REGULAR ARTICLE

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

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

[1] Max Planck Institute for Marine Microbiology, Bremen, Germany
[2] Max Planck Institute for Molecular Genetics, Berlin, Germany

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. Almost all enzymes of glycolysis and TCA cycle 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 2-D 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.

1 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 20–30×10^9 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 [1]. 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 [2]. 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 [3]. Thus *Planctomycetes* are considered as key players in carbohydrate metabolism in marine systems. The recent publication of its genome sequence [4] 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.

2-DE is one of the central and well-established techniques in proteomic research [5]. It allows separation of complex protein mixtures at highest resolution. Identification of 2-DE

Correspondence: Dr. Ralf Rabus, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany
E-mail: rrabus@mpi-bremen.de
Fax: +49-421-2028-580

separated proteins is most widely performed by determination of peptide mass fingerprints (PMFs) with MALDI-MS [6], and subsequent comparison of the obtained PMF with theoretically calculated peptide masses [7]. 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 2-D DIGE technology, which is based on covalent labeling of proteins with fluorescent dyes prior to electrophoresis and application of an internal standard, allows determination of differences in protein abundance with statistical confidence [8–11].

Our initial proteomic study revealed the specific formation of a dehydrogenase in response to growth with *N*-acetylglucosamine [12], 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.

2 Materials and methods

2.1 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 [12]. Cells were adapted to each of these eight carbohydrates over at least five passages. Cells were harvested in the exponential growth phase by centrifugation ($10\,000 \times g$, 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, 30 mM Tris-HCL, 2% CHAPS ph 8.5. Cell breakage was performed with the PlusOne® grinding kit (Amersham Biosciences, Freiburg, Germany) following the manufacturer's instructions. Removal of cell debris, DNA and membranes by centrifugation ($100\,000 \times g$, 1 h, 15°C) yielded the fraction of soluble proteins. The protein content of this fraction was determined using the method described by Bradford [13].

2.2 2-D DIGE

2-D DIGE was essentially carried out as recently described [11] with the exception that 200 pmol cyanine dye were used for labeling of 50 µg protein sample. Application of three

different fluorescent dyes (Cy2, Cy3 and Cy5), 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 labeled with Cy5. Extracts from cells adapted to growth with either ribose, xylose, *N*-acetylglucosamine, maltose, lactose, melibiose or raffinose represented the seven test states and were labeled with Cy3 in each case. For each comparative experiment, the internal standard was composed of equal amounts of reference state and one of the seven test states, and labeled with Cy2. A mixture of equal amounts of each of the three types of labeling preparations was loaded *per* gel. To achieve statistical confidence, five parallel gels were run for each of the seven comparative experiments. To obtained sufficient amounts of protein for MS analysis, separate preparative gels were run for each of the test substrates. These gels were fixed and stained with colloidal CBB (cCBB; modified from Doherty *et al.* [14]). Proteins of interest, as defined by the 2-D DIGE/DeCyder analysis, were excised from the cCBB stained gels for subsequent MS analysis.

2.3 Image acquisition and analysis

2-D 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 [11]. Image and statistical analyses were performed using the DeCyder™ software package (version 5.0; Amersham Biosciences). Images of colloidal Coomassie stained gels were collected with the Image Scanner (Amersham Biosciences).

