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

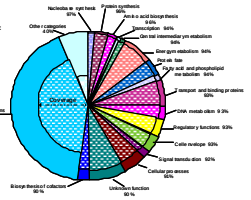
Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352

***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).



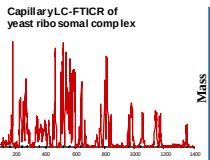
Higher characterization throughput at lower cost

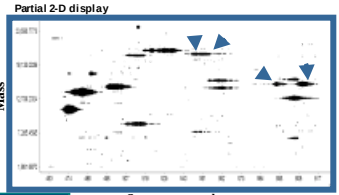
A component of the PNNL program is to develop an approach that provides both increased confidence, higher throughput, and a quantitative tool for characterizing protein complexes. We have initially explored the utility of characterizing protein complexes at the peptide level using AMT tags with Q-TOF 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.





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

Advantages of intact protein proteomics:

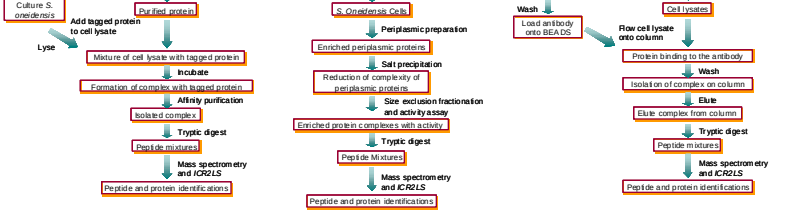
- Augments peptide level analyses
- Less complexity
- Much more information on protein modification states
- Potentially more quantitative, faster, and more sensitive

