CELL BIOLOGY by the numbers

CELL BIOLOGY by the numbers

Ron Milo
Rob Phillips

illustrated by
Nigel Orme

Garland Science
Vice President: Denise Schanck
Senior Editor: Summers Scholl
Senior Editorial Assistant: Katie Laurentiev
Production Editor: Natasha Wolfe
Cover Designer and Illustrator: Nigel Orme
Copyeditor: John Murdzek
Typesetting: Cenveo, Inc.
Proofreader: Elephant Editorial Services
Indexer: Bill Johncocks

Chapter Opening Images:
Chapter 1: Courtesy of Shutterstock.com
Chapter 2: Courtesy of Mark van Coller
Chapter 3: Courtesy of George Karbus
Chapter 4: Courtesy of Clark Little Photography
Chapter 5: Courtesy of Sean Davey
Chapter 6: Courtesy of Shutterstock.com

ISBN 978-0-8153-4537-4

The Library of Congress Cataloging-in-Publication data:

Milo, Ron, author.
Cell biology by the numbers / Ron Milo, Rob Phillips.
p. ; cm.
ISBN 978-0-8153-4537-4
[DNLM: 1. Cells–ultrastructure. 2. Cellular Structures. 3. Data Interpretation, Statistical. QU 350]
QH585.5.D38
572—dc23
2015033337

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 270 Madison Avenue, New York NY 10016, USA, and 2 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Garland Science
Taylor & Francis Group

Visit our website at http://www.garlandscience.com

This book is dedicated to our parents, Ada & Igal and Lee & Bob.

Table of Contents

Detailed Table of Contents

List of Estimates

Preface

> *I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.*
>
> **William Thomson (Lord Kelvin)** [Popular lectures and addresses, Vol. 1, Electrical Units of Measurement, 1883]

Though Lord Kelvin was unaware of the great strides that one can make by looking at bands on gels without any recourse to numbers, his exaggerated quantitative philosophy focuses attention on the possible benefits of bio logical numeracy.

One of the great traditions in biology's more quantitative partner sciences, such as chemistry and physics, is the value placed on centralized, curated quantitative data. Whether thinking about the astronomical data that describes the motions of planets or the thermal and electrical conductiv
ities of materials, the numbers themselves are a central part of the factual and conceptual backdrop for these fields. Indeed, often the act of trying to explain why numbers have the values they do ends up being an engine of discovery.

In our view, it is a good time to make a similar effort at providing defini tive statements about the values of key numbers that describe the lives of cells. One of the central missions of our book is to serve as an entry point that invites the reader to explore some of the key numbers of cell

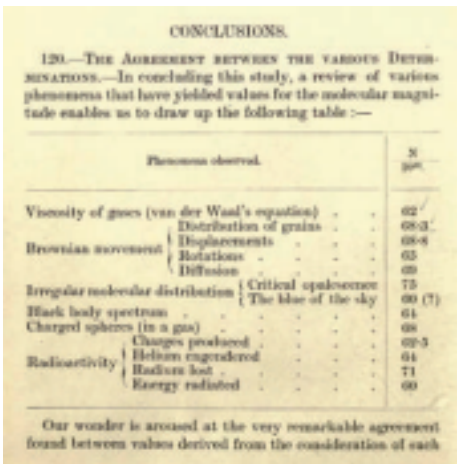
biol ogy. We hope to attract readers of all kinds—from seasoned researchers, who simply want to find the best values for some number of interest, to beginning biology students, who want to supplement their introductory course materials. In the pages that follow, we provide a broad collection of vignettes, each of which focuses on quantities that help us think about sizes, concentrations, energies, rates, information content, and other key quantities that describe the living world.

However, there is more to our story than merely providing a compendium of important biological numbers. We have tried to find a balance between presenting the data itself and reasoning about these numbers on the basis of simple estimates that provide both surprises and sanity checks. With

each vignette, we play with the interaction of two mindsets when thinking about cell biology by the numbers. First, we focus on trying to present in one place the relevant numbers for some particular biological structure or process. A second thrust is to "reason out" the numbers—to try and think about what determines their values and what the biological repercussions of those numbers might be. We are inspired by the so-called "Fermi prob
lems" made famous as a result of the simple estimates made by Enrico Fermi on subjects ranging from the number of piano tuners in a large American city to the advantages of having double windows for thermal insulation in winter. We were interested in the extent to which it is possi
ble to gain insights from a Fermi-inspired order-of-magnitude biology in which simple order-of-magnitude estimates serve as a sanity check on our understanding of biological phenomena.

When our hypothetical readers page to an entry of interest, be it the rate of translation or the number of genes in their favorite organism, we hope to greet them with a vignette that is at once entertaining and surprising. Rather than a dry elucidation of the numbers, as captured in our many tables, we use each vignette as a chance to tell some story that relates to the topic in question. We consider our book to be a quantitative companion to classic textbooks on molecular and cell biology and a source of enrichment for introductory and advanced courses. We thus aim to supply a quantitative component, which we consider an important complementary way of organ izing and viewing biological reality. We think that knowing the measure of things is a powerful and different way to get a "feel" for the organisms and their inner life.

Another reason for writing this book emerged from our own research. We often want to do "quick-and-dirty" analyses to estimate time scales, rates, energy scales, or other interesting biological parameters as a sanity check to see if some observation or claim makes sense. The issue is how to make it quick. Looking for key biological numbers using the internet or flipping through



CONCLUSIONS.

120.—The Agreement between the various Determinations.—In concluding this study, a review of various phenomena that have yielded values for the molecular magnitude enables us to draw up the following table :—

Phenomena observed.	N
Viscosity of gases (van der Waal's equation)	62
Brownian movement: Distribution of grains	68·3
Brownian movement: Displacements	68·8
Brownian movement: Rotations	65
Brownian movement: Diffusion	69
Irregular molecular distribution: Critical opalescence	75
Irregular molecular distribution: The blue of the sky	60 (?)
Black body spectrum	64
Charged spheres (in a gas)	68
Radioactivity: Charges produced	62·5
Radioactivity: Helium engendered	64
Radioactivity: Radium lost	71
Radioactivity: Energy radiated	60

Our wonder is aroused at the very remarkable agreement found between values derived from the consideration of such

textbooks is laborious at best and often futile. It is a common experience that even after hours of searching, we are left either with no result at all or a value with no reference to the experimental conditions that gave rise to that number, hence providing no sense of either the uncertainty or variability in the reported values. Our aspirations are for a biology that can boast the same kind of consistency in its data as is revealed in **Figure P-1**, which shows how in the early twentieth century

Figure P-1 The many measurements of Avogadro's number. The French physicist Jean Perrin in his book *Atoms* noted the broad diversity of ways to determine "atomic dimensions" and was justly proud of the consistent picture of the world that emerged from such different approaches.

a host of different methods yielded a surprisingly consistent set of val ues for Avogadro's number. Often in biology we are not measuring spe cific physical constants such as Avogadro's number, nevertheless, when measuring the same quantity under identical conditions we should find similar results. One of the points that will come up again in the first chapter is that reproducibility is required first as the basis for recogniz ing regularities. Then, once scientists are confident in their regularities, it becomes possible to recognize anomalies. Both regularities and anom alies provide a path to new scientific discoveries.

Our vision is that we need a sort of a "cheat sheet" for biology, just like those we got in high school for physical and chemical constants. We hope this book will serve as an extended cheat sheet or a brief version of the handbooks of the exact sciences—namely, those used prevalently in engineering, physics, and so on. Marc Kirschner, the head of the Sys
tems Biology department at Harvard University, compared doing biol ogy without knowing the numbers to learning history without knowing geography. Our aim is that our readers will find this book to be a useful atlas of important biological numbers with allied vignettes that put these numbers in context.

We are well aware that the particular list of topics we have chosen to con sider is subjective and that others would have made different choices. We limited our vignettes to those case studies that were consistent with our mutual interests and to topics where we felt we either knew enough or could learn enough to make a first pass at characterizing the state of the art in quantifying the biological question of interest.

The organization of the various numbers in the pages that follow is based upon roughly five different physical axes rather than biological context. First, we provide a narrative introduction to both the mindset and meth ods that form the basis for the remainder of the book. We offer our views on why we should care about the numbers described here, how to make back-of-the-envelope estimates, and simple rules on using significant dig its in writing out numbers. We then begin the "by-the-numbers" survey in earnest by examining the sizes of things in cell biology. This is followed by a number of vignettes whose aim is to tell us how many copies of the vari ous structures of interest are found.

Taking this kind of biological census is becoming increasingly important as we try to understand the biochemical linkages that make up the many pathways that have been discovered in cells. The third axis focuses on force and energy scales. The rates of pro cesses in biology form the substance of the fourth section of the book, fol lowed by different ways of capturing the information content of cells. As is often the case in biology, we found that our human effort at rational cate gorization did not fit nature's appetite for variety, and thus the last section

is a biological miscellany that includes some of our favorite examples that defy inclusion under the previous headings.

Unexpectedly to us, as our project evolved, it became ever more clear that there is a hierarchy of accuracy associated with the determination of the numbers we describe. For example, our first chapter deals with sizes of components in the cell, a relatively accurate and mature outgrowth of modern structural biology with its many different microscopies. Our sec
ond chapter on the cellular census ramps up the difficulty, with many of the numbers we report coming from very recent research literature, some of which show that calibrations of different methods, such as fluorescence techniques and those based upon antibodies, are not entirely consistent. Chapter 3, which deals with energy scales of various processes within the cell, suffers from challenges as severe as ambiguities in the definition of the quantities themselves. We thought hard about how to represent in writing the uncertainties associated with values that we collected from the literature. The guidelines we follow regarding how many significant digits to use are summarized in the opening chapter. It is our hope that attention to this issue of quantitative sanitation will become the norm among stu dents and researchers in biology.

Inspiration for the approach taken here of "playing" with the numbers has come from many sources. Some of our favorites, which we encourage our readers to check out, include: *Guesstimation* by Lawrence Weinstein and John Adam; John Harte's two books, *Consider a Spherical Cow* and *Consider a Cylindrical Cow*; Richard Burton's *Physiology by Numbers* and *Biology by Numbers*; *Why Big Fierce Animals Are Rare* by Paul Colinvaux; and Sanjoy Mahajan's fine books, *Street Fighting Mathematics* and *The Art of Insight in Science and Engineering: Mastering Complexity*. We are also big fans of the notes and homeworks from courses by Peter Goldreich, Dave Stevenson, and Stirl Phinney on "Order of Magnitude Physics." What all of these sources have in common is the pleasure and value of playing with numbers. In some ways, our vignettes are modeled after the examples given in these other books, and if we have in some measure succeeded in inspiring our readers as much as these others have inspired us, our book will be a success.

Acknowledgments

One of the great pleasures of writing a book such as ours is the many stim ulating and thoughtful interactions we have had with our colleagues. The most important such interaction was our long-standing collaboration with our illustrator Nigel Orme. We hope that it will become clear to our read ers that by thinking hard about how to visually represent some biological process, we can actually make discoveries about how living systems work. This is similar in spirit to an observation made by Darwin in his autobi ography—namely, by formulating our thoughts about systems mathemat ically we can see things that are completely invisible when couched in verbal language. When we couch our thoughts in careful visual language, it also reveals new ways of understanding biology. Nigel Orme has made us clearer thinkers, better communicators, and has helped us understand more about the world we live in. We are also deeply grateful to Uri Moran, who showed outstanding devotion in searching a wide range of scientific resources and scouring the original literature to find the numbers essential for making this book scientifically meaningful. He never tired of the many iterations we made both of individual vignettes and entire book drafts, and of guiding us through his wide knowledge and peaceful manner.

A number of hearty souls actually read the entire book from cover to cover and provided insights, suggestions, criticisms, and encouragements. This group consisted of Olaf Andersen, David Bensimon, Maja Bialecka-Fornal, Ian Booth, Dennis Bray, Rachel Brockhage, Avantika Lal, Molly Phillips, Rahul Sarpeshkar, and Carolina Tropini. We are deeply grateful to all of them for their insights, which both changed and improved our book sub stantially and from which we have learned so much. Others who dipped widely into our manuscript, making a host of important comments, include Niv Antonovsky, Wilfredo Ayala-López, Uri Barenholz, Bill Broad hurst, Robert Endres, Arren Bar Even, Wayne Fagerberg, Shai Fuchs, Akira Funahashi, Noriko Hiroi, Ian Kay, Stefan Klumpp, Tetsuya Kobayashi, Jenny Koenig, Patricia Marsteller, Stephen Martin, Steve Quake, Doug Rees, Maya Shamir, and Zhanchun Tu. Of course, it should be clear that mistakes in both the truth or logic of what we say is our fault and not theirs.

Much of the work presented here has been featured in a number of differ ent courses that the two of us have taught. Many students gave us motiva tion and feedback, and their willingness to serve as guinea pigs while we developed the ideas presented here is very much appreciated. They were members of courses focusing on the subject of Cell Biology by the Num bers given at the Weizmann Institute of Science, Caltech, and Cold Spring Harbor Laboratory in the years 2009–2015.

The two of us have been deeply lucky to have had the chance to work

together over these last seven years. We are grateful to Elliot Meyerowitz, who was the matchmaker and who saw the potential for like-minded souls to have a scientific collaboration, despite their 12,000-km separation and 10-hour time difference. Jon Widom also served as a constant source of support and insight during the incipient phases of our project.

Another source of extremely important support whose impact cannot be underestimated is the financial support that it takes to carry out such an enormous undertaking as the BioNumbers website and our book that grew from it. The NIH Director's Pioneer Award on one side of the Atlantic and the European Research Council on the other gave the essential sup
port to the labs of the authors that made this book possible. The Donna and Benjamin M. Rosen Center for Bioengineering at Caltech, under the excellent leadership of Niles Pierce and Frances Arnold, has been a source of generous and constant support, which has funded many critical parts of this project and which has made it possible for us to provide a copy of the book free of charge for those who are less fortunate than well-supported researchers. The development of the BioNumbers website itself was made possible through the vision, generosity, and expertise of Paul Jorgensen, Marc Kirschner, Daniel Kreiter, Ben Marks, Mason Miranda, Uri Simhony, Mike Springer, and Becky Ward. In addition, we are deeply grateful to the team at Garland Science who has joined us in this adventure, with special thanks to our editor Summers Scholl, who has been a steady supporter of the project, to Natasha Wolfe, who expertly served as production editor, and to John Murdzek for improving the readability and accuracy of our book through his copy editing.

Our list of specific scientific advisors is long and distinguished. These gen erous friends provided us insights on everything from the size of synapses to the scaling of the nucleus size to the physical limits of the senses. We are grateful to Yaarit Adamovich, Noam Agmon, Ariel Amir, Shira Amram, Tslil Ast, David Baltimore, Naama Barkai, Scott Barolo, Mark Bathe, Nathan Belliveau, Danny Ben-Zvi, Howard Berg, Michael Brenner, James Briscoe, Jan Brugues, Tabita Bucher, Richard Burton, Michael Clague, Bil Clemons, Adam Cohen, Athel Cornish-Bowden, Peter Curtin, Dan Davidi, Avigdor Eldar, Yuval Eshed, Daniel Fisher, Avi Flamholtz, Sarel Fleishman, Dan Fletcher, Michael Gliksberg, Dani Gluck, Matthew Good, Talia Harris, Yuval Hart, Elizabeth Haswell, Peter Henderson, Hermann-Georg Holzhütter, Joe Howard, KC Huang, Shalev Itzkovitz, Grant Jensen, Daniel Koster,

Eric Lander, Michael Laub, Heun Jin Lee, Avi Levy, Wendell Lim, Jen nifer Lippincott-Schwartz, Jeff Marlow, Wallace Marshall, Sarah Marzen, Markus Meister, Joe Meyerowitz, Lishi Mohapatra, Phil Mongiovi, Daniel Needelman, Ronny Neumann, Tom O'Halloran, Victoria Orphan, Linnaea Ostroff, Michal Perach, Matthieu Piel, Tzachi Pilpel, Rami Pugatch, Tobias Reichenbach, Joel Rothman, Michael Roukes, James Saenz, Dave Savage, Elio Schaechter, Declan Schroeder, Oren Schuldiner, Maya Schuldiner, Eran Segal, Pierre Sens, Joshua Shaevitz, Tom Shimizu, Tomer Shlomi, Alex Sigal, Thomas Simmen, Victor Sourjik, Todd Squires, Lucia C. Strader,

Lubert Stryer, Arbel Tadmor, Amos Tanay, Dan Tawfik, Julie Theriot, Elitza Tocheva, Katja Tummler, Antoine van Oijen, Alexander Varshavsky, Doug Weibel, Talia Weiss, Irving Weissman, Jeremy Young, Yossi Yovel, and Jesse Zalatan.

We would like to thank the following societies and publishers for grant ing permission to adapt or republish material appearing in the following figures: The American Association for the Advancement of Science for Figures 1-16, 2-7, 2-9, 2-18, 2-22, 2-24, 2-27, 3-18, 4-12, 4-37, 4-44, 5-4, and 5-10. The American Chemical Society for Figure 1-25. The American Institute of Physics for Figure 3-2. The American Physical Society for Figure 3-14. The American Society for Biochemistry and Molecular Biology for Figure 2-14. The American Society for Cell Biology for Figures 1-19 and 1-23. The American Society for Microbiology for Figures 1-14, 2-34, and 4-13. The Association for Research in Vision and Ophthalmology for Figure 2-37. The Bibliothèque de Genève for Figure 2-2. Elsevier Inc. for Figures 1-2, 1-5, 1-22, 1-26, 1-47, 2-21, 2-31, 2-35, 2-42, 4-2, 4-4, 4-14, 4-16, 4-24, 4-31, 4-34, 4-36, 4-38, 4-39, 4-46, 5-12, 6-2, and 6-9. John Wiley & Sons for Figures 1-12, 2-38, 2-43, 3-13, 4-18, 4-27, and 4-32. Macmillan Publishers Ltd. for Figures 1-6, 1-24, 1-29, 1-31, 2-13, 2-16, 2-19, 2-20, 2-29, 3-5, 3-10, 3-17, 4-30, 4-41, 5-8, 5-9, 6-3, and 6-4. The National Academy of Sciences for Figures 1-3, 1-20, 1-41, 2-6, 2-10, 2-15, 2-17, 2-36, 3-9, 3-19, and 6-6. The Royal Society of Chemistry for Figure 4-17. The Rockefeller University Press for Figure 1-46. The Society for General Microbiology for Figure 2-11. Springer Science + Business Media for Figures 1-21 and 2-12. The Trustees of the Natural History Museum, London for Figure 1-15.

In many ways that go far beyond the science, we were supported by Yoram Adler, Uri Alon, Efrat Binya Talmi, Efrat Cohen, Hernan Garcia, David Goodsell, Alon Gildoni, Alex Kizin, Jane Kondev, Ronen Levi, Yael Milo, Elad Noor, Daniel Ramot, Richard Sever, and Julie Theriot.

It is a long-held tradition that book acknowledgments end with heartfelt thanks to families. Of course, there is a reason for this and that is this: Though the lives of those who write books are so deeply enriched by the act of writing, they always know that there is something that matters much

more, and that is our loved ones. Both of us value our families above all and are deeply grateful to our spouses, Hilla and Amy, our parents, Ada and Igal Milo and Lee Cudlip and Bob Phillips, and our children, Geffen, Yaara, and Rimon and Molly and Casey, for bringing magic to our lives.

