The Processes of Life

An Introduction to Molecular Biology

Lawrence E. Hunter

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Syntax and Times Roman by SNP Best-set Typesetter Ltd., Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Hunter, Lawrence.
The processes of life : an introduction to molecular biology / Lawrence Hunter.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-262-01305-5 (hard cover : alk. paper) 1. Molecular biology. 2. Computational biology. 3. Evolution. I. Title.

[DNLM: 1. Molecular Biology. 2. Computational Biology. 3. Evolution.
QH 506 H945b′ 2009]
QH506.H86 2009
572.8–dc22

2008030905

10 9 8 7 6 5 4 3 2 1

Preface

I have to begin with an admission: The only biology course I ever took was in 10th grade, and I was too busy learning the facts of life in the hallways to have paid much attention in that classroom. So I feel obliged to tell the story of how I reached the point where I could write this book, and to thank all the people who helped me along the way.

In college, I was fascinated with the idea that computers could be a useful tool in understanding minds, and struck up what is now a 25-year-old friendship with *l'enfant terrible* of artificial intelligence, Roger Schank. A year after I had graduated, Roger called me back from an idyllic life in rural Hawaii because he had a project that he insisted I come to graduate school to work on. Roger can be very persuasive, and so I returned to lovely New Haven, CT, to meet Ray Yesner, MD, a great lung tumor pathologist, and Jerome Silbert, MD, a colleague of his with a passion for computers. While Roger knew that there was a good PhD thesis in modeling how Ray could organize and remember all that information, he couldn't have suspected that this would lead me to a lifelong love of biology and medicine.

As with some other loves, it didn't start off that well. Important as lung tumor pathology is, I found my thesis work depressing. In those days, despite dozens of distinctions pathologists could make by looking at tumors under a microscope, there were only two therapies to offer patients, and neither of them worked very well. Lung cancer is a horrible disease that kills just about everyone diagnosed with it in short order. It wasn't easy to look at those images, let alone to learn more about what was going on beneath the surface.

Fortunately for me, in addition to introducing me to Drs. Yesner and Silbert, Roger also introduced me to the U.S. National Library of Medicine (NLM), which funded my thesis work. I've always loved libraries, and with freshly minted PhD in hand, NLM was happy to have me as a staff scientist. DNA sequencing technology was advancing rapidly then, and the race to sequence

the Human Genome was getting started. For a computer scientist, gene sequences and protein structures and the biological questions about them were much easier to understand (and easier to write useful programs for) than patients and tumors. I worked hard to bring computer scientists and molecular biologists together and help launch contemporary computational biology. In my years at NLM, I met heroic molecular biologists like John Wootton and Harold Morowitz, who seemed to know at least a little something about every gene I ever encountered. Not only were they tremendously helpful colleagues, but their enthusiasm and even some of their knowledge was contagious. I was hooked.

The NLM is on the Bethesda campus of the National Institutes of Health (NIH). The breadth and quality of biomedical research there is probably unsurpassed anywhere in the world. My colleagues at NLM were a pleasure to work with and to learn from. NIH also attracted visits from many great scientists who gave fascinating lectures and showed a surprising willingness to talk with a curious computer scientist. I absorbed all I could. Although it was a struggle at first, once I became comfortable with the language of biology, the rest started to fall into place. As I asked questions of my colleagues, I started to hear the phrase that echoes in my ears to this day, "well, you're about right, but it's a little more complicated than that. . . ." I had a wonderful decade, surrounded by brilliant people and hearing firsthand about many important breakthroughs in biomedicine.

One of the reasons I ultimately left NIH for Colorado was to start a training program in computational biology. When I arrived in 2000, the fallout from the dot-com bust meant there were a lot of well-trained computer scientists looking for something new to do. However, they often struggled in my early computational biology classes, since many didn't know the first thing about biology. At that point, Harvey Greenberg, a friend and mathematician with an interest in things biological, encouraged me to create a course that would introduce molecular biology to the computer scientist. Since he was putting so much effort into creating the training program I wanted to see, I couldn't say no.

