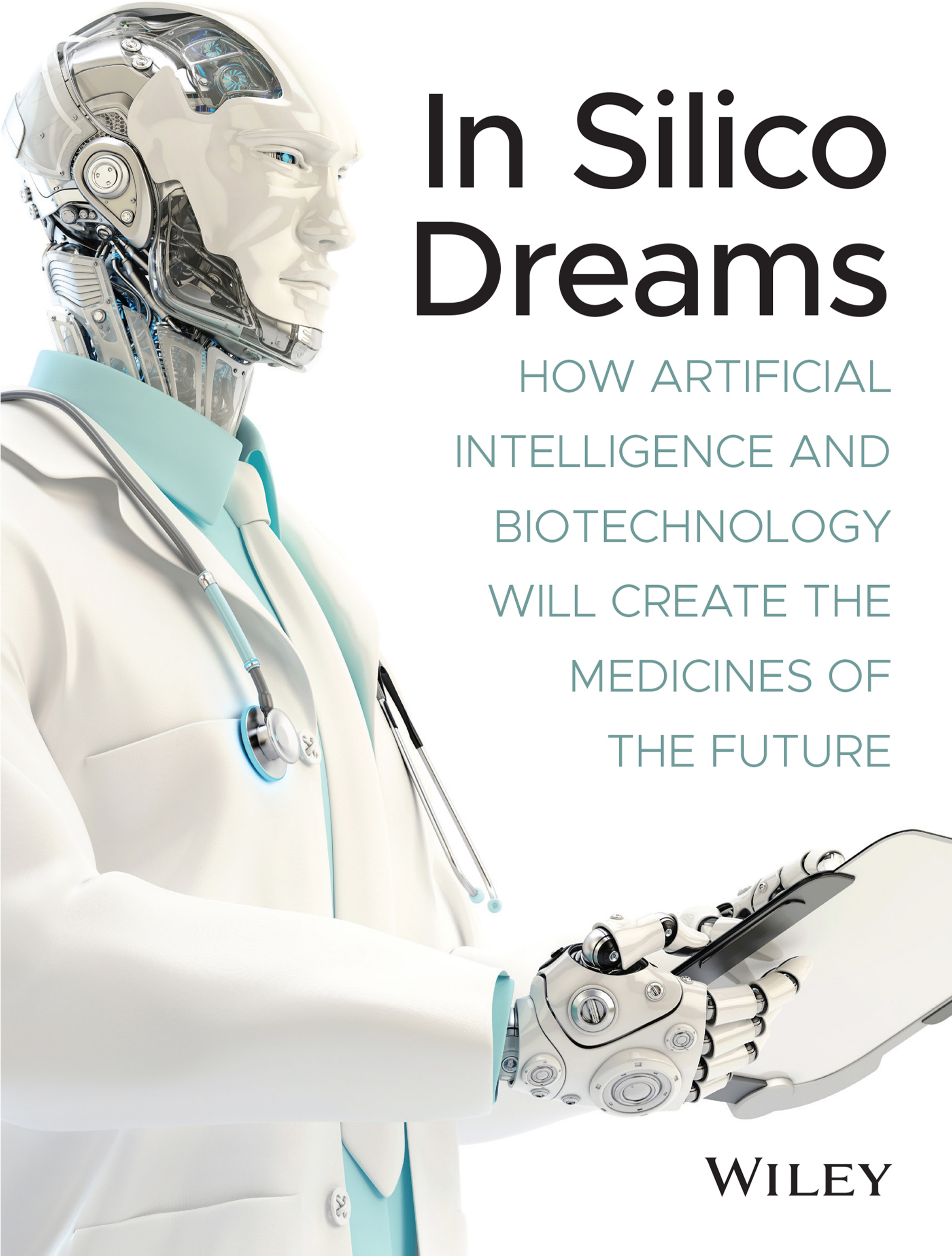
BRIAN HILBUSH

In Silico Dreams

HOW ARTIFICIAL INTELLIGENCE AND BIOTECHNOLOGY WILL CREATE THE MEDICINES OF THE FUTURE



WILEY

In Silico Dreams

In Silico Dreams

How Artificial Intelligence and Biotechnology Will Create the Medicines of the Future

