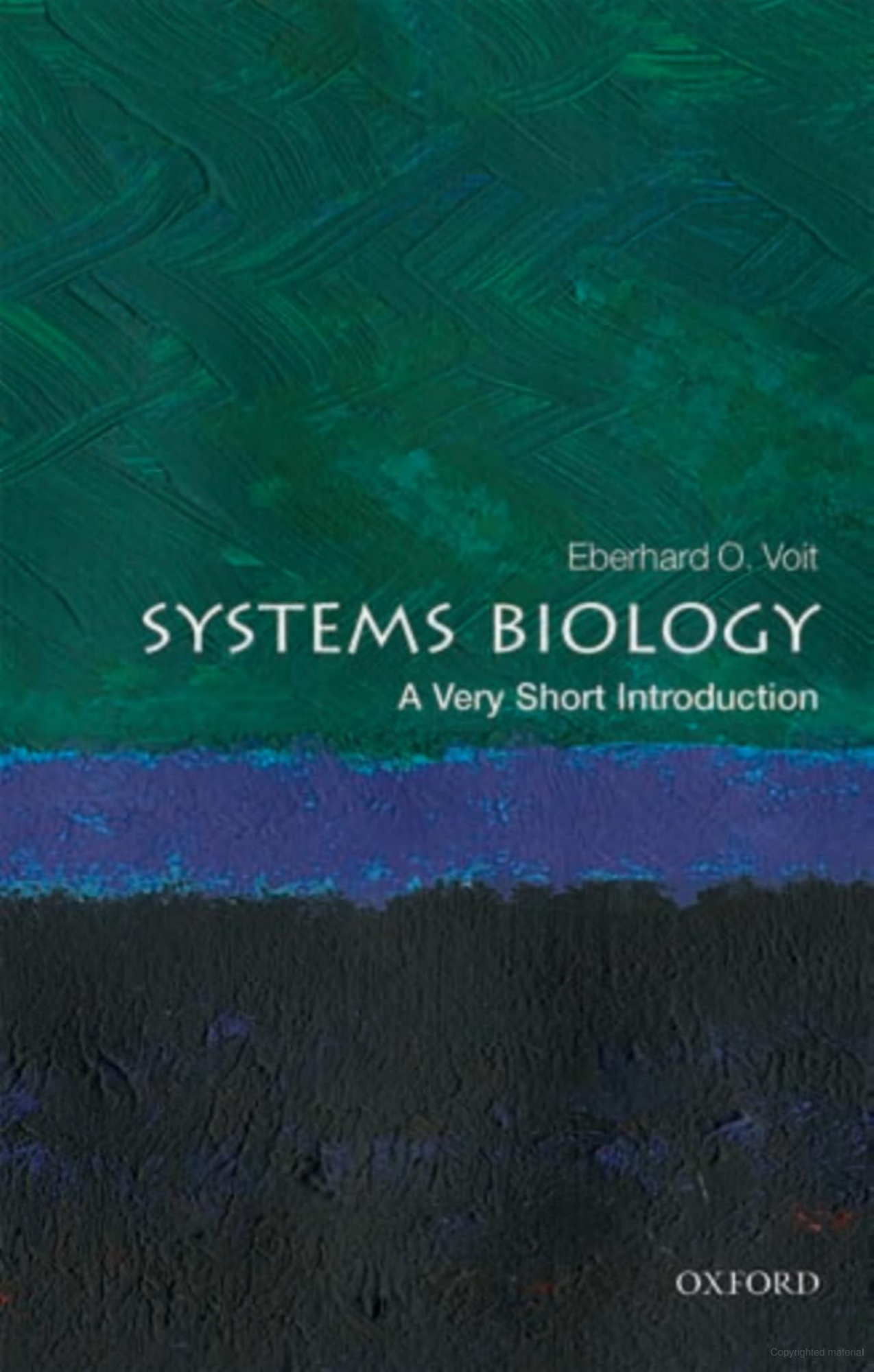


OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First edition published in 2020

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019957933

ISBN 978–0–19–882837–2

Printed in Great Britain by
Ashford Colour Press Ltd, Gosport, Hampshire

Contents

Acknowledgments

This little book on systems biology would not have become reality without the wisdom and vision of Oxford's Senior Commissioning Editor Latha Menon, who recognized the rising importance of systems biology. I was honored to receive her invitation to shape the impression of readers interested in this new field and am very grateful to her. My deep appreciation goes to several individuals who read a first draft and provided ample constructive feedback. I would like to thank, in particular, Alik Emelianov, a student at Westminster School in Atlanta, who ensured that the text can be understood by smart high schoolers. Similarly, I greatly appreciate Noune Sarvazyan's and Benedict Voit's numerous suggestions for improved readability of the first draft. My lab colleagues Jacob Davis, Luis Fonseca, and Carla Kumbale kept me on the straight and narrow with a number of loose scientific details. Last not least, Ann Voit very kindly read the whole volume carefully—three times!—in a truly dedicated effort to eradicate unusual formulations. A big thanks to all of them!