The Path to Biological Numeracy

> *…[I]n after years I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics, for men thus endowed seem to have an extra sense.*
>
> [**Charles Darwin,** Autobiography, 1887]

WHY WE SHOULD CARE ABOUT THE NUMBERS

This introduction sets the stage for what is to unfold in upcoming chapters. If you feel the urge to find some number of interest now, you can jump to any vignette in the book and come back later to this chapter, which presents both the overall logic and the basic tools used to craft biological numeracy. Each of the $\approx 10^2$ vignettes in the book can be read as a stand alone answer to a quantitative question on cell biology by the numbers. The formal structure for the remainder of the book is organized according to different classes of biological numbers, ranging from the sizes of things (Chapter 1) to the quantitative rules of information management in living organisms (Chapter 5) and everything in between. The goal of this section of the book is decidedly more generic, laying out the case for biological numeracy and providing general guidelines for how to arrive at these num bers using simple estimates. We also pay attention to the question of how to properly handle the associated uncertainty in both biological measure ments and estimates. We build on the principles developed in the physical sciences, where estimates and uncertainties are common practice, but in our case require adaptation to the messiness of biological systems.

What is gained by adopting the perspective of biological numeracy we have called “cell biology by the numbers”? The answer to this question can be argued along several different lines. For example, one

enriching approach to thinking about this question is by appealing to the many his
toric examples, where the quantitative dissection of a given problem is

what provided the key to its ultimate solution. Examples abound, whether from the classic discoveries in genetics that culminated in Alfred Sturte vant's map of the geography of the *Drosophila* genome or Hodgkin and Huxley's discoveries of the quantitative laws that govern the dynamics of nerve impulses. More recently, the sharpness of the questions, as formu lated from a quantitative perspective, has yielded insights into the limits of biological information transmission in processes ranging from bacte rial chemotaxis to embryonic development and has helped establish the nature of biological proofreading that makes it possible for higher fidelity copying of the genetic material than can be expected from thermodynam ics alone (some of these examples appear in a paper we wrote together[*]).

A second view of the importance of biological numeracy centers on the way in which a quantitative formulation of a given biological phenom enon allows us to build sharp and falsifiable claims about how it works. Specifically, the state of the art in biological measurements is beginning to reach the point of reproducibility, precision, and accuracy where we can imagine discrepancies between theoretical expectations and measure ments that can uncover new and unexpected phenomena. Further, bio logical numeracy allows scientists an "extra sense," as already appreciated by Darwin himself, to decide whether a given biological claim actually makes sense. Said differently, with any science, in the early stages there is a great emphasis on elucidating the key facts of the field. For example, in astronomy, it was only in light of advanced naked-eye methods in the hands of Tycho Brahe that the orbit of Mars was sufficiently well under stood to elucidate central facts, such as that Mars travels around the sun in an elliptical path with the sun at one of the foci. But with the maturity of such facts comes a new theoretical imperative—namely, to explain those facts on the basis of some underlying theoretical framework. For exam ple, in the case of the observed elliptical orbits of planets, it was an amaz ing insight to understand how this and other features of planetary orbits were the natural consequence of the inverse-square law of gravitation. We believe that biology has reached the point where there has been a suffi cient accumulation of solid quantitative facts that this subject, too, can try to find overarching principles expressed mathematically that serve as theory to explain those facts and to reveal irregularities when they occur. In the chapters that follow, we provide a compendium of such biological facts, often presented with an emphasis that might help as a call to arms for new kinds of theoretical analysis.

Another way to think about this quest for biological numeracy is to imagine some alien form coming to Earth and wishing to learn more about

[*] Phillips R & Milo R (2009) A feeling for the numbers in biology. *Proc Natl Acad Sci USA* 106:21465–21471.

what our society and daily lives look like. For example, if we could give the friendly alien a single publication, what such publication might prove most useful? Though different readers may come up with different ideas of their own, our favorite suggestion would be the report of the Bureau of Statistics, which details everything from income to age at marriage to level of education to the distributions of people in cities and in the country. The United Nations posts such statistics on their website: https://unstats. un.org/unsd/default.htm.

Hans Rosling has become an internet sensation as a result of the clever and interesting ways that he has found not only to organize data, but also to milk it for unexpected meaning. Our goal is to provide a kind of report of the bureau of statistics for the cell and to attempt to find the hidden and unexpected meaning in the economy and geography of the cell.

As an example of the kind of surprising insights that might emerge from this exercise, we ask our readers to join us in considering mRNA, the "blueprint" for the real workhorses of the cell, the proteins. Quickly, ask yourself: Which is larger, the blueprint or the thing being blueprinted? Our intuition often thinks of the blueprint for a giant skyscraper, and it is imme
diately obvious that the blueprint is but a tiny and flattened caricature of the building it "codes for." But what of our mRNA molecule and the protein it codes for? What is your instinct about the relative size of these two mol ecules? As we will show in the vignette entitled "Which is bigger, mRNA or the protein it codes for?" (pg. 43), most people's intuition is way off, with the mRNA molecule actually being substantially larger than the protein it codes for. This conclusion has ramifications, for example for whether it is easier to transport the blueprint or the machine it codes for.

Finally, we are also hopeful for a day when there is an increasing reliance in biology on numerical anomalies as an engine of discovery. As the meas urements that characterize a field become more mature and reproducible using distinct methodologies, it becomes possible to reliably ask the question of when a particular result is anomalous. Until the work of David Keeling in the 1950s, no one could even agree on what the level of CO_2 in the atmos phere was, let alone figure out if it was changing. Once Keeling was able to show the rhythmic variations in CO_2 over the course of a year, then questions about small overall changes in the atmospheric CO_2 concentration over time could be addressed. Perhaps more compellingly, Newton was repeatedly confounded by the 20% discrepancy between his calculated value for the speed of sound and the results from measurements. It was only many years later that workers such as Laplace realized that a treatment of the problem as an adiabatic versus isothermal process could explain that discrepancy. The recent explosion of newly discovered extrasolar planets is yet another exam
ple where small numerical anomalies are received with such confidence that they can be used as a tool of discovery. In our view, there is no reason at all to

believe that similar insights don't await those studying living matter once our measurements have been codified to the point that we know what is irregu lar when we see it. In a situation where there are factors of 100 distinguishing different answers to the same question, such as how many proteins are in an *E. coli* cell, there is little chance to discern even regularities, let alone having confidence that anomalies are indeed anomalous. Often, the great "effects" in science are named such because they were signaled as anomalous. For exam ple, the change in wavelength of an oncoming ambulance siren is known as the famed Doppler effect. Biochemistry has effects of its own, such as the Bohr effect, which is the shift in binding curves for oxygen to hemoglobin as a func tion of the pH. We suspect that there are many such effects awaiting discovery in biology as a result of reproducibly quantifying the properties of cells and then paying close attention as to what those numbers can tell us.

THE BIONUMBERS RESOURCE

As a reminder of how hard certain biological numbers are to come by, we recommend the following quick exercise for the reader. Pick a topic of par ticular interest from molecular or cell biology and then seek out the corre sponding numbers through an internet search or by browsing your favorite textbooks. For example, how many ribosomes are there in a human cell? Or, what is the binding affinity of a celebrated transcription factor to DNA? Or, how many copies are there per cell of any famous receptor, such as those of chemotaxis in bacteria or of growth hormones in mammalian cells? Our experience is that such searches are at best time-consuming, and often they are inconclusive or even futile. As an antidote to this problem, essentially all of the numbers presented in this book can be found from a single source— namely, the BioNumbers website (http://bionumbers.hms.harvard.edu/). The idea of this internet resource is to serve as an easy jumping-off point for accessing the vast biological literature in which quantitative data is archived. In particular, the data to be found in the BioNumbers database have been subjected to manual curation, have full references to the primary literature from which the data are derived, and provide a brief description of the method used to obtain the data in question.

As signposts for the reader, each and every time that we quote some num ber, it will be tied to a reference for a corresponding BioNumbers Iden tification (BNID). Just as our biological readers may be familiar with the PMID, which is a unique identifier assigned to published articles from the biological and medical literature, the BNID serves as a unique identifier of different quantitative biological data. For example, BNID 103023 points

us to one of several determinations of the number of mRNA per yeast

cell. The reader will find that both our vignettes and the data tables are filled with BNIDs, and by pasting this number into the BioNumbers website (or just Googling “BNID 103023”), the details associated with that particular quantity can be uncovered.

HOW TO MAKE BACK-OF-THE-ENVELOPE CALCULATIONS

The numbers to be found in the BioNumbers compendium and in the vignettes throughout this book can be thought of as more than simply data. They can serve as anchor points to deduce other quantities of interest and can usually be themselves scrutinized by putting them to a sanity test based on other numbers the reader may know and bring together by “pure thought.” We highly recommend the alert reader to try and do such cross tests and inferences. This is our trail-tested route to powerful numeracy. For example, in Chapter 4 we present the maximal rates of chromosome replication. But we might make an elementary estimate of this rate by using our knowledge of the genome length for a bacterium and the length of the cell cycle. Often such estimates will be crude (say, to within a factor of two), but they will be good enough to tell us the relevant order of mag nitude as a sanity check for measured values.

There are many instances in which we may wish to make a first-cut esti mate of some quantity of interest. In the middle of a lecture you might not have access to a database of numerical values, and even if you do, this skill of performing estimates and inferring the bounds from above and below as a way to determine unknown quantities is a powerful tool that can illu minate the significance of measured values.

One handy tool is how to go from upper and lower bound guesses to a concrete estimate. Let’s say we want to guess at some quantity. Our first step is to find a lower bound. If we can say that the quantity we are after is bigger than a lower bound x_L and smaller than an upper bound x_U, then a simple estimate for our quantity of interest is to take what is known as the geometric mean—namely,

$$x_{estimate} = \sqrt{x_L x_U}. \quad (0.1)$$

Though this may seem very abstract, in most cases we can ask ourselves a series of questions that allow us to guess reasonable upper and lower bounds

to within a factor of 10. For example, if we wish to estimate the length of an airplane wing on a jumbo jet, we can begin with “Is it bigger than 1 m?”. Yes. “Is it bigger than 5 m?” Yes. “Is it bigger than 10 m?” I think so, but I’m not sure. So, we take 5 m as our lower bound. Now the other end, “Is it smaller than 50 m?” Yes. “Is it smaller than 25 m?” I think so, but I’m not sure. So, we take 50 m as our upper bound.

Using 5 m and 50 m as our lower and upper bounds, respectively, we then estimate the wing size as [redacted] ≈ 15 m, the approximate square root of 250 m^2. If we had been a bit more bold, we could have used 10 m as our lower bound, with the result that our estimate for the length of the wing would be ≈22 m. In both cases we are accurate to within a factor of two compared with the actual value—that is, well within the target range of values we expect from "order-of-magnitude biology."

Let's try a harder problem, which will challenge the intuition of any one we know. What would you estimate is the number of atoms in your body? 10^{10} is probably too low—in fact, that sounds more like the num ber of people on Earth. 10^{20}? Maybe, but that vaguely reminds us of the exponent in Avogadro's number. 10^{80} sounds way too high, because such exponents are reserved for the number of atoms in the universe. 10^{40}? Maybe. So, [redacted] $\sim 10^{30}$. A more solid calculation is given later in the book using the Avogadro constant (can you see how to do it?), but it suffices to say that we are within about two orders of magnitude of the correct order of magnitude, and this is based strictly on educated guessing. We may object to pulling 10^{20} and 10^{40} out of thin air, but this is exactly the kind of case where we have extremely little intuition and thus have nothing to start with aside from vague impression. But we can still construct bounds by eliminating estimates that are too small and too large as we did above, and somewhat surprisingly, with the aid of the geometric mean, that takes us close to the truth. We probably have to try this scheme out several times to check if the advertised effectiveness actually works. The geometric mean amounts really to taking the normal arithmetic mean in log space (that is, on the exponents of 10). Had we chosen to take the normal mean on the values we guessed, our estimate would have been completely dominated by the upper bound we chose, which often leads to extreme overestimation.

One question worth asking is: How do we know whether our estimates are actually "right"? Indeed, often those who aren't used to making estimates fear of getting the "wrong" answer. In his excellent book, *Street Fighting Mathematics*, Sanjoy Mahajan makes the argument that an emphasis on this kind of "rigor" can lead, in fact, to mathematical "rigor mortis." The strategy we recommend is to think of estimates as successive approxima
tions, with each iteration incorporating more knowledge to refine what the estimate actually says. There is no harm in making a first try and getting a "wrong" answer. Indeed, part of the reason such estimates are worthwhile

is that they begin to coach our intuition so that we can *a priori* have a sense of whether a given magnitude makes sense or not without even resorting to a formal calculation.

As noted above, one of the most elusive, but important, skills is to be able to quickly and efficiently estimate the orders of magnitude associated with some quantity of interest. Earlier, we provided some of the conceptual rules that fuel such estimates. Here, we complement those conceptual rules with various helpful numerical rules that can be used to quickly find our way to an approximate but satisfactory assessment of some biological process of interest. We do not expect you to remember them all on first pass, but give them a quick look, and maybe a few of them will stick in the back of your mind when you need them.

Arithmetic sleights of hand

- $2^{10} \approx 1000$
- $2^{20} = 4^{10} \approx 10^6$
- $e^7 \approx 10^3$
- $10^{0.1} \approx 1.3$
- [redacted] ≈ 1.4
- [redacted] ≈ 0.7
- $\ln(10) \approx 2.3$
- $\ln(2) \approx 0.7$
- $\log_{10}(2) \approx 0.3$
- $\log_{10}(3) \approx 0.5$
- $\log_2(10) \approx 3$

Big numbers at your disposal

- Seconds in a year $\approx \pi \times 10^7$ (the approximate value of pi, a nice coincidence and an easy way to remember this value)
- Seconds in a day $\approx 10^5$

xxx

- Hours in a year $\approx 10^4$
- Avogadro's constant $\approx 6 \times 10^{23}$
- Cells in the human body $\approx 4 \times 10^{13}$

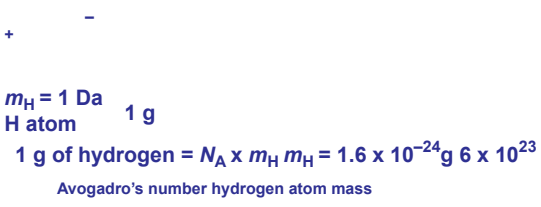
Rules of thumb

Just as there are certain arithmetical rules that help us quickly get to our order-of-magnitude estimates, there are also physical rules of thumb that can similarly extend our powers of estimation. We give here some of our favorites and you are most welcome to add your

own at the bottom of our list and also send them to us. Several of these estimates are represented pictorially as well. Note that here and throughout the book we try to follow the correct notation where "approximately" is indicated by the symbol ≈, and loosely means accurate to within a factor of two or so. The symbol ~ means "order of magnitude," so only to within a factor of 10 (or in a dif
ferent context it means "proportional"). We usually write approximately because we know the property value indeed roughly but to better than a factor of 10, so ≈ is the correct notation and not ~. In the cases where we only know the order of magnitude, we will write the value only as 10^x with
out extraneous significant digits.

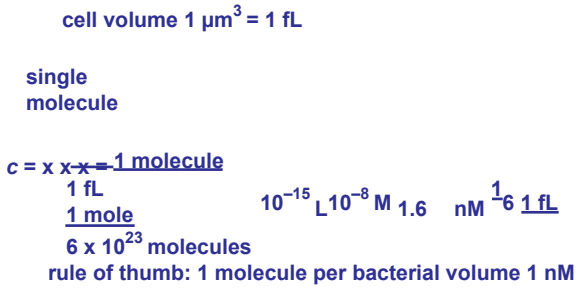
- 1 dalton (Da) = 1 g/mol ≈ 1.6×10^{-24} g (as derived in **Estimate 0-1**).
  - 1 nM is about 1 molecule per bacterial volume, as derived in **Estimate 0-2**, 10^1–10^2 per yeast cell, and 10^3–10^4 molecules per characteristic
  mammalian (HeLa) cell volume. For 1 μM, multiply by a thousand; for 1 mM, multiply by a million.
  - 1 M is about one per 1 nm^3.
- There are 2–4 million proteins per 1 μm^3 of cell volume.

Converting between daltons and grams



Estimate 0-1

Nanomolar in *E. coli* units



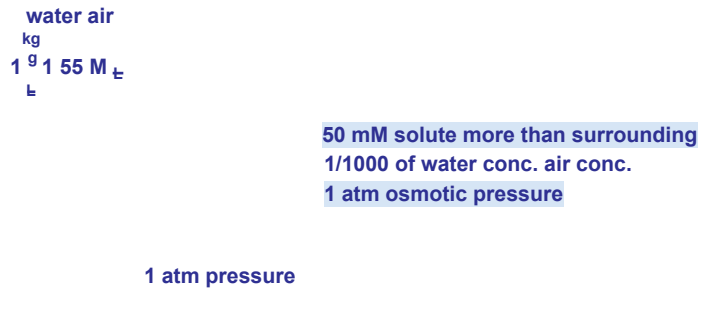
Estimate 0-2

- Concentration of 1 ppm (part per million) of the cell proteome is ≈ 5 nM.
- 1 mg of DNA fragments 1 kb long is ≈ 1pmol or ≈ 10^{12} molecules. •

Under standard conditions, particles at a concentration of 1 M are ≈ 1 nm apart.

- Mass of typical amino acid ≈ 100 Da.
- Protein mass [Da] ≈ 100 × Number of amino acids.
- Density of air ≈ 1 kg/m^3.
- Water density ≈ 55 M ≈ × 1000 that of air ≈ 1000 kg/m^3. • 50 mM osmolites ≈ 1 atm osmotic pressure (as shown in **Estimate 0-3**). • Water molecule volume ≈ 0.03 nm^3, (≈ 0.3 nm)^3.
- A base pair has a volume of ≈ 1 nm^3.
- A base pair has a mass of ≈ 600 Da.
- Lipid molecules have a mass of ≈ 500–1000 Da.
- 1 k_BT ≈ 2.5 kJ/mol ≈ 0.6 kcal/mol ≈ 25 meV ≈ 4 pN nm ≈ 4 × 10^{-21} J.
- ≈ 6 kJ/mol sustains one order of magnitude concentration difference [= k_BT ln(10) ≈ 1.4 kcal/mol].

Relating solute concentration to osmotic pressure



Estimate 0-3

- Movement across the membrane is associated with 10–20 kJ/mol per one net charge due to membrane potential.
- ATP hydrolysis under physiological conditions releases 20 k_BT ≈ 50 kJ/mol ≈ 12 kcal/mol ≈ 10^{-19} J.
- One liter of oxygen releases ≈ 20 kJ during respiration.
- A small metabolite diffuses 1 nm in ~ 1 ns.
- 1 OD_{600} ≈ 0.5 g cell dry weight per liter.
- There are ≈ 10^{10} carbon atoms in a 1 μm^3 cell volume.

RIGOROUS RULES FOR SLOPPY CALCULATIONS

One of the most important questions that all readers should ask them selves is: Are any of the numbers in this book actually “right”? What does it even mean to assign numbers to quantities such as sizes, concentra tions, and rates that are so intrinsically diverse? Cellular processes show immense variability, depending upon both the type of cell in question and the conditions to which it has been subjected.

