Post-translational Modifications of *Desulfovibrio vulgaris* Hildenborough Sulfate Reduction Pathway Proteins.

Sara P. Gaucher[1,5,†,§], Alyssa M. Redding[2,5,†], Aindrila Mukhopadhyay[3,5], Jay D. Keasling[2,3,4,5], Anup K. Singh[1,5,*]

Sandia National Laboratory, Livermore, CA, USA[1], Department of Chemical Engineering, University of California, Berkeley, CA, USA[2], Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA[3], Department of Bioengineering, University of California, Berkeley, CA, USA[4],Virtual Institute of Microbial Stress and Survival[5]

†These authors contributed equally to this work

§Currently at Amyris Biotechnologies, Emeryville, CA, USA

*****Corresponding author

Dr. Anup K Singh
Sandia National Laboratories
P.O. Box 969, Mail Stop 9291
Livermore, CA 94551-0969
Email: aksingh@sandia.gov
Phone: 925-294-1260

ABSTRACT

Recent developments in shotgun proteomics have enabled high-throughput studies of a variety of microorganisms at a proteome level and provide experimental validation for predicted open reading frames in the corresponding genome. More importantly, advances in mass spectrometric data analysis now allow mining of large proteomics datasets for the presence of post translational modifications (PTMs). Although PTMs are a critical aspect of cellular activity, such information eludes cell wide studies conducted at the transcript level. Here we analyze several mass spectrometric datasets acquired using two dimensional liquid chromatography tandem mass spectrometry, 2D-LC/MS/MS, for the sulfate reducing bacterium, *Desulfovibrio vulgaris* Hildenborough. Our searches of the raw spectra led us to discover several post translationally modified peptides in *D. vulgaris*. Of these, several peptides containing a lysine with a +42 Dalton (Da) modification were found reproducibly across all datasets. Both acetylation and trimethylation have the same nominal +42 Da mass, and are therefore candidates for this modification. Several spectra were identified having markers for trimethylation, while one is consistent with an acetylation. Surprisingly, these modified peptides predominantly mapped to proteins involved in sulfate respiration. Other highly expressed proteins in *D. vulgaris*, such as enzymes involved in electron transport and other central metabolic processes, did not contain this modification. Decoy database searches were used to control for random spectrum/sequence matches. Additional validation for these modifications was provided by alternate workflows, for example, two-dimensional gel electrophoresis followed by mass spectrometry analysis of the dissimilatory sulfite reductase γ-subunit (DsrC) protein. MS data for DsrC in this alternate workflow also contained the +42 Da modification at the same loci. Furthermore, the DsrC homolog in another sulfate reducing bacterium, *D. desulfuricans* G20, also showed similar +42 Da modifications in the same pathway. Here we discuss our methods and implications of potential trimethylation in the *D. vulgaris* sulfate reduction pathway.

Key words: iTRAQ, SRB, DsrC, trimethyl-lysine, acetylation, PTMs

INTRODUCTION

Desulfovibrio vulgaris Hildenborough belongs to the sulfate reducing class of bacteria (SRB). Recently, interest in the physiology of *D. vulgaris* has been heightened due to its potential in bioremediation applications at toxic and radioactive metal contaminated sites. Before bioremediation strategies can be implemented effectively, however, cellular models must be developed that capture the complex relationships between the environment and the desired metabolic activity. Techniques such as genome sequencing and transcriptomics were developed as a first step toward such comprehensive cellular analysis. Yet, following extensive study, it was determined that, while these techniques formed an important foundation, much of cellular behavior remained unexplained. Technological and analytical advances in mass spectrometry-based protein studies have resulted in high-throughput analyses of a variety of microorganisms at a proteomic level[1-6], with shotgun proteomics emerging as a popular and powerful method[7, 8]. Consequently vast amounts of mass spectroscopic data are now available that can be mined to obtain many different types of information about the source organisms. The use of rigorous data analysis methods allows for highly accurate high-throughput identification of peptides. At the most fundamental level, these studies provide valuable experimental validation for the presence of the predicted proteins. A critical feature inherently contained in these datasets that eludes other cell-wide profiling methods is post translational modifications (PTMs).

Proteomics studies have formed an important part of the several cell-wide studies focused on understanding cellular response in the anaerobic organism *D. vulgaris* to a variety of growth conditions[3, 9-11]. Utilizing the iTRAQ workflow and 2D-LC/MS/MS, a large number of shotgun proteomics datasets were collected. Using these datasets, Mascot MS/MS Ion searches, including decoy database searches, were conducted to identify potential PTMs in *D. vulgaris*. Recently, a number of similar studies have been reported in other organisms[12-14]. Many PTMs are known to play critical roles in bacterial regulation. Methylations are known to be involved in bacterial signal transduction, especially in chemotaxis systems[15], but have also been shown to be involved in protein translation[16, 17]. Acetylations in bacterial systems have been shown to regulate enzyme function[18] as well as participate in signal transduction[19]. Recent studies of acetylation in

eukaryotic systems have revealed many acetylation sites beyond those present in histones, suggesting that they play a role in many key metabolic functions[20, 21]. N-terminal protein modification has been shown to be involved in protein maturation and has been implicated in controlling protein half-life[22]. Additionally, oxidation of histidine residues has been shown to be a mechanism for providing transcriptional control[23]. Despite the fact that phosphorylations are known to be important in signal transduction mechanisms[24] and stasis[25], they are chemically labile and do not lend themselves to detection following the sample handling techniques utilized. Likewise, glycosylations and lipid modifications are less abundant in prokaryotes, but have been identified in various species and cause a variety of cellular effects[26, 27]. However, these modifications require specific isolation methods to be studied. All of these modifications have been found ubiquitously in all organisms, not only bacteria, and have been shown to have varied and diverse functions in protein activity and regulation[28-33]. Thirty-five amino acid modifications were selected for inspection in this study. One primary focus was the investigation of amino acid oxidations, given that two of the four samples had undergone oxidative stress. Protein methylation, acetylation, and N-terminal modifications were targeted for study because they are among the most commonly found[17, 34], chemically stable modifications and are known to be involved in biologically relevant phenomena. Additionally, *in vitro* modifications or adducts were examined to investigate the role of sample preparation in protein modification events. Many modifications expected to be chemically labile, present in low abundance, or requiring specific isolation methods were not considered here. The methods used to search the proteomics datasets of *D. vulgaris* for PTMs are outlined. Further, validation of these results was provided by conducting decoy searches and by mining orthologous datasets acquired using different workflows. Interestingly, many PTMs were discovered in the sulfate reduction pathway, and the implications of these modifications are discussed.

METHODS

Biomass production. *Desulfovibrio vulgaris* Hildenborough (ATCC 29579) was grown in a defined lactate (60 mM)/sulfate (50 mM) medium, LS4D[9], under a variety of different stress treatments or growth conditions. To minimize sub-culturing during

experimentation, *D. vulgaris* stocks stored at -80°C were used as a 10% (% is v/v unless otherwise indicated) inoculum into 100-200 mL of fresh LS4D medium to produce starter cultures, and the cells were grown to mid-log phase (OD_{600} 0.3 – 0.4). Fresh starter cultures were used as 10% inoculum into 1-3 L biomass production cultures and grown at 30°C. All production cultures were grown in triplicate. In each condition, both control and stress treated cultures were grown. Cells were allowed to grow to an optical density of approximately 0.3 prior to application of the treatment condition. The treatments were as follows: 250 mM nitrate for 8 hours[11] (Dataset 1), 0.1% oxygen exposure for 4 hours[35] (Dataset 2), and air stress at both 2 and 4 hours[35] (Dataset 3). Control cultures were treated with an equivalent volume of water (Dataset 1) or bubbled with prepurified nitrogen (Datasets 2 and 3). Biomass collection occurred immediately following the specified treatment time, and sample processing and chilling times for biomass collection were minimized by pumping samples through a metal coil immersed in an ice bath as described previously[9, 10]. The chilled samples were harvested via centrifugation, flash frozen in liquid nitrogen, and stored at -80°C until analysis. Additionally, a *D. vulgaris* sample grown for 96 hours under conditions promoting biofilm formation was compared to a 96-hour culture employing standard planktonic growth conditions, and proteomics data obtained from this sample (Dataset 4) was also used (Clark and Redding, unpublished data). No additional stress treatment was applied to this biofilm sample.

Proteomics sample preparation and data acquisition. Sample preparation, chromatography, and mass spectrometry for iTRAQ proteomics were performed as described previously[11]. Briefly, frozen cell pellets from triplicate 50 mL cultures were thawed and pooled prior to cell lysis. Cells were lysed via sonication in lysis buffer, composed of 4 M urea with 500 mM triethylammonium bicarbonate (TEAB), pH 8.5 (Sigma-Aldrich), and the clarified lysate was used as total cellular protein. Sample denaturation, reduction, blocking, digestion, and labeling with isobaric reagents were performed according to the manufacturer's directions (Applied Biosystems, Framingham, MA). For the nitrate stress experiment (Dataset 1), 150 μg of protein was labeled with 3 separate labeling reagent vials, whereas 80 μg of protein were labeled with a single reagent vial for all other samples. Strong cation exchange (SCX) chromatography was used to separate the iTRAQ-labeled samples into 21-23 salt fractions. Fractions were

desalted using C_{18} MacroSpin Columns (Nest Group, Southborough, MA), dried, and separated on a PepMap100 C_{18} reverse phase nano-LC-MS column (Dionex-LC Packings, Sunnyvale, CA) using an Ultimate HPLC with Famous Autosampler and Switchos Micro Column Switching Module coupled with an ESI-QTOF mass analyzer (QSTAR® Hybrid Quadrupole TOF, Applied Biosystems) as previously described[11]. A 2-hour gradient from 0-25% acetonitrile was used. Two product ion scans were collected for each cycle with a 1-s accumulation time. A threshold of 50 counts was required for ions to be selected for fragmentation. Parent ions and their isotopes were excluded from further selection for 1 minute, with a mass tolerance of 100 ppm. Typical mass accuracy achieved for these data sets was 50 ppm, thus peptide modification by acetylation versus trimethylation (discussed below), a difference of 0.036 Da, could not be distinguished based on mass measurement alone.

