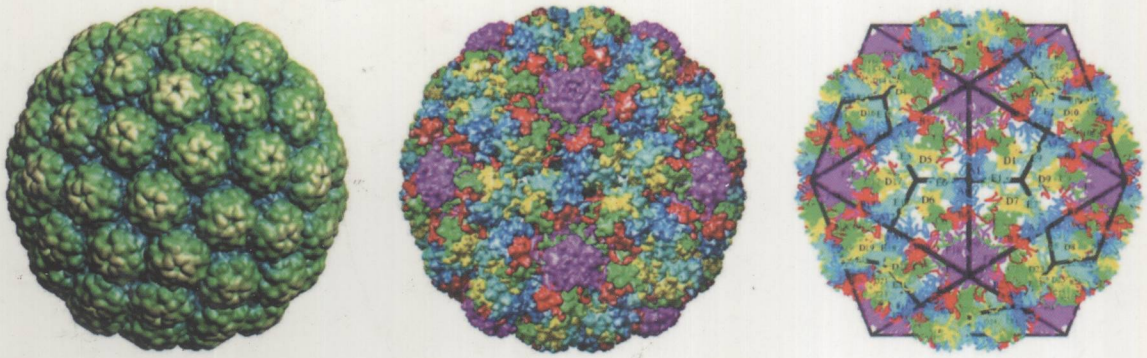
MEDICINAL PROTEIN ENGINEERING



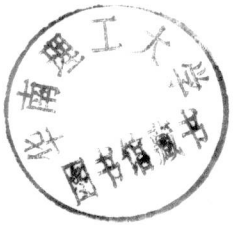
EDITED BY

YURY E. KHUDYAKOV

CRC Press
Taylor & Francis Group

MEDICINAL PROTEIN ENGINEERING

EDITED BY
YURY E. KHUDYAKOV



CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-0-8493-7368-8 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Medicinal protein engineering / editor, Yury E. Khudyakov.
p. ; cm.
"A CRC title."
Includes bibliographical references and index.
ISBN 978-0-8493-7368-8 (hardback : alk. paper)
1. Protein engineering. 2. Biomedical engineering. I. Khudyakov, Yury E. II. Title.
[DNLM: 1. Protein Engineering. 2. Biomedical Engineering. 3. Quantitative Structure-Activity Relationship. 4. Viral Vaccines. QU 450 M489 2009]

TP248.65.P76M43 2009
610.28--dc22 2008025423

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

MEDICINAL PROTEIN ENGINEERING

Preface

Knowing the small by way of the great, one goes from the shallow to the deep.

Miyamoto Musashi
The Book of Five Rings

The American Engineering Council for Professional Development defines engineering as "the creative application of scientific principles to design or develop structures, machines, apparatus, or manufacturing process..." With dry eloquence this definition clearly recognizes the roots of engineering in science and knowledge. Engineering is one of the most important tools developed by humans for using knowledge to control and adapt the environment to human needs. As a result, we live in the engineered world. Scientific exploration makes us aware of new opportunities to improve our welfare and engineering takes complete advantage of this knowledge to our benefit. Engineering started from inventing basic things for mere survival like making fire, and is now steadfastly penetrating all areas of our lives and life itself. Although engineering of biological objects is gaining a rather careful acceptance and sounds alien or controversial to many people (Chapter 19), it is actually as old as humankind and, since prehistoric times, it has been employed to breed animals and cultivate plants. The development of molecular biology led to the understanding of molecular mechanics of life and opened a new door for engineering of biological objects. Among many areas, this engineering promises unprecedented control over health. Although medicine is a very conservative discipline that cautiously and scrupulously evaluates all innovations before applying them to patients, it eagerly embraced and stimulated the development of genetic and protein engineering.

