Label Free Quantitation, with Biological and Technical Replicates, of a *Rhodopseudomonas palustris* Strain, which Produces Copious Quantities of Hydrogen Gas

Brian Erickson[2]; M. Thompson[2]; N.C. VerBerkmoes[1]; M. Shah[1]; D. Pelletier[1]; S. Romagnoli[3]; B. Zhang[1]; R. Laguna[3]; R.L. Hettich[1]; F.W. Larimer[1]; R.F. Tabita[3]
[1]Oak Ridge National Lab, Oak Ridge, TN; [2]Genome Science and Technology UT/ORNL, Oak Ridge, TN; [3]Department of Microbiology, The Ohio State University, Columbus, OH

OVERVIEW

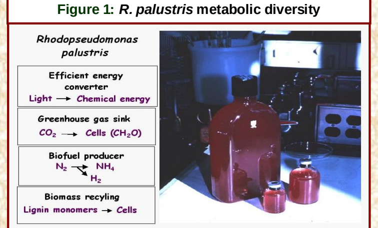
- Label free quantification has become an potential alternative to stable isotope labeling when growth media cannot be precisely controlled.
- The variation within technical replication and biological replication has not been well studied in label free "shotgun" proteomics.
- *Rhodopseudomonas palustris* is a purple nonsulfur anoxygenic phototrophic bacterium that is ubiquitous in soil and water samples and has great potential for biofuel production.
- We have created a genetic altered strain which can produce copious quantities of hydrogen gas under normal environmental conditions.
- The biological goal was to determine up and down-regulation of proteins important to this new mutant and its potential for hydrogen gas production.
- Our technical goal was to compare the variability between instrumental technical replicates and the true biological variability in living cell cultures.

INTRODUCTION

Rhodopseudomonas palustris is a purple nonsulfur anoxygenic phototrophic bacterium that is ubiquitous in the environment. *R. palustris* is of great interest due to its high metabolic diversity. While many bacterium are metabolically versatile, *R. palustris* is unique in its ability to catalyze more cellular processes than probably any known living organism (Figure 1). Furthermore, *R. palustris* is capable of producing hydrogen gas making it a potential biofuel producer and can act as a greenhouse gas sink by converting CO2 into cells. The genome of this microbe had been completed and annotated (Larimer et al, Nature Biotech, 2004).

We have created a mutant variant (strain 2044) which is de-repressed for nitrogenase biosynthesis in the presence of ammonia, enabling copious quantities of hydrogen gas to be produced. The wild-type strain was incapable of producing hydrogen under similar conditions.

Figure 1: *R. palustris* metabolic diversity



METHODS

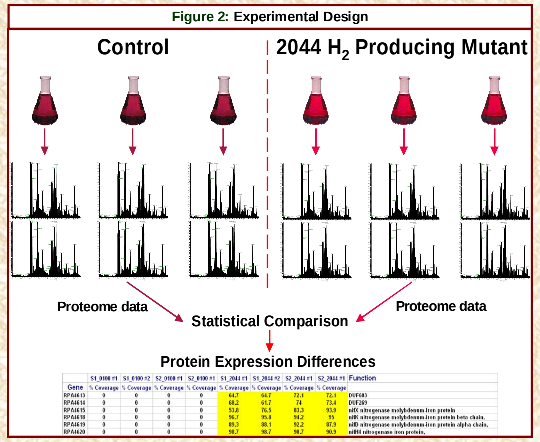
Cell Growth and Production of Protein Fractions:

- Wild-type and 2044 mutant *R. palustris* cells were grown in biological triplicates under anaerobic photoheterotrophic growth. Cells were grown in light to mid-log phase in defined mineral medium at 30° C with ammonium sulfate and succinate as fixed nitrogen and carbon sources.
- Cells were harvested, washed twice with Tris buffer, and disrupted with sonication. Two crude protein fractions were created by ultracentrifugation (100,000g for 1 hour creates membrane and crude fraction). Protein concentrations were determined by BCA and an equivalent 3mg was used for each digest. Protein fractions were denatured, reduced and digested with sequencing grade trypsin.

