Proteome Analysis of the Anoxygenic Photobacterium *Rhodopseudomonas palustris*

Nathan C. VerBerkmoes[1,3], Michael B. Strader[1], Patricia Lankford[2], Dale Pelletier[2], Loren Hauser[2,3],Mariam Land[2], Gregory B. Hurst[1], Steve Kennel[2], Caroline S. Harwood[4], Robert L. Hettich[1], Frank W. Larimer[2]

1) Chemical Sciences Division, Oak Ridge National Laboratory 2) Life Sciences Division, Oak Ridge National Laboratory 3)Graduate School of Genome Science and Technology UT-ORNL 4) Department of Microbiology, University of Iowa

Overview:

Purpose:

1. Develop an analytical method for rapid, simple and reproducible analysis of bacterial proteomes by "shotgun" LC-MS/MS. The goal is analysis of entire bacterial proteome on a single mass spectrometer in 3-5 days time.
2. Implement this methodology for a initial proteome analysis of R. *palustris,* grown under aerobic heterotrophic and anaerobic photoheterotrophic conditions.
3. Use initial proteome data to create initial set of Tandem Affinity Purification targets for large scale analysis of protein complexes from *R. palustris.*

Method:

- *R. palustris* CGA009 wild-type strain was grown to mid-log and/or stationary phase under photoheterotrophic or chemoheterotrophic conditions. LhaA (light harvesting apparatus assembly protein) mutant was grown under chemoheterotrophic conditions to mid-log phase.
- Four crude protein fractions were created from each growth state by ultracentrifugation (100,000g for 1 hour and then for 24 hours).
- All protein fractions digested trypsin and anaerobic membrane fractions were digested with CNBr/trypsin or Proteinase K.
- Protein fractions were analyzed by fully automated 1D LC-MS/MS with an ES-ion trap employing multiple mass range scanning.
- The resultant MS/MS spectra from each fraction were searched with SEQUEST against all predicted ORFs from *R. palustris.* The raw output files were filtered and sorted with DTASelect and growth states were compared with Contrast.

Results:

- Replicate proteome analysis of four crude fractions from both photoheterotrophic and chemoheterotrophic growth of the WT strain at mid-log phase. Over 120 proteins identified as showing significant change between the two states.
- Single analyses of the WT strain in stationary phase and a single analysis of the LhaA mutant under chemoheterotrophic growth at mid-log phase.
- The entire proteome analysis resulted in the overall identification of 2,633 proteins with liberal filtering constraints and 1,551 proteins with conservative filtering constraints representing 54.4% (liberal) and 32.0% (conservative) of the predicted proteome from genome annotation.
- Currently identified over 400 targets for tandem affinity purifications (TAP), representing all COGS functional categories and these genes as well as many others are currently in the process of being genetically tagged for large-scale affinity purification and MS analysis.

Introduction:

Rhodopseudomonas palustris **(Figure #1)** is a purple nonsulfur anoxygenic phototrophic bacterium that is ubiquitous in soil and water samples. *R. palustris* is of great interest due to its high metabolic diversity and ability to degrade simple aromatic hydrocarbons (lignin monomers). 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 #2)**. Specifically, this microbe is capable of photoheterotrophic and photoautotrophic growth, as well as chemoheterotrophic and chemoautotrophic growth. 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 annotation is near completion (Larimer et al, in preparation). Currently, there are ~4836 potential protein encoding genes in a 5.459Mb genome. Furthermore, *R. palustris* has multiple gene copies for some important protein components in these diverse metabolic states. We are currently employing "shotgun" proteomic techniques for analysis of the protein component of WT and various mutant strains of this microbe under photoheterotrophic and chemoheterotrophic growth.

