Mining Shotgun Proteomic Data for the Role of Post-Translational Modifications in the Adaptive Response of *Shewanella oneidensis* to Chromate Exposure

Melissa R. Thompson[1,2], Dorothea K. Thompson[3], and Robert L. Hettich[2]

[1]Graduate School of Genome Science and Technology, ORNL-UTK, [2]Chemical Sciences Division, Oak Ridge National Laboratory, [3]Department of Biological Sciences, Purdue University

OVERVIEW

- Protein function is often controlled by post-translational modifications (PTMs)
- PTMs provide clues to cellular activity and environmental response (i.e. chemotaxis)
- PTM-containing MS/MS spectra are present in high quality unidentified mass spectra
- Overall goal is to develop a non-biased global PTM search approach
- Two purposes of mining shotgun proteomic data for PTMs are to:
  - Determine if the number of protein identifications increase
  - Determine possible protein pathways regulated by modifications after an environmental perturbation

INTRODUCTION

- Small molecules covalently attached to amino acid residues is the category of post-translational modifications (PTMs) considered here
- PTMs have been implicated in regulating protein activity
  - Repression or stimulation of protein activity can be controlled by removal or addition of a PTM
- Some examples of PTMs are:
  - Methylation: implicated in gene expression regulation and protein-DNA interaction
    - Bacterial chemotaxis: methylation of methyl-accepting chemotaxis protein receptors (chemoreceptors) provide an adaptive mechanism for response to environmental cues
  - Acetylation: plays a role in bacterial chemotaxis through modifying CheY, enhancing activation of the flagellar switch
  - Oxidation: generally due to environmental causes
    - Highly reactive free radicals easily form during cellular metabolism
    - High abundance of protein in cytosol make them easy targets for free radical quenching
- *Shewanella oneidensis* MR-1 (a γ-proteobacteria) is very versatile in respiration and can use a variety of metal ions and small compounds as terminal electron acceptors
- *S. oneidensis* MR-1 was exposed to a sub-lethal dose of Cr(VI) and differentially expressed proteins based on semi-quantitation were determined[1]
- Data was examined further in order to identify PTMs that may play a role in Cr(VI) response in *S. oneidensis*
  - Cr(VI) reduction proceeds through a highly reactive Cr(V) species leading to the formation of free radicals which may lead to high levels of oxidized peptides
  - SO2426 (putative DNA-binding response regulator) was identified as highly up-regulated previously[1], but may be differentially modified instead
  - Predicted proteome contains 27 chemoreceptors, which may be differentially methylated in Cr(VI) exposed cells versus control growth conditions

EXPERIMENTAL DESIGN

- Experimental dataset from previous study[1]
  - *Shewanella oneidensis* MR-1 was grown aerobically in the presence or absence of 1 mM K_2CrO_4 and harvested 45 and 90 min post-shock
  - Cultures were lysed by sonication, fractionated into crude/solube and membrane fractions by high speed centrifugation, and digested with trypsin
  - The digests were separated online using two-dimensional liquid chromatography (SCX and RP) coupled to a LTQ mass spectrometer (operated in data-dependent mode for 24 hr)
- Computational searches and algorithms
  - Figure 1 is a flowchart of the bioinformatic pipeline used for PTM searches
  - DBDigger[2] was used during the initial method development due to its flexibility in handling multiple modifications simultaneously
    - Table 1 depicts the PTMs targeted here
  - Sequest[3] was used to aid in the conformation of PTM identifications
    - MS/MS spectra identified by DBDigger were extracted using DTAcopy (Vilmos Kertesz, ORNL) and searched by Sequest

Figure 1: Flowchart of the bioinformatic pipeline used to search for post-translational modifications

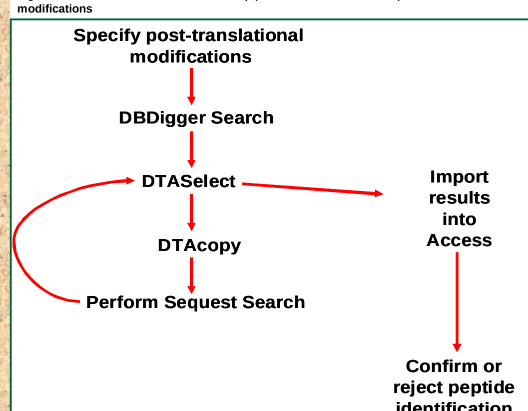

Table 1: Amino acid residues and corresponding post-translational modifications targeted