Reference	Description from FASTA	Gene	Mass Tag Count	Average Abundance	Approx. MW (Da)
Accession no.	**photosystem II core light harvesting protein**	**psbB**	26	2.33E6	**45000**
SLL1330	**unknown protein**		7	2.31E3	**10000**
SLL1638	**hypothetical protein**	**psbQ**	12	1.45E2	**11000**
SLR0144	hypothetical protein		4	1.3E78	34000
SLL0427	**photosystem II manganese-stabilizing polypeptide**	**psbO**	19	1.24E1	**31000**
SLL0528	**cytochrome c550**	**psbV**	11	1.18E3	**16000**
SLL1418	similar to photosystem II oxygen-evolving complex 23K protein PsbP	psbP	2	787.8	16828
SLL1390	**hypothetical protein**		7	723.3	**21000**
SLR1181	photosystem II D1 protein	psbA1	4	709.8	32361
SLL1099	elongation factor Tu	tufA	22	577.91	35472
SLL1194	**photosystem II 12 kDa extrinsic protein**	**psbU**	12	67.41	**10000**
SLR0172	**hypothetical protein**	**ycf48**	3	65.75	**30000**
SSR9592	hypothetical protein		3	64.45	47.46
SLL0951	**photosystem II CP43 protein**	**psbC**	22	62.67	**34000**
SLR1484	**cell division protein FtsH**	**ftsH**	21	61.82	**57000**
SLL1414	**hypothetical protein**		7	61.77	**24000**
SLR0012	ribulose bisphosphate carboxylase small subunit	rbcS	4	61.58	19758
SLR1051	enolase	acyl-carrier-protein	7	60.01	22407
SSL0428	putative transposase	ISY100_I	1	59.06	8807
SLR0228	**cell division protein FtsH**	**ftsH**	19	56.46	**59000**
SML0007	**photosystem II protein Y**	**psbY**	1	54.06	**3600**
SLL1887	**photosystem II D1 protein**	**psbA3**	6	54.03	**27000**
SLR1311	**photosystem II D1 protein**	**psbA2**	6	54.03	**27000**
SLL1398	**photosystem II reaction center W protein (Psb28/b?)**	**psbW, psb33, ycf79**	5	52.77	**10000**
SLR1128	hypothetical protein		14	50.96	28787.8
SLL1029	carbon dioxide concentrating mechanism protein CcmK	ccmK	5	49.95	9816
SLL0969	**photosystem II reaction center D2 protein**	**psbD**	9	47.53	**29000**
SLR0927	**photosystem II reaction center D2 protein**	**psbD2**	9	47.53	**29000**
SLR1622	soluble inorganic pyrophosphatase	ppa	6	47.07	25462
SLR0151	unknown protein		13	46.93	28320
SLL1894	prlin polypeptide/PhA1	psbA1	2	44.78	14254
SLR0009	ribulose bisphosphate carboxylase large subunit	rbcL	11	43.37	42976
SLL1463	cell division protein FtsH	ftsH	21	42.29	55317
SLL1545	glutathione S-transferase		8	41.13	24143
SSL2245	unknown protein		1	40.97	7467
SLL1306	hypothetical protein		5	39.85	14625
SLR0070	hypothetical protein		11	39.72	33504
SLL1543	hypothetical protein		1	39.21	13453
SLR1965	**photosystem II 11 kD protein (Psb27 b?)**	**psbZ**	7	37.81	**9100**
SLR1165	sulfate adenylyltransferase		13	37.24	35462
SMR0007	**photosystem II PsbL protein**	**psbL**	1	36.93	**4000**
SSL2598	**photosystem II Psb31 protein**	**psbH**	3	36.09	**5700**
SLR1855	unknown protein		20	34.92	56.350
SSR3451	**cytochrome b559 alpha subunit**	**psbE**	3	33.23	**7900**
SLR0752	enolase		13	28.77	37.740
SLR1218	hypothetical protein YCF39	ycf39	1	28.56	34.929
SLL1028	carbon dioxide concentrating mechanism protein CcmK	ccmK	5	28.69	9816
SLL1212	GDP-mannose 4,6-dehydratase		5	28.66	33506
SLL1183	hypothetical protein		2	28.52	23366
SLR0360	N-acetyl-gamma-glutamyl-phosphate reductase	argC	6	28.42	39962
SLL0860	phosphate-binding periplasmic protein precursor (PBP)		10	27.70	32463
SSL1498	hypothetical protein		2	27.28	5191
SLR0149	hypothetical protein		5	27.17	14913
SLR1619	hypothetical protein		3	27.16	22510
SLL1450	nitrate/nitrite transporter's system substrate-binding protein	nrtA	13	26.78	39717
SLL1854	hypothetical protein		1	24.30	11805
SLL1043	polyribonucleotide nucleotidyltransferase		18	24.14	63120
SLR0476	unknown protein		1	24.05	11675
SLL1852	nucleoside diphosphate kinase		3	22.97	13539
SLR1216	Mg2+ transport protein		5	22.80	43448
SLL1555	two-component hybrid sensor and regulator		4	22.59	33419
SLL1961	glutamate decarboxylase		4	21.51	43022
SLR0723	hypothetical protein		5	21.28	33803
SLR1618	unknown protein		3	21.23	23752.3
SLL0290	polyphosphate kinase	ppk	13	21.15	67303
SLL0116	fructose-bisphosphate aldolase, class II	fbaA, fda	8	21.01	31610
SLL0735	hypothetical protein		5	20.56	50350
SLR0203	rubredoxin	rub	1	20.42	10206
SLR1884	nucleic acid-binding protein, 30S ribosomal protein S1 homolog	nbp1, rps1b	3	20.00	27411
SLR1609	long-chain-fatty-acid CoA ligase		8	19.99	62364
SLL1161	hypothetical protein		2	18.21	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



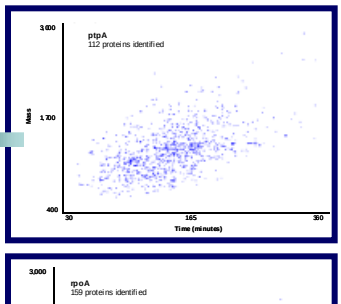
Tagged proteins generated to date for pull-down studies at PNNL

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	decaheme cytochrome C	SO1778
hcxK	Quinone-reactive NiFe hydrogenase small subunit precursor	SO2099
petA	ubiquinol-cytochrome C reductase iron-sulfur subunit	SO0608
	flavocytochrome C flavin subunit	SO3301
	GfoI/dhMocA family oxidoreductase	SO3120
	oxidoreductase 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
mfB	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 polymerase omega subunit	SO0360
hepA	RNA polymerase-associated protein	SO0575

