Chapter 2
Molecular Structure of Biological Systems

It is the intention of this section to familiarize the reader with some specific physical properties of biological systems on the molecular level. The overriding theme of this section is the controversy of thermal fluctuation against the forces of molecular orientation and organization.

Two kinds of physical behavior meet on the molecular level of biological structures: On the one hand, there are the characteristic properties of *microphysical* processes, based on the individual behavior of single small particles like atoms, molecules, or supramolecular structures. These processes are mostly stochastic. On the other hand, there are reactions which resemble *macrophysical* properties, the kind of behavior of "large" bodies. *Macrophysics* is ruled by the laws of classical physics, as for the example of classical mechanics. Our daily experiences with macrophysical systems teach us that their behavior is generally deterministic.

To explain this difference, let us consider a simple mechanical wheelwork. The knowledge of its design and construction allows a precise prediction of the behavior of the system. This prediction is based on the laws of classical mechanics. In contrast to this, a chemical reaction with a small number of molecules in a homogeneous phase depends on stochastic collisions of the individual molecules with each other. Since this process is stochastic, it is predictable only in a statistical way.

This stochastic behavior of molecular systems can be transformed into a deterministic one, if the number of participating stochastic events is large, or if the degrees of freedom of the single reactions are extremely limited. The increase of stochastic events can be realized either by an increasing number of participating molecules, by enlarging the volume for example, where the reaction takes place, or by an increase of the time interval of observation. This consideration indicates an interesting interplay between volume, time constants, and reliability of a biochemical reaction.

The limitation of the degree of freedom of a biochemical reaction is realized by a property of the system which is called *anisotropy*. In contrast to *isotropic* systems, like simple solutions, in anisotropic systems the mobility of molecules in various directions is not identical, but is restricted in some directions, and promoted in

others. This, for example, is the case for enzymatic reactions, where the participating enzymes are oriented in membranes, or if the reactions of charged or polar reactants occur in strong electric fields of electrical double layers.

In many fields the biological organism works as an amplifier of microphysical stochastics. A molecular mutation, for example, leads to a reaction chain, which finally ends with a phenomenological alteration of the organism. Or, as another example: a few molecular events in the pigments of an optical receptor can lead to perception and to reaction in behavior.

During the first steps in considering molecular mechanisms of biological systems, a further aspect is taken into consideration. Unfortunately, biologists often ignore that a qualitative jump has to be considered in the transition from the "visible" macrophysical structures, to the microphysical systems such as atoms or molecules. This includes not only the above-mentioned transition from the deterministic behavior of macroscopic systems to the stochastic behavior of single molecules, but many further aspects as well.

The biologists, for example, must acknowledge that the term "structure" now receives a new meaning. The visible "biological structure," as known in the fields of anatomy, morphology, and histology, now appears as concentration profiles, or as systems of electric charges, or electromagnetic fields. Instead of visible and measurable lengths, diameters, or distances, as common in the visible world, in the microphysical world so-called *effective parameters* are used. These sorts of parameters are exactly defined, and they can be measured with arbitrary exactness, but they do not correspond to some visible boundaries. A single ion, for example, has no diameter in the sense of the diameter of a cell, or a cell nucleus, which can be measured by a microscopic scale. In the following sections we will define effective parameters like crystal radius, hydration radius, and Debye–Hückel radius, which are very important parameters for functional explanations. We will consider this aspect in detail in Sect. 2.1.3.

It is not the intention of this book to describe the topics of molecular biology. However, the theoretical foundations and principles will be explained to make possible a link between structure and function at the molecular level and current biological thinking in these dimensions.

2.1 Thermal Molecular Movement, Order and Probability

In this section, the biophysics of molecular organization of biological systems will be discussed in the context of processes of thermal movements where a statistical mechanics approach can be applied. Stochastic phenomena are of great importance in molecular biophysics. Here, we are immediately confronted with the dialectics of arbitrary distribution on the one hand, and organized order on the other hand. This also touches upon problems of biological self-organization and stability of the resulting structures.

2.1.1   Thermodynamic Probability and Entropy

In 1854, Rudolf J. E. Clausius introduced the *entropy* (*S*) as a parameter of phenomenological thermodynamics, and defined it as the heat, added to a system in a reversible way in relation to the temperature (see Eq. 3.10). Later, in 1894, Ludwig Boltzmann used this parameter in the framework of statistical thermodynamics. In these circumstances, the entropy, and in context to this, the second principle of thermodynamics becomes more imaginable. Entropy appears as a kind of measure of disorder, or as a degree of random distribution, i.e., of missing order. The correlation between order and probability and, as will be explained later – information – is of great importance for the understanding of the principles of biological organization.

Let us start with the assumption that the entropy is a measure indicating the degree of randomization of a given distribution. We will consider a system of maximal entropy as a system in maximal disorder. Furthermore, let us demand that the entropy be an *extensive* parameter. Therefore, like volume, or mass, but in contrast to the *intensive* parameters such as for example temperature or density, the entropies S_1 and S_2 of two systems can be added, if these systems come together:

$$S_1 + S_2 = S \tag{2.1}$$

How can we now define a parameter, which indicates the degree of randomization or, on the contrary, a degree of disorder? What does order of organization mean? Of course, our daily experience shows that an ordered system spontaneously transforms into a disordered one, but not vice versa. This, actually, is the consequence of the second principle of thermodynamics.

Let us consider a very simple structure, just the distribution of four distinguishable spheres on two compartments of a box (Fig. 2.1). Let each of these spheres, independently of the three others, just by chance fall into one or the other compartment of the box. All of the 11 possibilities of the distribution, as indicated in Fig. 2.1, therefore, have the same degree of probability, because the probability of each sphere individually, to fall into compartment 1 or into compartment 2 is equal. Summarizing the patterns of distribution shows that there is only one way to
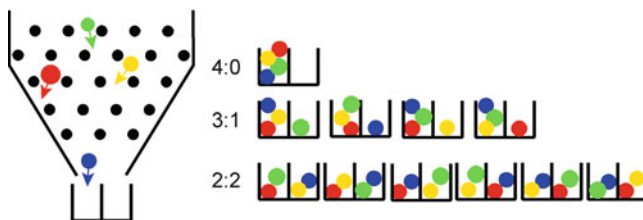
Fig. 2.1 All possibilities of the statistical distribution of four distinguishable spheres in two compartments of a box

realize the distributions 0:4 and 4:0. In contrast, there are four ways to realize the distributions 3:1 and 1:3, and, finally, six ways for equal distribution: 2:2.

Let us now ignore the fact that the spheres are distinguishable. Let us simply ask: How large is the probability that just by stochastic distributions one of the relations 4:0, 3:1, 2:2, 1:3, or 0:4 occur? Apparently, the probability of any kind of distribution will be larger, if it can be realized by a larger number of ways. The distribution mode 2:2, for example, is 6 times more likely, than the distribution 4:0, or 0:4. The number of ways which lead to the realization of a definite situation, in fact, seems to be a measure of the probability of the occurrence of it. We will designate this number of ways by the parameter W which we will call *thermodynamic probability*. The amount of W can be at least 1 and at maximum ∞, in contrast to the mathematical probability (P), which we will use in Sect. 2.1.2, and which ranges between 0 and 1.

Now, we come to the following sequence of conclusions: If W really is a measure of the probability of getting a definite distribution, and if an increase of the degree of order is the most uncertain result of a stochastic distribution and finally, if the entropy (S) is a parameter, indicating the degree of disorder – than S should be a function of W. If two situations with relative probabilities W_1 and W_2 are connected together, then the probability of this combined situation results from the product ($W_1 \cdot W_2$). Using Eq. 2.1, this means:

$$S = f(W) = S_1 + S_2 = f(W_1) + f(W_2) = f(W_1 \cdot W_2) \qquad (2.2)$$

This demand is met by the logarithmic function:

$$\ln A + \ln B = \ln(A \cdot B) \qquad (2.3)$$

Hence entropy is proportional to the logarithm of W:

$$S = k \ln W \qquad (2.4)$$

This is the Boltzmann equation of entropy. Boltzmann's constant k was defined as a universal constant later by Max Planck. It must have the same unit of measurement as entropy, and is as follows:

$$k = 1.380658 \cdot 10^{-23} \mathrm{J\,K^{-1}} = 8.6174 \cdot 10^{-5} \mathrm{eV\,K^{-1}}$$

This explanation was just based on the simplest experiment where four spheres were distributed randomly over two compartments. One step toward serious thermodynamics can be taken, considering for example the compartments of this box as molecules of a system, and the spheres, as quanta of energy, distributed among them. This complication, of course, means a transition of handling with larger numbers. If the number of elements and classes are increased, W cannot be

evaluated just by simple trial. It is possible to calculate this value using the following equation:

$$W = \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdot \ldots \cdot n_m!} \tag{2.5}$$

where n is the total number of all elements in the system (in case of Fig. 2.1 – the total number of spheres: $n = 4$); n_i (for $i = 1 \ldots m$) is the number of elements in each class of state (this means, the number n_1 in compartment 1 and n_2 in compartment 2); and m is the number of classes of state (namely: number of compartments in the box).

2.1.2   Information and Entropy

In 1948, C. E. Shannon introduced a parameter which in technical information theory has been proved as a useful measure of information content of a message. The information *(I)* of a message depends on the effort required to guess it by a highly systematic system of questions. Hence, information is some sort of degree of the actuality of a message.

It is not difficult to guess the result of the toss of a coin, since there are only two possibilities of equal probability. To guess a certain card in a full deck of playing cards is much more difficult. In this case, a much greater uncertainty factor has to be taken into account. Using a more systematic approach, a large number of yes-no questions have to be answered. Hence, the information content of a particular playing card is higher than that of a tossed coin. Should a deck consist of cards which are all the same, and should this be known to the challenged person, guessing will not make sense at all. The information content of each of these cards is zero. The probability by which possibilities are turned into reality, consequently, seems to become a measure of information.

In contrast to thermodynamics, in the information theory the mathematical term of probability *(P)* is used which is defined as follows:

$$P = \frac{\text{number of favorable cases}}{\text{greatest possible number of cases}}$$

On average, coins tossed a 100 times will land with heads up in 50 instances. Hence, the probability of heads facing up may be expressed by:

$$P = \frac{50}{100} = \frac{1}{2}$$

Conversely, the probability of throwing a "six" with some dice is only $P = 1/6$, whereas the probability of throwing one of the three even numbers would be higher: $P = 3/6 = 1/2$.

Whereas the thermodynamic probability (W) is always larger than 1, (cf. Sect. 2.1.1), the value of the mathematical probability lies between 0 and 1 ($0 \leq P \geq 1$). In this context, $P = 0$ means an impossibility, while $P = 1$ expresses absolute certainty.

The logical conclusions which led to the derivation of the Boltzmann equation (cf. Sect. 2.1.1) are the same as those on which the Shannon relation is based. Information (I) is a function of mathematical probability (P):

$$I = f(P)$$

The condition for the function f again, is satisfied by the logarithmic function, since here too, the multiplication rule for the calculation of probabilities must be valid (cf. Eq. 2.3). To obtain positive values of information, considering that $P \leq 1$, the negative value of the logarithmic function must be used. The information of a single event therefore is:

$$I = -K \ln P \tag{2.6}$$

This is the *Shannon equation* of information theory. The unit of I is determined by the unit of the always positive factor K. The bit (*binary digit*) is most commonly used. It expresses the number of binary yes-no decisions that are needed to determine a given message. For example, the one side of the coin can be guessed by one single decision, its information value, consequently, is 1 bit. Five binary decisions will be sufficient to guess a card from a deck. Hence, the information value of one card is 5 bits. The factor $K = 1/\ln 2 = 1.443$ must be used to calculate l in bits. In the information theory the logarithm to the base 2 ($\log_2$) is occasionally used:

$$I = -1.443 \ln P = -\log_2 P \quad (I \text{ in bit}) \tag{2.7}$$

A message usually consists of elements like symbols, words, or other structures of a set defined in advance. This concerns not only the letters of a word or the words in a written text, but the same approach can be applied also for various cases of biological information storage and transfer. So, for example, the information content of a DNA molecule is coded by the sequence of nucleic acids; the information content of a protein is given by their amino acids. In a similar way the information content of time-dependent structures like nervous signals can be evaluated too. In this case discrete patterns of neural spike trains are considered as elements of the message.

Assuming for example that a mammalian DNA molecule consists of 15,000 pairs of nucleotides and the four possible types of nucleoside bases have an equal probability of occurrence, then the information content of each single nucleotide

will, consequently, have a value of 2 bits. The information capacity of this DNA molecule therefore amounts to 30,000 bits.

In fact, however, in a DNA molecule the probability of the occurrence is not equal for all nucleotides. In this case, the information content of the whole message can be expressed by the sum of the evaluated probabilities (P_i) of the single elements (i) by the following equation:

$$I = -K \sum_i P_i \ln P_i \tag{2.8}$$

Estimates of information capacity in the whole genome of an organism in this way range up to about 10^{10} bit. The amount of actually stored information is even lower, if the redundancy in the storage of genetic information which is required for the protection of the information is taken into account.

It seems important to emphasize that the Shannon Information (I) as evaluated here, is the so-called *syntactic* information. It allows for example important conclusions about the maximum storage capacity of a DNA molecule of the whole genome. This Shannon Information (I), however, has nothing to do with the "information content" in the everyday use of this word, the so-called *semantic information*.

The difference between the syntactic and the semantic character of information can be illustrated best by the following example: A book, consisting just of stochastically jumbled letters, or even words according to Eq. 2.8 have a larger amount of syntactic information than such with a meaningful text, because in the stochastic sequence the probability of the occurrence of elements is lower than in an organized one. If P_i becomes smaller, its negative logarithm, and therefore I gets larger. Knowing some characters of a reasonable word, for example, the rest of them can usually easily be guessed – the probabilities of their occurrence therefore become larger, its content of syntactic information diminishes. If the same number of characters is mixed stochastically, guessing is more complicated – information is larger.

Transferred to the biological situation: The Shannon, i.e., the syntactic information of a polynucleotide with an arbitrary sequence is larger than that of a DNA with the same molecular length. In contrary to the *syntactic* information, the *semantic* information of a reasonable book, of course, is much larger than a book containing stochastically selected characters or even words. The same refers to the biological DNA in contrast to a polynucleotide with no biological function.

Does this mean that it is impossible to quantify biologically important information? Does it mean that the information concept is not applicable to biological systems at all? In fact, the semantic aspects of communication are irrelevant to most engineering problems and therefore not included in the definition of Shannon's I. Despite many attempts, quantification of semantic information has not yet been achieved.

Nevertheless, even the use of the syntactic information properties of biomolecules and nervous processes has proved to be a useful tool in modern molecular biology.

The enormous increase of the amount of data on the primary structure of biological macromolecules, especially DNA sequences, as well as data of neurophysiological recordings nowadays required mathematical methods for their evaluation, comparison, and storage. For this, bioinformatics was developed as a special discipline, based on these considerations.

Formally, the Boltzmann's constant (k) can be used in the Shannon equation (Eq. 2.6). In this case information is obtained formally in entropy units (JK^{-1}). Therefore the term "entropy" as Shannon-Entropy in bioinformatics was frequently used in the sense of information. This formal interconnection of Shannon's definition of information with Boltzmann's entropy triggered off a discussion about the relation between these two parameters.

The starting point for this discussion was the second principle of thermodynamics, predicting that isolated systems spontaneously try to reach a state of maximum disorder. The growth and the existence of a living system on the contrary is possible only by decreasing or at least by the conservation of entropy. Erwin Schrödinger (1944) made the frequently quoted statement: "The living system feeds on negative entropy." This is the reason, why sometimes the term "negentropy" was used.

The interconnection of information with entropy in terms of thermodynamics may be illustrated best by a thought experiment conceived by James Clerk Maxwell in 1881 and still discussed today (Fig. 2.2). Maxwell proposed a room, which is connected with another by an opening. This opening can be closed by means of a slide. Both rooms are filled with a gas which is, in the beginning, in equilibrium, for example of equal pressure and temperature. An intelligent creature, later called "Maxwell's demon," is able to handle the slide between the two rooms with ease. This "demon" can observe accurately the direction and the velocity of the molecules in his room. Velocity and direction of these particles in the beginning are statistically distributed. If a particle in the upper room flies accidentally toward the opening, the demon opens the slide to let the particle pass. As a result of such sorting the pressure in the lower room would rise.

The demon could also take another approach. For example, he could separate fast from slow particles. In this case, a difference in the temperature between the two rooms would occur. In both cases, the entropy of the whole system would be reduced and energy might be generated by an appropriate turbine. The result would be a "*perpetuum mobile*," a perpetual motion machine of the second order, as it would contradict the second principle of thermodynamics.

This apparent contradiction subsequently was the subject of a large number of scientific and philosophical publications. The problem was finally resolved by the following consideration: The "demon" requires information to carry out the sorting. He collects this information by "watching" the molecules. In order to "see" he needs light. For this, the demon in Fig. 2.2 symbolically carries a lit candle. Yet, a body will only be able to emit light in a state of nonequilibrium relative to its environment. This, however, contradicts the equilibrium condition at the beginning of the experiment. The same would apply to any other approach to acquisition of information. This resolves the apparent contradiction to the second law of thermodynamics.

Fig. 2.2 Maxwell's demon
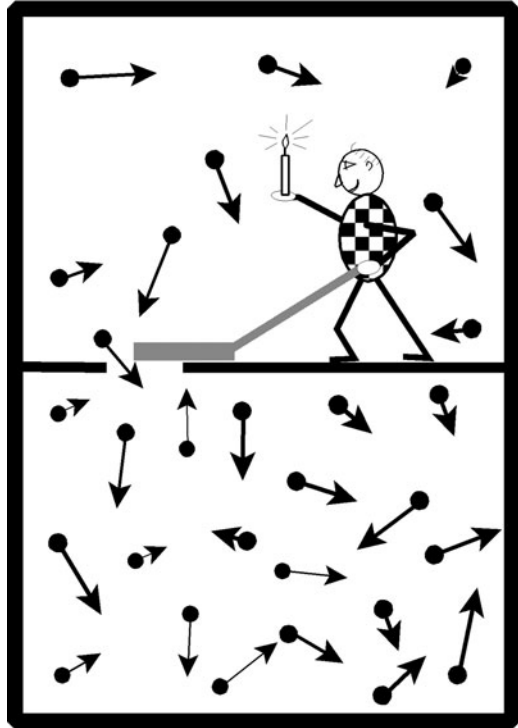
    Why do we discuss this thought experiment here, if it is clear that it will not work
as a perpetual moving machine? What has this to do with biophysics? In fact,
independent of this discussion about the virtual violation of the second law of
thermodynamics, Maxwell's demon demands particular interest in biophysics
because of its analogy to various functions of living systems. The living cell, too,
reduces its entropy at the expense of its environment, using information processing.
Yet, in the latter instance it is not the energy of statistical fluctuations, which is used
in this case. The biological system selects such molecules from its environment,
which are rich in free Gibbs energy of formation and correspondingly, with a low
content of entropy. It uses this energy and extrudes molecules with lower free
energy and larger entropy. The basic information for this process of selection, in
other words, the "software" for this process, is stored in the structure information of
the proteins, which are responsible for the recognition of these molecules, and
eventually for their metabolism. These proteins get this sort of semantic informa-
tion during the process of synthesis via the RNA, from the DNA of the genome.
    This example again raises the question of the relation between semantic and
syntactic information: what is the threshold value of information that is required to
control the processes of living systems? Or translated into the language of modern
computer science: how large must the simplest software for this sort of a biological

Maxwell demon be? What is the threshold information that carries out not only the metabolic function of the primordial organism but additionally its replication? How did the first accumulation of information in a molecule occur? Purely accidental combination of monomers to build their first functional macromolecule must be ruled out. The probability for this occasion is too low by far. Today, so-called *probiotic evolution* is assumed, i.e., chemical selection of developing polymers even before the stage of the biopolymer (see also Sect. 5.2.5).

Further Reading

Shannon and Weaver 1962; Bioinformatics in molecular biology: Kauffman 1993; Strait and Dewey 1996; Yockey 2005; Lesk 2002; Maxwell's demon: Leff and Rex 1990.

2.1.3 Biological Structures: General Aspects

In the previous section we introduced expressions like *order*, *structure*, and *organization* and discussed them in context with entropy and information, as well as with the statements of the second law of thermodynamics. This touches on a set of questions which are of central interest in biophysics and which will be mentioned in many sections of this textbook. Therefore it is necessary at this point, to explain some basic definitions and principal ideas.

What, really, is a *structure*? To the biologist, the term "structure," usually is related to the macroscopically or microscopically visible organization of an organism. This means, for example, the structure of an animal skeleton, structure of a cell, of a mitochondrion, etc. The term "molecular structure" already lies outside the limits of this view. It refers to a certain arrangement of atoms, without defined contours, which can be described just by means of wave mechanics. The same applies for the concentration profile of an electrical double layer (Fig. 2.43), and particularly for so-called "time structures," namely special time courses, like oscillations of a biological system (Figs. 5.3 and 5.16), like the shape of an electrocardiogram (Fig. 3.38), or like the sonogram of a bat's cry (Fig. 4.19). This means that the definition of the term "structure," which is used in biophysics has to be broader than that of the morphologists and cytologists. It must include these structures as well as those of metabolic networks, ecosystems, or others.

The best, and generalized definition of this term is given by the set theory of mathematics. Therefore: *a system is an aggregate of elements with certain interrelations between them.* The totality of these interrelations is called the *structure* of the system. This definition does not prescribe at all, what kind of elements, and what kind of interrelations these are. It is applicable to all kinds of systems including biological systems and structures. In biophysics, we are interested especially in *dynamic systems*, i.e., in such systems, the interrelations between their elements are *interactions*. In contrast to this, in *static systems* the elements have no interaction at all, but are just interrelated by formal *relations*. Examples for static

systems, are the system of natural numbers in mathematics, or the system of animal species or of plants in biology.

The elements of a metabolic network are the metabolites, and the interrelations between them, i.e., their interactions, are the steps of biochemical reactions. Correspondingly, an ecosystem is to be considered an interaction of individuals and populations depending on abiotic conditions.

Biological macromolecules can be considered as systems too. In this case, the following levels or organization occur depending on what will be assumed as their elements.

– *Primary structure* – linear sequence of monomers (= elements) in the linear molecular chain (= system). For example, a polypeptide chain: ...-serine-alanine-lysine-arginine-.
– *Secondary structure* – positioning in space of monomers (= elements) in a part of the molecule (= system) relative to each other. For example, the α-helix, or the *β*-sheet structure of an amino acid sequence in a protein.
– *Tertiary structure* – position in space of molecular regions of homogeneous secondary structures (= elements) in a molecule (= system). For example, intramolecular coordination of the position of several helical regions relative to each other or to a *β*-sheet.
– *Quaternary structure* – position in space of macromolecules (= elements) in a supramolecular complex (= system). For example, single proteins in a multienzyme complex.

When a salt is crystallized, a *periodic structure* forms, which is characterized by a periodic arrangement of their elements. Erwin Schrödinger (1944) called the biological structure an *aperiodic crystal*. This means a highly organized structure, the elements of which, are not simply repeated periodically. Sometimes one tries to evaluate this structural organization as *structural information*. As we pointed out in the previous Sect. 2.1.2, however, it is hard to quantify this parameter. Probably, the structural information should be measured as the effort which is necessary, to describe such a structure perfectly.

Consequently, the process of structure formation of biological systems, be it the development of life, its reproduction, or simply the biosynthesis of a macromolecule, all are accompanied by reduction of entropy. This appeared to be contrary to the second law of thermodynamics and has given rise to heated philosophical discussions in the past. The second law of thermodynamics actually postulates that in spontaneous processes occurring in isolated systems, the entropy strives towards a maximum. Yet, neither an organism nor its environment, i.e., the earth as a whole can be considered as an isolated system. The earth is constantly absorbing energy from the sun and is emitting this energy again. That continuous flow of energy maintains a permanent nonequilibrium state which manifests itself not only in a direct way in photosynthesis with subsequent metabolism of heterotrophic organisms, but also in the environment of life, for example, in flowing water, in alternation of light and darkness, and in changes in temperature, humidity, etc.

In fact, structures are also formed in inanimate nature under loss of entropy. Basically, a distinction must be made between two kinds of structure formation which at the same time explain the two ways of self-organization in living systems.

– *Equilibrium structures*, for example a salt crystal, formed spontaneously during evaporation of a solvent.
– *Nonequilibrium structures* (or: *dissipative structures*), for example an arrangement of cirrus clouds as the result of processes of air convection under special meteorological conditions.

The genesis of life which is based on the prebiotic formation of a first bio-macromolecule, and the subsequent evolution of all organisms, can be understood as the result of complicated processes that occur far from equilibrium. In this context, a major role is played by dissipative structures of the inanimate environment, such as periodic temperature variations, tides, etc. Substantial importance has to be attributed also to such nonequilibrium structures in the course of life itself. They are frequently represented by time structures, such as for example the heart rate, or other kinds of oscillations, which in some cases are associated with the so-called biological clock (for further explanation see Sects. 3.1.4, 5.2.4, and 5.2.5).

Equilibrium structures, such as inorganic salt crystals with a very primitive geometry, become very complex and variable in shape, when based on the sophisticated pattern of the primary structure of bio-macromolecules rather than on the relatively simple field of interaction of spherico-symmetric ions. The spontaneous folding of proteins and their arrangement to supramolecular structures, such as complex enzyme systems of even ribosomes, must be viewed from this aspect. Such processes are also referred to as *self-assembly*. More specific aspects relating to the formation of equilibrium and nonequilibrium structures will be given in subsequent chapters of this textbook.

Further Reading

Eigen 1971, 1992; Meinhardt 2008, 2009; Kauffman 1993; Strait and Dewey 1996.

2.1.4 Distribution of Molecular Energy and Velocity at Equilibrium

The Boltzmann equation of entropy (Eq. 2.4) as derived in Sect. 2.1.1, helps to illustrate the second law of thermodynamics, according to which, isolated systems spontaneously approach a state of maximum entropy. We demonstrated there that one can express this also in the following way: at equilibrium isolated systems reach a state of highest realization probability (maximum of W). Now, we will ask the question: what are the mean properties of the molecules at this equilibrium state? This is of fundamental importance for further considerations.

It is possible to answer this question on the basis of statistical thermodynamics, which helps to make predictions on the probability of energy distribution among the elements of a system in thermodynamic equilibrium. It should, however, be emphasized that a deterministic description of the properties of single particles lies outside the limits of this discipline. Nevertheless, it will be seen that even a statistical statement allows the drawing of important conclusions regarding reaction rates, stability of molecules, and many others.

