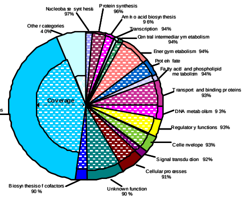
PNNL Protein Complex Characterization Efforts

Richard D. Smith, Joshua N. Adkins, Gordon A. Anderson, Deanna L. Auberry, Yuri A. Gorby, Eric A. Hill, Kim K. Hixson, Chiann-Tso Lin, Mary S. Lipton, Lye Meng Markillie, Uljana Mayer, Matthew E. Monroe, Ronald J. Moore, Sewite Negash, Ljiljana Paša-Tolić, Liang Shi, David L Springer, Thomas C. Squier, Eric F. Strittmatter, Keqi Tang, and Nikola Tolić
Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352
***Himadri B. Pakrasi**, Department of Biology, Washington University, St. Louis, MO 63130*

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

Identifying protein complexes using AMT tags

The database of *Shewanella* accurate mass and time (AMT) tags provides the basis for high throughput characterization of protein complexes at either the peptide level or intact protein level. The intact protein level analysis is enabled by the peptide level approach (by providing their initial identifications) and also complements the information obtainable at the peptide-level with additional information on protein modifications (e.g., chemical modifications, protein truncation). Regardless, the use of AMT tags can greatly speed the analysis and potentially allows characterization times of <5 minutes per complex. Proteome coverage by AMT tags is indicated by the pie chart (right).



Higher characterization throughput at lower cost

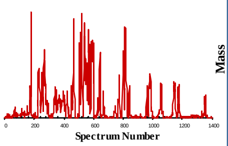
A component of the PNNL program is to develop an approach that provides both increased confidence, higher throughput, and a quantitative tool for characterizing protein complexes. We have initially explored the utility of characterizing protein complexes at the peptide level using AMT tags with Q-TOF instrumentation as an alternative to much less sensitive and lower throughput approaches based upon tandem MS (e.g., using ion trap mass spectrometers) or more expensive FTICR instrumentation that is needed for much more demanding "whole proteome" analyses.

This initial evaluation examined the highly active oxygen-evolving photosystem II (PSII) complex purified from the HT-3 strain of the cyanobacterium *Synechocystis* sp. PCC 6803. This initial study used high pressure capillary LC-Q-TOF instrumentation and used AMT tags that were generated from capillary LC-MS/MS analyses. The table below shows a partial listing of >120 proteins that were identified, along with a measure of their relative abundances based upon the integrated peak intensities for the corresponding peptides. The proteins highlighted in yellow were previously identified by Pakrasi and coworkers (Kashino et al., *Biochemistry* **2002**, *41*, 8004 – 8012). All proteins that were assigned based upon more than one peptide are done so with very high confidence.

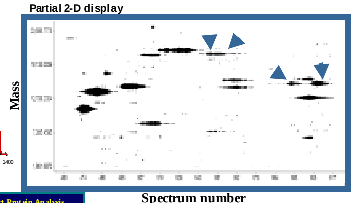
Value-added protein complex characterization at the intact protein level – initial demonstration

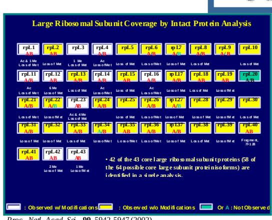
To evaluate our approach, we have initially studied the well characterized yeast large ribosomal subunit. The 43 proteins in this complex were previously identified at the peptide level using tandem MS, providing an expected set of tentative molecular weights. The constrained level of complexity associated with most protein complexes (ignoring obviously low-level contaminants) allows the detected masses to be assigned to the various proteins, as well as (in most cases) assignment of their modification states.