Initial results

Initial attempt of pull-down with ptpA

Reference	Description	MW(Da)	Average Abundance	QTOF Peptides Observed	Avg Peptide Prophet
SO3830	ribosomal protein S6	35	4.86E3	2	0.98
SO3552	ribosomal protein L21	11	3.36E1	4	1.00
SO2136	ribosomal protein L2	30	3.22E1	7	1.00
SO0223	ribosomal protein L7 &L12	13	3.05E7	3	0.97
SO0217	translation elongation factor Tu	43	2.22E8	20	0.98
SO3785	conserved hypothetical protein	25	1.56E9	5	0.96
SO2047	ribosomal protein L18	13	1.36E4	6	0.99
SO2155	ribosomal protein S4	23	1.36E4	9	1.00
SO2135	ribosomal protein S10	10	1.22E7	4	0.91
SO2697	integration host factor alpha subunit	11	1.18E4	5	0.98
SO3827	ribosomal protein L9	16	994.7	7	0.94
SO2156	DNA-directed RNA polymerase alpha subunit	36	981	12	0.98
SO3658	ribosomal protein S1	61	951	11	1.00
SO1207	ribosomal protein S15	10	923	3	1.00
SO3704	chaperonin GroEL	57	922	19	0.99
SO2021	ribosomal protein L1	25	837	4	0.97
SO2157	ribosomal protein L17	15	724	3	0.95
SO3828	ribosomal protein S8	15	698	6	0.97
SO2132	ribosomal protein L25	17	561	3	0.99
SO2160	ribosomal protein S10	12	523	14	1.00
SO2122	ribosomal protein L10	18	482	8	1.00
SO2178	cytochrome c551 peroxidase	36	476	9	0.99
SO2002	ribosomal protein L20	14	475	5	1.00
SO2150	ribosomal protein L15	15	465	7	0.90
SO0227	ribosomal protein S7	18	457	9	1.00
SO3640	ribosomal protein L13	16	452	5	0.98
SO2045	ribosomal protein S8	14	408	2	0.97
SO2644	ATP synthase F1 alpha subunit	87	339	9	0.98
SO1126	chaperone protein DnaK	69	343	23	0.99
SO2016	heat shock protein HtpG	72	318	8	1.00
SO3681	universal stress protein family	16	317	3	1.00
SO4410	glutamine synthetase type I	52	315	8	0.98
SO3537	ribosomal protein S20	10	305	4	1.00
SO2237	flagellin	28	303	7	0.97
SO3442	enolase	42	282	7	0.98
SO2228	translation elongation factor P	21	234	3	0.90
SO2148	ribosomal protein S5	18	225	12	0.97
SO4105	MSHA pilin protein MshA	18	223	4	0.96
SO2036	ribosomal protein L22	12	208	6	0.99
SO0632	aconitate hydratase 2	99	208	18	0.99
SO1960	ribosomal protein L19	13	205	4	0.99
SO2042	ribosomal protein L24	11	199	10	0.93
SO2143	ribosomal protein L5	20	196	7	1.00
SO2776	3-oxoacyl(acyl-carrier-protein) reductase	26	185	4	0.96
SO4436	pyruvate dehydrogenase complex E3 component, lipoamide dehydrogenase	51	165	8	1.00
SO2154	ribosomal protein S13	14	159	6	1.00
SO1490	alcohol dehydrogenase I	40	137	7	1.00
SO3531	2-oxoglutarate dehydrogenase E2 component, dihydrolipoamide succinyltransferase	43	115	7	1.00
SO3646	DNA-binding protein HNS family	15	101	4	0.97
SO2151	elongation factor G	11	100	4	0.97
SO1793	flagellar factor	48	95	9	0.99
SO3552	protein-export protein SecB	17	88	2	0.99
SO3688	aerobic respiration control protein ArcA	27	84	3	0.98
SO3816	hypothetical protein	35	73	2	0.97
SO0237	ribosomal protein S3	26	64	8	1.00
SO4748	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(Da)	Average Abundance	QTOF Peptides Observed	Avg Peptide Prophet
SO2237	flagellin	28	8.65E5	9	1.00
SO1797	DNA-binding protein HU-family	9	6.36E9	11	0.99
SO2401	integration host factor beta subunit	11	6.2E1	5	0.99
