and is typically the case in most tweezer experiments at low frequencies. The Laplace transform of the Langevin equation (15.33) can be taken to provide an expression of the power spectrum $Sx\ (\omega)$ of the fluctuations of the bead displacement:

$$S_x(\omega) = \frac{kT}{\pi^2\gamma(\omega_c^2\omega^2)} \qquad (15.35)$$

where kT is the thermal energy, γ is the drag coefficient, ω_c is the cornering frequency and ω is the frequency. The power spectrum of bead fluctuations can be easily determined experimentally using a numerical fast Fourier transform of the bead square displacement as a function of time (Figure 15.13). Equation (15.35) then allows the cornering frequency ω_c to be calculated and the spring constant (k) of the trap can be subsequently found using the equation:

$$\omega_c = \frac{k}{2\pi\gamma} \qquad (15.36)$$

The technique of atomic force microscopy (AFM) allows the force between the tip of a cantilever (with a small radius of curvature) and virtually any kind of surface to be measured (Figure 15.16. The AFM technique also has the significant advantage that the tip can be used to form an image of the surface. In a typical AFM experiment a small pyramid shaped tip is mounted on a cantilever which acts as a spring, with a spring constant ~ 0,1Nm. The cantilever is arranged on a piezo electric driver that moves the tip in the vertical direction whilst the resulting displacement is measured by reflecting a laser beam from the back of the cantilever onto a split photodiode. A range of feedback methods is used to control the position of the cantilever, for example, that holds the cantilever at a constant force. The detailed construction of an AFM is shown in the Figure 15.15.
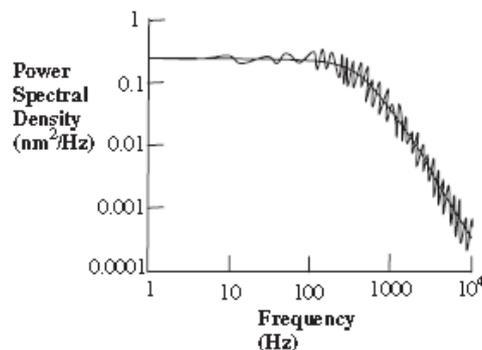
Figure 15.15. Power spectral density $< r^2(f) > Sx (\acute{\omega})$ of the position of a particle trapped in the laser of an optical tweezer set up as a function of frequency (f) [Reprinted with permission from K. Svoboda and S.M. Block, Ann. Rev. Biophys. Biomol. Struct. 23, 247–285, Copyright (1994) Annual Reviews]
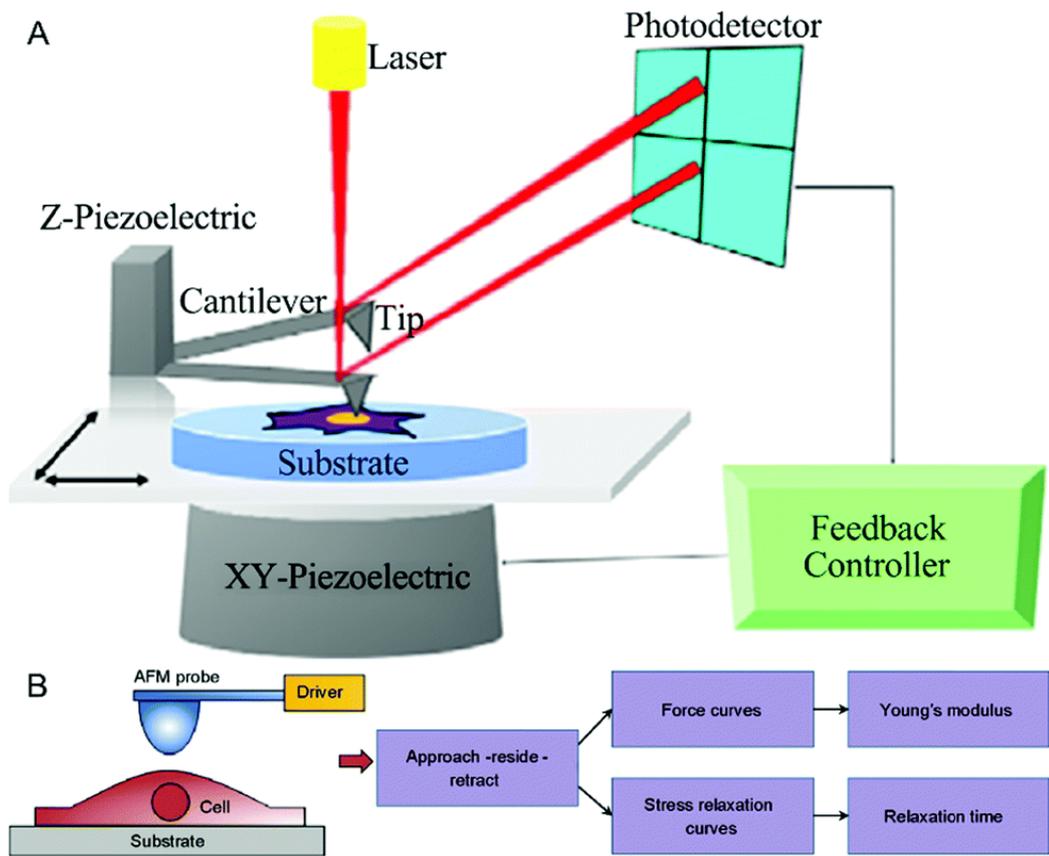

Figure 15.16 Schematic diagram of an Atomic Force Microscopy (AFM) experiment to study the surface of a biological material. Source: Recent advances in AFM-based biological characterization and applications at multiple levels, Journal of the Soft Matter.

AFM allows much larger forces to be applied to a sample than with optical/magnetic tweezers and imaging is also possible (Figure 15.17).
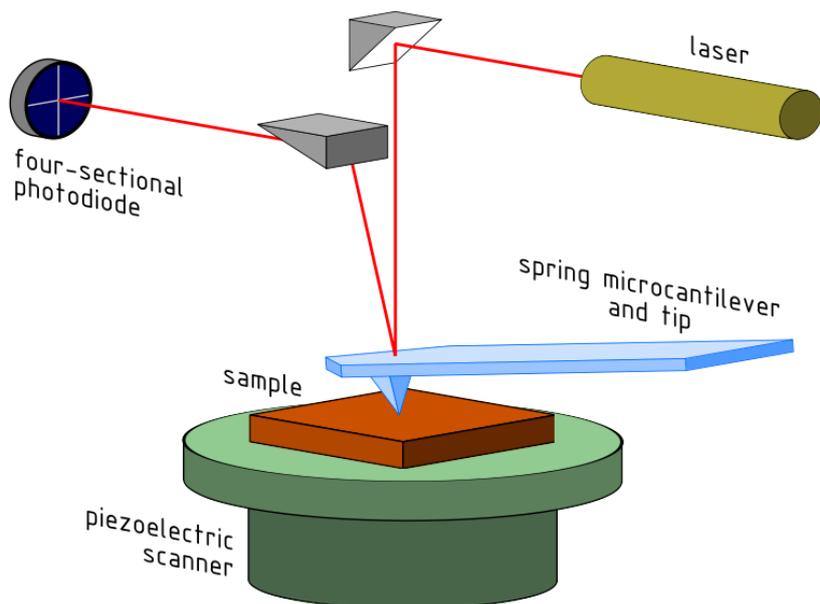
Figure 15.17 Detailed schematic diagram of an AFM apparatus. The feedback scheme can be used to hold the cantilever at a constant force on the sample. [Ref.: J.Yang, L.K.Tamm, A.P. Somlyo and Z. Shao, Journal of Microscopy, 1993,171, 183–198].
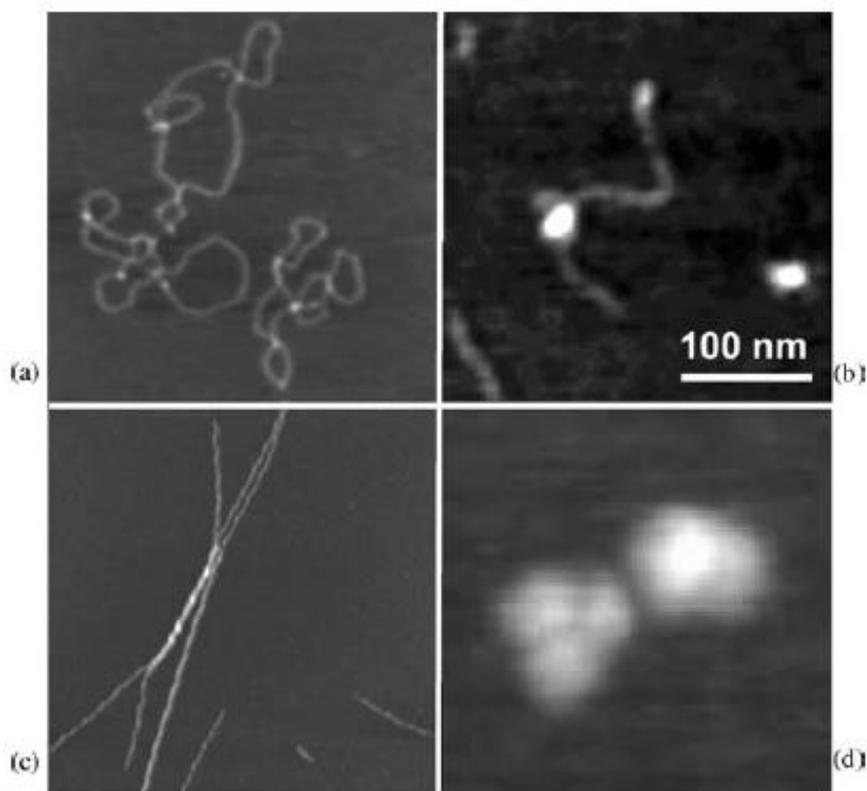
Figure 15.18. Atomic force microscopy image of circular DNA (a; b) and amyloid fibrils (c; d) [Ref.: Neil Thomson, University of Leeds, 2005].

However, it is much more difficult to model the viscoelastic response of materials close to surfaces with AFM due to the effects of lubrication hydrodynamics, and the cantilever geometry causes the sensitivity to bereduced compared to the tweezer techniques (important for single molecule applications). Soft surfaces can be perturbed (indented) during the process of image collection due to the contact with the cantilever and sensitive feedback systems have been implemented to reduce this damage(the so called 'non contact mode'). Generally, the magnitude of the cantilever displacement in response to a force (F) at the surface is given by:

$$F = K_{cantilever}\Delta x \qquad (15.37)$$

where Δx is the displacement of the tip and $K_{cantilever}$ is the springconstant of the cantilever.

A typical value for $K_{cantilever}$ is 10^3 Nm^{-1} and a typical tip radius is 3×10^8 m. Surface force apparatus (SFA) examine the forces between surfaces on macroscopic dimensions. The technique involves the measurement of the distance of separation as a function of the applied force of crossed cylinders coated with molecularly cleaved mica sheets (Figure 15.19).
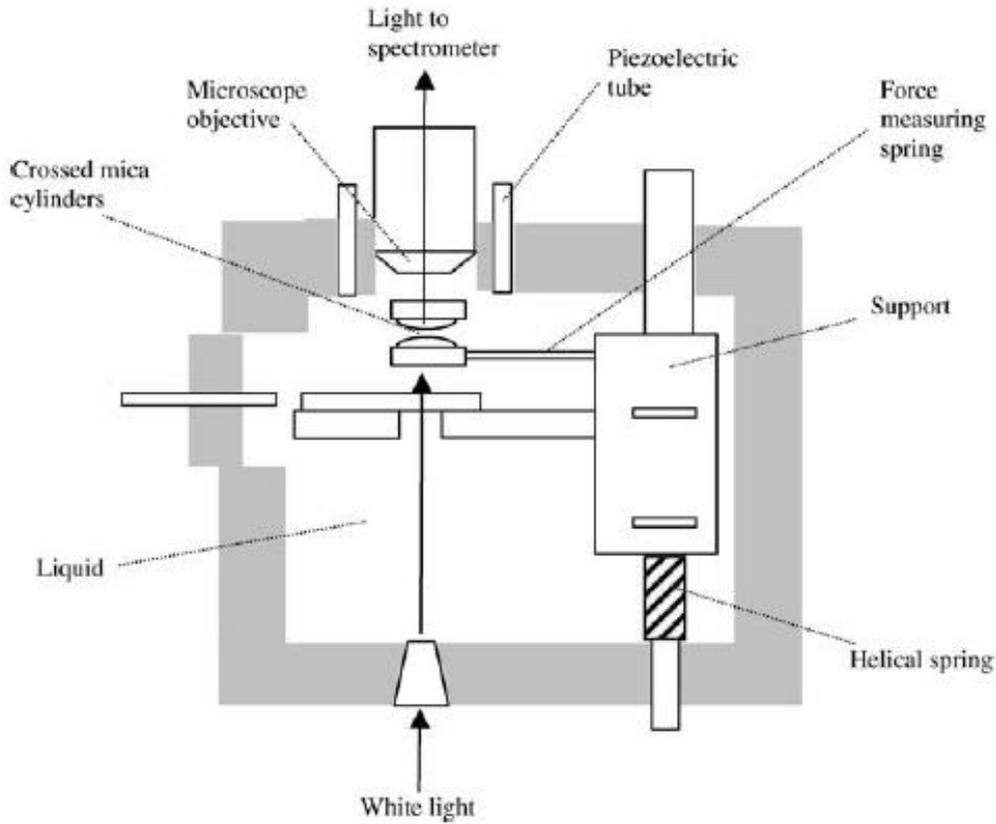
Figure 15.19. A schematic diagram of the arrangement of a surface force apparatus to measure mesoscopic forces. (The distance between the two mica cylinders is measured using an interferometric technique and the force is measured with a finely calibrated spring [Ref.: J.N. Israelachvili, Chemtracts-Analy. Phys. Chem, 1989, 1, 1–12])

The separation between the surfaces is measured interferometrically to a precision of 0,1 nm and the surfaces are driven together with piezo electric transducers with a resolution of 10^{-8}N. Much of the most accurate fundamental information on mesoscopic forces has been established using SFA. The split photodiode detector is a critical piece of technology for a series of force probe techniques that include AFM, glass fibres and optical tweezers. The detector allows fast accurate measurement of light intensities. It can provide sub-nanometer resolution of probe positions on the time scale of $100\,\mu s - 100\,s$ through comparison of the scattered light intensity projected onto the two sections of the split photodetector.