Brian Hilbush

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

ISBN: 978-1-119-74557-0
ISBN: 978-1-119-74564-8 (ebk)
ISBN: 978-1-119-74563-1 (ebk)

Library of Congress Control Number: 2021941419

Cover Image: © Vladyslav Otsiatsia/Getty Images
Cover Design: Wiley

About the Author

Brian Hilbush is a biotechnology and genomics expert with three decades of executive-level business and scientific management experience in the life sciences. Brian has worked at Fortune 500 companies, consulted for the pharmaceutical and semiconductor sectors, and commercialized technology platforms in genomics. Brian co-founded ModGene, where he built a genetics-based drug target discovery platform with a focus on neurodegenerative diseases. His technical expertise resides at the intersection of genomics, artificial intelligence, and bioinformatics. Brian has led technical teams developing software and analytics pipelines for human genomics studies, neuroanatomy, and gene expression. He has co-authored numerous scientific publications and was co-inventor on patents for novel treatments of Alzheimer's disease. Brian received his BS in cell and molecular biology from the University of Washington and his PhD in neuroscience from Stony Brook University in 1991. For more information on this book, please visit insilicodreams.com.

About the Technical Editor

Ted Coombs is a polymath futurist, technologist, and author of 24 technology books. His career in tech started in the 1970s as a laser engineer and roboticist. He began working in AI in 1983 for the Cleveland Clinic. He is one of the founders of modern computer forensics and today is also known for his knowledge of AIOps and DataOps. Ted is also an accomplished musician and fine artist.

Acknowledgments

This book was written entirely during the period spanning the COVID-19 pandemic. Prior to the outbreak, I discussed ideas for a book with Kenyon Brown, senior content acquisitions editor at Wiley, who provided suggestions on topics and content flow. It was out of these discussions that *In Silico Dreams* was formed; I thank Ken for his input at inception and guidance along the way. I also thank Gary Schwartz for expert editing and contributions to the flow of the book and to Ted Coombs for technical review and suggestions. I am also appreciative of the support of Christine O'Connor, managing editor; Saravanan Dakshinamurthy, Rachel Fogelberg, and the production staff; and Michael Trent for developing the cover design.

I gratefully acknowledge discussions, exchanges, and other communications that I had with individuals whose works and quotes are contained herein, including David Van Essen, Dan Felleman, Tushar Pandey, Philippe Schwabe, and Michael Wu for permission to use original figures, Emmanuelle Charpentier for clarifications on the potential of CRISPR technology, and many others in the STEM communities.

My intention in writing this book was to capture for the reader the intense pace of research and innovation occurring across many disciplines that are converging to impact medicine's future. Because of the enormous volume of work that has transpired over the past decade in artificial intelligence and biotechnology, it was impossible to cover all significant advances. I apologize for those omissions and for any errors in interpretation of technologies and research findings that I did cover as part of this project.

My personal journey inspired thoughts on many pages of this book. My meanderings across the once-separate fields of neuroscience, molecular biology, and computer science began in earnest while I was a freshman at the University of Washington in Seattle. I thank Donna Simmons for opening the door to neuroscience with a stint in her histology laboratory; Vijay Sarthy for allowing me to explore retinal neurochemistry and molecular biology in his research group; and Joel Levine for his patience and mentorship while I was a doctoral candidate in his lab, along with Simon Halegoua and many others at Stony Brook University. My first foray into industry was with James Morgan at the Roche Institute of Molecular Biology, a part of the pharmaceutical giant Hoffmann-La Roche in Nutley, New Jersey. I thank Jim for allowing me to plunge headlong into molecular biology.

I discovered the nascent power of combining computational approaches and software with biotechnology while working with Karl Hasel and Gregor Sutcliffe at Digital Gene Technologies in La Jolla, California. I thank Chuck for the opportunity and Greg for being a mentor and partner over the past two decades. To Graham Gaylard and the team at Real Time Genomics that provided me with a speed course in Reverend Bayes' theorem and machine learning, thank you.