Let us imagine a space that contains gas molecules of uniform mass (m), which at the beginning all have the same velocity (v), and consequently, the kinetic energy (E) which can be calculated by the following equation:

$$E = \frac{m}{2} v^2 \tag{2.9}$$

This equality of the kinetic energy of all molecules, in fact, is a highly improbable state. In this case, using Fig. 2.1, all molecules would belong to a single box, or a single class of properties n_i. Thus, according to Eq. 2.5: $W = n!/n_i! = 1$. This situation will change instantaneously. The molecules would exchange their energy by elastic collisions with each other, and soon a great number of energy states would be occupied. A simple mathematical example shows that W, and according to Eq. 2.4 also the entropy (S), will increase with a growing number of (m) of state classes, provided the following relation applies:

$$\sum_{i=1}^{m} n_i = n$$

Because of the law of conservation of energy (first principle of thermodynamics), the following condition must be satisfied at the same time:

$$\sum_{i=1}^{m} n_i E_i = \text{const} \tag{2.10}$$

The energy of this system, therefore, can be distributed randomly among all of its molecules. The total energy of the system, however, must always remain constant.

Now, let us look for a state with maximum probability, this means, with a maximum of entropy. Corresponding to the second principle of thermodynamics, this is actually the thermodynamic equilibrium. Applied to our example the question arises: How energy and, correspondingly, velocity will be distributed between the n particles, after a sufficiently long period of time if the system is isolated? Even under equilibrium conditions, of course, the energy of individual molecules will change permanently, but nevertheless, the statistical mode of energy distribution, or of the distribution of various degrees of molecular velocity at equilibrium becomes

stationary, i.e., time independent (for a detailed description of various kinds of stationary states see Sect. 3.1.4).

Considerations of this type lead to *Maxwell's equation of velocity distribution*:

$$\frac{dn(v)}{n_0 dv} = \frac{4}{\sqrt{\pi}} \left(\frac{m}{2kT}\right)^{3/2} v^2 \, e^{-\frac{m v^2}{2kT}} \tag{2.11}$$

The left part of this equation contains a relation which expresses the relative number (dn/n_0) of those molecules which are related to a particular velocity interval (dv). Its unit is s m^{-1}. This is a function expressing the probability of the distribution of the velocity, where m is the mass of a molecule (not to be confused with the numeral m in Eq. 2.5), and k is the Boltzmann constant. Multiplying in Eq. 2.11 the denominator and numerator of the expression within the brackets as well as that in the exponent with the Avogadro number (N = $6.023 \cdot 10^{23}$ mol^{-1}) one introduces molar, instead of molecular parameters:

$$M = N \cdot m \tag{2.12}$$

and

$$R = N \cdot k \tag{2.13}$$

(M – molar mass, R = 8.314 J K^{-1} mol^{-1} gas constant)

In Fig. 2.3, as an example, the velocity distribution of molecules of oxygen is depicted (M = 0.032 kg mol^{-1}). This curve is not symmetrical. The mean velocity (weighted arithmetic mean value) in general will be higher than the maximum value. The following relation applies:

$$v_{\max} = \sqrt{\frac{2kT}{m}} = \sqrt{\frac{2RT}{M}}; \quad \bar{v} = \sqrt{\frac{8kT}{\pi m}} = \sqrt{\frac{8RT}{\pi M}} \tag{2.14}$$
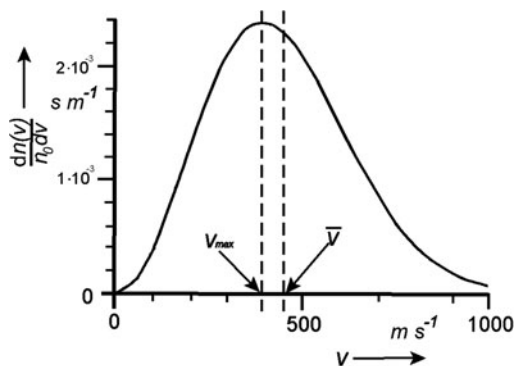
Fig. 2.3 Velocity distribution of O$_2$ molecules at 37°C corresponding to Eq. 2.11

These values increase with increasing temperature (T) and decrease with increasing molar mass (M).

Maxwell's velocity distribution is based on the Boltzmann function of energy distribution. It determines the number (n_i) of those particles which under equilibrium conditions of the system have an energy level of E_i:

$$n_i = C \; e^{-\frac{E_i}{kT}} \tag{2.15}$$

(C is a factor of standardization). The exponent of this equation as in Eq. 2.11 expresses the relation between energy E_i of the particle and the mean energy of thermal noise (kT). The Boltzmann constant (k) should be replaced by the gas constant (R) in cases where E_i expresses not the energy (in: J) of a single molecule, but the molar parameter (in: J mol^{-1}), as is customary in chemical thermodynamics.

The Boltzmann equation predicts the distribution of any form of energy. It expresses the situation, whereby lower states of energy tend to occur more often than states of higher energetic levels. Related to the total number of particles, it may also be written in the following form:

$$\frac{n_i}{n_{total}} = \frac{e^{-\frac{E_i}{kT}}}{\sum\limits_{i=0}^{\infty} e^{-\frac{E_i}{kT}}} \tag{2.16}$$

This equation provides the basis for many calculations in statistical thermodynamics. For example, it can be used to determine the percentage of molecules in a given system which are able to overcome a certain energy barrier (E_i). This aspect will be discussed in greater detail in the following chapter.

The use of this equation is allowed only in cases of thermodynamic equilibrium. At first glance it seems to limit its applicability to biological systems which are often far from equilibrium. The process of the adjustment of the Boltzmann distribution is, however, so fast that an imbalance of this kind has to be considered only in rare cases. On many occasions in this book we will indicate that in fact subsystems could be in equilibrium even in cases where the whole system is far from it (see especially in Sect. 3.1.4).

2.1.5   Energy of Activation, Theory of Absolute Reaction Rate

The temperature dependence of the rate of a chemical reaction was described by S. Arrhenius by the following equation:

$$k_R = A \; e^{-\frac{E_A}{RT}} \tag{2.17}$$

Here, k_R is the rate constant. The energy E_A was defined as the *activation energy*; A is an empirical factor.

In fact, this equation describes not only the temperature dependence of chemical reactions. It can also be used to depict corresponding temperature functions of other physicochemical processes, such as diffusion, kinetics of phase transitions, etc. At certain temperature intervals, even complicated biological processes can be fitted in this way, such as for example growth rate, or heart rate of poikilothermic animals.

To evaluate experiments, in which time-dependent parameters, like rate constants, reaction rates, frequencies of oscillations, etc., were measured as functions of temperature, a so-called *Arrhenius plot* is applied (see. Fig. 2.11). In this case the logarithmic form of Eq. 2.17 is used:

$$\ln k_R = \ln A - \frac{E_A}{RT} \tag{2.18}$$

Applying this kind of evaluation, $y = -\ln k_R$ is plotted against $x = 1/T$. If in the investigated process the activation energy itself is constant, then the measured points are located on a straight line:

$$y = ax + b \tag{2.19}$$

The slope of this straight line is, according to Eq. 2.18: $a = -E_A/R$. The relation: $b = \ln A$ and, consequently, the factor of A can be read from the extrapolated value of $x = 1/T = 0$.

Direct derivation of the logarithmic function to dT is more complicated. We need it later:

$$\frac{d \ln k_R}{dT} = \frac{E_A}{RT^2} \tag{2.20}$$

Real biological processes usually consist of a large number of single reactions, each with different activation energies. Hence, a temperature dependence must be expected which is much more complicated than that expressed by the Arrhenius equation (Eq. 2.17). Surprisingly, however, even in these cases functions will still be obtained which can be described by straight lines in the Arrhenius plot. This is caused by the fact that in complicated processes the rate of the overall process is determined by a simple rate-limiting reaction. The Arrhenius plot just reflects the activation energy of this rate-limiting process.

In some physiological papers, a so-called *Van't Hoff rule* of temperature dependence is used. According to this the rate of a reaction increases by 2–3 times, if the temperature rises by 10°C. This factor is referred to as the Q_{10} value. This rule, however, is not supported by thermodynamic considerations. Even in cases where the energy of activation is constant, corresponding to the Arrhenius equation, the Q_{10} value will increase with increasing temperatures.
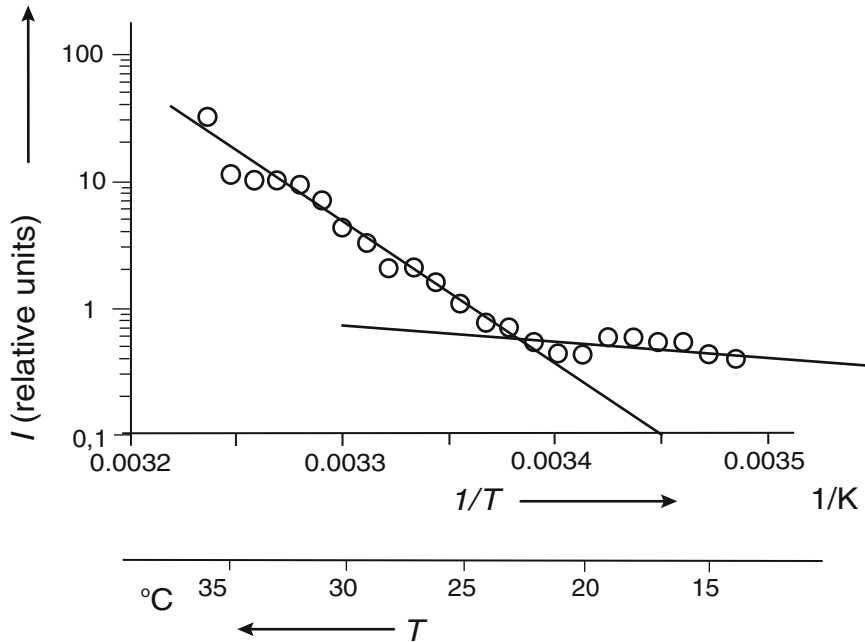
Fig. 2.4 An example of an Arrhenius plot: The logarithm of the heat-activated current (normalized to that at 25°C) in a temperature-sensitive TRPV4-channel is plotted versus the reciprocal value of the temperature. Compare with the plot of the same data directly versus temperature in Fig. 4.1 to calculate the Q_{10} corresponding parameters (Data from Watanabe et al. 2002)

In some reactions the energy of activation (E_A) itself is a function of temperature. In this case the straight line in the Arrhenius plot gets kinks (Fig. 2.4). The reasons for this can vary. In complex processes it is possible that at different temperatures, different reactions with different energies of activation will become rate limiting. In other cases such transitions may occur as the result of conformational changes of one of the components of the reaction, probably, conformational change in the enzyme itself. In reactions of membrane-bound enzymes, sometimes phase transitions of the enzyme near lipids may cause such effects.

The statistical mechanics allows a direct derivation of the Arrhenius equation. This is based on the *theory of absolute reaction rate* (or: *theory of transmission states*) as developed by H. Eyring. Here we can only explain some basic ideas of this approach.

Chemical reactions presuppose splitting of chemical bonds. To analyze such processes it is necessary to know the interaction energy of a given atom in the molecule as a function of the distance to each of its neighbors. This is an intricate function which can be expressed mathematically as an *n*-dimensional space corresponding to *n* possible ways of geometrical approach. In this space a line may be found which shows the way from state A to state B over the smallest peaks
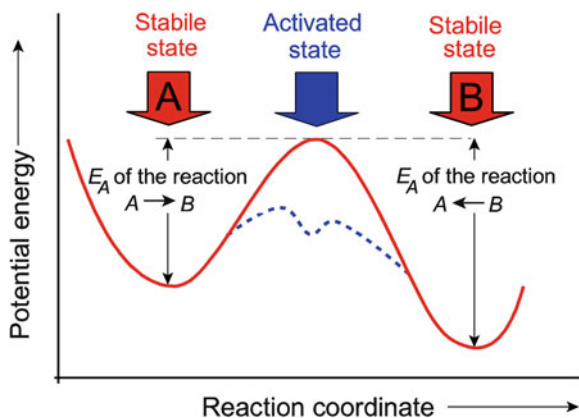
Fig. 2.5 Schematic potential energy diagram of a molecule in states A and B taken along the reaction coordinate. The solid curve represents the reaction without a catalyst; all the explanations correspond to this curve. The *blue dashed line* indicates how the energy of activation is modified in the presence of a catalyst (enzyme). In this case, the curve with a single maximum, dividing states A and B is transformed into another with two maxima, between which the minimal energy of an enzyme-substrate complex is positioned

of the activation energies. This line is called the *reaction coordinate*. It is possible to represent the energy level along this line in a 2-dimensional graph.

In Fig. 2.5 such a function is plotted with two minima, representing the two stationary states A and B. In these positions, the molecule will be in a stable state, which is however continuously deflected in a direction of the reaction coordinate by thermic collisions (the term *stability* will be explained in detail in Sect. 3.1.4). The effectiveness of those collisions depends on their energy, i.e., on the relation of their energy to the activation energy E_A. The energy of collisions, in fact, is thermic energy, which can be expressed by the factor kT (or in molar units, correspondingly in RT). The Boltzmann equation (Eq. 2.16) allows the calculation of the relative number of those molecules (n_i/n_{total}) which are able to overcome the energy barrier (E_i), which in our case is identical with the energy of activation (E_A). The relation E_i/kT (Eq. 2.16), of course, decreases with increasing temperature (T). This is illustrated in Fig. 2.6. While 36.8% of all particles have reached the energy level $E_i/kT = 1$, a value twice as high is only reached by 13.5%, and for $E_i/kT = 3$ there are only 5% reaching this level. The percentages drop fast with rising values of E_i/kT.

The theory of absolute reaction rate is based on the assumption that an activated transition state between the two states A and B exists. Its potential energy is identical with the energy maximum in Fig. 2.5. This activated state, of course, is highly unfavorable and unstable from the energetic point of view. At any point of time it will be assumed only by a very small number of molecules which will drop back from the energy maximum into the stable states of an energy minimum to the right and left of this maximum. An equilibrium constant (K^*) can be defined that characterizes the relation between activated and stable states according to the mass

Fig. 2.6 The relative number (n_i/n_{total}) of molecules as a function of the energy in kT-units. This curve corresponds to Eq. 2.16
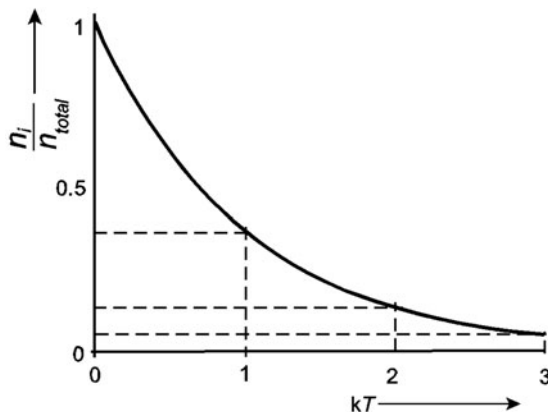
action law. The proportionality between the rate constant (k_R) and the number of molecules in the activated state can then be expressed as follows:

$$k_R = q\frac{kT}{h}K^*$$

(2.21)

The symbol * indicates the reference to an activated state. The relation between kT and Planck's constant **h** comes from kinetic quantum theory as derived by Wilson–Sommerfeld. The factor q gives the fraction of activated molecules which leave to the right or left of the energy maximum. For the case of symmetry of the relations, it becomes $q = 1$.

The phenomenological thermodynamics allows calculation of the relation between the molar free Gibbs energy of reaction (ΔG) and the equilibrium constant (K):

$$\Delta G = -RT \ln K$$

(2.22)

In the case of a reaction A→B, i.e., for a reaction in which the total number of molecules remains constant, and under the above-mentioned condition of $q = 1$, it is possible to calculate the conditions for the activated state (using the symbol *):

$$k_R = \frac{kT}{h}\, e^{-\frac{\Delta G^*}{RT}}$$

(2.23)

If one further considers the following relation:

$$\Delta G = \Delta H - T\Delta S$$

(2.24)

it follows that:

$$k_R = \frac{kT}{h}\, e^{\frac{\Delta S^*}{R}} e^{-\frac{\Delta H^*}{RT}}$$

(2.25)

The relation of this formula with the Arrhenius equation (Eq. 2.17) can be found by modification of the above equation in the following way:

$$\ln k_R = \ln T + \ln\left(\frac{k}{h} \; e^{\frac{\Delta S^*}{R}}\right) - \frac{\Delta H^*}{RT} \qquad (2.26)$$

and:

$$\frac{d \ln k_R}{dT} = \frac{1}{T} + \frac{\Delta H^*}{RT^2} = \frac{RT + \Delta H^*}{RT^2} \qquad (2.27)$$

Comparing Eq. 2.27 with Eq. 2.20 the following correlation is obvious:

$$E_A = RT + \Delta H^* \qquad (2.28)$$

For other values of q for reactions of synthesis or lysis, this relation can be modified.

In Eq. 2.25 especially the temperature-independent exponential term is of interest, which contains the *entropy of activation* (S^*). This is an indicator for conformation processes of the molecules participating in the reaction. These molecular conformations are prerequisites especially in cases where macromolecules are included in biochemical reactions. The parameter ΔS^* can become positive as well as negative. This depends on whether the molecule during the reaction gets a larger ($\Delta S^* < 0$) or a lower ($\Delta S^* > 0$) degree of order. Conversely, the entropy changes of the water molecules during this reaction must be considered too. We will discuss the entropy modifications of the water during reactions of hydration later in detail (Sect. 2.2.3, Fig. 2.19).

The acceleration of biochemical reactions by enzymes is based on a change in the reaction coordinate. The energy levels of the starting and the end product of the reaction will remain unaltered by this. However, with the help of an enzyme as a biological catalyst the transition from one state into another is possible with lower barriers of activation energy. This is indicated by the blue dashed curve in Fig. 2.5. First, an energy-substrate complex is formed, which requires less activation energy than that, which is needed to form the activated complex without the enzyme. This enzyme-substrate complex is then broken up again by a small input of activation energy, and substance B will be released. Of course, even with the help of an enzyme, the reaction A→B can take place spontaneously only, if the level of B is below that of A.

The concept of the theory of absolute reaction rate is not only applicable to processes of chemical reactions. In a similar manner, processes of diffusion can be described. Diffusion of a substance can be imagined as a reaction consisting entirely of location jumps from one position in the vicinity of a molecule to another. Under this circumstance the reaction coordinate will largely coincide with the direction of diffusion. Investigations of activation energy of diffusion processes

may provide the path of diffusion, for example, the mechanism of permeability of a molecule through a membrane.

Let us now discuss the problem of the life span of a bond and of a molecule. This question is directly connected to the above consideration. The Arrhenius equation also allows one to calculate the stability of a given molecule. In this case, however, not the reaction rate of molecular transformation is of interest, but its reciprocal value, namely the mean time, during which a bond with a given bonding energy will resist the attack of the energy of thermal noise (kT, resp. RT). In this case the energy of activation of a decay reaction (E_D) must be applied. The mean time of stability (t) of a molecule is inversely proportional to the rate of its transformation. Thus:

$$t = \tau \; e^{\frac{E_Z}{kT}} \tag{2.29}$$

This equation contains a time constant (τ), similar to the value A in Eq. 2.17. This factor depends on a number of structural properties of the molecule. It can only roughly be estimated in orders of magnitude. For smaller molecules, τ will lie between 10^{-14} and 10^{-13} s.

The characteristics of function (2.29) are illustrated in Fig. 2.7. Semi-logarithmic plotting ($\ln t$ against E_D) will result in a straight line. It can be clearly seen that even small changes of the activation energy (E_D) of the decomposition reaction will lead to dramatic changes of the life span (t). The range may stretch from 10^{-12} s to geological periods of time amounting 10^5 years.

Considering the energies of covalent bonds with several hundreds of kJ mol^{-1} one understands that they practically cannot be destroyed by thermal noise (RT) which at 300 K amounts to only 2.5 kJ mol^{-1}. However in fact, the word "cannot" does not exist in the language of statistical thermodynamics. As indicated in
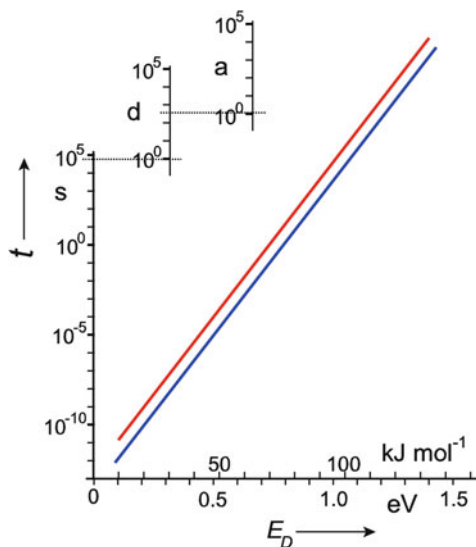
Fig. 2.7 The mean lifetime (t) of a bond dependent on the energy of activation of decay (E_D), corresponding to Eq. 2.29 ($T = 293$ K, *red line*: $\tau = 10^{-13}$, *blue line*: $\tau = 10^{-14}$ s)

Fig. 2.7, this just expresses the vanishing small probability of this process, resp. the extremely long life span of the corresponding bond. Even the binding energy of hydrogen bonds, between 13 and 25 kJ mol^{-1} is larger than RT in this temperature range. Nevertheless, the probability of their destruction by thermal noise is much larger and correspondingly their lifetime extremely short. Discussing the structure of water (Sects. 2.2.2, 2.2.3), we will show that in these cases the hydrogen bonds are destroyed and rebond at high frequency.

Not only the thermal noise (RT) influences the stability of molecular bonds but additionally quanta of various kinds of radiation. This will be discussed in detail in Sects. 4.7, 4.8, and 4.9 (see also Fig. 4.32).

As the mean life span of a molecule is only a question of probability, the concrete destiny of an individual molecule cannot be predicted. The assumption, however, can be made that by biological selection only DNA molecules with monomers of high bonding energy are left. Nevertheless, spontaneous molecular transformations take place which are known as mutations. A mutation results in a new molecule whose stability has not yet been approved by selection. The life span of this molecule can be substantially changed by a slight displacement of its bond energy. This may explain why remutations of mutation of already mutated molecules occur more often than could be expected statistically.
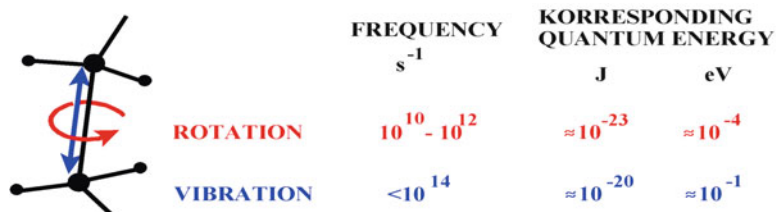
Further Reading

Blumenfeld and Tikhonov 1994; Westerhoff and van Dam 1987. The theory of absolute reaction rate: Eyring and Urry in: Waterman and Morowitz 1965.

## 2.1.6   Kinds of Thermal Molecular Movement

As a consequence of energy distribution in molecular systems, even under conditions of thermodynamic equilibrium, all of their parts are in vigorous thermal motion driven by the thermal energy kT. The energetic aspects of intermolecular interactions of proteins shall be considered in Sect. 2.2.5. Here we will discuss some general consequences of molecular movements and some methods for their quantification.

In general there exist three forms of molecular movements: vibrations, rotations, and translations: *Vibrations* are oscillations in the binding distances between the atoms in a molecule. The term *rotation* means not only the rotation of the whole molecule, but additionally, the spin of individual atoms or atomic groups around the axis of their bonds. The full translocation of a molecule or a part of it in space is meant by *translation*.

The frequency of these kinds of movements differs in a white range. As shown in Fig. 2.8, rotation can come in resonance with electromagnetic waves at frequencies between 10^{10} and 10^{12} Hz. In the case of vibrations, this frequency is higher and reaches about 10^{14} Hz. Both are in the range of infrared light (see Fig. 4.32) which is the basis for molecular analysis by infrared spectroscopy (see Sect. 4.8.1).

	FREQUENCY s^{-1}	KORRESPONDING QUANTUM ENERGY	
		J	eV
ROTATION	$10^{10} - 10^{12}$	$\approx 10^{-23}$	$\approx 10^{-4}$
VIBRATION	$< 10^{14}$	$\approx 10^{-20}$	$\approx 10^{-1}$

Fig. 2.8 Magnitudes of vibrational and rotational movement of a covalent C-C bond. The quantum energies are to be considered in relation to the energy of thermal movement which is $4.1 \cdot 10^{-21}$ J, resp. $2.6 \cdot 10^{-2}$ eV at $T = 300$ K
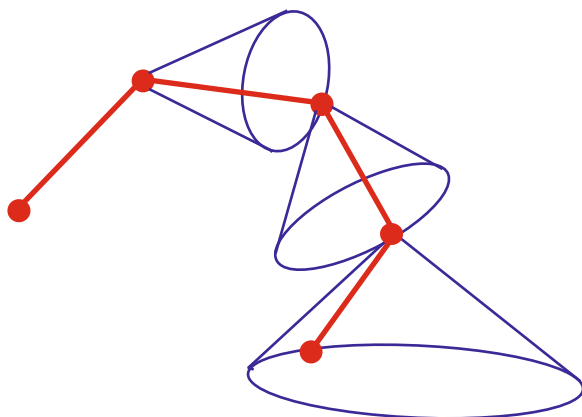
Fig. 2.9 The rotation cones of the atomic groups in a chain molecule

These frequencies allow us to estimate the quantum energies of these processes which appear to be near to that of the thermal energy kT. At a temperature of 300 K, the product kT equals $2.6 \cdot 10^{-2}$ eV. This means that for molecular vibrations and rotations the energy transfer is governed by discrete quantum steps. In contrast to this, translation requires quite small energy levels, of approximately 10^{-35} J or about 10^{-16} eV. Energy transfer of translational movement, therefore, can be considered to be continuous.

The intramolecular rotation of atoms in relation to each other has consequences for the structure of macromolecules. A covalent bond, rotating around its angle relative to another, consequently describes a cone. The next bond rotates, as well, but it does so quasi on the spot moving itself in an orbit of that lateral cone surface (Fig. 2.9). In this way chain molecules can assume a stochastic orientation unless other strong attracting or repelling interactions between the monomers are preventing this. The molecular structures resulting in this way can be calculated by the laws of statistical thermodynamics.

The so-called *random flight model* allows the simplest description of the behavior of a long molecule. In this case, the real molecule is described as a sequence of segments with the individual length l. These segments are connected to each other in such a way that the orientation of each is independent from the orientation of the former segment. This random flight chain is characterized by the length (l) and the
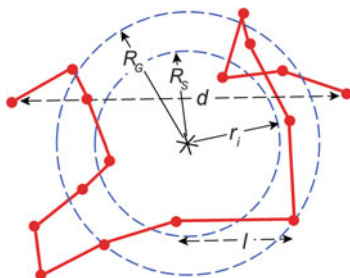
Fig. 2.10 Projection of a random flight chain, composed of 15 segments with a characteristic length *l*. * center of gravity of the molecule, *d* distance between the two ends of the chain, r_i distance of the *i*-th point from the center of gravity, R_G radius of gyration, R_S Stokes radius. This model corresponds to an unfolded polypeptide, composed of 150 amino acids, or a double strand DNA (According to data from G. Damaschun)

number (*n*) of its segments. The position of the endpoint of the segments can be expressed as vectors in a polar coordinate system which is based on the center of gravity of the molecule (Fig. 2.10). The distance between the endpoints of such a molecule (d_{max}) at maximal elongation is:

$$d_{max} = nl \tag{2.30}$$

The real distance of the endpoints of a random flight chain with large enough amounts of *n* segments, and after a sufficiently long time of observation, however, is distributed according to a Gauss statistic. As shown in Fig. 2.10, the mean of this distance is not identical with its maximum, i.e., the value of maximal probability. The mean value of this distance between the endpoints of the chain can be calculated by the following equation:

$$d = \sqrt{\langle r^2 \rangle} = l\sqrt{n} \tag{2.31}$$

In this case the expression $\langle r^2 \rangle$ is the mean of the squares of the distances as defined as follows:

$$\langle r^2 \rangle = \frac{\sum r_i^2}{n} \tag{2.32}$$

Comparing Eqs. 2.30 and 2.31 it can be seen that the mean distance (*d*) between the endpoints of the chain is smaller by $\sqrt{n}$ in relation to its maximum (d_{max}). Considering a chain consisting of 100 segments, the mean distance, therefore, will be only 10% of the maximal length, for chains of 1,000 segments, it will be only 3.2%.