The mean photocurrent $\{i\}$ at time t measured by a section of the split detector is:

$$(i) = n \int_0^\infty g(t)dt = nze \tag{15.38}$$

Where $g(t)$ is the photo current detected at a time (t) given by

$$g(t) = \frac{ze}{\tau_0}\exp\left(\frac{t}{\tau_0}\right) \tag{15.39}$$

249

where z is the total number of charges displaced upon absorption of a photon on the detector, e is the electronic charge, t_0 is the time constant of the detector and n is the total number of photons collected. The position of a probe (e.g. the cantilever with an AFM or the colloidal probe with optical tweezers) measured using a split photodiode is found by comparing the difference in current signals (Δi) between the two photodiodes. The displacement noise on the determination of the probe position (with standard deviation $\sigma x(f)$, where f is the frequency) quantifies the accuracy of the split diode in a particular geometry and can be calculated as:

$$\sigma_x^2(f) = \frac{d^2}{2qn} \qquad (15.40)$$

The resolution of a split diode experiment thus depends on the total number of photons collected by the detector (n), the efficiency of the detector for absorbing photons (q) and the spatial width of the detector (d). It does not depend on the electronic charge (e), the instrumentamplification (z), or the magnification.

Electrophoresis

Electrophoresis is a cheap, powerful tool for the analysis and separation of charged biological molecules such as proteins and nucleic acids. Electrophoresis can be used to measure the size of biopolymer molecules and also to deduce the chemical sequence of the chains. The force experienced by a particle (F) in an electric field (E) is given by Coulomb's law:

$$F = ZeE \qquad (15.41)$$

where Z is the number of charges on the particle and e is the electronic charge.

The mobility of a charged particle in an electric field is proportional to the ratio of the net charge on the particle (which provides the Coulombic force) to its frictional coefficient (f). Electrophoresis can be used to obtain information about either the relative charge or the relative size of charged molecules. For steady state electrophoretic motion, the frictional force (the frictional coefficient (f) multiplied by the velocity (v), $f(v)$ is balanced by the force due to the electric field. The electrophoretic mobility (U) with colloids is defined as:

$$U = \frac{v}{E} = \frac{Ze}{f} \qquad (15.42)$$

Combined with Stokes law for the frictional force, this equation becomes:

$$U = R\frac{Ze}{6\pi\eta} \qquad (15.43)$$

Thus the mobility of the colloids measured in an electrophoresis experiment can be related to the charge fraction (Z) and the radius of the particles (R).

Conceptually the simplest method of measuring the mobility of a colloidal particle in an electric field is by using moving boundary (free) electrophoresis (Figure 15.20). Particle velocities are measured directly with an optical microscope as they move in the electric field. However, this technique suffers from artefacts such as convection and multicomponent interactions. It is possible to circumvent these problems using gels and ion exchange papers, and these are the electrophoresis methods that are predominantly used today.
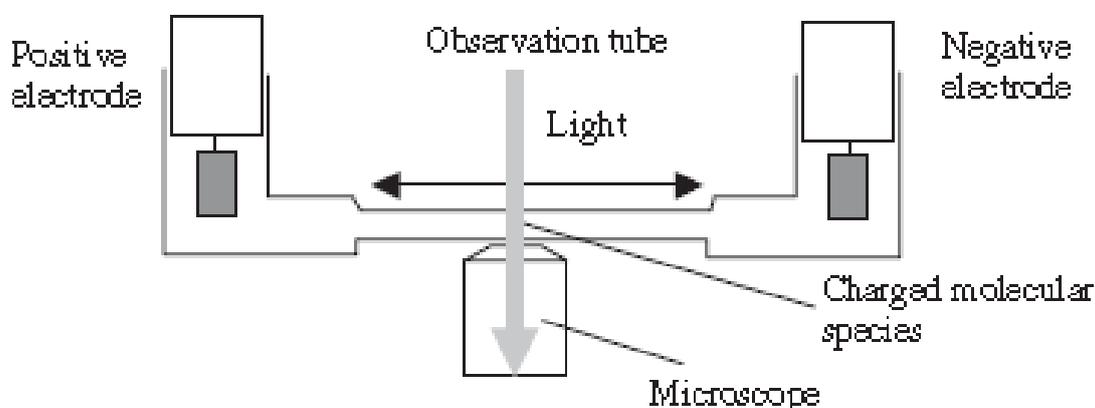
Figure 15.20. A moving boundary apparatus for the examination of free solution electrophoresis. The motion of the colloids that experience electrophoresis is measured with an optical microscope.

The charge on a protein depends on the pH of the buffer. Electrophoretic motion can be studied as a function of pH to calculate the isoelectric point. The isoelectric point is the pH at which the average enet charge on a macromolecule is zero. Isoelectric focusing (the pH is adjusted until there is no particle motility) can provide quantitative molecular information on charged macromolecules using simple table top apparatus.

The standard method of examining DNA chains is with gel electrophoresis (Figure 13.21). The use of the gel removes the problems with convection inherent in free boundary electrophoresis. The gel is placed across a constant applied voltage in a salt solution and the DNA chain are loaded onto the gel near the negative electrode. The gel is 'run' for a fixed period and the mobility of the chains (the time taken to travel a certain distance) can be simply related to their size. Surprisingly, detailed information on the complex topological nature of the gel is not required for quantitative predictions to be made on the molecular weight of DNA chains as they move across the gel.
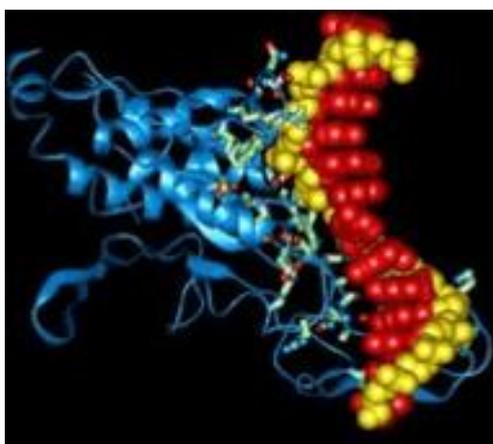
Figure 15.21. Capillary Electrophoresis. Source: Beckman – Coultier.

There are two common ways to locate DNA on a gel to measure the distance it has travelled. Ethidium bromide can be used to label the chains and fluoresces strongly under ultra violet light. Alternatively, it is possible to incorporate radiative phosphorus atoms that will darken a photographic film into DNA at one of its ends. An example is provided by electrophoresis with super coiled DNA molecules. Gel electrophoresis is a relatively easy method to separate closed super coiled DNA from the relaxed (cut) molecules. There is a large increase in the mobility of the super coiled DNA due to its compact form and it therefore experiences a reduced frictional coefficient (f) compared with the extended relaxed form For detailed sequencing of DNA chains restriction enzymes are used. These enzymes cut the DNA chains whenever they find the GAATTC sequence. If there are n such sequences there are n +1 bands that occur. Other specific enzyme/DNA reactions allow individual DNA molecules to be cut in different places and the resultant information can be combined to sequence chains of up to 400 base pairs. Isoelectric focusing is also possible with gel electrophoresis and can be used as an effective separation technique if the bands that contain the required charged molecules are cut out of the gel.

The theory of reptation is used to explain the ability of gel electrophoresis to separate DNA chains of different lengths (Figure 15.22).
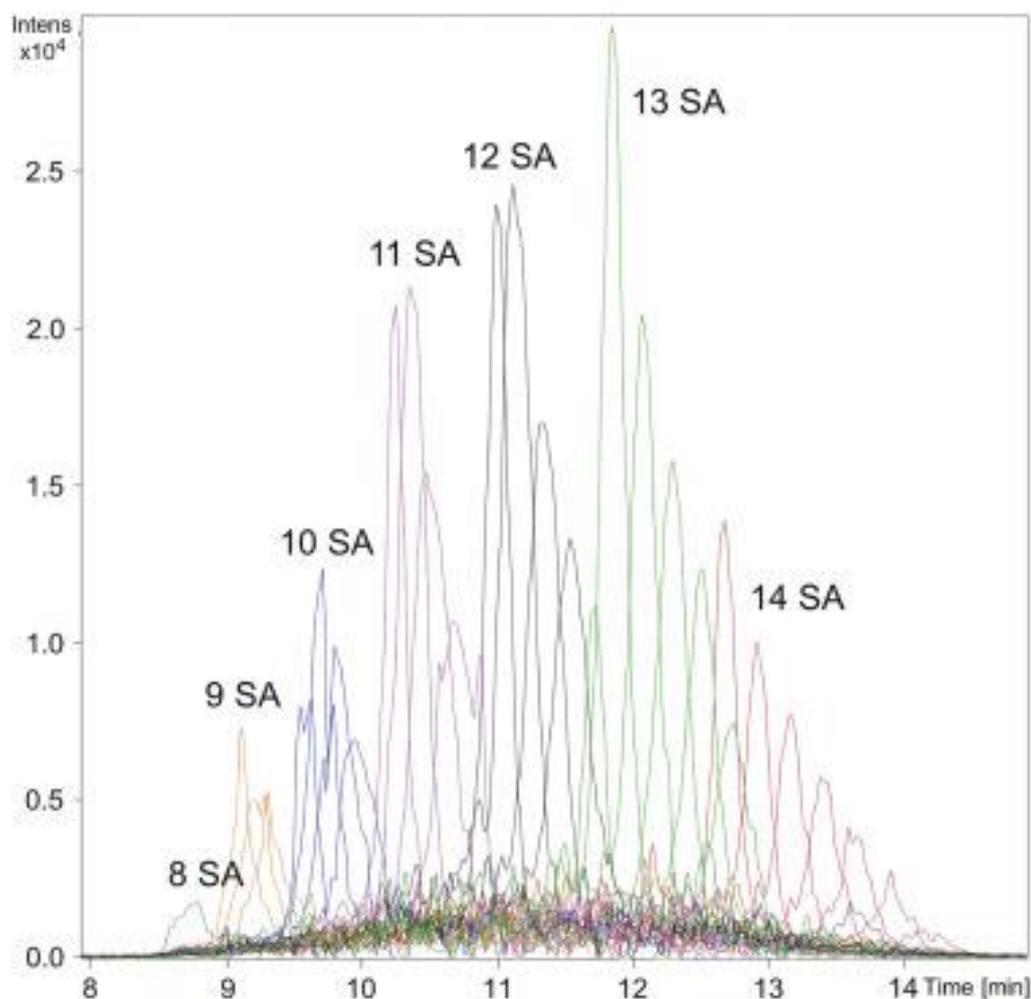
Figure 15.22. CE–MS analysis of erythropoietin reference substance of the European Pharmacopoeia. The traces are extracted ion electropherograms of the most abundant charge state (mostly [M + 15H]15 +-ions). Traces of the same color represent the extracted ion electropherograms of glycoforms with the same number of sialic acids. Within these groups, the glycoforms are separated by the number of HexHexNAc units (repeats).

The components of the electric force perpendicular to the axis of the tube are cancelled by the tube reaction force and the longitudinal components induce an electrophoretic motion of the chain along the tube (forced reptation).

In the limit of very strong electric fields the front end of the DNA chain moves forward and creates new parts of the tube (Figure 15.23 (a)). The stretching force is proportional to the number of monomers (N), since the total electric charge on the chain is also proportional to N. Furthermore, the coefficient of friction (m) for the whole chain is proportional to N (as for reptation, m ∞ N). Thus the speed of motion (v) in a strong electric field is independent of N (v = f /μ). The method of strong field electrophoresis is therefore not useful for the separation of DNA fragments.

However, weak electric field gel electrophoresis (Figure 15.23(b)) is much more successful, since the DNA molecules remain Gaussian coils.
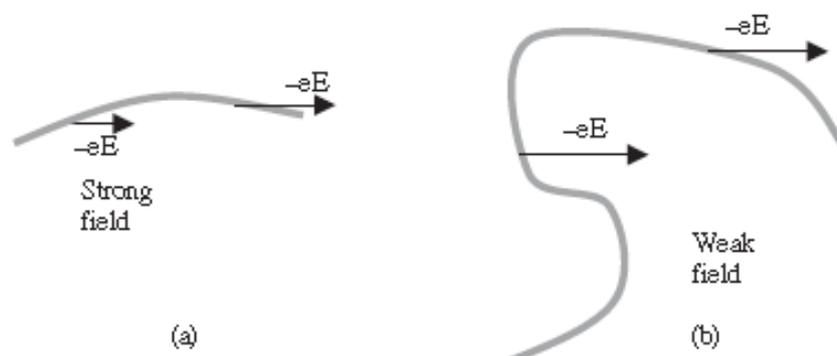
Figure 15.23. Schematic diagram indicating the difference in conformation of charged macromolecules during electrophoresis in a gel ((a) In a strong electric field the chains are completely stretched and electrophoresis is not a sensitive measure of the mobility; (b) in a weak field the chains adopt a Gaussian conformation and mobility measurements are much more successful. eE is the electrostatic force on a section of a chain.).

The force that the electric field exerts on the DNA molecules is proportional to the displacement of the chain parallel to the electric field $N^{\frac{1}{2}}$, since only the motion of segments of the chain parallel to the electric fieldare not restricted by the cross-links of the gel. The speed of reptation (v_γ)

$$v_\gamma = \frac{f}{\mu} \sim \frac{N^{\frac{1}{2}}}{N} \sim N^{\frac{1}{2}} \qquad (15.44)$$

where N is the number of monomers in the chain. The speed of the centre of mass motion (v) is a factor of $N^{\frac{1}{2}}$ slower than the speed of reputation and the speed of centre of mass reptation is therefore inversely

proportional to the chain length ($n - 1 = N$) in a weak field. Weak field electrophoresisthus provides a practical method for the separation of DNAchains. A more accurate calculation of the velocity of the DNA fragments in a weak field gives:

$$v = \frac{q}{3\eta}\left(\frac{1}{N} + w\left(\frac{EqL}{rT}\right)^2\right)E \qquad (15.45)$$

where E is the electric field vector, q is the charge per unit length of the DNA, N is the number of Kuhn segments in the DNA chain, L is the length of a Kuhn segment, h is the viscosity of the medium and w is a constant of the order of unity. The technique is, therefore, not very sensitive at separating long DNA chains (Figure 15.22), which is a big problem if micron long pieces of genomic DNA require sequencing.
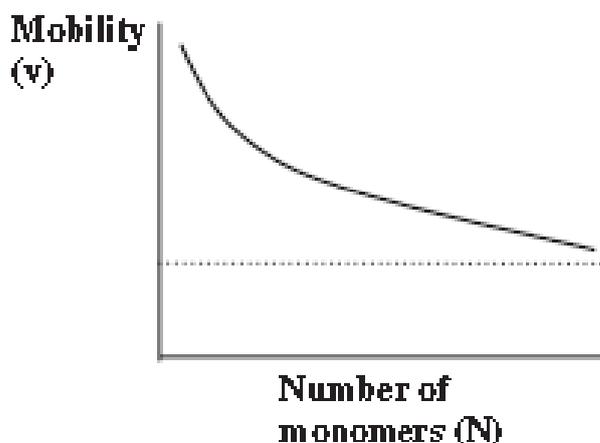
Figure 15.24 Dependence of the mobility of DNA fragments (v) on the number of monomers in a chain (N) during a gel electrophoresis experiment. The mobility becomes a less sensitive function of N as the chains increase in length.

