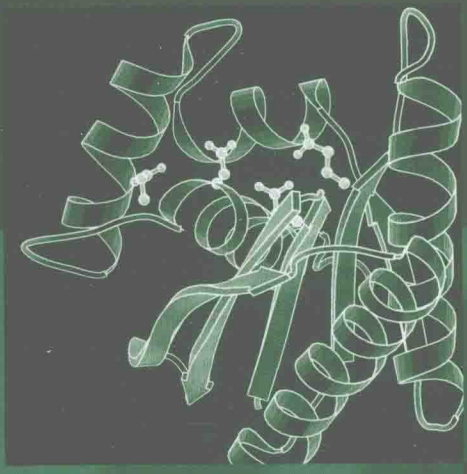
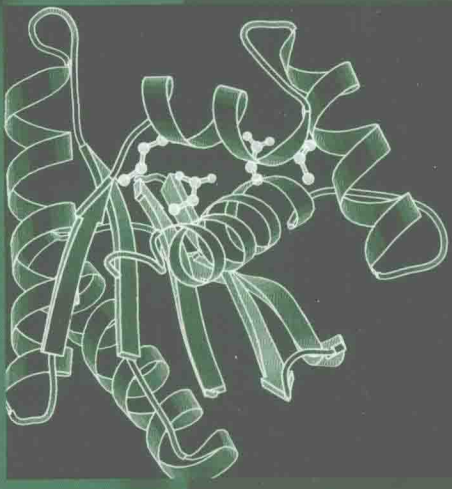
15 Nucleic Acids and Molecular Biology
Hans Joachim Gross (Ed.)

Practical Bioinformatics





Janusz M. Bujnicki (Ed.)

Springer

Janusz M. Bujnicki (Ed.)

Practical Bioinformatics

With 53 Figures, 10 of Them in Color, and 14 Tables

Springer

Dr. Janusz M. Bujnicki
Bioinformatics Laboratory
International Institute of
Molecular and Cell Biology
Trojdena 4
02-109 Warsaw
Poland

ISSN 0933-1891
ISBN 3-540-20613-2 Springer-Verlag Berlin Heidelberg New York

Library of Congress Cataloging-in-Publication Data

Practical bioinformatics / Janusz M. Bujnicki (ed.).
p. cm. -- (Nucleic acids and molecular biology ; vol. 15)
Includes bibliographical references and index.
ISBN 3-540-20613-2 (alk. paper)
1. Proteins–Research--Data processing 2. Amino acid sequence–Data processing. 3. Nucleic acids–Research--Data processing. 4. Nucleic sequence–Data processing. 5. Bioinformatics. I. Bujnicki, Janusz M. II. Series

QP260.N795 vol. 15
[QP551]
572.8 s–dc22
[572.8]

2003067348

Production and typesetting: Friedmut Kröner, 69115 Heidelberg, Germany
Cover design: *design & production* GmbH, 69126 Heidelberg, Germany

31/3150 YK – 5 4 3 2 1 0 – Printed on acid free paper

Nucleic Acids and Molecular Biology

15

Series Editor
H. J. Gross

Springer

Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Preface

The past decade has witnessed not only a flood of protein sequence and structure data generated by large-scale genomic sequencing and structural genomics projects, but also an ensuing growth of size and number of databases and computer programs designed to manage and process these data. The multitude of bioinformatic tools available to molecular biologists offers multiple solutions to various steps of process sequence–structure–function analyses. Often the choice of which tool to use depends more on its popularity among relatively naïve *users*, sometimes stemming from the availability of an intuitive web-server interface, rather than on an understanding of the underlying principles or on the user's ability to utilize all the information returned by the program, including the assessment of confidence of the results. Being educated and trained in molecular biology and biochemistry and self-taught in bioinformatics, I am interested in both the development of computational tools and their optimal application in the realm of experimental biology, especially in the studies of protein–nucleic acid interactions. Despite the abundance of literature on bioinformatics and on molecular biology of proteins that interact with nucleic acids, there are few (if any) timely volumes dedicated to the synthesis of these two research areas. Hence, I was delighted to accept the invitation to act as an editor of a "*Practical Bioinformatics*" volume of *Nucleic Acids and Molecular Biology* and to consolidate key bioinformatic methods for studying protein sequence–structure-function relationships into a convenient source.

This volume is mainly for the biochemist or molecular biologist who wants to analyze, search or manipulate protein structure or sequence data and to integrate these analyses with their experimental investigations to interpret the obtained results or to plan further studies better. Thus, the first part of the volume comprises reviews of methodology solicited from developers of bioinformatic software (with the emphasis on methods that explicitly utilize experimental information and/or are designed to guide experimental research), while the second part comprises useful strategies for studying protein function with the aid of bioinformatics, described in the form of "case

studies" by at-the-bench scientists. Methods and strategies range from protein structure prediction by template-dependent (comparative modeling, fold-recognition) and template-independent (ab initio) approaches, to prediction of protein–protein and protein–nucleic acid interactions, to identification of proteins exerting a defined function or prediction of the function for newly identified proteins. In the spirit of this series, all case studies involve analyses of proteins involved in interactions with nucleic acids – from ribosome assembly and structure, to posttranscriptional RNA modification, to DNA restriction and repair.

The bioinformatics field is a very fast-moving one, and every effort was made to produce this volume as rapidly as possible so the methods would be timely. In this regard, I am grateful to all the authors for taking their time to contribute and for adhering to a set of rigid deadlines; without their participation this volume would not have been possible. I hope that *Practical Bioinformatics* will serve as a useful compendium of methods both to newcomers in the field of bioinformatics-aided experimental molecular biology and biochemistry as well as to scientists actively engaged in research in this area.

Warsaw, July 2003 *Janusz M. Bujnicki*

Contributors

ALBER, FRANK

Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Genentech Hall, Suite N472D, 600 16th Street, University of California at San Francisco, San Francisco, California 94143–2240, USA

BUJNICKI, JANUSZ M. (e-mail: iamb@genesilico.pl)

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, 02-109 Warsaw, Poland

BYSTROFF, CHRISTOPHER (e-mail: bystrc@rpi.edu)

Department of Biology, Rensselaer Polytechnic Institute, Troy, New York 12180, USA

CATHERINOT, VINCENT

Centre de Biochimie Structurale, INSERM U554 - CNRS UMR5048 - Université Montpellier I. 15, Ave. Charles Flahault, 34060 Montpellier Cedex, France

COHEN-GONSAUD, MARTIN

Centre de Biochimie Structurale, INSERM U554 - CNRS UMR5048 - Université Montpellier I. 15, Ave. Charles Flahault, 34060 Montpellier Cedex, France

CRÉCY-LAGARD, VALÉRIE DE (e-mail: vcrecy@scripps.edu)

Molecular Biology Department, The Scripps Research Institute, BCC-379, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA

Cymerman, Iwona A.

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, 02-109 Warsaw, Poland

Douguet, Dominique

Centre de Biochimie Structurale, INSERM U554 - CNRS UMR5048 - Université Montpellier I. 15, Ave. Charles Flahault, 34060 Montpellier Cedex, France

Droogmans, Louis

Laboratoire de Microbiologie, Université Libre de Bruxelles, 1 av. E. Gryson, B-1070 Bruxelles, Belgium, and Laboratoire de Génétique des Procaryotes, Université Libre de Bruxelles, 12 rue des Professeurs Jeener et Brachet, 6041 Gosselies, Belgium

Eswar, Narayanan

Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Genentech Hall, Suite N472D, 600 16th Street, University of California at San Francisco, San Francisco, California 94143–2240, USA

Feder, Marcin

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, 02-109 Warsaw, Poland

Fischer, Daniel

Bioinformatics, Dept. Computer Science, Ben Gurion University, Beer-Sheva 84015, Israel

Friedhoff, Peter (e-mail: Friedhoff@chemie.bio.uni-giessen.de)

Institut für Biochemie, FB 08, Justus-Liebig Universität Giessen, Heinrich-Buff-Ring 58, 35395 Giessen, Germany

Grosjean, Henri

Laboratory of Structural Enzymology and Biochemistry, CNRS, 1 av. De la Terrasse, 91198 Gif-sur-Yvette, France

Kurowski, Michał A.

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, 02-109 Warsaw, Poland

Labesse, Gilles (e-mail: Labesse@cbs.cnrs.fr)

Centre de Biochimie Structurale, INSERM U554 - CNRS UMR5048 - Université Montpellier I. 15, Ave. Charles Flahault, 34060 Montpellier Cedex, France

Lapeyre, Bruno

Centre de Recherche de Biochimie Macromoléculaire du CNRS, 1919 Route de Mende 34293, Montpellier, France

Linge, Jens P.

Unité de Bio-Informatique Structurale, Institut Pasteur, 25–28 rue du docteur Roux, 75015 Paris, France

Mushegian, Arcady (e-mail: arm@stowers-institute.org)

Stowers Institute for Medical Research, 1000 E 50th St., Kansas City, Missouri 64110, USA

Nilges, Michael (e-mail: nilges@pasteur.fr)

Unité de Bio-Informatique Structurale, Institut Pasteur, 25–28 rue du docteur Roux, 75015 Paris, France

Pawłowski, Marcin

Bioinformatics Laboratory, International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, 02-109 Warsaw, Poland

Purushothaman, Suresh K.

Centre de Recherche de Biochimie Macromoléculaire du CNRS, 1919 Route de Mende 34293 Montpellier, France

Sali, Andrej (e-mail:sali@salilab.org)

Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Genentech Hall, Suite N472D, 600 16th Street, University of California at San Francisco, San Francisco, California 94143–2240, USA

Shao, Yu

Department of Biology, Rensselaer Polytechnic Institute, Troy, New York 12180, USA

Zharkov, Dmitry O. (e-mail: dzharkov@niboch.nsc.ru)

Institute of Chemical Biology and Fundamental Medicine, Siberian Division of Russian Academy of Sciences, Novosibirsk 630090, Russia

Contents