Alexander Vologodskii

The Basics of **Molecular Biology**

Springer

The Basics of Molecular Biology

Alexander Vologodskii

The Basics of Molecular Biology

Springer

Alexander Vologodskii
New York University
New York, NY, USA

ISBN 978-3-031-19403-0 ISBN 978-3-031-19404-7 (eBook)
https://doi.org/10.1007/978-3-031-19404-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book represents an attempt to describe the major features of life in a very short volume. It is addressed to people who have an education in areas different from biology and who want to receive some ideas about the world of molecular and cell biology. The book assumes a certain knowledge of physics, chemistry, and mathematics that corresponds to the level of high school. The primary attention is paid here to the most fundamental features of life. The book also briefly considers the developing areas of biology, which are especially important for medicine and the future of humankind.

Recommended Literature

There are many excellent textbooks on molecular and cell biology which describe the subject in detail. Three of these popular titles are shown below. Two other references are covering new fast-growing topics related to human genetics.

1. Alberts B, Heald R, Johnson A et al. (2022). Molecular biology of the cell. 7th ed. New York, NY: Norton & Company.
2. Watson J, Baker T, Bell S et al. (2013). Molecular biology of the gene. 7th ed. Menlo Park, CA: Benjamin Cummings.
3. Zlatanova J & van Holde K (2015). Molecular Biology: Structure and Dynamics of Genomes and Proteomes. New York, NY: Garland Science.
4. Reich D (2018). Who we are and how we got here. Ancient DNA and the new science of the human past. New York, NY: Pantheon Books.
5. Plomin R. (2019). Blueprint: How DNA Makes Us Who We Are. Cambridge, MA: The MIT Press.

New York, NY, USA　　　　Alexander Vologodskii

Introduction

We do not know how life appeared on our planet. However, we do know a lot about the organization of this life, and the goal of this book is to give a brief outline of its major features. Among natural sciences, biology stays apart from physics and chemistry in one very important respect. We can deduce, at least qualitatively, properties of physical objects from the small number of the basic laws of physics. We can predict the motion of planets around the Sun, the magnetic field around electric current, the way light passes through the lens, and so on. Although it is technically difficult to deduce the properties of atoms and molecules from the laws of quantum mechanics, we are confident that this is possible in principle. Biology is different. There is no way to deduce the way of life from general principles. There is an enormous amount of specific complex problems that have to be solved by living organisms, and these problems have many different solutions. Although each of these solutions has to follow the laws of physics and chemistry, it is impractical to search for all possible solutions to a particular problem in the biological world. Life could happen in many different ways. We, however, want to know the way of life on our planet, the one very special choice of many possibilities. To learn about this way, we have to study life experimentally. It makes the task more difficult. However, the evolution of life always had certain logic behind it, and finding this logic greatly helps in the study. We will try to articulate this logic through this book.

Although we cannot deduce much knowledge about life from the general laws of physics and chemistry, we are confident that life follows these laws. This belief will be at the base of the book. Analyzing basic processes of life, we will try to consider them as physical phenomena, wherever such consideration is useful.

Complexity is not the only striking feature of life. It is also incredibly diverse. Indeed, trees do not resemble animals, and animals seem to have little in common with bacteria. We know today, however, that the most fundamental features of all living organisms are nearly identical. All living organisms consist of cells. Some of them, like animals, have many trillions of cells in their bodies, while others consist of a single cell. But in all cases, the cell represents a unit of life. It has all the necessary elements for self-reproduction, and new cells can be only formed by the division of the existing cells. All basic elements involved in cell reproduction are essentially identical in all organisms living on our planet. Through all living organisms, hereditary information is kept in DNA molecules (molecules of deoxyribonucleic acid), and this information is coded in the same way! The same intracellular mechanisms are used to read this information and to use it for cell growth and reproduction. And all cells use the same mechanism for another major process, duplication of the hereditary information before the cell division. All cells are fundamentally similar inside, regardless of the striking difference in their appearances. These fundamental features of life clearly manifest that all living cells originate from a single colony that somehow emerged on Earth more than 3 billion years ago.

Acknowledgments

I would like to thank my friends Anshel Gorokhovsky and Igor Kulikov, and my sister Elena Vologodskaia for their remarks and suggestions which helped me greatly in preparing this book.

Contents

Chapter 1
The Major Processes in the Biological World

1.1 Cells as Basic Units of Life

It was in the middle of the nineteenth century when scientists concluded that cells represent fundamental units of life, although the term "cell" appeared much earlier. The major experimental basis of the theory was observations in the optical microscope. The size of the cells is in the range of 1–100 μm, so the limited resolution of the microscopes did not allow us to study the details of the cell organization. Therefore, it seems amazing that in 1839, Matthias Schleiden and Theodor Schwann formulated the cell theory. The theory stated that all living organisms are composed of one or more cells. They suggested that each cell contains all the needed hereditary information to produce new cells. In 1855, Rudolf Virchow added one more fundamental point there. He stated that new cells can be formed only by the division of preexisting cells. Today we know that these major statements of the cell theory are absolutely correct.