Capillary LC-FTICR of yeast ribosomal complex



Partial 2-D display





Proc. Natl. Acad. Sci. **99**, 5942-5947 (2002)

Advantages of intact protein proteomics:

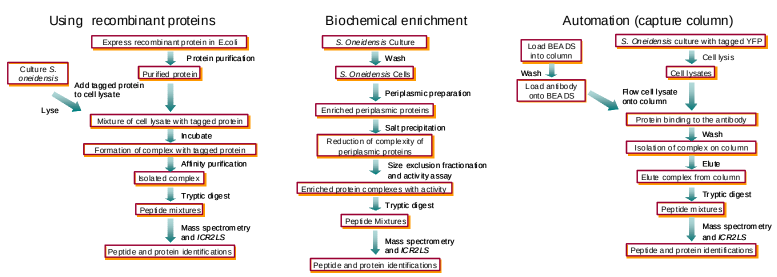
- Augments peptide level analyses
- Less complexity
- Much more information on protein modification states
- Potentially more quantitative, faster, and more sensitive

Reference	Description from FASTA	Gene	Mass Tag Count	Average MW	Approx. Abundance
SLR0906	**photosystem II core light harvesting protein**	**psbB**	26	23386	**45000**
SLL1130	**unknown protein**		7	23113	**10000**
SLL1638	**hypothetical protein**	**psbQ**	12	14362	**11000**
SLR0144	hypothetical protein		4	13678	24008
SLL0427	**photosystem II manganese-stabilizing polypeptide**	**psbO**	19	12461	**31000**
SLL0258	**cytochrome c550**	**psbV**	11	11983	**16000**
SLL1418	similar to photosystem II oxygen-evolving complex 23K protein PsbP	psbP	2	7878	16828
SLL1390	**hypothetical protein**		7	7213	**21000**
SLR1181	photosystem II D1 protein	psbA1	4	7098	32161
SLL1099	elongation factor Tu	tufA	22	6791	35472
SLL1194	**photosystem II 12 kDa extrinsic protein**	**psbU**	12	6741	**10000**
SLR0172	**hypothetical protein**	**guaB**	3	6675	**15000**
SSR0692	hypothetical protein		3	6445	4746
SLL0851	**photosystem II CP43 protein**	**psbC**	22	6267	**34000**
SLR1604	**cell division protein FtsH**	**ftsH**	21	6182	**57000**
SLL1414	**hypothetical protein**		7	6177	**24000**
SLR0012	ribulose bisphosphate carboxylase small subunit	rbcS	4	6158	10738
SLR1051	enoyl	acyl-carrier-protein	7	6001	22407
SSL0426	putative transposase	ISY100_t	1	5806	8867
SLR0228	**cell division protein FtsH**	**ftsH**	19	5646	**59000**
SML0007	**photosystem II protein Y**	**psbY**	1	5492	**3600**
SLL1867	**photosystem II D1 protein**	**psbA3**	6	5483	**27000**
SLR1311	**photosystem II D1 protein**	**psbA2**	6	5483	**27000**
SLL1398	**photosystem II reaction center W protein (Psb28b?)**	**psbW, psb13, ycf79**	5	5277	**10000**
SLR1128	hypothetical protein		14	5096	28978
SLL1029	carbon dioxide concentrating mechanism protein CcmK	ccmK	5	4995	9816
SLL0849	**photosystem II reaction center D2 protein**	**psbD**	9	4753	**29000**
SLR0927	**photosystem II reaction center D2 protein**	**psbD2**	9	4753	**29000**
SLR1622	soluble inorganic pyrophosphatase	ppa	6	4707	15482
SLR0151	unknown protein		13	4693	28320
SLL1694	pilin polypeptide PilA1	pilA1	2	4478	14254
SLR0009	ribulose bisphosphate carboxylase large subunit	rbcL	11	4337	42576
SLL1463	cell division protein FtsH	ftsH	21	4220	55317
SLL1545	glutathione S-transferase		8	4113	24143
SSL2245	unknown protein		1	4087	7467
SLL1106	hypothetical protein		5	3985	14625
SLR0670	hypothetical protein		11	3972	26504
SLL1543	hypothetical protein		1	3921	13433
SLR1645	**photosystem II 11 kD protein (Psb27 b?)**	**psbZ**	7	3781	**9100**
SLR1165	sulfate adenylyltransferase		13	3724	35442
SMR0007	**photosystem II PsbL protein**	**psbL**	1	3680	**4600**
SSL2598	**photosystem II PsbH protein**	**psbH**	3	3609	**5700**
SLR1855	unknown protein		20	3492	56780
SSR3451	**cytochrome b559 alpha subunit**	**psbE**	3	3323	**7800**
SLR0752	enolase		13	2877	37740
SLR1218	hypothetical protein YCF39	ycf39	1	2874	14339
SLL1028	carbon dioxide concentrating mechanism protein CcmK	ccmK	5	2869	9031
SLL1212	GDP-mannose 4,6-dehydratase		5	2866	33526
SLL1835	hypothetical protein		2	2852	23386
SLL0080	N-acetyl-gamma-glutamyl-phosphate reductase	argC	6	2842	30992
SLL0680	phosphate-binding periplasmic protein precursor (PBP)		10	2770	32463
SSL1498	hypothetical protein		2	2728	5191
SLR0149	hypothetical protein		5	2717	14913
SLR1619	hypothetical protein		3	2716	22510
SLL1450	nitrate/nitrite transport system substrate-binding protein	nrtA	13	2678	39717
SLL1654	hypothetical protein		1	2430	11805
SLL1043	polyribonucleotide nucleotidyltransferase		18	2414	63130
SLR0476	unknown protein		1	2405	11675
SLL1852	nucleoside diphosphate kinase		3	2297	13539
SLR1216	Mg2+ transport protein		5	2280	41448
SLL1555	two-component hybrid sensor and regulator		4	2259	33639
SLL1641	glutamate decarboxylase		4	2151	43022
SLR0723	hypothetical protein		5	2128	33603
SLR1618	unknown protein		3	2123	23523
SLL0290	polyphosphate kinase	ppk	13	2115	67003
SLL0018	fructose-bisphosphate aldolase, class II	fbaA, fda	8	2101	31610
SLL0735	hypothetical protein		5	2056	50350
SLR2033	rubredoxin	rub	1	2042	10196
SLR1984	nucleic acid-binding protein, 30S ribosomal protein S1 homolog	nbp1, rps1b	3	2000	27411
SLR1609	long-chain-fatty-acid CoA ligase		8	1909	62994
SLL1911	hypothetical protein		2	1821	7269

