

# Proteomics of Microbial Pathogens

*Edited by*

*Peter R. Jungblut and Michael Hecker*



WILEY-VCH Verlag GmbH & Co. KGaA



**Proteomics of Microbial  
Pathogens**

*Edited by  
Peter R. Jungblut and  
Michael Hecker*

## 1807–2007 Knowledge for Generations

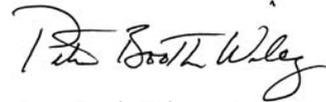
Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!



*William J. Pesce*  
President and Chief Executive Officer



*Peter Booth Wiley*  
Chairman of the Board

# Proteomics of Microbial Pathogens

*Edited by*

*Peter R. Jungblut and Michael Hecker*



WILEY-VCH Verlag GmbH & Co. KGaA

#### The Editors

**Dr. Peter R. Jungblut**  
Max-Planck Institut für  
Infektionsbiologie  
Schumannstr. 21–22  
10117 Berlin

**Prof. Dr. Michael Hecker**  
Institut für Mikrobiologie  
Ernst-Moritz-Arndt-Universität  
Friedrich-Ludwig-Jahn-Str. 15  
17487 Greifswald

#### Cover

Drawing of *E. coli* reproduced with kind  
permission of David S. Goodsell, The Scripps  
Research Institute

#### Cover-Design

A-Plus Design, Achim Bauer, Ludwigshafen

All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

**Library of Congress Card No.:**  
applied for

**British Library Cataloguing-in-Publication Data**  
A catalogue record for this book is available from the British Library.

**Bibliographic information published by the Deutsche Nationalbibliothek**  
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.d-nb.de>.

© 2007 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany  
Printed on acid-free paper

**Typesetting** X Con Media AG, Bonn  
**Printing** Strauss GmbH, Mörlenbach  
**Binding** Litges & Dopf Buchbinderei GmbH, Heppenheim

**ISBN:** 978–3-527–31759–2

## Table of Contents

<b>1</b>	<b>Genome and proteome analysis of <i>Chlamydia</i></b>	<b>1</b>
	<i>Brian B. S. Vandahl, Svend Birkelund and Gunna Christiansen</i>	
1.1	Introduction	1
1.1.1	<i>Chlamydia</i> biology	2
1.1.1.1	Diseases	2
1.1.1.2	The developmental cycle	3
1.2	<i>Chlamydia</i> genomes	5
1.2.1	Sequenced <i>Chlamydia</i> genomes	5
1.2.2	Chlamydial genes	6
1.2.3	Genome comparison	7
1.3	Proteome analysis of <i>Chlamydia</i>	9
1.3.1	Early <i>Chlamydia</i> proteome studies	10
1.3.2	<i>C. trachomatis</i> proteome studies	10
1.3.3	<i>C. pneumoniae</i> proteome studies	11
1.3.4	Identification of secreted proteins by comparative proteomics	13
1.3.5	Proteome studies of comc	15
1.3.6	Proteome comparison of <i>S. trachomatis</i> serovars	15
1.3.7	Proteome analysis of growth conditions	16
1.3.8	Considerations in proteomics	16
1.4	Concluding remarks	17
<b>2</b>	<b><i>Helicobacter pylori</i> vaccine development based on combined subproteome analysis</b>	<b>21</b>
	<i>Dirk Bumann, Peter R. Jungblut and Thomas F. Meyer</i>	
2.1	Introduction	21
2.2	Classical whole-cell inactivated <i>Helicobacter</i> vaccines	22
2.3	Subunit <i>Helicobacter</i> vaccines: Conventional antigen selection	22
2.4	Subunit <i>Helicobacter</i> vaccines: Global antigen selection based on proteomics	23
2.4.1	Proteomics as a tool for antigen characterization	23

2.4.2	The <i>Helicobacter</i> proteome	24
2.4.3	Criteria for promising antigen candidates	25
2.4.4	Identification of protective antigens based on multiple criteria	27
2.5	Concluding remarks	28

**3 Towards a comprehensive understanding of *Bacillus subtilis* cell physiology by physiological proteomics 31**

*Michael Hecker and Uwe Völker*

3.1	Introduction	32
3.2	Subproteomes vs. the total theoretical proteome	33
3.3	The vegetative proteome of growing cells	34
3.4	Proteomes of nongrowing cells – the adaptational network	39
3.5	Proteomic signatures – tools for microbial physiology and their practical application	48
3.6	Transcriptomics vs. proteomics – towards a second generation of proteomics	50
3.7	The interactome	52
3.8	The secretome	53
3.9	Post-translational modifications	53
3.10	Protein quality control/protein degradation at a proteomic scale	55
3.11	Gene expression network – horizontal and vertical approach	56
3.12	Concluding remarks	58

**4 Web-accessible proteome databases for microbial research 63**

*Klaus-Peter Pleißner, Till Eifert, Sven Buettner, Frank Schmidt, Martina Boehme, Thomas F. Meyer, Stefan H. E. Kaufmann and Peter R. Jungblut*

4.1	Introduction	64
4.2	Materials and methods	64
4.2.1	Data generation and data storage	64
4.2.2	Software tools	65
4.3	Results and discussion	65
4.3.1	Data management, analysis and presentation	65
4.3.2	2D-PAGE database	66
4.3.3	ICAT-LC/MS database	68
4.3.4	FUNC_CLASS database	68
4.3.5	Data analysis and visualization	69
4.4	Concluding remarks	73

<b>5</b>	<b>A targeted proteomics approach to the rapid identification of bacterial cell mixtures by matrix-assisted laser desorption/ionization mass spectrometry</b>	<b>75</b>
	<i>Bettina Warscheid and Catherine Fenselau</i>	
5.1	Introduction	75
5.2	Materials and methods	76
5.2.1	Chemicals	76
5.2.2	<i>Bacillus</i> strains	77
5.2.3	Vegetative cell digestion	77
5.2.4	MALDI-TOF MS and unimolecular decomposition product ion analysis	77
5.2.5	Database searches and identification of <i>Bacillus</i> species	78
5.3	Results and discussion	79
5.3.1	On-probe tryptic digestion of bacterial cells	79
5.3.1.1	<i>Bacillus subtilis</i> 168	79
5.3.1.2	<i>Bacillus globigii</i> and <i>sphaericus</i> 14577	79
5.3.1.3	<i>Bacillus cereus</i> T and <i>anthracis</i> Sterne	81
5.3.2	Partial sequencing for rapid identification of bacterial cells	83
5.3.2.1	Unimolecular decomposition product ion analysis (UDPIA)	83
5.3.2.2	Identification of bacteria	85
5.3.3	Identification of proteins and protein-families in bacterial cell digests	85
5.3.3.1	Flagellin and surface layer protein precursor	85
5.3.3.2	Cold shock and cold shock-like proteins	87
5.3.3.3	Ribosomal proteins	90
5.3.3.4	DNA-binding proteins	90
5.3.3.5	Heat shock proteins	91
5.3.3.6	Other stress related proteins and the prosthetic group of an acyl-carrier protein	92
5.3.4	Analysis of a 1:1 mixture of <i>B. globigii</i> and <i>B. sphaericus</i> 14577	92
5.4	Concluding remarks	94
<b>6</b>	<b>Protein identification and tracking in two-dimensional electrophoretic gels by minimal protein identifiers</b>	<b>97</b>
	<i>Jens Mattow, Frank Schmidt, Wolfgang Höhenwarter, Frank Siejak, Ulrich E. Schaible and Stefan H. E. Kaufmann</i>	
6.1	Introduction	98
6.2	Materials and methods	99
6.2.1	Materials	99
6.2.2	Two-dimensional electrophoresis	100
6.2.3	Mass spectrometry	100