Creating that course was an incredibly taxing—and incredibly rewarding—experience. I learned the truth of the aphorism about how you don't really know a topic until you've taught it. All that first semester, I worked late nights and long weekends, often finishing my preparations just moments before class. I had to go back to solidify my understanding of the basics, which I had largely intuited by listening to research presentations, and realized that there was still a great deal of interesting biology I didn't know. Answering students' questions in class was an exercise in humility; my most common

answer was "let me look that up and get back to you next class." My teaching assistants that year, Christiaan van Woudenberg and Raphael Bar-Or, gave me far more of an assist than they could have expected. The efforts and enthusiasm of those students kept me going. And my wife and young children made the first of many sacrifices that would ultimately be required of them for me to complete this book.

As I was putting the course together, it became clear that there was no textbook appropriate for it. The course was aimed at mature students who wanted to get up to speed on the whole of molecular biology in a hurry. The typical undergraduate route required years of prerequisites in order to ultimately digest one of the detailed 1,000+ page classic molecular biology textbooks, like *Molecular Biology of the Cell*. That wasn't going to work for these folks. The more gentle introductions, like Hoagland and Dodson's *The Way Life Works*, and Clark and Russell's *Molecular Biology Made Simple and Fun*, were really good, but didn't cover nearly the range of topics I thought the class needed to grasp.

After teaching the course for a couple of years, I came to the conclusion that I really needed to write that textbook. More and more people were interested in adding molecular biology to their professional lives. NIH was encouraging not only computer scientists, but physicists, mathematicians, engineers, ethicists, and others to join in the exciting world of postgenomic biomedicine. Although there was (and remains) a lot of interest, learning enough biology to even get started has been a significant barrier to entry. My goal for this book is to knock down that barrier and make molecular biology accessible to anyone who is seriously curious.

Of course, deciding to write a book is not the same thing as actually writing it, and a lot more people helped me get from the decision to this text. Debbie Kornblith painstakingly transcribed videotapes of my classroom presentations. Boris Tabakoff, my department chair, helped me arrange the sabbatical that it took for me to actually devote enough time to the project to get it done. My entire lab, and particularly my text mining group leader Kevin Cohen, suffered through my absence, and I appreciate all of their efforts; I hope to make it up to you all one day. Helen and Francis Weir and their kids, Gabriel and Dominic, hosted us on their lovely farm in Catalonia and took care of us like long-lost friends. Mary Queally, who owns the cottage in Fanore, Ireland, where I finished the book, was also a wonderful host.

Bob Prior, my editor at the MIT Press, was the right mixture of encouraging and forgiving, and ultimately was very generous with both time and resources. Even with the sabbatical I was two years late on delivering the manuscript, which Bob tells me is not a record. One of the things I am most grateful for

is the MIT Press's agreement to allow me to freely distribute the text of the book (without typesetting or figures), as part of a scientific project in text mining. Foresightful and generous publishers are rare, and I am very grateful to be involved with one of the best.

Once the text started flowing, I depended on the advice of many readers to help me improve the book and catch mistakes. My first victims were good friends I have had since those high school days, Alexis Pearce and Thair Peterson. Since neither have any science background (Alexis is a rabbi and Thair is a journalist), slogging through my early drafts was clearly an act of love. Their extensive comments really helped make the book much more understandable. Members of my lab who made substantial comments include Mike Bada, Greg Caporaso, Kevin Cohen, Cheryl Hornbaker, and Philip Ogren. Several faculty colleagues at the University of Colorado Denver School of Medicine, including Marilyn Coors, Michael Holers, Jennifer Richer, and Mark Yarborough, read sections and provided valuable comments. Of course, all remaining errors and infelicities are my responsibility.

Two people made even more substantial contributions to this work. Tzu Phang created wonderful illustrations, working closely with me to capture the spirit of the text in visual form. Mike Bada went through the Glossary using his comprehensive knowledge of biomedical ontologies to link concepts in the book to these increasingly valuable community resources.

Ultimately, the most important help I had in producing this book came from my wife, Jill, and my children, Max and Hailey. They were always supportive, even when it meant I wasn't spending time with them. It can be hard for a five-year-old living in a strange country to keep hearing that daddy has to work on his book, but Hailey was both patient when I was working, and always ready to play when I was done. Max, who has just discovered his own love of writing, was equally supportive, loving, and fun. Jill in particular had to cope with her new business taking off just as we were to leave on the sabbatical. Her offer to take the kids home a month early so I would have the time to finish was above and beyond the call of family duty—and without it, the book might never have been done. Thank you all, from the bottom of my heart.