One of the insights of recent years that has been confirmed again and again is that even within a clonal population of cells, there is wide cell-to-cell variability. Hence, both the diversity and intrinsic variability mean that the task of ascribing particular numbers to biological properties and processes is fraught with the danger of misinterpretation. One way to deal with this challenge is by presenting a range of values rather than "the value." Equally important, a detailed discussion of the environmental conditions under which the cells grew and when and how the measurement was taken and analyzed is in order. Unfortunately, this makes the discussion very cumbersome and is often resolved in textbooks and journals by avoiding concrete values alto
gether. We choose in this book to give concrete values that clearly do not give the "full" picture. We expect, and caution the reader to do the same, to think of them only as rough estimates and as an entry point to the lit erature. Whenever a reader needs to rely on a number for research rather than merely obtain a general impression, he or she will need to turn to the original sources. For most values given in this book, finding a different source that reports a value that is a factor of two higher or lower is the rule rather than the exception. We find that knowing the "order of magnitude" can be very useful, and we give examples in the text. Yet, awareness of the inherent variability is critical so as not to get a wrong impression or per
form inferences that are not merited by the current level of data.
Variety (and by extension, variability) is said to be the spice of life—it is definitely

evident at the level of the cell and should always be kept in the back of our minds when discussing values of biological properties.

How many digits should we include when reporting the measured value of biological entities such as those discussed throughout this book? Though this question might sound trivial, in fact there are many sub tle issues we had to grapple with that can affect the reader's capability to use these numbers in a judicious fashion. To give a concrete exam ple, say you measured the number of mitochondria in three cells and found 20, 26, and 34. The average is 26.666…, so how should you best report this result? More specifically, how many significant digits should you include to characterize these disparate numbers? Your spreadsheet software will probably entice you to write something like 26.667. Should it be trusted?

Before we dig deeper, we propose a useful conservative rule of thumb. If you forget everything we write below, try to remember this: It is usually a reasonable choice when reporting numbers in biology to use two signifi cant digits. This will often report all valuable information without the arti fact of too many digits giving a false sense of accuracy. If you write more than three, we hope some inner voice will tell you to think about what it means or just press the backspace key.

We now dive deeper. Significant digits are all digits that are not zero, plus zeros that are to the right of the first nonzero digit. For example, the num ber 0.00502 has three significant digits. Significant digits should supply information on the precision of a reported value. The

last significant digit— that is, the rightmost one—is the digit that we might be wrong about, but it is still the best guess we have for the accurate value. To find what should be considered significant digits, we will use a rule based on the precision (repeatability) of the estimate. The precision of a value is the repeatabil ity of the measurement, given by the standard deviation, or in the case of an average, by the standard error. If the above sentence confuses you, be assured that you are in good company. Keep on reading and make a men tal note to visit Wikipedia at your leisure for these confusing terms, as we do ourselves repeatedly.

Going back to the example above of counting mitochondria, a calculator will yield a standard deviation of 4.0552…. The rule we follow is to report the uncertainty with one significant digit. Thus 4.0552 is rounded to 4, and we report our estimate of the average simply as 26, or more rigorously as 26 ± 4. The last significant digit reported in the average (in this case, 6) is at the same decimal position as the first significant digit of the standard error (in this case, 4). We note that a leading 1 in some conventions does not count as a significant digit (for example, writing 123 with one significant digit will

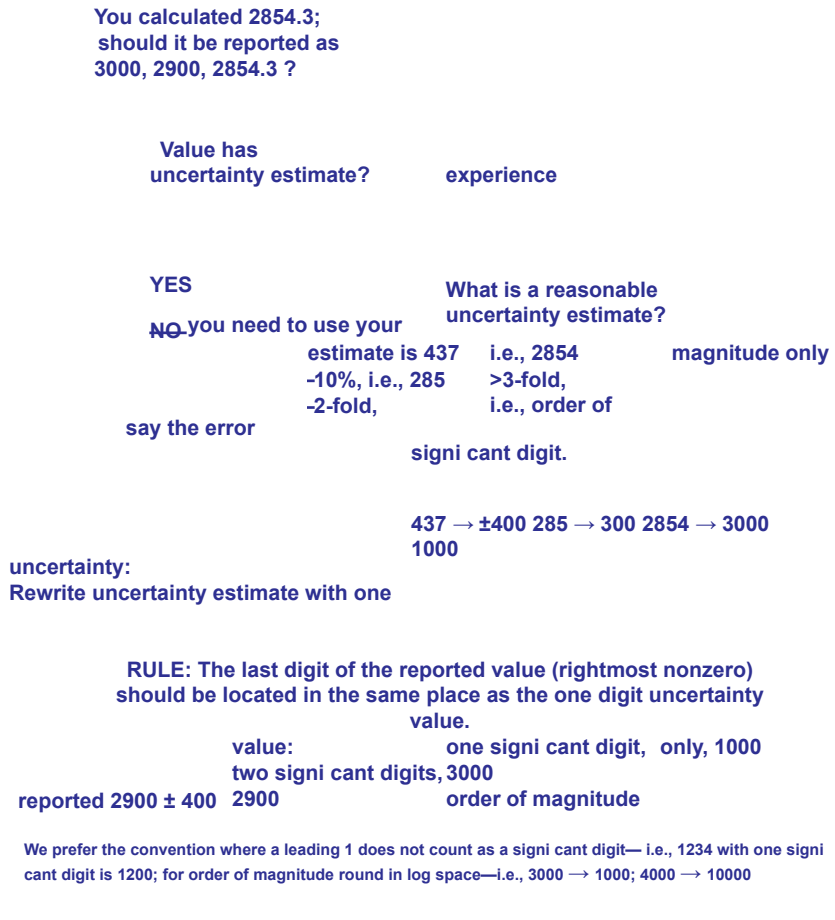
be 120) and that in some cases it is useful to report the uncertainty with two digits rather than just one, but that should not bother us further at this point. But be sure to stay away from using three or more digits in the uncertainty range. Anyone further interested can read a whole report (http://tinyurl.
com/nwte4l5) on the subject.

Unfortunately, for many measured values relating to biology, the impre cision is not reported. Precision refers to how much variation you have in your measurements, whereas accuracy refers to how different it is from the real value. A systematic error will cause an inaccuracy but not an impre cision. Precision you can know from your measurements, but for knowing accuracy you have to rely on some other method. You might want to add the distinction between accuracy and precision to your Wikipedia reading list, but bear with us for now. Not only is there no report of the imprecision (error) in most biological studies, but the value is often written with many digits—well beyond what should be expected to be significant given the biological repeatability of the experimental setting. For example, the aver age for the volume of a HeLa cell may be reported as 2854.3 μm^3. We find, however, that reporting a volume in this way is actually misleading, even if this is what the spreadsheet told the researcher. To our way of thinking, attributing such a high level of precision gives the reader a misrepresenta tion of what the measurement achieved or what value to carry in mind as a rule of thumb.

Because the uncertainty in such measurements is often not reported, we resort to general rules of thumb, as shown in **Estimate 0-4**. Based on reading many studies, we expect many biological quantities to be known with only twofold accuracy, in very good cases maybe to 10%, and in quite variable cases to within 5- or 10-fold accuracy. Low accuracy is usually not because of the tools of measurement, which have very good precision, but because systematic differences, say,

due to growth conditions being different, can lead to low accuracy with respect to any application where the value can be used. In this book, we choose to make the effort to report values with a number of digits that implicitly conveys the uncertainty. The rules of thumb we follow are summarized in Estimate 0-4 as a work flow to infer how many significant digits should be used in reporting a num ber based on knowing the uncertainty or by estimating the level of uncer tainty. For example, say we expect the reported HeLa cell average volume to have 10% inaccuracy (pretty good accuracy for biological data)—that is, about 300 μm^3. As discussed above, we report the uncertainty using one significant digit—that is, all the other digits are rounded to zero. We can now infer that the volume should be written as 2900 μm^3 (two significant digits). If we thought the value has a twofold uncertainty—that is, about 3000 μm^3, we would report the average as 3000 μm^3 (one significant digit).

How to report a value with the appropriate number of significant digits



Estimate 0-4

Finally, if we think there are very large imprecisions, say, to a factor of five or 10, we will resort to reporting only the order of magnitude (that is, 1000 μm^3), or better still, we will write it in a way that reflects the uncertainty as 10^3 μm^3. We indicate only an order of magnitude in cases when the expected imprecision is so large (practically, larger than threefold) that we cannot expect to have any sense of even one digit and have an estimate only of the number of digits in the accurate

value. The digit 1 is special in the sense that it doesn't mean necessarily a value of 1, but rather signifies the order of magnitude. So, in such a case the number can be thought of as reported with less than one significant digit. If you write 100, how do you know if this is merely an order of magnitude, or should be actually interpreted as precise to within twofold or maybe even 10% (that is, also the following zero is precise)? In one convention, this ambiguity can be solved by putting an underline for the last significant digit. So 1$\underline{00}$ shows the zero (and the 1) are significant digits, $\underline{10}$0 shows the 1 is a significant digit, whereas plain 100 is only to within an order of magnitude. We try

to follow this convention in this book. Trailing zeros are by custom used as a replacement for the scientific notation (as in 3×10^3). The scientific notation is more precise in its usage of digits but less intuitive in many cases. The trailing zeros should not be interpreted as indicating a value of exactly zero for those digits, unless specifically noted (for example, with an underline).

We often will not write the uncertainty, because in many cases it was not reported in the original paper the value came from, and thus we do not really know what it is. Yet, from the way we write the property value, the reader can infer something about our ballpark estimate based on the norms above. Such an implicit indication of the expected precision should be familiar, as in the following example borrowed from the excellent book *Guesstimation* by Lawrence Weinstein and John A. Adam. A friend gives you driving direc
tions and states you should be taking a left turn after 20 km. Probably when you reach 22 km and did not see a turn, you would start to get worried. But if the direction had been to take the turn after 20.1 km, you would probably have become suspicious before you reached even 21 km.

When aiming only to find orders of magnitude, we perform the rounding in log space—that is, 3000 would be rounded to 1000, while 4000 would be rounded to 10,000 because $\log_{10}(4) > 0.5$. We follow this procedure because our perception of the world, as well as many error models of measurement methods, are logarithmic (that is, we perceive fold changes rather than absolute values). Thus, the log scale is where the errors are expected to be normally distributed, and the closest round number should be found. When performing a series of calculations (multiplying, subtracting, etc.), it is often prudent to keep more significant digits than will be kept to report final results and perform the rounding only at the end result stage. This is most relevant when subtraction cancels out the leading digits making the following digits critical. We are under the impression that following such guidelines can improve the quantitative hygiene essential for properly using and interpreting numbers in cell biology.

to bear in mind that many of our vignettes are intimately connected. For example, when thinking about the rate of rotation of the flagellar motor that propels bacteria forward, as discussed in the rates chapter, we will do well to remember that the energy source that drives this rotation is the transmembrane potential discussed in the energy and forces chapter. Fur
ther, the rotation of the motor is what governs the motility speed of those cells, a topic with quantitative enticements of its own. Though we will attempt to call out the reticular attachments between many of our differ ent bionumbers, we ask the reader to be on constant alert for the ways in which our different vignettes can be linked up, many of which may harbor some new insights.

To set the cellular stage for the remainder of the book, in this brief sec tion, we highlight three specific model cell types that will form the basis for the coming chapters. Our argument is that by developing intuition for the "typical" bacterium, the "typical" yeast cell, and the "typical" mam malian cell, we will have a working guide for launching into more special ized cell types. For example, even when introducing the highly specialized photoreceptor cells, which are the beautiful outcome of the evolution of "organs of extreme perfection" that so puzzled Darwin, we will still have our "standard" bacterium, yeast, and mammalian cells in the back of our minds as a point of reference. This does not imply a naïveté on our side about the variability of these "typical" cells—indeed, we have several vignettes on these very issues. It is rather an appreciation of the value of a quantitative mental description of a few standard cells that can serve as a useful benchmark to begin the quantitative tinkering that adapts to the biological case at hand, much as a globe gives us an impression of the rel ative proportion of our planet that is covered by oceans and landmasses, and the key geographical features of those landmasses, such as mountain ranges, deserts, and rivers.

Figure 0-1 gives a pictorial representation of our three standard cell types and **Figure 0-2** complements it by showing the molecular census associ ated with each of those cell types. This figure goes hand in hand with **Table 0-1** and can be thought of as a compact visual way of capturing the various numbers housed there. In some sense, much of the remainder of our book focuses on asking the following questions: Where do the numbers in the following figures and table come from? Do they make sense? What do they imply about the functional lives of cells? In what sense are cells the "same" and in what sense are they "different"?

Figure 0-1A shows us the structure of the bacterium *E. coli*. Figure

0-2A shows its molecular census. The yeast cell shown in Figure 0-1B and Fig ure 0-2B reveals new layers of complexity beyond that seen in the stand ard bacterium, as we see that these cells feature a variety of internal

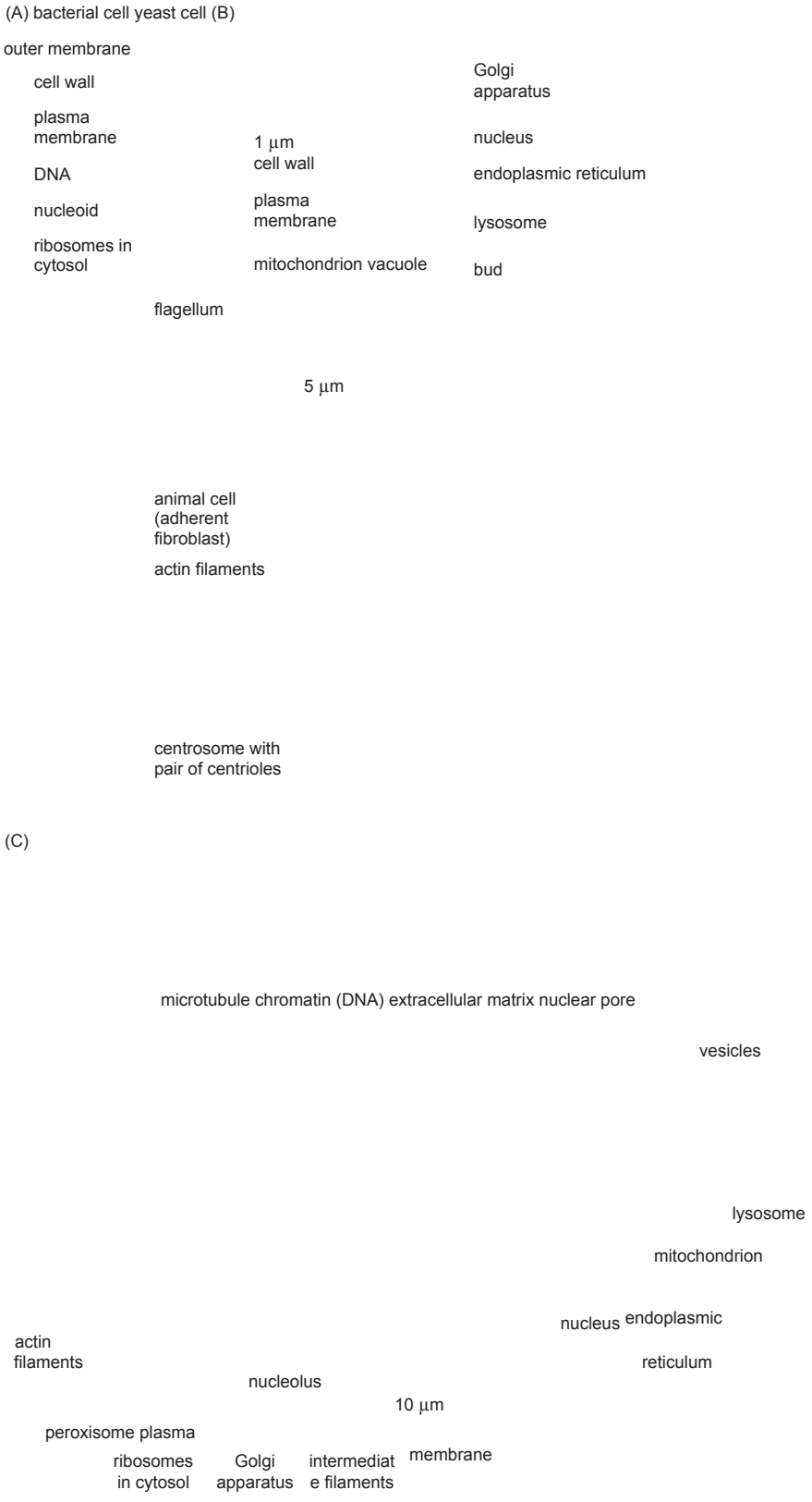


Figure 0-1 The standard cells. (A) A schematic bacterium revealing the characteristic size and compo nents of ***E. coli***. (B) A budding yeast cell showing its characteristic size, its organelles, and various classes of molecules present within it. (C) An adherent human cell. We note that these are very simplified schemat ics. For example, only a small fraction of ribosomes are drawn. Each cell is drawn to a different scale, as in dicated by the distinct scale bars in each schematic. The relative sizes of the bacterial and yeast cells at the same scale as the mammalian cell are shown in the bottom right. (A, and C, adapted from Alberts B, Johnson A, Lewis J et al. [2015] Molecular Biology of the Cell, 6th ed. Garland Science.)

bacterial cell (speci cally, ***E. coli***: $V \approx 1\ \mu m^3$; $L \approx 1\ \mu m$; $\tau \approx 1$ hour)
(A)

membrane
water
protein
lipid
+
2×10^{10}
protein
ion
inorganic
2×10^3 2×10^4
5×10^7
5×10^6 bp
5×10^5 10^8
3×10^6

mRNA ribosome DNA

yeast cell (speci cally, ***S. cerevisiae***: $V \approx 30\ \mu m^3$; L
$\approx 5\ \mu m$; $\tau \approx 3$ hours) (B)

10^7 10^9
+
3×10^9 10^5
6×10^{11} 10^8
12×10^7
bp
3×10^4

(C) mammalian cell (speci cally, HeLa: $V \approx 3000\ \mu m^3$; $L \approx 20\ \mu m$; $\tau \approx$

1 day) 10^9

6×10^{13} 2×10^5 2×10^{11} 10^6 10^{11}
10^{10}
+
3×10^9 bp

Figure 0-2 An order-of-magnitude census of the major components of the three model cells we employ often in the lab and in this book. A bacterial cell (***E. coli***), a unicellular eukaryote (the budding yeast ***S. cerevisiae***), and a mammalian cell line (such as an adherent HeLa cell).

membrane-bound structures. One of the key reasons that yeast cells have served as representatives of eukaryotic biology is the way they are divided into various compartments, such as the nucleus, the endoplasmic reticulum, and the Golgi apparatus. Further, their genomes are packed tightly within the cell nucleus in nucleoprotein complexes known as

property	*E. coli*	budding yeast	mammalian (HeLa line)
cell volume	0.3–3 μm^3	30–100 μm^3	2–4 × 10^6
number protein coding genes	4.6 Mbp	4–5 nm	12 Mbp
proteins per μm^3 cell	10^3–10^4	10^4–10^5	10^5–10^6
volume mRNA per cell	~10^6	~10^8	~10^{10}
proteins per cell	4300	6600	3.2 Gbp
mean diameter of protein		1000–10,000 μm^3	21,000
genome size			
regulator binding site length	10–20 bp	5–10 bp	
promoter length	~1000 bp	~1000 bp	~10
minimal doubling time			10^4–10^5 bp
gene length			10^4–10^6 bp (with introns)
ribosomes/cell		pM	
concentration of one protein per cell	~100 bp	~1000 bp	~1 nM
transitions between protein states (active/inactive)		~0.2 s	~0.1–1 pM
diffusion time of protein across cell ($D \approx 10\ \mu m^2/s$)	~0.01 s	~0.03 s	~1–10 s
diffusion time of small molecule across cell ($D \approx 100\ \mu m^2/s$)	~0.001 s	~1 min	~0.1–1 s
time to transcribe a gene	<1 min (80 nts/s)	~1 min	~30 min (incl. mRNA processing)
time to translate a protein	<1 min (20 aa/s)	30 min	~30 min (incl. mRNA export)
typical mRNA lifetime	3 min	0.3–3 h	10 h
typical protein lifetime	20 min	1 h	10–100 h
	1 h	~10^5	20 h
	~10^4	1–100 μs	~10^6
time scale for equilibrium binding of small molecule to protein (diffusion limited)	1–1000 ms (1 μM–1 nM affinity)		
time scale of transcription factor binding to DNA site	~1 s		
mutation rate	10^{-8}–10^{-10}/bp/replication		

Table 0-1 Typical parameter values for a bacterial ***E. coli*** cell, the single-celled eukaryote ***S. cerevisiae*** (budding yeast), and a mammalian HeLa cell line. These are crude characteristic values for happily dividing cells of the common lab strains. (Adapted from Alon U [2006] Introduction to Systems Biology. CRC Press. See full references at BNID 111494.)

nucleosomes, an architectural motif shared by all eukaryotes. Beyond its representative cellular structures, yeast has been celebrated

because of the “awesome power of yeast genetics,” meaning that in much the same way we can rewire the genomes of bacteria such as *E. coli*, we are now able to alter the yeast genome nearly at will. As seen in the table and figure, the key constituents of yeast cells can roughly be thought of as a scaled-up version of the same census results already sketched for bacteria in Figure 0-1A.