2.4 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. High reproducibility of protein patterns (Fig. 2) allowed co-detection of these proteins in parallel cCBB-stained gels. Selected protein spots were excised from these gels using an automatic excision workstation (Proteineer; Bruker Daltonics, Bremen, Germany) and delivered into 96-well poly-propylene microtiter plates (Costar Thermowell®, Cornis, NY, USA). Washing solutions and reagents were removed by centrifugation through the pierced (d <0.5 mm) well bottoms. Protection from contamination was achieved by placing the pierced 96-well microtiter plates (MTP) in a second 96-well MTP and covering with a lid. Following removal of all liquid 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
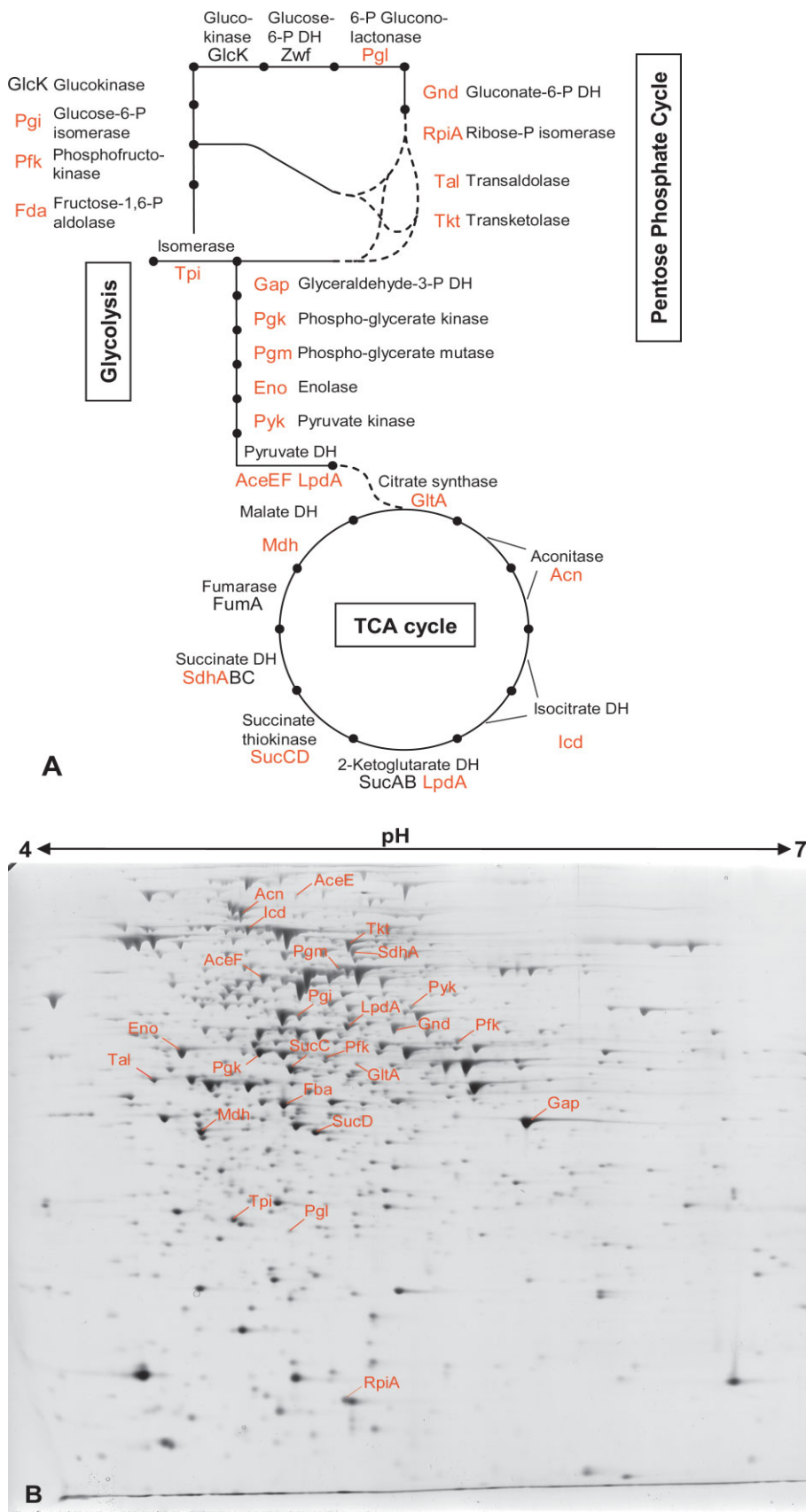
Figure 1. Central routes of carbohydrate metabolism. (A) Schematic representation of glycolysis, oxidative pentose phosphate cycle and the TCA cycle. Enzymes identified by 2-DE/MALDI-TOF MS are highlighted in red. In case of succinate dehydrogenase and 2-ketoglutarate dehydrogenase complex only subunits were detected. (B) Partial view of the *R. baltica* proteome during growth with *N*-acetylglucosamine as revealed on a silver-stained 2-D gel. Protein spots correlated with the carbohydrate catabolism shown in Fig. 2A are annotated with the abbreviation of the protein name.

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 evaporation of the ethanol. 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 MTP were placed in a humidified box and incubated for 4 h at 37°C.

2.5 MALDI-TOF MS and protein identification

Protein digests were prepared for MALDI using the α-cyano-4-hydroxycinnamic acid affinity sample preparation technique. Mass analysis of positively charged ions was performed on an Ultraflex LIFT and a Reflex III instrument (Bruker Daltonics) operated in the reflector mode and using delayed ion extraction. Positively charged ions in the mass range 700–3500 Da were analyzed.

Database searching was performed using the software MASCOT (Matrix Science, London, UK). 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 ($p < 0.05$).

2.6 Enzyme studies

2.6.1 General remarks

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, Daventry, UK) applying 2700 bar. Intact cells, cell debris and membranes were removed by centrifugation ($17\ 900 \times g$, 15 min, 4°C). The protein content of the soluble fraction was determined according to the method described by Bradford [13].