SO3827	ribosomal protein L9	16	4.8E4	9	0.95
SO2157	ribosomal protein L17	15	4.6E5	5	1.00
SO0217	translation elongation factor Tu	43	4.2E0	26	0.98
SO3235	ribosomal protein S19	10	3.2E7	7	0.97
SO3533	succinate-semialdehyde aldolase-class II, Calvin cycle subtype	38	2.7E9	4	1.00
SO1207	ribosomal protein S15	10	2.7E1	4	0.99
SO2047	ribosomal protein L18	13	2.5E9	4	1.00
SO3704	chaperonin GroEL	57	2.5E6	26	0.97
SO2021	ribosomal protein L1	25	1.9E9	12	0.96
SO3828	ribosomal protein S8	15	1.7E1	6	1.00
SO1126	chaperone protein DnaK	69	1.4E5	32	1.00
SO3640	ribosomal protein L13	16	1.3E5	4	1.00
SO0227	ribosomal protein S7	18	1.3E5	10	1.00
SO3115	2-oxoglutarate dehydrogenase E2 component, dihydrolipoamide succinyltransferase	43	1.2E6	8	1.00
SO2036	ribosomal protein L22	30	1.1E9	14	1.00
SO2132	ribosomal protein L25	17	1.0E0	5	0.98
SO3360	ribosomal protein L19	13	9E2	6	0.99
SO2145	ribosomal protein S8	14	9E3	5	1.00
SO0223	ribosomal protein L7/L12	13	9E1	10	0.98
SO3442	enolase	42	8E6	16	0.94
SO2036	ribosomal protein L22	12	8E6	5	1.00
SO4410	glutamine synthetase type I	52	8E5	8	1.00
SO2148	ribosomal protein S5	18	8E8	10	0.99
SO3146	DNA-binding protein HNS family	15	8E3	6	1.00
SO1793	flagellar factor	48	8E3	21	1.00
SO2230	ribosomal protein S10	12	7E5	14	1.00
SO2697	integration host factor alpha subunit	11	5E2	4	0.98
SO2238	ribosomal protein L1	23	5E7	12	1.00
SO1490	alcohol dehydrogenase II	40	5E1	9	0.98
SO3537	phosphoenolpyruvate synthase	97	4E1	20	0.98
SO2748	ATP synthase F1 alpha subunit	55	4E0	14	1.00
SO2228	translation elongation factor P	21	4E9	4	1.00
SO0222	ribosomal protein L10	18	4E9	11	1.00
SO2002	ribosomal protein L20	14	4E6	10	0.99
SO4436	pyruvate dehydrogenase complex E3 component, lipoamide dehydrogenase	51	4E2	15	1.00
SO2178	cytochrome c551 peroxidase	36	4E4	12	1.00
SO4432	aconitate hydratase 2	98	4E0	19	1.00
SO2016	heat shock protein HtpG	72	3E9	17	0.99
SO1524	heat shock protein GrpE	23	3E9	3	1.00
SO4105	MSHA pilin protein MshA	18	3E7	4	1.00
SO1842	translation elongation factor G	77	3E6	13	1.00
SO4751	ATP synthase F0B subunit	17	3E1	5	1.00
SO4434	pyruvate dehydrogenase complex E1 component, pyruvate dehydrogenase	99	3E0	12	0.98
SO3785	conserved hypothetical protein	25	3E5	8	1.00
SO3243	ribosomal protein L5	20	3E2	13	0.99
SO3253	ribosomal protein S13	13	2E4	6	0.99
SO3816	hypothetical protein	35	2E2	4	1.00
SO2256	DNA-directed RNA polymerase alpha subunit	36	2E5	15	0.97
SO2230	ribosomal protein L1	15	2E5	8	0.98
SO3681	universal stress protein family	16	2E1	5	1.00
SO3674	DnaK suppressor protein	17	2E6	2	1.00
SO2402	ribosomal protein S1	61	1E6	18	1.00
SO3529	ribosomal protein S2	27	1E1	3	0.98
SO1807	phage shock protein A	26	1E4	4	1.00
SO0227	ribosomal protein S3	26	1E7	14	1.00
SO3495	thirteen synthase beta subunit	17	1E4	2	0.98
SO2345	glyceraldehyde 3-phosphate dehydrogenase	36	94	7	1.00
SO3752	ribosomal protein S5	15	79	6	1.00
SO4747	ATP synthase F1 beta subunit	50	69	7	0.99
SO3552	protein-export protein SecB	17	63	2	0.94
SO0234	DNA-directed RNA polymerase beta subunit	150	52	16	0.98