MS/MS dataset analysis. The data mining approach is shown in Figure 1. Peak lists (m/z vs. intensity) were generated for the resulting spectra from every individual LC/MS/MS file using the Distiller program (version 2.1, Matrix Science Inc, Boston, MA), with parameters given in Supplementary Table 1. Peak lists from individual files were then merged for each stress condition to form Datasets 1, 2, 3 and 4, which were used for all subsequent analysis. The only exception was Dataset 3, where two of the LC/MS/MS files (strong cation exchange fractions 7 and 10) contained scans with no data, rendering the Distiller program unable to handle these files without manual intervention. Consequently, these two files were omitted from the analysis. A protein search database was generated from a FASTA file containing all the putative ORF sequences of *D. vulgaris* (obtained from microbesonline.org[36], 2005-02-08 upload, 3632 genes) appended with trypsin, bovine serum albumin, and ovalbumin. Identical search parameters were used in all cases, which were as follows: tryptic peptides with up to 1 missed cleavage site were considered, the precursor ion m/z tolerance was set at +/- 200 ppm, and the product ion m/z tolerance was set at +/- 0.2 Da. Modifications were considered either as static (where all occurrences of the given amino acid are expected to be modified, e.g., chemical alkylation of cysteine) or variable (where the given amino acid may be present in both modified and unmodified forms, e.g., oxidation of methionine). The choice of static and variable modifications varied according to the

search conducted. The datasets were submitted initially to a preliminary Mascot search against the *D. vulgaris* protein database. The preliminary search allowed for modification of primary amines by the iTRAQ reagent (lysine residues and peptide N-terminus) and alkylation of cysteines (methyl methane thiosulfonate for Datasets 1 and 2, and carbamidomethylation for Datasets 3 and 4) as static modifications. Both methionine oxidation to sulfoxide and iTRAQ labeling of Y were also allowed as variable modifications. Further searches used the same criteria with the following exceptions: where additional possible modifications on K, C, or protein N-terminal residues were allowed, iTRAQ labeling and alkylation were considered as variable modifications; additional modifications were considered as variable modification in individual searches as specified (Supplementary Table 2).

Potential amino acid modifications were identified from literature and on the web at unimod.org, ionsource.com, and abrf.org (Delta Mass database). In all, 35 modifications were selected for analysis. As Dataset 1 represented the most comprehensive analysis in terms of the number of proteins identified and spectra collected, MS/MS peak lists from Dataset 1 were submitted to Mascot in 22 additional searches with discrete PTMs of interest specified. Structures of the considered modifications are shown in Supplementary Figure 1, with accompanying parameters listed in Supplementary Table 2. Results (peptide sequences, modifications, match scores, etc) were exported to Excel, where sequences were filtered to retain only the top scoring match for each spectrum with an expectation value ≤ 0.05. This corresponds to a match score at the 95% confidence interval of the Mascot scoring scheme for a given search. Searches with and without additional modifications specified (searches 1 and 11, Supplementary Table 2) were also performed against a reversed sequence database generated from the *D. vulgaris* database to confirm that this score cutoff yielded an acceptable false discovery rate. Where multiple equivalently scoring matches from a given spectrum exceeded this threshold, the match with the highest protein score was retained. In cases where a given modification may occur at more than one location within the matched peptide sequence, only the highest scoring isomer is reported. The numbers of peptides from Dataset 1 having each amino acid modification and passing the filter criteria were counted, and the results are

shown in Table 1. A list of all modified peptide sequences indicated in Table 1 is given in Supplementary Table 3.

The modifications were further classified to distinguish between likely biological modifications (*in vivo*) as compared to those likely caused by sample handling (*in vitro*). *In vivo* modifications that returned the greatest number of peptides were identified as "target modifications", which were P(+18), K(+28), D/E(+14), K(+42), as well as N-terminal modifications of methylation, formylation, and acetylation, and initial methionine cleavage. Peak lists from the other three Datasets were then subjected separately to Mascot searches with the target modifications specified. Resulting peptides having the target modifications were exported to Excel and filtered as for Dataset 1. The numbers of peptides having each amino acid modification and passing the filter criteria were counted, and the results are shown in Table 2. A list of all modified peptide sequences indicated in Table 2 is given in Supplementary Table 4. Raw spectra were extracted from Dataset 1 for the peptides containing a K(+42) modification identified in at least two samples, and the sequence assignments were manually validated.

Peptides that passed the filtering criteria (top scoring match, expectation value ≤ 0.05) and were identified as protein N-terminal sequences for each dataset are listed in Supplementary Table 5. The number of each category of N-terminal modification was counted and is listed in Supplementary Table 6.

RESULTS

Amino acid modifications - post translational (in vivo) vs. artifacts (in vitro). This study used four 2D-LC/MS/MS iTRAQ proteomics datasets acquired as part of previous studies for *D. vulgaris*. The initial Mascot survey of the four datasets indicated that Dataset 1 was the most comprehensive in terms of number of spectra acquired and total proteins identified. Therefore, this dataset was used as the benchmark to evaluate the extent of known side reactions and other experimental artifacts, in addition to providing a first pass for profiling predominant modifications. The complete list of the 35 amino acid modifications considered (excluding protein N-terminal modifications), and the total number of unique sequences that passed the filtering thresholds from Dataset 1 is reported in Table 1. As shown, 6175 peptides were identified in Dataset 1 containing

only the expected modifications of K- and peptide N-terminal iTRAQ labeling, which represents about 18% of the 33,477 submitted queries. It is known that the iTRAQ reactive group, comprised of an NHS-ester, may also have a side reaction with tyrosine residues. Approximately 5% of the peptides returned were identified to have an iTRAQ-labeled Y, which is slightly higher than previously reported[37]. Prior to evaluating potential PTMs, it is critical that potential labeling side reactions and the efficiency of all sample preparation reactions be examined to ensure that the appropriate modifications are included in the search space without needlessly increasing it. This is important as increases in search space lead to increases in the threshold required to pass the filtering criteria, which may mask the presence of actual PTMs.

To evaluate the labeling efficiency of peptide primary amines (lysines and N-terminus), a separate Mascot search was performed, in which iTRAQ labeling of K or N-terminus was allowed to be variable. In this search, a total of 6220 unique peptide sequences passed the filtering thresholds. Of these, 309 (~5%) were identified as having an unlabeled peptide N-terminus. A total of 2665 sequences contained one or more lysines, and of these, only 14 (~0.5%) were identified as having unlabeled K. The overall trend observed is that iTRAQ labeling is highly efficient for K side chains, but peptide N-termini are slightly less reactive. This observation is important for further spectrum interpretation, as will be shown. Labeling of S or T by iTRAQ was negligible, as expected. No iTRAQ labeling of C was observed, which was consistent with the fact that the C residues are always reduced and blocked prior to labeling. The efficiency of cysteine modification was evaluated by performing a search allowing for variable methyl methane thiosulfonate (MMTS) modification of cysteine. In this case, only 7 peptides were returned having unmodified cysteines, indicating that alkylation had indeed proceeded to completion.

Several amino acid modifications that are most likely experimental artifacts (but cannot be completely ruled out as post-translational) were observed in Dataset 1; these include pyroglutamine formation from N-terminal Q (76 peptides), deamidation of N or Q (443 peptides total), and oxidation of M to methionine sulfoxide (118 peptides). Pyroglutamine formation is promoted by acidic conditions and likely occurred during the chromatography steps, which were performed at a pH of 3. Deamidation of N and Q is

known to occur at physiological pH, consequently deamidation could have occurred *in vivo* or *in vitro*. Interestingly, there appears to be a preference for deamidation of N (345 peptides) versus Q (98 peptides), which has been reported previously[38]. It is also possible that the deamidation observed was due to an experimental artifact where the ^{13}C peak was incorrectly selected for MS/MS analysis. However, this does not explain the observed preference for a +1 Da mass shift on N-containing peptides - an equivalent number of N and Q amino acids are present in the identified sequences, which should have produced a comparable number of N-and Q-containing peptides if the ^{13}C peak was selected for analysis. Methionine oxidation to methionine sulfoxide is known to occur during sample handling and storage; however, it can also occur *in vivo*[39].

Peptides identified with these modifications were then examined to determine if modified peptides tended to occur along with the unmodified counterpart. Artifacts of iTRAQ modified tyrosine and methionine sulfoxide were generally observed along with an unmodified counterpart, while this was not the case for pyroglutamine and deamidation modified peptides. In total, peptide sequences with such chemical modifications represent more than 10% of the sample. Because additional modifications were not considered in the current search set, the total fraction of *in vitro* modified peptides in this sample may be higher. Such a study highlights the importance of minimizing chemical side reactions so that the biological modifications are not masked.

Many modifications, such as methylations and certain oxidations/hydroxylations, are more likely to be post translational modifications rather than *in vitro* side-reactions. The most extensive of these reactions identified in Dataset 1 were hydroxyproline (34 peptides), methylated D/E (69 peptides), dimethylated K (21 peptides), and a +42 Da modification on K (18 peptides) (Table 1). The +42 Da modification is consistent with either acetylation or trimethylation, as they have the same nominal mass. It should also be noted that although the search for methylation of N and Q residues returned 23 peptides, on a per residue basis this is estimated to be only approximately 12 identifications (i.e., 23 matches divided by 2 amino acids). As is evident, the total number of modified peptides having biologically relevant modifications is far fewer compared to side-reactions. Following the identification of the most prevalent modifications in Dataset 1, the remaining three datasets were examined for these modifications (Table 2).