Proteins derived from pathogenic microorganisms and viruses have a broad array of applications in all three domains of medicine—diagnostics, prophylaxis, and therapy. Human proteins also may serve as diagnostic reagents and prophylactic agents for many complex diseases like cancer. There are three major sources of pathogen and human proteins for medicinal applications: (1) the human body, (2) tissue and cell cultures, and (3) recombinant DNA technology. The first two sources were exploited first. However, significant safety concerns and frequently prohibitive cost limited the use of these sources. On the other hand, the medicinal application of recombinant DNA technology to, for example, viral diseases provided an immediate and safe access to viral proteins that could be used for diagnostics and vaccines. One of the most widely known examples of such use of recombinant DNA is the hepatitis B virus surface antigen obtained from transformed yeast to create an efficient, safe, and affordable recombinant vaccine (Chapter 12). However, it was very rapidly realized that the straightforward approach employed for the development of a recombinant hepatitis B vaccine is not applicable to many other viruses such as the human immunodeficiency virus (Chapter 7) or the hepatitis C virus (Chapter 12). Alan R. Fersht and his colleagues found the solution to this problem around 25 years ago, presenting the concept of protein engineering (Chapter 9) as a new area for the application of molecular techniques and scientific principles to the design and construction of novel proteins with desired properties.

The remarkable developments over the last three decades in the automatic high-throughput chemical syntheses of long peptides, as well as in chemical and enzymatic syntheses of long DNA fragments, have built a solid technological foundation for protein engineering. However, the currently limited understanding of protein structure and of the relationship between structure and function has hindered the application of protein engineering to many complex medical problems.

Rational design of proteins is deep-rooted in the quantitative knowledge of their chemical, physical, and biological properties. In cases when sufficient knowledge of these properties was

secured, a significant progress in protein engineering was attained. Parts I and II of this book review some of the success stories in the development of vaccines, diagnostics, and therapeutics as well as problems associated with protein engineering in these fields.

Major strategies for engineering of proteins with predetermined biological properties can be classified into structural, functional, and focused approaches. The structural approach is a major focus of this book. This approach is as close to the real rational design of proteins as the modern state of science allows. However, because comprehensive knowledge of a protein's quantitative structure–activity relationship (QSAR) is often unavailable, this approach can be applied only with caution. Protein engineering requires the availability of simple building blocks with clearly established QSAR. Such blocks became available through research on the antigenic and immunogenic properties of proteins (Chapters 6 through 13). Building blocks for diagnostics and vaccines exist in the form of antigenic epitopes that can be modeled with short protein fragments and in the form of carrier molecules that can efficiently present antigenic epitopes to immunocompetent cells or antibodies.

There are two major approaches to using these building blocks in the engineering of proteins with predetermined immunological properties. One approach, based on construction of artificial multiple epitope proteins (MEP), is used mainly to develop reagents with improved diagnostically relevant properties as described in Chapter 13. Because antibody–antigen binding is a relatively simple protein–protein interaction process, the MEP approach was shown to be somewhat efficient for obtaining artificial diagnostic antigens. The structural design of vaccines is more complicated because the process of eliciting antibodies is significantly more complex. In this case, antigenic epitopes of interest are frequently inserted into carrier proteins that serve to stimulate an immunoresponse against the inserted epitopes. Chapters 6 and 8 through 11 describe this approach using virus-like particles to vaccine development.

The functional approach, also known as "directed evolution," requires no prior knowledge of a protein's QSAR. Rather, it is based on the availability of a representative library of peptides or proteins and a selection procedure for the peptide or protein with the desired activity. It is termed a "functional" approach because proteins are selected based on specific functions without regard to structure. This approach is only briefly reviewed in this book (Chapter 17). It is not protein engineering *per se*, because proteins generated using this approach are discovered rather than engineered. To a significant degree, the functional approach is used as a "magic wand" that, unfortunately, has a very limited efficiency. There are examples of reasonably successful applications of this strategy to antigens and antibodies described in Chapters 9, 15, and 17.

The focused approach applies mathematical modeling to gain new QSAR knowledge and uses this knowledge to design proteins with improved desired properties (Chapter 2). The mathematical QSAR model can be built using a limited set of protein variants displaying variations in their activity. Accurate quantitative models can be used to directly design a desired protein. Modestly accurate models can be used as a guide in a cyclic process of the virtual search of sequence space for proteins with improved properties, experimental construction and testing of novel proteins, and optimization of the model for a more accurate virtual exploration. Part I of this book reviews available Web computational resources (Chapter 1), machine learning (Chapters 2 and 4) and phylogenetic techniques (Chapter 3), as well as techniques for detection of coordinated substitutions (Chapter 5) that can be used to study protein properties and build mathematical models for engineering of novel vaccines, diagnostic reagents, and therapeutics.