A trick to increase the sensitivity of electrophoresis for the separation of long DNA chains is to periodically switch off or rotate by 90^0, the applied external field at the typical time for the renewal of the reputation tube (t ~N^3). In this case, electrophoretic motion only occurs for chains of the length (N) defined by the periodicity of the rotating electric field and the method can be used to select the longer chains. This technique of pulsed electrophoresis works very well.

The polymerase chain reaction (PCR) is a biochemical technique for amplifying short (10 000 base pairs) stretches of nucleic acid. It is often used in conjunction with electrophoresis methods in the process of genetic fingerprinting. Thus a single DNA molecule can be amplified to provide sufficient quantities of DNA to be sequenced using electrophoresis by means of the PCR technique.

SDS electrophoresis can be used to calculate the molecular weights of proteins. SDS is a surfactant that is an effective protein denaturant. It binds to all proteins qualitatively to the same degree and causes them

toadopt extended conformations. The apparent electrophoretic mobility (c) of a protein denatured with SDS at a particular gel concentration(c) is phenomenologically given by:

$$\ln u(c) = k_x c + \ln u(0) \tag{15.46}$$

where k_x depends on the extent of cross-linking of the gel and u(0) is a constant for a particular protein. The mobility (u(0) is related to the molecular weight (M) of the protein through a simple relationship

$$u(0) = b \quad \alpha og M \tag{15.47}$$

where b and a are standard constants. The molecular weight of a denatured protein can therefore be calculated from the measurement of its mobility on a gel at a series of different gel concentrations.

There have been a number of modern developments in electrophoretic techniques. Problems with convection in free boundary electrophoresis can be reduced by using a very fine bored capillary in capillary electrophoresis (Figure 15.25).
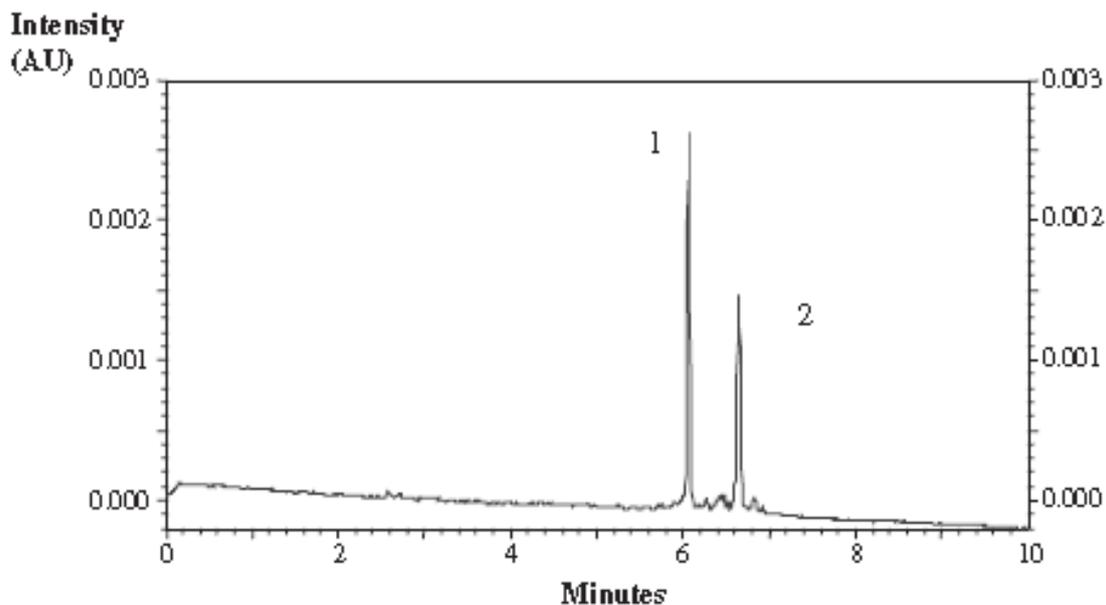
Figure 15.25 Typical data from a capillary electrophoresis experiment that shows the separation of two samples of short double stranded DNA: (1) is AAATTATATTAT/ATAATATAATTT and (2) is GGGCCGCGCCGC/GCGGCGCGGCCC. Sample (1) travels down the column faster than sample (2) [Ref.: I.I. Hamden, G.G. Skellern, R.D. Waigh, Journal of Chromatography, A, 806, (1), 165–168, 1998].

This is a useful microanalytical separation technique. The importance of convection in a fluidic system is described by a dimensionless group, the Rayleigh number (Ra), which is defined as:

$$R_\alpha = \frac{R^4 g}{\eta \alpha} \left(\frac{\Delta \rho}{\Delta \gamma} \right) \tag{15.48}$$

where R is the radius of the channel, g is gravity, η is the viscosity, α is the thermal diffusivity of the medium and $(\Delta \rho = \Delta r)$ is the density change per unit radial distance caused by heating. For small Rayleigh number

$(Ra < 1)$ convection is suppressed in an electrophoresis tube and this corresponds to a small capillary bore (R). Typically, capillary diameters for electrophoresis experiments are in the order of a few microns to provide low Rayleigh number dynamics for the charged molecules examined.

Etched obstacle arrays on silicon chips can also be used for electrophoresis (Figure 15.26). Silicon microarrays offer a number of advantages over standard gel techniques: smaller samples can be explored and the microstructure of the etched silicon can be better defined than with gels and, consequently, so too can the microfluidics.
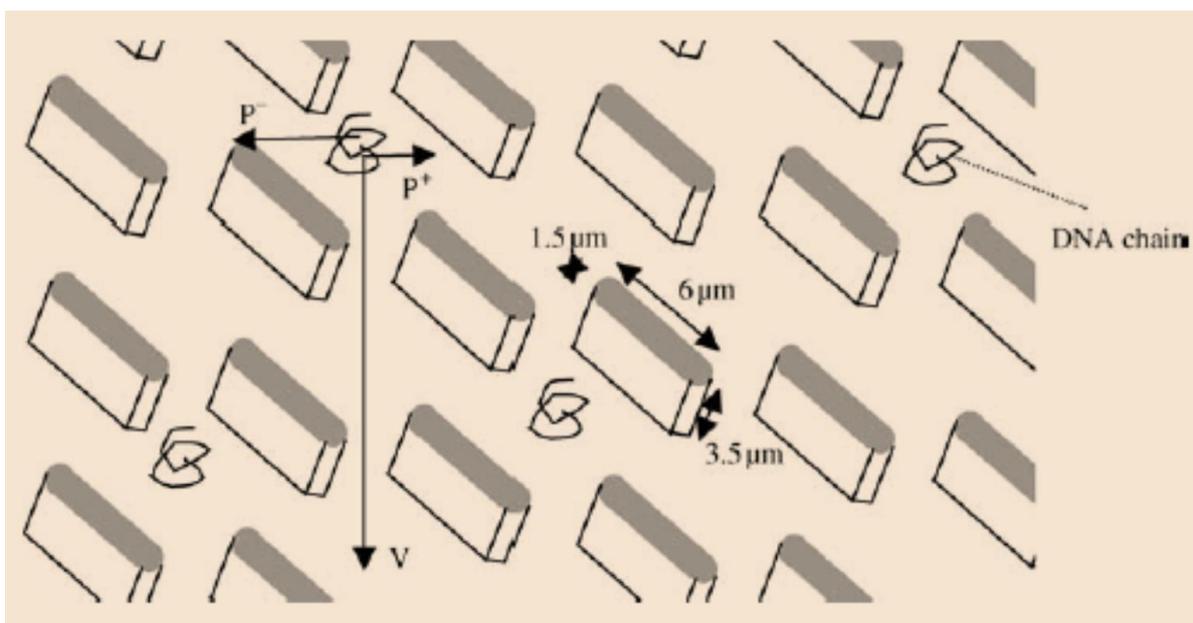
Figure 15.26. Etched microarrays used for electrophoresis experiments. (An electric field moves the DNA molecules vertically downwards with a velocity v. Due to the anisotropic nature of the obstacle orientation there is a larger probability for the molecules to change channels to the right (P⁺) than to the left (P⁻), and thisprobability is a function of the size of the chains [Ref.: C.F.Chou, R.H.Austin, O.Bakajin et al., Electrophoresis, 2000, 21, 81–90]

Sedimentation

Sedimentation is a key separation technique used to extract the particular biomolecule of interest from the complex soup of species found in the cell. Separation by sedimentation is a standard first step in a molecular biophysics experiment. External force acting on a mixture of suspended particles is used to separate them by means of their varying buoyancies with respect to the background solvent. In an analytical ultracentrifuge the radial acceleration provides the external force and causes the molecules to be separated as a function of both their density and shape (Figure 15.27).
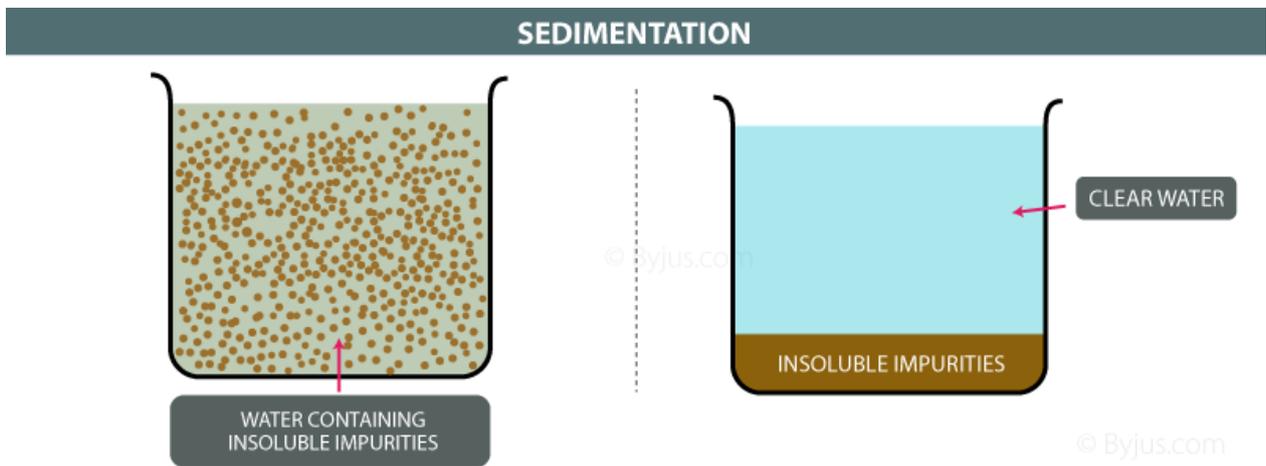
Figure 15.27. Sedimentation experiment (The sample is rotated about an axis at an angular velocity of v and macromolecules adjust their position with respect to the solution due to their relative densities.)

From simple Newtonian mechanics the radial force (F) on a suspended particle that is rotating in an ultracentrifuge is given by:

$$F = m^* \omega^2 \gamma \qquad (15.49)$$

where m* is the effective mass, $\acute{\omega}$ is the angular velocity and r is the radial distance of the particle from its centre of motion. From Archimedes principle, the mass of a particle (m) suspended in a solvent needs to be corrected by the density of the surrounding solvent and the effective mass of the particle (m*) in the solvent is given by:

$$m^* = m(1 \quad \bar{v}\rho_0) \qquad (15.50)$$

where v is the partial specific volume of the molecule and r0 is the solvent density. The velocity at which the particles move due to the centripetal force is given by equation (15.49) divided by the frictional resistance:

$$v = \frac{d\gamma}{dt} = m^* \omega^2 \frac{\gamma}{f\eta_0} \qquad (15.51)$$

where f is a frictional coefficient and η_0 is the solution viscosity. The variation of sedimentation velocity (v) with particle size and density forms the basis of a method to separate particles using sedimentation.