In addition, the N-terminal characteristics of the peptides in all four datasets were examined to assess if the initial methionine was intact and if the terminus was methylated, formylated, or acetylated. Overall, approximately 50% of the detected protein N-termini were identified having the initial methionine cleaved (Supplementary Table 6). Very few matches were returned with additional modifications, indicating that further protein N-terminal modification does not appear to be extensive in *D. vulgaris*.

Assessment of PTM assignment accuracy. While the probabilistic scoring functions built into search algorithms such as Mascot or SEQUEST aid the user in eliminating random sequence matches, a certain statistical percentage of false positive identifications are made and a certain percentage of spectra are discarded as false negatives. A Mascot expectation value of $p \leq 0.05$ was utilized as the cutoff threshold for accepting peptide results to minimize false positive results. Additionally, several reverse database searches were performed to estimate the resultant false positive rate. Specifically, the preliminary search and a search including the +42 modification on K (see searches 1 and 11, Supplementary Table 2) were repeated against a reversed sequence database generated from the *D. vulgaris* database. After applying the same filtering criteria, the number of reversed sequences reported was divided by the number of peptides reported in the original search. Based on this strategy, the false discovery rate was estimated to be 2-3%.

While reverse database searches are useful for determining the number of hits returned by random chance in a given search space, this false positive rate must be equally applied to all sequences returned by the search. Thus, this technique cannot be applied either to confirm or discount results for the subgroup of modified peptides returned by the search. In order to gauge the number of modified sequences typically returned at random by a given search, a different type of decoy search was employed. This series of searches was analogous to searches 11 and 12 (Supplementary Table 2), where iTRAQ modification to each peptide N-terminus and alkylation of cysteines were considered static modifications, while methionine oxidation to sulfoxide and iTRAQ labeling of Y were allowed as variable modifications. ITRAQ labeling of K was also allowed as a variable modification along with one of 17 different “false” modification mass shifts, ranging from 1 – 43 Da (see Supplementary Table 7). The masses searched

were chosen to minimize the chances that they may correspond to the addition of known chemical groups (i.e., 5, 13, 20 Da). Such a search strategy serves as a negative control where only random matches are returned, providing a baseline measurement for the number of hits identified by random chance. This is a similar approach to the PTM frequency matrix proposed by Pevzner and coworkers, applied on a smaller scale [40]. The median number of hits returned in these 17 decoy searches was 8 (Supplementary Table 7). This gives further validation that at least the majority of peptides identified with K(+28) and K(+42) modifications are nonrandom and deserve further scrutiny.

Of the predominant PTMs in the four datasets examined in this study, the +42 Da modification on lysine appeared most consistently on six distinct peptide(s) that were reproducible across the four datasets and passed manual spectrum evaluation. Additional modified proteins were identified having a +42 Da lysine modification in only one of the four datasets. These proteins are listed in Supplementary Tables 3 and 4, but were not investigated further. The six modified peptides of interest mapped to the following *D. vulgaris* proteins: ApsA, ApsB, Sat, DsrC, and RplK (Table 3). It is noteworthy that all of these proteins, with the exception of RplK, are involved in dissimilatory sulfate reduction in *D. vulgaris*. Based only on the nominal mass shift of 42 Da, the modification may either be an acetylation or a trimethylation. Previous mass spectrometry studies of peptides containing acetylated and trimethylated lysine residues have shown that particular MS/MS marker ions can be used to confirm the modified amino acid identity and to distinguish between the two modifications[41, 42]. In particular, peptides containing a lysine modified by an acetylation may have an immonium ion at m/z 126. The unmodified lysine immonium ion originally has a nominal mass of 101 Da, but then can undergo an ammonia-elimination reaction yielding an immonium ion at m/z 84. Trimethylated lysines, on the other hand, produce product ions which may undergo a neutral loss of 59 Da, corresponding to the loss of trimethyl-amine, which does not occur for acetylated peptides. Manual examination of the MS/MS data was performed for all six peptides (from Dataset 1) to confirm the location and, where possible, the type of modification on each of these peptides.

ApsB was identified with 52% sequence coverage, and one peptide was identified having a +42 Da modification. The MS/MS data for peptide SADSIMWTVK(+42)FR

(precursor ion m/z 542.90^{3+}) from ApsB are shown in Figure 2a. The b and y-ion series covers the majority of the peptide sequence and definitively localizes the 42 Da mass shift to the TVK residues within the peptide. The fact that the K was not labeled by the iTRAQ reagent strongly supports that it is the K and not the T or V with the +42 Da modification; the efficiency of iTRAQ labeling on lysine residues was found to be 99.5%, which would suggest that the K was already blocked from further modification. Furthermore, the sequence SADSIMWTVKFR contains a missed tryptic cleavage site. Although trypsin cleaves C-terminal to R and K, cleavage is known to be inhibited at modified K sites[43]. Indeed, an unmodified form of this peptide was observed as SADSIMWTVK, having an iTRAQ labeled lysine residue. Taken together, the data support a 42 Da modification on K10 within this peptide. The presence of 59 Da neutral losses from the y_6 and y_7 sequence ions in this peptide indicates that the modification is most likely trimethylation[42].

The DsrC protein was identified with 93% sequence coverage. Two modified peptides, LK(+42)EVYELFPSGPGK, and ESEGISDISPDHQK(+42)IIDFLQDYYK were observed on this protein. MS/MS data were observed both for the modified LK(+42)EVYELFPSGPGK sequence as well as an oxidized form, LK(+42)EVYELFPS(+16)GPGK; however, this sequence was never observed without the +42 Da mass shift. A spectrum from the oxidized form, having a parent molecular weight of m/z 637.28^{3+}, is shown in Figure 2b. The sequence coverage is quite high, and indicates definitively that the oxidation occurs on the S residue. In this case, the +42 modification is localized to the KE residues within the peptide. Again, following the above logic, if the K were not modified then it should have been iTRAQ labeled, and trypsin should have cleaved the LK residues away from the remaining peptide sequence. Further inspection of the modified MS/MS spectrum reveals the presence of several 59 Da neutral losses from the b ion series, indicating that the modification is a trimethylation on the N-terminal end of the peptide. The MS/MS data for ESEGISDISPDHQK(+42)IIDFLQDYYK are shown in Supplementary Figure 2a. Despite the length of this sequence, y and b-ion series are present for the entire sequence except the HQKII subsequence, which localizes the +42 Da mass shift to this region. Again, it is important to note the presence of the missed cleavage and lack of iTRAQ

label as indications that the K is the most likely modified residue. Both ESEGISDISPDHQK and IIDFLQDYYK were independently observed in their unmodified forms, suggesting that the missed cleavage was not a random event. However, in this case, the spectrum does not allow discrimination between acetylation and trimethylation.

Additional MS survey scans for the peptides SADSIMWTVKFR from ApsB and LKEVYELFPSGPGK from DsrC were identified having precursors 14 Da lower in mass than the +42 Da modified precursors. These precursor ions had also been subjected to MS/MS, and Mascot analysis returned the same sequences (SADSIMWTVKFR and LKEVYELFPSGPGK) but modified by +28 instead of +42 Da at the first K in each of these peptides. The +28 Da mass shift most likely corresponds to dimethylation, further substantiating that the +42 modification identified for these peptides corresponds to trimethylation.

The presence of the sulfate adenyltransferase, Sat, was confirmed by identifying this large protein with 83% coverage. The peptide VILSGTK(+42)LR, having a precursor ion molecular weight of m/z 586.87^{2+}, is shown in Figure 2c. Complete sequence coverage by the b and y-ion series localizes the modification to K7 in the peptide. Of note, this spectrum contains a fairly prominent ion at m/z 126, which has been previously shown to be a marker for acetylation[41, 42]. Additionally, this peptide spectrum does not contain any observed 59 Da neutral loss products.

ApsA was identified with 86% sequence coverage, and the MS/MS spectrum for the modified peptide DGYGPVGAWFLLFK(+42)AK, having a precursor molecular weight of m/z 700.36^{3+}, is shown in Supplementary Figure 2b. The sequence coverage is complete, definitively localizing the +42 Da modification to K14 in the peptide. An immonium ion at m/z 143 is also present, which corresponds to a +42 Da modified K. Again, the K modification has inhibited cleavage by trypsin. Although the amino acid sequence N-terminal to this peptide is RFK and the C-terminal sequence of this peptide is KAK, neither of the peptides FKDGYGPVGAWFLLFK or FKDGYGPVGAWFLLFKAK is observed, strongly supporting the conclusion that the missed cleavage is not a random event. In this case, the peptide sequence was not

observed with an unmodified counterpart. Unfortunately, the MS/MS spectrum does not provide evidence to discriminate between acetylation and trimethylation in this peptide.

One modified lysine residue was identified in RplK. The spectrum for TMEQK(+42)GMITPVVITVYADR is shown in Supplementary Figure 2c. This spectrum has a nearly complete series of fragment ions that localizes the modification to the QK residues of the peptide. Again, the peptide contains a missed cleavage site at the lysine residue, which is not iTRAQ labeled, suggesting that the lysine is the modified amino acid. A second peptide, LQIPAGAANPSPPVGPALGQHGLNIMAFCK(+42)EFNAK, was also returned by the Mascot search algorithm for multiple samples with an expectation value above the 0.05 threshold. However, the match did not pass manual inspection due to poor spectrum quality.

It should be acknowledged that several known PTMs have the same nominal mass shift as single amino acid substitutions. Single amino acid substitutions become especially prevalent when working with lab strains that have been repeatedly cultivated over many generations, leading to a divergence between the strain being examined and the sequenced strain used to generate the protein database. In the present work, this issue is expected to be minimal because the lab strains utilized in this study were obtained directly from the ATCC stock that provided the original sequence. Culturing protocols were specifically designed to ensure that the biomass used for any experiment was within three subcultures from the original ATCC stock. Furthermore, in many cases diagnostic marker ions were observed in the spectra that support the presence of proposed PTMs, as described above.

