MCP Papers in Press. Published on August 23, 2008 as Manuscript M800187-MCP200

LYSINE ACETYLATION IS A HIGHLY ABUNDANT AND EVOLUTIONARILY CONSERVED MODIFICATION IN E. COLI

Junmei Zhang[1*], Robert Sprung[1*], Jimin Pei[1, 2], Xiaohong Tan[3], Sungchan Kim[4], Heng Zhu[5], Chuan-Fa Liu[3], Nick V. Grishin[1, 2], Yingming Zhao[1]

Department of Biochemistry[1], Howard Hughes Medical Institute[2], UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390

Division of Chemical Biology and Biotechnology[3], School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

Department of Biochemistry[4], College of Medicine, Hallym University, Chuncheon, Kangwon-Do, Korea

Department of Pharmacology and Molecular Sciences[5], Johns Hopkins University School of Medicine, 733 North Broadway, Baltimore, Maryland 21205

* The authors contributed equally to this work.

Address correspondence to: Yingming Zhao, Ph.D, Department of Biochemistry, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, Texas 75390-9038

Fax: 214-648-2797; Tel: (214) 648-7947; Email: Yingming.Zhao@UTSouthwestern.edu

Running title: Lysine acetylation in *E. coli*

Summary

Lysine acetylation and its regulatory enzymes are known to have pivotal roles in mammalian cellular physiology. However, the extent and function of this modification in prokaryotic cells remain largely unexplored, thereby presenting a hurdle to further functional study of this modification in prokaryotic systems. Here we report the first global screening of lysine acetylation, identifying 138 modification sites in 91 proteins from *E. coli*. None of the proteins has been previously associated with this modification. Among the identified proteins are transcriptional regulators, as well as others with diverse functions. Interestingly, more than 70% of the acetylated proteins are metabolic enzymes and translation regulators, suggesting an intimate link of this modification to energy metabolism. The new dataset suggests that lysine acetylation could be abundant in prokaryotic cells. In addition, these results also imply that functions of lysine acetylation beyond regulation of gene expression are evolutionarily conserved from bacteria to mammals. Furthermore, we demonstrate that bacterial lysine acetylation is regulated in response to stress stimuli.

Introduction

Lysine acetylation is a dynamic, reversible, and regulatory post-translational modification in mammalian cells. Lysine-acetylation status and its regulatory enzymes have been shown to influence several fundamental cellular pathways in mammalian cells, including cell survival and apoptosis, cellular differentiation, and metabolism. Dysregulation of the modification is associated with aging (1) and a few diseases, such as cancer, cardiovascular diseases, and neurodegenerative diseases (1-5).

Emerging evidence suggests diverse non-nuclear roles of lysine acetylation and its regulatory enzymes, in addition to its well-recognized functions in DNA-templated processes. For example, Sirt1 modulates diverse cellular processes, such as fat metabolism, insulin production, glucose homeostasis, and cell survival, of which some are mediated through non-nuclear proteins (6-8). Three members of the sirtuin family (Sirt3, Sirt4, and Sirt5) are located in the mitochondrion (9). Activation of Sirt1, a mammalian ortholog of yeast Sir2, by resveratrol leads to diverse metabolic changes in animals (10). The roles of lysine acetylation in metabolism are further exemplified by the facts that the modification is present in more than 20% of mitochondrial proteins and is highly enriched among metabolic enzymes (11-14). While the inventory and biological functions of lysine-acetylation substrates in eukaryotic cells have begun to unfold, especially in histone proteins and transcription factors (15-19), the nature of lysine acetylation substrates in prokaryotic cells remains largely unknown.

The high abundance of lysine acetylation in mammalian mitochondrial proteins implies the possible widespread existence of the modification in prokaryotes, given the evolutionary lineage of eukaryotic mitochondria from bacteria (20). Acetyl-coenzyme A (CoA) synthetase (Acs), CheY, and Alba are the only proteins in prokaryotes known to be lysine-acetylated (21-25). In *S. enterica,* the lysine-acetylation status of Acs is regulated by CobB deacetylase, a Sir2 homolog in bacteria (21,26), as well as Pat acetyltransferase (27). Acs activation by CobB sirtuin deacetylation is required for the bacteria to grow on short-chain fatty acids such as acetate and propionate. Interestingly, both human Sirt2 and yeast Sir2 proteins could restore growth of CobB sirtuin-deficient strains of *S. enterica* on short-chain fatty

acids, suggesting that the sirtuins may have evolutionarily conserved roles in cellular metabolism (28). Despite evidence of the presence and roles of lysine acetylation in prokaryotes, the extent of lysine acetylation has not been carefully examined before.