Figure #1

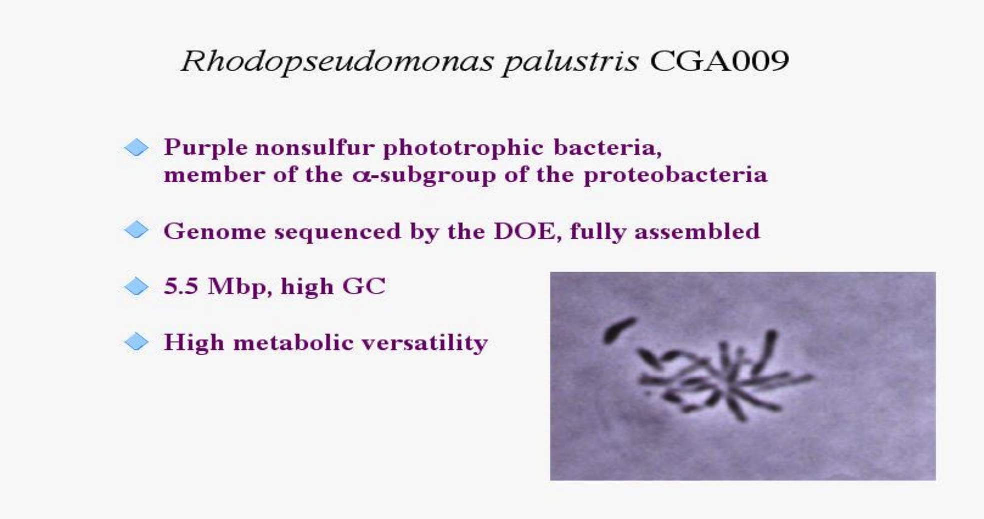
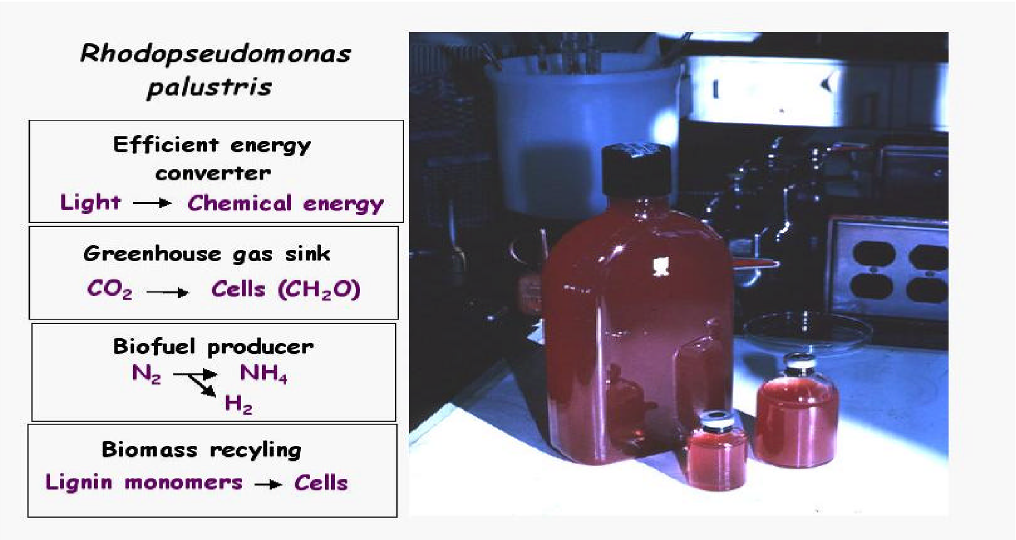


Figure #2



Experimental:

Cell Growth and Production of Protein Fractions:

- *R. palustris* CGA009 wild-type strain was grown anaerobically (photoheterotrophic growth) in light, and the WT stain and a LhaA (light harvesting apparatus assembly protein) mutant were grown aerobically (chemoheterotrophic growth) with shaking, in defined mineral medium at 30°C(Kim, M.-K. & Harwood, C. S.; FEMS Microbiol. Lett. 1991) to mid-log or stationary phase. 10mM succinate was provided as a carbon source.
- Cells were harvested, washed twice with buffer, and disrupted with sonication. Four crude protein fractions were created by ultracentrifugation (100,000g for 1 hour and then for 24 hours).
- Protein fractions were denatured, reduced and digested with sequencing grade trypsin. In separate analysis photoheterotrophic membrane fractions were digested with CNBr/trypsin or Proteinase K (Wu, C.C.; et al, Nature Biotech. 2003).