LC-MS/MS Analysis and Database searching:

- Triplicate biological replicates and duplicate technical replicates were performed on a 2-dimensional linear ion trap mass spectrometer (LTQ, Thermo Finnigan) (Figure 2). This analysis used a "shotgun" proteomics approach via a two-dimensional (2D) nano-LC MS/MS system with a split-phase column (RP-SCX-RP).
- Columns were packed as follows: approximately 3cm of SCX material (Luna SCX 5µm 100A Phenomenex) was first packed into a 100µm fused silica via a pressure cell followed by 3cm of C-18 RP material (Aqua C-18 5µm 200A Phenomenex, Torrance, CA). 50uL of each sample was then loaded off-line onto the dual phase column. The RP-SCX column was then positioned on the instrument behind a ~12cm c18 RP column (Aqua C-18 5µm 200A Phenomenex) also packed by pressure cell into Pico Frit tip (75µm with 15µm tip New Objective, Woburn, MA) positioned directly in the nanospray source on a LTQ (nanospray voltage 2.8kV). The samples were analyzed via a 24-hour MudPIT analysis detailed in Washburn; et al. Nature Biotech. 2001.
- For all LC/MS/MS data acquisition, the LTQ was operated in the data dependent mode with dynamic exclusion enabled, where the top five peaks in every full MS scan were subjected to MS/MS analysis. Two microscans were averaged for full scans and MS/MS scans. Dynamic exclusion was enabled with a repeat count of 1.
- From the genomic dataset, we created a database of 4,836 proteins that was used to identify MS/MS spectra. All MS/MS spectra from the LTQ datasets were searched with the SEQUEST algorithm. and filtered with DTASelect at the peptide level [Xcorrs of at least 1.8 (+1), 2.5 (+2) 3.5 (+3) were used in all cases]. Results of all replicate runs were compared with the Contrast program and evaluated based on matching of one peptide, or two or more peptides per protein. Only proteins identified with two fully tryptic peptides were considered for further biological considerations.

Figure 2: Experimental Design

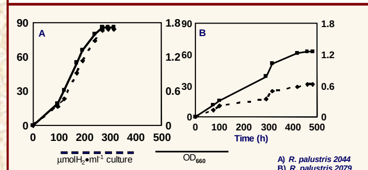


RESULTS

Hydrogen-Producing Mutant Strain of *R. palustris* (Strain 2044)

- If one disrupts the Calvin-Benson-Bassham redox pathway; e.g., by knocking out the RubisCO genes, one can select for strains that MUST induce other means to balance the redox potential.
- One way that the cell may adapt is to derepress nitrogenase, and then use this enzyme as a hydrogenase to remove excess reducing equivalents emanating from the oxidation of organic carbon.
- Nitrogenase-derepressed strains selected from strains compromised or unable to use CO_2 as electron acceptor (RubisCO knockouts) have now been isolated from *Rhodopseudomonas palustris*. Such strains use H+ as electron acceptor and produce H_2 via the hydrogenase activity of nitrogenase, which is derepressed in the presence of ammonia.
- *R. palustris* (strain 2044) is particularly active in H_2 production (Figure 3). To gain a better perspective on the consequences, and eventually elucidate the molecular basis for this switch in metabolism, proteomics studies were conducted with *R. palustris* strain 2044

Figure 3: Hydrogen production of nitrogenase –derepressed* strains of *R. palustris* (all cultures grown photoheterotrophically in the presence of 30 mM ammonia)



OD_{660}	*R. palustris 2044* (*cbbM*⁻)	*R. palustris 2079* (*cbbM⁻ cbbLS⁻*)
	µmolH2 · ml-1 culture	
0.5	23.4	12.1
1.0	46.8	24.1
1.5	70.3	28.8 (OD = 1.2)

Courtesy of Rick Laguna, OSU

Global Results

- *R. palustris* (strain 2044) and WT (designated 0100) were both grown in biological triplicates under identical photoheterotrophic conditions.
- Cells were lysed, fractionated and proteins were digested with trypsin and resulting peptide mixtures analyzed in technical duplicates by 2D-LC-MS/MS.
- Table 1 illustrates the proteins identification, peptide identifications and spectral count from each biological replicate and technical replicate. All results are for at least two unique fully tryptic peptides per protein.
- Table 2 illustrates the reproducibility of each technical replicate in overlap between proteins identifications at the 2 peptide level per protein.