Assessment of the +42 Da modification in SRP from alternate workflows. In order to ensure that the observed modifications were not artifacts associated with the iTRAQ workflow in the four datasets analyzed, selected mass spectrometry data from two alternate work flows were also analyzed. As was the case with the previous four datasets, this data was generated as part of separate experiments and was reevaluated as part of this study. These data further confirmed the presence of modified lysine residues in proteins ApsB, ApsA and DsrC. In the first workflow, two *D. vulgaris* strains were created

incorporating a Strep tag at the terminus of either ApsA or ApsB. A Strep-Tactin column (IBA, Göttingen, Germany) was used to isolate each of these proteins, along with any interacting proteins following standard procedures[44] (Chhabra *et al.*, in preparation). After reduction, alkylation and tryptic digestion, the resulting peptide mixtures were analyzed by LC/MS/MS and peptide sequences were identified by Mascot using the same criteria outlined in the Methods above. Three peptides were confirmed in these *D. vulgaris* samples: SADSIMWTVK(+42)FR from ApsB with a precursor m/z of 494.9^{3+}; DGYGPVGAWFLLFK(+42)AK from ApsA with precursors at both m/z 905.9^{2+} and 604.3^{3+}; and ESEGISDISPDHQK(+42)IIDFLQDYYK from DsrC having precursors at both m/z 961.4^{3+} and 721.3^{4+}. Note that because iTRAQ labeling was not performed for these samples, the precursor ions for each of these sequences appear at different m/z values than those reported in Figure 2 and Supplementary Figure 2, even when the same precursor charge state was observed. This is important confirmatory evidence because it precludes the possibility that a particular contaminant at a given m/z value was present in one case, and caused a false positive match to one of these modified sequences. In a second set of experiments, *D. vulgaris* proteins were resolved by two-dimensional electrophoresis. The spots were cut out, digested in-gel with trypsin, and the proteins were identified using a peptide mass fingerprint approach[3]. In particular, the DsrC protein was observed to contain the peptide LK(+42)EVYELFPSGPGK as well as the oxidized form of this peptide. Tandem mass spectra were acquired on this peptide and found to be consistent with the conclusions presented above, namely: the spectrum localized the 42 Da mass shift to the LK peptide N-terminal subsequence, the observation of several 59 Da neutral loss product ions indicates that the modification is most likely trimethylation, and the oxidized peptide version has the +16 Da modification within the PSG subsequence.

Assessment of the +42 Da modifications in D. desulfuricans G20, another SRB. The majority of the +42 Da PTM maps to sulfate reducing proteins in *D. vulgaris*, where an earlier analysis of the genes encoded by sulfate reducing bacteria suggested that these were among the signature genes[3]. To find out if this modification was more generally applicable to other SRB, the proteome of *Desulfovibrio desulfuricans* G20 was examined

to assess if the +42 Da modification could also be detected. G20 biomass was grown in LS4D, and the cellular lysis, denaturation, and reduction were performed using the same protocols used for *D. vulgaris*, with the exception that the cysteine residues were blocked with iodoacetamide for 30 min. Examination of the sulfate reducing homologs identified the +42 Da modification on two analogous peptides in G20, namely the LK(+42)QVYELFPSGPGK peptide from DsrC and the DGYGPVGAWFLLFK(+42)AK peptide from ApsA. Although the spectra produced in this study were unable to provide sufficient evidence to distinguish between trimethylation or acetylation, the presence of these modifications was established in *D. desulfuricans* G20, even where the peptide sequences differ slightly from those observed in *D. vulgaris*. Thus, the presence of these modifications in other homologous sulfate reducing pathways was confirmed.

Additional PTMs. Determining which modified peptides were observed in multiple datasets allowed targeted selection of seven peptides having a K(+42) modification of which six passed manual inspection/confirmation. A similar analysis was performed for the other modified peptides of interest, revealing 11 peptides containing K(+28), 23 peptides having P(+16), and 22 peptides with D/E(+14) mass shifts. The sequence data are summarized in Supplementary Table 8, where the peptides are grouped by modification type and sorted by sequence length. However, the median size of peptides observed with the K+42 modification was 16, while it was much smaller for the others. This may be because these other modifications do not inhibit trypsin activity, leading to shorter sequence lengths overall. Interestingly, in no other case were a significant number of modified proteins found to belong to members of the same pathway. While it may be that these are legitimate modifications containing interesting biological stories, we elected not to pursue them as part of this study.

DISCUSSION

Regulation at the protein level is being recognized as an integral component of cell wide functions and post translational modifications form an important part of these regulatory processes. However, very little is yet known about any PTMs in *D. vulgaris*. Thirty-five modifications of interest were identified from literature and databases and

four large proteomics datasets were screened for them. While several PTMs were identified using this data-mining technique, a +42 Da modification on lysine residues was the only modification that appeared consistently across multiple datasets. This +42 Da mass shift was reproducibly identified on 6 distinct peptides, where each modification was represented in at least two separate datasets. Decoy searches of lysine residues further confirmed the specificity of these results. Interestingly, the majority of these modifications mapped to proteins that were linked functionally, being members of the sulfate reduction pathway (SRP). The candidates involved in sulfate reduction appear to be highly abundant evidenced by both high absolute mRNA levels[45] and many peptide observations in proteomics datasets. However, the presence of this modification on the SRP proteins cannot be attributed to high abundance, as other pathways that appear to be similarly abundant did not display this modification. In fact, the only exception among the 1100 proteins identified was for RplK, the ribosomal subunit L11, which had one +42 modification site. Two factors supported the physiological relevance of this modification in the sulfate reduction pathway. This modification was observed in multiple datasets from orthologous work flows, and homologs of the SRP protein in the closely related *D. desulfuricans* also contained modified lysine residues at the same loci.

All organisms contain assimilatory sulfate reduction complexes, which enable the incorporation of sulfur into metabolites. However, in SRB where sulfate also serves as an electron acceptor and is postulated to be the primary source of energy, the dissimilatory SRP[46] is encoded by signature genes conserved across all SRB[3]. The annotated SRP in *D. vulgaris* is depicted in Figure 3. The first step, required in all sulfate utilization pathways, assimilatory or dissimilatory, is the conversion of sulfate into the activated intermediate adenylyl sulfate (APS), and is effected via the sulfate adenylyltransferase (Sat), also called ATP sulfurylase[47]. Once formed in the dissimilatory pathway, APS is reduced to sulfite via the APS reductase complex, comprised of ApsA and ApsB. For the final conversion of bisulfite to sulfide, the bisulfite reductase, also referred to as dissimilatory sulfite reductase (Dsr) or desulfoviridin[48], is the primary protein complex involved[49]. The most interesting discovery in this study is that each of the enzymatic steps in the sulfate reduction pathway delineated above contains at least one lysine residue with the +42 modification.

In almost all cases, the modified lysine loci as well as the neighboring sequence are highly conserved. Exceptions to this include the K78 of DsrC and the K382 of Sat; however, in these instances the locus harbors a positively charged residue.

It is difficult to assess the impact of a PTM on the function of a protein with no information about its structure. The only protein in this pathway that has been crystallized in any organism is DsrC[50-52]. To obtain a better understanding of the implications of the protein modifications, crystal structures from the homologous DsrC proteins were investigated. The *Archeoglobus fulgidus* DsrC has the highest homology to the *D. vulgaris* DsrC sequence and was used as the scaffold for modeling (Figure 4B). As the model shows, the modified lysines appear on two distinct faces of the three dimensional structure and are both present on helices. A helical wheel rendering of these predicted helices also show them to be amphiphilic with the modified lysine on the charged face (Figure 4C). Previous studies of protein methylation revealed that trimethylation can stabilize α-helix structures[53]. It is therefore interesting to speculate that the purpose of these modified lysines may be to stabilize the charged face of the helix. The DsrC enzyme in *D. vulgaris* forms part of the DsrABC complex[54], although this may not be the case in the *A. fulgidus*[50]. DsrC is also predicted to interact with the DsrMKJOP complex[54-56]. As both the PTMs point outward rather than inward, a modification at either of these positions may impact protein-protein interactions[57]. The regions important for these protein-protein interactions need to be elucidated to validate this hypothesis.

In this study, several of the +42 Da modifications observed appear to be trimethylations primarily associated with sulfate reduction in *D. vulgaris*. Many putative methyl- and acetyltransferases have been annotated in the *D. vulgaris* genome, providing potential pathways capable of generating these modifications. Although iTRAQ labeling allowed determination of total protein expression in stressed relative to control conditions[11, 35], the individual peptide data was not sufficient to determine whether these modifications occurred as a function of the stress treatments. Biochemical evidence needs to be found to establish the presence and physiological relevance of these PTMs unequivocally. For non-model organisms such as *D. vulgaris*, recent advances in genetic tools provide the means for such follow up work[58, 59]. The discovery of this modification

in multiple proteins of this critical pathway and in another sulfate reducing bacteria makes it an ideal candidate for such experimental validation.

ACKNOWLEDGEMENTS

We thank Dr. Subodh Nimkar (Applied Biosystems), Dr. Swapnil Chhabra (Lawrence Berkeley National Laboratory), and Dr. Christopher Petzold (QB3) for assistance with mass spectrometric data acquisition and insightful comments. We are grateful to Melinda Clark (Montana University) for *D. vulgaris* biofilm sample preparation and Dominique Joyner (Lawrence Berkeley National Laboratory) for providing the *D. desulfuricans* culture. We appreciate Dr. Judy Wall for engaging in helpful discussions and critically reviewing the manuscript. This work is part of the Virtual Institute for Microbial Stress and Survival (http://vimss.lbl.gov) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program through contract DE-AC02-05CH11231 between the Lawrence Berkeley National Laboratory and the US Department of Energy.