LC-MS/MS Analysis and Database searching:

- One-dimensional LC-MS/MS experiments were performed with an Ultimate HPLC (LC Packings, a division of Dionex, San Francisco, CA) coupled to an LCQ-DECA or LCQ-DECA XP ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with an electrospray source. Injections were made with a Famos (LC Packings) autosampler onto a 50ul loop. Flow rate was ~4ul/min with a 240min gradient for each run.
- A VYDAC 218MS5.325 (Grace-Vydac, Hesperia, CA) C18 column (300µm id x 15cm, 300Å with 5µm particles) or a VYDAC 238EV5.325 monomeric C18 (300µm id x 15cm, 300Å with 5µm particles) was directly connected to the Finnigan electrospray source with 100µm id fused silica.
- For all 1D LC/MS/MS data acquisition, the LCQ was operated in the data dependent mode with dynamic exclusion enabled, where the top four peaks in every full MS scan were subjected to MS/MS analysis.
- To increase dynamic range separate injections were made with a total of 4 or 5 separate m/z ranges scanned.
- The resultant MS/MS spectra from each fraction were searched with SEQUEST against all predicted ORFs from *R. palustris*. The raw output files were filtered and sorted with DTASelect and growth states were compared with Contrast (Tabb, D.L.; et al, Journal of Proteome Research, 2003).

Results

Initial *R. palustris* Proteome analysis to date:

- *R. palustris* WT and LhaA mutant were cultured under a variety of growth conditions.
- Cells were lysed by sonication and four crude fractions were prepared by centrifugation (membrane fraction, crude soluble fraction, pellet fraction and cleared soluble fraction).
- All four fractions were analyzed by shotgun proteomics with automated 1D-LC-MS/MS.
- In attempt to identify previously unidentified but known proteins involved in membrane associated complexes a CNBr/trypsin digest and Proteinase K digest was preformed on anaerobic photoheterotrophic membrane fraction (**see attached handouts for results**).
- Mid-log phase aerobic and anaerobic cells were run in duplicate to verify large scale differences, while the LhaA mutant, aerobic stationary phase and membrane digest were analyzed once.
- ~2 months of LC-MS/MS instrument time on a LCQ-DECA and DECA XPs (Thermo Finnigan).
- Over 200 individual 4.5 hour LC-MS/MS runs resulting in over 1.7 million MS/MS spectra
- ~194,600 peptides passed filters [Xcorrs of at least 1.8 (+1) 2.5 (+2) 3.5 (+3)].
- 2,633 unique protein identified with liberal filters **(Table #1).**
- 1,551 unique proteins identified with conservative filters **(Table #1).**

TABLE #1: Rhodopseudomonas palustris Proteome Analysis

Growth Condition:	# IDs liberal filters*	# IDs Conservative filters**
WT Chemoheterotrophic Stationary phase Run 1	1332	949
WT Chemoheterotrophic mid-log phase Run 1	1351	940
WT Chemoheterotrophic mid-log phase Run 2	1224	798
WT Photoheterotrophic mid-log phase Run 1	1267	885
WT Photoheterotrophic mid-log phase Run 2	1245	881
LhaA mutant Chemoheterotrophic mid-log phase Run 1	1288	923
WT Photoheterotrophic membrane CNBr/Trypsin Run 1	1307	521
WT Photoheterotrophic membrane Proteinase K Run 1	410	158
Total Proteins Identified	**9424**	**6055**
Total Non-redundant Proteins Identified	**2633**	**1551**