Figures 0-1C and 0-2C complete the trifecta by showing a “standard” mam malian cell. The schematic shows the rich and heterogeneous structure of such cells. The nucleus houses the billions of base pairs of the genome

and is the site of the critical transcription processes taking place as genes are turned on and off in response to environmental stimuli and over the course of both the cell cycle and development. Organelles, such as the endoplasmic reticulum and the Golgi apparatus, are the critical site of key processes, such as protein processing and lipid biosynthesis. Mitochon
dria are the energy factories of cells where in humans, for example, about our body weight in ATP is synthesized each and every day. What can be said about the molecular players within these cells?

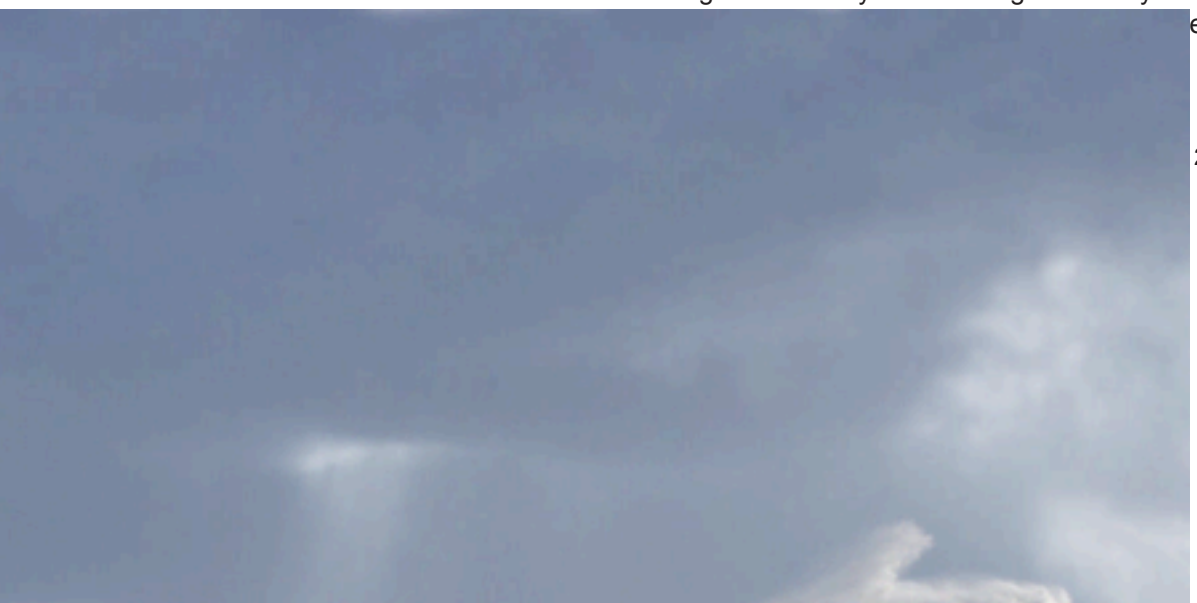
Given that there are several million proteins in a typical bacterium and these are the product of several thousand genes, we can expect the “aver age” protein to have about 10^3 copies. The distribution is actually very far from being homogenous in any such manner, as we will discuss in several vignettes in Chapter 2 on concentrations and absolute numbers. Given the rule of thumb from above that one molecule per *E. coli* corresponds to a concentration of roughly 1 nM, we can predict the “average” protein concentration to be roughly 1 mM. We will be sure to critically dissect the concept of the “average” protein, highlighting how most transcription factors are actually much less abundant than this hypothetical average protein and why components of the ribosome are needed in higher con
centrations. We will also pay close attention to how to scale from bacteria to other cells. A crude and simplistic null model is to assume that the abso lute numbers per cell tend to scale proportionally with the cell size. Under this null model, concentrations are independent of cell size.

Let’s exemplify our thinking for a mammalian cell that has 1000 times the volume of a bacterial cell. Our first-order expectation will be that the abso lute copy number will be about 1000 times higher and the concentration will stay about the same. The reader knows better than to take this as an immutable law. For example, some universal molecular players, such as ribosomes or the total amount of mRNA, also depend close to proportion ally on the growth rate—that is, inversely with the doubling time. For such a case we should account for the fact that the mammalian cell divides, say, 20 times slower than the bacterial cell. So, for these cases we need a dif ferent null model. But in the alien world of molecular biology, where our intuition often fails, any guidance (that is, null model to rely on) can help. As a teaser example, consider the question of how many copies there are of your favorite transcription factor in some mammalian cell line—say, p53 in

a HeLa cell. From the rules of thumb above, there are about 3 million proteins per μm^3 and a characteristic mammalian cell will be 3000 μm^3 in volume. We have no reason to think our protein is especially high in terms of copy number, so it is probably not taking one part in a hundred of the proteome (only the most abundant proteins will do that). So, an upper crude estimate would be 1 in 1000. This translates immediately into 3×10^6 proteins/$\mu m^3 \times 3000$ μm^3/1000 proteins/our protein $\sim$ 10 million copies

of our protein. As we shall see, transcription factors are actually on the low end of the copy number range, and something between 10^5–10^6 copies would have been a more accurate estimate, but we suggest this is definitely better than being absolutely clueless. Such an estimate is the crudest example of an easily acquired "sixth sense." We find that those who master the simple rules of thumb discussed in this book have a significant edge in street-fighting cell biology (borrowing from Sanjoy Mahajan's gem of a book on "street-fighting mathematics").

The logical development of the remainder of the book can be seen through the prism of Figure 0-1. First, we begin by noting the structures and their sizes. This is followed in the second chapter by a careful analysis of the copy numbers of many of the key molecular species found within cells. Already at this point, the interconnectedness of these numbers can be seen, for example, in the relationship between the ribosome copy number and the cell size. In Chapter 3, we explore the energy and force scales that mediate the interactions between the structures and molecu lar species considered in the previous chapters. This is then followed in Chapter 4 by an analysis of how the molecular and cellular drama plays out over time. The various structures depicted in Figure 0-1 exhibit order on many different scales, an order that conveys critical information to the survival and replication of cells. Chapter 5 provides a quantitative picture of different ways of viewing genomic information and on the fidelity of information transfer in a variety of different cellular processes. Our final chapter punctuates the diversity of cells beyond what is shown in Figure 0-1 by considering a variety of other



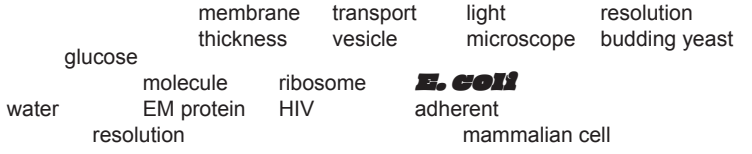
es that

Chapter 1: Size and Geometry

In this chapter, all of our vignettes center in one way or another on the same simple question: “How big?” JBS Haldane, when he wasn’t busy inventing population genetics or formulating the theory of enzyme kinet ics (among many other things), wrote a delightful essay entitled “On Being the Right Size.” There, he discusses how size is critical in understanding functional constraints on animals. For example, Haldane notes that when a human steps out of water, because of surface tension, he or she car ries roughly a pound of water. On the other hand, an insect would carry comparatively much more, covered by about its own weight in water. The functional implications are often dire. In this same spirit, we aim to char acterize the sizes of things in molecular and cellular biology with the hope of garnering insights into the kinds of functional implications explained by Haldane at larger scales.

Biological structures run the gamut in sizes from the nanometer scale of the individual macromolecules of life all the way up to the gigantic cyanobacte rial blooms in the ocean that can be seen from satellites. As such, biologists can interest themselves in phenomena spanning more than 15 orders of magnitude in length scale. Though we find all of these scales fascinating (and important), in this book we primarily focus on those length scales that are smaller than individual organisms, as depicted in **Figure 1-1**.

In moving from the intuitive macroscopic world into the microscopic domain, a critical intellectual linkage will often be provided by Avogadro’s number (see the Preface for historical efforts to determine its value). This important constant is defined as the number of hydrogen atoms contained in one gram of such atoms. With a value of about 6×10^{23}, this conversion



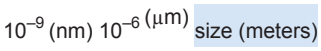


Figure 1-1 Range of characteristic sizes of the main biological entities relevant to cells. On a logarithmic scale we depict the range from single molecules, serving as the nuts and bolts of biochemistry, through molecular machines, to the ensembles that are cells.

factor reveals itself time and again, and the conversion was shown in the opening chapter.

One dilemma faced when trying to characterize biological systems is the extent to which we should focus on model systems. Often, the attempt to be comprehensive can lead to an inability to say anything concrete. As a result, we aim to give an intimate quantitative description of some com
mon model cells and organisms, punctuated here and there by an attempt to remind the reader of the much larger diversity that lies beyond. We sug gest that in those cases where we don't know better, it is very convenient to assume that all bacteria are similar to *E. coli.* We make this simplification for the sake of providing a general order-of-magnitude idea of the num bers that characterize most bacteria. In the same vein, our picture of a mammalian cell is built around the intuition that comes from using HeLa cells as a model system. We can always refine this crude picture when more information becomes available on, say, the volume or geometry of a specific cell line of interest. The key point is to have an order of magnitude to start with. A similar issue arises when we think about the changes in the properties of cells when they are subjected to different external condi tions. Here again, we often focus on the simplified picture of happily divid ing, exponentially growing cells, while recognizing that other conditions can change our picture of the "average" cell considerably. The final issue along this progression of challenges having to do with how to handle the diversity of biological systems is how we should deal with cell-to-cell var iation—that is, how much do individual cells that have the same genetic composition and face the same external conditions vary? This chapter addresses these issues through a quantitative treatment both for cell size and protein abundance.

The geometries of cells come in a dazzling variety of shapes and sizes. Even the seemingly homogeneous world of prokaryotes is represented by a sur prising variety of shapes and sizes. But this diversity of size and shape is not restricted only to cells. Within eukaryotic cells are found organelles with a similar diversity of form and a range of different sizes. In some cases, such as the mitochondria and chloroplasts (and perhaps the nucleus), the sizes of these organelles are similar to bacteria, which are their evolutionary ances tors through major endosymbiotic events. At smaller scales still, the mac romolecules of the cell come into relief and, yet again, it is found that there are all sorts of different shapes and sizes, with examples ranging from small peptides, such as toxins, to the machines of the central dogma to the assem blies of proteins that make up the icosahedral capsids of viruses.

follows function. In biology, this idea has been a part of a long "structural" tradition that includes the development of microscopy and the emergence of structural biology. We are often tempted to figure out the *relative* scales of the various participants in some process of interest. In many of the vignettes, we attempt to draw a linkage between the size and the biological function.

Interestingly, even from the relatively simple knowledge of the sizes of bio logical structures, we can make subtle functional deductions. For exam ple, what governs the burst size of viruses (that is, the number of viruses that are produced when an infected cell releases newly synthesized viruses)? Some viruses infect bacteria, whereas others infect mammalian cells, but the sizes of both groups of viruses are relatively similar, whereas the hosts differ in size by a characteristic volume ratio of 1000. This helps explain the fact that burst sizes from bacteria are about 100, whereas in the case of mammalian cells the characteristic burst size is in the thousands. Throughout the chapter, we return to this basic theme of reflecting on the biological significance of the many length scales we consider.

CELLS AND VIRUSES

How big are viruses?

In terms of their absolute numbers, viruses appear to be the most abun dant biological entities on planet Earth. The best current estimate is that there are a whopping 10^{31} virus particles in the biosphere. We can begin to come to terms with these astronomical numbers by realizing that this implies that for every human on the planet, there is nearly an Avogadro's number worth of viruses. This corresponds to roughly 10^{8} viruses to match every cell in our bodies. The number of viruses can also be contrasted with an estimate of $4–6 \times 10^{30}$ for the number of prokaryotes on Earth (BNID
104960). However, because of their extremely small size, the mass tied up in these viruses is only approximately 5% of the prokaryotic biomass. The assertion about the total number of viruses is supported by measurements using both electron and fluorescence microscopy. For example, if a sam
ple is taken from the soil or the ocean, electron microscopy
observations reveal an order of magnitude more viruses than bacteria (≈10:1 ratio, BNID

104962). These electron microscopy measurements are independently confirmed by light microscopy measurements. By staining viruses with fluorescent molecules, they can be counted directly under a microscope and their corresponding concentrations determined (for example, 10^7 viruses/mL).

Organisms from all domains of life are subject to viral infection, whether tobacco plants, flying tropical insects, or archaea in the hot springs of Yel lowstone National Park. However, it appears that those viruses that attack bacteria (that is, so called bacteriophages—literally, bacteria eater—see **Figure 1-2**) are the most abundant of all, with these viruses present in huge numbers (BNID 104839, 104962, 104960) in a host of different environments ranging from soils to the open ocean.

As a result of their enormous presence on the biological scene, viruses play a role not only in the health of their hosts, but in global geochemical cycles affecting the availability of nutrients across the planet. For exam ple, it has been estimated that as much as 20% of the bacterial mass in the ocean is subject to viral infection every day (BNID 106625). This can strongly decrease the flow of biomass to higher trophic levels that feed on prokaryotes (BNID 104965).

Viruses are much smaller than the cells they infect. Indeed, it was their remarkable smallness that led to their discovery in the first place. Researchers

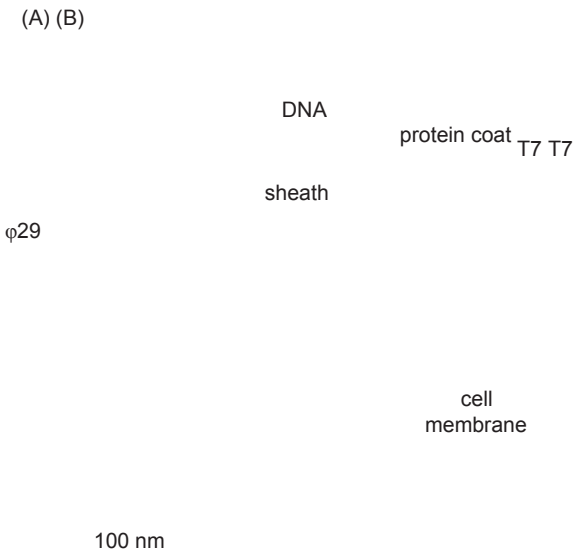


Figure 1-2 Geometry of bacteriophages. (A) Electron microscopy image of phi29 and T7 bacteriophages as revealed by electron microscopy. (B) Schematic of the
structure of a bacteriophage. (A, adapted from Grimes S, Jardine PJ & Anderson D [2002] ***Adv Virus Res*** 58:255–280.)

were puzzled by remnant infectious elements that could pass through filters small enough to remove pathogenic bacterial cells. This led to

the hypothesis of a new form of biological entity. These entities were subsequently identified as viruses.

Viruses are among the most symmetric and beautiful of biological objects, as shown in **Figure 1-3**. The figure shows that many viruses are charac terized by an icosahedral shape with all of its characteristic symmetries (that is, twofold symmetries along the edges, threefold symmetries on the faces, and fivefold rotational symmetries on the vertices). The outer pro tein shell, known as the capsid, is often relatively simple since it consists of many repeats of the same protein unit. The genomic material is contained

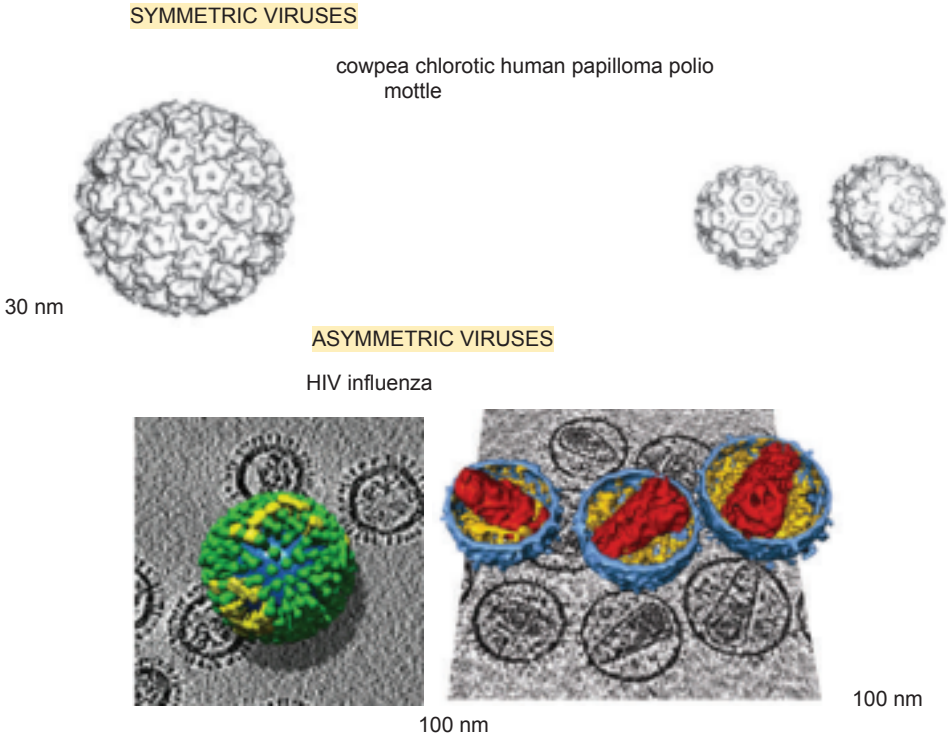


Figure 1-3 Structures of viral capsids. The regularity of the structure of viruses has enabled detailed, atomic-level analysis of their construction patterns. This gallery shows a variety of the different geometries explored by the class of nearly spherical viruses. HIV and influenza figures are 3D renderings of virions from the tomogram. (Symmetric virus structures adapted from Baker TS, Olson NH & Fuller SD [1999] ***Microbiol Mol Biol Rev*** 63:862–922. HIV structure adapted from Briggs JAG, Grünewald K, Glass B et al. [2006] ***Structure*** 14:15–20. Influenza virus structure adapted from Harris A, Cardone G, Winkler DC et al. [2006] ***Proc Natl Acad Sci USA*** 103:19123–19127.)