I was supported in so many ways by my incredible wife Melanie Greenberg and daughter, Sydney, who both gave me the space and encouragement to complete this book. To Melanie and Sydney, thank you for your kindness and love.

—Brian Hilbush

Contents at a Glance

Contents

Introduction

We have entered an unprecedented era of rapid technological change where developments in fields such as computer science, artificial intelligence (AI), genetic engineering, neuroscience, and robotics will direct the future of medicine. In the past decade, research organizations around the globe have made spectacular advances in AI, particularly for computer vision, natural language processing and speech recognition. The adoption of AI across business is being driven by the world's largest technology companies. Amazon, Google, and Microsoft offer vast, scalable cloud computing resources to train AI systems and platforms on which to build businesses. They also possess the talent, resources, and financial incentives to accelerate AI breakthroughs into medicine. These tech giants, including Apple, are executing on corporate strategies and product roadmaps that take them directly to the heart of healthcare. Every few weeks, a new AI tool is announced that performs a medical diagnostic procedure at human levels of performance. The pace of innovation in the tech sector is exponential, made possible by continual improvements and widespread availability of computing power, algorithmic design, and billions of lines of software programming code. Technology's influence on the sciences has been profound. Traditional disciplines such as biology and chemistry are being transformed by AI and data science to such an extent that new experimental paradigms have emerged for research and the pharmaceutical industry.

Biotechnology's growth and innovation cycles are equally impressive. Startling advances have been made to move the field from simple gene cloning experiments using viral and bacterial genetic material in test tubes to performing gene editing at precise locations in the human genome. A new generation of gene therapy and T-cell engineering companies are building tools to equip the immune systems of patients to destroy cancer. Explosive growth in data-generating capabilities from DNA sequencing instruments, medical imaging, and high-resolution microscopy has created a perfect storm of opportunities for AI and machine learning to analyze the data and produce biological insights. Out of this milieu, the first generation of tech-inspired startups has emerged, initiating the convergence of AI and biotechnology. These young companies are taking aim at the conventional path of drug development, with the brightest minds and freshest ideas from both fields providing a new base of innovation for the pharmaceutical industry.

This book tells the story of the impact of innovations in biology and computer science on the future of medicine. The creation of a new industry based on therapeutic engineering has begun. Nearly 200 years ago, Emmanuel Merck saw a commercial opportunity to produce the pain-killing substance from the opium poppy, which was in widespread use across Europe and beyond. He was inspired by Fredrich Sertürner′s innovative process for the extraction of the opiate alkaloid. Sertürner gave the newly purified narcotic substance the name morphium, after the Greek god of dreams. For thousands of years before these Germans helped to launch the pharmaceutical industry, medicinal compounds derived from nature had been concocted into noxious mixtures of uncertain potency by alchemists, physicians, or shamans in all cultures. With the elucidation of the rules of organic chemistry, the preparation and manufacturing of small molecule drugs and the practice of medicine would be forever changed.

The pharmaceutical industry began during the Industrial Revolution, drawing on a series of innovations in chemistry from the coal tar-based dye industry, along with other technological developments. This same rhythm of explosive innovation occurred again 100 years later in post–World War II laboratories in the United States and Britain. In the epochal years of 1952 and 1953, the foundations of computing, molecular biology, neuroscience, AI, and modern medicine arose almost at once, appearing in juxtaposition against the afterglow of the first thermonuclear bomb detonated in the Pacific. Science was literally blazing on all fronts.

Medicine has benefited enormously from the scientific discoveries and technologies born in the atomic age. Biotechnology has its roots in the principles and successes of molecular biology. The historic beginning was the discovery of the double helical structure of DNA in 1953, followed a generation later by the development of recombinant DNA technology in the 1970s. Therapeutics originating from biotechnology innovations now account for 7 of the top 10 drugs sold in the world.