These results show:

- Q-TOF instrumentation augmented by the use of LC elution time information provides sufficient specificity for application of AMT tag approaches.
- The AMT tag approach with LC-Q-TOF analysis provides sufficient specificity for protein complex characterization, along with high throughput, and preliminary quantitation.
- The use of quantitative information and multiple analyses (e.g., using different wash conditions) will be needed with this approach to better qualify which proteins are part of the complex in contrast to being non-specifically associated.

Complex generation and isolation



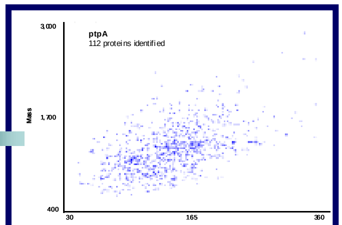
Tagged proteins generated to date for pull-down studies at PNNL

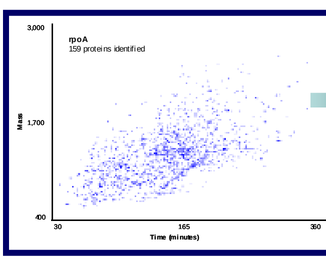
Gene	Description	Annotation
hydB	periplasmic Fe hydrogenase small subunit	SO3921
hydA	periplasmic Fe hydrogenase large subunit	SO3920
napA	periplasmic nitrate reductase	SO0848
omcA	decaheme cytochrome C	SO1779
omcB	deceheme cytochrome C	SO1778
hoxK	Quinone-reactive Ni/Fe hydrogenase small subunit precursor	SO2099
petA	ubiquinol-cytochrome C reductase iron-sulfur subunit	SO0608
	flavocytochrome C flavin subunit	SO3301
	Gfo/Idh/MocA family oxidoreductase	SO3120
	oxidireductase molybdopterin-binding	SO0715
nrfC	formate-dependent nitrite reductase	SO0483
ptpA	phosphotyrosine protein phosphatase	SO2208
ptpB	Tyrosine-specific protein phosphatase	SO3124
cpxP	Spheroplast protein y precursor	SO4476
msrA	methionine sulfoxide reductase (isoform A)	SO2337
msrB	methionine sulfoxide reductase (isoform B)	SO2588
eno	Enolase	SO3440
rnlB	ATP-dependent RNA helicase	SO0407
rpoD	RNA polymerase sigma-70 factor	SO1284
	Cytochrome c3	SO2727
rpoA	DNA-directed RNA polymerase alpha subunit	SO0256
rpoZ	DNA-directed RNA omega subunit	SO0360
hepA	RNA polymerase-associated protein	SO0575