Of course, animals consist of many very different types of cells, and all of them are different from single-cell bacteria. But one of the striking features of life, which Schleiden and Schwann could not suggest, is that the most fundamental processes inside all living cells follow the same universal mechanisms. In all cells, hereditary information is stored in DNA molecules that have an identical double-stranded structure in all cells. The same principle and mechanism are used for the duplication of DNA molecules during cell division. The genetic information stored in DNA is used to build proteins, and the same universal complex apparatus do it in all living cells. And the same code for recoding protein structures is used in all living organisms! So, life on our planet proceeds fundamentally in a single universal way, in remarkable agreement with Darwin's theory of evolution. Thus, if we want to understand the basic principles of life, we should, first of all, consider an organism consisting of a single cell. This universality of the way of life makes studying it much easier, of course. However, the organization and functioning of even the

A. Vologodskii, *The Basics of Molecular Biology*,
https://doi.org/10.1007/978-3-031-19404-7_1

simplest cell are incredibly complex, and it remains to be a mystery how the first cell appeared on Earth.

In this chapter, we outline the major universal biochemical processes of life, storing and duplication of genetic information, and its use for the synthesis of proteins, the key molecules of the cells. Our emphasis here will be on the most general principles of these processes. More detailed consideration of various issues will be given in the following chapters. It is essential that at the level of detail used in this chapter, the description is equally applied to all living cells on Earth.

1.2 Important Facts from Chemistry and Physics

There are features of great importance for the properties of biological macromolecules, which we will consider in this chapter. Although these are not features specific only to biological macromolecules, they do not receive enough attention in traditional chemistry courses.

1.2.1 The Flexibility of Polymer Chains

All molecules consist of atoms that are bound by chemical bonds. Chemistry considers, first of all, covalent bonds between atoms. These bonds are very strong and stable, so they are not destroyed by collisions with other molecules due to thermal motion at normal temperatures. A covalent bond between two atoms originates from sharing the atoms' electrons. If one pair of electrons, one from each atom, is shared, it creates a single covalent bond. If two pairs of electrons, two from each atom, are shared, this creates a double covalent bond. It is common, when drawing the structural formula of a molecule, to show single covalent bonds by single lines connecting the atoms. The double bonds are shown by double lines, correspondingly (Fig. 1.1). The covalent bonds are very strong, and the energy of a single bond is in the range of 50–120 kcal/mol.

The angles between adjacent bonds depend on the chemical structure of the molecule and do not change much due to thermal motion. However, larger molecules can change their shape, or *conformation*, due to the rotation of their parts around

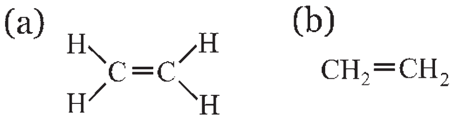

Fig. 1.1 The structural formula of ethylene. (**a**) The hydrogen atoms are connected to the carbon by the single covalent bonds, and the carbons are connected by the double bond. (**b**) A reduced drawing of the formula. Since hydrogen has a single electron, it can form only single bonds with other atoms, so the reduced drawing does not introduce any ambiguity

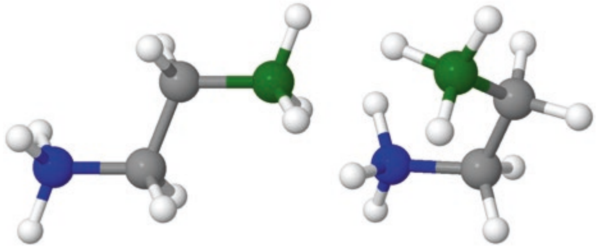

Fig. 1.2 Two different conformations of butane. Large spheres correspond to carbon atoms while small white spheres correspond to hydrogens. The conformations are obtained by rotation around the bond between two central atoms of carbon. The rotation does not change the angles between adjacent single bonds that connect carbon atoms. The image was obtained with the program ChemTube3D by N. Greeves

