

PREFACE

Proteomics is the study of all the proteins of a cell or an organism. It is the newly developed science for the study of proteins. It attempts to define the proteome, which is the entire protein content of an organism encoded by its genome; hence, the word is derived from protein and genome. Proteomics aims at describing the structure and function of the proteins of a cell at a large scale. This enables us to understand the structure and function of a cell and finally that of an organism. The science of proteomics has obvious applications to medicine through identification of proteins as marker(s) of a disease (i.e., diagnostics) or as targets of new drugs or as therapeutics (i.e., drugs) as well. Proteomics provides new tools for the understanding of proteins, which are the workhorse molecules of a cell that control all its biophysical and biochemical attributes. The one gene–one enzyme concept of Beadle and Tatum (1941) provided a unique tool for the study of proteins; this approach is being used every day, even to this date. Proteomics based on high-throughput technologies added a new dimension to the approach initiated by Beadle and Tatum. This book, therefore, examines proteomics beyond the one gene–one enzyme concept.

My research interest in genetics and the biochemistry of proteins goes back to the mid-1960s, when I began my association with the late Nobel Laureate Professor Edward L. Tatum at the Rockefeller University as a postdoctoral fellow supported by the Jane Coffin Childs funds for Medical Research. Beadle and Tatum together formulated the one-gene–one enzyme concept in 1941. George Beadle, Edward L. Tatum, and Joshua Lederberg shared the 1958 Nobel Prize in Physiology and Medicine for their respective

contributions to the development of the one-gene–one-enzyme concept in *Neurospora* and recombination in bacteria; Lederberg later became president of Rockefeller University. This theory of Beadle and Tatum established the conceptual scheme for the control of the structure and function of a protein by a gene.

At Rockefeller University, the laboratories of William Stein and Stanford Moore and that of Robert Bruce Merrifield were situated close to Tatum's laboratory. In their laboratories, the first large protein was sequenced and chemically synthesized. I remember having several discussions with these scientists about the structure and function of proteins. William Stein, Stanford Moore, and Gerry Edelman, all of whom were from Rockefeller University, and Christian Anfinsen of the National Institutes of Health (NIH) became Nobel Laureates in 1972. Later, Bruce Merrifield in 1984 and Günter Blobel in 1999, also from Rockefeller University, received Nobel Prizes, all of them for their contributions to protein chemistry, including the structure, function, synthesis, and intracellular transport of proteins. The goal of Stein and Moore at that time was to sequence more than 1000 proteins by the end of the 20th century. This goal was realized much faster with the science of genomics and with the application of mass spectrometry and other high-throughput technologies.

At Rockefeller University, I also had the opportunity to know Professor Frank H. Field, director of the mass spectrometry laboratory. Earlier, Dr. Field, in collaboration with Joe Franklin, had developed the first ionization technique for mass spectrometry. Dr. Field was helping Professor Tatum with the identification of chemical(s) emitted into the gas phase by a slow-growing morphological mutant of *Neurospora*. An exposure of this gaseous emission to the wild-type strain made it grow slowly like the mutant. This chemical, however, remained elusive to identification by mass spectrometry.

Soon after my arrival at Rockefeller University, I remember having a discussion with the Professor Victor Najjar on the one-gene–one-enzyme theory. Dr. Najjar, then a Professor at the Vanderbilt University and an editor of *Methods in Enzymology*, was visiting Rockefeller University on a sabbatical leave. During a discussion of my work with him, he became somewhat concerned after learning about the possible role of two genes in the control of an enzyme, phosphoglucosyltransferase, involved in the morphogenesis of a fungus *Neurospora* as my work indicated at that time. I believe this was perhaps because of his unfamiliarity with the literature in genetics and particularly that of the role of suppressor genes in controlling the structure of a protein encoded by another gene. He, therefore, thought that my findings were in contradiction to the original idea of the one-gene–one-enzyme hypothesis. However, I convinced Dr. Najjar that such findings make a difference only in semantics and not in the conceptual scheme of

the original one-gene–one-enzyme theory. I pointed out to him that these exceptions only strengthen the original one-gene–one-enzyme concept, just as certain observations such as the partial dominance, co-dominance, and epistasis, which on the surface seem to be in conflict with Mendelian rules of inheritance, actually lend support to the original ideas implicit in the rules of inheritance by Mendel.

Later that day, I discussed with Professor Tatum the exchange on the one-gene–one-enzyme theory during my conversation with Dr. Najjar. During our conversation, Professor Tatum immediately pointed out that the one-gene–one-enzyme hypothesis has already been modified to a one-cistron (gene)–one-polypeptide hypothesis: However, I was aware of this concept and told professor Tatum that I had already pointed out this modification to Professor Najjar. Professor Tatum also expressed that he expected additional modification to this theory because of the looming complexity of our genetic material as was being revealed by the nucleic acid hybridization experiments. He expressed to me that it was indeed a matter of semantics and that so long we understood what we were talking about, we lived with the limits of the conceptual scheme of the one-gene–one-enzyme hypothesis. Almost a decade later, Phillip Sharp from the Massachusetts Institute of Technology (MIT) revealed the split nature of the gene and received the Nobel Prize in 1990 for his work. Furthermore, the study of the structure of the immunoglobulin gene(s), which brought the Nobel Prize to Tonegawa, also from MIT in 1987, presented an extreme view of an exception to the one-gene–one-enzyme hypothesis. However, these findings affirmed the expectations of Professor Tatum that the one-gene–one-enzyme theory would be modified in view of the complexity of our genetic material. Despite the changes to this theory, it is important to note that almost all genes in prokaryotes and more than 50% of genes in higher eukaryotes obey the dictum of one-gene–one-enzyme theory. This theory still provides the basis for creation of mutants and knockouts crucial for the study of a protein structure and function and its role in controlling the phenotype of the organism. This theory is also the basis for the gene therapy approach for the treatment of human diseases.

