

DISS. ETH Nr. 18904

**GENERAL PRINCIPLES OF CELLULAR ORGANIZATION
IN THE GENOME-REDUCED BACTERIUM
*MYCOPLASMA PNEUMONIAE***

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH
for the degree

Doctor of Sciences

presented by

SEBASTIAN KÜHNER

Dipl. Biol., Ruprecht-Karls-University Heidelberg
born July 31, 1981
German citizen

Accepted on the recommendation of:
Prof. Dr. Ruedi Aebersold, examiner
Dr. Anne-Claude Gavin, co-examiner
Prof. Dr. Uwe Sauer, co-examiner

2010

Zusammenfassung

Eines der Ziele naturwissenschaftlicher Forschung ist es, eine selbstreplizierende Zelle in ihrer Gesamtheit zu verstehen. Kennt man die organisatorischen Prinzipien, denen grundlegende Bestandteile der Zelle, wie Transkriptom, Proteom und Metabolom unterliegen, so ist man diesem Ziel einen bedeutenden Schritt näher. In den vergangenen Jahren haben enorme technische Fortschritte bei analytischen Hochdurchsatzverfahren die systematische Untersuchung dieser einzelnen „Ome“ ermöglicht. Dies wiederum gibt uns heute die Gelegenheit, in einem Organismus parallel das Transkriptom und seine Komplexität, die Proteomorganisation in Proteinkomplexe sowie den Metabolismus und dessen Regulation global zu analysieren. Die vorliegende Doktorarbeit beschreibt die systematische Charakterisierung dieser zellulären Bestandteile in dem genomreduzierten Bakterium *Mycoplasma pneumoniae*, das zu den kleinsten sich selbstständig fortpflanzenden Lebensformen zählt. Der Fokus dieser Arbeit liegt auf der Analyse der Proteomorganisation, der Studie zu der ich, im Vergleich aller drei Teilstudien, den mit Abstand größten Teil beigetragen habe.

Zur Untersuchung der grundlegenden Prinzipien bakterieller Proteomorganisation führten wir eine proteomweite, auf chromatographischer Aufreinigung sowie massenspektrometrischer Analyse von Proteinkomplexen basierende Studie durch. Die Datenanalyse identifizierte 62 homopolymere und 116 heteropolymere, lösliche Proteinkomplexe, von denen die Mehrzahl bisher uncharakterisiert ist. Etwa ein Drittel der heteropolymeren Komplexe ist an der Bildung komplexerer Formen der Proteomorganisation beteiligt. Dazu zählt die Anordnung in größere Multiproteinkomplexeinheiten, welche zum Teil die sequentiellen Schritte eines biologischen Prozesses reflektieren. Des Weiteren beobachten wir, dass ein und dasselbe Protein Bestandteil mehrerer verschiedener Komplexe sein kann, was wir als Proteinmultifunktionalität interpretieren. Dies könnte eine mögliche Erklärung für die Genomreduzierung von *M. pneumoniae* darstellen. Die Kombination der Proteinkomplexdaten mit 484 Proteinstrukturen, elektronenmikroskopischen und elektronentomographischen Daten liefert zusätzliche strukturelle Details der Proteomorganisation dieses Bakteriums.

Um die grundlegenden Prinzipien des bakteriellen Metabolismus und seiner Regulation zu verstehen, wurde ein metabolisches Netzwerk aus 129 Enzymen, die insgesamt 189 verschiedene Reaktionen katalysieren, konstruiert. Die Nutzung dieses biochemischen Atlas ermöglichte es, ein Minimalmedium, bestehend aus 19 essentiellen Bestandteilen, zu entwickeln. Das metabolische Netzwerk von *M. pneumoniae* hat im Vergleich zu

komplexeren Bakterien eine stärker lineare Topologie und enthält einen höheren Anteil an multifunktionalen Enzymen. Allgemeine Eigenschaften wie die Konzentration von Metaboliten, der zelluläre Energiehaushalt, die Anpassungsfähigkeit und die globale Genexpression sind denen anderer Bakterien ähnlich.