List of illustrations

Chapter 1
What is systems biology all about?

Biology is the study of life. Its goal is to gain comprehensive knowledge of how cells and organisms work, individually or in groups. Biology is interested in all aspects of life, from plants and animals to the rich ecosystems of rainforests and oceans, from genes and proteins to cells and organs, from bacteria and viruses to humans. Biological research relies on its own observations, but it has also learned much from chemistry and physics, whose applications to biology eventually became subspecialties in their own right, as the fields of biochemistry and biophysics.

Systems biology is a new specialty area that actually has exactly the same goals and purposes as general biology, namely, to understand how life works. But in contrast to traditional biology, systems biology pursues these goals with a whole new arsenal of tools that come from mathematics, statistics, computing, and engineering, in addition to biology, biochemistry, and biophysics. Systems biology utilizes these tools to determine the specific roles of the many different components that we find in living organisms, how these components interact with each other, and how they all collaborate to create and support life. These components are often molecules, but they may also be smaller, like atoms, electrons, and protons, or much larger, like mitochondria or the nucleus within a cell, cells themselves, organs, or complex mixtures like blood.

A prominent feature of systems biology is its heavy use of the enormous capacity of modern computers to store, manage, analyze, and interpret huge amounts of data. To appreciate why it is necessary to deal with so many data at the same time, let's look at the seemingly simple example of how we get nourishment from food. The digestive system is of course responsible for this process. It consists of distinct parts, all with their own, clearly defined purposes. The teeth and saliva allow us to chew off and swallow food, the stomach churns the food and secretes acid and enzymes that break down the food into its chemical components, and the small and large intestines take up the nutrients before the rest is excreted. This chain of events does not seem to be overly complicated. However, looking more closely, we quickly find that numerous other systems are required for these processes to work properly. Muscles need to move the food and nutrients throughout the digestive system. The cardiovascular system transports the nutrients from the intestines through the bloodstream to even the most remote areas of the body. Seeing or smelling food can make us hungry, so our eyes and nose are involved. The brain and the nervous system tell us when we are hungry and when we have eaten enough and, without our conscious participation, manage the well-coordinated activities of the oesophagus, stomach, and intestines. Moreover, there are the liver, pancreas, and kidneys that all deal with nutrients and the removal of unwanted materials. All of a sudden, digesting food involves a lot of systems that must work together and are tightly and perfectly coordinated to make the process feasible.

The situation is actually even more complicated, because every single one of these systems consists of a variety of subsystems. In particular, we have known for a long time that all organs and tissues contain cells, and these cells have turned out to be very complicated, finely tuned systems in themselves. They contain subsystems and these in turn contain subsystems, and if we step back for a moment, we realize that life is driven by hierarchies of interacting systems, like a whole society of different Russian

Matryoshka nesting dolls. Systems biology tries to understand how each of these systems works and how all of them, within and beyond their own realms, work together.

Just imagine how many different molecules and processes are involved in something like digestion, and you will see very quickly that we need computers that keep thousands of details 'in mind' and don't forget them. The power of computers is only one of the reasons that systems biology has all of a sudden appeared in the limelight, seemingly coming out of nowhere. A second important reason is a large repertoire of novel and different machines and technologies that permit biological experiments and laboratory analyses that were unthinkable even a few decades ago. As just one example, we can determine with a single measurement how much or how little the activity of any or all of our genes is changed in response to some stimulus, such as being cold or hungry.

More important than these technical and computational aspects is the fact that systems biology looks at the living world in a new way that is radically different from traditional biology. The main goal of biology throughout the past century has been the identification and characterization of as many building blocks of life as possible, and we have amassed enormous amounts of knowledge about genes, proteins, metabolites, and other fundamental components. For instance, just in the field of immunology, a new paper is published about every ten minutes; twenty-four hours a day, seven days a week. That is a lot of new information.

Although we have learned much about the molecular building blocks of life, there are still very many things we simply do not understand. Pressing examples include autoimmune and neurodegenerative diseases like rheumatoid arthritis and Alzheimer's disease, whose root causes and details of disease progression remain unclear, in spite of intensive research over many decades.