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 pH-adjusted 0.5 M Tris/HCl buffers. All enzyme activities were measured in triplicate for each cell extract of the eight 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 minute.

2.6.2 6-Phosphofructokinase (EC 2.7.1.90)

6-Phosphofructokinase was indirectly 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.

2.6.3 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.

2.6.4 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.

2.6.5 Transaldolase (EC 2.2.1.2)

Transaldolase was measured by coupling the formation of glyceraldehyde-3-phosphate to the oxidation of NADH *via* triosephosphate isomerase (TIM) and glycerol-3-phosphate dehydrogenase. The assay mixture contained 84 mM Tris/HCl (pH 7.5), 5.4 mM fructose-6-phosphate, 0.7 mM erythrose-4-phosphate, 0.2 mM NADH, 16.7 U/mL TIM, 0.134 U/mL glycerol-3-phosphate dehydrogenase and 0.1 mg protein.

2.6.6 Isocitrate dehydrogenase (EC 1.1.1.42)

Isocitrate dehydrogenase was assayed by monitoring the reduction of $NADP^+$. The assay mixture contained 88 mM Tris/HCl (pH 8.5), 3.5 mM $MgCl_2$, 0.41 mM $NADP^+$, 0.55 mM isocitrate and 0.1 mg protein.

2.6.7 Malate dehydrogenase (EC 1.1.1.37)

Malate dehydrogenase was measured in the reductive direction following the oxidation of NADH. The assay mixture contained 95 mM phosphate buffer (pH 7.0), 0.5 mM oxaloacetate, 0.2 mM NADH and 0.1 mg protein.

2.6.8 β-Galactosidase (EC 3.2.1.23)

β-Galactosidase was tested according to the method described by Rose and Botstein [15]. Specific β-galactosidase activity was normalized to the total protein in each extract and equals (optical density at 420 nm × 1.7)/(0.0045 × extract volume × time).

3   Results and discussion

3.1 Reconstruction of central routes for carbohydrate degradation

Glycolysis, the oxidative branch of pentose phosphate cycle and the tricarboxic acid cycle form the main routes for carbohydrate catabolism in most organisms. The recently established 2-D map of soluble proteins allowed us to demonstrate that these pathways are also operative in *R. baltica* (Fig. 1). As known from proteomic studies with other bacteria, enzymes of these central pathways belong to the most abundant protein spots on 2-D gels [16, 17]. No evidence was obtained for the presence of the Entner-Doudoroff pathway, which agrees with the bioinformatical prediction [4].

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 (Table 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* harbors two PP$_i$-PFK, both of which were detected to be expressed constitutively in our proteomic study (data only shown for RB10591, Fig. 1).

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 2). This finding is in agreement with a normalized codon usage analysis according to Karlin and Mrázek [18], 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 2-D DIGE gels.

3.2 Carbohydrate-specific regulation of protein composition

3.2.1 Introductory remarks

The principle capacity of *R. baltica* to regulate expression of catabolic genes in response to substrate adaptation [12] 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 2-D DIGE technology (Fig. 2 and Table 2). The genomic context of regulated genes was analyzed (Fig. 3) on the basis of the recently published genome sequence of *R. baltica* [4]. The results from the adaptation experiments are presented and discussed in the following paragraphs.

3.2.2 Ribose

Among the mapped 1962 detected protein spots in the five parallel 2-D 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 2).

Table 1. Activities of enzymes from the central routes of carbohydrate degradation in substrate-adapted cells of *R. baltica* [a]

Growth substrate	Enzyme activity U/mg					
	PFK	GAPDH	Enolase	Transaldolase	Isocitrate DH	Malate DH
Ribose	0.263	0.058	0.068	0.037	0.054	0.273
Xylose	0.276	0.027	0.070	0.026	0.072	0.330
Glucose	0.254	0.045	0.092	0.030	0.043	0.193
NAG	0.307	0.047	0.040	0.058	0.139	0.997
Lactose	0.285	0.046	0.045	0.027	0.045	0.293
Maltose	0.269	0.022	0.050	0.031	0.053	0.281
Melibiose	0.248	0.020	0.015	0.023	0.029	0.208
Raffinose	0.254	0.022	0.024	0.028	0.052	0.305