Für die Transkriptionsanalyse wurde strangspezifische Mikrochiptechnologie „Tilingarrays“, ergänzt durch die direkte Sequenzierung der Transkripte, mit mehr als 252 Spottedarrays kombiniert. Es wurden 117 neue, meist nicht-codierende Transkripte, von denen 89 in antisense Konfiguration zu bekannten Genen liegen, detektiert. Weiterhin wurden 202 monocistronische sowie 139 polycistronische Operons identifiziert, wobei fast die Hälfte der letzteren ein stufenartiges Transkriptionsprofil aufweist. Unter variierenden Wachstumsbedingungen können sich die Operons in bis zu 447 kleinere transkriptionelle Einheiten aufteilen, was die Mannigfaltigkeit der alternativen Transkripte erklärt. Häufige antisense und alternative Transkripte sowie unterschiedliche Regulationsmechanismen verdeutlichen die hohe Dynamik dieses Transkriptom.

Die Informationen in dieser Doktorarbeit können als kleiner Teil eines Bauplans für eine minimale zelluläre Maschine angesehen werden. Die vier allgemeinen Schlußfolgerungen aus dieser Arbeit sind:

1. *M. pneumoniae* weist besonders in seiner Proteomorganisation und in seinem Metabolismus Aspekte der Multifunktionalität auf.
2. *M. pneumoniae* ist für einen nahezu minimalen Organismus, bezüglich seiner regulatorischen Prozesse, deutlich komplexer als erwartet.
3. *M. pneumoniae* ähnelt Eukaryoten, hinsichtlich Komplexität seines Transkriptoms, wesentlich stärker als erwartet.
4. *M. pneumoniae* ist auf Wirtsanpassung und nicht auf maximales Wachstum optimiert.

Summary

One aim of research in natural sciences is to understand a self-reproducing cell in its entity. Transcriptome, proteome and metabolome are at the heart of cellular organization. To know the general principles underlying these basic functional aspects of a cell is a key part on the way to comprehensively understand cellular life. In recent years systematic analysis of the individual “omes” has been facilitated via dramatically advancing analytical high-throughput techniques. This opened opportunities to globally investigate the transcription and its complexity, the proteome organization into protein complexes as well as the metabolism and its regulation of an entire organism. In the presented PhD thesis I describe an effort to systematically characterize these features for the genome-reduced organism *Mycoplasma pneumoniae*, which is among the smallest self-replicating forms of life on earth. My work hereby focuses on the analysis of the proteome organization, the part, among the three investigations, to which I contributed most substantially.

In order to study basic principles of bacterial proteome organization, we used tandem affinity purification followed by mass spectrometric analysis to investigate protein complexes in a proteome-wide screen. The analysis revealed 62 homomultimeric and 116 heteromultimeric soluble protein complexes, of which the majority are novel. About a third of the heteromultimeric complexes show higher levels of proteome organization, including assembly into larger, multi-protein complex entities, suggesting sequential steps in biological processes, and extensive sharing of components implying protein multifunctionality. Incorporation of structural models for 484 proteins, single particle EM and cellular electron tomograms provided supporting structural details for this proteome organization.

To understand basic principles of bacterial metabolism and its regulation, a manually curated metabolic network of 129 enzymes catalyzing 189 reactions was constructed. This metabolic map allowed the design of a defined, minimal medium with 19 essential nutrients. In summary, the *M. pneumoniae* metabolic network has a more linear topology and contains a higher fraction of multifunctional enzymes compared to more complex bacteria; general features such as metabolite concentrations, cellular energetics, adaptability and global gene expression responses are similar though.

For the transcriptome analysis we combined strand-specific tiling arrays, complemented by transcriptome sequencing, with more than 252 spotted arrays. We detected 117 previously undescribed, mostly non-coding transcripts, 89 of them in antisense configuration to known genes. We identified 202 monocistronic and 139 polycistronic operons; almost half of the latter show decaying expression in a staircase-like manner. Under various conditions, operons

could be divided into 447 smaller transcriptional units, resulting in many alternative transcripts. Frequent antisense transcripts, alternative transcripts, and multiple regulators per gene imply a highly dynamic transcriptome.

In essence this thesis dataset provides a blueprint of the minimal cellular machinery required for life. The four general conclusions of this work are that:

1. *M. pneumoniae* shows aspects of multifunctionality particularly in its proteome organization and metabolism.
2. *M. pneumoniae*, even though being an almost minimal organism, is more complex than expected, particularly in its regulation.
3. *M. pneumoniae* resembles eukaryotes more than expected, particularly in its transcription.
4. *M. pneumoniae* is optimized for host adaptation and not for growth.