# Physiologische und proteomische Untersuchungen am marinen Planktomyceten *Rhodopirellula baltica*

Dissertation
zur
Erlangung des Grades eines
Doktors der Naturwissenschaften
- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der
Universität Bremen vorgelegt von

Dörte Gade

Bremen 2004

Die Untersuchungen zur vorliegenden Doktorarbeit wurden in der Zeit von Juni 2001 bis Juni 2004 am Max-Planck-Institut für Marine Mikrobiologie in Bremen durchgeführt.

1. Gutachter: Prof. Dr. Friedrich Widdel
2. Gutachter: Priv.-Doz. Dr. Ralf Rabus

Tag des Promotionskolloqiums: 2. Juli 2004

# Inhaltsverzeichnis

## Teil II: Publikationen

# Abkürzungen

| | |
|---|---|
| 2DE | Zweidimensionale Gelelektrophorese |
| cDNA | copy Desoxyribonukleinsäure |
| DIGE | Difference Gel Electrophoresis |
| ESI | Elektrospray-Ionisation |
| ICM | Intracytoplasmatische Membran |
| IEF | Isoelektrische Fokussierung |
| IEP | Isoelektrischer Punkt |
| MALDI | Matrix-Assisted Laser Desorption/Ionisation |
| max. | maximal |
| Mb | Megabasen |
| MS | Massenspektrometrie |
| NAG | *N*-Acetylglucosamin |
| $OD_{600}$ | Optische Dichte gemessen bei 600 nm |
| ORF | Open Reading Frame (offener Leserahmen) |
| PAGE | Polyacrylamid-Gelelektrophorese |
| PHX | Predicted Highly Expressed (als hochexprimiert vorhergesagt) |
| PMF | Peptidmassen-Fingerabdruck |
| PSD | Post Source Decay |
| RB | *Rhodopirellula baltica* (Nomenklatur beim Genom) |
| S | Svedberg Einheit |
| SDS | Sodiumdodecylsulfat |
| s(p)p. | Spezies |
| TCA | Tricarbonsäure |
| TOF | Time of Flight |

# Zusammenfassung

*Rhodopirellula baltica* (ehemals *Pirellula* sp. Stamm 1) ist dem bislang wenig untersuchten bakteriellen Phylum *Planctomycetes* zugeordnet. Bekannte Mitglieder dieser Gruppe wurden vor allem aufgrund ihrer morphologischen Besonderheiten erkannt und isoliert. Hierzu zählen beispielsweise die intrazelluläre Kompartimentierung und die Knospenbildung zur Zellteilung. Diese für Bakterien seltene Art der Fortpflanzung führt zu einem Lebenszyklus, der sich durch den Wechsel zwischen sessilen Adultzellen (Rosetten) und mobilen Schwärmerzellen auszeichnet. Planktomyceten sind häufige Vertreter der mikrobiellen Lebensgemeinschaften in der marinen Wassersäule. Zudem sind sie auf die aerobe Verwertung von Kohlenhydraten spezialisiert. Es ist daher anzunehmen, dass sie eine wichtige Rolle bei der Remineralisierung von Kohlenhydraten im marinen Habitat spielen.

Phylogenetisch signifikante Unterschiede der 16S rDNA-Sequenzen bekannter *Pirellula* spp. (*P. staleyi*, *P. marina* und *Pirellula* sp. Stamm 1) erforderten die Beschreibung zweier neuer Gattungen (*Rhodopirellula* und *Blastopirellula*). Im Zuge dieser taxonomischen Neuordnung wurde *Pirellula* sp. Stamm 1 als Typusstamm der neuen Spezies *Rhodopirellula baltica* beschrieben und *P. marina* wurde als *Blastopirellula marina* in die zweite neue Gattung überführt. Klassische Wachstumsexperimente belegten die Spezialisierung von *P. staleyi*, *B. marina* und *R. baltica* auf die Verwertung von Kohlenhydraten als Energie- und Kohlenstoffquelle sowie ihr strikt aerobes Wachstumsverhalten.

Den Schwerpunkt der vorliegenden Arbeit bildeten proteomische Untersuchungen physiologischer Eigenschaften von *R. baltica*. Dabei diente die zweidimensionale Gelelektrophorese als zentrale Technik.

Zunächst wurde (**1**) eine aktuelle Methode zur Bestimmung von Unterschieden in Proteinhäufigkeiten etabliert und evaluiert. Auf der Basis (**2**) eines Mastergels und (**3**) substratadaptierter Zellen wurden die zentralen Wege des Kohlenhydratabbaus rekonstruiert sowie Enzymkandidaten für die hinführenden substratspezifischen Wege identifiziert. (**4**) Schließlich wurde Wachstumsphasen- bzw. Lebenszyklus-spezifische Proteinbildung untersucht. Das Vorliegen der vollständigen Genomsequenz von *R. baltica* ermöglichte hierbei eine Proteinidentifizierung durch Abgleich der mithilfe der Massenspektrometrie erhalten Peptidmassen mit denen, die durch *in silico* Verdau der Proteinsequenz aus der Datenbank generiert wurden.

1 Physiologische Anpassungen spiegeln sich in Veränderungen der Proteinzusammensetzung wider, welche mit Hilfe der Difference Gel Electrophoresis (DIGE)-Technologie quantitativ erfasst werden können. Präelektrophoretische, kovalente Markierung von Proteinen mit Fluoreszenzfarbstoffen (CyDyes™), Koseparierung und Verwendung eines internen Standards erlauben eine Quantifizierung mit hoher statistischer Sicherheit. Die hohe Sensitivität der dabei verwendeten Cy-Farbstoffe wurde durch den Vergleich mit konventionellen Farbstoffen (Coomassie und Silber) und dem Fluoreszenzfarbstoff SYPRO® Ruby nachgewiesen. Die DIGE-Technologie und das auf SYPRO® Ruby basierende System lieferten im gewählten Testansatz vergleichbare Abundanzunterschiede der untersuchten Proteinspots.

2 Es wurde ein Mastergel erstellt, auf dem 626 lösliche Proteine von *R. baltica* identifiziert und annotiert wurden, welche in einem pH-Bereich von 4 bis 7 separiert wurden. Ein Großteil dieser Proteine wurde mehrfach identifiziert. Dieses Mastergel wurde in Bezug auf die folgenden Punkte untersucht:

a) Verteilung der Proteine über den pH-Gradienten im Vergleich zum theoretischen (wie bzw. woraus vorhergesagt) Mastergel;
b) Nachweis der Expression vorhergesagter Proteine, insbesondere von hypothetischen Proteinen;
c) Vorkommen von Signalpeptid-tragenden Proteinen;
d) Vergleich von tatsächlichen Proteinabundanzen und theoretisch vorhergesagtem Expressionsniveau der codierenden Gene;
e) Identifizierung bislang nicht vorhergesagter Gene.

Darüber hinaus diente das Mastergel als Grundlage für die nachfolgenden funktionalen Untersuchungen.

3 Die Enzyme der Glykolyse, des oxidativen Pentosephosphat-Zyklus und des TCA-Zyklus konnten bis auf wenige Ausnahmen auf dem Mastergel nachgewiesen werden. Es gab keinerlei Hinweise auf den Entner-Doudoroff-Weg. Die meisten Enzyme der rekonstruierten Abbauwege bilden die größten Spots auf dem 2D Gel und gehören somit zu den häufigsten Proteinen in der Zelle.

Die Untersuchungen zu den peripheren Zuckerabbauwegen zeigten, dass die substratspezifisch induzierten Gene von *R. baltica* in zwei Kategorien eingeteilt werden können: zum einen Gene, die clustern und deren genetische Nachbarschaft eine Funktion im Abbau des untersuchten Substrates nahe legt (z. B. bei Ribose und *N*-Acetyl-glucosamin) und zum anderen Gene, welche scheinbar wahllos über das gesamte Genom

verteilt sind (z. B. bei Xylose) und deren Genkontext keinen Hinweis auf eine mögliche Funktion im Zuckerstoffwechsel gibt. Gene dieser zweiten Gruppe können jedoch auf regulatorischer Ebene miteinander verbunden sein. Im Falle von Raffinose und Melibiose konnten keine spezifischen Abbauenzyme bioinformatisch vorhergesagt werden. Beim Wachstum auf diesen beiden Zuckern wurden jedoch mehrere hypothetische Proteine gebildet. Möglicherweise handelt es sich dabei um Enzyme neuer peripherer Abbauwege für die genannten Kohlenhydrate.

4 In Wachstumskurven von *R. baltica* wurde eine lange stationäre Phase beobachtet. Sie deutet auf die Fähigkeit von *R. baltica* hin, Stress infolge von Substratbegrenzung oder hoher Zelldichte lange Zeit überdauern zu können. Entsprechend zeigten die proteomischen Analysen, dass Enzyme des Zentralstoffwechsels (z. B. Aconitase) in der späten stationären Phase herunterreguliert waren, während Enzyme der Stressantwort (z. B. Katalase) hochreguliert waren.

Beim Übergang in die stationäre Phase der Flüssigkultur bzw. beim Wachstum auf Agarplatte wurden außerdem diverse hypothetische Proteine vermehrt gebildet. Da unter diesen Wachstumsbedingungen der sessile Morphotyp vorherrscht, könnten diese spezifisch regulierten Proteine an der morphologischen Differenzierung beteiligt sein.

# Summary

*Rhodopirellula baltica* (former *Pirellula* sp. Strain 1) is affiliated with the poorly understood bacterial phylum *Planctomycetes*. Members of this group were mainly recognized and isolated due to their special morphological properties, such as the intracellular compartmentalization and the reproduction via budding. This mode of reproduction is uncommon among bacteria. In case of *R. baltica* it is characterized by an alternation between sessile adult cells (rosettes) and mobile swarmer cells. Planctomycets are common representatives of the microbial community in the marine water column. Additionally, they are specialized in the aerobic degradation of carbohydrates. Thus one can assume that they might play an important role in the remineralization of carbohydrates in marine habitats.

Phylogenetically significant differences in 16S rDNA sequences of known *Pirellula* spp. (*P. staleyi*, *P. marina* and *Pirellula* sp. strain 1) required the establishment of two new genera (*Rhodopirellula* and *Blastopirellula*). According to this taxonomic reorganization *Pirellula* sp. strain 1 was described as the type strain of the new species *Rhodopirellula baltica* and *P. marina* was transferred as *Blastopirellula marina* to the second new genus. Classical growth experiments demonstrated the specialization of *P. staleyi*, *B. marina* and *R. baltica* in utilization of carbohydrates as well as their strict aerobic growth behaviour.

The emphasis of this work lies on proteomic studies of the physiological properties of *R. baltica*. Here two-dimensional gel electrophoresis was used as a central technique.

To begin with (**1**) a recent method for determination of differences in protein abundances was established and evaluated. On the basis of (**2**) a master gel and (**3**) substrate adapted cells the central routes of carbohydrate degradation were reconstructed and candidates for enzymes of peripheral substrate specific pathways were identified. (**4**) Finally, growth phase and life cycle specific protein formation was investigated. The availability of the complete genome of *R. baltica* allowed protein identification via peptide mass fingerprinting.

**1** Changes in protein composition which reflect physiological adaptation can be quantified by the DIGE technology. Pre-electrophoretic covalent labeling of proteins with fluorescent dyes (CyDyes™), co-separation and the application of an internal standard allow quantification with high statistical confidence. The high sensitivity of the CyDyes™ was proven by a comparison with conventional stains (Coomassie and silver) and with the

fluorescent dye SYPRO® Ruby. In the applied test system the DIGE technology and the SYPRO® Ruby based system yielded comparable differences in abundance.

**2** A master gel was established containing 626 soluble proteins of *R. baltica*, which were separated within a pH range from 4 to 7. The majority of these proteins were identified several times. This master gel was examined according to the following points:

a) distribution of proteins along the pH gradient in comparison to the theoretical mastergel
b) evidence for the expression of predicted proteins, in particular of hypothetical proteins
c) occurrence of proteins carrying a signal peptide
d) comparison of real protein abundances with theoretical predicted expression levels of the coding genes
e) identification of unpredicted genes

Moreover, the mastergel served as a basis for the following functional studies.

**3** With a few exceptions all enzymes of glycolysis, oxidative pentose phosphate cycle and the TCA cycle could be identified on the master gel. No indications for the Entner-Doudoroff pathway were found. The majority of enzymes of the reconstructed metabolic pathways formed the largest spots on the 2D gel and thus belong to the most abundant proteins in the cell.

Studies on the peripheral sugar degradation pathways indicated that substrate specific induced genes can be divided in two categories: on the one hand genes that cluster and the genetic environment of which suggests a function in the degradation of the tested substrate (*e.g.* ribose and *N*-acetylglucosamine), and on the other hand genes that seem to be randomly distributed across the whole genome (*e.g.* xylose) and whose gene context gives no hint on a possible function in sugar metabolism. Genes of the second group may be linked on a regulatory level. In the case of raffinose and melibiose no specific enzymes for the degradation of these substrates were predicted on the bioinformatics level. During growth with these two sugars several hypothetical proteins were synthezised. These proteins may represent enzymes of new peripheral degradation pathways for these two carbohydrates.

**4** In growth curves of *R. baltica* a long stationary phase was observed. This finding points to the ability of *R. baltica* to survive stress resulting from substrate limitation and high cell densities for a long period of time. Correspondingly, the proteomic analysis revealed, that enzymes of the central metabolism (*e.g.* aconitase) were down-regulated in

the late stationary phase, whereas enzymes affiliated with stress response (*e.g.* catalase) were up-regulated.

Moreover, at the transition to stationary phase in liquid culture and during growth on agar plates, respectively, a number of hypothetical proteins were present at elevated amounts. Since the sessile morphotype predominates under these growth conditions, one may speculate that these specifically regulated proteins are involved in morphological differentiation.

# Teil I: Darstellung der Ergebnisse im Gesamtzusammenhang

## A Einleitung

Vor ca. zehn Jahren wurde das erste Genom eines Bakteriums (*Haemophilus influenzae*) vollständig sequenziert (Fleischmann et al. 1995). Seither haben sich die technischen und bioinformatischen Methoden der Genomforschung so rasant entwickelt, dass mittlerweile die Genomsequenzen von mehr als 160 Prokaryoten vorliegen (Stand Mai 2004; http://www.genomesonline.org). Obgleich die annotierte Genomsequenz gewissermaßen den Bauplan des Lebens für den jeweiligen Organismus darstellt, bedarf es der experimentellen Herangehensweise, um vorhergesagte und noch unbekannte Funktionen zu untersuchen und regulatorische Zusammenhänge aufzuzeigen. Dazu stehen in der postgenomischen Ära Methoden zur globalen Expressionsanalyse (Transkriptom-/ Proteomanalysen) zur Verfügung.

Die Mehrzahl aller bisherigen Genomprojekte beschäftigt sich mit biotechnologisch oder medizinisch relevanten Bakterien. In den vergangenen zehn bis 20 Jahren wurde die Bedeutung von Umweltbakterien für die globalen Stoffkreisläufe zunehmend erkannt. Entsprechend zog auch diese Gruppe von Bakterien in jüngster Zeit das Interesse der Genomforscher auf sich. Das REGX (Real Environmental GenomiX) Projekt, in das diese Arbeit eingebettet war, ist eines der ersten Umweltgenomprojekte. Ziel dieses Projektes war die Sequenzierung und funktionelle Genomanalyse von drei marinen Umweltbakterien: *Rhodopirellula baltica*, *Desulfobacterium autotrophicum* und *Desulfotalea psychrophila*. Das in dieser Arbeit untersuchte Bakterium *Rhodopirellula baltica* gehört zu dem bisher wenig untersuchten Phylum *Planctomycetes*, deren Vertreter sich durch eine Vielzahl morphologischer Besonderheiten auszeichnen. Hierzu gehören unter anderem die intrazelluläre Kompartimentierung, die Fortpflanzung durch Knospung und der Lebenszyklus, bei dem es zu einem Wechsel zwischen sessilen Adultzellen und mobilen Schwärmerzellen kommt. Planktomyceten gehören zu den häufigsten Mitgliedern der mikrobiellen Lebensgemeinschaften in der marinen Wassersäule (Llobet-Brossa et al. 1998; Glöckner et al. 1999).

In der vorliegenden Arbeit wurden zunächst im Rahmen der taxonomischen Beschreibung von *R. baltica* umfangreiche Wachstumsexperimente durchgeführt, um das physiologische Potential von *R. baltica* im Vergleich zu *Pirellula staleyi* und *Blastopirellula marina* zu erfassen. Dabei wurde die Spezialisierung dieser Bakteriengruppe auf den aeroben Kohlenhydratabbau deutlich.

Den Schwerpunkt der Arbeit bildeten detaillierte proteomische Untersuchungen auf der Basis definierter physiologischer Experimente. Damit wurden erstmals molekulare Einblicke in die Anpassung von *R. baltica* an wechselnde Umweltbedingungen (verändertes Substratangebot, Substratlimitierung) sowie in den Wachstumsphasen und während des Lebenszyklus gewonnen.

# 1. Allgemeine Eigenschaften der Planktomyceten

Bereits 1924 entdeckte N. Gimesi in einem Teich in Budapest einen koloniebildenden Mikroorganismus, den er als *Planctomyces bekefii* beschrieb. Der Umstand, dass er diesen planktischen Mikroorganismus den Pilzen zuordnete, spiegelt sich noch heute im Wortstamm „myces“ wider. Die Ordnung *Planctomycetales* beinhaltet eine in vielerlei Hinsicht ungewöhnliche Gruppe Gram-negativer knospender Bakterien. Bisher ist relativ wenig über Physiologie, Ökologie und Zellbiologie der Angehörigen dieser Bakteriengruppe bekannt. Nachdem die Planktomyceten jahrelang als eher exotische Organismengruppe angesehen wurden, haben in den letzten 20 Jahren molekularökologische Untersuchungen gezeigt, dass Planktomyceten in terrestrischen und aquatischen Habitaten ubiquitär vorkommen (Liesack und Stackebrandt 1992; Zarda et al. 1997; Ravenschlag et al. 2000).

## 1.1 Taxonomie

Die phylogenetische Stellung der Planktomyceten war über einen langen Zeitraum unklar. Analysen der 16S und 23S rDNA zeigten dann, dass die Planktomyceten eine monophyletische Gruppe bilden und eine eigene Hauptenwicklungslinie innerhalb der Bacteria darstellen (Woese 1985; Ward et al. 1995, 2000; Griepenburg et al. 1999). Die Ordnung *Planctomycetales* enthält eine einzige Familie, die *Planctomycetaceae* (Schlesner und Stackebrandt 1986). Die elf gültig beschriebenen Vertreter dieser Familie lassen sich teilweise anhand morphologischer Merkmale differenzieren und werden in die sechs Gattungen *Planctomyces* (Bauld und Staley 1976), *Pirellula* (Schlesner und Hirsch 1984),

*Blastopirellula*, *Rhodopirellula* (Schlesner et al. 2004), *Gemmata* (Giovannoni et al. 1987) und *Isosphaera* (Franzmann und Skerman 1984) eingeteilt.

Bei keiner der bisher isolierten Planktomyceten konnte Peptidoglykan nachgewiesen werden. Das Fehlen dieses Zellwandbestandteils stellt eine Besonderheit dar, die innerhalb der Bacteria ansonsten nur von den Chlamydien und den zellwandlosen Mycoplasmen bekannt ist. Die äußerst stabilen Zellwände der Planktomyceten bestehen aus Protein und sind reich an Prolin und Cystein/Cystin (Liesack et al. 1986). Die hohe Rigidität der Zellwände und ihre Beständigkeit gegenüber SDS wird auf die stabilisierende Wirkung von Disulfidbrücken zurückgeführt.

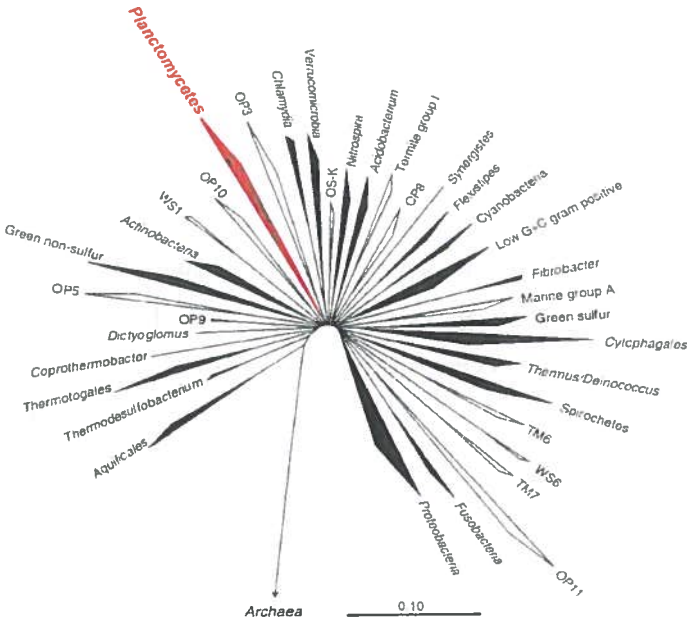



**Abb. 1**: Phylogenetischer Stammbaum der Domäne Bacteria basierend auf 16S rDNA-Sequenzierung. Phyla mit kultivierten Vertretern sind schwarz bzw. rot dargestellt und Entwicklungslinien, die nur durch Umweltsequenzen repräsentiert werden, erscheinen weiß. (Abbildung verändert nach Hugenholtz et al. 1998)

Ein weiterer chemotaxonomischer Marker, welcher die Abgrenzung der Planktomyceten von anderen Taxa unterstützt, ist das Vorkommen von Menachinon 6 (MK-6) als einzigem Chinon in der Atmungskette (Sittig und Schlesner 1993).

Weitere charakteristische Eigenschaften einiger Planktomyceten (z. B. *Rhodopirellula baltica*, *Planctomyces limnophilus*, *Gemmata obscuriglobus*, *Isosphaera pallida*) sind 5S rRNA-Moleküle, die sich mit einer Länge von 109 bis 111 Basen (Bomar et al. 1988) von den 118 Basen umfassenden „Minimal"-Strukturen der Archaea und

Bacteria unterscheiden (Erdmann und Wolters 1986). Des Weiteren wurde bei einigen Vertretern eine Entkopplung von 16S rRNA- und 23S rRNA-Genen nachgewiesen (Liesack und Stackebrandt 1989). Die eindrucksvolle phylogenetische Diversität dieser Bakteriengruppe wurde bereits in mehreren Studien aufgezeigt (Ward et al. 1995, 2000; Griepenburg et al. 1999).

Erst vor wenigen Jahren wurden Bakterien entdeckt, die Ammonium unter anaeroben Bedingungen zu molekularem Stickstoff oxidieren (Strous et al. 1999; Schmid et al. 2000). Phylogenetische Studien der 16S rDNA dieser Organismen zeigten, dass diese Annamox-Bakterien einen tief abzweigenden, monophyletischen Stamm innerhalb der Planktomyceten bilden. Ein chemotaxonomischer Marker dieser Bakterien, der auch zu ihrer Detektion eingesetzt wird, sind die so genannten Ladderane. Diese in der Natur zuvor noch nie nachgewiesenen Lipide bilden den Hauptbestandteil der Anammoxosommembran (Sinninghe Damsté et al. 2002; zu Anammoxosom siehe nächster Abschnitt).

## 1.2 Morphologie

Planktomyceten besitzen eine Reihe morphologischer Besonderheiten, aufgrund derer sie lichtmikroskopisch deutlich von anderen Bakterien unterschieden werden können. Die auffälligsten Merkmale einiger Vertreter dieses Taxons sind die Knospenbildung zur Zellteilung und die Bildung von Rosetten, welche aus aneinander gehefteten Einzelzellen bestehen. Schmidt und Starr entwickelten 1978 ein „Morphotypen-System", das zur Differenzierung verschiedener Planktomyceten diente.

Vertreter der Gattung *Planctomyces* zeichnen sich durch den Besitz eines nichtzellulären Stiels mit Holdfast-Substanz am vegetativen Pol aus. Diese Holdfast-Substanz befähigt die Zellen zur Anheftung an Oberflächen. Kraterförmige Strukturen und Fimbrien sind über den gesamten Zellkörper verteilt (Schmidt und Starr 1978). Die am reproduktiven Pol mittels Knospung gebildeten Tochterzellen besitzen bei einigen Stämmen eine Geißel und sind als Schwärmer mobil. Drei Arten dieser gestielten, knospenden Gattung sind auf der Basis der Untersuchung von Reinkulturen beschrieben worden: *Planctomyces maris* (Bauld und Staley 1976), *Planctomyces limnophilus* (Hirsch und Müller 1985) und *Planctomyces brasiliensis* (Schlesner 1989). *Planctomyces bekefii* (Gimesi 1924), *Planctomyces guttaeformis* und *Planctomyces stranskae* (Starr und Schmidt 1984) sind zwar gültig beschrieben, aber nicht in offiziellen Stammsammlungen hinterlegt. Sie konnten bisher nicht isoliert werden.

Die Gattungen *Pirellula, Blastopirellula* und *Rhodopirellula* beinhalten die gültig beschriebenen Arten *Pirellula staleyi* (Schlesner und Hirsch 1987), *Blastopirellula marina* und *Rhodopirellula baltica* (Schlesner et al. 2004). Mitglieder dieser Gattungen sind birnenförmig (daher der Name „Pirellula" = kleines Birnchen), oval oder rund. Die Holdfast-Substanz wird direkt am vegetativen Pol des Zellkörpers gebildet und ermöglicht den Zellen, sich an Oberflächen festzuheften oder durch Anheftung aneinander Rosetten zu bilden. Fimbrien und kraterförmige Strukturen befinden sich nur am reproduktiven Pol. Die durch Knospung am reproduktiven Zellpol gebildeten Tochterzellen lösen sich ab und können durch eine Geißel beweglich sein.

*Isosphaera pallida* (Giovannoni et al. 1987), der einzige beschriebene, thermophile Vertreter dieser Gattung, bewegt sich gleitend fort und bildet vielzellige Filamente, welche phototaktisch sind. *Gemmata obscuriglobus* (Franzmann und Skerman 1984), einziger beschriebener Vertreter dieser Gattung, hat eine runde bis ovale Zellform. Die Tochterzellen sind durch Geißelbüschel beweglich.

Es sei an dieser Stelle darauf hingewiesen, dass unterschiedliche Morphotypen innerhalb einer Kultur eines einzelnen Stammes beobachtet werden können. Diese stehen oft in Zusammenhang mit dem physiologischen Zustand der Kultur.

Eine weitere bemerkenswerte Besonderheit der Planktomyceten ist eine intracytoplasmatische Kompartimentierung, die sonst nur von Eukaryoten bekannt ist. In *P. staleyi*, *B. marina* und *R. baltica* unterteilt eine einschichtige intracytoplasmatische Membran (ICM) die Zelle in zwei separate Regionen, das Pirellulosom und die Polarkappen-Region (Lindsay et al. 1997). Das Pirellulosom enthält das kondensierte Nucleoid, ribosomenartige Partikel und den Großteil der RNA. Analoge Strukturen wurden auch in *Pl. maris, I. pallida, G. obscuriglobus* und „*Candidatus* Brocadia anammoxidans" nachgewiesen (Lindsay et al. 2001). In diesen Spezies wird der Bereich innerhalb der ICM als Riboplasma bezeichnet und der von DNA freie Teil zwischen ICM und Zellmembran als Paryphoplasma. In *G. obscuriglobus* wird das Nucleoid zusätzlich noch von einer Doppelmembran umschlossen. Die Entdeckung dieses dritten Kompartiments, das in seiner Struktur dem Nucleus der Eukaryoten gleicht, gab Anlass zur erneuten Diskussion über die Endosymbionten-Hypothese (Lindsay et al. 2001). Im Riboplasma der Anammox-Bakterien befindet sich ebenfalls ein drittes Kompartiment, das Anammoxosom. Innerhalb des Anammoxosoms, das interessanterweise DNA und tubuläre Strukturen enthält (Lindsay et al. 2001), findet räumlich getrennt vom Rest der Zelle die anaerobe Ammoniumoxidation statt. Die besondere biochemische Zusammensetzung der ICM

verhindert Diffusion und dient so unter anderem dem Schutz der Zelle vor toxischen Intermediaten des Anammoxprozesses (Sinninghe Damsté et al. 2002).

## 1.3 Ökologie und Physiologie

Da es seit der Beschreibung des ersten Planktomyceten durch Gimesi (1924) lange Zeit nicht gelang, eine Reinkultur dieser Bakterien zu isolieren, galten Planktomyceten jahrzehntelang als nicht kultivierbar. Seitdem Staley 1973 die erste Reinkultur erhielt, konnte durch verbesserte Isolierungsmethoden und den Einsatz neuer Medien eine große Anzahl Planktomyceten isoliert werden. Der überwiegende Teil der isolierten Planktomyceten stammt aus aeroben aquatischen Habitaten (Schlesner 1986, 1994; Staley et al. 1992). Molekularökologische Untersuchungen, wie z. B. Fluoreszenz *in situ* Hybridisierung und 16S rDNA-Sequenzierung, deuten auf eine wesentlich weiter reichende Verbreitung der Planktomyceten in der Umwelt hin. So wurde das Vorkommen von Planktomyceten auch in anoxischen Bereichen mariner und limnischer Sedimente (Gray und Herwig 1996; Miskin et al. 1999; Ravenschlag et al. 2000) sowie in Böden (Liesack und Stackebrandt 1992; Zarda et al. 1997) nachgewiesen. Tatsächlich gelang es vor einigen Jahren, *Gemmata* sp. aus Böden zu isolieren (Wang et al. 2002).

Diese Befunde deuten darauf hin, dass die Planktomyceten ubiquitär sind und einen wichtigen Anteil der natürlichen Bakterienpopulationen darstellen (Zarda et al. 1997; Hugenholtz et al. 1998; Neef et al. 1998). Das Wissen über Bedeutung und Funktion der Planktomyceten in den unterschiedlichen Ökosystemen ist dabei ähnlich lückenhaft wie die Kenntnisse über ihre stoffwechselphysiologischen Leistungen und Anpassungsfähigkeiten. Alle bisher in Reinkultur untersuchten Stämme sind chemoheterotroph und aerob. Während *Pl. maris*, *G. obscuriglobus*, *I. pallida*, *R. baltica*, und *B. marina* strikt aerob sind (Bauld und Staley 1976; Franzmann und Skerman 1984; Giovannoni et al. 1987; Schlesner et al. 2004), kann *Pl. limnophilus* Glucose vergären (Hirsch und Müller 1985).

Zucker und Zuckerderivate, wie z. B. *N*-Acetylglucosamin, sind die wichtigsten Kohlenstoff- und Energiequellen der Planktomyceten. Einige Stämme können mit *N*-Acetyglucosamin als einziger Kohlenstoff- und Stickstoffquelle wachsen. Über die Abbauwege der Zucker sowie deren Regulation ist bislang nichts bekannt. Aminosäuren können von den untersuchten Stämmen nicht verwertet werden und Peptide dienen als Stickstoff-, jedoch nicht als Kohlenstoffquelle (Schlesner et al. 2004).

Aufgrund hoher Abundanzen in marinen Lebensräumen (Glöckner et al. 1999; Gade et al. 2004), wie z. B. dem Vorkommen in „marine snow“ Partikeln (DeLong et al. 1993)

und ihrer Spezialisierung auf den Abbau von Kohlenhydraten, wird den Planktomyceten eine wesentliche Beteiligung im globalen Kohlenstoffkreislauf zugesprochen (Fuerst 1995).

Während alle bisher isolierten Planktomyceten organotroph sind, wurde vor einigen Jahren in Kläranlagen erstmals ein lithotropher Vertreter dieses interessanten Phylums entdeckt, der Ammonium anaerob zu oxidieren vermag (Strous et al. 1999; Schmid et al. 2000). Seither ist die anaerobe Ammoniumoxidation, bei der Ammonium mittels Nitrit zu elementarem Stickstoff umgesetzt wird, Gegenstand intensiver Forschung. Obwohl es bisher nicht gelang, Reinkulturen der daran beteiligten Organismen zu isolieren, konnten Vorkommen und Aktivität von Anammox-Planktomyceten bisher nicht nur in Reaktoren von Kläranlagen (Jetten et al. 2003) sondern auch in anoxischen Bereichen des Schwarzen Meeres (Kuypers et al. 2003) nachgewiesen werden. Daher ist der Anammoxprozess vermutlich auch von globaler Bedeutung für den Stickstoffkreislauf.

Die neuen Ergebnisse und das damit erneut aufkeimende Interesse an dieser besonderen Gruppe von Bakterien führten dazu, dass vier Planktomycetenstämme zur Genomesequenzierung ausgewählt wurden. Ein mariner Stamm, *Rhodopirellula baltica* (http://www.regx.de), zwei limnische Planktomyceten *Gemmata obscuriglobus* (http://www.tigr.org/tdb/mdb/mdbinprogress.html) und *Gemmata* sp. Wal-1 (http://www.genomesonline.org) sowie *Candidatus* "Kuenenia stuttgartiensis", ein bisher nur in Mischkultur vorliegender Annamox-Planktomycet (http://www-micrbiol.sci.kun.nl/research/nitrogen/nitrogen_genomics.html). Vergleichende Genom- und Funktionsanalysen werden in Zukunft weitere Einblicke in die Ökophysiologie der Planktomyceten ermöglichen.

## 2. *Rhodopirellula baltica*

*Rhodopirellula baltica* ist der einzige gültig beschriebene Vertreter der Gattung *Rhodopirellula* (Schlesner et al. 2004). Dieser marine Planktomycet wurde vor 30 Jahren aus dem Oberflächenwasser der Kieler Förde isoliert. Die Zellen sind durch einen bisher unbekannten Farbstoff rosa pigmentiert, birnenförmig und durchlaufen einen dimorphen Lebenszyklus (sessiler adulter versus mobiler juveniler Zelltypus). Adulte Zellen besitzen Fimbrien und kraterförmige Strukturen am reproduktiven Zellpol. Eine Holdfast-Substanz, welche zur Anheftung der Zelle an Oberflächen und/oder zur Rosettenbildung dient, wird am schmaleren vegetativen Zellpol ausgeschieden. Im natürlichen Habitat ist *R. baltica* fast ausschließlich in Rosetten oder assoziiert mit detritischen Partikeln zu finden (Gade et

al. 2004). Die Zellen wachsen relativ langsam mit Verdopplungszeiten von 10 – 14 h. Die Zellteilung erfolgt durch Knospung. Dabei bildet die oftmals sessile Mutterzelle am breiteren Zellpol eine Knospe, welche ein verkleinertes Spiegelbild der Mutterzelle darstellt. Die Knospe löst sich nach Erreichen einer bestimmten Größe von der Mutterzelle ab und ist mit Hilfe einer proximalen, subpolar inserierenden Geißel beweglich. Diese Schwärmerzelle besitzt keine Fimbrien und die kraterförmigen Strukturen sind über die gesamte Zelloberfläche verteilt. Später verliert die Tochterzelle ihre Geißel und bildet nach der Reifung zur adulten Zelle selber eine Knospe aus.

Zellen von *R. baltica* weisen eine intrazytoplasmatische Kompartimentierung auf. Das kondensierte Nucleoid und der Großteil der Ribosomen sind von einer einschichtigen Membran (ICM) umgeben. Im Vergleich zu anderen daraufhin untersuchten Planktomyceten weist *R. baltica* im elektronenmikroskopischen Bild zusätzlich zu einem zentralen Kompartiment (dem Pirellulosom vergleichbar) noch einige kleinere von einer Membran umschlossene Bereiche auf.

*R. baltica* ist strikt aerob und wächst chemoorganotroph, wobei der Stamm auf die Verwertung von Zuckern als Kohlenstoff- und Energiequelle spezialisiert ist. Das Genom (Glöckner et al. 2003) dieses Planktomyceten ist mit einer Größe von etwa 7.145 Mb eines der bislang größten komplett sequenzierten, zirkulären bakteriellen Genome. In den 7325 bioinformatisch vorhergesagten offenen Leserahmen (ORF) befinden sich unter anderem alle für die Glykolyse, den Tricarbonsäure-Zyklus (TCA-Zyklus) und den Pentosephosphat-Zyklus benötigten Gene und unerwartet ca. 110 Gene, welche für Sulfatasen kodieren. Letztere könnten dazu dienen, den Kohlenhydratanteil sulfatisierter Kohlenhydrate als Wachstumssubstrat verfügbar zu machen.

## 3. Funktionelle Genomanalyse

Seit es J. C. Venter 1995 gelang, das Genom des Bakteriums *Haemophilus influenzae* zu sequenzieren (Fleischmann et al. 1995), steigt die Anzahl der veröffentlichten bzw. aktuell bearbeiteten Genome exponentiell an; eine Entwicklung von der allgemein als „genomische Revolution“ gesprochen wird. Mittlerweile sind 165 bakterielle Genome vollständig sequenziert und zum Teil bereits veröffentlicht, während 489 weitere prokaryotische Genome in Arbeit sind (Stand Mai 2004; http://www.genomesonline.org). Nachdem die Genomforschung 2001 mit der Veröffentlichung der drei Milliarden Basenpaare des menschlichen Genoms (Lander et al. 2001) einen vorläufigen Höhepunkt erreichte, gelang es 2004 erstmals, bakterielle Genome direkt aus „Umwelt-DNA“ zu

rekonstruieren. Die Arbeitsgruppe von J. F. Banfield fügte die Genome von jeweils einer Spezies der Gattung *Leptospirillum* und *Ferroplasma* fast vollständig aus Proben eines natürlichen acidophilen Biofilms zusammen (Tyson et al. 2004). J. C. Venter und Mitarbeiter setzten Shotgun-Sequenzierung ein, um Umwelt-DNA aus der Sargassosee (nordwestlicher Atlantischer Ozean) zu untersuchen. Sie produzierten 1,045 Milliarden Basenpaare nicht redundanter Sequenzinformation aus schätzungsweise mindestens 1800 Spezies. Aus diesem gewaltigen Datensatz wurden mehr als 1,2 Millionen Gene vorhergesagt, von denen ca. zwei Drittel für Proteine unbekannter Funktion kodieren.

So ist in jüngster Vergangenheit ein explosionsartiges Wachstum der öffentlichen Sequenzdatenbanken zu verzeichnen, das jedoch zum erheblichen Teil auf genetische Information zurückzuführen ist, der bislang keine Funktion zugeordnet werden kann. Die wachsende Lücke zwischen Verfügbarkeit genetischer Information und biologischem Kenntniszuwachs zu schließen, ist Aufgabe der „funktionellen" Genomanalyse. Dieser Begriff umfasst vor allem Transkriptom- und/oder Proteomanalysen, die auf globaler Ebene die Untersuchung physiologischer und regulatorischer Prozesse erlauben. Besonders im Fall der hypothetischen Gene ermöglicht oftmals der physiologische Kontext ihrer Expression eine Funktionszuordnung.

Der Begriff Transkriptom bezeichnet alle Transkripte (mRNAs) einer Zelle, die unter definierten experimentellen Bedingungen gebildet werden (Velculescu et al. 1997). So erlaubt die Transkriptomanalyse zwischen konstitutiv exprimierten, selektiv induzierten und stillen Genen zu unterscheiden.

Eine in der Transkriptomanalyse vielfach angewandte Technik stellen die DNA-Microarrays dar. Sie erlauben eine schnelle quantitative Analyse der Genexpression. Auf so genannten „whole genome" Microarrays sind spezifische Sensoren (synthetische Oligonukleotide oder PCR-Produkte) für jedes vorhergesagte Gen immobilisiert. Sie werden mit dem zu analysierenden, fluoreszenzmarkierten cDNA-Pool (revers transkribierte mRNA) oder mit fluoreszenzmarkierter mRNA hybridisiert. Die Intensität der Hybridisierungssignale ist ein Maß für die Quantität der Genexpression. Durch Einsatz unterschiedlicher Fluoreszenzfarbstoffe können die charakteristischen Expressionsprofile unterschiedlicher biologischer Zustände mit einem Microarray bestimmt werden. Zu den gegenwärtigen technischen Limitierungen der Methode zählt die uneinheitliche Hybridisierungsqualität bei der gleichzeitigen Analyse vieler Transkripte.

Auf dem Informationsfluss von mRNA zum Protein findet noch eine Vielzahl regulatorischer Prozesse statt. Einzelne mRNA-Spezies werden unterschiedlich schnell

degradiert und mit variierender Effizienz translatiert. Somit gibt es keine einheitliche Korrelation zwischen mRNA und Proteinmenge (Anderson und Seilhamer 1997; Haynes et al. 1998; Gygi et al. 1999). Ebenso kann durch die mRNA-Analyse keine Aussage über posttranslationale Ereignisse getroffen werden. Viele Proteine werden nach der Translation modifiziert. Neu synthetisierte Polypeptide müssen erst richtig gefaltet, proteolytisch prozessiert oder modifiziert werden und möglicherweise an ihren spezifischen Wirkungsort gebracht werden. Phosphorylierungen, Acetylierungen, Methylierungen, Sulfatierungen und Glykosylierungen sind die häufigsten der mehr als 190 bekannten posttranslationalen Modifikationen (Krishna und Wold 1993). Keine davon ist durch Analyse der Genomsequenz zu erkennen und doch bestimmen sie die biologische Funktion der Proteine ganz entscheidend. Proteine bzw. Enzyme sind die Hauptakteure der Zelle. Da ihre aktuellen Mengen und Aktivitäten letztendlich den physiologischen Zustand und die Eigenschaften einer Zelle bestimmen, wurde im Rahmen dieser Arbeit die quantitative Proteomanalyse zur funktionellen Genomanalyse eingesetzt.

## 4. Proteomik

Das Arbeitsgebiet der Proteomik untersucht die Dynamik der Proteinzusammensetzung eines Organismus und umfasst dabei die Charakterisierung von Proteineigenschaften wie Abundanzen, posttranslationale Modifikationen und Lokalisation der Proteine. Ziel ist es dabei, einen globalen Überblick über die physiologischen und zellulären Prozesse sowie über die regulatorischen Netzwerke auf Proteinebene zu bekommen.

Da der Begriff „**Proteom**“ in der vorliegenden Arbeit häufig verwendet wird, soll er zunächst definiert werden:

*Das* **Proteom** (von engl. **prot**ein complement expressed by a gen**om**e, Wasinger et al. 1995) *umfasst die Gesamtheit aller von einer Zelle zu einem bestimmten Zeitpunkt und unter definierten Bedingungen synthetisierten Proteine.*

Das Proteom zeigt im Gegensatz zum statischen Genom eine ausgeprägte kontextabhängige Dynamik. Aus ökonomischen Gründen exprimiert ein Organismus möglichst nur diejenigen Gene, deren Produkte zum Überleben/Wachstum unter den jeweils vorliegenden Umweltbedingungen notwendig sind. Daher besitzt ein Organismus zwar nur ein Genom aber viele Proteome. Je nach Umweltbedingung variieren Lokalisation, Stabilität, Quantität und Art der posttranslationalen Modifkationen der Proteine. Die „Ein-Gen-ein-Protein-eine-Funktion“-Hypothese gilt lange als überholt. Nach heutigem Kenntnisstand dient in menschlichen Thrombozyten ein einziges Gen

durchschnittlich als Matrize für 2,3 verschiedene Proteine (O'Neill et al. 2002). Ebenso werden auch bei Bakterien aufgrund posttranslationaler Modifikationen in 2D Gelen etwa 1,4 Proteinspots pro ORF detektiert (Tonella et al. 1998, 2001). Ein Protein kann außerdem unterschiedliche Funktionen ausüben, oder die gleiche Funktion wie andere Proteine haben. Oftmals entfalten Proteine ihre Funktion nicht als Einzelmolekül, sondern als integraler Bestandteil von Proteinkomplexen. Für die globale Analyse von dynamischen Protein-Protein-Interaktionen und molekularen Netzwerken in der so genannten „funktionellen Proteomik" werden Techniken wie z. B. die Two-Hybrid Analyse (Fields und Song 1989; Uetz et al. 2000), die Tandemaffinitätsaufreinigung (Gavin et al. 2002) oder die Blue-Native-Gelelektophorese (Schagger und von Jagow 1991) eingesetzt.

Eine andere Fragestellung wird in der „Expressionsproteomik" verfolgt. Mit diesem Ansatz wird versucht, die relativen Häufigkeiten einzelner Proteine quantitativ zu erfassen, um so Stoffwechselwege und regulatorische Prozesse in Abhängigkeit von veränderten Umweltbedingungen aufzuklären. Die mittlerweile am häufigsten eingesetzten Techniken hierfür sind die zweidimensionale Gelelektrophorese in Kombination mit der Massenspektrometrie (MS). Der experimentelle Ablauf einer Proteomanalyse, wie sie in der vorliegenden Arbeit durchgeführt wurde, ist in Abbildung 2 dargestellt.

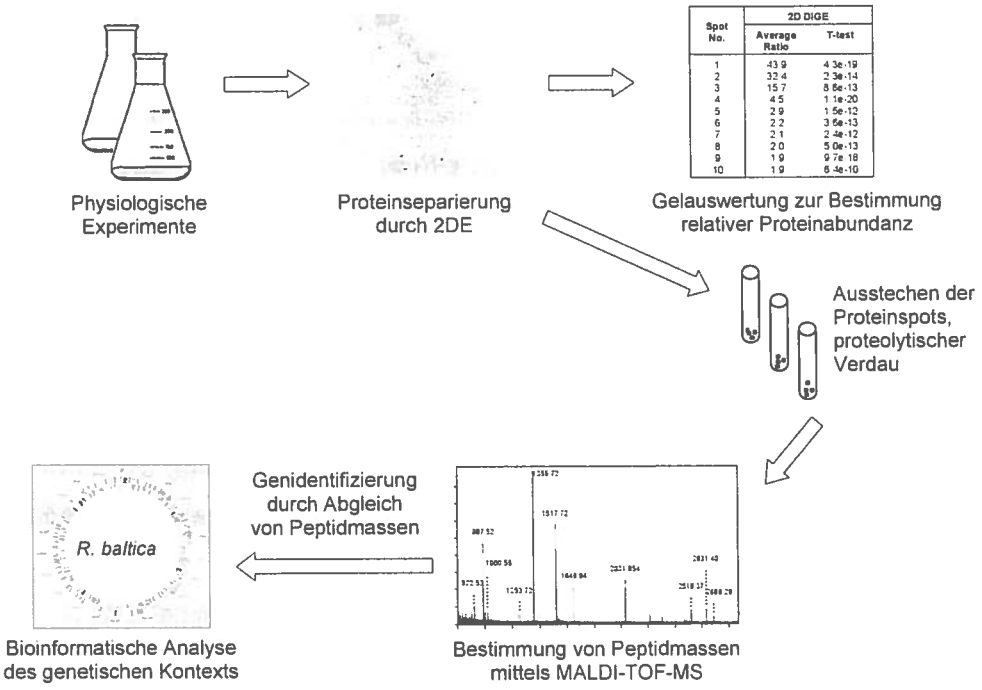



**Abb. 2** Ablaufschema einer quantitativen Proteomanalyse.

Schätzungen zufolge kann eine Zelle theoretisch zwischen 6 000 (*E. coli*; Tonella et al. 2001) und 30 000 (*Saccharomyces cerevisiae*; Corthals et al. 2000) verschiedene Proteinspezies synthetisieren. Keine der heutigen biochemischen Methoden kann diese Komplexität eines Proteoms im Ganzen erfassen, doch die rasanten Entwicklungen auf dem Gebiet der Proteomanalyse lassen dieses Ziel in greifbare Nähe rücken.

## 4.1 Zweidimensionale Gelelektrophorese

Die zweidimensionale Gelelektrophorese (2DE) kombiniert zwei Trennprinzipien, die auf verschiedenen physiko-chemischen Eigenschaften der Proteine beruhen. In der ersten Dimension, der isoelektrischen Fokussierung (IEF), werden die Proteine gemäß ihrer Isoelektrischen Punkte (IEP) in einem pH-Gradienten aufgetrennt. Der IEP entspricht dem pH-Wert, an dem die Nettoladung des Proteins Null ist. Demzufolge wandern Proteine so lange im elektrischen Feld, bis sie den pH-Wert erreichen, der ihrem IEP entspricht. Die IEF erlaubt keine vollständige Separierung, da Proteine unterschiedlicher Molekularmassen durchaus den gleichen IEP besitzen können. In der zweiten Dimension, einer Sodiumdodecylsulfat-Polyacrylamid-Gelelektrophorese (SDS-PAGE), werden die bereits IEF-separierten Proteine daher gemäß ihrer Größen aufgetrennt. Die Kombination dieser beiden Trennprinzipien erlaubt eine bisher mit anderen Methoden nicht erreichte Auflösung bei der Separierung von chemisch relativ gleichartigen Komponenten komplexer Proteinmischungen.

Die Ursprünge dieser Technik gehen auf eine Veröffentlichung aus dem Jahre 1970 zurück, in der H. Stegemann den kombinierten Einsatz von IEF und SDS-PAGE zur Proteinfraktionierung beschrieb. J. Klose und P. H. O'Farrell gelang es kurze Zeit später, das Auflösungsvermögen der 2DE durch den Einsatz denaturierender Bedingungen bei der Probenaufbereitung und während der IEF erheblich zu verbessern (Klose 1975; O'Farrell 1975). Noch heute ist die 2DE die am häufigsten in der Proteomforschung eingesetzte Technik (Rabilloud 2002), weil sie als einzige Methode die Möglichkeit bietet, komplexe Proteingemische schnell und hochauflösend zu trennen und gleichzeitig auf einem Gel darzustellen. In einem Gel können so bis zu 10 000 Proteine aufgetrennt und visualisiert werden (Klose und Kobalz 1995).

Die Einführung kommerzieller immobilisierter pH-Gradienten (Bjellqvist et al. 1982) für die IEF stellte einen weiteren Quantensprung in der Entwicklung der 2DE Technik dar. In den Anfängen der 2D-PAGE wurde die IEF mit freien Trägerampholyten durchgeführt. Durch die Separierung der freien Trägerampholyte im elektrischen Feld wird der pH-

Gradient erst während der Elektrophorese ausgebildet. Dieser so erzeugte pH-Gradient ist jedoch nur begrenzt stabil, was eine geringe Reproduzierbarkeit der IEF zur Folge hat. Bei immobilisierten pH-Gradienten wird durch die Kopolymerisation von Immobilinen (Acrylamidderivate mit Amino- und Carboxylgruppen unterschiedlicher pK-Werte) mit der Acrylamidmatrix ein stabiler pH-Gradient erzeugt. Polymerisierung auf Plastikfolien führte zu einer mechanischen Stabilisierung und leichteren Handhabung der IEF-Gele. So verbessern sie nicht nur die Reproduzierbarkeit der ersten Dimension, sondern auch die Auflösung. Es können mehrere Milligramm Protein aufgetrennt werden, wodurch auch die Detektion von Proteinen ermöglicht wird, die nur in geringer Kopienzahl in den Zellen vorliegen (Görg et al. 1988; Tonella et al. 2001). Durch den Einsatz gespreizter pH-Gradienten (Zoom-in-Gele) lassen sich mittlerweile auch Proteine mit nur geringfügig voneinander abweichenden IEPs separieren. Mehrere 2D Gele miteinander überlappenden pH-Gradienten können dann zu einem großen virtuellen Gel mit maximaler Auflösung zusammengesetzt werden.

Es ist heute möglich, Proteine mit IEPs zwischen 3 – 12 und einer Größe zwischen 10 und 200 kDa auf einem 2D Gel darzustellen (Görg et al. 1999). Probleme ergeben sich bei Proteinen, die auch unter denaturierenden Bedingungen nicht in Lösung gehen, wie z. B. Proteine mit mehreren transmembranen Domänen. Gemäß der Annahme, dass bis zu 25% aller Proteine eines Bakteriums integrale Membranproteine sind (Wallin und von Heijne 1998) und der Rest einen IEP zwischen 3 – 13 besitzt (Himmelreich et al. 1996), kann der überwiegende Anteil eines bakteriellen Proteoms auf einem oder mehreren 2D Gelen visualisiert werden.

Die Proteomanalyse mittels der 2DE Technik kann im Idealfall in zwei Phasen unterteilt werden. In der ersten Phase werden alle unter den zuvor vom Experimentator definierten Standardbedingungen von einem Organismus, einer Zelle oder einem Gewebe synthetisierten Proteine identifiziert und auf einem so genannten Mastergel dargestellt. Dieses Mastergel zeigt das „konstitutive“ Proteom und dient in Folgeversuchen als Landkarte, auf dem die Proteinkoordinaten wie Orte verzeichnet sind. In der zweiten Phase der Proteomanalyse werden Veränderungen des „konstitutiven“ Proteoms analysiert, die durch Veränderungen der Standardbedingungen hervorgerufen wurden. Hierbei kann es sich um Neubildung von Proteinen, relative Änderungen von Proteinabundanzen, posttranslationale Modifikationen, oder auch um Änderungen der subzellulären Lokalisation von Proteinen handeln.

## 4.2 Detektion von Proteinen in 2D Gelen

Nach der elektrophoretischen Trennung können die Proteine im 2D Gel mit Hilfe verschiedener Färbemethoden als Punkte (Proteinspots) sichtbar gemacht werden. Neben den seit langem etablierten Färbungen mit Coomassie Blau und Silber sind erst kürzlich Fluoreszenzfarbstoffe wie z. B. SYPRO® Ruby (Berggren et al. 2000; Yan et al. 2000) oder Ruthenium(II)-tris-(bathophenanthrolindisulfonat) (Rabilloud et al. 2001) und CyDyes™ (Ünlü et al. 1997) einsetzbar. Die genannten Verfahren unterscheiden sich nicht nur hinsichtlich Sensitivität, linearem Bereich und der damit verbundenen Quantifizierbarkeit von Proteinmengen, sondern auch in ihren Färbeprinzipien.

**Tabelle 1.** Detektionsgrenze, linearer Bereich und Massenspektrometrie-Kompatibilität der in der 2D Technik eingesetzten Färbemethoden.

| Farbstoff | Detektionsgrenze [ng] | Linearer Bereich Größenordnungen | MS-Kompatibilität |
|---|---|---|---|
| Coomassie Blau | 30 – 100 | 1 – 2 | + |
| kolloidales Coomassie Blau | 5 – 10 | 1 – 2 | + |
| Silber | 0,05 – 2 | 1 – 2 | (+)[1] |
| SYPRO® Ruby | 1 – 2 | 4 – 5 | + |
| CyDyes™ | 0,25 – 0,95 | 4 | + |
| Pro-Q® Diamond | 1 – 16 | 3 | + |
| Pro-Q® Emerald 300 | 1 – 10 | 3 | + |

[1] MS-kompatible Silberfärbungen weisen eine geringere Sensitivität auf.

Die Proteindetektion mit Coomassie, dem unempfindlichsten Färbeverfahren, beruht auf der nicht-kovalenten Anlagerung des Farbstoffs an hydrophobe Bereiche der Proteine. Aufgrund des geringen dynamischen Bereichs über maximal zwei Größenordnungen erlaubt diese Färbemethode nur semiquantitative Aussagen. Ein großer Vorteil dieser Methode ist, dass sie sehr gut mit den nachfolgenden Identifizierungsmethoden (z. B. Edman-Sequenzierung oder Massenspektrometrie) kompatibel und einfach anzuwenden ist.

Silber ist die empfindlichste nicht-radioaktive Färbemethode, die zurzeit standardmäßig in 2D Laboren eingesetzt wird. Die adsorbtiven Vorgänge bei der reduktiven Silberfärbung sind allerdings nur schlecht verstanden und stark von experimentellen Bedingungen abhängig. Außerdem erfasst die Silberfärbung auch Nichtproteinanteile der Probe. Eine Quantifizierung ist aufgrund des engen linearen

Bereichs (max. zwei Größenordnungen) nicht möglich. Die Proteinidentifizierung mittels MS-Analyse kann nur durchgeführt werden, wenn eine Färbung ohne Vernetzungsreagenz, wie z. B. Glutardialdehyd, durchgeführt wurde. Letzteres führt jedoch zu einer Verringerung der Sensitivität (Shevchenko et al. 1996).

SYPRO® Ruby besitzt eine Sensitivität, die der von MS-kompatiblen Silberfärbungen entspricht. Der Fluoreszenzfarbstoff lagert sich durch hydrophobe Wechselwirkungen an basische Aminosäuren an (Steinberg et al. 2000). Die Detektion gefärbter Proteine erfordert die Verfügbarkeit eines Fluoreszenzscanners. Der große lineare dynamische Bereich über vier bis fünf Größenordnungen und die Kompatibilität mit der Massenspektrometrie sind wichtige Vorteile gegenüber konventionellen Färbetechniken (Lopez et al. 2000).

Posttranslationale Modifikationen spielen eine wichtige Rolle in vielen zellulären Prozessen. Erst kürzlich wurden Fluoreszenzfarbstoffe (Pro-Q® Diamond phosphoprotein gel stain, Pro-Q® Emerald 300 glycoprotein gel stain) entwickelt, welche eine selektive Detektion phosphorylierter und glykosylierter Proteine direkt im Polyacrylamidgel ermöglichen (Steinberg et al. 2003; Schulenberg et al. 2003). Durch aufeinander folgendes Färben eines 2D Gels mit den beiden Farbstoffen und mit SYPRO® Ruby ist es möglich posttranslational modifizierte Proteine gezielt zu detektieren und gleichzeitig alle serparierten Proteine quantitativ zu erfassen. Pro-Q® Diamond und Pro-Q® Emerald 300 sind sehr sensitiv, haben einen dynamischen Bereich über ca. drei Größenordnungen und sind MS-kompatibel.

Im Gegensatz zu den bereits genannten postelektrophoretischen Färbetechniken gibt es Verfahren, in denen die Proteine bereits vor der elektrophoretischen Trennung kovalent mit Fluoreszenzfarbstoffen oder radioaktiv (Einbau von z. B. $^{35}$S-Methionin in neusynthetisierte Proteine) markiert werden.

Bei der DIGE-Technologie (difference gel electrophoresis) werden die Proteine bereits vor der 2DE kovalent an der ε-Aminogruppe von Lysinen mit Fluoreszenzfarbstoffen (CyDyes™) markiert. Die drei verfügbaren CyDyes™ sind sehr sensitiv und weisen einen breiten dynamischen Bereich auf. Ihre identischen Migrationseigenschaften („charge“ und „mass“ matching) im 2D Gel und deutlich voneinander unterscheidbare Absoptions- und Emissionsmaxima ermöglichen ein Multiplexing, bei dem in einem einzigen Gel gleichzeitig zwei Proben und ein interner Standard getrennt und nacheinander mit einem Fluoreszenzscanner detektiert werden.

Theoretisch können sich die Konzentrationsunterschiede zwischen dem am wenigsten und dem am häufigsten vorkommenden Protein in einer Zelle über bis zu sechs Größenordnungen erstrecken (Rabilloud 2002). Dieser Bereich wird von den gängigen Fluoreszenzmethoden mit ihrem dynamischen Bereich von vier bis fünf Größenordnungen nur bedingt abgedeckt.

Radioaktive Markierungen hingegen haben zumindest theoretisch den gleichen dynamischen Konzentrationsbereich wie Proteine. Dies hängt natürlich von der Markierungseffizienz der verwendeten chemischen oder biochemischen Verfahren ab. Für eine anschließende Identifizierung interessanter Proteine müssen jedoch zusätzlich konventionell gefärbte Gele eingesetzt werden.

## 4.3 Bildanalyse

Die Proteinmuster auf 2D Gelen sind sehr komplex (ca. 2 000 Spots pro Gel bei *R. baltica*). Nur durch den Einsatz spezieller Software ist es möglich, Unterschiede in Proteinmustern zweier Zustände umfassend zu ermitteln. Zu bewältigende Aufgaben bestehen dabei in der automatischen Spoterkennung, der Quantifizierung der Proteinspots und der möglichst präzisen Zuordnung korrespondierender Spots in mehreren Parallelgelen. Die kommerzielle DIGE-Technologie schließt eine spezielle Software (DeCyder™) ein, welche erstmals die Kodetektion von Spotmustern mit statistischen Analysen verknüpft. Damit ist es möglich, ermittelte Änderungen der Proteinabundanzen mit statistischen Signifikanzen zu belegen.

Eine voll automatisierte Bildanalyse existiert bislang noch nicht. Vielmehr sind noch immer umfangreiche manuelle Korrekturen bei der Auswertung der Gelbilder notwendig.

## 4.4 Identifizierung von Proteinen mittels Massenspektrometrie und Abgleich von Peptidmassen

Die Massenspektrometrie (MS) stellt heute eine der wichtigsten analytischen Methoden in den Biowissenschaften dar. Einen wesentlichen Beitrag dazu leistete die Entwicklung der schonenden und zugleich effizienten Ionisierungstechniken Matrix-Assisted-Laser-Desorption-Ionisation-MS (MALDI-MS) und Elektrospray-Ionisation-MS (ESI-MS), welche die Bestimmung der molekularen Massen von großen Biomolekülen ermöglichen und damit zu den wichtigsten Techniken in der routinemäßigen Peptid- und Proteinanalytik avancierten (Aebersold und Goodlett 2001; Mann et al. 2001). Aufgrund der Sensitivität im Femtomolbereich, der Schnelligkeit der Analysen und der möglichen

Automatisierbarkeit haben sie die langsame und arbeitsaufwendige N-terminale Sequenzanalyse (Edman-Sequenzierung) weitgehend ersetzt.

Zur Identifizierung mittels MS werden die separierten Proteine aus dem Gel ausgestochen, durch eine sequenzspezifische Protease, z. B. Trypsin, gezielt fragmentiert und aus dem Gel eluiert. Trypsin hydrolysiert die Polypeptidketten stets nach einer basischen Aminosäure (sofern die darauf folgende Aminosäure nicht Prolin ist) und generiert dadurch ein für das jeweilige Protein spezifisches Gemisch aus Peptiden mit definierter Länge und Aminosäurezusammensetzung. Die proteinspezifische Komposition dieser Peptidgemische hängt im Falle des tryptischen Verdaus nur von der individuellen Verteilung der beiden Aminosäuren Arginin und Lysin im jeweils analysierten Protein ab. Bei der MALDI-TOF-MS werden die zu analysierenden Peptide nach der Fragmentierung in eine kristalline Matrix eingebettet, durch einen kurzen Laserimpuls freigesetzt und ionisiert. Die sich daraufhin in der Gasphase befindlichen Ionen werden in einem elektrischen Feld mit gleicher kinetischer Energie beschleunigt und im Hochvakuum einer feldfreien Driftstrecke aufgrund ihrer Flugzeit getrennt. Die zeitliche Separierung der Ionen beruht darauf, dass bei gleicher Beschleunigungsenergie leichte Ionen schneller fliegen als schwere Ionen und daher den Detektor früher erreichen.

Bei der ESI-MS liegt der Analyt in flüssiger Form vor. Die Peptidlösung wird unter Atmosphärendruck durch eine dünne Kapillare gesprüht und dabei in Form elektrisch geladener Tröpfchen zerstäubt. Unter Anlegen von Hochspannung wird das Lösungsmittel kontinuierlich verdampft und die Peptide ionisiert. Die Masse der Peptidfragmente wird in diesem Fall nicht durch die Messung der Flugzeit, sondern aufgrund der Stärke der Ablenkung der Flugbahn der Ionen im elektrischen Feld bestimmt.

Das Ergebnis beider Methoden ist für jedes untersuchte Protein ein charakteristischer Satz von Peptidmassen-Fingerabdrücken (peptide mass fingerprints, PMF). Diese proteinspezifischen Massenleitern werden durch Datenbankvergleiche zur Proteinidentifizierung herangezogen. Dabei übersetzt eine Software jede in der verwendeten Datenbank hinterlegte Gensequenz in eine Proteinsequenz und berechnet für jedes dieser „theoretischen" Proteine die Massen aller Peptide, die durch theoretische Spaltung mit der experimentell verwendeten Protease entstehen würden. Durch den Abgleich dieser mittels *in silico* Verdau erhaltenen „theoretischen" Peptidmassen mit den experimentell erhaltenen Peptidmustern, lässt sich der überwiegende Teil der Proteine einem Gen in der Datenbank zuordnen und damit identifizieren. Die Güte der Übereinstimmung zwischen experimentellem und theoretischem PMF wird in einem

Qualitätsmaß („score“ und/oder „coverage“) ausgedrückt. Je mehr Übereinstimmungen gefunden werden, je genauer diese Übereinstimungen sind und je weniger Peptidmassen vorliegen, welche nicht zugeordnet werden können, desto höher ist die Qualität der Identifizierung.

Die beschriebenen Methoden zur massenspektrometrischen Proteinanalyse führen allerdings nur zum Erfolg, wenn die DNA- oder Proteinsequenz in den Datenbanken vorliegt. Ist das nicht der Fall, können massenspektrometrische *de novo* Sequenzierung oder auch klassische Methoden, wie z. B. Edman-Sequenzierung, zur Identifizierung eingesetzt werden, um diese voranzutreiben.

## 5. Ziele der vorliegenden Arbeit

Zu Beginn der vorliegenden Arbeit war *Rhodopirellula baltica* (ehemals *Pirellula* sp. Stamm 1) ein noch kaum untersuchtes Bakterium. Daher sollten klassische Wachstumsexperimente durchgeführt werden, um das physiologische Potential dieses Planktomyceten zu erfassen. Der Vergleich seiner physiologischen Eigenschaften mit denen von *Blastopirellula marina* und *Pirellula staleyi* sollten unter Einbeziehung chemotaxonomischer Marker schließlich eine gültige taxonomische Beschreibung ermöglichen.

Hauptziel der vorliegenden Arbeit war es, die physiologischen Eigenschaften und regulatorischen Anpassungsfähigkeiten von *R. baltica* auf molekularer Ebene zu untersuchen. Dies sollte durch einen integrierten physiologisch/proteomischen Ansatz unter Einbeziehung der genomischen Sequenz erreicht werden. Als zentrale experimentelle Methode sollte die zweidimensionale Gelelektrophorese eingesetzt werden. Um die bei veränderten physiologischen Bedingungen zu erwartenden Veränderungen im Proteinmuster quantitativ und statistisch abgesichert zu erfassen, sollte die aktuelle 2D DIGE-Technologie etabliert und evaluiert werden.

Als Basis für die differentiellen proteomischen Studien sollte ein Mastergel erstellt werden. Dies sollte möglichst viele der unter Standardbedingungen exprimierten löslichen Proteine mit einem isoelektrischen Punkt zwischen pH 4 und 7 enthalten und entsprechend der massenspektrometrischen Identifizierung annotiert sein. Das Mastergel sollte die Zuordnung der differentiell exprimierten Proteine in Folgeversuchen erleichtern.

Die zentralen Stoffwechselwege für den Kohlenhydratabbau in *R. baltica* sollten identifiziert und durch Aktivitätsbestimmungen ausgewählter Schlüsselenzyme untermauert werden. Des Weiteren sollten Enzyme der zuckerspezifischen peripheren Abbauwege identifiziert und ihre Regulation untersucht werden.

In einem weiteren Vorhaben sollten die Veränderungen des Proteoms im Verlauf der Wachstumskurve von *R. baltica* erfasst werden. Zum einen sollte damit die molekulare Stressantwort hinsichtlich Substratlimitierung und hoher Zelldichte untersucht werden, zum anderen sollten Proteine identifiziert werden, die möglicherweise bei der Ausdifferenzierung bestimmter Morphotypen des Lebenszyklus von *R. baltica* eine Rolle spielen.

# B Ergebnisse und Diskussion

Im Folgenden werden die Ergebnisse der Publikationen/Manuskripte übergreifend im Zusammenhang dargestellt und diskutiert. Für eine detaillierte Betrachtung aller Einzelergebnisse dieser Arbeit sei auf die angefügten Veröffentlichungen verwiesen. Ausführlich werden an dieser Stelle Aspekte dargestellt, die in den Veröffentlichungen nicht oder nur in begrenztem Umfang behandelt wurden.

## 1. Taxonomische Einordnung von *Rhodopirellula baltica*

Zu dem Zeitpunkt als die ersten genomischen und physiologischen Untersuchungen an *R. baltica* durchgeführt wurden, war der Organismus taxonomisch noch nicht gültig beschrieben. Bis zur aktuellen Beschreibung wurde dieser Planktomycet aufgrund seiner birnenförmigen Zellmorphologie und vorangegangener phylogenetischer Studien (Ward et al. 1995, 2000; Griepenburg et al. 1999) der Gattung *Pirellula* zugeordnet und als *Pirellula* sp. Stamm 1 bezeichnet. Da *Pirellula* sp. Stamm 1 bereits 1974 aus dem Oberflächenwasser der Kieler Förde isoliert wurde, war der Stamm schon in zahlreichen Studien untersucht worden, die sich mit Phylogenie, Zellwandaufbau (König et al. 1984; Liesack et al. 1986) sowie weiteren taxonomischen Markern wie Phospholipiden (Kerger et al. 1988; Sittig und Schlesner 1993) oder Polyaminen (Griepenburg et al. 1999) bei den Planktomyceten beschäftigten.

Zur taxonomischen Einordnung von *Pirellula* sp. Stamm 1 musste auch die taxonomische Position der Gattung *Pirellula* erneut evaluiert werden. Nach allgemein anerkannten Richtlinien sollten individuelle Spezies einer Gattung untereinander mehr als 95% 16S rDNA-Sequenzidentität aufweisen (Stackebrandt und Goebel 1994). Da die 16S rDNA der bereits in der Gattung *Pirellula* zusammengefassten Spezies *P. staleyi* und *P. marina* untereinander und ebenso zu *Pirellula* sp. Stamm 1 weniger als 90% Sequenzidentität aufwiesen (Tabelle 3, Publikation 3), wurden die beiden neuen Gattungen *Rhodopirellula* und *Blastopirellula* vorgeschlagen, um der phylogenetischen Diversität dieser drei Planktomyceten gerecht zu werden. *Pirellula* sp. Stamm 1 wurde als Typusstamm der neuen Spezies *Rhodopirellula baltica* beschrieben. Der Gattungsname nimmt hier auf die rosa Färbung der Kolonien und auf die birnenartige Zellform Bezug, während der Speziesname auf den Ort der Isolierung (Baltic Sea) hinweist. *P. marina* wurde als *Blastopirellula marina* in die neue Gattung *Blastopirellula* (knospende *Pirellula*)

überführt. Für *P. staleyi,* der ersten beschriebenen Spezies dieser Gruppe, blieb der Gattungsname erhalten.

