Introduction

This book covers some subjects that we find inspiring when teaching physics students about biology. The book presents a selection of topics centered around the physics/biology/chemistry of genes. The focus is on topics that have inspired mathematical modeling approaches. The presentation is rather condensed, and demands some familiarity with statistical physics from the reader. However, we attempted to make the book complete in the sense that it explains all presented models and equations in sufficient detail to be self-contained. We imagine it as a textbook for the third or fourth years of a physics undergraduate course.

Throughout the book, in particular in the introductions to the chapters, we have expressed basic biology ideas in a very simplified form. These statements are meant for the physics student who is approaching the biological subject for the first time. Biology textbooks are necessarily more descriptive than physics books. Our simplified statements are meant to reduce this difference in style between the two disciplines. As a consequence, the expert may well find some statements objectionable from the point of view of accuracy and completeness. We hope, however, that none is misleading. One should think of these parts as first-order approximations to the more complicated and complete descriptions that molecular biology textbooks offer. On the other hand, the physical reasoning that follows the simplified presentation of the biological system is detailed and complete.

The book is not comprehensive. Large and important areas of biological physics are not discussed at all. In particular we have not ventured into membrane physics and transport across membranes, signal transmission along neurons and sensory perception, to mention a few examples. While there are already excellent books and reviews on all these subjects, the reason for our limited choice of topics is more ambitious. The basic physics ideas that are relevant for molecular biology can be learned on a few specific examples of biological systems. The examples were chosen because we find them particularly suited to illustrate the physics.

We have chosen to place the focus on genes, DNA, RNA and proteins, and in particular how these build a functional system in the form of the λ-phage switch. We further elaborate with some larger-scale examples of molecular networks and with a short overview of current models of biological evolution. The overall plan of the book is to proceed from simple systems toward more complex ones, and from small-scale to large-scale dynamics of biological systems.

Chapter 1 gives some impression of important ideas in biology. To be more precise, the chapter summarizes those concepts which, we think, strike a physicist who approaches the field, either because they have no counterpart in physics, or, on the contrary, because they are all too familiar. The chapter grew out of discussions with biologists, and we normally use it as a first introductory lecture when we give the course. Of the subsequent chapters, we regard Chapter 7 on the λ-phage in *E. coli* as especially central: it deals with the interplay between elements introduced earlier in the book, and it contains a lot of the physics reasoning that the book is meant to teach.

In Chapter 2 we describe the physics of polymer conformations, emphasizing the interplay between energy and entropy and examining both the behavior of extended polymers and how compact configurations may be reached. In the next chapters we introduce and discuss the most important biological polymers: DNA, RNA and proteins. Although the covalent bonds forming the polymer backbone have binding energies $\Delta G > 1$ eV, the form and function of these biomolecules is associated to the much weaker forces perpendicular to the polymer backbone. These interactions are of order $k_{\mathrm{B}}T$, and it is the combined effect of many of these forces that forms the functional biomolecule. In Chapters 3–5 we characterize the stability of DNA, RNA and proteins, with emphasis on the cooperativity responsible for this stability.

Biological molecules can be used for various types of computations. Chapter 3 includes a section on DNA computation and DNA manipulation in the laboratory. This is in part a continuation of Chapter 2 (reptation), and also an introduction to the computational aspects of molecular replication (the PCR reaction). Chapters 4–6, on the other hand, focus on proteins and protein folding and thus the functional aspects are left to subsequent chapters. In this book we have addressed in considerable detail one of these aspects, namely how a protein may control the production of another protein (Chapter 7). As we explain in Chapter 7, genetic control involves mechanisms associated to both equilibrium statistical mechanics and to the timescales involved in complex formation and disruption. Topics in this chapter include a discussion of cooperativity, of target location by diffusion, of timescales in a cell and of stability of expressed genetic states.

Chapter 7 also forms a microscopic foundation for the large-scale properties of molecular networks, which we discuss in Chapter 8. Chapter 8 thus continues the subject of genetic regulation and molecular networks, in part by venturing into the

heat shock mechanism. This shows that protein folding is also a control mechanism in a living cell, and it introduces a type of genetic regulation that was not treated in the previous chapter: σ sub-units of RNAp, which control the expression of larger classes of genes. Chapter 8 also discusses the larger-scale properties of genetic regulatory networks, introducing a few recent physics attempts at modeling these.