I remember the events and the manner in which the field of protein chemistry progressed and then was later ignored with the coming of the genome projects and the science of genomics; it was finally revived and blossomed into the science of proteomics. The coming of genomics and the subsequent development of proteomics have completely changed our view regarding the philosophy of science and how we understand biology. Before genomics, we had a reductionist view of science, and the biology of an organism was thought to be understood in terms of the molecules only. We also used to do one thing at a time when deciphering one molecule

after another. Now, we are trying to understand all things at the same time because of our ability for high-throughput analyses; we are no longer reductionists, rather we are holists trying to understand the biology in terms of the interactions of a large number of molecules at once. The science of proteomics has thus ushered in the coming of a new branch of science called *systems biology* to obtain the ultimate understanding of an organism within a particular environment. An understanding of the environment is important because it can bring about changes in the structure and function of genes and gene products.

I write this book on the science of proteomics with the goal of bringing out its conceptual development starting from one-gene–one-enzyme theory and leading to its instrumentation-based methodologies and applications in medicine and biotechnology and the fact that life is sustained by the interactions of proteins. I take special effort in describing the nature and operation of these complex instrumentations involved in proteomics in a language readily understandable to students with an exclusive background in biology. I also provide an emphasis on biological methods in elucidating certain aspects of proteomics, which has been ignored in earlier treatises on the subject of proteomics. This book is written in a manner comprehensible to emerging scientists, including undergraduate and graduate students as well as postdoctoral trainees.

The book is organized into seven chapters, and many references, although some included at the end of the chapters, are not cited in the text to allow for the smooth flow of main concepts and easy reading of the subject matter. I hope that my efforts are successful.

I believe no such text that particularly addresses the needs of the biologist exists at this time. In this book, an attempt is made to give a biologist's view of the subject to non-biologists equally well, particularly bringing to their attention how biologists approached certain problems—for example, protein–protein interactions in the absence of advanced technologies such as bioinformatics. I also believe that this text is a contribution to this emerging branch of science of proteomics and to systems biology, and of course to scientists in these branches of science, leading to the appreciation of the developments in proteomics beyond the one-gene–one-enzyme concept of Beadle and Tatum that provided the conceptual scheme and the tool for understanding proteins in the living system.

This book is being published on the occasion of the 52nd anniversary of the awarding of the Nobel Prize to Beadle and Tatum in 1958 to reflect the progress made in the understanding of proteins, which was started by the conceptualization of the one-gene–one-enzyme hypothesis that provided the tool for analysis of proteins.

I would like to thank many colleagues for their help with this work. I would like to thank Professors Steve Threlkeld and J.J. Miller, both of McMaster University, for my fueling initial interest in genetics and Professor Stuart Brody of the University of California, San Diego, (formerly at the Rockefeller University) for my introduction to enzymology. In addition, I am grateful to Professor Philip Hanawalt of Stanford University and Professor Stuart Linn of the University of California, Berkeley for their support of my continued interest in the genetical biochemistry of proteins. I would also like to thank Professor David Reisman at the University of South Carolina for reading the manuscript in its entirety and for his many helpful comments. I am also thankful to Professors Michael Felder and Sanjib Mishra both at the University of South Carolina, Professor Narsingh Deo of the University of Central Florida, Professor David Gangemi of Clemson University, Professor Alexandru Almasan of the Cleveland Clinic, Dr. Narendra Singh of the U.S.C. Medical School, Professor R.P. Jha of Patna University, Professor K.M. Marimuthu of the Post Graduate School at Madras University, Professor Ramesh Maheshwari of the Indian Institute of Science, Prashant Jha and Dr. Kanchan Kumari for their support of my endeavors and to Dr. Richard Vogt of the University of South Carolina for help with the cover picture.

This work would not have been possible without the encouragement and show of infinite patience from Dr. Darla Henderson of John Wiley and Sons, particularly during periods of multiple personal challenges. I also thank Anita Lekhwani, the Senior Acquisition Editor of John Wiley and Sons, for her immense interest in this work and for her enthusiastic support and assistance that eased the submission of this manuscript and made its publication possible. I am also thankful to Christine Moore, Rebekah Amos, Sheree Van Vreede, and Kellsee Chu of John Wiley & Sons for assistance with the manuscript that helped its timely publication. I am grateful to Dr. Kevin H. Lee of the University of Delaware for the two-dimensional gel picture, Darryl Leza of NHGRI, NIH, for the protein structure picture, and to John Alam, Clint Cook and Michelle J. Bridge of the Dept. of Biological Sciences at the University of South Carolina for the diagrams and for their assistance in preparation of the manuscript.

Finally, I thank my wife, Purnima, and our son, Prakash, for their continuous support and interest in this work. I dedicate this work to Purnima and Prakash and above all to the memory of the mentors in my life, my parents and Professor E.L. Tatum. I am solely responsible for any and all errors that may be found in this book.