Here we report the first proteomics screening of lysine acetylation in *E. coli*. The proteomics study involves efficient affinity enrichment of lysine acetylated, tryptic peptides with anti-acetyllysine antibodies and subsequent peptide identification for nano-HPLC/mass spectrometric analysis. The screening identified 138 lysine acetylation sites in 91 proteins in *E. coli*, of which 25% have mammalian orthologues. Our results suggest that diverse groups of bacterial proteins are the substrates of lysine acetylation, including metabolic enzymes, stress response proteins, and transcription and translation factors. The lysine acetylation substrates in *E. coli* are highly enriched in metabolic enzymes (~53%) and proteins involved in translation (~22%), two processes that are intimately linked to cellular energy status. These data therefore reveal previously unappreciated roles of lysine acetylation in the regulation of prokaryotic biochemical pathways and imply that DNA-independent functions of lysine acetylation are evolutionarily conserved from prokaryotic to eukaryotic cells.

Experimental Procedures

Materials-The reagents used in this work include Protein A conjugated agarose beads from Amersham Biosciences (Uppsla, Sweden); Luria-Bertani (LB) medium from Life Technologies (Carlsbad, CA); iodoacetamide, C_{18} ziptips from Millipore Corp. (Bedford, MA); Luna C_{18} resin from Phenomenex (Torrance, CA). *E. coli* Strains MG1655 and JW1106 were both acquired from the Coli Genetic Resource Center at Yale University. MG1655 is the wild type and JW1106 is a CobB-deficient single-gene knockout of the Keio Collection (29).

Two anti-acetyllysine antibodies were used: an affinity-purified anti-acetyllysine polyclonal antibody from ImmuneChem Pharmaceuticals Inc. (Burnaby, British Columbia, Canada) and an anti-acetyllysine monoclonal antibody from Cell Signaling Technology (Boston, MA). We reasoned, because

the two antibodies were generated using different antigens (acetyllysine peptide library or acetyllysine residue, please see the vendors' information for details) and purified differently (30,31), they would likely have different binding specificities. Accordingly, a more diverse panel of acetyllyisne substrate peptides would be identified by using two antibodies for the described proteomics screening.

Preparation of cell lysate from E. coli-E. coli DH5 was grown aerobically in LB medium at 37 °C. The cultured cells were harvested during the exponential growth phase by centrifugation at 4500 × *g* for 10 min and washed twice by resuspension of the pellet in ice-cold PBS buffer (0.1 M Na_2HPO_4, 0.15 M NaCl, pH 7.2). The cells were resuspended in chilled lysis buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM DTT) and then sonicated with 12 short bursts of 10 s followed by intervals of 30 s for cooling. Unbroken cells and debris were removed by centrifugation at 4 °C for 30 min at 21,000 × *g*. The supernatant was divided into aliquots and stored at -80 °C until use.

In-solution tryptic digestion-Five milligrams of proteins were precipitated with acetone, followed by centrifugation at 22,000 × *g* for 10 min. The resulting pellet was digested according to a previously described procedure (32). The protein pellet was rinsed twice with cold acetone to remove residual salts, resuspended in 50 mM NH_4HCO_3 (pH 8.5) (the protein pellet was not completely re-dissolved but rather was suspended as small particles), and digested with trypsin (Promega, Madison, WI) at an enzyme-to-substrate ratio of 1:50 for 16 h at 37 °C to enhance the solubility of the proteins prior to reduction and alkylation. The tryptic peptides were reduced with 5 mM DTT at 50 °C for 30 min and then alkylated using 15 mM iodoacetamide at ambient temperature for 30 min in darkness. The reaction was terminated with 15 mM cysteine at ambient temperature for 30 min. To ensure complete digestion, additional trypsin at an enzyme-to-substrate ratio of 1:100 was added to the peptide mixture, and the mixture was incubated for an additional 3 h.

Affinity purification of lysine-acetylated peptides-The anti-acetyllysine antibodies from ImmuneChem Pharmaceuticals Inc. and Cell Signaling Technology were mixed at a ratio of 1:1 and then immobilized on protein A-conjugated agarose beads at 4-6 mg/ml by incubation at 4 °C for 4 h. The supernatant was

removed and the beads were washed three times with NETN buffer (50 mM Tris·HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% NP40).

The tryptic peptides obtained from in-solution digestion were re-dissolved in NETN buffer. Insoluble particles were removed by centrifugation. Affinity purification was carried out by incubating the peptides with 20 μl of anti-acetyllysine antibody-protein A-immobilized agarose beads at 4 °C for 6 h with gentle shaking. The beads were washed three times with 1 ml of NETN buffer and twice with ETN (50 mM Tris·HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA). The bound peptides were eluted from the beads by washing three times with 50 μl of 1% TFA. The eluates were combined and dried in a SpeedVac. The resulting peptides were cleaned with C18 ZipTips (Millipore, Bedford, MA) according to the manufacturer's instructions, prior to nano-HPLC/mass spectrometric analysis.