within the capsid. These genomes can be DNA or RNA, single-stranded or double-stranded (that is, ssDNA, dsDNA, ssRNA, or dsRNA) with charac teristic sizes ranging from 10^3–10^6 bases (BNID 103246, 104073). With some interesting exceptions, a useful rule of thumb is that the radii of viral cap sids themselves are all within a factor of 10 of each other, with the smaller viruses having a diameter of several tens of nanometers and the larger ones reaching diameters

of several hundreds of nanometers, which is on par with the smallest bacteria (BNID 103114, 103115, 104073). Represent ative examples of the sizes of viruses are listed in **Table 1-1**. The structures of many viruses, such as HIV, have an external envelope (resulting in the label "enveloped virus") made up of a lipid bilayer. The interplay between the virus size and the genome length can be captured via the packing ratio, which is the percent fraction of the capsid volume taken by viral DNA. For phage lambda it can be calculated to be about 40%, whereas for HIV it is more than 10 times lower (BNID 111591).

virus	size (nm)	genome size (nucleotides)	genome type, capsid structure	BNID
porcine circovirus (PCV)	40 × 300 30	153,000	dsDNA, icosahedral	106456, 106457
cowpea mosaic virus (CPMV)	45 × 54	170,000	ssRNA, roughly spherical	103246, 106442
cowpea chlorotic mottle virus (CCMV)	58	1,200,000 2,800,000	ssRNA, roughly spherical	104376, 104375, 106453
φX174 (*E. coli* bacteriophage)	58	circular ssDNA, icosahedral		103114, 111324
tobacco mosaic virus (TMV)	88–110	2 ssRNA molecules, icosahedral		109734
polio virus	80–120	3 ssRNA molecules, icosahedral		103122, 105770
φ29 (*Bacillus* phage)	120–150 125	ssDNA, icosahedral		109732, 109733
lambda phage	140	ssRNA, rod shaped		103114, 103115, 106441
T7 bacteriophage	500	ssRNA, icosahedral		104073, 105768
adenovirus (linear DNA)	500 × 1000	dsDNA, icosahedral (T3)		101849, 105769
influenza A	1760	dsDNA, icosahedral (with tail)		103114, 106458
HIV-1	9400			103246, 111424
herpes simplex virus 1	7900			105142, 105143
Epstein-Barr virus (EBV)	5400			109554, 109556
mimivirus	6400			
pandora virus	7500		dsDNA, icosahedral	
17	19,000		dsDNA, icosahedral	
28	49,000		dsDNA, icosahedral	
28	40,000		dsDNA, icosahedral	
32	36,000		106467, 106468	
	14,000	dsDNA, 55 genes, icosahedral (T7)	106454, 106455	
	9700			

Table 1-1 Sizes of key representative viruses. The viruses in the table are organized according to their size, with the smallest viruses shown first and the largest viruses shown last. The organization by size gives a different perspective than typical biological classifications that use features such as the nature of the genome—RNA or DNA, single-stranded (ss) or double-stranded (ds)—and the nature of the host. Values are diameters rounded to one or two significant digits.

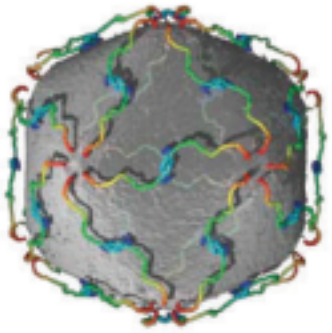
PRD1 capsid.

Some of the most interesting viruses have structures with less symmetry than those

Figure 1-4 The P30 protein dimer serves as a measuring tape to help create the bacteriophage

described above. Indeed, two of the biggest viral newsmakers, HIV and influenza, sometimes have irregular shapes, and even the structure from one influenza or HIV virus particle to the next can be different. Exam ples of these structures are shown in Figure 1-3. Why should so many viruses have a characteristic length scale of roughly 100 nm? If we consider the density of genetic material inside the capsid, a useful exercise for the motivated reader, it is found that the genomic material in bacterial viruses can take up nearly as much as 50% of the volume. Further, the viral DNA often adopts a structure that is close-packed and nearly crystalline to enable such high densities. Thus, in

CELLS AND VIRUSES 9



~~15 nm~~ P30, a molecular measuring tape

these cases, if we take as a given the length of DNA, which is tied in turn to the number of genes that viruses must harbor, the viruses show strong economy of size, minimizing the required volume to carry their genetic material.

To make a virus, the monomers making up the capsid can self-assemble; one mechanism is to start from some vertex and extend in a symmetric manner. But what governs the length of a facet? That is, what is the distance between two adjacent vertices that dictates the overall size of a virion? In one case, a nearly linear 83-residue protein serves as a molecular tape measure, helping the virus to build itself to the right size. The molecular players making this mechanism possible are shown in **Figure 1-4**. A dimer of two 15-nm-long proteins defines distances in a bacteriophage, which has a diameter of about 70 nm.

The recently discovered gigantic mimivirus and pandoravirus are about an order of magnitude larger (BNID 109554, 111143). The mechanism that serves to set the size of these viruses remains an open question. These viruses are larger than some bacteria and even rival some eukaryotes. They also contain genomes larger than 2 Mbp long (BNID 109556) and challenge our understanding of both viral evolution and diversity.

How big is an *E. coli* cell and what is its mass?

The size of a typical bacterium such as *E. coli* serves as a convenient stand ard ruler for characterizing length scales in molecular and cell biology. A rule

of thumb, based upon generations of light and electron microscopy

meas urements, for the dimensions of an *E. coli* cell is to assign it a diameter of about ≈1μm, a length of ≈2μm, and a volume of ≈1μm^3 (1 fL) (BNID 101788). The shape can be approximated as a spherocylinder—that is, a cylinder with hemispherical caps. Given the quoted diameter and length, we can compute a more refined estimate for the volume of ≈1.3 μm^3 (5π/12, to be accurate). The difference between this value and the rule-of-thumb value quoted above shows the level of inconsistency we live with comfortably when using rules of thumb. One of the simplest routes to an estimate of the mass of a bacterium is to exploit the ≈1 μm^3 volume of an *E. coli* cell and to assume it has the same density as water. This naïve estimate results in another standard value—
namely, that a bacterium such as *E. coli* has a mass of ≈1 pg (pico = 10^{-12}). Because most cells are about two-thirds water (BNID 100044, 105482), and the other components, like proteins, have a characteristic density of about 1.3 times the density of water (BNID 101502, 104272), the conversion from cellular volume to mass is accurate to about 10%.

One of the classic results of bacterial physiology emphasizes that the plas ticity in properties of cells derives from the dependence of the cell mass upon growth rate. Stated simply, faster growth rates are associated with larger cells. This observation refers to physiological changes where media that increase the growth rate also yield larger cells, as shown in **Figure 1-5**. This was also found to hold true genetically where long-term experimental evolution stud ies that led to faster growth rates showed larger cell volumes (BNID 110462).

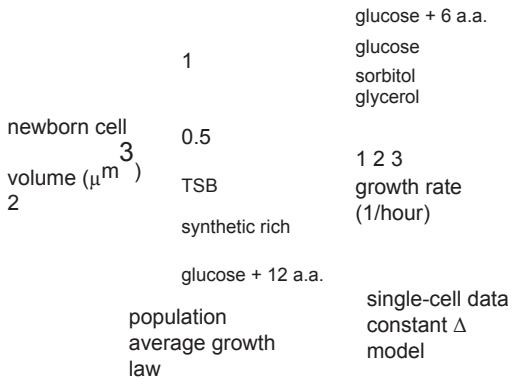


Figure 1-5 Relation between cell volume and growth rate. Using microscopy and microfluidic devices, cell volume can be measured at the single-cell level under various conditions, confirming that the average cell volume grows exponentially with growth rate. In contrast, variation among cells for a given condition scales differently. The variation in single-cell behavior is used to test models of cell size regulation. (Adapted from Taheri-Araghi S, Bradde S, Sauls JT et al. [2015] ***Curr Biol*** 25:385.)

Such observations help us dispel the myth of "the cell"—where people, often unwarily, use measurements about one cell to make inferences about other cell types or the same cell type under different conditions. Classic studies by Dennis and Bremer systematized these measurements and found that dry mass varies, as shown in **Table 1-2**, from an average value of 148 fg for cells dividing every 100 minutes to 865 fg for those with a 24-minute divi sion time, indicating over a fivefold difference, depend

generation time (min)

100

60	dry mass per cell (fg = 10^{-15}g)	430
40		640
30	150	870
24	260	

ing upon the growth rate. A similar trend has been seen in other organisms (for example, for budding yeast; BNID 105103). At about 70% water, these values correspond to a range between about 0.4 to 2.5 μm^3 in terms of volume. How can we rationalize the larger sizes for cells growing at faster rates? This question is under debate to this day. Explanations vary from suggesting it has an advantage in the way resource allocation is done* to claiming that it is actually only a side effect of having a built-in period of about 60 minutes from the time a cell decides it has accu mulated enough mass to begin the preparations for divi sion and until it finishes DNA replication and the act of division. This roughly constant "delay" period leads to an exponential dependence of the average cell mass on the growth rate in this line of reasoning.†

Table 1-2 Relation between bacterial mass and division time. The dry mass per cell is given as a function of the generation (doubling) time. Mass is suggested to increase roughly exponentially with growth rate, as originally observed by Schaechter M, Maaloe O & Kjeldgaard NO (1958) ***J Gen Microbiol*** 19:592–606. The cell dry weight was calculated using a value of 173 μg per OD460 unit of 1 mL (BNID 106437). The strain used is B/r, a strain commonly used in early bacterial physiology studies. (Values taken from Neidhardt FC [1996] ***Escherichia coli*** and ***Salmonella***: Cellular and Molecular Biology, Vol 1. ASM Press.)

Methods to measure cell volume range from the use of a Coulter Coun ter (BNID 100004), which infers volume based on changes in resistance of a small orifice as a cell passes in it, to more direct measurements using fluorescence microscopy that gauge cell lengths and diameters under different conditions (BNID 111480, 106577; see Figure 1-5). Surprisingly, the fact that different laboratories do not always converge on the same values may be due to differences in calibration methods or exact strains and growth conditions. An unprecedented ability to measure cell mass is achieved by effectively weighing cells on a microscopic cantilever. As illustrated in **Figure 1-6A**, fluid flow is used to force a cell back and forth in the hollowed-out cantilever. The measurement exploits the fact that the cell mass affects the oscillation frequency of the cantilever. This frequency can be measured to a phenomenal accuracy and used to infer masses with femtogram precision. By changing the liquid flow direction, the cell is trapped for minutes or more, and its mass accumulation rate is meas ured continuously at the single-cell level. In the initial application of this

* Molenaar D, van Berlo R, de Ridder D, & Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5:323.

† Amir A (2014) Cell size regulation in bacteria. *Phys Rev Lett* 112:208102.

t_1

t_2 flow traps and moves cells in and out of cantilever

infer mass from tiny shifts in cantilever oscillation frequency

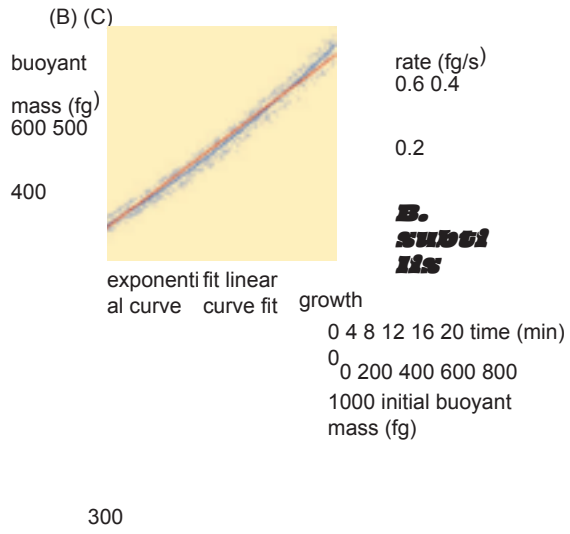
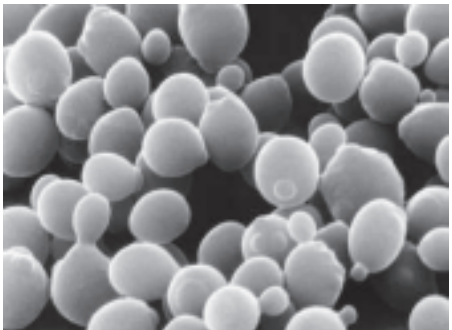


Figure 1-6 Using buoyant mass to measure the growth of single cells. (A) A micron scale cantilever oscillates at high frequency, and the mass of cells can be determined from changes in the oscillation frequency. (B) Measured over time, this results in a single-cell mass accumulation curve as shown. (C) Relation between growth rate and buoyant mass for ***B. subtilis*** cells. A comparison between the predictions of linear and exponential growth models are shown as best fits. The similarity demonstrates how close the two models are over a range of only a twofold increase over the course of the cell cycle. Cell dry weight is about four times the buoyant mass. (Adapted from Godin M, Delgado FF, Son S et al. [2010] ***Nat Methods*** 7:387–390. With permission from Macmillan Publishers Ltd.)

technique, it was shown that single cells that are larger also accumulate mass faster, shedding light on a long-standing question: Is cell growth lin ear with time or more appropriately described by an approximately expo nential trend? The differences can be minute, but with these revolutionary capabilities it was clearly seen that the latter scenario better represents the situation in several cell types tested, as shown in **Figure 1-6B**.

How big is a budding yeast cell?

The budding yeast, *Saccharomyces cerevisiae* has served as the model eukaryote in much the same way that *E. coli* has served as the representative



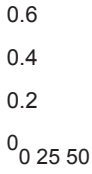
prokaryote. Due to its importance in making beer and baking bread (thus also called brewer's or baker's yeast), this easily accessible and simply cultured organism was also an early favorite of sci

entists, as interestingly recalled by James A. Bar nett in his paper “Beginnings of microbiology and biochemistry: the contribution of yeast research.” These cells are significantly larger than common bacteria and, as such, are a convenient single-celled organism to study under the microscope. In large part due to the ease with which its genome can be manipulated, yeast has remained at the forefront of biological research and, in 1996, was the first eukaryotic organism to have its genome completely sequenced. Another feature that makes yeast handy for geneticists is their dual lifestyle as either hap loids or diploids. Haploid cells have only one copy

Figure 1-7 Electron micrograph of budding yeast cells. (Courtesy of Ira Herskowitz and E. Schabtach.)

of each chromosome, just like a human female egg cell, whereas diploid cells have two copies of each chromosome, just like somatic cells in our bodies. Haploids are analogous to our gametes—the egg cell and sperm cells. The haploid/diploid coexistence in budding yeast enables scientists to easily transfer mutations and study their effects.

A simple rule of thumb for the dimensions of yeast cells is to think of them as spheres with a diameter of roughly 4 μm for haploids and roughly 6 μm for diploids, as shown in **Figure 1-7** (BNID 101796). To put the rela tive sizes of yeast and bacteria in everyday terms, if we think of a world in which *E. coli* is the size of a human, then yeast is about the size of an elephant. Prominent components of the cell volume include the nucleus, which takes up about 10% of the total cell volume (BNID 100491, 103952), the cell wall, often ignored but making up 10–25% (BNID 104593, 104592) of the total dry mass, and the endoplasmic reticulum and vacuole, which are usually the largest organelles.



One of the ideas that we repeatedly emphasize in a quantitative way is cell-to-cell variability and its role in establishing the different behaviors of cells in response to different environmental cues. As yeast replicate by budding off small daughter cells from a larger mother, any population has a large range of cell sizes spread around the median, as shown in **Figure 1-8**. The hap loid strain shown has a median cell volume of 42 ± 2 μm^3 (BNID 100427). Another common metric is the 25th–75th percentile range, which here is ≈30–60 fL. The median cell size itself is highly dependent on genetic and environmental factors. A diploid cell is almost twice as big as its hap loid progenitors at ≈82 μm^3 (BNID 100490). This reflects the more general observation from cell biology that median cell size tends to grow propor tionally to ploidy (DNA content). Yeasts, where ploidy can be manipu lated to higher than two, serve as useful test cases for

Figure 1-8 Histogram of distribution of cell sizes for wild-type budding yeast cells. (Adapted from Jorgensen P, Nishikawa JL, Breitkreutz BJ & Tyers M [2002] *Science* 297:395–400.)

illuminating this phenomenon.

Beyond the bulk DNA content, the median cell volume can differ by more than twofold in different strains of *S. cerevisiae* that evolved in different parts of the world, or more recently, in different industries utilizing them. Finally, like *E. coli*, median cell size in yeast is correlated with growth rate—the better the environmental conditions and growth rate, the larger the cells (BNID 101747). An intriguing open question is whether there is an evolutionary advantage of shifting cell size in response to environ mental conditions. Recent measurements (BNID 100490) have probed how sensitive yeast cell size is to single-gene deletions. In some of these deletion mutants, the median volume was only 40% of the wild-type size, whereas in others it was larger than wild type by >70%. These observations reveal strong coupling between size regulation and the expression of criti
cal genes. It still remains largely unknown how genetic and environmental changes shift the median cell size in yeast.

How big is a human cell?

A human is, according to the most recent estimates, an assortment of $3.7 \pm 0.8 \times 10^{13}$ cells (BNID 109716), plus a similar complement of allied microbes. The identities of the human cells are distributed amongst more than 200 different cell types (BNID 103626, 106155) that perform a stag
gering variety of functions. The shapes and sizes of cells span a large range, as shown in **Table 1-3**. Size and shape, in turn, are intimately tied to the function of each type of cell. Red blood cells need to squeeze through nar row capillaries, and their small size and biconcave disk shape achieve that while also attaining a high surface-area-to-volume ratio. Neurons need to transport signals and, when connecting our brains to our legs, can reach lengths of over a meter (BNID 104901), but with a width of only about 10 μm. Cells that serve for storage, like fat cells and oocytes, have very large volumes.

The different shapes also enable us to recognize the cell types. For exam ple, the leukocytes of the immune system are approximately spherical in shape, while adherent tissue cells on a microscope slide resemble a fried egg, with the nucleus analogous to the yolk. In some cases, such as the

cell type	BNID	average volume (μm^3)		
sperm cell		fibroblast		cardiomyocyte
red blood cell		HeLa, cervix		megakaryocyte
lymphocyte		hair cell (ear)		fat cell
neutrophil		osteoblast		oocyte
beta cell		alveolar		30
enterocyte		macrophage		100
				200

300	30,000	103725, 105879
1000	600,000 4,000,000	108242
1400	109891, 109892	108088
2000	107600	103566
3000	111788	108243
4000	108241	110129
4000	109227	107668
5000	111216	101664
15,000	108244	

Table 1-3 Characteristic average volumes of human cells of different types. Large cell–cell variation of up to an order of magnitude or more can exist for some cell types, such as neurons or fat cells, whereas for others the volume varies by much less (for example, red blood cells). The value for a beta cell comes from a rat, but we still present it because average cell sizes usually change relatively little among mammals.

different types of white blood cells, the distinctions are much more subtle and are only reflected in molecular signatures.