a) Cultures of *R. baltica* were adapted to growth with the listed carbohydrate substrates over at least five passages
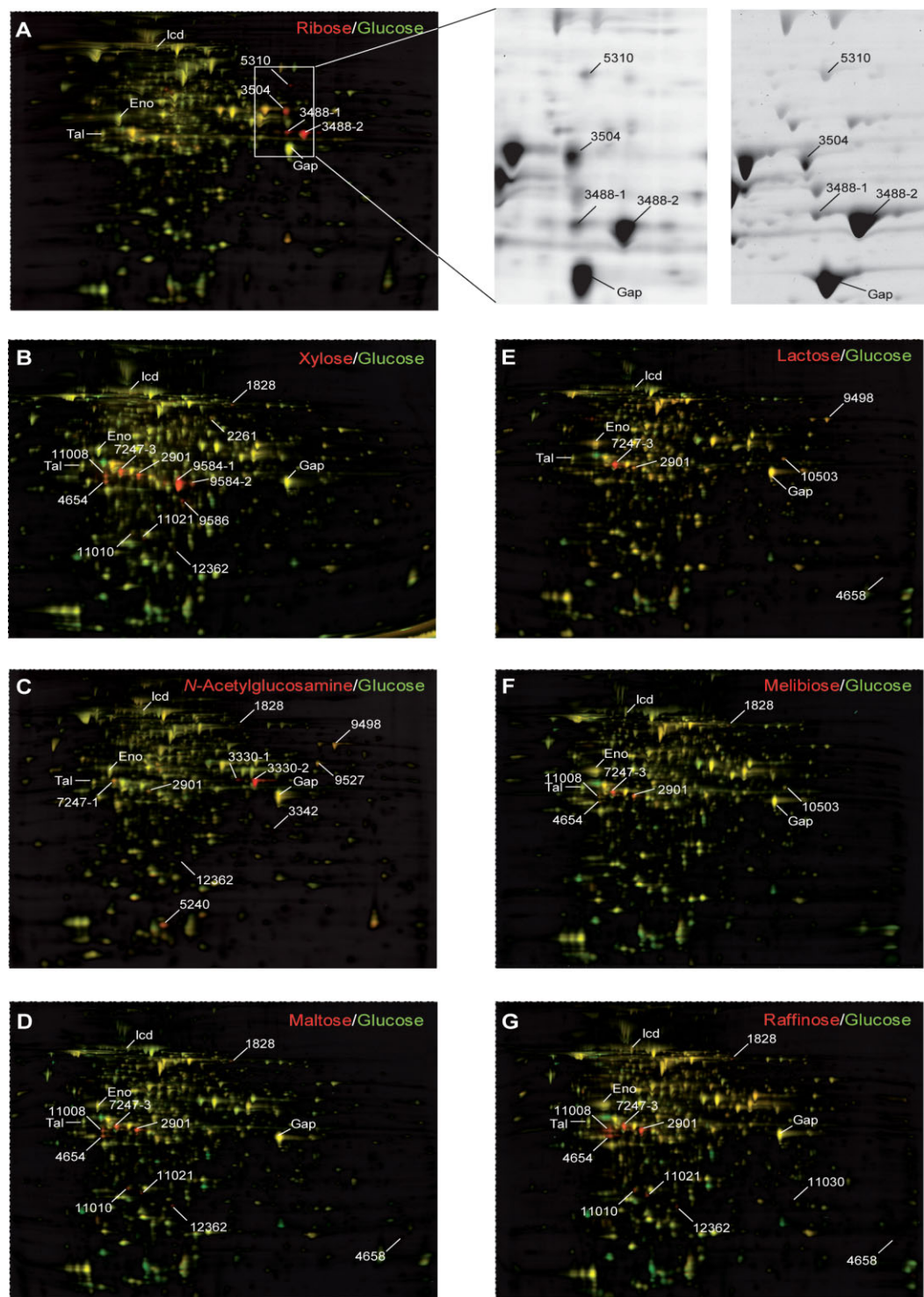
Figure 2. Overlay images of 2-D DIGE gels. Protein spots appearing in red are up-regulated when cells were grown with ribose (A), xylose (B), *N*-acetylglucosamine (C), maltose (D), lactose (E), melibiose (F), and raffinose (G), respectively. In contrast, spots appearing in green are more abundant when 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 two gel sections in the top row correspond to the boxed area in the ribose gel (A). The left section shows the Cy3-image and the right section represents a cCBB-stained ribose gel, from which the spots were excised for MS identification.