The geometric dimension of such a molecule can be described by the mean of the square distance of all atoms of the molecule from the common center of gravity of

the molecule. This gives the so-called *radius of gyration* (R_G). For a random flight chain with $n \to \infty$ it gives:

$$R_G = \sqrt{\frac{\langle r^2 \rangle}{6}} = l\sqrt{\frac{n}{6}} \qquad (2.33)$$

The radius of gyration, therefore, is proportional to the square root of the length of the chain ($\sqrt{n}$). For compact molecules, however, the monomers of which are not arranged in this way, one gets $R_G \sim \sqrt[3]{n}$, and for rod-shaped molecules, $R_G \sim n$.

Furthermore, it is possible, to define a so-called *Stokes radius* (R_S) for macromolecules, which is measurable experimentally. This is an example of a typical effective parameter. It predicts that the molecule moves in the water like a macroscopic sphere with a hydrophilic surface. In this case, Stokes' law could be applied. It gives the frictional force ($\mathbf{F}$) of a sphere with radius r, moving in a fluid with viscosity η by a velocity $\mathbf{v}$:

$$\mathbf{F} = 6\pi\eta r\,\mathbf{v} \qquad (2.34)$$

(The bold type characterizes the vector character of force and velocity).

The Stokes radius, therefore, is a measurable parameter, which in the same way as the radius of gyration does not correspond to any "visible" radius of the molecule. Both are effective parameters which, however, are well defined (see Fig. 2.10). We will come back to this point in context with the introduction of the hydration radius of ions in Sect. 2.2.2. The relation of the experimentally determined parameters R_G and R_S gives important hints to the configuration of a given macromolecule. If the molecule really resembles a random flight chain, one gets: $R_G/R_S = 1.51$. If the molecules have a spherical, but compact structure, this relation becomes 0.8. In the case of root-shaped molecules, this parameter depends on the relation of the radius to the length of it.

Real molecular chains, composed of covalently bound atoms (Fig. 2.9) are not usually as flexible as predicted by the random flight model. Nevertheless, this model is also applicable in these cases. For this, such monomeric constituents were connected together as segments, matching the conditions of the random flight model. For stiffer molecules one has to choose segments with a larger characteristic length. In the case of a quite flexible molecule of poly-L-glycine, the characteristic segments are formed by approximately four amino acids with a characteristic segment length of $l = 1.38$ nm. For the stiffer molecule of poly-L-alanine, in contrast, 12 amino acids must be combined in one segment with $l = 4.10$ nm.

A repulsive force between the atoms, for example because of electrical charges of equal sign, results in an additional expansion of coiled molecules. A similar limitation of the random flight model is given by the exclusion principle, i.e., by the fact that two atoms cannot be at the same location at the same time. Therefore, the characteristic length of the segment of a DNA molecule as a strong polyelectrolyte is $l = 90$ nm.
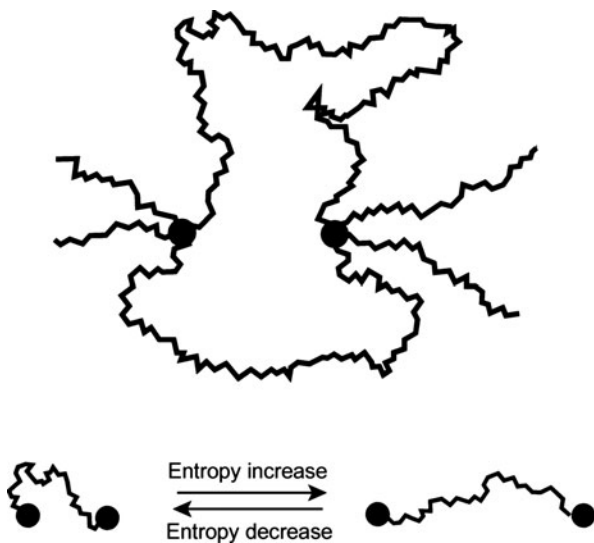
Fig. 2.11 The molecular mechanism of rubber or entropy elasticity. If the distance of the two points of fixation increases, the entropy content of the molecule decreases (schematically illustrated in the elongation of the chain in the lower part of the figure). A force is generated with the tendency to recover the previous situation (the state of maximal entropy) (According to data from G. Damaschun)

If such a chain molecule is fixed at two points (Fig. 2.11) then attractive or repulsive forces are developed. This occurs, if the distance between these points is larger or smaller than the parameter d (Eq. 2.31). The result is the so-called rubber, or *entropy elasticity*. It is caused by the decrease of the entropy of the molecular chain during stretching. In entropy elastic materials, the points of fixation are formed by interconnections of the molecules. This property is the most common type of elasticity of biological materials. We will come back to this in Sect. 3.6.3.

The third kind of molecular movement, mentioned above, *translation*, means a movement in space. This translational movement, driven by thermic hits is important not only for single molecules, but also for microscopically visible particles. This was first observed in 1827 by the Scottish botanist Robert Brown, who observed the stochastic movement of gymnosperm pollen in suspension. This movement was later named *Brownian motion*. The reason for this stochastic translocation of microscopically visible particles is the hits resulting from thermic translations of the surrounding molecules. The movement results from the vector sum of these hits (Fig. 2.12). If the particles are sufficiently small, then a real vector occurs as the sum. For larger particles, this mean vector force becomes zero.

The spatial translocation of the particle, in fact, is stochastic as well as the molecular translocations themselves. Observing the position of the particle over a sufficiently long time, there will be no translocation at all. As a measure of the Brownian motion, therefore, the mean of the squares of the translocations is used. For this, the position of the particle will be registered at definite time intervals

Fig. 2.12 A snapshot indicating the thermic hits by surrounding molecules on a microscopically small particle (*blue arrows*). The individual hits, drawn as fine vectors, induce by vector addition a resting force (*red vector*) which affects the Brownian motion of the particle
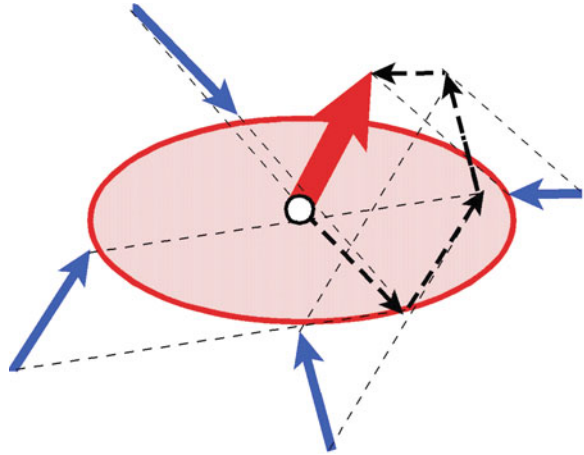
Fig. 2.13 Method to evaluate Brownian motion. The position of a particle was marked at definite intervals of time (*Δt*). This two-dimensional picture subsequently was projected on an arbitrarily drawn axis, which allows us to measure the distances x_i
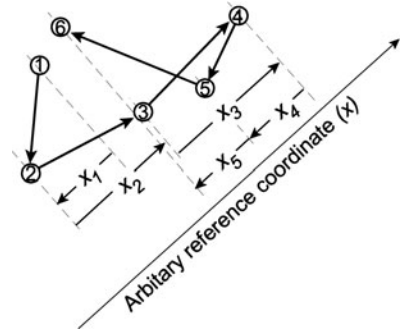
Δt (Fig. 2.13). Subsequently, when the distance x_i is measured, the particle has moved in a projection of an arbitrarily drawn axis. If these values are measured n times, the parameters x_i for $i = 1 \ldots n$ are obtained. The square of displacement $\langle x^2 \rangle$ can be calculated in analogy to Eq. 2.32 in the following way:

$$\langle x^2 \rangle = \frac{\sum x_i^2}{n} \tag{2.35}$$

This parameter corresponds to the square of the mean path length of the moving particle:

$$\bar{x} = \sqrt{\langle x^2 \rangle} \tag{2.36}$$

The intensity of Brownian motion depends on the shape and size of the particle as well as on the temperature and the viscosity of the medium. Einstein and

Smoluchowski derived an equation for spherical particles, meeting the conditions of the Stokes equation (Eq. 2.34):

$$\langle x^2 \rangle = \frac{RT\Delta t}{3N\pi\eta r} = \frac{kT\Delta t}{3\pi\eta r} \tag{2.37}$$

The Einstein–Smoluchowski Eq. 2.37, therefore connected the measurable square of displacement ($\langle x^2 \rangle$) of Eq. 2.35 in relation to the radius of the particle (r), the viscosity of the fluid (η), and the temperature (T), respectively the energy of thermal movement ($RT/N = kT$). Δt is the time interval in which the values x_i were measured.

This equation indicates that the square of the mean translocation length which was achieved by Brownian motion per time interval was inversely proportional to the particle radius (r) and the viscosity (η). For example: a particle with a diameter $r = 1\ \mu m = 10^{-6}$ m moves in water ($\eta = 0.00089$ N s m^{-2}) at $T = 298$ K during $\Delta t = 1$ s at a mean distance of

$$\bar{x} = \sqrt{\frac{1.38 \cdot 10^{-23} \cdot 298 \cdot 1}{3 \cdot \pi \cdot 0.00089 \cdot 10^{-6}}} = 0.7 \cdot 10^{-6} m$$

The size of the particle (r), as well as the mean translocation ($\bar{x}$) are just within the resolution of a normal microscope. Measuring ($\bar{x}$), and knowing (r), it is possible to measure the viscosity (η). This was sometimes used to measure the viscosity of cytoplasm. Conversely it is possible to measure the diameter of very small particles by dark field microscopy in fluids with known viscosity, measuring the parameter $\bar{x}$.

The Einstein–Smoluchowski equation is not only important for these methods. With the help of this equation it is also possible to calculate the time a particle needs to cross a certain distance by diffusion. If, for example, at one point in the cell a vesicle was created and is bound at another definite location by specific receptors, the mean time this particle needs to cross this distance by stochastic movements can be calculated (considering, however, the cytoplasm as an isotropic and homogeneous viscous medium!). Phenomenologically such an event appears to be the result of directed transport. (This, by the way, is an example of a kind of biological Maxwell demon, as discussed in Sect. 2.1.2!)

This translational movement is the basis of all processes of diffusion. Einstein found the connection between the square of displacement ($\langle x^2 \rangle$) and the diffusion coefficient (D) of a molecule in solution:

$$D = \frac{\langle x^2 \rangle}{2\Delta t} \tag{2.38}$$

Introducing Eq. 2.37, it gives

$$D = \frac{RT}{6N\pi\eta r} = \frac{kT}{6\pi\eta r} \qquad (2.39)$$

We will come back to deriving Fick's law of diffusion in Sect. 3.3.1 using the approaches of phenomenological thermodynamics (Eq. 3.130).

From this equation the conclusion can be drawn that the diffusion constants of two molecules with similar molecular shapes are inversely proportional to the square root of their molar masses (M):

$$\frac{D_1}{D_2} = \frac{\sqrt{M_2}}{\sqrt{M_1}} \qquad (2.40)$$

Further Reading

Eyal and Bahar 2008; Tinoco et al. 2002.

2.2 Molecular and Ionic Interactions as the Basis for the Formation of Biological Structures

In Sect. 2.1 the relationship between order and life was discussed, and in this context, the molecular basis of biological structures was considered from the point of view of statistical thermodynamics. Emphasis was placed on the energy of thermal movement and thermal noise (kT). These forces of stochastic movement appear to be opposed by intermolecular forces which tend to create order, building up molecular and supramolecular structures. Therefore, the expression E_i/kT (see: Eqs. 2.16, 2.17, 2.29) characterizes the relationship between any sort of structure-forming energy and the destructive energy of thermal noise.

Now, we will consider in particular those forces that contribute to the formation of molecular and supramolecular equilibrium structures. The forces operating in these so-called self-assembling processes are electrostatic interactions and weak bonds, i.e., bonds with energies of the same order as those of the thermal noise (kT). These bonds, therefore, can be split by thermic hits in a realistic period of time. Such weak bonds, for example, are hydrogen bonds as well as the whole complex of van der Waals interactions.

2.2.1 Some Foundations of Electrostatics

Because electrical phenomena are of major importance at all levels of biological organization (see: Sect. 3.5.1, Fig. 3.33) an account will be given of some fundamental parameters, definitions, and implications.

The basis of electrostatic considerations is electric charges. In the CGS-System electric charges are defined by mechanical parameters. So, for example, the electrostatic unit (eu) is an amount of charge which interacts with an opposite polarized charge of the same size at a distance of 1 cm by an attractive force of 1 dyn. This definition eliminates the need for defining specific electrical units. In the *Système International d'Unités* (SI), however, the application of mechanical parameters has been abandoned, and the charge is defined by the unit for current – the ampere (A), as coulomb (C) = ampere (A) · second (s). In this book, we will use this now generally accepted international system of units.

In electrochemical as well as in biophysical calculations the smallest electric charge is the charge of a univalent ion or a corresponding charged group. The *Faraday constant* (F) gives the number of charges per mole of singly charged ions, an amount, which is usually called *val*:

$$F = 9.6485 \cdot 10^4 C \text{ val}^{-1}$$

Dividing this molar amount by the Avogadro number (N), gives the absolute value of the charge of a single ion (*e*), the same value as that of a positron:

$$e = \frac{F}{N} = 1.6021 \cdot 10^{-19} \text{ C}$$

Using the units of the SI-system, a conversion factor must be applied, to calculate mechanical parameters resulting from electrical interactions. This conversion factor is the so-called *permittivity of free space* (ε_0), which amounts to:

$$\varepsilon_0 = 8.854 \cdot 10^{-12} C \text{ V}^{-1} \text{m}^{-1}$$

Coulomb's law makes it possible to calculate the repulsion force (**F**) of two charges (q_1, q_2) at distance *r* from each other:

$$\mathbf{F} = \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon r^2} \tag{2.41}$$

This equation additionally contains the *dielectric constant* (ε), which is also called *permittivity*. In contrast to ε_0, which unfortunately is usually written with the same symbol, the dielectric constant ε is just a number without dimension. It is a factor indicating, how much the interaction force between two charges will be diminished, if not vacuum, but a real substance is in between. The dielectric constant for pure water at 18°C, for example, amounts to 81.1. For various kinds of glass, it is between 5 and 7, and for arranged layers of lipids about 3.5.

Using Coulomb's law (Eq. 2.41) one can also calculate the energy of electrostatic binding (*E*) of two charges in a medium with $\varepsilon > 1$. With the inverted sign, it equals the work required to move both charges from distance $r = \infty$ to distance $r = r_i$:

$$E = -W = \int_{\infty}^{r_1} \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon\,r^2}\,dr = -\frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon\,r} \tag{2.42}$$

Any charge in space generates an electric field which is characterized by the gradient of the electric potential. The *electric potential* (ψ) is a state parameter of a particular point in space. It is defined as the work that is required to move a positive charge from infinite distance to point r. The unit of the potential is volt (V). In the case of the electric field around a charge q, the electric potential is a radial function of the distance r from this point. According to this definition and Eq. 2.42, the potential ψ_i at distance r_i from charge q is given by:

$$\psi_i = \frac{q_i}{4\pi\varepsilon_0\varepsilon r_i} \tag{2.43}$$

It is important to stress that the electrical potential is a scalar parameter in space, similar, say, to pressure or temperature. Later we will discuss surface potentials, Donnan potentials, diffusion potentials, etc. These are all just electrical potentials. Their names only indicate the reason for their occurrence. At any given time, at one and the same point in space, there can be only one value of an electrical potential, as there can be only one single temperature at one time at one point!

Points of equal potentials in space can be interconnected by *equipotential lines*.

Perpendicular to these equipotential lines electrostatic forces occur. Correspondingly, an electric field strength ($\mathbf{E}$) as a vector parameter can be defined as follows:

$$\mathbf{E} = -\mathrm{grad}\,\psi = -\nabla\psi \tag{2.44}$$

The differential operator "grad" as well as the so-called *Nabla operator* ∇ means: derivation of the potential ψ to all three dimensions of space. It transforms the scalar ψ into vector $\mathbf{E}$.

Figure 2.14 shows the equipotential lines and the field vectors around a positively charged point. The directions of these vectors correspond with the vectors of the electric field strength ($\mathbf{E}$). According to convention, these arrows are always directed from the positive toward the negative pole. Similar illustrations can be seen in Figs. 3.35 and 3.37.

Considering the gradient of the electric field in just one direction (x) of space, Eq. 2.44. becomes:

$$\mathbf{E}_x = -\frac{d\psi}{dx}\mathbf{i} \tag{2.45}$$

In this case $\mathbf{i}$ is the unit vector in the direction of the x coordinate.

In Fig. 2.15 the potential and the field strength in the x-direction are depicted in the vicinity of a homogeneous membrane which separates two different electrolyte solutions (phase I and phase II). This demonstrates that points of discontinuity of

Fig. 2.14 Equipotential lines
and field vectors around a
positively charged point
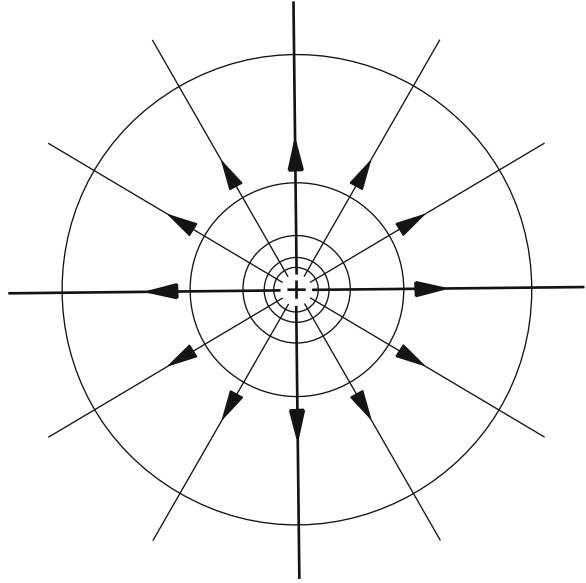
Fig. 2.15 The electrical
potential [$\psi(x)$] and the
electric field strength [$\mathbf{E}(x)$]
near a homogeneous
membrane which separates
two electrolyte solutions with
different composition
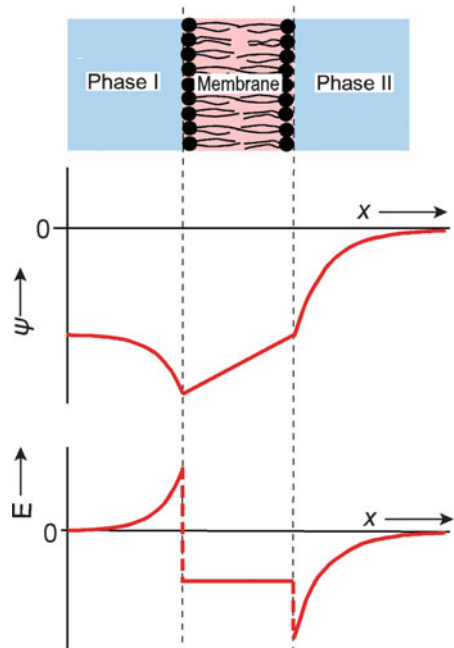
the field strength can occur, with a reversal of the field direction. In this case, the reason for these discontinuities are surface charges on the phase boundaries. The field vectors and the equipotential lines in this example would indicate a set of parallels, perpendicular to each other. Actually, the biological membrane is not homogeneous, but it is a mosaic of various compounds with different surface charges and various dielectric constants. Even its cross section is not homogeneous. Its electrostatic structure is far more complicated, as we will see in Sect. 2.3.6, Fig. 2.48.

The unit of field strength, according to its definition (Eqs. 2.44, 2.45) is $V\ m^{-1}$. As an example, let us calculate the field strength across a cell membrane with a thickness of 7 nm $= 7 \cdot 10^{-9}$ m. If the electrical potential difference between the two surfaces amounts to $\Delta\psi = 50$ mV $= 5 \cdot 10^{-2}$ V, the following field strength results:

$$|E_i| = \frac{5 \cdot 10^{-2}}{7 \cdot 10^{-9}} = 7.14 \cdot 10^6\ Vm^{-1}$$

The term $|E_i|$ means the magnitude of the field strength in the x-direction. Later we will discuss the biophysical consequence of this large electric field strength (Sects. 2.3.6 and 3.4.3).

Using Eqs. 2.41, 2.43, and 2.44. One can calculate the force (F), acting on a charge (q) in an electric field (E):

$$F = Eq \tag{2.46}$$

Furthermore, dipoles must be considered, which play an important role in electrostatics. A dipole is a structure that contains two equal charges (q) of opposite polarity, positioned at a certain distance (l) from each other. Consequently, an electric dipole, as a whole, is electrically neutral. A dipole moment (μ) is defined as the product of charge (q) and distance (l) between them:

$$\mu = ql \tag{2.47}$$

Dipoles can be formed by ampholytic molecules, this means, in molecules carrying simultaneously positive as well as negative charges. They may also result from polarization phenomena of covalent bonds.

According to Eq. 2.47, the unit of a dipole is C m or, correspondingly: A s m. In molecular studies, sometimes the Debye unit (D) is preferred because its amount is in the order of real molecules. The relation between these two units is:

$$1\ D = 3.3 \cdot 10^{-30} Cm$$

In Table 2.1 dipole moments of several compounds are listed. The dipole moments of chloroform, methanol, and water are caused exclusively by polarizations of covalent bonds. In the case of proteins and nucleic acids, the values

Table 2.1 Dipole moments of various substances (After Netter 1959)

	D	$\cdot 10^{-30}$ C m
Chloroform	1.05	3.47
Methanol	1.68	5.54
Water	1.85	6.11
Urea	8.6	28.4
Glycocoll	15	50
Phosphatidylcholine head group[a]	20	66
α-helix (l = 1.3 nm)[b]	63.1	208
Proteins[c]	100–1,000	330–3,300
DNA from calf thymus[d]	32,000	105,600

[a]After Raudino and Mauzerall 1986
[b]See text
[c]After Pethig 1979
[d]After Takashima 1963

listed in this table are to be considered just as orders of magnitude. In these cases charged groups are responsible, whose distance apart, according to the size of these molecules, can be considerable. Single α-helices of proteins have a dipole moment, corresponding to a charge $q = e/2$ on both ends of them. Considering a mean length of such an α-helix of $l = 1.3$ nm corresponding to Eq. 2.47, a value of $\mu = 1.602 \cdot 10^{-19} \cdot 1.3 \cdot 10^{-9} = 2.08 \cdot 10^{-28}$ C m = 63.1 D results.

Ampholytic molecules exhibit electrostatic neutrality in the vicinity of the isoelectric point, since just at this pH-value equal numbers of negative and positive charges are dissociated (for example: -COO$^-$ and -NH$_3^+$ groups, see Sect. 2.2.8). Because of the reversibility of this process, however, a rapid alteration of dissociation and association may take place. A molecule of this type can be considered as a *fluctuating dipole* the charges of which oscillate in a frequency range of about 10^7 s^{-1}.

If a noncharged molecule is influenced by a strong electric field, shifts of electrons may take place, and an *induced dipole* occurs. The resulting dipole moment depends on the external field strength (**E**), and on the ability of the molecule to polarize, a property which is described by a factor α:

$$\mu_{\text{ind}} = \alpha\text{E} \tag{2.48}$$

Such types of induction are measurable as *protein electric response signals*.

At the beginning of this chapter, the dielectric constant (ε) had already been introduced. Now we can characterize this parameter in greater detail on the basis of dipole properties. The electric field strength (**E**) of a plate capacitor is proportional to its surface charge density (σ):

$$\mathbf{E} = \frac{\sigma}{\varepsilon_0\varepsilon} \tag{2.49}$$

The surface charge density expresses the charges per area ($\sigma = q/A$). The introduction of a dielectric between the two plates of the capacitor means an increase of the dielectric constant (ε). As can easily be seen in Eq. 2.49, an increase of the dielectric constant (ε) has the same effect as a decrease of the surface charge density (σ). This is not only a mathematical relation. In fact, it is caused by the polarization of the dielectrics under the influence of the electric field of the capacitor. Therefore, there is a correlation between the dielectric constant and the dipole moment of a substance.

2.2.2   The Structure of Water and Ionic Hydration

The properties of water and its interactions with biological molecules has been a very active area of experimental and theoretical research. Recently Raman, infrared, and NMR spectroscopy, magnetic resonance dispersion, dielectric relaxation, neutron scattering, and various time-resolved fluorescence methods have allowed insights into the molecular properties of water. The liquid water itself, all the more, its interactions with ions and molecules are now known to be so complicated that it stands out as a challenging problem. A complete picture of the complex interactions of hydration water molecules that accounts for all these data is still lacking.

The physicochemical properties of water are quite different from those of chemically similar compounds, such as for example H_2Te, H_2Se, or H_2S. Astonishing values could be predicted extrapolating the properties of these substances to H_2O. This would predict a melting point of $-100°C$ and boiling point of $-30°C$. Water therefore would be a gas under normal life conditions. The same deviations between predicted and existing values would be found in relation to other physicochemical properties, such as molar heat of evaporation, heat capacity, surface tension, dielectric constant, etc.

Another characteristic feature is the so-called *temperature anomaly* of water, which is of particular relevance for the generation and preservation of life. As illustrated in Fig. 2.16, water has its highest density at a temperature of $4°C$. This is the reason why, in contrast to other melting substances, ice swims on the surface, and the temperature of water near the bottom of lakes will not become colder than $4°C$. If water had the same properties as other comparable substances, the cold water would always sink and the lakes would freeze downside up. Life in such a lake would not survive the winter.

In general, there are two molecular properties which are responsible for these effects: the dipole moment of water, and the ability of water molecules to build intermolecular hydrogen bonds.