Cancer chemotherapy treatments entered into clinical practice in the early 1950s, landmarked by the FDA's approval of methotrexate in 1953. These therapies provided a rational basis for attacking cancer cells selectively and sparked a decades-long search for new chemotherapeutics. As importantly, clinicians became critical in the evaluation of these and other new drugs in clinical trials, taking a seat at the table alongside medicinal chemists and pharmacologists as decision-makers in industry.

In neuroscience, Alan Hodgkin and Andrew Huxley's unifying theory of how neurons fire action potentials was published in 1952. The Hodgkin-Huxley model stands as one of biology's most successful quantitative models, elegantly tying together experimental and theoretical work. The framework led to the search for the ion channels, receptors, and transporters that control ionic conductance and synaptic activity, which together formed the basis of 50 years' worth of neuroscience drug discovery.

Modern computing and AI began with the work of its seminal figures starting in the 1930s and was anchored by successful operation of the first stored program, electronic digital computer—the MANIAC I—in 1952. Historian George Dyson framed the significance of this moment well in his 2012 book, *Turing's Cathedral: The Origins of the Digital Universe*, (Vintage, 2012), stating that "The stored-program computer conceived by Alan Turing and delivered by John von Neumann broke the distinction between numbers that mean things and numbers that do things. The universe would never be the same." AI pioneers who had hopes for machine intelligence based on neural networks would need another 60 years and a trillion-fold increase in computing performance to have their dreams realized.

The science and technologies sparking the biotech and digital revolutions developed in parallel over the past 50 years and within the past decade have acquired powerful capabilities with widespread applications. The convergence of these technologies into a new science will have a profound impact on the development of diagnostics and medicines and nonpharmaceutical interventions for chronic diseases and mental health. The recent advances in AI and biotechnology together will be capable of disrupting the long-standing pharmaceutical industry model via superiority in prediction, precision, theory testing, and efficiency across critical phases of drug development. Not too far into the future, with any luck, the *in silico* dreams of scientists and its impact on medicine will be realized.

What Does This Book Cover?

The book ties together historical background with the latest cutting-edge research from the fields of biotechnology and AI, focusing on important innovations affecting medicine. Several chapters also contain highlights of the crop of new businesses engaged in the latest gene and cell therapy, along with those founded on AI-based therapeutic discovery and engineering. An in-depth look at the history of medicines sets the stage for understanding the pharmaceutical industry today and the evolution of therapeutic discovery for tomorrow.

Chapter 1, "The Information Revolution's Impact on Biology," begins with an overview of milestones in technology innovation that are central to modern biology and biomedical applications. The first section covers the success of genomics in tackling the deluge of genome sequencing information during the COVID-19 pandemic and biotech's utilization of the data for creating a vaccine against SARS-CoV-2. The next section details the recent paradigm shift in biology, describing how the field is moving toward a more quantitative discipline. Another major thrust of the chapter is the role of computational biology in human genome sequencing, and its potential for medicine in the 21st century.

Chapter 2, "A New Era of Artificial Intelligence," covers the history of AI's development and the major milestones leading up to the stunning advances in deep learning. The role of neuroscience in formulating some of the ideas around artificial neural networks and the neurobiological basis of vision are discussed. An introduction to various approaches in machine learning is presented along with current deep learning breakthroughs. A first look at AI applications in medicine is also given. The chapter ends with a brief look at current limitations of AI.

Chapter 3, "The Long Road to New Medicines," travels all the way back to the Stone Age to reveal humanity's first random experimentations to find nature's medicines. The first section outlines the progression of therapeutic discovery through four eras: botanicals, chemical therapeutics, biotherapeutics, and therapeutic engineering. The next section delves into the industrial manufacturing of medicines and the rise of the modern pharmaceutical industry. The chapter describes the birth of chemotherapeutic drugs and antibiotics and the impact of war on their development. A segment is devoted to the development of cancer therapeutics, including immunotherapy. The latter sections cover the pharmaceutical business model of the 21st century and the role of biotechnology in drug discovery innovation.