Um diese aus phylogenetischer Sicht gerechtfertigte Aufspaltung der Gattung *Pirellula* zusätzlich noch durch vergleichende phänotypische und taxonomische Daten zu unterstützen, wurden *P. staleyi*, *B. marina* und *R. baltica* hinsichtlich markanter Unterscheidungsmerkmale in ihrer Physiologie, ihren Polyaminmuster und Phospholipide vergleichend untersucht. Im Rahmen der vorliegenden Arbeit wurden klassische Wachstumsexperimente durchgeführt, in denen die drei oben genannten Typusstämme hinsichtlich ihrer Substratspektren und der Fähigkeit zu anaerobem Wachstum untersucht wurden. Die Ergebnisse dieser Studien zeigten, dass sich die phylogenetische Diversität von *P. staleyi*, *B. marina* und *R. baltica* nicht in ihren physiologischen Eigenschaften widerspiegelt. Alle drei Stämme nutzten überwiegend Kohlenhydrate als Energie- und Kohlenstoffquelle und sind somit als ausgesprochene Zuckerspezialisten anzusprechen. Chondroitinsulfat konnte nur von den zwei marinen Organismen, *B. marina und R. baltica,* als Wachstumssubstrat verwertet werden. Diese Beobachtung ist interessant, da 110 Gene für putative Sulfatasen im Genom von *R. baltica* entdeckt wurden (Glöckner et al. 2003). Es ist davon auszugehen, dass die Sulfatasen in marinen Organismen dazu verwendet werden, sulfatisierte Kohlenhydrate (z. B. Carrageen) durch Abspaltung der Sulfatgruppe für den weiteren Abbau zugänglich zu machen.

Weiterhin sind *P. staleyi*, *B. marina* und *R. baltica* strikt aerob. Unter fermentativen Wachstumsbedingungen sowie mit Nitrat als terminalem Elektronenakzeptor fand kein Wachstum statt. Dieses Ergebnis stand im Widerspruch zur Erstbeschreibung von *B. marina* (damals noch *P. marina*, Schlesner 1986), in welcher der Organismus als fakultativ anaerob beschrieben wurde. In die überarbeitete aktuelle Beschreibung von *B. marina* ist dieses neue Ergebnis aufgenommen worden.

Obwohl im Genom von *R. baltica* alle Gene für eine Milchsäuregärung vorhergesagt wurden (Glöckner et al. 2003), gelang es bisher nicht, diesen Organismus unter fermentativen Bedingungen zu kultivieren. Da bisher kein Gen für eine anaerobe Ribonukleotid-Reduktase im Genom gefunden wurde, steht zu vermuten, dass die Milchsäuregärung, wenn sie überhaupt durchgeführt wird, dem Zweck der Aufrechterhaltung des Stoffwechsels aber nicht dem Wachstum dient (Publikation 3). In Anbetracht der physiologischen Uniformität von *P. staleyi*, *B. marina* und *R. baltica* sind letztendlich auch taxonomische Marker (Fettsäuren, Phospholipide, Polyamine und

Aminosäurezusammensetzung der Zellwände) und morphologische Charakteristika zur Unterscheidung der drei Gattungen herangezogen worden (Publikation 3, Tabelle 3).

## 2. Evaluierung von in der 2D Technik angewandten Färbemethoden

Die zweidimensionale Gelelektrophorese mit anschließender Identifizierung ausgewählter Proteinspezies (MALDI-TOF-MS, PMF) bildet die Grundlage für die qualitativen und quantitativen Untersuchungen des Proteoms von *R. baltica* nach physiologischen Anpassungsexperimenten. Die 2DE ist bislang die höchstauflösende Technik zur Trennung komplexer Proteinmischungen. Bis zu 10 000 Proteine können in einem Gel reproduzierbar aufgetrennt werden, ohne dass Kenntnisse über Zusammensetzung der Proteinmischung und Beschaffenheit der Proteine vorliegen müssen (Klose und Kobalz 1995). Diese Eigenschaft und die Kombination mit der Massenspektrometrie machte die 2DE zu einer der Kerntechniken in der aktuellen Proteomforschung. In der vorliegenden Arbeit wurden diese Techniken eingesetzt, um erstmals differentielle Genexpression in *R. baltica* auf Proteinebene zu verfolgen und dadurch mehr über Physiologie, Zellentwicklung und regulatorische Netzwerke dieses marinen Planktomyceten zu erfahren.

Das Ziel funktioneller Proteomanalyse ist jedoch nicht nur die reproduzierbare Auftrennung von Proteinen, sondern vielmehr deren anschließende Detektion und die Quantifizierung von Änderungen der Proteinabundanzen. Letztere spiegeln die molekulare Anpassung eines Organismus auf variable Umwelteinflüsse wider. Limitationen der 2D Technik ergeben sich vor allem auf der Ebene der Detektionsmethoden und damit verbunden bei der zuverlässigen Quantifizierbarkeit der Proteine. Für beide Schritte sind zwei Parameter von entscheidender Bedeutung, nämlich die Sensitivität und der lineare dynamische Bereich der eingesetzten Färbemethode. Eine hohe Sensitivität ermöglicht die Detektion von Proteinen, die nur in geringer Kopienzahl in der Zelle vorliegen. Ein breiter dynamischer Bereich ist ebenso wichtig, denn nur wenn das Signal eines Proteinspots proportional zur Proteinmenge ist, lassen sich zuverlässige Aussagen über Abundanzveränderungen treffen.

Zu Beginn dieser Arbeit wurden hauptsächlich drei Färbetechniken zur Proteinvisualisierung eingesetzt, welche unterschiedliche Vor- und Nachteile aufweisen. Die am häufigsten eingesetzte Färbemethode war die Silberfärbung, welche mit 0,05 – 2 ng je Proteinspot die höchste Sensitivität aufweist (Publikation 2, Tabelle 1). Der lineare dynamische Bereich erstreckt sich jedoch nur über ein bis zwei Größenordnungen, wodurch bei hohen Proteinkonzentrationen schnell der Sättigungsbereich der Färbung

erreicht ist. Die Proteinmenge lässt sich in solchen Fällen nicht quantifizieren. Des Weiteren ist die Färbung stark von verschiedenen experimentellen Parametern, wie z. B. Färbedauer und Temperatur, abhängig und somit wenig reproduzierbar. Die kolloidale Coomassie Färbung ist bezüglich der Reproduzierbarkeit viel zuverlässiger, jedoch um etwa eine Größenordnung weniger sensitiv als die Silberfärbung. Aufgrund des vergleichbar engen dynamischen Bereichs eignet sich auch diese Färbemethode nicht zur Quantifizierung. Im Gegensatz zu Silber beeinträchtigt der Coomassie Farbstoff jedoch die Proteinidentifizierung mittels MS nicht und wird daher häufig eingesetzt.

Erste Quantifizierungen wurden mit dem Fluoreszenzfarbstoff SYPRO® Ruby durchgeführt. Der Farbstoff ist ähnlich sensitiv wie Silber, weist jedoch einen weitaus größeren dynamischen Bereich von vier bis fünf Größenordnungen auf. Dieses Detektionsverfahren ist zudem schnell und zuverlässig. Die mit einem Fluoreszenzscanner generierten Gelbilder wurden mit einer Software (Image Master™) wie folgt ausgewertet.

Um statistisch sinnvolle, quantitative Aussagen über die Änderung von Proteinabundanzen treffen zu können, müssen mindestens drei Parallelgele von einer Probe analysiert werden. Die Proteinmuster der Parallelgele sind einander ähnlich jedoch aufgrund von Gel-zu-Gel Variationen nie identisch. In einem einzelnen Experiment werden die Proteome zweier Proben, welche unterschiedliche physiologische Zustände reflektieren, miteinander verglichen. Bei drei parallelen Gelen pro Probe enthält ein einzelnes Experiment sechs Gele.

Das für die Bildauswertung eingesetzte Softwareprogramm erkennt die einzelnen Spots und ordnet durch Mustererkennung übereinstimmende Spots in parallelen Gelen einander zu. Dieser Vorgang wird Matching genannt und erzeugt aus den Parallelgelen ein virtuelles Mastergel. Dieses Mastergel enthält folglich alle Spots, denen übereinstimmende Spots aus den Parallelgelen zugeordnet werden konnten. Für die zu vergleichenden Proben wird jeweils ein virtuelles Mastergel generiert. Durch den anschließenden Vergleich dieser Mastergele lassen sich dann für jeden Spot relative Veränderungen in der Abundanz errechnen und den untersuchten Proben zuordnen. Ein Nachteil dieses Verfahrens ist, dass bereits im ersten Schritt der Spoterkennung in den einzelnen Gelen oftmals umfangreiche manuelle Korrekturen vorgenommen werden müssen, um nicht erkannte Spots zu berücksichtigen oder abweichende Spotgrenzen in Parallelgelen einander anzugleichen.

Einen völlig neuen Ansatz zur quantitativen Untersuchung differentieller Proteinmuster stellt die DIGE (difference gel electrophoresis)-Technologie dar. Diese Technik wurde in der Arbeitsgruppe von J. S. Minden entwickelt (Ünlü et al. 1997) und

von Amersham Biosciences kommerzialisiert. Bei diesem Verfahren werden die Proteine bereits vor der 2DE mit einem Fluoreszenzfarbstoff markiert, welcher kovalent an die ε-Aminogruppen der Lysine bindet. Die Markierungsreaktion wird so durchgeführt, dass theoretisch jedes Protein maximal nur ein Farbstoffmolekül trägt und dass nur 1 – 2% jeder Proteinspezies überhaupt markiert wird. Dieser Vorgang wird als Minimalmarkierung bezeichnet. Es stehen drei Farbstoffe (Cy2, Cy3 und Cy5) zur Verfügung, die über diskrete nicht überlappende Absorptions- und Emissionsmaxima verfügen. Die drei Farbstoffe besitzen sehr ähnliche Molekulargewichte („mass matching") und kompensieren durch ihre einfache positive Ladung den Verlust einer positiven Ladung bei der Bindung an das Lysin („charge matching"). „Mass" und „charge matching" stellen sicher, dass das Trennverhalten der markierten Proteine nicht beeinflusst wird.

Diese Eigenschaften der Cy-Farbstoffe ermöglichen es, Proteine aus drei Proben gleichzeitig in einem Gel aufzutrennen (Multiplexing) und somit Gel-zu-Gel Variationen auszuschließen. Die Erzeugung von Gelbildern erfolgt dann mit einem Fluoreszenzscanner, der mit drei Lasern ausgestattet ist, um die Cy-Farbstoffe entsprechned ihrer Absorptionsmaxima gezielt anzuregen. In drei aufeinander folgenden Scanvorgängen werden so bis zu drei digitale Bilder von einem Gel generiert.

Das innovative Element dieser Methode liegt im Vergleich zu „traditionellen" Färbetechniken in der gleichzeitigen Trennung zweier zu vergleichender Proteinproben zusammen mit einem internen Standard in ein und demselben Gel. Der interne Standard enthält ein Aliquot aller im Versuch zu untersuchenden Proben und somit alle Proteinspezies. Die Spotdetektion durch die DeCyder™ Software erfolgt auf der Basis dieses internen Standards. Beim anschließenden Vergleich von Parallelgelen dient der interne Standard als Anker zum Matchen der Gelbilder. Durch Kodetektion und statistische Analyse werden Unterschiede in der Proteinhäufigkeit mit großer statistischer Signifikanz erfasst.

Die DIGE-Technologie stellt ein noch relativ neues Verfahren dar, das bisher nur in wenigen medizinischen Studien (Gharbi et al. 2002; Skynner et al. 2002; Zhou et al. 2002) und beim Modellbakterium *Escherichia coli* (Yan et al. 2002; Alban et al. 2003) eingesetzt wurde. Zur Evaluierung von 2D DIGE für die differentielle, quantitative Proteomanalyse bei einem Umweltbakterium, wurde dieses System für die Detektion und Quantifizierung von Proteinen eingesetzt, die spezifisch für den *N*-Acetylglucosamin-Metabolismus in *R. baltica* sind. In einer vorangegangenen Studie (Publikation 1) wurden Proteinmuster von Zellen verglichen, die an das Wachstum mit *N*-Acetylglucosamin bzw. Glucose

angepasst waren. Dabei wurden *N*-Acetylglucosamin-spezifische Proteine detektiert und auf der Basis einer Färbung mit SYPRO® Ruby quantifiziert.

Das Evaluierungsexperiment basierte zunächst auf zehn parallelen 2D DIGE Gelen. In fünf dieser Gele wurde Probe 1 (Proteine von auf Glucose angezogenen Zellen) mit Cy3 und Probe 2 (Proteine von auf *N*-Acetylglucosamin angezogenen Zellen) mit Cy5 markiert; bei den anderen fünf Gelen wurde mit der Markierung umgekehrt verfahren. Die Proteine des internen Standards (gepoolte Aliquots von Probe 1 und Probe 2) wurden mit Cy2 markiert und die Bildauswertung erfolgte mit der DeCyder™ Software. Zum Vergleich der DIGE-Technologie mit dem SYPRO® Ruby System wurde ein weiteres Experiment mit jeweils vier SYPRO® Ruby gefärbten Gelen pro Anzuchtbedingung durchgeführt. Die Bildanalyse wurde in diesem Fall mit der Image Master 2D Elite™ Software durchgeführt.

Bis auf wenige Ausnahmen stimmten die Tendenzen der mit den zwei Systemen ermittelten Unterschiede der Proteinabundanzen überein (Publikation 2, Tabelle 3). Bei der Betrachtung der absoluten Werte unterscheiden sich die relativen Abundanzen jedoch um das 1 – 3,7-fache. Dieses Ergebnis zeigte sehr deutlich, dass absolute Werte für die Änderungen der Proteinabundanz nicht vergleichbar sind, wenn sie mit unterschiedlichen Färbe-/Auswertsystemen ermittelt wurden. Im Sinne konsistenter Daten sollte folglich stets das gleiche System verwendet werden.

Um Cy-Farbstoff-abhängige Variation auszuschließen, wurden die zwei Sätze aus den fünf Parallelgelen miteinander verglichen. Es zeigte sich, dass die Wahl des Farbstoffs für die Markierung einer bestimmten Probe keinen Einfluss auf die anschließend berechneten Abundanzunterschiede hat.

In einem weiteren Experiment wurde die Sensitivität der Cy-Farbstoffe mit der von Silber, SYPRO® Ruby und kolloidalem Coomassie verglichen. Dazu wurden pro Farbstoff drei parallele Gele untersucht, die alle mit der gleichen Proteinmenge (50 µg) beladen waren. Anschließend wurde die Anzahl der detektierten Spots verglichen. Sensitivität wurde hierbei als die maximale Anzahl detektierbarer Spots in einem 2DE Gel definiert. Wie erwartet, wurde die geringste Zahl an Proteinspots auf den Coomassie gefärbten Gelen detektiert, während Silber und SYPRO® Ruby die sensitivsten Färbemethoden darstellten. Mit den Cy-Farbstoffen konnten noch 90% der mit Silber detektierten Spots nachgewiesen werden (Publikation 2, Abb. 4).

Um den Einfluss der kovalent gebundenen Cy-Farbstoffe auf das Migrationsverhalten der markierten Proteine im Gel zu überprüfen, wurden 2D DIGE Gele postelektrophoretisch mit SYPRO® Ruby und Coomassie nachgefärbt. Es zeigte sich, dass

das Proteinspotmuster unter allen Färbemethoden nahezu identisch war (Publikation 3, Abb. 4). Dabei konnte beobachtet werden, dass einige Proteine nicht mit allen Farbstoffen nachgewiesen werden können. Dieses Phänomen ist bereits aus der Literatur bekannt (Görg et al. 2000). Im Fall der DIGE-Technologie können nur Proteine detektiert werden, die mindestens ein Lysin enthalten. Diese Grundvoraussetzung erfüllen ca. 96% aller annotierten Proteine im Genom von *R. baltica.*

Zudem erschienen einige wenige sehr prominente Spots in den DIGE Gelen (mit allen drei Cy-Farbstoffen) vertikal in die Länge gezogen. Hier lag die Vermutung nahe, dass diese Proteine einen überdurchschnittlich hohen Lysinanteil haben und trotz der Minimalmarkierung mehr als ein Lysinrest pro Proteinmolekül markiert wurde. Die Identifizierung der Proteine bestätigte diese Vermutung allerdings nicht. Die zwei plakativsten Beispiele sind in Abbildung 3 dargestellt. RB2627 enthält 21 und RB10727 19 Lysinreste. Die Spots beider Genprodukte erscheinen in die Länge gezogen, während RB8246 mit 18 und RB12844 mit 22 Lysinresten in der Peptidkette als kompakte Punkte im 2D DIGE Gel erscheinen. Eine Erklärung für dieses Phänomen ist bisher nicht gefunden worden. (Dieser Abschnitt enthält Daten, die bisher nicht veröffentlicht wurden.)

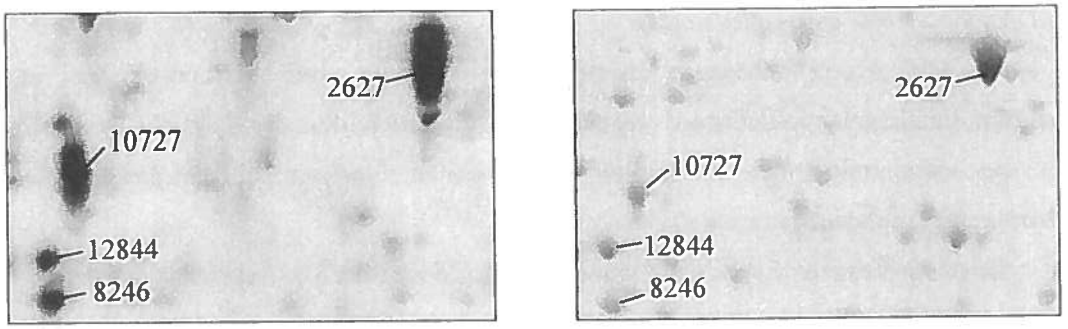



**Abb. 3** Ausschnitte von einem 2D DIGE Gel. Links der Cy3 Kanal, rechts der gleiche Bildausschnitt, des mit SYPRO® Ruby nachgefärbten Gels. Die im Text diskutierten Spots sind mit den RB Nummern gekennzeichnet.

Die Auswertung der Gele mit der DeCyder™ Software war, verglichen mit der Auswertung mit der Image Master Software, in einem deutlich kürzeren Zeitraum durchzuführen. Zusätzlich wurden alle Daten mit statistischen Werten unterlegt. Die Ergebnisse der DIGE Analysen zeigten, dass eine Anzahl von fünf parallelen Gelen ausreicht, um gute statistische Absicherungen der Abundanzunterschiede zu erreichen. Doch auch der Einsatz dieser Software erfordert manuelle Korrekturen beim Matchen der Parallelgele. Dabei ist das Maß der Eingriffsmöglichkeiten von Seiten des

Wissenschaftlers sehr beschränkt. Im Gegensatz zu anderen Programmen können Spotbegrenzungen nicht manuell erzeugt oder verändert werden. So bleibt das Maß an anwenderspezifischer Manipulation gering.

Aufgrund der Überlegenheit der DIGE-Technologie gegenüber anderen Färbe- und Quantifizierungsmethoden wurde diese Methode in allen weiteren Experimenten als Standardtechnik zur quantitativen Untersuchung differentieller Proteinmuster eingesetzt. Zur Identifizierung mittels MALDI-TOF-MS wurden die differentiell regulierten Proteine jedoch weiterhin aus Coomassie gefärbten Gelen ausgestochen. Der Grund dafür war das Fehlen eines Roboters für das automatische Ausstechen von Spots anhand der Koordinatenliste, welche bei der DeCyder™ Auswertung entsteht.

## 3. Physiologische Proteomanalysen bei *Rhodopirellula baltica*

Die Sequenzierung des Genoms von *R. baltica* stellte eine wichtige Voraussetzung für die erfolgreiche Proteomanalyse dieses Organismus dar, denn sie erlaubt die schnelle Identifizierung von 2DE-separierten Proteinen durch Abgleich von Peptidmassen-Fingerabdrücken (PMF). Das 7,145 Mb große zirkluläre Genom von *R. baltica* enthält 7325 vorhergesagte ORFs (Glöckner et al. 2003). Mit den 2384 Genen, denen aufgrund von Sequenzvergleichen mit bereits bekannten Proteinen eine Funktion zugewiesen werden konnte, war es möglich, den katabolen und anabolen Stoffwechsel von *R. baltica in silico* zu rekonstruieren (http://www.regx.de). Ungeklärt ist hingegen die Funktion von 4941 hypothetischen bzw. konserviert hypothetischen Genen. Über 50% des annotierten Genoms besteht aus diesen Genen, die keine Homologie zu einem Protein bekannter Funktion haben, und somit spezifisch für *R. baltica* zu sein scheinen. Diese Proteine sind möglicherweise in Prozesse wie Zelldifferenzierung und Zellentwicklung involviert, über die weder bei *R. baltica* noch bei anderen Planktomyceten molekulare Kenntnisse vorliegen. Die Herausforderung für die funktionelle Proteomanalyse besteht vor allem in der Klärung der Frage, welche Proteine zu welchem Zeitpunkt und unter welchen Umweltbedingungen gebildet werden. Im Fall der hypothetischen Proteine verspricht dieser Ansatz auch die Möglichkeit der Funktionszuweisung.

### *3.1 Trennung und Darstellung der löslichen Proteine von R. baltica*

In einer ersten Phase der Proteomanalyse bei *R. baltica* wurde eine 2DE Gelkarte der löslichen Proteine mit einem IEP zwischen pH 4 und pH 7 erstellt. Dazu wurden die löslichen Proteine unter Laborstandardbedingungen angezogener Kulturen auf einem 2D

Gel aufgetrennt, durch Coomassie Färbung visualisiert und anschließend mittels MALDI-TOF-MS und PMF identifiziert (Publikation 4). Es gelang, 626 Proteinspots zu identifizieren und diese 558 ORFs zuzuordnen. Die 626 Proteine wurden dann auf einem Mastergel gekennzeichnet. So entstand eine proteomische Landkarte, die stark vereinfacht den physiologischen „Grundzustand" der Zellen auf Proteinebene widerspiegeln sollte. Diese 2DE Gelkarte bildete eine breite Wissensbasis für weiterführende Studien, in denen die Veränderungen dieses Grundzustand-Proteoms als Folge externer Stimuli untersucht wurden.

Die 558 nicht redundanten Proteine konnten in drei Hauptkategorien aufgeteilt werden: Proteine, denen eine Funktion zugewiesen werden konnte (366), konservierte hypothetische Proteine (95), die Homologe in den Datenbanken aufweisen und hypothetische Proteine (97), die bisher nur von *R. baltica* bekannt sind. Die 366 Proteine, deren Funktion bekannt war, wurden zusätzlich in funktionelle Kategorien unterteilt (Abb. 4).

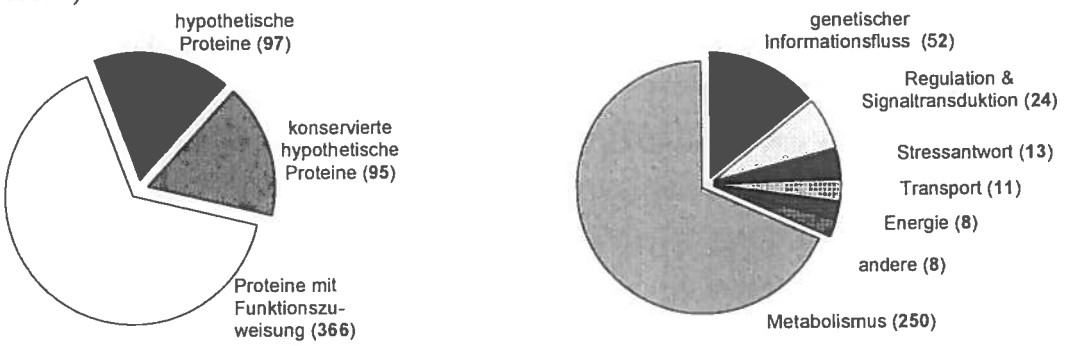



**Abb. 4** Gruppierung der identifizierten 558 nicht redundanten Proteine nach funktionellen Kategorien. Das rechte Diagramm stellt die Gesamtheit aller Proteine mit Funktionszuweisung dar, welche noch in weitere sieben Subkategorien unterteilt wurde.

Diese globale proteomische Studie lieferte einen ersten experimentellen Nachweis, welche vorhergesagten Gene in *R. baltica* tatsächlich exprimiert werden. Dies ist von besonderer Bedeutung für die vielen hypothetischen Proteine.

Die abundantesten Proteinspots auf der 2D Gelkarte repräsentierten Enzyme, denen Funktionen in der Aufrechterhaltung des Zentralstoffwechsels der Zelle (engl. housekeeping) zugewiesen werden konnten. Das Genprodukt von RB7247, ein Protein mit höchster Ähnlichkeit zu einer Glutamin-Synthetase II, war mit einem Gesamtproteinanteil von etwa 3,4% das abundanteste Protein im Mastergel. Das Enzym bildete drei Spots auf dem Gel, die sich in IEP und Molekulargewicht unterschieden. Generell waren fast alle

abundanten Spots auf dem 2D Gel auf der Basis des „Codon usage“ (Karlin und Mrázek 2000) auch als hochexprimiert (PHX) vorhergesagt.

Die Zuordnung der 626 Proteinspots zu 558 ORFs entspricht etwa 1,12 Proteinspots pro Gen. Hinweise auf posttranslationale Modifikationen wurden vor allem im hochmolekularen Bereich gefunden, wo identifizierte Proteine in mehreren Spots mit gleichem Molekulargewicht aber unterschiedlichen IEPs vorkamen. In mindestens drei Fällen konnte durch die Färbung mit einem phosphoproteinspezifischen Fluoreszenzfarbstoff nachgewiesen werden, dass es sich hierbei um Phosphorylierungen handelt (Publikation 5, Abb. 3).

Die Zellen von *R. baltica* weisen eine ähnliche intrazelluläre Kompartimentierung auf, wie sie auch von anderen beschriebenen Planktomyceten bekannt ist (Lindsay et al. 1997, 2001; Publikation 3, Abb. 1). Das als Pirellulosom bezeichnete innerste Kompartiment ist der Ort der Proteinbiosynthese. Es enthält das Riboplasma mit dem kondensierten Nucleoid und den Ribosomen und ist durch eine einschichtige Membran vom äußeren so genannten Paryphoplasma getrennt. Nach ihrer Synthese im Pirellulosom müssen die Proteine durch gezielte Translokation in die verschiedenen intrazellulären Kompartimente verteilt werden. Gemäß der Signalpeptidhypothese trägt die Mehrzahl aller sekretierten Proteine am N-Terminus ein Signalpeptid (Blobel 2000). Aufgrund des Vorhandenseins der intracytoplasmatischen Kompartimente ist es nicht verwunderlich, dass 16% der Proteine von *R. baltica* ein Signalpeptid tragen. Von den 558 identifizierten Proteinen des Mastergels wurde für 146 (26%) ein Signalpeptid vorhergesagt. Da beim Zellaufschluss das Riboplasma nicht vom Paryphoplasma getrennt wurde, sind diese Proteine vermutlich entweder im Paryphoplasma lokalisiert oder auf dem Weg in die extrazelluläre Matrix. Enzyme mit vermeintlicher „Housekeeping-Funktion“ trugen kein Signalpeptid, was darauf hindeutet, dass der Zentralstoffwechsel tatsächlich im Pirellulosom stattfindet. Unter den Signalpeptid-tragenden Proteinen, denen eine Funktion zugewiesen werden konnte, befinden sich Enzyme zum Abbau extrazellulärer Makromoleküle, oder Signalproteine, wie z. B. 1,4-β-Xylanase bzw. Phosphoprotein-Phosphatasen.

In einem virtuellen 2D Gel wurden jedem im *R. baltica* Genom vorhergesagten Protein aufgrund seines theoretischen IEPs und Molekulargewichts eine Position im Gel zugewiesen (Abb. 5). Dabei stellte sich eine Unterteilung des theoretischen Proteoms in drei Bereiche heraus. Die Proteine bilden Cluster im sauren Bereich um pH 5, im neutralen Bereich um pH 6,5 und im alkalischen Bereich zwischen pH 8 und pH 12.

Computergenerierte 2D Gele bakterieller und archaeeller Gesamtproteome weisen typischerweise Proteincluster im alkalischen und im sauren Bereich auf, während nahezu keine Proteine im neutralen Bereich zu finden sind (VanBogelen et al. 1999; Schwartz et al. 2001). Der Grund hierfür liegt darin, dass im Cytoplasma ein nahezu neutraler pH-Wert herrscht und Proteine nahe ihres IEPs die geringste Löslichkeit aufweisen. Im Gegensatz dazu existiert in eukaryotischen Proteomen, ebenso wie bei *R. baltica,* ein drittes Proteincluster im neutralen Bereich. Es wird vermutet, dass die Unterteilung in zwei bzw. drei Cluster die subzelluläre Lokalisation von Proteinen reflektiert. Während cytoplasmatische Proteine einen IEP um pH 5 aufweisen und der IEP integraler Membranproteine um pH 9 liegt, haben bei Eukaryoten die Proteine aus dem Zellkern einen IEP um pH 7 und bilden daher im 2D Gel ein Cluster im neutralen pH-Bereich (Schwartz et al. 2001). Ob die Dreiteilung des Gesamtproteoms in *R. baltica* auf die Lokalisation der Proteine in der Zelle zurückzuführen ist, muss allerdings noch durch weitere Untersuchungen geklärt werden. Insbesondere im Zusammenhang mit den Signalpeptiden kommt künftigen Untersuchungen der Subproteome von Pirellulosom, Paryphoplasma und Sekretom große Bedeutung zu.

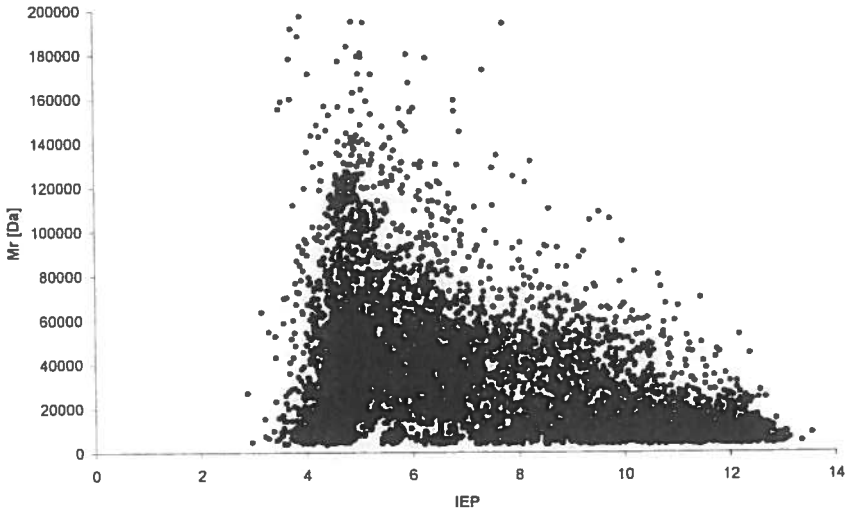



**Abb. 5** Theoretisches 2D Gel des Gesamtproteoms von *R. baltica.* Die theoretischen Molekularmassen aller 7325 annotierten Proteine wurden gegen ihren theoretischen IEP aufgetragen.

Schließlich gelang es, durch Verwendung eines maximalen ORF-Satzes, der alle theoretisch möglichen ORFs von *R. baltica* enthält, vier neue ORFs zu identifizieren. Diese Gene waren bis dato nicht im veröffentlichen Datensatz des Genoms enthalten. Die

neu identifizierten Gene kodieren Proteine mit bisher unbekannter Funktion. Das Ergebnis dieses Versuchs zeigte, dass die Genomik in der ORF-Vorhersage effektiv durch die Proteomik unterstützt und verbessert werden kann.

## 3.2 Rekonstruktion der zentralen Wege des Zuckerstoffwechsels in *R. baltica*

Aufgrund der weitreichenden Verbreitung der Planktomyceten in marinen Habitaten und ihrer Assoziation mit absinkenden phytodetritischen Aggregaten („marine snow") (DeLong 1993) wird den Planktomyceten eine Schlüsselrolle beim Abbau von biologischen Polymeren und damit im globalen Kohlenstoffkreislauf zugeschrieben (Fuerst 1995). Da Kohlenhydrate den Hauptbestandteil der Biomasse darstellen, ist ihre Remineralisierung für die Stoffkreisläufe von elementarer Bedeutung. In dieser Arbeit wird *R. baltica* als Modellorganismus für den in der Natur wichtigen mikrobiologischen Abbau von Kohlenhydraten untersucht. Dazu wurde eine Reihe von physiologischen Anpassungsexperimenten durchgeführt, mit denen die Regulation kataboler Stoffwechselwege untersucht werden sollte.

Untersuchungen des Substratspektrums von *R. baltica* ergaben, dass der Organismus überwiegend Zucker und Zuckerderivate als Kohlenhydrat- und Energiequellen nutzt, während Aminosäuren, organische Säuren und Alkohole nicht verwertet werden (Publikation 3).

Zu den 558 identifizierten Proteinspots der 2D Gelkarte zählten auch viele Enzyme der Glykolyse, des oxidativen Pentosephosphat-Zyklus und des TCA-Zyklus. Auf der Basis dieses Mastergels konnte nunmehr auch auf Proteinebene eine Rekonstruktion dieser zentralen katabolen Zuckerstoffwechselwege vorgenommen werden (Abb. 6). Die quantitative Auswertung von 2D DIGE Gelen belegte, dass unter anderem Enolase, Glycerinaldehyd-3-phosphat-Dehydrogenase, Malat-Dehydrogenase und Transaldolase zu den abundantesten Proteinen auf den 2D Gelen gehörten. Ein Befund, der sich mit der bioinformatischen Expressionsvorhersage für diese Enzyme bei *R. baltica* deckt. Auch in vorangegangenen proteomischen Studien bei anderen Bakterien gehörten Enzyme des Zentralstoffwechsels zu den abundantesten auf 2D Gelen (Büttner et al. 2001; Liao et al. 2003).

Um den aus der Proteomanalyse abgeleiteten Enzymen des Kohlenhydratabbaus Aktivitäten zuzuweisen, wurden die spezifischen Aktivitäten ausgewählter Schlüsselenzyme des zentralen Zuckerstoffwechsels in Zellextrakten unterschiedlicher

adaptierter Anzuchten gemessen. (Die Enzymtests wurden von T. Stührmann durchgeführt.) Dazu wurden Zellen von *R. baltica* in definiertem Mineralmedium an das Wachstum auf Ribose, Xylose, Glucose, *N*-Acetylglucosamin, Lactose, Maltose, Melibiose bzw. Raffinose als alleinige Kohlenstoff- und Energiequelle adaptiert. Anschließend wurden die spezifischen Aktivitäten der folgenden Schlüsselenzyme in den verschiedenen Rohextrakten gemessen: 6-Phosphofruktokinase, Glycerinaldeyd-3-phosphat-Dehydrogenase und Enolase (Glykloyse); Transaldolase (oxidativer Pentosephosphat-Zyklus), Isocitrat-Dehydrogenase und Malat-Dehydrogenase (TCA-Zyklus). Es konnte gezeigt werden, dass die Gykolyse, der oxidative Pentosephosphat-Zyklus und der TCA-Zyklus unter allen getesteten Anzuchtbedingungen funktionstüchtig waren. Weiterhin unterschieden sich die gemessenen spezifischen Aktivitäten der Enzyme kaum unter den verschiedenen Anzuchtbedingungen.

Eine Besonderheit stellt die pyrophosphatabhängige Phosphofruktokinase in *R. baltica* dar. Die bioinformatische Vorhersage dieses Enzyms konnte durch die enzymatischen Tests bestätigt werden. Die Aktivität der Phosphofruktokinase konnte ausschließlich mit Pyrophosphat gemessen werden, während sowohl ADP als auch ATP nicht als Phosphoryldonor dienen konnten. Darüber hinaus besitzt *R. baltica* zwei pyrophosphatabhängige Phosphofruktokinasen in seinem Genom, die beide konstitutiv exprimiert werden (Publikation 5).

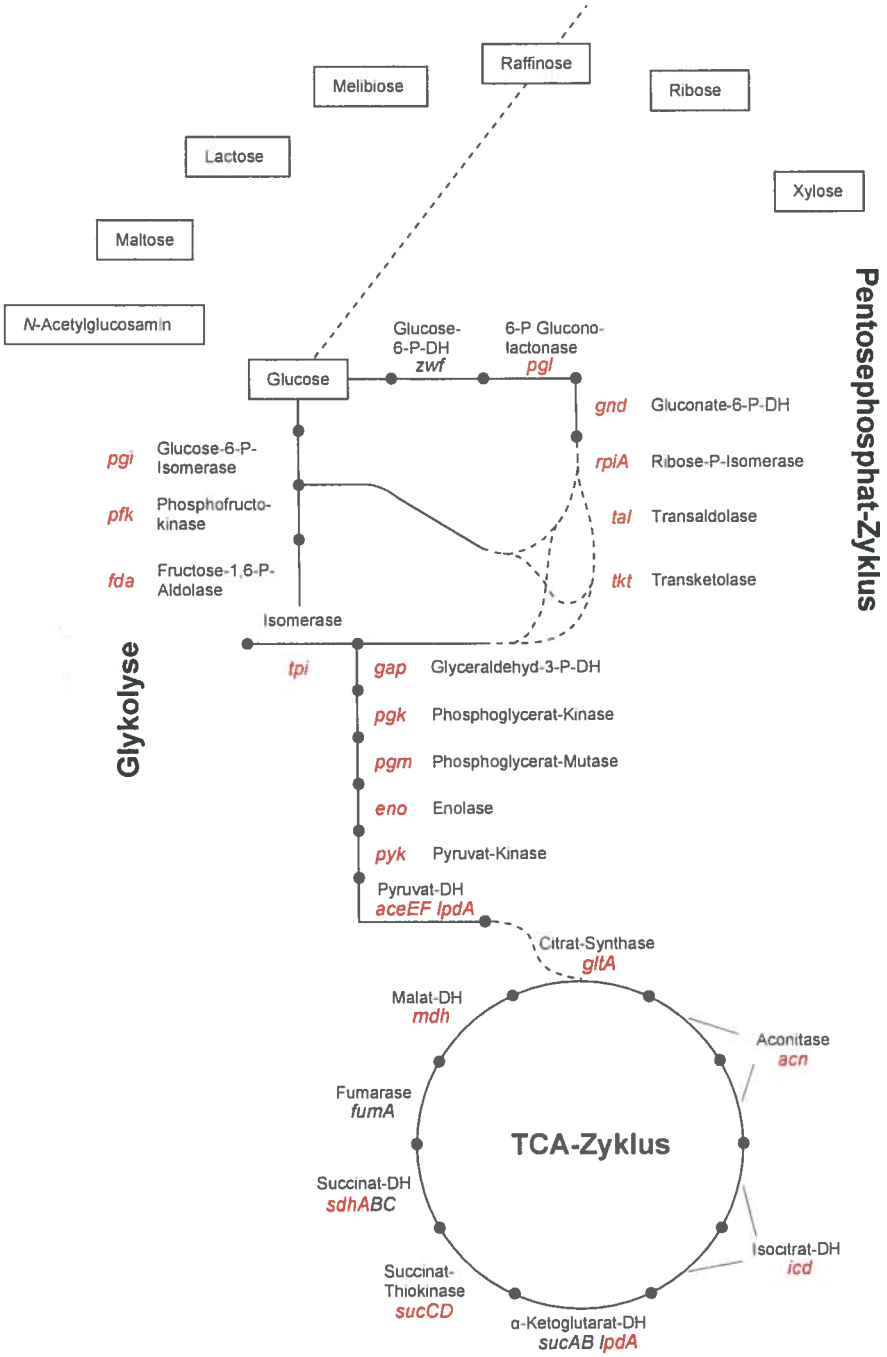



**Abb. 6** Metabolische Rekonstruktion des zentralen Kohlenhydratstoffwechsels in *R. baltica*. Die Genprodukte der in rot dargestellten Gene konnten anhand der 2D Gelelektrophorese separiert und mittels MALDI-TOF-MS und PMF identifiziert werden.

## 3.3 Untersuchungen zur Regulation des Zuckerstoffwechsels in *R. baltica*

Beim Abbau von Kohlenhydraten nehmen *R. baltica* und seine Verwandten im natürlichen Habitat vermutlich eine Schlüsselstellung ein. Das in der Natur am häufigsten vorkommende Polysaccharid Chitin unterstützt etwa 10% der bakteriellen Produktion in marinen Systemen (Kirchmann et al. 1999). *N*-Acetylglucosamin, das Monomer des Chitins, kann von *R. baltica* als Stickstoff- und auch als Kohlenstoffquelle verwertet werden (Publikation 3).

Differentielle physiologische und proteomische Studien wurden bei *R. baltica* erstmals am Beispiel des *N*-Acetylglucosaminstoffwechsels etabliert (Publikation 1). Erste Untersuchungen von dessen Regulation wurden mit Zellsuspensionen von Kulturen durchgeführt, die an das Wachstum mit Glucose bzw. mit *N*-Acetylglucosamin angepasst waren. Wurde solchen Suspensionen eine äquimolare Mischung aus Glucose und *N*-Acetylglucosamin angeboten, vermochte *R. baltica* beide Substrate gleichzeitig zu verwerten. Von Standardbakterien, wie z. B. *E. coli*, ist hingegen seit langem bekannt, dass die Verwertung verschiedenster Kohlenhydrate (z. B. auch *N*-Acetylglucosamin) in Gegenwart von Glucose, dem bevorzugten Substrat, reprimiert wird. Dieses Ernährungsverhalten wird auf molekularer Ebene durch den Prozess der Katabolitrepression hervorgerufen. In *E. coli* werden die spezifischen Enzyme des *N*-Acetylglucosaminabbaus im *nag*-Operon kodiert. Die Expression dieses *nag*-Operons wird in Gegenwart von *N*-Acetylglucosamin durch einen Komplex aus cyclischem AMP und cAMP receptor protein (CRP) induziert. Bei gleichzeitigem Vorhandensein der bevorzugten Glucose wird die Expression des *nag*-Operons jedoch mittels des NagC Repressors reprimiert (Katabolitrepression). Dies ist für die Zelle ökonomischer, da eine energetisch kostspielige Synthese von *N*-Acetylglucosamin-abbauenden Enzymen in Gegenwart der leichter verwertbaren Glucose vermieden wird. *R. baltica* ist ein langsam wachsendes Umweltbakterium, das im Gegensatz zu *E. coli* die überwiegende Zeit unter substratlimitierenden Bedingungen lebt. Daher erscheint eine strikte Regulation des Kohlenhydratkatabolismus als weniger sinnvoll. Vielmehr verwertet *R. baltica* verfügbare Zucker gleichzeitig.

Entsprechend dieser Ernährungsstrategie verwerteten Zellsuspensionen von *R. baltica* *N*-Acetylglucosamin ohne merkliche Adaptionsphase, unabhängig davon auf welchem der beiden Zucker die vorausgegangene Adaptation erfolgte. Untersuchungen auf proteomischer Ebene zeigten jedoch, dass *N*-Acetylglucosamin-spezifische Proteine in 2D

Gelen der an Glucose adaptierten Zellen nicht nachzuweisen waren. Dies belegt die Induzierbarkeit kataboler Gene und deutet darauf hin, dass die Neusynthese substratspezifischer Enzyme offenbar schneller abläuft, als das zeitliche Auflösungsvermögen der Suspensionsexperimente zu detektieren vermag (Publikationen 1 und 2).

Um die substratabhängige Regulation des Zuckerkatabolismus in *R. baltica* auf breiter Basis zu untersuchen, wurden Kulturen auf das Wachstum mit verschiedenen Zuckern adaptiert und die substratspezifischen Proteinmuster mit der 2D DIGE-Technologie untersucht (Publikation 5). Dazu wurden die Zellen in definiertem Mineralmedium angezogen, welches als Stickstoffquelle 1 mM Ammoniumchlorid und als Kohlenstoffquelle jeweils 10 mM der folgenden Zucker enthielt: Glucose, *N*-Acetylglucosamin, Xylose, Ribose, Maltose, Lactose, Melibiose und Raffinose. *N*-Acetylglucosamin diente gleichzeitig als Kohlenstoff- und Stickstoffquelle, daher wurde in diesem Fall kein Ammoniumchlorid zugesetzt. Ziel dieses Versuchs war es, substratspezifisch gebildete Enzyme zu identifizieren, die für die hinführenden (peripheren) Reaktionen des jeweiligen Zuckers zu den gemeinsamen, zentralen Abbauwegen (Glykolyse oder oxidativer Pentosephosphat-Zyklus) verantwortlich sind (Abb. 6). Proteinextrakte der auf Glucose angezogen Zellen dienten hierbei als Referenzzustand, mit dem die Proteinmuster der anderen sieben Anzuchten verglichen wurden.

Die Ergebnisse dieser vergleichenden 2D DIGE Analysen zeigten, dass in Abhängigkeit der zur Anzucht verwendeten Zucker die Bildung bestimmter Proteine induziert wurde. Die abundantesten substratspezifischen Proteine wurden aus parallelen mit Coomassie gefärbten Gelen ausgestochen und anhand ihrer PMFs identifiziert. Zunächst wurden mögliche Funktionen der identifizierten Proteine durch Homologiesuche ermittelt. In einem zweiten Schritt wurde die genetische Umgebung der kodierenden Gene auf weitere Hinweise zum Kohlenhydratstoffwechsel untersucht. Im Folgenden werden die Befunde nach Zuckersubstrat geordnet vorgestellt.

Für **Ribose** konnte gezeigt werden, dass zwei der vier substratspezifischen und identifizierten Proteinspots Produkte eines einzigen Gens (RB3488) waren. Hierbei muss es sich um eine posttranslationale Modifikation handeln, die ausschließlich eine kleine Verschiebung des IEPs hervorruft (Publikation 5, Abb. 2). Dieses Protein weist große Ähnlichkeit zu einer zinkabhängigen Alkohol-Dehydrogenase auf. Die beiden anderen Ribose-spezifischen Proteine sind Paraloge (RB3504, RB5310), denen bisher keine

Funktion zugewiesen werden konnte. Für RB3504 und RB3488 weist der genomische Kontext auf eine Funktion im Katabolismus von Ribose hin (Publikation 5, Abb. 4 A). Die ORFs flankieren die in einem Cluster angeordneten Gene eines Ribose ABC-Transporters sowie der Ribokinase. Das genetische Umfeld von RB5310 deutet allerdings nicht auf eine Beteiligung dieses Proteins im Ribosestoffwechsel hin. In dieser Studie konnten weder die Ribokinase noch lösliche Untereinheiten des Ribose-spezifischen ABC-Transporters auf den 2D Gelen identifiziert werden.

Bei Wachstum mit ***N*-Acetylglucosamin** konnten die zwei am stärksten induzierten Proteine als eine Dehydrogenase (kodiert von RB3330) und ein hypothetisches Protein unbekannter Funktion (kodiert von RB3342) identifiziert werden. Beide ORFs liegen in unmittelbarer Nachbarschaft zu einem ORF, welcher für eine Glucosamin-6-phosphat-Isomerase kodiert (Publikation 5, Abb. 4 C). Anders als bei *E. coli* sind die Gene für den Abbau von *N*-Acetylglucosamin im Genom von *R. baltica* nicht in einem Operon organisiert, sondern über das gesamte Genom verstreut. Obwohl die beiden Schlüsselenzyme des *N*-Acetylglucosaminabbaus (Glucosamin-6-phosphat-Deacetylase und Glucosamin-6-phosphat-Isomerase) bisher nicht auf den 2D Gelen identifiziert werden konnten, weist das genetische Umfeld der induzierten Genprodukte auf ihre Bedeutung im *N*-Acetylglucosaminabbau hin. Ebenso wurde die Expression eines weiteren hypothetischen Proteins (RB5240) induziert, dessen genetische Nachbarschaft jedoch keinen Aufschluss über die Funktion dieses Proteins im *N*-Acetylglucosaminkatabolismus gibt.

Ein wichtiger Befund war die spezifische Induktion einer Glutamin-Synthetase II (RB7247) in *N*-Acetylglucosamin-adaptierten Zellen. Wie bereits in Kapitel 3.1 erwähnt, bildete das Produkt dieses ORFs drei abundante Spots auf den 2D Gelen. Interessanterweise wird bei Wachstum mit *N*-Acetylglucosamin einer dieser drei Proteinspots spezifisch hochreguliert, d. h., er ist in den sechs anderen Anzuchtbedingungen relativ zur Glucoseanzucht herunterreguliert (zu erkennen an der grünen Farbe des Spots in Publikation 5, Abb. 2). Im Gegensatz dazu ist ein anderer der drei Proteinspots dieses ORFs bei Anzucht mit Xylose, Maltose, Lactose, Melibiose und Raffinose hoch abundant. Der dritte Proteinspot von RB7247 zeigte hingegen unter allen untersuchten Anzuchtbedingungen eine gleichbleibende Abundanz. In ihrem elektrophoretischen Laufverhalten unterscheiden sich die drei Proteinspots von RB7247 vor allem durch ihre IEPs. Daher kam eine posttranslationale Phosphorylierung der Glutamin-Synthetase II als Ursache in Frage. Diese Möglichkeit wurde überprüft, indem

2D Gele von *N*-Acetylglucosamin-Anzuchten mit dem phosphoproteinspezifischen Fluoreszenzfarbstoff Pro-Q® Diamond angefärbt wurden. Anschließend wurden die gleichen Gele mit SYPRO® Ruby nachgefärbt, um die Gesamtheit der Proteine im Gel zu visualisieren. Überraschend zeigte sich, dass sowohl die unter *N*-Acetylglucosamin-induzierte Proteinspezies als auch die konstitutiv exprimierte Proteinspezies phosphoryliert sind. Eine unabhängige Bestätigung dieser posttranslationalen Modifikation durch massenspekrometrische Methoden war im Rahmen der vorliegenden Arbeit nicht mehr möglich.

*R. baltica* besitzt nur eine Glutamin-Synthetase, welche zum Typ II gehört. Untersuchungen der Glutamin-Synthetase II bei *Rhizobium melioti* (*Sinorhizobium melioti*) zeigten, dass dieses Enzym auf transkriptioneller Ebene reguliert wird und dass seine Transkription bei Ammoniummangel im Medium induziert wird (Shatters et al. 1989). Über eine posttranslationale Modifikation und deren Funktion ist bisher nichts bekannt. Mit Ausnahme der *N*-Acetylglucosamin-Anzuchten enthielt das Medium 1 mM Ammoniumchlorid als Stickstoffquelle. Eine Besonderheit von *R. baltica* und den meisten Planktomyceten ist es, ihren Stickstoffbedarf aus dem *N*-Acetylglucosamin allein decken zu können. Ob die unterschiedlich stark abundanten Proteinspezies der Glutamin-Synthetase II durch die verschiedenen Mengen an frei verfügbarem Ammonium in den Medien bedingt sind, muss durch weitere Versuche geklärt werden.

Es wurden vier Proteinspots identifiziert, welche spezifisch in **Xylose** adaptierten Zellen induziert waren. Zwei der Proteinspots konnten einem ORF (RB9584) zugeordnet werden. Die mögliche posttranslationale Modifikation hatte auch hier nur einen Effekt auf den IEP. RB9584 kodiert für ein Protein mit großer Ähnlichkeit zu einer zinkabhängigen Chinon-Oxidoreduktase und befindet sich direkt neben dem ebenfalls induzierten RB9586, welches für eine weitere Oxidoreduktase kodiert. Der genetische Kontext der beiden ORFs ergab hierbei keinen Hinweis auf eine mögliche Verknüpfung der Genprodukte mit dem Kohlenhydratstoffwechsel (Publikation 5, Abb. 4 B). Des Weiteren wurde eine Carboxypeptidase (RB2261) unter diesen Anzuchtbedingungen spezifisch induziert. Auch hier ergab der Genkontext keinen Aufschluss über eine Bedeutung dieses Proteins im Zuckerstoffwechsel. Acht weitere hochregulierte Proteine wurden identifiziert, welche jedoch nicht spezifisch für das Wachstum auf Xylose waren, da sie ebenfalls unter anderen Anzuchtbedingungen induziert waren (Tabelle 1). Für vier dieser Proteine (RB1828, RB2901, RB11008, RB11010) war eine Funktionszuweisung nicht möglich. Zudem wurden eine putative Zuckerphosphat Epimerase/Isomerase, die auch bei Wachstum auf

Maltose, Melibiose und Raffinose vermehrt exprimiert war, und die Untereinheit eines ABC-Transporters identifiziert. Da beide Proteine auch bei Wachstum auf anderen Zuckern heraufreguliert waren, könnte es sich um ein allgemeines Enzym des Zuckerabbaus und einen unspezifischen Zuckertransporter handeln.

Ähnlich wie die Gene des *N*-Acetylglucosaminstoffwechsels liegen auch die vorhergesagten Gene des Xylosestoffwechsels bei *R. baltica* verstreut im Genom. Die zwei Enzyme, welche die ersten Abbauschritte der Xylose bis zur Einschleusung in den Pentosephosphat-Zyklus als Xylulosephosphat katalysieren, sind die Xylose-Isomerase und die Xylulosekinase. Während Erstere als konstitutiv unter allen untersuchten Anzuchtbedingungen nachgewiesen wurde, konnte Letztere aufgrund eines IEPs von 7,7 nicht auf den pH 4 – 7 Gelen detektiert werden.

Für das Wachstum mit **Maltose** konnten neun regulierte Proteine identifiziert werden. Im Gegensatz zur Situation bei den zuvor besprochenen Monosacchariden ist keines dieser Proteine Maltose-spezifisch, sondern auch unter anderen Anzuchtbedingungen heraufreguliert. Fünf dieser Proteine sind als hypothetisch (RB2901, RB11008) bzw. konserviert hypothetisch (RB1828, RB4658, RB11010) annotiert. Der genetische Kontext dieser fünf Gene implizierte keine Beteiligung ihrer Produkte im Zuckerstoffwechsel. Die Induktion eines Gens (RB11021), dessen Protein für das ATP-Bindeprotein eines ABC-Transporters kodiert, lässt vermuten, dass Maltose über diesen Weg in die Zellen aufgenommen wird. Das gleiche Protein ist auch bei Wachstum auf Xylose und Raffinose induziert.

Für den Abbau von **Lactose** wurde im Genom von *R. baltica* das bei *E. coli* bekannte *lac*-Operon nicht nachgewiesen. Jedoch wurde für RB3405 die Funktion einer β-Galactosidase vorhergesagt. Da dieses Protein nur eine wenig ausgeprägte Ähnlichkeit zu einem kryptischen Gen der α-Untereinheit einer *E. coli* β-Galactosidase besitzt, ist diese Vorhersage kritisch zu betrachten. Die spezifische Aktivität dieses Enzyms konnte in den Rohextrakten der auf Lactose angezogenen Zellen mit den herkömmlichen Tests nicht bestimmt werden. (Der Versuch wurde von T. Stührmann durchgeführt.) Dieses Ergebnis deutet darauf hin, dass *R. baltica* einen neuen Typ von β-Galactosidase besitzt und dass dieses Enzym möglicherweise nach einem bislang unbekannten Mechanismus arbeitet. Interessant ist, dass *R. baltica* alle Enzyme zum weiteren Abbau der durch die Spaltung der Lactose freigesetzten Galactose fehlen. Dieser Befund war von besonderer Bedeutung, weil neben der Lactose auch Melibiose und das Trisaccharid Raffinose je ein Molekül Galactose enthalten und Galactose ein gutes Wachstumssubstrat für *R. baltica* darstellt.

Somit konnte unter den fünf identifizierten Proteinspots die in Lactose adaptierten Zellen induziert sind, neben der bereits diskutierten Glutamin-Synthetase II (RB7247), einer Arylsulfatase (RB9498) und zwei hypothetischen Proteinen (RB2901, RB 4658), deren Genkontext keinerlei Auskunft über die Funktionen der Proteine gab, nur eine NADH-abhängige Oxidoreduktase (RB10503) identifiziert werden.

Obwohl im Genom von *R. baltica* keine Transporter für die spezifische Aufnahme von **Melibiose** und keine Enzyme zu dessen Abbau vorhergesagt sind, stellt Melibiose ein gutes Wachstumssubstrat für *R. baltica* dar. Sechs Proteine wurden identifiziert, die in Melibiose-adaptierten Zellen hochreguliert waren. Darunter befanden sich erneut die bereits erwähnten Glutamin-Synthetase II, Zuckerphosphat Isomerase/Epimerase (RB4654) und die NADH-abhängige Oxidoreduktase (RB10503). Die ORFs der drei anderen regulierten Proteine (RB1828, RB2901, RB11008) sind bisher ohne Funktionszuweisung und auch in den Anzuchten auf Xylose, Maltose und Raffinose heraufreguliert.

**Raffinose** besteht aus einem Molekül Melibiose und einem Molekül Fructose. Für die Verwertung der gebildeten Melibiose fehlt wiederum die Melibiase, welche die Spaltung der $\alpha$ 1 $\rightarrow$ 6 glycosidischen Bindung zwischen der Galactose und der Glucose katalysiert. Von den bei Wachstum mit Raffinose induzierten Proteinen wurden zehn identifiziert. Lediglich eins von ihnen (die Untereinheit eines ABC Transporter, RB11030) wurde substratspezifisch gebildet. Möglicherweise wird die Raffinose spezifisch über diesen Transporter in die Zellen aufgenommen. Allerdings war unter den anderen neun Proteinen die Untereinheit eines weiteren ABC-Transporters. Beide Transportproteine weisen untereinander eine Sequenzähnlichkeit von 53% auf.

**Tabelle 2.** Relative Abundanzen der Proteine, welche unter den angegebenen Zuckeranzuchten im Vergleich zur Glucose-Anzucht hochreguliert werden. Die kodierenden Gene und die durch BLASTP-Analysen vorhergesagten Funktionen sind dargestellt.

| ORF | vorhergesagte Funktion | Average ratio | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Raffinose | Melibiose | Lactose | Maltose | NAG | Xylose | Ribose |
| 1828 | konserviertes hypothetisches Protein | 2,9 | 2,7 | | 3,7 | 1,9 | 3,9 | |
| 2261 | Carboxypeptidase | | | | | | 2,0 | |
| 2901 | hypothetisches Protein | 8,3 | 4,1 | 2,7 | 7,5 | 1,9 | 4,6 | |
| 3330 | Dehydrogenase | | | | | 43,9 | | |
| 3330 | Dehydrogenase | | | | | 15,7 | | |
| 3342 | konserviertes hypothetisches Protein | | | | | 32,4 | | |
| 3488 | Zn-abhängige ADH (Sorbitol DH) | | | | | | | 43,4 |
| 3488 | Zn-abhängige ADH (Sorbitol DH) | | | | | | | 12,5 |
| 3504 | konserviertes hypothetisches Protein | | | | | | | 8,1 |
| 4654 | Zuckerphosphat-Isomerase/Epimerase | 9,0 | 2,5 | | 8,1 | | 4,3 | |
| 4658 | konserviertes hypothetisches Protein | 4,4 | | 9,8 | 6,4 | | | |
| 5240 | hypothetisches Protein | | | | | 4,5 | | |
| 5310 | konserviertes hypothetisches Protein | | | | | | | 23,6 |
| 7247 | Glutamin-Synthetase II | | | | | 1,8 | | |
| 7247 | Glutamin-Synthetase II | 3,0 | 2,8 | 5,1 | 2,8 | | 2,9 | |
| 9498 | Arylsulfatase | | | 2,1 | | 1,8 | | |
| 9527 | hypothetisches Protein | | | | | 1,7 | | |
| 9584 | Zn-abhängige Chinon-Oxidoreduktase | | | | | | 19,8 | |
| 9584 | Zn-abhängige Chinon-Oxidoreduktase | | | | | | 4,2 | |
| 9586 | Oxidoreduktase | | | | | | 13,2 | |
| 10503 | NADH-abhängige Oxidoreduktase | | 2,1 | 2,3 | | | | |
| 11008 | hypothetisches Protein | 9,0 | 2,5 | | 8,1 | | 4,3 | |
| 11010 | konserviertes hypothetisches Protein | 3,4 | | | 4,8 | | 2,1 | |
| 11021 | ABC-Transporter (ATP-Bindeprotein) | 3,9 | | | 4,3 | | 2,7 | |
| 11030 | ABC-Transporter (ATP-Bindeprotein) | 4,9 | | | | | | |
| 12362 | Peroxiredoxin | 3,9 | | | 5,8 | 2,9 | 2,9 | |

Zusammenfassend lassen sich die regulierten Gene hinsichtlich ihrer genomischen Umgebung in zwei Kategorien einteilen. Die eine Kategorie umfasst Gene in der genomischen Nachbarschaft weiterer potentieller Gene für den Abbau des untersuchten Substrats. Beispiele dafür waren die Ribose- bzw. *N*-Acetylglucosamin-spezifischen Proteine. Die andere Kategorie umfasst Gene, welche offenbar zufällig über das gesamte Genom verteilt sind und deren genomische Kontexte keinen Hinweis auf die Funktion der Proteine im Zuckerstoffwechsel geben. Beispiel dafür waren Xylose- bzw. Raffinose-spezifische Proteine.

Während die Gene für bestimmte Abbauwege in *E. coli* oder *Bacillus subtilis* in Operons organisiert sind, ist dies in *R. baltica* nur selten der Fall. Vielmehr sind in *R. baltica* die Gene eines Stoffwechselwegs oftmals über das gesamte Genom verteilt (Glöckner et al. 2003). Möglicherweise hängt dies mit der überdurchschnittlichen Größe des Genoms zusammen. Neuere Untersuchungen deuten darauf hin, dass mit wachsender Genomgröße die Anzahl der in Operons organisierten Gene abnimmt (Ermolaeva et al. 2001). In der vorliegenden Arbeit konnte gezeigt werden, dass solche im Genom verstreuten Gene funktionell verbunden sein können und auch gemeinsam reguliert werden. Über die zugrunde liegenden Regulationsmechanismen ist bisher nichts bekannt. Zusätzlich konnte gezeigt werden, dass *R. baltica* auf Substraten wie Raffinose oder Melibiose wächst, obwohl die bioinformatische Analyse entsprechende Abbauwege nicht in der Genomsequenz vorhergesagt hat. Darüber hinaus konnten für diese Substrate spezifisch induzierte hypothetische Proteine identifiziert werden. Da ihnen bislang keine Funktion zugeordnet werden konnte, besteht die Möglichkeit, dass sie neuen Reaktionswegen im peripheren Kohlenhydratstoffwechsel von *R. baltica* angehören.

## 3.4 Untersuchungen der Wachstumsphasen und des Lebenszyklus bei R. baltica

### 3.4.1 Differentielle Proteinbildung im Verlauf einer Wachstumskurve

Der Lebenszyklus von *R. baltica* zeichnet sich durch das Alternieren von sessilen knospenden Mutterzellen und mobilen Schwärmerzellen aus und ähnelt darin dem Lebenszyklus von *Caulobacter crescentus* (Jacobs-Wagner 2004). Dieses Süßwasserbakterium gilt als Modellorganismus für zelluläre Entwicklung bei Bakterien. Im Labor konnte gezeigt werden, dass die zwei Morphotypen von *R. baltica* in unterschiedlichen Phasen der Wachstumskurve einer diskontinuierlichen Zellkultur dominieren. Während Schwärmer und Einzelzellen hauptsächlich in der exponentiellen

Wachstumsphase zu finden sind, können die sessilen, Rosetten-bildenden Zellen dem Ende der exponentiellen Phase und der stationären Phase zugeordnet werden. Über den Ablauf und die Kontrolle der Zelldifferenzierung ist bisher nur bekannt, dass in *R. baltica* andere Mechanismen verwirklicht sein müssen als in *Caulobacter*. Das CtrA-Protein ist bei *Caulobacter* für die zentrale Kontrolle des Zellteilungszyklus verantwortlich. Ein Homolog von CtrA wird im Genom von *R. baltica* nicht kodiert. Des Weiteren sind bis auf *ftsK* auch keine Gene vorhanden, die für bekannte Wege der Zellteilung benötigt werden (Glöckner et al. 2003).

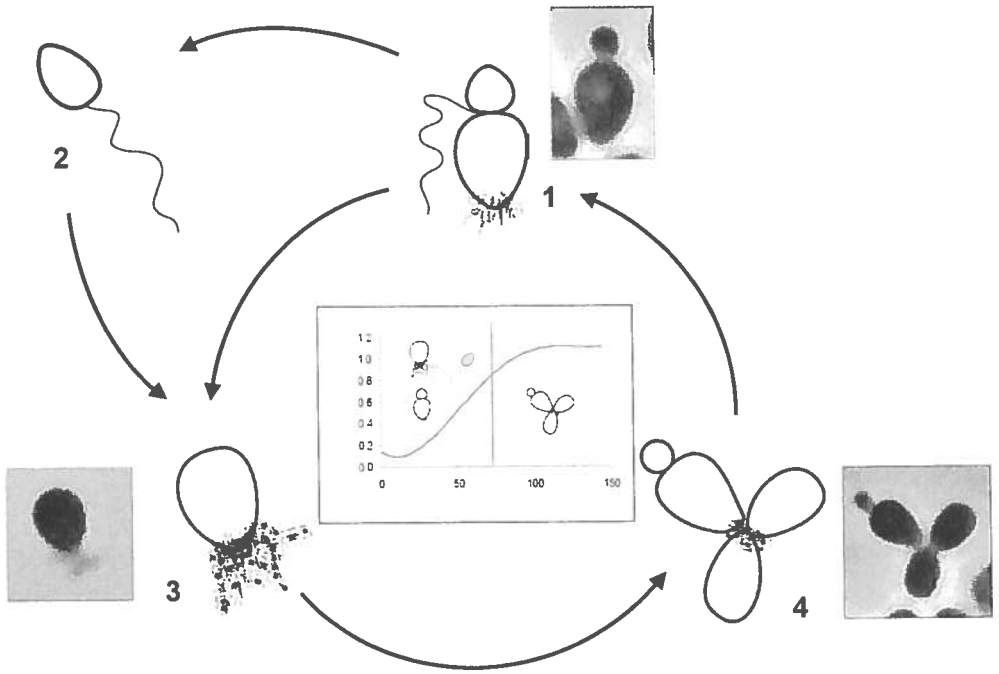



**Abb. 7** Zellzyklus von *R. baltica* im Verlauf einer Wachstumskurve. **1** Mutterzelle mit Knospe, die bereits eine Geißel trägt und sich bald als Schwärmerzelle (**2**) ablösen wird. **3** Adulte Zelle mit Holdfast-Substanz. **4** Rosette aus drei Zellen, die mit ihrer Holdfast-Substanz aneinanderhaften. Eine der drei Zellen bildet bereits erneut eine Knospe.

Zur Untersuchung des Lebenszyklus von *R. baltica* wurden Zellen in Erlenmeyerkolben mit Minimalmedium angezogen (Publikation 6). Das Medium enthielt *N*-Acetylglucosamin als einzige Kohlenstoff- und Stickstoffquelle. Unter diesen Anzuchtbedingungen besitzt *R. baltica* eine Generationszeit von 14 h. Das Wachstum der Kultur wurde durch Messung der optischen Dichte bei 600 nm ($OD_{600}$) verfolgt. Um physiologische und morphologische Änderungen während der Wachstumsphasen auf proteomischer Ebene zu verfolgen, wurden Zellkulturen über einen Zeitraum von 18 Tagen

in unterschiedlichen Phasen der Wachstumskurve beprobt (Abb. 8) und die Proben mit der 2D DIGE-Technologie untersucht. Als Referenz dienten Zellen aus der mittleren exponentiellen Phase ($OD_{600}$ 0,47), deren Proteinmuster mit denen der anderen Proben ($OD_{600}$ 0,37, 0,52, 0,82, 0,97, 1,02 und $0{,}98_{alt}$) verglichen wurde. Um sicherzustellen, dass die geernteten Kulturen nicht im Wachstum limitiert waren, wurden von jeder Kultur 20 ml steril entnommen und bis zur finalen optischen Dichte weiter inkubiert. Alle Kulturen mit denen in dieser Weise verfahren wurde, zeigten das in Abbildung 8 dargestellte Wachstumsverhalten. Da die Versuche in diskontinuierlichen Flüssigkulturen durchgeführt wurden, waren die Zellen mit zunehmender optischer Dichte auch Stress aufgrund von Substratlimitierung und hoher Zelldichte ausgesetzt. Diese in der stationären Phase vorherrschenden und für das Zellwachstum ungünstigen Zustände ähneln den Umweltbedingungen, unter denen *R. baltica* auch im natürlichen Habitat überleben muss.

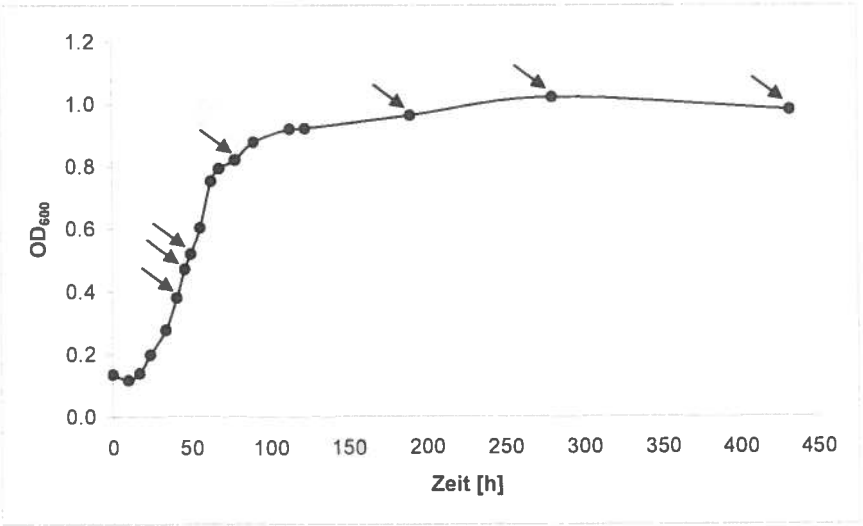



**Abb. 8** Wachstumskurve von *R. baltica* auf *N*-Acetylglucosamin. Die Zeitpunkte der Probennahme ($OD_{600}$ 0,37, 0,52, 0,82, 0,97, 1,02 und $0{,}98_{alt}$) sind mit schwarzen Pfeilen gekennzeichnet. Die Referenzprobe wurde bei $OD_{600}$ 0,47 geerntet und ist in rot dargestellt.