The dipole property of the water molecule results from an asymmetric distribution of the common electron pair in the covalent O–H bonds. In other words, the molecular orbitals of the bonds are shifted toward oxygen, as the larger of the two covalently bonded atoms. Hence, in both H–O bonds of the water molecule a strong polarization occurs. The common valence electrons in these bonds are strongly
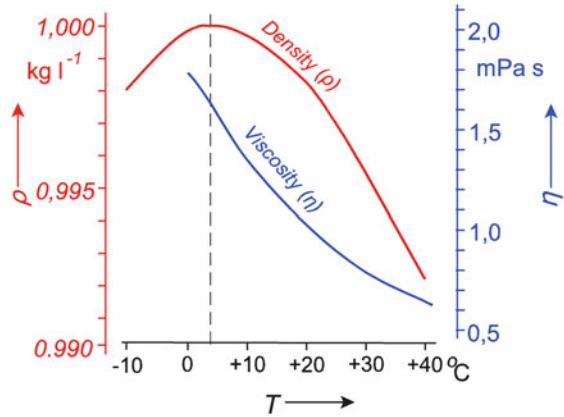
Fig. 2.17 The tetrahedral
arrangement of water
molecules in ice
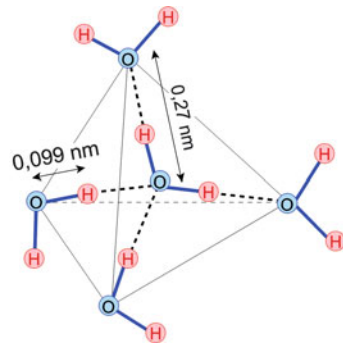
attracted to the oxygen atom, which therefore becomes negative in relation to the
protons. The positive charges of the two H atoms lead to a repulsion of each other.
This increases the angle between them up to a value of 104.5°. The water molecule
as a whole becomes a dipole which is directed towards the bisector of this angle.

This polarization of the H–O bonds in the water molecule not only has
consequences for its dielectric behavior but is also responsible for its ability to
build hydrogen bonds. They can be established between two water molecules as
well as between water and other molecules. Even if the lifetime of these hydrogen
bonds is only about 0.5 ps, these kinds of interactions result in quickly fluctuating
arrangements, which usually are termed *water structure*.

The structure of ice has received the most attention from investigators. Several
ice structure types were found, corresponding to various conditions. The basis of all
of them is a tetrahedral structure as shown in Fig. 2.17. It is built by oxygen atoms
which are always connected by H-bonds to two hydrogen atoms of other molecules.
The continuation of this kind of structure builds up a three-dimensional network
which finally leads to the crystalline structure of ice.

Fig. 2.18 Schematic two-
dimensional illustration of
the arrangement of water
molecules which partly are
connected to clusters
(*green region*) and partly
occur more or less separately
(*blue region*). In the same
scale, a Na⁺- and a K⁺-ion
is included. The *blue circle*,
surrounded by the full line
indicates the crystal radius;
the *red circles* with *dashed
lines* indicate the size of the
Stokes radius
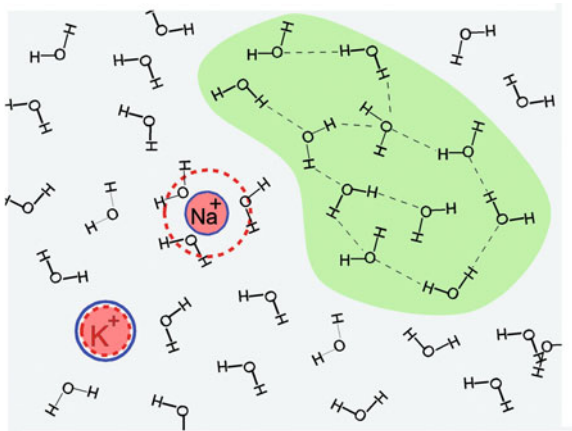
These elements of ice structure will not be destroyed fully if the ice melts. Liquid water should be conceived as a rapidly fluctuating mixture of the intermolecular bonding types found in the polymorphism of ice. On average, in this case, one water molecule makes about 3.5 hydrogen bonds. Nemethy and Scheraga proposed already in 1962 that water clusters diminished with increasing temperature. Near the melting point, about 90–650 water molecules are clustered together. Near the boiling temperature, there are only 25–75 molecules connected in such a cluster (the numbers differ according to different authors). In this way one can understand the temperature dependence of the viscosity of water, as depicted in Fig. 2.16.

This property of water molecules is inherent to the mixed cluster model of fluid water as shown in Fig. 2.18. This model suggests that the tetrahedral structure of water, as depicted in Fig. 2.17 is not the only possible geometrical arrangement of the cluster. The strongly oriented crystalline structure of ice is disturbed in floating water. Irregularities as well as supramolecular structures based on nontetrahedral geometry occur.

These structures are to be considered as being in permanent movement. The lifetime of an H-bond is around 0.5 ps $= 0.5.10^{-12}$ s. The mean lifetime of a single cluster is approximately 10^{-11}–10^{-10} s. Therefore, during the oscillation, every H-atom will form 100–1,000 times a H-bond with the same oxygen atom before it connects to another one. Because of these dynamics, the name *flickering cluster* is used.

In biological systems the water structure is to a great extent determined by its interactions with ions and organic molecules.

The interaction of ions with water molecules leads to their solvation. Inorganic ions interact with water molecules exclusively electrostatically. In Table 2.1 the dipole moment of water is listed as 1.85 D. The electric field around an ion therefore, will lead to a more or less strong orientation of water molecules. This orientating force is in competition with influences from other water molecules and of course, with the destructive activity of thermic noise. As we discussed in Sect. 2.2.1 (Fig. 2.14), the electric field strength decreases with increasing distance from

Table 2.2 Several characteristic parameters of alkali ions

Ions	Relat. atomic mass	Crystal radius	Hydration radius	Equivalent conductivity at infinite dilution	Relative residence time of a H_2O molecule t_i/t
	M	r_k (nm)	r_H (nm)	(Sm^{2}val^{-1})	
Li^+	6.94	0.069	0.240	0.00387	2.60
Na^+	22.99	0.098	0.185	0.00501	1.27
K^+	39.10	0.133	0.126	0.00735	0.54
Rb^+	85.48	0.148	0.119	0.00778	
Cs^+	132.91	0.169	0.120	0.00772	0.59

the center of charge. Correspondingly, two areas at different distances from the ion must be considered: Close to the ion is a region of *primary hydration*. This is a small amount of water molecules that are strongly oriented in the electric field of the ion. The region of *secondary hydration* follows. At this distance the electric field is always too weak to orient the water molecules. Conversely, it is strong enough to disturb the normal water structure. In Fig. 2.18 this is indicated by the destruction of the darkly marked water cluster near the Na^+-ion. In this, it is like a fault zone with fully disturbed water structure.

To quantify this situation several parameters can be used. They are illustrated in Table 2.2. The amount of water molecules, being more or less affected by the ion is indicated by the *solvation number* and is evaluated by NMR-measurements. Investigations of coefficients of self-diffusion of water, including its activation energy, make it possible to determine the relative residence time of water molecules in the vicinity of ions. *Self-diffusion* means the diffusion of water in water, or in solutions. If t_i stands for the mean time in which the H_2O molecule resists near an ion, and t stands for the time interval, during which such a molecule will permanently stand at any point near another water molecule, the relation t_i/t characterizes the degree of demobilization of these molecules near an ion. If $t_i/t > 1$ then obviously an increase of the structural degree of water near the ion occurs, if $t_i/t < 1$ the ion is a structure-breaking, or *chaotropic* agent.

An important parameter to characterize the hydration shell around an ion is the *Stokes radius*. It corresponds to the effective radius which we introduced in Sect. 2.1.6 for spherical molecules. Measuring the equivalent conductivity of an ion at infinite dilution, it is possible, to calculate its electrophoretic mobility, i.e., its mobility in relation to the applied field strength (see: Eq. 2.84). Equation 2.46 allows one to calculate the force that is applied to an ion in a given electric field ($\mathbf{E}$). Under conditions of stationary movement, this driving force is fully compensated by the friction which results from the interaction of the moving ion with the surrounding water. Now let us postulate that this frictional force can be calculated by the Stokes law (Eq. 2.34). This means the ion is assumed to be a macroscopic sphere with a hydrophilic surface. "Macroscopic" means – large enough that water can be considered as a continuum, not mentioning its molecular structure. In this case, knowing the viscosity (η), the driving force ($\mathbf{F}$), and the mobility ($\mathbf{v}$) of the

ion, the Stokes equation makes it possible to calculate an equivalent radius (r) of it as a sphere.

This *Stokes radius* is a typical example of an *effective parameter*. It is exactly defined by its way of measurement and calculation. Conversely, it does not exist as a visible or at least imaginable border. As indicated in Fig. 2.18, it is in no form identical for example to the thickness of a single layer of water molecules. In fact, the Stokes equation is not the proper law to calculate the real movements of an ion in water as the conditions listed above are not realized in this case. Nevertheless this parameter, because of its exact definition, is an important measure for the hydration properties of ions. However, when using it for example to explain membrane permeability of ions or to discuss other similar mechanisms it is always necessary to be aware of the conditions of its definition.

As illustrated in Table 2.2, the solvation numbers as well as the Stokes radius of the alkali ions decrease with increasing atomic mass. In the case of lithium, the hydration radius is much larger than the crystal radius, whereas conversely for potassium it is already smaller. In this case K^+ and Cs^+ are called *structure-breaking* or *chaotropic* ions. The region of secondary hydration, i.e., the fault zone of destroyed water structure is larger than that of electrically oriented molecules. In contrast to this, in cases of small ions, like Li^+ and Na^+ the region of primary hydration is far larger. These are *structure-building* or *cosmotropic* ions. This category also includes the two-valent cations Mg^{2+} and Ca^{2+} which are not mentioned in the table.

What is the reason for these differences? Both ions, Na^+ as well as K^+ carry the same single positive charge. The Bohr's radius of these ions, i.e., the effective radius of the outer electron shell which is the same as the crystal radius, however, is larger for K^+ than for Na^+, because of the additional electron orbital in the case of potassium. Where there is a larger distance from the charged nucleus of the ion, the electric field strength will become lower. The ions did not alter the hydrogen bonding network outside their direct vicinity. No long-range structure-making or structure-breaking effects for either cosmotropes or chaotropes take place.

Further Reading

Griffith and Scheraga 2004; Israelachvili 1994; Lynden-Bell et al. 2010; Schmid 2001; Zielkiewicz 2005; Wernet et al. 2004.

2.2.3 *Interaction of Water with Macromolecules*

Water is an essential compound of all living systems. The specific properties of the water molecule, as listed in the previous chapter, are the preconditions for development and existence of life. We will discuss this in many chapters of this book and consider them over different levels of biological organization.

In general, the chemical activity of the surrounding water influences the functions of proteins at least in the same way as the pH of the medium or some specific ligands. Membrane transport proteins for example, periodically

change their water binding during opening and closing. The hemoglobin molecule adjusts its water content during binding and releasing of oxygen. Similar processes occur in various enzymic reactions. We are only now beginning to understand the various types of water binding reactions and their kinetic parameters. Various theoretical approaches and experimental techniques lead to different models of these kinds of interaction.

One of the phenomenological approaches concerns interactions of water with nonpolar molecules or nonpolar groups of a molecule, which is called *hydrophobic interaction*. This is a dominant factor in the stabilization of biological membranes and other supramolecular structures.

Hydrophobic molecules, in general, behave in relation to the surrounding water like an air bubble, producing a surface tension. From the phenomenological point of view, water molecules surround a nonpolar solute without sacrificing much of the H-bonding; rather, H-bonding networks are stabilized by the presence of other molecules. In view of the tangential orientation of the first water sphere around this solute, a cage is formed which can be compared to an elastic net compressing these hydrophobic structures.

In this way, hydrophobic interactions influence not only the entropy of the macromolecules, but also the entropy of the surrounding water. Because of the structure-building property of the hydrophobic interface, the entropy of the water near the surface is lower than that in the bulk phase. In case of conformational changes of macromolecules, or of processes of supramolecular organization, the surface available for interaction with water molecules may decrease. This happens for example if lipid molecules aggregate to form membranes or vesicles. In this case the water between the molecules will be extruded.

From a thermodynamic point of view, the entropy of the enclosed macromolecule decreases as the result of proper orientation, but simultaneously, the entropy of the surrounding water increases because of the loss of structure-building surfaces. In general, therefore, even if a macromolecular structure with a higher degree of organization results, the entropy of the whole system increases. According to the second principle of thermodynamics therefore, these kinds of processes may occur spontaneously. They are called *entropy-driven*. This is an important mechanism leading to the spontaneous formation of secondary and tertiary structures of macromolecules (Fig. 2.19).

It was found that the desorption of water from surfaces leads to a certain increase in its partial molar volume (Fig. 2.19). Obviously, the more highly organized water molecules at surfaces are packed closer together. In this respect, we should mention the so-called *Le Chatelier principle*. It predicts that if a constraint is applied to a system in equilibrium, the system adjusts to a new equilibrium that tends to counteract the constraint. An increase of the hydrostatic pressure, in this sense, could be counteracted by the decrease of the partial volume of water in the system. In Fig. 2.19 this would mean: an increase of pressure leads to an increase of hydration, and therefore to a decrease in organization of the macromolecular, or the supramolecular structure. Hence, an increase of hydrostatic pressure should destroy structures that are connected by hydrophobic interactions.
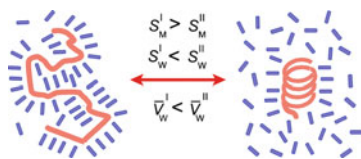
Fig. 2.19 The orientation of water molecules, schematically drawn as bars, near the surfaces of macromolecules. In contrast to the decrease of entropy of the macromolecule during formation of the helix ($S_M' > S_M''$), the entropy of water during this process increases ($S_W' < S_W''$). This leads to an increase of the entropy in the whole system. At the same time the partial molar volume of water increases
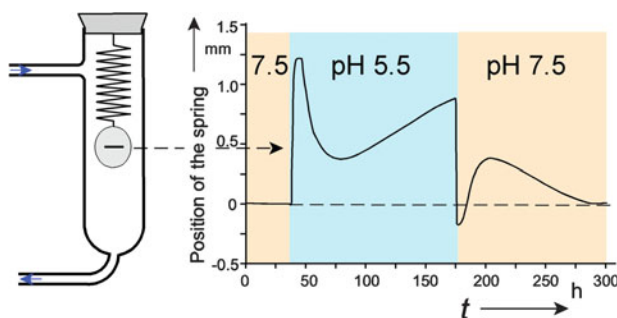

Fig. 2.20 Measurement of the kinetics of a hydration process, using a dialysis sac hanging on a quartz spring. The curve indicates results obtained on the protein of the tobacco mosaic virus (According to data from Lauffer 1975)

In fact, extremely high pressures of the order of 10–100 MPa are necessary for such reactions. Membranes, microtubules, and other supramolecular structures will disintegrate under these conditions. In some cases high hydrostatic pressures were applied to study the physiological properties of membranes. In this context it must be remembered that in the deepest parts of the ocean, i.e., at about 10 km, a hydrostatic pressure of 100 MPa occurs!

Figure 2.20 illustrates how the shift of the partial volume of water can be measured directly during a dissociation–association reaction. For this, in a vessel with exactly controlled temperature the macromolecular sample, enclosed in a dialysis sac, hangs on a very sensitive quartz spring. The surrounding solution of this dialysis sac is exchanged continuously. A modification of the density of the macromolecules will be reflected in a change of the length of the spring. As indicated in the figure, the observable changes are minimal. In this case the change during the pH-dependent polymerization of proteins of the tobacco mosaic virus is demonstrated. The time which is necessary to arrive at equilibrium is quite long.

In fact, there are various kinds of interaction between water and organic molecules. Charged groups of these molecules may interact electrostatically, like ions as described in Sect. 2.2.2, i.e., as a *hydration of the first order*. Attachment of water molecules by hydrogen bonds in contrast is called *hydration of the second order*.
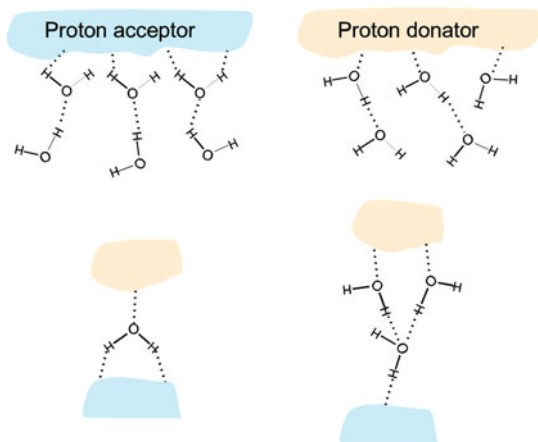
In view of this kind of interaction, two types of molecular surfaces must be considered: proton-donator and proton-acceptor surfaces. As depicted in Fig. 2.21, this difference is responsible for the orientation of the attached water molecules. It may be transmitted by a few layers of water molecules. This kind of hydration can also lead to an interaction between two surfaces. Surfaces with identical orientation of water molecules repel each other, surfaces with opposite orientation of water molecules show attractions at low distances.

In this context Ramachandran already in 1968 coined the term *water bridges* (sometimes even called *Ramachandran bridges*). These connections are important for a number of intermolecular protein interactions forming a ternary structure, as for example the triple helix of collagen. These models proposed two types of water bridges: single water bridges, connecting the molecules over a shorter distance, and double bridges, including three water molecules (Fig. 2.21). Furthermore, dielectric water clusters exist around these bridges, and in some cases inside the protein volume encapsulated bulk water is sequestered.

These circumstances have been established by stoichiometric hydration models, based on predictions from known protein structures and supported by experimental data. A number of methods are available to measure the kind of protein hydration, such as gravimetric measurements (relating the mass of a hydrated protein to one which was dried to equilibrium in vacuum), determination of rehydration rate (indicating various time constants of rehydration), differential scanning calorimetry (direct recording of the energy of the hydration–dehydration process), various proton NMR titration methods, and finally osmotic methods as described in Sect. 3.2.2.

The behavior of water molecules localized in this way near the protein surface, i.e., the time scales of rotational and translation dynamics, as well of the kinetics of its exchange and self-diffusion, shows a wide range of time scales. A complete picture that accounts for all the data is still lacking. In contrast to the hydrogen bond lifetime of 0.5 ps in pure water, NMR and magnetic resonance dispersion show that most water molecules near the protein surface are slowed approximately onefold to

twofold. Quasi-elastic neutron scattering experiments typically show two-step kinetics, with fast (ps) and slow (100 ps–ns) relaxation times. In no case could a memory effect be found, i.e., a conservation of structural properties of water lasting longer times, as argued by homeopaths and other advocates of alternative medicine.

There are also some other peculiarities of the physical properties of water near surfaces in relation to the bulk water. In a layer of 0.1–0.2 nm, for example, the effective dielectric constant of water decreases dramatically up to the value of 10. Furthermore, because of the water structure, this region may become anisotropic. This means, for example, that diffusion rates become higher than the average value in the direction parallel to the surface and lower in the direction normal to it.

Further Reading

Benz and Conti 1986; Fullerton and Cameron 2007; Makarov et al. 1998; Parsegian 2002; Raschke 2006; Takashima 1986.

2.2.4   *Ions in Aqueous Solutions, the Debye–Hückel Radius*

In the previous sections, we already discussed the electric field that was induced by a single charged point, and which influences molecules in their vicinity. When considering extremely short distances, calculating for example force and energy of ionic bonds, or the influence on water dipoles nearby, it is possible to consider only this single charged point and to use Coulomb's law (Eq. 2.41) for calculation. In the case of larger distances from the source of the field, however, the influence of other charges must also be considered. In a 100-mM solution of NaCl, for example, the mean distance between the ions only amounts to about 2 nm. This means that under physiological conditions the interactions of ions must be considered as a set of charged points.

These kinds of calculations are possible on the basis of the theory of strong electrolytes, the *Debye–Hückel theory*. This theory considers an ionic cloud around an arbitrarily chosen central ion. This central ion is taken as the center of polar coordinates. It attracts ions with opposite charges (counterions) and repels those with identical charges (co-ions). In this way an electric field will build up, generated by the central ion, as well as by the ions surrounding it. This field is spherically symmetric and can be described by the function $\psi(r)$.

At a distance r from the central ion, other ions will be attracted or repelled by the central ion, depending on their charge. We will denote the number of charges of an ion by z_i, for example: $z_{Na} = +1$, $z_{SO_4} = -2$, etc. The electrostatic energy of an ion at a particular point in space with an electrical potential ψ amounts to $z_i e \psi$.

Boltzmann's law of energy distribution (Eq. 2.16) makes it possible to calculate the concentration (c_i) of the ions i at a point with the potential ψ. In this case, the energy of an ion in the electric field ($z_i e \psi$) is related to the energy of thermal noise (kT). The orientation of the ions in the electric field is opposed by thermic fluctuations.

$$c_i(\psi) = c_{i0}\, e^{-\frac{z_i e\psi}{kT}} \tag{2.50}$$

In this equation, c_{io} is the concentration of the ion i, far away from the influence of the central ion, which means in the bulk solution. To answer our question, we additionally should know the real amount of potential ψ at this point, because it depends on the concentration of the ions in the ionic cloud. Furthermore, we are not interested in function $c_i(\psi)$, but would rather know the concentration and potential distribution in space, i.e., the functions $c_i(r)$ and $\psi(r)$. This is possible using the *Poisson equation*. This is a partial differential equation of the second order, which gives function $\psi(r)$ in a space as a result of a charge density $\rho(r)$:

$$\nabla^2\psi = -\frac{1}{\varepsilon_0\varepsilon}\rho \tag{2.51}$$

$\nabla^2\psi$ means the second derivation of the potential by the coordinates of the space. This Nabla operator has already been used in Sect. 2.2.1 (Eq. 2.44). For ε the dielectric constant of water can be used.

The charge density (ρ, in $C\,m^{-3}$) can be calculated from the ion concentration c_i of the ions in the cloud by:

$$\rho = \sum_{i=1}^{n} c_i z_i N e = F \sum_{i=1}^{n} c_i z_i \tag{2.52}$$

Introducing Eq. 2.50 in 2.52, and both in 2.51, one obtains the *Poisson–Boltzmann equation*:

$$\nabla^2\psi = -\frac{F}{\varepsilon_0\varepsilon} \sum_{i=1}^{n} c_{i0} z_i\, e^{-\frac{z_i e\psi}{kT}} \tag{2.53}$$

The expression (2.53) is a partial differential equation, which can be solved analytically only under simplified conditions. Let us consider the following situation: Let the solution contain only a single kind of ions i, let the potential be very low ($z_i e\psi \ll kT$) and simply a spherical symmetric function from the central ion. For this simplified case, the solution of Eq. 2.53 becomes:

$$\psi_r = \frac{z_i e}{4\pi\varepsilon_0\varepsilon r}\, e^{-\kappa r} \tag{2.54}$$

Comparing this equation with the simple Coulomb's law (Eq. 2.43), it shows that the influence of the potential of the central ion will be diminished by the factor $e^{\kappa r}$ (considering: $e^{-\kappa r} = 1/e^{\kappa r}$). The extent of this decrease depends on the concentration of the ions in the solution (see Fig. 2.22). This becomes understandable when

considering the definition of the *Debye–Hückel parameter* (κ) in the following equation:

$$\kappa = \sqrt{\frac{e^2 N}{\varepsilon_0 \varepsilon k T} \sum_{i=1}^{n} c_{io} z_i^2} = \sqrt{\frac{F^2}{\varepsilon_0 \varepsilon R T} \sum_{i=1}^{n} c_{io} z_i^2} = \sqrt{\frac{2 F^2 I}{\varepsilon_0 \varepsilon R T}} \qquad (2.55)$$

The first expression of this equation can easily be transformed into the second one, using the relation between the Faraday constant (F) and the charge of a single electron (e) as: $F = eN$, as well as the relation: $R = kN$. In the last expression of Eq. 2.55, the *ionic strength* (I) was introduced.

$$I = \tfrac{1}{2} \sum_{i=1}^{n} c_{io} z_i^2 \qquad (2.56)$$

When calculating the value of κ for real conditions, it can be seen, that this Debye–Hückel parameter has the measure: m^{-1}. (For this of course, the concentration measure: mole per m^3 must be used!). The parameter $1/\kappa$ therefore is a measure of distance. Accordingly, it is called the *Debye–Hückel length* or *Debye–Hückel radius* of the ionic cloud. Considering Fig. 2.22 it is obvious that this value $1/\kappa$ again is not a "visible" radius, but just an effective parameter.

In the following example we will demonstrate the calculation of the ionic strength of a solution: Let us consider a physiological solution with the following composition: 105 mM NaCl, 5 mM KCl, 25 mM Na_2HPO_4, 2 mM $CaCl_2$ (mM representing the usual measure of concentration: mmol l^{-1}). Let us assume a full dissociation of the Na_2HPO_4. Using Eq. 2.56, one obtains:

$$\begin{aligned}
I &= \tfrac{1}{2}(c_{Na} z_{Na}^2 + c_K z_K^2 + c_{Cl} z_{Cl}^2 + c_{HPO_4} z_{HPO_4}^2 + c_{Ca} z_{Ca}^2) \\
&= \tfrac{1}{2}(0.155 \cdot 1^2 + 0.005 \cdot 1^2 + 0.114 \cdot 1^2 + 0.025 \cdot 2^2 + 0.002 \cdot 2^2) \\
&= 0.191 \text{ mol } l^{-1}
\end{aligned}$$

Fig. 2.22 The electrical potential (ψ) as a function of the distance (r) from a central ion. The *blue line* ignores the influence of the ionic cloud according to Coulomb's law (Eq. 2.43); the *red lines* show the potential $\psi(r)$ which is diminished by the ionic clouds in solutions of various concentrations of a one-one-valent salt ($z = \pm 1$) according to Eq. 2.54
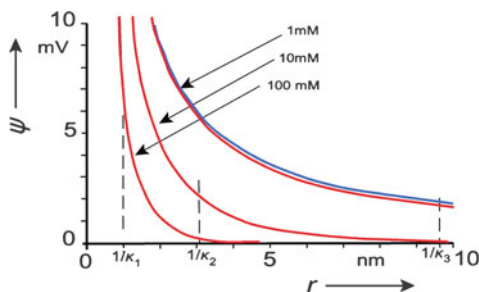
According to Eq. 2.55, the Debye–Hückel radius (in: nm) at temperature $T = 298$ K can easily be calculated by the following equation, using as concentration measure $mol \cdot l^{-1} = M$:

$$\frac{1}{\kappa} = \frac{0.304}{\sqrt{I}} \quad \text{(in nm)} \tag{2.57}$$

This means for the case of the above considered solution:

$$\frac{1}{\kappa} = \frac{0.304}{\sqrt{0.191}} = 0.696 \quad \text{(in nm)} \tag{2.58}$$

Summarizing these considerations, we conclude that with the Poisson–Boltzmann equation (Eq. 2.53) and its integration, the electrical potential can be calculated dependent on the distance from a central ion. In this case the screening effect of the ionic cloud is considered. The degree of this screening, in fact, increases with increasing ionic strength of the solution. The Debye–Hückel radius ($1/\kappa$) is defined as a measure of the thickness of the ionic cloud, which, in contrast to the crystal radius or the hydration radius, is not a property of a specific ion but just depends on the ionic strength of the solution. The Debye–Hückel radius is an effective parameter in the same way as the others (see Sect. 2.2.2).