6.2.4	Compilation of a theoretical dataset comprising all tryptic peptides of <i>M. tuberculosis</i> H37Rv	100
6.2.5	Comparison of MALDI spectra by the program MS-Screener	101
6.2.6	Determination of exogenous contaminant masses	101
6.2.7	Generation of template spectra	102
6.3	Results and discussion	102
6.3.1	Proteome analysis of <i>M. tuberculosis</i> H37Rv CSN	102
6.3.2	Determination of exogenous contaminant masses by MS-Screener	102
6.3.3	The MPI approach revealed HspX-specific peptide masses in multiple spectra	106
6.3.4	The MS-Screener analysis revealed truncated variants of Tuf previously not identified by PMF	111
6.3.5	Frequency of tryptic peptides with similar $m/z$ ratios in <i>M. tuberculosis</i> H37Rv	113
6.3.6	In mass spectrometry it is important to consider detection probabilities of proteins and peptides	114
6.4	Concluding remarks	117
<b>7</b>	<b>Continued proteomic analysis of <i>Mycobacterium leprae</i> subcellular fractions</b>	<b>121</b>
	<i>Maria Angela M. Marques, Benjamin J. Espinosa, Erika K. Xavier da Silveira, Maria Cristina V. Pessolani, Alex Chapeaurouge, Jonas Perales, Karen M. Dobos, John T. Belisle, John S. Spencer and Patrick J. Brennan</i>	
7.1	Introduction	121
7.2	Materials and methods	125
7.2.1	Preparation and separation of <i>M. leprae</i> proteins	125
7.2.2	MS	126
7.2.3	Isolation of basic proteins of <i>M. leprae</i>	127
7.3	Results and discussion	127
7.4	Concluding remarks	138
<b>8</b>	<b>CFP10 discriminates between nonacetylated and acetylated ESAT-6 of <i>Mycobacterium tuberculosis</i> by differential interaction</b>	<b>141</b>
	<i>Limei Meng Okkels, Eva-Christina Müller, Monika Schmid, Ida Rosenkrands, Stefan H. E. Kaufmann, Peter Andersen and Peter R. Jungblut</i>	
8.1	Introduction	141
8.2	Materials and methods	142
8.2.1	Protein samples	142
8.2.2	2-DE blot overlay	143
8.2.3	Mass spectrometry	143
8.3	Results	144
8.3.1	High resolution separation of acidic ST-CF proteins	144

8.3.2	Mass analysis of ESAT-6 spots	144
8.3.3	Interaction of recombinant CFP10 with ESAT-6 spots	149
8.4	Discussion	150
<b>9</b>	<b>The cell wall subproteome of <i>Listeria monocytogenes</i></b>	<b>153</b>
	<i>Jessica Schaumburg, Oliver Diekmann, Petra Hagedorff, Simone Bergmann, Manfred Rohde, Sven Hammerschmidt, Lothar Jänsch, Jürgen Wehland, and Uwe Kärst</i>	
9.1	Introduction	153
9.2	Material and methods	154
9.2.1	Bacterial strain and growth conditions	154
9.2.2	Serial extraction of cell wall proteins	155
9.2.3	Aminopeptidase C assay	155
9.2.4	SDS-PAGE Western blotting, and N-terminal sequencing	155
9.2.5	2-D-PAGE	155
9.2.6	Gel staining and protein identification by mass spectrometry	156
9.2.7	Immunoelectron microscopy	156
9.2.7.1	Postembedding labeling studies	156
9.2.7.2	Field emission scanning electron microscopic immunolabeling	157
9.2.8	Overlay blot	157
9.2.9	Ligand fishing	158
9.2.10	Cloning and purification of proteins	158
9.2.11	Kinetic analyses using SPR detection	159
9.2.12	Bioinformatic analysis	159
9.3	Results	160
9.3.1	Prediction and validation of cell wall-associated proteins	160
9.3.1.1	Prediction of exported proteins	160
9.3.1.2	Validation of the cell wall subproteome	160
9.3.2	Analysis of protein processing and localization	167
9.3.3	Analysis of plasminogen-binding proteins	168
9.4	Discussion	173
9.4.1	Analysis of proteins identified from surface extracts	173
9.4.2	Analysis of protein processing and localization in surface extracts	174
9.4.3	Possible function of plasminogen binding in virulence	175
9.5	Concluding remarks	177
<b>10</b>	<b>Low virulent strains of <i>Candida albicans</i>: Unravelling the antigens for a future vaccine</b>	<b>181</b>
	<i>Elena Fernández-Arenas, Gloria Molero, César Nombela, Rosalía Díez-Orejas and Concha Gil</i>	
10.1	Introduction	182
10.2	Materials and methods	183

10.2.1	Microorganism and culture conditions	183
10.2.2	Mice	183
10.2.3	Systemic infection conditions and generation of immune sera	183
10.2.4	T cell purification and passive immunization	184
10.2.5	2-DE	184
10.2.5.1	Protoplast lysate preparation	184
10.2.5.2	Analytical and micropreparative 2-DE	185
10.2.6	Immunoblot analyses	185
10.2.7	MALDI-TOF and MALDI-TOF MS analyses of spots	186
10.2.8	Database search	187
10.3	Results	187
10.3.1	Vaccination assays with different mutant strains and generation of immune sera	187
10.3.2	Importance of cellular immunity in vaccination with <i>C. albicans</i> CNC13	188
10.3.2.1	Role of Th1/Th2 cytokines in CNC13 immune response	188
10.3.2.2	Protection induced by passive transfer of sensitized CNC13 lymphocytes	189
10.3.3	Profile of <i>C. albicans</i> immunoreactive proteins in the different mutant strains	190
10.3.3.1	Detection and identification of the immunoreactive proteins	190
10.4	Discussion	194
10.4.1	Low virulent <i>C. albicans</i> strains as a tool to study the host immune response	194
10.4.2	<i>C. albicans</i> <i>hog1</i> mutant induces protection in a vaccination assay	196
10.4.3	<i>C. albicans</i> new antigenic proteins	197
10.4.4	Antibody profile linked to successful vaccination against systemic candidiasis	197
10.5	Concluding remarks	198
<b>11</b>	<b>Proteomic analysis of the sarcosine-insoluble outer membrane fraction of the bacterial pathogen <i>Bartonella henselae</i></b>	<b>203</b>
	<i>Thomas A. Rhomberg, Olof Karlberg, Thierry Mini, Ursula Zimny-Arndt, Ulrika Wickenberg, Marlene Röttgen, Peter R. Jungblut, Paul Jenö, Siv G. E. Andersson and Christoph Dehio</i>	
11.1	Introduction	203
11.2	Materials and methods	205
11.2.1	Strains and culture conditions	205
11.2.2	Enrichment of <i>B. henselae</i> OMPs	205
11.2.3	Protease exposure	205
11.2.4	1-D SDS-PAGE	206