Chapter 4, "Gene Editing and the New Tools of Biotechnology," begins by introducing the timeline and brief history of the development of precision genome engineering tools. A significant portion of the chapter covers molecular biology and biological information flow, with a history of recombinant DNA technology. The second-generation biotechnology tools from the bacterial CRISPR-Cas systems are outlined and presented as important genome editing strategies. A companion section reviews clinical trials of CRISPR-Cas engineered therapies. A final section describes the mRNA vaccine platforms and innovations leading up to its success against the SARS-CoV-2 virus.

Chapter 5, "Healthcare and the Entrance of the Technology Titans," provides a look at how each of the technology giants—Amazon, Apple, Google, and Microsoft—are making moves to enter the healthcare sector. The first section describes digital health and investment activity in this newly emerging area, along with the drivers of healthcare technology innovation. A series of vignettes presents the ability of each tech giant to disrupt and play a role as new participants in healthcare, with a look at their competitive advantages in the healthcare landscape.

Chapter 6, "AI-Based Algorithms in Biology and Medicine," explores how AI technology is already impacting biomedical research and medicine today and potential routes for the future. Two sections provide in-depth coverage of deep learning algorithms for cancer and brain diseases. The final sections review regulatory approval of AI-based software as a medical device and the challenges faced in implementation of clinical AI.

Chapter 7, "AI in Drug Discovery and Development," dives into the use of AI and machine learning in drug discovery. A brief survey of *in silico* methods in drug discovery and development is presented, followed by a section on computational drug design with AI tools. A subsequent section introduces biotechnology companies that are creating a new base of innovation for the industry. A final section summarizes where AI is deployed currently across pharmaceutical discovery and development.

Chapter 8, "Biotechnology, AI, and Medicine's Future," begins with a discussion of convergence and how a new discovery engine based on hypothesis generation and evaluation by AI might work across biology, pharma, and medicine. The next section looks at how experimental approaches and computational methods together power biology by forming a new tech stack. AI's potential for neuroscience and the value of brain studies for AI and medicine are presented around the theme of motor control behavior and the brain. The chapter ends with a look at the landscape of companies arrayed against the range of technologies being developed to engineer therapeutics.

Reader Support for This Book

If you believe you've found a mistake in this book, please bring it to our attention. At John Wiley & Sons, we understand how important it is to provide our customers with accurate content, but even with our best efforts an error may occur.

To submit your possible errata, please email it to our Customer Service Team at `wileysupport@wiley.com` with the subject line "Possible Book Errata Submission."

CHAPTER

1

The Information Revolution's Impact on Biology

I think the biggest innovations of the twenty-first century will be at the intersection of biology and technology. A new era is beginning, just like the digital one. . ."

Steve Jobs in an interview with Walter Isaacson

The transformative power of the information revolution has reverberated across all industry sectors and has profoundly altered economies, political landscapes, and societies worldwide. Among the scientific disciplines, physics, astronomy, and atmospheric sciences were the first to benefit directly from the development of mainframe computing and supercomputers from the 1960s onward. The most dramatic advances came after the development of semiconductor electronics, the personal computer, and the Internet, which further accelerated the information revolution. These historic innovations were the catalysts for producing amazing new technological capabilities for biological sciences and for the biotechnology and pharmaceutical industries.

Scientific progress in biology is highly dependent on the introduction of new technologies and instrumentation with ever-increasing resolution, precision, and data gathering capacities and output. Table 1.1 breaks down the milestones in technology innovation by decade.

For much of the twentieth century, biology borrowed equipment from physics to visualize cellular and macromolecular structures and measure atomic-scale dimensions. In the pre-digital era, the observations and experimental data of scientists were captured on paper and physical media such as magnetic drums, tapes, X-ray films, and photographs.

