

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For other titles published in this series, go to
www.springer.com/series/7651

Proteome Bioinformatics

Edited by

Simon J. Hubbard

Faculty of Life Sciences, The University of Manchester, Manchester, UK

Andrew R. Jones

Department of Pre-clinical Veterinary Science, University of Liverpool, Liverpool, UK

Editor

Simon J. Hubbard
Faculty of Life Sciences
The University of Manchester
Manchester M13 9PT
UK
simon.hubbard@manchester.ac.uk

Andrew R. Jones
Department of Pre-clinical Veterinary Science
University of Liverpool
Liverpool L69 7ZJ
UK
andrew.jones@liv.ac.uk

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-60761-443-2 e-ISBN 978-1-60761-444-9
DOI 10.1007/978-1-60761-444-9
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009941056

© Humana Press, a part of Springer Science+Business Media, LLC 2010

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Humana Press, c/o Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of going to press, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Humana Press is a part of Springer Science+Business Media (www.springer.com)

Preface

The field of proteomics moves rapidly. New methods, techniques, applications, standards, models and software appear almost on a daily basis. Accompanying this are plenty of texts on the experimental side of the field and a few appearing on the informatic and data analysis side. This latterly includes one in the *Methods in Molecular Biology* series tackling the specific analysis of “Mass spectrometry data in proteomics” in MMB vol. 376. This current collection builds on this, but takes a broader view of proteome data analysis covering data analysis essentials, but also the databases and data models, as well as practical considerations for analysing database search results, annotating genomes, and speeding up searches. It also digs deeper into some topics, such as decoy database searching and aspects of signal processing in proteomic mass spectrometry. The aim of the volume is to provide the reader with a mix of reviews and methodology chapters, which build from the essentials of database searching in proteomics, on through specific data processing challenges to databases, data standards and data models.

The computational challenges facing proteomics are many and should not be underestimated. The direction in post-genome science is one of increasing complexity and larger, richer datasets. Proteomics is no exception and most active proteomics labs retain or work closely with bioinformaticians and computer scientists as they play an increasingly important role in the day-to-day running of the lab. This book covers all the essential topics that need to be considered and should help the novice and expert alike address their data analysis and management problems.

Manchester, UK
Liverpool, UK

Simon J Hubbard
Andrew R Jones

Contents

<i>Preface</i>	<i>v</i>
<i>Contributors</i>	<i>ix</i>
1 An Introduction to Proteome Bioinformatics	1
<i>Andrew R. Jones and Simon J. Hubbard</i>	
2 Bioinformatics Methods for Protein Identification Using Peptide Mass Fingerprinting	7
<i>Zhao Song, Luonan Chen, and Dong Xu</i>	
3 Computational Approaches to Peptide Identification via Tandem MS	23
<i>Simon J. Hubbard</i>	
4 Scoring and Validation of Tandem MS Peptide Identification Methods	43
<i>Markus Brosch and Jyoti Choudhary</i>	
5 Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics	55
<i>Joshua E. Elias and Steven P. Gygi</i>	
6 Understanding and Exploiting Peptide Fragment Ion Intensities Using Experimental and Informatic Approaches	73
<i>Ashley C. Gucinski, Eric D. Dodds, Wenzhou Li, and Vicki H. Wysocki</i>	
7 Spectral Library Searching for Peptide Identification via Tandem MS	95
<i>Henry Lam and Ruedi Aebersold</i>	
8 <i>De Novo</i> Sequencing Methods in Proteomics	105
<i>Christopher Hughes, Bin Ma, and Gilles A. Lajoie</i>	
9 Cross Species Proteomics	123
<i>J.C. Wright, R.J. Beynon, and S.J. Hubbard</i>	
10 Gene Model Detection Using Mass Spectrometry	137
<i>Bindu Nanduri, Nan Wang, Mark L. Lawrence, Susan M. Bridges, and Shane C. Burgess</i>	
11 Signal Processing in Proteomics	145
<i>Rene Hussong and Andreas Hildebrandt</i>	
12 A High-Performance Reconfigurable Computing Solution for Peptide Mass Fingerprinting	163
<i>Daniel Coca, Istvan Bogdan, and Robert J. Beynon</i>	
13 Mining Proteomic MS/MS Data for MRM Transitions.	187
<i>Jennifer A. Chem (Mead), Luca Bianco, and Conrad Bessant</i>	
14 OpenMS and TOPP: Open Source Software for LC-MS Data Analysis	201
<i>Knut Reinert and Oliver Kohlbacher</i>	
15 Trans-Proteomic Pipeline: A Pipeline for Proteomic Analysis.	213
<i>Patrick G.A. Pedrioli</i>	