11.2.5	Protein solubilization and protein quantitation	206
11.2.6	2-D NEPHGE	206
11.2.7	MALDI-TOF-MS	207
11.2.8	Database query	208
11.2.9	<i>In silico</i> analysis	208
11.3	Results	208
11.3.1	Enrichment of <i>B. henselae</i> OMPs	208
11.3.2	1-D SDS-PAGE of <i>B. henselae</i> OMPs and protein assignment by PMF	209
11.3.3	2-D NEPHGE of <i>B. henselae</i> OMPs and protein assignment by PMF	211
11.4	Discussion	215
<b>12</b>	<b>The influence of <i>agr</i> and <math>\sigma^B</math> in growth phase dependent regulation of virulence factors in <i>Staphylococcus aureus</i></b>	<b>225</b>
	<i>Anne-Kathrin Ziebandt, Dörte Becher, Knut Ohlsen, Jörg Hacker, Michael Hecker and Susanne Engelmann</i>	
12.1	Introduction	225
12.2	Material and methods	227
12.2.1	Bacterial strains and culture conditions	227
12.2.2	Preparation of the extracellular protein fraction	227
12.2.3	Analytical and preparative PAGE	227
12.2.4	Quantitation of protein spots	228
12.2.5	Transcriptional analyses	228
12.3	Results	228
12.3.1	Growth phase dependent regulation of extracellular proteins	228
12.3.2	The influence of <i>agr</i> and $\sigma^B$ on the extracellular proteome	236
12.3.3	The influence of <i>agr</i> and $\sigma^B$ on the transcription of virulence genes	239
12.3.4	Effect of $\sigma^B$ on the transcription of <i>agr</i> and <i>sarA</i>	239
12.4	Discussion	239
12.5	Concluding remarks	245
<b>13</b>	<b>Comparative proteome analysis of cellular proteins extracted from highly virulent <i>Francisella tularensis</i> ssp. <i>tularensis</i> and less virulent <i>F. tularensis</i> ssp. <i>holarctica</i> and <i>F. tularensis</i> ssp. <i>mediaasiatica</i></b>	<b>249</b>
	<i>Martin Hubálek, Lenka Hernychová, Martin Brychta, Juraj Lenčo, Jana Zechovská and Jiří Stulík</i>	
13.1	Introduction	250
13.2	Material and methods	251
13.2.1	Bacterial cultures and sample preparation for 2-DE	251
13.2.2	Two-dimensional gel electrophoresis	252
13.2.3	Statistical analysis	252

13.2.4	In-gel digestion	253
13.2.5	Mass spectrometric identification	253
13.3	Results and discussion	254
13.4	Concluding remarks	265
<b>14</b>	<b>Proteome comparison of <i>Vibrio cholerae</i> cultured in aerobic and anaerobic conditions</b>	<b>267</b>
	<i>Biao Kan, Hajar Habibi, Monika Schmid, Weili Liang, Ruibai Wang, Duochun Wang and Peter R. Jungblut</i>	
14.1	Introduction	267
14.2	Materials and methods	269
14.2.1	Strains and culture	269
14.2.2	Sample preparation for 2-DE	269
14.2.3	2-DE	269
14.2.4	MS	270
14.2.5	Electron microscope	270
14.3	Results and discussion	270
14.3.1	Identified protein species more abundant in aerobic culture	274
14.3.2	Identified protein species which are more abundant in anaerobic culture	275
14.4	Concluding remarks	276
<b>15</b>	<b>Induction of <i>Mycobacterium avium</i> proteins upon infection of human macrophages</b>	<b>279</b>
	<i>Lara Brunori, Federico Giannoni, Luca Bini, Sabrina Liberatori, Cristiane Frota, Peter Jenner, Ove Fredrik Thoresen, Graziella Orefici and Lanfranco Fattorini</i>	
<b>16</b>	<b>Proteomics-based identification of novel <i>Candida albicans</i> antigens for diagnosis of systemic candidiasis in patients with underlying hematological malignancies</b>	<b>289</b>
	<i>Aida Pitarch, Joaquín Abian, Montserrat Carrascal, Miguel Sánchez, César Nombela and Concha Gil</i>	
16.1	Introduction	290
16.2	Materials and methods	291
16.2.1	Human serum samples	291
16.2.2	Preparation of <i>C. albicans</i> protoplast lysates	292
16.2.3	Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE)	292
16.2.4	Immunoblot analysis	294
16.2.5	Peptide sample preparation for MS and MS/MS analyses	294
16.2.5.1	In-gel digestion	294
16.2.5.2	Desalting	294

16.2.6	Mass spectrometric analysis (MS and MS/MS)	295
16.2.6.1	Peptide mapping by MALDI-TOF-MS	295
16.2.6.2	Peptide fragmentation and sequencing by nanoESI-IT-MS	295
16.2.7	Database searching	296
16.3	Results	296
16.3.1	Mapping of <i>C. albicans</i> immunogenic proteins	296
16.3.1.1	Overall 2-D <i>C. albicans</i> antigen recognition pattern	296
16.3.1.2	Identification of <i>C. albicans</i> immunoreactive proteins by peptide mass fingerprinting (PMF)	297
16.3.1.3	Peptide sequencing of <i>C. albicans</i> immunoreactive proteins	298
16.3.1.4	Reference 2-DE <i>C. albicans</i> antigen pattern display on the Net	305
16.3.2	Comparison of 2-D <i>C. albicans</i> antigen recognition patterns obtained with serum samples from patients with and without systemic candidiasis	309
16.3.3	Differences in the 2-D <i>C. albicans</i> antigen recognition profile associated with infection progression	311
16.4	Discussion	312
16.4.1	<i>C. albicans</i> housekeeping enzymes can stimulate the human immune system during systemic candidiasis	316
16.4.1.1	Heat shock proteins (Hsps)	317
16.4.1.2	Metabolic proteins	318
16.4.1.3	Elongation factors and ribosomal proteins	318
16.4.1.4	Miscellaneous proteins	319
16.4.2	Natural anti- <i>Candida</i> antibodies might be correlated with differentiation of the human immune response	319
16.4.3	Serum levels of specific anti- <i>Candida</i> antibodies could be useful for the clinical follow-up of systemic candidiasis	320
	Index	325



## Preface

### Proteomics of microbial pathogens

Infectious diseases still plague mankind. According to the World Health Report 2004, 19.1% of the deaths estimated in 2002 were caused by infectious diseases. Aids, tuberculosis and malaria each contributed more than 2% to this figure. In June 2006, 387 completely sequenced genomes (<http://www.genomesonline.org/>) have been published in total, 352 of them from bacteria, an important prerequisite for the analysis of the proteomes of these organisms. In total 940 ongoing bacterial genome projects were reported.

The first successful proteome studies revealed vaccine candidates with promising results in animal models. Immunoproteomics resulted in the detection of antigens which may be used for diagnostics and vaccine candidate prediction. So it can be assumed that proteomics will make a marked contribution to the improvement of worldwide health within the next few years.