SUPPORTING INFORMATION AVAILABLE

Supplementary Figures 1-2 and Supplementary Tables 1-8 are available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Fang, R.; Elias, D. A.; Monroe, M. E.; Shen, Y.; McIntosh, M.; Wang, P.; Goddard, C. D.; Callister, S. J.; Moore, R. J.; Gorby, Y. A.; Adkins, J. N.; Fredrickson, J. K.; Lipton, M. S.; Smith, R. D., Differential label-free quantitative proteomic analysis of *Shewanella oneidensis* cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol Cell Proteomics* **2006,** *5*(4), 714-25.

(2) Lipton, M. S.; Romine, M. F.; Monroe, M. E.; Elias, D. A.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Fredrickson, J.; Hixson, K. K.; Masselon, C.; Mottaz, H.; Tolic, N.; Smith, R. D., AMT tag approach to proteomic characterization of *Deinococcus radiodurans* and *Shewanella oneidensis*. *Methods Biochem Anal* **2006,** *49*, 113-34.

(3) Chhabra, S. R.; He, Q.; Huang, K. H.; Gaucher, S. P.; Alm, E. J.; He, Z.; Hadi, M. Z.; Hazen, T. C.; Wall, J. D.; Zhou, J.; Arkin, A. P.; Singh, A. K., Global Analysis of Heat Shock Response in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* **2006,** *188*(5), 1817-28.

(4) Ram, R. J.; Verberkmoes, N. C.; Thelen, M. P.; Tyson, G. W.; Baker, B. J.; Blake, R. C., 2nd; Shah, M.; Hettich, R. L.; Banfield, J. F., Community proteomics of a natural microbial biofilm. *Science* **2005,** *308*(5730), 1915-20.

(5) Rosen, R.; Ron, E. Z., Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom Rev* **2002,** *21*(4), 244-65.

(6) Luo, Q.; Tang, K.; Yang, F.; Elias, A.; Shen, Y.; Moore, R. J.; Zhao, R.; Hixson, K. K.; Rossie, S. S.; Smith, R. D., More sensitive and quantitative proteomic measurements using very low flow rate porous silica monolithic LC columns with electrospray ionization-mass spectrometry. *J Proteome Res* **2006,** *5*(5), 1091-7.

(7) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd, An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **2001,** *73*(23), 5683-90.

(8) MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., Shotgun identification of protein modifications from protein complexes and lens tissue. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99*(12), 7900-7905.

(9) Mukhopadhyay, A.; He, Z.; Alm, E. J.; Arkin, A. P.; Baidoo, E. E.; Borglin, S. C.; Chen, W.; Hazen, T. C.; He, Q.; Holman, H.-Y.; Huang, K.; Huang, R.; Joyner, D. C.; Katz, N.; Keller, M.; Oeller, P.; Redding, A.; Sun, J.; Wall, J.; Wei, J.; Yang, Z.; Yen, H.-C.; Zhou, J.; Keasling, J. D., Salt stress in *Desulfovibrio vulgaris* Hildenborough: an integrated genomics approach. *J. Bacteriol.* **2006,** *188*(11), 4068-4078.

(10) Mukhopadhyay, A.; Redding, A. M.; Joachimiak, M. P.; Arkin, A. P.; Borglin, S. E.; Dehal, P. S.; Chakraborty, R.; Geller, J. T.; Hazen, T. C.; He, Q.; Joyner, D. C.; Martin, V. J. J.; Wall, J. D.; Yang, Z. K.; Zhou, J.; Keasling, J. D., Cell wide responses to low oxygen exposure in *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* **2007**, JB.00368-07.

(11) Redding, A. M.; Mukhopadhyay, A.; Joyner, D. C.; Hazen, T. C.; Keasling, J. D., Study of nitrate stress in *Desulfovibrio vulgaris* Hildenborough using iTRAQ proteomics. *Brief Funct Genomic Proteomic* **2006,** *5*(2), 133-143.
(12) Thompson, M. R.; Thompson, D. K.; Hettich, R. L., Systematic Assessment of the Benefits and Caveats in Mining Microbial Post-Translational Modifications from Shotgun Proteomic Data: The Response of *Shewanella oneidensis* to Chromate Exposure. *Journal of Proteome Research* **2008,** *Web Release Jan. 2*.
(13) Vasilescu, J.; Smith, J. C.; Zweitzig, D. R.; Denis, N. J.; Haines, D. S.; Figeys, D., Systematic determination of ion score cutoffs based on calculated false positive rates: application for identifying ubiquitinated proteins by tandem mass spectrometry. *Journal of Mass Spectrometry* **2007,** *9999*(9999), n/a.
(14) Gupta, N.; Tanner, S.; Jaitly, N.; Adkins, J. N.; Lipton, M.; Edwards, R.; Romine, M.; Osterman, A.; Bafna, V.; Smith, R. D.; Pevzner, P. A., Whole proteome analysis of post-translational modifications: Applications of mass-spectrometry for proteogenomic annotation. *Genome Research* **2007,** *17*(9), 1362-1377.
(15) Szurmant, L.; Ordal, G. W., Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiology and Molecular Biology Reviews* **2004,** *68*(2), 301-319.
(16) Young, C. C.; Bernlohr, R. W., Elongation Factor-Tu Is Methylated in Response to Nutrient Deprivation in *Escherichia-coli*. *Journal of Bacteriology* **1991,** *173*(10), 3096-3100.
(17) Polevoda, B.; Sherman, F., Methylation of proteins involved in translation. *Molecular Microbiology* **2007,** *65*(3), 590-606.
(18) Gardner, J. G.; Grundy, F. J.; Henkin, T. A.; Escalante-Semerena, J. C., Control of acetyl-coenzyme A synthetase (AcsA) activity by acetylation/deacetylation without NAD(+) involvement in *Bacillus subtilis*. *Journal of Bacteriology* **2006,** *188*(15), 5460-5468.
(19) Barak, R.; Yan, J.; Shainskaya, A.; Eisenbach, M., The chemotaxis response regulator CheY can catalyze its own acetylation. *Journal of Molecular Biology* **2006,** *359*(2), 251-265.
(20) Kim, S. C.; Sprung, R.; Chen, Y.; Xu, Y. D.; Ball, H.; Pei, J. M.; Cheng, T. L.; Kho, Y.; Xiao, H.; Xiao, L.; Grishin, N. V.; White, M.; Yang, X. J.; Zhao, Y. M., Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Molecular Cell* **2006,** *23*(4), 607-618.
(21) Zhang, X. H.; Yuan, Z. G.; Zhang, Y. T.; Yong, S.; Salas-Burgos, A.; Koomen, J.; Olashaw, N.; Parsons, J. T.; Yang, X. J.; Dent, S. R.; Yao, T. P.; Lane, W. S.; Seto, E., HDAC6 modulates cell motility by altering the acetylation level of cortactin. *Molecular Cell* **2007,** *27*(2), 197-213.
(22) Gonzales, T.; RobertBaudouy, J., Bacterial aminopeptidases: Properties and functions. *Fems Microbiology Reviews* **1996,** *18*(4), 319-344.
(23) Lee, J. W.; Helmann, J. D., The PerR transcription factor senses H_2O_2 by metal-catalysed histidine oxidation. *Nature* **2006,** *440*(7082), 363-7.
(24) Stock, A. M.; Robinson, V. L.; Goudreau, P. N., Two-component signal transduction. *Annual Review of Biochemistry* **2000,** *69*, 183-215.
(25) Freestone, P.; Nystrom, T.; Trinei, M.; Norris, V., The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J Mol Biol* **1997,** *274*(3), 318-24.

(26) Upreti, R. K.; Kumar, M.; Shankar, V., Bacterial glycoproteins: Functions, biosynthesis and applications. *Proteomics* **2003,** *3*(4), 363-379.
(27) Hayashi, S.; Wu, H. C., Lipoproteins in Bacteria. *Journal of Bioenergetics and Biomembranes* **1990,** *22*(3), 451-471.
(28) Eichler, J.; Adams, M. W. W., Posttranslational protein modification in Archaea. *Microbiology and Molecular Biology Reviews* **2005,** *69*(3), 393-425.
(29) Villar-Garea, A.; Imhof, A., The analysis of histone modifications. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2006,** *1764*(12), 1932-1939.
(30) Kimura, A.; Matsubara, K.; Horikoshi, M., A decade of histone acetylation: Marking eukaryotic chromosomes with specific codes. *Journal of Biochemistry* **2005,** *138*(6), 647-662.
(31) Latham, J. A.; Dent, S. Y. R., Cross-regulation of histone modifications. *Nature Structural & Molecular Biology* **2007,** *14*(11), 1017-1024.
(32) Morgunkova, A.; Barlev, N. A., Lysine methylation goes global. *Cell Cycle* **2006,** *5*(12), 1308-1312.
(33) Yang, X. J.; Gregoire, S., Metabolism, cytoskeleton and cellular signalling in the grip of protein N-epsilon- and O-acetylation. *Embo Reports* **2007,** *8*(6), 556-562.
(34) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Integrating "top-down" and "bottom-up" mass spectrometric approaches for proteomic analysis of *Shewanella oneidensis*. *Journal of Proteome Research* **2002,** *1*(3), 239-252.
(35) Mukhopadhyay, A.; Redding, A. M.; Joachimiak, M. P.; Arkin, A. P.; Borglin, S. E.; Dehal, P. S.; Chakraborty, R.; Geller, J. T.; Hazen, T. C.; He, Q.; Joyner, D. C.; Martin, V. J. J.; Wall, J. D.; Yang, Z. K.; Zhou, J.; Keasling, J. D., Cell wide responses to low oxygen exposure in *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* **2007**, 5996-6010.
(36) Alm, E. J.; Huang, K. H.; Price, M. N.; Koche, R. P.; Keller, K.; Dubchak, I. L.; Arkin, A. P., The MicrobesOnline Web site for comparative genomics. *Genome Res* **2005,** *15*(7), 1015-22.
(37) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** *3*(12), 1154-69.
(38) Robinson, A. B.; Scotchle.Jw; McKerrow, J. H., Rates of Nonenzymatic Deamidation of Glutaminyl and Asparaginyl Residues in Pentapeptides. *Journal of the American Chemical Society* **1973,** *95*(24), 8156-8160.
(39) Lagerwerf, F. M.; vandeWeert, M.; Heerma, W.; Haverkamp, J., Identification of oxidized methionine in peptides. *Rapid Communications in Mass Spectrometry* **1996,** *10*(15), 1905-1910.
(40) Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A., Identification of post-translational modifications by blind search of mass spectra. *Nature Biotechnology* **2005,** *23*(12), 1562-1567.
(41) Kim, J. Y.; Kim, K. W.; Kwon, H. J.; Lee, D. W.; Yoo, J. S., Probing lysine acetylation with a modification-specific marker ion using high-performance liquid