In der für diesen Versuch aufgenommenen Wachstumskurve können die verschiedenen Wachstumsphasen deutlich voneinander abgegrenzt werden. Nach einer ausgeprägten exponentiellen Wachstumsphase befindet sich die Kultur ab einer $OD_{600}$ von 0,85 am Übergang in die stationäre Phase. Letztere verläuft bemerkenswerter Weise über mindestens 14 Tage. Zum Zeitpunkt der letzten Probennahme nach 18 Tagen war die optische Dichte leicht auf $OD_{600}$ 0,98 zurückgegangen, was auf einen möglichen Beginn der Absterbephase hindeutet. Diese Probe wurde als $OD_{600}$ $0{,}98_{alt}$ bezeichnet.

In Proben der exponentiellen Wachstumsphase ($OD_{600}$ 0,37 und 0,52) wurden nur wenige differentielle Proteine detektiert. Dies hängt vermutlich mit den unvermindert guten Wachstumsbedingungen in dieser Phase zusammen. Die zwei in dieser Phase differentiell gebildeten hypothetischen Proteine könnten Funktionen in der Zellteilung oder Zelldifferenzierung der beiden Morphotypen haben. Da sich die Zellen in den Kulturen asynchron teilen, enthält jede Probe stets ein Gemisch aus unterschiedlichen Wachstumsstadien. Dadurch ist eine genaue Zuordnung der Proteine zu einem bestimmten Zelltypus nicht möglich.

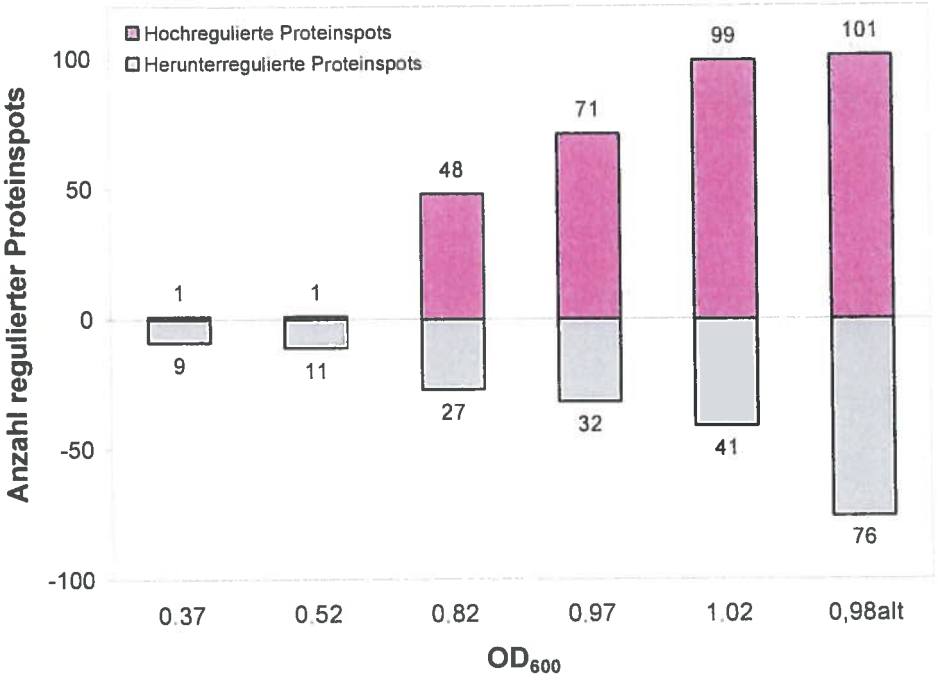



**Abb. 9** Anzahl differentiell gebildeter Proteinspots in verschiedenen Wachstumsphasen. Als Referenz dienten Zellen, die in der exponentiellen Wachstumsphase bei $OD_{600}$ 0,47 geerntet wurden.

Erste drastische Änderungen des Proteoms konnten bei einer $OD_{600}$ von 0,82 beobachtet werden. In dieser Wachstumsphase waren 48 Proteine hochreguliert und 27 Proteine herunterreguliert (Abb. 9). Der Zeitpunkt dieser Probennahme stimmt mit dem Übergang der Kultur in die stationäre Phase überein, welche durch hohe Zelldichten, Nährstoffmangel und zahlreiche Stressfaktoren gekennzeichnet ist. Von Modellorganismen, wie z. B. *E. coli*, ist bekannt, dass die Zellen ihren Stoffwechsel und ihre Proteinausstattung massiv verändern, um sich den ungünstigen Bedingungen

anzupassen. Zellwachstum und Zellvermehrung werden eingestellt und Proteine synthetisiert, welche die Zellen vor Stressfaktoren, wie z. B. toxischen Sauerstoffspezies, schützen. Mit diesen Anpassungen wird eine Langzeitüberdauerung ermöglicht.

Im Verlauf der stationären Phase konnten Änderungen in den katabolen Stoffwechselwegen beobachtet werden. Proteine, die spezifisch für das Wachstum auf *N*-Acetylglucosamin sind (RB3330, RB3342, Publikationen 1 und 5), waren um den Faktor vier herunterreguliert, was den vollständigen Verbrauch von *N*-Acetylglucosamin im Medium widerspiegelt. Da *N*-Acetylglucosamin auch die einzige Stickstoffquelle darstellte, waren mit abnehmender Verfügbarkeit auch die Glutamin-Synthetase II (RB7247), ein ABC-Transporter (RB11021) und ein Regulatorprotein für die Stickstoffassimilation (RB4487) in den Zellen der späten stationären Phase weniger abundant.

Zusätzlich war die Aconitase aus dem TCA-Zyklus herunterreguliert, wohingegen zwei Proteine aus dem oxidativen Pentosephosphat-Zyklus (Gluconolactonase und Transaldolase) in diesen Zellen abundanter als in den exponentiell wachsenden Zellen waren. Dieses Ergebnis lässt vermuten, dass in der stationären Phase aufgrund eines reduzierten TCA-Zyklus, verstärkt der Pentosephosphat-Zyklus zur Energiegewinnung genutzt wird (Yoon et al. 2003).

Im Gegensatz dazu waren einige Hydrolasen, Dehydrogenasen und Oxidoreduktasen spezifisch in der stationären Phase stark hochreguliert (z. B. eine Hydrolase (RB3689) um den Faktor 30). Diese Enzyme könnten bei der Umstellung des Stoffwechsels auf die Bedingungen der stationären Phase von noch unbekannter Bedeutung sein.

Die Abundanz einiger Enzyme, die an der Biosynthese von Aminosäuren wie Methionin, Leucin, Valin, Isoleucin und Prolin beteiligt sind, sank beim Eintritt in die stationäre Phase. Da Prolin ein Hauptbestandteil der Zellwand von *R. baltica* ist (Liesack et al. 1986) wird es bei langsamerem Wachstum bzw. Ausbleiben der Zellteilung nicht mehr in großen Mengen benötigt. Im Gegensatz dazu war jedoch die Expression von Enzymen, welche an der Biosynthese von Lysin und Tryptophan beteiligt sind, hochreguliert. Von *E. coli* ist bekannt, dass bei hohen Zelldichten vermehrt Lysin und andere Aminosäuren synthetisiert werden (Liu et al. 2000) und dass die Induktion des Tryptophan-Operons gesteigert wird (Yoon et al. 2003). Über die physiologische Bedeutung, die diesen Aminosäuren in der stationären Phase zukommt, ist bislang nichts bekannt.

Die Expression von Chaperonen wie DnaK und GroEL war in der späten stationären Phase herunterreguliert. Die gleichzeitig sinkenden Konzentrationen des Translations-Elongations-Faktors EF-P, einiger tRNA Synthetasen und die Umstellung des Aminosäure-metabolismus weisen auf eine Reduzierung des Zentralstoffwechsels hin. Während Wachstum nicht mehr möglich ist, werden überlebensnotwendige metabolische Abläufe aufrechterhalten. Interessanterweise waren einige Enzyme, wie z. B. die DNA-Polymerase, andere Elongations-Faktoren (EF-Tu, EF-Ts, EF-G) und ATP-Synthasen, über den gesamten Verlauf der Wachstumskurve in unveränderten Mengen vorhanden. Dies deutet darauf hin, dass die Enzyme entweder ständig neu synthetisiert werden oder aber eine längere Lebensdauer als andere Proteine aufweisen.

Der alternative Sigmafaktor H ist zum Ende der stationären Phase um den Faktor 3,6 hochreguliert. Sigma H gehört zu den Sigmafaktoren mit extrazytoplasmatischer Funktion, welche in der Regulation der Interaktionen von Bakterien mit ihrer Umwelt, besonders bei Anpassung an Stress, eine Rolle spielen (Manganelli et al. 2002). In der stationären Phase leiden die Zellen von *R. baltica* zunehmend an oxidativem Stress. Dies zeigt sich in der stetig zunehmenden Konzentration von Proteinen, welche die Entgiftung reaktiver Sauerstoffspezies, wie sie bei hohen Zelldichten auftreten, katalysieren (z. B. Katalase, Superoxiddismutase, Alkylhydroperoxidasen). Ein weiterer alternativer Sigmafaktor ($\sigma^B$ oder $\sigma^S$), der von *B. subtilis* bzw. *E. coli* beim Übergang in die stationäre Phase vermehrt exprimiert wird und welcher für die Expression von Proteinen der generellen Stressantwort zuständig ist (Bernhardt et al. 1997; Hegge-Aronis 1999), konnte nicht auf den Gelen identifiziert werden. Da das homologe Protein in *R. baltica* (RB3727) einen theoretischen IEP von 6,7 hat, kann das Protein im Schmier des Randbereiches der pH 4 – 7 Gele vermutet werden.

Die Verfolgung von Änderungen im Proteom als Reaktion auf externe Stimuli lässt Rückschlüsse auf den physiologischen Zustand der Zelle und die dort ablaufenden molekularen Umstrukturierungen nur zu, wenn auch eine Funktionszuweisung für die als reguliert ermittelten Genprodukte vorliegt. Die Proteine, welche für Knospung, Zelldifferenzierung und deren Regulation verantwortlich sind, verbergen sich jedoch vermutlich hinter hypothetischen Proteinen. In der vorliegenden Arbeit wurden 13 hypothetische Proteine identifiziert, deren relative Abundanzen in den untersuchten Wachstumsstadien stark variierten. Diese 13 Proteine könnten für die Differenzierung der Morphotypen oder aber in der Stressantwort eine Rolle spielen. Der Zeitpunkt zu dem diese Proteine am prominentesten sind, gibt dabei einen direkten Hinweis auf ihre

mögliche Funktion. So können Proteine, welche am Ende der stationären Phase hochreguliert werden, eine Funktion in der Stressantwort oder in der Rosettenbildung erfüllen, während Proteine, die in der exponentiellen Phase hoch abundant sind, eine putative Funktion in der Zellvermehrung oder in der Differenzierung der Schwärmerzellen haben könnten.

### 3.4.2 Differentielle Proteinbildung bei Wachstum auf festem Medium

Um die proteomischen Unterschiede zwischen angeheftetem Zelltypus, wie er in Rosetten oder oberflächenassoziierten Zellen vorkommt, und frei lebenden Schwärmer- und Einzelzellen der exponentiellen Wachstumsphase zu bestimmen, wurden zusätzlich Anzuchten von *R. baltica* auf Agarplatten untersucht. Auch hier diente *N*-Acetylglucosamin aus Gründen der Vergleichbarkeit als einzige Kohlenstoff- und Stickstoffquelle. Variationen der Proteinmuster wurden mit der 2D DIGE-Technologie untersucht. Als Referenz dienten, wie bereits im vorangegangenen Experiment, exponentiell wachsende Zellen ($OD_{600}$ 0,47). Zum einen sollten Proteine detektiert werden, welche spezifisch für das exponentielle Wachstum in Flüssigkulturen sind, da einige von ihnen verantwortlich für die Differenzierung von Schwärmerzellen bzw. des adulten einzelligen Morphotypus sein sollten. Zum anderen könnten Proteine, die spezifisch bei Wachstum auf festem Medium gebildet werden, zur Ausbildung des sessilen Morphotypus beitragen.

Die Identifizierung der regulierten Proteine zeigte, dass auch die Zellen in den Kolonien oxidativem Stress ausgesetzt sind. Allerdings wurden hier andere Enzyme zur Eliminierung reaktiver Sauerstoffspezies eingesetzt als in den Zellen der stationären Flüssigkultur. Glutathion-Peroxidase (RB2244) und ein Peroxiredoxin (RB12362) waren bei Wachstum auf festem Medium hochreguliert (Average ratio 4,1 bzw. 2,7), während diese in der stationären Phase der Flüssigkultur in deutlich geringeren Mengen vorhanden waren (Average ratio -4,3 bzw. 1). Im Gegensatz dazu schien ein universelles Stressprotein (RB11179) spezifisch für die Stressantwort in Zellen der Flüssigkultur zu sein (Publikation 6, Tabelle 1).

Besonders auffällig war ein hypothetisches Protein (RB6941), dessen Abundanz in der stationären Phase herunterreguliert war und das zudem in 11,4-fach geringerer Menge in Zellen der Kolonien vorlag. Dieses Protein könnte somit aufgrund seiner spezifischen Bildung während der exponentiellen Wachstumsphase der Flüssigkulturen in der Differenzierung von Schwärmerzellen involviert sein. Ein weiteres hypothetisches Protein

(RB5240), welches spezifisch in den angehefteten Zellen der Agarplatten exprimiert wurde, könnte für oberflächenassoziiertes Wachstum verantwortlich sein.

Diese Ergebnisse liefern erste Hinweise darauf, dass Proteine bisher unbekannter Funktion, aufgrund ihrer Bildung unter bestimmten Wachstumsbedingungen entsprechenden zellulären Prozessen zugeordnet werden können. Diese Proteine sind daher hervorragende Kandidaten für eine weiterführende molekulare Funktionsuntersuchung.

Im natürlichen Habitat ist *R. baltica* meist einem Mangel an Nährstoffen ausgesetzt. Die in dieser Arbeit durchgeführten Untersuchungen zeigen, dass dieser Umweltorganismus in der Lage ist längere Perioden der Substratlimitation zu überleben. Eine Strategie, um in der Natur eine verbesserte Nährstoffversorgung zu gewährleisten, könnte die Anheftung an Substratpartikel, wie z. B. „marine snow", darstellen. Mit zunehmender Zelldichte muss *R. baltica* den damit verbundenen Stress bewältigen. Bei erneuter Substratlimitation können die Schwärmerzellen neue Lebensräume aufsuchen. Somit ist der Lebenszyklus von *R. baltica* von essentieller Bedeutung für das Überleben dieses Planktomyceten in der Natur. (Die Daten dieses Kapitels sind in Publikation 6 enthalten.)

## 4. Ausblick

In der vorliegenden Arbeit konnte durch Einsatz der 2DE Technik erstmals ein Mastergel der löslichen Proteine mit einem IEP zwischen pH 4 und pH 7 von *R. baltica* erstellt werden. Dadurch gelang es, die bioinformatische Vorhersage für 558 ORFs zu bestätigen und die Annotation des Genoms durch die Identifizierung weiterer bislang nicht vorhergesagter Proteine zu verfeinern. Es wäre wichtig, diese Proteinkarte in zukünftigen Studien noch zu erweitern. Dazu könnte durch Einsatz überlappender enger pH-Gradienten die Zahl der detektierbaren bzw. identifizierbaren Proteine erhöht werden. Insbesondere im Hinblick auf die große Anzahl hypothetischer Proteine mit einem IEP größer 9 und einem Molekulargewicht kleiner 20 kDa wäre es sinnvoll, auch im alkalischen Bereich proteomische Untersuchungen durchzuführen. Damit könnte der Frage nachgegangen werden, ob es sich um eine Übervorhersage (Overprediction) bei der Annotation handelt, oder ob diese Gene tatsächlich exprimiert werden.

Alle auf dem Mastergel verzeichneten Genprodukte sind Teil des Proteoms, das *R. baltica* unter den zuvor definierten Standardlaborbedingungen synthetisiert. Die in den physiologischen Experimenten als differentiell erkannten Proteine verschafften erstmals Einblicke in das regulatorische Potential eines Planktomyceten. Dabei wurden

schwerpunktartig die substratabhängige Induktion kataboler Gene des Kohlenhydratstoffwechsels und die molekulare Anpassung an die Stressbedingungen erarbeitet. Darüber hinaus wurden Proteinkandidaten für die Differenzierung der verschiedenen Morphotypen von *R. baltica* identifiziert. Für künftige habitatorientierte Studien konnte eine Fülle validierter Targets für einen *in situ* Nachweis differentieller Genexpression ermittelt werden.

Über der Hälfte der vorhergesagten Gene von *R. baltica* konnte bioinformatisch keine Funktion zugewiesen werden. In den durchgeführten Proteomanalysen wurde eine große Anzahl hypothetischer Proteine identifiziert, welche unter den untersuchten physiologischen Bedingungen gebildet und zum Teil differentiell reguliert wurden. Der physiologische Zusammenhang, in dem diese Proteine spezifisch gebildet wurden, gibt einen ersten konkreten Hinweis auf die Funktion dieser Proteine. Diese Funktionsvorschläge bilden eine solide Basis für weiterführende Untersuchungen, z. B. durch gezielte Mutagenese.

Die Verknüpfung der bisher identifizierten Proteine mit der bioinformatischen Signalpeptidvorhersage ergab erste Hinweise auf die Lokalisation der Proteine in der Zelle. Um die subzelluläre Lokalisation der Proteine und damit die Güte dieser Vorhersagen zu überprüfen, müssen die einzelnen Kompartimente (z. B. das Pirellulosom) vom Rest der Zelle isoliert und deren Subproteome separat untersucht werden. So könnte auch die Frage beantwortet werden, ob die Unterteilung des theoretischen Proteoms aufgrund der IEPs in drei Cluster auf die intrazelluläre Kompartimentierung zurückzuführen ist.

Im Hinblick auf die von den Zellen ausgeschiedenen Enzyme, welche den extrazellulären Abbau von Biopolymeren katalysieren, ist die Untersuchung des so genannten Sekretom von Bedeutung. Durch diese Untersuchungen kann Aufschluss über die Bedeutung der 110 Sulfatasen, von denen für 59 ein Signalpeptid vorhergesagt wurde, erlangt werden (T. Lombardot, persönliche Mitteilung). Das Sekretom von *R. baltica* wird zurzeit im Rahmen des REGX 2 Projekts an der Universität Greifswald in der Arbeitsgruppe von PD Dr. T. Schweder untersucht.

# C Literaturverzeichnis

**Aebersold, R., and D. R. Goodlett.** 2001. Mass spectrometry in proteomics. Chem. Rev. **101**:269-295.

**Alban, A., S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie.** 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics **3**:36-44.

**Anderson, L., and J. Seilhamer.** 1997. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis **18**:533-537.

**Bauld, J., and J. Staley.** 1976. *Planctomyces maris* sp. nov.: a marine isolate of the *Planctomyces-Blastocaulis* group of budding bacteria. J. Gen. Microbiol. **97**:45-55.

**Berggren, K., E. Chernokalskaya, T. H. Steinberg, C. Kemper, M. F. Lopez, Z. Diwu, R. P. Haugland, and W. F. Patton.** 2000. Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulphate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis **21**:2509-2521.

**Bernhardt, J., U. Völker, A. Völker, H. Antelmann, R. Schmid, H. Mach, H., and M. Hecker.** 1997. Specific and general stress response in *Bacillus subtilis* – a two-dimensional protein electrophoresis study. Microbiology **143**:999-1017.

**Bjellqvist, B., K. Ek, P. G. Righetti, E. Gianazza, A. Görg, R. Westermeier, and W. Postel.** 1982. Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J. Biochem. Biophys. Methods **6**:317-339.

**Blobel, G.** 2000. Protein targeting (Nobel Lecture). Chembiochem. **1**:86-102.

**Bomar, D., S. Giovannoni, and E. Stackebrandt.** 1988. A unique type of eubacterial 5S rRNA in members of the order *Planctomycetales*. J. Mol. Evol. **27**:121-125.

**Büttner, K., J. Bernhardt, C. Scharf, R. Schmid, U. Mäder, C. Eymann, H. Antelmann, A. Völker, U. Völker, and M. Hecker.** 2001. A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. Electrophoresis **22**:2908-2935.

**Corthals, G. L., V. C. Wasinger, D. F. Hochstrasser, and J.-C. Sanchez.** 2000. The dynamic range of protein expression: A challenge for proteomic research. Electrophoresis **21**:1104-1115.

**DeLong, E. F., D. G. Franks, and A. L. Alldredge.** 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. **38**:924-934.

**Erdmann, V. A., and J. Wolters.** 1986. Collection of published 5S, 5,8S and 4,5S ribosomale RNA sequences. Nucl. Acids Res. **14** Suppl:r1-r59.

**Ermolaeva, M. D., O. White, and S. L. Salzberg.** 2001. Prediction of operons in microbial genomes. Nucl. Acids Res. **29**:1216-1221.

**Fields S., and O. Song.** 1989. A novel genetic system to detect protein-protein interactions. Nature **340**:245-246.

**Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, J. F. Dougherty, J. M. Merrick, et al.** 1995. Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**:496-512.

**Franzmann, P. D., and V. B. D. Skerman.** 1984. *Gemmata obscuriglobus*, a new genus and species of the budding bacteria. Antonie van Leeuwenhoek **50**:261-268.

**Fuerst, J. A.** 1995. The planctomycetes: emerging models for microbial ecology, evolution and cell biology. Microbiology **141**:1493-1506.

**Gade, D., H. Schlesner, F. O. Glöckner, R. Amann, S. Pfeiffer, and M. Thomm.** 2004. Identification of planctomycetes with order-, genus, and strain-specific 16S rRNA-targeted probes. Microb. Ecol. Mar 4 Epub ahead of print.

**Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon et al.** 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415**:141-147.

**Gharbi, S., P. Gaffney, A. Yang, M. J. Zvelebil, R. Cramer, M. D. Waterfield, and J. F. Timms.** 2002. Evaluation of two-dimensional difference gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol. Cell. Proteomics **1**:91-98.

**Gimesi, N.** 1924. Hydrobiologiai Tanulmányok (Hydrobiologische Studien). I. *Planktomyces bekefii* Gim. nov. gen. et sp., p. 1-8. Budapest: Kiadja a Magyar Ciszterci Rend.

**Giovannoni, S. J., E. Schabtach, and R. W. Castenholz.** 1987. *Isosphaera pallida* gen. nov. and comb. nov., a gliding, budding eubacterium from hot springs. Arch. Microbiol. **147**:276-284.

**Glöckner, F. O., B. M. Fuchs, and R. Amann.** 1999. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. Appl. Environ. Microbiol. **65**:3721-3726.

**Glöckner, F. O., M. Kube, M. Bauer, H, Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann, and R. Reinhardt.** 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc. Natl. Acad. Sci. USA **100**:8298-8303.

**Görg, A., W. Postel, and S. Günther.** 1988. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **9**:531-546.

**Görg, A., C. Obermaier, G. Boguth, and W. Weiss.** 1999. Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: Wide pH gradients up to pH 12, longer separation distances and simplified procedures. Electrophoresis **20**:712-717.

**Görg, A., C. Obermaier, G. Boguth, A Harder. B. Scheibe, R. Wildgruber, and W. Weiss.** 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **21**:1037-1053.

**Gray, J. P., and R. P. Herwig.** 1996. Phylogenetic analysis of the bacterial communities in marine sediments. Appl. Environ. Microbiol. **62**:4049-4059.

**Griepenburg, U., N. Ward-Rainey, S. Mohamed, H. Schlesner, H. Marxen, F. A. Rainey, E. Stackebrandt, and G. Auling.** 1999. Phylogenetic diversity, polyamine pattern and DNA base composition of members of the order *Planctomycetales*. Int. J. Syst. Bacteriol. **49**:689-696.

**Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold.** 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. **19**:1720-1730.

**Haynes, P. A., S. P. Gygi, D. Figeys, and R. Abersold.** 1997. Proteome analysis: biological assay or data archive? Electrophoresis **19**:1862-1871.

**Hengge-Aronis, R.** 1999. Interplay of global regulators and cell physiology in the general stress response of *Escherichia coli*. Curr. Opin. Microbiol. **2**:148-152.

**Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkl, B.-C. Li, and R. Hermann.** 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucl. Acids Res. **24**:4420-4449.

**Hirsch, P., and M. Müller.** 1985. *Planctomyces limnophilus* sp. nov., a stalked and budding bacterium from freshwater. Syst. Appl. Microbiol. 6:276-280.

**Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180**:4765-4774.

**Jacobs-Wagner, C.** 2004. Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter*. Mol. Microbiol. **51**:7-13.

**Jetten, M., O. Sliekers, M. Kuypers, T. Dalsgaard, L. van Niftrik, I. Cirpus, K. van de Pas-Schoonen, G. Larvik, B. Thamdrup, D. Le Paslier, H. J. Op den Camp, S. Hulth, L. P. Nielsen, W. Abma, K. Third, P. Engström, J. G. Kuenen, B. B. Jørgensen, D. E. Canfield, J. S. Sinninghe Damsté, N. P. Revsbech, J. Fuerst, J. Weissenbach, M. Wagner, I. Schmidt, M. Schmid, and M. Strous.** 2003. Anaerobic ammonium oxidation by marine and freshwater planctomycetes-like bacteria. Appl. Microbiol. Biotechnol. **63**:107-114.

**Karlin, S., and J. Mrázek.** 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol. **182**:5238-5250.

**Kerger, B. D., C. A. Mancuso, P. D. Nichols, D. C. White, T. Langworthy, M. Sittig, H. Schlesner, and P. Hirsch. 1988.** The budding bacteria, *Pirellula* and *Planctomyces*, with atypical 16S rRNA and absence of peptidoglycan, show eubacterial phospholipids and uniquely high portions of long chain beta-hydroxy fatty acids in the lipopolysaccharide lipid A. Arch. Microbiol. **149**: 255-260.

**Kirchman, D. L., and J. White.** 1999. Hydrolysis and mineralization of chitin in the Delaware Estuary. Aquat. Microb. Ecol. **18**:187-196.

**Klose, J.** 1975 Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutation in mammals. Humangenetik **26**:231-243.

**Klose, J., and U. Kobalz.** 1995. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. Electrophoresis **16**:1034-1059.

**König, H., H. Schlesner, and P. Hirsch.** 1984. Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on a *Prosthecomicrobium* sp. Arch. Microbiol. **138**:200-205.

**Krishna R. G., and F. Wold.** 1993. Post-translational modification of proteins. Adv. Enzymol. Relat. Areas Mol. Biol. **67**:265-298.

**Kuypers, M. M., A. O. Sliekers, G. Larvik, M. Schmid, B. B. Jørgensen, J. G. Kuenen, J. S. Sinninghe Damsté, M. Strous, and M. S. Jetten.** 2003. Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. Nature **422**:608-611.

**Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, et al. (International Human Genome Sequencing Consortium).** 2001. Initial sequencing and analysis of the humane genome. Nature **409**:860-921.

**Liao, X., T. Ying, H. Wang, J. Wang, Z. Shi, E. Feng, K. Wei, Y. Wang, X. Zhang, L. Huang, G. Su, and P. Huang, P.** (2003) A two-dimensional proteome map of Shigella flexneri. Electrophoresis **24**:2864-2882.

**Liesack, W., H. König, H. Schlesner, and P. Hirsch.** 1986. Chemical composition of the peptidoglycan-free envelopes of budding bacteria of the *Pirella/Planctomyces* group. Arch. Microbiol. **145**:361-366.

**Liesack, W., and E. Stackebrandt.** 1989. Evidence for unlinked *rrn* operons in the planctomycete *Pirellula marina*. J. Bacteriol. **171**:5025-5030.

**Liesack, W., and E. Stackebrandt.** 1992. Occurrence of a novel group of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. **174**:5072-5078.

**Lindsay, M. R., R. I. Webb, and J. A. Fuerst.** 1997. Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus *Pirellula*. Microbiology **143**:739-748.

**Lindsay, M., R. I. Webb, M. Strous, M. S. M. Jetten, M. K. Butler, R. J. Forde, and J. A. Fuerst, J. A.** 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch. Microbiol. **175**:413-429.

**Liu, X., C. Ng, and T. Ferenci, T.** 2000. Global adaptations resulting from high population densities in *Escherichia coli* cultures. J. Bacteriol. **182**:4158-4164.

**Llobet-Brossa, E., R. Rosselló-Mora, and R. Amann.** 1998. Microbial community composition of Wadden Sea sediments as revealed by fluorescence *in situ* hybridisation. Appl. Environ. Microbiol. **64**:2691-2696.

**Lopez, M. F., K. Berggren, E. Chernokalskaya, A. Lazarev, M. Robinson, and W. F. Patton.** 2000. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein

detection in two-dimensional gels and identification by peptide mass profiling. Electrophoresis **21**:3673-3683.

**Manganelli, R., M. I. Voskuil, G. K. Schloonik, E. Dubnau, M. Gomez, and I. Smith.** 2002. Role of the extracytoplasmic-function σ factor $\sigma^H$ in *Mycobacterium tuberculosis* global gene expression. Mol. Microbiol. **45**:365-374.

**Mann, M., R. C. Hendrickson, and A. Pandey.** 2001. Analysis of proteins and proteomes by mass spectrometry. Annu. Rev. Biochem. **70**:437-473.

**Miskin, I. P., P. Farrimond, and I. M. Head.** 1999. Identification of novel bacteria lineages as active members of microbial populations in a freshwater sediment using a rapid RNA extraction procedure and RT-PCR. Microbiology **145**:1977-1987.

**Neef, A., R. Amann, H. Schlesner, and K.-H. Schleifer.** 1998. Monitoring a widespread bacterial group: *in situ* detection of planctomycetes with 16S rRNA-targeted probes. Microbiology **144**:3257-3266.

**O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**:4007-4021.

**O'Neill, E. E., C. J. Brock, A. F. von Kriegsheim, A. C. Pearce, R. A. Dwek, S. P. Watson, and H. F. Hebestreit.** 2002. Towards complete analysis of the platelet proteome. Proteomics **2**:288-305.

**Rabilloud, T., J. M. Strub, S. Luche, A. van Dorsselaer, and J. Lunardi.** 2001. A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels. Proteomics **1**:699-704.

**Rabilloud, T.** 2002. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but still climbs up the mountains. Proteomics **2**:3-10.

**Ravenschlag, K., K. Sahm, and R. Amann.** 2000. Quantitative molecular analysis of the microbial community in marine arctic sediments (Svalbard). Appl. Environ. Microbiol. **67**:387-395.

**Schagger, H. and G von Jagow.** 1991. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal. Biochem. **199**:223-231.

**Schlesner, H.** 1986. *Pirella marina* sp. nov., a budding, peptidoglycan-less bacterium from brackish water. Syst. Appl. Microbiol. **8**:177-180.

**Schlesner, H.** 1989. *Planctomyces brasiliensis* sp. nov., a halotolerant bacterium from a salt pit. Syst. Appl. Microbiol. 12:159-161.

**Schlesner, H.** 1994. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from various aquatic habitats using dilute media. Syst. Appl. Microbiol. **17**:135-145.

**Schlesner, H., and P. Hirsch.** 1984. Assignment of ATCC 27377 to *Pirella* gen. nov. as *Pirella staleyi* comb. nov. Int. J. Syst. Bacteriol. **34**: 492-492.

**Schlesner, H., and E. Stackebrandt.** 1986. Assignment of the genera *Planctomyces* and *Pirella* to the new family *Planctomycetaceae* fam. nov. and description of the order *Planctomycetales* ord. nov. Syst. Appl. Microbiol. **8**:174-176.

**Schlesner, H., and P. Hirsch.** 1987. Rejection of the genus name *Pirella* for pear-shaped budding bacteria and proposal to create the genus *Pirellula* gen. nov. Int. J. Syst. Bacteriol. **37**:441.

**Schlesner, H., C. Rendsmann, B. J. Tindall, D. Gade, R. Rabus, S. Pfeiffer, and P Hirsch.** 2004. Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA/DNA-hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov., and an emended description of the genus *Pirellula*. Int. J. Syst. Evol. Microbiol. in press.

**Schmid, M., U. Twachtmann, M. Klein, M. Strous, S. Juretschko, M. Jetten, J. W. Metzger, K.-H. Schleifer, and M. Wagner.** 2000. Molecular evidence for genus level diversity of bacteria capable of catalysing anaerobic ammonium oxidation. Syst. Appl. Microbiol. **23**:93-106.

**Schmidt, J. M., and M. P. Starr.** 1978. Morphological diversity of freshwater bacteria belonging to the *Blastocaulis-Planctomyces* group as observed in natural populations and enrichments. Curr. Microbiol. **1**:325-330.

**Schulenberg, B., B. Arnold, and W. F. Patton.** 2003. An improved mechanically durable electrophoresis gel matrix that is fully compatible with fluorescence-based protein detection technologies. Proteomics **3**:1196-1205.

**Schwartz, R., C. S. Ting, and J. King.** 2001. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. Genome Res. **11**:703-709.

**Shatters, R. G., J. E. Somersville, and M. L. Kahn.** 1989. Regulation of glutamine synthetase II activity in *Rhizobium melioti* 104A14. J. Bacteriol. **171**:5087-5094.

**Shevchenko, A., M. Wilm, O. Vorm, and M. Mann.** 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. **68**:850-858.

**Sinninghe-Damsté, J. S., M. Strous, W. I. Rijpstra, E. C. Hopmans, J. A. Geenevasen, A. C. van Duin, L. A. van Niftrik, and M. S. Jetten.** 2002. Linearly concatenated cyclobutane lipids form a dense bacterial membrane. Nature **419**:708-712.

**Sittig, M., and H. Schlesner.** 1993. Chemotaxonomic investigation of various prosthecate and/or budding bacteria. Syst. Appl. Microbiol. **16**:92-103.

**Skynner, H. A., T. W. Rosahl, M. R. Knowles, K. Salim, L. Reid, R. Cothliff, G. McAllister, and P. C. Guest.** 2002. Alterations of stress related proteins in genetically altered mice revealed by two-dimensional differential in-gel electrophoresis analysis. Proteomics **2**:1018-1025.

**Stackebrandt, E. and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44**: 846-849.

**Staley, J. T.** 1973. Budding bacteria of the *Pasteuria-Blastobacter* group. Can. J. Microbiol. **19**:609-614.

**Staley, J. T., J. A. Fuerst, S. Giovannoni, and H. Schlesner.** 1992 The order *Planctomycetales* and the genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera*. In: Balows, A., H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (eds.). The Prokaryotes 2nd edition, Springer-Verlag New York, pp 3710-3731.

**Starr, M. P., and J. M. Schmidt.** 1984. *Planctomyces stranskae* (*ex* Wawrik 1952) sp. nov., nom. rev. and *Planctomyces guttaeformis* (*ex* Hortobágyi 1965) sp. nov., nom. rev. Int. J. Syst. Bacteriol. **34**:470-477.

**Stegemann, H.** 1970. Proteinfraktionierung in Polyacrylamid und die Anwendung auf die genetische Analyse bei Pflanzen. Angew. Chem. **82**:640.

**Steinberg, T. H., E. Chernokalskaya, K. Berggren, M. F. Lopez, Z. Diwu, R. P. Haugland, and W. F. Patton.** 2000. Ultrasensitive fluorescence protein detection in isoelectric focusing gels using a ruthenium metal chelate stain. Electrophoresis **21**:486-496.

**Steinberg, T. H., B. J. Agnew, K. R. Gee, W.-L. Leung, T. Goodman, B. Schulenberg, J. Hendrickson, J. M. Beechem, R. P. Haugland, and W. F. Patton.** 2003. Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. Proteomics **3**:1128-1144.

**Strous, M., J. A. Fuerst, E. H. M. Kramer, S. Longemann, G. Muyzer, K. T. van de Pas-Schoonen, R. Webb, J. G. Kuenen, and M. S. M. Jetten.** 1999. Missing lithotroph identified as new planctomycete. Nature **400**:446-449.

**Tonella, L., B. J. Walsh, J. C. Sanchez, K. Ou, M. R. Wilkins, M. Tyler, S. Frutiger, A. A. Gooley, I. Pescaru, R. D. Appel, J. X. Yan, A. Bairoch, C. Hoogland, F. S. Morch, G. J. Hughes, K. L. Williams, and D. F. Hochstrasser.** 1998. '98 *Escherichia coli* SWISS-2DPAGE database update. Electrophoresis **19**:1960-1971.

**Tonella, L., C. Hoogland, P.-A. Binz, R. D. Appel, D. F. Hochstrasser, and J.-C. Sanchez.** 2001. New perspectives in the *Escherichia coli* proteome investigation. Proteomics **1**:409-423.

**Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**:37-43.

**Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg.**

2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature **403**:623-627.

**Ünlü, M., M. E. Morgan and J. S. Minden.** 1997. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electrophoresis **18**:2071-2077.

**VanBogelen, R. A., E. E. Schiller, J. D. Thomas, and F. C. Neidhardt.** 1999. Diagnosis of cellular status of microbial organisms using proteomics. Electrophoresis **20**:2149-2159.

**Velculescu, V. E., L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrei, D. E. Jr. Bassett, P. Hieter, B. Vogelstein, and K. W. Kinzler.** 1997. Characterization of the yeast transcriptome. Cell **88**:243-251.

**Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffmann, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**:66-74.

**Wallin, E., and G. von Heijne.** 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaeon, and eukaryotic organisms. Protein Sci. 7:1029-1038.

**Wang, J., C. Jenkins, R. I. Webb, and J. A. Fuerst.** 2002. Isolation of *Gemmata*-like and *Isosphaera*-like planctomycetes bacteria from soil and freshwater. Appl. Environ. Microbiol. **68**:417-422.

**Ward, N., F. A. Rainey, E. Stackebrandt, and H. Schlesner.** 1995. Unraveling the extent of diversity within the order *Planctomycetales*. Appl. Environ. Microbiol. **61**:2270-2275.

**Ward, N. L., F. A. Rainey, B. P. Hedlund, J. T. Staley, W. Ludwig, and E. Stackebrandt.** 2000. Comparative phylogenetic analyses of members of the order *Planctomycetales* and the division *Verrucomicrobia*: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny. Int. J. Syst. Bacteriol. **50**:1965-1972.

**Wasinger, V. C., S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphrey-Smith.** 1995. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. Electrophoresis **16**:1090-1094.

**Woese, C. R., E. Stackebrandt, T. J. Macke, and G. E. Fox.** 1985. A phylogenetic definition of the major eubacterial taxa. Syst. Appl. Microbiol. **6**:143-157.

**Yan, J. X., R. A. Harry, C. Spilbey, and M. J. Dunn.** 2000. Postelectrophoretic staining of proteins separated by two-dimensional gel electrophoresis using SYPRO dyes. Electrophoresis **21**:3657-3665.

**Yan, J. X., A. T. Devenish R., Wait, T. Stone, S. Lewis, and S. Fowler.** 2002. Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of *Escherichia coli*. Proteomics **2**:1682-1698.

**Yoon, S.H., M.-J. Han, S. Y. Lee, K. J. Jeong, and J.-S. Yoo.** 2003. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnology and Bioengineering **81**:753-767.

**Zarda, B., D. Hahn, A. Chatzinotas, W. Schönhuber, A. Neef, R. Amann, and J. Zeyer.** 1997. Analysis of bacterial community structure in bulk soil by *in situ* hybridisation. Arch. Microbiol. **168**:185-192.

**Zhou, G., H. Li, D. DeCamps, S. Chen, H. Shu, Y. Gong, M. Flaig, J. W. Gillespie, N. Hu, P. R. Taylor, M. R. Emmert-Buck, L. A. Liotta, E. F. Petricoin III and Y. Zhao.** 2002. 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. Mol. Cell. Proteomics **1**:117-123.

# Teil II: Publikationen

## A Publikationsliste mit Erläuterungen

Die Dissertation beruht zum großen Teil auf den folgenden sechs Publikationen. Die angefügten Erläuterungen zeigen meinen Beitrag an der jeweiligen Arbeit auf.

1. **Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1 by a proteomic approach**
   Ralf Rabus, Dörte Gade, Roger Helbig, Margarete Bauer, Frank Oliver Glöckner, Michael Kube, Heinz Schlesner, Richard Reinhardt, Rudolf Amann

   Proteomics (2002) **2**:649-655

   Durchführung der Wachstumsexperimente und der Suspensionsversuche sowie der HPLC-Analytik. Redaktionelle Mitarbeit beim Erstellen des Manuskripts.

2. **Evaluation of the 2D DIGE method for protein profiling**
   **Soluble proteins of the marine bacterium *Pirellula* sp. strain 1**
   Dörte Gade, Jürgen Thiermann, Dieter Markowsky, Ralf Rabus

   Journal of Molecular Microbiology and Biotechnology (2003) **5**:240-251

   Durchführung aller Experimente und Mitarbeit bei der computergestützten Datenauswertung. Redaktionelle Mitarbeit bei der Erstellung des Manuskripts.

3. **Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA/DNA-hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov., and an emended description of the genus *Pirellula***
   Heinz Schlesner, Christina Rendsmann, Brian J. Tindall, Dörte Gade, Ralf Rabus, Stefan Pfeiffer, Peter Hirsch

   International Journal of Systematics and Evolutionary Microbiology
   accepted for publication

   Mitwirkung bei den Wachstumexperimenten und redaktionelle Mitarbeit bei der Erstellung des Manuskripts.

4. **Towards the proteome of the marine bacterium *Rhodopirellula baltica*: mapping the soluble proteins**

   Dörte Gade, Dorothea Theiss, Daniela Lange, Ekaterina Mirgorodskaya, Thierry Lombardot, Frank Oliver Glöckner, Michael Kube, Richard Reinhardt, Rudolf Amann, Hans Lehrach, Ralf Rabus, Johan Gobom

   Manuscript in preparation

   Durchführung der 2D Gelelktrophorese sowie redaktionelle Mitwirkung am Manuskript.

5. **Proteomic reconstruction of carbohydrate catabolism and regulation in the marine bacterium *Rhodopirellula baltica***

   Dörte Gade, Johan Gobom, Ralf Rabus

   Manuscript in preparation

   Entwicklung des Konzepts zusammen mit Ralf Rabus. Durchführung aller physiologischen Experimente sowie der 2D Gelelektrophorese und der computergestützen Datenauswertung. Erstellung des Manuskripts unter redaktioneller Mitarbeit von Ralf Rabus.

6. **Growth phase/cycle dependent regulation of protein composition in *Rhodopirellula baltica***

   Dörte Gade, Torben Stührmann, Richard Reinhardt, Ralf Rabus

   Manuscript in preparation

   Entwicklung des Konzepts zusammen mit Ralf Rabus. Durchführung aller physiologischen Experimente. Die 2D Gelelektrophorese und die computergestütze Datenauswertung wurde in Zusammenarbeit mit Torben Stührmann durchgeführt. Erstellung des Manuskriptes mit Ralf Rabus.

# B Publikationen

1

# Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1 by a proteomic approach

Ralf Rabus[1*], Dörte Gade[1], Roger Helbig[1], Margarete Bauer[1], Frank Oliver Glöckner[1], Michael Kube[2], Heinz Schlesner[3], Richard Reinhardt[2] and Rudolf Amann[1]

Proteomics (2002) 2:649-655

[1] Max-Planck-Institut für Marine Mikrobiologie, Bremen, Germany
[2] Max-Planck-Institut für Molekulare Genetik, Ressourcen Zentrum, Berlin, Germany
[3] Institut für Allgemeine Mikrobiologie der Universität Kiel, Germany
[*] Corresponding author

## Abstract

*Pirellula* sp. strain 1 is a marine bacterium that can grow with the chitin monomer *N*-acetylglucosamine as sole source of carbon and nitrogen under aerobic conditions, and that is a member of the bacterial division *Planctomycetes*. As basis for the proteomic studies we quantified growth of strain 1 with *N*-acetylglucosamine and glucose, revealing doubling times of 14 and 10 h, respectively. Studies with dense cell suspensions indicated that the capacities to degrade *N*-acetylglucosamine and glucose may not be tightly regulated. Proteins from soluble extracts prepared from exponential cultures grown either with *N*-acetylglucosamine or glucose were separated by two-dimensional gel electrophoresis and visualized by fluorescence staining (Sypro® Ruby). Analysis of the protein patterns revealed the presence of several protein spots only detectable in soluble extracts of *N*-acetylglucosamine grown cells. Determination of amino acid sequences and peptide mass fingerprints from tryptic fragments of the most abundant one of these spots allowed to identify the coding gene on the genomic sequence of *Pirellula* sp. strain 1. This gene showed similarities to a dehydrogenase from *Bacillus subtilis*, and is closely located to a gene similar to glucosamine-6-phosphate isomerase from *B. subtilis*. Genes of two other proteins expressed during growth on *N*-acetylglucosamine as well as on glucose were also identified and found to be similar to a glyceraldehyde-3-phosphate-dehydrogenase and a NADH-dehydrogenase, respectively. Thus the coding genes of three proteins expressed during growth of *Pirellula* sp. strain 1 on carbohydrates were identified and related by sequence similarity to carbohydrate metabolism.

## Introduction

Carbohydrates play an important role in the global carbon cycle of marine systems. Estimates on the annual primary production in marine systems are in the order of 20 - 30 × $10^9$ tons of carbon. Most of the organic carbon is decomposed by aerobic microorganisms in the oxic water column (Rullkötter 1999). About half of this carbon is present as dissolved organic carbon (DOC), up to 50% of which can be constituted of polysaccharides (Benner et al. 1992). Chitin is one of the most abundant polysaccharides in nature and has been estimated to support 10% of bacterial production in marine systems (Kirchman and White 1999). Hydrolytic cleavage of the β-1,4-glycosidic bonds during degradation of chitin yields the monomer, *N*-acetylglucosamine (Gooday 1994).

*Pirellula* sp. strain 1 is an aerobic, marine bacterium that was isolated from the water column of the Kiel Fjord (Germany) (Schlesner 1994). With respect to the range of substrates utilized as carbon sources, *Pirellula* sp. strain 1 is apparently specialized in the metabolism of carbohydrates (*e.g.* *N*-acetylglucosamine and glucose), since amino acids, short-chained fatty acids and alcohols do not support growth (Schlesner, Gade, Rabus unpublished results). *Pirellula* and phylogenetically related bacteria are frequently isolated from aquatic habitats (Schlesner 1994) and quantitatively abundant in marine habitats, as revealed by fluorescence in situ hybridisation (Llobet-Brossa et al. 1998; Glöckner et al. 1999). Thus, *Pirellula* sp. strain 1 may represent an important member of the microbial community in marine systems involved in the degradation of carbohydrates, in particular of the chitin-monomer *N*-acetylglucosamine.

The genus *Pirellula* belongs to the order *Planctomycetales* (Schlesner and Stackebrandt 1986; Staley et al. 1992; Griepenburg et al. 1999) which is part of the phylogenetically deep branching phylum *Planctomycetes* (Stackebrandt et al. 1984; Hugenholtz et al. 1998). Special morphological characteristics of *Pirellula* sp. strain 1 and other related bacteria of this phylum are the peptidoglycan-less, proteinaceous cell envelope (Stackebrandt et al. 1984; König et al. 1984; Liesack et al. 1986) and a membrane engulfed intracytoplasmic structure termed pirellulosome (Lindsay et al. 1997; Lindsay et al. 2001). Due to the above described properties, *Pirellula* sp. strain 1 has been selected for genome sequencing within the REGX project, which focuses on environmentally relevant marine bacteria (www.regx.de).

Past investigations on *Pirellula* sp. strain 1 and other members of the *Planctomycetes* have focused on characterizing nutritional properties and special morphological features.

To date *Pirellula* sp. strain 1 has not been studied in detail on a physiological or biochemical/molecular level. In this study, we determined growth curves with *N*-acetylglucosamine and glucose as the basis for investigations about the expression of capacities to degrade the two carbohydrates in dense cell suspensions and for defined proteomic analysis. Amino acid sequences and peptide mass fingerprints of constitutively and differentially expressed proteins detected by 2DE were used to identify the coding genes on the genomic sequence of *Pirellula* sp. strain 1.

## Materials and Methods

### Media and cultivation

Defined mineral media were prepared as described by Staley (Staley 1968) and Schlesner (Schlesner 1994). The media did not contain any additions of complex substrates such as yeast extract or peptone generally used for the cultivation of members of the Planctomycetes. *N*-acetylglucosamine (10 mM), glucose (10 mM) and ammonium (1 mM) were added from sterile concentrated stock solutions. Transfer of cultures was performed under a sterile hood. Purity of cultures was regularly controlled by microscopic examination and fluorescence in situ hybridisation of samples with probe PIRSt1-197 specific for *Pirellula* sp. strain 1 (Gade, Schlesner, Thomm unpublished results). Incubation was carried out at 25°C in the dark on a rotary shaker (90 rpm).

For determination of growth curves, *Pirellula* sp. strain 1 was cultivated in 30 mL medium in 250 mL Erlenmeyer flasks. Controls were carried out that did not contain cells or substrate. Samples (1.5 mL) were withdrawn for measuring the optical density at 600 nm and for determination of the protein content according to the method described by Bradford (Bradford 1976).

Dense cell suspensions were prepared from cultures that had been transferred at least 5 times on *N*-acetylglucosamine and glucose, respectively. Cultures (750 mL in 2 L Erlenmeyer flasks) were harvested in the exponential phase ($OD_{600}$ of 0.3, corresponding to approx. 10 µg protein per mL). Cultures were centrifuged at 10 000 g for 20 min at 4°C and cell pellets were washed in substrate-free, sterile mineral medium. Dense cell suspensions (30 mL, approx. 120 µg protein per mL) were supplemented with *N*-acetylglucosamine and/or glucose (total sugar concentration was 5 mM) and ammonium (0.5 mM) and incubated at 25°C on a rotary shaker (120 rpm). Samples (1 mL) taken from the dense cell suspensions were centrifuged (10,000 g, 10 min, 4°C) and the filtered (0.2

µm filter) supernatant was stored at –20°C until measurement of sugar concentrations by HPLC.

Chemicals used were of analytical grade. Water for all media and electrophoresis solutions was 18 MΩ/cm (MembraPure, Bodenheim, Germany).

**Chemical analysis**

The concentration of *N*-acetylglucosamine and glucose was determined by HPLC (Sykam, Fürstenfeldbruck, Germany). The system contained a Carbohydrate, $H^+$ column (300 × 8 mm; Sierra Separation Inc., Reno, Nevada, USA), and a RI detector (ERC Inc., Tokyo, Japan). The temperature of the column was 60°C and the flow rate of the eluent (5 mM $H_2SO_4$) was 0.6 mL/min. Under these conditions, glucose eluted from the column after a retention time of 10.3 min, and *N*-acetylglucosamine after 13.1 min. The detection limit of the system was at about 0.25 mM; the dynamic range was up to 10 mM. Data acquisition and processing was performed with the Pyramid software (Axxiom Chromatography, Moorpark, USA).

**Gel electrophoresis and staining**

**Sample preparation**

Samples for 2DE were obtained from exponential cultures (500 mL, $OD_{600}$ of 0.3) of *Pirellula* sp. strain 1 grown on either *N*-acetylglucosamine or glucose and ammonium. Cultures were harvested by centrifugation (10,000 g, 20 min, 4°C). Cells were washed with 100 mM Tris/HCl pH 7.5 containing 5 mM $MgCl_2$. Cell pellets (approx. 300 mg wet weight) were rapidly frozen in liquid nitrogen and stored at -80°C until cell breakage and 2DE. Cell pellets were resuspended in 1 mL lysis buffer (9 M urea, 2% [w/v] carrier ampholytes (Pharmalyte; Amersham Biosciences, Freiburg, Germany)), as described by Görg et al. (Görg et al. 2000). Cell breakage was carried out by repeated freezing in liquid nitrogen and thawing in combination with ultrasonication (20% cycle, 151.2 µm; Bandelin, Berlin, Germany). Intact cells, cell debris and membranes were removed by centrifugation (100,000 g, 1 h, 4°C), yielding the soluble extract. The protein content of the soluble extract was determined by the method described by Bradford (Bradford 1976).

**Two-dimensional gel electrophoresis**

Separation of proteins of the soluble extract by isoelectric focusing was carried out with the IPGphor system using commercial 24 cm long IPG strips with an immobilized pH gradient of 4-7 (Amersham Biosciences, Freiburg, Germany). The amount of protein loaded into the IPG strip holders was 50 to 100 µg, if gels were to be stained with Sypro® Ruby, and 500 to 800 µg, if gels were to be stained with colloidal Coomassie Brilliant Blue (CBB). Soluble extracts were diluted in freshly prepared rehydration buffer containing 8 M urea, 0.5% [w/v] CHAPS, 0.5% carrier ampholytes and 15 mM DTT (Görg et al. 2000). The sample volume was 450 µL and IPG strips were covered with 1.5 mL mineral oil. The conditions for rehydration of the IPG strips, sample entry and IEF were as follows: the temperature was set constant at 20°C and 50 µA were applied per IPG strip. The voltage settings were modified from (Görg et al. 2000): (1) 30 V for 7 h, (2) 60 V for 6 h, (3) 200 V for 1 h, (4) 1000 V for 1 h, a gradient to 8000 V for 0.5 h and (6) 8000 V for 8 h, resulting in a total of about 75 000 Vh. Moist filter papers were laid between the electrodes and the IPG gel after completion of step (2).

After isoelectric focusing the IPG strips were incubated in 15 mL equilibration buffers containing 65 mM DTT and 260 mM iodoacetamide, respectively, for 15 min each at room temperature. The composition of the equilibration buffer was as described by Görg et al. (Görg et al. 2000).

For separation of proteins by SDS-PAGE an electrophoresis system with a peltier-based temperature control was used (Genomic Solutions Inc.; Ann Arbor, Michigan, USA). Gels were made of 375 mM Tris-Base / Tris-HCl, 0.1% SDS and 12.5% of the high tensile strength acrylamide, Duracryl™ (Patton et al. 1992). Gels (five parallels) were run overnight at 10°C using a running buffer containing 25 mM Tris Base, 192 mM glycine and 0.1% SDS; the settings at the power supply were 500 V, 150 mA and 10 W. Electrophoresis was stopped when the bromophenol blue marker reached the end of the gel.

**Staining and image analysis**

After SDS-PAGE, gels were fixed in an aqueous mixture of 40% (v/v) methanol and 10% (v/v) glacial acetic acid and of 45% (v/v) ethanol and 10% (v/v) glacial acetic acid, respectively, for staining with Sypro® Ruby (Berggren et al. 2000) and CBB (Neuhoff et al. 1990). Fixed gels were incubated in 300 mL of staining solution for at least 5 h to overnight. Gels stained with Sypro® Ruby were destained in a mixture of 10% (v/v) methanol and 6% (v/v) glacial acetic acid for 1 h.

Data acquisition of Sypro® Ruby stained gels was performed using a fluorescence scanner (Typhoon; Amersham Biosciences). Excitation from a green visible light laser was at 532 nm and signal recording between 595 and 625 nm using a 610 band pass filter (610 DF30). With respect to identification and quantification of protein spots, digitalized gels were further analyzed with the Image Master 2D software (Amersham Biosciences).

**Amino acid and mass spectrometric analysis**

Protein spots stained with CBB were excised manually. Tryptic digest of proteins and separation of peptide fragments by HPLC was carried out as described before (Jeno et al. 1995). Determination of amino acid sequences by Edman-degradation (Hunkapillar et al. 1983) and of peptide masses by MALDI-TOF-MS (Corthals et al. 2000) was carried out by TopLab (Martinsried, Germany).

**Protein identification and genomic analysis**

Protein identification and genome analysis were based on the genomic sequence of *Pirellula* sp. strain 1. The genomic sequence was determined with a shot-gun approach as part of the microbial genome sequencing effort of both Max-Planck-Institutes as members of the REGX-consortium (www.regx.de).

Analysis of open reading frames in the genomic sequence was carried out using the PEDANT-Pro software package (Frishman and Mewes 1997); Biomax, Martinsried, Germany). Identification of expressed genes on the basis of amino acid sequences was achieved by applying the BLAST program (Altschul et al. 1990). Additionally, peptide mass fingerprints were mapped to the coding genes using the MS-Digest program (Clauser et al. 1999). For matching of experimentally determined peptide masses with theoretical values differences of 50 ppm (spot 1 and 2) and 100 ppm (spot 3) were allowed as thresholds. Homology search with identified genes was based on BLASTP (Clauser et al. 1999) and PFAM domains (Sonnhammer et al. 1997).

## Results and discussion

### Growth with carbohydrates

Growth of *Pirellula* sp. strain 1 has not been quantitatively investigated so far. Here we present growth curves (Figure 1) with *N*-acetylglucosamine and glucose (plus ammonium as nitrogen source). Optical density and protein content increased congruently with both carbohydrates. It should be noted, however, that at the end of the exponential growth phase (Figure 1 A; approx. at $OD_{600}$ 0.5) optical density still increased, while protein content remained constant. This may reflect formation of rosettes or slime, which may not require detectable protein synthesis. In order to obtain samples of exponential cells for 2DE analysis, cultures were harvested at an optical density of 0.3. Doubling times of 14 and 10 h were calculated for growth with *N*-acetylglucosamine and glucose, respectively, which agrees well with those reporte for other members of the *Planctomycetes* (Bauld and Staley 1997).

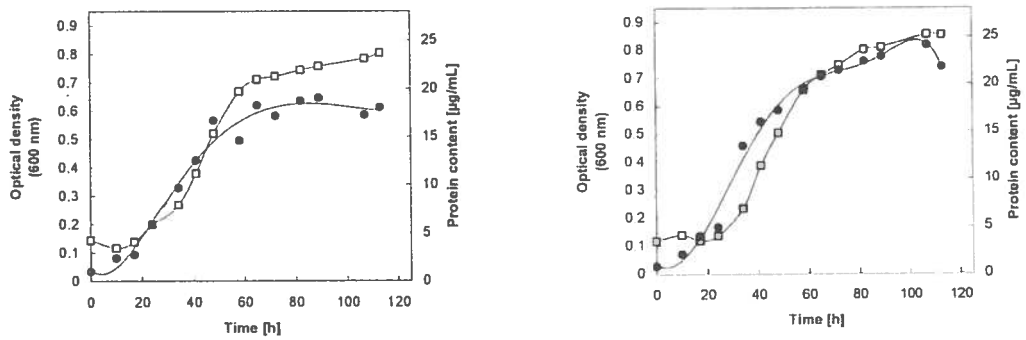



**Figure 1.** Growth of *Pirellula* sp. strain 1 with different carbohydrates in defined mineral media. (A) With *N*-acetylglucosamine (10 mM) as sole source of carbon and nitrogen. (B) With glucose as carbon (10 mM) and ammonium (1 mM) as nitrogen source. Growth was monitored by measuring the optical density (□) and the protein content (●).

### Expression of capacities to degrade *N*-acetylglucosamine and glucose

As a first approach to study the possibility of differential expression of the capacities to degrade different carbohydrates, the utilization of *N*-acetylglucosamine and glucose was studied in dense suspensions of cells adapted to either of the two carbohydrates.

When glucose or *N*-acetylglucosamine were added individually to the dense cell suspensions, the utilization pattern of both sugars was very similar irrespective of the carbohydrate used for initial growth (Figure 2 A and B). On the basis of the time resolution

of these in vivo experiments, *Pirellula* sp. strain 1 appeared to be simultaneously adapted to the utilization of glucose and *N*-acetylglucosamine.

In another set of experiments, the utilization of *N*-acetylglucosamine and glucose was studied, when both carbohydrates were added as an equimolar mixture. Both carbohydrates were simultaneously utilized, regardless of the carbohydrate used before for cultivation. Apparently, the presence of glucose did not repress the utilization of *N*-acetylglucosamine in *Pirellula* sp strain 1 (Figure 2 C and D).

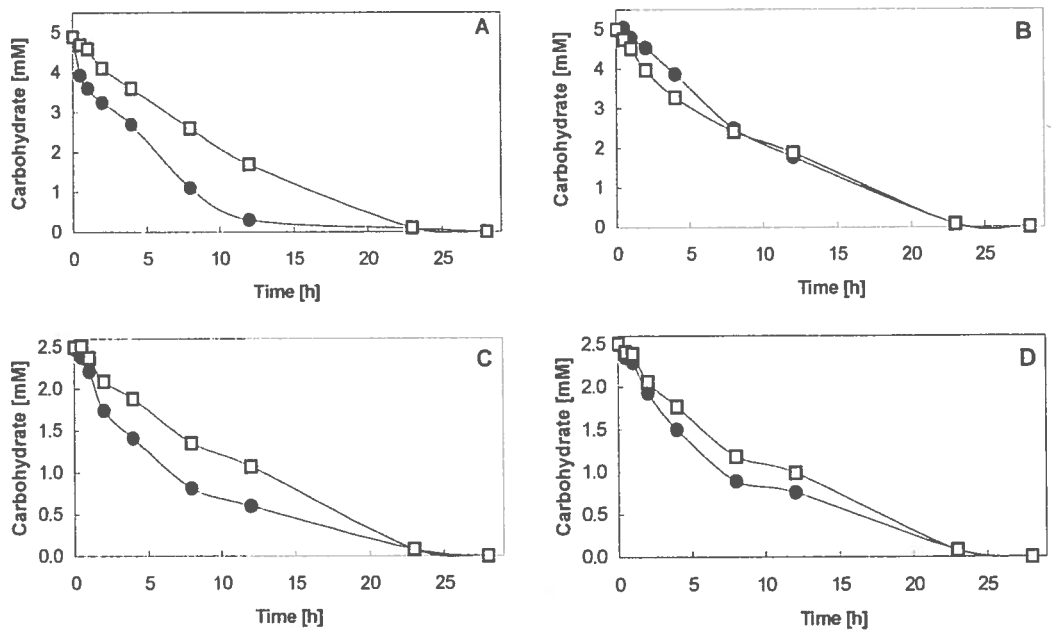



**Figure 2.** Simultaneous adaptation of *Pirellula* sp. strain 1 to utilization of carbohydrates in dense cell suspensions. Cells were pregrown on glucose and $NH_4^+$ (A and C) or *N*-acetylglucosamine (B and D). Cell suspensions were supplemented with *N*-acetylglucosamine (□) or glucose and $NH_4^+$ (●). Carbohydrates were added individually to separate cell suspensions (A and B), or as equimolar mixtures to the same cell suspension (C and D).

These in vivo findings contrast the known regulation of *N*-acetylglucosamine metabolism in well-studied *Escherichia coli*. In this organism *N*-acetylglucosamine is taken up via a PTS-system (Postma et al. 1996) and transformed by deacetylation and deamination/isomerisation into fructose-6-phosphate, which is channeled into glycolysis (Lin 1996). Transcription of the corresponding genes (*nag* regulon (Plumbridge 1989)) is activated by the cyclic AMP catabolite activator protein complex in the presence of *N*-acetylglucosamine (Plumbridge 1990) and repressed by the NagC repressor in the presence of glucose (Plumbridge 1991). One may speculate that slow growing environmental

organisms, such as *Pirellula* sp. strain 1, do not exert a tight regulation on the metabolism of different carbohydrates as known from *E. coli.*

**Two-dimensional gel electrophoresis**

Cell breakage of *Pirellula* sp. strain 1 requires unusually harsh conditions, probably due to the rigid proteinaceous cell wall, the presence of extracellular polysaccharides and the pronounced formation of tightly packed cell assemblages (rosettes). The following 3-step procedure yielded the highest concentration of soluble proteins (about 8 mg protein from 100 mg wet cell mass). First, frozen cells were thawed and resuspended in lysis buffer and cells were thinned out from rosettes by mild sonication. Second, these cell suspensions were repeatedly frozen in liquid nitrogen and thawed at 30°C. Third, cell suspensions were then subjected to several rounds of sonication. Microscopic examination showed that the formerly dark opaque cell bodies had turned translucent upon this treatment.

Soluble proteins were separated according to their isoelectric points using 24 cm long immobilized pH gradients (4-7) and then by molecular weight with 12.5% acrylamide gels. After electrophoresis gels were stained with the fluorescence dye Sypro® Ruby. Analysis of the digitalized gels revealed the presence of about 800 separated protein spots, the majority of which was located in the acidic range (Figure 3). Gels obtained under the same electrophoretic conditions from protein extracts of *N*-acetylglucosamine or glucose grown cells were compared using an evaluation/matching software (Image Master 2D). In the pH range of 5 to 6 (see boxed area in Figure 3) eight proteins were detected that were clearly differentially expressed in response to the carbohydrate used for growth (Figure 4 A and B). Three of these proteins (spots 2, 5 and 10) were higher expressed in soluble extracts from *N*-acetylglucosamine grown cells than in those from glucose grown cells (ratios in protein content were 38.7, 4.1 and 7.6, respectively). Four proteins (spots 6 – 9) were exclusively detected in soluble extracts from *N*-acetylglucosamine grown cells. One protein (spot 4) appeared to be higher expressed during growth with glucose than with *N*-acetylglucosamine (ratio in protein content was 3.3). In view of the simultaneous adaptation to utilization of *N*-acetylglucosamine and glucose (Figure 2), the observed differences in protein patterns may indicate that activation of gene expression and de novo protein synthesis might actually be too rapid to be detectable by the cell suspension experiments. For further analysis (see following sections) we selected three proteins (spots 1 – 3) which were abundant and represented different levels of expression. Spots 1 and 3 were considered as constitutively expressed during growth on both carbohydrates (ratios in

protein content were 1.2 and 1.6, respectively), whereas spot 2 was specifically expressed during growth with *N*-acetylglucosamine (see above).

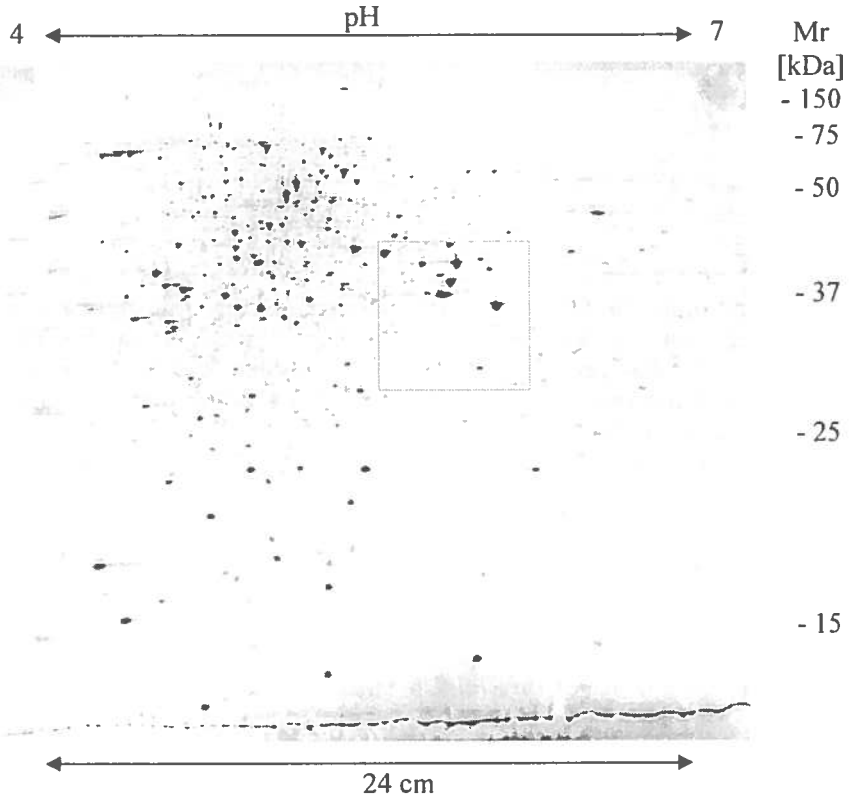



**Figure 3.** Partial view of the *Pirellula* sp. strain 1 proteome during growth with *N*-acetylglucosamine as revealed by two-dimensional gel electrophoresis. A 24 cm long pH 4-7 gradient was used for IEF in the first dimension and a 12.5% SDS-PAGE in the second dimension. Protein load was 50 µg and staining of gels was with the fluorescence dye Sypro® Ruby.

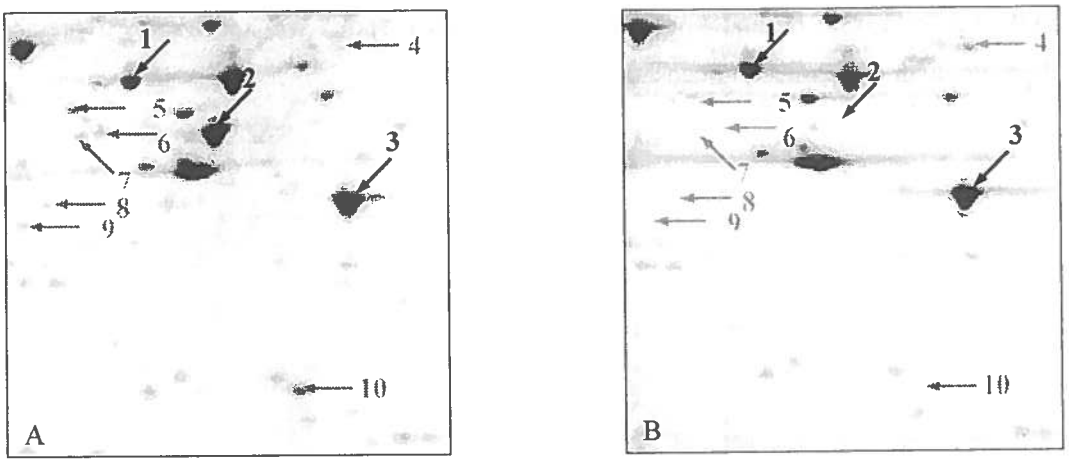



**Figure 4.** Details of the *Pirellula* sp. strain 1 proteome revealing differential expression of proteins in response to growth with different carbohydrates. (A) *N*-acetylglucosamine and (B) glucose and $NH_4^+$. Conditions for 2DE and staining were the same as described for the gel shown in Figure 3. The selected gel portion shown in (A) corresponds to the boxed area in Figure 3. Proteins spots (1-3) selected for further analysis (Table 1) are indicated by black arrows and additional differentially expressed protein spots are labeled with grey arrows.

**Protein identification**

Edman-degradation of an HPLC separated tryptic peptide of the CBB-stained, excised protein „spot 1" provided the sequence YYLLPEDK. Comparison of this peptide sequence with the genomic sequence of *Pirellula* sp. strain 1 allowed the unambiguous identification of an ORF, hereafter termed ORF A, with the peptide being located at positions 370-377. As a second line of identification, the masses of 16 tryptic peptides, determined by MALDI-TOF-MS, could be matched with the calculated masses of the in silico tryptic peptides of ORF A. A selection of the matched peptides is presented in Table 1. The matched peptides on a whole contained 43% of the total number of amino acids encoded by ORF A.