When a centrifugal field is applied to a solution of molecules a moving boundary is formed between the solvent and the solute. This boundary travels down the sample cell with a velocity determined by the sedimentation velocity of the macromolecules. Concentration gradients can be accurately measured using ultra violet absorption (Figure 15.26) and therefore the sedimentation velocities can be calculated.
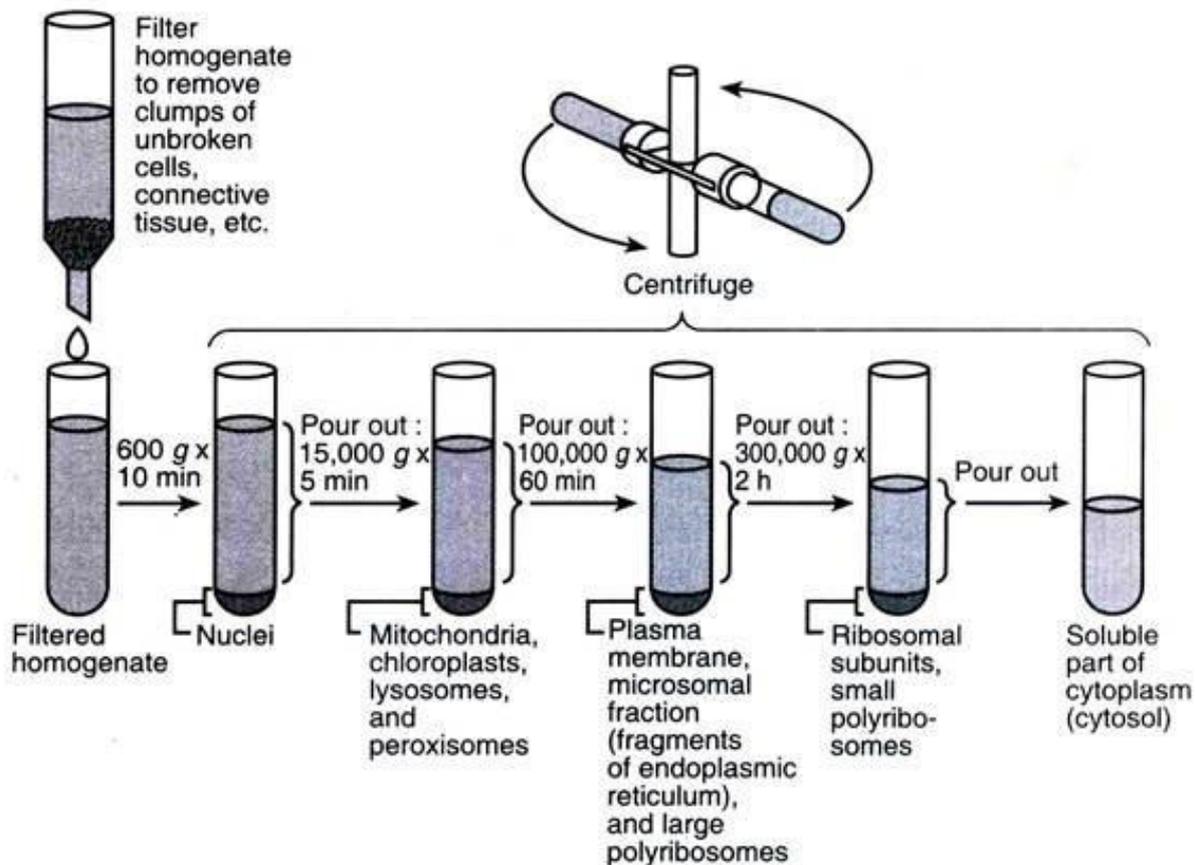
Figure 15.26. Experimental arrangement of a modern analytic ultracentrifuge. Optical spectrometry methods and tools are widely used to measure particle concentrations.

The velocity of sedimentation $\left(\frac{d\gamma_b}{dt}\right)$ is equal to the rate of motion of the boundary:

$$\frac{d\gamma_b}{dt} = \gamma_b \omega^2 s \qquad (15.52)$$

where s is the sedimentation coefficient, equal to the velocity of sedimentation divided by the centrifugal strength $(\gamma_b \omega^2)$ at the radius at which the boundary occurs (γ_b). Integration of equation (15.52) provides an expression for the position of the boundary as a function of time:

$$\ln\left[\frac{\gamma_b(t)}{\gamma_b(t_0)}\right] = \omega^2 s\,(t\ \ t_0) \qquad (15.53)$$

where t_0 is a reference time at which the boundary is found at $\gamma_b(t_0)$. Diffusion broadens the boundary as it progresses down the column and the rate of motion allows the sedimentation coefficient to be calculated from equation (15.53). The sedimentation coefficient depends on the size, shape and degree of hydration of a macromolecule. Fortunately, for globular proteins there is a well defined relationship between the sedimentation coefficient and the molecular weight due to their spherical geometry.

It is also possible to make focusing measurements with sedimentation experiments if the particles are suspended in a dense salt, e.g. a solution of cesium chloride (CsCl) or cesium sulfate (CsSO4).
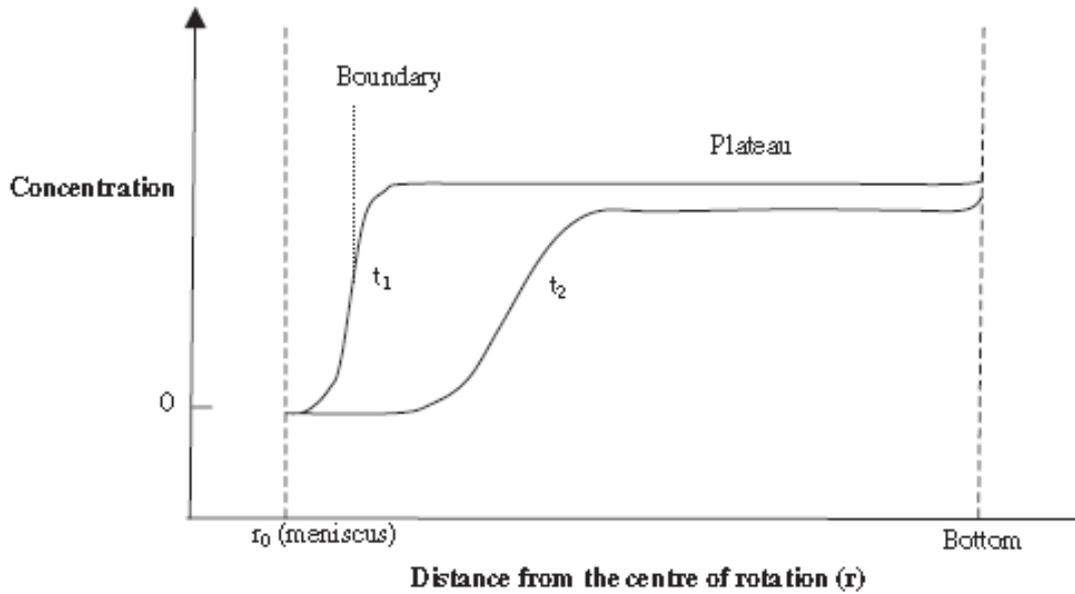
Figure 15.29. Schematic diagram showing the progress of a moving boundary sedimentation experiment at two different time steps (t_1 and t_2). The boundary moves towards the centre of rotation (the cell bottom) as a function of time.

Particles collect together in a narrow band at the point of matching buoyancy. The calculation of the sedimentation profile during a centrifugation experiment is an elegant illustration of the predictive power of equilibrium statistical mechanics. The work required $\{E(\gamma)\,E(\gamma_0)\}$ to lift a particle from a radius γ_0 to a radius γ in a centrifugal force field is equal to the work done against the centrifugal force:

$$E(\gamma)E(\gamma_0) = \int_{\gamma_0}^{\gamma} m'\gamma\omega^2 \, d\gamma = m'\omega^2 \frac{(\gamma_0^2\gamma^2)}{2} \tag{15.54}$$

where $m*$ is the effective mass of the particles adjusted for the solvent density. In thermal equilibrium the range of concentrations $c(\gamma)$ as a function of the radius is given by a simple Boltzmann distribution ($e^{E/kT}$) and therefore:

$$\frac{c(\gamma)}{c(\gamma_0)} = \exp m'\omega^2 \frac{(\gamma_0^2\gamma^2)}{2kT} \tag{15.55}$$

The density near the radius of a particular band (γ_b) can be expressed as a Taylor expansion:

$$\rho(\gamma) = \rho(\gamma_b) + \left.\frac{\partial\rho}{\partial\gamma}\right|_{\gamma-\gamma_b} (\gamma - \gamma_b) + \dots \tag{15.56}$$

Substitution of the density expansion in equation (13.50) allows equation (13.55) to be expressed as:

$$\frac{c(\gamma)}{c(\gamma_b)} = \exp\left[m'\gamma_b\omega^2\bar{v}\rho' \frac{(\gamma - \gamma_b)^2}{2kT}\right] \tag{15.57}$$

where $\rho'(\gamma_b)$ is the density gradient defined by:

$$\rho'(\gamma_b) = \frac{\partial \rho}{\partial \gamma}\bigg|_{\gamma - \gamma_b} \tag{15.58}$$

And mis the true mass of the particles. The concentration profile at radius (γ_b) at which the band of particles occurs is therefore a Gaussian distribution with standard deviation (σ_γ) and is given by:

$$\sigma_\gamma = \frac{kT}{[m'\gamma_b\omega^2\bar{v}\rho'(\gamma_b)]^{\frac{1}{2}}} \tag{15.59}$$

The band of particles is narrow and well - focused for particles of large mass (m), in high centrifugal fields (large $\gamma_b\acute{\omega}^2$) and in steep density gradients $\{large\rho'(\gamma_b)\}$. Sedimentation focusing is thus another extremely useful technique for particle separation.

Rheology

All real materials demonstrate behavior intermediate between the idealized cases of solids and liquids. Rheology is the study of this phenomenon of viscoelasticity and rheometers are instruments for measuring the rheology of materials.

There are two broad categories of techniques for measuring the viscoelasticity of a material. Firstly, there are bulk methods where the response of a macroscopic amount of a material to an externally applied stress or strain is recorded. These bulk methods have traditionally been used to examine the viscoelasticity of biological samples. Secondly, there is the measurement of the viscoelasticity of a sample as a function of length scale using microrheology techniques. Here probes are typically

injected into the system of interest; the probes can be passive (e.g. markercolloids) or active (e.g. magnetic colloids). The motion of the probes is recorded with a video camera or measured with light scattering and the resultant fluctuation spectrum of the particle displacements is related to the viscoelasticity of the material in which they are embedded.

In bulk rheology experiments a series of different geometries can be used and each tends to have different advantages in terms of the mechanism of sample loading, the time window that can be explored and the sensitivity of the measurements (Figure 15.30). Different geometries for rheometers require different corrections to analyze the dependence of the stress on the strain. How the geometry grips the sample is also important and, combined with the type of force or displacement transducers, this determines the sensitivity of the measurements. Bulk rheometers measure the large - scale viscoelastic properties of assemblies of biological molecules. Rheometers can function in linear or non-linear modes. Non-linear rheology corresponds to large deformations and deformation rates, in which both Deborah and Peclet numbers are appreciable. In drag flow rheometry the velocity or displacement of a moving surface is measured simultaneously with the force on another surface that moves in response to its motion. Couette's original

concentric cylinder drag flow rheometer was a controlled strain apparatus. The angular velocity of the outer cup was fixed and the torque on the inner cylinder was measured from the deflection of a suspended wire. The measured variable in a controlled strain rheometer is the torque. Couette measured the twist in a torsion bar whereas modern electronic rheometers use a linear variable differential transducer to do the same job (Figure 15.31, 15,32).
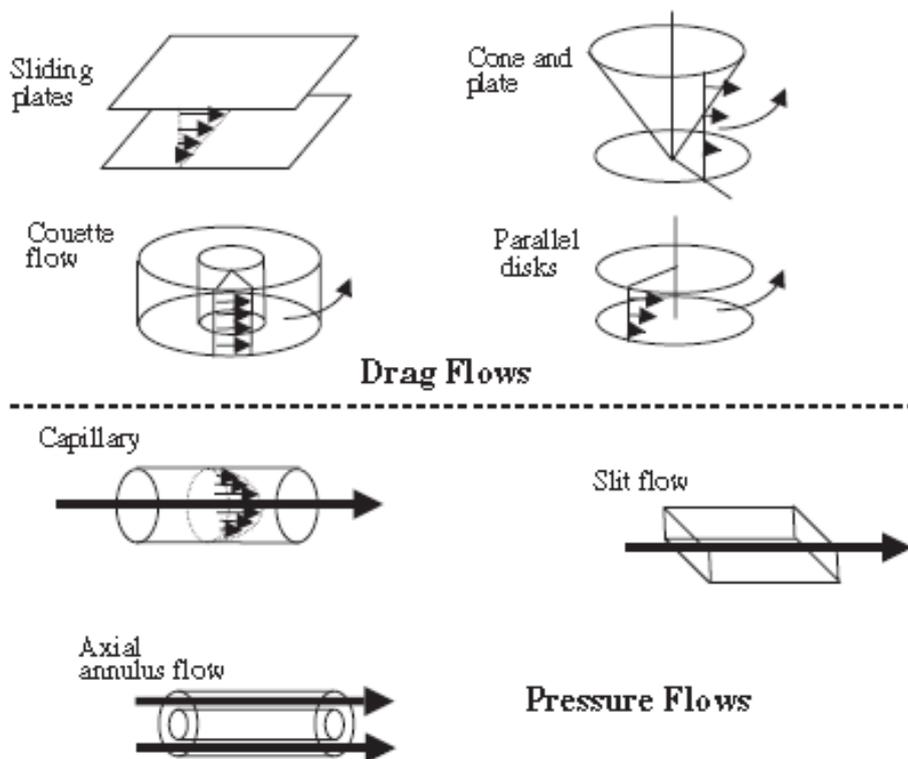