Chapter 9 discusses evolution, with emphasis on the interplay between randomness and selection from the smallest to the largest scales. The chapter introduces concepts such as neutral evolution, hill climbers and co-evolution, and uses these concepts to discuss questions related in part to the concept of punctuated equilibrium, and in part to the origin of life in the form of autocatalytic networks. Thus Chapter 9 introduces some simple models that allow us to discuss the nature of the history leading to the emergence of life, and in particular aims at stressing the importance of interactions and stochastic events on all scales of the biological hierarchy.

In the Appendix we have a short introduction to statistical mechanics, including the fluctuation–dissipation theorem and the Kramers escape problem; it is meant to render the book self-contained from the point of view of the physics.

1

What is special about living matter?

Kim Sneppen & Giovanni Zocchi

Life is self-reproducing, persistent (we are $\sim 4 \times 10^9$ years old), complex (of the order of 1000 different molecules make up even the simplest cell), "more" than the sum of its parts (arbitrarily dividing an organism kills it), it harvests energy and it evolves. Essential processes for life take place from the scale of a single water molecule to balancing the atmosphere of the planet. In this book we will discuss the modeling and physics associated, in particular, to the molecules of life and how together they form something that can work as a living cell. First we briefly review some basic concepts of living systems, with emphasis on what makes biological systems so different from the systems that one normally studies in physics.

Conceptually, molecular biology has provided us with a few fundamental/universal mechanisms that apply over and over. Some concepts, like evolution, do not have counterparts in physics. Others, like the role of stochastic processes, are, on the contrary, quite familiar to a physicist.

(1) **Biology is the result of a historical process.** This means that it is not possible to "explain" a biological system by applying a few fundamental laws in the same way that is done in physics. A hydrogen atom could not be different from what it is, based on what we know of the laws of nature, but an *E. coli* cell could. In evolution, it is much easier to modify existing mechanisms than to invent new ones. Thus on evolutionary timescales nearly everything comes about by cut and paste of modules that are already working. We will end the book with a chapter dedicated to evolutionary concepts and models.

(2) **The molecules of life are polymers.** At the molecular scale, life is made of polymers: DNA, RNA and proteins. Even membranes are built of molecules with large aspect ratios. Perhaps mechanics at the nano-scale can work only with polymers, molecules that are kept together by strong forces along their backbone, while having the property of forming specific structures by utilizing the much weaker forces perpendicular to the backbone. In molecular biology we witness nano-mechanics at work with polymers. We will discuss polymers in Chapter 2, and thereby introduce concepts necessary

4

for understanding DNA (Chapter 3), proteins (Chapters 4–5) and polymers in action
(Chapter 6).

(3) **Genetic code.** Information is maintained on a one-dimensional, double-stranded DNA
molecule, which will be discussed in Chapter 3. Thus the one-dimensional nature of
the information mirrors the one-dimensional nature of the polymers that make life
work. The DNA strands open for *copying* and *transcribing*, by separating the double-
stranded DNA into two single strands of DNA that each carry the full information.
The copying is done by DNA polymerase using the complementarity of base pairs.
Similarly the genetic code is read by RNA polymerase and ribosomes that again use
the matching of complementary base pairs to translate codons into amino acids. This
is usually summarized in terms of the central dogma

$$DNA \rightarrow RNA \rightarrow protein \tag{1.1}$$

This is highly simplified: proteins modify other proteins, and most importantly proteins
provide both positive and negative feedback on all the arrows in (1.1). If one has only
DNA in a test tube, nothing happens. One needs proteins to get DNA $\rightarrow$ RNA, etc.
Then Eq. (1.1) should be supplemented at least by an arrow from protein to DNA.
Thus it is not always clear where the start of this loop is, and the whole scheme has to
be extended to the complicated molecular networks discussed in Chapter 8.