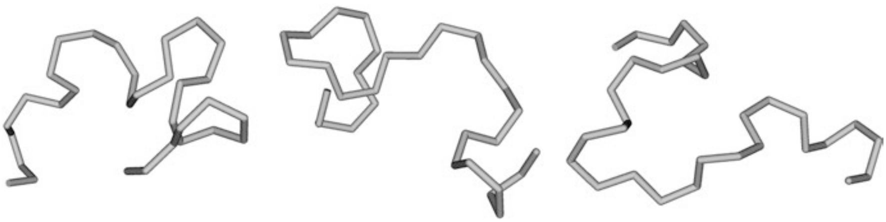

Fig. 1.3 Conformations of polyethylene $CH_3 - (CH_2)_{29} - CH_3$. Three randomly picked conformations of the molecule's backbone are shown. The carbon atoms are located at the chain vertices. The polymer molecule can adopt a huge number of possible conformations due to rotation around C–C bonds. C–H bonds are not shown here

single bonds of the backbone. Such rotation around single bonds is restricted only by possible collisions of spatially close groups of atoms. For example, butane, whose structural formula is CH_3–CH_2–CH_2–CH_3, has distinguished conformations specified by the mutual arrangements of four carbon atoms. Two of these conformations are shown in Fig. 1.2.

Rotation around the single bonds creates a huge amount of possible conformations of polymer chains. This is illustrated in Fig. 1.3, where three randomly chosen conformations of polyethylene, a polymer with a repeating motive $-CH_2-$, are shown. This flexibility of polymers due to rotation around single bonds in the backbone is extremely important for the properties of nucleic acids and proteins.

1.2.2 Noncovalent Interactions Between Atoms

There are a few types of noncovalent bonds that appear in biological molecules in water solutions. These bonds are weak. Thermal motion causes permanent collisions of molecules in solution, and such collisions can easily destroy noncovalent bonds. However, if a few such bonds are formed between two groups of atoms, the probability of their simultaneous destruction becomes low. Thus, weak noncovalent

Fig. 1.4 Diagram of a hydrogen bond. The hydrogen atom forms a covalent bond with atom A_1 and a hydrogen bond with atom A_2. Atoms A_1 and A_2 can be either nitrogen or oxygen atoms

bonds are capable of stabilizing certain conformations (structures) of the macromolecules and complexes of those molecules. Noncovalent bonds are extremely important in the functioning of biological macromolecules. They can be divided into four groups.

Ionic Bonds These bonds correspond to the electrostatic attraction between oppositely charged atoms. Although the strength of ionic bonds is greatly reduced in water, they are the strongest noncovalent bonds. Still, their strength is ten times lower than the strength of covalent bonds. The length of these bonds varies but exceeds 0.25 nm.[1]

Hydrogen Bonds Hydrogen bonds are created by the partial sharing of a hydrogen atom between two electronegative atoms (nitrogens and oxygens). The bond strength has the largest value when it is directed along the covalent bond between the hydrogen and another atom, as shown in Fig. 1.4. The length of the hydrogen bond is close to 0.2 nm. In water solutions, the hydrogen bonds are about three times weaker than ionic bonds.

Van der Waals Bonds This bond can appear between any two uncharged atoms which are not bound by a covalent bond. The atoms repel one another if the distance between them is shorter than a certain limit. We can say that this critical distance, r_0, corresponds to the sum of the radiuses of the atoms. The value of r_0 is usually close to 0.35 nm. It turns out that if the distance between the atoms exceeds r_0 (but smaller than $2r_0$), they attract each other, creating a van der Waals bond. Such bonds are a hundred times weaker than covalent bonds. Still, if many van der Waals bonds are formed simultaneously between atoms of a large polymer molecule (or two molecules), they can substantially stabilize a particular conformation of the molecule or a complex of two molecules.

Hydrophobic Interaction Water molecules are highly polar, and they form a dense network of hydrogen bonds. These bonds greatly reduce the total energy of the water solution. The interfaces between water molecules and nonpolar groups of other molecules destruct the network and therefore increase the solution energy. If nonpolar groups contact one another rather than water molecules, the distraction will be smaller, and the energy of the solution will be lower. Therefore, contacts between nonpolar groups are stabilized in water solutions. Thus, there are no hydro-

[1] 1 nm equals 10^{-9} m. It is a common unit of length in molecular scale.

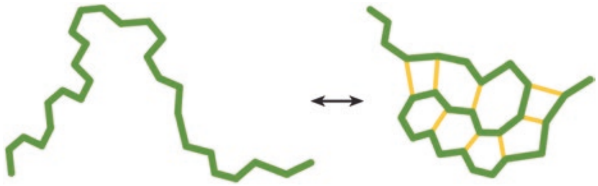

Fig. 1.5 Weak noncovalent interactions can stabilize some conformations of flexible chain molecules. One randomly picked conformation of the chain, which is not stabilized by weak bonds, is shown on the left. A compact conformation of the same chain, shown on the right, is stabilized by eight weak interactions (shown by yellow or by light gray in some formats)

phobic bonds, but there is an attraction between nonpolar molecules in water solutions due to so-called hydrophobic interaction.