Protein engineering is unavoidably connected to genetic engineering. Some protein applications require a careful matching between the designed proteins and genetic constructs. This intricate matching is thoroughly surveyed in Chapter 20. Many chapters in Parts II and III frequently touch on plant systems used for the expression of engineered proteins. The interest of protein engineers in plant expression is quite understandable. Besides many advantages meticulously illustrated in Chapters 18 and 19, plant expression systems allow for circumventing some technological problems related to protein purification, storage, and implementation that are not directly linked to the targeted

protein properties. As such, these systems open the door for an intriguing prospect of protein engineering "without proteins."

Medicinal protein engineering is an immense field for molecular and computational exploration. Attempts to describe this field in one book will inevitably be selective. Nevertheless, I hope that many current practitioners of protein engineering as well as many computational and experimental researchers with an aspiration for exceptionally thrilling and extremely challenging tasks will find this volume instructive and stimulating. Protein engineering for medicine is only going to be more exciting. In his book *The Romantic Generation* Charles Rosen wrote, "There are details of music which cannot be heard but only imagined." This is the wonderful way of art. However, it is not what we expect from science. This book was conceived with the hope that all the ideas the readers may "imagine" after going through this volume will be definitely "heard" at the lab bench, at the computer, or, better yet, at the patient's bed.

Since I opened this preface with a quotation from *The Book of Five Rings* by Miyamoto Musashi I would like to close by borrowing another timeless gem from this philosopher and peerless fighter that in my mind appropriately emphasizes the power of knowledge—"When one is in full combat gear, one does not think of small things."

Yury E. Khudyakov
Centers for Disease Control and Prevention
Atlanta, Georgia

Editor

Yury E. Khudyakov, PhD, is chief of the Molecular Epidemiology and Bioinformatics Laboratory, Laboratory Branch (LB), Division of Viral Hepatitis (DVH), Centers for Disease Control and Prevention (CDC), Atlanta, Georgia. He received his MS in genetics from the Novosibirsk State University, Novosibirsk, Russia and his PhD in molecular biology from the D.I. Ivanovsky Institute of Virology, Academy of Medical Sciences, Moscow, Russia. He started his research career in the laboratory of gene chemistry at the M.M. Schemyakin Institute of Bioorganic Chemistry, Academy of Sciences, Moscow, Russia. He was a research fellow in the laboratory of viral biochemistry and chief of the genetic engineering section in the laboratory of chemistry of viral nucleic acids and proteins at the D.I. Ivanovsky Institute of Virology, Moscow, Russia. In 1991, he joined the Hepatitis Branch (HB), Division of Viral and Rickettsial Diseases (DVRD)/CDC, Atlanta, Georgia National Academy of Sciences, United States as a National Research Council research associate. Since 1996, he has served as chief in the Developmental Diagnostic Unit, Molecular and Immunodiagnostic Section/HB/DVRD/NCID/CDC and later as chief of Computational Molecular Biology Activity and deputy chief of Developmental Diagnostic Laboratory, LB/DVH/CDC.

Dr. Khudyakov's main research interests are molecular epidemiology of viral diseases, development of new diagnostics and vaccines, molecular biology and evolution of viruses, and bioinformatics.

Dr. Khudyakov has published over 110 research papers and book chapters. He has also edited a book *Artificial DNA*, CRC Press (2002). He has authored several issued and pending patents. He is a member of the editorial board for the *Journal of Clinical Microbiology*.

Contributors

Michael Aidoo
Laboratory Branch
Division of HIV/AIDS Prevention
National Center for HIV, Hepatitis, STD, and Tuberculosis Prevention
Centers for Disease Control and Prevention
Atlanta, Georgia

Ekaterina Alekseeva
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Martin F. Bachmann
Cytos Biotechnology AG
Schlieren, Switzerland

Rick A. Bright
Novavax, Inc.
Rockville, Maryland

Salvatore Butera
Laboratory Branch
Division of HIV/AIDS Prevention
National Center for HIV, Hepatitis, STD, and Tuberculosis Prevention
Centers for Disease Control and Prevention
Atlanta, Georgia

Stephen A. Cammer
Virginia Bioinformatics Institute
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

David Campo
Molecular Epidemiology and Bioinformatics Laboratory
Division of Viral Hepatitis
Centers for Disease Control and Prevention
Atlanta, Georgia

Michael Czar
Virginia Bioinformatics Institute
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Zoya Emilova Dimitrova
Molecular Epidemiology and Bioinformatics Laboratory
Division of Viral Hepatitis
Centers for Disease Control and Prevention
Atlanta, Georgia

Dennis Ellenberger
Laboratory Branch
Division of HIV/AIDS Prevention
National Center for HIV, Hepatitis, STD, and Tuberculosis Prevention
Centers for Disease Control and Prevention
Atlanta, Georgia

Lilia Milkova Ganova-Raeva
Molecular Epidemiology and Bioinformatics Laboratory
Division of Viral Hepatitis
Centers for Disease Control and Prevention
Atlanta, Georgia

Alma Gedvilaite
Institute of Biotechnology
Vilnius, Lithuania

María José Gómara
Department of Peptide and Protein Chemistry
Institute for Chemical and Environmental Research (IIQAB-CSIC)
Barcelona, Spain

Elmars Grens
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Isabel Haro
Department of Peptide and Protein Chemistry
Institute for Chemical and Environmental Research (IIQAB-CSIC)
Barcelona, Spain

Reimar Johne
Federal Institute for Risk Assessment
Berlin, Germany

Andris Kazaks
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Tatyana Kozlovska
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

James Lara
Molecular Epidemiology and Bioinformatics Laboratory
Division of Viral Hepatitis
Centers for Disease Control and Prevention
Atlanta, Georgia

Alla Likhacheva
NPO Diagnostic Systems
Nizhny Novgorod, Russia

Miguel Angel Gómez Lim
Centro de Investigación y de Estudios Avanzados del I.P.N.
Departamento de Ingeniería Genética de Plantas
Irapuato, Mexico

Baiba Niedre-Otomere
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Anna Obriadina
NPO Diagnostic Systems
Nizhny Novgorod, Russia

Velta Ose
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Néstor O. Perez
Probiomed S.A. de C.V.
Mexico City, Mexico

Paul Pumpens
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Michael Anthony Purdy
Molecular Epidemiology and Bioinformatics Laboratory
Division of Viral Hepatitis
Centers for Disease Control and Prevention
Atlanta, Georgia

Peter M. Pushko
Novavax, Inc.
Rockville, Maryland

Till A. Röhn
Cytos Biotechnology AG
Schlieren, Switzerland

Lívia M.G. Rossi
NL Biotechnologies LLC
Little Rock, Arkansas

Kestutis Sasnauskas
Institute of Biotechnology
Vilnius, Lithuania

Gale E. Smith
Novavax, Inc.
Rockville, Maryland

James M. Smith
Laboratory Branch
Division of HIV/AIDS Prevention
National Center for HIV, Hepatitis, STD, and Tuberculosis Prevention
Centers for Disease Control and Prevention
Atlanta, Georgia

Terrence M. Tumpey
Influenza Division
National Center for Immunization
and Respiratory Diseases
Centers for Disease Control
and Prevention
Atlanta, Georgia

Tatyana Ulanova
NPO Diagnostic Systems
Nizhny Novgorod, Russia

Rainer G. Ulrich
Friedrich-Loeffler-Institut
Federal Research Institute for Animal Health
Institute for Novel and Emerging
Infectious Diseases
Greifswald, Germany

Tatyana Voronkova
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Jonny Yokosawa
NL Biotechnologies LLC
Little Rock, Arkansas

Anna Zajakina
Latvian Biomedical Research and Study Centre
University of Latvia
Riga, Latvia

Andris Zeltins
Latvian Biomedical Research and Study Centre
Riga, Latvia

Contents

Part I

Computational Approaches

1 Web Resources for Protein Analysis

Stephen A. Cammer and Michael Czar

CONTENTS