Mature female egg cells are among the largest cell types, with a ≈120 μm diameter. Other large cell types include muscle fiber cells (which merge together to form syncytia, where multiple nuclei reside in one cell) and meg akaryocytes (the bone marrow cells responsible for the production of blood platelets). Both of these cell types can reach 100 μm in diameter (BNID 106130). Red blood cells, also known as erythrocytes, are some of the small est and most abundant of human cells. These cells have a characteristic biconcave disk shape with a depression where the nucleus was lost in mat uration and have a corresponding diameter of 7–8 μm (BNID 100509) and a volume of ≈100 μm^3 (BNID 101711, 101713). Sperm cells are even smaller, with a volume of about 20–40 μm^3 (BNID 109892, 109891).

Certain human cell lines have been domesticated as laboratory work horses. Perhaps the most familiar of all are the so-called HeLa cells, an example of which is shown dividing in **Figure 1-9**. Such immortal can cer cell lines divide indefinitely, alleviating the need to sacrifice primary animal tissue for experiments. These cell lines have been used for studies on topics such as the molecular basis of signal transduction and the cell cycle. In these cell types, the cell volumes are captured by a rule-of-thumb

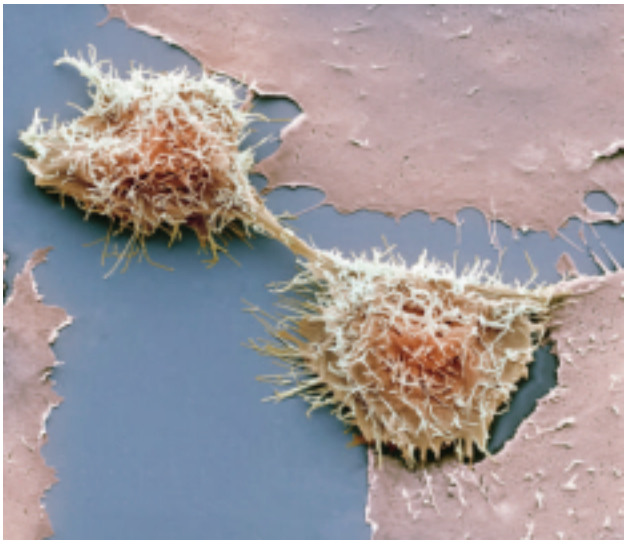


Figure 1-9 Dividing HeLa cells as seen by a scanning electron micrograph (colored). The image was taken during cell division (cytokinesis). The transient connecting midbody formed by microtubules can be seen. (Courtesy of Steve Gschmeissner/Photo Researchers, Inc.)

value of 2000 μm^3, with a range of 500–4000 μm^3 (BNID 100434). HeLa cells adhere to the extracellular matrix and, like many other cell types on a microscope slide, spread to a diameter of ≈40 μm (BNID 103718, 105877, 105878), but only a few thin μm in height. When grown to confluence, they press on each other to compact the diameter to ≈20 μm such that in one of the wells of a 96-well plate, they create a monolayer of ≈100,000 cells. As in bacteria and yeast, average cell size can change with growth condi tions. In the case of HeLa cells, a greater than twofold decrease in volume was observed when comparing cells three days and seven days after split ting and re-plating (BNID 108870, 108872). A snapshot of the variability of mammalian cells was achieved by a careful microscopic analysis of a mouse lymphocyte cell line, as shown in **Figure 1-10**. The distribution is

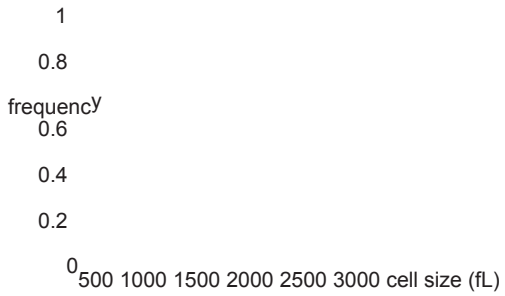


Figure 1-10 Distribution of cell sizes for L1210, a mouse lymphoblast cell line. The cell volumes are reported in units of fL (1 fL = 1 μm^3). (Adapted from Tzur A, Kafri R, LeBleu VS et al. [2009] ***Science*** 325:167–171.)

centered at about 1000 μm^3 with a variance of about 300 μm^3. To put these cellular sizes in perspective, if we think of *E. coli* as having the size of a human being, then a HeLa cell is about the size of a blue whale.

Our examination of the sizes of different cell types will serve as a jump ing-off point for developing intuition for a variety of other biological num bers we will encounter throughout the book. For example, when thinking about diffusion, we will interest ourselves in the time scale for particular molecules to traverse a given cell type, and this result depends in turn upon the size of those cells.

How big is a photoreceptor?

One of the greatest charms of biology is the overwhelming diversity of liv ing organisms. This diversity is reflected, in turn, by the staggering array of different types of cells found in both single-celled and multicellular organisms. Earlier, we celebrated some of the most important "model" cells, such as our standard bacterium, *E. coli* and our single-celled eukary ote, the yeast *S. cerevisiae*. Studies of these model systems have to be tem pered by a realization of both the great diversity of single-celled organisms themselves, as shown in the vignette on cell size diversity (pg. 21), as well as of the stunning specializations in different cell types that have arisen in multicellular organisms. The cells that make possible the sense of vision discussed in this vignette are a beautiful and deeply studied example of such specializations.

There is perhaps no sense that we each take more personally than our vision. Sight is our predominant means of taking in information about the world around us, a capacity made possible as a result of one of evolution's greatest inventions—namely, the eye—as shown in **Figure 1-11**. Eyes and the cells that make them up have been a central preoccupation of scientists of all kinds for centuries, whether in the hands of those like Hermann von Helm
holtz, who designed instruments such as the opthalmoscope to study eyes of living humans, or those like Charles Darwin and his successors, who have mused on how evolution could have given rise to such specialized organs. Chapter VI of *The Origin of Species* is entitled "Difficulties on Theory" and is used by Darwin as a forum to explain what he referred to as a "crowd of difficulties" that "will have occurred to the reader." Darwin says that some of these difficulties are "so grave that to this day I can never reflect on them without being staggered; but, to the best of my judgment, the greater num ber are only apparent, and those that are real are not, I think, fatal to my theory." One of the most significant of those difficulties was what Darwin

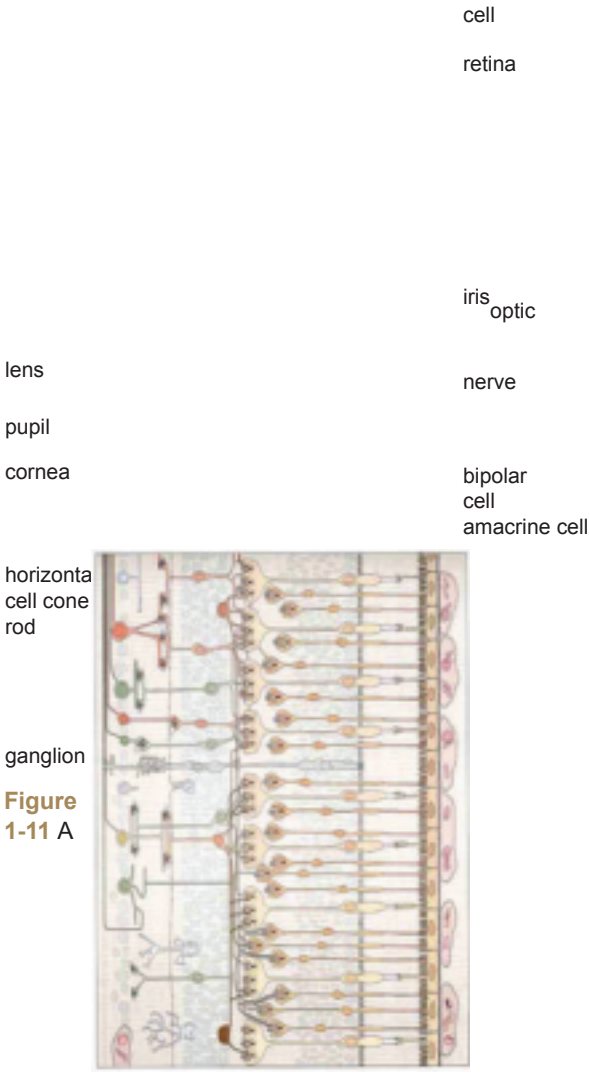


Figure 1-11 A multi-scale view of the retina. The schematic on the left shows the entire eye. The magnified view on the right illustrates the organization of the different cell types in the retina ranging from the photoreceptors that receive light to the ganglion cells that communicate electrical impulses as a result of stimulation of photoreceptors by light. (Adapted from Rodieck RW [1998]. The First Steps of Seeing. Sinauer Associates.)

thought of as “organs of extreme perfection,” such as our eye. He goes on to say, “To suppose that the eye, with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree. Yet reason tells me, that if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist; if further, the eye does vary ever so slightly, and the variations be inherited, which is certainly the case; and if any variation or modification in the organ be ever useful to an animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, can hardly be considered real.” Our under standing of the long evolutionary history of eyes continues to evolve itself, and a current snapshot can be attained by reading a recent review.[‡]

What are these organs of extreme perfection like at the cellular level? Fig ure 1-11 provides a multi-scale view of the human eye and the cells that

‡ Lamb T, Collin SP, & Pugh EN Jr. (2007) Evolution of the vertebrate eye: opsins, photore ceptors, retina and eye cup. *Nat Rev Neurosci* 8:960–976.

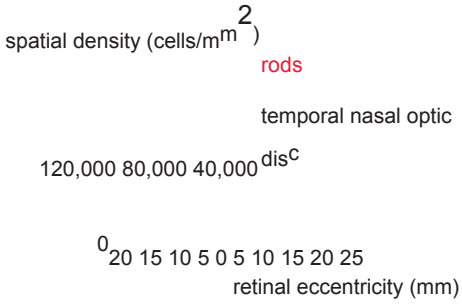


Figure 1-12 Distribution of rods and cones in the vertebrate retina. Note that if we consider 100,000 rods/mm^2 as the typical areal density, this corresponds to 10 μm^2 per rod cell, which jibes nicely with our simple estimate made above. (Adapted from Rodieck, RW [1988] The primate retina. In Comparative Primate Biology, Vol 4 [HD Steklis and J Erwin eds], pp. 203-278. Alan R. Liss Inc.)

make it work, giving a sense of the complexity and specialization that so staggered Darwin. Our focus here is on the retina, the 100–300-μm thick (BNID 109683) structure at the back of the eye. The mammalian retina harbors two types of photoreceptor cells, called rods and cones. Rods are mostly used for night vision, whereas cones enable color vision using three types of pigments. As seen in Figure 1-11, in addition to the rods and cones, the retina is also populated by layers of cells, such as horizontal cells, bipolar cells, amacrine cells, and the ganglion cells that convey the information derived from the visual field to the brain itself. One of the sur prising features of the human eye is that the photoreceptors are actually located at the back of the retina, whereas the other cells responsible for processing the data and the optic nerve that conveys that information to the brain are at the front of the retina, thus blocking some of the photons in our visual field. This seems a strange feature for an organ considered a glaring example of optimality in nature. Indeed, in cephalopods, like the squid and octopus, the situation is reversed, with the nerve fibers routing behind rather than in front of the retina. Further, the human eye structure is not optimal not only in this respect, but also in the aberrations it fea tures, many of which are corrected by the cells downstream of the photo receptors.§

The distribution of rods and cones throughout the retina is not uniform. As shown in **Figure 1-12**, cones have the highest density at a central part of the retina called the fovea and thus enable extremely high resolution. To get a feeling for the optical properties of this collection of photore
ceptors, it is perhaps useful to consider a comparison with digital cam eras. We are used to cameras with 10 million pixels per image.

§ Liang J & Williams DR (1997) Aberrations and retinal image quality of the normal human eye. *J Opt Soc Am A* 14:2873-2883.

a photoreceptor is much more functionally potent than a pixel, it is still interesting to contemplate how many photoreceptor cells we have and how this value compares to what we find in our cameras. To produce a naïve estimate of the number of photoreceptors in the human retina, we need a rough sense of how much area is taken up by each such cell. A human rod cell is ≈2 microns in diameter (BNID 107894), which is a few times the wavelength of light. If we maximally stacked them, we could get 500 by 500 such cells in a square millimeter—that is, ≈250,000 rods/ mm^2. Figure 1-12 reports experimental values that confirm this is close to reality. To finish the estimate of the total number of receptors decorating the back surface of the eye, we consider the eyeball to be a hemisphere of 2–3 cm in diameter, as shown in Figure 1-11 (BNID 109680), implying an area of roughly 10^9 μm^2. The number of photoreceptors can be estimated as (10^9 μm^2/retina)/(4 μm^2/photoreceptor), which yields ≈200 million photoreceptors in each of our eyes and is of the same order of magnitude as estimates based on current knowledge and visualization techniques (BNID 105347, 108321). Digital cameras still have a long way to go before they reach this number, not to speak of the special adaptation and pro cessing that each cell in our eye can perform and a digital pixel cannot.

The anatomy of these individual photoreceptors is remarkable. As seen in **Figure 1-13**, a typical photoreceptor cell such as a rod is roughly 100 μm

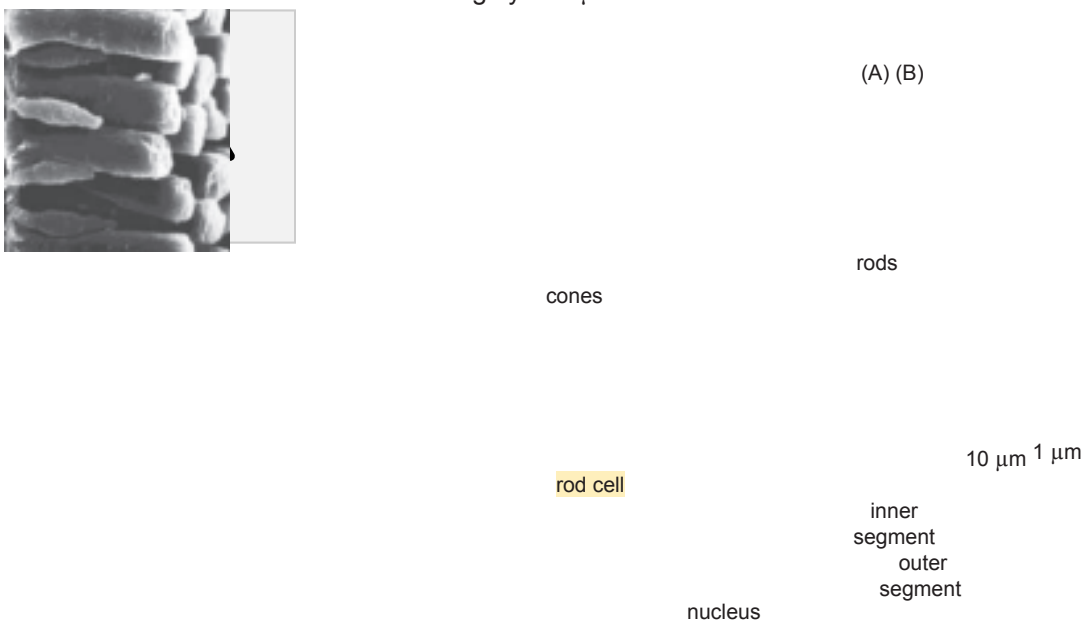


Figure 1-13 Anatomy of rods and cones. The schematic shows some of the key anatomical features of a photoreceptor cell. (A) A scanning electron micrograph illustrating the organization of rods and cones in the retina of a salamander. (B) Electron micrograph of the membrane disks of the outer segment of the photoreceptor. In both rods and cones, the proteins holding the light-absorbing retina are homologous opsins—that is, rhodopsins in

rods and three types of spectrally distinct photopsins in cones. (A, courtesy of Scott Mittman and David R. Copenhagen. B, adapted from Dowling J [2012] The Retina. Belknap Press.)

long (BNID 108246, 109684) and is characterized by a number of special ized features, such as the roughly 25-micron-long “outer segment” (BNID 107894, 107895) shown in Figure 1-13B that is filled with the rhodopsin molecules that absorb light. At the opposite extremity of these cells are the synapses—the structures used to communicate with adjacent cells. Synapses are crucial to the signal cascade that takes place following the detection of a photon by a photoreceptor cell. As seen in Figure 1-13, the outer segments of a photoreceptor rod cell are characterized by stacks of membrane discs. These discs are roughly 10 nm thick and are stacked in a periodic fashion with a spacing of roughly 25 nm. Given that the outer segment is ≈25,000 nm long, this means that there are roughly 1000 such discs in each of the $\approx 10^8$ rod cells in the vertebrate retina (with about 10^8 rhodopsin molecules per rod cell, as discussed in the vignette entitled “How many rhodopsin molecules are in a rod cell?”; pg. 142). These 1000 effective layers increase the cross section available for intercepting pho tons, thus making our eyes such “organs of extreme perfection.”

What is the range of cell sizes and shapes?

Cells come in a dazzling variety of shapes and sizes. As we have already seen, deep insights into the workings of life have come from focused studies on key “model” types such as *E. coli*, budding (baker’s) yeast, and certain human cancer cell lines. These model systems have helped develop a precise feel for the size, shape, and contents of cells. However, undue focus on model organisms can give a deeply warped view of the diversity of life. Stated simply, there is no easier way to dispel the myth of “the cell”—that is, the idea that what we say about one cell type is true for all others—than to show examples of the bizarre gallery of different cell types found both in unicellular and multicellular organisms.

In this vignette, we are interested in the broad question of the diversity of cell size and shape. Some representative examples summarizing the diver sity of shapes and sizes in the microbial world are shown in **Figure 1-14**. Though this figure largely confirms our intuitive sense that microbial cells are usually several microns in size, the existence of the giant *Thiomarga rita namibiensis* belies such simple claims in much the same way that the Star-of-David shape of *Stella humosa* is at odds with a picture of bacteria as nothing more than tiny rods and spheres.