Nearly all biological macromolecules are linear chemical chains. Due to the backbone flexibility, they can adopt many different conformations. Among them, there can be conformations that are well stabilized by noncovalent bonds between segments of the macromolecules (Fig. 1.5). Such conformations are critically important for cell life. During their functioning inside the cell, the majority of the macromolecules must change their conformations. These changes are only possible because the conformations of macromolecules and their complexes are stabilized by weak noncovalent bonds. This conformational flexibility would not be possible if the three-dimensional (3D) structures of the molecules are stabilized by covalent chemical bonds rather than by weak interactions. Without this conformational flexibility of the macromolecules, life would not be possible. We will return to these issues and their consequences again and again in this book.

1.3 Structure of DNA and Inheritance of Genetic Information

1.3.1 The General Principles

It is generally accepted that modern biology started in 1953 when Watson and Crick suggested the structure of DNA. Their model was based on the data of X-ray diffraction from DNA fibers. This kind of data does not allow unambiguous reconstruction of a molecule structure. One has to design a hypothetical structure, calculate its diffraction pattern, and compare it with the diffraction pattern. If the calculated pattern does not match the experimental data, a different structure has to be designed. In this way, one can, eventually, find a good solution. Using the experimental data of Franklin, Watson and Crick were the first who suggested the correct DNA structure. It was already known at that moment that DNA is the molecule that carries and transmits hereditary information. Therefore, everybody knew that

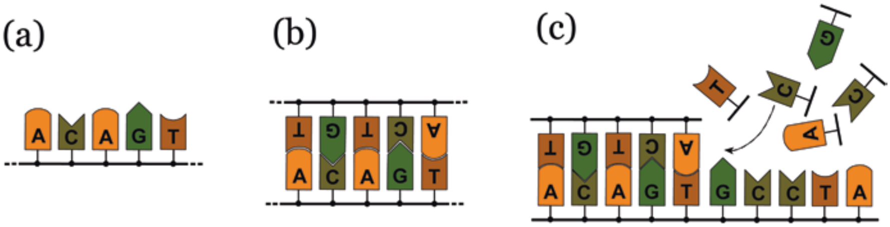

Fig. 1.6 Recording and replicating genetic information. (**a**) The information is coded in the linear string of four different elements (*bases*), A, C, G, and T, which are the side groups of the single-stranded DNA chain (shown by the black line). (**b**) There is a strict correspondence between the sequences of the strands in the double-stranded DNA since A in one strand can be paired only with T in the other strand and G can be paired only with C. Thus, the sequence of the elements in one strand completely specifies the sequence in the other, complementary strand. (**c**) During the *replication* process (making two copies of DNA molecules from one), DNA strands are separated, and the complementary strands are built on each of the parental strands. For simplicity, the synthesis of only one daughter double-stranded DNA is shown

finding the DNA structure is crucial for biology. Still, the structure itself exceeded expectations because it immediately explained the physical principle of inheritance. We will consider this structure in the next subsection, after explaining the principle of recording and duplicating the hereditary information on the simple diagrams.

The DNA structure represents a double helix formed by two single-stranded polynucleotide chains (the repeating units of the chain are called nucleotides). The strands are kept together by weak interactions and can be separated by various means, such as interaction with other molecules or elevated temperatures. The backbone of each chain consists of identical repeating units. Each repeating unit has a side element, A, C, G, or T (Fig. 1.6a). The hereditary, or genetic, information is coded by the sequence of these elements, similar to the strings of letters we use to record words and sentences.

Thus, in principle, the genetic information is written in a very natural way, as a sequence of four letters. We consider later in this chapter conversion of this information into the structure of proteins, which is a key biological process. There is another critical task for the cell, however. The DNA molecule (or molecules) has to be duplicated before the cell division, so one copy of each molecule goes to each of the dividing cells. Remarkably, the physical principle of the latter process immediately follows from the structure of the double helix.

The chains of the double-stranded DNA are kept together by many noncovalent bonds. Simultaneous formation of these bonds is possible because the surface of one chain complements a similar surface of the other chain (Fig. 1.6b). This complementary exists only if A in one strand is paired across the helix with T in the other strand, and G is paired with C. Thus, only AT and GC base pairs can be incorporated into the structure without its distortion, and only these pairs exist in the double-stranded DNA. The complementarity rule means that the sequence of elements in one strand completely specifies the sequence in the other strand.

This principle of DNA structure offers a simple and very elegant principle of duplicating genetic information. First, the parental strands of the double helix have to be separated. Then, the new complementary strands have to be synthesized from