(4) **Computation.** A living cell is an incredible information-processing machine: an
E. coli transcribes about 5×10^6 genes during 1/2 h, i.e. about 10 Gb/h of informa-
tion. All this within a 1 μm^3 cell, coded by about 5×10^6 base pairs. The information
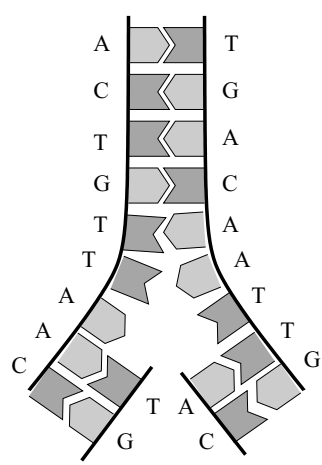density far outnumbers that in any computer chip, and even a million *E. coli* occupy

Figure 1.1. Information in life is maintained one-dimensionally through a double-
stranded polymer called DNA. Each polymer strand in the DNA contains exactly
the same information, coded in form of a sequence of four different base pairs.
Duplication occurs by separating the strands and copying each one. This interplay
between memory and replication opened 4 billion years of complex history.

much less space than a modern CPU, thus beating PCs on computation speed as well. The levels of computation in a living system increase when one goes to eukaryotes and especially to multi-cellular organisms (where each cell must have encoded awareness of its social context also). The simplest organisms (e.g. the prokaryote *M. pneumono-miae* with 677 genes) can manage essentially without transcription control. Larger genome size prokaryotes typically need a number of control units that grow with the square of number of genes. We discuss modeling of processes within living cells in Chapter 7 and, to some extent, also in Chapters 6 and 8.

(5a) **Life is modular.** It is build of parts that are build of parts, on a wide range of scales. This facilitates robustness: if a process doesn't work, there are alternative routes to replace it. Molecular-scale examples include the secondary, tertiary and quaternary structures of proteins (complexes of proteins); they may include network modules, such as sub-cellular domains, that each facilitate an appropriate response to external stimuli. Most importantly, the minimum independent living module is the cell.

(5b) **Life is NOT modular.** Life is more than the sum of its parts. Removing a single protein species often leads to death for an organism. Another observation is that the number of regulatory proteins for prokaryotes increases with the square of the number of proteins that should be regulated. Thus regulatory networks are an integrated system, and not modular in any simple way. This is the subject for the chapter on networks.

(6) **Stochastic processes play an essential role** from molecules to cells; in particular, they include mechanisms driven by Brownian noise, trial-and-error strategies, and the individuality of genetically identical cells owing to their finite number of molecules. An example of a trial-and-error mechanism is microtubule growth, attachment and collapse (see Chapter 6). Individuality of cells has been explored by individual cell measurements of gene expression, and variability of cell fate has been associated with fluctuations in gene expressions. An example of such stochasticity includes the lysis–lysogeny decision in temperate phages; see Chapter 7.

(7) **Biological physics is "$k_B T$-physics".** The relevant energy scale for the molecular interactions that control all biological mechanisms in the cell is $k_B T$, where T is room temperature and k is the Boltzmann constant ($k_B N_A = R$, where N_A is Avogadro's number and R is the gas constant; $1\ k_B T = 4.14 \times 10^{-14}$ ergs $= 0.62$ kcal/mole at $T = 300\ K$). This is not true for most of the systems described in a typical physics curriculum, for example:
- the hydrogen atom, with an energy scale ~ 10 eV, whereas $k_B T_{room} \simeq 1/40$ eV;
- binding energies of atoms in metals; covalent bonds: energy ~ 1 eV;
- macroscopic objects (pendulum, billiard ball), where even a 1 mg object moving with a speed of 1 cm/s has an energy $\sim 10^{-10}$ J $\sim 10^9$ eV (1 eV $= 1.602 \times 10^{-19}$ J).

The approach is therefore different. For example, in the solid state one starts with a given structure and calculates energy levels. Thermal energy may be relevant to kick carriers in the conduction band, but $k_B T$ is not on the brink of destroying the ordered structure.

Soft-matter systems often self-assemble in a variety of structures (e.g. amphiphilic molecules in water form micelles, bilayers, vesicles, etc.; polypeptide chains fold to

form globular proteins). These ordered structures exist in a fight against the disruptive effect of thermal motion. The quantity that describes the disruptive effect of thermal motion is the entropy S, a measure of microscopic disorder that we review in the Appendix. So for these systems energy and entropy are both equally important, and one generally considers a free energy $F = E - TS$. The language and formalism of thermodynamics are effective tools in describing these systems. For example: free-energy differences are just as "real" as energy differences; therefore entropic effects can result in actual forces, as we discuss in Chapter 2.