Table 1.1: Milestones in Technology Innovation by Decade

DECADE(S)	TECHNOLOGY	FIELD	APPLICATIONS	REFERENCE
1900s–1930s	Electron microscopy	Physics	Structural biology	Chapter 1
	X-ray crystallography	Physics	Structural biology	Chapter 1
	Electrocardiogram	Physics	Cardiology	Chapter 6
1940s	Broad spectrum antibiotics	Microbiology	Medicine	Chapter 3
	NMR	Physics	Biology, chemistry, drug discovery	Chapter 4
1950s	Confocal microscopy	Physics	Biology	Chapter 7
	Artificial intelligence	Computer science	Information technology	Chapter 2
1960s	Laser	Physics	Information technology and instrumentation	Chapter 1
	Ultrasound	Physics	Medicine	Chapter 7
	Semiconductor electronics	Tech industry	Information technology and instrumentation	Chapter 2
	Integrated circuits	Tech industry	Information technology and instrumentation	Chapter 2
1970s	MRI, PET, CT	Physics and chemistry	Medical imaging and biology	Chapter 6
	Recombinant DNA	Molecular biology	Biology, drug discovery, therapeutics	Chapter 4
	Single-channel recording	Biophysics	Biology and drug discovery	Chapter 3
	Monoclonal antibodies	Immunology	Drug discovery and diagnostics	Chapter 3
	RAM memory	Tech industry	Information technology and instrumentation	Chapter 1

DECADE(S)	TECHNOLOGY	FIELD	APPLICATIONS	REFERENCE
1980s	PCR	Molecular biology	Biology, drug discovery, diagnostics	Chapter 1
	DNA sequencing	Molecular biology	Biology, drug discovery, diagnostics	Chapter 1
	DNA synthesis	Chemistry	Biology and drug discovery	Chapter 4
	MALDI-TOF/ESI MS	Biophysics	Biology and drug discovery	Chapter 3
	Personal computing	Tech industry	Information technology	Chapter 1
1990s	fMRI	Physics	Medical imaging	Chapter 5
	Two-photon microscopy	Physics	Biology	Chapter 4
	Transgenic technology	Molecular biology	Biology and drug discovery	Chapter 3
	RNAi	Molecular biology	Biology, drug discovery, therapeutics	Chapter 1
	Internet	Tech industry	Information technology	Chapter 1
2000s	CAR-T	Immunology	Therapeutics	Chapter 4
	Stem cell reprogramming	Immunology/ hematology	Biology, drug discovery, therapeutics	Chapter 4
	Next-generation sequencing	Molecular biology	Biology, drug discovery, diagnostics	Chapter 1
	Optogenetics	Biophysics	Biology	Chapter 3
	Cloud computing	Tech industry	Information technology	Chapter 2

Continues

Table 1.1 **(*continued*)**

DECADE(S)	TECHNOLOGY	FIELD	APPLICATIONS	REFERENCE
2010s	Cryo-EM	Biophysics	Drug discovery	Chapter 1
	CRISPR gene editing	Molecular biology	Gene therapy, drug discovery, diagnostics	Chapter 4
	Single-cell sequencing	Molecular biology	Biology and drug discovery	Chapter 8
	Deep Learning	Computer science	Biology, drug discovery, diagnostics	Chapter 2
	Quantum computing	Tech industry	Information technology	Chapter 8

The advent of microprocessor-based computing brought about the realization of analog-to-digital data conversion and, along with that, random access memory (RAM) on semiconductor circuits. Digitization of data streams and availability of petabyte-scale data storage have been immensely important for conducting modern science, not only allowing researchers to keep pace with the deluge of information, but also enabling network science and the widespread sharing of research data, a fundamental feature of scientific progress.

The information revolution's impact on biology continues unabated into the twenty-first century, providing computing power that continues to grow exponentially, producing sophisticated software for data acquisition, analysis, and visualization, and delivering data communication at speed and scale. New disciplines have been launched during this era in large part due to the introduction of technologies and instruments combining computation with high resolution. Some of the most important are DNA synthesizers and sequencers, which led to genomics and computational biology, fMRI for computational neuroscience, cryo electron microscopy (cryo-EM), NMR, and super-resolution microscopy for structural biology and several compute-intensive spectroscopic techniques (for instance, MS/ MALDI-TOF, surface plasmon resonance, and high-performance computing) that opened up computational drug discovery. Medicine has similarly advanced in the twentieth century by the application of computational approaches and breakthroughs in physics that were combined to create an array of imaging technologies.