*Liberal Filters: At least 1 peptide per gene, Xcorrs of at least 1.8 (+1), 2.5 (+2) 3.5 (+3)
**Conservative Filters: At least 2 peptides per gene, Xcorrs of at least 1.8 (+1), 2.5 (+2) 3.5 (+3)

Gene duplication in *R. Palustris*:

- Genome annotation of *R. palustris* suggest the presence of many homologous genes for critical metabolic functions such as protein folding (two copies of GroEL and GroES), energy production (two homologous NADH dehydrogenase complexes), four sets of genes for terminal oxidases, as well as form I and form II Rubisco for carbon dioxide fixation and it has genes for three nitrogenase enzymes.
- Due to its inherent accuracy at the molecular level proteome analysis by Mass Spectrometry has the unique ability to identify which genes are actually being expressed even if sequence homology is fairly high.
- For example we have identified that *R. palustris* expresses both copies of GroEL and GroES under all growth conditions monitored to date.
- Evidence from affinity purification suggest all four copies form into a single protein complex (Hurst el al, MPQ 315)

GroEL and GroES Gene Duplication Identified by Proteome Analysis

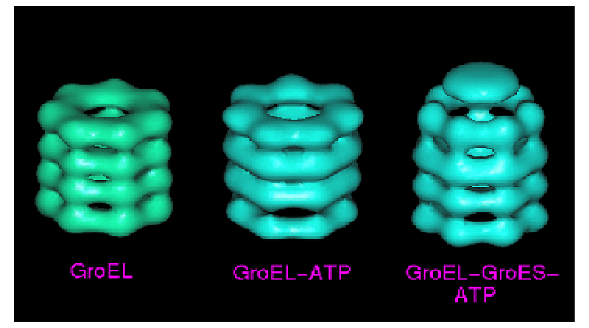
Gene 5426 GroEL 1
91.6% Coverage, 59 Unique Peptides

Gene 4709 GroEL 2
89.6% Coverage, 65 Unique Peptides

Gene 5425 GroES 1
67.3% Coverage, 10 Unique Peptides

Gene 4708 GroES 2
93.3% Coverage, 18 Unique Peptides

GroEL and GroES



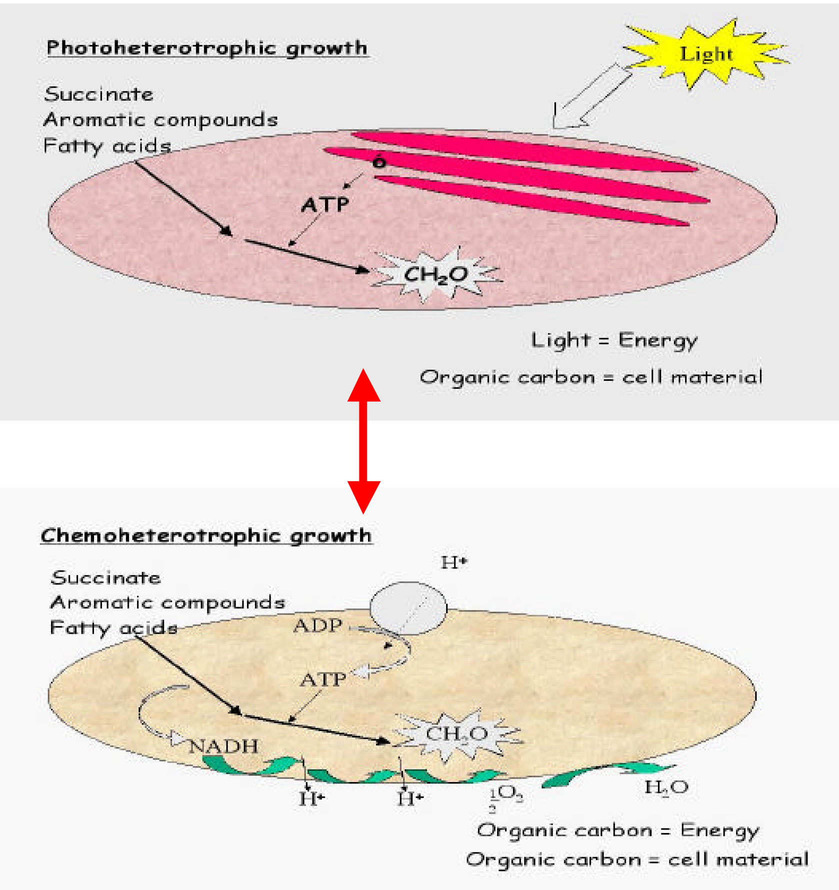
14 copies GroEL (~57kDa each)
7 copies GroES (~10kDa each)