Further reading

Berg, H. C. (1993). *Random Walks in Biology*. Princeton: Princeton University Press.
Boal, D. H. (2002). *Mechanics of the Cell*. Cambridge University Press.
Bray, D. (2001). *Cell Movements: From Molecules to Motility*. Garland Publishing.
Crick, F. H. C. (1962). The genetic code. *Sci. Amer.* **207**, 66–74; *Sci. Amer.* **215**, 55–62.
Eigen, M. (1992). *Steps Towards Life*. Oxford University Press.
Godsell, D. (1992). *The Machinery of Life*. Springer Verlag.
Gould, S. J. (1991). *Wonderful Life, The Burgess Shale and the Nature of History*. Penguin.
Howard, J. (2001). *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Associates.
Kauffman, S. (1993). *The Origins of Order*. Oxford University Press.
Lovelock, J. (1990). *The Ages of Gaia*. Bantam Books/W. W. Norton and Company Inc.
Pollack, G. H. (2001). *Cells, Gels and the Engines of Life*. Ebner & Sons Publishers.
Ptashne, M. & Gann, A. (2001). *Genes & Signals*. Cold Spring Harbor Laboratory.
Raup, D. (1992). *Extinction: Bad Genes or Bad Luck*? Princeton University Press.
Schrödinger, E. (1944). *What is Life*? Cambridge University Press.

2

Polymer physics

Kim Sneppen & Giovanni Zocchi

Living cells consist of a wide variety of molecular machines that perform work and localize this work to the proper place at the proper time. The basic design idea of these nano-machines is based on a one-dimensional backbone, a polymer. That is, these nano-machines are not made of cogwheels and other rigid assemblies of covalently interlocked atoms, but rather are based on soft materials in the form of polymers – i.e. one-dimensional strings. In fact most of the macromolecules in life are polymers. Along a polymer there is strong covalent bonding, whereas possible bonds perpendicular to the polymer backbone are much weaker. Thereby, the covalent backbone serves as a scaffold for weaker specific bonds. This opens up the possibility (1) to self-assemble into a specific functional three-dimensional structure, (2) to allow the machine parts to interact while maintaining their identity, and (3) to allow large deformations. All three properties are necessary ingredients for parts of a machine on the nano-scale. In this chapter we review the general properties of polymers, and thus hope to familiarize the reader with this basic design idea of macromolecules.

Almost everything around us in our daily life is made of polymers. But despite the variety, all the basic properties can be discussed in terms of a few ideas. Some of these properties are astounding: consider a metal wire and a rubber band. The metal wire can be stretched about 2% before it breaks; its elasticity comes from small displacements of the atoms around a quadratic energy minimum. The rubber band, on the other hand, can easily be stretched by a factor of 4. Clearly its elasticity must be based on an entirely different effect (it is in fact based on entropy); see also Fig. 2.1.

Polymers are long one-dimensional molecules that consist of the repetition of one or a few units (the "monomers") bound together with covalent bonds. You can think of beads on a string. Figure 2.2 shows three examples; the first two are synthetic polymers, the third represents the primary structure of proteins. What is radically different between these molecules and all others is that the number of
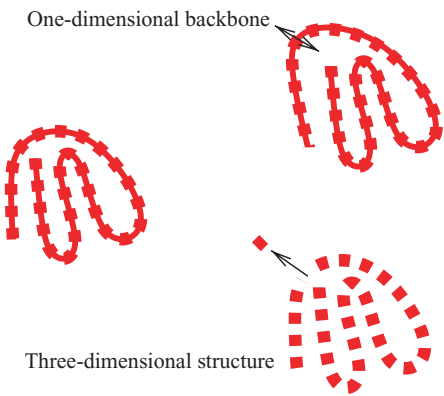
8