As a consequence of transporting biology from a "data poor" to a "data rich" science, the information revolution has delivered its most fundamental and unexpected impact: causing a paradigm shift that has turned biology into a quantitative science. Biological science and biomedical research are now benefiting from the tools of data science, mathematics, and engineering, which had been introduced in "big science" endeavors, that is, projects such as the Human Genome, Proteome, and Microbiome Projects[1–3]; the Brain Initiative and Human Brain Project[4,5]; international consortia such as the Cancer Genome Atlas and International Cancer Research Consortium[6–7]; and precision medicine and government-backed population health projects like All of Us in the United States, the UK Biobank, and GenomeAsia 100K.[8–10] These projects propelled forward the development of the "omics" technologies, most importantly genomics, epigenomics, proteomics, and metabolomics, with new quantitative approaches imagined and inspired by the information revolution to manage and analyze "big data."

Chapter 1 of this book will explore the information revolution's impact on biology. Computing's massive influence on industry has also been referred to as the third industrial revolution. Subsequent chapters of the book will deal with aspects of the fourth industrial revolution,[11] described as the confluence of highly connected sensors comprising the Internet of Things (IoT), machine learning and artificial intelligence, biotechnology, and digital manufacturing that is creating the future. Over the next few decades, the technologies powering the fourth industrial revolution will bring about *in silico biology*. Similar to the economic transformations occurring in banking, manufacturing, retail, and the automotive industry, the pharmaceutical industry is poised to see enormous returns to scale by embracing the coming wave of innovations.

A Biological Data Avalanche at Warp Speed

The breathtaking speed with which the worldwide medical and scientific communities were able to tackle the coronavirus pandemic was a direct consequence of the information revolution. The Internet and wireless communication infrastructures enabled immense amounts of data from viral genome sequencing efforts and epidemiological data to be shared in real time around the world. Digital technologies were essential for gathering, integrating, and distributing public health information on a daily basis. In the private sector, computationally intensive drug discovery pipelines used artificial intelligence algorithms and biotechnology innovations to accelerate compound screening, preclinical testing, and clinical development. Nearly every government-backed initiative, industry partnership, and global collaborative research effort was powered by cloud-based computing resources. The mountain of data provided insights into the nature of the disease and inspired hope that treatments and effective countermeasures would arrive soon.

The biomedical research and drug development communities went into emergency action almost immediately, sensing that time was of the essence, but also spotting opportunities for business success and scientific achievement. Researchers in thousands of laboratories were formulating new therapeutic hypotheses based on incoming data from viral genome sequence, virus-host interactions, and healthcare systems.

During the initial phase of the pandemic, an unprecedented trove of knowledge became available via a torrent of publications. The rapid publishing of more than 50,000 documents announcing early research results on medRxiv and bioRxiv servers provided an invaluable platform for reviewing new studies on anything from clinical and biological investigations of viral replication to complex, multinational clinical trials testing an array of potential therapeutics and vaccines. Unfortunately, there was also an urgency to vet ideas, drugs, and important public health policy measures in real time. Many of these failed, were premature, or became conspiracy theory fodder. Technology can only do so much, and COVID-19 has shown us that science, too, has its limits. In addition, choices in the political sphere have enormous consequences on outcomes of scientific endeavors, public health, and the future of medicine.

The tragic irony of the SARS-CoV-2 outbreak is that China had an impressive defense to protect against a second SARS-type epidemic event, built around information technology. The Contagious Disease National Direct Reporting System[12] was engineered to facilitate reporting of local hospital cases of severe flu-like illnesses and deadly pathogens such as bubonic plague and to deliver notification to authorities within hours of their occurrence anywhere in China. Health officers in charge of nationwide disease surveillance in Beijing could then launch investigations, deploy experts from the China Centers for Disease Control, and set strategy for regional and local municipalities to deal with escalating public health situations. The design had a weak link that proved disastrous—hospital doctors and administration officials, fearing reprisals and hoping to contain the damage, decided not to use the alarm system.