Systems biology readily recognizes that detailed knowledge of the molecules and structures of life is crucial, and that traditional research absolutely needs to continue, but it stresses the point that this knowledge alone is insufficient. In addition to looking for all individual parts of cells and organisms, we need methods of putting the parts back together. We need new ways of thinking and new methods of analysis that allow us to understand how the building blocks interact and what controls and regulates these interactions.

You may ask: Have we not always put data and information together? The answer is: *Yes* and *No*. *Yes*, we have always asked what different pieces of information collectively mean and whether we might find something new that we had not seen before. But we never had the tools to study huge numbers of observations simultaneously. As a consequence, we often did not even ask questions that modern systems biology attempts to address, especially if they involve many different pieces of information. To appreciate the slow but important changes in the ways biological research has been performed, let us briefly review how we got where we are today in biology and why systems biology adds a genuinely new perspective. How has the study of life changed over time, what do we know and understand, what exactly is missing, and what is yet to be discovered? A whirlwind tour through the evolution of biology will illustrate the roots and contexts from which the field of systems biology has emerged.

Biology is probably almost as old as humankind. Early forays into manipulating the living world around us included the domestication of wolves, which scientists think occurred between 20,000 and 30,000 years ago, as well as the beginnings of farming, maybe 12,000 years ago. As early as about 7,000 years ago, the first beer was brewed. Agriculture did not just mean planting and harvesting, it also required a basic understanding of seasons and life cycles, along with rudimentary record keeping of how much and how quickly the tribe grew or shrank, in order to

estimate how much food needed to be secured to survive the winter or other adverse conditions.

A very important driver of biological knowledge has always been medicine. Chinese and Egyptian medicine took root about 5,000 years ago, presumably based on animal observations and on trial and error with people. Much later, we know of Hippocrates (*c.*460–*c.*370 BC), the Greek 'father of medicine,' whom one could also consider an early systems biologist, because he treated the human body as one fully connected system. He taught that health was determined by the whole of an organism and its surroundings, including diet, lifestyle, family history, and environmental factors. Not having knowledge of chemistry, he proposed that health was the result of the right balance of four bodily fluids, called *humors*, and identified these as blood, black bile, yellow bile, and phlegm. In some sense, we have come full circle today, emphasizing the importance of lifestyle and a balanced diet for healthy living, while considering environmental exposures and a 'bad' family history as risk factors. However, in comparison with the present, Hippocrates' 'system biology' had no foundation in physics, chemistry, and biology, whereas we now have extensive information about the processes occurring inside the body, and the field of epigenetics is beginning to reveal the interactions between genes, environment, and lifestyle on a molecular level.

As the father of Western thinking, Aristotle (384–322 BC) created the foundation of the philosophy of science and had a special interest in biology. His approach was quite modern sounding: he proposed to investigate facts as opposed to religious beliefs, to determine whether structures observed in one organism, such as a bird, were also found in other birds and maybe even in fish, and what the reason for their existence might be. He was also the first documented scientist to propose what has become a popular cliché and is also a central theme in modern systems biology, namely, that the whole can be greater than the sum of its parts.

This notion of *synergism* ('working together') is a critical observation that transcends all of biology and will come up many times in this book.

After the demise of the Greek and Roman Empires, the medieval Islamic world engaged in astronomy, mathematics, and medicine, while the Western world entered the dark ages, where there were no major breakthroughs in the sciences, from what we know. Instead, mysticism and alchemy were on the rise, with their quest for a universal elixir for eternal youth and the goal of making gold out of mercury or lead. The Scientific Revolution of the 17th century brought real change with two great advances in biology. First, the microscope was invented and paved the way to an entirely new world of life: microbiology. For us today, it is hard to imagine a world without recognizing microbes, but nobody at the time had as much as an inkling of the existence of the incredible variety and complexity of this 'invisible world.' The second important novelty in the 17th century was a firm rule set for valid research, called the *scientific method*. This method demands that exact science follow well-defined steps of formulating a hypothesis, testing this hypothesis with experiments, analyzing the results, and, based on the results and their interpretation, formulating a new hypothesis. The scientific method is still a cornerstone of research today, although there are some recent alternative strategies that we will discuss in Chapters 3 and 4. During the 18th century, enlightenment, rational thought, and science had finally displaced alchemy. Biology in the 19th century was dominated by physiology, which focused on the various bodily systems governing health and disease, such as the nervous system, the cardiovascular system, the digestive system, and so on. Clearly, the physiology of the 19th century was a direct precursor of systems biology.