Here we look at some of the trends in this field. As there are so many microorganisms currently under investigation, it is not possible to present a comprehensive overview of microbial proteomics. Proteomics technology has been automated within recent years: spot picking, digestion, LC-MS/MS and database searches have increased throughput but produced new bottlenecks in quality control and data evaluation. Microorganisms are ideal models for the application of these new technologies. Bacteria with genomes containing 600 to 7000 predicted genes present a medium-sized complexity which can be used to apply proteomic techniques with a good chance of obtaining an overview of a substantial part of the proteome in combination with prefractionation procedures. Standardization is now an important theme in proteomic technology but the multiple properties of organisms and proteins make standardizing sample preparation nearly impossible. Even related bacteria need different procedures for sample preparation, as outlined in this book in the example of *Mycobacterium leprae*. It may be estimated that in one biological situation more than 50% of the predicted proteins may be identified for genomes such as *Mycoplasma pneumoniae* containing less than 1000 genes, 30% for those containing less than 2500 genes and only 10% for those containing more than 4000 genes. Subfractionation contributes to the number of accessible proteins, but in the future throughput has to be increased further to allow the presentation of the proteome in a kind of film with changing environ-

mental conditions. Only then may more complete proteomes become accessible. Bioinformatics accompanies proteomics through all the technological steps, allowing the data obtained to be stored in a database. A microbial proteomics database system was set up at the Max Planck Institute for Infection Biology (<http://www.mpiib-berlin.mpg.de/2D-PAGE/>) and by June 2006 it contains 18 bacterial species and 4889 identified spots. Peptide mass fingerprinting data are stored for *Helicobacter pylori* and isotopic labelling results are represented for *Mycobacterium tuberculosis* LC/MS data. Proteomics of microorganisms allow the scientist to start with a hypothesis-free global approach and focus early on the hypotheses elaborated from this first step. In the first few years we learned that posttranslational modifications play a more important role than expected in bacteria, and the resulting protein species composition may be directly visualized by 2-DE/MS but not by LC/MS which has other advantages such as higher throughput and sensitivity potentials. At the moment, for example, the impact of more than 10 ESAT-6 protein species in *Mycobacterium tuberculosis* remains unclear. Proteome analysis at the protein species level is a task for the future.

We wish to thank the authors for their contributions, the referees for their prompt reviewing of the manuscripts and the publishers for their help in producing this book. We also take this opportunity to thank the "Bundesministerium für Bildung und Forschung" in Germany for financing the project "New methods to access the complete proteome of bacteria" and the European Union for support in developing the European Bacterial Proteome Database within the project "Comparative analysis of proteome modulation in human pathogenic bacteria for the identification of new vaccines, diagnostics and antibacterial drugs" (QLRT-1999-31536). Several articles in this book were supported by these two initiatives.

Peter R. Jungblut  
Max Planck Institute for Infection Biology  
Core Facility Protein Analysis  
Berlin

Michael Hecker  
Ernst-Moritz-Arndt-Universität Greifswald  
Institut für Mikrobiologie  
Greifswald

## 1

**Genome and proteome analysis of *Chlamydia*\***

Brian B. S. Vandahl, Svend Birkelund and Gunna Christiansen

It has been difficult to study the molecular biology of the obligate intracellular bacterium *Chlamydia* due to lack of genetic transformation systems. Therefore, genome sequencing has greatly expanded the information concerning the biology of these pathogens. Comparing the genomes of seven sequenced *Chlamydia* genomes has provided information of the common gene content and gene variation. In addition, the genome sequences have enabled global investigation of both transcript and protein content during the developmental cycle of chlamydiae. During this cycle *Chlamydia* alternates between an infectious extracellular form and an intracellular dividing form surrounded by a phagosome membrane termed the chlamydial inclusion. Proteins secreted from the chlamydial inclusion into the host cell may interact with host cell proteins and modify the host cell's response to infection. However, identification of such proteins has been difficult because the host cell cytoplasm of *Chlamydia* infected cells cannot be purified. This problem has been circumvented by comparative proteomics.

## 1.1

**Introduction**

*Chlamydia* is an obligate intracellular bacterium comprising a number of important animal and human pathogens causing infections with serious sequelae. *Chlamydia trachomatis* is a cause of ocular and genital infections. *Chlamydophila pneumoniae* (previously *Chlamydia pneumoniae*) causes respiratory diseases and has been associated with asthma and atherosclerosis. Sequelae are primarily due to an inflammatory response, which may be sustained by bacteria persisting in the infected organism due to a special intracellular nonreplicative state [1] but delayed-type hypersensitivity may also be involved.

\* Originally published in *Proteomics* 2004, 10, 2831–2842

Molecular biological studies of *Chlamydia* have been hampered by the lack of genetic transformation systems. Therefore, sequencing of the genomes of several *Chlamydia* species has been especially important for chlamydial research. The *C. trachomatis* serovar D genome was published in 1998 [2] and in 1999 the first *C. pneumoniae* genome followed [3]. Today *Chlamydia* is one of the most extensively sequenced microorganisms with seven published genomes including four from different isolates of *C. pneumoniae* (<http://www.ncbi.nlm.nih.gov:80/PMGifs/Genomes/org.html>). Besides direct analysis of genome sequences (genomics), global investigation of transcripts (transcriptomics) and protein content (proteomics) are developed based on the genome sequences.

### 1.1.1

#### ***Chlamydia* biology**

Traditionally *Chlamydia* was the only genus in the family of *Chlamydiaceae* which was the only family in the order of *Chlamydiales*. Since the introduction of *C. pneumoniae* in 1989 [4] there were four species, distinguished mainly by serology: *C. pneumoniae*, *C. trachomatis*, *C. pecorum* and *C. psittaci*. In 1999 a new taxonomy was suggested [5], introducing more genera and species based on phylogenetic relationships with requirement of > 95% 16S rRNA identity within a genus. The suggested taxonomy placed *C. trachomatis* in the genus *Chlamydia* and divided the *C. trachomatis* into three species. The remaining *Chlamydia* species were placed in the new genus *Chlamydophila*, and *C. psittaci* was divided into a number of different species. Similar developmental biology, similar genome size and genome organization of *C. trachomatis* (1.0 Mb) and *C. pneumoniae* (1.2 Mb) [3], representing *Chlamydia* and *Chlamydophila*, respectively, indicate basic similarities but differences are also found [6]. In the present review the new taxonomy will be followed with respect to species names, but *Chlamydia* will be used as a unifying term describing both of the suggested genera *Chlamydophila* and *Chlamydia*.