Table 1: Proteins Identified by Sample

Sample	Protein IDs		Peptide IDs		Spectral Count	
	Sequest 1-peptide	Sequest 2-peptide	Sequest 1-peptide	Sequest 2-peptide	Sequest 1-peptide	Sequest 2-peptide
Sample1 0100 Run1	2463	1881	19869	19291	43582	42931
Sample1 0100 Run2	2397	1795	17870	17274	40054	39358
Sample2 0100 Run1	2331	1742	19148	18561	40839	40127
Sample2 0100 Run2	2442	1863	19170	18593	41420	40687
Sample3 0100 Run1	2469	1898	19476	18907	48917	48160
Sample3 0100 Run2	2517	1872	18931	18289	43348	42543
Sample1 2044 Run1	2443	1815	19936	19317	41069	40341
Sample1 2044 Run2	2395	1804	19429	18842	36373	35706
Sample2 2044 Run1	2481	1895	20126	19544	41189	40469
Sample2 2044 Run2	2478	1890	20619	20039	47644	46934
Sample3 2044 Run1	2637	2061	20654	20084	45200	44473
Sample3 2044 Run2	2647	2050	20133	19540	46609	45893
Average	**2475.00**	**1880.50**	**19613.42**	**19023.42**	**43020.33**	**42301.83**
STD	**92.02**	**94.84**	**784.36**	**791.85**	**3588.72**	**3574.56**
Total Protein Identifications	**3814**	**2850**				

Table 2: Reproducibility between replicates

Sample	2-Peptide Filter
Sample1 0100	79.80%
Sample2 0100	74.60%
Sample3 0100	80.40%
Sample1 2044	78.10%
Sample2 2044	76.90%
Sample3 2044	80.20%

Manual Analyses Of Large Scale Differences

- The 2044 strain was compared to the control (0100 strain) by manual analyses across all biological replicates and technical replicates.
- We required a reproducible change of at least 40% sequence coverage or 5 unique peptides as well as a 2x difference in spectral count to call a protein as differentially expressed.
- Manual analyses validated 60 proteins up-regulated in the mutant and 49 proteins down-regulated with reproducible differences in every biological and technical replicate.
- Table 3 highlights some of the large scale differences, proteins highlighted yellow are known to be involved in nitrogen fixation and thus hydrogen production. Proteins highlighted red are unknowns. Proteins highlighted green is the hydrogenase complex The protein highlighted blue was knocked-out. Numbers in table are percent coverage.

Automated Analyses Of Large Scale Differences

- Label-free LC-MS/MS "shotgun" proteomics was developed for protein profiling and has been proposed recently for quantitative studies.
- Label free quantitation methods must use other means than direct labeling to infer up and down regulation of proteins. Intrinsic values such as peptide count (number of identified peptides), spectral count (number of MS/MS spectra obtained from a protein), and percent sequence coverage (total percentage of the protein sequence covered by tryptic peptides) and peak area (area under the peak for each eluting peptide of a protein) have been used as measures of abundance.
- Previous studies of label free quantitation have suggested that spectral counts are the most reproducible and reliable measurement of protein abundance (Liu et al. Anal Chem., 2004 and Zhang et al 2006 submitted).
- Table 4 illustrates the spectral counts for some proteins from Table 3 from the 2044 mutant, for these proteins virtually no spectra were found from the control.