Some of the most dramatic examples of cellular diversity include the beau tiful and symmetrical coccolithophore, *Emiliana huxleyi* (**Figure 1-15**),

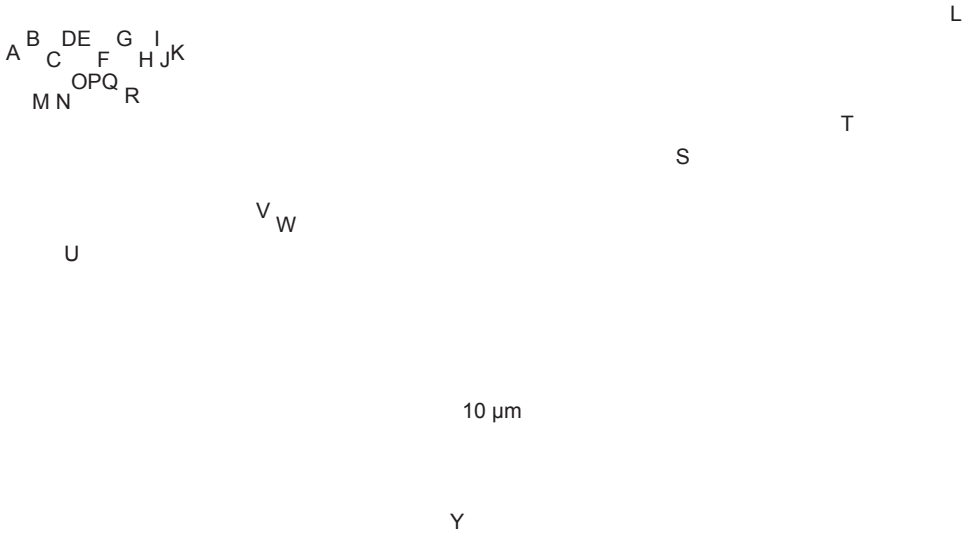


Figure 1-14 A gallery of microbial cell shapes. These drawings are based upon microscopy images from the original literature. (A) ***Stella strain*** IFAM1312 (380); (B) ***Microcyclus*** (a genus since renamed Ancylobacter) ***flavus*** (367); (C) ***Bifidobacterium bifidum***; (D) ***Clostridium cocleatum***; (E) ***Aquaspirillum autotrophicum***; (F) ***Pyroditium abyssi*** (380); (G) ***Escherichia coli***; (H) ***Bifidobacterium sp.***; (I) transverse section of ratoon stunt-associated bacterium; (J) ***Planctomyces sp.*** (133); (K) ***Nocardia opaca***; (L) chain of ratoon stunt associated bacteria; (M) ***Caulobacter sp.*** (380); (N) ***Spirochaeta halophila***; (O) ***Prosthecobacter fusiformis***; (P) ***Methanogenium cariaci***; (Q) ***Arthrobacter globiformis*** growth cycle; (R) gram-negative ***Alphaproteobacteria*** from marine sponges (240); (S) ***Ancalomicrobium*** sp. (380); (T) ***Newskia ramosa*** (133); (U) ***Rhodomicrobium vannielii***; (V) ***Streptomyces sp.***; (W) ***Caryophanon latum***; (X) ***Calothrix sp.***; (Y) A schematic of part of the giant bacterium, ***Thiomargarita namibiensis*** (290). All images are drawn to the same scale. (Adapted from Young KD ***obiol Mol Biol Rev*** 70:660–703.)

whose exoskeleton shield is very prominent and makes up the chalk rocks we tread on and much of the ocean floor, though its functional role is still not clear; the richly decorated protozoan, *Oxytricha* (for more single-cell

Figure 1-15 Scanning electron microscopy image of a collection of ***E. huxleyi*** cells, which illustrates their solid exterior. Each of these structures contains a single eukaryotic cell on its interior. 2 µm (Courtesy of Jeremy R. Young.)

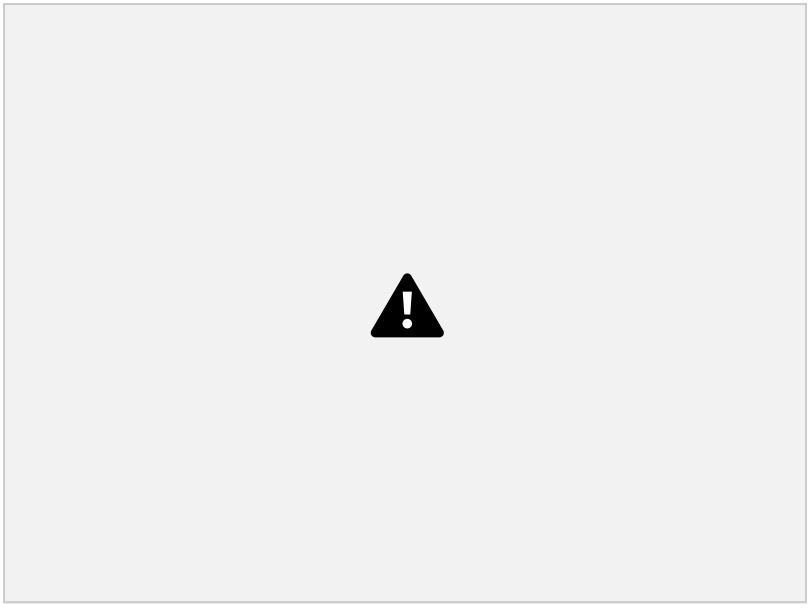


Figure 1-16 Protist diversity. This figure illustrates the morphological diversity of free-living protists. The various organisms are drawn to scale relative to the head of a pin about 1.5 mm in diameter. (Adapted from Finlay BJ [2002] *Science* 296:1061– 1063.)

protists, see the diversity and relative scale depicted in **Figure 1-16**); and the sprawling geometry of neurons, which can have sizes of over a meter (even in our own bodies). One of the most interesting classes of questions left in the wake of these different examples concerns the mechanisms for the establishment and maintenance of shape and the functional consequences of different sizes and shapes.

Perhaps the most elementary measure of shape is cell size, with sizes run ning from sub micron to meters, exhibiting roughly a seven order-of-mag nitude variability in cell sizes across the different domains of life, as shown in **Figure 1-17**. Though prokaryotes are typically several microns in size, sometimes they can be much larger. Similarly, even though eukaryotes typi cally span the range from 5 to 50 microns, they too have a much wider range of sizes, with the eggs of eukaryotes and the cells of the nervous system pro viding a measure of just how large individual cells can be. One of the most interesting challenges that remains in understanding the diversity of all of these sizes and shapes is gathering a sense of their functional implications and the evolutionary trajectories that gave rise to them.

(E)

(F)

(A) (B)

(D)

(G)

(C)

Figure 1-17 Cartoons of several different types of cells all referenced to a standard ***E. coli*** ruler of 1 micron width drawn in gray. (A) The protist, ***Giardia lamblia***, (B) a plant cell, (C) a budding yeast cell, (D) a red blood cell, (E) a fibroblast cell, (F) a eukaryotic nerve cell, and (G) a rod cell from the retina.

ORGANELLES

How big are nuclei?

One of the most intriguing structural features of eukaryotic cells is that they are separated into many distinct compartments, each characterized by differences in molecular composition, ionic concentrations, membrane potential, and pH. In particular, these compartments are separated from each other and the surrounding medium (that is, the cytoplasm or extra
cellular solution) by membranes that themselves exhibit a great diversity of lipid and protein molecules, with the membranes of different compartments also characterized by different molecular compositions. Given the central role of genomes in living matter, there are few organelles as important as

nucleus nucleus (A) (B)

Figure 1-18 Nuclear size. (A) Electron microscopy image of a yeast cell, which reveals the roughly 2-micron-sized nucleus. (B) A portion of a rat liver cell that shows part of the nucleus and a variety of surrounding organelles, such as the endoplasmic reticulum, the mitochondria, and the Golgi apparatus. (C) Fluorescence image of a human fibroblast cell with the roughly 10-micron nucleus labeled in green. (D) Light microscopy image of a human epithelial sheet. The dark ovals are the cell nuclei stained with silver. (A, adapted from Alberts B, Johnson A, Lewis J et al. [2015] Molecular Biology of the Cell, 6th ed. Garland Science. B, adapted from an electron micrograph in Fawcett DW [1966] The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders. With permission from Elsevier Inc. C, and D, adapted from Phillips R, Kondev J, Theriot J & Garcia H [2013] Physical Biology of the Cell, 2nd ed. Garland Science.)

the eukaryotic nucleus, home to the chromosomal DNA that distinguishes one organism from the next. As seen in **Figure 1-18**, using both electron and light microscopy, it is possible to determine nuclear size variation with typical diameters ranging between 2 and 10 microns, though in exceptional cases such as oocytes, the nuclear dimensions are substantially larger.

One feature of organellar dimensions is their variability. We have already seen the range of sizes exhibited by yeast cells in an earlier vignette. **Figure 1-19** takes up this issue again by revealing the typical sizes and variability for the nuclei in haploid and diploid yeast cells, complement
ing the data presented earlier on cell size. For haploid yeast cells, the mean nuclear volume is 3 μm^3 (BNID 104709). With a genome length of

0 0 10 20 30 40

cell area (μm²)

Figure 1-19 Nuclear size for haploid and diploid yeast cells. The cross-sectional areas of the nuclei are plotted as a function of the cross-sectional areas of the cells themselves. (Adapted from Jorgensen P, Edgington NP, Schneider BL et al. [2007] ***Mol Biol Cell*** 18:3523–3532. With permission from the American Society for Cell Biology.)

12 Mbp (BNID 100459), the DNA takes up roughly 0.3% of the nuclear volume. We can arrive at this estimate based on the rule of thumb that a base pair has a volume of ≈1 nm^3 (BNID 103778), in which case the DNA occupies roughly 0.01 μm^3. A similar value is found for diploid yeast. In contrast, for the yeast spore, the nuclear volume is an order of magnitude smaller—namely, 0.3 μm^3 (BNID 107660) or about 3% of the nuclear vol
ume—indicating a much more dense packing of the genomic DNA.

These estimates for nuclear fraction are agnostic about the higher-level chromatin structure induced by nucleosome formation. In nucleosomes, 147 base pairs of DNA are wrapped roughly 1.75 times around an octamer of histone proteins, making a snub disk roughly 10 nm across (BNID
102979, 102985). In **Figure 1-20**, we show the so-called 30-nm fiber. When we travel 10 nm along the fiber, about six nucleosomes are packed in a staggered manner, and thus we have included on the order of 1000 bp. We can estimate the total volume taken up by yeast genomic DNA when in this structure by multiplying the area of the effective circular cross section by the height of the structure, resulting in

$$V = \pi(15\ \text{nm})^2 \times (10\ \text{nm}/1000\ \text{bp}) \times (10^7\ \text{bp}) \approx 10^8\ \text{nm}^3 = 0.1\ \mu\text{m}^3. \quad (1.1)$$

Given the volume of the yeast nucleus of roughly 4 μm^3, this implies a packing fraction of ≈2%, and is consistent with our earlier estimate, which was based on the volume of a base pair.

Questions about nuclear size in eukaryotes have been systematically investigated in other organisms besides yeast. It has been hypothesized

(A) (B)

histone octamer

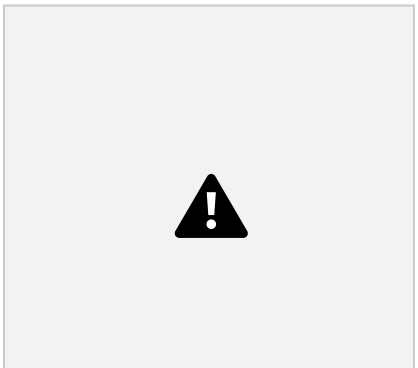
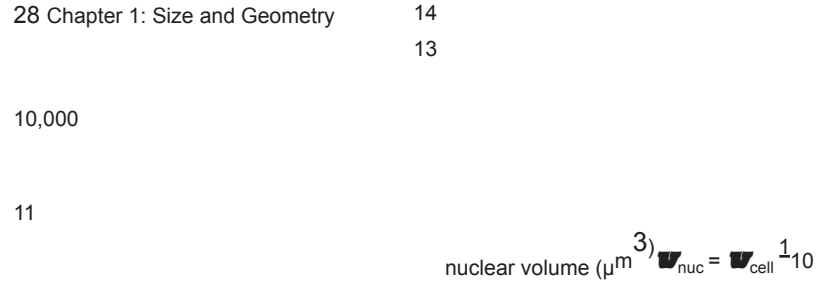


Figure 1-20 DNA packing into higher-level compact structures. (A) Schematic illustrating how multiple nucleosomes can be arranged into a solenoidal structure. The histone octamer shown in yellow and the DNA as a red strand. (B) Models of nucleosome packing based upon high-resolution cryo-electron microscopy images of arrays of nucleosomes. In these ***in vitro*** experiments, nucleosome arrays were generated by using purified histones and specific DNA molecules of known sequence. (B, adapted from Robinson P, Fairall L, Huynh V & Rhodes D [2006] ***Proc Natl Acad Sci USA*** 103:6506–6511.)

that there is a simple linear relationship between the mean diameter of a plant meristematic cell (the plant tissue consisting of undifferentiated cells from which growth takes place) and the diameter of its nucleus. Such ideas have been tested in a variety of different plant cells, as shown in **Fig**
ure 1-21, for example. In the experiments summarized there, the nuclear and cell volumes of 14 distinct species of herbaceous angiosperms, includ ing some commonly known plants such as chickpeas and lily of the valley, were measured, resulting in a simple relationship of the following form (BNID 107802)

$$V_{nuc} \approx 0.2\, V_{cell}\,. \quad (1.2)$$

The observations reported here raise the question of how the relative size of the nucleus compared to the whole cell is controlled. This is especially compelling since the nucleus undergoes massive rearrangements dur ing each and every cell cycle as the chromosomes are separated into the daughter cells. We remind the reader that a relatively stable ratio is a com mon observation rather than a general law. In mammalian cells, this ratio can be very different between cell types. For example, in resting lympho cytes, the nucleus occupies almost the whole cell, while in macrophages or fat cells, the ratio of nucleus to cell volume is much smaller.



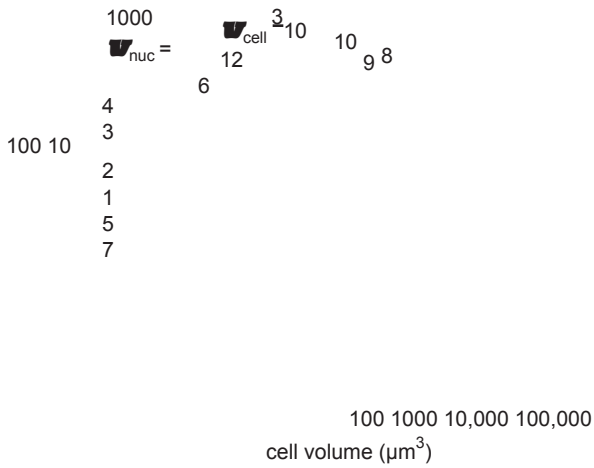


Figure 1-21 The relationship between the nuclear volume and cell volume in apical meristems of 14 herbaceous angiosperms. 1 ***Arabidopsis thaliana***, 2 ***Lobularia maritime*** (Sweet Alison), 3 ***Hypericum virginicum*** (Marsh St. John's wort), 4 ***Cicer arietinum*** (chickpea), 5 ***Nelumbo lutea***, 6 ***Spinacia oleracea*** (spinach), 7 ***Cyanotis pilosa***, 8 ***Anemone pulsatilla*** (Meadow Anemone), 9 ***Tradescania navicularis*** (day flower), 10 ***Convallaria majalis*** (lily of the valley), 11 ***Fritillaria lanceolata*** (chocolate lily), 12 ***Fritillaria camtschatcensis***, 13 ***Lilium longiflorum*** (Easter lily)(4×), 14 ***Sprekelia formosissima*** (Aztec lily). (Adapted from Price HJ, Sparrow AH & Nauman AF [1973] ***Experientia*** 29:1028–1029.)

How big is the endoplasmic reticulum of cells?

The endoplasmic reticulum, known to its friends as the ER, is often the largest organelle in eukaryotic cells. As shown in **Figure 1-22**, the struc ture of the ER is made up of a single, continuous membrane system, often spreading its cisternae and tubules across the entire cytoplasm. In addi tion to its exquisite and beautiful structure, it serves as a vast processing unit for proteins, with ≈20–30% of all cellular proteins passing through it as part of their maturation process (BNID 109219). As another indication of the demands made on the ER, we note that a mature secreting B cell can secrete up to its own weight in antibody each day (BNID 110220), all of which must first be processed in the ER. The ER is also noted for producing most of the lipids that make up the cell's membranes. Finally, the ER is the main calcium deposit site in the cell, thus functioning as the crossroads for various intracellular signaling pathways. Serving as the equivalent of a corporate mailroom, the ER activity, and thus its size, depends on the state of the cell.

When talking about the "size" of organelles such as the ER, there are several different ways we can characterize their spatial extent. One perspective is

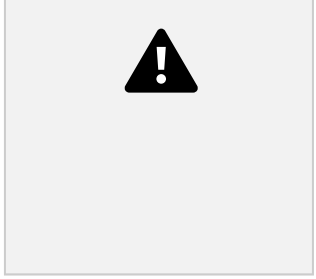


Figure 1-22 Structure of the endoplasmic reticulum. The left panel shows a thin section electron micrograph of the region surrounding the nucleus in an acinar cell that comes from the pancreas of a bat. The schematic illustrates the connected membrane morphology of the ER, which is contiguous with the nuclear membrane. (Adapted from Fawcett DW [1966] The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders.)

to compare the total membrane area tied up in the organelle of interest rel ative to that of the plasma membrane, for example. A second way of char acterizing the spatial extent of the organelle is by appealing to the volume enclosed within the organelle of interest and comparing it to the total cell volume. As can be inferred from the electron micrograph image of an aci nar cell from the pancreas in charge of secretion (Figure 1-22), the undulat ing shape of the convoluted ER membrane ensures that its surface area is actually 10–20 times larger than the outer surface area of the cell itself (the plasma membrane). The distribution of membrane surface area among different organelles in liver and pancreatic cells is quantified in **Table 1-4**. The table shows that the membrane area allocation is dominated by the ER (as much as 60%), followed by the Golgi and mitochondria. The cell plasma membrane in these mammalian cells tends to be a small fraction of less than 10%. In terms of volume, the ER can comprise >10% of the cel lular volume, as shown in **Table 1-5**.

In recent years, the advent of both fluorescence microscopy and tomo graphic methods in electron microscopy have made it possible to con struct a much more faithful view of the full three-dimensional structure of these organelles. One of the insights to emerge from these studies is the recognition that they are made up from a few fundamental structural units—namely, tubules, which are 30–100 nm in diameter (BNID 105175, 111388), and sheets, which bound an internal space known as the ER lumen, as shown in Figure 1-22. As with studies of other organelles, such

percentage of total cell membrane			
membrane type	**liver hepatocyte**	**pancreatic exocrine cell**	
plasma			
rough ER			
smooth ER			

Golgi apparatus	peroxisome	5
mitochondria outer	endosome	60 <1
mitochondria inner	2	10 4
nucleus inner	35 16 7	17
secretory vesicle	7	0.7 3
lysosome	32 0.2 — 0.4 0.4	— — —
	0.4	

Table 1-4 The percentage of the total cell membrane of each membrane type in two model cells. The symbol "—" indicates that the value was not determined. (Adapted from Alberts B, Johnson A, Lewis J et al. [2015] Molecular Biology of the Cell, 6th ed. Garland Science.)

as the mitochondria, early electron microscopy images were ambiguous since in cross section, even planar cisternae have a tubular appearance. The more recent three-dimensional membrane reconstructions have clarified such issues by making it possible to actually see tubular struc
tures unequivocally and to avoid mistaking them for cuts through pla nar structures. These more detailed studies have revealed that the ER's fundamental structures are spatially organized, with the sheets being predominant in the perinuclear ER and tubules found primarily at the peripheral ER. Thus, it appears that the various parts of the cell "see" dif ferent ER architecture. The ER is in contact with most organelles through membrane contact sites. For example, the mitochondria–ER contact site is composed of a complex of membrane proteins that span either

intracellular compartment	percentage of total cell volume
cytosol	50–60 20
mitochondria	10
rough ER cisternae	6
smooth ER cisternae plus	6
Golgi cisternae nucleus	1
peroxisomes	1
lysosomes	1
endosomes	

Table 1-5 The volume fraction occupied by different intracellular compartments in a liver hepatocyte cell. (Adapted from Alberts B, Johnson A, Lewis J et al. [2015] Molecular Biology of the Cell, 6th ed. Garland Science.)

anaphase telophase cytokinesis early interphase

ER marker (Sec61β)
chromosome marker (H2B)

10 μm

Figure 1-23 Structural dynamics of the endoplasmic reticulum during the cell cycle. Confocal images of HeLa cells. The chromosomes are labeled in red using a fusion of a fluorescent protein with histone H2B. The ER is labeled in green by virtue of a fusion of a fluorescent protein to a molecular member of the ER segregation apparatus (Sec61β-GFP). The sequence of images shows the changes in ER morphology as a function of time during the cell cycle. (Adapted from Lu L, Ladinsky MS & Kirchhausen T [2009] *Mol Biol Cell* 20:3471–3480.)

organelle. Similar contacts are found between the ER and the vacuole, peroxisome, and cell membrane.