Table 2. Abundance of proteins identified as being highly up-regulated in cells grown with one of the listed carbohydrates. Glucose-grown cells served as reference state. Coding genes and putative functions based on BLASTP analysis are displayed

RB	Putative function	sp[a]	Average ratio[b]						
			Raffinose	Melibiose	Lactose	Maltose	NAG[c]	Xylose	Ribose
1828	Hypothetical protein	+	2.92	2.71		3.72	1.9	3.92	
2261	Carboxypeptidase	−						2.00	
2901	Hypothetical protein[d]	+	8.27	4.10	2.66	7.46	1.9	4.58	
3330-1	Dehydrogenase[e]	−					15.7		
3330-2							43.9		
3342	Conserved hypothetical protein	−					32.4		
3488-1	Zn-containing ADH (sorbitol DH)	−							12.47
3488-2		−							43.39
3504	Hypothetical protein[f]	−							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	Hypothetical protein[f]	+							23.63
7247-1	Glutamine synthetase II	−					1.8		
7247-3	Glutamine synthetase II	−	3.02	2.79	5.07	2.84		2.85	
9498	Arylsulfatase	+			2.05		1.8		
9527	Hypothetical protein	+					1.7		
9584-1	Zn-dependent quinone oxidoreductase	−						19.78	
9584-2		−						4.23	
9586	Oxidoreductase	−						13.21	
10503	NADH-dependent oxidoreductase[e]	−		2.09	2.31				
11008	Hypothetical protein[d]	+	8.95	2.45		8.10		4.27	
11010	Conserved hypothetical protein	+	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	6-Phosphofructokinase	−	−1.23	−1.00	1.16	1.01	−1.31	−1.27	−1.17
2627	Glyceraldehyde-3-phosphate 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

a) sp: signal peptide prediction
b) On the basis of five parallel gels changes in protein abundance were regarded as significant when the average ratio was higher than 2 and had a *t*-test value lower than 10^{-5}
c) Data were taken from [11]
d) RB2901 and RB11008 are paralogs
e) RB3330 and RB10503 are paralogs
f) RB3504 and RB5310 are paralogs

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 post-translational modification only affects the p*I* (Fig. 1A). 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, 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 (Fig. 3A). The order of these genes is only slightly different from those in *E. coli* [19]. Although the theoretical M_r

and p*I* suggest that transporter subunits or ribokinase could in principle be resolved under the applied electrophoretic conditions, they could not be identified in this study. The genomic context of *RB5310* gives no hint to a possible function of its product in ribose catabolism.

3.2.3 Xylose

The 2-D DIGE master gel of the xylose experiment contained 1950 protein spots. Thirty-eight of these proteins were found to be more abundant during growth with xylose and twelve of these proteins were identified by PMF (Table 2).
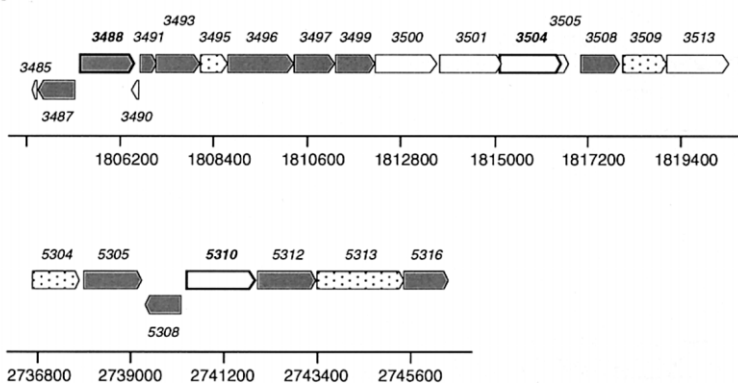
Figure 3. Genomic context of carbohydrate induced genes in *R. baltica*. Color coding: *thick lines*, up-regulated ORF; *grey*, ORF of proteins with assigned putative functions; *dotted*, ORF of conserved hypothetical proteins; *white*, ORF of hypothetical proteins. Putative functional assignments for predicted ORF in displayed genome sections 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*.
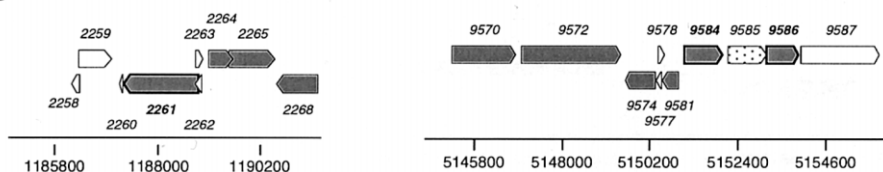
(B) Genes induced by xylose. 2261, carboxypeptidase-related protein; 2264, large-conductance mechanosensitive channel; 2265, lipase/esterase; 2268, peptide methione sulfoxide reductase, *msrB*; 9570, hydrogenase-4 transcriptional activator, *hyfR*; 9572, L-sorbosone dehydrogenase; 9574, 3-oxoacyl reductase, *fabG*; 9581, cinnamoyl ester hydrolase; 9584, Zn-containing oxidoreductase; 9586, oxidoreductase.
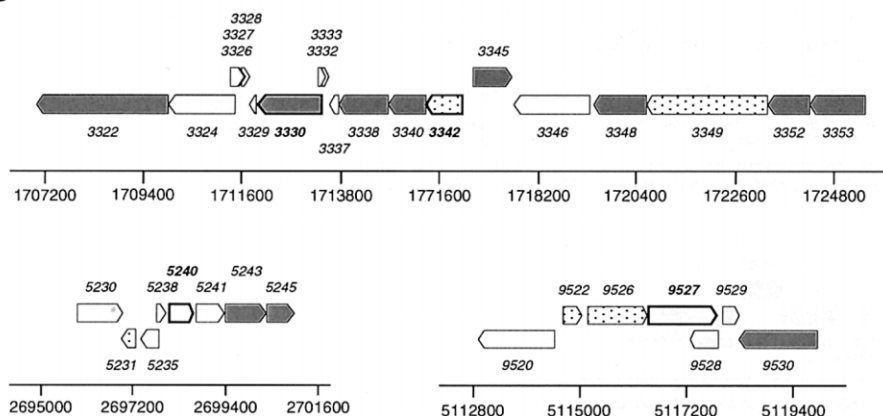
(C) Genes induced by *N*-acetylglucosamine. 3322, cytochrome *c*; 3330, dehydrogenase; 3338, myo-inositol-2-dehydrogenase, *idh*; 3340, glucosamine-6-phosphate isomerase 2, *nagB*; 3345, activator protein MtlR; 3348, *N*-acetylglucosamine-2-epimerase; 3352, *N*-acetylneuraminate lyase, *nanA*; 3353, sialidase precursor; 5243, endo-1,4-beta-xylanase B, *xynB*; 5245, hemolysin III; 9530, arylsulfatase.