#### 1.1.1.1 Diseases

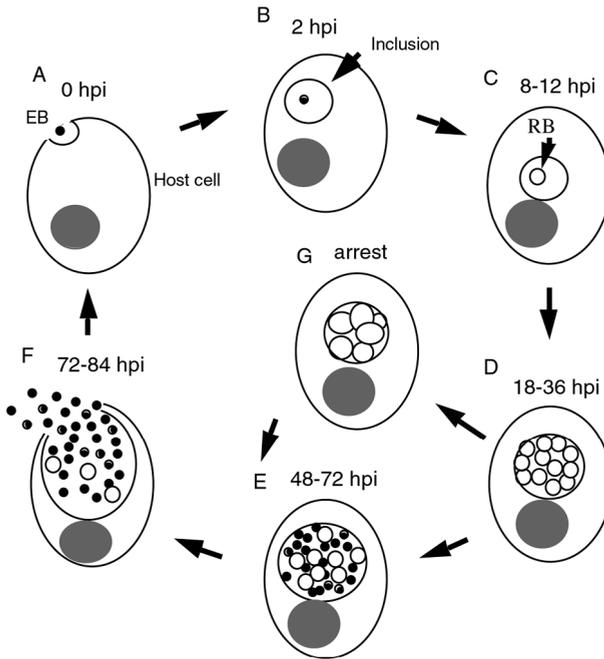
The main human pathogenic chlamydiae are *C. trachomatis* and *C. pneumoniae*, but also bird pathogenic *C. psittaci* can cause severe pneumonia, psittacosis, if transferred to humans [7]. *C. trachomatis* is divided into three groups of serovars: (i) serovars A–C are endemic in developing countries and the cause of trachoma, which may lead to blindness by scarring of the cornea [7]; (ii) serovars D–K are sexually transmitted and cause urethritis, cervicitis and salpingitis. It is the most widespread sexually transmitted bacterial disease and infections are often asymptomatic. The infection may cause sterility and increased risk for ectopic pregnancy by scarring of the fallopian tubes if it spreads from the cervix [7]; (iii) serovars L1–L3 are also sexually transmitted but cause lymphogranuloma venereum (LGV). LGV is a more severe infection as it readily spreads to the lymphatic system and becomes systemic [7]. Serovars A–K are known as the trachoma biovar and L1–3 as the LGV biovar.

*C. pneumoniae* is a respiratory pathogen that causes acute and chronic respiratory diseases. Most infections are asymptomatic, but about 30% cause more severe pneumonia, bronchitis or other upper airway illness [8]. About 10% of the cases of community acquired pneumonia in adults and about 5% of the cases of bronchitis and sinusitis are caused by *C. pneumoniae* [9]. Persistent infections have been described [10] and there are indications that treatment may not eliminate the organism [11]. *C. pneumoniae* has been associated with chronic lung diseases [8] and as a possible risk factor for the development of atherosclerosis [12]. *C. pneumoniae* has been detected in atherosclerotic lesions [13] and studies have shown that atheromatous plaques are commonly infected with *C. pneumoniae*. Animal studies suggest that *C. pneumoniae* can accelerate atherosclerosis-like disease [14, 15]. However, other studies fail to detect *C. pneumoniae* in plaques and many studies find no significant association by serology [16, 17]. At present it is not clear whether there is an increased risk of coronary artery disease due to *C. pneumoniae* infection and if there is, the increase may be small.

#### 1.1.1.2 The developmental cycle

*Chlamydia* is a Gram-negative, obligate intracellular bacterium, characterized by a biphasic developmental cycle. The developmental cycle (Fig. 1) in which the bacteria alternate between an infectious, extracellular form, the elementary body (EB) and a noninfectious intracellular replicating form, the reticulate body (RB) is unique for chlamydiae [18–20]. EBs are small rigid bodies of about 300 nm in diameter that are traditionally described as being metabolically inactive with their DNA packed by histone-like proteins [21, 22]. They are adapted for extracellular survival with a heavily disulfide cross-linked outer membrane, that provides osmotic stability. RBs are about 1  $\mu\text{m}$  in diameter with an outer membrane that is permeable for transport of host cell nutrients and the DNA is unpacked as in other bacteria.

Infectious EBs attach to a susceptible host cell by which they are phagocytosed. The exact mechanism is not known but the uptake is thought to be induced by the bacteria. Inside the phagosome, named the inclusion, the EBs develop into RBs, which divide by binary fission. This includes unpacking of the DNA and reduction of the disulfide bridges of the outer membrane [23], but it is not known what triggers these events. After multiple divisions, the RBs begin conversion into EBs, including packing of the DNA and synthesis of late outer membraneproteins that are disulfide bridged. Ultimately, a new generation of infectious EBs is released upon disruption of the host cell. The bacteria stay inside the inclusion throughout the intracellular stage, which lasts for 72–96 h for *C. pneumoniae* grown in cell culture. The inclusion membrane grows by the acquisition of lipids derived from the host cell [24–26]. It is modified by the insertion of chlamydial proteins, the so-called inclusion membrane proteins (inCs), and prevented from fusion with lysosomes [27, 28].



**Fig. 1** The developmental cycle of *Chlamydia*. Hours post infection (hpi) are listed for *C. pneumoniae* in cell culture. A, the infectious EB adheres to a host cell and is taken up by endocytosis. B, *Chlamydia* modifies the phagosome, the chlamydial inclusion, to escape the endocytic pathway. C, the EB develops into the metabolically active RB. D, the RBs divide by binary fission and the inclusion grows by incorporation of host cell derived lipids. E, after multiple divisions, the RBs reorganize into EBs. F, ultimately, a new generation of infectious EBs is released by lysis of the host cell. G, low nutrient availability, IFN- $\gamma$  mediated tryptophan starvation or other stressful conditions can trigger a persistent state with abnormal nondividing RBs. These RBs can be reactivated to enter the developmental cycle when the conditions are again suited for growth. Redrawn from [8].

The developmental cycle of *C. pneumoniae* can be arrested by interferon-gamma (IFN- $\gamma$ )-induced tryptophan catabolism of the host cell [29]. Tryptophan starvation leads to a nonproductive infection in which enlarged aberrant RBs evolve. These abnormal RBs do not divide and do not mature into EBs, but the developmental cycle can be reactivated [30, 31]. Also *C. trachomatis* can enter a persistent state [32] and in addition to cytokines, limited nutrient availability [33] and treatment with antibiotics that fail to eradicate the infection have been shown to trigger this state [34, 35].

## 1.2 *Chlamydia* genomes

### 1.2.1 Sequenced *Chlamydia* genomes

The first *Chlamydia* genome sequences of *C. trachomatis* [2] and *C. pneumoniae* [3] are from the *Chlamydia* Genome Project (CGP) (<http://chlamydia-www.berkeley.edu:4231/>). The sequenced genomes provide insight into genome organization and metabolic pathways of *Chlamydia* and form a basis for further research in gene regulation and protein expression [36]. Genome sequences of *C. muridarum* (previously *C. trachomatis* MoPn) [37], three other isolates of *C. pneumoniae* [37–39] and most recently that of *C. caviae* (previously *C. psittaci* GPIC) [40] have been published. An overview of the sequenced genomes is given in Tab. 1 where the number of predicted protein encoding open reading frames (ORFs) is the number given in the respective references. The number of ORFs is dependent on what sequence length is considered minimum for an expected protein and the cut-off varies slightly between sequencing projects.

*C. trachomatis* serovar D and *C. muridarum* contain a plasmid, and in *C. caviae* and *C. pneumoniae* AR39 a bacteriophage was found. The genomes of *C. trachomatis* D and *C. muridarum* (human and mouse genital pathogens, respectively), are very similar with an average of about 10% difference between orthologous genes [37]. Most differences between these genomes were found in the replication termination region (RTR) [40] including those in *C. trachomatis* D genes involved in tryptophan synthesis, which are missing in *C. muridarum*.