Table 3

Down regulated in 2044 mutant strain

Gene	% Coverage												Function
	S1 0100 #1	S1 0100 #2	S2 0100 #1	S2 0100 #2	S3 0100 #1	S3 0100 #2	S1 2044 #1	S1 2044 #2	S2 2044 #1	S2 2044 #2	S3 2044 #1	S3 2044 #2	
RPA0139	39.3	40.7	42.7	49.2	50.1	44.9	2.6	0	9.4	8	7.6	5.2	Mmethyl-accepting chemotaxis receptor/sensory transducer
RPA0143	66.1	45.5	82.6	62	59.5	64.5	0	0	0	0	0	0	cheY1 response regulator receiver, CheY1
RPA0584	27.2	40.2	43.1	42.2	31.4	45.4	0	0	0	0	0	0	Transcriptional accessory protein
RPA0677	13.9	22.2	8.9	19.7	15.1	20.9	0	0	6.4	0	0	0	Hypothetical protein
RPA1496	20.7	24.2	3.3	16	32.6	23.4	0	0	0	0	0	0	Possible monooxygenase
RPA2096	19.7	21.6	26.5	18.9	26.5	30	0	0	0	0	0	0	cobQ2 cobyric acid synthase
RPA2627	36.1	28.3	19.1	26.2	17	26.7	0	0	0	0	0	0	Putative carboxylesterase
RPA3577	31.1	24.3	27.3	23.6	26.4	28.1	0	0	6	0	10.1	6.2	thiC thiamin biosynthesis protein thiC1
RPA3978	34.5	47.8	45.7	42.6	43.9	52	14.1	0	0	0	0	0	wpbE putative aminotransferase (aminosugar biosynthesis)
RPA4580	57.3	29	14	21.2	46.6	24.8	0	0	0	0	11	0	Uncharacterized protein family UPF0065:Tat pathway signal
RPA4596	20.2	41.6	42.8	33.7	29.6	32.9	0	0	0	0	0	0	fabG5 putative 3-oxoacyl-acyl carrier protein reductase
RPA4642	75.3	81.4	88.1	83.1	85.9	85.3	0	0	0	0	0	0	cbbA fructose-bisphosphate aldolase