HPLC/MS/MS Analysis-HPLC/MS/MS analysis was performed in an integrated system that includes an Agilent 1100 series nanoflow LC system (Agilent, Palo Alto, CA) and a LTQ 2D trap mass spectrometer (Thermo Electron, Waltham, MA) equipped with a nanoelectrospray ionization (NSI) source. One μl of tryptic peptides in buffer A (97.95% water/2% acetonitrile/0.05% acetic acid) was manually injected and separated in a capillary HPLC column (11 cm length × 75 μm I.D.) packed in-house with Luna C18 resin (5 μm particle size, 100 Å pore diameter, Phenomenex, Torrance, CA). Peptides were eluted from the column with a gradient of 6.0% to 90% buffer B (90% acetonitrile/9.95% water/0.05% acetic acid) in a 2 h LC/MS/MS analysis. The eluted peptides were electrosprayed directly into the LTQ ion trap mass spectrometer. LC/MS/MS was operated in a data-dependent mode such that the ten strongest ions in each MS scan were subjected to collisionally activated dissociation (CAD) with a normalized CAD energy of 35%.

Protein sequence database search and manual verification-Tandem mass spectra were used to search the *E. coli* entries (51,059 *E. coli* sequences) of the NCBI-nr database (updated July 31th, 2006 with a total of 3,841,279 sequences). Only the *E. coli* subset of the database was used for search because we are only interested in *E. coli* acetylation in the current study and all the cell lysate was from *E. coli*. The search

engine MASCOT (version 2.1, Matrix Science, London, U.K.) was used for database search, and extract_msn.exe version 4.0 was used for peaklist generation. A low cutoff of peptide score 20 was selected to maximize the identification of lysine-acetylated peptides. Trypsin was specified as the proteolytic enzyme and up to 6 missed cleavage sites per peptide were allowed. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine and acetylation of lysine as variable modifications. Charge states of +1, +2 or +3 were considered for parent ions. Mass tolerance was set to ±4.0 Da for parent ion masses and ± 0.6 Da for fragment ion masses. Acetylated lysine containing peptides identified with a Mascot score of 25 were manually verified by the method previously described (33).

Western blotting analysis-E. coli MW1655 and JW1106 were grown, harvested and lysed as described above. Protein concentration was determined using the Bradford assay (Bio-Rad, Hercules, CA). Forty μg of proteins from *E. coli* cells was resolved by 10% SDS-PAGE and ttransferred to a PVDF membrane. The membrane was blocked with 5% milk at ambient temperature for 1 hr. Then the membrane was incubated with anti-acetyllysine monoclonal antibody (0.8 μg/ml in TBST with 3% BSA; from Cell Signaling Technology) overnight at 4 ºC. After washing with TBST four times for 5 min each, the membrane was incubated with HRP-conjugated anti-mouse IgG (1 μg/ml in TBST with 3% BSA) at ambient temperature for 2 hr. The ECL system (PerkinElmer LAS, Inc., Boston, MA) was used for signal detection. To carry out the competition experiment, the anti-AcK antibody was pre-incubated with 3% BSA (acetylated or non-acetylated) at ambient temperature for 2 hr before it was incubated with the membrane.

Hypoxic treatment of E. coli-Hypoxic treatment of *E. coli* was carried out as previously described (34). *E. coli* MG1655 was grown in LB medium in 125 ml flasks. When its optical density at 600 nm (OD_{600}) reached 0.3, it was exposed to one of the following conditions: nitrogen only, 5% air or 25% air. For these three experiments, nitrogen was bubbled through the cell suspension for 15 min and the flasks were sealed tightly immediately. For the latter two experiments, 5% or 25% air was then introduced back to the

flasks by a syringe needle and further sealing was applied immediately. The cells were allowed to continue to grow until their OD_{600} reached 0.6. The cells were harvested and prepared for western blotting analysis as described above.

Starvation of E. coli-Starvation of *E. coli* was carried out as previously described (35). *E. coli* MG1655 was grown in LB medium as described above. When its OD_{600} reached 0.3, starvation was imposed by sudden depletion of all carbon sources as follows: cells were centrifuged at 4500 × *g* for 10 min at 4° C and washed twice with two volumes of sterile ice-cold M9 minimal medium (no carbon source). The cells were resuspended in the same volume of M9 minimal medium and incubated overnight. The cells were harvested and prepared for western blotting analysis as described above.

Structure analysis of E. coli lysine acetylation sites-For each acetylated *E. coli* protein, BLAST was run against a database of domain sequences with known structures from the SCOP90 representative set of ASTRAL compendium (version 1.71) (36,37). Database size of BLAST was set to the size of the protein nr database as of April 6th, 2007 (39,280,211,952 letters) to impose a stringent e-value cutoff. Hits with an e-value less than 0.001 were analyzed. We identified homologous structures for 69 of the 91 acetylated proteins (~76%). For these proteins, we mapped the positions of acetylated lysines to the model structures using BLAST local alignments and visually inspected the crystal structures to determine the role of the conserved lysine in substrate and protein binding or catalytic activity.