Amino Acid	PTM[a]	Amino Acid Mass	Amino Acid + PTM mass
Lysine	$+CH_2$	128.09	142.11
Lysine	$+C_2H_4$	128.09	156.12
Lysine	$+C_3H_6$	128.09	170.14
Lysine	$+COCH_2$	128.09	170.10
Lysine	$+C_2O_2C_2H_4$	128.09	210.11
Arginine	$+CH_2$	156.10	170.12
Arginine	$+C_2H_4$	156.10	184.13
Arginine	$+C_3H_6$	156.10	198.15
Arginine	$+COCH_2$	156.10	198.11
Arginine	$+C_2O_2C_2H_4$	156.10	238.12
Glutamic Acid	$+CH_2$	129.04	143.06
Methionine	+O	131.04	147.03
Methionine	$+O_2$	131.04	163.02
Cysteine	+O	103.01	119.00
Cysteine	$+O_2$	103.01	134.99
Cysteine	$+O_3$	103.01	150.99
Tryptophan	+O	186.08	202.07
Tyrosine	+O	163.06	179.05

[a] Empirical formula for post-translational modification

METHOD DEVELOPMENT

Table 2: False discovery rates for 45min control run1 (no Cr(VI) added) dataset at various filter threshold levels (DBDigger values for +1, +2, +3 ions)

No PTM	2 peptide filter level: Protein			Peptide			4 peptide filter level: Protein			Peptide		
25, 30, 45	158[a]	2111[b]	14.8%	333[c]	25943[d]	2.6%	9	1443	1.2%	17	24390	0.1%
27, 32, 47	68	1933	7.0%	136	24111	1.1%	7	1383	1.0%	8	22817	0.1%
29, 34, 49	20	1816	2.2%	37	22643	0.3%	3	1326	0.5%	3	21482	0.0%

Monomethylation	2 peptide filter level: Protein			Peptide			4 peptide filter level: Protein			Peptide		
25, 30, 45	427	2560	33.4%	1069	28112	7.6%	51	1526	6.7%	224	25660	1.7%
27, 32, 47	215	2146	20.0%	503	25388	4.0%	21	1420	3.0%	75	23685	0.6%
29, 34, 49	96	1924	10.0%	217	25330	1.5%	9	1343	1.3%	25	21952	0.2%
31, 36, 51	50	1818	5.5%	106	21764	1.0%						
35, 40, 55	15	1689	1.8%	32	19089	0.3%						

[a] Number of proteins identified from reverse database entries
[b] Total number of proteins identified in the search
[c] Number of peptides identified from reverse database entries
[d] Total number of peptides identified in the search

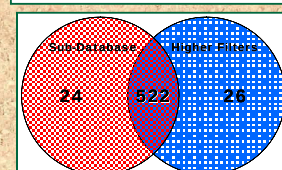

Figure 2: Venn diagram comparing peptide identifications searching for monomethylations on lysines and arginines with the 45 min control run1 dataset by two different search methods. Circle on the left is a modification search with the sub-database and the circle on the right represents a modification search with the full-size database and stringent filters. Value in the middle (522) is the number of modified peptide identifications shared between the two methods, which is a 90% overlap between them.