Initial results

Initial attempt of pull-down with ptpA

Reference	Description	MW (kDa)	Average Abundance	QTOF Peptides Observed	Avg Peptide Prophet
SO3930	ribosomal protein S6	15	4860	7	0.98
SO3652	ribosomal protein L21	11	3361	4	1.00
SO0234	ribosomal protein L2	30	3221	7	1.00
SO0223	ribosomal protein L7/L12	13	3057	3	0.97
SO0217	translation elongation factor Tu	43	2228	20	0.98
SO3765	conserved hypothetical protein	25	1549	5	0.96
SO0247	ribosomal protein L18	13	1344	6	0.99
SO0255	ribosomal protein S4	23	1344	9	1.00
SO0235	ribosomal protein S19	10	1227	4	0.91
SO2087	integration host factor alpha subunit	11	1184	5	0.98
SO3927	ribosomal protein L9	16	994	7	0.94
SO0256	DNA-directed RNA polymerase alpha subunit	36	981	12	0.98
SO2402	ribosomal protein S1	61	951	11	1.00
SO1207	ribosomal protein S15	10	923	3	1.00
SO0704	chaperonin GroEL	57	922	19	0.99
SO0221	ribosomal protein L1	25	837	4	0.97
SO0257	ribosomal protein L17	15	724	3	0.95
SO3939	ribosomal protein S9	15	658	6	0.97
SO2112	ribosomal protein L25	17	561	3	0.99
SO0230	ribosomal protein S10	12	523	14	1.00
SO0222	ribosomal protein L10	18	482	8	1.00
SO2178	cytochrome c551 peroxidase	36	476	9	0.99
SO2302	ribosomal protein L20	14	475	5	1.00
SO0250	ribosomal protein L15	15	465	7	0.90
SO0227	ribosomal protein S7	18	457	9	1.00
SO3940	ribosomal protein L13	16	452	5	0.98
SO0245	ribosomal protein S8	14	408	2	0.97
SO2644	phosphoenolpyruvate synthase	87	359	9	0.98
SO1126	chaperone protein DnaK	69	343	23	0.99
SO2016	heat shock protein HtpG	72	319	8	1.00
SO3681	universal stress protein family	16	317	3	1.00
SO4410	glutamine synthetase type I	52	315	8	0.99
SO3537	ribosomal protein S20	10	305	4	1.00
SO3237	flagellin	28	303	7	0.97
SO3440	enolase	42	262	7	0.98
SO2328	translation elongation factor P	21	234	3	0.90
SO0248	ribosomal protein S5	18	225	12	0.97
SO4105	MSHA pilin protein MshA	18	223	4	0.96
SO0236	ribosomal protein L22	12	208	6	0.99
SO0432	aconitate hydratase 2	93	208	18	0.99
SO1360	ribosomal protein L19	13	205	4	0.99
SO0242	ribosomal protein L24	11	199	10	0.93
SO0243	ribosomal protein L5	20	196	7	1.00
SO2776	3-oxoacyl-(acyl-carrier-protein) reductase	26	185	4	0.96
SO0426	pyruvate dehydrogenase complex E3 component, lipoamide dehydrogenase	51	185	8	1.00
SO0254	ribosomal protein S11	14	159	6	1.00
SO1490	alcohol dehydrogenase II	40	137	7	1.00
SO1931	2-oxoglutarate dehydrogenase E2 component, dihydrolipoamide succinyltransferase	43	115	7	1.00
SO3146	DNA-binding protein H-NS family	15	101	4	0.97
SO2401	integration host factor beta subunit	11	100	4	0.97
SO1793	trigger factor	48	95	9	0.99
SO0052	protein-export protein SecB	17	88	2	0.99
SO3988	aerobic respiration control protein ArcA	27	84	3	0.98
SO0691	hypothetical protein	15	73	2	0.97
SO0237	ribosomal protein S3	26	64	8	1.00
SO4749	ATP synthase F1 alpha subunit	55	63	7	1.00
SO1524	heat shock protein GrpE	23	61	2	1.00