Figure 15.30. Schematic diagram of the flow geometries commonly encountered in rheological experiments.(The geometries are separated between drag flows where the surfaces move relative to one another, and pressure flows in which the flow rate is determined by the pressure drop across the pipe [Adapted with permission from C.W. Macosko, Rheology, Principles, Measurements and Applications, Copyright (1994) Wiley-VCH]).
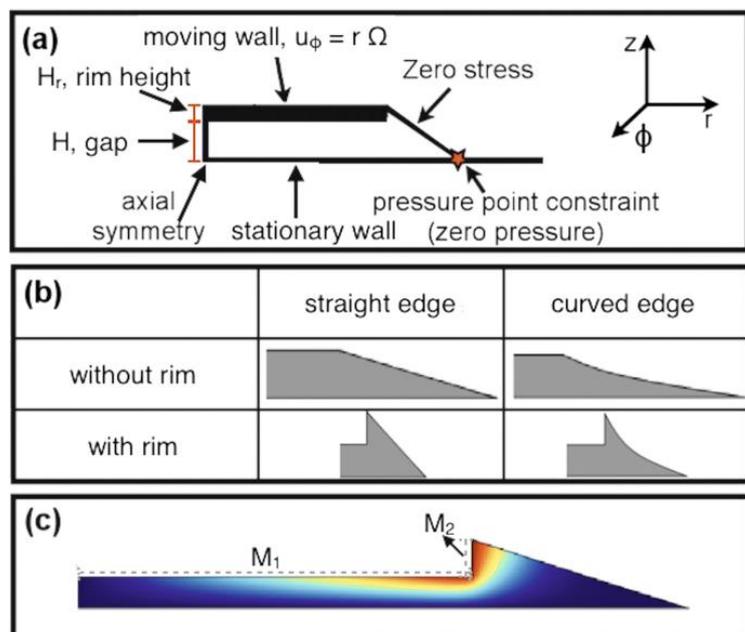
Figure 15.31. Schematic representation of simulation boundary conditions. b Four different configurations of the sample edge at the plate rim (not to scale). c Velocity distribution in the sheared sample with the two torque contributions indicated, where M1 is due to the sample underneath the upper plate and M2 is due to the sample in contact with the outer rim of the upper plate. Source: Ruth Cardinaels, Naveen Krishna Reddy, Christian Clasen. Quantifying the errors due to overfilling for Newtonian fluids in rotational rheometry, Published: 06 July 2019 Rheologica Acta volume 58, pages525–538(2019)
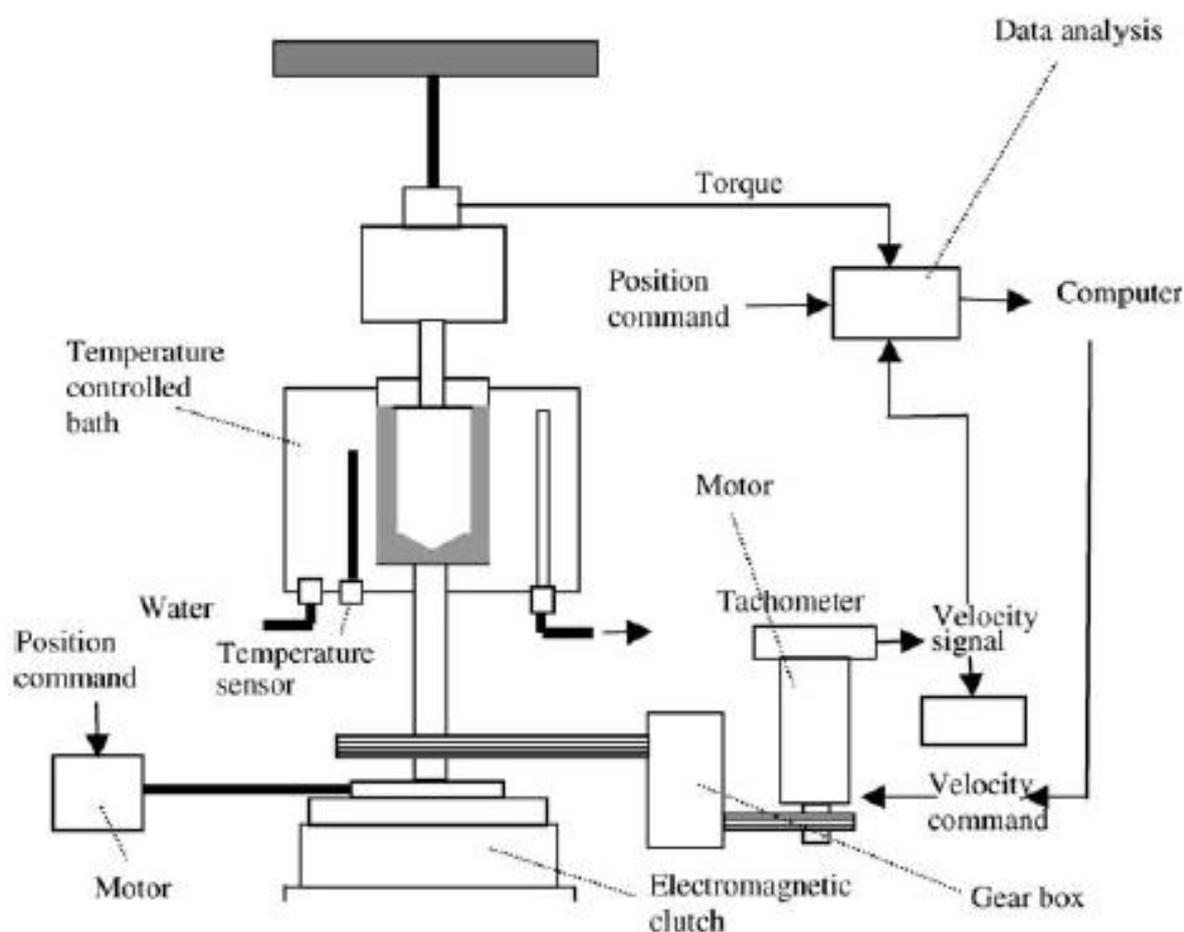
Figure 15.32 Schematic diagram of a modern shear rheometer. [Ref.: L. Bohlin, in Progress in Trends in Rheology, II, Eds H. Gieseku and M.F. Hibberd, Steinkopf, 1988, 151].

More sophisticated modern rotary rheometers measure normal stresses (the stresses normal to the direction of shear). It is an experimental challenge to find steady state behaviour with normal stresses, as they are easily disturbed by fluctuations in the temperature and axis of rotation. Rheometers thus need to be machined to high precision. Often commercial rheometers are mechanically accurate to within 2 mm over the 25mm cup diameter. The control of the torque and the subsequent measurement of the angular motion in a controlled stress rheometer is also a standard technique in rotational rheometry. Furthermore, it is important to control the temperature, pressure and humidity to make accurate rheological measurements with biological specimens.

The most commonly measured linear viscoelastic material function is the complex shear modulus, $G^*(\acute{\omega})$. There are three standard techniques used to measure G^*: in the shear wave propagation method the time for a pulsed deformation to travel through a sample is measured; the sample can be made to oscillate at its resonant frequency, and the response at this single frequency observed; and the forced response to a sinusoidal oscillation in stress/strain can be measured in terms of the resultant strain/stress (Figure 15.33). Forced resonance devices are better suited to low elasticity materials such as polymer solutions and soft biomaterials.
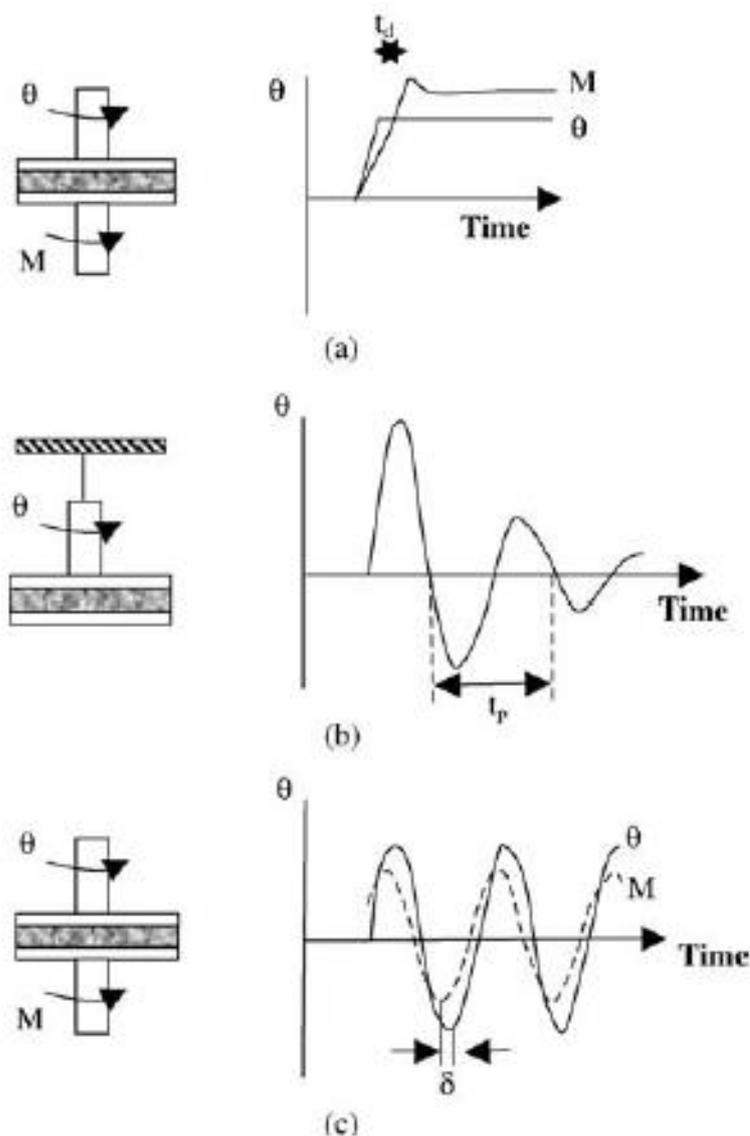
Figure 15.33. Methods for the measurement of the shear modulus: (a) wave speed, (b) resonance and (c) forced oscillations [Adapted with permission from C.W. Macosoko, Rheology, Principles, Measurements and Appications, Copyright (1994) Wiley-VCH]).

There are two basic design types of pressure driven rheometers. One features control of the pressure and measures the flow rate (e.g. capillary rheometers) and the other uses a controlled flow rate and measures the pressure drop. Such capillary type geometries have direct analogues in biological circulatory systems (e.g. blood flow) which motivate the analysis.

Microrheology has experienced a number of important recent developments. Often biological samples, which are homogeneous on the macro scale, are inhomogeneous on the micron scale (e.g. inside living cells); a range of microrheology techniques have been developed to measure this behaviour. The range of frequencies and moduli that can typically be accessed using the different microrheology techniques are shown in Figure 15.34. In particular, the measurable frequency range can be increased by many orders of magnitude using microrheology when compared with standard bulk rheology methods.

Particle tracking microrheology is, practically, the simplest microrheology technique to implement. It requires a video camera, optical microscope, oil immersion objective and digital recording apparatus.
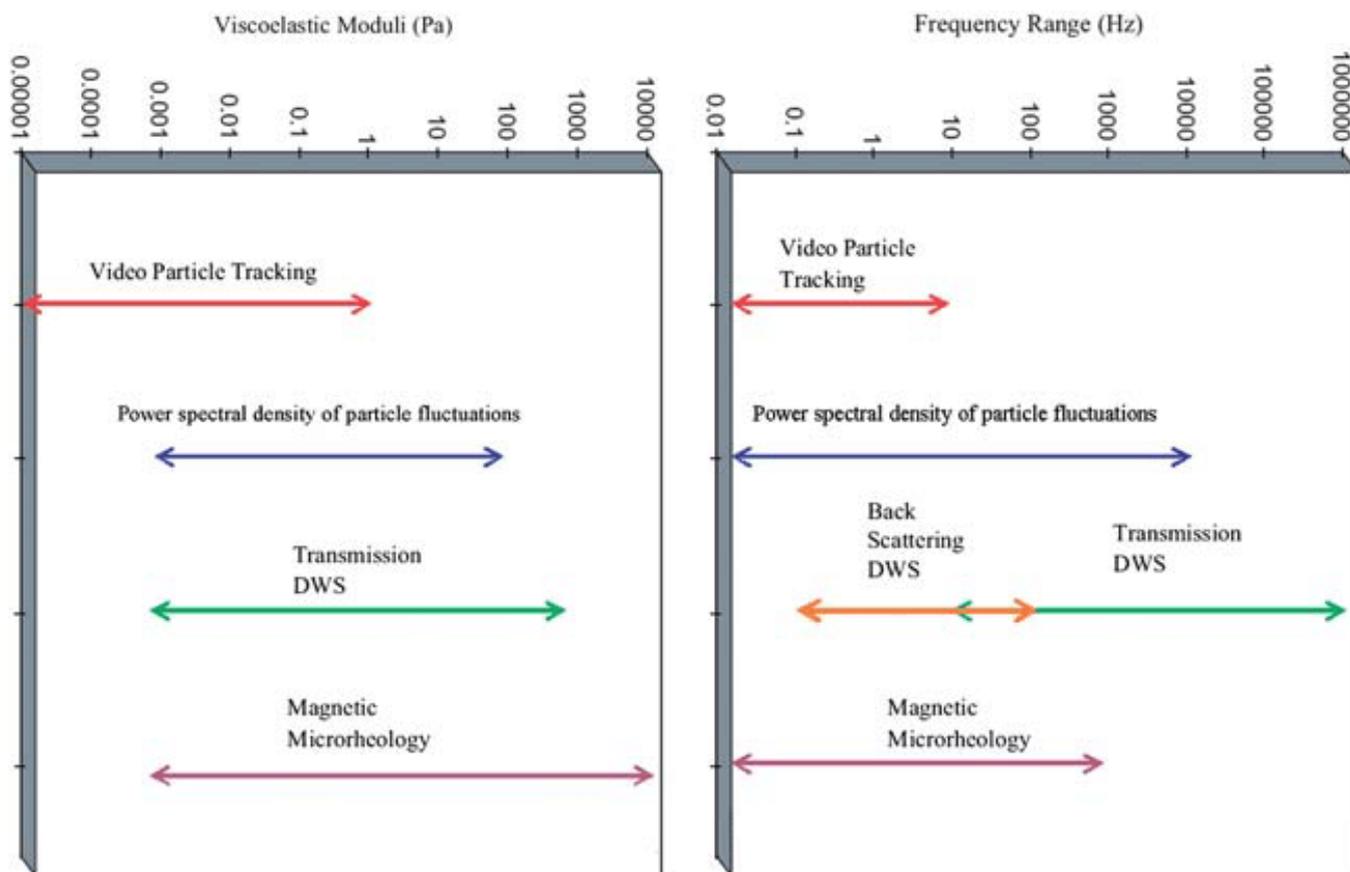
Figure 15.34 Typical ranges for the shear viscoelastic moduli (elastic G 0 (u) and viscous G 00 (u) modulus) and frequency that can be measured in fluid soft materials using microrheological methods. (Modified figure from Fig. 1 in T. A. Waigh, Rep. Prog. Phys., 2005, 68, 685.)