Image obtained from
http://bioc02.uthscsa.edu/~seale/Chap/chap.html
The Chaperonin Home Page
Helen Saibil

Initial comparision between aerobic and anaerobic states:

- For the *R. palustris* microbial cell project the photoheterotrophic anaerobic growth state has been designated the standard culture.
- An initial comparision was made between aerobic chemoheterotrophic growth and anaerobic photoheterotrophic growth by a qualitative comparision between duplicate analysis of each growth state **(Figure #3).**
- Proteins that showed a replicate change of at least 30% sequence coverage and/or 4 or more unique peptides between the two states were highlighted.
- Initial analysis of these proteins indicated changes in over 120 protein species between the two states (some highlighted in **Table #2**).
- Many of these were in expected metabolic pathways (**see highlights below**).
- Further analysis will include a true quantitative comparision by N-15 labeling as well as a direct comparision between membrane fractions from each growth state digested with CNBr/trypsin.

Figure #3



Highlighted Major Differences Between Aerobic and Anaerobic *R. palustris* Proteome
(Based on Sequence Coverage and # Peptides Observed)

Aerobic UP
- -Cytochrome c oxidase
- -Ethanol dehydrogenase (quinohemeprotein)
- -Iron uptake systems
- -Branched-chain amino acid transport systems
- -Flagellar proteins
- -12 Conserved Hypotheticals and 5 Pure Hypotheticals

Anaerobic UP
- -Photosynthesis apparatus
- -Reductive TCA cycle
- -Reductive pentose phosphate pathway
- -Hypothetical 6 gene operon(or4566-4571)
- -Ribonucleotide reductase
- -13 Conserved Hypotheticals and 12 Pure Hypotheticals

Table #2
Consist of Aerobic #2 and #3 vs Anaerobic photosynthetic experiment #1 and #2
RULES for acceptance:
At least a 30% change in sequence coverage or 4 or more unique peptides difference

UP in Aerobic

		% Coverages			
Gene	**Function**	**Aerobic #2**	**Aerobic #3**	**Anaerobic #1**	**Anaerobic #2**
461	periplasmic iron binding protein FbpA precursor	71	56	25	28
483	conserved hypothetical protein	89	89	0	0
1554	cytochrome c oxidase subunit II	22	17	0	0
1801	hypothetical protein	31	19	4	0
2319	TrapT family, dctP subunit, C4-dicarboxylate periplasmic binding protein	55	32	10	10
2605	conserved hypothetical protein	45	45	0	0
2606	conserved hypothetical protein	44	19	0	0
2869	possible translation initiation inhibitor	42	42	0	0
3440	putative nitroreductase	47	32	0	5
3679	Branched-chain amino acid transport system substrate- binding protein	34	23	2	2
3688	putative outer membrane receptor for iron transport	28	14	1	0
3696	ABC transporter, ATP-binding protein	31	20	0	0
3697	Sugar uptake ABC transporter periplasmic solute-binding protein precursor	42	44	6	6
3926	alcohol dehydrogenase	49	44	0	0
4897	putative TonB-dependent iron siderophore receptor	16	17	0	0
4916	putative aldolase	26	23	0	4
4982	putative long-chain-fatty-acid CoA ligase	31	21	0	0
4997	Branched-chain amino acid transport system substrate-binding protein	78	67	15	38
5818	Branched-chain amino acid binding protein, ABC transport system	68	62	37	38
7191	putative cytochrome c	58	46	0	0