D
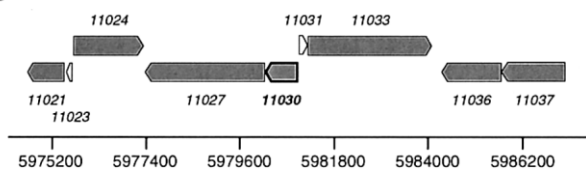
(D) Genes induced by raffinose. 11021, ABC transport system ATP-binding protein; 11024, enterochelin esterase, *fes*; 11027, ABC transport system integral membrane protein; 11030, ABC transport system ATP-binding protein; 11033, ABC transport system integral membrane protein; 11036, nitrogen regulation protein NR(I), *ntrC*; 11037, nitrogen regulation protein, *ntrB*.

E
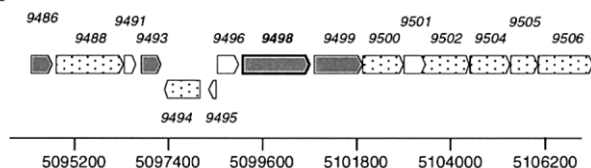
(E) Genes induced by two different carbohydrates. 9486, hypothetical 161 kDa transcriptional regulator; 9493, acetyltransferase; 9498, arylsulfatase; 9499, MoxR-related protein, *moxR*; 10488, sensory transduction histidine kinase; 10491, two-component system regulatory protein, *cheY*; 10500, phosphoglycerate kinase, *pgk*; 10503, NADH-dependent oxidoreductase; 10507, alpha-glucosidase.

F
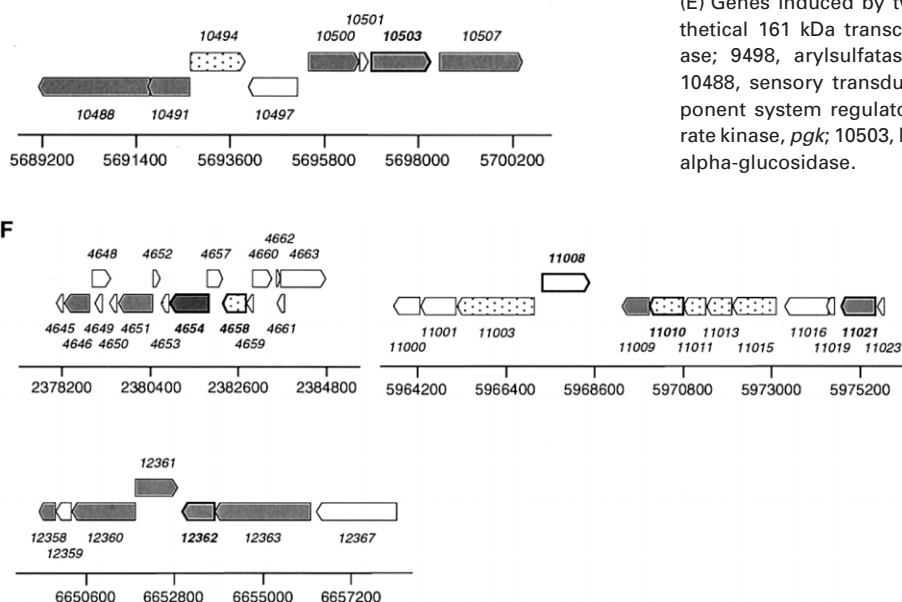
(F) Genes induced by three or four different carbohydrates. 4646, cytochrome *c* oxidase, subunit III, *cox3*; 4651, thiamine biosynthesis protein, *thiG*; 4654, sugar phosphate isomerase/epimerase; 11009, iron-uptake factor, *piuB*; 11021, ABC transport system ATP-binding protein; 12358, flagellar motor switch protein (fragment), *fliN*; 12360, levanase precursor; 12361, sugar kinase; 12362, peroxiredoxin; 12363, ATP-dependent DNA helicase, *pcrA*.