The fluctuation–dissipation theory is used to relate the fluctuations in the displacements of tracer particles embedded in a material to its viscoelastic response. The fluctuation spectrum of the mean square displacements of colloidal particles embedded in a viscoelastic material is calculated as a function of time (Figure 15.35).
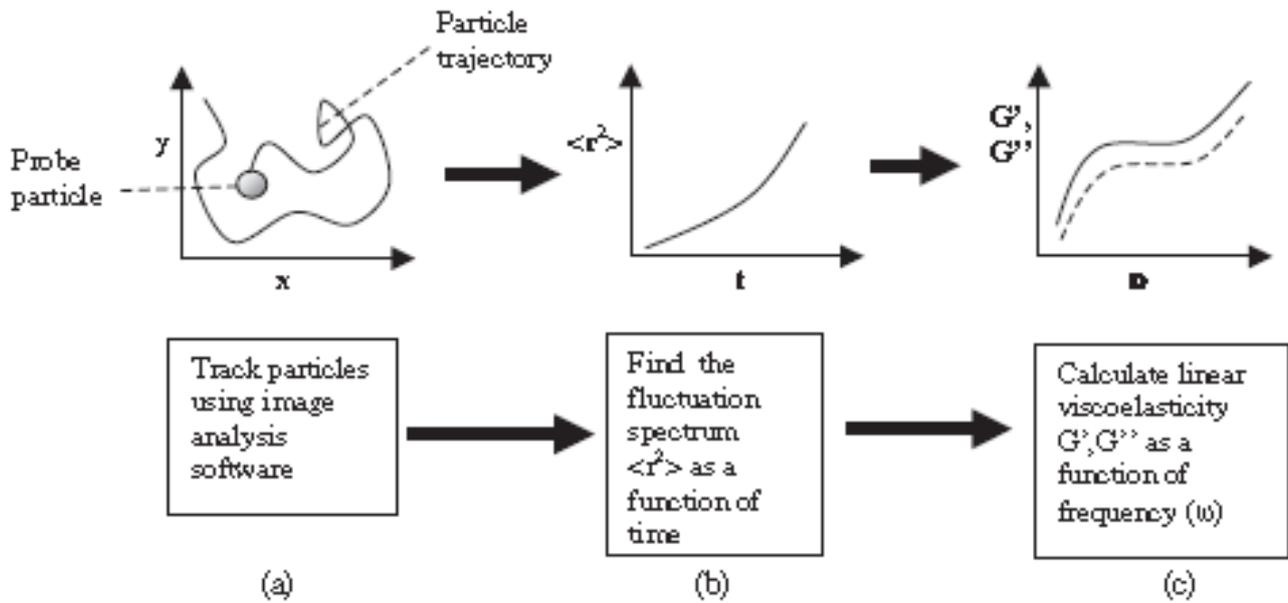
Figure 15.35. The strategy used in passive particle tracking micorheology experiments: {(a) The trajectory of a fluctuating colloidal sphere is recorded, (b) the mean square displacement (γ^2 (t) fluctuations are calculated and (c) the shear moduli (G^0,G^{00}) are found using the generalised Stokes–Einstein equation}.

For simple viscous liquids one would expect a linear dependence of the mean square displacement in two dimensions of the embedded probes (γ^2) under the microscope on time (t):

$$(\gamma^2) = 4Dt \qquad (15.60)$$

A viscoelastic component is introduced as a sub-linear diffusive process at short times modifying equation (13.60) ($r^2 \sim t$, $a < 1$) and includes the information on the viscoelasticity of the material. The linear viscoelastic shear moduli (G*) of the material can subsequently be calculated from the mean square displacement using the generalized Stokes– Einstein equation (compare with equation (5.10) :

$$G(s) = \frac{kT}{\pi a^2 s \left(\gamma^2(s)\right)} \qquad (15.61)$$

where γ^2 (s) is the Laplace transform of (γ^2 (t), $G(s)$ is the Laplace transform of the relaxation modulus $G(t)$ and a is radius of the probe particle. The Laplace frequency (s) has been introduced to provide a compact solution of the fluctuation dissipation theorem. $G'(\omega)$ and $G''(v)$ can be determined mathematically by Fourier transform of the relaxation modulus as a function of time $G(t)$.

It is also possible to determine the linear viscoelastic spectrum by the analysis of the mean square fluctuations in angular displacement of a probe as a function of time, and again a generalized Stokes–Einstein equation for rotational motion is used to calculate the complex shear modulus.

Laser deflection techniques allow the high frequency viscoelastic behavior of materials to be probed. Back focal plane interferometry is a particularly sensitive laser deflection method for measuring the small fluctuations (nanometres) of probe spheres that occur at high frequencies (Figure 15.36).
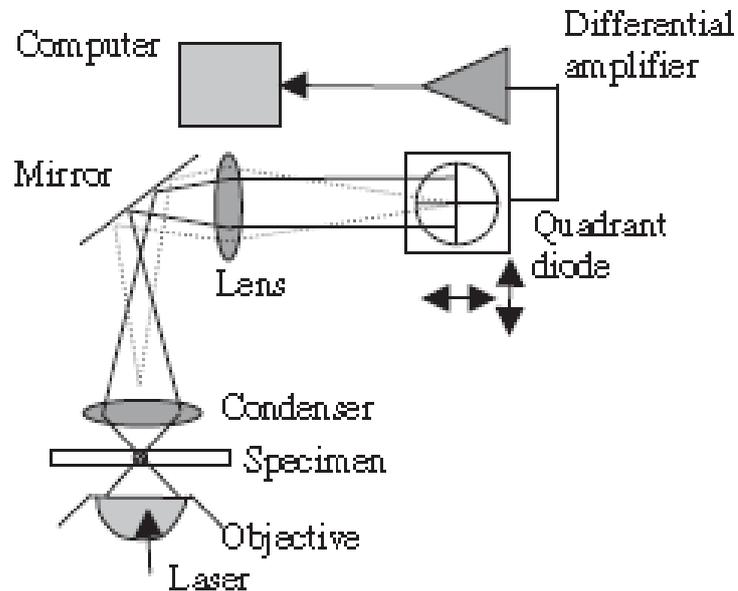

Figure 15.36. The fluctuations of the displacement of a single bead embedded in a biological specimen can be followed with scattered laser light projected onto a quadrant diode. This provides a sensitive laser scattering microrheology technique.

Multiply-scattered laser light from colloidal spheres also can be used to yield the fluctuation spectrum of colloidal spheres embedded in biological specimens through the technique of diffusing wave spectroscopy (DWS, Figure 15.37).
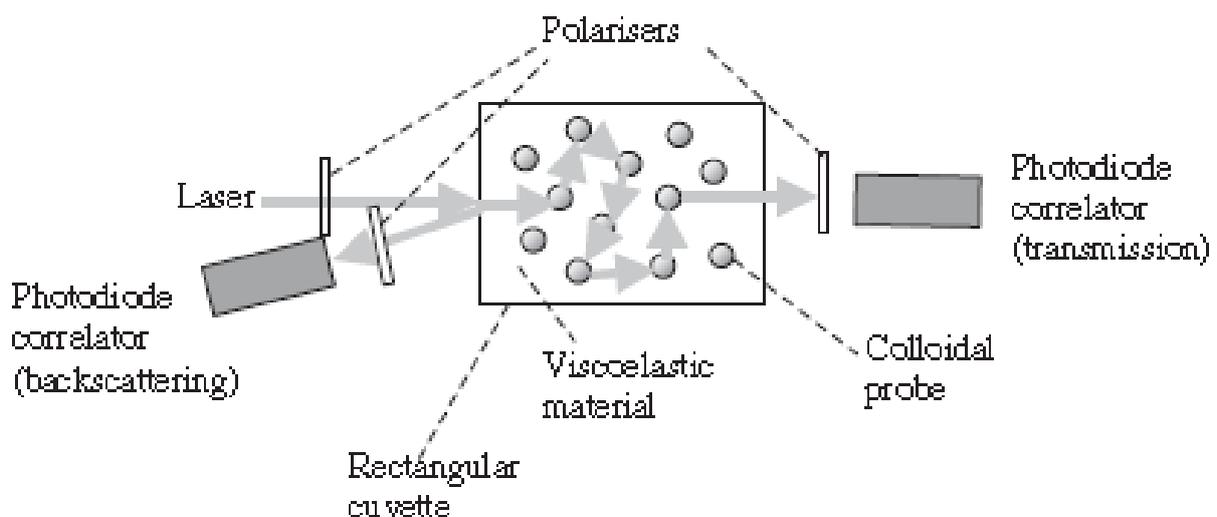
Figure 15.37 Schematic diagram of a diffusing wave spectroscopy experiment. (Coherent laser light is multiply-scattered from a dense suspension of colloidal particles. Analysis of the resultant correlation functions can provide the high frequency visoelasticity of a biological specimen).

The intensity correlation function is measured as the autocorrelation of the scattered intensity and is used to construct the mean square displacement $r^2(t)$ of the probe spheres. The viscoelastic moduli can then be calculated in a similar manner to the particle tracking technique. DWS microrheology is useful for ultra high frequency viscoelastic measurements, since the process of multiple-scattering amplifies the sensitivity of the measurements to small particle displacements (A°) and allows particle motions to be detected at high frequencies (MHz). Single scattering photon correlation spectroscopy techniques can also be used at lower colloidal concentrations and provide information on particle motion at slightly lower frequencies.

DWS microrheology methods enable a very wide range of frequencies for the linear rheology of solution state biological materials to be accessed (Figure 15.38).
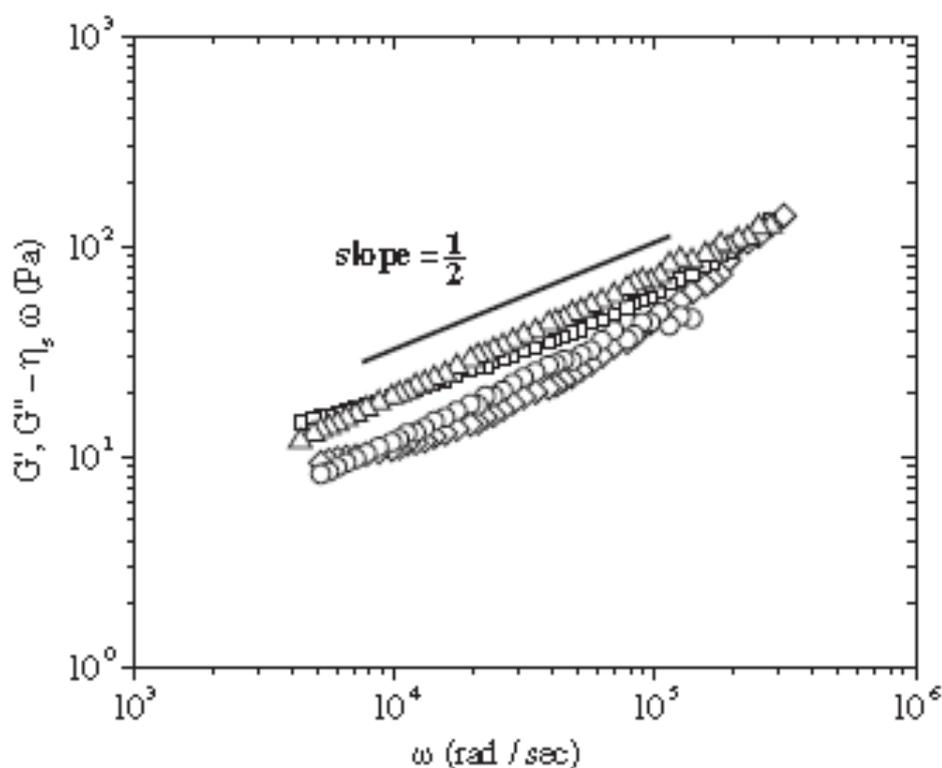
Figure 15.38. High frequency linear viscoelasticity from aggrecan solutions.(The slope of G'~ G''~$\omega^{1/2}$ indicates that Rouse modes are present in these flexible polyelectrolyte solutions at high concentrations. Two separate concentrations are shown 2 mg/ml and 12 mg/ml [Ref.: A.P.Papagiannopoulos T.A.Waigh, T.Hardingham and M.Heinrich, Biomacromolecules, 2006, 7, 2162–2172]).

Both transmission and back scattering geometries are possible for DWS experiments (Figure 15.38). Optical tweezers also find many applications in microrheology studies and are particularly well suited for measuring the elasticity of membranes due to their low modul. Further reduction in sample volumes for nano and pico rheology are possible, but data analysis often becomes more difficult and can reduce the sensitivity of the methods. Examples of submicrolitre rheometers that are currently being investigated are

backscattered DWS with optical fibres (sensitive to picolitre volumes), fluorescent correlation spectroscopy (sensitive to picolitre volumes) and oscillatory AFM (sensitive to nanolitre volumes). Other micromechanical techniques specialize solely in the measurement of the elasticity of biological systems and neglect the behaviour of the viscosity. These include micropipette aspiration, steady state deformation using AFM and the use of internal markers to drive or record the deformation of the cytoplasm, e.g. magnetic beads or fluorescent markers.

Tribology

A range of tribometers have been developed to quantify frictional behavior at surfaces. In a typical modern device adapted for the measurement of solid–solid friction mediated by a thin viscoelastic film, forces are obtained in a direct manner through the measurement of the deflection of a spring with nanometre resolution (Figure 15.39). The stiffness of the employed bending beam is known exactly. The instrument is calibrated both in the normal and tangential direction.
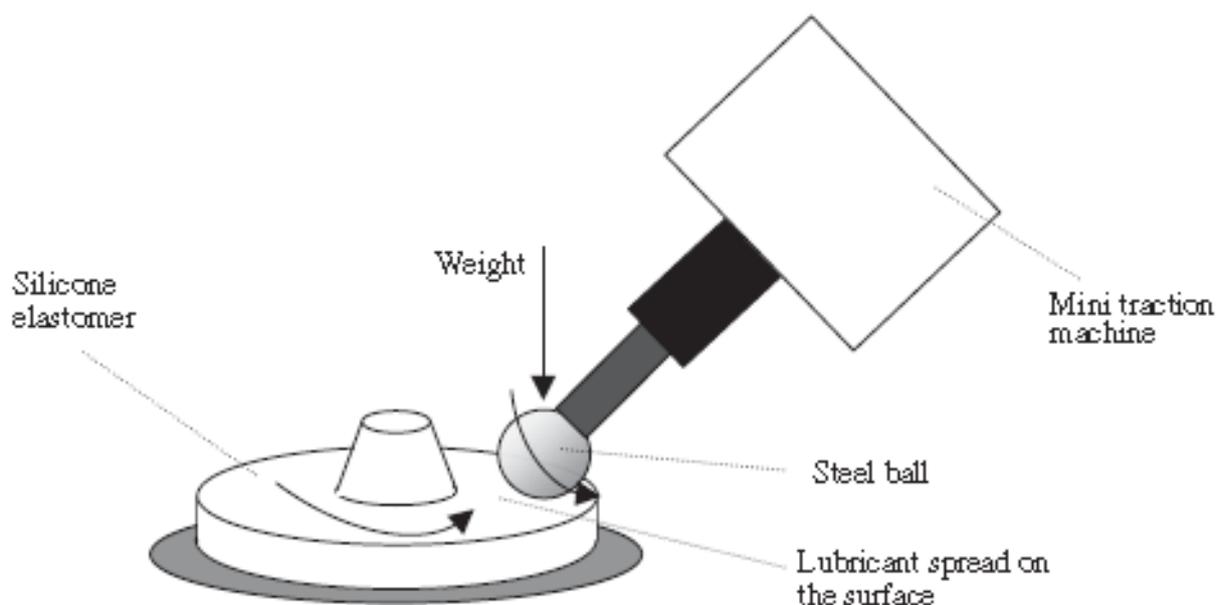
Figure 15.39. Modern ball on plate tribometer used to measure the friction coefficient of thin viscoelastic films as a function of shear rate (Stribeck curves). [Ref.: J. de Vicente, J.R. Stokes and H.A. Spikes, Tribology International, 2005, 38, 515–526].