16	Informatics and Statistics for Analyzing 2-D Gel Electrophoresis Images	239
	<i>Andrew W. Dowsey, Jeffrey S. Morris, Howard B. Gutstein, and Guang-Zhong Yang</i>	
17	Automated Generic Analysis Tools for Protein Quantitation Using Stable Isotope Labeling	257
	<i>Wen-Lian Hsu and Ting-Yi Sung</i>	
18	An Overview of Label-Free Quantitation Methods in Proteomics by Mass Spectrometry.	273
	<i>Jason W.H. Wong and Gerard Cagney</i>	
19	The PeptideAtlas Project.	285
	<i>Eric W. Deutsch</i>	
20	Using the PRIDE Proteomics Identifications Database for Knowledge Discovery and Data Analysis	297
	<i>Philip Jones and Lennart Martens</i>	
21	Molecular Interactions and Data Standardisation.	309
	<i>Sandra Orchard and Samuel Kerrien</i>	
22	Mass Spectrometer Output File Format mzML.	319
	<i>Eric W. Deutsch</i>	
23	Managing Experimental Data Using FuGE	333
	<i>Andrew R. Jones and Allyson L. Lister</i>	
24	Proteomics Data Collection (ProDaC): Publishing and Collecting Proteomics Data Sets in Public Repositories Using Standard Formats	345
	<i>Christian Stephan, Martin Eisenacher, Michael Kohl, and Helmut E. Meyer</i>	
25	Computational Resources for the Prediction and Analysis of Native Disorder in Proteins	369
	<i>Melissa M. Pentony, Jonathan Ward, and David T. Jones</i>	
	<i>Index.</i>	395

Contributors

- RUEDI AEBERSOLD • *Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland; The Institute for Systems Biology, Seattle, WA, USA*
- CONRAD BESSANT • *Bioinformatics Group, Cranfield University, Cranfield, Bedfordshire, UK*
- ROBERT J. BEYNON • *Department Veterinary Preclinical Sciences, University of Liverpool, Crown Street, Liverpool, UK*
- LUCA BIANCO • *Bioinformatics Group, Cranfield University, Cranfield, Bedfordshire, UK*
- ISTVAN BOGDAN • *Department of Automatic Control & Systems Engineering, University of Sheffield, UK*
- SUSAN M. BRIDGES • *Department of Computer Science and Engineering, the Institute for Digital Biology, Mississippi State University, Mississippi State, MS, USA*
- MARKUS BROSCHE • *The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
- SHANE C. BURGESS • *College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA; Institute for Digital Biology, Mississippi State University, Mississippi State, MS, USA; Life Sciences and Biotechnology Institute, Mississippi State University, Mississippi State, MS, USA*
- GERARD CAGNEY • *Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4, Ireland*
- LUONAN CHEN • *Department of Computer Science, University of Missouri, Columbia, MO, USA*
- JYOTI CHOUDHARY • *The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*
- DANIEL COCA • *Department of Automatic Control & Systems Engineering, University of Sheffield, UK*
- ERIC W. DEUTSCH • *Institute for Systems Biology, Seattle, WA, USA*
- ERIC D. DODDS • *Department of Chemistry and Biochemistry, The University of Arizona, Tucson, AZ, USA*
- ANDREW W. DOWSEY • *Institute of Biomedical Engineering, Imperial College London, UK*
- MARTIN EISENACHER • *Medizinisches Proteom-Center (MPC), Ruhr-Universitaet Bochum, Bochum, Germany*
- JOSHUA E. ELIAS • *Department of Cell Biology, Harvard Medical School, Boston, MA, USA*
- ASHLEY C. GUCINSKI • *Department of Chemistry and Biochemistry, The University of Arizona, Tucson, AZ, USA*
- HOWARD B. GUTSTEIN • *Department of Anesthesiology and Pain Management & Department of Biochemistry and Molecular Biology, The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA*

- STEVEN P. GYGI • *Taplin Biological Mass Spectrometry Facility, Harvard Medical School, Boston, MA, USA*
- ANDREAS HILDEBRANDT • *Center for Bioinformatics, Saarland University, Saarbrücken, Germany*
- WEN-LIAN HSU • *Institute of Information Science, Academia Sinica, Taiwan*
- SIMON J. HUBBARD • *Faculty of Life Sciences, University of Manchester, Michael Smith Building, Manchester, UK*
- CHRISTOPHER HUGHES • *Department of Biochemistry, University of Western Ontario, London, ON, Canada*
- RENE HUSSONG • *Center for Bioinformatics, Saarland University, Saarbrücken, Germany*
- Andrew R. Jones • *Department of Pre-clinical Veterinary Science, Faculty of Veterinary Science, University of Liverpool, Liverpool, UK*
- DAVID JONES • *Department of Computer Science, University College London, Gower Street, London, UK*
- PHILIP JONES • *EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
- SAMUEL KERRIEN • *EMBL – European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK*
- MICHAEL KOHL • *Medizinisches Proteom-Center (MPC), Ruhr-Universitaet Bochum, Bochum, Germany*
- OLIVER KOHLBACHER • *Wilhelm Schickard Institute for Computer Science, Tübingen University, Tübingen, Germany*
- GILLES A. LAJOIE • *Department of Biochemistry, University of Western Ontario, London, ON, Canada*
- HENRY LAM • *Department of Chemical and Biomolecular Engineering, University of Science and Technology, Hong Kong*
- MARK L. LAWRENCE • *College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA; Institute for Digital Biology, Mississippi State University, Mississippi State, MS, USA; Life Sciences and Biotechnology Institute, Mississippi State University, Mississippi State, MS, USA*
- WENZHOULI • *Department of Chemistry and Biochemistry, The University of Arizona, Tucson, AZ, USA*
- ALLYSON L. LISTER • *CISBAN and School of Computing Science, Newcastle University, Newcastle upon Tyne, UK*
- BIN MA • *David Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada*
- LENNART MARTENS • *EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
- JENNIFER A. MEAD • *Bioinformatics Group, Cranfield University, Cranfield, Bedfordshire, UK*
- HELMUT E. MEYER • *Medizinisches Proteom-Center (MPC), Ruhr-Universitaet Bochum, Bochum, Germany*
- JEFFREY S. MORRIS • *Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA*

- BINDU NANDURI • *College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, USA; Institute for Digital Biology, Mississippi State University, Mississippi State, MS, USA*
- SANDRA ORCHARD • *EMBL – European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK*
- PATRICK G. A PEDRIOLI • *Institute of Biochemistry, Swiss Federal Institute of Technology Zürich (ETHZ), Zürich, Switzerland*
- MELISSA PENTONY • *Department of Computer Science, University College London, Gower Street, London, UK*
- KNUT REINERT • *Department Computer Science and Mathematics, Free University of Berlin, Berlin, Germany*
- ZHAO SONG • *Department of Computer Science, University of Missouri, Columbia, MO, USA*
- CHRISTIAN STEPHAN • *Medizinisches Proteom-Center (MPC), Ruhr-Universitaet Bochum, Bochum, Germany*
- TING-YI SUNG • *Institute of Information Science, Academia Sinica, Taiwan*
- NAN WANG • *Department of Computer Science and Engineering, the Institute for Digital Biology, Mississippi State University, Mississippi State, MS, USA*
- JONATHAN WARD • *Department of Computer Science, University College London, Gower Street, London, UK*
- JASON W.H. WONG • *UNSW Cancer Research Centre, University of New South Wales, Sydney, NSW, Australia*
- JAMES C. WRIGHT • *Department Veterinary Preclinical Sciences, University of Liverpool, Crown Street, Liverpool, UK*
Faculty of Life Sciences, University of Manchester, Manchester, UK
- VICKI H. WYSOCKI • *Department of Chemistry and Biochemistry, The University of Arizona, Tucson, AZ, USA*
- DONG XU • *Department of Computer Science, University of Missouri, Columbia, MO, Institute of Biomedical Engineering, Imperial College London, UK*