G
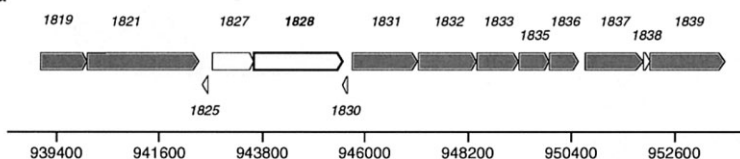
(G) Genes induced by five different carbohydrates. 1819, UDP-glucose-4-epimerase, *galE*; 1821, serine/threonine-protein kinase, *pknB*; 1831, Na$^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrA*; 1832, Na$^+$-translocating NADH:ubiquinone oxidoreductase subunit, *nqrB*; 1833, Na$^+$-translocating 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*.
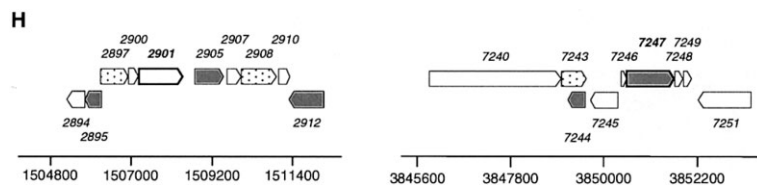
(H) Genes induced by six of the seven different carbohydrates. 2895, biopolymer transport protein, *exbD*; 2905, signal peptidase-like protein; 2912, lipoprotein; 7244, peptidyl-prolyl isomerase, *ppiB*; 7247, glutamine synthetase II, *glnII*.

The genome of *R. baltica* revealed the existence of a low affinity xylose transporter (*xylE*; *RB10309*) [20] 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 indicated 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, suggesting the presence of 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*. The protein with the second highest average ratio (13) is 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 the xylose catabolism of *R. baltica*.

The xylose isomerase (XylA) is constitutively formed 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.

3.2.4 *N*-acetylglucosamine

The 2-DIGE results for quantitative changes in protein abundances of this experiment (Table 2) 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 [11]. 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 [12]. The gene product shows high similarity to a dehydrogenase and forms two spots on the gel which slightly differ in their p*I*. Another *N*-acetylglucosamine specific protein was identified as product of *RB3342* and found to be of unknown function. *RB3330* and *RB3342* are in close proximity to *nagB* (glucosamine-6-phosphate deaminase; Fig. 3C). 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* [21], whereas a PTS-transport system for *N*-acetylglucosamine could not be detected. Unlike in *E. coli*, the *nag* genes do not cluster in a single operon-like structure; in fact, they 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, expression of this gene results in three different protein spots on the gel. As mentioned above one of them RB7247–1 is up-regulated when the cells were grown with *N*-acetylglucosamine. RB7247-3 is up-regulated under all other tested growth conditions (ribose being the only exception) while RB7247-2 (not shown in Fig. 2) does not seem to change its abundance at all.

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 (Fig. 3C).

3.2.5 Maltose

From the 1950 spots that were detected on the five parallel 2-D DIGE gels of the maltose experiment, 15 proteins were up-regulated. Nine of them were detected on cCBB-stained gels and identified by MS analysis (Table 2).

Many bacteria have uptake systems for disaccharides, *e.g.*, *B. subtilis* and *E. coli* [19, 22] 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 hydrolyze the disaccharide less economically with an amylomaltase encoded by *RB4161*, which however has not been 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.

3.2.6 Lactose

Among the mapped 2111 detected protein spots in the 2-D 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 2).

The typical *lac* operon observed in, *e.g.*, *E. coli* [19] 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 [2]. Thus 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. Those among them that are also abundant during growth with other carbohydrates, *e.g.* maltose or xylose, probably serve other functions.