Moving into the 20th century, and pursuing the concepts of physiology, an old notion of the philosopher René Descartes re-emerged, which asserted that living organisms were merely

complicated machines. Specifically, this assertion suggested that to understand an organism, one had to understand its organs; to understand these, one had to understand tissues and cells; and understanding cells required knowledge of all molecular components. This strategy of increasingly more detailed investigation is called *reductionism*, because it attempts to reduce life all the way down to its ultimate building blocks. The Holy Grail of reductionism is the identification and characterization of all parts of a cell, collectively called its *molecular inventory*. The quest for identifying this inventory ushered in the field of molecular biology, which became the undisputed biological highlight of the 20th century and constitutes the second important root of today's systems biology, complementing physiology. Scientists learned about DNA and its fundamental role in genetics, and invented ever more sophisticated methods to characterize genes, proteins, and metabolites (natural chemical compounds). In parallel, methods like electron microscopy and molecular imaging began to make structures inside cells visible with a resolution unbelievable even a few decades ago. There is absolutely no doubt that the strategy of reductionism has been extremely successful.

A major breakthrough in molecular biology, with manifestations in many areas, happened around the year 2000: it became possible to parallelize and automate many measurement processes that had formerly required a lot of effort. At the time, the scientific community was already in the process of sequencing essentially all genes in a number of organisms, including humans. And since the genes are the carriers of our own blueprint and contain much of the information that makes us who we are, a big piece of the molecular inventory puzzle had been solved. With the new techniques, and heavily supported by robots, it became possible to measure to what degree specific genes were actually expressed in a sample of cells. Other impressive *high-throughput* methods, like mass spectrometry, flow and mass cytometry, as well as molecular imaging, began to allow the characterization of very large

inventories of proteins and metabolites, eventually even in single cells. Chapter 3 presents more details regarding these incredible advances. The collective result today is that we can routinely identify thousands of expressed genes, proteins, and metabolites even from tiny biological samples. Given time and resources, there is a good chance that we will soon have identified, quantified, and characterized a very significant portion of all types of molecules in many cell types. The Holy Grail of reductionism is emerging at the horizon.

Are all problems solved then? The answer is a resounding *No!* Just knowing all parts certainly does not mean that we understand what they do, how they play their unique roles, and what coordinates and controls them. Just imagine someone had taken the engine of a modern car apart and presented you with all its pieces. Would you be able to rebuild the engine? Probably not. But consider this: While a car engine consists of a lot of parts, a cell easily has tens of thousands of times as many, and a lot of them we do not even know yet!

So, biology is complicated, and we cannot even make reliable predictions of how a lowly bacterium will respond if we put it into a new environment. Why exactly is that? There are many reasons, which certainly are associated with the sheer numbers of different molecules, but also with their chemical features and physical structures, with uncounted processes running simultaneously in all cells, and with the complexity of interactions between all these components. This realization of complexity and unreliable predictability leads us directly into the realm of systems biology. Systems biology has begun to address these issues with new combinations of methods, including mathematics, computing, and engineering. Systems biology is still in its infancy, and that makes it a very exciting area of research. We have taken but the first steps of a long and very exhilarating path that is opening up in front of us. And we can only dream where this path may lead us.

Chapter 2
Exciting new puzzles

Systems biologists want to understand how biological systems operate within their natural surroundings. These 'systems' may be whole cells, organisms, or even populations, but they are more often comprised of biological molecules and their interactions. Because even seemingly simple systems in biology are in truth complicated, it is no surprise that investigating them together with their surroundings often means that an enormous amount of different types of information is involved. And as one might expect, this information is almost never complete. In fact, the typical situation is a collection of important core data, interspersed in a sea of gaping holes.

Systems biology has two closely interacting branches, and they address this issue in different, complementary ways. *Experimental systems biologists* use many different types of laboratory techniques to fill some of the holes with new measurements, while *computational systems biologists*, or *systems modelers* for short, depend on mathematics and computing to *infer* what is most likely happening in the holes. We will take a detailed look at both approaches in the next chapters, but because the general concepts of experimentation are relatively intuitive, I'll outline here why we need computational approaches. The answer, in one word, is 'complexity.'