Results and Discussion

Proteomics screening of lysine acetylation. Lysine acetylation is more difficult to identify by a candidate approach than protein phosphorylation due to the low radioactivity of [^{14}C]-acetyl-CoA and the weak binding affinity of anti-acetyllysine antibody. A lack of protein substrates represents one of the major bottlenecks for characterization of its biological functions. To begin the systematic study of lysine acetylation in prokaryotes, we carried out the first proteomics screening of lysine-acetylated substrates in bacteria. The goals of this study were (i) to determine the spectrum and extent of lysine acetylation in

bacteria, (ii) to identify novel lysine-acetylation substrates and lysine-acetylation sites that could provide candidate proteins for further functional studies, and (iii) to define the molecular pathways that are likely to be affected by lysine acetylation.

The proteomics of lysine acetylation was carried out as previously reported (11), and consisted of four steps: (i) The protein lysate of *E. coli* was proteolytically digested by trypsin, (ii) the resulting tryptic peptides were subjected to affinity purification by anti-acetyllysine antibody, (iii) the isolated, lysine-acetylated peptides were then analyzed by nano-HPLC/MS/MS for peptide identification and precise localization of lysine-acetylation sites, and (iv) the peptide candidates were further manually evaluated to ensure the accuracy of the identification (Fig. 1A,C). The raw spectrum of each acetylated peptide can be found in Supplemental Table 1.

The strategy described here, integration of immunoisolation with mass spectrometry for characterization of biomolecules, can be traced back more than seventeen years ago (38). Peptides released from immunoisolated complexes have been analyzed by both MALDI-TOF mass spectrometry and electrospray tandem mass spectrometry for epitope mapping (39,40) , sequencing MHC-binding peptides (41), and analysis of disease-related peptides (42). In addition to sequence-specific antibodies, pan-antibodies, such as anti-phosphotyrosine antibody, have also been used to isolate and to identify tyrosine-phosphorylated proteins on a global scale in response to extracellular stimulation (43,44). Isolation of modified peptides from tryptic digests by immunoaffinity purification using a pan-antibody is much simpler than that of the corresponding proteins for three obvious reasons. First, the modified residue will not be buried in peptides. In contrast, the modified residue may not be accessible for antibody binding in the context of proteins due to protein folding or non-covalent interactions (e.g., phosphotyrosine with SH_2 domain). Second, a protein typically has more complex domains with variant surface properties, leading to higher non-specific binding during immunoprecipitation. Finally, a significant portion of proteins are denatured and subsequently precipitate during immunoisolation, therefore leading to more contaminant proteins. In contrast, peptides are more difficult to precipitate due to their small size and a lack of hydrophobic core structure. Nevertheless, immunoisolation using tryptic

peptides has been successfully combined with mass spectrometry for proteomics of protein modifications such as tyrosine phosphorylation and lysine acetylation (11,45).

Western blotting analysis demonstrates that proteins of a wide molecular weight range can be lysine-acetylated (Fig. 1B). Subsequent proteomics screening identified 138 lysine-acetylation sites among 91 proteins (Supplemental Table 2). To the best of our knowledge, none of these bacterial proteins has been associated with the modification before.

Lysine-acetylated protein groups in E. coli. An unbiased screen of a large set of lysine acetylation substrates can provide insight into cellular pathways that would not be apparent through single-protein analysis. This has been exemplified by our recent study on proteomics of lysine acetylation in mammalian cells (11). We therefore attempted to assign each lysine-acetylation substrate to a functional group based on Gene Ontology (GO) molecular functions or biochemical process groups, or previously published literature (Fig. 2A).

Several of the protein groups defined in this way are of particular interest. For example 48 of the 91 lysine-acetylation substrates (~53%) are metabolic enzymes, including 4 TCA cycle proteins, 7 glycolytic enzymes and several enzymes involved in the metabolism of nucleotides and amino acids (Table 1). A few translational regulators are found to be lysine acetylation substrates in mammalian cells. However, 22% of lysine-acetylation substrates identified here are proteins that are either subunits of translational machinery or are enzymes associated with translational processes, such as the aminoacyl tRNA synthetases (Supplemental Table 2). Proteins involved in stress response represented ~5% of lysine acetylation substrates (Supplemental Table 2). Transcription factors are well-known examples of lysine-acetylation targets in eukaryotic cells. Our screening identified 2 bacterial transcriptional regulators and one RNA polymerase subunit as substrates of lysine acetylation (Supplemental Table 2).

Lysine-acetylated substrate proteins in bacteria. The identification of metabolic enzymes, which have previously been shown to be lysine- acetylated in the mammalian mitochondrion, represented a critical validation of functional roles of the modification in energy metabolism. Identification of a large number of metabolic enzymes and TCA proteins is reminiscent of lysine acetylation substrates identified in

mitochondria, providing evidence supporting the importance of the modification in the regulation of energy metabolism.

Glycolytic enzymes: We detected 7 of 9 glycolytic enzymes, catalyzing the key reactions to degrade glucose to pyruvate, as subjects of lysine acetylation (Fig. 3). These proteins include phosphoglucose isomerase, aldolase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase, and pyruvate kinase. Three of these enzymes were also found to be acetylated in mammalian samples, suggesting potentially conserved functional consequences for the modification in the regulation of glycolyic flux.