Figure 2.1. Illustration of the self-healing properties of a device with a one-dimensional backbone. Thermal or other fluctuations may dislodge a single element, but if attached to a backbone it typically will move back into the correct position (from Hansen & Sneppen, 2004).
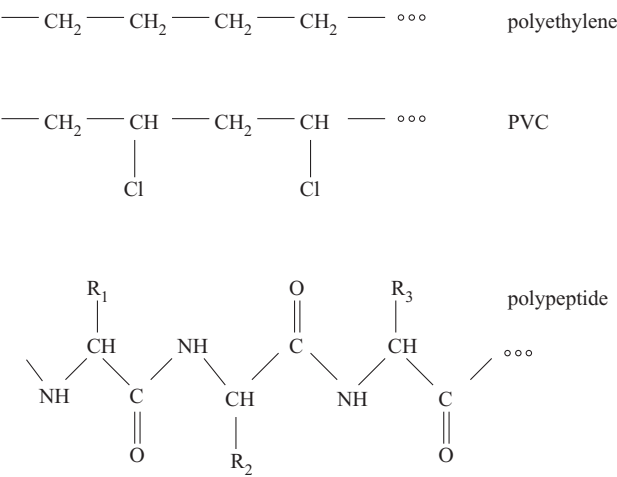

Figure 2.2. Examples of polymers.

monomers, N, is large, typically $N \sim 10^2 - 10^4$ (but note that for DNA N can be $\sim 10^8$). The single most dramatic consequence is that the molecule becomes flexible. We normally think of the relative motion of atoms within a small molecule, say CO_2, in terms of vibrational modes. A polymer, however, can actually bend like a string! There are more consequences. Perpendicular to the strong (covalent) forces along the one-dimensional backbone, weaker forces may come into play; forces that would be insignificant if the atoms were not brought together by the backbone bonds. But given that the backbone forces these monomers together, the cooperative

Polymerization

$$CH_2 = CH_2 + CH_2 = CH_2 \longrightarrow CH_3 - CH_2 - CH = CH_2$$

Polycondensation

$$R_1 - \overset{\overset{\text{O}}{\|}}{C} - OH + OH - R_2 \longrightarrow R_1 - \overset{\overset{\text{O}}{\|}}{C} - O - R_2$$
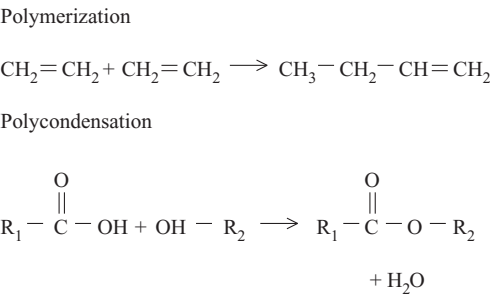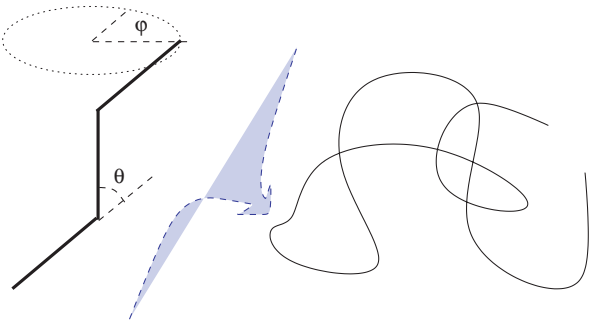
$$+ H_2O$$

Figure 2.3. How polymers are formed.

Figure 2.4. One mechanism for polymer flexibility: bond rotations.

binding of many of these weaker forces, both within the same molecule and between different molecules, allows the enormous number of specific interactions found in the biological world.

In this chapter we will study the simplest polymers, consisting of many identical monomers ("homopolymers"). This allows us to gain insight into the interplay between the one-dimensional polymer backbone and the possible three-dimensional conformations of the molecule.

Polymers are formed by polymerization (e.g. polyethylene) or by polycondensation (e.g. polypeptides); see Fig. 2.3. The single most important characteristic of polymers is that they are flexible. The simplest mechanism for their flexibility comes from rotations around single bonds. Figure 2.4 shows three links of, say, a polyethylene chain; the C atoms are at the vertices and the segments depict the C–C bonds. The bond angle θ is fixed, determined by the orbital structure of the carbon, but ϕ rotations are allowed. As a result, on a scale much larger than the monomer size a snapshot of the polymer chain may look as depicted on the right in the figure, i.e. a coil with random conformation. For other polymers, for example double-stranded DNA, the chemical structure does not allow bond rotations as in