One of the deceiving aspects of images like those shown in Figure 1-22 is that they give the illusion that these structures are static. However, given the cell's imperative to reproduce itself, it is clear that during the process of cell division, when the nuclear envelope dissolves away, the ER must undergo substantial rearrangement as well, cutting it in two parts to later re-engulf the two nuclei to be. Beautiful recent studies have made it pos
sible to watch the remodeling of the ER structure during the cell cycle in real time, as shown in **Figure 1-23**. By making a stack of closely spaced confocal images, it is possible to gain insights into the three-dimen
sional structure of the organelle over time. In these images, we see that during interphase the ER is reticular (netlike). To appreciate the tangled arrangement of organellar membranes even more deeply, **Figure 1-24** provides a reconstructed image using X-ray microscopy of the ER and other ubiquitous membrane systems in the cell. In this cell type and under

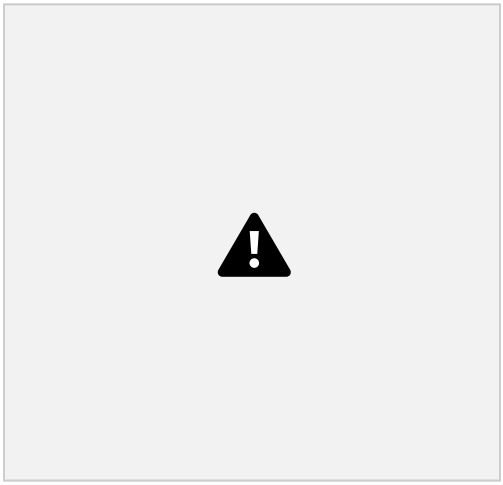


Figure 1-24 X-ray microscopy image of cellular ultrastructure highlighting the endoplasmic reticulum. This image is a volumetric rendering of images of a mouse adenocarcinoma cell. The numbers represent percent of the volume occupied by the different compartments. (Adapted from Schneider G, Guttmann P, Heim S et al. [2010] *Nature Meth* 7:985–987.)

mitochondria
endoplasmic reticulum vesicles
external
13% 17% 3%
2%
65%
lysosomes

these growth conditions, the reconstruction reveals that the mitochon dria and lysosomes are more dominant in terms of volume than the ER. The cytoplasm itself occupies more than one-half of the volume, even if it is deemed transparent in these reconstructions, which take a wide slice (depth of focus) and project it into a dense two-dimensional

image. Struc tural images like these serve as a jumping-off point for tackling the utterly mysterious microscopic underpinnings for how the many complex mem brane structures of the ER and other organelles are set up and change dur ing the course of the cell cycle.

How big are mitochondria?

Mitochondria are famed as the energy factories of eukaryotic cells, the seat of an array of membrane-bound molecular machines synthesizing the ATP that powers many cellular processes. It is now thought that mitochondria in eukaryotic cells came from some ancestral cell that took up a prokar
yote through a process such as endocytosis or phagocytosis and, rather than destroying it, opted for a peaceful coexistence in which these former bacteria eventually came to provide the energy currency of the cell. One of the remnants of this former life is the presence of a small mitochondrial genome that bears more sequence resemblance to its prokaryotic precur
sors than to its eukaryotic host.

Beyond their fascinating ancestry, mitochondria are also provocative due to their great diversity in both size and shape. Though probably familiar to many for the morphology depicted in **Figure 1-25**, with its character istic micron-sized, bacterium-like shape and series of internal lamellae shown in magnified form in Figure 1-25, mitochondria have in fact a host of different structural phenotypes. These shapes range from onion-like morphologies to reticular structures such as those shown in Figures 1-25C and 1-25D, in which the mitochondrion is one extended object, to a host of other bizarre shapes that arise when cells are exposed to an oxygen-rich environment or that emerge in certain disease states. These reticular mito
chondria can spread over tens of microns.

As shown in **Figure 1-26**, electron microscopy images of mitochondria encourage their textbook depiction as approximately spherocylindrical in shape (that is, cylinders with hemispherical caps), with a length of roughly two microns and a diameter of roughly one micron. These organelles are characterized by two membrane systems that separate the space into three distinct regions—namely, the space external to the mitochondrion, the intermembrane space between the mitochondrial inner and outer

compact reticular

(A)

nucleus Golgi apparatus
lysosome

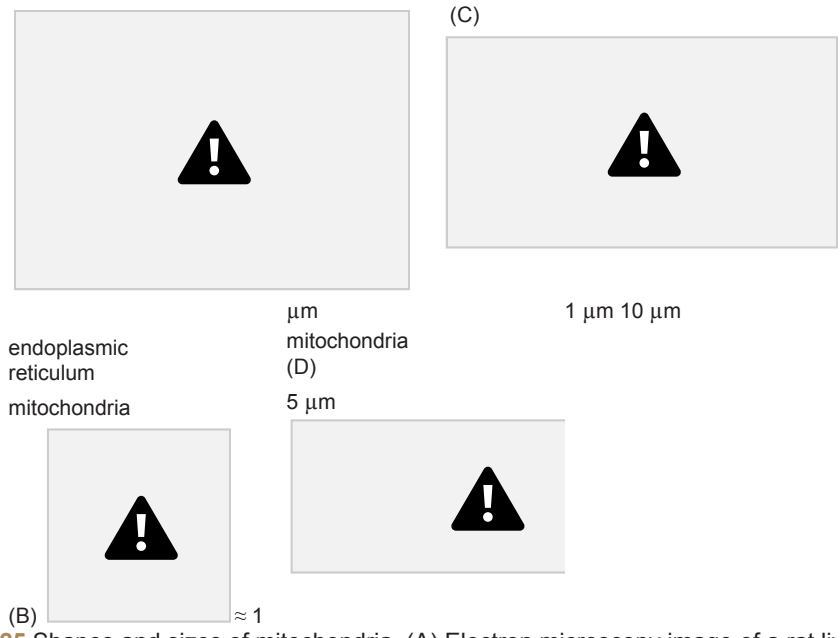


Figure 1-25 Shapes and sizes of mitochondria. (A) Electron microscopy image of a rat liver cell, which highlights many of the important organelles and illustrates the size and shape of mitochondria. (B) Cryo electron microscopy reconstruction of the structure of a lamellar mitochondrion. (C) Reticular structure of mitochondria in a budding yeast cell. Bud scars are labeled separately in blue. (D) Reticular mitochondrial network in a PtK2 kangaroo rat cell. The mitochondria are visible in green and were labeled with an antibody against the proteins responsible for the transport of proteins across the mitochondrial membranes. The tubulin of the microtubules are labeled in red and the nucleus is shown in blue. (A, adapted from Fawcett DW [1966] The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders. With permission from Elsevier Inc. B, courtesy of Terry Frey and Guy Perkins. C, adapted from Egner A, Jakobs S & Hell SW [2002] ***Proc Natl Acad Sci USA*** 99:3370–3375. With permission from the National Academy of Sciences. D, adapted from Schmidt R, Wurm CA, Punge A et al. [2009] ***Nano Lett*** 9:2508–2510.)

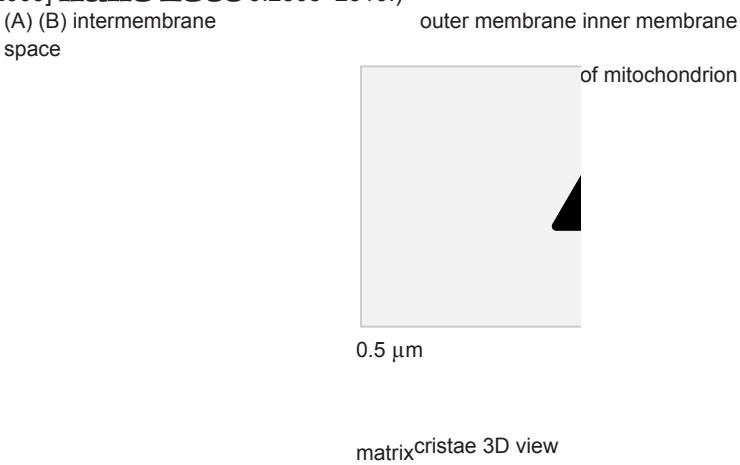


Figure 1-26 The structure of a mitochondrion. (A) Electron microscopy image of a mitochondrion from the pancreas of a bat. (B) Schematic illustrating the three membrane spaces relevant to mitochondria as well as the connectivity. (A, adapted from Fawcett DW [1966] The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders.)

membranes, and the matrix, which is the volume enclosed by the inner

membrane. Different mitochondrial morphologies all respect this basic organizational connectivity.

How many mitochondria are in a cell? A characteristic order of magni tude for yeast would be 10^1 and for mammalian cells it would be 10^3–10^4, but beware that the very idea of "counting" mitochondria can be tricky in many cases, because the mitochondria are sometimes reticular and indis tinct, peanut-shaped objects. For example, yeast cells grown on ethanol contain a larger number (20–30; BNID 103070) of small, discrete mito chondria, whereas when these same cells are grown on glucose, they con tain a smaller number (~3; BNID 103068) of large, branched mitochondria. These distinct morphologies do not significantly affect the fraction of the cellular volume occupied by the mitochondria and probably relates to the different demands in a respiratory-versus-respiro-fermentative lifestyle.

How big are chloroplasts?

Chloroplasts play a key role in the energy economy of the cells that harbor them. Chloroplasts are less well known than their mitochondrial counter parts, though they are usually much larger and play a key role in producing the reduced compounds that store the energy that is then broken down in mitochondria. Chloroplasts have the pivotal role in the biosphere of car rying out the chemical transformations linking the inorganic world (CO_2) to the organic world (carbohydrates). This feat of chemical transformation enables the long-term storage of the fleeting sun's energy in carbohydrates and its controlled release in energy currencies such as ATP and NADPH. Those same carbon compounds also serve to build all of the biomass of cells as a result of downstream metabolic transformations.

Chloroplasts in vascular plants range from being football- to lens-shaped and, as shown in **Figure 1-27**, have a characteristic diameter of ≈4–6 microns (BNID 104982, 107012), with a mean volume of ≈20 μm^3 (for corn seedling; BNID 106536). In algae, they can also be cup-shaped, tubular, or even form elaborate networks, paralleling the morphological diversity found in mitochondria. Though chloroplasts are many times larger than most bacteria, in their composition they can be much more homogene ous, as required by their functional role, which centers on carbon fixation. The interior of a chloroplast is made up of stacks of membranes, in some ways analogous to the membranes seen in the rod cells found in the visual systems of mammals. The many membranes that make up a chloroplast are fully packed with the apparatus of light capture, photosystems, and

chloroplast envelope vacuole

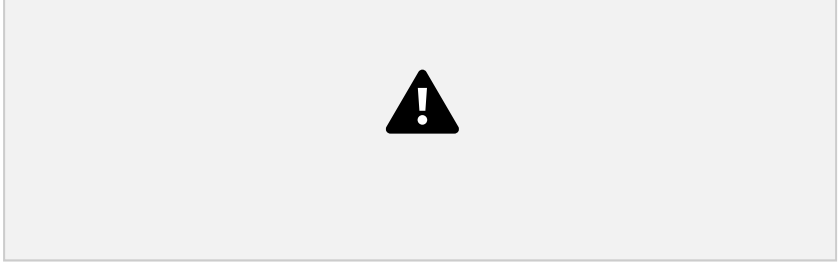


Figure 1-27 Electron micrograph of a chloroplast. The light reactions occur in the membrane-bound compartment called the thylakoid. There are usually about 40–60 stacks of disks termed grana per chloroplast (BNID 107013) covering 50–70% of the thylakoid membrane surface (BNID 107016). Each single stack has a diameter of 0.3–0.6 μm (BNID 107014). The sugar produced is stored in starch granules.
(Adapted from Alberts B, Johnson A, Lewis J et al. [2015] Molecular Biology of the Cell, 6th ed. Garland Science.)

related complexes. The rest of the organelle is packed almost fully with one dominant protein species—namely, Rubisco, the protein serving to fix CO_2 in the carbon fixation cycle. The catalysis of this carbon fixation reaction is relatively slow, thus necessitating such high protein abundances.

The number of chloroplasts per cell varies significantly between organ isms. Even within a given species, this number can change significantly, depending upon growth conditions. In the model algae, *Chlamydomonas reinhardtii*, there is only one prominent cup-shaped chloroplast per cell, whereas in a typical photosynthetic leaf cell (mesophyll) from plants such as *Arabidopsis* and wheat, there are about 100 chloroplasts per cell (BNID 107030, 107027, 107029). A vivid example from a moss is s

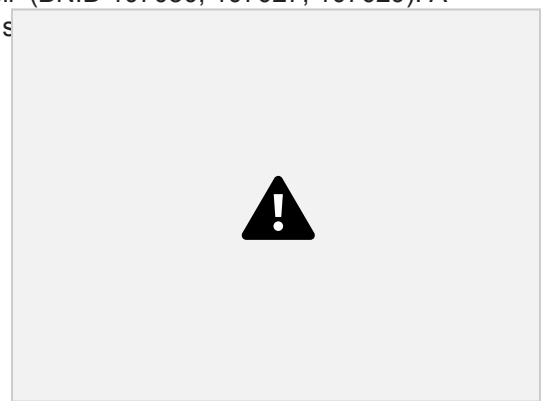


Figure 1-28 Chloroplasts in the moss ***Plagiomnium affine***, found in old-growth boreal forests in North America, Europe, and Asia, growing in moist woodland and turf. The-lamina cells shown here are elongated, with a length of about 80 microns and a width of 40 microns. These cells, as with most plant cells, have their volume mostly occupied by large vacuoles, so the cytoplasm and chloroplasts are at the periphery. Chloroplasts also show avoidance movement, in which they move from the cell

surface to the side walls of cells under high light conditions to avoid photodamage. (Courtesy of Ralf Wagner.) 10 μm

in **Figure 1-28**. Each chloroplast has tens to hundreds of copies (BNID 107105, 107107, 107108) of the chloroplast genome, which is ≈100 kbp in length (BNID 105918). This creates a fascinating challenge of how to balance the expression of genes that are coded in the chloroplast genome at thousands of gene copies per cell with the expression of genes that have a single copy in the main nuclear genome. In some cases, such as the protein Rubisco, they form a complex at one-to-one stoichiometric ratios!

Much evidence points to the idea that chloroplasts originated in a process of endosymbiosis—that is, they were originally free-living cells (probably photosynthetic cyanobacteria) that were engulfed (or enslaved) a billion years ago (BNID 107041) by cells that became their new hosts. With time, these originally distinct cells forged a tight collaboration in which most genes transferred from the engulfed cell to the host nucleus, in much the same way that the mitochondrial genome obtained its tiny size. From genomes that probably originally contained over 3000 genes, only about 130 genes remain in the chloroplasts of contemporary plants (BNID 106553, 106554).

These processes of engulfment followed by adaptation can still be observed today. Through a process known as kleptoplasty, different organisms (ranging from dinoflagellates to sea slugs) are able to digest algae while keeping the chloroplasts of these algae intact. These captured plastids are kept functional for months and are used to "solar power" these organisms. Not only the act of engulfing, but also the slower process of adaptation between the host and the organelle, can be observed. In one study it was determined that in 1 out of ~10,000 pollen grains, a reporter gene is transferred from the chloroplast to the nuclear genome (BNID 103096). How can such a low value be assessed reliably? A drug-resistance gene that can only function in the nucleus was incorporated into the chloroplasts of tobacco plants. Pollen from these plants was used to pollinate normal plants. Next, 250,000 seeds were screened and 16 showed resistance to the drug. Now here is the catch: Chloroplast genomes are transferred only through the mother. The pollen has only nuclear genes. The only way for the resistance gene to arrive through the pollen was shown to be through infiltration from the chloroplast genome into the nuclear genome. Measuring the rate of this process gives some insight into how genomes of organelles can be so small. It leaves open the question of what is the selective advantage of transferring the genomic information from the organelle's DNA to the central cell repository in the nucleus.

All told, chloroplasts are organelles of great beauty and sophistication. Their intriguing evolutionary history is revealed in their compact genomes. Structurally, their stacked membrane systems provide a critical system for capturing light and using its energy for the synthesis of the carbohydrates that are at the center of food chains across Earth.

How big is a synapse?

So far in the book, we have mainly focused on the sizes of individual cells and the organelles within them. Multicellularity, however, is all about partnerships between cells. A beautiful example of our own multicellu larity is provided by the cells in our nervous system. These cells are part of a vast and complex array of interactions that are only now beginning to be mapped. The seat of interactions between neighboring neurons are synapses, the interface between cells in which small protrusions adopt a kissing configuration, as seen in **Figure 1-29** and **Figure 1-30** for the cases of a neuromuscular junction and a synapse in the brain, respectively. These synapses are responsible for the propagation of information from one neuron to the next. Interestingly, information transmission in the nervous system is partly electrical and partly chemical. That is, when an action potential travels along a nerve, it does so by transiently changing the transmembrane potential from its highly negative resting value to a nearly equal positive potential. When the action potential reaches the syn apse, this leads to vesicle fusion and subsequent release of chemical sig
nals (neurotransmitters), which induce channel gating in the neighboring cell with which it has formed the synapse. This results, in turn, in an action potential in the neighboring cell. As seen in Figures 1-29 and 1-30, the synapse is composed of a presynaptic terminal on the axon of the trans
mitting neuron and a postsynaptic terminal with a so-called synaptic cleft between them. The total number of such synapses in the human brain has been vaguely stated to be in the range of 10^{13}–10^{15} (BNID 106138, 100693), with every cubic millimeter of cerebral cortex having about a billion such synapses (BNID 109245).

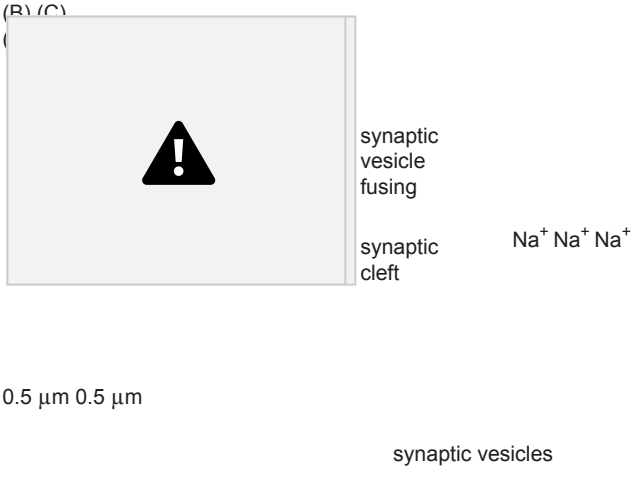


Figure 1-29 Structure of a neuromuscular junction. (A) Electron microscopy image of a nerve terminal and its synapse with a neighboring cell in a neuromuscular junction. (B) Cryo-electron microscopy reconstruction image of a fraction of the presynaptic neuron showing the synaptic vesicles it harbors for future release. (C) Schematic of a synapse. Note that the synaptic cleft, vesicles, etc., are not drawn to scale. (A, and B, adapted from Rizzoli SO & Betz WJ [2005] *Nat Rev Neurosci* 6:57–69.)