Pyruvate dehydrogenases: Pyruvate is one of the major products of glycolysis. It can be cleaved by either pyruvate dehydrogenase or pyruvate formate-lyase, of which both were found to be lysine acetylated. The oxidative decarboxylation of pyruvate by pyruvate dehydrogenase generates acetyl-CoA and one NADH, and is therefore functionally restricted to respiratory metabolism under aerobic conditions. All three subunits of pyruvate dehydrogenase are subjects of lysine acetylation. On the other hand, non-oxidative cleavage of pyruvate to acetyl CoA and formate is catalyzed by pyruvate formate-lyase and this enzyme is functional only anaerobically. The activities of the two enzymes are usually mutually exclusive. Pyruvate formate-lyase is a homodimer that catalyzes the conversion of pyruvate to acetyl-CoA, which is in turn used to produce ATP through acetyl phosphate. The ATP produced in this manner is the single ATP source in *E. coli* cells under anaerobic conditions with pyruvate as the sole carbon and energy source (46). In *E. coli* cells, this protein's activity is tightly regulated at both the transcriptional level and by post-translational modification.

TCA cycle proteins: Four of the eight TCA cycle proteins, citrate synthase, isocitrate dehydrogenase, and succinate dehydrogenase catalytic subunits, 2-ketoglutarate dehydrogenase (E3 dihydrolipoamide dehydrogenase subunit shared by both 2-ketoglutarate dehydrogenase and pyruvate dehydrogenase) are lysine acetylated (Fig. 3). Among the four proteins, citrate synthase and isocitrate dehydrogenase are known to be regulated at the level of transcription and protein modification. The synthesis of citrate synthase is subject to catabolite repression. It is repressed by glucose and anaerobiosis, and induced by

acetate and oxygen. Isocitrate dehydrogenase catalyzes the conversion of isocitrate to α-ketoglutarate and CO_2, which can be a rate-limiting step in the citric acid cycle, and is also involved in the regulation of carbon flux at the branch point between the TCA and glyoxylate cycles. Protein phophorylation is also known to regulate the protein's functions (47,48).

Enzymes involved in the metabolism of amino acids and nucleotides are also lysine acetylated. These proteins include 7 proteins involved in amino acid metabolism and 7 proteins in nucleotide metabolism.

One enzyme involved in nucleotide metabolism, deoxyribosephosphate aldolase, is of particular interest. The enzyme catalyzes the reversible reaction between glyceraldehyde-3-phosphate and acetaldehyde to form 2-deoxyribose-5-phosphate. We identified this enzyme as being acetylated at lysine 167. This residue has been implicated in the enzyme's catalytic mechanism, forming a Schiff base with the substrate (49). Acetylation of this active site lysine would therefore be expected to block enzymatic activity. Such acetylation of active site lysine may represent a common regulatory mechanism among the aldolase family of proteins. Our previous survey of lysine acetylation among mammalian proteins identified acetylation of lysine 146 of the glycolytic enzyme fructose-1, 6-bisphosphate aldolase (11), which is also the key active site catalytic residue involved in Schiff base formation.

Transcription and translation factors: Transcription factors and histones are founding members of lysine acetylation substrates in mammalian cells. Reminiscent of this, two transcription factors (cAMP receptor protein and trp repressor binding protein) and one subunit of RNA polymerase (rpoB) are lysine-acetylation substrates. In addition to transcriptional regulators, 20 proteins involved in translational regulation are subjects of lysine acetylation. These proteins include translational elongation factors, ribosomal proteins as well as aminoacyl tRNA synthetases. Lysine acetylation has not been detected in ribosomal proteins nor aminoacyl tRNA synthetases in mammals, suggesting that the modification of such protein may have been lost during evolution.

A large portion of cellular energy is used in protein synthesis. Therefore, translational machinery needs to be synchronized with energy availability. It is not surprising that the protein expression level for

about half of the aminoacyl tRNA synthetases and for translational factors are under metabolic control (50). Our screening identified lysine acetylation at 5 aminoacyl tRNA synthetases, 11 ribosomal subunits, and three translational machinery proteins, implying that post-translational modifications might provide an alternative avenue to modulate the activity of cellular translational machinery.

Stress response proteins: In eukaryotic cells, activation of the sirtuin family of deacetylases promotes cell survival and resistance to stress, implying that the lysine acetylation status of proteins may be altered in response to stress. Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan (51). Our previous screening identified heat shock proteins as well as ROS regulators such as SOD as substrates of lysine acetylation (11). Like eukaryotic cells, a significant portion of stress response proteins in *E. coli* are also substrates of lysine acetylation, including chaperones and proteins involved in regulation of free radical reduction (such as superoxide dismutase, alkyl hydroperoxide reductase and thioredoxin). In *E. coli*, loss of alkyl hydroperoxide reductase and SOD leads to an increased sensitivity to hydrogen peroxide and alkyl hydroperoxide oxidative stress (52,53).