From a tryptic peptide of protein „spot 2" the sequence GFAIIGAGSSGST was determined by Edman-degradation. This peptide could be specifically located to the positions 338-354 in the translated protein sequence of an ORF, hereafter termed ORF B. MALDI-TOF-MS analysis (Tab 1) allowed to match the masses of 14 tryptic peptides with theoretically generated tryptic peptide masses of ORF B. The total number of amino acids from the matched peptides corresponded to 40% of the amino acids encoded by ORF B.

Peptide sequences derived by Edman-degradation from 4 tryptic peptides of protein „spot 3" could all be identified in the translated protein sequence of an ORF, hereafter termed ORF C. The sequences LTFR, IVALAERDRC, STADKPGYDSHLAAGAK and VVLSAPAK could be located at positions 16-19, 73-82, 105-121 and 123-130,

respectively. Based on MALDI-TOF-MS analysis (Table 1) the masses of 11 tryptic peptides could be matched with the theoretical peptide masses generated from the translated protein sequence of ORF C. The matched peptides contained 32% of the amino acids encoded by ORF C.

**Table 1.** Selection of MALDI-TOF-MS determined masses of tryptic peptides from CBB-stained, excised protein spots and comparison to the theoretical masses derived from translated ORFs of *Pirellula* sp. strain 1.[a)]

| Protein "Spot 1" (476 aa) ORF A | | | | Protein "Spot 2" (451 aa) ORF B | | | | Protein "Spot 3" (342 aa) ORF C | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Position | Theor. mass | Exper. mass | Δ Da | Position | Theor. mass | Exper. mass | Δ Da | Position | Theor. mass | Exper. mass | Δ Da |
| 54-81 | 2898.402 | 2898.424 | + 0.022 | 72-101 | 3117.498 | 3117.460 | - 0.038 | **73-82** | **1139.654** | **1139.644** | **- 0.010** |
| 264-277 | 1524.879 | 1524.917 | + 0.038 | 162-177 | 1832.875 | 1832.939 | + 0.064 | **105-121** | **1688.824** | **1688.967** | **+ 0.143** |
| 313-336 | 2727.346 | 2727.427 | + 0.081 | 227-235 | 1204.546 | 1204.602 | + 0.056 | **123-130** | **784.493** | **784.432** | **- 0.061** |
| **370-377** | **1040.530** | **1040.514** | **- 0.016** | **338-354** | **1655.839** | **1655.928** | **+ 0.089** | 221-232 | 1223.736 | 1223.732 | - 0.004 |
| 452-467 | 1876.92 | 1876.98 | + 0.06 | 430-447 | 2015.983 | 2015.976 | - 0.007 | 261-276 | 1700.853 | 1700.962 | + 0.109 |

[a)] Peptide masses shown in bold are congruent with amino acid sequences determined by Edman degradation.

## Genome analysis

The ORF A codes for protein „spot 1, as shown in the preceeding section. Based on BLASTP analysis, ORF A showed highest similarity to an NADH-dependent dehydrogenase (e-value of $1e^{-09}$; accession number AF 039207). Adjacent to ORF A were genes that showed similarity to the phosphoglycerate mutase from *Aquifex aeolicus* and the ribulose-phosphate-3-epimerase from *Vibrio cholerae*. The latter two proteins are apparently involved in carbohydrate metabolism with ribulose-phosphate-3-epimerase suggesting the presence of the oxidative pentose-phosphate cycle in *Pirellula* sp. strain 1.

Protein „spot 2" encoding ORF B was similar to a dehydrogenase from Bacillus halodurans (e-value of $1e^{-06}$, accession number AP00154) according to BLASTP analysis. Upstream of ORF B another ORF was located that displayed similarity to a hypothetical protein of *Bacillus subtilis* (e-value of $2e^{-44}$, accession number 031458). Interestingly, this protein contained PFAM domains characteristic of glucosamine-6-phosphate-isomerase. This might link protein „spot 2", which was specifically expressed during growth with *N*-acetylglucosamine, to the metabolism of this carbohydrate. A second ORF found in the proximity of ORF B was similar to a mannitol transcription regulator MtlR, providing another hint for an involvement of protein „spot 2" in carbohydrate metabolism. BLASTP analysis of protein „spot 3" encoding ORF C revealed similarity to the glyceraldehyde-3-phosphate dehydrogenase from *Mycobacterium tuberculosis* (e-value of $2e^{-99}$, accession number O06822). This result suggested that protein „spot 3" was involved in glycolysis.

## Concluding remarks

This study presents the first investigation into the metabolism of *N*-acetylglucosamine of the marine bacterium *Pirellula* sp. strain 1 combining quantitative growth and substrate utilization studies with a proteomic/genomic approach. The coding genes of three proteins detected by 2DE could be unambiguously identified among the predicted ORFs of the genome of *Pirellula* sp. strain 1. Analysis of sequence similarities of these 3 expressed genes and neighbouring genes suggested a potential function of the 3 proteins in carbohydrate metabolism. Future studies will focus on identifying the complete protein set required in *N*-acetylglucosamine degradation and elucidating the degree of regulation of this catabolic route in *Pirellula* sp. strain 1. Moreover, future research will aim at identifying the coding genes for all proteins of this bacterium separable by 2DE.

## Acknowledgements

We would like to thank A. Beck for help with sequence assembly and analysis, T. Lombardot and H. Teeling for help with the ORF-finding, R. Appel, D. Lange, T. Stührmann and J. Wulf for technical assistance, K. Champion, A. Görg, and U. Völker for helpful discussions about 2DE, and T. Halder for mass-spectrometrical analysis. This work was supported by the BMBF and the Max Planck Society.

## References

**Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipmann.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.

**Bauld, J., and J. T. Staley.** 1976. *Planctomyces maris* sp. nov.: a marine isolate of the *Planctomyces-Blastocaulis* group of budding bacteria. J. Gen. Microbiol. **97:**45-55.

**Benner, R., J. D. Pakulski, M. McCarthy, J. I. Hedges, and P. G. Hatcher.** 1992. Bulk chemical characteristics of dissolved organic matter in the ocean. Science **255:**1561-1564.

**Berggren, H., E. Chernokalskaya, T. H. Steinberg, C. Kemper, M. F. Lopez, Z. Diwu, R. P. Haugland, and W. F. Patton.** 2000. Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis **21:**2509-2521.

**Bradford, M. M.** 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248-254.

**Clauser, K. R., P. Baker, and A. L. Burlingame.** 1999. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal. Chem. **71:**2871-2882.

**Corthals, G. L., S. P. Gygi, R. Aebersold, and S. D. Patterson.** 2000. In: Rabilloud, T. (Ed.). Proteome research: Two-dimensional gel electrophoresis and identification methods. Springer-Verlag Berlin. pp. 197-231.

**Frishman, D., and H. W. Mewes.** 1997. PEDANTic genome analysis. Trends Genet. **13:**415-416.

**Glöckner, F. O., B. M. Fuchs, and R. Amann.** 1999. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. Appl. Environ. Microbiol. **65:** 3721-3726.

**Gooday, G. W.** 1994. In: Ratledge C (Ed.). Biochemistry of microbial degradation. Kluwer Academic Publishers. Dordrecht. pp. 279-312.

**Görg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss.** 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **21:**1037-1053.

**Griepenburg, U., N. Ward-Rainey, S. Mohamed, H. Schlesner, H. Marxsen, F. A. Rainey, E. Stackebrandt, and G. Auling.** 1999. Phylogenetic diversity, polyamine pattern and DNA base composition of members of the order *Planctomycetales*. Inter. J. Syst. Bacteriol. **49:**689-696.

**Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180:**4765-4774.

**Hunkapiller, M. W., R. M. Hewick, J. M. Dreyer, and L. E. Hood.** 1983. High-sensitivity sequencing with a gas-phase sequenator. Meth. Enzymol. **91:**399-413.

**Jeno, P., T. Mini, S. Moes, E. Hintermann, and M. Horst.** 1995. Internal sequences from proteins digested in polyacrylamide gels. Anal. Biochem. **224:**75-82.

**Kirchman, D. L., and J. White.** 1999. Hydrolysis and mineralization of chitin in the Delaware Estuary. Aquat. Microb. Ecol. **18:**187-196.

**König, E., H. Schlesner, and P. Hirsch.** 1984. Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on *Prosthecomicrobium* sp. Arch. Microbiol. **138:**200-205.

**Liesack, W., H. König, H. Schlesner, and P. Hirsch.** 1986. Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella/Planctomyces* group. Arch. Microbiol. **145:**361-366.

**Lin, E. C. C.**. 1996. In: Neidhardt, F. C. (Ed.). *Escherichia coli* and *Salmonella*. American Society for Microbiology Press. Washington DC. pp. 307-342.

**Lindsay, M. K., R. I. Webb, M. Strous, M. S. M. Jetten, M. K. Butler, R. J. Forde, and J. A. Fuerst.** 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch. Microbiol. **175:**413-429.

**Lindsay, M. R., R. I. Webb, and J. A. Fuerst.** 1997. Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycetes bacteria of the genus *Pirellula*. Microbiology **143:**739-748.

**Llobet-Brossa, E., R. Rosselló-Mora, and R. Amann.** 1998. Microbial community composition of Wadden Sea sediments as revealed by fluorescence in situ hybridization. Appl. Environ. Microbiol. **64:**2691-2696.

**Neuhoff, V., R. Stamm, I. Pardowitz, N. Arold, W. Ehrhardt, and D. Taube.** 1990. Essential problems in quantification of proteins following colloidal staining with coomassie brilliant blue dyes in polyacrylamide gels, and their solution. Electrophoresis **11:**101-117.

**Patton, W. F., M. F. Lopez, P. Barry, and W. M. Skea.** 1992. A mechanically strong matrix for protein electrophoresis with enhanced silver staining properties. BioTechniques **12:**580-585.

**Plumbridge, J. A.** 1989. Sequence of the *nag*BACD operon in *Escherichia coli* K12 and pattern of transcription within the *nag* regulon. Mol. Microbiol. **3:**505-515.

**Plumbridge, J. A.** 1990. Induction of the *nag* regulon of *Escherichia coli* by *N*-acetylglucosamine and glucosamine: role of the cyclic AMP-catabolite activator protein complex in the expression of the regulon. J. Bacteriol. **172:**2728-2735.

**Plumbridge, J. A.** 1991. Repression and induction of the *nag* regulon of *Escherichia coli* K-12: the roles of *nagC* and *nagA* in maintenance of the uninduced state. Mol. Microbiol. **5:**2053-2062.

**Postma, P. W., J. W. Lengeler, and G. R. Jacobson.** 1996. In: Neidhardt, F. C. (Ed.). *Escherichia coli* and *Salmonella*. American Society for Microbiology Press. Washington DC. pp. 1149-1174.

**Rullkötter, J.** 1999. In: Schulz, H. D., Zabel, M. (eds.). Marine Geochemistry. Springer-Verlag Berlin. pp. 129-172.

**Schlesner, H.** 1994. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from various aquatic habitats using dilute media. Syst. Appl. Microbiol. **17:**135-145.

**Schlesner, H., and E. Stackebrandt.** 1986. Assignment of the genera *Planctomyces* and *Pirella* to a new family *Planctomycetaceae* fam. nov. and description of the order *Planctomycetales* ord. nov. Syst. Appl. Microbiol. **8:**174-176.

**Sonnhammer, E. L. L., S. R. Eddy, and R. Durbin.** Pfam: a comprehensive database of protein domain families based on seed alignments. 1997. Proteins **28:**405-420.

**Stackebrandt, E., W. Ludwig, W. Schubert, F. Klink, H. Schlesner, T. Roggentin, and P. Hirsch.** 1984. Molecular genetic evidence for early evolutionary origin of budding peptidoglycan-less eubacteria. Nature **307:**735-737.

**Staley, J. T.** 1968. *Prosthecomicrobium* and *Ancalomicrobium*: new prosthecate freshwater bacteria. J. Bacteriol. **95:**1921-1942.

**Staley, J. T., J. A. Fuerst, S. Giovannoni, and H. Schlesner.** 1992. In: Balows, A., H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (eds.). The Prokaryotes 2nd edition, Springer-Verlag New York, pp 3710-3731.

2

# Evaluation of the 2D DIGE method for protein profiling
# Soluble proteins of the marine bacterium *Pirellula* sp. strain 1

Dörte Gade[1], Jürgen Thiermann[2], Dieter Markowsky[2] and Ralf Rabus[1*]

Journal of Molecular Microbiology and Biotechnology (2003) 5:240-251

[1] Max-Planck-Institut für Marine Mikrobiologie, Bremen, Germany
[2] Amersham Biosciences Europe GmbH, Freiburg, Germany
* Corresponding author

## Abstract

Two-dimensional gel electrophoresis (2DE) is a central tool of proteome research, since it allows separation of complex protein mixtures at highest resolution. Quantification of gene expression at the protein level requires sensitive visualization of protein spots over a wide linear range. 2D DIGE (two-dimensional difference gel electrophoresis) is a new fluorescent technique for protein labeling in 2DE gels. Proteins are labeled prior to electrophoresis with fluorescent CyDyes™ and differently labeled samples are then co-separated on the same 2DE gel. We evaluated 2D DIGE for detection and quantification of proteins specific for glucose or *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1. The experiment was based on 10 parallel 2DE gels. Detection and comparison of the protein spots were performed with the DeCyder™ software that uses an internal standard to quantify differences in protein abundance with high statistical confidence; 24 proteins differing in abundance by a factor of at least 1.5 (T-test value < $10^{-9}$) were identified. For comparison another experiment was carried out with four SYPRO® Ruby stained 2DE gels for each of the two growth conditions; image analysis was done with the ImageMaster™ 2D Elite software. Sensitivity of the CyDye fluors was evaluated by comparing Cy2, Cy3, Cy5, SYPRO Ruby, silver, and colloidal Coomassie staining. Three replicate gels, each loaded with 50 µg of protein, were run for each stain and the gels were analyzed with the ImageMaster software. Labeling with CyDyes allowed detection of almost as many protein spots as staining with silver or SYPRO Ruby.

## Introduction

The advent of genomics revolutionized the possibilities of biological experimentation. With the entire genetic information of a given study organism available, analysis of gene expression is not limited to selected genes anymore, but can be performed on the global level. This opens new avenues for functional analysis of the complexity of physiological and cellular aspects of living cells. Cells regulate gene expression, in order to adapt to changing environmental conditions in an economical manner. Differential gene expression is ultimately reflected in the level of proteins, which are the catalytically active molecules that bring cells to life. Therefore profiling of the proteome, *i.e.* the sum of proteins synthesized by a genome under a set of defined conditions, is of particular importance for advancing our understanding of the complexity of biological systems.

Two-dimensional gel electrophoresis (2DE) is thus far the most widely used method for proteome research since it allows the highest level of resolution for separation of complex protein mixtures (Klose 1999; Görg et al. 2000). Following electrophoretic separation, staining of proteins is instrumental for detection and quantification. Three principle staining techniques are most commonly used, which differ with respect to sensitivity, linear range and compatibility with downstream mass spectrometric (MS) analysis (Table 1; (Lauber et al. 2001; Patton 2002)). (i) Colloidal Coomassie Brilliant Blue (cCBB) is regarded as a less sensitive stain, but as probably best compatible with MS-analysis. (ii) Silver staining is presently the most sensitive, generally used staining technique. However, this stain does not allow reliable quantification and poses difficulties on MS analysis. (iii) Since the mid 90ties, SYPRO stains, in particular SYPRO Ruby, became increasingly popular because of the easy staining procedure, high sensitivity and broad linear range (Nishihara and Champion 2002). All of these staining techniques have in common, that protein staining is performed after electrophoretic separation. Furthermore, differential analysis requires separate gels for different samples, resulting in gel-to-gel variations. Therefore warping is required to overlay and compare gels with different samples. Typically, for each sample several gels have to be run in order to generate an "electronic average gel", which is then the basis for determining changes in protein quantity. The resolution of minor differences in protein quantity is particularly demanding.

**Table 1.** Detection limit, linear range and compatibility with mass spectrometric analysis of protein stains.

| Stain | Detection limit[a] [ng] | Linear range[a] [orders of magnitude] | MS-compatible[a] |
|---|---|---|---|
| Conventional Coomassie Blue | 30-100 | 1 | + |
| Colloidal Coomassie Blue | 5-10 | 1 | + |
| Silver | 0.05-2 | 1[b] | +[c]/- |
| Sypro Ruby | 1-2 | 4[d] | + |
| CyDyes | 0.25-0.95 | 4 | +[e] |

[a] Data compiled from literature [3-5, 8, 17, 20, 22, 23].
[b] Recently silver staining was reported to be linear over a range of approximately 200-fold [23].
[c] MS-compatibility of silver staining can be achieved by special staining protocols [25].
[d] Recently SYPRO Ruby stain was reported to be linear over a range of approximately 150,000-fold [23].
[e] During the 2D DIGE process more than 95% of the total protein remain unlabeled. Thus post-staining might be required to make the protein spots detectable for picking. Thus MS-compatibility depends on the dye used for post-staining.

Ünlü et al. (Ünlü et al. 1997) described a modified system for 2DE using cyanine dyes, which was further developed by Amersham Biosciences and is designated two-dimensional difference gel electrophoresis (2D DIGE). Recently, a similar approach employing Alexa Fluor dyes was reported (Eggeling et al. 2001). In contrast to the above mentioned post-electrophoresis staining techniques, in case of 2D DIGE, labeling is carried out prior to electrophoresis and different samples can be co-separated on one gel (termed multiplexing), whereby gel-to-gel variations are avoided. The 2D DIGE process consists of four major steps (see Figure 1): (i) the labeling reaction, (ii) the electrophoretic co-separation of different samples, (iii) the acquisition of separate images for each sample run on a gel, and (iv) the software-based analysis of images to identify spots and determine differences in protein abundance. (i) The individual samples are covalently labeled with one of three cyanine dyes (CyDyes: Cy2, Cy3, Cy5) prior to electrophoretic separation. In the labeling reaction, the CyDyes (as their *N*-hydroxysuccinimidyl (NHS) esters) form an amide with the epsilon amino group of lysine residues. The conditions for labeling are tailored such (low CyDye to protein ratio; termed minimal labeling) that theoretically only 3% of each protein species in the analyzed protein mixture and statistically only one lysine residue per peptide chain is labeled. Samples are labeled with either Cy3 or Cy5. As an internal standard, equal amounts of all samples are mixed and labeled with Cy2. The labeled samples and the internal standard are mixed prior to electrophoresis. (ii) 2DE is essentially carried out according to conventional protocols. Since the CyDyes carry a positive charge that compensates for the positive charge of the lysine amino group lost

during formation of the amide bond (termed charge matching), only little shift of the *pIs* is observed relative to the unlabeled proteins. Labeling with any of the three CyDyes will increase the molecular mass of the proteins by approx. 500 Da. This shift is most pronounced with proteins of low molecular weight. Since these shifts in molecular mass are almost identical for the three CyDyes they do not affect image analysis. (iii) Image acquisition is based on the different excitation and emission wavelengths of the three cyanine fluorphores (see Table 1). Thus, from the different samples run on the same gel separate images are generated, which are devoid of gel-to-gel variations. (iv) Therefore, comparison of the different images of one gel can be performed without warping. Amersham Biosciences developed a specific software, DeCyder, for the analysis of gels generated by 2D DIGE. For each sample (labeled with Cy3 or Cy5) spot boundaries are defined separately. Both sets of spot-boundaries (termed mapsets) are then superimposed on the Cy2-labeled internal standard. Thus using the two sample-specific mapsets, the ratios of spot volumes between an individual sample and the internal standard can be determined. Subsequently, for each spot the change in protein abundance between the different samples can be calculated as the ratio of the two "sample : internal standard" ratios ([C3:Cy2]:[Cy5:Cy2]). Using the "Batch"-modus of DeCyder software, parallel gels can be analyzed, allowing to determine changes in protein quantities with high statistical confidence. The 2D DIGE process allows to limit gel-to-gel variations and to reduce the number of gels.

Only very recently, first applications of the 2D DIGE method for biomedical systems (Gharbi et al. 2002; Skynner et al. 2002; Zhou et al. 2002) and the standard bacterium *Escherichia coli* (Yan et al. 2002; Alban et al. 2003) were reported. Here we apply, for the first time, the 2D DIGE system under optimal technological conditions for an environmental bacterium; *i.e.* all three available CyDyes are used, with Cy2 as label for the internal standard, the Typhoon™ 9400 scanner equipped with three lasers is used for image acquisition and the DeCyder software is used for differential in-gel analysis (DIA) and batch processing. In this study we evaluated the 2D DIGE system for protein profiling in the marine bacterium *Pirellula* sp. strain 1, currently described as "Rhodopirellula baltica". This bacterium was isolated from the water column of the Kiel Fjord, Germany (Schlesner 1994) and found to be catabolically specialized on the utilization of carbohydrates (Schlesner, Gade, Rabus unpublished). Using a proteomic approach we recently identified a putative dehydrogenase which was specifically synthesized during growth with *N*-acetylglucosamine (Rabus et al. 2002). The degradation of *N*-acetylglucosamine is of

environmental relevance in the marine system, since it is the monomer of one of the most abundant biopolymers, chitin (Kirchmann and White 1999). *Pirellula* sp. strain 1 is assumed to be a representative of those bacteria that contribute significantly to the degradation of carbohydrates in marine systems. This bacterium has an unusually large genome size of approx. 7 Mb and the sequence of which has recently been determined by the REGX-consortium (www.regx.de) Here we analyzed protein patterns of cells grown with *N*-acetylglucosamine or glucose with the 2D DIGE system and with SYPRO Ruby staining. In addition, we compared the sensitivity of CyDyes with that of SYPRO Ruby, silver and cCBB, using again samples from *N*-acetylglucosamine grown cells.

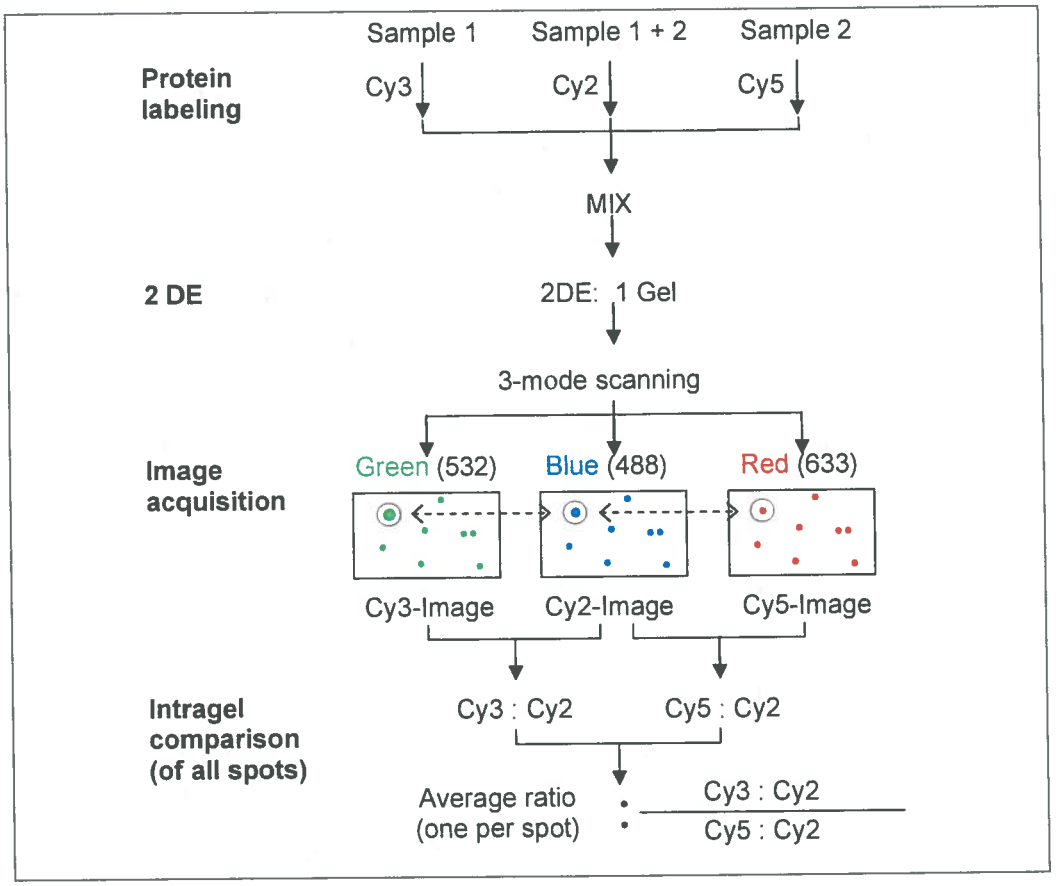



**Figure 1.** Schematic representation of the 2D DIGE workflow for the differential analyses of protein abundance in two different samples (1 and 2). (i) The two samples are labeled with Cy3 and Cy5, respectively. Equal amounts of both samples are mixed and labeled with Cy2 to be used as internal standard. (ii) The three labeling preparations are then mixed prior to simultaneous electrophoretic separation on the same 2DE gel. (iii) For each of the three cyanine fluorphores an independent image is acquired by successive scanning of the gel three times with the fluorphore-specific excitation and emission wavelengths (see Table 2). (iv) During intragel comparison for each spot the ratios of normalized volumes between samples and internal standard are determined. Differences in protein abundance are then determined by calculating the average ratios ([Cy3:Cy2] : [Cy5:Cy2]). Statistical confidence is achieved by analysis of parallel gels (intergel comparison).

## Materials and Methods

### Media and cultivation

*Pirellula* sp. strain 1 was grown on mineral medium with *N*-acetylglucosamine (10 mM) or glucose (10 mM) and ammonium (1 mM) as described before (Rabus et al. 2002). The study organism is currently described as the type strain of a new genus and species, "Rhodopirellula baltica" (Schlesner et al. unpublished).

### Sample preparation

Cells from exponential cultures of *Pirellula* sp. strain 1 were harvested by centrifugation (10,000 *g*, 20 min, 4°C). The pellets were washed with 100 mM Tris/HCl pH 7.5 containing 5 mM $MgCl_2$. Cell pellets (about 100 mg wet weight each) were immediately frozen in liquid nitrogen and stored at –80 °C until cell breakage and 2 DE. The pellets were resuspended in 1 ml lysis buffer (7 M urea, 2 M thiourea, 2% DTT, 2% CHAPS, 0.5% carrier ampholytes (Amersham Biosciences, Freiburg, Germany). Cell breakage was carried out by applying the PlusOne™ Sample Grinding Kit from Amersham Biosciences following the instructions of the manufacturer. Cell debris, DNA and membranes were removed by centrifugation (100,000 *g*, 1 h, 4 °C) and the protein content of this fraction was determined by the method described by Bradford (Bradford 1976).

### Protein labeling with cyanine dyes

Pellets of *Pirellula* sp. strain 1 cells grown with glucose or *N*-acetylglucosamine were solubilized in lysis buffer (7 M urea, 2 M thiourea, 30 mM Tris/HCl, 2% CHAPS; final pH 8.5). DTT and carrier ampholytes were omitted from the sample at this stage since the respective thiol and primary amine groups could potentially react with the NHS esters of the cyanine dyes. Cell breakage was carried out as described above. The pH of the samples was adjusted to 8.5 by careful addition of 50 mM NaOH, as this has been shown to be the optimal pH for the labeling reaction (Tonge et al. 2001). The labeling was performed following the manufacturers recommended protocol. Briefly, a stock solution of 1 ng fluor dye/µl (in dimethyl formamide (DMF)) was further diluted to a working solution of 400 pmol /µl in anhydrous DMF. Finally, 50 µg of protein of either sample were labeled with 400 pmol dye. Samples were vortexed and incubated on ice for 30 min in the dark. The reaction was stopped by adding 1 µl of 10 mM lysine per 400 pmol dye used. The samples were vortexed and incubated for 10 min on ice in the dark. An equal volume of

2 × 2DE sample buffer (as lysis buffer above, plus 2% carrier ampholytes and 2% DTT) was added to the labeled protein yielding a final concentration of 7 M urea, 2 M thiourea, 2% CHAPS, 1% carrier ampholytes and 1% DTT. Labeled samples were combined and rehydration buffer (7 M urea, 2 M thiourea, 2% CHAPS, 0.4% DTT and 0.5% carrier ampholytes) was added to a final volume of 450 µl. This mixture was immediately used for isoelectric focussing. A pool of the two samples was prepared and labeled with Cy2 to be used as internal standard on each gel. For this 25 µg protein of each sample (cells grown with *N*-acetylglucosamine and glucose, respectively) were mixed and the labeling was carried out as described above. For 2DE equal amounts of each of the three labeled preparations (*N*-acetylglucosamine, glucose, and internal standard composed of both) were mixed to be run on the same gel. The total amount of protein loaded on each gel was 150 µg.

**Two-dimensional gel electrophoresis and protein visualization**

Isoelectric focussing (IEF) was carried out as described before (Rabus et al. 2002) using the IPGphor™ system and commercial 24-cm-long IPG strips (linear pH gradient of 4 – 7; Amersham Biosciences). Prior to SDS-PAGE, IPG strips were equilibrated with DTT- and iodoacetamide-containing buffers (Görg et al. 2000). For second dimension separation the Ettan™ Dalt II system (Amersham Biosciences) was used. Gels were made of 375 mM Tris/HCl, 0.1% SDS and 12.5% Duracryl). Twelve parallel gels were run at 25°C with a running buffer composed of 25 mM Tris, 192 mM glycine and 0.1% SDS. Electrophoresis was conducted at 180 W and stopped when the bromophenol blue dye front reached the bottom of the gel.

After SDS-PAGE cyanine dye-labeled protein gels were scanned directly between the low fluorescence glass plates using the Typhoon 9400 scanner (Amersham Biosciences). All gels were scanned with a resolution of 100 microns and the photomultiplier tube set at 500 V. Emission filters and laser combinations are listed in Table 2.

Protein spots in gels with unlabeled samples were visualized using colloidal Coomassie Brilliant Blue (cCBB; modified from Doherty et al. (1998), Fast Blue Coomassie stain (Genomic Solutions, Ann Arbor, Michigan, USA), silver (modified from Heukeshoven and Dernick 1988) and SYPRO Ruby (Molecular Probes Inc., Eugene, Oregon, USA). For comparison of protein detection with SYPRO Ruby and silver staining, SYPRO Ruby stained gels were destained over night in 10% (v/v) methanol and 7% (v/v)

glacial acetic acid, and then stained with silver nitrate. Whereas images of SYPRO Ruby stained gels were also acquired with the fluorescence scanner Typhoon, those of silver and cCBB stained gels were collected with the Image Scanner (Amersham Biosciences).

**Table 2.** Emission filters and laser combinations used for detection of fluorescent labels.

| Fluorphore | Emission Filter (nm) | Laser |
|---|---|---|
| Cy2 | 520 BP[a] 40 | Blue2 (488) |
| Cy3 | 580 BP 30 | Green (532) |
| Cy5 | 670 BP 30 | Red (633) |
| SYPRO Ruby | 610 BP 30 | Green (532) or Blue1 (457) |

[a] BP, bandpass

**Image Analysis**

Determination of protein abundance and statistics based on 2D DIGE (see Table 3) were carried out with the DeCyder software package (version 4.0; Amersham Biosciences) according to the following steps. (i) From a single gel three datasets are produced during scanning, corresponding to the Cy2, Cy3 and Cy5 images. First step of spot detection is the creation of electronically merged images for each of the Cy3/Cy2 and Cy5/Cy2 pairs. Spot maps are then created for each merged image (co-detection algorithm). Here the number of estimated spots can be defined by the user. Then the spotmap is superimposed onto the original image pair of a merged image. This results in identical boundaries for the same spots in the original images. (ii) Normalized ratios of volumes (Cy3:Cy2 and Cy5:Cy2) are calculated for each spot and correlated to the gel-specific internal standard (Cy2). (iii) For inter-gel comparison, corresponding spots were matched between all 20 Cy2-images (10 each for Cy3:Cy2 and Cy5:Cy2 pairs, respectively). (iv) After correlation with the internal standard across all Cy2-images, the average ratio of abundance was calculated for each spot. (v) Statistical analysis was performed by using the "standardized log abundance". The calculated T-test value implicated the degree of statistical confidence of the observed differences in abundance. Intra-gel comparison would be facilitated, if spot boundaries were defined for the Cy2-labeled internal standard and then superimposed on the individual samples labeled with either Cy3 or Cy5 (resulting in one instead of two mapsets).

**Table 3.** Ratios of protein abundance and statistical confidence of proteins identified as being differently abundant in cells grown with *N*-acetylglucosamine versus glucose.

| Spot No.[a] | 2D DIGE | | SYPRO Ruby |
|---|---|---|---|
| | Average Ratio[b] | T-test[c] | Average Ratio[d] |
| 1 | 43.9 | 4.3e-19 | 12.0 |
| 2 | 32.4 | 2.3e-14 | 36.5 |
| 3 | 15.7 | 8.6e-13 | 5.9 |
| 4 | 4.5 | 1.1e-20 | 3.8 |
| 5 | 2.9 | 1.5e-12 | 1.9 |
| 6 | 2.2 | 3.6e-13 | 1.6 |
| 7 | 2.1 | 2.4e-12 | 2.0 |
| 8 | 2.0 | 5.0e-13 | 1.8 |
| 9 | 1.9 | 9.7e-18 | 1.5 |
| 10 | 1.9 | 6.4e-10 | 1.0 |
| 11 | 1.9 | 3.3e-14 | 2.6 |
| 12 | 1.9 | 6.3e-12 | 2.7 |
| 13 | 1.8 | 2.0e-12 | 1.5 |
| 14 | 1.8 | 1.9e-14 | 2.1 |
| 15 | 1.7 | 5.3e-14 | 1.6 |
| 16 | 1.6 | 3.5e-10 | 1.6 |
| 17 | 1.6 | 5.3e-10 | 1.8 |
| 18 | -1.7 | 7.6e-12 | 1.2 |
| 19 | -1.8 | 5.2e-11 | -2.1 |
| 20 | -1.8 | 1.0e-13 | 1.1 |
| 21 | -1.9 | 1.3e-13 | -1.4 |
| 22 | -2.0 | 2.8e-14 | -2.7 |
| 23 | -2.1 | 1.4e-14 | -2.0 |
| 24 | -2.6 | 3.7e-14 | -2.8 |

[a] Spot numbers are the same as used in Figure 2 A and B.
[b] Average ratios were calculated from 10 parallel 2D DIGE gels by the DeCyder software. Differential in-gel analysis was based on an "estimated number of spots" of 2500 as guide for spot detection. If for a batch of 5 gels, this number was increased to 4000, the values of average ratio showed differences of +/- 10% in some cases.
[c] The Student's T-Test was used according to the Null-Hypothesis. Here, the probability is calculated, that the abundances of a given protein spot under different conditions is the same.
[d] Average ratios were calculated from two set of gels (each containing 4 gels), by comparison of electronic average gels using the ImageMaster software.

For the analysis of the other gels the ImageMaster software (version 4.01; Amersham Biosciences) was used. When SYPRO Ruby stained gels were analyzed for comparison with results obtained by 2D DIGE / DeCyder (see Table 3), the following steps were applied. (i) The contrast of images was adjusted to allow optimal perception of spots. (ii) Parameters of the "spot detection wizard" were defined such that detection of artifacts (*e.g.* dust particles) was limited. The same parameters were applied for all eight gels analyzed. (iii) Spot finding was manually edited to remove artifacts and to optimize spot boundaries

(steps i to iii were also carried out, when sensitivity of the used stains was determined; Figure 4). (iv) In the reference gel the detection of spots was finally limited to those spots that were identified before by application of DeCyder software to differ in abundance, whereas in all other seven gels detection of all spots was conserved. (v) Then all gels were matched to the selected spots in the reference gel and the matches were manually edited. For example, if a spot could only be detected in the gel set representing one of the two studied growth conditions, its spot boundary was manually inserted in images of the other gel set where the spot was missing. This made it possible to calculate the spot volume ratios for all spots. (vi) Background subtraction was carried out according to "lowest on boundary".

Gel-to-gel variations may cause differences in volumes of the same spots in parallel gels, despite the application of the same amount of total protein to each gel. These differences are leveled out by "normalization", since they do not reflect biological changes. In the DeCyder software normalization is performed after spot detection and calculation of ratios in abundance for each spot in all gels (for details see Alban et al. 2003). In case of 10 analyzed gels a total of 20 Cy3:Cy2 and Cy5:Cy2 ratios were calculated for each spot. During standardization all of these spot-specific ratios were related to one common value for the internal standard (Cy2), which is set at 0 on the log scale for abundance (see Figure 3, bottom horizontal panel). When ImageMaster software was used for image analyses, normalization was carried out differently. For each gel of a matchset, the sum of the total spot volumes was determined and set at 100%. Then the shares of individual spot volumes in the sum of the total spot volumes of a gel were calculated. Share values for individual spots should be constant in a matchset, *i.e.* they should not be affected by varying staining intensities in parallel gels. Changes in protein abundance are then determined by calculating the ratio between sample-specific average share values.

## Results and Discussion

In a previous study we observed by 2DE analysis that cells of *Pirellula* sp. strain 1 grown with *N*-acetylglucosamine contain proteins that are apparently less abundant in cells grown with glucose (Rabus et al. 2002). In the present study we used this inducible *N*-acetylglucosamine metabolism as a model system to evaluate the 2D DIGE system for detailed analysis of changes in protein quantities in this bacterium. Such a proteomic tool will be of great benefit for future studies with this bacterium, when global investigations are able to benefit from the fully sequenced genome.

**Determination of changes in protein quantities using the 2D DIGE system**

In this part of the study we analyzed a batch of 10 parallel gels, consisting of two subsets of 5 gels each. All of these gels were run simultaneously. In the first subset extracts of soluble proteins from *N*-acetylglucosamine grown cells were labeled with Cy3, whereas protein extracts from glucose grown cells were labeled with Cy5. This labeling order was reversed in the second subset of gels. In both subsets, a mixture of protein extracts from both types of samples was labeled with Cy2 and used as internal standard. By changing the labeling order for Cy3 and Cy5, we tried to take into account a possible bias resulting from the different CyDyes. The statistical analysis was carried out with 1688 spots detected essentially in all gels using the DeCyder software. On the basis of a matchset from the 10 gels, changes in protein abundance were regarded as significant, when minimal values for average ratio of 1.5 and maximal values for the T-test of $10^{-9}$ were reached. With this conservative setting of parameters we determined 24 protein spots, the abundance of which differed in cells grown with *N*-acetylglucosamine or glucose. Average ratios ranged from 1.5 to 44 and T-test values from $10^{-9}$ to $10^{-20}$ (Figure 2, Table 3). For all 24 proteins, consistency of matching was manually verified throughout the entire matchset. Details of the DeCyder output are exemplified for three spots (no. 1, 17, 23) in Figure 4. For comparison image analysis was carried out with reduced stringency for the statistical parameters. If the thresholds for average ratio and T-test were altered to values of 1.2 and 0.05, respectively, as previously reported in other studies (Gharbi et al. 2002; Tonge et al. 2001), approximately 387 spots appeared to have different abundance in the two analyzed growth conditions. Thus parameters for statistical stringency defined by the user have a major influence on the outcome of the image analysis. Despite the assumed analytical reliability of the large number of 387 differential protein spots (corresponding to approximately 23% of all detected proteins), a biological meaning should not be directly inferred. It should be taken into account that the only biological difference between the two analyzed samples is the substrate used for growth. From a physiological point of view, mainly alterations on the level of catabolic/anabolic enzymes and some regulatory proteins should occur. However, one would not expect this to translate into such high numbers of differentially abundant proteins. An assessment of the biological significance of the observed changes in protein quantities requires the identification of the respective proteins and further studies.

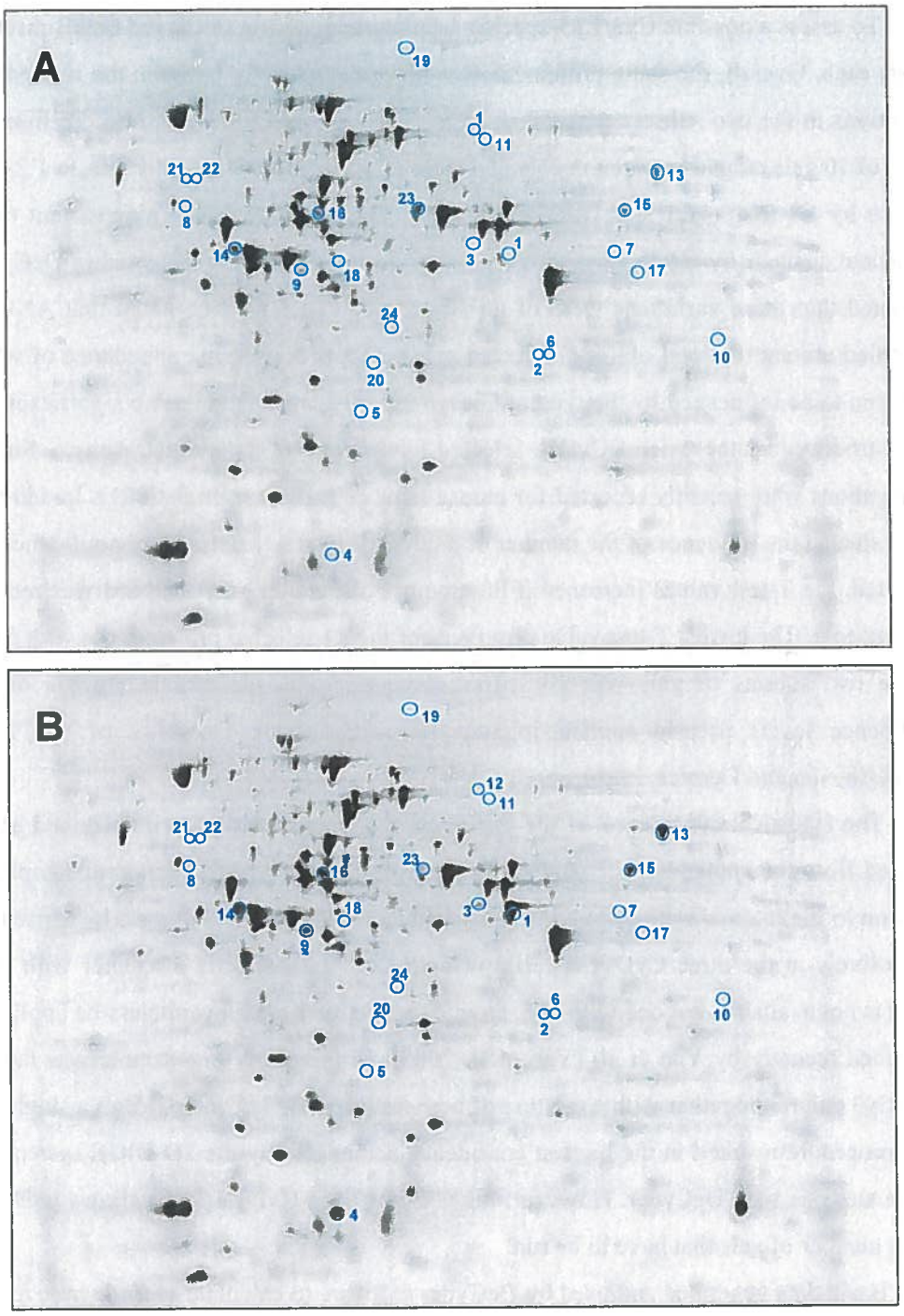



**Figure 2.** Cyanine fluorphor-specific images of a 2DE gel (2D DIGE) of protein samples of *Pirellula* sp. strain 1. Soluble proteins of cells grown with glucose (**A**) and *N*-acetylglucosamine (**B**), were labeled with Cy5 (**A**) and Cy3 (**B**), respectively. In both cases 50 µg of protein were used and isoelectric focussing was carried out with a pH gradient of 4-7. Both samples were simultaneously separated on the same gel. Excitation of the cyanine fluorphores and collection of the emitted fluorescent signals were as indicated in Table 2. Numbers indicate proteins that are differently abundant by a factor of at least 1.5-fold (T-test value $<10^{-9}$ in 10 parallel gels) in the two studied growth conditions: numbers correspond to those used in Figure 3 and Table 3 (listing average ratios and statistical confidence).

To assess a possible Cy3/Cy5-specific labeling bias we also analyzed both subsets of 5 gels each. Overall, the same protein spots were found to differ between the two growth conditions in the two subsets. Variations of the 24 protein spots selected before from the batch of 10 gels ranged between unchanged (1.0) and 1.4-fold (20 spots by 1.0- to 1.2-fold, 3 spots by 1.3-fold and 1 spot by 1.4-fold). Variations did not follow a consistent trend. Statistical analysis by means of ANOVA (analysis of variance; implemented in DeCyder) indicated that these variations were of no significance. It should be noted that ANOVA identified among the total of 1688 detected protein spots only 3, the abundance of which appeared to be influenced by the type of CyDye used for labeling. Thus no significant bias was introduced by the type of CyDye selected for labeling of a particular sample. Similar observations were recently reported for mouse liver cells (Tonge et al. 2001). In addition, we evaluated the influence of the number of analyzed gels on the statistical confidence. As expected, the T-test values increased if the number of parallel gels analyzed was reduced from 10 to 5. The lowest T-test value observed for the 24 selected protein spots with either of the two subsets of gels was $10^{-5}$. This is significantly better than the 5% or 1% confidence levels usually applied in statistical analysis on the basis of the T-test (Heukeshoven and Dernick 1988).

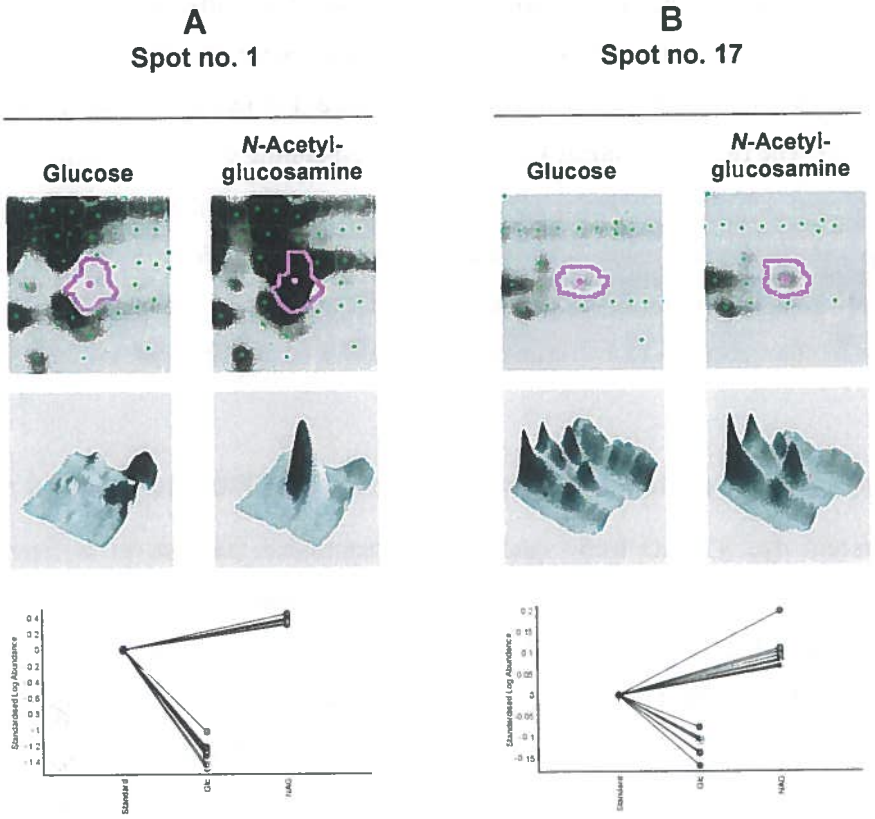
The statistical confidence of the differences in protein abundance described above resulted from the application of an internal standard. To analyze two different samples in addition to the internal standard in a single gel, the image acquisition has to be carried out successively at the three CyDye-specific wavelengths (Table 2). If a scanner with three lasers is not available, but one with two, an internal standard can nevertheless be applied as described recently by Yan et al. (Yan et al. 2002). In this study one sample was labeled with Cy3 and run together with a mixture of both samples labeled with Cy5 on a single gel. This procedure resulted in the highest confidence achievable by the 2D DIGE system and image analysis with DeCyder. However, the use of a third CyDye (Cy2) allows reduction of the number of gels that have to be run.

Basic data generated and used by DeCyder software to calculate average ratio and T-test values for changes in protein abundance can be visually inspected for each spot in each analyzed gel (Figure 3). Spot boundaries defined in the Cy3 and Cy5 channel can be displayed via zoom-in views of the corresponding gel areas to control matching. A 3D view allows to examine the three-dimensional shape of spots and to evaluate the quality of electrophoretic separation and spot detection. A "Graph View" displays the abundance of the selected protein spot under the analyzed conditions in all gels relative to the internal

standard, providing a visual impression of the statistical confidence. The 3D view is especially useful for a refined control of spot boundaries in areas of insufficient electrophoretic separation (Figure 3 B and C). In Figure 3 C respective data is presented for spot no. 23. The two-dimensional view of the corresponding gel section shows that spot no. 23 is part of a large spot. Only the 3D view allows recognition of spot no. 23 as a distinct spot incompletely separated from the neighboring spot. Such information is not directly provided by the ImageMaster software, but has to be manually retrieved by application of the "Profile"-tool displaying two-dimensional profiles of the selected gel areas.

**Determination of changes in protein quantities based on SYPRO Ruby stained gels**

The fluorescent dye SYPRO Ruby qualifies for quantitative analysis of differences in protein abundance, since it combines a low detection limit with a wide linear range (Table 1). The first applications of SYPRO Ruby for 2DE analysis were reported only a few years ago (Berggren et al. 2000; Lopez et al. 2000). Since then this protein stain was used for quantitative differential analysis of protein profiles in standard bacteria such as *Escherichia coli* (Nishihara and Champion 2002) or environmental bacteria (Rabus et al. 2002). With respect to detection limit and linear range, SYPRO Ruby is similar to CyDyes. Some differences in the current application protocols and general characteristics of the two types of fluorescent dyes are obvious. (i) Labeling with CyDyes is carried out prior to and staining with SYPRO Ruby after electrophoresis. (ii) The availability of different CyDyes and their pre-electrophoretic labeling principle allows to run different samples on a single gel. (iii) CyDyes are covalently bound to the ε-amino group of lysine residues, whereas SYPRO Ruby is suggested to interact preferentially with basic amino acids until saturation (Steinberg et al. 2000). (iv) Image analysis of CyDye-stained proteins can involve an internal standard, if DeCyder software is used. This is not possible with SYPRO Ruby stained gels. Considering these differences, the question arises whether application of CyDyes and SYPRO Ruby, respectively, has different effects on the changes in protein abundance determined.



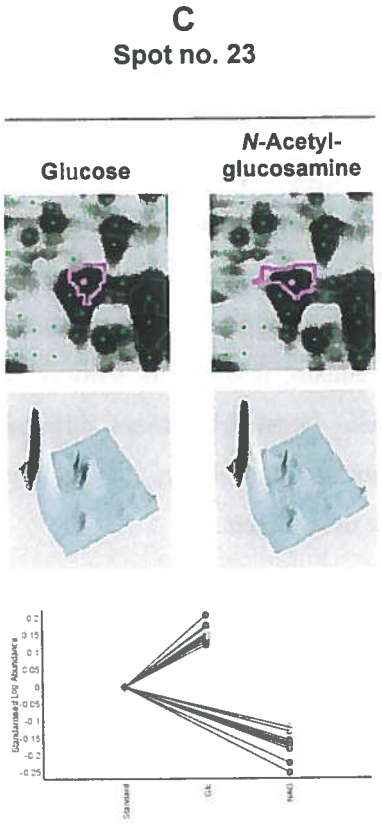




**Figure 3.** Information on differently abundant proteins extracted from acquired gel images by DeCyder software. The three vertical panels (**A**, **B** and **C**) compile the information on the abundance of three selected spots (1, 17 and 23) in cells grown with glucose and *N*-acetylglucosamine, respectively (spot numbering corresponds to that used in Figure 2 and Table 3). The top horizontal panel displays partial views of the images of the 2DE gels shown in Figure 2. Spot boundaries defined by DeCyder software are indicated as green lines. The spot boundaries of the selected protein spots are highlighted with purple lines. The middle horizontal panel displays the 3D images of the selected spots. The 3D images were rotated to allow optimal perception. The bottom horizontal panel shows scattered blot representations ("Graph View") of the log abundances of the selected spots under the two biological conditions (Glc, growth with glucose; NAG, growth with *N*-acetyl-glucosamine) in the analyzed set of 10 parallel 2DE gels. The log of abundance is displayed at the y-axis and related to the normalized standard.

We compared these two types of fluorescent dyes in the model system of the present study, namely the inducible *N*-acetylglucosamine metabolism in *Pirellula* sp. strain 1. For SYPRO Ruby staining, four gels were analyzed for each of the two studied growth conditions (growth with *N*-acetylglucosamine or glucose); all 8 gels were run simultaneously. The protein load was 50 µg per gel. Image analysis was performed with the ImageMaster software, as described in the Materials and Methods section. Within the set of 24 proteins analyzed the variance of the volumes determined for a single spot in the parallel gels was different. The coefficients of variance for the normalized volumes ranged between 102% (spot no. 10 in "Glucose"-gels) and 1% (spot no. 4 in *N*-acetyl-glucosamine-gels), indicating that the reliability of these values is not comparable. After spot detection, the 24 spots identified by the 2D DIGE process as differently abundant (see preceding paragraph and Table 3) were selected for quantitative analysis after normalization. Results (listed in Table 3) show that the changes in abundance determined with the two types of fluorescent dyes were overall fairly similar. In most cases, differences in abundance varied 1 to 1.5-fold. The maximal difference in abundance of 3.7-fold was observed for spot no. 1. These results indicate that the absolute values for changes in abundance of a particular protein can differ, when different types of fluorescent dyes are used. Therefore, within a given experiment quantitative data derived from different types of stains and/or methods of image analysis should not be mixed.

**Sensitivity of CyDyes in comparison with that of other protein stains**

Sensitivity of protein stains is often defined as the minimum amount of protein detectable with a particular stain (see Table 1). Sensitivity can also be understood as the maximal number of protein spots detectable with a particular dye on a 2DE gel. We evaluated the latter definition of sensitivity for the stains used in this study. For each stain (Cy5, Cy3, Cy2, SYPRO Ruby, silver, and colloidal Coomassie) we analyzed 3 parallel gel images. To allow optimally comparable conditions, solubilization of proteins was carried out with the lysis buffer specific for labeling with CyDyes and all gels were loaded with 50 µg protein. After electrophoresis and staining, respectively, the same section in the low molecular range and middle of the pH gradient (4-7) was selected for spot detection in all gels (Figure 4). In this area of 2DE gels protein spots are generally separated best and spot boundaries can be defined most easily. Image analysis was performed with the ImageMaster software. For spot detection the "spot detection wizard" was used, followed by manual control and adjustments. The maximal number of detectable spots in the selected area was determined

for each gel and an average value was calculated from the three parallel gels analyzed for each stain (Figure 4). With 459 and 443 detected spots, silver and SYPRO Ruby appeared to be most sensitive for staining proteins of *Pirellula* sp. strain 1. The CyDyes allowed detection of only somewhat fewer proteins, namely 418 (Cy5), 399 (Cy2) and 387 (Cy3). This agrees with an earlier report by Tonge et al. (Tonge et al. 2001). As expected, Coomassie stained gels revealed lower numbers of detectable spots than the aforementioned stains. However, if gels were stained with colloidal Coomassie, essentially according to the method described by Doherty et al. (1998), considerably more protein spots were detected (269) than with the commercial staining kit (168). As expected from the described detection limits of the used stains (Table 1), spots specifically detected with the sensitive dyes were those of low abundance. Comparing these numbers it should be considered, that the sensitivity of a particular stain can vary between different protein species, and that comparison of absolute numbers of detection limits reported in different studies is often hampered by the variety of protocols used for staining.

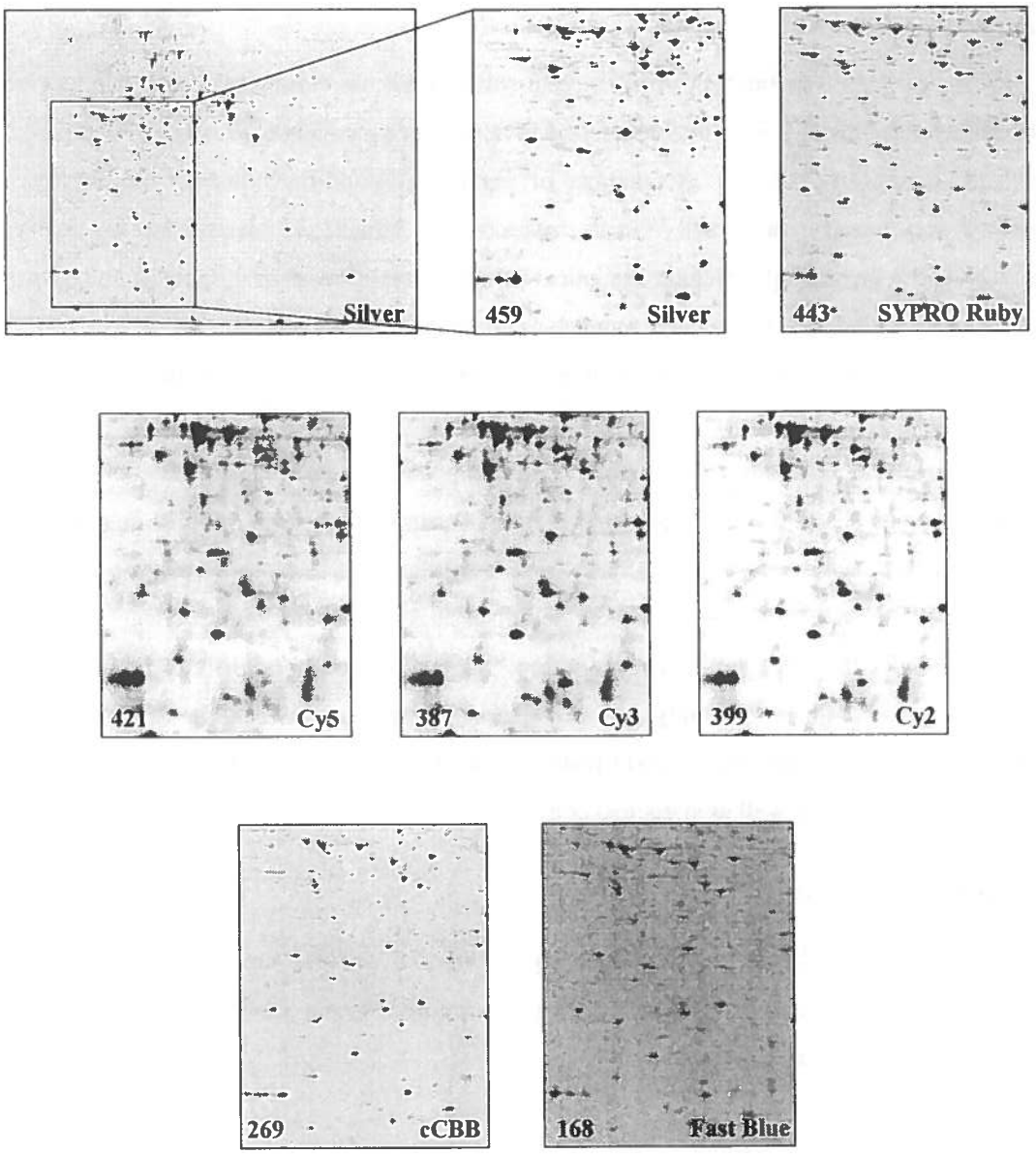



**Figure 4.** Partial views of 2DE gels from protein samples of *Pirellula* sp. strain I stained with different dyes. In each case the total amount of protein used for 2DE was 50 µg. Applied stains were silver nitrate, the fluorescent dyes SYPRO Ruby, Cy5, Cy3, Cy2, and colloidal Coomassie ((cCBB) according to Doherty et al. (1998) and (Fast Blue) using the commercial kit Fast Blue). The type of stain is indicated in the bottom right corner of each gel section. For each stain the maximal number of detectable protein spots was determined within the shown section of 3 parallel gels; the average numbers are indicated in the bottom left corner of each gel section. The darker image of the gel section stained with Fast Blue reflects the higher background produced by this staining kit.

## Conclusions

The 2D DIGE system together with DeCyder software allows determination of differences in protein abundance. The degree of statistical confidence, however, depends on user defined parameters, such as the number of parallel gels, or the minimal and maximal values for average ratio and T-test, respectively, which are demanded to justify significance. In case of stringent statistical parameters the number of recognized spots differing in abundance was greatly reduced as compared to the scenario with less stringent parameters. Even if the differences in protein abundance may be significant from an analytical point of view, the biological meaning has to be verified by means of protein identification and correlation with the genomic and physiological context. The comparative analysis of batches of 5 and 10 gels in this study, demonstrated that 5 parallel gels should be sufficient to achieve results with reasonable statistical confidence. To take into account also sample-to-sample variations one may also analyze several independent samples obtained under the same biological condition. The co-separation of different samples in combination with an image analysis implementing an internal standard by the 2D DIGE system clearly surpassed the SYPRO Ruby / ImageMaster system with respect to work- and time-efficiency as well as statistical confidence.

## Acknowledgements

We would like to thank Rudolf Amann and Friedrich Widdel for general support of proteomic work at our institute. This study was supported by the BMBF, the Max Planck Society and Amersham Biosciences.

## References

**Alban, A., S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie.** 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics **3**:36-44.

**Berggren, K., E. Chernokalskaya, T. H. Steinberg, C. Kemper, M. F. Lopez, Z. Diwu, R. P. Haugland, and W. F. Patton.** 2000. Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. Electrophoresis **21**:2509-2521.

**Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

**Doherty, N. S., B. H. Littman, K. Reilly, A. C. Swindell, J. M. Buss and N. L. Anderson.** 1998. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis **19**:355-363.

**Eason, C., C.W. Coles and G. Gettinby.** 1989. Mathematics and statistics for the biosciences. John Wiley & Sons. New York.

**Eggeling, F. von, A. Gawriljuk, W. Fiedler, G. Ernst, U. Claussen, J. Klose, and I. Römer.** 2001. Fluorescent dual colour 2D-protein gel electrophoresis for rapid detection of differences in protein pattern with standard image analysis software. Int. J. Mol. Med. **8**:373-377.

**Gharbi, S., P. Gaffney, A. Yang, M. J. Zvelebil, R. Cramer, M. D. Waterfield, and J. F. Timms.** 2002. Evaluation of two-dimensional difference gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol. Cell. Proteomics **1**:91-98.

**Görg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss.** 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **21**:1037-1053.

**Heukeshoven, J., and R. Dernick.** 1988. Improved silver staining procedure for fast staining in PhastSystem Development Unit. I. Staining of sodium dodecyl sulfate gels. Electrophoresis **9**:28-32.

**Kirchman, D. L., and J. White.** 1999. Hydrolysis and mineralization of chitin in the Delaware Estuary. Aquat. Microb. Ecol. **18**:187-196.

**Klose, J.** 1999. Large-gel 2-D electrophoresis. Methods Mol. Biol. **112**:147-172.

**Lauber, W. M., J. A. Carroll, D. R. Dufield, J. R. Kiesel, M. R. Radabaugh, and Malone J. P.** 2001. Mass spectrometry compatibility of two-dimensional gel protein stains. Electrophoresis **22**:906-918.

**Lopez, M. F., K. Berggren, E. Chernokalskaya, A. Lazarev, M. Robinson, and W. F. Patton.** 2000. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein

detection in two-dimensional gels and identification by peptide mass profiling. Electrophoresis **21**:3673-3683.

**Nishihara, J. C., and K. M. Champion.** 2002. Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. Electrophoresis **23**:2203-2215.

**Patton, W. F.** 2002. Detection technologies in proteome analysis. J. Chromatogr. B. **771**:3-31.

**Rabus, R., D. Gade, R. Helbig, M. Bauer, F. O. Glöckner, M. Kube, H. Schlesner, R. Reinhardt, and R. Amann.** 2002. Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain I by a proteomic approach. Proteomics **2**:649-655.

**Rabus, R., V. Brüchert, J. Amann, and M. Könneke.** 2002. Physiological response to temperature changes of the marine, sulfate-reducing bacterium *Desulfobacterium autotrophicum*. FEMS Microbiol. Ecol. **42**:409-417.

**Schlesner, H.** 1994. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from various aquatic habitats using dilute media. Syst. Appl. Microbiol. **17**:135-145.

**Skynner, H. A., T. W. Rosahl, M. R. Knowles, K. Salim, L. Reid, R. Cothliff, G. McAllister, and P. C. Guest.** 2002. Alterations of stress related proteins in genetically altered mice revealed by two-dimensional differential in-gel electrophoresis analysis. Proteomics **2**:1018-1025.

**Steinberg, T. H., E. Chernokalskaya, K. Berggren, M. F. Lopez, Z. Diwu, R. P. Haugland and W. F. Patton.** 2000. Ultrasensitive fluorescence protein detection in isoelectric focusing gels using a ruthenium metal chelate stain. Electrophoresis **21**:486-496.

**Tonge, R., J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, and M. Davison.** 2001. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics **1**:377-396.

**Ünlü, M., M. E. Morgan, and J. S. Minden.** 1997. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electrophoresis **18**:2071-2077.

**Yan, J. X., A. T. Devenish R., Wait, T. Stone, S. Lewis, and S. Fowler.** 2002. Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of *Escherichia coli*. Proteomics **2**:1682-1698.

**Zhou, G., H. Li, D. DeCamps, S. Chen, H. Shu, Y. Gong, M. Flaig, J. W. Gillespie, N. Hu, P. R. Taylor, M. R. Emmert-Buck, L. A. Liotta, E. F. Petricoin III, and Y. Zhao.** 2002. 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. Mol. Cell. Proteomics **1**:117-123.

3

# Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA/DNA-hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov., and an emended description of the genus *Pirellula*

Heinz Schlesner[1]*, Christina Rendsmann[1], Brian J. Tindall[2], Dörte Gade[3], Ralf Rabus[3], Stefan Pfeiffer[4] and Peter Hirsch[1]

Manuscript accepted for publication in
International Journal of Systematics and Evolutionary Microbiology

[1] Institut für Allgemeine Mikrobiologie der Universität Kiel, Kiel, Germany
[2] DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany
[3] Max-Planck-Institut für Marine Mikrobiologie, Bremen, Germany
[4] Zentrale Mikroskopie der Universität Kiel, Kiel, Germany
* Corresponding author

## Abstract

Ninety-seven strains of budding bacteria originating from various aquatic habitats and morphologically resembling planctomycetes were investigated taxonomically. Taxonomic differentiation was based on DNA/DNA-hybridization, physiological properties and chemotaxonomic tests. Nineteen hybridization groups, containing 79 of the tested strains, were established. Eighteen strains, however, did not fit into any of these groups. We describe *Rhodopirellula baltica* gen. nov., sp. nov. (strain SH $1^T$ = IFAM $1310^T$ = DSM $10527^T$ = NCIMB 13988), and transfer *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov. (SH $106^T$ = IFAM $1313^T$ = DSM $3645^T$ = ATCC $49069^T$). An emended description of the genus *Pirellula* is also provided. Differentiation between *Rhodopirellula baltica*, *Blastopirellula marina* and *Pirellula staleyi* was achieved by the integration of morphological, physiological, chemotaxonomic and genetic characteristics.

## Introduction

The budding peptidoglycan-less bacteria of the phylum *Planctomycetes* (Garrity and Holt 2001) comprise the order *Planctomycetales*, with the single family *Planctomycetaceae* (Schlesner and Stackebrandt 1986) and the four genera *Planctomyces*, *Pirellula*, *Gemmata*, and *Isosphaera* (Staley et al. 1992). The planctomycetes were considered as unculturable bacteria for many decades. Following the first report of a pure culture (Staley 1973), however, a large number of strains have been isolated from a variety of aquatic habitats differing considerably in salinity, pH, or the availability of nutrients (Schlesner 1994). Other sources of strains were tissue cultures (Fuerst et al. 1991) and postlarvae of the giant tiger prawn *Penaeus monodon* (Fuerst et al. 1997) and the tissue of the Mediterranean sponge *Aplysina aerophoba* (Gade et al. in press). Recently, the isolation of *Gemmata*-like bacteria from soil was reported (Wang et al. 2002). Detailed information on the history and biology of the planctomycetes can be found in the excellent review of Fuerst (1995).

Studies on the evolutionary position of these organisms indicated a great inter- and intrageneric heterogeneity in the 16S rDNA sequences (Stackebrandt et al. 1986; Ward et al. 1995; Fuerst et al. 1997; Griepenburg et al. 1999). Sequence similarity values of the 16S rDNA of *Pirellula staleyi*, *Pirellula marina* (here described as *Blastopirellula marina*) and *Pirellula* sp. SH $1^T$ (here described as *Rhodopirellula baltica*) were below 90 % (Table 3), supporting the differentiation into three genera. At present it is generally accepted that individual species of the same genus should have 16S rDNA sequence similarity of more than 95 % (Devereaux et al. 1990; Fry et al. 1991; Stackebrandt and Goebel 1994). The 16S rDNA sequence heterogeneity was further supported by analysis of the cell wall components (König et al. 1984; Liesack et al. 1986), phospholipids (Kerger et al. 1988; Sittig and Schlesner 1993) and polyamines (Griepenburg et al. 1999).

An unusual characteristic of planctomycetes appears to be their large genome sizes. Initial investigations of genome sizes of budding bacteria was performed using DNA renaturation kinetics, and indicated that *Pirellula* spp. and *Planctomyces* spp. had considerably larger genomes than most of the other strains of budding bacteria examined, which belong to the genera *Hyphomicrobium*, *Hyphomonas*, *Filomicrobium* and *Pedomicrobium* (Kölbel-Boelke et al. 1985). These results have been confirmed by the recent publication of the complete genome sequence of *Rhodopirellula baltica* (Glöckner et al. 2003; www.regx.de). With a genome size of 7.145 Mb it is one of the largest circular bacterial genomes known to date. In addition, ongoing genome sequencing projects with

*Gemmata obscuriglobus* (www.tigr.org) and *Gemmata* sp. Wa 1-1 (wit.integratedgenomics.com) support the unusually large genome sizes of species belonging to the planctomycetes. It is generally assumed that early genomes were probably small in size and that larger genomes arose later in evolution as a result of events such as gene duplication or the acquisition of additional genes from other organisms. While it is also known that "organisms" such as obligate intracellular symbionts or plastids may also reduce their genome size by gene loss – evolutionary reversal – a pre-requisite is that the ancestors must have passed through a stage where the genome size was larger (Anderson and Anderson 1999; Blanchard and Lynch 2000; Gil et al. 2002). While planctomycetes have been variously described as rapidly evolving or ancient, it would appear that the large genome size, at least is probably not a "primitive" feature.