It must not be forgotten when using the Debye–Hückel equation that already a number of simplifications were used for its derivation and for integrating the Poisson–Boltzmann equation (Eq. 2.53). Furthermore, several other interactions were ignored. In fact, already in solutions of physiological ionic strengths, a number of experimental data, such as the activity, or osmotic coefficients of ions, show significant deviations from the theoretical predictions. Much effort has gone into extension of this theory, from adding empirical correction factors, to theoretical extensions using advances in the statistical mechanical theory of liquids, or quantum mechanical approaches. The most limiting condition of the Debye–Hückel theory is the ignorance of specific interactions of the ions with other charged groups and molecules. This concerns a number of electrodynamic interactions which will be explained in the next chapter.

Further Reading

Loehe and Donohue 1997.

2.2.5   Intermolecular Interactions

Intermolecular interactions, beginning with various enzyme reactions, receptor properties up to the self-assembly of biological structures represent an intricate process consisting of an interplay between a variety of different intermolecular forces. In Sects. 2.2.2 and 2.2.3 we have already explained the role of hydrogen bonds in building

the structure of water, in interactions among ions and between ions and water dipoles, as well as the role of hydrophobic interactions in the process of self-organization of macromolecules. In the following the spectrum of these interactions will be extended to include additional electrostatic and electrodynamic forces. In fact, any kind of interaction between molecules in an aqueous medium is the sum of several forces which differ by their sign (i.e., repulsive or attractive), and by their distance function.

Let us first consider a dipole interacting with an ion. In this case, the energy of electrostatic binding (E_{ID}), can be calculated using Coulomb's law (Eq. 2.41). In the simplest case it results from the sum of all electrostatic components, namely the interaction of the ion (q_1) with one ($-q_2$), as well as with the other ($+q_2$) charge of the dipole. Considering a linear orientation of the dipole in relation to the ion (Fig. 2.23a), it gives:

$$E_{ID} = \left[-\frac{q_1(-q_2)}{4\pi\varepsilon_0\varepsilon r} \right] + \left[-\frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon(r + l)} \right] \qquad (2.59)$$

We will consider simply that the dipole $\mu = q_2 l$ contains two charges with equal size but opposite signs. A simple algebraic rearrangement of Eq. 2.59, and the introduction of the dipole moment (μ) results in:

$$E_{ID} = \frac{q_1 q_2 l}{4\pi\varepsilon_0\varepsilon r(r + l)} = \frac{q_1 \mu}{4\pi\varepsilon_0\varepsilon r(r + l)} \qquad (2.60)$$

Assuming: $r > l$, Eq. 2.60 will be simplified to:

$$E_{ID} \approx \frac{q_1 \mu}{4\pi\varepsilon_0\varepsilon r^2} \qquad (2.61)$$

In contrast to the energy of interaction of two ions (Eq. 2.42) as a function of $1/r = r^{-1}$, this equation shows that the interaction energy of an ion with a dipole decreases with r^{-2}.

The mutual interaction between two dipoles (E_{DD}) can be calculated in a similar way. Let us again look at the simplest case of collinear orientation of both dipoles, as depicted in Fig. 2.23b. Using the same approach, one gets:

$$E_{DD} = -\frac{1}{4\pi\varepsilon_0\varepsilon} \left[\frac{q_1 q_2}{l_1 + r} + \frac{q_1(-q_2)}{l_1 + r + l_2} + \frac{(-q_1)q_2}{r} + \frac{(-q_1)(-q_2)}{r + l_2} \right] \qquad (2.62)$$

Fig. 2.23 Ion–dipole (a) and dipole–dipole (b) interaction. q charge, l length of the dipoles, r distance
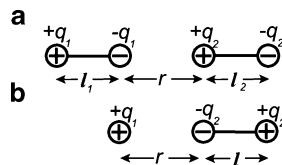
Again we will rearrange this equation and introduce the dipole moments μ_1 and μ_2:

$$
\begin{aligned}
E_{DD} &= \frac{q_1 l_1 q_2 l_2}{4\pi\varepsilon_0\varepsilon} \cdot \frac{l_1 + l_2 + 2r}{(l_1 + r)(l_2 + r)(l_1 + l_2 + r)r} \\
&= \frac{\mu_1\mu_2}{4\pi\varepsilon_0\varepsilon} \cdot \frac{l_1 + l_2 + 2r}{(l_1 + r)(l_2 + r)(l_1 + l_2 + r)r}
\end{aligned}
\tag{2.63}
$$

Using the same approach: $r > l$, gives:

$$
E_{DD} \approx \frac{\mu_1\mu_2}{4\pi\varepsilon_0\varepsilon} \cdot \frac{2r}{r^4} = \frac{2\mu_1\mu_2}{4\pi\varepsilon_0\varepsilon r^3}
\tag{2.64}
$$

This sequence of possible electrostatic interactions can be extended including interactions with induced dipoles (see Sect. 2.2.1). In this case, a more complicated function of the distance must be expected because the induction of a dipole itself depends on the electric field (Eq. 2.48). For this case, a function can be calculated, including r^{-4}. If an interaction of two dipoles is considered, inducing each other, a function with r^{-6} will be obtained.
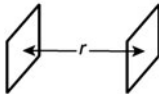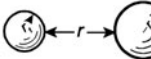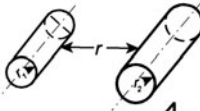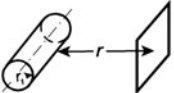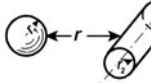
This brings us to the wide field of *van der Waals interactions*. The plural of this expression indicates that it includes a number of different intermolecular forces. Van der Waals interactions in general, are attractive forces between molecular components, whole molecules, or supramolecular particles that are not simply based on electrostatic interactions. They are rather attributed to electromagnetic interplay, occurring by fluctuations of charges. The reason for such charge fluctuations in molecules may be different. There can be thermal molecular translations, or oscillations in the thermic level of electrons, as illustrated in the Jablonski diagram (Fig. 4.34, Sect. 4.8.2). Furthermore, fluctuations occur in the electron structure of molecules which can be calculated by quantum mechanics. This is the basis of the so-called *dispersion forces*, or *London-dispersion forces*, sometimes also called *London–van der Waals interactions*. These oscillations occur at frequencies near to that of visible light.

The understanding of van der Waals forces was substantially advanced by F. London, H. C. Hamaker, and E. M. Lifschitz. Nevertheless, many aspects are unclear even today. The central question concerns the character of this interaction, and its energy as a function of interparticle distance (r). Furthermore, it still seems unclear whether in this kind of interactions a molecular specificity can exist which might be related to its frequency.

The energy of van der Waals interactions is governed by the *Hamaker constant*, which results from quantum mechanical calculations. The distance relation of this energy depends strongly on particle geometry. This is illustrated in Fig. 2.24. There is a difference whether it is an interaction between a spherical and/or a cylindrical particle and additionally, there are different functions $f(r^{-n})$ dependent on the total

Fig. 2.24 The distance
function [$f(r^{-n})$] of the van
der Waals energy of
interaction between particles
of various shapes (parameters
after Israelachvili 1994)

distance between these particles. This means at the same time that simple hyperbolic functions for this kind of interactions are just rough approximations.

To understand the relevance of van der Waals forces, it is necessary to consider them in context with electrostatic interactions. Here we touch upon a central concept of colloid chemistry, especially the concept which is connected with the names of B. V. **D**erjaguin, L. D. **L**andau, E. J. W. **V**erwey, and J. Th. G. **O**verbeek. This *DLVO theory* was originally developed for modeling the behavior of lyophobic colloids (see: Verwey and Overbeek 1948). To extend this approach to proteins, they are considered to be hard, polarizable spheres with charges, uniformly distributed over their surface. Water is treated as a structureless, continuum dielectric medium. The basic ideas of this approach, considering the Debye–Hückel theory of ionic clouds, as well as the van der Waals properties, can be listed as follows:

– The interaction force between two particles results from the sum of the particular forces, each one decreasing with increasing distance in different ways.
– In contrast to generally attractive van der Waals forces, electrostatic interactions can also become repulsive.
– The intensity of electrostatic interactions, in contrast to van der Waals forces, depends to a high degree on environmental conditions.

Fig. 2.25 Functions: $y = r^{-n}$
to illustrate the character of
short- and long-distance
interactions
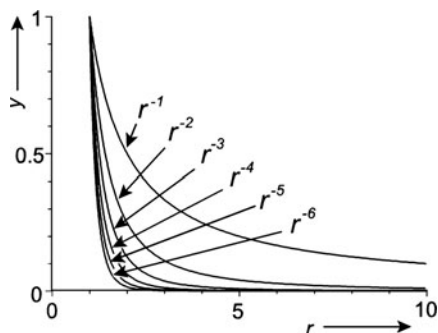
Fig. 2.26 The total energy
of interaction (E) (*red line*)
between two particles as a
sum of three functions
(*blue lines*): One type of
electrostatic repulsion (r^{-3}),
and two kinds of van der
Waals attractions (r^{-1} and
r^{-6}). The point r_{min} indicates
the distance of the secondary
minimum
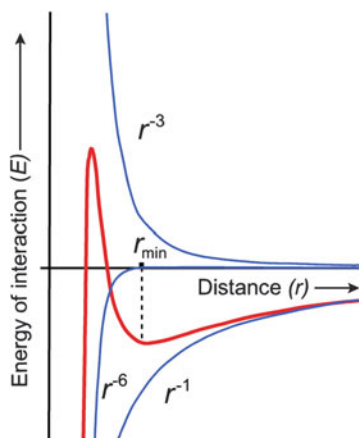
To illustrate the consequences of these postulates, we must imagine that the
distance function of the interaction energy $[E(r)]$ results as a sum of various
functions, shown schematically in Fig. 2.25. There are so-called *short-distance
interactions*, decreasing with r^{-5} or r^{-6}, and *long-distance interactions*, the curves
of which do not decrease as fast (r^{-1}). Some of these curves can be looked at upside
down, representing repulsive forces, which consequently, must be subtracted by the
others. Additionally, the dependence of the electrostatic forces on the ionic strength
of the solution is to be considered, as illustrated in Fig. 2.22, as well as the various
distance functions of the van der Waals forces, as depicted in Fig. 2.24. Hence, the
shape of the resulting sum can be quite complicated.

Figure 2.26 illustrates this situation for a more or less simple case of the interac-
tion of two particles in a solution. In this case the interaction is supposed to be the
sum of an electrostatic repulsion (r^{-3}) and two kinds of van der Waals attraction
forces, one as a long (r^{-1}), and another as a short-distance interaction (r^{-6}). This
curve demonstrates at least qualitatively some basic properties of macromolecular
interactions. It shows that two particles may attract each other at larger distances,
even if they carry charges of the same sign, if additionally a long-ranging van der

Waals attraction occurs. This is demonstrated in Fig. 2.26 at distance $r > r_{min}$. The point r_{min} marks the position of minimum interaction energy. The attractive force ($\mathbf{F} = \mathrm{d}E/\mathrm{d}r$) at this point becomes zero. If this distance is decreased a repulsion force appears. The particles therefore, will rest at this point, which is called the distance of the *second minimum*. If in any way this repulsion force were to be overcome, attractive interactions again would take place and the particles would aggregate at the distance of the *primary minimum*. In Fig. 2.26 this primary minimum is not shown. This would predict an inclusion of further components of interactions at very low distances. So for example, the influence of Bohr's radii must have been considered, the extremely strong repulsion forces in the case of interactions of the electron shells.

As already mentioned, the electrostatic interaction strongly depends on the ionic conditions of the medium. This is understandable in view of Eq. 2.54 (Sect. 2.2.4), as well as Fig. 2.22. Increasing the ionic strength, for example, the energy barrier in the region $r < r_{min}$ will become diminished. Finally, thermic hits could be sufficient to let the particles cross this barrier and to agglutinate. This effect is experimentally verified. An increase of the salt concentration of a colloid solution leads to the agglutination of the particles. This is known as the *salt-out effect*, in contrast to the *salt-in* process, leading to a re-stabilization of a colloidal system.

Despite the general success of DLVO theory, there are many situations where it is unable to provide even a proper qualitative description of macromolecular interaction. The reason for this, particularly considering protein–protein interactions, is the ignorance of a number of circumstances, such as for example:

– Protein–ion interactions cannot be simply expressed by ionic strength. The limitations of the Debye–Hückel theory as mentioned in the previous section must be considered. This particularly concerns specific properties of anions. In this respect, the DLVO theory is unable to predict any dependence on salt type.
– The proteins cannot be taken simply as charged spheres. The charged groups are rather distributed not homogeneous. This results in an anisotropic character of protein–protein interactions. In fact, weak protein–protein interactions should be treated more like molecular recognition processes.
– Hydrophobic interactions must be considered. Protein–protein interactions are very sensitive to surface properties, such as protonation states of the charged residues on the surfaces, or of surface hydrophobicity.

In 1888, the German pharmacologist Franz Hofmeister, investigating the colligative properties of colloids, had already determined that the ability of an ion to induce protein precipitation strongly depends on its type. A *Hofmeister series* was developed as a ranking of the salting-out effectiveness of various ions for globular proteins. The effectiveness is much greater for anions than for cations in the decreasing order: $SO_4^{-2} > HPO_4^{-2} > CH_3COO^- > Cl^- > Br^- > I^- > SCN^-$, and correspondingly: $Li^+ > Na^+ \sim K^+ > NH_4^+ > Mg^{2+}$. This sequence refers to properties of anions and cations, on a wide range of phenomena. Particularly the phosphate is of physiological interest, with its position at one end of the Hofmeister sequence, and its strong adsorption at biological surfaces.

The Hofmeister series cannot be explained by the classical DLVO theory. One reason for this is that van der Waals interactions and other electrodynamic interactions until recently were not taken into account for the description of ion–ion interactions. Already at physiological concentrations the Coulomb interactions are more or less screened. This is not the case for the dispersion forces which react under these conditions. In the case of physiological solutions, where the Debye–Hückel length amounts to only about 0.6 nm (see Eq. 2.58 in Sect. 2.2.4), dispersion forces may become important even for ions.

In molecular biology these kinds of interplay between various forces of inter-molecular interactions are very important. This is one way to control supramolecular processes. The influence is possible either by control of the dielectric constant, or by changing the ionic conditions. The local dielectric constant depends on the arrangement of molecules, on their polarization and on the water structure in their vicinity. Mostly, these processes, however, are controlled by changing local ionic compositions.

In this context the influence of multivalent ions is of particular interest. Let us mention the role of calcium in muscle contraction, nerve excitation, exocytosis, cell adhesion, and many other processes. In colloid chemistry the so-called *Schulze–Hardy rule* is an empirically determined relation which later was also verified theoretically. This rule predicts that the concentration of a salt, which would be necessary to overcome the energy barrier between the second and the first minimum, is inversely proportional to the 6th degree of its valency. This means that the critical concentration produced by a univalent ion relates to this concentration of a bivalent one, like 1^6–2^6, which means: 1–1/64. Consequently, 2-mM Ca^{++} would be as effective as 128-mM K^+ or Na^+. This is a relation that reminds us of the real concentrations of in vivo conditions. It must however be considered that this would not explain the difference in the behavior of Ca^{++} in contrast to Mg^{++}. Other reasons must also be taken into consideration.

Further Reading

For van der Waals interactions in molecular biology: Israelachvili 1994; Curtis and Lue 2006; Hofmeister effect: Kunz et al. 2004; Zhang and Cremer 2006.

2.2.6   Structure of Proteins

In the preceding sections problems relating to the information content of biomacro-molecules were discussed, the idea of random coil formation by molecular chains as a state of maximum entropy was introduced, as well as the role of water in the entropy balance of macromolecular systems, and the broad spectrum of intermo-lecular interactions was outlined. Knowledge of these aspects is an important prerequisite to understand the dynamic structure of proteins as a polypeptide chain with a unique spatial structure guaranteeing its specific function in the living cell.

The first X-ray crystallographic structure of a protein in 1959 (M. F. Perutz and J. Kendrew – Nobel Prize 1962) revealed a surprising amount of irregularities, and an apparent lack of symmetry as compared to the simple double-stranded DNA structure observed 5 years earlier. The process by which a polypeptide chain acquires its stable three-dimensional structure to achieve its specific biological activity, the so-called *native* state, i.e., the process of *protein folding*, remains a key problem of molecular biophysics (see Sect. 2.2.7). As a central dogma of molecular biology, all the necessary information for the protein to achieve the native secondary and tertiary structures is buried in the amino acid sequence per se.

Proteins are composed of α-L-amino acids which are covalently linked to each other by peptide bonds, eventually forming long chains (see Fig. 2.27). The peptide bond is formed by the carboxyl group of the α-C-atom of one, and the amino group of the following monomer. By a particular type of resonance interaction the corresponding C–N bond is shorter than a usual C–N bond, and resembles the resonance of the C=O double bond. As a result of this partial enol character of the peptide bond the group of six atoms – C_α-CONH-C_α – is planar. Therefore, the polypeptide appears as a sequence of planar structures that are twisted only by the rotating bonds with angles φ and ψ (see Fig. 2.27). Nevertheless, depending on the specificity of amino acids, steric constraints only allow distinct angles of rotation. The allowed angles φ and ψ determined by the particular van der Waals distances are usually plotted in the so-called *Ramachandran diagram*.

The linear sequence of amino acids as demonstrated in Fig. 2.27 is called the *primary structure* of the protein. The symbols **R** in this figure stand for the side chains, responsible for the specific properties of the 20 naturally occurring amino acids. By interactions of various parts of these chains, a three-dimensional arrangement is established – the *secondary structure* of the molecule.
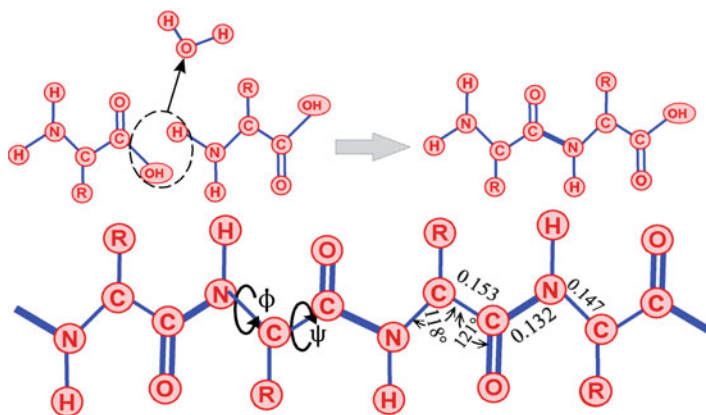

Fig. 2.27 *Above*: formation of the peptide bond between the amino acids, as a reaction of condensation, releasing a molecule of water (see text). *Below*: parameters of a polypeptide chain: R – side chains characterizing the particular amino acids, φ and ψ – angles of rotating bonds

Depending on the parameters of angles φ and ψ, two principal types of secondary structures are possible: the α-helix and the β-pleated sheet. By repeating angles of φ and ψ the *β-pleated sheet* is formed. The stability of these sheets increases through parallel or antiparallel arrangements of several sheets that are connected noncovalently by intermolecular hydrogen bonds. The side chains **R** in this case are oriented perpendicular to the plane of the monomers. Mostly the folded elements themselves form a right-hand spiral.

The *α-helix* is a rigid arrangement of the polypeptide chain that is twisted by right-hand torsion of the angles $\varphi = -57°$ and $\psi = -47°$ (see Fig. 2.28). Therefore it is a right-handed spiral, where each amino acid corresponds to a 100° turn. Hence, there are 3.6 amino acids per turn. The side chains (**R**) are oriented perpendicular to the helical axis, avoiding steric interference with the polypeptide backbone. In contrast to the β-pleated sheet, the α-helix is stabilized not by *inter*molecular, but by *intra*molecular hydrogen bonds in such a way that the peptide N–H bond of the *n*-th residue is connected with the C=O group of the (*n*-4)-th residue.

In general there are also arrangements of other polypeptide helices possible. They are usually described in an n_m-nomenclature. In this case *n* means the number of residues per helical turn, and *m* is the number of atoms, including H, in the ring that is closed by the hydrogen bond. The α-helix as the most abundant type in this nomenclature is a 3.6_{13}-helix, corresponding to 3.6 amino acids per turn. Approximately 15–20% of all helical structures are so-called 3_{10}-helices, which also are right-handed containing however only three residues per turn. They are typically only three to five residues long compared with a mean of 10–12 residues for 3.6_{13}-helices. These 3_{10}-helices have been proposed to be intermediates in the folding/unfolding dynamics of α-helices. In principle, also a left-handed 4.4_{16}-, a so-called
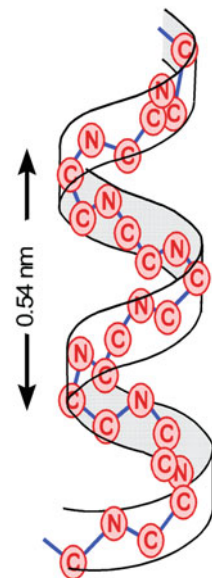
Fig. 2.28 The right-handed α-helix (for more clarity, only the backbone-forming atoms of the polypeptide chain are depicted)

0.54 nm

π-*helix* is possible. It is, however, unlikely to be observed in proteins because only glycine is likely to adopt this chirality.

The α-helix is a typical secondary structure of the membrane-spanning part of most integral membrane proteins (see also Figs. 2.39 and 3.32). Essentially, amino acids with hydrophobic side chains stabilize this secondary structure in the lipid bilayer. Sometimes these proteins span the membrane by several α-helices stabilizing each other (for more detail see Sect. 2.3.3).

One has always tried to find correlations between amino acid sequences and their secondary structures, to derive rules for structure prediction. In fact, the secondary structure of a protein is determined through the collective interactions between the elements of the molecule: the specific properties of a single amino acid alone do not by themselves exclusively determine which type of secondary structure will be seen. For example, the five negatively charged glutamic acids in five places in the chain A subunit of bovine pancreatic ribonuclease are organized in different secondary structures: GLU 2 and GLU 49 are in a coil configuration, GLU9 is in the α-helix, and GLU111 is in a β-sheet.

There are, however, specific sequences of amino acids that tend to form helical structures, while others form β-sheets or unstructured regions connecting different secondary structures. For example, collagen as a typical structural protein forms a triplet-helix as a tertiary structure. This is already encoded in the primary structure by a repeating sequence $(GLY-X-Y)_n$ where GLY stands for glycine, and X and Y resemble any other amino acids. Mostly these are proline and hydroxyproline. This repeated occurrence of glycine in every third position stabilizes the triplet-helix via hydrogen bonds. In some cases amino acids with the ability to form disulfide bridges provide a kind of starting point for the formation of super helices. In this case, as in the formation of α-helices, the process resembles a zip fastener. In a similar way myosin forms a left-hand double super-helix. The turn of this helix is predicted by the arrangement of amino acids in the corresponding α-helices of the secondary structure.

Further Reading

Voet and Voet 2011.

2.2.7 Protein Folding and Protein Dynamics

A central question in molecular biophysics is the mechanism of protein folding, i.e., the formation of its tertiary structure. Already in 1968 Cyrus Levinthal showed that protein folding cannot be considered simply as a stochastic process of reversibly connecting and disconnecting bridges between monomers to eventually find a structure of minimal free energy. This can easily be reduced to absurdity, considering the time which would be necessary for this process: Let us take a protein with 300 amino acids, each of which should have eight rotational positions. In this case $8^{300} = 10^{270}$ positions of the whole molecule are possible. Even in a case where, by

reason of steric hindrance the number of positions would be reduced, and even if the conformational changes were extremely fast, the time needed to arrive at the correct structure would be more than astronomically long. Even a polypeptide chain of only 150 residues could adopt 10^{68} possible conformations. Effectively folding into the native structure would require about 10^{52} years if we assume a mean time of 1 ns for each conformation. Actually, the folding time of a protein is between 0.1 and 1,000 s.

To solve this "*Levinthal paradox*," a minimal set of parameters must be found directing the process of protein folding to a suitable reaction coordinate. First the two principal pathways of protein synthesis must be considered: on one hand the *cotranslational* folding, where the native conformation of protein emerges directly during the biosynthesis on the ribosome, and on the other hand the *post-translational* folding, where the polypeptide chain starts to fold only after its full synthesis by the ribosome. Many experiments in vivo and on model systems have demonstrated that almost all proteins fold cotranslationally. The possibility of post-translational folding is supported by denaturation–renaturation experiments because many denatured small globular proteins restore their initial state quickly after removal of the denaturing agent. However, refolding of an unfolded protein probably is not an optimal model to study the protein folding in vivo, because the protein is not completely unfolded even under denaturing conditions.

Furthermore, it should be considered that proteins fold in different compartments of the cell. Some of them fold in the cytosol near the ribosomes, others are transported to different cellular locations, such as for example to the lumen of the endoplasmic reticulum across membranes. The cell compartments where proteins do fold are either aqueous or even hydrophobic such as the interior of the membrane. The targeting to these compartments is directed by the amino-terminal signal sequences.

In the last decades a number of fast spectroscopic techniques with a dramatically improved time resolution have been developed, such as photochemical triggering, temperature and pressure jump, ultrarapid mixing methods, etc. They allowed insights into the fast processes of protein folding. The first observable event in the folding pathway of a number of proteins, lasting a few milliseconds, is a collapse of the flexible disordered polypeptide chain into a partly organized globular state, which is called the *molten globule*. In this structure, hydrophilic parts of the chain are exposed externally to the aqueous environment, whereas other parts build hydrophobic centers inside the globule. Hence, a kind of microphase separation occurs. The generic collapse increases the packing density, and results in an extensive amount of entropy reduction of the entire molecule. Nevertheless, this is a typical entropy-driven process because the entropy changes of hydration of polar and nonpolar groups are both positive. The resulting conformational entropy loss caused by backbone packing, as well as side-chain interaction will be compensated by an increase in entropy of the surrounding water molecules (see Fig. 2.19).

To solve the Levinthal paradox, an energy landscape was defined for the system of folding as a mapping of the chain conformation to its internal energy, along with particular rules defining what transient intermediates are accessible from a given configuration, and how the protein can switch between them. Since the native

structure of a protein is the lowest in energy, one should expect the shape of the landscape for a protein to have a funnel-like topography, at least in the vicinity of the native structure. The extremely high degrees of freedom of a polypeptide chain make the funnel multidimensional. It was proposed to reduce this multidimensional picture into one with few relevant coordinates (see Fig. 2.29).

In fact, protein folding is a hierarchic and multiphasic process consisting of a number of particular reactions with quite different time constants. There is no singular way through the multidimensional energy funnel, but rather a large number of bifurcations and bypasses. This is an explanation for differences in the time course to eventually acquiring a native protein.

Various cellular mechanisms have been found to ensure a faster and proper protein folding. In the living cell this process is mostly assisted by the so-called *chaperones*. This is an ancient French word, standing for an older woman accompanying a young girl to guarantee her virtue. Chaperones, such as for example heat shock, or chaperonin proteins facilitate protein folding, and recognize and repair misfolded proteins. By reversible binding to unfolded polypeptides and separating them from bulk cytosol they dramatically accelerate folding and prevent misfolding and precipitation. Folding of about one third of all proteins in living cells is coordinated by such cofactors.

