

Preface

Over the 10-year period between 1995 and 2005, DNA sequencing costs decreased fifty-fold, primarily as a result of increasing throughput due to incremental advances in tools, technologies and process improvements. As impressive as that accomplishment is the fact that the current outlay for sequencing the three billion base pairs of the DNA found in human or mammal genomes is \$10 million dollars (sources: The National Human Genome Research Institute (NHGRI), The National Institute of Health, and genomics.energy.gov). When new high-throughput techniques and the computational machinery needed to analyze data generated are created and implemented, further cost reductions are expected.

Increasing the capacity to capture and analyze proteomic data is paramount on the research scene. To the casual observer, it would appear that proteomic researchers could simply adopt automation technology developed for genomics research. However, technology transfer between genomics and proteomics would need to be modified to account for post-translational modifications (PTM), which are specific to proteomics. Also, proteomic datasets are expanding exponentially while genomic datasets are saturated. Managing a much larger, and continually growing database presents its own set of organizational opportunities.

The inspiration for *'Automation in Genomics and Proteomics: An Engineering Case-Based Approach'*, was two Massachusetts Institute of Technology (MIT) and Harvard interdisciplinary special studies courses, *Bioinformatics and Proteomics: An Engineering-Based Problem Solving Approach* and numbered 6.092/HST.480, respectively, and the article 'Automation, parallelism, and robotics for proteomics' created by Alterovitz, Liu, Chow and Ramoni. *Bioinformatics and Proteomics: An Engineering-Based Problem Solving Approach* focused on bioinformatics and proteomics with engineering-based approaches. While the 'Automation, parallelism and robotics for proteomics' [1] article discussed various technologies and methods being applied within the proteomics field that facilitate automation to achieve cost- and time-saving benefits and link proteomics-based information with germane research areas.

The book, '*Automation in Genomics and Proteomics: An Engineering Case-Based Approach*', is the product of an international effort that spanned six countries and involved 18 public, private and academic institutions. The textbook addresses automation technology currently in the areas of analysis, design and integration and describes current technological limitations. The underlying biology concepts are also delineated. By disseminating knowledge of leading experts in the field, the textbook underscores that a collaborative effort across many disciplines is required to advance automation. This book's distinctive engineering-oriented coverage makes the material more intuitive for a technical audience, and is intended for an upper-level undergraduate elective or as a graduate-level foundation class. It can also be used as an industry reference tool. This book is an excellent stand-alone text for an introductory/motivational seminar or course on the subject, or it can serve as a complementary text to traditional texts.

'*Automation in Genomics and Proteomics: An Engineering Case-Based Approach*', is divided into four sections. In Section 1, the fundamental biology is introduced from an engineering perspective. The first chapter, *The Central Dogma: from DNA to RNA, and to Protein*, presents the needed molecular and cellular biology background and can be treated within a review session, if an introductory biology course is a prerequisite. Molecules and bioprocesses that are related to protein biosynthesis are the focus of this chapter, where DNA is the source of genetic information and amino acids are the raw materials. In the second chapter, *Genomes to Proteomes*, the book moves from the genomics to proteomics by discussing the automatic annotation and the generation of high-quality gene models, the set-up and execution of quantitative and statistically rigorous global proteomic experiments, and proteomics in a biological context.

Sections 2 and 3 focus on design and analysis via automation:

- Chapter 3, *High-Throughput DNA Sequencing*, reviews the current state of traditional dideoxy sequencing workflows and the development of next-generation technologies that are poised to revolutionize genomics.
- Chapter 4, *Modeling a Regulatory Network using Temporal Gene Expression Data: Why and How?*, introduces methodologies to classify genes according to their expression measurements across a set of conditions, and presents sophisticated procedures to infer regulatory networks from gene expression data.
- Chapter 5, *Automated Prediction of Protein Attributes and Its Impact on Biomedicine and Drug Discovery*, systematically explores recent progress in computational methods and, for those prediction methods with web-servers currently available, a step-by-step instruction is presented.
- Chapter 6, *Molecular Interaction Networks: Topological and Functional Characterizations*, introduces both basic concepts and current research trends in network biology.
- Chapter 7, *DNA Synthesis*, evaluates the standard methods, chemistry and instrumentation of DNA and gene synthesis and the applications, strategies and the trends of applying high-throughput gene synthesis in synthetic biology to design and engineer biological systems.
- Chapter 8, *Computational and Experimental RNA Nanoparticle Design*, discusses some of the recent advances in the ability to computationally design RNA nanoparticles, drawing upon a wide array of known structural motifs as well as the issues that are involved in experimentally constructing and testing their validity.

- Chapter 9, *New Paradigms in Droplet-Based Microfluidics and DNA Amplification*, addresses the confluence of new genetic and microfluidic technologies that will be used to automate *in vitro* evolution, genomics and molecular and cellular screening on a scale that was previously impractical.
- Chapter 10, *Synthetic Networks*, demonstrates several interesting synthetic networks and provides valuable engineering tools to study motifs, modularity and robustness of cellular networks. Also in this chapter, the current understanding of cellular networks, synthetic network construction and remaining challenges towards automating biochemical processes using synthetic circuitry are reviewed.

Chapters 11 and 12 are included in Integration, the fourth and final section of *Automation in Genomics and Proteomics: An Engineering Case-Based Approach*. In this section, Chapter 11, *Molecular Modeling of CYP Proteins and Its Implication for Personal Drug Design*, addresses the existing computational methods for modeling 3-D protein structures with Shanghai Molecular Modeling (SAMM), the molecular modeling software developed at Shanghai Jiaotong University, while Chapter 12, *Recent Progress of Bioinformatics in Membrane Protein Structural Studies*, examines the recent progress of computational work and automation process in membrane protein structural studies.

In Chapter 13, *Trends in Automation for Genomics and Proteomics*, the book concludes with a field summary and an exploration of future avenues of research. For those interested in additional resources, source code and related materials: the book's internet site can be accessed at: <http://bcl.med.harvard.edu/proj/automation>.

As an international effort, there are many people whose contributions were critical to the publication of this work. The editors would like to thank the contributing authors to the text, including: Masahiko Sisido and Takashi Ohtsuki (Chapter 1); Scott E. Baker, Ellen A. Panisko, Igor Grigoriev, Don S. Daly and Bobbie-Jo Webb-Robertson (Chapter 2); Tarjei S. Mikkelsen (Chapter 3); Sophie Lebre and Gaelle Lelandais (Chapter 4); Kou-Chen Chou (Chapters 5, 11 and 12); Jake Chen and Xiaogang Wu (Chapter 6); Jingdong Tian (Chapter 7); Isil Severcan, Cody Geary, Luc Jaeger, Eckart Bindewald, Wojciech Kasprzak and Bruce A. Shapiro (Chapter 8); Jonathan Rothberg, Michael Samuels, John Leamon, Ronald Godiska, Thomas Schoenfeld and David Mead (Chapter 9); Jongmin Kim (Chapter 10); Cheng-Cheng Zhang, Jing-Yi Yan, Jing-Fang Wang and Dong-Qing Wei (Chapter 11); Hong-Bin Shen, Jun-Feng Wang, Li-Xiu Yao and Jie Yang (Chapter 12); and Dmitriy Sonkin (Chapter 13). The editors would especially like to thank the Wiley commissioning editor, Paul Deards, who invited us to write this book.

Thank you to William H. Down and Jonathan Dreyfuss for reviewing and editing the manuscript. A special thanks also to the anonymous book proposal and book draft reviewers.

Gil Alterovitz, PhD
Roseann Benson
Marco F. Ramoni, PhD

Reference

1. G. Alterovitz, J. Liu, J. Chow, and M.F. Ramoni. (2006) Automation, parallelism, and robotics for proteomics. *Proteomics*, 6(14), 4016–22.