3.2.7 Melibiose

In the five parallel 2-D DIGE gels of the melibiose experiment 1809 spots were detected, 10 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 2).

Although no genes for melibiose transport and degradation were predicted in the genome, this carbohydrate 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 contexts of their coding genes give 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.

3.2.8 Raffinose

From the 1925 detected proteins in the 2-D 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 (Table 2).

As for melibiose, all genes generally known to be needed for the initial step of raffinose degradation are not predicted in the genome of *R. baltica*. Two subunits of ABC-transporters (RB11021 and RB11030) are highly abundant. While RB11030 seems to be specific for raffinose, RB11021 is also present in cells grown with maltose and xylose. Interestingly, these two proteins share 53% sequence similarity.

3.2.9 General observations

In contrast to cells grown with monosaccharides, cells grown with di- and trisaccharides exhibited less pronounced differential proteomes 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 2, Fig. 3). In fact, up-regulation of proteins is usually induced by two or more different di- or trisaccharides (Fig. 3E–H). One hypothetical protein (RB2901) is up-regulated under all tested growth conditions with the only exception being ribose (Table 2, Fig. 3H). 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 2).

4   Concluding remarks

With respect to the genomic context, two groups of regulated genes could be distinguished: first, 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* [4]. 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 [23]. The proteomic approach used in this study revealed that proteins, the genes of which are not organized in clusters, are functionally related and co-ordinately regulated in *R. baltica*. 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.

5   References

[1] Rullkötter, J. in:Schulz, H. D., Zabel, M.. (Eds.), *Marine Geochemistry*, Springer, Berlin 1999, pp. 129–172.

[2] Schlesner, H., Rendsmann, C., Tindall, B. J., Gade, D. *et al.*, *Int. J. Syst. Evol. Microbiol.* 2004, *54*, 1567–1580.

[3] DeLong, E. F., Franks, D. G., Alldredge, A. L., *Limnol. Oceanogr.* 1993, *38*, 924–934.

[4] Glöckner, F. O., Kube, M., Bauer, M., Teeling, H. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 8298–8303.

[5] Görg, A., Obermaier, C., Boguth, G., Harder, A. *et al.*, *Electrophoresis* 2000, *21*, 1037–1053.

[6] Karas, M., Hillenkamp, F., *Anal. Chem.* 1988, *60*, 2299–2301.

[7] Mann, M., Højrup, P., Roepstorff, P., *Biol. Mass Spectrom.* 1993, *22*, 338–345.

[8] Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 91–98.

[9] Alban, A., David, S. O., Bjorkesten, L., Andersson, C. *et al.*, *Proteomics* 2003, *3*, 36–44.

[10] Knowles, M. R., Cervino, S., Skynner, H. A., Hunt, S. P. *et al.*, *Proteomics* 2003, *3*, 1162–1171.

[11] Gade, D., Thiermann, J., Markowsky, D., Rabus, R., *J. Mol. Microbiol. Biotechnol.* 2003, *5*, 240–251.

[12] Rabus, R., Gade, D., Helbig, R., Bauer, M. *et al.*, *Proteomics* 2002, *2*, 649–655.

[13] Bradford, M. M., *Anal. Biochem.* 1976, *72*, 248–254.

[14] Doherty, N. S., Littman, B. H., Reilly, K., Swindell A. C. *et al.*, *Electrophoresis* 1998, *19*, 355–363.

[15] Rose, M., Botstein, D., *Methods Enzymol.* 1983, *101*, 167–180.

[16] Büttner, K., Bernhardt, J., Scharf, C., Schmid, R. *et al.*, *Electrophoresis* 2001, *22*, 2908–2935.

[17] Liao, X., Ying, T., Wang, H., Wang, J., Shi, Z. *et al.*, *Electrophoresis* 2003, *24*, 2864–2882.

[18] Karlin, S., Mrázek, J., *J. Bacteriol.* 2000, *182*, 5238–5250.

[19] Lin, E. C. C., in: Neidhardt, F. C. (Ed.), *Escherichia coli and Salmonella*, American Society for Microbiology Press, Washington DC 1996, pp. 307–342.

[20] Davis, E. O., Henderson, P. J. F., *J. Biol. Chem.* 1987, *262*, 13928–13932.

[21] Plumridge, J. A., *Mol. Microbiol.* 1989, *3*, 505–515.

[22] Tangney, M., Buchanan, C. J., Priest, F. G., Mitchell, W. J., *FEMS Microbiol. Lett.* 1997, *76*, 191–196.

[23] Ermolaeva, M. D., White, O., Salzberg, S. L., *Nucleic Acids Res.* 2001, *29*, 1216–1221.