- 45 min control run1 dataset was used for method development
  - Table 2 and Figure 2 illustrate results using this dataset
- The Monomethylation PTM was chosen for initial testing because this PTM is believed to be common and the resulting mass shift (14 Da) is small
  - Preliminary searches (Table 2) demonstrated that monomethylation results in a high false discovery rate
    - Other PTMs did not demonstrate the high false discovery rate in the total peptide dataset (Tables 3 and 4)
- Two methods were tested for global PTM searching of shotgun proteomic data
  - Method 1: Search data with no modifications (filter threshold of 1 peptide, score of 25 for +1, 30 (+2), and 45 (+3) then create a sub-database from the identified proteins and search with modifications specified
    - MacCoss *et al*[4] performed a PTM study on yeast Cdc2p complex and human lens tissue using a sub-database, demonstrating the application of such a method
    - In our study, the database size was 2,751 forward protein sequences with matching reverse sequence (total database size 5,502 entries)
    - PTM search filter thresholds: 2 peptides, 25 (+1), 30 (+2), and 45 (+3) had false discovery rate of 7.1%; and 2 peptides 29 (+1), 34 (+2), and 49 (+3) had false discovery rate of 1.8%
  - Method 2: Search data with the full-size database and increase the filter threshold levels
    - Full-size database was 4,798 forward protein sequences with matching reverse sequences (total database size 9,596 entries)
    - PTM search filter thresholds: 2 peptides, 25 (+1), 30 (+2), and 45 (+3) had false discovery rate of 7.6%; and 2 peptides 29 (+1), 34 (+2), and 49 (+3) had false discovery rate of 1.9%
  - No difference in the false discovery rates for the two methods and the overlap of modified peptides between them is 90%; decided to use full-size database for the rest of the searches

POST-TRANSLATIONAL MODIFICATIONS RESULTS

Table 3: Total peptide identifications using DBDigger adding PTMs

Dataset	Monomethylation	FDR[a]	Monoacetylation	FDR	Oxidation	FDR
45 min Control	25703	2.1%	25392	1.7%	28138	1.2%
45 min Cr	25338	2.1%	24808	1.6%	26758	1.1%
90 min Control	23490	2.3%	23140	0.7%	26784	0.7%
90 min Cr	25338	1.6%	25209	0.7%	26717	0.9%

[a]FDR: false discovery rate

Table 4: Peptide false discovery rates of PTM peptides only

Modification	45 min Control	FDR[a]	45 min Cr	FDR	90 min Control	FDR	90 min Cr	FDR
Monomethylation	1147	53.9%	1167	62.4%	1356	45.7%	902	53.8%
Dimethylation	1058	N/A[b]	852	N/A	951	N/A	890	N/A
Trimethylation	889	N/A	769	N/A	798	N/A	720	N/A
Monoacetylation	709	62.1%	594	65.3%	619	59.0%	557	65.5%
Diacetylation	779	N/A	669	N/A	667	N/A	620	N/A
Oxidation	5567	8.0%	3940	8.6%	7068	4.2%	3209	9.3%

[a]FPR: false discovery rate [b]N/A: not available

- Table 3 depicts total peptide identifications and the false discovery rate for post-translational modification searches with the full-size forward and reverse databases
- Table 4 depicts the PTM peptides extracted from the total dataset (Table 3) and the results illustrate a caveat of post-translational modification searches
  - From the total false-positive peptides identified, most have PTMs
  - The false discovery rate is high when considering the subset of modified peptide identifications only
  - PTM search provides a putative set of PTM-containing peptides, but manual validation is required to verify the true identifications
- Tables 5 and 6 are the functional category distributions of modified peptides in the DBDigger set
- Table 7 is the Sequest search results of the MS/MS spectra first identified by DBDigger as containing at least one PTM
  - Filter thresholds: 1 peptide required for protein identification, +1 (1.8), +2 (2.5) +3 (3.5)
  - Much lower number of peptide identifications found; could be due Sequest's inflexibility with searching for PTMs

Table 5: Functional category distribution of PTM peptides identified in the 45 min dataset

Number	Functional Category	Control Mono-methylation	Cr Mono-methylation	Control Di-methylation	Cr Di-methylation	Control Tri-methylation	Cr Tri-methylation	Control Mono-acetylation	Cr Mono-acetylation	Control Di-acetylation	Cr Di-acetylation	Control oxidation	Cr oxidation
1	Amino acid biosynthesis	21	23	40	25	25	18	26	13	18	14	93	76
2	Biosynthesis of cofactors, prosthetic groups, and carriers	16	13	17	33	24	14	10	16	16	14	105	43
3	Cell envelope	80	58	40	36	35	36	25	24	38	30	383	255
4	Cellular processes	62	93	89	59	75	49	54	53	62	48	369	254
5	Central intermediary metabolism	4	3	6	4	10	3	6	0	9	7	20	17
6	DNA metabolism	44	38	74	63	55	56	40	39	63	41	102	75
7	Energy metabolism	139	132	83	64	75	57	55	56	65	63	1098	720
8	Fatty acid and phospholipid metabolism	28	23	27	13	19	19	13	7	19	12	180	97
9	Hypothetical proteins	209	202	264	173	214	182	173	129	162	164	666	458
10	Mobile and extrachromosomal element functions	9	21	23	13	23	25	19	12	17	12	17	12
11	Protein fate	91	132	53	78	54	55	43	46	65	54	585	465
12	Protein synthesis	192	201	99	73	62	76	66	61	61	43	944	823
13	Purines, pyrimidines, nucleosides, and nucleotides	31	25	27	21	20	22	25	16	17	13	170	95
14	Regulatory functions	21	18	45	35	47	39	18	26	28	21	78	45
15	Signal transduction	7	12	17	16	11	15	21	5	13	15	15	2
16	Transcription	46	44	36	28	34	22	21	22	26	22	199	118
17	Transport and binding proteins	59	78	47	43	45	28	33	26	41	38	239	207
18	Unknown function	88	51	71	75	61	53	61	43	59	58	304	178
	Total	**1147**	**1167**	**1058**	**852**	**889**	**769**	**709**	**594**	**779**	**669**	**5567**	**3940**

Table 6: Functional category distribution of PTM peptides identified in the 90 min dataset

Number	Functional Category	Control Mono-methylation	Cr Mono-methylation	Control Di-methylation	Cr Di-methylation	Control Tri-methylation	Cr Tri-methylation	Control Mono-acetylation	Cr Mono-acetylation	Control Di-acetylation	Cr Di-acetylation	Control oxidation	Cr oxidation
1	Amino acid biosynthesis	23	26	25	15	12	21	18	14	12	11	91	76
2	Biosynthesis of cofactors, prosthetic groups, and carriers	30	20	29	20	21	20	20	16	11	19	162	51
3	Cell envelope	89	50	40	46	22	27	24	22	34	17	520	152
4	Cellular processes	104	64	87	51	69	47	45	53	62	41	545	208
5	Central intermediary metabolism	1	8	6	6	5	7	3	4	4	5	25	26
6	DNA metabolism	54	39	66	62	44	43	35	30	38	51	133	73
7	Energy metabolism	187	93	83	76	79	66	60	59	58	51	1337	583
8	Fatty acid and phospholipid metabolism	35	21	16	20	11	14	19	10	11	18	267	100
9	Hypothetical proteins	201	169	203	198	169	173	125	113	125	139	947	352
10	Mobile and extrachromosomal element functions	15	17	23	17	20	21	22	15	19	10	26	8
11	Protein fate	151	82	69	61	59	54	34	36	71	53	759	368
12	Protein synthesis	220	135	76	84	62	75	56	43	46	42	998	657
13	Purines, pyrimidines, nucleosides, and nucleotides	32	16	26	17	16	17	15	10	22	13	188	87
14	Regulatory functions	26	22	33	45	51	27	28	24	21	28	106	48
15	Signal transduction	11	6	22	27	18	13	12	20	12	18	36	5
16	Transcription	49	45	44	29	33	19	15	24	29	23	244	128
17	Transport and binding proteins	55	53	36	55	43	29	31	18	40	31	302	158
18	Unknown function	73	36	67	61	64	47	57	46	52	50	382	129
	Total	**1356**	**902**	**951**	**890**	**798**	**720**	**619**	**557**	**667**	**620**	**7068**	**3209**

Table 7: Sequest search results of MS/MS spectra identified by DBDigger as containing a PTM

Dataset	Monomethylation: MS/MS count	Identified	Dimethylation: MS/MS count	Identified	Trimethylation: MS/MS count	Identified	Monoacetylation: MS/MS count	Identified	Diacetylation: MS/MS count	Identified	Oxidation: MS/MS count	Identified
45 min Control	1363	314	967	172	776	123	887	123	749	73	6044	3480
45 min Cr	1357	406	790	136	669	116	719	95	647	73	4371	2352
90 min Control	1558	456	885	159	682	111	716	78	628	74	7388	4770
90 min Cr	1059	248	808	140	628	89	680	74	580	70	3549	1896

CONFORMATION OF PTMs

Figure 4: Distribution of PTMs in SO2426 under control and Cr(VI) exposure conditions

45 min Cr Sample

MILILVWCLEMSRILLVDDDPLFRVWLTDVLKAQG
HEVKCAVNGVEGLTCIRSFAPDIVMLDLVMPQMD
GFELLQARECMTPIMMLSARDNEEDRIRSYELGA
DDFLTKPFSIKELLVRLQAIERRLVQRPKPPMAVE
QVATIQSVRFDETAYKITIGKHAVELTQTEFRLFKY
LFERKGQVITKQELQKSVLQKDLGRFDRNLDMHI
SNTRRKLANTRLPRTLINTVRGQGYSFSA

46.8%

45 min Control Sample

MILILVWCLEMSRILLVDDDPLFRVWLTDVLKAQG
HEVKCAVNGVEGLTCIRSFAPDIVMLDLVMPQMD
GFELLQARECMTPIMMLSARDNEEDRIRSYELGA
DDFLTKPFSIKELLVRLQAIERRLVQRPKPPMAVE
QVATIQSVRFDETAYKITIGKHAVELTQTEFRLFKY
LFERKGQVITKQELQKSVLQKDLGRFDRNLDMHI
SNTRRKLANTRLPRTLINTVRGQGYSFSA

24.9%

90 min Cr Sample

MILILVWCLEMSRILLVDDDPLFRVWLTDVLKAQG
HEVKCAVNGVEGLTCIRSFAPDIVMLDLVMPQMD
GFELLQARECMTPIMMLSARDNEEDRIRSYELGA
DDFLTKPFSIKELLVRLQAIERRLVQRPKPPMAVE
QVATIQSVRFDETAYKITIGKHAVELTQTEFRLFKY
LFERKGQVITKQELQKSVLQKDLGRFDRNLDMHI
SNTRRKLANTRLPRTLINTVRGQGYSFSA

48.1%

90 min Control Sample

MILILVWCLEMSRILLVDDDPLFRVWLTDVLKAQG
HEVKCAVNGVEGLTCIRSFAPDIVMLDLVMPQMD
GFELLQARECMTPIMMLSARDNEEDRIRSYELGA
DDFLTKPFSIKELLVRLQAIERRLVQRPKPPMAVE
QVATIQSVRFDETAYKITIGKHAVELTQTEFRLFKY
LFERKGQVITKQELQKSVLQKDLGRFDRNLDMHI
SNTRRKLANTRLPRTLINTVRGQGYSFSA

0.0%

KEY
Red = monomethylated
Blue = dimethylated
Green = trimethylated/ monoacetylated
Fuchsia = diacetylated
Cyan = conserved aspartate putative phosphorylation site

- SO2426 has differential modifications at the C-terminus under various growth conditions
  - The percentages (i.e. 46.8% under 45 min Cr sample) are the amount of sequence coverage identified in the previous study, which considered only unmodified peptides[1]
  - The amino acids underlined were part of peptides identified with a at least one PTM
    - The control samples appear to have more of the total protein sequence covered by PTM containing peptides, which may be why their sequence coverage was initially lower in the unmodified search
  - The C-terminus of the protein is a putative DNA-binding domain that contains many more methylated amino acids under the 45 min Cr dataset than the corresponding control sample
    - Methylation on lysines and arginines changes their point-charge, which may affect interactions with DNA

Table 7: Monomethylated Peptides in Chemoreceptors

Protein	45min Control #peptides	45min Control #spectra	90min Control #peptides	90min Control #spectra	45min Cr #peptides	45min Cr #spectra	90min Cr #peptides	90min Cr #spectra
SO0584	2	2	2	2				
SO0987			2	2			2	2
SO1056	4	6	10	15				
SO1144	2	3						
SO1278	2	2						
SO2083			4	4				
SO2117					2	2		
SO2123	3	3	2	2			2	2
SO2240			6	21				
SO2323					3	3		
SO3282					2	2	2	2
SO3396			3	3	3	3	2	2
SO3582			2	2	2	2	4	4
SO3642			2	10	3	5	4	5
SO3838	8	11	8	17	2	4	4	4
SO3890			3	3				
SO4053					3	3		
SO4454	2	2			2	2		
SO4466	3	3						
SO4557	2	2						
SOA0106	3	3	5	5	3	3		
Total	**31**	**37**	**49**	**86**	**25**	**29**	**20**	**21**

Figure 5: MS/MS Spectra of a Monomethylated Peptide from SO2117 Chemoreceptor

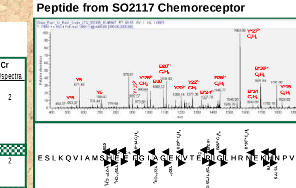

Figure 6: Protein Sequence of SO2117 Chemoreceptor in MR-1 with Putative Monomethylated Residues Highlighted in Bold Green Font at the C-terminus of the Protein

MDGVIAVFAIAIGYFVGLITFWIVSPVLAVAIMLFWQLFRQQQSTLATGTSQTE
QLNLADLGWQTHQHLYQEAITTVNECENNINNVLMVQGEAVNLLGDSFEGLGRL
MTEQTGFISDLVKSTDDAEIFHSEQMKLFANNTSQTLERFIQSTIEMSASTMEL
LEKVNLIYESMPQAMKALQDIDQISSQTNLLALNAAIEAARAGEAGRGFAVVAD
EVRALSNRSGGFSESIQKQLRSIQNQIEQLTQQASKVAAQDMSYIIDAKKDIDT
ALKQIILKAEKDSKVIARIDSSAHELETAISDTIRGLQFSDITSQSLSHTNQSL
YHLKESLKQVIAMSH**EE**FGIAGEKVT**E**RIGLHRNEKHNPVSRNEITSGEIELF

- Monomethylation on glutamic acid residues
  - Total of 27 methyl-accepting chemotaxis protein receptors (chemoreceptors) are predicted in the *S. oneidensis* proteome
    - 21 chemoreceptors with at least one monomethylated glutamate residue were identified
    - 3 chemoreceptors were identified with at least one modified peptide only under Cr exposure conditions (green highlight in Table 8) and belonged to proteins SO2117, SO3282, and SO4053
  - SO2117 was found modified solely under Cr exposure at the 45 min time-point with 2 peptides identified from 2 MS/MS spectra located at the C-terminus of the protein in both the crude and membrane fractions (Figure 5)
    - The primary identification describes E340, E341, and E351 as monomethylated, however other modified residues of minor abundance were identified as isomers of the peptide
    - The location of three putative of the the modifications have been mapped to the chemoreceptor's C-terminus (Figure 6) aiding in the validation of these PTMs as true- and not false-positives

CONCLUSIONS

- Post-translational modifications (PTMs) are important in regulating protein activity in bacteria
- MS/MS spectra of PTM peptides are present in shotgun proteomic data as unidentified spectra
- Previous acute shock study[1] data is searched with DBDigger specifying biologically relevant PTMs
- For bacteria, increasing charge state scoring filters gives same results as creating a sub-database first (Table 2 and Figure 2)
  - The small proteome size of the bacteria provides an advantage over eukaryotes
- A <3.0% peptide false discovery rate for the dataset was achieved (Table 3)
  - For PTM peptides only, false discovery rate jumps to ~50% for some modifications (Table 4), stressing the need for further validation
- Differential PTMs on SO2426, a putative DNA-binding response regulator, were manually validated (Figure 4)
- Monomethylation of glutamate residues of chemoreceptors were identifed (Table 7)
- Searching for PTMs in shotgun proteomic data on a medium resolution instrument is difficult but feasible and biologically relevant

ACKNOWLEDGMENTS

- This research was supported by the U.S. Department of Energy, Office of Science, Environmental Remediation Sciences Program.
- M. Thompson acknowledges support from the ORNL-UTK Genome Science and Technology Graduate School.
- We thank Vilmos Kertesz, Oak Ridge National Laboratory for DTAcopy.
- Oak Ridge National Laboratory is managed and operated by the University of Tennessee-Battelle, L.L.C., under contract DE-AC05-00OR22725 with the U.S. Department of Energy.
- Project website: http://compbio.ornl.gov/shewanella_metal_stress/

REFERENCES

1. Brown *et al.* Molecular and Cellular Proteomics. 2006. 5: 1054-1071.
2. Tabb *et al.* Analytical Chemistry. 2005. 77: 2464-2474.
3. Eng *et al.* Journal of the American Society for Mass Spectrometry. 1994. 5: 976-989.
4. MacCoss *et al.* Proceedings of the National Academy of Sciences. 2002. 99: 7900-7905.