The *C. pneumoniae* genomes are more than 99.9% identical and the few differences are mainly found in *pmp* [37] and *ppp* genes [41, 42]. A double-stranded circular DNA, the replicative form of a bacteriophage was found upon sequencing the *C. pneumoniae* AR-39 genome [37]. The phage of *C. pneumoniae* AR-39 was suggested as contributing to pathogenicity [43], and a similar phage was identified in *C. abortus* [6].

**Tab. 1** Genome size and number of ORFs

Genome	Reference	Base pairs	ORFs	Plasmid/Phage
<i>C. trachomatis</i> D	[2]	1,042,519	894	7493 bp plasmid
<i>C. pneumoniae</i> , CWL029	[3]	1,230,230	1073	–
<i>C. pneumoniae</i> , AR39	[37]	1,229,853	1052	4524 bp phage
<i>C. pneumoniae</i> , J138	[38]	1,226,565	1072	–
<i>C. muridarum</i>	[37]	1,069,412	924	7501 bp plasmid
<i>C. caviae</i> (GPIC)	[40]	1,173,390	1009	7966 bp phage

*C. pneumoniae* TW-183 has also been sequenced [39] but is not contained in this table as no paper has yet been published on the results

## 1.2.2

**Chlamydial genes**

The environment of *Chlamydia* can be considered hostile, since the host cell will attempt to eradicate the bacteria, or friendly, since the bacteria have access to nutrients from the host cell. Analyzing the genome sequences of *Chlamydia* by comparing metabolic pathways and energy systems to those of free-living bacteria reveal many consequences of the availability of nutrients. However, the defense systems implicated by the intracellular nature do not appear as readily from the genome sequences since these may be unique for *Chlamydia*.

No genes are found that encode proteins involved in *de novo* purine and pyrimidine synthesis and the ability to synthesize amino acids is greatly reduced. Correspondingly, a large number of genes encoding different transport proteins have been identified, including many ABC transporters which are primarily involved in transport of smaller peptides and amino acids [2]. Also in good agreement with the intracellular and thus isolated nature of *Chlamydia*, no genes involved in DNA uptake were identified and no insertion sequences were found [2].

Chlamydiae have traditionally been described as energy parasites obtaining ATP from their host cells [19, 44], and the genomes of *C. trachomatis* and *C. pneumoniae* confirmed the presence of two genes CT065/Cpn0351 and CT495/Cpn0614 (CTXXX and CpnXXXX refer to *C. trachomatis* and *C. pneumoniae* gene numbers, introduced by the CGP [2, 3]), homologous to genes encoding ATP transporting proteins from *Rickettsia prowazekii* [45]. The orthologs from *C. trachomatis* L2 were cloned and used to express functional nucleoside phosphate transporters (npt) in *Escherichia coli*, one (CT065) exchanging ADP for ATP, the other (CT495) transporting all four ribonucleoside triphosphates [46]. Surprisingly, also genes encoding a wide range of ATPases as well as phosphoglycerate kinase, pyruvate kinase, and succinate thiokinase were identified, suggesting the capability of *Chlamydia* to produce ATP itself [2]. This ability may be important in the early and late stages of the developmental cycle where *Chlamydia* supposedly cannot obtain ATP from the host cell [44]. Genes encoding the proteins of an intact glycolytic pathway (although it is questionable whether an enzyme with fructose-1,6-diphosphate aldolase activity is present or this is circumvented), a partial TCA cycle, a complete glycogen synthesis and degradation system, and genes involved in aerobic respiration were also found [2]. Furthermore, proteins encoded by many of these genes were shown to be present in EBs [47] and pyruvate kinase, phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase and glucose-6-phosphate dehydrogenase were shown to complement *E. coli* mutants when expressed recombinantly [48]. During the intracellular stage, *Chlamydia* may store glycogen that is used to fuel the chlamydiae in the beginning and the end of the developmental cycle together with stored pools of ATP [48].

Four groups of chlamydial proteins have been indicated as especially interesting and important results of the genome project [49]. These groups were (i) peptidoglycan synthesis proteins; (ii) type III secretion proteins; (iii) inclusion membrane proteins (incls); and (iv) polymorphic membrane proteins (pmps).

The presence of a nearly full set of genes involved in peptidoglycan synthesis was unexpected since a peptidoglycan layer is not detected in EBs. However, *Chlamydia* is sensitive to beta-lactam antibiotics and peptidoglycan has been suggested to play a role in the division of RBs [50] supported by the finding of three amidases with probable peptidoglycan degrading activity.

The finding of type III secretion system genes was expected as such genes had earlier been found in *C. caviae* (*C. psittaci* GPIC) [51]. The type III secretion system is known from other Gram-negative bacteria to facilitate the transport of molecules from the bacterial cytosol into a future host cell by penetration of the host cell membrane with a surface protrusion that is thought to function as a channel. Surface projections of both EBs and RBs observed in electron microscopy [52, 53] thought to be involved in nutrient uptake, were speculated to be such type III needles when type III genes in *Chlamydia* were found [54, 55].

Inclusion membrane proteins are chlamydial proteins that are inserted into the inclusion membrane. Such proteins were first identified in *C. caviae* and termed incA, B and C [56, 57]. Homologs of these were found in the genomes of all sequenced chlamydiae but have not been found in any other organism. Several additional incs have since been identified and all of these share a characteristic bilobed hydrophobic region, even though no sequence motif is apparent [58]. Thirty-three genes encoding proteins with this hydrophobicity pattern have been identified in the *C. trachomatis* genome and 93 in the *C. pneumoniae* CWL029 genome [59].

Another group of *Chlamydia* specific proteins found in the genome was the family of polymorphic membrane proteins (pmps). These were initially identified in *C. abortus* (ovine abortion subtype of *C. psittaci*) being immunogenic proteins present in the outer membrane [60]. Nine *pmp* genes were identified in *C. trachomatis*, 17 in *C. caviae* and 21 in *C. pneumoniae*. The pmps are defined by being predicted outer membrane proteins containing repeated sequences of GGAI and FxxN [61] and by protein structure analysis they are predicted to be autotransporters [47, 62]. Incs and pmps are likely to be pivotal for *Chlamydia* biology indicated by the fact that 37.4% of the *Chlamydia* specific coding sequence of *C. pneumoniae* is constituted by *inc* and *pmp* genes (18.9% and 17.5%, respectively) [49].

### 1.2.3

#### Genome comparison

Genome sequences are thus available for *C. trachomatis* serovar D, *C. muridarum*, *C. caviae* and four isolates of *C. pneumoniae* (CWL029, AR39, J138 and TW-183), all of these share the unique developmental cycle but they are diverse in tissue tropism; *C. trachomatis* serovar D infects the genital tract of humans, *C. pneumoniae* infects the human respiratory tract; *C. caviae* the conjunctiva of guinea pigs and *C. muridarum* is a mouse pathogen. Hence, genome comparisons may reveal differences that are important for pathogenicity and tissue specificity.

Comparison of the *C. caviae* genome [40] to those of *C. pneumoniae* and *C. muridarum* showed that only 68/1009 *C. caviae* genes were not found in any of the other *Chlamydia* genomes, but differential expression of genes shared by the different

organisms may contribute to pathogenicity differences. Seven hundred and ninety-eight genes were found in all genomes and may be the minimum set of genes required for the basic growth and development of *Chlamydia*. Out of the 798 shared genes, 183 could not be found in any other of 70 published microbial genomes in the TIGR database [40]. Investigation of these genes, which include the *inc* and *pmp* genes, may elucidate functions that are specifically related to the intracellular characteristics of *Chlamydia* and its developmental cycle.