UP in Anaerobic

Gene	**Function**	**% Coverage**	**% Coverage**	**% Coverage**	**% Coverage**
201	hypothetical protein	0	0	70	53
530	putative two-component response regulator	0	0	23	59
877	possible+E2677 pyruvate-flavodoxin oxidoreductase	0	0	42	42
1232	putative 6-phosphofructokinase	0	0	20	12
1361	conserved hypothetical protein	0	0	36	36
1630	conserved hypothetical protein	0	0	31	34
1683	possible protocatechuate 4,5-dioxygenase small subunit (AB035121)	9	9	71	75
1861	putative 2-oxoglutarate ferredoxin oxidoreductase, alpha subunit	0	0	45	43
1863	putative 2-oxoglutarate ferredoxin oxidoreductase, gamma subunit	0	0	20	13
1896	conserved hypothetical protein	0	0	63	51
2067	hypothetical protein	0	0	72	72
2099	putative Mg chelatase subunit BchI	5	9	28	38
2132	protochlorophyllide reductase subunit BchN	0	0	31	24
3645	conserved hypothetical protein	0	0	82	80
4055	hypothetical protein	0	0	42	46
4566	**hypothetical protein (operon)**	0	0	27	27
4568	**hypothetical protein (operon)**	0	0	46	38
4569	**hypothetical protein (operon)**	0	0	26	13
4570	**hypothetical protein (operon)**	0	0	33	37
4571	**hypothetical protein (operon)**	0	0	18	16
5044	putative coproporphyrinogen oxidase III	4	2	67	63
5075	hypothetical protein	0	0	79	79
5162	hypothetical protein	0	0	50	60
5901	possible transmembrane protein	0	0	23	18
6425	fructose-1,6-bisphosphatase	2	0	28	25
6426	phosphoribulokinase (phosphopentokinase) (PRK)	16	25	53	46
6427	transketolase	25	17	68	78
6428	fructose-bisphosphate aldolase	37	51	85	85
6429	ribulose-bisphosphate carboxylase form II	7	30	53	66

Initial TAP targets generated from *R. palustris* Proteome analysis

- In the recently funded Genomes to Life Center for Molecular and Cellular Systems at ORNL, the core goal will be to build a research program for the high-throughput identification and characterization of protein complexes from primarily bacterial species.
- The initial attempts at developing methodology for this program will focus on protein complexes affinity purified with Tandem Affinity Purification (TAP) Tags from *R. palustris.* The components of these complexes with then be rapidly identified by LC-MS/MS.
- One of the initial needs for the system is a list of potential target proteins, which will be genetically tagged and serve as the initial bait proteins in the initial pull-down experiments.
- While currently the process is being done a few proteins at a time, in the near future, genetic tagging and protein purification will be done in 96-well format with robotic automation.
- Thus there will be a critical need for a large number of starting bait proteins. While these could be chosen based on relevant biological information it may make sense to choose a number of bait proteins based on established whole proteome data.
- We have observed a characteristic fractionation of known components of protein complexes into the two pellet fractions **(Figure #4).** The specific fractionation information as well as the fact that detected proteins are likely to be highly abundant allows for the generation of possible initial bait proteins, which may have a greater likely hood of initial success in TAP purifications.
- Based on fractionation data of our entire proteome set we built a subset list of proteins as potential TAP targets, that were detected by at least four unique peptides and showed preferentially fractionation to one or both of the pellet fractions **(Figure #5).**