Up regulated in 2044 mutant strain

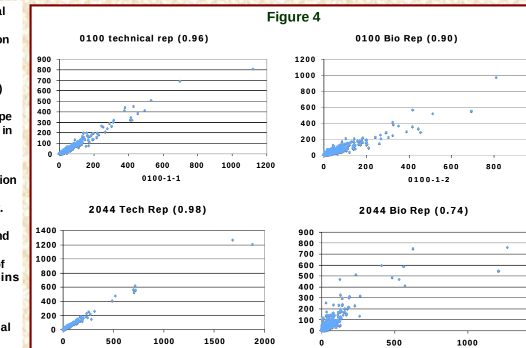
Gene	% Coverage												Function
	S1 0100 #1	S1 0100 #2	S2 0100 #1	S2 0100 #2	S3 0100 #1	S3 0100 #2	S1 2044 #1	S1 2044 #2	S2 2044 #1	S2 2044 #2	S3 2044 #1	S3 2044 #2	
RPA0005	0	0	0	0	0	0	24.2	52.2	45.4	45.2	40.3	42.5	hpd 4-hydroxyphenylpyruvate dioxygenase
RPA0274	16.1	0	28.6	0	0	0	84.8	83	89.3	89.3	73.2	96.4	glnK2 GlnK, nitrogen regulatory protein P-II
RPA0761	0	0	0	0	0	0	54.1	46.5	37.8	28.2	37.8	42.3	Possible oligopeptide ABC transporter,
RPA0762	0	0	0	0	0	0	16.7	17.9	7.1	8.8	16	18	Possible oligopeptide ABC transporter
RPA0962	0	0	0	0	0	0	31.4	32.2	46.8	37.6	55.1	45.9	hupS hydrogenase small chain
RPA0963	8.4	0	0	0	0	0	44.9	45.1	62	62	63.3	59.6	hupL hydrogenase large chain
RPA0969	0	0	0	0	14.6	9.7	19.1	19.1	13.2	44.4	51.7	57.3	hupH hydrogenase expression/formation protein
RPA1274	0	0	0	0	0	16	39.9	23.9	45.4	48.5	46.4	41.3	Possible Dps protein family starvation-inducible DNA-binding protein
RPA1717	0	0	0	0	0	0	85.2	70.3	67.2	39.1	54.7	47.7	Hypothetical protein
RPA1927	0	0	0	0	0	0	69.6	69.6	59.3	70.1	66	66	Hypothetical protein
RPA1928	0	0	0	0	0	0	95.7	93	93	94	91.7	94	Ferredoxin-like protein [2Fe-2S]
RPA2123	0	17.5	5.4	39	34.1	15.7	83	97.3	84.8	81.6	84.8	85.2	Conserved unknown protein
RPA3308	0	0	0	0	0	0	58	53.8	55	54.4	34.9	48.5	ycfI, putative structural proteins
RPA3669	0	0	0	0	0	0	53.4	50.2	27.7	50.7	65.5	55.9	putative urea short-chain amide uptake ABC transporter periplasmic
RPA3773	0	0	0	0	52.4	0	90.5	86.9	85.7	73.8	76.2	73.8	Conserved unknown protein
RPA3943	0	0	0	0	0	0	26.2	28.6	47.6	55.4	33.3	45.2	Conserved hypothetical protein
RPA4128	0	0	35	0	0	0	46.2	45	76.2	35	41.2	26.2	Conserved hypothetical proteinn
RPA4137	22.4	0	0	0	10.5	0	85.3	72	77.6	81.8	60.1	59.4	Conserved unknown proteinn
RPA4602	0	0	0	0	0	0	75.5	45.9	43.9	43.9	43.9	64.3	fixX ferredoxin like protein
RPA4603	0	0	0	0	0	0	82.8	75.6	76.6	79.5	84.6	86	fixC nitrogen fixation protein
RPA4604	0	0	0	0	0	0	67.9	69	75.5	77.2	70.4	71.5	fixB electron transfer flavoprotein alpha chain protein
RPA4605	0	0	0	0	0	0	89	89.3	88.6	91.1	87.5	87.5	fixA electron trnasfer flavoprotein beta chain
RPA4608	0	0	0	0	0	0	18.3	32.4	22.9	25.4	26.3	41.2	nifS2 nitrogenase cofactor synthesis protein
RPA4612	0	0	0	0	0	0	65	65	63.1	80.6	64.1	70.9	fdxB ferredoxin 2[4Fe-4S] III
RPA4613	0	0	0	0	0	0	64.7	64.7	72.1	72.1	64.7	64.7	DUF683
RPA4614	0	0	0	0	0	0	68.2	61.7	74	73.4	60.4	65.6	DUF269
RPA4615	0	0	0	0	0	0	53.8	76.5	83.3	93.9	61.4	87.9	nifX nitrogenase molybdenum-iron protein
RPA4618	0	0	0	0	0	0	96.7	95.8	94.2	95	96.7	96.7	nifK nitrogenase molybdenum-iron protein beta chain
RPA4619	0	0	0	0	0	0	89.3	88.1	92.2	87.9	87.5	88.5	nifD nitrogenase molybdenum-iron protein alpha chain
RPA4620	0	0	0	0	0	0	98.7	98.7	98.7	90.9	95	83.6	nifH4 nitrogenase iron protein
RPA4621	0	0	0	0	0	0	27.7	28	20.7	26.3	30.3	36	Conserved hypothetical protein
RPA4623	0	0	0	0	0	0	86.4	86.4	87.9	97	86.4	87.9	conserved hypothetical protein
RPA4626	0	0	0	0	0	0	41.4	34.1	23.4	38.7	25.3	42.5	Hypothetical Protein
RPA4631	0	0	0	0	0	0	15.6	15.6	93.8	93.8	93.8	56.2	fer1 ferredoxin 2[4Fe-4S].
RPA4633	0	0	0	0	0	0	31.9	30.5	23.5	26.8	23.8	24.5	fixR2 short-chain dehydrogenase