In the light of these studies on the genomes of various planctomycetes it is becoming important to re-examine our appreciation of the biology and diversity of this group of organisms. In this study, we investigated the taxonomic positions of ninety-seven pigmented and unpigmented *Pirellula*-, *Planctomyces*- and *Gemmata*-like strains isolated from various aquatic habitats, including the type-strains of *Pirellula staleyi*, *Blastopirellula marina*, *Planctomyces maris* and *Gemmata obscuriglobus*. *Rhodopirellula baltica* was selected as representative of a group of isolates that were genetically related at the species level. This group contained 22 pigmented isolates from brackish water and 2 strains isolated from the tissue of the sponge *Aplysina aerophoba* (Gade et al. in press). Here we describe *Pirellula* sp. SH $1^T$ as the type strain of *Rhodopirellula baltica* gen. nov., sp. nov. Due to the diversity of the respective strains at the genetic and phenotypic level we propose that *Pirellula marina* be excluded from the genus *Pirellula* and be transferred to the new genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov. This also means that the description of the genus *Pirellula* must be emended to reflect the changed circumscription of this genus.

## Methods

**Media for cultivation.** Various types of media were used for the cultivation of strains with different nutritional requirements. Media differed essentially with respect to carbon and nitrogen sources and with respect to salinity (Table 1).

**Table 1.** Media used for cultivation. Components [%].

| Medium | Peptone | Yeast extract | Glucose | *N*-Acetylglucosamine | Casaminoacids | ASW [#] |
|---|---|---|---|---|---|---|
| M1 *[†] | - | - | - | 0.2 | - | - |
| M13 [‡] | 0.025 | 0.025 | 0.025 | - | - | 25 |
| M13a [§] | 0.025 | 0.025 | 0.025 | - | - | 50 |
| M13f | 0.075 | 0.075 | 0.5 | - | - | 25 |
| M20 [‡¦] | - | - | 0.25 | - | 0.1 | 25 |
| M20c [¦] | - | - | 0.25 | - | 0.1 | - |
| M22 [§] | 0.025 | 0.025 | 0.025 | - | - | - |
| M30 * | - | - | - | 0.2 | - | 25 |
| M30a * | - | - | - | 0.2 | - | 50 |
| M31 * | - | - | - | 0.2 | - | - |
| M40 [¶] | - | - | - | - | 0.1 | 25 |
| M40c [¶] | - | - | - | - | 0.1 | - |

* Described in Staley et al. (1992).
[†] Medium M1 contained $CaCO_3$ to keep an alkaline pH. This medium was solidified with gellan gum Gelrite™ (Kelco, San Diego, USA), rather than with agar (Staley *et al.*, 1992).
[‡] Described by Schlesner (1986).
[§] Described by Schlesner (1994).
[¦] Vitamin-free medium for the purpose of testing for vitamin requirement.
[¶] Medium M40 represents a modification of medium M9 (Schlesner 1986). This medium was used to test for growth substrates. Casamino acids were added as nitrogen source. They did not support growth of the tested strains, when added as sole substrate.
[#] The composition of artificial seawater medium (ASW) was described by Lyman and Fleming (1940).

**Sampling, enrichment and isolation of strains.** Experimental procedures were as previously described (Schlesner 1994).

**Microscopy.** Morphological studies by phase contrast microscopy and electron microscopy were performed as described earlier (Schlesner 1986; Gade et al. in press). For cryofixation and -substitution, bacteria were fixed by high pressure freezing (Hohenberg et al. 1994). Samples in 1 % (w/v) osmium tetroxide in acetone were dehydrated in a freeze substitution unit (AFS Leica, Bensheim, Germany) at temperatures of -90 °C, -60 °C and -30 °C (for 8 h each). Finally, the temperature was increased to 4 °C. The samples were infiltrated with Epon 812 (Epon Kit 45359, Fluka, Germany) by incubation with 30 and

70% (w/v) resin in acetone for 2 h each and then in pure resin for 12 h at 20 °C. Polymerization occurred at 60 °C for 24 h.

**Physiological studies.** Growth experiments with liquid media were essentially carried out as described by Schlesner (1986, 1994). Cultivation was done with 100 ml Erlenmeyer flasks containing 50 ml medium. Cultures were incubated at 25 °C for 3 weeks on a shaker. To test for carbon sources (0.1%, w/v), which supported growth, medium M40 was used for *R. baltica* (SH $1^T$) and *B. marina* (DSM $3645^T$), and medium M40c for *P. staleyi* (ATCC $27377^T$). Anaerobic growth was tested under fermentative conditions and in the presence of nitrate as electron acceptor. Techniques for preparation of media and cultivation under anoxic conditions were performed as previously described (Widdel and Bak 1992). Ascorbate was added to the media as additional reductant (Rabus and Widdel 1995).

Hydrolysis of gelatine was visualized by flooding the plates with hot (70 – 80 °C) saturated ammonium sulfate instead of mercury chloride. Lipase activity was tested according to Kouker and Jaeger (1987), using plates that contained trioleic acid and Rhodamine B. Activities for hydrolysis of starch, casein, esculin and DNA were tested according to standard procedures (Smibert and Krieg 1994). Formation of hydrogen sulfide from thiosulfate was tested as previously described (Schlesner 1986).

Salinity tolerance was studied with liquid media containing increasing portions of artificial seawater (ASW; Lyman and Fleming 1940) to give final concentrations of 0, 6, 12, 25% and then in 25%-steps to the maximum of 300% ASW (100% ASW = 34.5‰ salinity).

**Phospholipids and quinones.** These cell components were extracted from lyophilized cells and analyzed by thin layer chromatography as described by Sittig and Hirsch (1992). In addition polar lipids, respiratory lipoquinones, and fatty acids were extracted and analyzed as described previously (Tindall 1990a, b; Strömpl et al. 1999).

**DNA base ratio and DNA/DNA-hybridization.** Determination of DNA base ratio and DNA/DNA-hybridization experiments were performed as described earlier (Rathmann 1992; Gade et al. in press).

## Results and discussion

### Strains investigated and their habitats

We investigated the taxonomic position of ninety-seven strains that were isolated in this study and that resembled planctomycetes morphologically. Several known species were also included (Table 2). Initially we categorized all strains mainly according to pigmentation of colonies, the motility of daughter cells and DNA/DNA-hybridization groups (Table 2). The new isolates were obtained from a variety of aquatic habitats, including freshwater, brackish and marine water. Most habitats had a slightly alkaline pH (around 8), while water bodies of chalk mines reached maximal pH values of 11.6 (Schlesner 1994). Hypertrophic water samples were retrieved from the Apetlon village ponds (Austria) and the Schrevenpark Pond (Kiel, Germany). The following section deals with the categorization of the entire set of investigated strains.

**Table 2.** Strains investigated in this study, their growth medium, pigmentation, motility, and DNA base ratio. Strains are arranged according to their DNA/DNA hybridization groups. DNA of underlined strains was labeled. nd = not determined.

| Strain | Source | Culture medium | Pigmented colonies | Motile daughter cells | G+C content (mol%) | Hybridization group |
|---|---|---|---|---|---|---|
| ***Rhodopirellula baltica*** | | | | | | |
| SH 1 (IFAM 1310) | Kiel Fjord[1] | 13a | + | + | 55 | I |
| SH 12 (IFAM1428) | Kiel Fjord | 13a | + | + | nd | I |
| SH 26 (IFAM 1429) | Kiel Fjord | 13a | + | + | nd | I |
| SH 28 (IFAM 1430) | Kiel Fjord | 13a | + | + | 55 | I |
| SH 29 (IFAM 1431) | Kiel Fjord | 13a | + | + | 57 | I |
| SH 32 (IFAM 1695) | Kiel Fjord | 13a | + | + | nd | I |
| SH 39 (IFAM 1449) | Kiel Fjord | 13a | + | + | 55 | I |
| SH 121 (IFAM 1433) | Kiel Fjord | 13a | + | + | nd | I |
| SH 123 (IFAM 1434) | Kiel Fjord | 13a | + | + | nd | I |
| SH 126 (IFAM 1435) | Kiel Fjord | 13 | + | + | nd | I |
| SH 155 (IFAM 1436) | Kiel Fjord | 13a | + | + | 55 | I |
| SH 156 (IFAM 1437) | Kiel Fjord | 13a | + | + | 56 | I |
| SH 157 (IFAM 1438) | Kiel Fjord | 13a | + | + | 56 | I |
| SH 159 (IFAM 1455) | Kiel Fjord | 13a | + | + | 56 | I |
| SH 165 (IFAM 1451) | Kiel Fjord | 13a | + | + | 53 | I |
| SH 188 (IFAM 1580) | Kiel Fjord | 13a | + | + | nd | I |
| SH 190 (IFAM 1582) | Kiel Fjord | 13a | + | + | nd | I |
| SH 198 (IFAM 1735) | Kiel Fjord | 13a | + | + | nd | I |
| 99/2 | Kiel Fjord | 13a | + | + | 55 | I |
| SH 385 (IFAM 3255) | Aquarium[2] | 13a | + | + | nd | I |
| SH 386 (IFAM 3184) | Aquarium | 13a | + | + | nd | I |
| SH 398 (IFAM 3246) | Aquarium | 13a | + | + | 55 | I |
| SH 400 (IFAM 3187) | Aquarium | 13a | + | + | nd | I |
| SH 796 | *Aplysina aerophoba* | 13a | + | + | 54 | I |
| SH 797 | *Aplysina aerophoba* | 13a | + | + | 54 | I |
| ***Blastopirellula marina*** | | | | | | |
| *B. marina* DSM 3645[T] | | 13a | - | + | 57 | II |
| SH 150 (IFAM 1557) | Kiel Fjord | 13a | - | + | nd | II |
| SH 152 (IFAM 1547) | Kiel Fjord | 13a | - | + | nd | II |
| SH 166 (IFAM 1570) | Kiel Fjord | 13a | - | + | 55 | II |
| SH 168 (IFAM 1453) | Kiel Fjord | 13a | - | + | 54 | II |
| SH 405 (IFAM 3258) | Sugar processing plant[3] | 13 | - | + | 56 | II |
| SH 406 (IFAM 3277) | Sugar processing plant | 13 | - | + | 56 | II |
| SH 452 (IFAM 3298) | Fjord Schlei[4] | 13 | - | + | nd | II |

**Table 2.** continued

| Strain | Source | Culture medium | Pigmented colonies | Motile daughter cells | G+C content (mol%) | Hybridization group |
|---|---|---|---|---|---|---|
| ***Pirellula staleyi*** | | | | | | |
| *Pirellula staleyi* ATCC 27377 [T] | | 31 | - | + | 57 | III |
| SH 353 (IFAM 3411) | Chalk mine[5] | 1 | - | + | nd | III |
| SH 355 (IFAM 3191) | Chalk mine | 1 | - | + | 56 | III |
| SH 356 (IFAM 3281) | Chalk mine | 1 | - | + | nd | III |
| SH 479 (IFAM 3323) | Pebble wash water[6] | 31 | - | + | 60 | III |
| SH 488 (IFAM 3411) | Lake Felder See[7] | 1 | - | + | nd | III |
| **Others** | | | | | | |
| SH 140 (IFAM 1319) | Kiel Fjord | 13a | + | - | 56 | IV |
| SH 410 (IFAM 3275) | Aquarium | 13a | + | - | 58 | IV |
| SH 411 (IFAM 3186) | Aquarium | 13a | + | - | nd | IV |
| SH 520 (IFAM 3483) | Winogradski column[8] | 30 | + | - | 56 | IV |
| SH 143 (IFAM 1358) | Strande | 13a | + | + | 56 | V |
| SH 453 (IFAM 3390) | Karlsminde Beach[9] | 30 | + | + | 56 | V |
| SH 455 (IFAM 3300) | Karlsminde Beach | 30 | + | + | 54 | V |
| SH 158 (IFAM 1452) | Kiel Fjord | 13a | + | + | 55 | VI |
| SH 282 (IFAM 3016) | Schrevenpark Pond[10] | 22 | + | + | 54 | VII |
| SH 287 (IFAM 3011) | Lake Mondsee[11] | 22 | + | + | 54 | VII |
| SH 241 (IFAM 3207) | Schrevenpark Pond | 22 | + | + | 64 | VIII |
| SH 269 (IFAM 3050) | Schrevenpark Pond | 22 | + | + | 66 | VIII |
| SH 116 (IFAM 1432) | Kiel Fjord | 13a | - | + | 55 | IX |
| SH 118 (IFAM 1447) | Kiel Fjord | 13a | - | + | 55 | IX |
| SH 203 (IFAM 1999) | Gypsum mine[12] | 13 | - | + | 52 | X |
| SH 218 (IFAM 2078) | Gypsum mine | 13 | - | + | 55 | X |
| SH 238 (IFAM 2076) | Schrevenpark Pond | 13 | - | + | 53 | X |
| SH 239 (IFAM 2077) | Schrevenpark Pond | 13 | - | + | nd | X |
| SH 245 (IFAM 2296) | Campus pond, University Kiel | 22 | - | + | 55 | X |
| SH 248 (IFAM 2297) | Campus pond, University Kiel | 22 | - | + | nd | X |
| SH 255 (IFAM 2299) | Schrevenpark Pond | 22 | - | + | nd | X |
| SH 267 (IFAM 3013) | Schrevenpark Pond | 22 | - | + | nd | X |
| SH 277 (IFAM 2247) | Campus pond, University Kiel | 13 | - | + | nd | X |
| SH 280 (IFAM 3014) | Schrevenpark Pond | 22 | - | + | nd | X |
| SH 286 (IFAM 3010) | Lake Mondsee | 22 | - | + | 56 | X |
| SH 217 (IFAM 1945) | Lake Fuhlensee | 13 | - | + | 62 | XI |
| SH 221 (IFAM 2001) | Lake Fuhlensee | 13 | - | + | nd | XI |
| SH 460 (IFAM 3202) | Garbagedump[13] | 31 | - | + | 65 | XI |
| SH 292 (IFAM 3017) | Lake Fuhlensee | 13 | - | + | 61 | XII |
| SH 293 (IFAM 3015) | Lake Fuhlensee | 13 | - | + | 61 | XII |
| SH 295 (IFAM 3018) | Lake Fuhlensee | 13 | - | + | 61 | XIII |
| SH 139 (IFAM 1318) | Kiel Bight | 13a | + | + | 57 | XIV |
| SH 302 (IFAM 3198) | Chalk mine | 13 | + | + | 52 | XV |
| SH 440 (IFAM 3302) | Surface and groundwater lagoon[14] | 31 | + | + | 56 | XV |
| SH 441 (IFAM 3245) | Chalk mine | 31 | + | + | nd | XV |
| SH 458 (IFAM 3391) | Sugar processing plant | 31 | + | + | 56 | XVI |
| SH 404 (IFAM 3288) | Surface and groundwater lagoon | 31 | + | + | 70 | XVII |
| SH 449 (IFAM 3304) | Chalk mine | 1 | + | + | 68 | XVIII |
| SH 380 (IFAM 3182) | Aquarium | 30 | + | + | nd | XIX |
| SH 382 (IFAM 3183) | Aquarium | 30 | + | + | 57 | XIX |
| SH 240 (IFAM 3001) | Campus Pond | 31 | + | + | 59 | - |
| SH 279 (IFAM 3051) | Schrevenpark Pond | 22 | + | + | nd | - |
| SH 331 (IFAM 3248) | Village pond Apetlon[15] | 31 | - | + | nd | - |
| SH 367 (IFAM 3262) | Village pond Apetlon | 30 | + | + | 68 | - |
| SH 381 (IFAM 3269) | Aquarium | 30 | + | + | nd | - |
| SH 389 (IFAM 3270) | greenhouse "Nebelhaus"[16] | 13 | - | + | nd | - |
| SH 391 (IFAM 3241) | Aquarium | 13 | - | + | 54 | - |
| SH 392 (IFAM 3393) | Aquarium | 13a | - | + | nd | - |

**Table 2**. continued

| Strain | Source | Culture medium | Pigmented colonies | Motile daughter cells | G+C content (mol%) | Hybridization group |
|---|---|---|---|---|---|---|
| **Others (continued)** | | | | | | |
| SH 420 (IFAM 3285) | Sugar processing plant | 31 | - | + | nd | - |
| SH 423 (IFAM 3293) | Sugar processing plant | 31 | - | - | nd | - |
| SH 427 (IFAM 3291) | Surface and groundwater lagoon | 31 | + | + | nd | - |
| SH 431 (IFAM 3252) | Sugar processing plant | 13 | - | + | nd | - |
| SH 439 (IFAM 3395) | Surface and groundwater lagoon | 31 | + | + | nd | - |
| SH 461 (IFAM 3267) | Leakage water from garbage[10] | 1 | + | + | nd | - |
| SH 462 (IFAM 3396) | Karlsminde, pond[17] | 30 | + | + | nd | - |
| SH 468 (IFAM 3203) | Sugar processing plant | 30 | + | + | nd | - |
| SH 474 (IFAM 3394) | Lake Fuhlensee[18] | 1 | + | + | nd | - |
| SH 531 (IFAM 3322) | Strohbrück[19] | 1 | + | - | 62 | - |
| SH 567 (IFAM 3405) | Lake Heidensee[20] | 1 | + | + | 54 | - |
| SH 592 (IFAM 3421) | Groß Bakau[21] | 1 | + | + | nd | - |
| *Planctomyces maris* ATCC 29201$^T$ | | 13a | - | + | 51 | |
| *Gemmata obscuriglobus* DSM 5831$^T$ | | 629* | + | + | 64 | |

[1] part of the Baltic Sea
[2] public aquarium of the Institute for Marine Research at University of Kiel. The basins in the aquarium contained water from the Kiel Fjord.
[3] waste water aeration lagoons of the sugar processing plant near Schleswig, Germany. Highly eutrophic.
[4] part of the Baltic Sea
[5] Lägerdorf, Holstein, Germany
[6] gravel pit Rastorfer Kreuz near Kiel
[7] Holstein, Germany, highly eutrophic
[8] made up with sediment and water from the Kiel Fjord
[9] beach at the Baltic Sea near Eckernförde, Schleswig-Holstein, Germany
[10] pond in a public park in Kiel
[11] pond in a public park in Kiel, highly eutrophic because of flocks of of water fowl being fed by visitors of the park
[12] Klein-Nordende, Holstein, Germany
[13] garbage depositing plant Rastorfer Kreuz near Kiel
[14] within 10
[15] hypertrophic village pond east of Lake Neusiedler See, Burgenland, Austria
[16] Botanical garden, Kiel University. Greenhouse with high humidity but moderate temperature
[17] Pond near Karlsminde camping area, about 50 m from the shore of the Baltic Sea on the northern coast of
the Eckernförde Bight, Germany
[18] shallow hypertrophic lake, slightly brackish, north of Kiel, Germany
[19] experiment in a pilot project to convert cattle manure to fertilizer
[20] west of Plön, Germany
[21] fire pond, west of Preetz, Germany
* see DSMZ Catalogue of Strains, 6$^{th}$ ed., 1998 (or www.dsmz.de)

**DNA base ratios and DNA/DNA-hybridization groups**

The broad range of the G+C-content from 52 to 69 mol % (Table 2) indicates the genetic diversity of this group. The G+C-value for most strains was in the range reported for the genera *Planctomyces* (50 – 58 mol %), *Pirellula* (54 – 57 mol %), or *Gemmata* (64 mol %; Staley et al. 1992). Strains SH 404 and SH 449, however, displayed considerably higher values; 69 and 68 mol %, respectively.

DNA/DNA-hybridization experiments with labeled DNA of 20 selected strains (underlined in Table 2) resulted in 19 hybridization groups (HGs). Definition of HGs was on the basis of 70% DNA binding (different strains belonging to the same species; Wayne et al. 1987; Stackebrandt and Goebel 1994). Six HGs were represented by one strain only. The DNA of 18 strains did not hybridize with labeled DNA of any of the 20 selected strains, with *Planctomyces maris* (ATCC 29201$^T$) or with *Gemmata obscuriglobus* (DSM 5831$^T$). These results also point to the genetic heterogeneity among the studied strains. Pigmentation proved to be a useful taxonomic marker, since individual HGs contained exclusively either pigmented or unpigmented strains.

Hybridization group I contained 25 pigmented strains of *Rhodopirellula baltica*, which displayed the typical *Pirellula* morphology. All of them possessed motile daughter cells. The majority of strains (19) were isolated from the Kiel Fjord over a period of 25 years, suggesting that these bacteria belong to the autochtonous microbial community of this habitat. Interestingly, two novel strains were recently isolated from the Mediterranean sponge *Aplysina aerophoba* (Gade et al. in press), indicating the widespread occurrence of this species. In addition, DeLong and coworkers (1993) obtained molecular clones of 16S rDNA genes with high similarity to *R. baltica* from marine snow (Pacific Ocean).

Hybridization group II consisted of 8 strains of *Blastopirellula marina*, including the type strain (DSM 3645$^T$). The main difference to HG I was the lack of pigmentation. While four strains were isolated from the Kiel Fjord, others originated from other brackish habitats.

Hybridization group III consisted of 6 strains of *Pirellula staleyi*, including the type strain (ATCC 27377$^T$). Like HG II colonies were unpigmented. While members of the hybridization groups I and II originated from brackish to marine habitats, group III strains were isolated from freshwater habitats in northern Germany. *P. staleyi* originated from Lake Lansing (Michigan, USA; Staley 1973).

A detailed differentiation of the 19 hybridization groups according to salinity tolerance, hydrolysis of polymers and presence of phosphatidylcholine is summarized in

Table 4. Strains within an individual hybridization group displayed only minor phenotypic differences. Although the presence of menaquinone 6 (MK-6) as the major respiratory quinone did not allow differentiation within the group it allows this group to be distinguished from all other prokaryotes which produce menaquinones of longer chain length. Similar results were found by Sittig and Schlesner (1993). Detailed studies on the fatty acids and polar lipids of a range of strains indicated that it was possible to differentiate distinct groups, pointing to the evolutionary heterogeneity of this group. Among the 97 strains investigated here only 6 had immotile daughter cells. In the following sections we focus on the taxonomic differentiation between *R. baltica*, *B. marina* and *P. staleyi*.

**Morphological characteristics**

The pink coloured *R. baltica* (SH $1^T$) was isolated from the Kiel Fjord, from the same sample as the unpigmented *B. marina* (ex *Pirellula marina*; strain SH $106^T$ = IFAM $1313^T$ = DSM $3645^T$; Schlesner 1986). The morphology of *R. baltica* resembled that of other members of the *Planctomycetales*, in particular of the group of *Pirellula*-like bacteria, as indicated by the mode of budding, the presence of fimbriae and crateriform structures and the holdfast-substance excreted directly from the smaller cell pole (Schlesner and Hirsch 1984; Figure 1 A). In contrast to *B. marina* (Schlesner 1986) the bud was not bean-shaped but was a smaller mirror image of the mother cell. Thin cryosubstituted sections of cells (Figure 1 B) showed membranous structures surrounding the nuclear material and the majority of ribosomes. Analysis of the cross sections of strain SH $1^T$ as well as of SH 796 (Gade et al. in press) revealed several small structures in addition to a large central one. This microscopic appearance differs from the pirellulosome structure described for *P. staleyi* and *B. marina* (Lindsay et al. 2001).

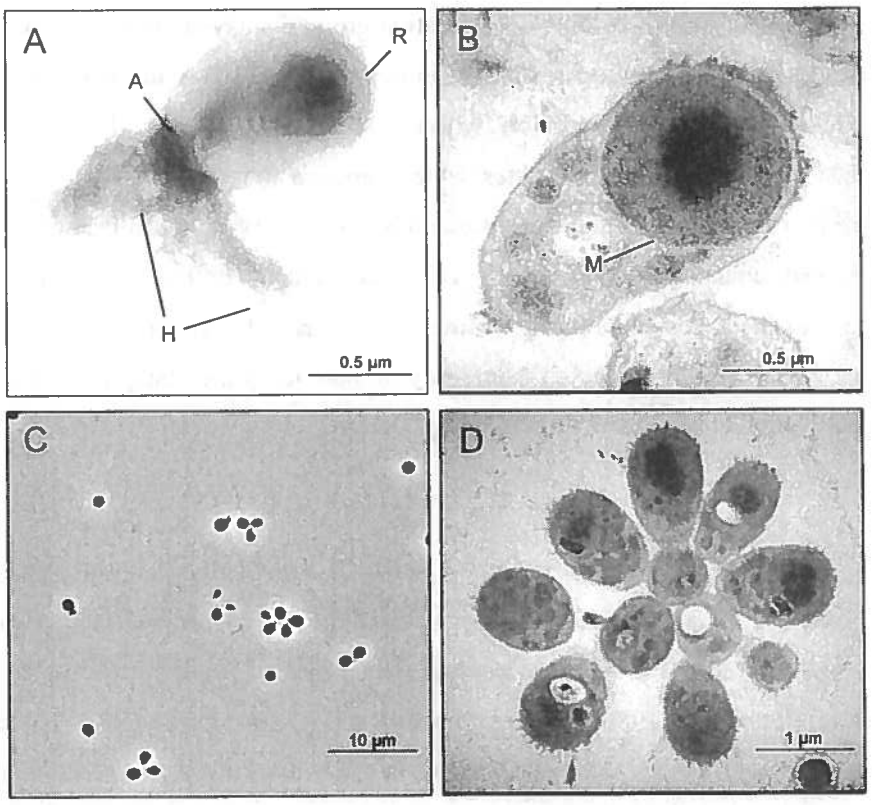



**Figure 1.** Microscopic images of *R. baltica* and related strains. **(A)** Electron-microscopic image of a single cell displaying the polar organization of *R. baltica* cells. *R*, reproduction pole; *A*, attachment pole; *H*, holdfast substance. **(B)** Electron-microscopic image displaying the intracellular compartmentalization. *M*, membrane engulfing the pirellulosome-like structures. **(C)** Light-microscopic image of rosette-like aggregates. **(D)** Electron-microscopic image of a rosette showing the attachment of cells via the attachment poles.

## Physiological properties

The pigment of *R. baltica* could be extracted with methanol and ethanol but not with chloroform, ether or petrolether. The absorption spectrum of a methanol extract had a maximum at 495 nm and two shoulders at 460 and 520 nm and thus showed similarity to carotene. The sulfuric acid test for polyene carotenoids, however, was negative, and carotene is also soluble in chloroform and petroleum ether.

*R. baltica* and *B. marina* can be considered as marine bacteria since they do not grow in freshwater media. Growth of *R. baltica* occurred in media containing rising concentrations (12 – 175%) of ASW (100% ASW corresponds to a salinity of 34.5‰). Similar values were observed for *B. marina.* Essential components of ASW were $Ca^{2+}$, $Na^{+}$ and $Cl^{-}$. In contrast, the freshwater bacterium *P. staleyi* tolerated only up to 50% ASW. *R. baltica* required the addition of vitamin $B_{12}$ to the medium, while *B. marina* and *P. staleyi* were able to grow in vitamin-free media (Table 3).

**Table 3.** Characteristics useful for the differentiation between *P. baltica, B. marina* and *P. staleyi*. + = positive, main component; - = negative, not present; (+) = present in less amounts

| Characteristics | *R. baltica* strain SH1$^T$ | *B. marina* DSM 3645$^T$ | *P. staleyi* ATCC 27377$^T$ |
|---|---|---|---|
| Cell size [µm] | 1.0 – 2.5 × 1.2 – 2.3 | 0.7 – 1.5 × 1.0 – 2.0 | 0.9 – 1.0 × 1.0 – 1.5 |
| Pigmentation | pink to red | unpigmented | unpigmented |
| Salinity tolerance (% ASW) | 12 – 200 | 12 – 175 | 0 – 50 |
| Vitamin requirement | $B_{12}$ | - | - |
| Carbon source utilization | | | |
| Fucose | - | + | + |
| Glycerol | + | + | - |
| Glutamic acid | - | + | - |
| Chondroitin sulfate | + | + | - |
| Hydrolysis of casein | - | - | + |
| Lipase | - | + | - |
| Polyamines [µmol (g dry weight)$^{-1}$] * | | | |
| Putrescine | 19.5 | - | - |
| Cadaverine | 15.3 | - | - |
| Spermidine | -- | - | 0.8 |
| *sym*-homospermidine | 19.6 | 11.3 | 50.2 |
| Phosphatidylcholine | + | - | - |
| Cell wall amino acids (molar ratios) | | | |
| Threonine | 9.0† | 3.8‡ | 3.0† |
| Glutamate | 36.3† | 11.3‡ | 9.0† |
| Cysteine | 9.2† | 1.2‡ | 3.6† |
| Valine | 8.2† | 2.4‡ | 1.7† |
| Fatty acids (in rel. %) | | | |
| 14:0 | 0.5 | - | 4.9 |
| 15:0 | 0.5 | 5.9 | - |
| i16:0 | - | 4.9 | - |
| 16:1 Δ9 | 8.0 | 4.1 | 3.5 |
| 16:0 | 39.2 | 27.5 | 33.8 |
| 17:1 Δ9 | 4.0 | - | 14.4 |
| 17:0 | 1.2 | - | 5.3 |
| 18:1 Δ9 | 40.8 | 26.6 | 26.6 |
| 18:1 Δ11 | 1.6 | 2.3 | 2.0 |
| 18:0 | 4.3 | 2.5 | 3.3 |
| 19:1 Δ11 | - | 2.6 | - |
| 19:0 | - | 2.7 | - |
| 20:1 Δ11 | - | 1.2 | 15.7 |
| Similarity of 16S rDNA sequences (%)§ | 85.0<br>100 | 87.5<br>87.1 | 100<br>85.0 |

* results taken from Griepenburg et al. 1999
† results taken from Liesack et al. 1986
‡ results taken from König et al. 1984
§ results taken from Ward et al. 1995

**Table 4.** Characters useful for differentiation between DNA/DNA hybridization groups.
+ = positive or main component, (+) = weak or present in lower amounts, - = negative or not present, n.d. = not determined., PC phosphatidylcholine.

| HG | Strain | ASW [%] | Casein | DNA | Esculin | Gelatin | Starch | $H_2S$ from thiosulfate | Lipase pH 7 | PC |
|---|---|---|---|---|---|---|---|---|---|---|
| I | *R. baltica* | 12.5 – 200 | - | + | + | + | + | + | - | + |
| II | *B. marina* | 12.5 – 175 | - | + | + | + | + | + | + | - |
| III | *P. staleyi* | 0 – 50 | + | - | + | + | + | + | - | - |
| IV | SH 140 | 12.5 – 150 | - | + | + | + | + | + | - | + |
| V | SH 453 | 12.5 – 325 | + | + | + | + | - | + | - | + |
| VI | SH 158 | 25 – 175 | - | - | - | + | + | + | + | + |
| VII | SH 282 | 0 – 25 | - | - | + | - | - | - | - | - |
| VIII | SH 269 | 0 – 25 | + | - | + | + | + | - | - | + |
| IX | SH 116 | 12.5 – 175 | + | + | + | + | + | + | + | - |
| X | SH 245 | 0 – 125 | - | - | + | + | + | + | - | - |
| XI | SH 217 | 6 – 125 | + | + | + | + | + | + | - | - |
| XII | SH 292 | 0 – 100 | - | - | - | + | + | + | - | - |
| XIII | SH 295 | 0 – 50 | + | + | - | + | + | - | - | + |
| XIV | SH 139 | 25 – 175 | + | - | + | + | + | + | - | + |
| XV | SH 302 | 0 – 50 | + | + | + | + | + | - | - | + |
| XVI | SH 458 | 0 – 25 | + | + | + | + | - | + | - | - |
| XVII | SH 404 | 0 – 25 | + | (+) | + | + | - | - | - | - |
| XVIII | SH 449 | 12.5 – 50 | - | - | - | + | - | - | - | + |
| XIX | SH 382 | 12.5 – 175 | - | + | + | + | - | + | + | + |

Despite the 16S rDNA sequence difference between *R. baltica*, *B. marina* and *P. staleyi*, their physiological properties were very similar. Substrates serving as carbon and energy sources were mainly carbohydrates. *N*-Acetylglucosamine also served as a nitrogen source. An excellent carbon source for *R. baltica* and *B. marina* was chondroitin sulfate. All three organisms displayed catalase and cytochrome oxidase, but no urease activity; they produced $H_2S$ from thiosulfate, but did not produce acetoin or indole. Initial tests on the mesophilic cells of *R. baltica* indicated that they appeared to be strictly aerobic, since they were unable to use nitrate as electron acceptor or to grow fermentatively with glucose. In agreement with a strictly aerobic metabolism, the genome of *R. baltica* revealed no evidence for an anaerobic ribonucleotide reductase. Thus the predicted capacity of lactic acid fermentation (Glöckner et al. 2003) may serve maintenance only.

**Chemotaxonomic characteristics**

Chemotaxonomic markers proved to be more useful for the differentiation between *R. baltica*, *B. marina* and *P. staleyi* than the physiological properties. Chemotaxonomic markers which have been analyzed were the fatty acids and phospholipids (Kerger et al. 1988; Sittig and Schlesner 1993), the amino acid composition of cell walls (König et al. 1984; Liesack et al. 1986) and the polyamine patterns (Griepenburg et al. 1999). In contrast, the analysis of the quinone profile was not useful within the group, since all planctomycetes investigated so far possess menaquinone MK-6 as the only quinone (Sittig and Schlesner 1993; Tindall and Schlesner unpublished). However, the presence of this short chain lipoquinone is useful for delineating this group and also in distinguishing it from other menaquinone-producing prokaryotes with longer isoprenoid side chains. It should be noted that the chemical composition of the cells (polar lipids, fatty acids, and respiratory lipoquinone composition) provides a way of differentiating organisms within this group, but also indicates that the planctomycetes are chemically distinct from any other taxa examined to date. Detailed analyses of the fatty acids examined in this study are given in Table 3. All species produce 16:1 Δ9, 16:0, 18:1 Δ9, and 18:0 fatty acids. While these fatty acids are fairly common in members of the α-, β-, and γ-subclasses of the *Proteobacteria*, this combination, together with the presence of MK-6, clearly distinguishes these species from these major evolutionary groups. Among the strains examined the presence/absence of 14:0, 15:0, i-16:0, 17:1 Δ9, 17:0, 20:1 Δ11 can be used to distinguish between different taxa. While Kerger et al. (1988) reported on the presence of hydroxy fatty acids in planctomycetes, they only deduced that they originated from lipopolysaccharides, providing no direct proof. In this study the methods used would detect the presence of lipopolysaccharide derived hydroxy fatty acids present in *Escherichia coli*. Thus we conclude that, the absence of measurable amounts of hydroxy fatty acids is indicative of the absence of significant amounts of lipopolysaccharides in the cell wall, despite the fact that these organisms are Gram-negative. Sittig and Schlesner (1993) provided the first indications on the chemical heterogeneity of this group, but the full significance of this is now evident when different (phenotypic and genetic) data sets are integrated.

The polar lipid patterns, determined by two dimensional thin layer chromatography are shown in Figure 2. The polar lipid compositions of the three type strains are clearly different. Investigations of a wider range of strains (Sittig and Schlesner 1993; Tindall and

Schlesner unpublished) indicate that this chemical diversity also correlates well with the 16S rDNA diversity. Thus the presence of phosphatidylcholine in *R. baltica* is not just a feature of this species, but is also to be found in other strains which group with this species in 16S rDNA studies. Similarly the polar lipid pattern of *P. staleyi* and *B. marina* also indicate features which allow them not only to be differentiated from *R. baltica*, but also from one another. Using spray reagents which visualise all lipid-like material, it is evident that a significant percentage of the cellular lipids are novel. To date there are no indications that these unidentified lipids are present in any other taxa outside the planctomycetes.

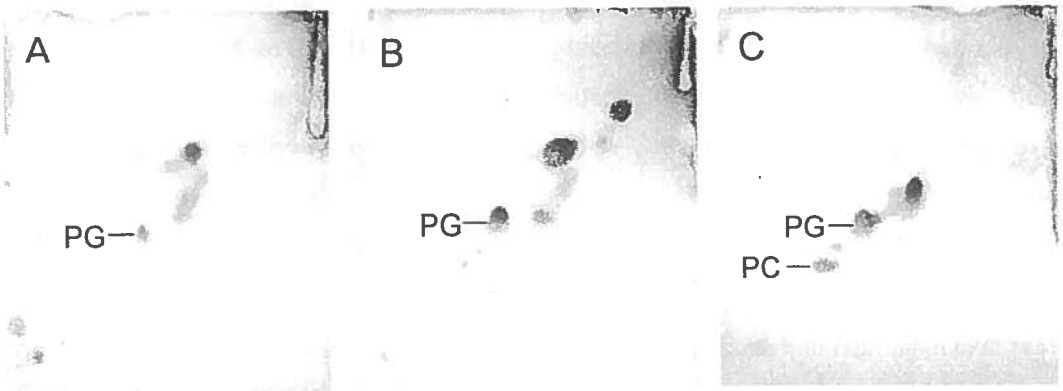



**Figure 2.** Two-dimensional thin layer chromatograms of the polar lipids. **(A)** *Blastopirellula marina*, **(B)** *Pirellula staleyi* and **(C)** *Rhodopirellula baltica*. All polar lipids were stained with 5% (v/v) ethanolic molybdophosphoric acid. Solvents: chloroform:methanol:water (65:25:4 v/v/v), first dimension; chloroform:methanol:acetic acid:water (80:12:15:4 v/v/v/v), second dimension. PG = phosphatidylglycerol; PC = phosphatidylcholine. Unlabeled lipids could not be fully characterized or assigned to known structures.

### Resistance to antibiotics

All three strains were resistant to ampicillin and penicillin (1000 µg/ml), cephalotin (100 µg/ml), streptomycin (500 µg/ml) and cycloserine (100 µg/ml), but not to tetracycline (10 µg/ml were lethal).

### Genome of *R. baltica*

The complete sequence of the 7.145 Mb genome of *R. baltica* (SH $1^T$) was only recently reported (Glöckner et al. 2003). Sequence analysis revealed the complete genetic blueprint for glycolysis, pentose phosphate cycle and TCA cycle, which agrees with the specialization in carbohydrate utilization, found in the present study for this strain. A surprising finding was the presence of more than 100 genes possibly coding for sulfatases. One may speculate that growth with chondroitin sulphate, as found here, may require the activity of a specific sulfatase liberating the carbohydrate moiety. First insights into the

regulation of carbohydrate metabolism were recently obtained by a proteomic approach (Rabus et al. 2002).

The presence of known phospholipids (phosphatidylcholine and phosphatidylglycerol), together with the presence of novel compounds (of as yet unknown structures) will provide a stimulus for examining the biosynthesis and gene regulation of these cellular components, as well as locating the genes responsible. The presence of phosphatidylcholine in *R. baltica* as the sole nitrogen-containing phospholipid is interesting, since it is generally accepted that phosphatidylcholine is synthesized via progressive methylation of phosphatidylethanolamine (Pieringer 1989). Polar lipid patterns containing phosphatidylethanol and phosphatidylcholine are typical of certain actinomycetes and some major evolutionary groups within the α-subclass of the *Proteobacteria* (Ratledge and Wilkinson 1989). However, there are currently no reliable reports of prokaryotes containing phosphatidylcholine as the sole nitrogen-containing phospholipid, and this may be indicative of an alternative pathway leading to the synthesis of phosphatidylcholine in these organisms.

In re-examining the taxonomy of the *Pirellula* group within the planctomycetes we have attempted to integrate as much of our current knowledge about this group as possible. Although DNA/DNA-hydridization studies are generally considered to be problematic from a methodological standpoint, and with respect to determining which fragments of DNA are involved in binding, this method still serves as one of the best ways of gaining an indirect insight into overall similarities between genomes. The results presented in this work clearly indicate that the strains currently available in pure culture constitute not only a diverse range of strains based on 16S rDNA studies, but that they are represented by numerous, different species. Differences in16S rDNA sequence of > 3% (Stackebrandt and Goebel 1994) are generally indicative of different species (when the species groups are tested by DNA/DNA-hybridization). Thus differences in the 16S rDNA sequence are indicative of differences at the "genetic level", which will certainly be evident in the genomes of the organisms concerned. Unfortunately, this genetic (and of course evolutionary) diversity does not appear to be reflected in the physiology of the organisms concerned, as experimentally determined by classical phenotypic tests. While there are clearly marine and freshwater strains, their ability to utilize the range of substrates tested would suggest that members of this group are rather "uniform". However, this uniformity at the level of substrate utilization does not tell us anything about the potential diversity in the underlying biochemical pathways, nor in the structural diversity of the enzymes

concerned. Further studies of genomes of different species and different strains within the planctomycetes will help to elucidate this point.

In contrast to the apparently "limited" physiological diversity, the diversity in chemical composition indicates that the use of this polyphasic taxonomic approach, as in other prokaryotes, reflects the evolutionary diversity which we also detect via constrained elements, such as the ribosome. Thus, strains which share a high degree of genetic similarity (as reflected by DNA/DNA-hybridization and high 16S rDNA similarity values) are difficult to distinguish chemically, but with increasing genetic diversity (as reflected by decreasing DNA/DNA-hybridization values and decreasing 16S rDNA sequence similarity) the chemical differences become increasingly evident. In such instances changes in the chemical composition may reflect either changes in regulatory mechanisms or in the biochemical pathways leading to the synthesis of the end products. Such changes are as significant in the evolution of the cell as the changes in the sequences of genes, such as the 16S rDNA. Taking these aspects into consideration we suggest that it is possible to examine their significance in the evolution of prokaryotes and the taxonomy upon which this is based.

While the planctomycetes have been described as rapidly evolving or ancient, it is evident from the genome size that this is probably representative of a later stage in evolution and it is debatable whether this is a feature of a "primitive" group of organisms. Arguments centering on the fact that the planctomycetes are rapidly evolving (Liesack et al. 1992), are based on Simpson's work on "tempo and mode" in evolution (Simpson 1944). Key arguments in favour of rapid evolution are the large 16S rDNA differences between species within the genus *Planctomyces* and *Pirellula* as well as the low 16S rDNA similarity values between genera (Liesack et al. 1992). In addition, idiosyncrasies in the 16S rDNA sequence, taken together with other peculiarities of the planctomycetes are used in support of the hypothesis that they are rapidly evolving. However, both Simpson (1944) and Mayr (1969) have drawn on a geological time scale in evaluating rates of evolution, a feature usually missing in the majority of gene sequence comparisons. The problem is compounded by the fact that calibration rates are ultimately based on a known fossil record (Doolittle 1997; Doolittle et al. 1996; Feng et al. 1997; Lee 1999), and that even then rates may vary between lineages (Ochman et al. 1999; Solits et al. 2002). In addition, there are no reliable reports of "molecular fossils" which are older than 50-100 million years (Poinar et al. 1996; Willerslev et al. 2003). Such problems have been highlighted by Sneath (1974) and are beginning to be discussed again.

In recent years, detailed three dimensional structures of ribosomes have been published, which indicated the significance of not only secondary structure (in both the proteins and RNA molecules), but also the diverse close interaction between RNA-RNA and RNA-proteins (Ban et al. 2000; Brimacombe 2000; Schluenzen et al. 2000; Wimberly et al. 2000). Thus, the idiosyncrasies in the 16S rDNA sequence should be considered in the light of these interactions, and not based on the gene sequence alone. Similarly the fact that the planctomycetes are characterised by a set of other peculiarities (to which we can also add the polar lipid composition) certainly serves to underline the uniqueness of this group of organisms, but are not necessarily indicative of rapid rates of evolution. Morse et al. (1996) have also called into question the fact that *Oenococcus oeni* is rapidly evolving.

Based on the data presented here we consider the planctomyces to be diverse in a range of properties, and that this diversity is a reflection of extensive evolution. While we cannot easily say whether evolution of this group has taken place over short or long periods of geological time, it is evident that the current taxonomy does not reflect the diversity which we find in this group. We, therefore, propose a number of taxonomic changes, based on the closer study of a limited number of strains, which may also further serve as the basis for evaluating the systematics of this group of organisms.

**Properties differentiating the genera *Rhodopirellula*, *Blastopirellula* and *Pirellula***

Properties of the type strains of *R. baltica*, *B. marina* and *P. staleyi* that allow differentiation of the three genera are summarized in Table 3. Pigmented cells are only observed with *R. baltica*. Growth of *R. baltica* and *B. marina* requires high concentrations of sodium chloride and calcium, whereas *P. staleyi* has only limited tolerance to ASW. Even though all type strains grow with carbohydrates, some difference can be noted. *R. baltica* and *P. staleyi* are not able to utilize fucose and chondroitin sulfate, respectively. Only *P. staleyi* can utilize glutamic acid and hydrolize casein, while lipase activity can only be observed with *B. marina*. Polyamine patterns differ between the type strains. While all contain *sym*-homospermidine, putrescine and cadaverine are only found in *R. baltica*, and spermidine only in *P. staleyi*. Only *R. baltica* possesses phosphatidylcholine. Differences are also observed with respect to the molar ratios of the cell wall amino acids. The fatty acid patterns of the three organisms differ as follows: *R. baltica* and *P. staleyi* lack i16:0, 19:1Δ11 and 19:0, which are found in *B. marina*. In contrast, 17:1Δ9 and 17:0 are present in the former two while absent in the latter. The intracellular compartmentalization is formed by different structures. The pirellulosomes described for

*B. marina* and *P. staleyi* (Lindsay et al. 2001) show a single large structure in the electronmicroscopic image, whereas multiple smaller structures are visible in addition to a large one in case of *R. baltica*. These considerable differences between the type strains are also reflected on the genetic level as indicated by DNA/DNA-hybridization (Table 2) and the low similarity value of less than 90% of the 16S rDNA sequences (Table 3).

**Emended description of the genus *Pirellula* Schlesner and Hirsch 1987**

The description of the genus *Pirellula* is largely based on physiological, biochemical, and morphological properties, and it would be appropriate to emend the description to take into account both additional data and our changing appreciation of the taxonomy of this group. The biochemical, physiological, and morphological characteristics are given in Schlesner and Hirsch (1984). The major polyamine is *sym*-homospermidine. The major respiratory lipoquinone present is MK-6. The major phospholipid present is phosphatidylglycerol. A number of other lipids are present, which have characteristic Rf values, but whose structures are not currently known. However, the lipid pattern is characteristic of this genus. The major fatty acids present are 14:0, 16:1 Δ9, 16:0, 18:1 Δ9, 18:1 Δ11, 18:0, and 20:1 Δ9. It is also evident that the 16S rDNA sequence similarity values are of significance in delineating this genus. Nevertheless the extent cannot be defined at present, since strains with less than 95% sequence similarity to members of this genus probably should be placed in separate genera. The type species of the genus is *Pirellula staleyi*.

**Description of *Rhodopirellula* gen. nov.**

*Rhodopirellula* (Rho.do.pi.rel´lu.la Gr. neut. n. rhodon, a rose; N.L. fem. n. Pirellula, name of a bacterial genus; N.L. fem. n. Rhodopirellula, a red Pirellula).

Cells are ovoid, ellipsoidal or pear-shaped, occurring singly or in rosettes by attachment at the smaller cell pole. Buds are formed at the broader cell pole. Buds may have a single flagellum inserted subpolarly at the proximal pole. Adult cells are immobile. Crateriform structures and fimbriae are found in the upper cell region. Colour of colonies is pink to red. Non-sporulating. Strictly aerobic. Catalase- and cytochrome oxidase-positive. The proteinaceous cell wall lacks peptidoglycan. The major polyamines are putrescine, cadaverine, and *sym*-homospermidine. The major menaquinone is MK-6. The major fatty acids are 16:1 Δ9, 16:0, 17:1 Δ9, 17:0, 18:1 Δ9, 18:1 Δ11, and 18:0. The major phospholipids present comprise phosphatidylcholine and phosphatidylglycerol. Additional,

unidentified polar lipids are also present. This genus is a member of the phylum *Planctomycetes,* order *Planctomycetales*, family *Planctomycetaceae*, as currently defined primarily on the basis of 16S rDNA sequence analysis. The type species is *R. baltica.*

**Description of *Rhodopirellula baltica* sp. nov.**

*baltica* (bal´ti.ca. *L. fem.* adj. pertaining to the Baltic Sea, the place of isolation).

Cell size 1.0 – 2.5 × 1.2 – 2.3 µm. A single flagellum is subpolarly inserted at the proximal pole. Colonies are round, smooth and pink to red in colour. Growth is optimal between 28 and 30 °C. Growth is not observed above 32 °C. Vitamin $B_{12}$ and seawater are required for growth. The bacterium is strictly aerobic. Glucose is not fermented, nitrate can not serve as electron acceptor.

Carbon sources utilized are: cellobiose, fructose, galactose, glucose, lactose, lyxose, maltose, mannose, melibiose, melezitose, raffinose, rhamnose, ribose, sucrose, trehalose, xylose, dextrin, *N*-acetylglucosamine, glycerol, esculin, amygdalin, gluconate, glucuronate, salicin and chondroitin sulfate. Carbon sources not utilized are: fucose, sorbose, methylamine, methylsulfonate, methanol, ethanol, erythritol, adonitol, arabitol, dulcitol, inositol, mannitol, sorbitol, acetate, adipate, benzoate, caproate, citrate, formate, fumarate, glutarate, lactate, malate, 2-oxoglutarate, phthalate, propionate, pyruvate, succinate, tartrate, alanine, arginine, asparagine, aspartate, cysteine, cystine, glutamine, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, norleucine, ornithine, phenylalanine, proline, serine, threonine, tryptophane, tyrosine, valine, urea, indole, inulin, and pectin. Peptone, casamino acids, yeast extract, gelatine, ammonium, nitrate, and *N*-acetylglucosamine are utilized as a nitrogen source, nicotinate or urea are not utilized. Esculin, gelatin, and starch are hydrolyzed, alginate, casein, cellulose, chitin, or tween 80 are not hydrolized. No hemolytic activity was found with blood of horse, calf or sheep. Activity of catalase and cytochrome oxidase, but not of urease is observed. $H_2S$, but not acetoin or indole is produced. Sensitive to tetracycline, resistant to streptomycin, ampicillin, and penicillin. DNA base composition: 53 – 57 mol % G+C (55 mol % for the type strain). The chemical composition is identical to that of the genus description. Main habitat: Brackish water of Kiel Fjord (Baltic Sea). Type strain: SH $1^T$ = IFAM $1310^T$ = DSM $10527^T$ = NCIMB $13988^T$.

**Description of *Blastopirellula* gen. nov.**

*Blastopirellula* (Blas.to.pi.rel'lu.la Gr. masc. n. blastos, bud, shoot; N.L. fem. n. Pirellula, name of a bacterial genus; N.L. fem. n. Blastopirellula, a budding Pirellula).

Cells are ovoid, ellipsoidal or pearshaped, occurring singly or in rosettes by attachment of the smaller cell pole. Buds are formed at the broader, proximal cell pole. Adult cells are immobile. Crateriform structures and fimbriae are found in the upper cell region. Colonies are greyish to brownish white. Non-sporulating. Strictly aerobic. Catalase and cytochrome oxidase positive. The proteinaceous cell wall lacks peptidoglycan. The major polyamine is *sym*-homospermidine. The major menaquinone is MK-6. The major phospholipid present is phosphatidylglycerol. Additional unidentified polar lipids are also present, at least one of which appears to be identical (in Rf value) to one of the major components present in members of the genus *Pirellula*. The major fatty acids present are 15:0, i-16:0, 16:1 Δ9, 16:0, 17:1 Δ9, 17:0, 18:1 Δ9, 18:1 Δ11, 18:0, 19:1 Δ11, and 20:1 Δ9. This genus is a member of the phylum *Planctomycetes*, order *Planctomycetales*, family *Planctomycetaceae*, as currently defined primarily on the basis of 16S rDNA sequence analysis. The type species is *Blastopirellula marina*.

**Description of *Blastopirellula marina* comb. nov.**

ma.ri'na (L. fem. adj. marina, of or belonging to the sea, marine).

The description of *Blastopirellula marina* comb. nov. is the same as that published for *Pirellula marina* (Schlesner 1986) with the following additions: The species is apparently strictly aerobic as glucose is not metabolized anaerobically by either fermentation or with nitrate as electron acceptor. The chemical composition is identical to that of the genus description. Type strain: SH $106^T$ = IFAM $1313^T$ = DSM $3645^T$ = ATCC $49069^T$.

## Acknowledgements

We are indebted to Michaela Rathmann for expert analysis of DNA/DNA-hybridizations. We would like to thank Helga Marxsen (Kiel) and Daniela Lange (Bremen) for technical assistance. BJT would like to thank Carmela Belloch, CECT, University of Valencia, Spain for stimulating discussions on the interpretation of 16S rDNA data. This study was supported by the Deutsche Forschungsgemeinschaft (Schl 231/1-1 and 1-2) and the Max Planck Society.

## References

**Anderson, J. O., and S. G. E. Anderson.** 1999. Genome reduction is an ongoing process in Rickettsia. Mol. Biol. Evol. **16:**1178-1191.

**Ban, N., P. Nissen, P. B. Hansen, P. Moore, and T. A. Steitz.** 2000. The complete atomic structure of the large of the ribosomal subunit at 2.4 Å resolution. Science **289:**905-920.

**Blanchard, J. L., and M. Lynch.** 2000. Organellar genes why do they end up in the nucleus? Trends in Genetics. **16:**315-320.

**Brimacombe, R.** 2000. The bacterial ribosome at atomic resolution. Structure **8:**195-200.

**DeLong, E. F., D. G. Franks, and A. I. Alldredge.** 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. **38:**924-934.

**Devereux, R., S.-H. He, C. L. Doyle, S. Orkland, D. A. Stahl, J. LeGall, and W. B. Whitman.** 1990. Diversity and origin of *Desulfovibrio* species: phylogenetic definition of a family. J. Bacteriol. **172:**3609-3619.

**Doolittle, W. F.** 1997. Fun with genealogy. Proc. Natl. Acad. Sci. USA **94:**12751–12753.

**Doolittle, R. F., D. F. Feng, S. Tsang, G. Cho, and E. Little.** 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470-477.

**Feng, D. F., G. Cho, and R. F. Doolittle.** 1997. Determining divergence times with a protein clock: Update and reevaluation. Proc. Natl. Acad. Sci. USA **94:**13028–13033.

**Fuerst, J. A.** 1995. The planctomycetes: emerging models for microbial ecology, evolution and cell biology. Microbiology **141:**1493-1506.

**Fuerst, J. A., S. K. Sambhi, J. L. Paynter, J. A. Hawkins, and H. G. Atherton.** 1991. Isolation of a bacterium resembling *Pirellula* species from primary tissue culture of the giant tiger prawn (*Penaeus monodon*). Appl. Environ. Microbiol. **57:**3127-3134.

**Fuerst, J. A., H. G. Gwillian, M. Lindsay, A. Lichanska, C. Belcher, J. E. Vickers, and P. Hugenholtz.** 1997. Isolation and molecular identification of planctomycete bacteria from postlarvae of the giant tiger prawn *Penaeus monodon*. Appl. Environ. Microbiol. **63:**254-262.

**Fry, N. K., S. Warwick, N. A. Saunders, and T. M. Embley.** 1991. The use of 16S rRNA analyses to investigate the phylogeny of the familiy *Legionellaceae*. J. Gen. Microbiol. **137:**1215-1222.

**Gade, D., H. Schlesner, F. O. Glöckner, R. Amann, S. Pfeiffer, and M. Thomm.** 2004. Identification of planctomycetes with order-, genus- and strain-specific 16S rRNA-targeted probes. Microb. Ecol. In press.

**Garrity, G. M., and J. G. Holt.** 2001. The road map to the manual. In: Boone, D. R. and R. W. Castenholz (eds). Bergey's Manual of Systematic Bacteriology 2nd edition., Springer-Verlag New York pp. 119-166.

**Gil, R., B. Sabater-Muñoz, A. Latorre, F. J. Silva, and A. Moya.** 2002. Extreme genome reduction in *Buchnera* spp.: Toward the minimal genome needed for symbiotic life. Proc. Natl. Acad. Sci. USA **99:**4454–4458.

**Glöckner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann, and R. Reinhardt.** 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc. Natl. Acad. Sci. USA **100:**8298-8303.

**Griepenburg, U., N. Ward-Rainey, S. Mohamed, H. Schlesner, H. Marxsen, F. Rainey, E. Stackebrandt, and G. Auling.** 1999. Phylogenetic diversity, polyamine pattern and DNA base composition of members of the order *Planctomycetales*. Int. J. Syst. Bacteriol. **49:**689-696.

**Hohenberg, H., K. Mannweiler, and M. Müller.** 1994. High-pressure freezing of cell suspensions in cellulose capillary tubes. J. Microscopy **175:**34-43.

**Kerger, B. D., C. A. Mancuso, P. D. Nichols, D. C. White, T. Langworthy, M. Sittig, H. Schlesner, and P. Hirsch.** 1988. The budding bacteria, *Pirellula* and *Planctomyces*, with atypical 16S rRNA and absence of peptidoglycan, show eubacterial phospholipids and uniquely high portions of long chain beta-hydroxy fatty acids in the lipopolysaccharide lipid A. Arch. Microbiol. **149:**255-260.

**Kölbel-Boelke, J., R. Gebers, and P. Hirsch.** 1985. Genome size determinations for 33 strains of budding bacteria. Int J Syst Bacteriol **35:**270-273.

**König, H., H. Schlesner, and P. Hirsch.** 1984. Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on a *Prosthecomicrobium* sp. Arch Microbiol **138:**200-205.

**Kouker, G., and K. E. Jaeger.** 1987. Specific and sensitive plate assay for bacterial lipases. Appl. Environ. Microbiol. **53:**211-213.

**Lee, M. S. Y.** 1999. Molecular clock calibrations and metazoan divergence dates. J. Mol. Evol. **49:**385-391.

**Liesack, W., H. König, H. Schlesner, and P. Hirsch.** 1986. Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella/Planctomyces* group. Arch. Microbiol. **145:**361-366.

**Liesack, W., R. Söller, T. Stewart, H. Hass, S. Giovannoni, and E. Stackebrandt.** 1992. The influence of tachyletic (rapidly) evolving sequences on the topology of phylogenetic trees – intrafamily relationships and the phylogenetic position of *Planctomycetaceae* as revealed by comparative analysis of 16S ribosomal RNA sequences. Syst. Appl. Microbiol. **15:**357-362.

**Lindsay, M. R., R. I. Webb, M. Strous, M. S. M. Jetten, M. K. Butler, R. J. Ford, and J. A. Fuerst.** 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch. Microbiol. **175:**413-429.

**Lyman, J., and R. H. Fleming.** 1940. Composition of sea water. J. Marine Research (Sears Foundation) **3:**134-146.

**Mayr, E.** 1969. Principles of systematic zoology. McGraw-Hill, New York.

**Morse, R., M. D. Collins, K. O'Hanlon, S. Wallbanks, and P. T. Richardson.** 1996. Analysis of the β subunit of DNA-dependent RNA polymerase does not support the hypothesis inferred from 16S rRNA analysis that *Oenococcus oeni* (formerly *Leuconostoc oenos*) is a tachytelic (fast-evolving) bacterium. Int. J. Syst. Bacteriol. **46:**1004-1009.

**Ochman, H., S. Elwyn, and N. A. Moran.** 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. USA **96:**12638-12643.

**Pieringer, R. A.** 1989. Biosynthesis of non-terpenoid lipids. In Microbial Lipids, vol. 2, pp.51-114. Edited by C. Ratledge and S. G. Wilkinson. Academic Press, London.

**Poinar, H. N., M. Höss, J. L. Bada, and S. Pääbo.** 1996. Amino acid racemization and the preservation of ancient DNA. Science **272:**864-866.

**Rabus, R., D. Gade, R. Helbig, M. Bauer, F. O. Glöckner, M. Kube, H. Schlesner, R. Reinhardt, and R. Amann.** 2002. Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1 by a proteomic approach. Proteomics **2:**649-655.

**Rabus, R., and F. Widdel.** 1995. Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria. Arch. Microbiol. **163:**96-103.

**Rathmann, M.** 1992. Taxonomische Untersuchungen an *Pirellula*-ähnlichen Bakterien aus verschiedenen aquatischen Standorten. Diplomarbeit. Christian-Albrechts-Universität zu Kiel.

**Ratledge, C., and S. G. Wilkinson.** 1989. In Microbial lipids vol. 1. Academic Press. London.

**Schlesner, H.** 1986. *Pirella marina* sp. nov., a budding, peptidoglycan-less bacterium from brackish water. Syst. Appl. Microbiol. **8:**177-180.

**Schlesner, H.** 1994. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from various aquatic habitats using dilute media. Syst. Appl. Microbiol. **17:**135-145.

**Schlesner, H., and P. Hirsch.** 1984. Assignment of ATCC 27377 to *Pirella* gen. nov. as *Pirella staleyi* comb. nov. Int. J. Syst. Bacteriol. **34:**492-495.

**Schlesner, H., and P. Hirsch.** 1987. Rejection of the genus name *Pirella* for pear-shaped budding bacteria and proposal to create the genus *Pirellula* gen. nov. Int. J. Syst. Bacteriol. **37:**441.

**Schlesner, H., and E. Stackebrandt.** 1986. Assignment of the genera *Planctomyces* and *Pirella* to a new family *Planctomycetaceae* fam. nov. and description of the order *Planctomycetales* ord. nov. Syst. Appl. Microbiol. **8:**174-176.

**Schluenzen, F., A. Tocilji, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath.** 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. Cell **102:**615-623.

**Simpson, G. G.** 1944. Tempo and mode in evolution. Columbia University Press. New York.

**Sittig, M., and P. Hirsch.** 1992. Chemotaxonomic investigation of budding and/or hyphal bacteria. Syst. Appl. Microbiol. **15:**209-222.

**Sittig, M., and H. Schlesner.** 1993. Chemotaxonomic investigation of various prosthecate and/or budding bacteria. Syst. Appl. Microbiol. **16:**92 103.

**Smibert, R. M., and N. R. Krieg.** 1994. Phenotypic characterization. In Methods for General and Molecular Bacteriology, pp. 607-654. Edited by P. Gerhardt, R.G.E. Murray, W.A. Wood, and N.R. Krieg. Washington, D.C.: American Society for Microbiology.

**Sneath, P. H. A.** 1974. Phylogeny of micro-organisms. In Evolution in the microbial world, pp 1-39. Edited by M. J. Carlilie and J. J. Skehel. Soc. Gen. Microbiol. Symposium 24. Cambridge University Press. Cambridge.

**Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough.** 2002. Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. Proc. Natl. Acad. Sci. USA **99:**4430–4435.

**Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:**846-849.

**Stackebrandt, E., A. Fischer, P. Hirsch, T. Roggentin, and H. Schlesner.** 1986. The phylogeny of an ancient group of budding peptidoglycan-less eubacteria: The genera *Planctomyces* and *Pirella*. Endocyt. C. Res. **3:**29-40.

**Staley, J. T.** 1973. Budding bacteria of the *Pasteuria-Blastobacter* group. Can. J. Microbiol. **19:**609-614.

**Staley, J. T., J. A. Fuerst, S. Giovannoni, and H. Schlesner.** 1992 The order *Planctomycetales* and the genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera*. In: Balows, A., H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (eds.). The Prokaryotes 2nd edition, Springer-Verlag New York, pp 3710-3731.

**Strömpl, C., B. J. Tindall, G. N. Jarvis, N. Lünsdorf, E. R. B. Moore, and H. Hippe.** 1999. A re-evaluation of the taxonomy of the genus *Anaerovibrio*, with the reclassification of *A. glycerini* as *Anaerosinus glycerini* gen. nov., comb. nov., and *A. burkinabensis* as *Anaeroarcus burkinabensis* gen. nov., comb. nov. Int. J. Syst. Bacteriol. **49:**1861-1872.

**Tindall, B. J.** 1990a. A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. Syst. Appl. Microbiol. **13:**128-130.

**Tindall, B. J.** 1990b. Lipid composition of *Halobacterium lacusprofundi*. FEMS Microbiol. Letts. **66:**199-202.

**Wang, J., C. Jenkins, R. I. Webb, and J. A. Fuerst.** 2002. Isolation of *Gemmata*-like and *Isosphaera*-like planctomycete bacteria from soil and freshwater. Appl. Environ. Microbiol. **68:**417-422.

**Ward, N., F. A. Rainey, E. Stackebrandt, and H. Schlesner.** 1995. Unravelling the extent of diversity within the order *Planctomycetales*. Appl. Environ. Microbiol. **61:**2270-2275.

**Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper.** 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. **37:**463-464.

**Widdel, F., and F. Bak.** 1992. Gram-negative mesophilic sulphate-reducing bacteria. In Balows, A., M. Dworkin, W. Harder, K. H. Schleifer, and H. G. Trüper (eds). The Prokaryotes 2nd edition, Springer-Verlag New York, pp. 3352-3378.

**Willerslev, E., A. J. Hansen, J. Binladen, T. B. Brand, T. P. Gilbert, B. Shapiro, M. Bunce, C. Wiuf, D. A. Gillichinsky, and A. Cooper.** 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science **300:**791-795.

**Wimberly, B. T., D. E. Brodersen, W. M. Jr. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan.** 2000. Structure of the 30S ribosomal subunit. Nature **407:**327-339.

4

# Towards the proteome of the marine bacterium *Rhodopirellula baltica*: mapping the soluble proteins

Dörte Gade[1], Dorothea Theiss[2], Daniela Lange[1], Ekaterina Mirgorodskaya[2], Thierry Lombardot[1], Frank Oliver Glöckner[1], Michael Kube[2], Richard Reinhardt[2], Rudolf Amann[1], Hans Lehrach[2], Ralf Rabus[1*], Johan Gobom[2]

Manuscript in preparation

[1] Max-Planck-Institut für Marine Mikrobiologie, Bremen, Germany
[2] Max-Planck-Institut für Molekulare Genetik, Ressourcen Zentrum, Berlin, Germany
[*] Corresponding author

## Abstract

The marine bacterium *Rhodopirellula baltica*, a member of the phylum *Planctomycetes*, has distinct morphological properties and contributes to remineralization of biomass in the natural environment. On the basis of its recently determined complete genome we investigated its proteome by two-dimensional gel electrophoresis (2-DE) and established a reference 2-DE gel for the soluble protein fraction. Approximately 1000 protein spots were excised from a colloidal Coomassie-stained gel (pH 4 – 7), analyzed by matrix-assisted laser desorption/ionization mass spectrometry and identified by peptide mass fingerprinting. The non-redundant dataset contained 626 distinct protein spots, corresponding to 558 different genes. The identified proteins were classified into role categories according to their predicted functions. The experimentally determined and the theoretically predicted proteomes were compared. Proteins, which were most abundant in 2-DE gels and also predicted to be highly expressed, could be linked mainly to housekeeping functions in glycolysis, tricarboxic acid cycle, amino acid biosynthesis, protein quality control and translation. Absence of predictable signal peptides indicated a localization of these proteins in the intracellular compartment, the pirellulosome. Among the identified proteins, 146 contained a predicted signal peptide suggesting their translocation. Some proteins were detected in more than one spot on the gel, indicating posttranslational modification. In addition to identifying proteins present in the published sequence database for *R. baltica*, an alternative approach was used, in which the mass spectrometric data was searched against a maximal ORF set, allowing the identification of four previously unpredicted ORFs. The 2-DE reference map presented here will serve as framework for further experiments to study differential gene expression of *R. baltica* in response to external stimuli or cellular development and compartmentalization.

## Introduction

Since the pioneering determination of the *Haemophilus influenzae* (Fleischmann et al. 1995) and *Mycoplasma pneumoniae* (Himmelreich et al. 1996) genomes, more than 170 complete genomes from bacteria have been reported (for detailed information see *e.g.* www.genomesonline.org and www.tigr.org). Even though a given genome represents the blueprint of life, there is a need for functional analysis on the transcriptional and proteomic level in order to define (1) which of the predicted genes can be expressed in principle, and (2) the physiological conditions inducing their expression. In contrast to the numerous publicly available genome sequences, only few proteomes (protein maps) have been reported to date. Moreover, only a limited number of proteins is usually identified and annotated. Among the comprehensive protein maps are the very well studied standard bacteria *Escherichia coli* (Tonella et al. 1998, 2001; Molloy et al. 2000) and *Bacillus subtilis* (Ohlmeier et al. 2000; Büttner et al. 2001), some pathogens like *e.g. Mycoplasma pneumoniae* (Regula et al. 2000, 2001; Ueberle et al. 2002), *Staphylococcus aureus* (Cordwell et al. 2002, Hecker et al. 2003), *Haemophilus influenzae* (Fountoulakis et al. 1998a, 1998b; Langen et al. 2000) and *Pseudomonas aeruginosa* (Nouwens et al. 2002), and some biotechnologically relevant bacteria such as *Corynebacterium glutamicum* (Hermann et al. 2001) and *Streptomyces coelicolor* (Hesketh et al. 2002).

Protein maps are most often constructed by applying two-dimensional gel electrophoresis (2-DE) in combination with mass-spectrometric (MS) analysis. 2-DE is a well-established technique for high resolution separation of proteins from complex mixtures (Görg et al. 2001). Electrophoretically separated proteins are excised from stained 2-DE gels and cleaved enzymatically (*e.g.* by trypsin) to defined fragments. The masses of the generated peptides, determined by MS, constitute a peptide mass fingerprint (PMF) of the protein. For protein identification, the PMF is compared to sets of masses calculated for each protein sequence in a database, based on the known cleavage specificity of the protease used (Pappin et al. 1993; Henzel et al. 1993; Mann et al. 1993). Matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS; Karas et al. 1988) has become the most widely used technique for protein identification. Automation allows a high throughput at the level of spot excision, sample processing and MS analysis.