Because chaperones simply act by binding reversibly to specific sequences of the polypeptide, not modifying their chemical structure, it is not correct to use the term "structure-forming enzymes" for them. Such are for example the enzymes protein-disulfide-isomerases and poly-*cis/trans*-isomerases which also promote the process of folding. In contrast to chaperones, they modify the polypeptides directly. They are necessary, because about 93% of the peptide bonds are primarily synthesized in *trans*-form. To generate a native protein, a *trans–cis* transformation is necessary, which generally proceeds quite slowly but can be significantly accelerated by these enzymes.

Most investigations into protein folding are performed by in vitro studies. As already mentioned above, the process of refolding of denatured proteins is not directly comparable with in vivo protein folding, because a number of structural elements remain preserved even after denaturation. A second difference between studies in vitro versus such in vivo are the specific conditions in the cell. Most in vitro studies are performed in buffered solution with $<1\%$ protein concentration. In contrast, concentrations of these solutes in the living cell can reach hundreds of grams per liter. This crowding environment can have a significant impact on the stability and the native structure of a protein. Hence, it can significantly change the energy landscape of protein folding.

The difference of the free molar Gibbs energy (ΔG) between the native versus the denatured state of a protein, as depicted for example in Fig. 2.29, is sometimes below 50 kJ mol^{-1}. Considering that about 100 noncovalent interactions between amino acids of neighboring chains must be disrupted during denaturation, their interaction energies must be lower than the energy of thermal motion ($RT = 2.5$ kJ mol^{-1} for $T = 300$ K). The stability of the protein structure, therefore, can be explained only as a phenomenon of cooperativity.
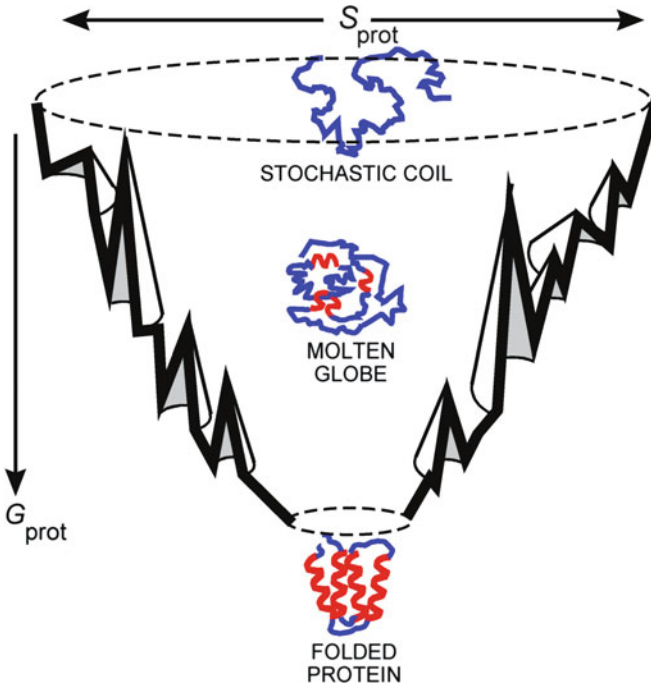
Fig. 2.29 Three steps of protein folding on a schematic energy landscape which has an overall funnel topography. The funnel is a three-dimensional projection of a many-dimensional space corresponding to the intricate connectivities between configurational states. G_{prot} the free energy of the conformational state, S_{prot} a measure of the configurational entropy of the protein (Corresponding to Plotkin and Onuchic 2002)

Conversely, the protein as a functional unit in the biological system not only needs sufficient structural stability, but must rather be considered as a highly dynamic structure. Obviously, in native proteins small conformational transitions in the form of stochastic fluctuations are the precondition for their function as enzymes or transport proteins. Further modifications may be controlled by ligands, by changes of the local ionic conditions, by temperature, or in the case of membrane proteins by the electric fields. In particular, the electric membrane voltage regulates ion channels, causing the action potentials in neurons (see Sect. 3.4.4).

These fluctuations can be expressed as the average square of displacement of various parameters such as the molecular internal energy ($<\Delta U^2>$), or volume ($<\Delta V^2>$) or even as amplitude of motion of an individual part of the molecule ($<\Delta x^2>$). They are functions of the energy of thermal noise (kT), the molecular mass (m), the molecular volume (V), the isochoric heat capacity (C_V), as well as the isothermic volume compressibility (β_T):

$$\langle \Delta U^2 \rangle = kT^2 m C_V \tag{2.65}$$

$$\langle \Delta V^2 \rangle = kTV\beta_T \tag{2.66}$$

Let us consider a globular protein with a molecular mass of M = 25,000, the mass (m) of an individual molecule therefore is:

$$m = \frac{M}{N} = 4.15 \cdot 10^{-20}\text{g} = 4.15 \cdot 10^{-23}\text{kg}$$

Assuming further the following case: $V = 3.2 \cdot 10^{-26}$ m^3, $C_V = 1.34$ kJ kg^{-1}, $T = 298$ K, as well as $\beta_T = 2 \cdot 10^{-11}$ Pa^{-1}, from Eq. 2.65 results $\Delta U = 2.61 \cdot 10^{-19}$ J per molecule. Multiplying this by the Avogadro number (N $= 6.022 \cdot 10^{23}$ mol^{-1}), a molar energy of 157 kJ mol^{-1} results for the case that all molecules would fluctuate synchronously. This is already close to the energy that would be sufficient for the denaturation of the protein. The deviations in the volume of this molecule are given by Eq. 2.66 for: $\Delta V = 5.1 \cdot 10^{-29}$ m^3, i.e., about 0.16% of its total volume.

This rather simple approach, however, does not reflect the true detail of the functional dynamics of the protein. Recently, methods have become available that allow us to determine fluctuations of selected parts of the molecule. For example, elastic incoherent neutron scattering (EINS), provides information about the motion of individual nuclei, especially by replacing selected hydrogen atoms by deuterium. These investigations show that the mean-square amplitudes of motion ($<\Delta x^2>$) in particular segments of the molecule, as for example the active center of myoglobin, are smaller than those of the rest of the protein. These parts of lower motion can be considered as an aperiodic solid, whereas the outside of the protein has semiliquid-like properties.

It is a question of biological optimization: why do proteins not have a higher degree of stability? In fact, proteins must have enough structural stability for maintenance of their specific three-dimensional conformations which is required for their function. But this stability must not prevent rapid and precise alterations of protein subdomains as required for function. Furthermore, a protein with a short half-life, especially, if it is needed at high concentrations, must be expeditiously synthesized demanding large energy costs. Conversely, a protein with an abnormally long half-life may be difficult to remove from the cell when its activity is no longer required.

Further Reading

Chen et al. 2008; Frauenfelder and McMahon 1998; Kloss et al. 2008; Muñoz 2007; Pain 2000; Plotkin and Onuchic 2002.

2.2.8 Ampholytes in Solution, the Acid–Base Equilibrium

We already noted that charges in biomolecules significantly influence the formation of their molecular and supramolecular structures. This aspect will be emphasized in

the next section when we discuss properties of interfaces and biological membranes. Furthermore, in Sects. 3.2.4 and 3.2.5 we will discuss the role of fixed charges in the cytoplasm to explain the occurrence of Donnan potentials. These aspects demand knowledge about the pH dependence of these sorts of charges.

An acid, per definition, is a proton donor, i.e., a compound that is dissociating hydrogen ions from particular groups. A base, conversely, is a proton acceptor and thus is able to associate them. According to this definition, water is a weak acid, dissociating with the following scheme:

$$H_2O \rightleftharpoons H^+ + OH^-$$
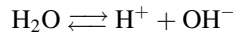
This equilibrium leads to the definition of the ionic product of water and, finally, to the pH-value:

$$pH = -\log c_H \tag{2.67}$$

Strictly of course, not the concentration (c_H), but rather the activity of the protons must be considered but in view of the extremely small concentration of these protons, this makes no difference.

In the same way, the dissociation of any other acid can be written:

$$HA \rightleftharpoons H^+ + A^-$$
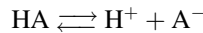
The equilibrium constant of this reaction may be written as:

$$K = \frac{c_A c_H}{c_{HA}} \tag{2.68}$$

In the same way, as in the case of pH, it is possible to define a pK-value:

$$pK = -\log K = -\left(\log c_H + \log \frac{c_A}{c_{HA}}\right) \tag{2.69}$$

or using Eq. 2.67:

$$pK = pH - \log \frac{c_A}{c_{HA}} \tag{2.70}$$

This is the *Henderson–Hasselbalch equation* which makes it possible to understand the buffer properties of various substances and their titration curves.

Substances that contain acid, as well as basic groups, are called *Ampholytes*; the first example we will mention is the amino acids. The buffer capacity, as well as the dissociation property of these substances can be investigated by titration experiments. This furthermore allows us to calculate the dynamics of molecular charges of these molecules. In Fig. 2.30 the charge of glycine is demonstrated as a function of the pH
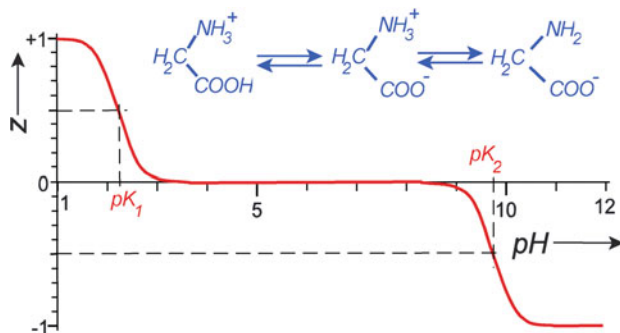
Fig. 2.30 The average number of charges (z) of glycine plotted against pH in the solution (pK$_1$ = 2.35, pK$_2$ = 9.78)

in the solution. For this, the mean number of elementary charges (z) per molecule is used. In the preceding sections we learned that dissociation and association are stochastic processes. Hence these amounts of z can be considered as an average in time or as a mean value of a large number of molecules.

Glycine is a cation at extremely low pH, and on the contrary an anion at extremely high pH. In the pH region in between, both groups of this molecule are dissociated. In this case the term zwitterion is used. This expression comes from German "Zwitter" or hermaphrodite, and means a molecule that is electrically neutral but carries positive as well as negative charges. In this case it has a considerable dipole moment and a maximal number of charges. This curve allows us to determine two pK-values which always correspond to the pH of the medium, where 50% of the charges are dissociated (in Fig. 2.30 corresponding to $z = +1/2$, and $z = -1/2$), as well as an isoelectric point which is positioned exactly in the middle between these two pH-values.

The pH-dependent charge of ampholytes of course also strongly influences the structure of the surrounding water. This is the reason why the effective volume of these molecules was often modified by the pH of the medium (see Sect. 2.2.3, Fig. 2.21).

Ampholytes are typical constituents of buffer solutions. Buffers are substances that are able to bind protons, or lose them, stabilizing in this way the pH of the solution. A buffer capacity can be defined, which corresponds to the slope of the titration curve. It is maximal near the pK-values.

Whereas the Henderson–Hasselbalch equation can easily be applied to binary ampholytes with two pK-values at considerable distance from each other, this becomes impossible in the case of polyampholytes, such as proteins with a large amount of different amino acids. This is illustrated in Fig. 2.31, showing the titration curve of hemoglobin. This curve can be considered as a sum of many sigmoidal curves with different properties. Near the isoelectric point, this curve can be fitted by the simple function:
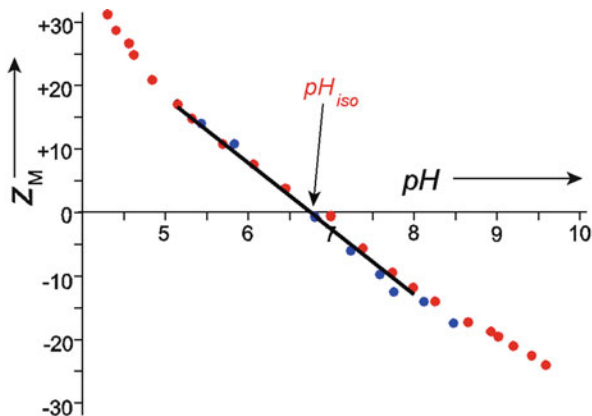
Fig. 2.31 The mean number of charges of horse hemoglobin as a function of the pH (*blue points* – oxidized, *red points* – deoxidized hemoglobin). The line between pH 5 and pH 8 corresponds to Eq. 2.71 with a buffer capacity of $z_{M0} = 10.2$ val mol^{-1}, and an isoelectric point pH$_{iso} = 6.68$ (According to measurements of German and Wyman 1937)

$$z_M = -z_{M0}(\text{pH} - \text{pH}_{iso}) \tag{2.71}$$

(z_{M0} – buffer capacity, pH$_{iso}$ – isoelectric point)

The titration curve of course indicates only the effective charges of molecules as a function of the pH in the solution. In the case of polymer molecules, the interaction of their monomers must be considered, depending on their mutual distance in the molecule. Furthermore, near the place of dissociation a local pH may exist, differing from that of the bulk solution. In this case, the term *intrinsic* pK is used, meaning the true pK of specific dissociable groups in a polymer, considering the real local proton activity. This, in fact, is possible to measure. Using nuclear magnetic spin resonance (NMR) for example, the dissociation of specific carboxyl groups can be studied, traced by the isotope ^{13}C. Investigations of polypeptides indicated that considerable differences may exist between the effective and the intrinsic pK of some groups even if the influences of neighboring amino acids are just slow. The local pH near functional groups in a protein can deviate significantly from that of the bulk solution. This is caused by fixed charges and therefore can be predicted by local electric potentials in the sense of the Boltzmann- or Nernst-equations (Eq. 3.112).

The different content of amino acids with acidic and basic residues in proteins is used to separate them by isoelectric focusing. For this a gel, composed of poly-acrylamide, starch, or agarose is used with an immobilized pH gradient. In this gradient an electrophoresis is performed. If the protein is in a pH region below its isoelectric point, where it is positively charged it migrates towards the cathode. Arriving in this way in regions of higher pH value, the charge will decrease until the protein reaches its isoelectric point where the migration ceases. Isoelectric focusing can resolve proteins that differ in isoelectric points by as little as 0.01pH-numbers.

In the case of two-dimensional gel electrophoresis, proteins are first separated in one dimension by isoelectric focusing in a pH-gradient and subsequently in a second dimension by their electrophoretic mobility at constant pH, separating them according to their molecular weight or the length of their polypeptide chain.

Further Reading

On electrostatic properties of proteins: Matthew 1985; Pethig and Kell 1987.

2.3   Interfacial Phenomena and Membranes

After explaining various kinds of molecular interactions in the previous section, we will now explore the next level of biological organization, considering self-assembly and molecular dynamics of supramolecular structures, in particular of membranes.

Biological membranes do not only surround the living cells but lead also to separation of numerous intracellular compartments. It is a well-organized structure fulfilling a broad spectrum of physiological functions:

– As a surface, it forms a dynamic matrix for enzymatic reactions, receptor processes, and immunological recognition;
– As a barrier of diffusion it controls the ionic composition of the compartments by highly specific transporters and channels;
– As an electrically isolating leaflet it contains various passive and active electric devices, controlling membrane potential as well as membrane near electrodynamic conditions;
– As a mechanical structure it maintains the integrity of the compartments, and is a determinant of cell shape and cell movement as well as the displacement of organelles.

We will discuss the underlying biophysical mechanisms of membrane functions in several sections of this book in detail. Here, we consider its molecular structure and dynamics as well as the processes of self-organization. In some cases I cannot avoid mentioning topics that will only be explained in detail in later sections.

2.3.1   Surface and Interfacial Tensions

Each molecule of a homogeneous isotropic phase interacts energetically with all of its surrounding neighbors. If these forces are averaged for a sufficiently long period of time, all the moments of interaction compensate each other. Considering however a molecule that is not in the bulk phase, but at the phase boundary, or at the liquid–air surface, the situation is quite different (Fig. 2.32). In this case these forces of interaction with molecules in one phase differ to those in the other phase. A molecule of a liquid for example at the liquid–gas surface senses a stronger

Fig. 2.32 Interaction forces
of molecules at the air–water
surface (S), and in the bulk
solution (B)
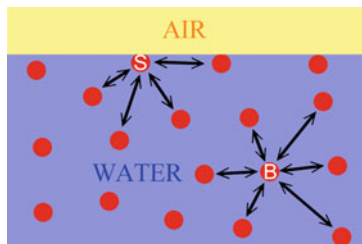
attraction to its neighbor molecules in the liquid than to molecules of the gas phase.
Thus, the energy of the surface molecules is higher than that of the molecules in the
bulk phase. This is the reason why water droplets in air spontaneously form a sphere
as the shape of minimal relative surface and therefore minimal surface energy.

To increase the surface of a liquid against air or against a gas that is saturated by
the vapor of this liquid, means to increase the number of molecules with a higher
level of energy. The energy that is required to enlarge the surface of a liquid phase
in this way by 1 m^2 is called the *specific surface energy*, which from the physical
point of view, is the same as the *surface tension* γ (J·m^{-2} = N·m·m^{-2} = N·m^{-1}).

Water as a polar liquid with strong intermolecular interactions (see Sect. 2.2.2)
has a high surface tension which decreases with increasing temperature. At 25°C it
amounts to 0.0728 N m^{-1}. The surface tension of organic liquids, for example, of
benzol is only 0.0282 N m^{-1}, and of ethanol only 0.0223 N m^{-1} at the same
temperature.

For biological reactions the surface tension at the water–air interface is interest-
ing from two aspects: On one hand it is important for organisms living directly at
the surface of a lake or the sea. These are microorganisms and small algae, forming
the so-called *neuston*, as well as various higher plants, and finally various insects
using hydrophobic properties of their legs to run on the water surface. On the other
hand surface tension plays an important role in the biomechanics of the lung, where
a highly solvated surface is in contact with the air.

To explain this situation, the behavior of bubbles or vesicles formed by a liquid
should be considered in detail: In the same way as the surface tension tends to
decrease the surface of a droplet, leading to a spherical shape, it tends to decrease
the volume of a gas bubble, generating an internal pressure (Fig. 2.33a). The
pressure difference $\Delta p = p_i - p_e$ can be calculated by the Laplace equation:

$$\Delta p = \frac{2\gamma}{r} \tag{2.72}$$

This means that the internal pressure of a vesicle is proportional to the surface
tension of the liquid (γ) which forms it, and inversely proportional to its radius (r).

In most cases it is not the *surface tension* liquid–air, but rather the *interfacial
tensions* between two liquid phases that are important, or, in other words, the
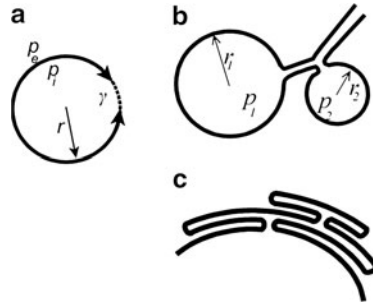
Fig. 2.33 (**a**) A vesicle with a surface tension γ tends to decrease the internal volume and generates an internal pressure $\Delta p = p_i - p_e$; (**b**) two vesicles of different size, built at the same interface generate different internal pressures which are inversely proportional to their radius. In the case of interconnection, the smaller vesicle should collapse in favor of the larger one. (**c**) The change in the surface of the surfactant layer by folding

specific energy of interfacial molecules. In the ideal case the interfacial tension of two liquids is equal to the difference between their surface tensions.

Earlier, these considerations sometimes uncritically were transformed to understand various cell physiological properties such as phagocytosis, exocytosis, cell division, amoeboid movement, cell shape, etc. Ludwig Rhumbler for example already in 1898 demonstrated that a drop of chloroform in water could "phagocytose" a thin, shellac-covered glass fiber in a manner similar to that of an amoeba, ingesting a filamentous alga. D'Arcy Thompson, in his classic book "On Growth and Form," first published in 1917, discussed in detail the problem of cell shape in relation to water droplets or even cellular ensembles in relation to the structure of foam.

Those comparisons, however, as visually impressive as they may be, are misleading in explaining cellular processes. In the case of a liquid droplet, surface molecules can always be drawn back into the bulk, or extruded again in the surface layer very quickly. This process is reversible; the mechanically deformed drop will re-assume its spherical shape as soon as the external mechanical disturbance is discontinued. In contrast, a living cell is surrounded by a membrane. An increase of the membrane area by inclination of additional molecules from the cytoplasm is a process which is much slower.

Conversely, the problem of surface tension is crucial, to understand the function of the alveoli in the mammalian lung. These are spherical outcroppings of the respiratory bronchioles covered by an epithelial layer, the diameter of them periodically varying between 0.05 and 0.1 mm. They are responsible for gas exchange with the blood.

The challenge of the system is to ensure their periodic change of size, and the stability of coexistence of vesicles of different diameter, connected to each other. As indicated in Fig. 2.33b, according to Eq. 2.72, the pressure in the smaller vesicles would be higher than in the larger vesicles. This would lead to a collapse of smaller vesicles in favor of the larger ones if they were connected to each other.

To avoid this, the surface tension of the water–air interface in the alveoli of the lung is strongly reduced by a surface-active material known as *pulmonary surfactant*, permanently secreted by Type II cells of the alveolar epithelium. This is a complex mixture of lipids and proteins that forms a monolayer at the alveolar liquid–water interface. It decreases the surface tension water–air from approximately 0.007 N m^{-1} to nearly zero. The periodic change of the alveolar volume and their surface during inhalation requires the property of this layer to be expanded quickly and reversibly. For this, a reservoir was formed, reversibly folding and unfolding these layers (Fig. 2.33c). Despite intensive research in this field, this process is still unclear in many aspects.

Further Reading

Historical papers: Rhumbler 1898; Thompson 1966. For the surfactant of the lung: Lalchev et al. 2008; Veldhuizen and Haagsman 2000.

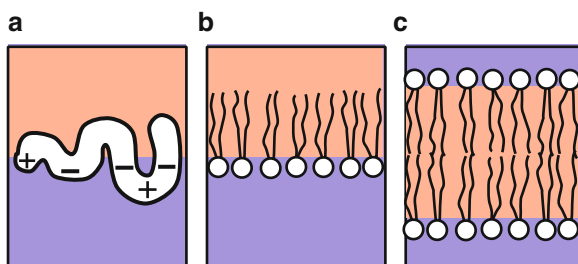## 2.3.2   Orientation of Molecules at Phase Boundaries; Self-Assembly of Membranes

The spontaneous orientations of molecules at phase boundaries are of particular interest for questions of self-organization and stability of biological membranes. Let us first consider the behavior of molecules at the boundary between two phases, one being hydrophillic (water), the other hydrophobic (for example oil). A protein tends to orient its polar groups as much as possible toward the aqueous phase, and its nonpolar, i.e., hydrophobic groups toward the oil (Fig. 2.35a). This leads to an orientation of the molecule at the phase boundary. If the polar groups of the molecule are distributed homogeneously over the whole molecule, then it could become unfolded by orientation in such a boundary. The protein, now being a fibril will spread over this interface. Some of these proteins form amphipathic helices or even stretches of α-helix, one side of which is polar (charged) and the other hydrophobic. In this case it orients parallel to the boundary layer.

The most important constituents of biological membranes are phospholipids. This is a group of compounds that consist of two fatty acids and a phosphate group, all bound to a polyhydric alcohol such as glycerol. The phosphate group is usually esterified with a nitrogen-containing alcohol. Structure and formula of a simple phospholipid is depicted schematically in Fig. 2.34. Two long hydrophobic fatty acid chains and a single phosphate group esterified with an ethanolamine residue are linked to a central glycerol molecule. The phosphate and the amino group of the ethanolamine residue represent the hydrophilic part of the molecule.

The large variety of biological lipids is realized by the diversity of the fatty acids as well as by their different head groups. In the case of phosphatidylcholine, the head group is polar but does not carry net charges at physiological conditions. The negative charge of the phosphate is compensated by a positive charge of the amino group. Phosphatidylserine, in contrast, contains an additional negative charge,
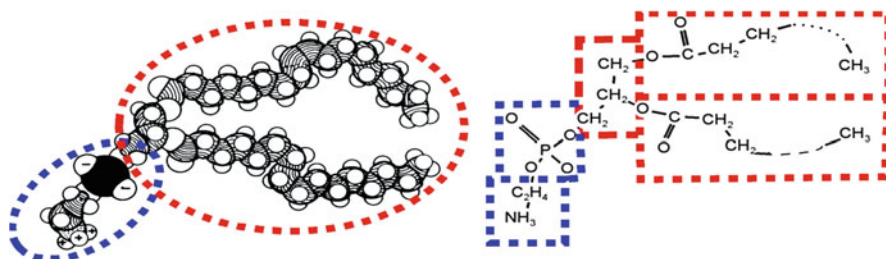
Fig. 2.34 Model and formula of a simple phospholipid (phosphatidylethanol amine with olein- and palmitin acid-rests). The hydrophilic part of the molecule is surrounded with *blue*, the hydrophobic with *red lines* (after Haggis 1965 redrawn)


Fig. 2.35 Arrangement of molecules in phase boundaries (*red area* – hydrophobic phase, *blue area* – hydrophilic phase): (**a**) orientation of a protein with hydrophobic and hydrophilic regions; (**b**) orientation of a monolayer of phospholipids (this corresponds also to the lipid layer at a water–air surface); (**c**) a lipid bilayer in a homogeneous aqueous phase. In the region of the fatty-acid chains, oriented against each other, a hydrophobic microphase is formed

namely that of a carboxyl group. This is the reason why it significantly influences the surface charges of biological membranes.

Phospholipids strongly orient themselves at interfaces. In this case the hydrophobic chains of fatty acids are directed parallel to each other and perpendicular to the interface (Fig. 2.35b). In this way, monomolecular layers, so called *monolayers*, are formed. The same occurs at the surface of an aqueous solution, i.e., at the water–air interface. Phospholipids that are oriented in such a way can be considered as a kind of two-dimensional phase.

Depending on the surface concentration of these molecules, it resembles gaseous, liquid, or solid states. Like in three-dimensional phases, phase diagrams as a function of pressure versus concentration, for example intermolecular distances, can be obtained. For this, a so-called *Langmuir trough* is used (Fig. 2.36). It consists of a temperature-controlled flat rectangular trough with a mobile barrier, as a part of a film balance. Phase diagrams can be obtained, measuring the resulting force resp. the lateral pressure (p' as force per length of the barrier) by shifting this barrier, i.e., increasing the lateral concentration of the molecules in the layer.

Fig. 2.36 Langmuir trough and a corresponding phase diagram of a layer of surfactant molecules
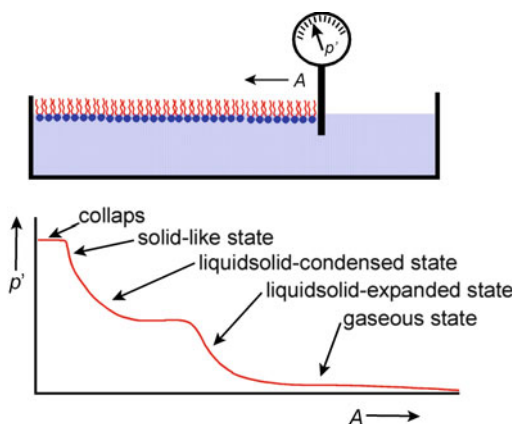
Whereas at low lateral concentration the molecules like in a gaseous state are moving more or less freely, they are always touching one another being concentrated to the liquid state. Depending on the degree of hydration of their head groups, an expanded liquid state can be differentiated from a condensed one, which finally is transformed to a solid-like state, without hydration of the head groups. A collapse finally occurs if several lipid layers slip one upon the others, forming multilayers. The shape of these phase diagrams reflects the properties of the flexible hydrocarbon chains as well as that of the polar head groups of the molecules. These kinds of investigations allow important inferences to be drawn on the properties of biological membranes.