The most prominent *C. caviae* specific genes compared to *C. pneumoniae* are the genes required for tryptophan synthesis found in the RTR. *C. caviae* appears to be able to synthesize tryptophan from anthranilate, which is a very early precursor [40]. *C. trachomatis* possess a more limited set of tryptophan synthesis genes [2] and the genital and LGV serovars can produce tryptophan from the intermediate precursor indole, whereas the ocular serovars A and C have a truncated *TrpA* and serovar B lacks the *trpA* operon [63] similar to that which is found for *C. pneumoniae* [3]. A *tox* gene similar to cytotoxic genes from enterobacteria has been found in *C. caviae* and *C. muridarum*, the product of which may be secreted by the type III secretion system in order to inhibit actin polymerization [40]. In addition, a gene with homology to an invasin/intimin family protein was identified but the gene is interrupted by two frame shifts [40]. Specific genes found in *C. pneumoniae* that are absent from *C. caviae* include a uridine kinase, two 3-deoxy-D-manno-octulosonic acid (KDO) transferases, and two genes involved in biotin synthesis. In addition, 168 genes with unknown function are present in *C. pneumoniae* but not in any other *Chlamydia* [40].

Comparing *C. pneumoniae* to *C. trachomatis*, 80% of the predicted protein encoding genes have an ortholog in *C. trachomatis* [3]. From the 214 genes found in *C. pneumoniae* but not in *C. trachomatis*, most have no known function, but those that have include genes for purine and pyrimidine salvage pathways and completion of the biotin synthase pathway. A prominent difference is the expansion of the *pmp* gene family from nine members in *C. trachomatis* to 21 members in *C. pneumoniae* [61]. The *C. trachomatis* *pmp* genes are located in two clusters *pmpA-C* and *pmpE-H* except for one gene, *pmpD*. Most of the difference between *C. pneumoniae* and *C. trachomatis* is accounted for by expansion of *pmpG* to 13 *pmps* (*pmp1-13*) in *C. pneumoniae* [61]. The amino acid identity between *pmp1-13* is 34–55%.

The *C. pneumoniae* genomes elucidated that several *pmp* genes contain frame shifts, and these vary between isolates, as listed in Tab. 2. Furthermore, at least *pmp10* was shown to be differentially expressed between chlamydiae within the same cell, and this is likely due to a polyG tract that varies in length [64]. Based on the relatively high variability in the *pmp* gene family, considering the otherwise very conserved sequences between isolates, it has been speculated that the *pmps* may function in surface variation of *Chlamydia* as seen in other pathogenic bacteria [65].

Another gene family in *C. pneumoniae* that shows remarkable variation is the recently identified *Cpn1054* family or *C. pneumoniae* polymorphic protein (ppp) family [42]. *Cpn1054* was initially identified as one of eleven paralogous genes located in four hyper-variable regions in *C. pneumoniae* CWL029, one of which is situated between *pmp1* and *pmp2* [66]. The genes were predicted to encode *inc*

**Tab. 2** Variation in *C. pneumoniae* polymorphic membrane protein (*pmp*) genes

<i>pmp</i>	CWL029	AR39	J138
2			frame shift
3	frame shift	frame shift	frame shift
4	frame shift	frame shift	+1 frame shift
5	frame shift	frame shift	frame shift
6		393 bp del.	393 bp del.
10	frame shift		
12	truncation	truncation	truncation
17	frame shift	frame shift	frame shift

proteins by the presence of the characteristic bilobed hydrophobic motif. Many of the genes contain stop mutations that differ between sequenced strains and as in *pmp10*, a poly-G tract was identified in the 5' end of *cpn1054* [66]. Recently, poly-G tracts present in seven of eleven 1054 family members were analyzed by sequencing of a number of clinical isolates [67]. Five out of seven were found to vary in all investigated isolates, and functional analysis of protein products from this gene family will be interesting.

### 1.3

#### Proteome analysis of *Chlamydia*

The genome sequence reveals the coding capacity of an organism and thus what proteins it theoretically can produce. The coding capacity is informative, but does not reveal information about when, where and in what quantities the genes are transcribed and whether the possibly resultant proteins are modified or secreted. The direct investigation of proteins in their post-translationally modified and processed form present in a given biological compartment at a specific time and in a defined environment is the task of proteomics.

Proteomics is used to describe any large-scale investigation of proteins and can be approached in many ways but in principle it involves two steps: separation of the proteins in a sample and subsequent identification of these proteins. The perfect proteome study would provide a quantitative measure of every single protein present in the investigated sample. Unfortunately, such a study is so far not possible. Novel quantitative mass spectrometric techniques come close, but these are still in the development phase. Today 2-D gels as a separation tool coupled to mass spectrometry protein identification provides the most comprehensive way of analyzing complex protein mixtures [68].

## 1.3.1

**Early *Chlamydia* proteome studies**

In 1985, 2-DE was used to compare the protein content of outer membrane preparations from *C. trachomatis* serovars L2 and F [69]. Chlamydiae were selectively radiolabeled by [<sup>35</sup>S]methionine incorporation in the presence of the inhibitor of eukaryotic ribosomes, cycloheximide. EBs were purified and *Chlamydia* outer membrane complex (COMC) was prepared by sarkosyl extraction [70]. The COMC was solubilized in a 2-D buffer based on urea with NP-40 as detergent and mercaptoethanol as reducing agent and subjected to 2-D PAGE where the first dimension was carried out in tube gels in which the pH gradient was established during focusing. Three proteins, major outer membrane protein (MOMP), a 60 kDa protein and a 12 kDa protein were observed for *C. trachomatis* F, whereas the 60 kDa protein was missing for *C. trachomatis* L2. However, by NEPHGE it could be concluded that the 60 kDa protein was also present in *C. trachomatis* L2, but migrating more basic than in serovar F [69].

Improvements in 2-D PAGE, including IPG strips for the first dimension, means that today more proteins can be resolved in 2-D gels of COMC [71]. Lambden *et al.* [72] identified the 60 kDa large cysteine-rich outer membrane protein, OmcB (Omp2), and the 12 kDa small cysteine-rich protein, OmcA (Omp3) to be developmentally regulated and transcribed as a polycistronic mRNA late in the developmental cycle. A model of the COMC architecture has been proposed [73] in which Omp2 is localized in the periplasmic space, disulfide cross-linked to Omp3, which is suggested to be anchored in the outer membrane by its lipid moiety. Comparing COMC from different *C. trachomatis* serovar to *C. pneumoniae* and *C. caviae* [74] showed that Omp2 from *C. trachomatis* L2 was resolved in the gels, but migrated one pH unit more basic than Omp2 of *C. trachomatis* F and two pH units more basic than *C. trachomatis* D. No additional proteins were identified for any of the species even though high molecular bands were observed by 1-D SDS gels for *C. trachomatis* serovar D [74].