The force measurement has a resolution of nN in the range nN to mN in both directions. The springs for force measurement are often made of photostructurable glass and the spherical ball probes are made of silicon or steel with a well -defined diameter. Interferometers can be used to measure the deflection of the spring. With biological specimen challenges are presented by their non-planarity and the requirement for hydrated environments, e.g. the cartilage in articulated joints. AFMs are sometimes used to measure frictional forces, since they are not confined to planar specimens. However, quantitative measurements of frictional coefficients with AFM continue to be challenging, since it is difficult to infer both the normal and

frictional forces simultaneously using light scattered from the back of a cantilever. Drag flow rheometers can also be adapted to provide high precision frictional measurements (e.g. a plate–plate rheometer with a section of material attached to eitherplate), but the utility of the technique is dependent on the geometry of the specimens matching that of the cell.

Solid Properties

Solid materials with high elasticity and minimal flow behaviour require a separate set of techniques for their measurement, since extremely large forces must be applied to provide significant sample displacements (Figure 15.40).
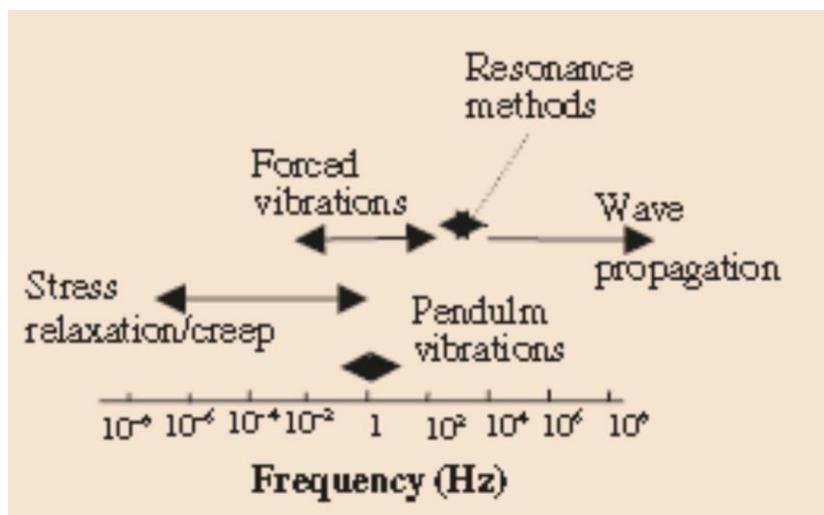
Figure 15.40. Methods for the measurement of the linear viscoelasticity of solid materials over different frequency ranges.[Ref.: Becker, Mater. Plast. Elast., 1969, 35, 1387].

Dynamic mechanical testing apparatus (DMTA) are compressional analogues of oscillatory shear rheology that are often used on solid biomaterials. DMTA can provide the complex Young's modulus (E) of a material in compression or extension as a function of frequency. Highly anisotropic biomaterials provide a challenge for the experimentalist, since a large number of parameters need to be measured to fully characterize the stress response of the oriented material. The relative orientation of the applied stress and resultant strain needs to be carefully monitored. Other material properties are also important for the mechanical properties of biomaterials such as how samples buckle under compressive stress (measured using three point Euler buckling apparatus) and indentation tests for fracture mechanics.

Examples of Modern Methods and Technics

Nuclear Magnetic Resonance (NMR) Spectroscopy

Machine called an NMR Spectrometer, which processes using electromagnetic radiation, such as radio waves, **and** pulses a radio wave at the sample, which excites the atoms within the sample. At specific frequencies, the atoms will resonate, and return a signal. These signals are specific to certain kinds of atoms and change depending on how the atoms are bonded. The returning frequencies detected by the sensitive machine are therefore specific to each molecule. This information can be used to determine the size and shape of the original molecule.

Each singular atom gives off enough of a resonance shift to be traced by NMR machinery, and this is greatly dependent on and affected by the effective magnetic field at the nucleus. This resonance changes as an atom forms more connections with other atoms. Using these resonance patterns to begin to understand the details of a molecule's three-dimensional structure and the functional groups that form it.

Figure 15.41. Nuclear Magnetic Resonance (NMR) equipment

Compared with mass spectrometry, NMR spectroscopy contains several advantages. First of all, there is no need to purify or derive a substance before analysis. NMR is able to measure and observe all of the atoms within t he sample. Second, NMR is a quantitative measurement and therefore does not need as much analysis as mass spectrometry to decipher the results. The resonance readings we obtain from NMR spectroscopy lends us the ability to decipher an unknown compound's molecular structure and its "purity." Purity is defined as a substance which contains only one kind of molecule. Sometimes, solvents or other chemicals may stay within a sample and decrease its purity. If we were testing a mystery substance, then, we would compare the spikes in our NMR readings to the vast spectral libraries in

existence and then make inferences about the compounds basic structure. NMR readings will reflect the relative proportion of said atom in our sample's **s**keleton. Therefore, we can estimate how many hydrogen atoms or methane groups are in our unknown compound based on the relative length of the spike compared to those given off by the other atoms. NMR spectroscopy also gives us information about the relative position of our atoms. A hydrogen atom can give off several different resonance signals depending on its neighboring atoms or groups. For example, a hydrogen atom located next to a polar group, such as an oxygen-containing carboxyl group, will give off a higher NMR reading than a hydrogen atom neighbored by non-polar methane groups.

The guiding principle behind NMR lies in the fact that nuclei have two special properties: they can spin and they are charged. These properties cause nuclei to react to a magnetic field. If we were to run an electric current or apply an external magnetic field, this would allow an upward energy transfer to a higher state. This energy transfer is reflected at a certain wavelength and radio frequency. Once the spin returns to its baseline, the emitted energy from the drop will be read by the NMR machinery. Likewise, the application of an external magnetic field can cause a nucleus to either spin in the same direction, or against the direction, of the magnetic field. A nucleus with the opposite spin will have higher energy. A nucleus spinning in the same direction will have lower energy. Furthermore, the resonant frequency as it applies to NMR spectroscopy will be affected by the electron shielding. The general rule is that the more electronegative the nucleus is, the higher the resonant frequency expected. Further, the more electronegative, or "electron withdrawing" a group is the lower the chemical shift. On the other hand, the most "electron donating" groups will have the highest chemical shifts. The rise in chemical shift can be due to many factors, including the delocalization of current that occurs in aromatic groups that can distribute current across the groups.
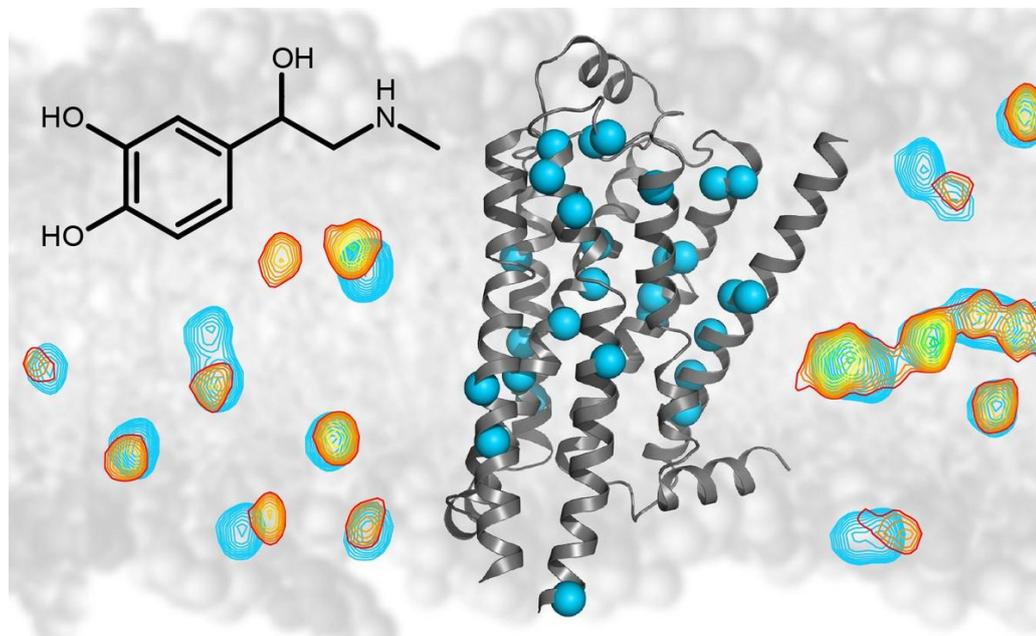
Figure 15.42. Nuclear magnetic resonance (NMR) spectroscopy provides detailed information about the structure, dynamics and the interactions of molecules. Illustration of structural data that can be obtained from NMR experiments. Source: Biozentrum - Universität Basel. Research Group Stephan Grzesiek,

NMR spectroscopy has a wide variety of applications in the study of molecules and molecular processes across many disciplines including chemistry, biochemistry, medicine, pharmacy, physics, engineering, plant biology and soil science, sports science and marine archaeology. NMR spectroscopy uses methodology from which magnetic resonance imaging (MRI) is based.

NMR uses a fundamental property of atomic nuclei to observe nuclei from specific isotopes that are NMR (e.g. ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P). This property is spin angular momentum and is usually referred to as 'spin'. Not all isotopes exhibit spin, for example, ^{12}C - the most abundant carbon isotope, has a spin of zero and is NMR inactive. This forces NMR practitioners to observe the nuclei from the 1.09% abundant ^{13}C isotope.

NUCLEAR: NMR detects the excitation and relaxation of 'spins' of atomic nuclei from stable isotopes: for example ^{1}H is the isotope hydrogen that is 99.995% abundant in nature.

MAGNETIC: NMR spectrometers are essentially large radio transceivers that can excite and manipulate the spin angular momentum from an isotope of interest whilst it is residential in a magnetic field.

RESONANCE: The process of resonance provides a transfer of radio frequency based energy to the nuclei of interest to enable direction.

The power of NMR spectroscopy exists in the wealth of information available from manipulating nuclear 'spins'. The most common piece of information is the **chemical shift** where chemical environment and structure dictate the positions and nature of NMR peaks in a spectrum. This is used to confirm the bonding arrangement or nature of a chemical compounds or alternatively, it can help solve structures of new compounds, natural products and biomolecules. NMR has powerful applications in monitoring complex mixtures including metabolomics mixtures and can be applied to the study of biological and biomedical processes. This is assisted by the ability of NMR to measure quantitative amounts when set-up correctly.

Biological NMR Solution-state biological NMR spectroscopy applications are available 500 and 600 MHz with each spectrometer providing specific solutions.

500 MHz: biological diffusion, peptide NMR, ligand drug screening
600 MHz: metabolomics, ligand drug screening, peptide NMR, natural products, vitamins and metabolic

pathway intermediate analysis, protein triple-resonance (H/C/N NMR), protein interactomics and structural biology.

Capabilities exist at 600 MHz to use 5 mm and 1.7 mm TXI probes with sample volumes of 0.040, 0.250 and 0.650 mL. We can optimize experimentation for your needs including the study of biomolecular inter-atomics (protein-protein, protein-ligand, protein-DNA, protein-RNA) and biomolecular dynamics cross a variety of timescales: s, ms, μs, ns, ps. We can advise on recombinant isotopic enrichment of proteins and peptides in addition to custom designing biological NMR experiments for single samples, detailed projects or larger scale drug/pharmaceutical screening libraries.

Metabolomic experimentation exists at 600 MHz using the 24-sample changer with baseline optimized 1D 1H NMR for profile analysis. Natural product and pathway intermediate analysis (eg. Vitamins and metabolic precursors) analysis is available that can utilize small sample volumes.

Gamma knife method and tool

Recent advances have improved the effectiveness, decreased the complications, and expanded the implications of radiation therapy. These advances include three-dimensional conformal radiation therapy, intensity-modulated radiation therapy, stereotactic radiotherapy, brachytherapy, and radioimmunotherapy. Each of these modalities has improved radiation targeting, thereby limiting radiation exposure of healthy tissues. The way radiation therapy is administered has also changed. Although traditional external beam radiation therapy is administered daily over several weeks, stereotactic radiotherapy may be administered as a one-time treatment. Radioimmunotherapy is administered intravenously. Contemporary radiation techniques also have distinct toxicity profiles. The high radiation doses employed during stereotactic radiotherapy have been associated with obliteration or obstruction of tubular structures, such as bronchi and bile ducts, limiting its use near these tissues. Radioimmunotherapy may be complicated by anaphylactic reactions during and following infusions. As more patients are diagnosed with cancer and as these patients live longer, primary care physicians will increasingly care for those who have received radiation therapy.


Figure 15.43. Scheme of radio isotopic surgery.

Pituitary adenomas represent 10 to 20% of all intracranial lesions. One of the most challenging pituitary adenomas from a treatment standpoint is ACTH-secreting adenomas. These tumors can be difficult to discern on MR imaging, and, even in the hands of an experienced neurosurgical team, some patients may not achieve remission after transsphenoidal or transcranial resection of the adenoma.

Historically, external beam radiation therapy was the adjuvant treatment of choice for recurrent or residual pituitary adenomas. More contemporary data has demonstrated serious limitations of conventional radiotherapy in treating Cushing's disease– the time to hormonal normalization is slow and the rate of post-procedural hypopituitarism is high. Although Cushing's disease is a serious health problem if left untreated, many physicians became concerned that with radiation therapy that the "cure" was worse than the disease.

Stereotactic radiosurgery was investigated as a treatment option for persistent Cushing's disease. The advantages of stereotactic radiosurgery over conventional radiation therapy include a higher biologic equivalent dose, steep fall off, image guidance, and small treatment volume. Moreover, the serious risks of hypopituitarism, stroke, and other radiation induced injury appear to be substantially less with radiosurgery.