Over the last 1 to 2 decades, the impact of microbial activity on environmental processes has been increasingly recognized. This led recently to the initiation of genome projects on environmentally relevant bacteria. Genomes of such bacteria in conjunction with

functional analysis will provide new insights into the molecular basis of microbial activity (and its control) in the natural environment. The first examples are the complete genome sequences of *Synechocystis* sp. (Kaneko et al. 1996; www.kazusa.or.jp), *Caulobacter crescentus* (Nierman et al. 2001;www.tigr.org) and *Rhodopirellula baltica* (Glöckner et al. 2003; www.regx.de). With 7.145 Mb and 7,325 open reading frames, the genome of *R. baltica* represents one of the largest bacterial genomes sequenced so far. In the case of *C. crescentus*, a protein map with 295 identified proteins has only very recently been reported (Vohradsky et al. 2003), whereas 57 membrane proteins were identified from *Synechocystis* sp. strain PCC6803 (Huang et al. 2002). *R. baltica* is a marine, aerobic bacterium that has been isolated from the Baltic Sea (Schlesner 1994). It belongs to the phylogenetic distinct group of *Planctomycetes* (Schlesner et al. 2004), members of which are known to be globally distributed and suggested to be involved in carbon remineralization. Interest in this group of bacteria also comes from their unusual morphological properties. The cells reproduce via budding and display a complex life cycle. *R. baltica* cells can occur in two morphotypes, *i.e.* as single motile cells or attached to each other in aggregates.

Peptidoglycans appear to be absent from the proteinaceous cell wall. Cells are organized in membrane-defined compartments including a membrane-engulfed nucleoid, termed pirellulosome (Lindsay et al. 1997, 2001).

To study molecular physiology, cellular development and compartmentalization of this bacterium, we established a 2-DE map of soluble proteins in the pH range of 4 to 7. The master gel contains 626 gene products, which were identified by PMF.

## Materials and Methods

### Growth of cells and preparation of soluble proteins

Cells of *R. baltica* (DSM 10527) were grown in mineral medium with ribose (10 mM), glucose (10 mM) or *N*-acetylglucosamine (10 mM) as sole source of organic carbon (Rabus et al. 2002).

Harvesting of cells was essentially performed as previously described (Rabus et al. 2002). Cells were harvested in the exponential growth phase by centrifugation (10,000 × g, 15 min, 4 °C). The pellets were washed with 100 mM Tris/HCl pH 7.5 containing 5 mM $MgCl_2$. Cell pellets were directly frozen in liquid nitrogen and stored at -80 °C until cell breakage and 2-DE. Prior to cell breakage, pellets were resuspended in 1 mL lysis buffer (7 M urea, 2 M thiourea, 2% DTT, 2% CHAPS, 0.5% carrier ampholytes; Amersham Biosciences, Freiburg, Germany). Cell breakage was performed with the PlusOne® grinding kit (Amersham Biosciences) following the manufacturers instructions. Removal of cell debris, DNA and membranes by centrifugation (100,000 × g, 1 h, 15 °C) yielded the fraction of soluble proteins. The protein content of this fraction was determined using the method described by Bradford (Bradford 1976).

### Two-dimensional gel electrophoresis, staining and image acquisition

Two-dimensional gel electrophoresis (2-DE) was essentially performed as described before (Görg et al. 2000; Rabus et al. 2002; Gade et al. 2003). In brief, isoelectric focusing (IEF) was performed using the IPGphor™ system and 24 cm long IPG strips (linear pH gradient from 4 to 7; Amersham Biosciences), followed by equilibration of the gels with DTT and iodoacetamide. The second dimension separation was then performed using the Ettan™ Dalt system (Amersham Biosciences) and gels made of 375 mM Tris/HCl, 0.1% SDS and 12.5% Duracryl (Genomic Solutions; Ann Arbor, Michigan, USA). The protein load for preparative gels was 400 µg. Proteins were visualized using colloidal Coomassie (method modified from Dorothy et al. 1998). For image acquisition the gels were digitalized with the Image Scanner (Amersham Biosciences).

### Gel sample excision and processing

Excision and processing of the gel samples for PMF was performed as described previously (Nordhoff et al. 2001), with some modifications. Protein spots were sampled from

the gel using an automatic excision workstation (Proteineer, Bruker Daltonics, Bremen, Germany). The excision head was equipped with a single needle with a diameter of 2 mm. The excised gel spots were delivered into 96-well poly-propylene microtiter plates (Costar Thermowell®, Cornis, NY, USA), pre-treated by punching two holes (d < 0.5 mm) in the bottom of each well. This preparation allows removal of the washing solutions and reagents used throughout the digestion procedure by simple flow-through centrifugation, while retaining the gel particles in the wells. To protect the pierced 96-well MTPs from environmental contamination, they were placed in a second 96-well MTP and covered by a lid. The second MTP also serves as collector for liquid removed by centrifugation. To ensure that no liquid from the collection MTP reaches or contaminates the pierced MTP, a spacer was placed between these two MTPs. Following excision, all liquid was removed from the gel pieces by centrifugation and the sample plates were stored at -80°C prior to further processing.

Prior to digestion the gel particles were washed by incubation for 2 × 30 min in 100 µL 50% ethanol (v/v). Following removal of the washing solution by centrifugation, residual water was expelled from the gel particles by incubation for 5 min in 100% ethanol. The sample plates were then placed without lid in a laminar flow-bench for 15 min to allow the ethanol to evaporate. An aliquot of freshly prepared, cooled trypsin (Roche, recombinant porcine) solution (5 µL, 10 ng/µL, 50 mM $NH_4HCO_3$, pH 7.8) was added to each sample. The sample plates were immediately placed in a refrigerator and incubated at 4 °C for 30 min. Thereafter, an aliquot of digestion buffer (50 mM $NH_4HCO_3$, pH 7.8) was added to each sample, and the MTPs were placed in a humidified box and incubated for 4 h at 37°C.

**Matrix-assisted laser desorption/ionization-mass spectrometry**

Protein digests were prepared for MALDI using the α-cyano-4-hydroxycinnamic acid affinity sample preparation technique described previously (Gobom et al. 2001). Mass analysis of positively charged ions was performed on an Ultraflex LIFT and a Reflex III instrument (Bruker Daltonics, Bremen, Germany) operated in the reflector mode and using delayed ion extraction. Positively charged ions in the mass range 700 – 3,500 Da were analyzed.

**Data processing and protein identification**

The success rate and confidence of protein identification by peptide mass fingerprinting depends to a high degree on the accuracy of the mass measurement. High mass accuracy by MALDI-TOF-MS was achieved by using internal reference compounds for spectra calibration. To calibrate the large number of spectra acquired in this study, the following procedure was developed. Firstly, the acquired MALDI-TOF spectra were calibrated externally using a polynomial function according to a previously described procedure (Gobom et al. 2002). This calibration ensures a maximum error of 500 ppm over the entire MALDI sample support. For a subsequent internal mass correction, each spectrum was searched for signals corresponding to known reference compounds. Three peptides (Angiotensin I, $MH^+$ 1296.68; Neurotensin 1-13, $MH^+$ 1,672.9150; ACTH 18-39, $MH^+$ 2,465.1989; monoisotopic mass values), which were mixed into the MALDI matrix solution, and two abundant signals corresponding to trypsin autoproteolysis ($MH^+$ 842.510 and 2,211.1045, respectively) were used as internal references. For the spectra in which at least three of these compounds were detected, a linear regression of the relative errors for the reference signals versus their calculated m/z values was calculated. If the standard deviation of the regression line was below 10 ppm, the regression function was used for correction of the externally calibrated mass values. If a detected calibrant had a relative error >2 standard deviations it was discarded and the linear regression calculated again.

In some cases a sufficient number of reference masses was not detected, and in other cases, an analyte signal with a molecular mass close to the reference compound was erroneously selected as a calibrant. An example of the latter is shown in Figure 1 A. Here, an analyte signal that partially overlaps with the trypsin autoproteolysis signal of m/z 842.510 was erroneously selected as a calibrant. In the given example, the resulting standard deviation of the linear regression was 17.7 ppm, and the calibration thus discarded. Out of 384 spectra acquired on one MALDI sample support, 190 fulfilled the criteria for internal mass correction. The remaining 194 spectra were calibrated with background signals of unknown identity, as follows: using the internally calibrated spectra, a histogram was constructed of the abundance of signals with mass differences within $\Delta$m/z 0.05 (Figure 1 B). Mass values within this interval, detected in >25 spectra in the data set (indicated by a dotted line) were averaged and added to the list of reference masses. Using the new list of internal reference masses, the internal correction procedure was repeated with the remaining 194 spectra, this time with the requirement that at least six signals in each spectrum should match values in the calibrant list. Following this second round of internal correction, all the

remaining spectra were successfully calibrated. For example, in the spectrum for which the internal correction failed, shown in Figure 1 A, 17 signals matched mass values in the calibration list. The new error plot for the spectrum is shown in Figure 1 C.

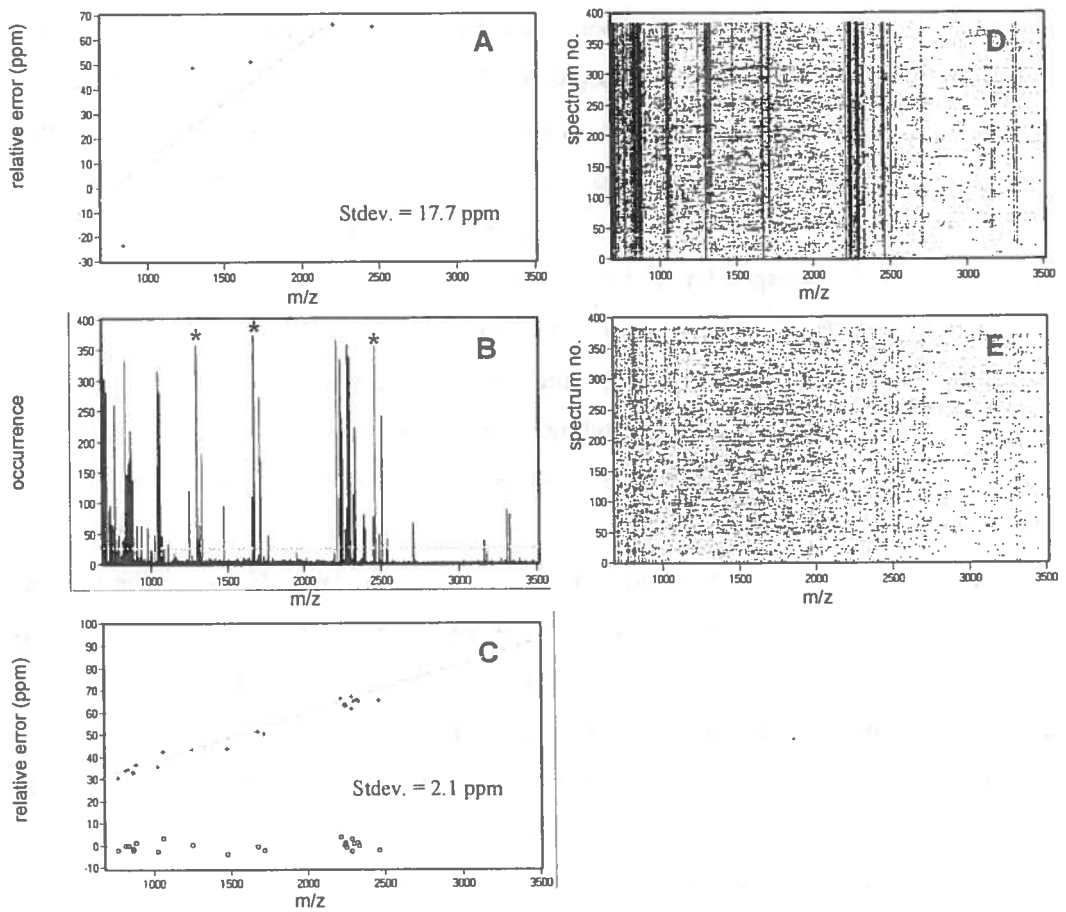



**Figure 1.** Calibration and peak filtering of MALDI-TOF-MS data. (A) Plot of the relative deviations vs. m/z for experimentally determined masses matched to a list of internal calibrant masses. In the shown example, a peptide signal that partially overlaps with an internal calibrant mass at m/z 842.51 was erroneously assigned as a calibrant. This is reflected by an unusually large standard deviation (>10 ppm) of the linear regression calculated for the assigned calibrants, and the calibration was thus discarded. (B) Histogram of signals with $\Delta m$ <0.05 in successfully calibrated spectra acquired on one MALDI sample support. Signals detected in >25 spectra (indicated by a dotted line) were used as internal calibrants in a second round of calibration. (C) Plot of the relative deviations vs. m/z (same sample as in *a*) following calibration using the background signals detected in *b* as calibrants. (D) Plot of the elution time vs. m/z of all assigned monoisotopic signals in the calibrated data set (signal plot). Background signals from trypsin autoproteolysis and sample contaminants appear as vertical lines in the plot. (E) Signal plot following removal of background signals as described in the text.

The presence of background signals in the spectra decreases the specificity of the database search. A peak plot of the 384 spectra from the previous example is shown in Figure 1 D. Calibrants and background signals appear as vertical lines in the plot. Background signals were assigned as described in the previous section, and removed from the data set. In addition, sodium- and potassium-cationized molecular ions, appearing as satellite signals to the protonated peptide molecular ion signal with $\Delta m/z$ 21.982 and 38.090, respectively, were removed. The peak plot following background filtering is shown in Figure 1 E.

Database searching was performed using the software MASCOT (Matrixscience, London, UK; Perkins et al. 1999). The published ORF set of *R. baltica* (BX119912) was searched using the following settings: mass error tolerance: 50 ppm; fixed modifications: Cys-carbamidomethylation; variable modifications: oxidation; one tolerated missed cleavage. Under these conditions, a probability based MOWSE score >51 was considered significant ($p < 0.05$).

**Generation of theoretical 2-DE gels**

The published ORF set of *R. baltica* (Acc. BX119912) was used to create the theoretical 2-DE gels. Molecular weight ($M_r$) and isoelectric point (p*I*) were calculated for each predicted protein using the program *pepstats* from emboss (www.hgmp.mrc.ac.uk/Software/EMBOSS; Rice et al. 2000). The annotation of the published ORF set was scanned for the keywords "conserved hypothetical" and "hypothetical" in the product key of the description, generating the *conserved hypothetical* and *hypothetical* groups. The remaining proteins were sorted into the group *assigned function.*

**Construction of a Maximal ORF Set**

In order to identify proteins encoded by genes that are not present in the published ORF set of *R. baltica* (BX119912), the following strategy was employed. Based on the genomic sequence of *R. baltica*, a new ORF set was constructed by means of a PERL script according to the following steps. Firstly, the positions of all stop codons in the genome were determined. For each stop codon, all theoretically possible reading lengths with a minimal ORF length of 102 bases were calculated by extending their sequences from the stop codon to all possible start codons detectable until the next stop codon. The resulting ORF list, denoted *Maximum ORF Set* (MOS), comprised 578,949 sequences and represents the maximal coding capacity of the genome. The MOS was translated into amino acid sequences

and used as database for protein identification by PMF using data from all three analyzed 2-DE gels, as described in the *Results* section.

**Signal peptides and gene expression levels predictions**

Signal peptides were predicted by analyzing each theoretical protein encoded by the *R. baltica* genome with the program SignalP 2.0 (Nielsen et al. 1997). From this data set proteins were extracted which corresponded to identified 2-DE separated proteins by means of a custom PERL script (using the GenDB system; Meyer et al. 2003). Proteins with SignalP scores >0.75 were considered as potentially translocated.

Expression level prediction based on codon usage optimization was calculated for each gene in the *R. baltica* genome according to the method described by Karlin et al. (2001). Highly expressed reference genes including ribosomal proteins, translation factors and chaperonins were extracted from the published annotation of *R. baltica* (Glöckner et al. 2003).

## Results and discussion

### Comparison of theoretical and experimental proteome

Three different theoretical proteome maps of *R. baltica* were created: one for proteins with *assigned function* (Figure 2 A), one for *conserved hypothetical* proteins (Figure 2 B) and one for *hypothetical* proteins (Figure 2 C). Proteins with *assigned function* are homologous to proteins with known functions. *Conserved hypothetical* proteins cannot be assigned to any function, however they have homologs in genomes of other organisms. *Hypothetical proteins* are also of unknown function, but they are to date not known from any organism other than *R. baltica*.

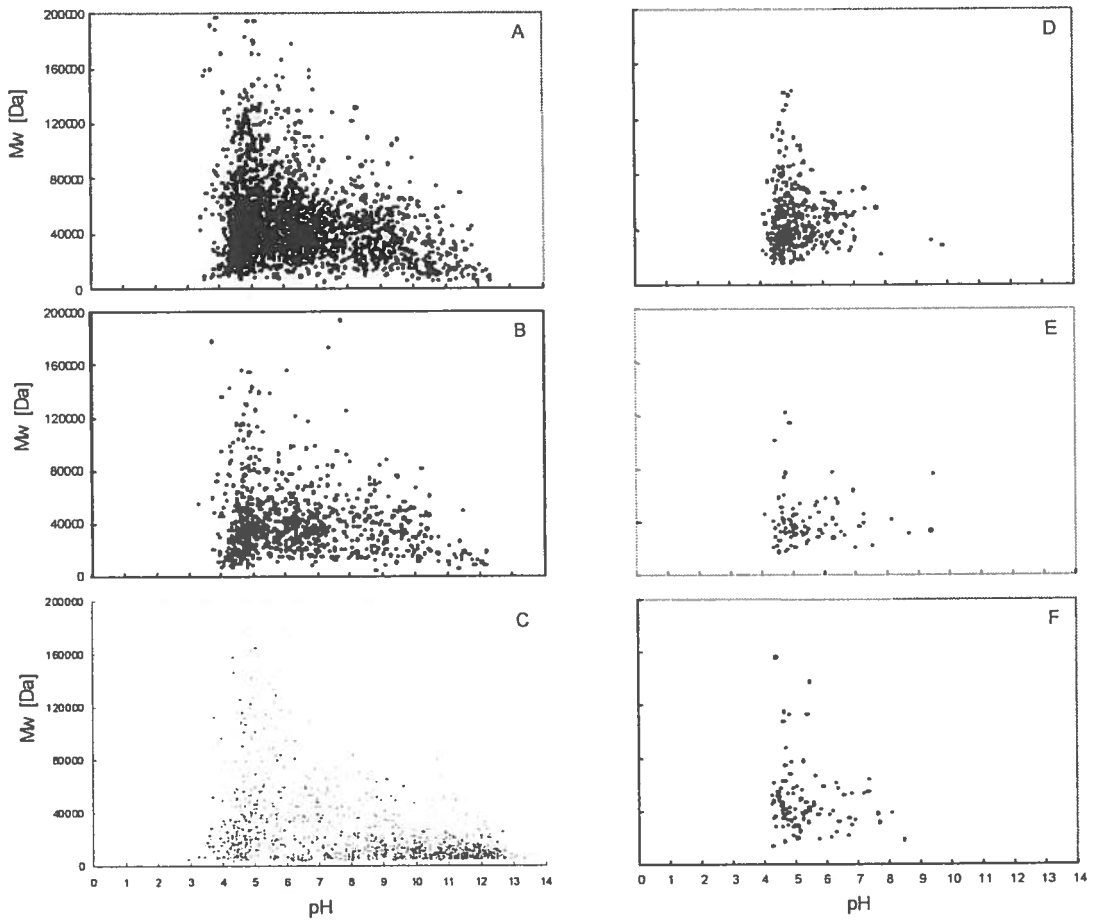



**Figure 2.** Theoretical 2-DE gels of proteins predicted from the genome of *R. baltica* (A – C) and experimentally identified in this study (D – F). Isoelectric points and molecular weights were calculated using the "pepstats" program module of emboss. Proteins with functional assignment (A and D), conserved hypothetical proteins (B and E), hypothetical proteins (C and F).

An overlay of these three maps represents the complete theoretical proteome map predicted from the annotated genome sequence. Remarkably, it shows a different isoelectric distribution pattern than those of previously reported bacterial and archaeal proteomes (VanBogelen et al. 1999; Schwartz et al. 2001). Typically, prokaryotic theoretical proteome maps display a bimodal distribution with two protein-rich areas in the acidic and alkaline ranges, separated by a pronounced protein-depleted area around pH 7. In contrast, *R. baltica* displays a trimodal distribution with a third area of protein abundance in the neutral range. A protein peak around pH 7 has previously only been described for eukaryotic proteomes (Schwartz et al. 2001). It is assumed that the bi- or trimodality of protein p*I*s reflects the subcellular localization of the proteins. While cytoplasmic proteins typically have p*I* values of around 5, integral membrane proteins tend to have p*I* values of around 9. Proteins belonging to these two groups can be found in all genomes in large numbers. The nuclear proteins apparently form the third cluster in eukaryotic proteomes (Schwartz et al. 2001). While a large number of proteins with neutral isoelectric points are predicted, the analyzed 2-DE gel (Figure 2 D-F, Figure 3) reveals only a limited number of proteins close to pH 7. Notably, the theoretical proteome of *R. baltica* contains a large number of predicted proteins with p*I*s higher than 10, while the alkaline proteins of well-studied bacteria such as *E. coli* center around p*I* 9.

Functions could be assigned to only 32% of the predicted proteins of *R. baltica*. Out of these, the majority is predicted to fall into the acidic region of the theoretical 2-DE gel. Thus, the applied IEF conditions are apparently well suited to study this group of proteins. For the conserved hypothetical proteins (amounting to 14% of the predicted proteins) a similar situation was observed.

More than half (54%) of all predicted proteins belong to the hypothetical proteins, which are unique to *R. baltica*. However, these proteins are apparently underrepresented in the set of proteins identified in this study. The theoretical 2-DE gel displayed in Figure 1 C reveals that the majority of hypothetical proteins have theoretical p*I*s above 7, in fact, 37% of them have p*I*s of greater than 10. Since a pH gradient from 4 to 7 was used in this study these alkaline proteins could not be detected. Remarkably, many of the predicted alkaline proteins have rather low molecular weights (below 10 kDa), probably hampering their isolation by conventional 2-DE. A contribution of orf overprediction to the high number of hypothetical proteins cannot be excluded at present. Nevertheless, it is tempting to speculate that *R. baltica* recruits hypothetical proteins for specific functions, *e.g.* in cellular development or translocation of proteins and solutes across the complex membrane structure.

**Master gel**

The soluble protein fraction of *R. baltica* grown under standard conditions was visualized using 2-DE with immobilized pH gradients from 4 to 7. This fraction should represent the major part of the cytosolic proteins. Under these conditions, approximately 2000 proteins of *R. baltica* can be separated and detected, when highly sensitive protein stains such as silver or fluorescent dyes are applied (see accompanying publication). Figures 3 and 4 show the colloidal Coomassie-stained master gel from *R. baltica* cells grown with ribose. The image shown in Figure 3 serves as a roadmap to the master gel sections displayed in Figure 4 A-D. From the approximately 1000 excised gel samples, 626 proteins, represented by different spots on the gel, were identified by means of PMF ($p < 0.05$). Since some 30 proteins occurred as at least two spots, the actual number of distinct identified ORFs amounted to 558. The identified proteins were annotated in the master gel sections with the published gene numbers (Figure 4 A-D) and grouped according to functional categories (Table 1). Predicted functions of each identified protein are given in Table 2. To verify the identifications of the master gel, 2-DE and MS analysis of cells grown with glucose and *N*-acetylglucosamine, respectively, were analyzed in parallel (Table 2). Among the 558 identified gene products 301 (54%) were identified from at least two independent 2-DE gels.

Newly developed software was used for processing of the calibrated mass spectrometric data. This included filtering of sodium and potassium adduct signals, filtering of non-peptide derived masses, filtering of signals derived from known contaminants such as trypsin autoproteolysis products, and statistical filtering of frequently occurring m/z values representing unknown gel sample contaminants. This processing improves the quality of the input data for the database search, thereby increasing the number of identified proteins and their respective scores. For example, in a subset of 384 samples prepared on one MALDI target (Figure 1), 205 spectra (53%) resulted in a significant identification score when filtering of the data was not applied. With filtering, the number of significant identification results increased to 262 (68%). Concomitantly, the average MOWSE score of the identified proteins increased from 99 to 119, thereby improving the certainty of the identification results.

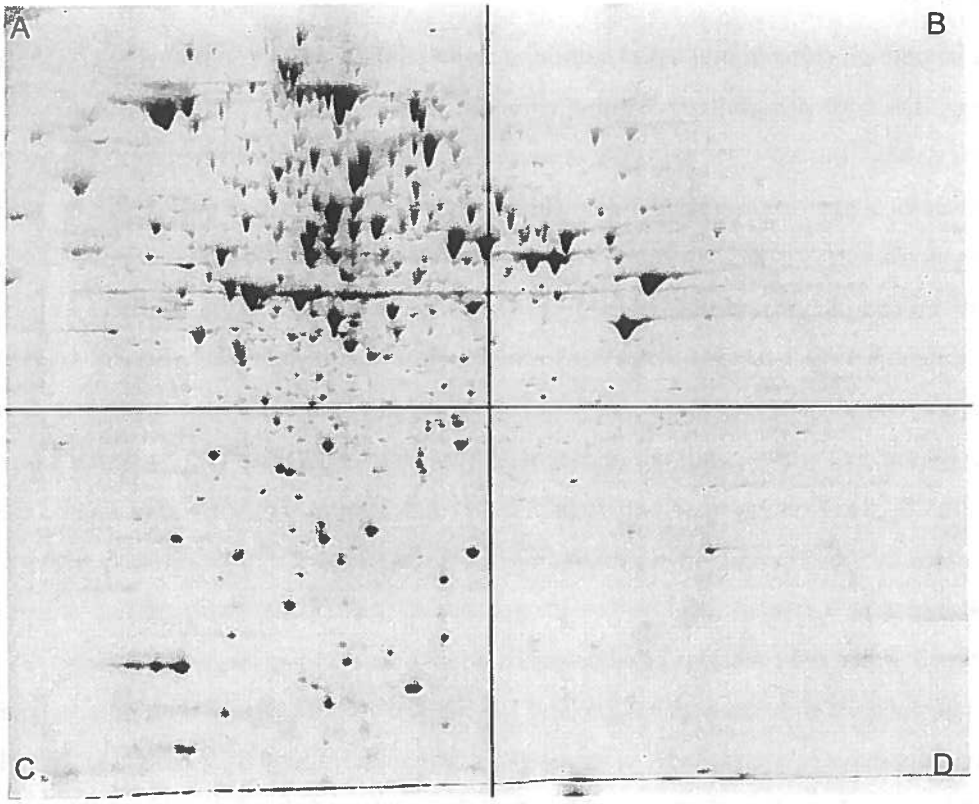
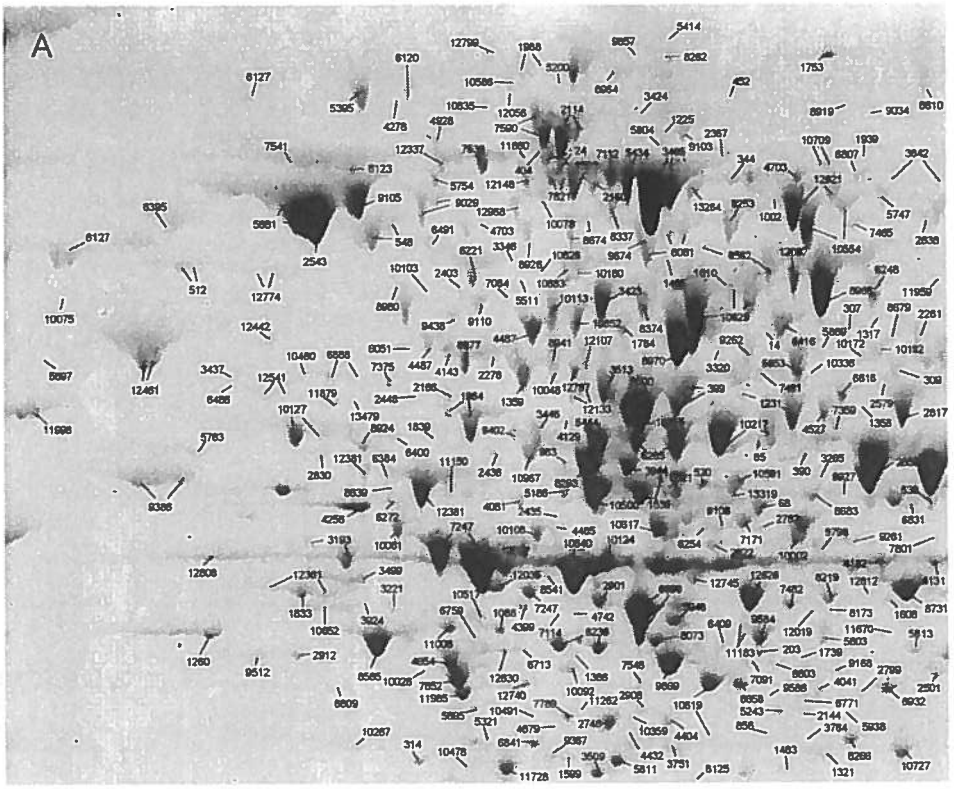
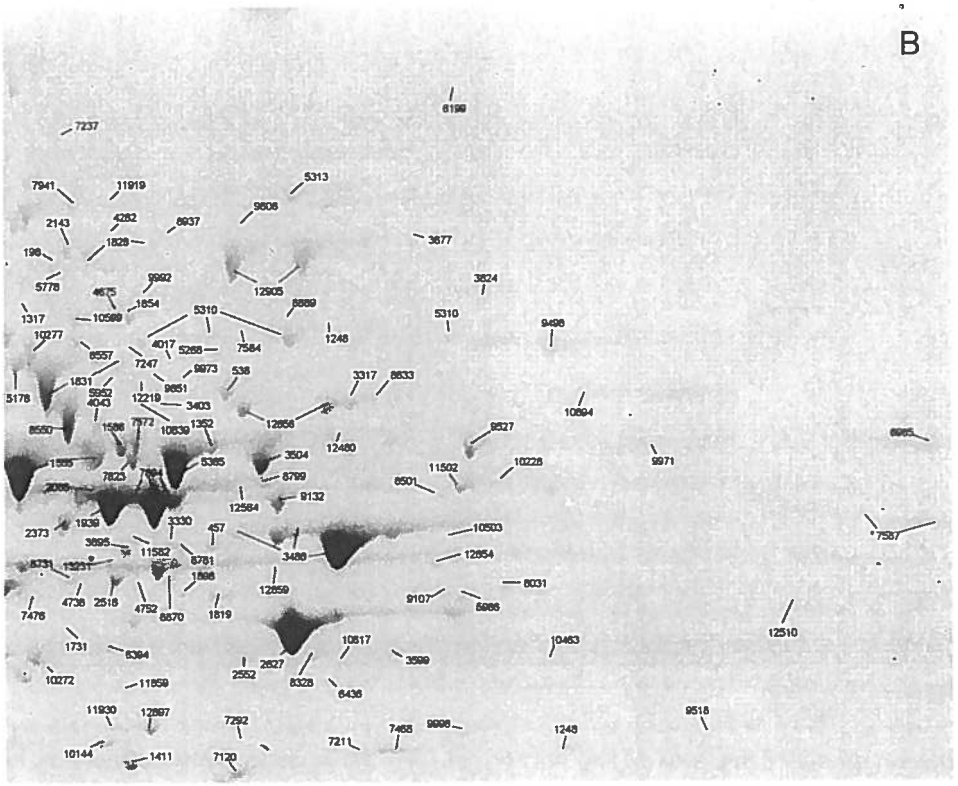



**Figure 3.** Master gel of soluble proteins of *R. baltica* grown with ribose. The pH gradient is from 4 to 7. Assignment of identified ORF to the corresponding protein spots is shown in Figure 4. For improved readability, the gel is divided into four sections (A – D).

**Table 1.** Distribution of identified proteins among functional groups.

| Category | number of identified proteins | % of identified proteins |
|---|---|---|
| metabolism | 250 | 45 |
| genetic information processing | 52 | 9 |
| regulation and signal transduction | 24 | 4 |
| stress response | 13 | 2 |
| energy | 8 | 1 |
| transport | 11 | 2 |
| conserved hypothetical proteins | 95 | 17 |
| hypothetical proteins | 97 | 17 |
| others | 8 | 1 |


A


B

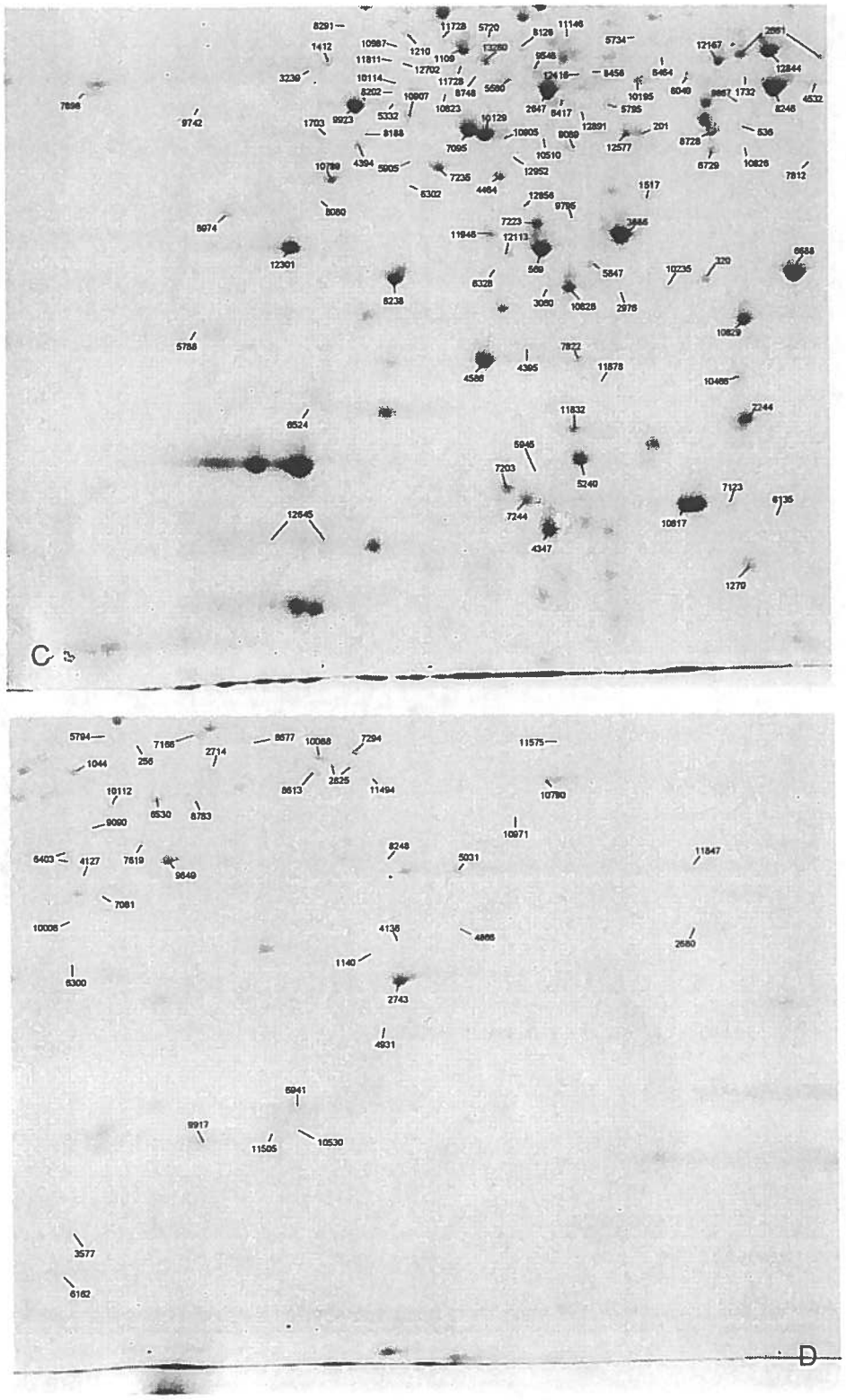



**Figure 4.** Annotated sections of the master gel (**A** – **D**) refer to the four sections in Figure 3. Assigned numbers represent genes. Identified proteins are listed in Table 2 according to functional classes.

**Proteins with functional assignment**

Functions could be assigned to the majority (366) of the identified proteins. These proteins were classified according to their predicted functions into nine categories (see Table 1). The category "metabolism" was further divided into eight sub-categories (see Table 2). The sub-category "sulfatases" was used, since the presence of 110 sulfatase encoding genes was one of the major unexpected findings from the annotation of the *R. baltica* genome (Glöckner et al. 2003), ten of which were identified on the gel.

The pattern of identified proteins displayed in Figure 3 is typical of exponentially growing cells, with the most abundant proteins involved mainly in housekeeping functions, *e.g.* GAP DH (RB2627) of glycolysis, malate dehydrogenase (RB7652) of tricarboxic acids cycle, glutamate synthase (RB5653) of amino acid biosynthesis, protease (RB9402) of protein quality control and translation. The fact that several proteins existed as more than a single spot could point to thus far unknown posttranslational modifications. In particular, proteins with high molecular weight formed chains with the same $M_r$ but differing p*I*s.

**Hypothetical and conserved hypothetical proteins**

About 17% of the identified proteins represent predicted hypothetical proteins that are apparently unique to *R. baltica*. Thus the present study for the first time provides experimental evidence that genes coding for hypothetical proteins are actually expressed under standard growth conditions and consequently have to be considered relevant for the physiology of *R. baltica*. Conserved hypothetical proteins constitute about the same percentage of identified proteins. In both cases no functions could be assigned. Nevertheless, the hypothetical proteins could be of particular interest with respect to the cell cycle and unusual morphological features of *R. baltica*, which may require the activity of thus far unknown proteins.

**Identification of unpredicted proteins**

Initial analysis of the *R. baltica* genome sequence with three different ORF prediction programs (Orpheus, Glimmer and Critica) generated a non-redundant set of 13,331 predicted ORFs. Manual removal of presumably overpredicted ORFs resulted in the published set of 7,325 ORFs (BX119912; Glöckner et al. 2003). Thus the possibility exists that ORFs were initially not predicted or were erroneously removed during manual refinement. This possibility was compounded by the observation that several PMFs with abundant signals did not result in identification of a protein.

As a first attempt to identify proteins encoded by genes that were not present in the predicted ORF set, the PMF data was searched against an amino acid sequence database translated from a *Maximal ORF Set* (MOS); a highly redundant set consisting of 578,949 sequence entries, designed to contain all possible genes and all possible reading lengths thereof. To reduce the number of false positive results, identifications for which the experimental and calculated molecular weight differed by >30% were discarded. This database search retrieved 4 proteins with scores >51, which are listed in Table 3. All of the newly identified genes code for hypothetical proteins, which are surrounded by further hypothetical or conserved hypothetical proteins in the genomic context. For example, ORF 9191 from MOS was identified with a Mascot score of 101 and a sequence coverage of 62%. The position of the corresponding spot on the 2-DE gel was used as a guide to suggest the ORF length by defining the probable start codon. The product of ORF 9191 is therefore predicted to have a molecular weight of 25 kDa. These results indicate that PMF is not necessarily restricted to identification in protein databases, but can also be used to refine ORF prediction. However, future analysis should include MS/MS to verify the identity of the additional proteins.

**Table 3.** Proteins (new ORFs) specifically identified from the Maximal ORF Set (MOS).

| ORF-No. | Start | Stop | Length [aa] | Mascot score | Predicted function | Genetic context |
|---|---|---|---|---|---|---|
| pir.6532c | 1798993 | 1799565 | 290 | 61 | hypothetical | Methionine aminopeptidase, hypotheticals; other strand: ribose-regulated sugar-ADH |
| pir.8508 | 2358829 | 2359248 | 139 | 60 | hypothetical | Mostly hypotheticals |
| pir.9191c | 2546400 | 2546921 | 173 | 101 | hypothetical | Mostly hypotheticals, downstream of possible adenylate cyclase |
| pir.15895 | 4437587 | 4438426 | 279 | 56 | hypothetical | Mostly hypotheticals; upstream of D-tyrosyl-tRNA(Tyr)deacylase |

**Signal peptides and protein localization**

As observed with other described *Planctomycetes*, cells of *R. baltica* contain membrane separated intracytoplasmic compartments (Lindsay et al. 2001). The internal region is termed pirellulosome and contains the riboplasm with ribosome-like particles and the condensed nucleoid (Figure 5). The region between the intracytoplasmic and cytoplasmic membranes contains the paryphoplasm that harbours some RNA but no ribosome-like particles. The finding that ribosome-like particles are confined to the riboplasm suggests that protein biosynthesis only takes place in this compartment. Due to the cellular compartmentalization in *R. baltica* an extensive protein translocation can be expected.

According to the signal hypothesis (Blobel 2000) the majority of secreted proteins has a signal peptide, which is found in 1160 (16%) of the predicted proteins in *R. baltica*. Out of the 558 identified proteins annotated in the master gel 146 (26%) possess a signal peptide (Table 2). Since the applied methods for cell breakage did not separate riboplasmic from paryphoplasmic proteins one can conclude that the 146 signal peptide containing proteins have potentially been secreted and are actually localized in the paryphoplasm or cell wall associated.

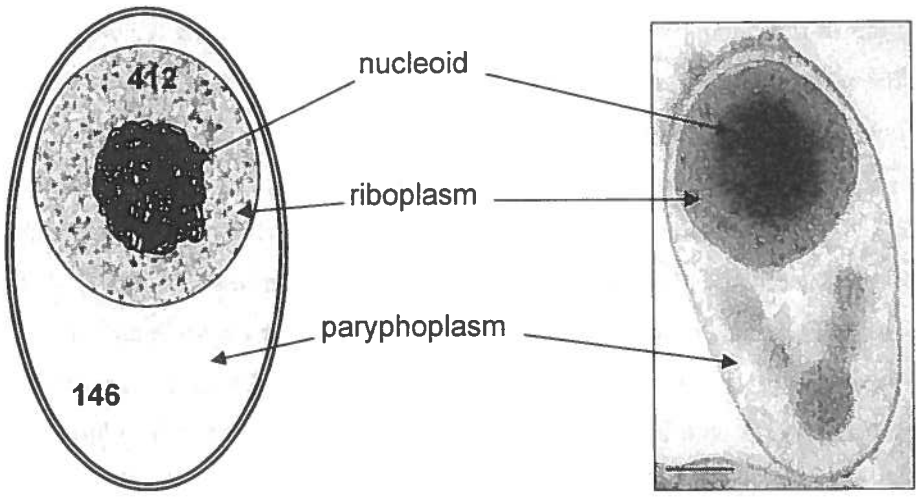



**Figure 5.** Intracellular compartmentalization of *R. baltica* and possible location of identified proteins (Bar = 0.2 µm).

For 58% (56 proteins) of the hypothetical proteins a signal peptide was predicted. Thirtysix (about 38%) of the 95 conserved hypothetical proteins are secreted according to the signalP prediction. Secreted proteins with functional assignment are mainly dehydro-

genases, hydrolases for extracellular macromolecules or involved in signal transduction (phosphoprotein kinases). In contrast, the enzymes performing housekeeping functions seem to be confined to the riboplasm (no signal peptide). Interestingly, 9 of the 10 identified sulfatases have a signal peptide prediction. The *R. baltica* genome encodes 110 sulfatases, which are suggested to function in extracellular degradation of sulfated glycopolymers such as *e.g.* carrageen (Glöckner et al. 2003). Thus the identified sulfatases could be in the process of being excreted, since proteins already excreted to the extracellular space would have been lost under the applied conditions of cell harvesting. Expression of sulfatase encoding genes might not require the presence of sulfated substrates, since the studied *R. baltica* cells were grown with ribose as only source of organic carbon. In some cases the correctness of the signal peptide prediction is questionable, *e.g.* for the elongation factor G and lysyl-tRNA synthase. Both enzymes play an important role in protein synthesis, a process that should exclusively take place in the riboplasm. Thus the presence of a signal peptide alone does not allow defining the exact target region of translocation. Future research on secreted proteins (secretome) will have to differentiate the different compartments present in *R. baltica* cells.

**Predicted highly expressed (PHX) genes**

Among the 30 most abundant proteins on the master gel of *R. baltica,* 27 were encoded by genes that were predicted to be highly expressed (PHX) according to codon usage adaptation. Thus, a correlation between experimentally determined protein abundance and codon usage features as it already has been shown for fast-growing bacteria (Karlin et al. 2001) could also be observed for *R. baltica*, a slowly-growing environmental bacterium (doubling times between 10 – 14 hours; Rabus et al. 2002). As expression predictions based on codon usage adaptation reflect gene expression under most environmental conditions, one would expect some variation in expression levels when the cells are grown under specific laboratory conditions with defined substrates. Indeed, there are some exceptions where the genes of proteins appearing as highly abundant on 2-DE gels are not PHX; this applies mainly for proteins that were *e.g.* specifically induced during growth with ribose (see adjacent paper) or proteins affiliated with lipid metabolism.

## Conclusions

With more than 550 identified gene products, the present study established a solid proteomic framework for further analysis of differential gene expression in *R. baltica*. Considering the nutritional specialization of this bacterium on the utilization of carbohydrates, we will be able to reconstruct the major catabolic routes which are operative in *R. baltica* and to learn about the potential of this bacterium to regulate the expression of catabolic genes in response to the availability of respective growth substrates (see accompanying publication). The master gel will also be beneficial for the identification of proteins involved in cell cycle and development. Such proteins should be related to the two morphotypes (single cells versus aggregates) as well as to different growth stages.

## Acknowledgements

We thank Alfred Beck and Thomas Kreitler for bioinformatics support. We are indepted to Friedrich Widdel for continuous support of proteomic work at the MPI in Bremen. This study was supported by the Max Planck Society.

**Table 2.** Predicted functions of proteins annotated in the master gel (Fig. 4 A – D). Information on quality (**Score**) / reproducibility (*) of identification, signal peptide prediction (**sp**) and prediction level of gene expression (**PHX**) are provided. See text for details.

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Metabolism (250 proteins)** | | | | |
| | C-compound and Carbohydrate | | | | |
| 201 | sugar phosphate isomerase/epimerase | - | - | 190 | 1 |
| 307 | NAD-dependent malic enzyme | - | + | 201 | 2 |
| 344 | xanthan lyase | + | - | 132 | |
| 399 | glucose-6-phosphate isomerase | - | - | 139 | 1 |
| 548 | 1,4-alpha-glucan branching enzyme | - | - | 233 | 2 |
| 856 | L-lactate/malate dehydrogenase | + | - | 131 | |
| 1210 | hexulose-6-phosphate isomerase | - | + | 105 | 1 |
| 1231 | dihydrolipoamide dehydrogenase | - | + | 135 | 2 |
| 1358 | ADP-glucose pyrophosphorylase | - | - | 198 | |
| 1412 | inositol monophosphatase | - | - | 124 | 1 |
| 1593 | isocitrate dehydrogenase | - | - | 192 | 2 |
| 1988 | glucose dehydrogenase | + | - | 283 | |
| 2114 | aconitate hydratase | - | + | 178 | 2 |
| 2160 | alpha-amylase | - | + | 229 | |
| 2373 | formaldehyde dehydrogenase | - | + | 213 | 2 |
| 2403 | D-mannonate oxidoreductase | - | - | 150 | |
| 2518 | GDP-mannose 4,6 dehydratase | - | + | 162 | 2 |
| 2627 | glyceraldehyde 3-phosphate dehydrogenase | - | + | 230 | 2 |
| 2638 | glycogen branching enzyme | - | - | 133 | 1 |
| 2658 | xylose isomerase | - | + | 196 | 2 |
| 2817 | 6-phosphogluconate dehydrogenase | - | + | 194 | 2 |
| 3193 | transaldolase | - | + | 233 | 2 |
| 3239 | D-tagatose 3-epimerase | - | + | 54 | 1 |
| 3265 | glucose-fructose oxidoreductase | - | - | 117 | |
| 3423 | pyruvate dehydrogenase, E2 component | - | + | 144 | 2 |
| 3424 | pyruvate dehydrogenase, E1 component | - | + | 288 | |
| 3488 | sorbitol dehydrogenase | - | - | 244 | |
| 3499 | ribokinase | - | - | 124 | |
| 4131 | alcohol dehydrogenase | - | - | 108 | 2 |
| 5200 | alpha-amylase | - | - | 129 | |
| 5243 | endo-1,4-beta-xylanase B | - | - | 72 | 2 |
| 5321 | myo-inositol catabolism protein IolH | - | - | 99 | |
| 5948 | alcohol dehydrogenase | - | - | 157 | 2 |
| 6061 | phosphomannomutase | - | - | 310 | 2 |
| 6254 | mannose-1-phosphate guanylyltransferase | - | - | 80 | |
| 6394 | 2-hydroxyacid dehydrogenase | - | - | 162 | 2 |
| 6683 | citrate synthase | - | - | 150 | 2 |
| 6690 | fructose-1,6-bisphosphate aldolase | - | + | 130 | 2 |
| 6729 | deoxyribose-phosphate aldolase | - | + | 55 | 1 |
| 6759 | methenyltetrahydromethanopterin cyclohydrolase | - | - | 60 | |
| 6807 | sialic acid-specific 9-O-acetylesterase | + | - | 131 | |
| 6841 | UDP-N-acetylglucosamine pyrophosphorylase | - | - | 142 | 1 |
| 6977 | UDP-N-acetylhexosamine pyrophosphorylase | - | - | 158 | 2 |
| 7095 | triosephosphate isomerase | - | + | 109 | 2 |
| 7294 | glucose 1-dehydrogenase | - | - | 159 | 1 |
| 7572 | 6-phosphofructokinase, pyrophosphate-dependent | - | + | 105 | 1 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | C-compound and Carbohydrate (continued) | | | | |
| 7652 | malate dehydrogenase | - | + | 138 | 2 |
| 8073 | alpha-L-arabinofuranosidase II | + | - | 151 | 2 |
| 8248 | carboxymethylenebutenolidase | - | - | 112 | |
| 8541 | endoglucanase | - | + | 147 | 2 |
| 8562 | phosphoglycerate mutase | - | - | 82 | |
| 8731 | 2-keto-3-deoxygluconate kinase | - | + | 91 | 2 |
| 8924 | phosphonopyruvate decarboxylase 1 | - | + | 131 | 2 |
| 8941 | ketoglutarate semialdehyde dehydrogenase | - | - | 67 | 2 |
| 9089 | 6-phosphogluconolactonase | - | - | 117 | 2 |
| 9651 | sialic acid-specific 9-O-acetylesterase | - | - | 80 | |
| 10002 | glucose dehydrogenase | - | + | 133 | 2 |
| 10048 | sialic acid-specific 9-O-acetylesterase | - | + | 140 | |
| 10092 | hydratase, aerobic aromate catabolism | - | - | 74 | 2 |
| 10124 | polyvinylalcohol dehydrogenase | + | - | 147 | |
| 10127 | PQQ-dependent glucose dehydrogenase | + | - | 89 | 2 |
| 10144 | endo-1,4-beta-xylanase B | + | - | 121 | |
| 10172 | aldehyde dehydrogenase | - | + | 218 | 2 |
| 10277 | pyruvate kinase | - | - | 138 | 2 |
| 10500 | phosphoglycerate kinase | - | + | 142 | 2 |
| 10503 | lipopolysaccharide biosynthesis protein BplA | - | + | 142 | 2 |
| 10554 | succinate dehydrogenase subunit A | - | + | 238 | 2 |
| 10591 | PPi-phosphofructokinase | - | + | 164 | 2 |
| 10617 | succinyl-CoA synthetase beta subunit | - | + | 99 | 2 |
| 10619 | succinyl-CoA synthetase alpha subunit | - | + | 103 | 2 |
| 10817 | ribose 5-phosphate epimerase | - | + | 81 | |
| 12361 | ribokinase family sugar kinase | - | - | 175 | 2 |
| 12381 | enolase | - | + | 212 | 2 |
| 12740 | gluconolactonase precursor | - | - | 101 | |
| 12921 | transketolase | - | + | 151 | 2 |
| 13260 | alcohol dehydrogenase | - | - | 73 | |
| 13264 | acetyl-coenzyme A synthetase | - | - | 220 | 2 |
| | Amino Acids and Proteins | | | | |
| 1225 | dipeptidyl peptidase IV | + | - | 158 | |
| 1317 | 2-isopropylmalate synthase | - | - | 147 | |
| 1359 | serine protease | - | - | 156 | 2 |
| 1411 | dihydrodipicolinate synthase | - | + | 178 | 2 |
| 1732 | beta-alanine synthetase | + | - | 73 | |
| 1898 | dehydroquinate synthase | - | - | 101 | |
| 2261 | carboxypeptidase-related protein | - | + | 187 | 1 |
| 2278 | 3-phosphoshikimate 1-carboxyvinyltransferase | - | - | 131 | |
| 2552 | *N*-acetyl-gamma-glutamyl-phosphate reductase | - | + | 172 | 1 |
| 2661 | UDP-*N*-acetylenolpyruvoylglucosamine reductase | - | - | 63 | 2 |
| 2746 | dihydrodipicolinate synthase | - | - | 100 | |
| 3824 | L-aspartate oxidase | - | - | 176 | |
| 3842 | dipeptidyl peptidase IV | + | - | 220 | |
| 4282 | matrix metalloproteinase-11 | - | - | 71 | |
| 4394 | proteinase | - | + | 65 | |
| 4928 | aminopeptidase | - | - | 244 | |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | Amino Acids and Proteins (continued) | | | | |
| 5444 | S-adenosylmethionine synthetase | - | + | 234 | 2 |
| 5560 | tryptophan synthase alpha chain | - | - | 126 | 2 |
| 5653 | NADH-glutamate synthase small chain | - | + | 134 | 1 |
| 5720 | amidohydrolase | - | - | 74 | |
| 5986 | ornithine carbamoyltransferase | - | - | 151 | |
| 6248 | phosphoglycerate dehydrogenase | - | + | 149 | 2 |
| 6285 | adenosylhomocysteinase | - | + | 152 | 2 |
| 6300 | glutamine amido-transferase | - | - | 119 | |
| 6821 | aspartate aminotransferase | - | + | 88 | 1 |
| 6932 | cysteine synthase | - | - | 126 | 2 |
| 7359 | gamma-glutamyl phosphate reductase | - | + | 99 | 1 |
| 7375 | aminopeptidase T | - | - | 115 | 1 |
| 7584 | glycine dehydrogenase (decarboxylating) subunit 2 | - | - | 154 | |
| 7587 | aminotransferase-glycine cleavage system T protein | - | - | 238 | |
| 7590 | proteinase | - | - | 174 | 2 |
| 7823 | transaminase | - | - | 123 | |
| 7941 | cysN/cysC bifunctionyal enzyme | - | - | 245 | 2 |
| 8080 | phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase | - | - | 127 | |
| 8126 | branched-chain amino acid aminotransferase | - | + | 111 | 2 |
| 8219 | aspartate aminotransferase | - | - | 135 | 2 |
| 8262 | proline dehydrogenase | - | - | 147 | |
| 8293 | argininosuccinate synthase | - | + | 144 | 2 |
| 8633 | acetylornithine aminotransferase | - | + | 119 | |
| 8926 | aspartokinase | - | + | 93 | |
| 9029 | metalloproteinase | - | - | 200 | 2 |
| 9107 | chorismate mutase | - | - | 150 | |
| 9402 | protease | - | + | 170 | 1 |
| 9674 | X-pro dipeptidyl-peptidase | + | - | 134 | 1 |
| 9795 | aspartate-semialdehyde dehydrogenase | - | + | 117 | 2 |
| 9857 | 5-methyltetrahydrofolate-homocysteine methyltransferase | - | - | 117 | |
| 9869 | acetohydroxy acid isomeroreductase | - | + | 170 | 2 |
| 10112 | imidazole glycerol phosphate synthase subunit hisF | - | - | 221 | 2 |
| 10114 | indole-3-glycerol phosphate synthase | - | + | 161 | |
| 10180 | peptidase | - | + | 173 | 2 |
| 10272 | pteridine reductase | - | - | 136 | 1 |
| 10287 | dihydropicolinate synthase | - | + | 166 | 2 |
| 10586 | aminopeptidase | - | - | 165 | 1 |
| 10826 | ATP-dependent clp protease proteolytic subunit | - | - | 120 | 2 |
| 10829 | ATP-dependent clp protease proteolytic subunit | - | + | 88 | 1 |
| 10894 | threonine synthase precursor | - | + | 109 | 2 |
| 11847 | methionine sulfoxide reductase | + | - | 81 | |
| 11878 | methionine sulfoxide reductase | - | PA | 56 | |
| 11879 | periplasmic serine proteinase | - | + | 91 | |
| 11919 | dihydroxy-acid dehydratase | - | - | 59 | |
| 11959 | dihydrodipicolinate reductase | - | - | 114 | 2 |
| 12087 | dihydroxy-acid dehydratase | - | + | 94 | 2 |
| 12107 | cytosol aminopeptidase | - | + | 165 | 2 |
| 12113 | carbamoyl-phosphate synthase large chain | - | + | 177 | 2 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | Amino Acids and Proteins (continued) | | | | |
| 12133 | succinyl-diaminopimelate desuccinylase | - | + | 178 | 1 |
| 12148 | periplasmic tail-specific proteinase | + | + | 287 | 2 |
| 12337 | prolyl endopeptidase | + | + | 240 | 2 |
| 12510 | phospho-2-dehydro-3-deoxyheptonate aldolase | - | - | 143 | |
| 12597 | 3-isopropylmalate dehydrogenase | - | - | 131 | |
| 12656 | 3-isopropylmalate dehydratase large subunit | - | - | 226 | 2 |
| 12905 | acetolactate synthase III precursor | - | + | 172 | 2 |
| | Nucleotides | | | | |
| 256 | formyltetrahydrofolate deformylase | - | - | 145 | |
| 1386 | nucleoside hydrolase | + | - | 128 | |
| 1784 | UDP-glucose 6-dehydrogenase | - | - | 195 | 2 |
| 1819 | UDP-glucose 4-epimerase | - | - | 176 | 2 |
| 3751 | UDP-glucose 4-epimerase | - | - | 146 | 2 |
| 4043 | glucose-1-phosphate thymidylyltransferase | - | - | 89 | |
| 4752 | dihydroorotate dehydrogenase | - | - | 101 | |
| 5395 | phosphoribosylformylglycinamidine synthase II | - | + | 367 | 2 |
| 5603 | ATP phosphoribosyltransferase | - | - | 189 | 2 |
| 5695 | beta-alanine synthetase | - | - | 173 | 1 |
| 5847 | adenine phosphoribosyltransferase | - | - | 92 | 2 |
| 6135 | phosphoribosylaminoimidazole carboxylase catalytic subunit | - | - | 67 | |
| 6302 | ADP-ribose pyrophosphatase | - | - | 79 | |
| 6328 | adenylyl cyclase | - | - | 107 | |
| 6524 | hypoxanthine-guanine phosphoribosyltransferase | - | - | 79 | |
| 6616 | phosphoribosylamine-glycine ligase | - | + | 99 | 1 |
| 7468 | methylentetrahydrofolate cyclohyolase | - | - | 59 | |
| 8374 | GMP synthase | - | + | 265 | 2 |
| 8613 | phosphoribosylformylglycinamidine synthase I | - | - | 133 | 2 |
| 8748 | dihydroorotate dehydrogenase | - | + | 65 | |
| 10113 | bifunctional purine biosynthesis protein purH | - | + | 224 | 2 |
| 10192 | dihydroorotase | + | + | 96 | 2 |
| 10510 | cytidylate kinase | - | + | 89 | |
| 11832 | nucleoside diphosphate kinase | - | + | 71 | 2 |
| 12745 | phosphoribosylformylglycinamidine cyclo-ligase | - | - | 109 | 2 |
| | Lipids, Fatty Acids and Isoprenoids | | | | |
| 314 | malonyl CoA-acyl carrier protein transacylase | - | + | 101 | 2 |
| 320 | 3-oxoacyl-(acyl-carrier-protein) synthase | - | - | 173 | 2 |
| 1586 | 3-oxoacyl-(acyl-carrier-protein) synthase II | - | - | 104 | 2 |
| 1839 | thiamine biosynthesis lipoprotein apbE | - | - | 136 | 2 |
| 2144 | geranylgeranyl pyrophosphate synthetase precursor | - | - | 78 | |
| 2579 | ethanolamine utilization protein EutE | - | + | 67 | |
| 2825 | glycerophosphodiester phosphodiesterase | - | - | 113 | 1 |
| 4527 | 3-oxoacyl-(acyl-carrier protein) synthase | - | + | 177 | 1 |
| 6272 | 3-oxoacyl-(acyl-carrier protein) synthase | - | - | 97 | |
| 6464 | sulfolipid biosynthesis protein | - | - | 85 | |
| 7171 | 3-oxoacyl-(acyl-carrier protein) synthase | - | + | 219 | 2 |
| 7812 | enoyl-CoA hydratase/isomerase | - | - | 95 | |
| 8125 | trans-2-enoyl-ACP reductase | - | + | 75 | 2 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | Lipids, Fatty Acids and Isoprenoids (continued) | | | | |
| 8550 | biotin carboxylase | - | + | 212 | 2 |
| 10466 | probable beta-hydroxyacyl-ACP dehydratase | - | + | 72 | 1 |
| 10790 | enoyl-(acyl-carrier-protein) reductase (NADH) | - | + | 127 | 2 |
| 11008 | cycloartenol synthase | + | - | 242 | 2 |
| 12812 | 3-oxoacyl-(ACP) synthase III | - | - | 60 | |
| | Vitamins, Cofactors, and Prosthetic Groups | | | | |
| 24 | L-sorbosone dehydrogenase | - | - | 197 | 2 |
| 309 | magnesium protoporphyrin chelatase | - | - | 95 | 2 |
| 536 | pyridoxal phosphate biosynthetic protein | - | - | 128 | 1 |
| 2143 | 1-deoxy-D-xylulose 5-phosphate synthase | - | - | 226 | |
| 6809 | thiamine-monophosphate kinase | - | - | 69 | 1 |
| 6831 | glutamate-1-semialdehyde 2,1-aminomutase | - | - | 89 | 1 |
| 6964 | L-sorbosone dehydrogenase | + | - | 123 | |
| 9090 | 3-methyl-2-oxobutanoate hydroxymethyltransferase | - | - | 64 | |
| 10006 | pyridoxamine oxidase | - | - | 71 | |
| 11582 | cysteine desulfurase | - | - | 86 | 1 |
| 12480 | riboflavin biosynthesis protein RibA | - | + | 91 | |
| | Sulfatases | | | | |
| 198 | *N*-acetylgalactosamine-4-sulfatase precursor | + | - | 129 | |
| 1610 | arylsulfatase | + | - | 100 | |
| 2367 | sulfatase | + | + | 121 | |
| 3403 | *N*-acetylgalactosamine 6-sulfatase | + | - | 65 | |
| 3877 | aryl-sulfate sulphohydrolase | + | - | 137 | |
| 4017 | sulfatase | - | - | 114 | |
| 7481 | arylsulphatase A | + | + | 162 | |
| 9498 | arylsulfatase | + | + | 139 | 2 |
| 10599 | sulfatase 1 precursor | + | - | 161 | |
| 11502 | alkylsulfatase | + | - | 60 | 2 |
| | Inorganic Compounds | | | | |
| 5869 | bacterioferritin comigratory protein | - | - | 79 | |
| 6049 | adenylylsulfate kinase | - | - | 110 | |
| 7247 | glutamine synthetase II | - | + | 136 | 2 |
| 7465 | sulfite reductase | - | + | 65 | |
| 11670 | ferric enterobactin esterase-related protein | - | - | 117 | |
| | Others | | | | |
| 203 | oxidoreductase | - | - | 154 | 2 |
| 1555 | NADH-dependent dehydrogenase | + | + | 154 | 2 |
| 1608 | esterase | + | - | 125 | 2 |
| 1939 | oxidoreductase | - | + | 212 | |
| 2242 | oxidoreductase | - | + | 164 | 1 |
| 3317 | NADH-dependent dehydrogenase | + | - | 108 | |
| 3330 | dehydrogenase | - | + | 104 | 1 |
| 3405 | hydrolase | + | - | 275 | 2 |
| 4404 | oxidoreductase | - | - | 124 | 2 |
| 4432 | oxidoreductase | + | - | 67 | 1 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | Others (continued) | | | | |
| 5332 | phosphoesterase | - | - | 53 | |
| 5365 | NADH-dependent dehydrogenase | + | + | 165 | 2 |
| 6199 | dehydrogenase | + | + | 94 | 2 |
| 6985 | NADH-dependent dehydrogenase | - | + | 86 | |
| 7081 | oxidoreductase | - | - | 71 | |
| 7482 | CDP-tyvelose epimerase | - | + | 183 | 2 |
| 7548 | syringomycin biosynthesis enzyme 2 | - | - | 133 | 2 |
| 8679 | oxidoreductase | - | - | 121 | |
| 8728 | oxidoreductase | - | + | 119 | 2 |
| 8781 | NADH-dependent oxidoreductase | - | + | 182 | |
| 8799 | NADH-dependent dehydrogenase | - | - | 159 | 2 |
| 8937 | NADH-dependent dehydrogenase | - | - | 94 | |
| 9168 | nucleotide sugar epimerase | - | - | 68 | 2 |
| 9584 | oxidoreductase | - | - | 74 | |
| 9586 | oxidoreductase | - | - | 116 | |
| 9971 | NADH-dependent dehydrogenase | + | + | 116 | 1 |
| 10652 | C-methyltransferase | - | - | 103 | 1 |
| 10967 | oxidoreductase | - | + | 91 | |
| 10971 | dehydrogenase | - | - | 96 | 1 |
| 11146 | hydrolase | - | - | 99 | |
| 11859 | hydrolase | - | - | 167 | |
| 12019 | oxidoreductase | - | - | 74 | 1 |
| 12564 | NADH-dependent dehydrogenase | + | - | 139 | 1 |
| | **Transport (11 proteins)** | | | | |
| 1248 | ATPase component; multidrug transport system | - | + | 117 | 2 |
| 1517 | ATP-binding protein, lipoprotein releasing system | - | - | 75 | |
| 4866 | ATP-binding protein, lipoprotein releasing system | - | - | 69 | |
| 5795 | PTS system, fructose-specific IIABC component | - | - | 114 | 2 |
| 6236 | ATP-binding protein, ABC-transport system | - | - | 147 | 2 |
| 7166 | ATP-binding protein, ABC-transport system | - | - | 89 | 1 |
| 7211 | ATP-binding protein, phosphate transport | - | - | 186 | |
| 9998 | ATP-binding protein, ABC-transport system | - | + | 67 | |
| 10709 | periplasmic dipeptide transport protein precursor | + | - | 203 | |
| 11930 | ATP-binding protein, ABC-transport system | - | - | 158 | 2 |
| 12859 | ATP-binding protein, oligopeptide transport | - | - | 124 | |
| | **Others (8 proteins)** | | | | |
| 3895 | internalin | + | - | 93 | |
| 4879 | nodulin-26 | - | - | 117 | 1 |
| 10228 | twitching motility protein PilB, biogenesis of pili | - | + | 118 | |
| 10338 | FlbA protein, biogenesis of flagellae | - | - | 88 | |
| 10463 | ferredoxin-NADP reductase | - | - | 145 | |
| 10905 | phosphoesterase PH1616 | - | + | 132 | 1 |
| 10907 | phosphoesterase PH1616 | - | - | 143 | 2 |
| 12774 | type IV fimbrial assembly protein PilB | - | - | 269 | 1 |
| | **Genetic Information Processing (52 proteins)** | | | | |
| 539 | competence-damage inducible protein CinA | - | - | 93 | 1 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Genetic Information Processing (continued)** | | | | |
| 1270 | translation initiation inhibitor | - | + | 65 | |
| 1485 | DNA polymerase beta family | - | - | 193 | |
| 1964 | DNA-directed RNA polymerase alpha chain | - | + | 191 | 2 |
| 2543 | 30S ribosomal protein S1 | - | + | 159 | 2 |
| 3446 | peptidyl-prolyl cis-trans isomerase cyp2 | - | - | 100 | 2 |
| 3886 | ribosome recycling factor | - | - | 106 | 2 |
| 4143 | glutamyl-tRNA amidotransferase subunit A | - | - | 127 | 2 |
| 4395 | macrophage infectivity potentiator (map) protein | - | + | 83 | |
| 4675 | cysteinyl-tRNA synthetase | - | - | 108 | 2 |
| 5178 | prolyl tRNA synthetase | - | - | 218 | 2 |
| 5414 | DNA-directed RNA polymerase beta chain | - | + | 53 | |
| 5434 | translational elongation factor G | - | + | 126 | 1 |
| 5681 | trigger factor | - | + | 142 | 2 |
| 5697 | thiol-disulfide interchange protein | - | - | 52 | |
| 5747 | arginyl-tRNA synthetase | - | - | 205 | 1 |
| 5754 | DnaK, chaperone | - | - | 97 | 1 |
| 5778 | alkaline phosphatase | - | - | 234 | |
| 5804 | polyribonucleotide nucleotidyltransferase | - | + | 56 | |
| 5813 | alkaline phosphatase D | + | - | 105 | |
| 6123 | protein disulfide-isomerase | - | + | 201 | 2 |
| 6436 | tryptophan-tRNA synthetase | - | - | 79 | 1 |
| 7112 | phenylalanyl-tRNA synthetase beta chain | - | - | 131 | |
| 7114 | phenylalanyl-tRNA synthetase alpha chain | - | - | 129 | 1 |
| 7237 | DNA mismatch repair protein MUTS | - | + | 98 | |
| 7244 | peptidylprolyl cis-trans isomerase | - | - | 61 | 1 |
| 7821 | translational elongation factor G | + | + | 253 | 2 |
| 7894 | translational elongation factor Tu | - | + | 187 | 2 |
| 8253 | aspartyl-tRNA synthetase | - | + | 249 | 2 |
| 8328 | CMP-binding protein | - | + | 151 | |
| 8649 | peptidylprolyl cis-trans isomerase | + | + | 100 | |
| 8889 | alkaline phosphatase D precursor | + | - | 132 | |
| 8919 | leucyl-tRNA synthetase | - | + | 172 | |
| 8966 | 60 kDa chaperonin | - | + | 198 | 2 |
| 8970 | 60 kDa chaperonin | - | + | 148 | 2 |
| 8974 | GrpE chaperone | - | + | 57 | |
| 9103 | ATPases with chaperone activity, ATP-binding subunit | - | - | 155 | 1 |
| 9105 | DnaK | - | + | 168 | |
| 9917 | single-strand binding protein | - | - | 110 | |
| 9923 | 50S ribosomal protein L25 | - | + | 117 | 2 |
| 9927 | ATP-dependent Clp protease ATP-binding subunit | - | + | 89 | |
| 10108 | DNA polymerase III, beta chain | - | - | 212 | 1 |
| 10129 | macrophage infectivity potentiator (map) protein | + | + | 76 | |
| 10629 | GroEL, heat shock protein | - | + | 172 | 2 |
| 10640 | translational elongation factor Ts | - | + | 211 | 2 |
| 10852 | glutamyl-tRNA amidotransferase subunit B | - | + | 114 | 2 |
| 10883 | lysyl-tRNA synthetase | + | + | 153 | 2 |
| 12577 | translation elongation factor EF-P | - | - | 90 | |
| 12626 | DNA-directed RNA polymerase alpha chain | - | + | 287 | 2 |
| 12799 | DNA polymerase I | - | + | 205 | |
| 12854 | methionyl-tRNA formyltransferase | - | - | 159 | 2 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Regulation and Signal Transduction (24)** | | | | |
| 983 | phosphoprotein kinase | + | - | 189 | |
| 1140 | response regulator | - | - | 101 | |
| 1321 | transcription repressor | - | - | 93 | |
| 1483 | sensor histidine kinase/response regulator | - | - | 58 | |
| 2743 | nitrate/nitrite regulatory protein NarP | - | + | 159 | 2 |
| 4081 | regulatory protein | - | - | 146 | 2 |
| 4136 | regulatory components of sensory transduction system | - | + | 53 | |
| 4487 | nitrogen assimilation regulatory protein | - | - | 168 | 2 |
| 5905 | phosphoprotein phosphatase | - | - | 60 | |
| 6403 | response regulator | - | - | 119 | 1 |
| 6486 | phosphoprotein kinase | + | - | 165 | |
| 6491 | RNA polymerase subunit sigma-54 | - | - | 134 | |
| 6603 | MoxR-related protein | - | + | 164 | 1 |
| 7123 | response regulator | - | - | 56 | |
| 7541 | phosphoprotein kinase | + | + | 171 | 2 |
| 7898 | transcription antiterminator NusG | - | + | 109 | 1 |
| 8173 | MoxR-related protein | - | - | 113 | |
| 9108 | MoxR-related protein | - | - | 109 | 2 |
| 9110 | phosphoprotein kinase | - | - | 162 | 2 |
| 10491 | two-component system regulatory protein | - | - | 71 | |
| 10517 | methanol dehydrogenase regulation homolog YeaC | - | - | 134 | 1 |
| 10839 | phosphoprotein kinase | - | - | 101 | |
| 11660 | phosphoprotein kinase | + | - | 218 | |
| 12952 | two-component system, regulatory protein | - | - | 69 | |
| | **Energy (8 proteins)** | | | | |
| 1831 | $Na^+$-translocating NADH:ubiquinone oxidoreductase NqrA | - | - | 220 | 2 |
| 1833 | $Na^+$-translocating NADH:ubiquinone oxidoreductase NqrC | + | + | 106 | |
| 4399 | quinone oxidoreductase | - | - | 185 | 1 |
| 7084 | pyrophosphatase | - | - | 165 | |
| 10215 | $H^+$-transporting ATP synthase alpha chain | - | + | 124 | 2 |
| 10217 | $H^+$-transporting ATP synthase beta chain | - | + | 307 | 2 |
| 11946 | thermophilic NAD(P)H-flavin oxidoreductase | - | - | 125 | 2 |
| 11985 | quinone oxidoreductase | - | - | 191 | 1 |
| | **Stress Response (13 proteins)** | | | | |
| 390 | alkylhalidase, dehalogenase | - | - | 167 | |
| 2244 | glutathione peroxidase | + | + | 75 | |
| 2799 | general stress protein 69 | - | - | 107 | 2 |
| 4586 | thiol peroxidase | - | + | 92 | 2 |
| 6384 | thioredoxin related protein | + | - | 116 | |
| 6688 | superoxide dismutase, Mn family | - | + | 86 | 2 |
| 7223 | thioredoxin reductase | - | + | 76 | 2 |
| 8238 | peroxiredoxin 2 | - | + | 72 | 2 |
| 8674 | thioredoxin | + | + | 145 | |
| 8870 | multi drug resistance protein | - | + | 127 | 1 |
| 10727 | manganese-containing catalase | - | - | 73 | 2 |
| 11150 | xenobiotic reductase B | - | - | 114 | 1 |
| 12541 | thioredoxin | + | - | 214 | 2 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Conserved Hypothetical Proteins (95 proteins)** | | | | |
| 85 | conserved hypothetical protein | + | - | 114 | |
| 452 | conserved hypothetical protein | - | - | 215 | 1 |
| 457 | conserved hypothetical protein | - | - | 183 | |
| 520 | conserved hypothetical protein | - | - | 145 | 1 |
| 538 | conserved hypothetical protein | - | - | 232 | 2 |
| 569 | conserved hypothetical protein | + | + | 74 | 2 |
| 1044 | conserved hypothetical protein | - | - | 221 | 1 |
| 1109 | conserved hypothetical protein | + | + | 70 | 2 |
| 1703 | Maf protein | - | - | 56 | |
| 1731 | conserved hypothetical protein | - | - | 75 | |
| 1739 | conserved hypothetical protein | - | - | 67 | |
| 1753 | conserved hypothetical protein | - | + | 76 | 2 |
| 1854 | conserved hypothetical protein | - | - | 144 | |
| 2435 | conserved hypothetical protein | + | - | 157 | |
| 2680 | conserved hypothetical protein | + | - | 91 | 1 |
| 2714 | conserved hypothetical protein | - | - | 122 | |
| 2908 | conserved hypothetical protein | + | - | 76 | |
| 2912 | conserved hypothetical protein | + | - | 143 | 2 |
| 2976 | conserved hypothetical protein | - | - | 102 | |
| 3221 | conserved hypothetical protein | - | - | 161 | |
| 3509 | conserved hypothetical protein | + | - | 183 | |
| 3599 | ring canal kelch protein | + | - | 133 | |
| 3924 | conserved hypothetical protein | - | - | 107 | 1 |
| 3944 | conserved hypothetical protein | - | + | 167 | 2 |
| 4127 | conserved hypothetical protein | + | - | 99 | |
| 4129 | conserved hypothetical protein | - | + | 105 | 2 |
| 4278 | conserved hypothetical protein | - | - | 81 | |
| 4347 | conserved hypothetical protein | - | + | 109 | |
| 4485 | conserved hypothetical protein | - | - | 86 | 2 |
| 4532 | conserved hypothetical protein | + | - | 83 | |
| 4654 | conserved hypothetical protein | - | + | 124 | 2 |
| 4738 | conserved hypothetical protein | - | - | 130 | 2 |
| 4742 | conserved hypothetical protein | - | - | 89 | 1 |
| 5186 | conserved hypothetical protein | - | - | 55 | 1 |
| 5313 | conserved hypothetical protein | - | + | 157 | |
| 5511 | conserved hypothetical protein | - | - | 161 | |
| 5788 | conserved hypothetical protein | + | + | 58 | |
| 5952 | conserved hypothetical protein | + | - | 133 | |
| 6120 | TolB protein | + | - | 105 | 2 |
| 6395 | conserved hypothetical protein | - | - | 65 | 2 |
| 6409 | conserved hypothetical protein | + | - | 113 | |
| 6416 | conserved hypothetical protein | - | - | 133 | 1 |
| 6417 | conserved hypothetical protein | - | - | 81 | 2 |
| 6530 | conserved hypothetical protein | - | - | 70 | |
| 7091 | conserved hypothetical protein | - | - | 84 | |
| 7120 | conserved hypothetical protein | + | - | 102 | 1 |
| 7292 | conserved hypothetical protein | - | - | 86 | |
| 7538 | conserved hypothetical protein | - | + | 158 | 1 |
| 7619 | conserved hypothetical protein | - | - | 79 | 2 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Conserved Hypothetical Proteins (continued)** | | | | |
| 7789 | TolB protein precursor | - | - | 61 | |
| 7822 | conserved hypothetical protein | - | PA | 96 | 2 |
| 8031 | conserved hypothetical protein | + | - | 122 | 1 |
| 8188 | conserved hypothetical protein | - | - | 58 | |
| 8202 | conserved hypothetical protein | - | + | 86 | |
| 8246 | conserved hypothetical protein | + | - | 184 | 2 |
| 8266 | conserved hypothetical protein | - | + | 60 | 1 |
| 8291 | conserved hypothetical protein | - | - | 53 | |
| 8456 | conserved hypothetical protein | - | - | 73 | |
| 8501 | conserved hypothetical protein | + | - | 158 | |
| 8557 | conserved hypothetical protein | - | - | 130 | |
| 8565 | conserved hypothetical protein | - | + | 136 | 2 |
| 8639 | conserved hypothetical protein | + | - | 168 | 2 |
| 8677 | conserved hypothetical protein | - | - | 152 | |
| 8783 | conserved hypothetical protein | - | - | 102 | |
| 9132 | conserved hypothetical protein | - | + | 138 | 2 |
| 9261 | conserved hypothetical protein | - | + | 139 | |
| 9262 | conserved hypothetical protein | - | + | 113 | 1 |
| 9367 | conserved hypothetical protein | + | - | 64 | |
| 9386 | FixW protein | + | + | 64 | 2 |
| 9438 | conserved hypothetical protein | + | - | 134 | 2 |
| 9546 | conserved hypothetical protein | + | - | 91 | 1 |
| 9606 | conserved hypothetical protein | + | - | 78 | |
| 9849 | conserved hypothetical protein | + | + | 175 | |
| 9992 | conserved hypothetical protein | - | + | 151 | |
| 10028 | conserved hypothetical protein | - | - | 77 | 1 |
| 10061 | conserved hypothetical protein | - | - | 123 | 2 |
| 10078 | conserved hypothetical protein | + | - | 299 | |
| 10088 | conserved hypothetical protein | - | - | 100 | 1 |
| 10103 | conserved hypothetical protein containing kelch-motif | + | + | 102 | |
| 10195 | conserved hypothetical protein | + | + | 90 | 2 |
| 10235 | conserved hypothetical protein | - | - | 56 | |
| 10359 | conserved hypothetical protein | - | - | 129 | |
| 10478 | conserved hypothetical protein | + | - | 140 | 2 |
| 10789 | conserved hypothetical protein | - | + | 96 | |
| 10987 | conserved hypothetical protein | - | - | 113 | |
| 11183 | conserved hypothetical protein | - | - | 181 | 2 |
| 11262 | conserved hypothetical protein | - | - | 101 | 1 |
| 11494 | conserved hypothetical protein | + | - | 60 | |
| 11505 | conserved hypothetical protein | + | + | 68 | |
| 11728 | conserved hypothetical protein | - | + | 177 | 2 |
| 11811 | conserved hypothetical protein | + | PA | 70 | |
| 11998 | conserved hypothetical protein | + | + | 156 | 2 |
| 12056 | conserved hypothetical protein containing TPR domain | + | - | 281 | 2 |
| 12301 | conserved hypothetical protein | + | + | 79 | |
| 12891 | conserved hypothetical protein | - | - | 95 | |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Hypothetical proteins (97 proteins)** | | | | |
| 14 | hypothetical protein | - | - | 212 | |
| 68 | hypothetical protein | - | - | 91 | 1 |
| 404 | hypothetical protein | - | - | 56 | |
| 512 | hypothetical protein | + | - | 219 | 1 |
| 1002 | hypothetical protein | + | - | 275 | |
| 1088 | hypothetical protein | - | - | 150 | 1 |
| 1260 | hypothetical protein | + | - | 108 | 2 |
| 1352 | hypothetical protein | - | - | 132 | 1 |
| 1535 | hypothetical protein | - | + | 152 | 1 |
| 1599 | hypothetical protein | + | - | 95 | 1 |
| 1828 | hypothetical protein | + | - | 210 | |
| 2088 | hypothetical protein | + | - | 78 | |
| 2166 | hypothetical protein | - | - | 128 | |
| 2436 | hypothetical protein | - | + | 213 | 2 |
| 2448 | hypothetical protein | + | - | 65 | |
| 2501 | hypothetical protein | - | - | 133 | 2 |
| 2647 | hypothetical protein | + | + | 113 | 2 |
| 2782 | hypothetical protein | - | - | 79 | |
| 2822 | hypothetical protein | + | + | 65 | 1 |
| 2830 | hypothetical protein | + | + | 174 | 2 |
| 2901 | hypothetical protein | + | - | 68 | 2 |
| 3060 | hypothetical protein | + | - | 83 | |
| 3320 | hypothetical protein | + | - | 259 | |
| 3346 | hypothetical protein | + | - | 101 | |
| 3437 | hypothetical protein | + | - | 88 | |
| 3479 | hypothetical protein | - | - | 76 | |
| 3500 | hypothetical protein | + | - | 71 | |
| 3504 | hypothetical protein | - | - | 184 | |
| 3513 | hypothetical protein | + | - | 102 | |
| 3577 | hypothetical protein | + | - | 81 | |
| 3764 | hypothetical protein | + | - | 67 | |
| 4041 | hypothetical protein | - | - | 146 | 2 |
| 4182 | hypothetical protein | + | + | 98 | 2 |
| 4256 | hypothetical protein | + | - | 91 | 1 |
| 4464 | hypothetical protein | + | - | 119 | 2 |
| 4703 | hypothetical protein | + | - | 213 | |
| 4931 | hypothetical protein | + | + | 76 | |
| 5031 | hypothetical protein | + | - | 86 | 2 |
| 5240 | hypothetical protein | + | + | 58 | 1 |
| 5268 | hypothetical protein | - | - | 217 | |
| 5310 | hypothetical protein | + | - | 202 | |
| 5734 | hypothetical protein | - | - | 68 | |
| 5763 | hypothetical protein | + | - | 143 | |
| 5794 | hypothetical protein | - | - | 110 | |
| 5811 | hypothetical protein | + | - | 183 | 2 |
| 5938 | hypothetical protein | - | - | 95 | |
| 5945 | hypothetical protein | - | - | 53 | |
| 6051 | hypothetical protein | - | - | 90 | |
| 6127 | hypothetical protein | + | - | 105 | 2 |