Initial attempt of pull-down with rpoA

Reference	Description	MW (kDa)	Average Abundance	QTOF Peptides Observed	Avg Peptide Prophet
SO3237	flagellin	28	8635	9	1.00
SO1797	DNA-binding protein HU family	9	6399	11	0.99
SO2401	integration host factor beta subunit	11	6261	5	0.99
SO3927	ribosomal protein L9	16	4804	9	0.95
SO0257	ribosomal protein L17	15	4695	5	1.00
SO0217	translation elongation factor Tu	43	4290	26	0.98
SO0235	ribosomal protein S19	10	3267	7	0.97
SO0933	fructose-bisphosphate aldolase class II, Calvin cycle subtype	38	2759	4	1.00
SO1207	ribosomal protein S15	10	2721	4	0.99
SO0247	ribosomal protein L18	13	2519	4	1.00
SO0704	chaperonin GroEL	57	2150	26	0.97
SO0221	ribosomal protein L1	25	1939	12	0.96
SO3930	ribosomal protein S6	15	1781	6	1.00
SO1126	chaperone protein DnaK	69	1425	32	1.00
SO3940	ribosomal protein L13	16	1395	4	1.00
SO0227	ribosomal protein S7	18	1375	10	1.00
SO1931	2-oxoglutarate dehydrogenase E2 component, dihydrolipoamide succinyltransferase	43	1208	8	1.00
SO0234	ribosomal protein L2	30	1169	14	1.00
SO2112	ribosomal protein L25	17	1000	5	0.98
SO1360	ribosomal protein L19	13	982	6	0.99
SO0245	ribosomal protein S8	14	953	5	1.00
SO0223	ribosomal protein L7/L12	13	911	10	0.98
SO3440	enolase	42	896	16	0.94
SO0236	ribosomal protein L22	12	886	5	1.00
SO4410	glutamine synthetase type I	52	875	8	1.00
SO0248	ribosomal protein S5	18	828	10	0.99
SO3146	DNA-binding protein H-NS family	15	823	6	1.00
SO1793	trigger factor	48	803	21	1.00
SO0230	ribosomal protein S10	12	775	14	1.00
SO2087	integration host factor alpha subunit	11	592	4	0.98
SO0231	ribosomal protein L3	23	587	12	1.00
SO1490	alcohol dehydrogenase II	40	571	9	0.98
SO2644	phosphoenolpyruvate synthase	87	481	20	0.98
SO4749	ATP synthase F1 alpha subunit	55	470	14	1.00
SO2328	translation elongation factor P	21	469	4	1.00
SO0222	ribosomal protein L10	18	459	11	1.00
SO2302	ribosomal protein L20	14	456	10	0.99
SO0426	pyruvate dehydrogenase complex E3 component, lipoamide dehydrogenase	51	452	15	1.00
SO2178	cytochrome c551 peroxidase	36	444	12	1.00
SO0432	aconitate hydratase 2	93	430	19	1.00
SO2016	heat shock protein HtpG	72	399	17	0.99
SO1524	heat shock protein GrpE	23	399	3	1.00
SO4105	MSHA pilin protein MshA	18	377	4	1.00
SO0842	translation elongation factor G	77	346	13	1.00
SO4751	ATP synthase F0B subunit	17	341	5	1.00
SO0424	pyruvate dehydrogenase complex E1 component, pyruvate dehydrogenase	99	340	12	0.98
SO3765	conserved hypothetical protein	25	325	8	1.00
SO0243	ribosomal protein L5	20	312	13	0.99
SO0253	ribosomal protein S13	13	274	6	0.99
SO0691	hypothetical protein	15	272	4	1.00
SO0256	DNA-directed RNA polymerase alpha subunit	36	235	15	0.97
SO0220	ribosomal protein L11	15	235	8	0.98
SO3681	universal stress protein family	16	221	5	1.00
SO0874	DnaK suppressor protein	17	206	2	1.00
SO2402	ribosomal protein S1	61	186	18	1.00
SO1629	ribosomal protein S2	27	181	3	0.96
SO1807	phage shock protein A	26	174	4	1.00
SO0237	ribosomal protein S3	26	147	14	1.00
SO3466	riboflavin synthase beta subunit	17	134	2	0.96
SO2345	glyceraldehyde 3-phosphate dehydrogenase	36	94	7	1.00
SO0250	ribosomal protein L15	15	79	6	1.00
SO4747	ATP synthase F1 beta subunit	50	69	7	0.99
SO0052	protein-export protein SecB	17	63	2	0.94
SO0224	DNA-directed RNA polymerase beta subunit	150	52	16	0.98