In the case of homogeneous aqueous phases, phospholipids reach a state of minimal surface tension if they orient their hydrophobic portions toward one another (Fig. 2.37a). There is a large diversity in lipid aggregates in aqueous solutions. Using special methods, for example sonication of lipid suspensions with ultrasound, it is possible to produce structures with lipid double layers (*bilayers*) and vesicles, containing inside aqueous solutions (*liposomes*). There are two types of liposomes: unilamellar liposomes covered only by a single bilayer (Fig. 2.37b), and multilamellar liposomes (Fig. 2.37c) containing numerous lipid bilayers, separated from each other by a thin water layer.

In this context it must be mentioned that surface tension at least partly, is the phenomenological expression of processes which on the molecular level are the result of the dynamics of hydration, i.e., of water structure and hydrophobic bonds. All the lipid structures mentioned here can be destroyed by extremely high pressure in the sense of the Le Chatelier principle, as discussed in Sect. 2.2.3.

It is also possible to prepare in aqueous phases planar lipid membranes (Fig. 2.35c). If such a membrane is formed, the fatty acid chains of the lipids have built a hydrophobic microphase. During the last decades ingenious methods have been developed to produce bimolecular membranes with high structural and chemical specificity. These are the so-called BLMs, which is an abbreviation for:

Fig. 2.37 Self-assembly of lipids in aqueous solutions: (**a**) micelle; (**b**) unilamellar liposome; (**c**) sector of a multilamellar liposome. The *blue areas* indicate local microphases of water
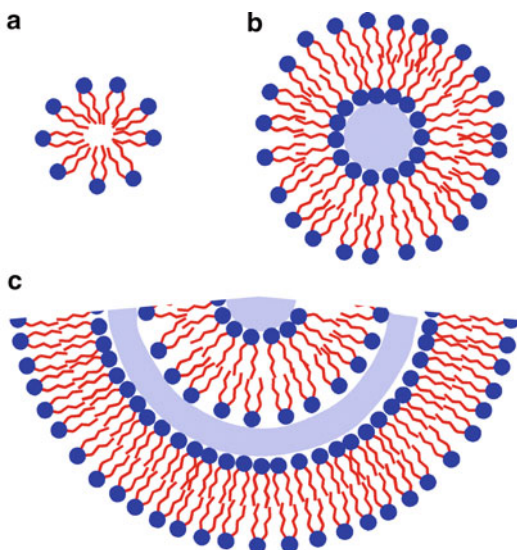
Fig. 2.38 BLM in a hole of a barrier between two aqueous phases
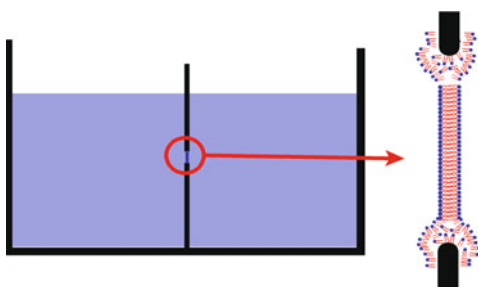
Bimolecular Lipid Membrane. By various techniques it is possible to expand such membranes in a small hole of a barrier, dividing two aqueous phases (Fig. 2.38). In this case various electrical and electrochemical measurements are possible. Usually, these membranes are at first multilamellar, showing interference of visible light. Later the excessive lipids concentrate on the edges of the hole, and the membranes become bilayers with a thickness of less than 10 nm. In this case no interference of visible light is possible; they become optically nondetectable. For this reason the abbreviation BLM is also interpreted as "Black Lipid Membrane." The real existence of a barrier in the hole in this case can be proved only by its electrical resistance.

These artificial membranes are very important tools to investigate properties of biological membranes and their constituents. It is possible to introduce specific molecules into these membranes, such as for example transport and signal proteins, and investigate their specific characteristics under various conditions. Their lipid

composition can be varied as well as the ionic composition of the solution on both sides of the membrane. Using specific techniques it is even possible to mimic the lipid asymmetry of biological membranes.

Further Reading

Butt et al. 2006; Hianik and Passechnik 1995.

### 2.3.3 The Molecular Structure of Biological Membranes

Measurements of electric conductivity of cell suspensions already indicated a long time ago that the cell must be surrounded by an electrically isolating layer with a specific capacity of about $10 \, mF \, m^{-2}$ (for more detail, see Sects. 2.3.6 and 3.5.3). It corresponds to a double layer of lipids with a thickness of ca. 8 nm. This led to a number of ideas and models of membrane composition, and its molecular structure. In 1972 Singer and Nicolson proposed a *fluid-mosaic model* of the cell membrane which in fact reflects a number of experimental properties. It considers a lipid bilayer as a two-dimensional solvent with integral membrane proteins, embedded in it, that are free to move. This model has been modified by recent experimental results. In fact the real concentration of proteins in the membrane is larger as suggested by Singer and Nicolson. Furthermore, many membrane compounds are not freely mobile, but rather connected to the cytoskeleton, and the plasma membrane of eukaryotic cells contains a variety of inhomogeneities such as protein and lipid domains.

The lipid composition of the inner and outer membrane leaflet of biological membranes is highly specific. This concerns the particularity of their fatty acids, as well as their polar head groups. There is also an asymmetric distribution of lipids between the inner and the outer layer, maintained by specific transport proteins, so-called *flippases*.

The fluidity of the membrane, i.e., the lateral movement of its constituents is strongly determined by the length of the aliphatic fatty acid chains, as well as by their degree of saturation (see also Sect. 3.6.3). The property of the polar head groups determines their mutual distance in the membrane surface, i.e., their density of packing.

The proteins are organized in the membrane according to their hydrophobic (apolar), and hydrophilic (polar) regions (Fig. 2.39). Most of the functional proteins, such as transporters, penetrate the hydrophobic core of the double layer of the membrane by helical parts. The number of membrane-spanning parts of a protein may range from one to more than 12. This is dictated by the dimension of the hydrophobic part of the proteins and has to be matched to the hydrophobic thickness of the lipid bilayer. In order to compensate for the mismatch, the lipid molecules closest to the protein stretch out or compress in order to cover the hydrophobic core of the protein (Fig. 2.39a). This leads to a perturbed region around the protein. Another possibility is an inclination of the helices in the

Fig. 2.39 Various kinds of protein interaction with the lipid double layer: (**a**, **b**) integral membrane proteins that span the bilayer; (**c**) protein, anchored via a lipid extended conformation; (**d**) amphiphilic protein partially penetrating the bilayer; (**e**) electrostatically bound protein
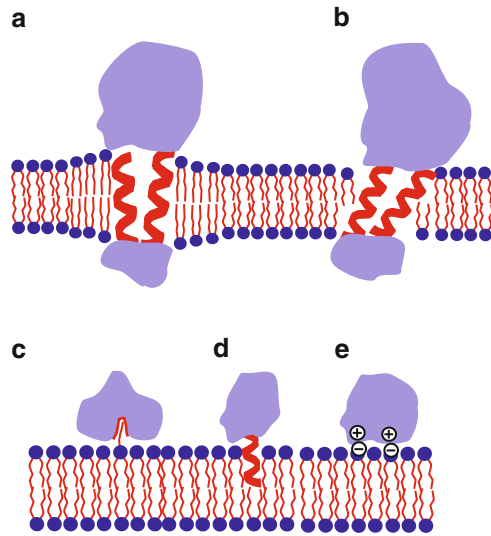
membrane (Fig. 2.39b). Nonpenetrating proteins can be fixed to the membrane in various ways (see Fig. 2.39c–e).

Many proteins are exposed to the external surface of the cell membrane as *glycoproteins*. Typically, they are located with their C-terminal end in the cytoplasm. Their N-terminal end, which is modified by various carbohydrates, extends several nanometers into the outer environment of the cell. These carbohydrates in some cases represent the receptor-sides of the proteins, and are the location of immunological reactions. Preferentially, at the end of the protein chains, monomers of the *N-acetylneuraminic acid* (also called: *sialic acid*) are located which carry a dissociable carboxyl group. In this way, these are the most important carriers of fixed negative surface charges of the cell (see Fig. 2.48). Depending on the type of cell, there are between 1 and 10 groups of sialic acid per 1 nm^2 membrane area. Hence, the glycoproteins form a loose external layer of the cell, the so-called *glycocalyx* or *surface coat*. The charges, located at the ends of these filaments, modify the thickness of this structure electrostatically (for detail, see Sect. 2.3.6).

The lateral distribution of membrane constituents according to the Singer–Nicolson model could be random only if all pairwise interaction energies are within the range of the thermal energy kT. Considering the diversity of lipid and protein species and the variety of intermolecular interactions such as hydrogen bonds, electrostatics, and hydrophobic effects, this condition is hardly fulfilled. In contrast, it should have been expected that regions of biased composition would exist.

Recent investigations indicate that membrane proteins may be organized in large functional complexes. In common with proteins, lipids also tend to group together by specific lipid–lipid and lipid–protein interactions. Some lipids form specific complexes with proteins. As indicated in Fig. 2.39a, at the rim of proteins and protein complexes, different thicknesses of the lipid bilayer may exist.

Lateral movement of proteins has been recently investigated by single-particle tracking experiments. For this gold-labeled molecules were moved in the membrane surface until a barrier was encountered. Some receptor proteins in this way show a lateral compartmentalization in 300–600-nm diameter domains. These results led to the *membrane-skeleton fence* model, including the possibility of successive movements (*hops*) to adjacent compartments.

From the truncation experiments it was estimated that the barriers to this lateral mobility were located 2–3 nm below the cytoplasmic leaflet. Obviously it consists of a network of cytoplasmatic proteins as part of the cytoskeleton with connections to the inner side of the membrane. It is best investigated in human erythrocytes, where it consists mostly of spectrin. This is a heterodimeric protein (molecular mass: 260 and 220 kDa) that forms rods of approximately 100 nm length, aggregating head-to-head, to form tetramers of double length. These spectrin rods are connected to each other by the globular actin molecules (molecular mass: 100 kDa). This leads to an *anchored-protein picket* model assuming that close to the cytoplasmic leaflet the actin membrane skeleton meshwork anchors various transmembrane proteins.

Another important class of membrane inhomogeneities is *caveolae*. These are flask-shaped membrane invaginations of about 60 nm diameter containing mainly *caveolin*, a protein that binds cholesterol. Caveolae have been implicated in processes like cholesterol transport and endocytosis.

Studies on lipid marker molecules indicated lipid microdomains in the exoplasmic leaflet, so-called *lipid rafts*. These are areas ranging from several tens of nanometers in biological membranes to almost a micrometer in model membranes that are enriched in cholesterol and lipids with saturated acyl chains, such as sphingolipids. A tight packing of these constituents results in a liquid-ordered phase, which separates the lipid rafts from the surrounding. The biological functions of lipid rafts range from membrane trafficking and sorting to a dynamic role in signal transduction.

The membrane therefore, can be considered as a two-dimensional compartmentalized structure, which is more mosaic-like than fluid.

Further Reading

Engelman 2005; Lommerse et al. 2004.

2.3.4   Mechanical Properties of Biological Membranes

The mechanical properties of biological membranes are very important to understand a number of physiological cell functions such as cell movement, cell division, volume alterations, vesiculation, membrane fusion, etc. In the same way, the mechanical properties of tissues, as well as the streaming properties of the blood are controlled, which we will discuss in Sect. 3.6.

It must be emphasized that parameters, for example viscosity, elasticity, strain, etc., in fact, are macrophysical quantities, which are defined for homogeneous phases and materials (see Sect. 3.6). Hence, they are not proper parameters for

biological membranes at the submicroscopic scale, and also not for highly organized supramolecular structures. Nevertheless, in some cases it may be convenient to use these parameters, but then they must be considered *effective parameters*, which we explained in the introduction to Sect. 2, and which we have already used in many other cases.

The effective membrane viscosity, for example, is a parameter that can be measured by monitoring the rotation or translocation of a marker molecule. This is possible through the use of special fluorescence methods or electron spin resonance techniques (ESR). Considering the membrane as a homogeneous phase, and the marker molecule as a macroscopic body with a particular shape, one can formally calculate the viscosity, applying the usual phenomenological equations for movement. This measurement is reproducible, and the parameter is exactly defined by the experimental method, and by the applied theoretical approach. The problem which arises through the use of equations that are not adequate for the movement of a molecule in a heterogeneous phase is not relevant here, because it is included in the definition. It is an "effective" parameter, since it is defined by observing a particular effect. Conversely, this means that measuring these effective viscosities by different methods, one gets different parameters, but not a general viscosity according to the original physical definition.

This is the general difference between molecular and macroscopic types of measurement. In the case of homogeneous liquids, the resulting viscosity is independent of the method used to obtain it. In fact: a "viscosity" in the sense of macroscopic materials does not exist in supramolecular structures such as biological membranes. The same applies to the other parameters mentioned above.

Applying different markers, the parameters measured in this way depend on the regions where the marker was located, for example, at different locations in the membrane depth, i.e., at different positions in the fatty acid of a lipid, or even at different lateral positions of the membrane. These differences indicate the inhomogeneity perpendicular to the plane of the membrane, as well as the anisotropy of the mechanical properties of the biological membrane in general.

For some biomechanical questions, such as for example deformation of erythrocytes in small capillaries (Sect. 3.7.2), osmotic swelling of cells, etc., some general properties of the membrane are of interest, independently of its inhomogeneity.

In general, the cell membrane is easily deformable by shearing forces without increasing the area (Fig. 2.40c). In contrast to this it is nearly impossible to extend the area of the membrane (Fig. 2.40b). The stress (σ'), which must be applied, to extend the area A of a membrane by the amount ΔA can be calculated by

$$\sigma' = Y' \frac{\Delta A}{A} \tag{2.73}$$

Y' is the specific modulus of elasticity, a coefficient which is related to the thickness of the membrane. The symbol Y stands for *Young's modulus*, another name for the same parameter (see Sect. 3.6.3). For the membrane of erythrocytes a

Fig. 2.40 Possible kinds of
passive molecular
displacement and mechanical
deformations of biological
membranes: (**a**) translation
and rotation of membrane
components; (**b**) planar
extension of the membrane
area; (**c**) shear deformation
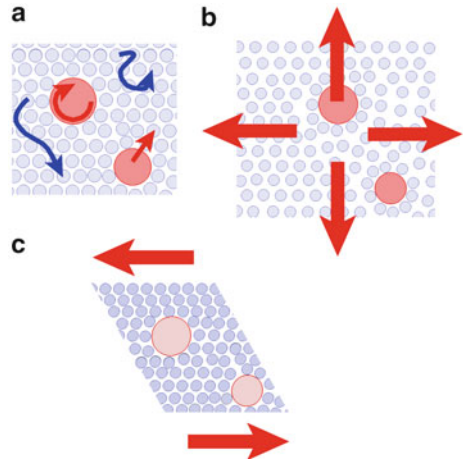
value of $Y' = 0.45$ Nm^{-1} was estimated, for lymphocytes 0.64 Nm^{-1}. To find a relationship between membrane properties and properties of macroscopic materials, the parameter Y' must be divided by the thickness of the membrane (d), which is approximately $8 \cdot 10^{-9}$ m. In this case, for erythrocytes it results in:

$$Y = Y'/d = 5.6 \cdot 10^7 \text{Nm}^{-2} = 56 \text{ MPa}$$

The membrane will rupture if it expands more than 1–2% of its area.

In Sect. 3.6.3 we will differentiate between two kinds of elastic materials, resembling steel, in contrast to rubber elasticity. Steel elasticity occurs in crystalline materials and is caused by modifications of the crystal structure. It is characterized by a large Young's modulus (for steel about $2 \cdot 10^5$ MPa) and a low expandability. In contrast, rubber elasticity is a typical behavior of macromolecular structures and based on the stretching of molecules, as illustrated in Fig. 2.11. It indicates a Young's modulus of only around several MPa.

In fact, the mechanical behavior of biological membranes already tends toward the values of steel-elastic materials. This is understandable considering that any increase of the membrane area leads to an increase of the distance between the lipid head groups, analogous to the ion atoms in steel. Properties of rubber elasticity are given only by the intrinsic proteins.

The membrane therefore resembles a material with a high degree of flexibility in plan, but with minimal ability for area extension (Fig. 2.40). Hence, it has properties very unlike a rubber sheet, which can easily be expanded on account of its thickness. Living cells, therefore, can swell only by spherulation, by unfolding of the membrane, or by inclusion of new membrane material. Each swelling of the cell, which extends the surface of a sphere, or increases the capacity of unfolding, immediately leads to cytolysis.

   As we will see later, using the elastic modulus, it is also possible to calculate forces that oppose the bending of a material (Sect. 3.6.3). Applying this to the biological membrane, again, the inhomogeneity of the membrane must be considered. According to the heterogeneity of the lipid composition and the differences in the distribution of the proteins, the bending resistance of the membrane differs from place to place. Recently, atomic force microscopy (AFM) has been used for such investigations. AFM consists of a microscale cantilever with a sharp tip at its end. Using this, periodic indents of the membrane at particular locations can be produced. Scanning the membrane in pixels approx. 70 nm apart, and by using a 20-nm tip radius cantilever deflected by 50-nm a relative pattern of regions with different stiffness can be obtained.

   In addition to these passive mechanical properties, biological membranes per se are bent. As illustrated in Fig. 2.41, this active bending is realized in different ways. One possibility is the bilayer couple mechanism. In this case bending occurs by inclusion of additional molecules in one of the membrane leaflets. Another possibility is the introduction of conically shaped molecules. A factor (f) characterizes this property. In the case of lipids, it relates the area of the head group (A_H) to the mean area (A_F) which is required for the region of fatty acids (Fig. 2.42). A_F can be
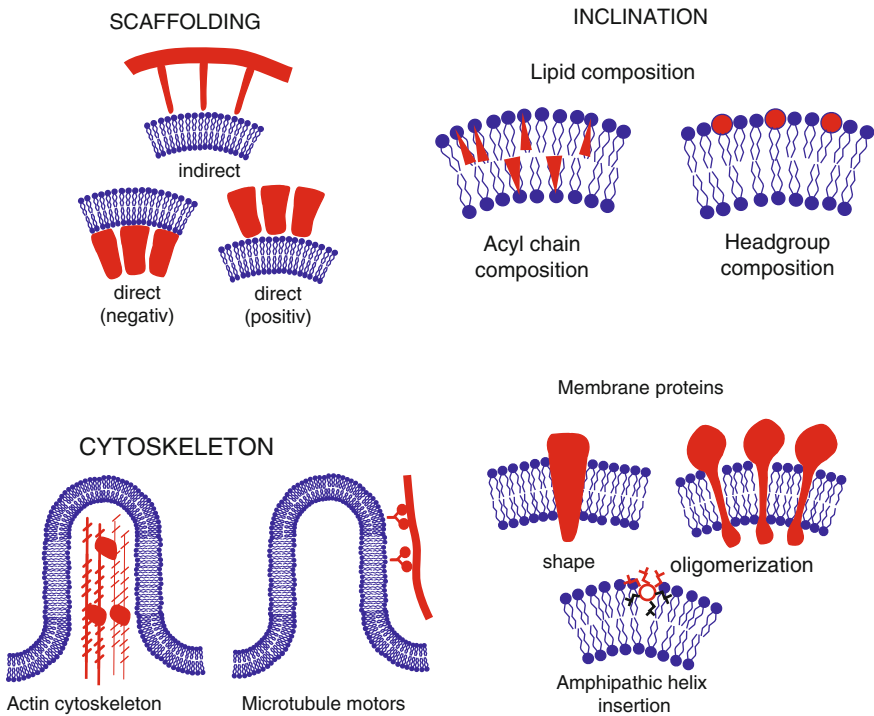
Fig. 2.41 Possible kinds of active membrane bending (*blue* – phospholipids, *red* – proteins) (Redrawn after McMahon and Gallop 2005)

Fig. 2.42 The effective
shape of a phospholipid
(explanations in text)
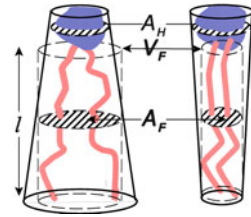
calculated by the mean length of the fatty acids (l) and the effective volume of the
region of fatty acids (V_F) using: $A_F = V_F/l$. The shape factor f, therefore, is defined
as

$$f = \frac{V_F}{lA_H} \tag{2.74}$$

The effective volume of the fatty acid chains (V_F) strongly depends on their
phase conditions, i.e., on their degree of order. Dynamic variations are possible as
well as phase transitions through temperature modification. Other parameters of the
molecule directly depend on their composition. So, for example, lysophospholipids,
containing only one single fatty acid chain, are to a large degree conical with a
dominant A_H area; in this case, $f < 1/3$. In contrast, phosphaticylethanolamine
containing usually a large amount of unsaturated fatty acids indicates f > 1, with
a dominant A_F. Mostly lipids are more or less cylindrical, with $1/2 < f < 1$. The
area of the head group is in fact a parameter which includes the mean distance to the
neighboring molecule, which can also be influenced by electrostatic interactions.

Figure 2.41 shows further possible types of active membrane bending. This
concerns not only inclusions of other molecules in the leaflet, but also bending by
external or internal scaffolding proteins. Also very important is the assembly and
disassembly of the cytoskeleton of the membrane. In many areas of high membrane
curvature, such as filopodia, pseudopodia, phagocytic cups, etc., actin filaments are
involved in the generation and remodeling. In some cases kinesin motors attached
to Golgi membranes are involved in processes of fenestration or generation of some
transport processes.

Further Reading

Devaux and Herrmann 2011; McLaughlin and Murray 2005; McMahon and Gallop
2005; Roduit et al. 2008; Schmid-Schönbein et al. 1986.

2.3.5   *Electrical Double Layers and Electrokinetic Phenomena*

An interface carrying fixed charges induces in its vicinity an electric field, and
accordingly, modifies the local ionic conditions, analogous to a point charge, as
described in Sect. 2.2.4. In the simplest case, a double layer occurs with the fixed

charges on one side, and the electrostatically attracted mobile counterions on the opposite side. This concept was originally put forward by Herrmann von Helmholtz. It only applies for extremely high concentrations of mobile ions in solution. In general, the thermal movement repels a great part of the ions from their positions near the fixed charges. This leads to the formation of a *diffuse double layer*, the electrical potential of which is declining exponentially with the distance from the surface.

Figure 2.43 illustrates the generalized model of Stern, which includes the Helmholtz double layer as well as its diffuse part. To avoid overloading of the picture, only the fixed charges (in this case proposed as negative), and the compensating mobile counterions are depicted. Of course, there are also mobile anions present, and there are a great number of both cations and anions that should be included to give a true picture. This picture indicates an enlarged concentration of cations near the surface for the case of negative surface charges, and the electroneutrality of the total system, including the bulk solution.

Despite the thermal noise, some of the mobile ions directly oppose the fixed charges, forming the so-called *Helmholtz layer*. Between the surface and the end of the Helmholtz layer, the potential is proposed to decline linearly from ψ_0 to ψ_H. Further, the potential declines according to the model of the diffuse double layer which was calculated by Gouy and Chapman.

The theory of the diffuse double layer again is based on the Poisson–Boltzmann equation, which we have already explained in context with the Debye–Hückel theory of ion clouds (Sect. 2.2.4.). This equation makes it possible to calculate the distribution of mobile charges in a given electric field:

Fig. 2.43 Schematic illustration of an electrical double layer. *Above*: fixed charges and mobile counterions near a surface; *beneath*: the function of the electrical potential, according to the model of Stern. The charge of the surface in this particular case is enforced additionally by a chemically adsorbed negatively charged ion. ψ_0 surface potential, ψ_H potential at the end of the Helmholtz layer, x_H thickness of the Helmholtz layer, $1/\kappa$ Debye–Hückel length as a measure of the effective thickness of the diffuse double layer
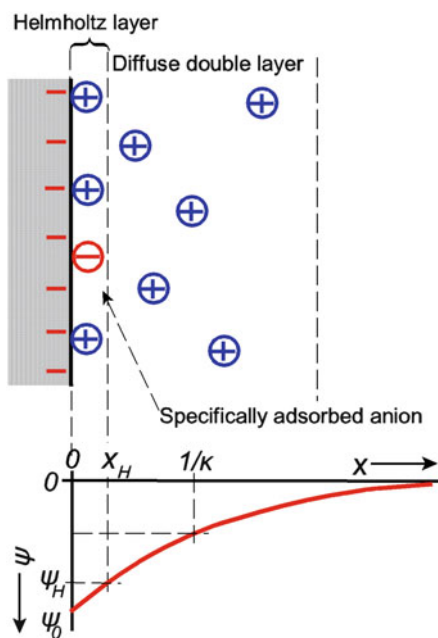
$$\nabla^2\psi = -\frac{F}{\varepsilon_0\varepsilon} \sum_{i=1}^{n} c_{io}z_i \, e^{-\frac{z_i e\psi}{kT}} \tag{2.75}$$

In this case the amount of fixed charges is compensated by the excess of charges in the whole double layer.

Applying this equation for one-one-valent electrolytes, like KCl or NaCl, it reduces to:

$$\nabla^2\psi = -\frac{Fc_0}{\varepsilon_0\varepsilon}\left(e^{-\frac{e\psi}{kT}} - e^{\frac{e\psi}{kT}}\right) = \frac{2Fc_0}{\varepsilon_0\varepsilon}\sinh\frac{e\psi}{kT} \tag{2.76}$$

In this case, a rearrangement was taken, using the function: $\sinh x = (e^x - e^{-x})/2$.

To solve the Poisson–Boltzmann equation, i.e., to calculate the function $\psi(x)$, one needs to define two initial conditions. For our case this means:

$$\psi(\infty) = 0$$
$$\psi(0) = \psi_H$$

whereas the end of the Helmholtz layer (ψ_H) is taken as the starting point of the function [$\psi(0)$]. The algebraic integration of this equation again, is possible only after linearization. For $x > x_H$ one gets the following function $\psi(x)$:

$$\psi(x) = \psi_H e^{-\kappa x} \tag{2.77}$$

In this equation, κ again is the Debye–Hückel parameter, which was already introduced in Sect. 2.2.4 (Eq. 2.55). The Debye–Hückel length ($1/\kappa$) is the distance, where the potential ψ_H is dropped by the factor e:

$$\frac{\psi(1/\kappa)}{\psi_H} = \frac{1}{e} = 0.368 \tag{2.78}$$

The amount of this parameter, as a function of the ionic strength (I) is given in Eqs. 2.55 and 2.57.