**Table 2.** continued

| ORF | Putative Function | sp | PHX | Score | * |
|---|---|---|---|---|---|
| | **Hypothetical Proteins (continued)** | | | | |
| 6162 | hypothetical protein | - | + | 66 | |
| 6221 | hypothetical protein | + | - | 161 | 2 |
| 6400 | hypothetical protein | + | - | 103 | 1 |
| 6610 | hypothetical protein | - | + | 370 | |
| 6658 | hypothetical protein | + | + | 160 | 2 |
| 6713 | hypothetical protein | - | - | 164 | 2 |
| 6771 | hypothetical protein | + | - | 220 | |
| 6888 | hypothetical protein | + | - | 130 | 2 |
| 6941 | hypothetical protein | + | + | 52 | |
| 7203 | hypothetical protein | - | + | 157 | 2 |
| 7235 | hypothetical protein | + | + | 194 | |
| 7476 | hypothetical protein | + | - | 95 | |
| 7801 | hypothetical protein | - | - | 70 | |
| 8337 | hypothetical protein | - | - | 144 | |
| 8750 | hypothetical protein | + | - | 112 | |
| 8980 | hypothetical protein | + | - | 106 | |
| 9034 | hypothetical protein | + | - | 176 | |
| 9101 | hypothetical protein | + | - | 108 | 2 |
| 9512 | hypothetical protein | + | - | 82 | 1 |
| 9518 | hypothetical protein | + | - | 65 | |
| 9527 | hypothetical protein | + | + | 215 | 1 |
| 9742 | hypothetical protein | - | - | 86 | |
| 9798 | hypothetical protein | - | + | 110 | |
| 9867 | hypothetical protein | - | - | 107 | |
| 9973 | hypothetical protein | - | - | 118 | |
| 10075 | hypothetical protein | - | - | 103 | |
| 10460 | hypothetical protein | + | - | 89 | 1 |
| 10530 | hypothetical protein | - | - | 128 | |
| 10823 | hypothetical protein | + | - | 143 | 2 |
| 10835 | hypothetical protein | - | - | 275 | 1 |
| 11575 | hypothetical protein | + | - | 92 | 1 |
| 12035 | hypothetical protein | - | - | 120 | 2 |
| 12167 | hypothetical protein | - | + | 124 | 1 |
| 12219 | hypothetical protein | - | - | 91 | |
| 12416 | hypothetical protein | + | - | 162 | 1 |
| 12442 | hypothetical protein | - | - | 165 | |
| 12461 | hypothetical protein | + | + | 154 | 2 |
| 12489 | hypothetical protein | + | - | 113 | |
| 12630 | hypothetical protein | + | - | 90 | 2 |
| 12645 | hypothetical protein | - | + | 114 | |
| 12702 | hypothetical protein | + | - | 62 | |
| 12787 | hypothetical protein | - | - | 168 | 2 |
| 12808 | hypothetical protein | + | - | 95 | |
| 12844 | hypothetical protein | - | - | 134 | 2 |
| 12897 | hypothetical protein | + | - | 62 | |
| 12968 | hypothetical protein | + | - | 119 | 1 |
| 13231 | hypothetical protein | - | - | 105 | |
| 13319 | hypothetical protein | - | + | 102 | |

## References

**Blobel, G.** 2000. Protein targeting (Nobel Lecture). Chembiochem **1**:86-102.

**Bradford, M. M.** 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

**Büttner, K., J. Bernhardt, C. Scharf, R. Schmid, U. Mäder, C. Eymann, H. Antelmann, A. Völker, U. Völker, and M. Hecker.** 2001. A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. Electrophoresis **22**:2908-2935.

**Cordwell, S. J., M. R. Larsen, R. T. Cole, and B. J. Walsh.** 2002. Comparative proteomics of *Staphylococcus aureus* and the response of methicillin-resistent and methicillin-sensitive strains to Triton X-100. Microbiol. **148**:2765-2781.

**Dorothy, N. S., B. H. Littmann, K. Reilly, A. C. Swindell, J. M. Buss, and N. L. Anderson.** 1998. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis **19**:355-363.

**Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. Rd. Science **269**:496-512.

**Fountoulakis, M., B. Takacs and H. Langen.** 1998a. Two-dimensional map of basic proteins of *Haemophilus influenzae*. Electrophoresis **19**:761-766.

**Fountoulakis, M., J. F. Juranville, D. Roder, S. Evers, P. Berndt, and H. Langen.** 1998b. Reference map of the low molecular mass proteins of *Haemophilus influenzae*. Electrophoresis **19**:1819-1827.

**Gade, D., J. Thiermann, D. Markowsky, and R. Rabus.** 2003. Evaluation of two-dimensional difference gel electrophoresis for protein profiling. J. Mol. Microbiol. Biotechnol. **5**:240-251.

**Glöckner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann, and R. Reinhardt.** 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc. Natl. Acad. Sci. USA **100**:8298-8303.

**Gobom, J., M. Mueller, V. Egelhofer, D. Theiss, H. Lehrach, and E. Nordhoff.** 2002. A calibration method that simplifies and improves accurate determination of peptide molecular masses by MALDI-TOF MS. Analytical Chemistry **74**:3915-3923.

**Gobom, J., M. Schuerenberg, M. Mueller, D. Theiss, H. Lehrach, and E. Nordhoff.** 2001. Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. Analytical Chemistry **73**:434-438.

**Görg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss.** 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **21**:1037-1053.

**Hecker, M., S. Engelmann, and S. L. Cordwell.** 2003. Proteomics of *Staphylococcus aureus* – current state and future challenges. J. Chromatogr. B **787**:179-195.

**Henzel, W. J., T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley, and C. Watanabe.** 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc. Natl. Acad. Sci. **90**: 5011-5015.

**Hermann, T., W. Pfefferle, C. Baumann, E. Busker, S. Schaffer, M. Bott, H. Sahm, N. Dusch, J. Kalinowski, A. Pühler, A. K. Bendt, R. Kramer, and A. Burkovski.** 2001. Proteome analysis of *Corynebacterium glutamicum*. Electrophoresis **22**:1712-1723.

**Hesketh, A. R., G. Chandra, A. D. Shaw, J. J. Rowland, D. B. Kell, M. H. Bibb, and K. F. Chater.** 2002. Primary and secondary metabolism, and posttranslational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*. Mol. Microbiol. **46**:917-932.

**Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkl, B. C. Li, and R. Herrmann.** 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucl. Acids Res. **24**:4420-4449.

**Huang, F., I. Parmryd, F. Nilsson, A. L. Persson, H. B. Pakrasi, B. Andersson, and B. Norling.** 2002. Proteomics of *Synechocystis* sp. strain PCC 6803: identification of plasma membrane proteins. Mol. Cell. Proteomics **1**:956-966.

**Karlin, S., J. Mrázek, A.Campell, and D. Kaiser.** 2001. Characterizations of highly expressed genes of four fast-growing bacteria. J. Bacteriol. **183**:5025-5040.

**Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata.** 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. **3**:109-136.

**Karas, M., and F. Hillenkamp.** 1988. Laser desorption ionization of proteins with molecular masses exceeding 10000 Daltons. Anal. Chem. **60**: 2301-2303

**Langen, H., B. Takács, S. Evers, P. Berndt, H.-W. Lahm, B. Wipf, C. Gray, and M. Fountoulakis.** 2000. Two-dimensional map of the proteome of *Haemophilus influenzae*. Electrophoresis **21**:411-429.

**Lindsay, M., R. I. Webb, and J. A. Fuerst.** 1997. Pirellulosomes: a new type of membrane bounded cell compartment in planctomycete bacteria of the genus *Pirellula*. Microbiology **143**:739-748.

**Lindsay, M., R. I. Webb, M. Strous, M. S. M. Jetten, M. K. Butler, R. J. Forde, and J. A. Fuerst.** 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch. Microbiol. **175**:413-429.

**Mann, M., P. Højrup, and P. Roepstorff.** 1993. Use of mass-spectrometric molecular-weight information to identify proteins in sequence databases. Biol. Mass Spectrom. **22**:338-345.

**Meyer, F., A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, and A. Puhler.** 2003. GenDB--an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. **31**:2187-2195.

**Molloy, M. P., B. R. Herbert, M. B.Slade, T. Rabilloud, A. S. Nouwens, K. L. Williams, and A. A. Gooley.** 2000. Proteomic analysis of the *Escherichia coli* outer membrane. Eur. J. Biochem. **276**:2871-2881.

**Nielsen H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10**:1-6.

**Nierman, W. C, T. V. Feldblyum, M. T. Laub, I. T. Paulsen, K. E. Nelson, J. A. Eisen, J. F. Heidelberg, M. R. Alley, N. Ohta, J. R. Maddock, I. Potocka, W. C. Nelson, A. Newton, C. Stephens, N. D. Phadke, B. Ely, R. T. DeBoy, R. J. Dodson, A. S. Durkin, M. L. Gwinn, D. H. Haft, J. F. Kolonay, J. Smit, M. B. Craven, H. Khouri, J. Shetty, K. Berry, T. Utterback, K. Tran, A. Wolf, J. Vamathevan, M. Ermolaeva, O. White, S. L. Salzberg, J. C. Venter, L. Shapiro, and C. M. Fraser.** 2001. Complete genome sequence of *Caulobacter crescentus*. Proc. Natl. Acad. Sci. U.S.A. **98**:4136-4141.

**Nordhoff, E., V. Egelhofer, P. Giavalisco, H. Eickhoff, M. Horn, T. Przewieslik, D. Theiss, U. Schneider, H. Lehrach, and J. Gobom.** 2001. Large-gel two-dimensional electrophoresis-matrix assisted laser desorption/ionization-time of flight-mass spectrometry: An analytical challenge for studying complex protein mixtures. Electrophoresis **22**:2844-2855

**Nouwens, A. S., M. D. Willcox, B. J. Walsh, and S. J. Cordwell.** 2002. Proteomic comparison of the membrane and extracellular proteins from invasive (PAO1) and cytotoxic (6206) strains of *Pseudomonas aeruginosa*. Proteomics **2**:1325-1346.

**Ohlmeier, S., C. Scharf, and M. Hecker.** 2000. Alkaline proteins of *Bacillus subtilis*: First steps towards a two-dimensional alkaline master gel. Electrophoresis **21**:3701-3709.

**Pappin, D. J. C., P. Højrup, and A. J. Bleasby.** 1993. Rapid identification of proteins by peptide-mass fingerprinting. Curr. Biol. **3**:327-332.

**Perkins, D. N., D. J. Pappin, D. M. Creasy, and J. S. Cottrell.** 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20**:3551-3567.

**Rabus, R., D. Gade, R. Helbig, M. Bauer, F. O. Glöckner, M. Kube, H. Schlesner, R. Reinhardt, and R. Amann.** 2002. Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1 by a proteomic approach. Proteomics **2**:649-655.

**Regula, J. T., B. Ueberle, G. Boguth, A. Görg, M. Schnölzer, R. Herrmann, and R. Frank.** 2000. Towards a two-dimensional proteome map of *Mycoplasma pneumoniae*. Electrophoresis **21**:3765-3780.

**Regula, J. T., G. Boguth, A. Görg, F. Mayer, R. Frank, and R. Herrmann.** 2001. Defining the mycoplasma "cytoskeleton": the protein composition of the Triton X-100 insoluble fraction of the bacterium *Mycoplasma pneumoniae* determined by 2-D gel electrophoresis and mass spectrometry. Microbiol. **147**:1045-1057.

**Rice, P., I. Longden, and A. Bleasby.** 2000. "EMBOSS: The European Molecular Biology Open Software Suite". Trends in Genetics **16**:276-277.

**Schlesner, H.** 1994. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from various aquatic habitats using dilute media. System Appl. Microbiol. **17**:135-145.

**Schlesner, H., C. Bartels, B. J. Tindall, D. Gade, R. Rabus, S. Pfeiffer, and P. Hirsch.** 2004. Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA/DNA-hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb nov., and an emended description of the genus *Pirellula*. Int. J. Syst. Evol. Microbiol. (in press).

**Schwartz, R., C. S. Ting, and J. King.** 2001. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. Genome Res. **11**:703-709.

**Tonella, L., B. J. Walsh, J. C. Sanchez, K. Ou, M. R. Wilkins, M. Tyler, S. Frutiger, A. A. Gooley, I. Pescaru, R. D. Appel, J. X. Yan, A. Bairoch, C. Hoogland, F. S. Morch, G. J. Hughes, K. L. Williams, and D. F. Hochstrasser.** 1998. '98 *Escherichia coli* SWISS-2DPAGE database update. Electrophoresis **19**:1960-1971.

**Tonella, L., C. Hoogland, P.-A. Binz, R. D. Appel, D. F. Hochstrasser, and J.-C. Sanchez.** 2001. New perspectives in the *Escherichia coli* proteome investigation. Proteomics **1**:409-423.

**Ueberle, B., R. Frank, and R. Herrmann.** 2002. The proteome of the bacterium *Mycoplasma pneumoniae*: comparing predicted open reading frames to identified gene products. Proteomics **2**:754-764.

**VanBogelen, R. A., E. E. Schiller, J. D. Thomas, and F. C. Neidhardt.** 1999. Diagnosis of cellular status of microbial organisms using proteomics. Electrophoresis **20**:2149-2159.

**Vohradsky, J., I. Janda, B. Grünenfelder, P. Berndt, D. Röder, H. Langen, J. Weise,r and U. Jenal.** 2003. Proteome of *Caulobacter crescentus* cell cycle publicly accessible on SWICZ server. Proteomics **3**:1874-1882.

5

# Proteomic reconstruction of carbohydrate catabolism and regulation in the marine bacterium *Rhodopirellula baltica*

Dörte Gade[1], Johan Gobom[2], Ralf Rabus[1*]

Manuscript in preparation

[1] Max-Planck-Institut für Marine Mikrobiologie, Bremen, Germany
[2] Max-Planck-Institut für Molekulare Genetik, Berlin, Germany
* Corresponding Author

## Abstract

The marine bacterium *Rhodopirellula baltica* is a model organism for aerobic carbohydrate degradation in marine systems, where polysaccharides represent the dominant components of biomass. On the basis of the genome sequence and a 2-D map of soluble proteins, the central catabolic routes of *R. baltica* were reconstructed. All enzymes of glycolysis and TCA-cycle (with the exception of fumarase) were identified. In addition, almost all enzymes of the oxidative branch of the pentose phosphate cycle were detected. This proteomic reconstruction was corroborated by determination of selected enzymatic activities. To study substrate-dependent regulation in *R. baltica*, cells were adapted to growth with eight different carbohydrates and profiled with 2-DE for changes in protein patterns. Relative abundances of regulated proteins were determined using the DIGE technology and protein identification was achieved by PMF. Most of the up-regulated proteins were either dehydrogenases/oxidoreductases or proteins of unknown function which are unique for *R. baltica*. For only some of the regulated proteins, the coding genes are located in a physiologically meaningful genomic context. *E.g.* a ribose-induced alcohol dehydrogenase is encoded within an operon-like structure together with genes coding for a ribose-specific ABC-transporter. However, most of the regulated genes are randomly distributed across the genome. Some of the regulated proteins appear to be posttranslationally modified by phosphorylation.

## Introduction

The carbon cycle in natural environments depends on the remineralization of biomass. In marine systems biomass is mainly produced by phototrophic microorganisms in the upper layer of the water column (annual production of $20 - 30 \times 10^9$ tons of carbon). Dead biomass is settling down to the sediment, often aggregating as particles known as marine snow. During this process, organic carbon is already decomposed by aerobic microorganisms (Rullkötter et al. 1999). Since polysaccharides are the major components of biomass, carbohydrate degradation is particularly relevant for carbon turnover.

The marine bacterium *Rhodopirellula baltica* belongs to the phylum *Planctomycetes*. Growth studies with *R. baltica* and related members of this group revealed their nutritional specialization in aerobic carbohydrate utilization (Schlesner et al. 2004). Moreover, members of the *Planctomycetes* were identified as abundant members of the microbial communities in the marine water column and found to be associated with marine snow (DeLong et al. 1993). Thus *Planctomycetes* are considered as keyplayers in carbohydrate metabolism in marine systems. The recent publication of its genome sequence (Glöckner et al. 2003) and establishment of a 2-D map of soluble proteins (see accompanying manuscript) qualify *R. baltica* as a promising model system for studying the environmentally important catabolism of carbohydrates by this group of bacteria. For this purpose we applied a combined physiological/proteomic approach.

Two-dimensional gel electrophoresis (2-DE) is one of the central and well-established techniques in proteomic research (Görg et al. 2000). It allows to separate complex protein mixtures at highest resolution. Identification of 2-DE separated proteins is most widely performed by determination of peptide masses with matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS) (Karas et al. 1988), and subsequent comparison of the obtained peptide mass fingerprints (PMF) with theoretically calculated peptide masses (*e.g.* Mann et al. 1993).

For economic reasons, microorganisms regulate gene expression in response to changing environmental conditions, *e.g.* changes in substrate availability. Such regulatory processes are well understood in standard bacteria, such as *Escherichia coli* or *Bacillus subtilis*. However, little is known about the adaptive responsiveness of environmental bacteria. Differential gene expression results in changes in the cellular protein composition. Quantitative resolution of changes in protein patterns by 2-DE requires reduction of gel-to-gel variations and sensitive protein dyes with a broad dynamic range. The DIGE (differ-

ence gel electrophoresis) technology, which is based on covalent labelling of proteins with fluorescent dyes prior to electrophoresis and application of an internal standard, allows to determine differences in protein abundance with statistical confidence (Gharbi et al. 2002; Alban et al. 2003; Knowles et al. 2003; Gade et al. 2003).

Our initial proteomic study revealed the specific formation of a dehydrogenase in response to growth with *N*-acetylglucosamine (Rabus et al. 2002), indicating the principle capacity of *R. baltica* to regulate the expression of catabolic genes. The main objective of this study was to extend our knowledge about the regulation of peripheral enzymes which channel various carbohydrates into the central catabolic routes, which we also reconstructed.

## Materials and Methods

### Growth of cells and preparation of protein extracts

Cells of *R. baltica* were grown in mineral medium with either ribose, xylose, glucose, *N*-acetylglucosamine, maltose, lactose, melibiose or raffinose (each 10 mM) as sole source of organic carbon as previously described (Rabus et al. 2002). Cells were adapted to each of these 8 carbohydrates over at least 5 passages. Details on preparation of extracts of soluble proteins are provided in the accompanying paper (Gade et al.).

### Two-dimensional difference gel electrophoresis (2-D DIGE)

2-D DIGE was essentially carried out as recently described (Gade et al. 2003) with the exception that 200 pmol cyanine dye were used for labelling of 50 µg protein sample. Application of three different fluorescent dyes (Cy2, 3 and 5), each displaying discrete absorption and emission maxima, allows co-separation of three different protein samples in a single gel: reference state, test state and internal standard. The protein extract of glucose grown cells served as reference state and was labelled with Cy5. Extracts from cells adapted to growth with either ribose, xylose, *N*-acetylglucosamine, maltose, lactose, melibiose or raffinose represented the 7 test states and were labelled with Cy3 in each case. For each comparative experiment, the internal standard was composed of equal amounts of reference state and one of the 7 test states, and labelled with Cy2. A mixture of equal amounts of each of the three labelling preparations was loaded per gel. To achieve statistical confidence, five parallel gels were run for each of the 7 comparative experiments. To obtained sufficient amounts of protein for MS analysis, separate preparative gels were run with each of the 7 test substrates. These gels were fixed and stained with colloidal Coomassie Brilliant Blue (cCBB; modified from Doherty et al. 1998). Proteins of interest, as defined by the DIGE/DeCyder analysis, were excised from the cCBB stained gels for subsequent MS analysis.

### Phosphoproteome

Soluble proteins of *N*-acetylglucosamine grown cells were separated by 2-DE as described above and phosphorylated proteins were visualized using Pro-Q™ Diamond phosphoprotein gel stain (Molecular Probes, Eugene, OR; USA). Staining was performed by fixing the gels in 50% methanol, 10% TCA (3 h to overnight), washing three times with water (15 min each), incubating with Pro-Q Diamond phosphoprotein gel stain (4 h), and destaining

with successive washes with a 50 mM sodium acetate solution (pH 4.0), containing 15% (v/v) 1,2-propanediol. Following image acquisition, gels were stained with SYPRO® Ruby (Rabus et al. 2002) allowing comparison of phosphoproteins and total protein profiles.

**Image acquisition and analysis**

DIGE gels were scanned directly between the glass plates at a resolution of 100 µm using the Typhoon™ 9400 scanner (Amersham Biosciences). The Cy2, Cy3 and Cy5 images of each gel were generated by subsequent scanning with the appropriate combination of laser and emission filter, as detailed earlier (Gade et al. 2003). Image and statistical analyses were performed using the DeCyder™ software package (version 5.0; Amersham Biosciences).

Images of gels stained with Pro-Q Diamond phosphoprotein gel stain were collected with the Typhoon 9400 scanner using a green laser (532 nm) and an emission filter of 580 BP 30. SYPRO Ruby stained gels were digitalized with the Typhoon 9400 scanner using a blue laser (457 nm) and an emission filter of 610 BP 30.

Images of colloidal Coomassie stained gels were collected with the Image Scanner (Amersham Biosciences).

**Gel sample excision und processing**

Among the up-regulated proteins determined by 2-D DIGE and DeCyder analysis, the reasonably abundant ones were selected for identification. They were automatically excised from the cCBB stained 2-DE gels, transferred to 96-well microtiter plates and stored at -20°C until further analysis by mass-spectrometry.

**MALDI-TOF-MS**

Tryptic digest and MALDI-TOF-MS analysis of protein samples, data processing and protein identification via PMF were carried out as described in the accompanying manuscript (Gade et al.).

**Enzyme studies**

Cell growth and harvesting was performed as described above. Cell pellets (100 – 200 mg wet weight) were resuspended in 1 mL Tris/HCl (0.1 M, pH 7.5). Cells were disrupted using the One Shot® System (Constant Systems Ltd., England) applying 2700 bar. Intact cells, cell debris and membranes were removed by centrifugation (17,900 *g*, 15 min, 4 °C).

The protein content of the soluble fraction was determined according the method described by Bradford (1976).

All enzyme assays were carried out aerobically at 30 °C in glass cuvettes with 1 mL assay mixtures. Enzyme activities were recorded by measuring the reduction of $NAD^+/NADP^+$ or the oxidation of NADH/NADPH, respectively. The pH optimum for each enzyme in the assay was determined by measuring its activity in a pH range from 6.5 to 8.5 in intervals of 0.5, using correspondingly adjusted 0.5 M Tris/HCl buffers. All enzyme activities were measured in triplicates for each cell extract of the 8 growth conditions, except for β-galactosidase which was only measured in extracts from glucose and lactose grown cells. One unit (U) of enzyme activity is defined as 1 μmol substrate consumed or product formed per min.

**6-Phosphofructokinase (EC 2.7.1.90)**

6-Phosphofructokinase was determined by measuring the formation of fructose-1,6-bisphosphate (FBP) from fructose-6-phosphate. The reaction was coupled to the oxidation of NADH via FBP aldolase, triosephosphate isomerase (TIM), and glycerol-3-phosphate dehydrogenase. The assay mixture contained 100 mM Tris/HCl (pH 7.5), 30 mM $MgCl_2$, 10 mM fructose-6-phosphate, 2 mM pyrophosphate, 0.3 mM NADH, 0.54 U/mL FBP aldolase, 20 U/mL triosephosphate isomerase, 0.34 U/mL glycerol-3-phosphate dehydrogenase and 0.1 mg protein.

**Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)**

Glyceraldehyde-3-phosphate dehydrogenase was assayed in the direction of 1,3-bisphosphoglycerate formation recording the reduction of $NAD^+$. The assay mixture contained 100 mM Tris/HCl (pH 8.0), 50 mM $MgCl_2$, 20 mM potassium phosphate, 2 mM fructose-1,6-bisphosphate (FBP), 2.5 mM $NAD^+$, 0.9 U/mL FBP aldolase and 0.1 mg protein.

**Enolase (EC 4.2.1.11)**

Enolase was tested by coupling the formation of phosphoenolpyruvate to the oxidation of NADH via pyruvate kinase and lactate dehydrogenase. The assay mixture contained 83 mM Tris/HCl (pH 8.0), 3.3 mM $MgSO_4$, 1.1 mM ATP, 0.9 mM glycerate-2-phosphate, 0.2 mM $NADH^+$, 18.5 U/mL lactate dehydrogenase, 2.7 U/mL pyruvate kinase and 0.1 mg protein.

In addition to the proteomic reconstruction, the activities of selected enzymes were determined: 6-phosphofructokinase, glyceraldehyde-3-phosphate dehydrogenase and enolase of glycolysis; transaldolase of the oxidative pentose phosphate cycle; isocitrate dehydrogenase and malate dehydrogenase of TCA cycle. The individual activities of the selected enzymes were rather similar, regardless of which of the eight carbohydrates was used for growth. Therefore only the enzymatic activities determined for ribose grown cells are shown in Figure 1.

Activity of phosphofructokinase (PFK) could only be measured with pyrophosphate, but not with ATP or ADP as phosphoryl donor. This finding agreed with the bioinformatical prediction of a pyrophosphate-dependent enzyme ($PP_i$-PFK). In fact, the genome of *R. baltica* harbours two $PP_i$-PFKs, both of which were detected to be expressed constitutively in our proteomic study (data only shown for RB10591, Figure 2).

Enolase, glyceraldehyde-3-phosphate dehydrogenase, malate dehydrogenase and transaldolase are among the most abundant proteins detected in the 2-DE gels. Moreover, they are present under all investigated growth conditions (Table 1). This finding is in agreement with a normalized codon usage analysis according to Karlin and Mrázek (2000), which predicts *eno*, *gap*, *mdh* and *tal* to be highly expressed in *R. baltica*. In contrast, isocitrate dehydrogenase, which is also constitutively formed, appears as a less abundant spot on the DIGE gels.

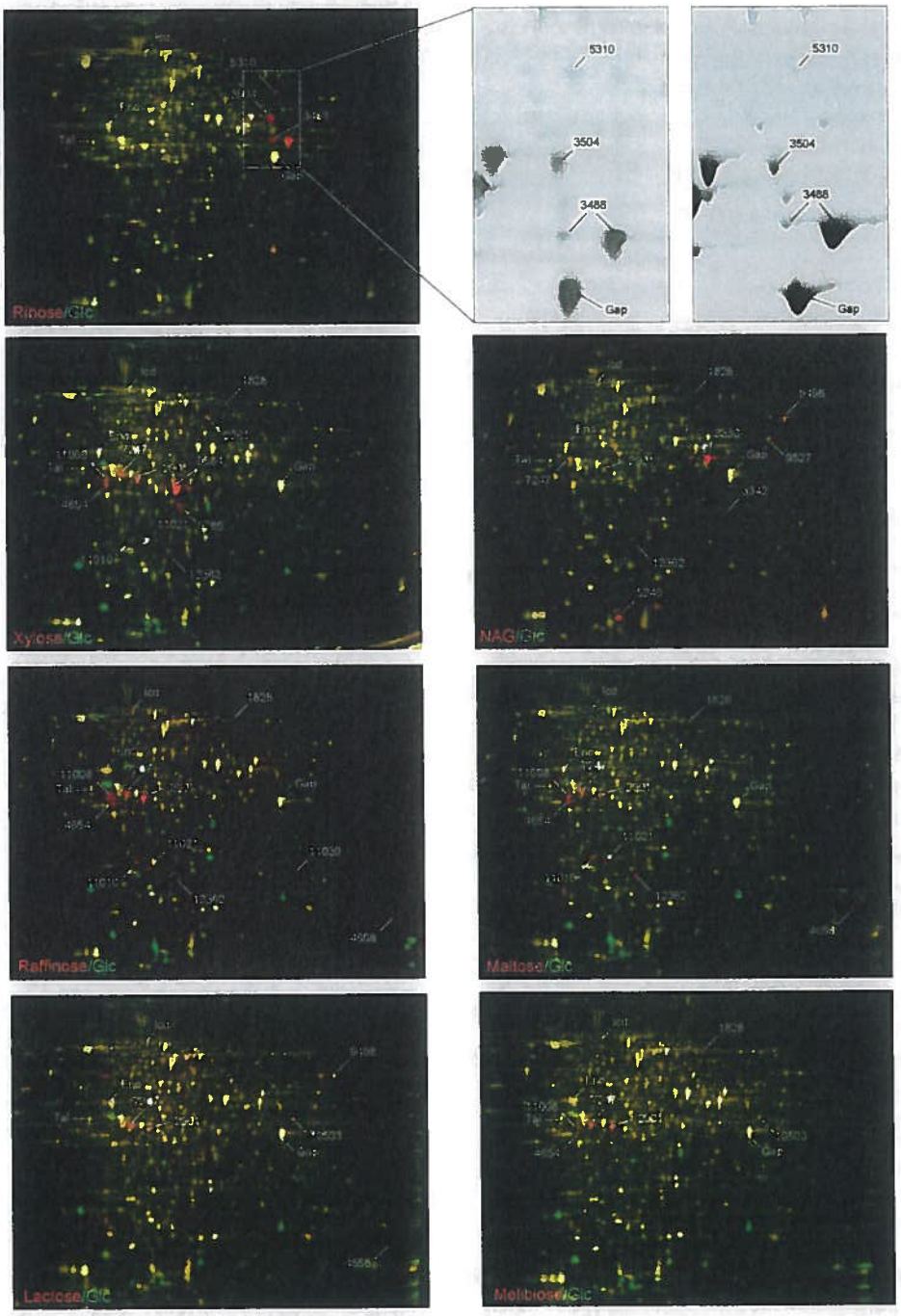



**Figure 2.** Overlay images of 2-D DIGE gels. Protein spots appearing in red are upregulated when cells were grown with raffinose, maltose, lactose, melibiose, xylose, *N*-acetylglucosamine and ribose, respectively. Whereas spots appearing in green are more abundant when cells were grown with the reference carbohydrate glucose. Yellow spots do not differ in relative abundance. Identified proteins which differ in abundance are annotated with their ORF numbers. Four proteins of the central metabolic pathways (Gap, Eno, Tal, Icd) that do not differ in abundance under the compared growth conditions are also marked. The selected two gel portions in the top row correspond to the boxed area in the ribose gel. The left section shows the ribose (Cy3) panel and the right section a colloidal Coomassie stained ribose gel, from which the spots were excised.

**Table 1.** Abundance of proteins identified as being highly upregulated in cells grown with one of the above mentioned carbohydrates relative to glucose. Coding genes and putative functions based on BLASTP analysis are displayed.

| | | | Average ratio[1] | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| RB | putative function | s p[2] | Raffinose | Melibiose | Lactose | Maltose | NAG[3] | Xylose | Ribose |
| 1828 | conserved hypothetical protein | + | 2.92 | 2.71 | | 3.72 | 1.9 | 3.92 | |
| 2261 | carboxypeptidase | - | | | | | | 2.00 | |
| 2901 | hypothetical protein[4] | + | 8.27 | 4.10 | 2.66 | 7.46 | 1.9 | 4.58 | |
| 3330 | dehydrogenase[5] | - | | | | | 43.9 | | |
| 3330 | | | | | | | 15.7 | | |
| 3342 | conserved hypothetical protein | - | | | | | 32.4 | | |
| 3488 | Zn-containing ADH (sorbitol DH) | - | | | | | | | 43.39 |
| 3488 | | - | | | | | | | 12.47 |
| 3504 | conserved hypothetical protein[6] | - | | | | | | | 8.08 |
| 4654 | sugar phosphate isomerase/epimerase | - | 8.95 | 2.45 | | 8.10 | | 4.27 | |
| 4658 | conserved hypothetical protein | - | 4.40 | | 9.80 | 6.41 | | | |
| 5240 | hypothetical protein | + | | | | | 4.5 | | |
| 5310 | conserved hypothetical protein[6] | + | | | | | | | 23.63 |
| 7247 | glutamine synthetase II | - | | | | | 1.8 | | |
| 7247 | glutamine synthetase II | - | 3.02 | 2.79 | 5.07 | 2.84 | | 2.85 | |
| 9498 | arylsulfatase | + | | | 2.05 | | 1.8 | | |
| 9527 | hypothetical protein | + | | | | | 1.7 | | |
| 9584 | Zn-dependent quinone oxidoreductase | - | | | | | | 19.78 | |
| 9584 | | - | | | | | | 4.23 | |
| 9586 | oxidoreductase | - | | | | | | 13.21 | |
| 10503 | NADH-dependent oxidoreductase[5] | - | | 2.09 | 2.31 | | | | |
| 11008 | hypothetical protein[4] | + | 8.95 | 2.45 | | 8.10 | | 4.27 | |
| 11010 | conserved hypothetical | + | 3.44 | | | 4.84 | | 2.07 | |
| 11021 | ABC transporter (ATP-binding protein) | - | 3.94 | | | 4.26 | | 2.65 | |
| 11030 | ABC transporter (ATP-binding protein) | - | 4.91 | | | | | | |
| 12362 | peroxiredoxin | + | 3.86 | | | 5.79 | 2.9 | 2.87 | |
| 10591 | PFK | - | -1.23 | -1.00 | 1.16 | 1.01 | -1.31 | -1.27 | -1.17 |
| 2627 | GAP DH | - | -1.16 | 1.06 | 1.36 | 1.21 | 1.15 | -1.13 | -1.32 |
| 12381 | enolase | - | -1.14 | 1.15 | 1.39 | 1.19 | -1.33 | -1.31 | -1.27 |
| 1593 | isocitrat DH | - | -1.05 | -1.10 | -1.30 | -1.15 | -1.16 | 1.12 | -1.22 |
| 3193 | transaldolase | - | -1.18 | -1.12 | 1.10 | -1.05 | 1.07 | 1.05 | -1.07 |

[1] On the basis of 5 parallel gels changes in protein abundance were regarded as significant if the average ratio was higher than 2 and had a t-test value of less than $10^{-5}$.
[2] sp: signal peptide prediction
[3] Data were taken from Gade et al. 2003.
[4] RB2901 and RB11008 are paralogs.
[5] RB3330 and RB10503 are paralogs.
[6] RB3504 and RB5310 are paralogs.

**Carbohydrate-specific regulation of protein composition**

The principle capacity of *R. baltica* to regulate expression of catabolic genes in response to substrate adaptation (Rabus et al. 2002) was investigated in the present study on a broader scale. Using glucose-grown cells as reference state, catabolic regulation was investigated with cells adapted to growth with ribose, *N*-acetylglucosamine, xylose, maltose, lactose melibiose, and raffinose, respectively. Quantification of changes in protein abundance was achieved by application of the DIGE technology (Figure 2 and Table 1). The genomic context of regulated genes was analyzed (Figure 4) on the basis of the recently published genome sequence of *R. baltica* (Glöckner et al. 2003). The results from the adaptation experiments are presented and discussed in the following paragraphs.

**Ribose**

Among the mapped 1962 detected protein spots in the 5 parallel DIGE gels, 13 proteins were specifically up-regulated during growth with ribose. Four of them could be detected on the corresponding cCBB stained gels and identified by MALDI-TOF-MS (Table 1).

Two of the identified spots could be correlated to one and the same gene (*RB3488*). As the molecular weight of both protein spots is identical, the implicated posttranslational modification only effects the p*I* (Figure 2A). This protein is highly similar to a Zn-containing alcohol dehydrogenase. The other two identified up-regulated proteins are encoded by *RB3504* and *RB5310*, and represent paralogous, conserved hypothetical proteins. The genomic context of *RB3488* and *RB3504* corroborates an involvement of their products in ribose catabolism, as they frame the genes for the ribose ABC-transporter and ribokinase (Figure 4 A). The order of these genes is only slightly different from those in *E. coli* (Lin 1996). Although the theoretical $M_r$ and p*I*s suggest that transporter subunits or ribokinase could in principle be resolved under the applied electrophoretic conditions, they could not be identified so far. The genomic context of *RB5310* gives no hint to a possible function of its product in ribose catabolism.

***N*-acetylglucosamine**

The DIGE results for quantitative changes in protein abundances of this experiment (Table 1) were taken from a previous study, where 24 of the 1688 detected proteins spots were found to be up-regulated during growth with *N*-acetylglucosamine (Gade et al. 2003). In the present study, ten of these protein spots could be detected on corresponding cCBB stained gels and identified by MS analysis.

A protein that is about 40-fold up-regulated is encoded by *RB3330*. This protein has already been in the focus of a previous study (Rabus et al. 2002). The gene product shows high similarity to a dehydrogenase and it forms two spots on the gel which slightly differ in their p*I*s. Another *N*-acetylglucosamine specific protein was identified as gene product of *RB3342* and to be of unknown function. *RB3330* and *RB3342* are in close proximity to *nagB* (glucosamine-6-phosphate deaminase, Figure 4 B). This finding points to an involvement of these genes in *N*-acetylglucosamine metabolism. So far neither NagA nor NagB could be identified in our proteomic approach.

The genome of *R. baltica* revealed the existence of all genes reported for the *nag* operon in *E. coli* (Plumridge 1989), whereas a PTS-transport system for *N*-acetylglucosamine could not be identified. Unlike in *E. coli* not all *nag* genes cluster in an operon-like structure but are distributed all over the genome, indicating a different mode of regulation.
Another up-regulated protein is encoded by *RB7247*. This protein shows very high similarity to a glutamine synthetase II. Interestingly, the product of this gene forms three different spots on the gel. As mentioned above one of them is up-regulated when the cells were grown with *N*-acetylglucosamine. The second protein spot is up-regulated under all other tested growth conditions (ribose being the only exception). The third protein spot of *RB7247* does not seem to change its abundance under all investigated growth conditions (Figure 3). As two of the three RB7247 protein spots differ only in p*I* staining of the gels with a phosphoprotein gel stain allowed to demonstrate that phosphorylation may be responsible for this shift. To our surprise the two gene products of *RB7247* that differ only with respect to their p*I*s seem to be phosphorylated whereas the third RB7247 protein spot is not phosphorylated (Figure 3 A and C). in contrast, the two protein spots encoded by *RB3330* apparently did not result from phosphorylation (data not shown).

Two specifically up-regulated proteins of unknown function are encoded by *RB5240* and *RB9527*. The genomic contexts of these genes however do not reveal any function of these proteins in carbohydrate catabolism.

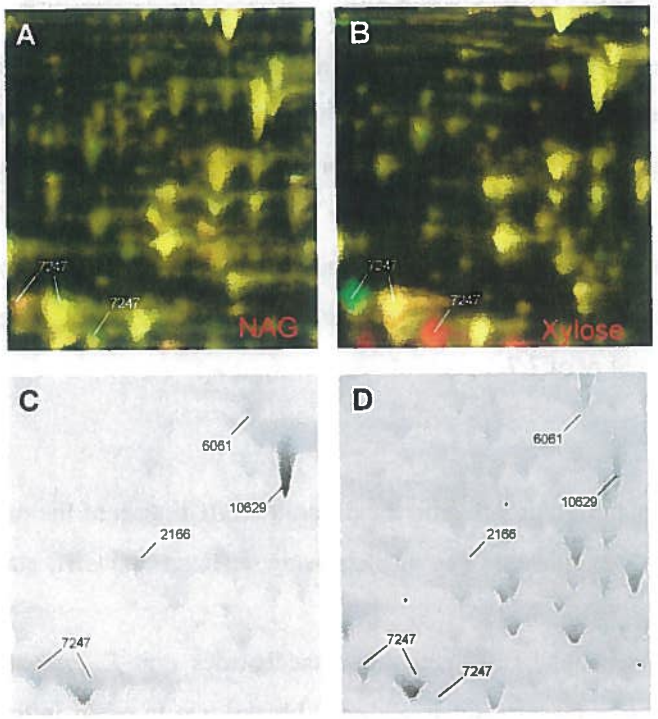



**Figure 3.** Top row. Partial view of the *R. baltica* soluble proteome during growth with *N*-actelyglucosamine (A) and Xylose (B) as revealed by 2-D DIGE. RB7274 (Glutamine synthetase II) appears as three differently regulated spots on the gel. Bottom row. Partial view of the *R. baltica* soluble proteome during growth with *N*-actelyglucosamine as revealed by 2-DE. (C) Phosphorylated proteins were visualized with the Pro-Q Diamond phosphoprotein gel stain. (D) Identical section of the same gel after staining with SYPRO Ruby.

## Xylose

The DIGE master gel of the xylose experiment contained 1950 protein spots. 38 of these proteins were found to be more abundant during growth on xylose and twelve of these proteins were identified by PMF (Table 1).

The genome of *R. baltica* revealed the existence of a low affinity xylose transporter (*xylE*; *RB10309*) (Davis et al. 1987) and all enzymes needed to channel the sugar into the central metabolic pathways. Unlike in *E. coli* or *B. subtilis* these genes are not organized in an operon. While no xylose-specific ABC-transporter was predicted from the genome the proteomic data showed the up-regulation of an ATP-binding protein belonging to an ABC transporter (RB11021). This protein is also up-regulated in cells grown with either raffinose or maltose which may be indicative for an unspecific sugar transporter.

The protein showing a specific up-regulation of about 20-fold is encoded by *RB9584*. This protein is most similar to a Zn-dependent quinone oxidoreductase and forms two spots on the gel which only differ in their p*I*s. The protein with the second highest average

ratio (13) is related to RB9586, which is similar to another oxidoreductase. *RB9584* and *RB9686* are transcribed in the same direction and are separated by *RB9585* coding for a putative monooxygenase. Although the functions of these enzymes during xylose degradation are not clear, their high abundances indicate important roles in xylose catabolism of *R. baltica.*

The xylose isomerase (XylA) has been found to be constitutively expressed under all investigated conditions, whereas the xylulose kinase (XylB) could not be detected on the 4-7 gels due to its predicted p*I* of 7.7.

**Maltose**

From the 1950 spots that were detected on the 5 parallel DIGE gels of the maltose experiment15 proteins were up-regulated. Nine of them were detected on cCBB stained gels and identified by MS analysis (Table 1).

Many bacteria have uptake systems for disaccharides *e.g. E. coli* and *B. subtilis* (Tangney et al. 1992; Lin 1996) and intracellular breakdown is often initiated by phosphorylytic cleavage. In *R. baltica* the high abundance of an ABC-transporter subunit (RB11021) suggests that maltose may be taken up via this system. However, as *R. baltica* does not have a maltose phosphorylase it seems to hydrolyse the disaccharide less economically with an amylomaltase encoded by *RB4161*, which however was not detected on the gels so far.

None of the nine identified proteins are exclusively up-regulated in maltose grown cells. Interestingly, there are some similarities in the patterns of up-regulated proteins between maltose, melibiose and raffinose adapted cells.

Besides a putative sugar phosphate isomerase/epimerase encoded by *RB4654* the other identified proteins could not be affiliated with the degradation of this disaccharide.

**Lactose**

Among the mapped 2111 detected protein spots in the DIGE gels of the lactose experiment 14 proteins were up-regulated. Five of them were detected on cCBB stained gels and identified by MS analysis (Table 1).

The typical *lac* operon observed in *e.g. E. coli* (Lin 1996) can not be found in the genome of *R. baltica.* Nevertheless for the degradation of lactose a β-galactosidase (RB3405) was predicted from the genome sequence. As this protein shows only weak similarity to a cryptic gene in *E. coli* coding for the α-subunit of a β-galactosidase one has to be critical

about this prediction. The specific activity of this enzyme could not be measured under conventional conditions, indicating that the enzyme may represent a new type of β-galactosidase.

Interestingly, the genes coding for enzymes needed for the first steps of galactose degradation (*galK* and *galT*) are absent from this genome. This finding is even more surprising as galactose is an excellent carbon source for this bacterium (Schlesner et al. 2004). Thus the degradation of this compound might involve alternative reactions. Since degradation of lactose, melibiose and raffinose share the liberation of galactose, commonly up-regulated hypothetical proteins might be involved in this process. These among them that are also abundant during growth with other carbohydrates, *e.g.* maltose or xlose, probably serve other functions.

**Melibiose**

In the five parallel DIGE gels of the melibiose experiment 1809 spots were detected. Ten of which were up-regulated during growth with melibiose. Six of these proteins could be detected on cCBB stained gels and identified by MALDI-TOF-MS (Table 1).

Although no genes for melibiose transport and degradation were predicted in the genome, this carbohyrate is an excellent growth substrate for *R. baltica*. As in case of maltose or lactose adapted cells, proteins exclusively up-regulated during growth with melibiose could not be detected. Three up-regulated proteins are of unknown function (RB1828, RB2901 and RB11008). The genetic context of there coding genes gives no hint to their possible functions in carbohydrate catabolism. For two other proteins an involvement in the degradation of carbohydrates appears likely, a putative sugar phosphate isomerase (RB4654) and a putative NADH dependent oxidoreductase (RB10503). RB10503 is a paralog of RB3330, the dehydrogenase that is specifically induced in *N*-acetylglucosamine adapted cells.

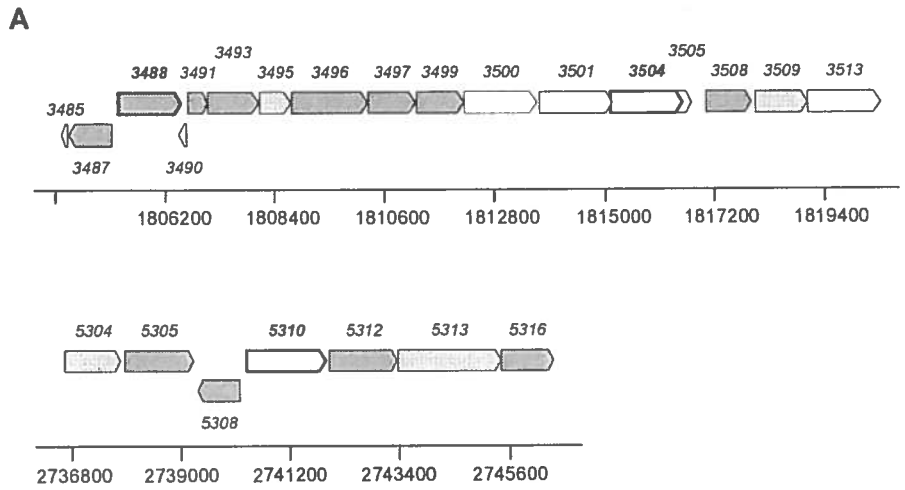
**Raffinose**

From the 1925 detected proteins in the DIGE master gel of the Raffinose experiment 16 proteins were up-regulated. Ten of them could be detected on cCBB stained gels and identified by PMF (Table1).

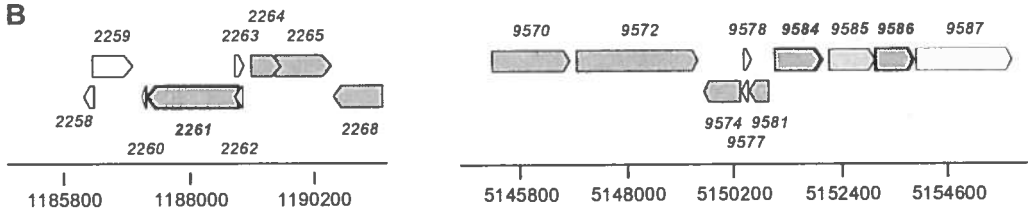
As for melibiose, all genes generally known to be needed for the initial steps of raffinose degradation are not predicted in the genome of *R. baltica*. Despite the presence of an α-glucosidase coding gene (*RB10507*) in *R. baltica* the first step has to be carried out by other enzymes following different reactions. Two subunits of ABC-transporters (RB11021 and RB11030) are highly abundant. While RB11030 seems to be specific for melibiose RB11021 is also present in cells grown with maltose and xylose. Interestingly, these two proteins share 53% sequence similarity.

The following general observations were made during the present study. In contast to cells grown with monosaccharides, cells grown with di- and trisaccharide exhibited less pronounced differential proteoms according to 2-D DIGE analysis. With one exception (RB11030 in raffinose grown cells) no proteins were detected which were highly up-regulated exclusively during growth with one particular di- or trisaccharide (Table 1; Figure 4). In the case of these substrates the up-regulation of certain proteins is induced by two or more carbohydrates used for growth (Figure 4E-H). One hypothetical protein (RB2901) is up-regulated under all tested growth conditions with the only exception being ribose. The gene product of RB2901 has a predicted signal peptide and is a paralog of RB11008, which is up-regulated under four of the six above mentioned growth conditions (Table 1). However ribose and glucose (growth substrate of the reference state) are the only substrates in this study that can directly be degraded via the central metabolic routes.

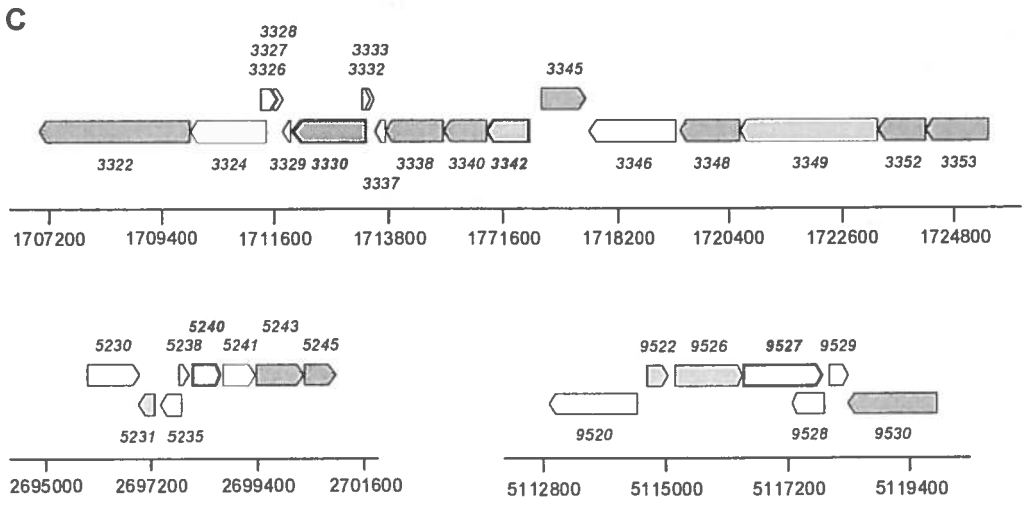
**Figure 4.** Genomic context of carbohydrate induced genes in *R. baltica*. Color coding: *thick lines*, up-regulated proteins; *grey*, ORFs of proteins with assigned putative functions; *white*, ORFs of hypothetical proteins; *dotted*, ORFs of conserved hypothetical proteins.
Putative functional assignments for predicted ORFs are as followed:



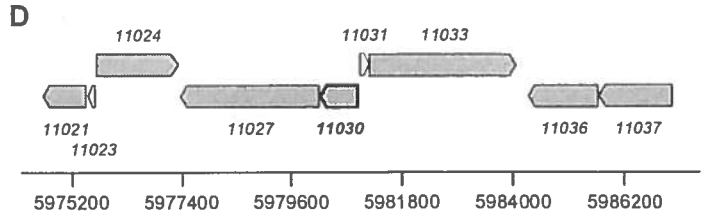

**(A) Genes induced by ribose. 3487**, melibiose operon regulatory protein, *melR*; **3488**, Zn-containing alcohol dehydrogenase; **3491**, high affinity ribose transport protein, *rbsD*; **3493**, ribose ABC transport system periplasmic ribose-binding protein, *rbsB*; **3496**, ribose ABC transport system ATP-binding protein, *rbsA*; **3497**, ribose ABC transport system permease protein, *rbsC*; **3499**, ribokinase, *rbsK*; **3508**, UDP-glucose-4-epimerase, *galE2*; **5305**, arylsulfatase A; **5308**, melibiose operon regulatory protein, *melR*; **5312**, pectate lyase, *pel*; **5316**, pectate lyase, *pel*.



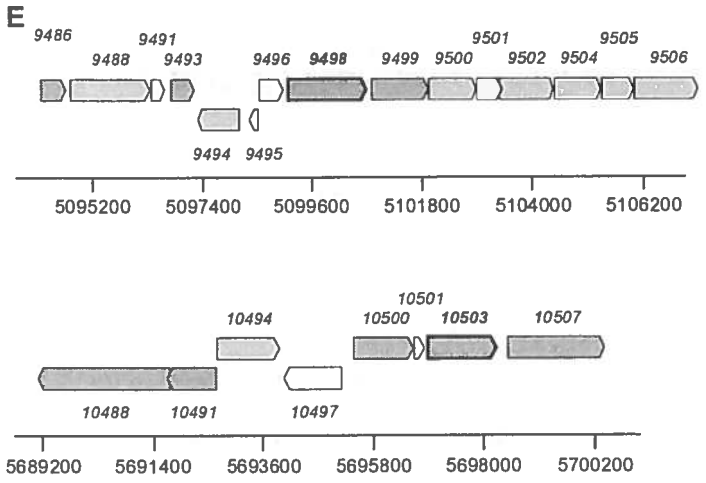

**(B) Genes induced by xylose. 2261**, carboxypeptidase-related protein; **2264**, large-conductance mechanosensitive channel; **2265**, lipase/esterase; **2268**, peptide methione sulfoxide reductase, *msrB*; **9570**, hydrogenase-4 transcriptional activator, *hyfR*; **9572**, L-sorbosone dehydrogenase; **9574**, 3-oxoacyl reductase, *fabG*; **9581**, cinnamoyl ester hydrolase; **9584**, Zn-containing oxidoreductase; **9586**, oxidoreductase.



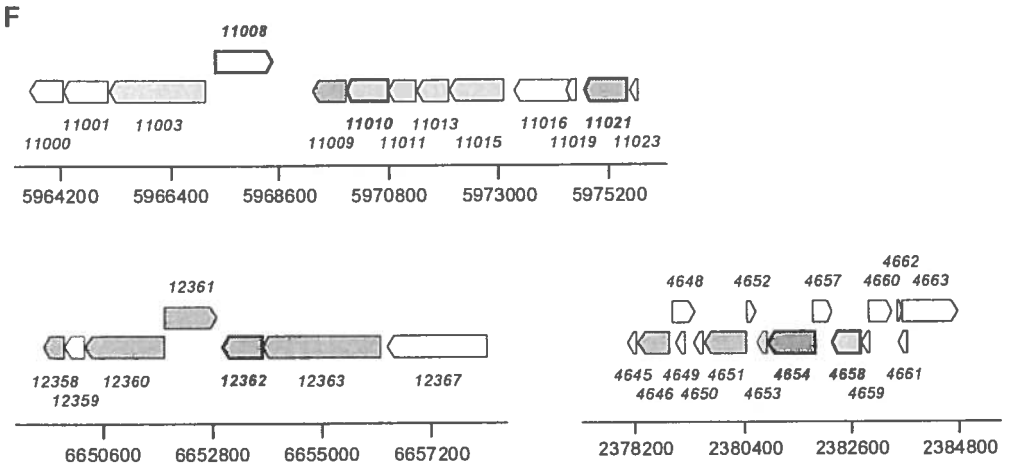

**(C) Genes induced by *N*-acetylglucosamine. 3322**, cytochrome *c*; **3330**, dehydrogenase; **3338**, myo-inositol 2-dehydrogenase, *idh*; **3340**, glucosamine-6-phosphate isomerase 2, *nagB*; **3345**, activator protein MtlR; **3348**, *N*-acetylglucosamine-2-epimerase; **3352**, *N*-acetylneuraminate lyase, *nanA*; **3353**, sialidase precursor; **5243**, endo-1,4-beta-xylanase B, *xynB*; **5245**, hemolysin III; **9530**, arylsulfatase.



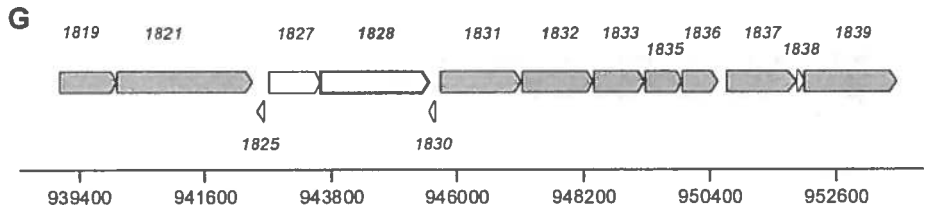

**(D) Genes induced by raffinose. 11021**, ABC transport system ATP binding protein; **11024**, enterochelin esterase, *fes*; **11027**, ABC transport system integral membrane protein; **11030**, ABC transport system ATP binding protein; **11033**, ABC transport system integral membrane protein; **11036**, nitrogen regulation protein NR(I), *ntrC*; **11037**, nitrogen regulation protein, *ntrB*.



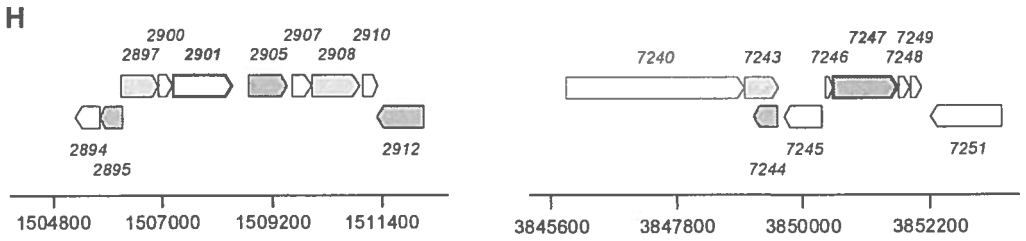

**(E) Genes induced by two different carbohydrates. 9486**, hypothetical 161 kDa transcriptional regulator; **9493**, acetyltransferase; **9498**, arylsulfatase; **9499**, MoxR-related protein, *moxR*; **10488**, sensory transduction histidine kinase; **10491**, two-component system regulatory protein, *cheY*; **10500**, phosphoglycerate kinase, *pgk*; **10503**, lipopolysaccharide biosynthesis protein BplA; **10507**, alpha-glucosidase.




**(F) Genes induced by three or four different carbohydrates. 4646**, cytochrome *c* oxidase, subunit III, *cox3*; **4651**, thiamine biosynthesis protein, *thiG*; **4654**, sugar phosphate isomerase/epimerase; **11009**, iron-uptake factor, *piuB*; **11021**, ABC transport system ATP binding protein; **12358**, flagellar motor switch protein (fragment), *fliN*; **12360**, levanase precursor; **12361**, sugar kinase; **12362**, peroxiredoxin; **12363**, ATP-dependent DNA helicase, *pcrA*.




**(G) Genes induced by five different carbohydrates. 1819**, UDP-glucose-4-epimerase, *galE*; **1821**, serine/threonine-protein kinase, *pknB*; **1831**, $Na^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrA*; **1832**, $Na^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrB*; **1833**, $Na^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrC*; **1835**, $Na^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrD*; **1836**, $Na^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrE*; **1837**, $Na^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrF*; **1839**, thiamine biosynthesis lipoprotein, *apbE*.




**(H) Genes induced by six of the seven different carbohydrates. 2895**, biopolymer transport protein, *exbD*; **2905**, signal peptidase-like protein; **2912**, lipoprotein; **7244**, peptidyl-prolyl isomerase, *ppiB*; **7247**, glutamine synthetase II, *glnII*.

## Conclusions

With respect to the genomic context, two types of regulated genes could be distinguished: firstly, genes, which are located in a physiologically meaningful genomic context, *e.g.* ribose or *N*-acetylglucosamine specific genes and secondly, genes that are randomly distributed across the genome, *e.g.* xylose specific genes. The latter is in agreement with the previous *in silico* observation that functionally related genes are often not clustered in operon-like structures in the 7.145 Mb genome of *R. baltica* (Glöckner et al. 2003). This contrasts the genetic organization known from standard bacteria such as *E. coli* or *B. subtilis*, both having considerably smaller genome sizes than *R. baltica*. Apparently, the number of

genes organized in clusters decreases with increasing genome size (Ermolaeva et al. 2001). The proteomic approach used in this study revealed that in *R. baltica* proteins, the genes of which do not cluster, are functionally related and co-ordinately regulated. The regulatory mechanisms behind these findings are yet to be explored. In addition, *R. baltica* was found to grow with melibiose and raffinose although in both cases the known key enzymes for the initial degradation steps are not encoded in the genome. Moreover several proteins are up-regulated independently of the carbohydrate used for growth. Several of the identified up-regulated genes encode unknown proteins, suggesting that *R. baltica* uses thus far unknown routes for peripheral carbohydrate catabolism.

## Acknowledgements

We would like to thank Torben Stührmann for carrying out the enzyme studies. We are indebted to Friedrich Widdel for continuous support of proteomic work at the MPI in Bremen. This study was supported by the Max Planck Society.

## References

**Alban, A., S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie.** 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics **3**:36-44.

**Bradford, M. M.** 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248-254.

**Büttner, K., J. Bernhardt, C. Scharf, R. Schmid, U. Mäder, C. Eymann, H. Antelmann, A. Völker, U. Völker, and M. Hecker.** 2001. A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. Electrophoresis **22**:2908-2935.

**Davis, E. O., and P. J. F. Henderson.** 1987. The cloning and DNA sequence of the gene *xylE* for xylose-proton symport in *Escherichia coli* K12. J. Biol. Chem. **262**:13928-13932.

**DeLong, E. F., D. G. Franks, and A. L. Alldredge.** 1993. Phylogenetic diversity of aggregate attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. **38**:924-934.

**Doherty, N. S., B. H. Littman, K. Reilly, A. C. Swindell, J. M. Buss, and N. L. Anderson.** 1998. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis **19**:355-363.

**Ermolaeva, M. D., O. White, and S. L. Salzberg.** 2001. Prediction of operons in microbial genomes. Nucleic Acids Res. **29**:1216-1221.

**Gade, D., J. Thiermann, D. Markowsky, and R. Rabus.** 2003. Evaluation of two-dimensional difference gel electrophoresis for protein profiling. J. Mol. Microbiol. Biotechnol. **5**:240-251.

**Gharbi S., P. Gaffney, A. Yang, M. J. Zvelebil, R. Cramer, M. D. Waterfield, and J. F. Timms.** 2002. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol. Cel. Proteomics **1**:91-98.

**Glöckner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann, and R. Reinhardt.** 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc. Natl. Acad. Sci. USA **100**:8298-8303.

**Görg, A., C. Obermaier, G. Boguth, A Harder. B. Scheibe, R. Wildgruber, and W. Weiss.** 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **21**:1037-1053.

**Karas, M., and F. Hillenkamp.** 1988. Laser desorption of proteins with molecular masses exceeding 10000 daltons. Anal. Chem. **60**:2299-2301.

**Karlin, S., and J. Mrázek.** 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol. **182**:5238-5250.

**Knowles, M. R., S. Cervino, H. A. Skynner, S. P. Hunt, C. de Felipe, K. Salim, G. Meneses-Lorente, G. McAllister, and P. Guest.** 2003. Multiplex proteomic analysis by two-dimesional differential in-gel electrophoresis. Proteomics **3**:1162-1171.

**Liao, X., T. Ying, H. Wang, J. Wang, Z. Shi, E. Feng, K. Wei, Y. Wang, X. Zhang, L. Huang, G. Su, and P. Huang.** 2003. A two-dimensional proteome map of Shigella flexneri. Electrophoresis **24**:2864-2882.

**Lin, E. C. C. 1996. In: Neidhardt F. C.** (ed.). *Escherichia coli* and *Salmonella*, American Society for Microbiology Press. Washington DC. pp. 307-342.

**Mann, M., P. Højrup, and P. Roepstorff.** 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrom. **22**:338-345.

**Plumridge, J. A.** 1989. Sequence of the nagBCAD operon in Escherichia coli K12 and pattern of transcription within the nag regulon. Mol. Microbiol. **3**:505-515.

**Rabus, R., D. Gade, R. Helbig, M. Bauer, F. O. Glöckner, M. Kube, H. Schlesner, R. Reinhardt, and R. Amann.** 2002. Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1 by a proteomic approach. Proteomics **2**:649-655.

**Rose, M., and D. Botstein.** 1983. Construction and use of gene fusions to lacZ (beta-galactosidase) that are expressed in yeast. Methods. Enzymol. **101**:167-180.

**Rullkötter, J. 1999. In: Schulz, H. D., and M. Zabel.** (eds.). *Marine Geochemistry*. Springer-Verlag. Berlin. pp. 129-172.

**Schlesner, H., C. Rendsmann, B. J. Tindall, D. Gade, R. Rabus, S. Pfeiffer, and P. Hirsch.** 2004. Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA/DNA-hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov., and an emended description of the genus *Pirellula*. Int. J. Syst. Evol. Microbiol. in press.

**Tangney, M., C. J. Buchanan, F. G. Priest, and W. J. Mitchell.** 1997. Maltose uptake and its regulation in *Bacillus subtilis*. FEMS Microbiol. Lett. **76**:191-196.

6

# Growth phase/cycle dependent regulation of protein composition in *Rhodopirellula baltica*

Dörte Gade[1], Torben Stührmann[1], Richard Reinhardt[2] and Ralf Rabus[1*]

Manuscript in preparation

[1] Max-Planck-Institut für Marine Mikrobiologie, Bremen, Germany
[2] Max-Planck-Institut für Molekulare Genetik, Ressourcen Zentrum, Berlin, Germany
[*] Corresponding author

## Abstract

The life style of the marine bacterium *Rhodopirellula baltica* is based on a pronounced growth cycle. Multiplication of cells via budding produces motile swarmer cells. Older cells become immotile and tend to attach to each other forming large aggregates, so-called rosettes. While the former are abundant in early growth phases, the latter dominate during the stationary growth phase. Moreover, stress and starvation, which are characteristic conditions of the stationary phase, affect survival in the natural environment. In the present study, changes in protein composition correlating to different growth phases were quantitatively monitored by applying the 2-D DIGE technology. The number of regulated proteins (average ratio >2) increased from early (10) to late stationary growth phase (179), with average ratios reaching maximal values of 40. About 110 of these regulated protein spots were analysed by MALDI-TOF-MS and identified by PMF. Results indicate an opposing regulation of TCA and oxidative pentose phosphate cycle, a down-regulation of several enzymes involved in amino acid biosynthesis and an up-regulation of the alternative sigma factor $\sigma^H$ in stationary phase. Interestingly, 26 (conserved) hypothetical proteins, *i.e.* proteins of unknown function, were up- or down-regulated in the stationary phase. Several proteins were specifically regulated during growth on solid surface (agar plates). These proteins could be promising candidates for the cellular development observed in *R. baltica*.

## Introduction

*Rhodopirellula baltica* was isolated from the water column of the Baltic Sea and taxonomically grouped with the bacterial phylum *Planctomycetes* (Schlesner 1994; Schlesner et al. 2004). Members of this group are abundant in aquatic habitats (Staley et al. 1992; Ravenschlag et al. 2000, Gade et al. 2004) and considered as environmentally relevant. *Planctomycetes* share several morphologically unique properties, such as a peptidoglycan less proteinaceous cell wall (König et al. 1984; Liesack et al. 1986), intracellular compartmentalization (Lindsay et al. 2001) and a mode of reproduction via budding. The latter results in a cell cycle (Figure 1), which alternates between motile and sessile morphotypes and resembles the one known from *Caulobacter crescentus* (Jacobs-Wagner 2004). Interest in the *Planctomycetes* recently resulted in the publication of the 7.145 Mb genome of *R. baltica* (Glöckner et al. 2003).

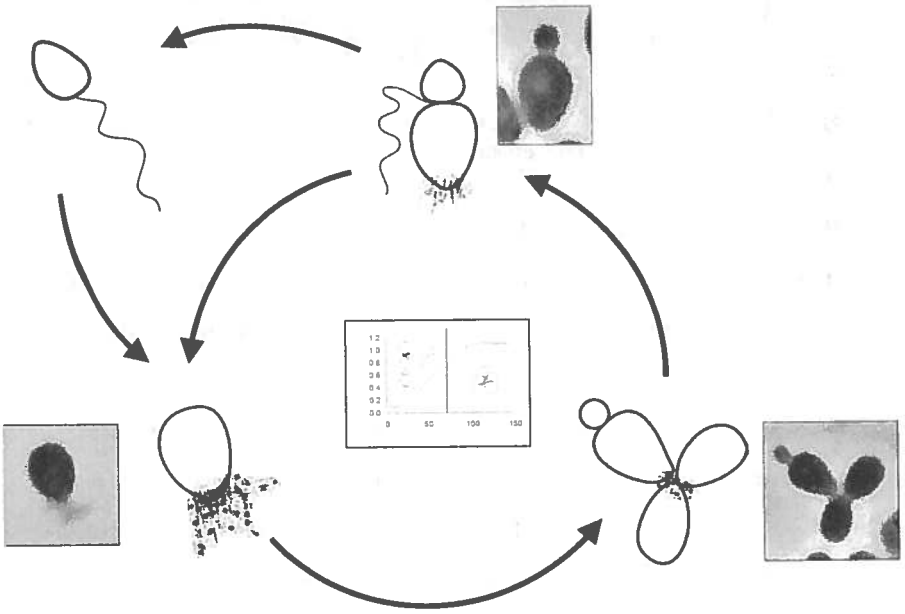


**Figure 1.** Schematic drawing of the life cycle of *R. baltica*. For clarity reasons cell structures typical for this planctomycete. such as fimbriae and crateriform structures, are not shown. See text for details.

The different morphotypes of *R. baltica* were thus far only studied by microscopy (Schlesner 1994; Schlesner et al. 2004). The pear-shaped adult cells of *R. baltica* are non-motile and display a polar organization. From the smaller cell pole a holdfast substance of yet unknown chemical composition is excreted. It mediates attachment of cells to surfaces or to each other forming rosettes (Figure 2 D).

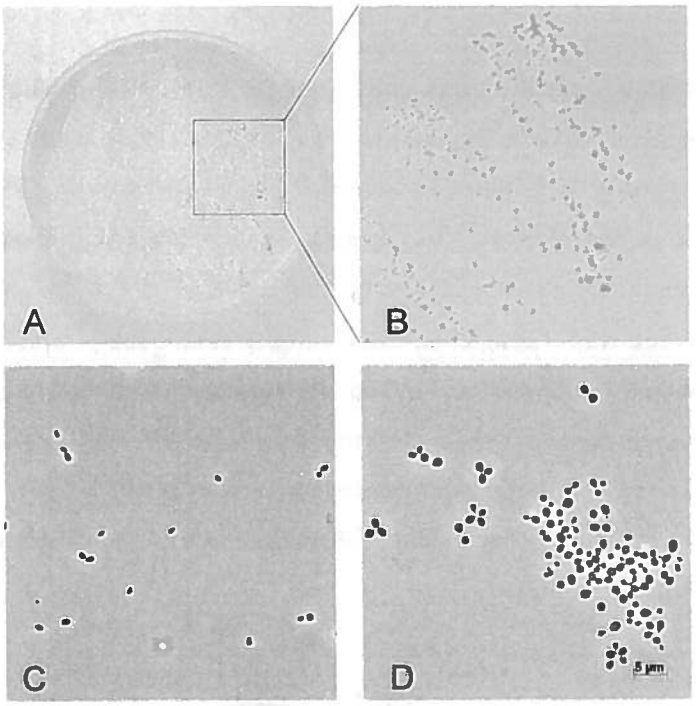



**Figure 2.** (**A** and **B**). Colonies of *R. baltica* grown on solid medium. (**C**) Microscopic picture of cells sampled in exponential phase. (**D**) Cell sampled in stationary phase.

The broader cell pole is the site of budding. The bud is a mirror image of the mother cell, except that it is smaller in size and has a single flagellum subpolarly inserted at the proximate pole. At a certain developmental stage the bud (daughter cell) is released as a motile swarmer. When differentiating into an adult cell, the swarmer aborts its flagellum (Figure 1). Only adult cells possess fimbriae that originate from so-called crateriform structures at the broader (reproductive) cell pole. While these crateriform structures are restricted to the broader cell pole in adult cells, they are observed across the complete cell surface of buds. The cellular mechanisms regulating these morphological changes during cell cycle of *R. baltica* are currently not understood. However, we could correlate distinct morphological states to different growth phases of *R. baltica* cultures by microscopic examination. During early stages of exponential growth numerous budding mother cells and swarmers are present. As the log phase proceeds, the number of swarmers decreases, and single cells begin to attach to each other forming rosettes. The latter are the dominating morphotype in late exponential and stationary phase, where swarmers are rarely observed. In the natural environment, *R. baltica*-like bacteria are rarely observed as single cells, but are rather attached to surfaces, *e.g.* marine snow particles (DeLong et al. 1993). Thus, at-

tachment and aggregate formation appear to be important elements of the natural life style of *R. baltica.*

The transition from exponential to stationary phase is generally characterized by a wide range of stress conditions such as high cell densities and nutrient depletion (starvation). Standard bacteria like *Escherichia coli* and *Bacillus subtilis* respond with complex physiological adaptations to such changes in growth conditions (Liu et al. 2000; Yoon et al. 2003; Bernhardt et al. 1997, 2003). It is generally assumed that bacteria have to survive extended periods of growth arrest due to unfavourable conditions in the natural environment. Therefore investigations of the stationary growth phase of environmental bacteria might provide insights into the suite of proteins required for survival in the natural habitat.

The described microscopic observations formed the basis of the present proteomic study. State of the art two-dimensional difference-in-gel-electrophoresis (2-D DIGE; Gharbi et al. 2002; Alban et al. 2003; Gade et al. 2003) was applied to profile changes in the protein composition of cells characterizing different morphotypes and stages in cell cycle, respectively, or reflecting adaptation to stress. Identification of regulated proteins was based on the recently completed master gel of soluble proteins (Gade et al. a in preparation) or achieved by matrix-assisted laser desorption/ionization time-of-flight-mass spectrometry (MALDI-TOF-MS) and analysis of generated peptide mass fingerprints (PMF).

## Materials and Methods

### Cultivation of cells and preparation of protein extracts

Cells of *R. baltica* were grown in batch cultures (500 ml) with mineral medium containing 10 mM *N*-acetylglucosamine as sole source of carbon and nitrogen (Rabus et al. 2002) or on corresponding solidified medium (18 g agar per l). The growth curve was determined by measuring the optical density at 600 nm ($OD_{600}$). Cultures were harvested after different incubation times to represent different optical densities of the growth curve (Figure 3). Harvesting was achieved by centrifugation (10,000 *g*, 20 min, 4°C) and washing of cells with 0.1 M Tris/HCl (pH 7.5). As control for the viability of cultures, subsamples (20 ml) were withdrawn from each non-stationary culture prior to harvesting and further incubated. All subsamples showed the same growth behaviour as exemplified in Figure 3. In late stationary phase, cell density decreased slightly. For clarity, the corresponding sample was designated $0.98_{old}$. Cells grown on solid medium where harvested after six days by abrading the colonies with a sterile rubber spatula and resuspended in 0.1 M Tris/HCl pH 7.5

containing 5 mM $MgCl_2$. This cell suspension was centrifuged (10,000 *g*, 10 min, 4°C). Harvested cells were rapidly frozen in liquid nitrogen and stored at -80 °C until cell breakage and further analysis. Preparation of extracts of soluble proteins was carried out as previously described (Gade et al. 2003).

**Two-dimensional difference-in-gel-electrophoresis (2-D DIGE)**

2-D DIGE was essentially carried out as described by Gade et al. (2003) with the only exception that 50 µg of each protein sample were labelled with 200 pmol cyanine dye.

The 2-D DIGE technology is based on pre-electrophoretic labelling of three sample types with different fluorescent dyes, allowing their co-separation in a single gel. Thus an individual experiment involves per gel: reference state, test state and internal standard. To achieve statistical confidence five parallel gels were run per experiment.

In the present study, protein extracts from cells harvested at $OD_{600}$ 0.47 were used as reference state and labelled with Cy5. Protein extracts from cells harvested at the other optical densities ($OD_{600}$ 0.38, 0.52, 0.82, 0.97, 1.02, $0.98_{old}$) or growth on agar plates represented the test states and were each labelled with Cy3. All test states were related to the reference state. For each individual experiment the internal standard was composed of equal amounts of reference state and the corresponding test state. The internal standards were labelled with Cy2.

Regulated proteins were detected by analysis of digitalized gel images with the DeCyder™ software (see below). To identify regulated proteins, separate preparative (400 µg protein load) gels were run. These gels were post-electrophoretically stained with colloidal Coomassie Brilliant Blue (cCBB) according to the method described by Doherty et al. (1998).

To assess biological variations, independent cultures were harvested at optical densities ($OD_{600}$) of 0.47, 0.82 and 1.02 and also analyzed by 2-D DIGE.

**Image acquisition and analysis**

DIGE gels were scanned directly between the glass plates at a resolution of 100 µm using the Typhoon™ 9400 scanner (Amersham Biosciences) with the CyDye-specific settings for excitation wavelength and emission filter as described before (Gade et al. 2003). Determination of protein abundance and statistical analysis were performed using the DeCyder software package (version 5.0; Amersham Biosciences).

### Gel sample excision und processing

Proteins that were recognized by 2-D DIGE to be up- or down-regulated, were co-detected in the preparative cCBB-stained gels by visual comparison of protein patterns. Those among them that were not present in the recently completed 2-D master gel, were selected for identification. They were manually excised, transferred to 96-well microtiter plates, subjected to tryptic digestion and MALDI-TOF-MS analysis, and identified by PMF as described before (Rabus et al. 2002). MS analysis were carried out by TopLap (Martinsried, Germany).

## Results and discussion

To study changes in protein composition that reflect adaptations to different growth stages, *R. baltica* was grown with *N*-acetylglucosamine as sole source of carbon and nitrogen. Under these conditions, *R. baltica* has a known doubling time of 14 h (Rabus et al. 2002). The growth curve for the present study (Figure 3) displays a pronounced linear phase, the transition to the stationary phase at around $OD_{600}$ 0.9, and a long stationary phase (at least 14 days). Cultures of *R. baltica* were harvested at optical densities of 0.38, 0.47, 0.52, 0.82, 0.97, 1.02 and $0.98_{old}$. Continued incubation of subsamples from each harvested, non-stationary culture yielded the same course of optical density as exemplified in Figure 3, demonstrating reproducible cultivation in each case. In addition, cultures grown on agar plates and harvested after six days of incubation were also analyzed to mimic growth during attachment on solid surfaces, such as *e.g.* marine snow particles. The $OD_{600}$ 0.47 culture served as common reference for the liquid cultures and the plate culture. An average ratio of >2 with a T-test value of $<10^{-5}$ was used as threshold for significant changes in protein abundance. The protein profiles of the tested samples were analyzed by 2-D DIGE, in order to detect changes in protein abundances that correlate with physiological adaptations to nutrient limitation or to different stages/morphotypes of the cell cycle.

### Growth in liquid culture

The total of protein spots detected as differentially regulated is shown in Figure 4 for each investigated growth state. Across the linear growth phase, as represented by the $OD_{600}$ 0.38 and 0.52 samples, only few proteins (10 and 12, respectively) appeared as regulated, indicating a rather constant overall protein composition. This reflects probably the favourable nutritional conditions encountered throughout the linear growth stage. The few observed

changes in the protein profiles may be attributed to different stages in cell division or morphotypes present in the respective cell population. This would agree with the applied batch cultivation, where cell division is not synchronized. The first pronounced change of the proteome was observed in the $OD_{600}$ 0.82 sample, which represents the transition to the stationary phase. Here, 48 and 27 proteins were up- and down-regulated, respectively. Since the transition phase is usually characterized by the onset of nutrient depletion and stress, *e.g.* due to high cell density, cells need to prepare for long term survival under suboptimal or even unfavourable conditions. Monitoring growth in the present study revealed that *R. baltica* survived for at least 14 days in the stationary phase, without indications for entry into a death phase (Figure 3). Apparently, *R. baltica* is well adapted for survival under nutrient limiting conditions, as often encountered in the natural environment. The most pronounced changes of the proteome were observed in late stationary phase, with 101 and 76 proteins being up- or down-regulated.

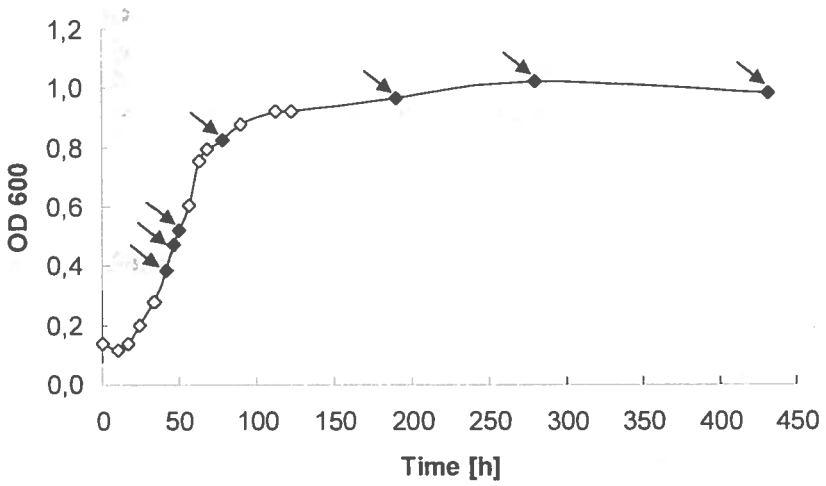



**Figure 3.** Growth of *R. baltica* in batch culture as determined by measuring optical density at 600 nm. Batch cultures were sampled at $OD_{600}$ 0.37, 0.47, 0.52, 0.82, 0.97, 1.02, and $0.98_{old}$ for 2-D analysis as indicated by arrows.

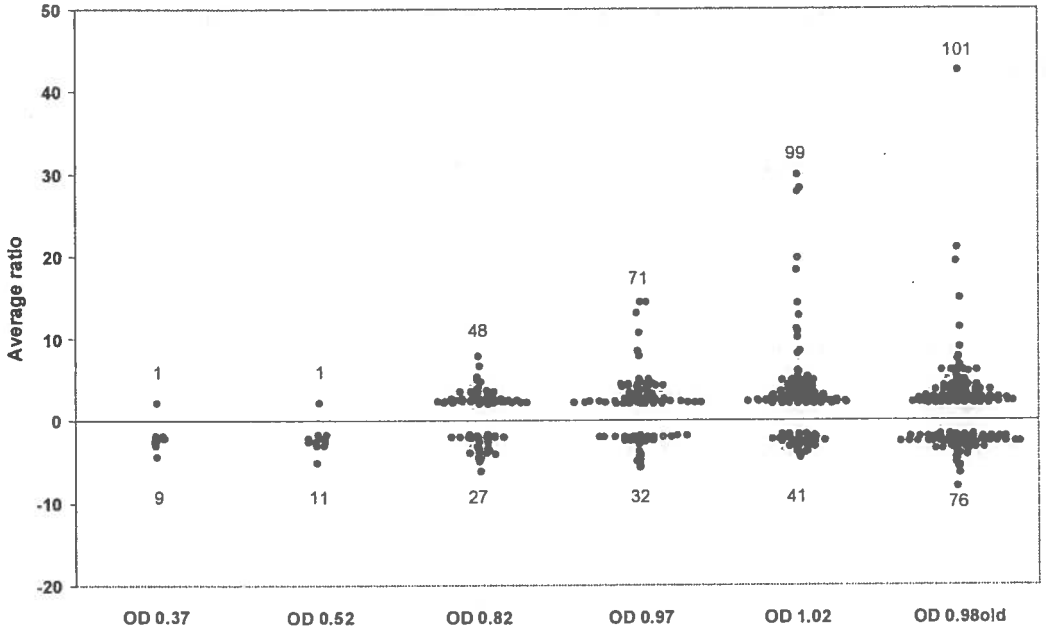



**Figure 4.** Proteins with abundances changing in response to different growth stages. Criteria for significance as applied for image analysis with DeCyder software: average ratio >2; T-test value $<10^{-5}$. Vertical columns represent the different growth stages (Figure 3), as indicated by the optical densities reached at time point of harvesting. Each dot represents a single protein spot on the 2-D gel, that was determined to be significantly less or more abundant in comparison to the reference state ($OD_{600}$ 0.47).

Identified proteins changing in relative abundance were grouped according to their putative functions and listed in Table 1. These proteins were manually excised from preparative cCBB stained gels, analyzed by MALDI-TOF-MS and identified by PMF. While 38 protein spots, corresponding to 32 open reading frames, were identified for the first time in this study, 72 proteins were also identified by comparing the protein patterns with the recently published master gel of the soluble proteome of *R. baltica* (Gade et al. a in preparation). The annotated 2-D DIGE gels (Figure 5) display all proteins listed in Table 1.

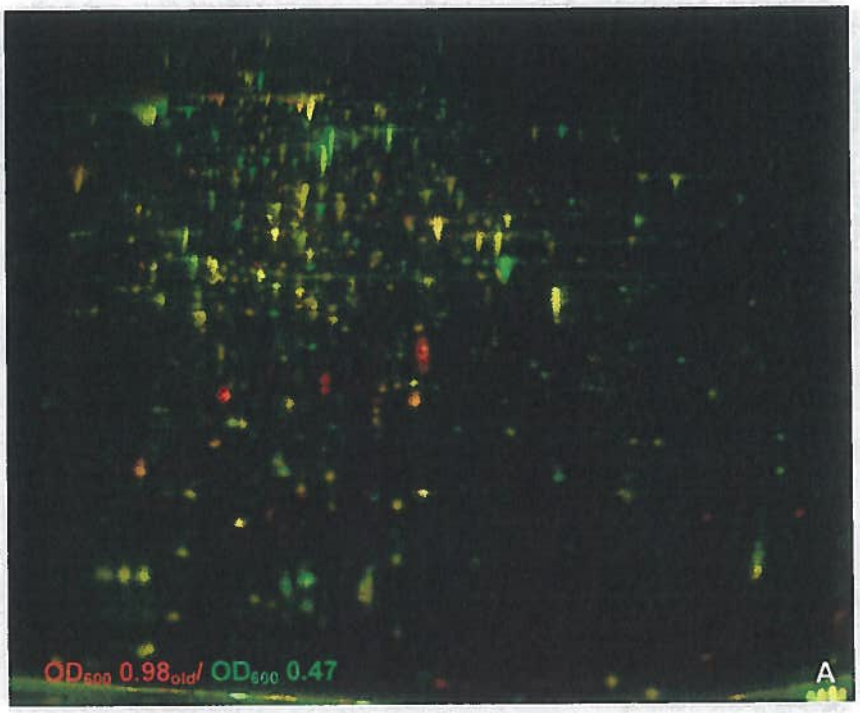
Regulated proteins of *R. baltica* differed not only with respect to the extent of changes in abundance, but also in the time courses of these changes. Some examples are illustrated in Figure 6. The conserved hypothetical protein RB1828 is equally abundant in the linear ($OD_{600}$ 0.52) and late stationary phase ($OD_{600}$ $0.98_{old}$), while it appears to be specifically down-regulated during the transition to stationary phase ($OD_{600}$ 0.82 to 0.97). The relative abundances of the hydrolase RB3689 and the conserved hypothetical protein RB11107 continuously increase during the linear phase until a maximum in early stationary phase ($OD_{600}$ 1.02). Both proteins are then somewhat less abundant in the late station-

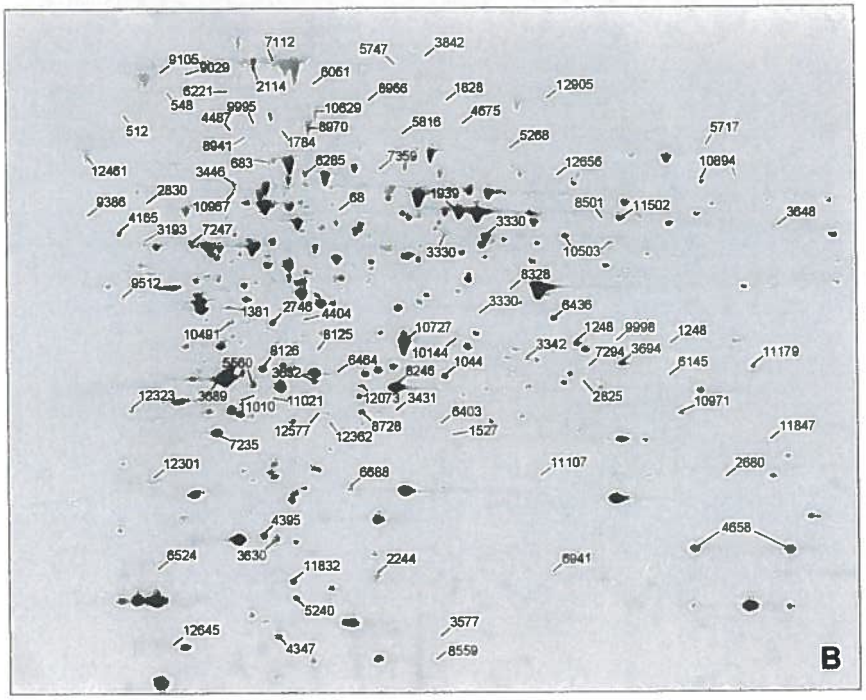
ary phase ($OD_{600}$ $0.98_{old}$). The universal stress response protein encoded by *RB11179* is rapidly up-regulated between $OD_{600}$ 0.52 and 0.82, reaching its maximum in abundance already at the beginning of the transition to stationary phase. From there its abundance remains constant until a slight decrease towards the end of the growth curve.

In the following sections, regulation of selected proteins will be described and discussed according to their functional classification.

**Carbon and nitrogen metabolism.** In the present study, *R. baltica* was grown with *N*-acetylglucosamine as sole source of organic carbon and nitrogen. Under these conditions the dehydrogenase RB3330 was previously shown to be specifically induced (Rabus et al. 2002). This enzyme was found to be 4-fold down-regulated during stationary phase, agreeing with depletion of *N*-acetylglucosamine. This was paralleled by the down-regulation of an ABC-transporter (RB11021), nitrogen assimilation regulatory protein (RB4487) and glutamine synthase II (RB7247), most likely in response to the depletion of the nitrogen source. Aconitase (RB2114) from the tricarboxylic acid cycle (TCA) was also down-regulated during stationary phase, while gluconolactonase (RB6145) and transaldolase (RB3193) from the pentose phosphate cycle were found to be up-regulated at the same time. This finding suggests that due to reduced TCA cycle activity the pentose phosphate cycle may play a more important role (Yoon et al. 2003).

Metabolic adaptation during entry into the stationary phase is also reflected by the formation of some hydrolases, dehydrogenases and oxidoreductases. A hydrolase encoded by *RB3689* (Figure 6) could be of particular relevance, since its relative abundance increased a 30-fold.


OD600 0.98old/ OD600 0.47
A


B

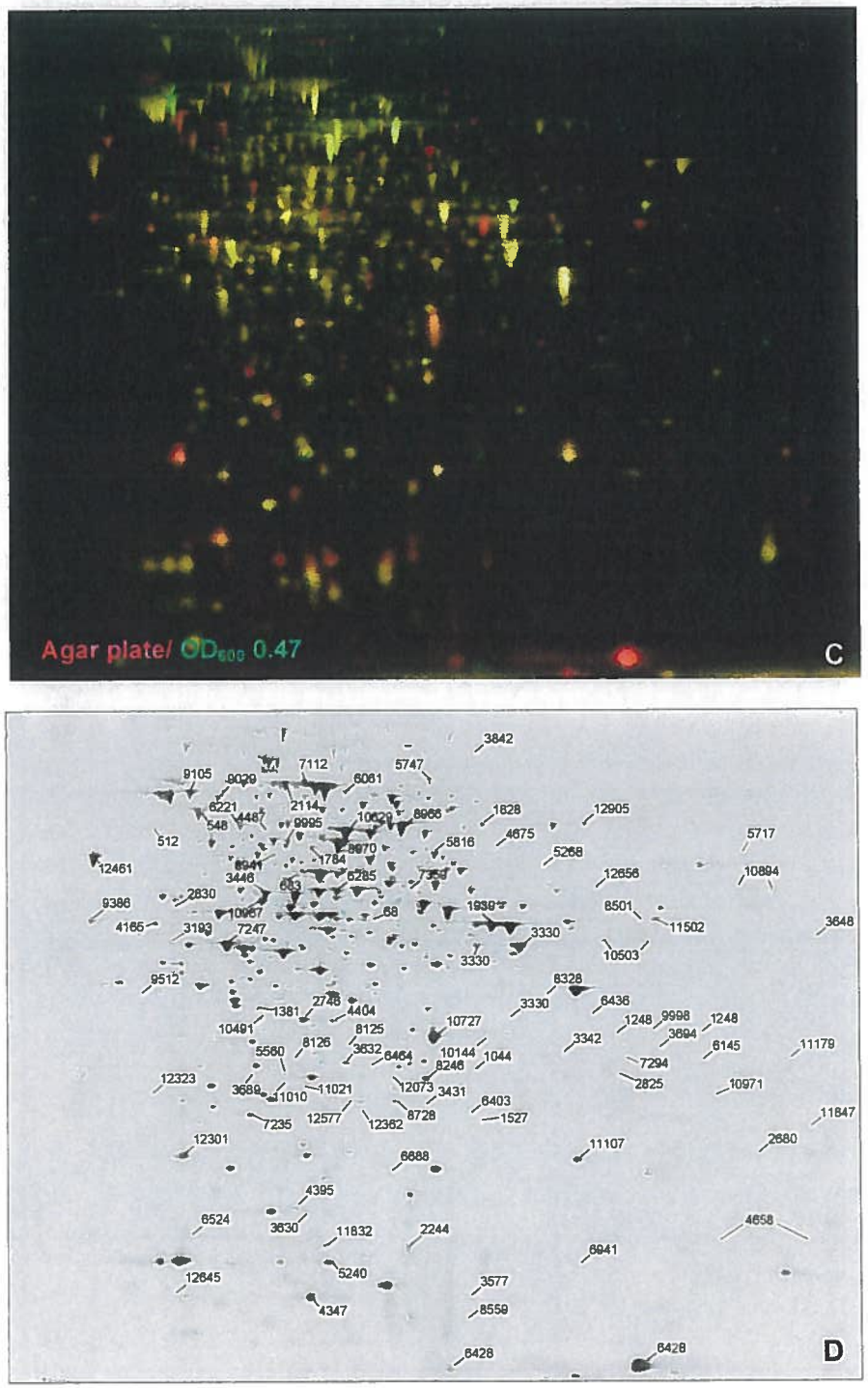



**Figure 5.** (**A** and **C**) Overlay images of 2-D DIGE gels. Protein spots appearing in red are up regulated under the tested conditions indicated in the bottom left corner. Whereas spots appearing in green are more abundant in cells sampled at an $OD_{600}$ of 0.47 (reference state). Yellow spots do not differ in relative abundance. Silver stained gels from cells sampled at $OD_{600}$ $0.98_{old}$ (**B**) or cells grown on agar plates (**D**). Identified proteins that were found to differ in abundance by DeCyder (Table 1) are indicated by their ORF numbers.

**Amino acid biosynthesis.** The abundance of several enzymes involved in the biosynthesis of amino acids such as methionine, leucine, valine, isoleucine and proline decreased as cells entered the stationary phase. The lower abundance of the enzyme γ-glutamyl phosphate reductase (RB7359), which is essential for proline biosynthesis, particularly agrees with the deceleration of growth and reproduction, since proline is one of the main constituents of the proteinaceous cell wall of *R. baltica* (Liesack et al. 1986).

Interestingly, the relative abundance of dihydrodipicolinate synthase (RB2746), which is involved in lysine biosynthesis, increased. Metabolic studies with *E. coli* revealed an intracellular increase of amino acids, *e.g.* lysine, in response to high cell densities (Liu et al. 2000). Also tryptophane seems to be more abundant during stationary phase, since the abundances of tryptophane synthase α-chain (RB5560) and tryptophane tRNA ligase (RB6436) increased. This finding agrees with the recently reported up-regulation of the tryptophane operon in *E. coli* during stationary phase (Yoon et al. 2003). The physiological meaning of elevated levels of lysine and tryptophane is presently unknown.

**Lipids.** As the relative abundances of sulfolipide biosynthesis protein (RB6464) and trans-2-enoyl-ACP reductase (RB8125) increased, one may speculate that biosynthesis of sulfated saturated fatty acids specifically occurs during stationary phase.

**Alternative sigma factors.** In the late stationary phase ($OD_{600}$ $0.98_{old}$) the alternative sigma factor $\sigma^H$ (RB4165) is 3.6-fold up-regulated. This extracytoplasmic function (ECF) sigma factor is a member of a subfamily of alternative sigma factors belonging to the $\sigma^{70}$ class. ECF sigma factors are known to be involved in the regulation of bacterial interactions with the extracellular environment, including adaptation to stress (Missiakas and Raina 1998). In *B. subtilis* $\sigma^H$ controls gene expression during sporulation and adaptation to nutrient depletion (Britton et al. 2002). In *Mycobacterium tuberculosis* $\sigma^H$ is involved in response to oxidative stress and heat shock (Manganelli et al. 2002). The present study with *R. baltica* reveals the increased formation of proteins required for detoxification of reactive oxygen species during high cell densities, such as catalase (RB10727), superoxide dismutase (SOD; RB6688) and alkylhydroperoxidases (RB3630 and RB4395). This finding indicates that *R. baltica* is exposed to oxidative stress during stationary phase. During all tested growth states, a distinct protein spot with constant abundance was identified as SOD. However, only samples from stationary phase contained a second spot also representing SOD. Both SOD spots have the same apparent molecular weight, while differing slightly in their isoelectric point (Figure 5).

Another alternative sigma factor, $\sigma^B$ (or $\sigma^S$), that plays a key role in general stress response of *B. subtilis* (Bernhardt et al. 1997) and/or *E. coli* (Hengge-Aronis 1999), was not detected on the gels. Since the $\sigma^S$ homolog of *R. baltica* (RB3727) has a theoretical p*I* of 6.7, it may be hidden in the smear at the neutral edge of the pH 4-7 gels.

**Protein biosynthesis and other essential functions.** The abundance of chaperons like DnaK (RB9105) and GroEL (RB10629) decreased in the late stationary phase. Together with the down-regulation of translational elongation factor EF-P (RB12577), tRNA synthetases (RB4675 and RB5747) and amino acid metabolism (see above), these findings indicate that protein biosynthesis is reduced to a level allowing maintenance of essential metabolic functions in the absence of growth. However, some proteins like the elongation factors EF-Tu, EF-Ts and EF-G, ATP-synthase, DNA-polymerase are still present in the same amounts as during linear growth. This finding suggests that these proteins are either continuously produced or that they have a long turn over time.

**Conserved hypothetical proteins.** These are proteins that have homologs in other organisms, but could not be assigned to any function at present. Thirteen conserved hypothetical proteins were identified, which changed their abundance during growth. One protein encoded by *RB11107* is unique among them, since it appears as three distinct spots on the gel; they differ with respect to their p*I*s (Figure 5). The relative abundances of these three spots increase dramatically during entry into the stationary phase, but drops to the level present in growing cells in late stationary phase (Table. 1).

**Hypothetical proteins.** These proteins could neither be assigned to any function, nor be found as homologs in other organisms. The latter indicates that they could be unique for *R. baltica* or other members of the *Planctomycetes*. In the present study, 14 hypothetical proteins were identified, which displayed changed abundances in different growth stages. If one assumes that these 14 hypothetical proteins do not function in adaptation to nutrient limitation or stress, they may represent promising candidates involved in the cell cycle and/or morphotype differentiation. The abundance of the hypothetical proteins RB3577, RB5268, RB7235, RB8559, and RB12461 increased in stationary phase (2.2 to 2.7-fold). They might be involved in rosette formation or attachment. In contrast, the hypothetical proteins encoded by *RB5240*, *RB6221*, and *RB6941* are more abundant in growing cells and may therefore be involved in proliferation or swarmer differentiation. The assumption that *R. baltica* recruits hypothetical proteins for these cellular processes agrees with the ab-

sence of nearly all cell division genes known from *E. coli* and *B. subtilis* in its genome (only exception being *ftsK*; Glöckner et al. 2003).

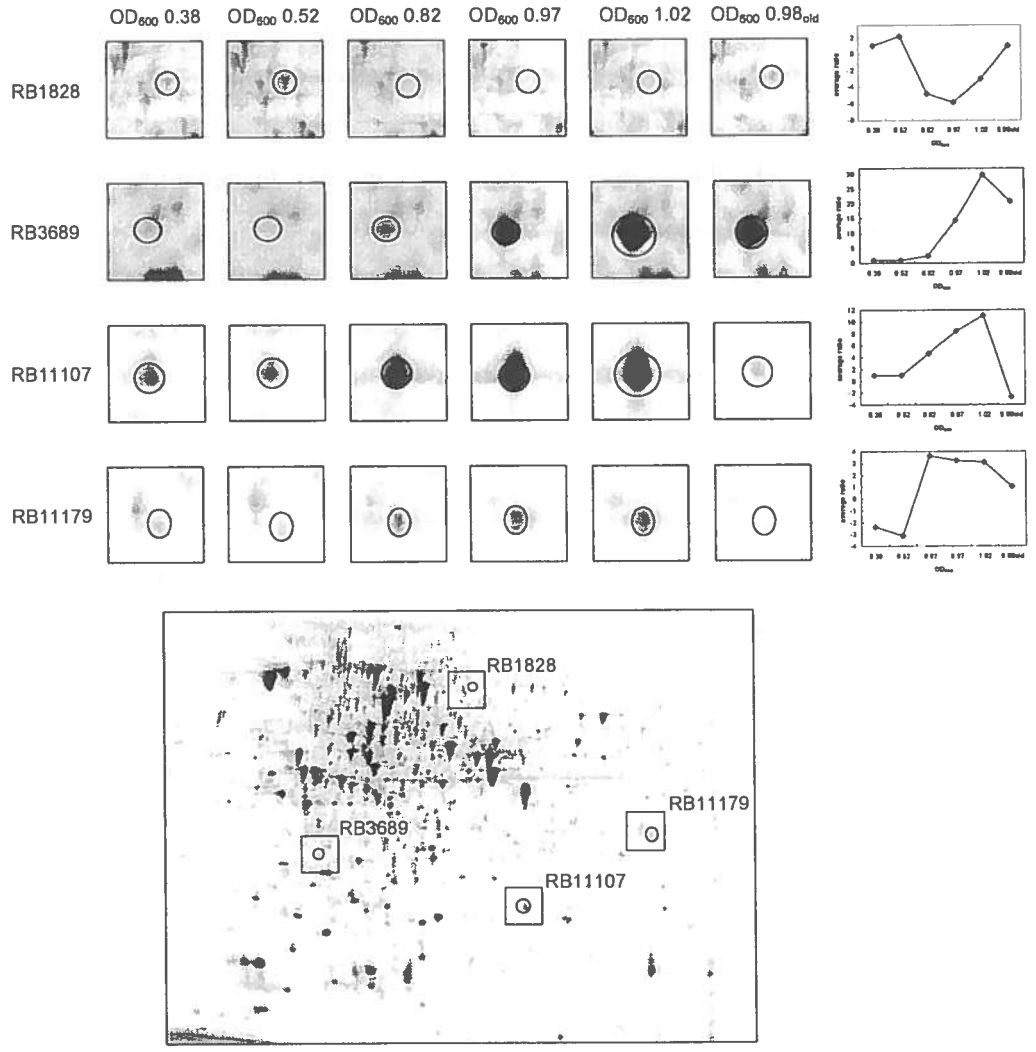



**Figure 6.** Changes in protein abundances of selected proteins in *R. baltica* throughout the batch culture experiment. The large gel depicted reflects the reference state sampled at an $OD_{600}$ 0.47.

To assess biological variation five additional cultures were harvested for each of the following optical densities $OD_{600}$ 0.47 (reference state), $OD_{600}$ 0.8 and $OD_{600}$ 1.1 (test states). One 2-D DIGE gel was run per harvested culture resulting in a set of 15 gels. In each case average ratios were determined for the same 14 spots, which were selected from the total of regulated proteins determined in the main experiment. For each of the two test states average ratios of the 14 spots were compared between this and the main experiment.

The average deviation of changes in abundance determined by DeCyder were 29.9% for $OD_{600}$ 0.8 and 33.7% for $OD_{600}$ 1.1.

**Growth on agar plates**

During growth on agar plates *R. baltica* forms colonies, which exhibit an intense pinkish colour due to the high cell density (Figure 2A and B). The protein profile of cells grown on agar plates was compared with the one of cells actively growing in liquid media ($OD_{600}$ 0.47). This approach was chosen to follow several lines of protein search.

The hypothetical protein encoded by *RB5240* forming two spots on the 2-D gel was 4.7-fold up-regulated in cells grown on plates. In contrast, this protein displayed basically constant abundances in all growth stages investigated for liquid cultures. The more basic one of the two spot forms was even less abundant in cells harvested at $OD_{600}$ higher than 0.82. Thus protein RB5240 might be involved in solid surface associated growth.

The hypothetical protein RB6941 was 11.4-fold less abundant in cells grown on plates than in the reference state from liquid culture, suggesting that this protein might be correlated to the free-living (probably motile) cell and involved in swarmer differentiation.

The elevated abundance of catalase (RB10727) and glutathione peroxidase (RB2244) in cells grown on plates indicates that cells in colonies are exposed to oxidative stress. Interestingly, glutathione peroxidase, a protein which acts on DNA and membrane damaging hydroperoxides, was regulated in opposite direction in cells from liquid cultures. Here the protein was even down-regulated in cells harvested between $OD_{600}$ 0.82 and 1.02. While the abundance of the universal stress protein RB11179 increased in stationary liquid cultures, it was 5.2-fold reduced in cells grown on plates. These results suggest that stress response differs in stationary liquid and plate cultures.

Two oxidoreductases (RB1939 and RB3648) and an α-glucan branching enzyme (RB548) were specifically up-regulated in cells grown on plates. These findings suggest metabolic differences between liquid and plate cultures and may point to the formation of glycogen as storage compound.

**Table 1.** Abundances of proteins identified as being up- or down-regulated during the six tested growth stages and the cells grown on solid medium relative to cell harvested at $OD_{600}$ 0.47. Coding genes and putative functions based on BLASTP analysis are displayed.

| ORF | putative function | optical density | | | | | | agar |
|---|---|---|---|---|---|---|---|---|
| | | 0.37 | 0.52 | 0.82 | 0.97 | 1.02 | 0.98old | plate |
| | **hypothetical proteins** | | | | | | | |
| 68 | hypothetical protein | | | | -2.8 | -2.8 | -2.9 | |
| 512 | hypothetical protein | | | 2.2 | | | | 2.3 |
| 2830 | hypothetical protein | 2.2 | | | | | | |
| 3577 | hypothetical protein-signal peptide prediction | | | | | 2.2 | 2.5 | |
| 5240 [1] | hypothetical protein-signal peptide and transmembrane prediction | | | | | | | 4.7 |
| 5240 [2] | hypothetical protein-signal peptide and transmembrane prediction | | | -3.8 | -4.1 | -2.4 | -3.4 | 2.5 |
| 5268 | hypothetical protein | | | | | | 2.6 | 2.6 |
| 6221 | hypothetical protein-signal peptide and transmembrane predicition | | | -5.2 | -5.2 | -3.0 | -5.4 | |
| 6941 | hypothetical protein-transmembrane prediction | | | -4.3 | -4.9 | -2.6 | -3.1 | -11.4 |
| 7235 | hypothetical protein-signal peptide prediction | | | | | 2.1 | 2.2 | |
| 8559 | hypothetical protein-signal peptide and transmembrane predicition | | | | | 2.1 | 2.7 | |
| 9512 | hypothetical protein-signal peptide predicition | | | | | | 2.3 | |
| 12461 | hypothetical protein | | | | | | 2.2 | |
| 12645 | hypothetical protein | | | 2.5 | | | | |
| | **conserved hypothetical proteins** | | | | | | | |
| 1828 | conserved hypothetical protein | | 2.1 | -4.8 | -5.9 | -3.0 | | 5.1 |
| 3342 | conserved hypothetical protein | | | | -2.1 | | -3.6 | |
| 4347 | conserved hypothetical protein | | | 3.1 | 3.2 | 3.0 | -2.9 | |
| 6428 [1] | conserved hypothetical protein-putative fasciclin domain | | | | | | 5.1 | 9.2 |
| 6428 [2] | conserved hypothetical protein-putative fasciclin domain | | | | 2.7 | 3.0 | | 7.4 |
| 8246 | conserved hypothetical protein | | | | | | 2.8 | |
| 11010 | conserved hypothetical protein | | -2.4 | -2.2 | -2.5 | | | |
| 11107 [3] | conserved hypothetical protein | | | 4.7 | 8.5 | 11.1 | -2.7 | |
| 11107 [1] | conserved hypothetical protein | | | 6.5 | 10.6 | 27.9 | | |
| 11107 [2] | conserved hypothetical protein | | | 2.5 | 4.6 | 10.8 | | |
| 12073 | conserved hypothetical protein | | | | | | 6.0 | |
| 12301 | conserved hypothetical protein | | | 2.1 | 2.0 | 2.4 | 2.8 | 3.3 |
| 12323 | conserved hypothetical protein | | | 2.2 | 2.0 | 2.1 | 3.5 | 2.5 |

**Table 1** continued.

| ORF | putative function | optical density | | | | | | agar |
|---|---|---|---|---|---|---|---|---|
| | | 0.37 | 0.52 | 0.82 | 0.97 | 1.02 | 0.98old | plate |
| | **C-compound metabolism** | | | | | | | |
| **548** | 1,4-alpha-glucan branching enzyme | | | | | | | 3.6 |
| **683** | aldehyde dehydrogenase | | | | | | 2.5 | |
| **2114** | aconitate hydratase | | | | | -2.2 | -3.7 | |
| **3193** | transaldolase | | | | | 2.1 | 2.6 | |
| **5816** | alpha L-fucosidase | | | | | | | 8.0 |
| **6061** | phosphomannomutase | | | | | | -2.7 | |
| **6145** | gluconolactonase precursor | | | | | | 2.2 | |
| **7294** | glucose 1-dehydrogenase | | | | | | -2.2 | |
| **8501** | alpha rhamnosidase | | | | | -4.8 | -3.6 | |
| **8941** | ketoglutarate semialdehyde dehydrogenase | | | | | | -4.1 | |
| **10144** | endo-1,4-beta-xylanase B | | | | -2.2 | | | 3.0 |
| | **amino acid metabolism** | | | | | | | |
| **1381** | similar to deoxyhypusine synthase (Dhs) | | | | 2.7 | 2.5 | | |
| **2746** | dihydrodipicolinate synthase | | | 5.1 | 3.3 | 2.5 | | |
| **3431** | ATP-dependent protease La 1 | | | | | 2.4 | | |
| **3842** | dipeptidyl peptidase IV | | | | | | -2.5 | 2.1 |
| **5560** | tryptophan synthase alpha chain | | | | | 4.7 | 5.9 | |
| **6285** | adenosylhomocysteinase | | | | | -2.3 | -3.0 | |
| **7112** | phenylalanyl-t-RNA synthetase beta chain | | | | | | -2.9 | |
| **7359** | gamma-glutamyl phosphate reductase | | | -2.2 | -3.6 | -4.0 | -3.6 | |
| **8126** | branched-chain amino acid aminotransferase | | | | | 2.0 | 3.5 | |
| **9029** | zinc metalloproteinase | | | | | | -6.8 | |
| **10894** [1] | threonine synthase precursor | | | | | 2.3 | | |
| **10894** [2] | threonine synthase precursor | | | | | | -3.6 | |
| **12656** | 3-isopropylmalate dehydratase large subunit | | | | | | -3.1 | |
| **12905** | acetolactate synthase III precursor | | | | | | -3.7 | |

**Table 1.** continued.

| ORF | putative function | optical density | | | | | | agar |
|---|---|---|---|---|---|---|---|---|
| | | 0.37 | 0.52 | 0.82 | 0.97 | 1.02 | 0.98old | plate |
| | **nucleotide metabolism** | | | | | | | |
| **1784** | UDP-glucose 6-dehydrogenase | | | | | | -2.8 | |
| **4658 [1]** | inosine monophosphate dehydrogenase-related protein | | | | | 14.2 | 42.5 | 6.7 |
| **4658 [2]** | inosine monophosphate dehydrogenase-related protein | | | | | | 19.2 | 13.7 |
| **5717** | soluble pyridine nucleotide transhydrogenase | | | | | 2.1 | | |
| **6524** | hypoxanthine-guanine phosphoribosyltransferase | | | 2.5 | 3.7 | 3.2 | 4.0 | 2.3 |
| **11832** | nucleoside diphosphate kinase (NDK) | | | -2.1 | -2.4 | -1.7 | -2.0 | |
| | **fatty acid and lipid metabolism** | | | | | | | |
| **2825 [1]** | glycerophosphodiester phosphodiesterase | | | | | | -2.3 | |
| **2825 [2]** | glycerophosphodiester phosphodiesterase | | | | | | -3.5 | |
| **6464** | sulfolipid biosynthesis protein SdqC | | | | | 2.5 | 3.6 | |
| **8125** | trans-2-enoyl-ACP reductase | | | | | 4.8 | 5.9 | |
| | **metabolism** | | | | | | | |
| **1044** | metal-dependent hydrolases | | | | 2.2 | 3.3 | 3.5 | |
| **1939** | oxidoreductase | | | | | | | 4.0 |
| **3330 [2]** | dehydrogenase | | | | | | -4.4 | |
| **3330 *** | dehydrogenase | | | | | 2.8 | | |
| **3330 [1]** | dehydrogenase | | | | | 2.5 | | |
| **3632 *** | short chain dehydrogenase | | | | 3.8 | 6.1 | 7.4 | |
| **3632** | short chain dehydrogenase | | | | 4.4 | 8.0 | 8.8 | |
| **3648** | oxidoreductase | | | -4.2 | | | | 5.7 |
| **3689** | hydrolase of unknown specifity | | | 2.3 | 14.3 | 29.7 | 20.8 | 3.3 |
| **4404** | dehydrogenase of the short-chain dehydrogenase family | | | | | -2.5 | -6.1 | |
| **8728** | oxidoreductase | | | | 3.0 | 3.1 | | |
| **10503 [1]** | NADH-dependent oxidoreductase | | | | 2.3 | 2.5 | | |
| **10503 [2]** | NADH-dependent oxidoreductase | | | | | | 3.5 | |
| **10967** | oxidoreductase | | | | | | -3.2 | |
| **10971** | short-chain dehydrogenase | | | | | 2.5 | | |

**Table 1.** continued.

| ORF | putative function | optical density | | | | | | agar |
|---|---|---|---|---|---|---|---|---|
| | | 0.37 | 0.52 | 0.82 | 0.97 | 1.02 | 0.98old | plate |
| | **inorganic compounds** | | | | | | | |
| 3694 | non-heme chloroperoxidase | | | | | 2.2 | 4.4 | |
| 7247 | glutamine synthetase II | | | | | | -2.2 | 2.3 |
| 11502 | alkylsulfatase | | | | | | | 2.9 |
| | **regulation** | | | | | | | |
| 4165 | RNA polymerase sigma-H factor | | | | | | 3.6 | |
| 4487 | nitrogen assimilation regulatory protein | | | | | | -5.7 | |
| 6403 | response regulator | | | | | | -2.8 | |
| 10491 | two-component system regulatory protein | | | | 2.1 | 2.4 | 2.7 | |
| | **translation and protein quality control** | | | | | | | |
| 2680 | protein disulfide isomerase | | | | | 3.4 | 3.3 | |
| 3446 | peptidyl-prolyl cis-trans isomerase cyp2 | | | | | | | 3.3 |
| 4675 | cysteinyl-tRNA synthetase | | | | -2.1 | -2.2 | | |
| 5747 | arginyl-tRNA synthetase | | | | | | -2.4 | |
| 6436 | tryptophan t-RNA ligase | | | | | | 2.1 | |
| 8328 | CMP-binding protein | | | | | 2.7 | 3.0 | |
| 8966 | 60 kDa chaperonin | | | | | | -2.8 | |
| 8970 | 60 kDa chaperonin 5 | | | | | | -2.5 | |
| 9105 | DnaK | | | | | | -2.2 | |
| 9386 | thiol-disulfide interchange protein | | | -2.4 | -2.7 | | | |
| 10629 | GroEL | | | | | | -3.2 | |
| 12577 | translation elongation factor EF-P | | | | | | -2.3 | |
| | **other** | | | | | | | |
| 1527 | platelet-activating factor acetylhydrolase IB gamma subunit | | | | | | | 4.1 |
| | **transport** | | | | | | | |
| 1248 [1] | ABC-type multidrug transport system, ATPase component | | | | | 2.1 | | |
| 1248 [2] | ABC-type multidrug transport system, ATPase component | | | | | | -4.5 | |
| 9995 | iron-regulated ABC-transport system | | | | | | -3.3 | |
| 9998 | ATP-binding protein, ABC-transport system | | | | | | -2.9 | |
| 11021 | ATP-binding protein, ABC-transport system | -2.1 | -2,08 | | | | | |

Table 1. continued.

| ORF | putative function | optical density | | | | | | agar |
|---|---|---|---|---|---|---|---|---|
| | | 0.37 | 0.52 | 0.82 | 0.97 | 1.02 | 0.98old | plate |
| | **stress response** | | | | | | | |
| **2244** | glutathione peroxidase | | | -3.6 | -5.9 | -4.3 | | 4.1 |
| **3630** | alkylhydroperoxidase AhpD domain protein | | | | 2.7 | 4.0 | 4.2 | |
| **4395** | alkylhydroperoxidase AhpD domain protein | | | 2.6 | 4.1 | 4.6 | 4.7 | |
| **6688** | superoxid dismutase, Mn familiy | | | | | 2.0 | 3.1 | |
| **10727** | manganese-containing catalase | | | 3.4 | 4.2 | 4.4 | 5.9 | 2.9 |
| **11847** | peptide methionine sulfoxide reductase | | | | 4.1 | 5.0 | | -2.6 |
| **11179** | universal stress protein | -2.4 | -3,14 | 3.6 | 3.2 | 3.1 | | -5.2 |
| **12362** | peroxiredoxin | | -2,4 | -4.1 | -2.7 | | | 2.7 |

In some cases the genes formed more than one spot on the gels. These spots of the same gene product were marked as follows
[1] the more acidic spot of one protein species forming two spots with the same Mr
[2] the more basic spot of one protein species forming two spots with the same Mr
[3] the middle of three spots of the same protein species; all having the same Mr but differing in p*I*
* protein species with the same p*I* but a higher Mr

## Conclusions

In the present study, questions about nutrient limitation and general stress, as well as about growth stages and cell cycle of *R. baltica* lead to hypothesis-driven physiological experiments, which were followed up by global proteomic analysis. This approach allowed identification of numerous specifically regulated proteins, which represent promising candidates for further molecular studies.

Proteins of unknown function constitute a major fraction of the theoretical proteome (67%) of *R. baltica*. First evidence for their principle formation was recently provided by the master gel of soluble proteins, where they amounted to 34% of all identified proteins (Gade et al. a in preparation). Unknown proteins are also involved in carbohydrate metabolism of *R. baltica* as determined by proteomic analysis of cells adapted to growth with various carbohydrates (Gade et al. b in preparation). In the present study, the abundance of numerous unknown proteins was revealed to be specifically regulated during stationary phase (26) or in response to growth on solid surfaces (1). Thus the combined physiological/proteomic studies indicate that physiological properties, morphotypes and cellular processes typical for *R. baltica* require the involvement of thus far unknown proteins.

## Acknowledgements

We are thankful to H. Schlesner (Kiel) for stimulating discussions. We thank D. Lange (Bremen) for technical assistance, A. Beck (Berlin) for bioinformatical support and T. Halder (Martinsried) for MS analysis. We are indepted to Friedrich Widdel for continuous support of proteomic work at the MPI in Bremen. This study was supported by the Max Planck Society and the BMBF.

## References

**Alban, A., S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie.** 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics **3**:36-44.

**Bernhardt, J., U. Völker, A. Völker, H. Antelmann, R. Schmid, H. Mach, and M. Hecker.** 1997. Specific and general stress response in *Bacillus subtilis* – a two-dimensional protein electrophoresis study. Microbiology **143**:999-1017.

**Bernhardt, J., J. Weibezahn, C Scharf, and M. Hecker.** 2003. *Bacillus subtilis* during feast and famine: visualization of the overall regulation of protein synthesis during glucose starvation by proteome analysis. Genome Res. **13**:224-237.

**Britton, R.A., P. Eichenberger, J. E. Gonzalez-Pastor, P. Fawcett, R. Monson, R. Losnik, and A. D. Grossman.** 2002. Genome-wide analysis of the stationary-phase sigma factor (Sigma-H) regulon in *Bacillus subtilis*. J. Bacteriol. **184**:4881-4890.

**DeLong, E. F., D. G. Franks, and A. I. Alldredge.** 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol. Oceanogr. **38**:924-934.

**Doherty, N. S., B. H. Littman, K. Reilly, A. C. Swindell, J. M. Buss and N. L. Anderson.** 1998. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis **19**:355-363.

**Gade, D., J. Thiermann, D. Markowsky, and R. Rabus.** 2003. Evaluation of two-dimensional difference gel electrophoresis for protein profiling. J. Mol. Microbiol. Biotechnol. **5**:240-251.

**Gade, D., H. Schlesner, F. O. Glöckner, R. Amann, S. Pfeiffer, and M. Thomm.** 2004. Identification of planctomycetes with order-, genus, and strain-specific 16S rRNA-targeted probes. Microb. Ecol. Mar 4 Epub ahead of print.

**Gharbi, S., P. Gaffney, A. Yang, M. J. Zvelebil, R. Cramer, M. D. Waterfield, and J. F. Timms.** 2002. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol. Cel. Proteomics **1**:91-98.

**Glöckner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann, and R. Reinhardt.** 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc. Natl. Acad. Sci. USA **100**:8298-8303.

**Hengge-Aronis, R.** 1999. Interplay of global regulators and cell physiology in the general stress response of *Escherichia coli*. Cur. Opin. Microbiol. **2**:148-152.

**Jacobs-Wagner, C.** 2004. Regulatory proteins with a sense of direction: cell cycle signalling network in *Caulobacter*. Mol. Microbiol. **51**:7-13.

**König, H., H. Schlesner, and P. Hirsch.** 1984. Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on a *Prosthecomicrobium* sp. Arch. Microbiol. **138**:200-205.

**Liesack, W., H. König, H. Schlesner, and P. Hirsch.** 1986. Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella/Planctomyces* group. Arch. Microbiol. **145**:361-366.

**Lindsay M., R. I. Webb, M. Strous, M. S. M. Jetten, M. K. Butler, R. J. Forde, and J. A. Fuerst.** 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch. Microbiol. **175**:413-429.

**Liu, X., C. Ng, and T. Ferenci.** 2000. Global adaptations resulting from high population densities in *Escherichia coli* cultures. J. Bacteriol. **182**:4158-4164.

**Manganelli, R., M. I. Voskuil, G. K. Schloonik, E. Dubnau, M. Gomez, and I. Smith.** 2002. Role of the extracytoplasmic-function σ factor $\sigma^H$ in *Mycobacterium tuberculosis* global gene expression. Mol. Microbiol. **45**:365-374.

**Missiakas, D., and S. Raina.** 1998. The extracytoplasmic function sigma factors: role and regulation. Mol. Microbiol. **28**:1069-1066.

**Rabus, R., D. Gade, R. Helbig, M. Bauer, F. O. Glöckner, M. Kube, H. Schlesner, R. Reinhardt, and R. Amann.** 2002. Analysis of *N*-acetylglucosamine metabolism in the marine bacterium *Pirellula* sp. strain 1 by a proteomic approach. Proteomics **2**:649-655.

**Ravenschlag, K., K. Sahm, and R. Amann.** 2000. Quantitative molecular analysis of the microbial community in marine arctic sediments (Svalbard). Appl. Environ. Microbiol. **67**:387-395.

**Schlesner, H.** 1994. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from various aquatic habitats using dilute media. Syst. Appl. Microbiol. **17**:135-145.

**Schlesner, H., C. Rendsmann, B. J. Tindall, D. Gade, R. Rabus, S. Pfeiffer, and P. Hirsch.** 2004. Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA/DNA-hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov., and an emended description of the genus *Pirellula*. Int. J. Syst. Evol. Microbiol. in press

**Staley, J. T., J. A. Fuerst, S. Giovannoni, and H. Schlesner.** 1992 The order *Planctomycetales* and the genera *Planctomyces*, *Pirellula*, *Gemmata* and *Isosphaera*. In: Balows, A., H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (eds.). The Prokaryotes 2nd edition, Springer-Verlag New York, pp 3710-3731.

**Yoon, S.H., M.-J. Han, S. Y. Lee, K. J. Jeong, and J.-S. Yoo.** 2003. Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnol. Bioeng. **81**:753-767.

# C Weitere Publikationen

1. **Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1**
   Frank Oliver Glöckner, Michael Kube, Margarete Bauer, Hanno Teeling, Thierry Lombardot, Wolfgang Ludwig, Dörte Gade, Alfred Beck, Katja Borzym, Katja Heitmann, Ralf Rabus, Heinz Schlesner, Rudolf Amann, Richard Reinhardt

   Proc. Natl. Acad. Sci. USA (2003) **100**:8298-8303.

## Danksagung

Zu guter Letzt möchte ich allen Personen danken, ohne die diese Arbeit nicht in der vorliegenden Form hätte entstehen können.

Herr Prof. Widdel hat mir nicht nur die Möglichkeit gegeben, als einzige Aerobe unter Anaeroben in seiner Abteilung diese Arbeit anzufertigen, sondern auch die Proteomforschung am MPI in Bremen unterstützt. Hierfür und für die Erstellung des Erstgutachtens sei ihm ganz herzlich gedankt.

Außerordentlich dankbar bin ich Priv. Doz. Ralf Rabus für die zahlreichen Denkanstöße („Denk da noch mal drüber nach.") und für die Unterstützung beim Finden und Verfolgen des „roten Fadens". Als direkter Betreuer hat er maßgeblich zum Gelingen dieser Arbeit beigetragen.

Herrn Dr. Heinz Schlesner danke ich dafür, dass er bereits vor Jahren mein Interesse an den Planktomyceten geweckt hat und sein umfangreiches Wissen über diese besondere Bakteriengruppe mit mir teilte. Er stand mir stets mit Rat und Tat zur Seite.

Dr. Johan Gobom war bereit, meine Probe auch ohne Finanzierung weiterhin zu analysieren. Dafür danke ich ihm sehr und bin bereit, die teilweise ausufernden Wartezeiten zu vergessen.

Torben Stührmann sei gedankt für die Anfertigung der wunderschönen DIGE Gele aus der Wachstumskurve und seine kurzweilige wenn leider auch nur kurzzeitige Gesellschaft im 2D Keller. Großes Kino!

Thierry Lombardot danke ich für die Einführung in die PEDANT Software und für das Entwirren des einen oder anderen Datenwustes.

Die Firma Amersham Biosciences ermöglichte es mir, die ersten DIGE Gele in ihrem Labor in Freiburg zu scannen. Dr. Jürgen Thiermann und Dieter Markowsky unterstützten mich tatkräftig bei diesen nächtlichen Aktionen. Danke!

An dieser Stelle möchte ich meinen Kollegen aus der Abteilung Mikrobiologie für ihre Hilfsbereitschaft und das gute Arbeitsklima danken. Besonders Daniela Lange und meinen Bürokollegen danke ich für das Teilen der kleinen und großen Sorgen.

Der größte Dank geht an Till Andrießen, für seine Geduld, sein Verständnis und die unermüdliche geleistete Aufbauarbeit besonders in der Endphase der Doktorarbeit. Nicht vergessen werden darf an dieser Stelle meine Familie, der ich zutiefst dankbar bin, weil sie mir gerade in den letzten Jahren jede nur erdenkliche Form an Unterstützung zukommen ließ.