Figure #4

Establishing TAP targets

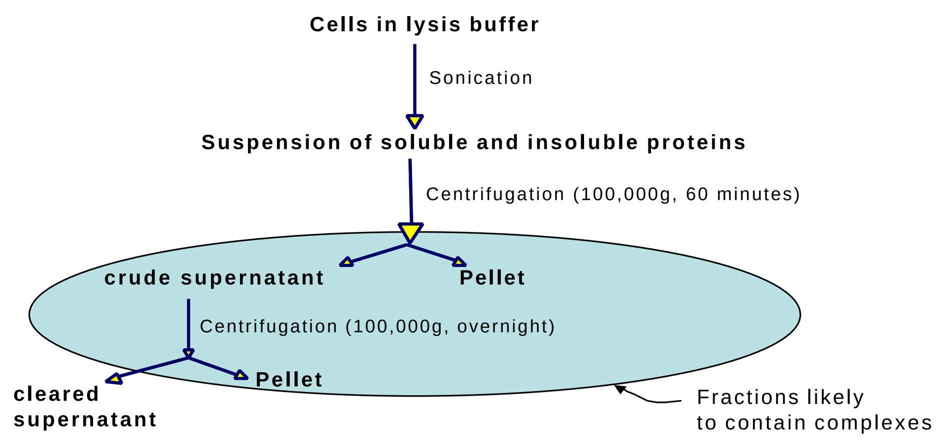
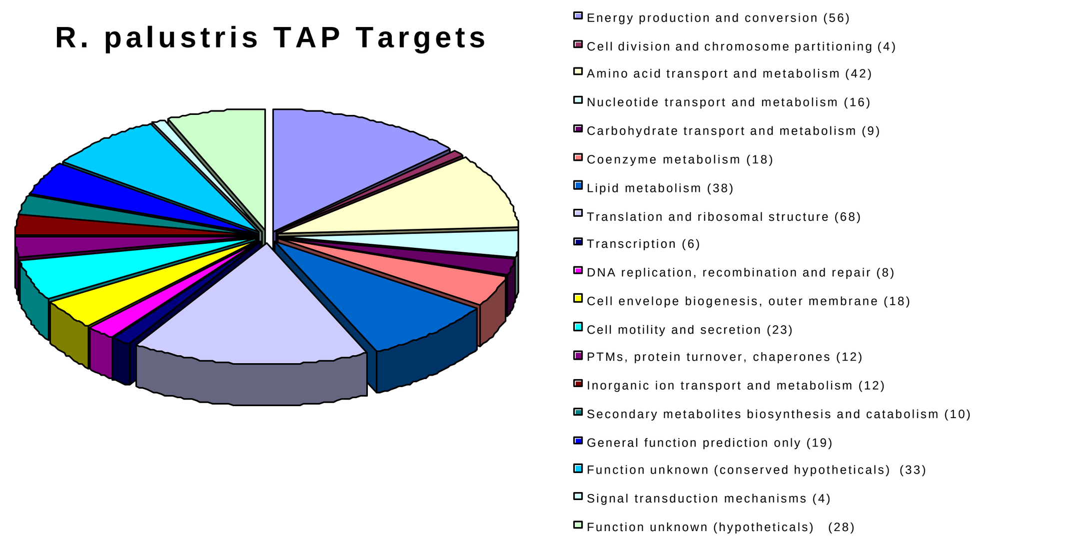


Figure #5

424 Established TAP targets



Conclusions

- Initial proteome analysis of *R. palustris* has been completed by automated 1D-LC-MS/MS resulting in the identification of 2,633 proteins with liberal filtering constraints and 1,551 proteins with conservative filtering constraints representing 54.4% (liberal) and 32.0% (conservative) of the predicted proteome.
- Initial qualitative comparision between aerobic and anaerobic growth indicated over 120 proteins showing significant change.
- Future work will include analysis of various mutants as well as nitrogen fixation and hydrogen production conditions.
- Future work will also include more accurate quantification of differences in growth states by metabolic N-15 labeling.

Research was sponsored by U.S. Department of Energy, Office of Biological and Environmental Research. Oak Ridge National Laboratory is operated and managed by University of Tennessee -Battelle, LLC. for the U.S. Department of Energy under contract DE-AC05-00OR22725.
We thank Grace-Vydac for LC columns & LC-Packings-Dionex for HPLC system