What does "linearization" of Eq. 2.76 mean, and what are the limitations of this simplification? In fact, the basis for such approaches is the expansion of these functions in series, using the following equations:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots \tag{2.79}$$

and:

$$\sinh x = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots \tag{2.80}$$

For the case:

$$\frac{ze\psi}{kT} << 1 \tag{2.81}$$

one can apply these equations, using only the first terms of the series:

$$e^{-\frac{ze\psi}{kT}} \approx 1 - \frac{ze\psi}{kT}, \quad \text{and}: \quad \sinh \frac{e\psi}{kT} \approx \frac{e\psi}{kT}. \tag{2.82}$$

Let us investigate the scope where this simplification is possible. For $T = 298$ K one can calculate:

$$z\psi \frac{e}{kT} = z\psi \frac{1.60218 \cdot 10^{-19}}{1.3807 \cdot 10^{-23} \cdot 298} = 38.94z\psi \tag{2.83}$$

Using, for example, a potential $\psi = 0.01$ V and a charge number $z = 1$, one gets: $e^{-0.3894} = 0.6774$, and according to Eq. 2.79: $1-0.3894 = 0.6106$. Correspondingly: $\sinh 0.3894 = 0.3993$. Both calculations, really, indicate some differences, which will become larger, if the electrical potential exceeds the value of 10 mV. This is important also for the application of this linearization in the previous Sect. 2.2.4., as well as in the following considerations.

Beside the calculation of the diffusion layer it is necessary to know the relation between the potentials ψ_0 and ψ_H. In general, the difference between these two parameters will become smaller, as the ionic strength of the solution becomes lower. In solutions of extremely low ionic strength, this difference can be neglected.

In fact, besides the screening of fixed charges by mobile ions, other processes must be considered, such as dipole orientations, interactions with water molecules or other processes of adsorption, based on van der Waals interactions. In this way, it is even possible that against electrostatic repulsion, co-ions were absorbed at the interface, increasing its charge density. In this case the Helmholtz potential (ψ_H) can even become larger than the surface potential (ψ_0).

We will discuss the relevance of electrical double layers for functions of biological membranes in detail in Sect. 2.3.6. The most important consequence are local ionic concentrations, including, local pH values (Fig. 2.44). This can be calculated according to the Nernst equilibrium (Eq. 3.112, Sect. 3.2.2). The relevance of these local conditions is understandable, considering the function of membrane-bound enzymes.

A number of electromechanical interactions are caused by these electrical double layers near charged surfaces which are summarized by the term: *electrokinetic phenomena*. On one hand, an externally applied electric field may induce mechanical effects, like movement or streaming; on the other hand, a mechanical

Fig. 2.44 Electrical potential (ψ), pH, as well as Na$^+$- and Cl$^-$-concentrations (c) as a function of distance (x), near a negatively charged surface in a diffuse double layer. The curves consider the following parameters: $\psi_0 = -50$ mV, pH$_\infty = 7.0$, $c_\infty = 10$ mM, $T = 298$ K (the index ∞ means: at $x \rightarrow \infty$). In this case the Debye–Hückel-length ($1/\kappa$) amounts to 3 nm (dissociation constants of salt and water are not taken into account)
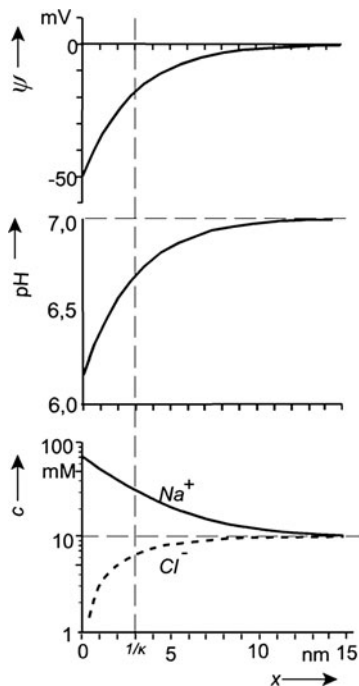
movement may lead to the occurrence of electrical potential differences. In Fig. 2.45 these phenomena are depicted schematically.

Electrophoresis means the movement of charged molecules, particles, or cells in an externally applied electric field. The displacement of particles and cells can be measured by microscopic observation. Their mobility (b) is defined as the relation between their velocity ($\mathbf{v}$) and the electric field strength ($\mathbf{E}$):

$$b = \frac{\mathbf{v}}{\mathbf{E}} \tag{2.84}$$

In an electric field the particles themselves move in the opposite direction to the ions of their double layers, caused by their opposite polarization. The movement of ions furthermore includes a movement of bound water. This leads to an increased frictional force, diminishing the resulting electrophoretic mobility.

The classical theory of cell electrophoresis is based on the model which was already proposed by Smoluchowski, considering the movement of large spherical particles. "Large" in this sense means that the diameter of the particle, and the corresponding radius of bending, must be large in relation to the Debye–Hückel parameter ($1/\kappa$). Smoluchowski postulated that a shearing plane exists, caused by the electrokinetically induced streaming. At this plane, the electroneutrality of the system is disturbed. The electrical potential at this distance from the surface is

Fig. 2.45 Schematic representation of electrokinetic phenomena of particles (*above*), and surfaces of tubes or capillaries (*beneath*); *left*: the occurrence of electrical potentials as a result of mechanical movement (*arrows*); *right*: the induction of movement (*above*) or streaming (*beneath*) as a result of an applied electric field (*arrows*)
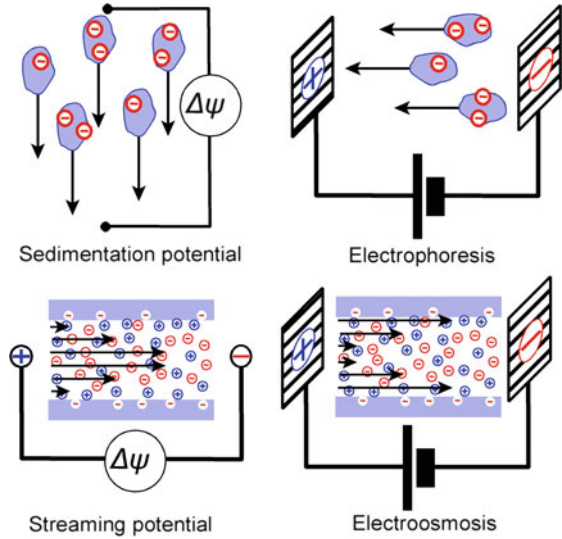
called *electrokinetic potential*, or ζ-*potential*. It can be calculated by the measured electrophoretic mobility (*b*) using the following equation:

$$\zeta = \frac{\eta V}{\varepsilon_0 \varepsilon \mathbf{E}} = \frac{\eta b}{\varepsilon_0 \varepsilon} \tag{2.85}$$

where η is the viscosity of the medium. Surprisingly, the radius (*r*) of the particle is not included in this equation. Sometimes, this equation contains an additional factor, considering stronger deviations from the shape of a sphere.

Of course, the above-introduced ζ-potential is not identical with the surface potential (ψ_0). With some approximation, however, it can be considered as equal to the Helmholtz potential (ψ_H) (see Fig. 2.43). In the same way, as discussed above, the ζ-potential increases, by constant surface charge density (σ_0), if the ionic strength in the solution is decreasing (Fig. 2.46).

It is possible to calculate the surface potential (ψ_0) and the surface charge density (σ_0) from the ζ-potential only with certain approximations. For this, the equation is used, derived from the Gouy–Chapman theory of electrical double layers:

$$\sigma_0 = \sqrt{8\varepsilon_0 \varepsilon RT} \sqrt{I} \sinh \frac{F\zeta}{2RT} \tag{2.86}$$

In Fig. 2.46 the dependence of the ζ-potential is depicted as a function of the ionic strength of a one-one-valent electrolyte, which in this case is equal to its concentration. In this case a surface charged density of $\sigma_0 = 0.014$ C m^{-2} was chosen which corresponds to that of a human erythrocyte.

Fig. 2.46 The ζ-potential (depicted as a negative value) as a function of the concentration of an one-one-valent electrolyte for a surface with a constant surface charge density ($\sigma_0 = 0.014$ C m^{-2}) according to Eq. 2.86
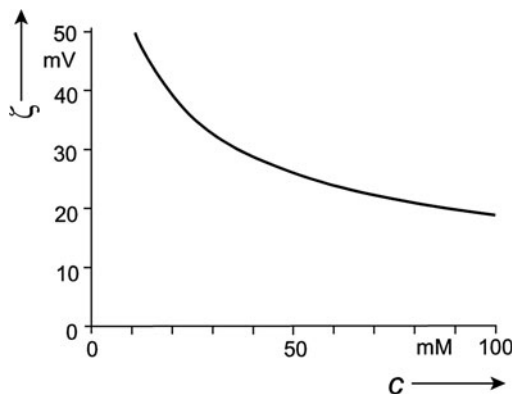
Equation 2.86 can be simplified expanding it as a series, corresponding to Eq. 2.80. If the Debye–Hückel parameter (κ) was used according to Eq. 2.55, we have

$$\sigma_0 = \zeta\kappa\varepsilon\varepsilon_0 \tag{2.87}$$

which, when combined with Eq. 2.85, results in

$$\sigma_0 = \eta\kappa b \tag{2.88}$$

Introducing this kind of linearization we found that this simplification can be accepted as a reasonable approximation only for potentials ≤ 10 mV. As indicated, however, in Fig. 2.46, this limit is exceeded for most cases of biological membranes.

Furthermore, all equations were derived considering the fixed charges as located on a planar boundary. The thickness of the Helmholtz layer, which amounts only to a few tenths of a nanometer explains what the word "planar" means. The roughness in the order of a molecular diameter is always enough to exceed this assumption.

To measure the electrophoresis a piece of equipment is used, which allows observation of the movement of cells in a homogeneous dc-field microscopically (Fig. 2.47). Usually, for electrophoretic measurements, a field strength of the order of 1 kV m^{-1} is used. The chamber needs a sufficient temperature control, because according to Eq. 2.85 the electrophoretic movement depends on the viscosity of the solution. Furthermore, it must be considered that an electro-osmotic flow occurs, caused by the surface charges of the glass, which mostly are negative. Hence the chamber is closed, this flux must return through the center of the chamber. Between these two flows of opposite directions, neutral planes must exist, where the streaming velocity becomes zero. This is the location, where it is possible to measure a correct electrophoretic mobility of cells. These planes are located at a distance, 21.1% of the total thickness of the chamber on both sides of it (see Fig. 2.47).
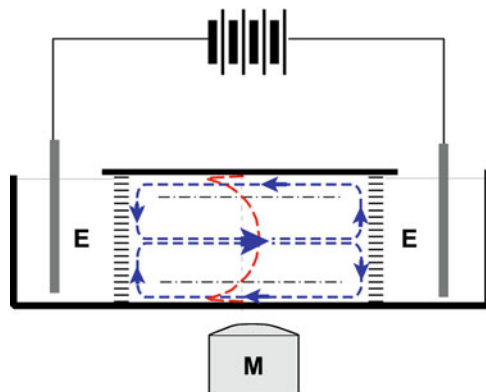
Fig. 2.47 Schematic illustration of equipment to measure cell electrophoresis; **M** microscope; **E** chamber for electrodes, which are connected to the measuring chamber by a conducting medium. The *blue arrows* indicate the direction of the electro-osmotic flow. The *red line* shows the flow profile in the chamber. –·–·–·–·– the neutral planes

Since the development of the method of cell electrophoresis many cells and organisms have been investigated. Under physiological conditions, nearly all biological objects indicate a ζ-potential between -10 and -70 mV. Only at pH < 4 in some cases, do these potentials become positive. This holds also for bacteria, yeast, and plant cells, which are surrounded by a cell wall. The cell wall is a thick (in relation to the membrane) porous ion exchange layer with fixed negative charges.

In contrast to cell electrophoresis, which is used routinely to measure surface charges, the reversible effect, i.e., the *sedimentation potential*, sometimes also called the *Dorn effect*, seems not to be so important. In this case, a small electrical potential difference occurs in suspensions of quickly sedimenting charged particles.

An electrical potential difference which occurs if a solution is streaming along a charged surface, or inside a capillary with surface charges, is called the *streaming potential*. The reason for this is the superposition of the streaming profile and the profile of mobile charges in an electrical double layer. In detail, the theory of streaming potential is quite complex. In the center of a capillary, for example, the streaming velocity is maximum, and the concentration of the co-ions of the surface charges – minimum. Therefore, these co-ions will flow faster than the counterions near the surface. This leads to a certain charge differentiation at the ends of the capillary. The resulting electric field, however, enforces the movement of counterions and lowers the movement of the co-ions. The effect, therefore, is lowered by this reaction. In capillaries with an extremely small diameter, additionally, a Donnan effect must be considered. This means, that the amount of surface charges always shifts the total electrolyte compositions in this fine capillary toward a Donnan equilibrium (see Sect. 3.2.4). The total amount of counterions, therefore, is larger than that of the co-ions. This can even lead to a streaming potential with opposite polarity. It is possible to measure the streaming potential in biological

tissue, if a flow of solutions is induced by osmotic gradients. The potential difference, measured in such experiments is only of the order of a few millivolts.

The reversible effect of streaming potential is *electro-osmosis*. In this case, based on the same considerations, near a charged surface, or in a capillary, a flow of solution occurs if a dc-field is applied.

Recently, the combination of electrophoresis and electro-osmosis has been used to measure the ζ-potential of the intercellular space in tissue. For this, small spots of fluorescent probes were injected into tissue slides and their movement under the influence of an applied electric field was observed microscopically. This movement is the result of the electrophoresis of the marker molecules driven by their own charges, plus their transport by the electro-osmotic flow, generated by the surface charges of the intercellular cleft. The electrophoretic mobility of the applied molecules can be measured separately, which allows us to evaluate the electro-osmotic effect separately. A neutral compound without an electrophoretic mobility in a cleft with negative charges would move toward the cathode (see Fig. 2.45). One difficulty is the tortuosity of the path in the intercellular space. This problem can be overcome by simultaneously using two fluorophores of different sizes and each differently charged. As the result of such experiments, the ζ-potential of brain tissue has been found to be in the same range as that of isolated cells, i.e., around -25 mV.

Furthermore, these kinds of investigations are important to give a quantitative insight into the interplay between different phenomena and permanent properties influencing the process of *iontophoresis*. This is a noninvasive technology for the delivery and monitoring of drugs through the skin epidermis under the influence of an applied electric field to increase and modulate the rate of transport.

The physiological role of electrokinetic phenomena is not very clear. It cannot be ruled out that electrophoresis of charged components in the membrane, or of proteins in intercellular spaces occurs under the influence of in vivo generated potentials (see Sect. 3.5.2). A hypothesis exists, whereby microorganisms can move electrophoretically in their own generated field. Streaming potentials occur in bone during mechanical deformations. In this case, it is induced by the flow in the small canals of the bone, the so-called canaliculi (see Sect. 3.5.2).

Further Reading

Theory of electrical double layers: Donath and Voigt 1986; Glaser 1996; Voigt and Donath 1989; cell electrophoresis: Bauer 1994; streaming potential in bones: Rubinacci et al. 2002; self-electrophoresis of microorganisms: Lammert et al. 1996; electrokinetic phenomena in tissue: Guy et al. 2008; iontophoresis: Imanidis and Luetolf 2006.

2.3.6 The Electrostatic Structure of the Membrane

After explaining some structural and mechanical properties of biological membranes in previous Sections, and after having discussed ionic properties near

fixed charges, we will now introduce the membrane as an electrical structure. This is a crucial point for many biophysical considerations, and it will pervade this whole book. Electrical properties of the biological membrane will be mentioned in the context of many topics. Whereas we introduce here, first of all the electrostatic properties of the membrane, we will continue this discussion later in Sects. 3.4.3 and 3.5.2, including discussion of electrodynamic processes, such as generation of electrical currents by processes coupled with cell metabolism. In Sect. 3.5.3 the explanation of passive membrane properties will be continued, including its behavior in electromagnetic fields. This knowledge will be applied in Sect. 3.5.5 to introduce techniques of cell manipulation with electric fields. Finally, in Sects. 4.5, 4.6, and 4.7 we will use this model to discuss the influence of weak electric and electromagnetic fields on cells and organisms.

In contrast to the external medium and to the cell plasma, the cell membrane has a high electrical resistance and a low dielectric constant. The membrane is an extremely thin hydrophobic electrically isolating interface between two aqueous phases. It can be considered to be like an electrical RC-system with certain capacity (C), and an electric resistance (R) (see Figs. 3.27 and 3.39).

The specific capacity (C_{sp}) can be calculated from the membrane thickness (Δx) and the dielectric constant (ε):

$$C_{sp} = \frac{\varepsilon_0 \varepsilon}{\Delta x} \tag{2.89}$$

The specific capacity of the cell membrane is relatively constant, because both parameters, the dielectric constant (ε), as well as the membrane thickness (Δx) cannot vary to a great extent. For the normal cell membrane it amounts approximately to 10 mF m^{-2} (for methods used to measure this parameter, see Sects. 3.5.3 and 3.5.4). Using this value, the rearranged equation (Eq. 2.89) allows calculation of the mean dielectric constant of the membrane, assuming: $\Delta x = 8 \cdot 10^{-9}$ m:

$$\varepsilon = \frac{C_{sp} \Delta x}{\varepsilon_0} = \frac{10^{-2} \cdot 8 \cdot 10^{-9}}{8.854 \cdot 10^{-12}} = 9.0$$

This value is surprisingly large, because the dielectric constant of lipids amounts only to $\varepsilon = 3.5$. The reason for this difference is the heterogeneity of the membrane, in particular, its content of proteins.

The specific capacity of the biological membrane is a very important parameter, indicating the relation between the amount of charges, resp. their distribution density (σ, in C m^{-2}), which are required to generate a membrane potential difference ($\Delta\psi$, in V):

$$C_{sp} = \frac{\sigma}{\Delta\psi} \tag{2.90}$$
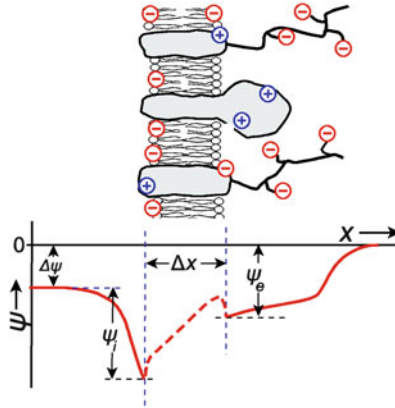
Fig. 2.48 The fixed charges of a cell membrane and the corresponding electrical potential (ψ). The transmembrane potential ($\Delta\psi$) is indicated as the difference between the potentials of the internal and external phases. The parameters ψ_i and ψ_e indicate the internal and external surface potentials; Δx means the thickness of the membrane. Inside the membrane itself, the function $\psi(x)$ depends on the positions at which the membrane is cut. Therefore, the function in this region is marked only by a simple connection between the surface potentials, considering only the dipole effect of the polar lipid groups

In Sect. 3.4.3 we will use this relation to explain the role of the membrane as an accumulator of electrochemical energy.

Despite the usefulness of these rough estimations it must be mentioned that the real molecular structure of the membrane and its surface, and therefore the real dielectric properties of it, in fact, are more complicated. In Sect. 2.3.3 we already described the membrane as a dynamic structure containing fixed charges in a two-dimensional pattern of the membrane area and in a charge profile perpendicular to this. This particular profile along an x-coordinate, cutting the membrane area perpendicularly, is depicted schematically in Fig. 2.48. This picture does not contain the mobile ions in the double layers. It assumes the existence of a transmembrane potential ($\Delta\psi$) *as* a potential difference between the inner and the outer phase of the cell. This could be a result of a Donnan equilibrium, of a diffusion potential, or finally directly generated by electrogenic transporters. These circumstances will be discussed later in Sect. 3.4 in detail.

The fixed surface charges of the outer part of the cell membrane mainly are the result of dissociations of carboxyl groups of neuraminic acids (also called: sialic acids), positioned at the ends of the glycoproteins. The isoelectric point of these neuraminic acids is near pH 2.6. For most membranes the surface density of negative charges is between -0.01 and -0.02 C m^{-2}. Human erythrocytes carry on their surface, which is about 140 μm^2, a number of approximately 10^7 dissociated neuraminic acid groups.

Additional carriers of electric charges are the polar groups of several phospholipids. The negatively charged phosphatidylserine is of particular importance.

In most cells it is located exclusively at the inner leaflet of the membrane. In human erythrocytes its surface charge density amounts approximately to -0.09 C m^{-2}.

As already mentioned in Sect. 2.3.3, proteins are located in the membrane in such a way that their polar groups are oriented towards the water phase (Fig. 2.39). Hence, they also will contribute to the surface charge of the membrane by charges of both sign. These charges of membrane proteins are very important controlling various protein functions, such as for example the ion transport through the membrane.

These considerations indicate that the models of electrical double layers, resulting from charges of a planar surface (Fig. 2.43) are just rough approximations for the case of the real cell membrane. In fact the glycocalyx forms an external layer covering the surface of most cells. The fixed charges of it are to be considered as charges in a definite space, but not directly on the plane of the surface. It must be therefore introduced into the Poisson–Boltzmann equation (see Eq. 2.75) as space density (ρ, in C m^{-3}), instead of the surface charge density (σ, in C m^{-2}):

$$\nabla^2\psi = -\frac{1}{\varepsilon_0\varepsilon}\left(\rho + F\sum_{i=1}^{n} c_{i0}z_i e^{-\frac{z_i e\psi}{kT}}\right) \tag{2.91}$$

Even for the simplest functions $\rho(x)$, which are to be introduced into Eq. 2.91, an integration of this equation is possible only by iteration. The resulting function $\psi(x)$ (Fig. 2.48), of course differs from that of a classical double layer (Fig. 2.43). The effective range of the electrical potential in this case is predicted first by the thickness of the glycocalyx, and only indirectly, by the Debye–Hückel length ($1/\kappa$). This thickness, however, itself is determined by electrostatic interactions of its charges. In solutions of high ionic strengths, these charges are more or less shielded, and the glycoproteins are more closely attached to the membrane. In solutions with lower ionic strength, however, the fixed charges repulse each other, the glycoprotein filaments, therefore are packed more loosely, and the glycocalyx becomes thicker. Hence, the function $\rho(x)$, i.e., the distribution of the charges of the neuraminic acids, itself depends on $\psi(x)$. The surface structure of the cell therefore is controlled by the cell itself, as well as by the conditions of its surroundings.

In the case of cells in tissue these glycoproteins are packed in the intercellular cleft. As pointed out in the previous Sect. 2.3.5 it is possible to obtain a picture of this situation, measuring the electrokinetic movement of fluorescent dyes in this region. The Donnan osmotic properties of this layer control its dimension. In Sect. 3.5.2 the intercellular cleft will be considered as a region for extracellular ionic currents (see Fig. 3.34).

Not only the external, but also the internal membrane surface carries fixed charges. These are in the first line the charges of phosphatidylserine, and additionally, a number of surface-associated proteins, which are not depicted in Fig. 2.48. In contrast to the surface charges at the outside of the cell, which can be measured by

cell electrophoresis, the charges of the cytoplasmatic side cannot be measured easily.

In Fig. 2.48 the function $\psi(x)$ inside the membrane, the *inner membrane potential*, is depicted only schematically. In fact, this function strongly depends on the place, where the cut was drawn. If it is taken through the lipid phases, the influence of the lipid head groups as dipoles must be considered, as schematically depicted in this figure. The profile through a protein region, of course is quite different from this.

Beside this electrical profile in a direction perpendicular to the membrane surface, a lateral mosaic of charges exists, which is partly controlled by the cytoskeleton. This charge distribution has been investigated by electron microscopy, and recently also by atomic force microscopy. It is also possible to measure the lateral electrophoretic movement of charged membrane proteins, marked by fluorescence labels if the cells are observed in strong external fields. The investigation of the dynamics of this lateral electrostatic structure of biological membranes and its regulatory function is, however, only at the beginning.

Although we will discuss the role of the membrane electric field, as well as the application of external fields to cell membranes in detail later in Sects. 3.4, 3.5, as well as 4.5–4.7, we will nevertheless already mention here the biological relevance of this electrostatic structure, and its role in several cell physiological processes. In general, on one hand, the profile of the electrical potential near the membrane, directly creates a profile of specific ion conditions. On the other hand, its first derivative, the electric field strength, affects some processes directly.

Inside the membrane, and in its vicinity, the electric field strength is of the order of 10^7 V m^{-1}. In Fig. 2.48 the course of the potential inside the membrane is marked only by a simple connection between the surface potentials, considering just the dipole effect of the polar lipid groups. It is a dotted line, because the real function $\psi(x)$ in this region strongly depends on the location, where the membrane was cut. This function, representing the intermembrane potential, would be quite different, if the x-axis were to cut a protein. The local charge densities, as well as the local dielectric constants in these cases would differ strongly. Furthermore, it must be considered that the function $\psi(x)$ will change dramatically, if the transmembrane potential $\Delta\psi$ is depolarized during the action potential in a time course of about one millisecond (see Fig. 3.29). Hence, we come to the conclusion that inside and near the membrane, an extremely strong electric field occurs which is not only a one-dimensional function $\psi(x)$, as depicted in Fig. 2.48, but which rather must be considered as a three-dimensional time-dependent parameter.

It is clear that this field must influence all polar or polarizable molecules in this region. This particularly concerns membrane proteins containing special segments acting as voltage sensors. Nerve cells are the classical example of a feedback circle, generating the action potential: An alteration of the membrane permeability, caused by any influences, changes the diffusion potential $\Delta\psi$. The resulting modification of the field strength in the membrane changes the permeability of the Na$^+$-, and subsequently K$^+$-channels, which again influences the membrane potential. We will explain these reactions in detail in Sect. 3.4.4. Potential sensitive transporters occur in all cells.

Not only proteins, but also lipids are influenced by the field strength inside the membrane. The field influences the orientation of the polar head groups of the lipids, and in this way modifies the effective area which they occupy in the membrane. We already discussed the consequence of this modification in Sect. 2.3.2. The field may also induce an intrinsic membrane pressure, which may alter the phase transition in the hydrophobic region of the membrane. Such processes are the basis of mechanisms of electromechanical coupling.

The electrical potential near the membrane generates particular microconditions near charged groups. Dependent on the ionic strength, and therefore on the Debye–Hückel length ($1/\kappa$), a lateral mosaic of local concentrations, and local pH values can occur. In Fig. 2.44 the local ionic concentrations and local pH values are illustrated as a function of distance from a charged surface. It must be emphasized that the field changed the concentration of multivalent ions, as for example Ca^{++} to a larger degree than univalent ions like K^+ of Na^+. In this context, the role of local charges must be considered for the control of membrane-bound enzymes and transporters.

Recently it was shown that this extremely high electric field strength near the membrane also influences the dissociation constant of molecules and molecular complexes. This process is applicable to a wide range of biological complexes including small molecules (norepinephrine-morphine sulfate), protein-protein complexes (insulin-glucagon), and small molecule-protein complexes (epinephrine-bovine serum albumin).

Further Reading

Electrostatic structure: Cevc 1990; Glaser 1996; Honig et al. 1986; McLaughlin 1989; role of membrane lipids: Starke-Peterkovic et al. 2005; surface mapping with atomic force microscopy: Heinz and Hoh 1999; effects of membrane near electric field: Dillon et al. 2006; lateral distribution of charges: McLaughlin and Murray 2005.