Several respiratory components of the mammalian mitochondrion are closely related to those in *E. coli*. Mitochondrial NADH dehydrogenase (complex I), succinate dehydrogenase (complex II), and cytochrome c oxidase (complex IV) are mammalian homologs of *E. coli* proteins that are lysine acetylated.

Evolutionary conservation of lysine acetylation between E. coli and mammalian cells. The mitochondrion is considered to have evolved from α-proteobacteria. Phylogenetic studies suggest that about one tenth of yeast mitochondrial proteins could be evolved from ancestral free-living α-proteobacteria whereas the remaining 90% were recruited from the nuclear genome of the eukaryotic host (54).

The results of our screen for lysine- acetylated proteins in *E. coli* suggest that regulatory programs tuning the activities of metabolic enzymes may be conserved between *E. coli* and mammalian cells. This concept is supported by the overlap in protein identities between the mammalian and *E. coli* datasets and the evidence of a common regulatory modification observed among the active site lysine of

aldolase family members. Twenty-two of the 91 *E. coli* lysine-acetylated proteins (~25%) have mammalian homologs, based on BLAST sequence alignment with 25% sequence identity as the cutoff score (Supplemental Table 2). We believe that the number is significantly underestimated because of the low percentage of lysine-acetylated proteins identified in mammalian cells that was caused by the wide dynamic range and limited sensitivity in our previous proteomics screening (11). Identification of a large number of lysine-acetylated orthologs between *E. coli* and mammalian cells, the abundance of lysine acetylation in both mitochondrion and *E. coli*, and the evolutionary linkage between the mitochondrion and *E. coli* suggest that the modification is likely to be evolutionarily conserved from bacteria to mammalian cells.

Lysine acetylation motifs in bacterial substrate proteins and bioinformatics analysis. Conserved protein sequence motifs are associated with some post-translational modifications such as protein phosphorylation. Structural studies of GCN5 HAT and H3 tail peptides suggest a recognition site of G-K-X-P (55). The motifs for lysine acetylation remain largely unknown. The dataset of acetylation sites identified from *E. coli* proteins allows us to conduct a preliminary analysis of motif preference (Fig. 2B). Such analysis identified histidine and tyrosine as preferred amino acid resides at the +1 position.

In contrast to the *E. coli* lysine-acetylation motifs, mammalian lysine-acetylation substrates in general show less preference for residues surrounding acetyllysine residues (11). This could be caused by averaged motif information from diverse acetyltransferases. Alternatively, substrates may be recognized through the three dimensional structure of the substrate proteins rather than their linear sequence. However, when considering the mitochondrial subset of lysine acetylation substrates identified in the mammalian dataset, a preference for histidine and tyrosine at the +1 position was also observed (11). The preference of amino acid residues flanking acetyllysine residues in *E. coli* proteins suggests that the modification might be catalyzed by a limited subset of acetyltransferases with unique substrate preferences. In addition, related acetyltransferases may be at work in the case of mitochondrial acetylation. Our results suggest the possibility that linear lysine acetylation motifs exist among lysine-acetylated proteins similar to motifs observed for phosphorylation.

To evaluate the structural features of acetyllysines in the substrate proteins, we first identified those proteins for which a homologous crystal structure was available. BLAST local alignment was used to map the position of the acetylated lysine onto the crystal structure. Visual inspection of the available crystal structures using the Pymol program suggested that most of the acetylation sites (95%) identified occur on the surface of proteins, rather than at active site or ligand binding residues. However, in some instances, modification sites were found within oligomerization interfaces, such as in serine hydroxymethyltransferase (PDB id - 1dfo) (56), isocitrate dehydrogenase (PDB id - 1hqs) (57) and IMP cyclohydrolase (PDB id - 1pkx) (58). This suggests that acetylation may affect the oligomerization of proteins or the formation of protein complexes. Such effects of lysine acetylation are known to occur in eukaryotes through the specific binding of acetylated lysine by bromodomains.

Lysine acetylation is a regulatory post-translational modification in E. coli. To test if lysine acetylation profiles can be regulated in *E. coli,* we carried out Western blotting analysis. The protein lysates from a wild-type *E. coli* strain (MG1655) and a CobB-deficient strain (JW1106) were prepared and resolved in SDS-PAGE to detect lysine-acetylated proteins by Western blotting analysis using an anti-acetyllysine antibody (Fig. 4A). Competition experiments using acetylated BSA suggest that most of the signals detected are specific to acetyllysine (Fig. 4A, lanes 1 and 2 vs lanes 3 and 4). Interestingly, the signals for four major bands increased between the wild type and CobB-deficent strain, while the majority was unchanged (Fig. 4A), suggesting the existence of other deacetylases in addition to CobB.