chromatography/electrospray-mass spectrometry with collision-induced dissociation. *Analytical Chemistry* **2002,** *74*(21), 5443-5449.
(42) Zhang, K. L.; Yau, P. M.; Chandrasekhar, B.; New, R.; Kondrat, R.; Imai, B. S.; Bradbury, M. E., Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. *Proteomics* **2004,** *4*(1), 1-10.
(43) Rice, R. H.; Means, G. E.; Brown, W. D., Stabilization of Bovine Trypsin by Reductive Methylation. *Biochimica Et Biophysica Acta* **1977,** *492*(2), 316-321.
(44) Schmidt, T. G.; Skerra, A., The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols* **2007,** *2*, 1528-1535.
(45) Arkin, A. P. Absolute Expression Estimates. http://genomics.lbl.gov/VIMSS/expression.html November, **2007**
(46) Peck, H. D.; Legall, J.; Vanbeeumen, J., Biochemistry of Dissimilatory Sulfate Reduction. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **1982,** *298*(1093), 443-466.
(47) Akagi, J. M., Respiratory Sulfate Reduction. In *Sulfate-Reducing Bacteria*, Barton, L. L., ed., Plenum Press: New York, **1995**; Biotechnology Handbooks Vol. 8, p 89-108.
(48) Lee, J. P.; Legall, J.; Peck, H. D., Isolation of Assimilatory-type and Dissimilatory-type Sulfite Reductases from *Desulfovibrio-vulgaris*. *Journal of Bacteriology* **1973,** *115*(2), 529-542.
(49) Hansen, T. A., Metabolism of Sulfate-Reducing Prokaryotes. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **1994,** *66*(1-3), 165-185.
(50) Mander, G. J.; Weiss, M. S.; Hedderich, R.; Kahnt, J.; Ermler, U.; Warkentin, E., X-ray structure of the gamma-subunit of a dissimilatory sulfite reductase: fixed and flexible C-terminal arms. *FEBS Lett* **2005,** *579*(21), 4600-4.
(51) Cort, J. R.; Mariappan, S. V.; Kim, C. Y.; Park, M. S.; Peat, T. S.; Waldo, G. S.; Terwilliger, T. C.; Kennedy, M. A., Solution structure of *Pyrobaculum aerophilum* DsrC, an archaeal homologue of the gamma subunit of dissimilatory sulfite reductase. *Eur J Biochem* **2001,** *268*(22), 5842-50.
(52) Weiss, M. S.; Mander, G.; Hedderich, R.; Diederichs, K.; Ermler, U.; Warkentin, E., Determination of a novel structure by a combination of long-wavelength sulfur phasing and radiation-damage-induced phasing. *Acta Crystallographica Section D* **2004,** *60*(4), 686-695.
(53) Bello, J., Helix Formation in Poly (N-Epsilon,N-Epsilon,N-Epsilon-Trimethyl-L-Lysine) and Poly(L-Lysine) - Dependence on Concentration and Molecular-Weight. *Biopolymers* **1992,** *32*(2), 185-188.
(54) Pierik, A. J.; Duyvis, M. G.; Vanhelvoort, J.; Wolbert, B. G.; Hagen, W. R., The 3rd Subunit of Desulfoviridin-type Dissimilatory Sulfite Reductases. *European Journal of Biochemistry* **1992,** *205*(1), 111-115.
(55) Pires, R. H.; Venceslau, S. S.; Morais, F.; Teixeira, M.; Xavier, A. V.; Pereira, I. A. C., Characterization of the Desulfovibrio desulfuricans ATCC 27774 DsrMKJOP complex - A membrane-bound redox complex involved in the sulfate respiratory pathway. *Biochemistry* **2006,** *45*(1), 249-262.

(56) Sander, J.; Engels-Schwarzlose, S.; Dahl, C., Importance of the DsrMKJOP complex for sulfur oxidation in *Allochromatium vinosum* and phylogenetic analysis of related complexes in other prokaryotes. *Archives of Microbiology* **2006,** *186*(5), 357-366.
(57) Pang, C. N. I.; Hayen, A.; Wilkins, M. R., Surface accessibility of protein post-translational modifications. *Journal of Proteome Research* **2007,** *6*(5), 1833-1845.
(58) Wall, J. D.; Hemme, C. L.; Rapp-Giles, B.; Ringbauer, J. A., Jr.; Casalot, L.; Giblin, T., Genes and genetic manipulations of *Desulfovibrio*. In *Biochemistry and physiology of anaerobic bacteria*, Ljungdahl, L. G.; Adams, M. W. W.; Barton, L. L.; Ferry, J. G.; Johnson, M. K., ed., Springer: New York, **2003**; Vol. p 85-98.
(59) Bender, K. S.; Yen, H.-C.; Wall, J. D., Analysing the metabolic capabilities of *Desulfovibrio* species through genetic manipulations. In *Biotechnology & genetic engineering reviews*, Harding, S. E., ed., Nottingham University Press: Lavoisier, France, **2006**; Vol. 23, p 157-174.

FIGURE CAPTIONS

Figure 1. Flowchart for identifying protein post-translational modifications. Four LC/MS/MS datasets (approximately 86,040 spectra) from *Desulfovibrio vulgaris* Hildenborough were peakpicked using MatrixScience Distiller and subjected to a preliminary Mascot search. Dataset 1 was identified as the most comprehensive file, and was used to screen all other modifications. The efficiencies of iTRAQ labeling and cysteine blockage were evaluated to ensure that the most appropriate search space was utilized. MS/MS peak lists were submitted to Mascot in 23 separate searches using different static and variable modifications. *In vivo* modifications with the greatest number of peptides were identified as "target modifications": P(+18), K(+28), D/E(+14), K(+42). Peak lists from the other three datasets were individually submitted to separate Mascot searches with the target modifications specified. Raw spectra were extracted from Dataset 1 for modified peptides identified in at least two datasets and sequence assignments were manually validated.

Figure 2. QStar ESI-MS/MS data (A) Peptide (iTRAQ)SADSIMWTVK(+42)FR from ApsB. The peptide fragmentation pattern localizes the modification to the TVK residues within the peptide. The fact that the K10 of this peptide is not iTRAQ labeled (0.5% occurrence) and that this represents a missed trypsin cleavage site both suggest that K10 has been modified. Further, the presence of y_6-59 and y_7-59 ions is consistent with this modification being a trimethylation event occurring on the C-terminal end of the peptide. (B) Peptide (iTRAQ)LK(+42)EVYELFPS(+16)GPGK(iTRAQ) from DsrC. This fragmentation series localizes the modification to the KE amino acids within the peptide. This peptide likewise contains a missed tryptic cleavage and the K2 remains unlabeled by iTRAQ reagents, suggesting that the lysine is the labeled residue. In this case, the b-ion series contains several -59 losses (b_6, b_7, b_8), which is indicative that the modification is a trimethylation on the N-terminal end of the peptide. This peptide also has an oxididation on the S10 residue, which is clearly identified from the ion series. The unoxidized form of this peptide was also identified, although this peptide was never observed without the +42 modification. (C) Peptide (iTRAQ)VILSGTK(+42)LR from Sat. In this spectrum, the fragmentation series is complete and clearly localizes the modification to the K7

residue. Unlike the previous two examples, there is a strong immonium ion at 126.1, which has been shown to be a marker for acetylation. Consistent with this, there are no 59 Da losses present in the spectrum.

Figure 3. A diagram of the known sulfate reduction pathway (SRP) in *D. vulgaris*. The chemical structures of intermediates of the SRP are shown. Enzymes completing the intermediate steps are given. The percent coverage for each protein is shown, along with the corresponding modified peptides that were identified. As can be seen, every major member of this pathway has at least one peptide containing a modified lysine residue.

Figure 4. DsrC Modeling (A) A ClustalW alignment of DsrC proteins from multiple organisms is shown. *D. vulgaris* and *D. desulfuricans* G20 are shown, and the high homology between them can be noted. The additional proteins in the alignment were selected based on the fact that crystal structures have been described for these DsrC proteins. (B) Model of DvH DrsC using the *Archeoglobus fulgidus* DsrC as template, as it has the highest homology of the crystallized DsrC proteins. As can be observed, both of the modified lysine residues point away from the DsrC protein itself. (C) A view of the residues involved in the helices containing the modified residues. Both of these helices are amphipathic.

Figure 1.

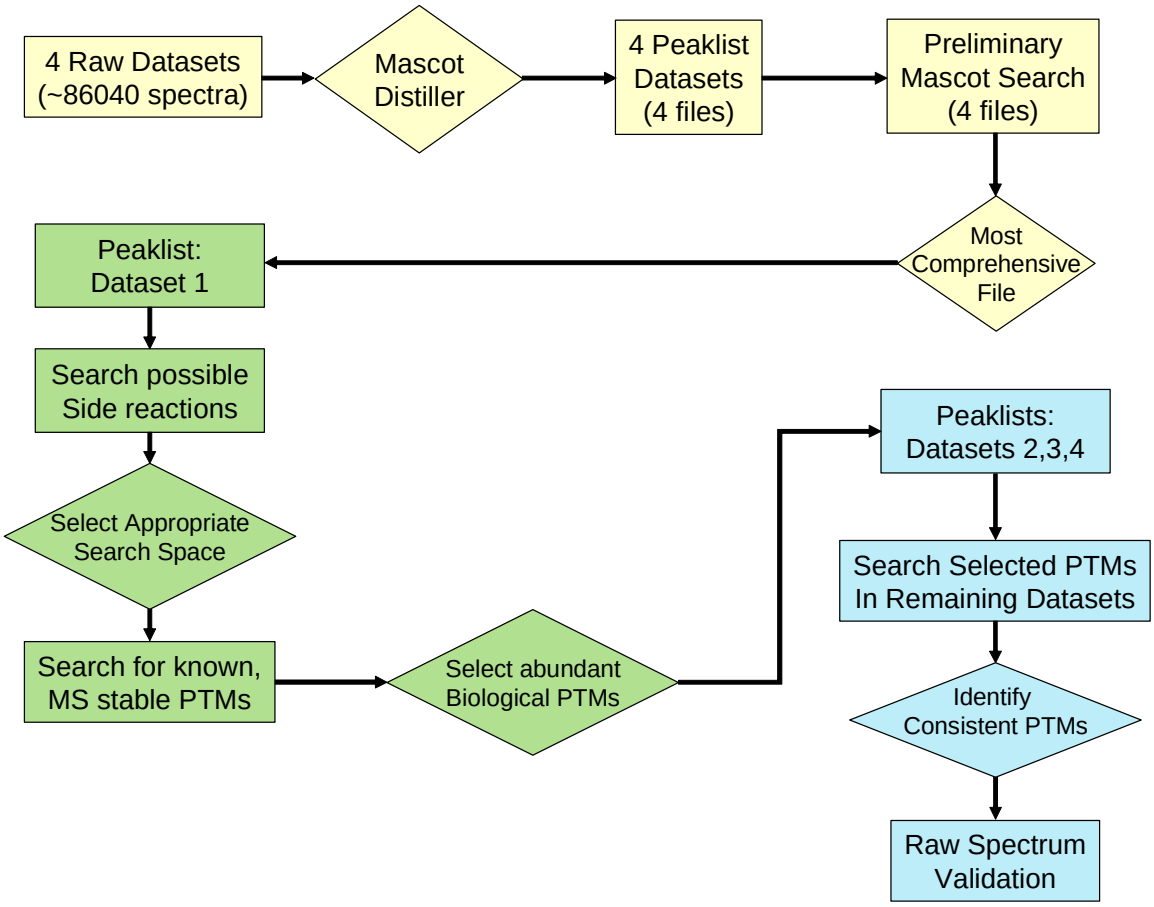

4 Raw Datasets (~86040 spectra)
Mascot Distiller
4 Peaklist Datasets (4 files)
Preliminary Mascot Search (4 files)
Most Comprehensive File
Peaklist: Dataset 1
Search possible Side reactions
Select Appropriate Search Space
Search for known, MS stable PTMs
Select abundant Biological PTMs
Peaklists: Datasets 2,3,4
Search Selected PTMs In Remaining Datasets
Identify Consistent PTMs
Raw Spectrum Validation

Figure 2

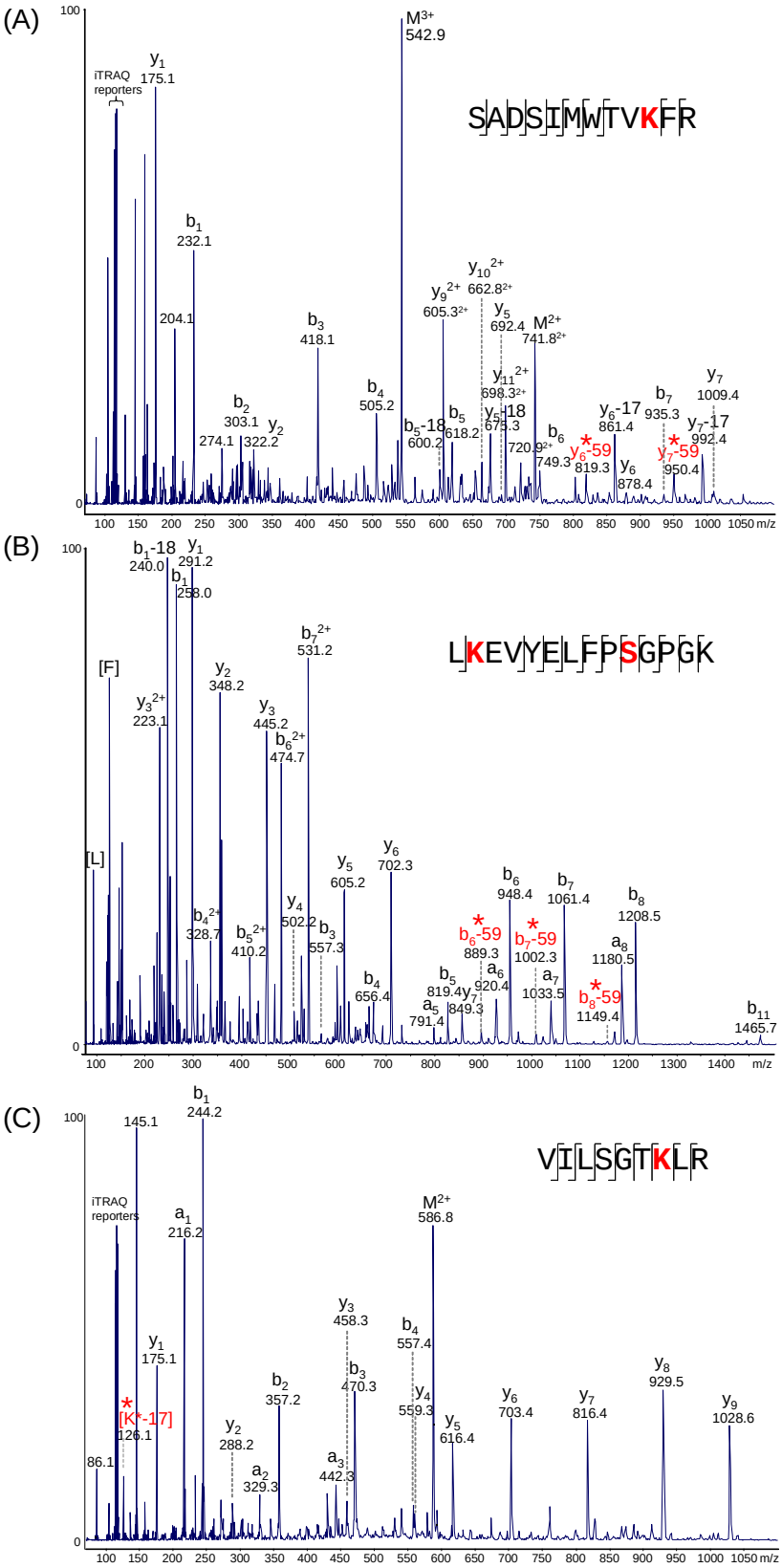

(A)
SADSIMWTVKFR
iTRAQ reporters
M3+ 542.9
y1 175.1
b1 232.1
204.1
b3 418.1
y10 2+ 662.8 2+
y9 2+ 605.3 2+
y5 692.4
M2+ 741.8 2+
y11 2+ 698.3 2+
b2 303.1
274.1
322.2
y2
b4 505.2
b5-18 600.2
b5 618.2
y5-18 675.3
b6 720.9 2+
749.3
y6-59 819.3
y6 878.4
y6-17 861.4
b7 935.3
y7 1009.4
y7-17 992.4
y7-59 950.4
(B)
LKEVYELFPSGPGK
b1-18 240.0
y1 291.2
b1 258.0
[F]
b2 2+ 531.2
y2 348.2
y3 2+ 223.1
y3 445.2
b6 2+ 474.7
[L]
y5 605.2
y6 702.3
b6 948.4
b7 1061.4
b8 1208.5
b4 2+ 328.7
b5 2+ 410.2
y4 502.2
b3 557.3
b6-59 889.3
b7-59 1002.3
a8 1180.5
b4 656.4
b5 819.4
a6 920.4
a5 791.4
y7 849.3
a7 1033.5
b8-59 1149.4
b11 1465.7
(C)
VILSGTKLR
b1 244.2
145.1
iTRAQ reporters
a1 216.2
M2+ 586.8
y3 458.3
b4 557.4
y1 175.1
b3 470.3
b2 357.2
y4 559.3
y8 929.5
y6 703.4
y7 816.4
y9 1028.6
[K*-17] 126.1
y2 288.2
y5 616.4
86.1
a3 442.3
a2 329.3

Figure 3.

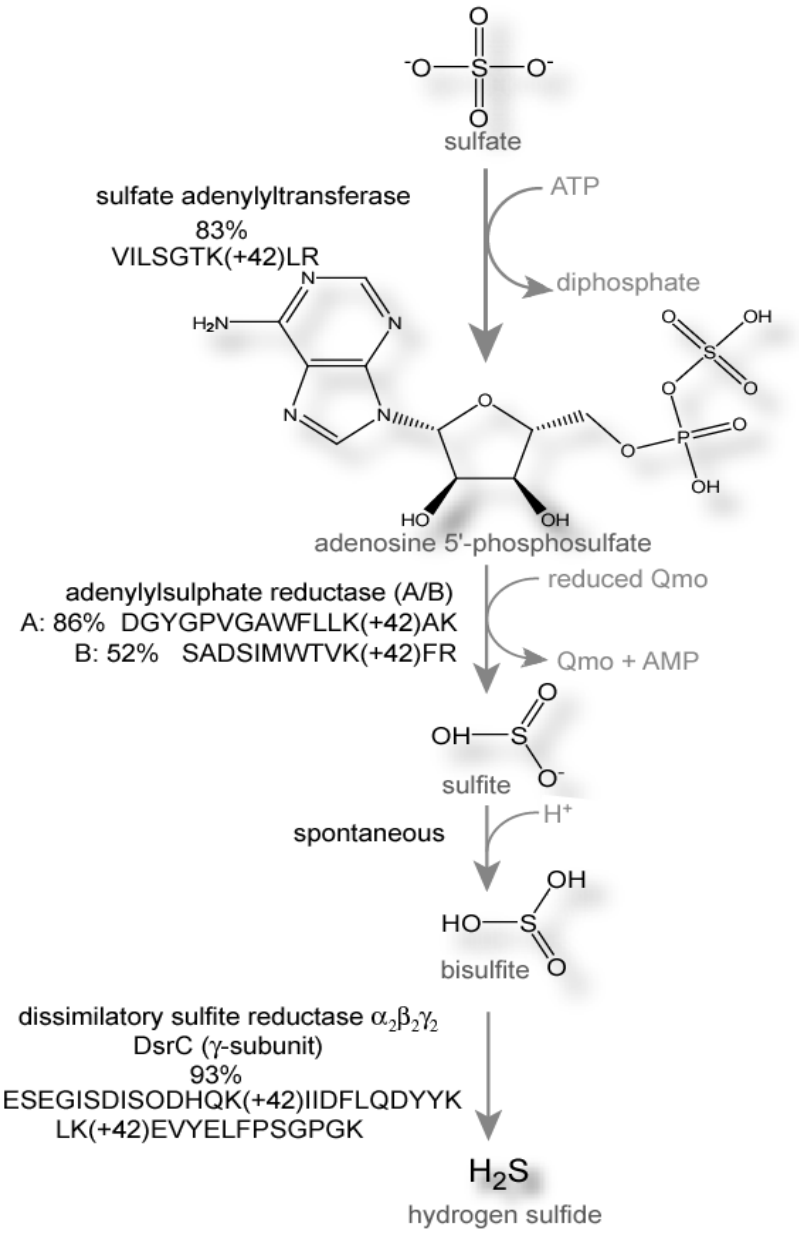
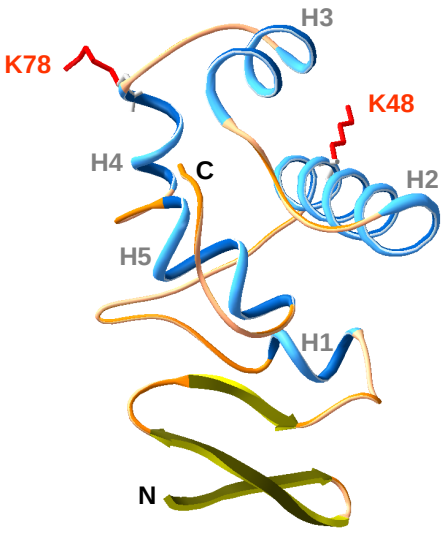

sulfate
sulfate adenylyltransferase
83%
VILSGTK(+42)LR
ATP
diphosphate
adenosine 5'-phosphosulfate
adenylylsulphate reductase (A/B)
A: 86% DGYGPVGAWFLLK(+42)AK
B: 52% SADSIMWTVK(+42)FR
reduced Qmo
Qmo + AMP
sulfite
spontaneous
H+
bisulfite
dissimilatory sulfite reductase α2β2γ2
DsrC (γ-subunit)
93%
ESEGISDISODHQK(+42)IIDFLQDYYK
LK(+42)EVYELFPSGPGK
H2S
hydrogen sulfide

Figure 4.

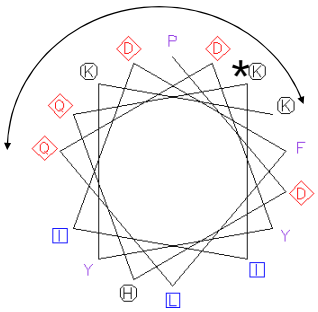
(A)

```
D. vulgaris             --------------------MAE-----VTYKGKSFEVDEDGFLLRFDDWCPEWVEYVK- 34
D. desulfuricansG20     --------------------MAE-----VSFQGKTFEVDEDGFLLRFDEWCPEWVEYVK- 34
A. fulgidus             --------------------MP-----ELEVKGKKLRLDEDGFLQDWEEWDEEVAEALAK 35
P. aerophilum           --------------------MPVKCPGEYQVDGKKVILDEDCFMQNPEDWDEKVAEWLAR 40
A. vinosum              MGSSHHHHHHSSGLVPRGSHMADT----IEVDGKQFAVDEEGYLSNLNDWVPG-VADVM- 54

D. vulgaris             ESEGISD---ISPDHQKIIDFLQDYYKKNGIAPMVRILSKNTGFKL-------KEVYELF 84
D. desulfuricansG20     ESEGIAE---ITEDHQKIIDFLQDYYRKNGIAPMVRILSKNTGFKL-------KQVYELF 84
A. fulgidus             DTRFSPQPIELTEEHWKIIRYLRDYFIKYGVAPPVRMLVKHCKKEV-RPDCNLQYIYKLF 94
P. aerophilum           ELEGIQK---MTEEHWKLVKYLREYWETFGTCPPIKMVTKETG-------FSLEKIYQLF 90
A. vinosum              AKQDNLE---LTEEHWDIINFLREYYEEYQIAPAVRVLTKAVGKKLGKEKGNSKYLYSLF 111

D. vulgaris             PSGPGKGACKMAGLPKPTGCV 105
D. desulfuricansG20     PSGPGKGACKMAGLPKPTGCV 105
A. fulgidus             PQGPAKDACRIAGLPKPTGCV 115
P. aerophilum           PSGPAHGACKVAGAPKPTGCV 111
A. vinosum              PYGPAKQACRFAGLPKPTGCV 132
```

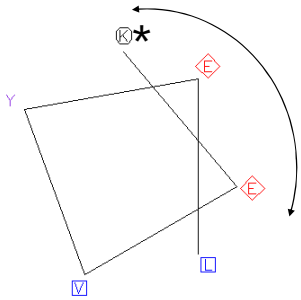
(B)



(C)



H2: PDHQK*IIDFLQDYYKK



H4: K*LKEVYEL

Table 1. Summary of amino acid modifications found in Dataset 1 searches. Each unique peptide sequence was counted only once, where the same sequence modified in 2 different ways would count as 2 sequences. Supplementary table 2 lists the parameters used in each search specified.

Putative Modification	Amino Acid	Delta Mass (Da)	# Unique Sequences	Results from Search #
iTRAQ	K and/or N-term	144	6175	1
	Y	144	318	1
	S	144	1	9
	T	144	2	9
	C	144	0	3
pyroglu	N-term Q	-17	76	4
deamidation	N	1	345	5
	Q	1	98	6
sodium adduct	C-term	22	4	7
	DE	22	17	7
potassium adduct	C-term	38	5	8
	DE	38	47	8
oxidation	M	16	118	1
	M	32	7	10
	C	16	1	10
	C	32	12	10
	C	48	2	10
	P	16	34	16
	F	16	11	20
	H	16	7	21
	W	16	1	22
methylation	K	14	11	12
	K	28	21	12
	R	14	11	15
	R	28	15	15
	H	14	13	13
	DE	14	69	14
	NQ	15	23	17
	C	14	5	19
trimethylation or acetylation	K	42	18	11

Table 2. Summary of amino acid modifications found for Datasets 2-4. Each unique peptide sequence was counted only once, where the same sequence modified in 2 different ways would count as 2 sequences. Supplementary Table 2 lists the parameters used in each search specified.

Putative Modification	Amino Acid	Delta Mass (Da)	# Unique Sequences			Results from Search #
			Biofilm	Air	O_2	
	K and/or N-term	144	4573	2745	610	1
iTRAQ	Y	144	55	29	13	1
	M	16	567	168	35	1
oxidation	P	16	75	55	30	16
	K	14	10	5	3	12
	K	28	28	33	6	12
methylation	DE	14	57	80	25	14
trimethylation or acetylation	K	42	16	9	5	11

Table 3. Specific sequences found modified with K(+42) in multiple samples passing manual inspection.

Protein	Sequence	Sample	Score
ApsB adenylylsulphate reductase β DVU0846	SADSIMWTVK(+42)FR	biofilm	8.30E-03
		nitrate	7.70E-03
		O_2	1.50E-03
ApsA adenylylsulphate reductase α DVU0847	DGYGPVGAWFLLFK(+42)AK	air	5.50E-09
		biofilm	6.70E-09
		nitrate	6.70E-09
		O_2	1.60E-08
Sat sulfate adenylyltransferase DVU1295	VILSGTK(+42)LR	air	7.20E-04
		nitrate	4.70E-04
DsrC dissimilatory sulfite reductase γ DVU2776	ESEGISDISPDHQK(+42)IIDFLQDYYK	air	7.10E-05
		biofilm	3.70E-04
		nitrate	2.30E-06
	LK(+42)EVYELFPSGPGK	biofilm	4.00E-04
		nitrate	7.50E-04
		O_2	2.20E-04
RplK ribosomal protein L11 DVU2924	TMEQK(+42)GMITPVVITVYADR	biofilm	2.70E-08
		nitrate	4.70E-07

TABLE OF CONTENTS SYNOPSIS AND GRAPHIC

Several proteomic mass spectrometric datasets were acquired using orthogonal LC/MS/MS workflows in the sulfate-reducing organism, *Desulfovibrio vulgaris* Hildenborough. These datasets were mined for post translational modifications, leading to the discovery of many modified peptides. Several peptides contained acetylated or trimethylated lysine residues in proteins belonging to the sulfate reduction pathway (SRP). The SRP is encoded by genes unique to sulfate reducing bacteria, where sulfate serves as the primary electron acceptor.