Shortly after radiosurgery's development, Lars Leksell treated patients with pituitary tumors using the Gamma Knife; original localization was achieved through use of plain radiographs of the sella. With more recent advances in neuro-imaging including Positron Emission Tomography (PET) scans and fat saturation MRI, the Gamma Knife and other radiosurgical devices have been utilized to control adenoma growth and normalize hormone over-production for many Cushing's patients.

Gamma Knife radiosurgery (GKRS) represents a minimally invasive neurosurgical option for the treatment of Cushing's disease, and it is usually reserved as a second line treatment following an initial, unsuccessful attempt at a microsurgical cure. GKRS involves multiple isocenters of different beam diameter to achieve a dose plan that conforms to the irregular three-dimensional volumes of most mass lesions. The total number of isocenters varies depending upon the size and shape of the adenoma and the proximity of surrounding at risk structures (e.g. the optic nerves or chiasm). The most recent version of the Gamma Knife, the Perfexion, combines advances in dose planning, collimation of 192 simultaneous beams, and robotic engineering, and it eliminates the need to set coordinates manually for each isocenter.
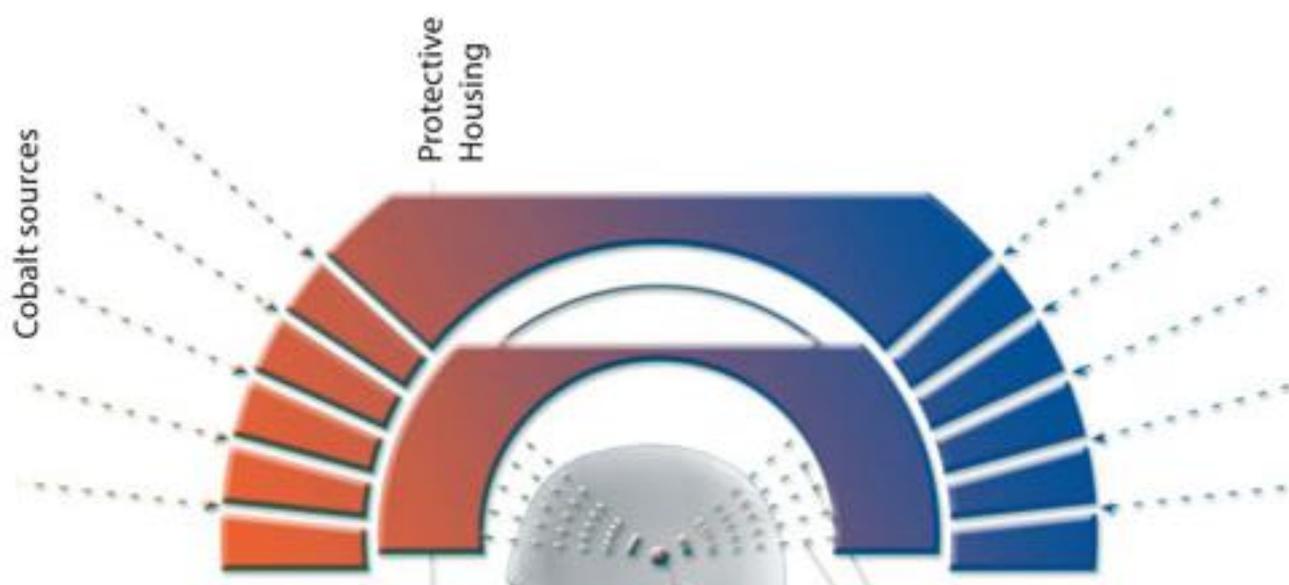

Figure 15.44. Scheme of Gamma Knife radiosurgery.

The goals of stereotactic radiosurgery in the treatment of Cushing's disease pituitary tumors are to inactivate adenoma cells, controlling tumor growth and to normalize hormone overproduction. Ideally, these goals are met without damaging the residual normal pituitary gland and surrounding vascular and neuronal structures.

Most centers define an endocrine remission for Cushing's disease as a 24 hour urine free cortisol in the normal range associated with the resolution of clinical stigmata or a series of normal post-radiosurgical

serum cortisol levels obtained throughout the day. In our institutional experience of treating 107 patients with Cushing's disease, the remission rate was 53% and the vast majority achieved adenoma volume control (i.e. shrinkage or no growth of the pituitary adenoma). In general, most patients who ultimately demonstrated hormonal normalization did so within the first 2 years following radiosurgery. Recent work suggests that ketaconazole and other cortisol suppressive medications may adversely affect the endocrine outcome following GKRS. As such, a temporary cessation of such cortisol lowering agents at the time of radiosurgery seems prudent.

Potential complications following Gamma Knife radiosurgery for Cushing's disease include delayed hypopituitarism which occurs in approximately 20-30% of cases and an extremely rare chance of visual dysfunction. Careful longitudinal follow up of Cushing's patients including serial MRI's, neurological, and endocrine evaluations are important.

Final Remarks

Is the textbook Applied Biophysics (Biophysics for Engineers) good enough and suitable for the engineer physicists, chemists, bio and biomedical engineers, etc.? It is Shakespeare' like question.

This book represents our view in the use of combination of fundamental knowledge in biophysics with novel achievements in this really interdisciplinary and rapidly developed science and technology.

Among the goals of the book are:

• To make all the material including a complete set of different chapters available to students on the World-Wide Web.

• To enhance the presented chapters of the book as a learning tool compared to a conventional textbook.

• To present a complex subject to students in several different ways so that each student can use the learning techniques best suited to that individual.

• To produce course material that might be appropriate for distance learning or self-paced courses.

The current set of book's chapters cover a one semester and/or two semester courses of Biophysics at Georgian Technical University (Biomedical engineering, Engineering physics, etc.).

Using our personal teaching experience students get very little more than an introduction to the material out of the lecture and prefer to learn from the textbook and homework. Some students claim they cannot learn from the textbook and rely on lectures to get their basic understanding. Some prefer a rather verbose exposition of the material in the text, while others prefer a concise discussion largely based on equations.

We should give the possibilities to our students to learning complex subjects from either a lecture or a textbook.

An important aspect of the design of the notes was to maintain a concise basic treatment of the physics, biology chemistry, mathematics with derivations and examples available. This format will allow more examples than are practical in a textbook.

The primary teaching though probably still works best at the blackboard (electronic board). One thing that our classrooms now more less are, and in future must be fully equipped for switching from one mode to the other.

In a future class, with the notes fully prepared, the author of the current book will plan to decrease the formal lecture time and add practice and/or discussion session time, with students working moving at their own pace using computers. Similar sessions would be possible at a distance. The formal lecture could be taped and available in bite size pieces inside the lecture notes.

This is author's vision to future classes where we will prepare well thinking, creative persons, doing the modern science and technology which can help to strengthen the potential of young people to develop the modern knowledge also in biophysics, which is becoming more and more important field of mankind safety and security.

References and Main Information Sources

- Tom A. Waigh, Applied Biophysics, 2007 John Wiley & Sons Ltd, The Atrium, England, 2007, 421p.
- Applied Biophysicshttps://www.biophysics.com. Label-free Real-time Automated Cell-based assays powered by the technology of ECIS. Developed by Nobel laureate Ivar Giaever and Charles Keese.
-
- Carlos A.Leguizamón, Concept of energy in biological systems, Bulletin of Mathematical Biology, Volume 37, March 1975, Pages 565-572.
- Energy in Biological Systems, Topic2, Lecture notes, Simon Frazer University, Faculty of Science, Department of Biological Sciences. British Columbia, Canada. 2015. 40p.
- Rob Phillips, Jane Kondev, Julie Theriot, Hernan Garcia, Physical Biology of the Cell.2nd Edition. 2012. Garland Science, Taylor and Francis Group LLC: New York. 1057p.
- Joël Janin, Physical biology of the cell, Second Edition, Université Paris-Sud 11, October 2013, Crystallography Reviews 19(4), DOI: 10.1080/0889311X.2013.830112,
- Gabriel Popescu, Lauren P. Deflores, and Joshua C. Vaughan George, Kamran Badizadegan, Hidenao. Iwai Fourier phase microscopy for investigation of biological structures and dynamics, November 1, 2004 / Vol. 29, No. 21 / Optics letters 2503.
- Iuliana Oita, Hadewych Halewyck, Bert Thys, Bart Rombaut, Yvan Vander Heyden, Debby Mangelings, Microfluidics in macro-biomolecules analysis: macro inside in a nano world., Anal Bioanal Chem (2010) 398:239–264. DOI 10.1007/s00216-010-3857-7.
- Alfred Driessen Life and quantum biology, an interdisciplinary approach. arXiv preprint arXiv:1109.2584, 2011 - arxiv.org
- Karoly Jakab, Adrian Neagu, Vladimir Mironov, Roger R. Markwald, and Gabor Forgacs, Engineering biological structures of prescribed shape using self-assembling multicellular systems, Communicated by Joel L. Lebowitz, Rutgers, The State University of New Jersey, Piscataway, NJ, January 8, 2004, PNAS, March 2, 2004, vol. 101, no. 9, 2864–2869.
- Pierfrancesco Palazzo, Theoretical arguments on exergy method and non-equilibrium in nuclear processes. Sapienza University di Roma, Doctoral Thesis, 11 Febbraio 2019, 280.

- Biophysics, Lecture notes, Johns Hopkins University, Department biophysics and biophysical chemistry Johns Hopkins University,725 N. Wolfe Street, WBSB 608DBaltimore, MD 21205-2185, USA.

- Thomas M. Nordlund, Peter M. Hoffmann, Quantitative Understanding of Biosystems: An Introduction to Biophysics. CRC Press, Mar 4, 2011 - Medical - 588 p.

- Kate Jeffery, Robert Pollack, and Carlo Rovelli, On the Statistical Mechanics of Life: Schrödinger Revisited. Entropy 2019, 21, 1211, 18p. doi:10.3390/e21121211.

- Brigita Urbanc, Biophysics, Lecture notes, Drexel University, Physics department Philadelphia, 2012.

- Edward A Codling, Michael J Plank and Simon Benhamou. Random walk models in biology. Review, Published by: Royal Society Interface J. R. Soc. Interface (2008) 5, 813–834. doi:10.1098/rsif.2008.0014.

- Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002.

- Lin Frank Song, Arkajyoti Sengupta, Kenneth M. Merz Jr. Thermodynamics of Transition Metal Ion Binding to Proteins. J. Am. Chem. Soc. 2020, 142, 13, 6365–6374, https://doi.org/10.1021/jacs.0c01329

- Kenneth R. Foster,Thermal and Nonthermal Mechanisms of Interaction of Radio-Frequency Energy with Biological Systems, IEEE transactions on plasma science, vol. 28, no. 1, February 2000, 15p.

- Biochemistry, Biophysics and Structural Biology at Berkeley: Course: Josh Deutsch, an introduction to biophysics, lecture notes, 2017.

- Katherine Lamos, Biological Physics at Princeton, Physics department, Princeton University, Biophysics, Princeton Lectures on Biophysics (Volume 1) Princeton, NJ 08544, United States.

- William Bialek, Biophysics: Searching for Principles, Lecture notes in biophysics

  Princeton University Press 2012, Princeton University, New Jersey, USA

- K. Michaelian, Dominion of the same physical laws as chemical, transport, and mechanical processes, Journal of Theoretical Biology 237 (2005) 323–335.

- Zhang T, Mu Y (2012) Initial Binding of Ions to the Inter helical Loops of Divalent Ion Transporter CorA: Replica Ex-change Molecular Dynamics Simulation Study. PLoS ONE 7(8): e43872. https://doi.org/10.1371/journal.pone.0043872.

- Helen Watson, Biological membranes, Essays Biochem (2015) 59: pp 43–69.https://doi.org/10.1042/bse0590043

- Paata J. Kervalishvili, Medical Physics and Biophysics, Lecture notes, GRUNI, 2019, Tbilisi, 205p.

- Paata J. Kervalishvili, Tamar N. Bzhalava, Optical Spectroscopy of Bionanoobjects, GTU, Tbilisi, 2018, 285p.

- Paata J. Kervalishvili, Essential Themes of Medical Physics and Biophysics in Fifteen Lectures. Publishing House "Universal", Tbilisi, 2020, 230p.

- Jan 31, 2020 — The XXIV International School of Pure and Applied Biophysics, jointly organized by the Italian Society for Pure and Applied Biophysics (SIBPA).

- Nuclear spin selectivity in enzymatic catalysis: A caution for applied biophysics. Arch Biochem Biophys. 2019 May 30, 667:30-35. doi: 10.1016/j.abb.2019.04.005.

Here we represent the little part of the information sources existing in different literature: research papers, inventions, communications, other different editions.

We used them directly or indirectly, apply to the knowledge developed by different groups from different universities, research institutions, companies, biomedical establishments, etc.

We are very thankful to our colleagues for information about their current achievements, their comments, remarks and suggestions.

We also apologize for mistakes and incorrections we made. We do hope that scientific community as well as teachers and students and all public will forgive us for them.

Contents

About Author

Paata J. Kervalishvili, was born in Tbilisi, Georgia. He started his university studies at Georgian Technical University and received his B.S. and M.S. degrees in engineering physics. His Ph.D and Doctor in Engineering was awarded in 1978 by the Research and Technology Center - "Institute Giredmet" in Moscow. In 1984 he was conferred a Degree of Doctor of Physics and Mathematics at main center of Soviet Atomic Program "Kurchatov Institute". At 1989 he received the USSR Professor state title.

Prof. Kervalishvili until 1992 worked as a Scientist and Director of various research and technology development centers within the Ministry Middle Machinery (State Committee for Atomic Energy) of former Soviet Union.

After working in the USA he returned to Georgia in 1994 to work in the Parliament and was appointed to the leading positions of the State Committee for Science and Technology and Ministry of Economy of Georgia where he worked till 2000.

While working on the accomplishment of science and technology projects in US, France, England, Italy, Germany, Greece and Georgia, Paata Kervalishvili also worked as a Professor and Chair of several prominent universities.

Dr. P. Kervalishvili conducted a range of research in several fields of physics and technology such as condensed matter, molecular, bio and quantum physics, nuclear and laser technologies, novel materials, sensory systems, energy and information technologies. He is an author of more than 500 papers, inventions and several renowned books.

Prof. Kervalishvili is a member of more than 20 top-level academies and research councils. Currently he is the President of Georgian Academy of Natural Sciences and President of the Euro Mediterranean Academy of Arts and Sciences.

More than 50 prestigious prizes and state orders were awarded to Dr. Kervalishvili.