The challenge of complexity

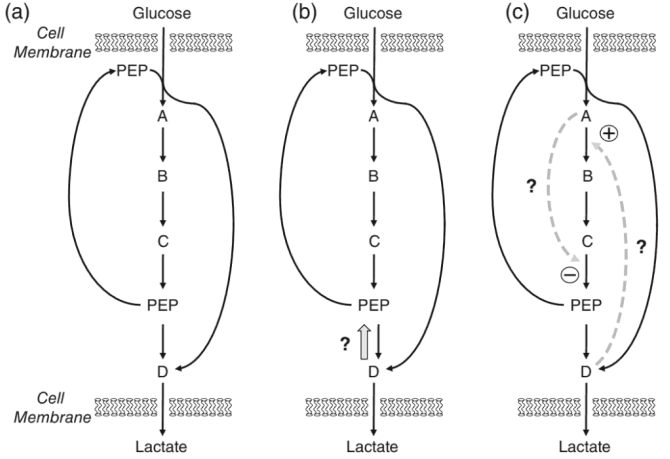
The overall task of the systems modeler is to extract information from available data and then to piece this information together, thereby generating genuinely new insights and narrowing gaps in knowledge. In many ways, one could compare the process with a huge Sudoku, where given numbers in certain locations permit true inferences regarding the missing numbers, if one just works on the puzzle hard enough. The big difference between Sudoku and computational systems biology (CSB) is that systems modelers usually don't even know whether the available information is actually sufficient to solve the puzzle, or at least parts of it. Then again, discovering even a partial solution for the first time can be exhilarating, and not knowing beforehand is therefore tantalizing and very exciting for true problem solvers. Moreover, even if one does not find a comprehensive solution, there can still be a meaningful advance if the computational analysis can identify which specific pieces of missing information would be most helpful in propelling the field forward, either by designing new experiments or collecting data from the literature or appropriate databases. The overriding challenge of this endeavor is the complexity of biological systems. I will not venture into trying to give a precise definition of complexity, but the following illustrations will hopefully create a good impression of what it is and why it challenges our minds.

Two situations are arguably of greatest interest to the systems modeler. The first situation deals with complexity: a system is so complicated that we cannot wrap our heads around it and it is therefore impossible to make reliable predictions on how this system would respond if something internal or in its surroundings were changed. The second situation consists of confusing or counterintuitive observations that traditional biology cannot explain. This situation can be found not only in large systems, but also in rather small ones.

A simple, yet true case study may illustrate this situation. The star of the story is the bacterium *Lactococcus lactis*, which is used in yoghurts, cheeses, fermented products, and other applications. *Lactococcus* is actually so important for the dairy industry that the US State of Wisconsin declared it in 2010 its State Bacterium, the first of its kind. The favorite food of this bacterium is the sugar glucose, which it takes up for energy and converts into lactate, an important ingredient in cheese. The conversion of glucose into lactate is the result of several biochemical reactions that collectively form a *pathway* called *glycolysis*, which translates into 'the break-down of glucose.' Humans also employ glycolysis for energy production, using the common molecule ATP (adenosine triphosphate) for the first reaction. *Lactococcus* instead uses a compound called PEP, which is short for phosphoenolpyruvate. The interesting aspect for our case study is that PEP itself is a molecule that is generated during glycolysis. Somewhat simplified, the pathway of glycolysis in *Lactococcus* is depicted in Figure 1(a). The bacterium takes up glucose by moving it across the cell membrane and immediately turning it into compounds A and D, using PEP in the process. Afterwards, A is converted into B, then C, which becomes PEP, from which D is produced. D is converted into lactate, which leaves the cell. So, that looks easy enough.

Now the puzzle begins. Imagine that glucose runs out, which happens frequently in nature. The obvious consequence is that no more A, B, C, D, and PEP can be produced, and all material flows through the system like through a chain of pipes and ends up as lactate. Now suppose that, after a while, glucose becomes available again, and *Lactococcus* would of course like to take it up, but it has a dilemma: all PEP is used up, and without PEP, no A can be produced. Ever again. If you don't think that sounds right, you are absolutely correct: After all, the bacterium has been around for a very long time. So, what's missing?

The way a systems modeler approaches the issue is to think about possible mechanisms with which the bacterium could possibly



1. Glycolytic pathway in the dairy bacterium *Lactococcus lactis*. (a) The bacterium takes up glucose from its surroundings across its cell membrane, and with the help of the compound PEP, converts it ultimately into lactate. (b) Could it be that there is a reverse reaction from D to PEP? (c) Would an inhibition signal from A or an activation signal from D solve the puzzle discussed in the text?

remedy the situation. Any promising possibility is formulated as a specific hypothesis and converted into a computational model, which typically consists of specifically chosen equations. We will discuss this process in Chapter 4 in greater detail. Now the modeler performs mathematical and computational tests with this model. Together, these tests are designed to show whether the hypothesis, if true, permits *Lactococcus* to take up glucose again. If the model confirms that the hypothesis is a possible solution to the problem, the systems modeler presents the hypothesis to an experienced microbiologist for validation. If not, the search goes on for better hypotheses or for additional data.

For example, one could think that some of the compound D could be converted back into PEP (Figure 1(b)). The argument could be that, while the diagram in Figure 1(a) does not exhibit this option,