## 1.3.2

***C. trachomatis* proteome studies**

The first proteome study on whole *Chlamydia* aimed at identifying early proteins in *C. trachomatis* L2 by pulse labeling with [<sup>35</sup>S]methionine at 2–4 h post-infection (hpi), 8–10 hpi, 14–16 hpi and 28–30 hpi [75]. Seven proteins were detected earlier than MOMP, four of which were labeled at 2–4 hpi. Three of these were identified by colocalization with proteins detected by immunoblotting with known antibodies. These were the heat shock proteins DnaK and GroEL and the ribosomal protein S1. The remaining four proteins were not identified. Early transcription of the *groEL* gene has recently been confirmed by transcript analysis [76, 77], but *dnaK* was designated a late gene in [78]. However, the designation “late” was based on lower transcription at earlier points in time than 24 hpi and higher transcription at later points in time and this does not exclude early transcription.

A second global study [79] aimed at providing a basis for the development of a protein database of *C. trachomatis* proteins. This was the first *Chlamydia* study to use IPG strips. Approximately 600 spots were separated in the area from pH 4–9 and 10–120 kDa in silver-stained gels. The very good resolution compared to earlier studies can be ascribed to the use of IPGs but also the substitution of mercaptoethanol with dithioerythritol (DTE), and NP-40 with CHAPS may have contributed to the superior results. Mercaptoethanol will more readily migrate out of the first dimension gel than DTE due to its charge, and removal of reducing agent will cause reoxidation and precipitation of proteins. A combination of immunoblotting with known antibodies and *N*-terminal sequencing was used to identify nine known proteins [79]. Seven sequences were obtained from yet uncharacterized proteins distributed in different areas of a 2-DE map and even though the gels showed very good resolution, the study like all other pregenomic proteome studies, suffered from the lack of identification methods for unknown proteins.

In the pregenomic area, 2-DE was most appropriate in studies where antibodies were available for identification of the proteins. One such study demonstrated the superiority of 2-DE in comparison to 1-DE with respect to the resolution of different isoelectric isoforms [80]. A family of high molecular weight *C. abortus* proteins detected by post-abortion sera from sheep were shown to be identical to immunogenic putative outer membrane proteins (POMPs). As the proteins had similar molecular weight, they could not have been distinguished in 1-D gels.

Western blotting of 2-D gels has also been applied to identify immunogenic proteins in *C. trachomatis* using sera from 17 patients suffering from genital inflammatory disease [81]. Fifty-five immunogenic proteins were detected with frequencies varying from 17 to 1. Eight proteins could be identified by colocalization with previously determined proteins. In addition, *N*-terminal sequences were obtained for nine proteins from which six could be identified in the genome sequence. Omp2, GroEL, MOMP and DnaK were the most frequently recognized proteins. These are known antigens, but also previously unknown antigens were detected such as elongation factor TU and ribosomal proteins.

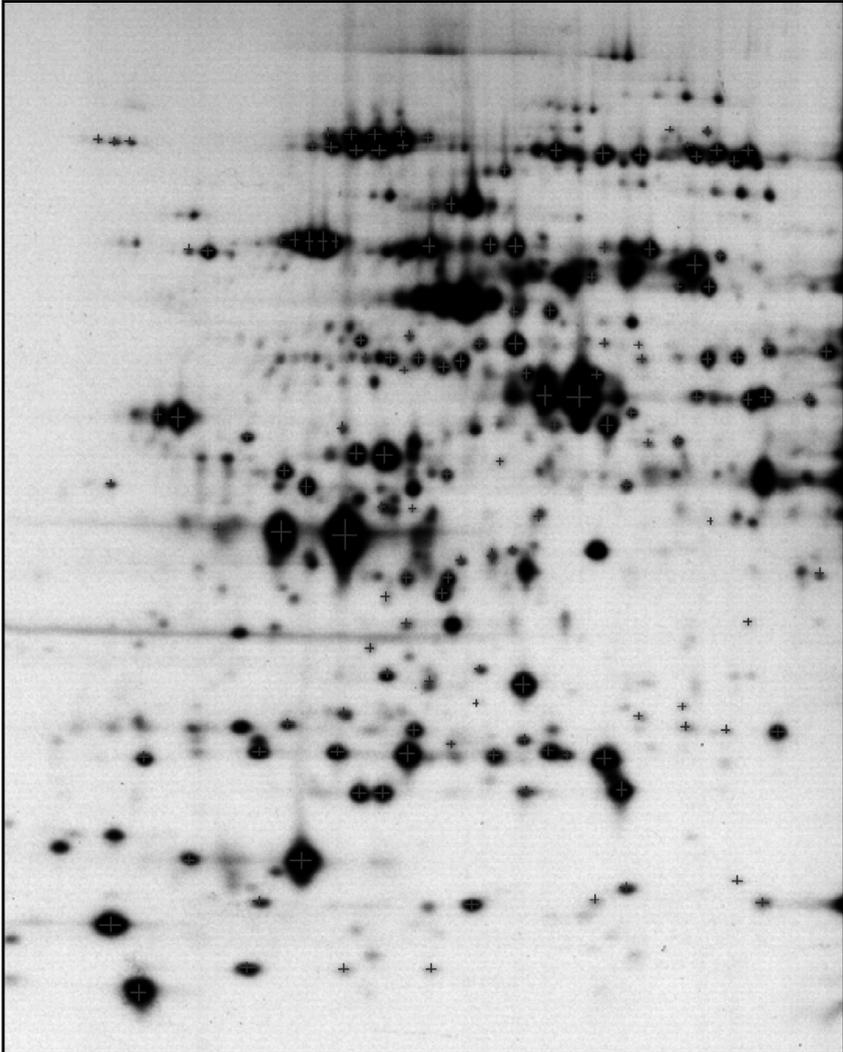
### 1.3.3

#### ***C. pneumoniae* proteome studies**

The first comprehensive proteome map of *Chlamydia* in the postgenomic area was that of *C. pneumoniae* [47] (Fig. 2). Like Bini *et al.* [79] this study used IPGs in the first dimension and thiourea was incorporated into the 2-DE buffer to obtain the best possible recovery of hydrophobic proteins. Mass spectrometry was used to identify 263 protein spots representing 167 different genes and all identifications were published on the internet at <http://www.gram.au.dk> in a searchable form. Data for pH 4–7 (Fig. 2) was also included in the bacterial proteome database at the Max Planck Institute for Infection Biology at <http://www.mpiib-berlin.mpg.de/2D-PAGE/>. The proteome map can thus serve as a reference for 2-D PAGE studies performed in other laboratories. A good agreement between predicted and observed number of proteins was observed in the acidic region, whereas recovery in

**Features of the Overview:**

1. cursor over cross: protein name appears for identified spots
2. mouse click on sector: zoom into this sector
3. mouse click on cross: show protein information and hyperlinks



**Fig. 2** Screen dump of the clickable IPC4-7 proteome map of *C. pneumoniae* available at [www.mpiib-berlin.mpg.de/2D-PAGE/EBP-PAGE/index.html](http://www.mpiib-berlin.mpg.de/2D-PAGE/EBP-PAGE/index.html). Crosses represent identified proteins.

the basic region was poor. The use of basic strips (pH 6–11) did not significantly improve the number of resolved protein spots, but gave a better spatial distribution which is important when proteins are to be excised for further analysis.