To further demonstrate that lysine acetylation profiles may be altered in response to stress, lysine acetylation profiles were analyzed by Western blotting analysis using the *E. coli* cell lysates from cells that were either starved or subjected to hypoxic conditions. Starvation has no obvious effect on the lysine acetylation profile in *E. coli* (Fig. 4B). However, when *E. coli* was exposed to hypoxia, significant changes were observed (Fig. 4C), suggesting that lysine acetylation is a dynamic and regulated process in *E. coli*.

It should be noted that our approach is capable of detecting only the most abundant substrate proteins. It is possible that changes in lysine acetylation for other less abundant proteins and substrate proteins with sequence motifs not recognized by the antibodies escaped our detection.

Several important points have emerged from this comprehensive protein screening of protein lysine acetylation. First, a large number of previously unknown lysine-acetylation substrates exist in prokaryotic species. Given the key roles of many of these substrate proteins in cellular functions and high sequence conservation among prokaryotic species, it is highly likely the modification is conserved and shares similar functions among prokaryotic species other than *E. coli*. Second, two preferred residues, histidine and tyrosine, at the +1 position of acetyllysine were identified, implying that linear sequence might be important for substrate binding and/or activity of acetyltransferase(s) in bacteria. This finding is intriguing given that a similar preference was observed among lysine-acetylated substrates identified in a screen of mitochondrial proteins, suggesting the existence of a related set of acetyltransferases in mitochondria (11). Third, the spectrum of lysine-acetylation substrates is evolutionarily well-conserved. A comparison of lysine-acetylation substrates from *E. coli* and mammalian cell lines suggests that three groups of substrate proteins are enriched in lysine acetylation, including metabolic enzymes, transcriptional/translational regulators, and stress response proteins. Identification of a high number of ribosomal proteins and aminoacyl tRNA synthetases was a surprise to us as these proteins are abundant in eukaryotic cells and lysine acetylation was not detected among these proteins in mammalian cells. Lysine acetylation of these proteins might have been lost during evolution.

Like our initial studies of lysine acetylation in mammalian cells (11), our study has limitations. First, only protein substrates present in medium to high abundance were identified. We might miss a significant portion of lysine acetylation substrates that are either of low abundance or modified at a low stoichiometry. Second, dynamic analysis has not been carried out under diverse genetic backgrounds and nutrient sources. Third, substrates of known acetyltransferases and deacetylases have not been defined. The answers for these questions await future research by quantitative proteomics of lysine acetylation with high sensitivity (e.g., protein pre-fractionation before proteomics studies).

Identification of a large number of lysine acetylation substrate raises many interesting questions. Only one acetyltransferase and one deacetylase have been described in prokaryotes. Identification of a large number of lysine acetylation substrate proteins raises the possibility of additional enzymes in bacteria that regulate the modification status. The recent discovery of a novel yeast acetyltransferase with no apparent sequence similarity with existing ones further supports such a possibility (59). Recently, we identified two novel *in vivo* lysine modifications, lysine propionylation at K5, K8, and K12 and lysine butyrylation at K5 and K12 of histone H4 (60). These lysine residues were found to be acetylated and methylated before (www.histone.com). Interestingly, two acetyltransferases, p300/CBP, can carry out *in vitro* lysine propionylation and lysine butyrylation *in vitro* on histones H3 and H4. It would be intriguing to know if the lysine-acetylated substrates can also be lysine-propionylated or lysine-butyrylated and the identities of the regulatory enzymes.

Our results provide a large number of future research opportunities in prokaryotic biology. The large datasets will provide protein leads for further genetic and molecular biological studies to test roles of lysine acetylation sites in cellular physiology. Prokaryotic species, such as *E. coli,* have been genetically engineered to boost protein expression or for fermentation of special chemicals, such as industrial ethanol and glycerol. Such genetic engineering typically takes advantage of overexpression of specific proteins and/or improving the activities of an enzyme with an optimized sequence. Nevertheless, these genetically engineered species have not fully taken advantage of the regulatory lysine-acetylation pathway. Given the essential roles of lysine acetylation in diverse metabolic pathways and importance of acetyl Co-A and NAD (the co-factors for lysine acetylation regulatory enzymes) in cellular metabolism, the modification is likely to play an important role in bacterial physiology. Therefore, understanding, capturing, and optimizing the modification landscape might provide a novel approach for the engineering of industrial bacterial species with high efficiency. The bioterror threat posed from certain bacteria (such as Bacillus anthracis) and the emergence of drug resistant bacteria remind us that the biology of prokaryotes remains to be further studied. Unfortunately, post-translational modifications in prokaryotes have been overlooked in the past. Generation of lysine-acetylation datasets reminds us of the need for

further studies of modification pathways in prokaryotes, of which some might provide good avenues for therapeutic intervention.

References

1. Haigis, M. C., and Guarente, L. P. (2006) *Genes Dev* **20**(21), 2913-2921
2. Hake, S. B., Xiao, A., and Allis, C. D. (2004) *Br J Cancer* **90**(4), 761-769
3. McKinsey, T. A., and Olson, E. N. (2004) *Trends Genet* **20**(4), 206-213
4. Yang, X. J. (2004) *Nucleic Acids Res* **32**(3), 959-976
5. Kouzarides, T. (2000) *Embo J* **19**(6), 1176-1179
6. Guarente, L., and Picard, F. (2005) *Cell* **120**(4), 473-482
7. Cohen, H. Y., Miller, C., Bitterman, K. J., Wall, N. R., Hekking, B., Kessler, B., Howitz, K. T., Gorospe, M., de Cabo, R., and Sinclair, D. A. (2004) *Science* **305**(5682), 390-392
8. Cohen, T., and Yao, T. P. (2004) *Sci STKE* **2004**(245), pe42
9. Michishita, E., Park, J. Y., Burneskis, J. M., Barrett, J. C., and Horikawa, I. (2005) *Mol Biol Cell* **16**(10), 4623-4635
10. Baur, J. A., Pearson, K. J., Price, N. L., Jamieson, H. A., Lerin, C., Kalra, A., Prabhu, V. V., Allard, J. S., Lopez-Lluch, G., Lewis, K., Pistell, P. J., Poosala, S., Becker, K. G., Boss, O., Gwinn, D., Wang, M., Ramaswamy, S., Fishbein, K. W., Spencer, R. G., Lakatta, E. G., Le Couteur, D., Shaw, R. J., Navas, P., Puigserver, P., Ingram, D. K., de Cabo, R., and Sinclair, D. A. (2006) *Nature* **444**(7117), 337-342
11. Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., Grishin, N. V., White, M., Yang, X. J., and Zhao, Y. (2006) *Mol Cell* **23**(4), 607-618

12. Schwer, B., Bunkenborg, J., Verdin, R. O., Andersen, J. S., and Verdin, E. (2006) *Proc Natl Acad Sci U S A* **103**(27), 10224-10229

13. Hallows, W. C., Lee, S., and Denu, J. M. (2006) *Proc Natl Acad Sci U S A* **103**(27), 10230-10235

14. Lombard, D. B.; Alt, F. W.; Cheng, H. L.; Bunkenborg, J.; Streeper, R. S.; Mostoslavsky, R.; Kim, J.; Yancopoulos, G.; Valenzuela, D.; Murphy, A.; Yang, Y.; Chen, Y.; Hirschey, M. D.; Bronson, R. T.; Haigis, M.; Guarente, L. P.; Farese, Jr., R. V.; Weissman, S.; Verdin, E.; Schwer, B. (2007) *Mol Cell Biol* **27**(24), 8807-8814

15. Pesavento, J. J., Kim, Y. B., Taylor, G. K., and Kelleher, N. L. (2004) *J Am Chem Soc* **126**(11), 3386-3387

16. Boyne, M. T., 2nd, Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2006) *J Proteome Res* **5**(2), 248-253

17. Medzihradszky, K. F., Zhang, X., Chalkley, R. J., Guan, S., McFarland, M. A., Chalmers, M. J., Marshall, A. G., Diaz, R. L., Allis, C. D., and Burlingame, A. L. (2004) *Mol Cell Proteomics* **3**(9), 872-886

18. Garcia, B. A., Hake, S. B., Diaz, R. L., Kauer, M., Morris, S. A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B. D., Allis, C. D., and Hunt, D. F. (2006) *J Biol Chem*

19. Beck, H. C., Nielsen, E. C., Matthiesen, R., Jensen, L. H., Sehested, M., Finn, P., Grauslund, M., Hansen, A. M., and Jensen, O. N. (2006) *Mol Cell Proteomics* **5**(7), 1314-1325

20. Gray, M. W., Burger, G., and Lang, B. F. (1999) *Science* **283**(5407), 1476-1481

21. Starai, V. J., Celic, I., Cole, R. N., Boeke, J. D., and Escalante-Semerena, J. C. (2002) *Science* **298**(5602), 2390-2392

22. Barak, R., and Eisenbach, M. (2001) *Mol Microbiol* **40**(3), 731-743

23. Bell, S. D.; Botting, C. H.; Wardieworth, B. N.; Jackson, S. P.; White, M. F. (2002) *Science* **296**(5565) 148-151

24. Zhao, K.; Chai, X.; Marmorstein, R. (2003) *J Biol Chem* **278**(28), 26671-26077

25 Yang, X. J. (2004) *Nucleic Acids Res* **32**(3), 959-976

26. Blander, G., and Guarente, L. (2004) *Annu Rev Biochem* **73**, 417-435

27. Starai, V. J., and Escalante-Semerena, J. C. (2004) *J Mol Biol* **340**(5), 1005-1012

28. Starai, V. J., Takahashi, H., Boeke, J. D., and Escalante-Semerena, J. C. (2003) *Genetics* **163**(2), 545-555

29. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) *Mol Syst Biol* **2**, 2006 0008

30. Qiang, L., Xiao, H., Campos, E. I., Ho, V. C., Li, G. (2005) *J Immunoassay Immunochem* **26**(1):13-23

31. Zhang, H., