



Swiss Institute of
Bioinformatics

Bioinformatics

A Swiss Perspective

Ron D Appel • Ernest Feytmans

editors

Bioinformatics

A Swiss Perspective

editors

Ron D Appel

Ernest Feytmans

Swiss Institute of Bioinformatics, Switzerland

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

BIOINFORMATICS

A Swiss Perspective

Copyright © 2009 by World Scientific Publishing Co. Pte. Ltd. and
the Swiss Institute of Bioinformatics

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-283-877-3

ISBN-10 981-283-877-5

Typeset by Stallion Press

Email: enquiries@stallionpress.com

Printed in Singapore by Mainland Press Pte Ltd

Bioinformatics

A Swiss Perspective

Foreword

Computational sciences have their roots in the development of increasingly powerful computers over the last few decades. Rather rapidly, the instrumentation and the newly developed methodology with the underlying algorithms became widely appreciated and used as novel research strategies serving in many different fields of academic investigation, particularly in natural sciences and engineering sciences, but also in social sciences and the humanities. Computational sciences have been recognized for their invaluable contributions to data collection, data storage, data handling, and data analysis, thus leading to efficient strategies of modeling, prediction, and design of molecular structures and of their functional properties that are often of immediate relevance for the medical sciences. Computational comparisons of DNA sequences from different organisms provide invaluable insights into past evolutionary developments, and this has become a powerful new tool in the systematics of living organisms.

The present book on computational biology testifies to the impact that this field of investigation exerts on many pending questions in the life sciences. Particular reference is thereby made to pioneering contributions that have their roots in Switzerland. In the early 1990s, I had the privilege, in my position as a member of the Swiss Science Council, to visit a group of scientists working at the University of Geneva on computer-assisted handling and storage of data related to protein sequences. The prospective value of their work became immediately obvious. A very positive recommendation that was then formulated by the Swiss Science Council may have influenced the support provided by the political leaders, and this may have also facilitated the creation of the

Swiss Institute of Bioinformatics (SIB) 10 years ago. The SIB is an active, virtual network of scientists working in different Swiss institutions in various fields of bioinformatics. While pivotal contributions of these researchers found wide recognition, this development also favored the access of the Swiss pioneers in bioinformatics to related, often complementary studies conducted in other countries and other continents.

May this overview of important aspects of bioinformatics further contribute to strengthen international contacts and serve as a testament to such a fruitful development for the basic as well as for the applied sciences.

Werner Arber

Professor Emeritus for Molecular Microbiology
Biozentrum, University of Basel, Switzerland
Nobel Laureate in Medicine 1978

Preface

Biological research and recent technological advances have resulted in an enormous increase in research data that requires large storage capacities, powerful computing resources, and accurate data analysis algorithms. Bioinformatics is the field that provides these resources to life science researchers. Originally the prerogative of a few biologists who were computer buffs and a few computer scientists who had a passion for life science, bioinformatics has developed into a fully-fledged science with its own specialists, pregraduate and postgraduate studies, as well as international conferences and journals.

In Switzerland, bioinformatics started in the 1980s, when a handful of enthusiastic scientists started developing databases and computational tools that rapidly became accessed and used by researchers from all over the world. While essential for thousands of scientists both in Switzerland and abroad, these computational biology resources were developed for many years without any substantial means. It was only in 1998 that the need for a Swiss bioinformatics infrastructure was recognized and that the Swiss Institute of Bioinformatics (SIB) was created. Even then, the SIB founders were ahead of their time, building an organization of exceptional quality and dynamism. Starting with 20 bioinformaticians in five groups at universities and research institutes in the Lake Geneva area, they were rapidly joined by new professors in major Swiss universities, thus growing into an institution of national importance, recognized worldwide for its state-of-the-art work.

After more than 10 years of existence, the SIB is regarded as one of the leading bioinformatics institutions in the world. Organized as a federation of bioinformatics research groups from Swiss universities and research institutes, the SIB provides services to the life science community

that are highly appreciated worldwide, and coordinates research and education in bioinformatics nationwide. The SIB plays a central role in life science research both in Switzerland and abroad by developing extensive and high-quality bioinformatics resources that are essential for all life scientists. It contributes to the economy and quality of life through the global distribution of its products, by providing state-of-the-art tools to the industry, and by its involvement in pregraduate and postgraduate teaching programs. Knowledge developed by SIB members in areas such as genomics, proteomics, and systems biology is directly transformed by academia and industry into innovative solutions to improve global health.

This astounding concentration of talent in a given field is unusual and unique in Switzerland. This book gives an insight into some of the key areas of activity in bioinformatics in our country, covering both research work and major infrastructural efforts in genome and gene expression analysis, investigations on proteins and proteomes, evolutionary bioinformatics, and modeling of biological systems.

We are grateful to the authors of all chapters for their efforts, patience, and goodwill, without which it would not have been possible to publish this book. We are particularly indebted towards our colleague Dr. Patricia Palagi for her careful proofreading of the whole book; and are grateful to Drs. Lydie Bougueleret, Janet James, Tania Lima, and Ms. Nicole Zaghia, who went over the text of our authors whose English is not always as outstanding as their science. We also acknowledge support from the Swiss State Secretariat for Education and Research, the Swiss universities, federal institutes of technology and research institutes, the Swiss National Science Foundation, as well as several international funding bodies such as the European Research and Development Programmes and the US National Institutes of Health. Finally, we are thankful to all members of the Swiss Bioinformatics Institute, whose relentless work constitutes the core of bioinformatic excellence in Switzerland.

We hope that this book will help readers understand some of the many facets of bioinformatics today, and encourage new scientists to get on board with such a fascinating and challenging field.

Ron D. Appel and Ernest Feytmans

July 2008

List of Contributors

Prof. Amos Bairoch

Department of Structural Biology and Bioinformatics
Faculty of Medicine
University of Geneva

Founder Member and Group Leader, Swiss-Prot group
Swiss Institute of Bioinformatics
CMU – 1, rue Michel Servet
1211 Geneva 4
Switzerland

Prof. Sven Bergmann

Department of Medical Genetics
Faculty of Biology and Medicine
University of Lausanne

Group Leader, Computational Biology Group
Swiss Institute of Bioinformatics
Quartier UNIL-Bugnon, Rue du Bugnon 27
1005 Lausanne
Switzerland

Dr. Gabriel Jīvasattha Bittar

Fondation Jīvasattha and Jīvarakkhī
Buddhāyatana, Warrawee Road, American River
Kangaroo Island
SA 5221
Australia

Dr. Philipp Bucher

Swiss Institute for Experimental Cancer Research
Ecole Polytechnique Fédérale de Lausanne

Group Leader, Computational Cancer Genomics
Swiss Institute of Bioinformatics
EPFL SV — AAB Station 15
1015 Lausanne
Switzerland

Dr. Michel Cuendet

Molecular Modeling Group
Swiss Institute of Bioinformatics
University of Lausanne
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Dr. Mauro Delorenzi

NCCR Molecular Oncology
Group Leader, Bioinformatics Core Facility
Swiss Institute of Bioinformatics
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Manuel Gil

Centre for Computational Biology and Bioinformatics
Department of Computer Science
Swiss Federal Institute of Technology, Zürich

Computational Biochemistry Research Group
Swiss Institute of Bioinformatics
CAB F 61.2
Universitätstrasse 6
8092 Zürich
Switzerland

Dr. Darlene R. Goldstein

Institut de mathématiques
Ecole Polytechnique Fédérale de Lausanne

Swiss Institute of Bioinformatics
Bâtiment MA, Station 8
1015 Lausanne
Switzerland

Prof. Gaston H. Gonnet

Center for Computational Biology and Bioinformatics
Department of Computer Science
Swiss Federal Institute of Technology, Zürich

Group Leader, Computational Biochemistry Research Group
Swiss Institute of Bioinformatics
CAB H 66, Universitätstrasse 6
8092 Zürich
Switzerland

Dr. Aurélien Grosdidier

Molecular Modeling Group
Swiss Institute of Bioinformatics
University of Lausanne
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Dr. Lukasz Jaskiewicz

Computational and Systems Biology
Biozentrum
University of Basel

RNA Regulatory Networks
Swiss Institute of Bioinformatics
University of Basel, Klingelbergstrasse 50/70
4056 Basel
Switzerland

Dr. Evgenia V. Kriventseva

Department of Structural Biology and Bioinformatics
Faculty of Medicine, University of Geneva
CMU – 1, rue Michel Servet
1211 Geneva 4
Switzerland

Dr. Lydie Lane

Co-director CALIPHO team, Swiss-Prot group
Swiss Institute of Bioinformatics
CMU – 1, rue Michel Servet
1211 Geneva 4
Switzerland

Dr. Frédérique Lisacek

Group Leader, Proteome Informatics Group
Swiss Institute of Bioinformatics
CMU – 1, rue Michel Servet
1211 Geneva 4
Switzerland

Prof. Olivier Michielin

Multidisciplinary Oncology Center
University Hospital, Lausanne
Ludwig Institute for Cancer Research
Lausanne Branch
Group Leader, Molecular Modelling Group
Swiss Institute of Bioinformatics
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Dr. Patricia M. Palagi

Proteome Informatics Group
Swiss Institute of Bioinformatics
CMU – 1, rue Michel Servet
1211 Geneva 4
Switzerland

Prof. Manuel C. Peitsch

Director, Computational Sciences and Bioinformatics
Philip Morris R&D
Neuchatel
Switzerland

Chairman of the Executive Board
Swiss Institute of Bioinformatics
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Prof. Marc Robinson-Rechavi

Department of Ecology and Evolution
University of Lausanne

Group Leader, Evolutionary Bioinformatics Group
Swiss Institute of Bioinformatics
Quartier Sorge, Bâtiment Biophore
1015 Lausanne
Switzerland

Dr. Ute F. Röhrig

Multidisciplinary Oncology Center,
Lausanne University Hospital

Ludwig Institute for Cancer Research
Lausanne Branch

Molecular Modelling Group
Swiss Institute of Bioinformatics
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Prof. Ivo F. Sbalzarini

Institute of Theoretical Computer Science

Swiss Federal Institute of Technology, Zurich

Group leader, Computational Biophysics Laboratory

Swiss Institute of Bioinformatics

CAB G 34

Universitätstrasse 6

8092 Zürich

Switzerland

Prof. Torsten Schwede

Biozentrum, University of Basel

Group Leader, Computational Structural Biology

Swiss Institute of Bioinformatics

Klingelbergstrasse 50/70

4056 Basel

Switzerland

Dr. Bernhard Pascal Sonderegger

Swiss Institute of Experimental Cancer Research

Ecole Polytechnique Fédérale de Lausanne

Computational Systems Biology Group

Swiss Institute of Bioinformatics

AAB 0 18, Station 15

1015 Lausanne

Switzerland

Prof. Christian von Mering

Group Leader, Institute of Molecular Biology

University of Zurich

Group Leader, Bioinformatics/Systems Biology Group

Swiss Institute of Bioinformatics

Room Y55-L76

Winterthurerstrasse 190

8057 Zurich

Switzerland

Robert M. Waterhouse

Division of Cell and Molecular Biology
Faculty of Natural Sciences
Imperial College London
London SW7 2AZ
United Kingdom

Pratyaksha Wirapati

NCCR Molecular Oncology
Bioinformatics Core Facility
Swiss Institute of Bioinformatics
Quartier Sorge, Bâtiment Génopode
1015 Lausanne
Switzerland

Prof. Mihaela Zavolan

Biozentrum
University of Basel

Group Leader, RNA Regulatory Networks
Swiss Institute of Bioinformatics
Klingelbergstrasse 50/70
4056 Basel
Switzerland

Prof. Evgeny M. Zdobnov

Department of Genetic Medicine and Development
University of Geneva

Division of Cell and Molecular Biology
Faculty of Natural Sciences
Imperial College London, UK

Group Leader, Computational Evolutionary Genomics Group
Swiss Institute of Bioinformatics
CMU – 1, rue Michel Servet
1211 Geneva 4
Switzerland

Dr. Vincent Zoete

Molecular Modelling Group

Swiss Institute of Bioinformatics

Quartier Sorge, Bâtiment Génopode

1015 Lausanne

Switzerland

Contents

<i>Foreword</i>		v
<i>Preface</i>		vii
<i>List of Contributors</i>		xiii
<hr/>		
SECTION I	GENES AND GENOMES	1
<hr/>		
Chapter 1	Methods for Discovery and Characterization of DNA Sequence Motifs	3
	<i>Philipp Bucher</i>	
Chapter 2	Comparative Genome Analysis	33
	<i>Robert M. Waterhouse, Evgenia V. Kriventseva and Evgeny M. Zdobnov</i>	
Chapter 3	From Modules to Models: Advanced Analysis Methods for Large-Scale Data	59
	<i>Sven Bergmann</i>	
Chapter 4	Integrated Analysis of Gene Expression Profiling Studies — Examples in Breast Cancer	85
	<i>Pratyaksha Wirapati, Darlene R. Goldstein and Mauro Delorenzi</i>	