Table 4

Gene #	Controls	S1 2044 #1	S1 2044 #2	S2 2044 #1	S2 2044 #1	S3 2044 #1	S3 2044 #2	Average	Std Dev.
RPA0274	0	51	40	45	59	26	45	44.33	11.09
RPA0962	0	17	21	84	75	220	164	96.83	80.56
RPA0963	0	131	135	181	197	122	135	150.17	30.87
RPA0969	0	5	7	2	8	14	14	8.33	4.84
RPA1927	0	75	52	45	42	127	131	78.67	40.69
RPA1928	0	700	569	415	583	400	483	525.00	114.38
RPA3943	0	4	4	5	7	8	8	6.00	1.90
RPA4128	0	3	7	12	3	4	2	5.17	3.76
RPA4137	0	19	24	27	26	10	13	19.83	7.08
RPA4602	0	105	96	34	69	26	56	64.33	32.05
RPA4603	0	182	185	302	342	468	455	322.33	125.04
RPA4604	0	103	78	241	309	373	470	262.33	153.21
RPA4605	0	137	141	245	280	439	529	295.17	159.31
RPA4608	0	11	14	8	8	13	18	12.00	3.85
RPA4612	0	20	30	9	16	19	19	18.83	6.79
RPA4613	0	15	10	21	17	23	22	18.00	4.98
RPA4614	0	40	34	27	33	26	30	31.67	5.16
RPA4615	0	14	16	27	44	31	43	29.17	12.83
RPA4618	0	1880	1212	549	599	945	1161	1057.67	488.37
RPA4619	0	1686	1273	765	780	798	848	1025.00	376.54
RPA4620	0	489	410	599	631	1109	887	687.50	262.64
RPA4621	0	14	10	7	10	10	18	11.50	3.89
RPA4623	0	24	27	16	24	26	32	24.83	5.23
RPA4626	0	13	8	3	8	8	8	8.00	3.16
RPA4631	0	9	7	9	18	9	11	10.50	3.89
RPA4633	0	21	18	9	13	10	8	13.17	5.27

- We created scatter plots based on spectral count for all technical replicate pairs and biological replicate pairs. Linear regression was calculated for all the pairs.
- Technical replicates showed very high reproducibility in both the wild type (0100) and mutant (2044). Biological replicates showed high reproducibility in the wild type (0100), while relatively low reproducibility in the mutant (2044) (Figure 4 shows four representative plots) .
- We used the G test with William's correction to identify proteins that are differentially expressed in the wild type and the mutant. As we are testing 2841 proteins simultaneously, we used the Benjamini and Hochberg correction for the multiple test adjustment. At the False Discovery Rate of 0.01, we identified 511 and 561 proteins that are up or down regulated in the mutant, respectively.
- Almost all proteins identified by the manual method were also identified by the automated method.

Figure 4



CONCLUSION

- We have completed a large scale "shotgun" proteomics study of a hydrogen producing mutant of *R. palustris*.
- This study included biological triplicates with technical replicates.
- We analyzed the differences between the control and mutant proteome by manual and automated methods.
- Both identified strong up-regulation of proteins involved in nitrogen fixation and hydrogen production as expected as well as many unknown proteins.
- The automated method found many more differentially expressed proteins than the manual method but these should be manually verified.

ACKNOWLEDGMENTS

- This research was supported by grants from the DOE Microbial Genome Program and DOE Genomes to Life Program.
- We thank Dr. David Tabb and the Yates Proteomics Laboratory at Scripps Research Institute for DTASelect/Contrast software, the Institute for Systems Biology for proteome bioinformatics tools used in analysis of the MS data, and Dr. Frank Larimer at ORNL for providing computational resources.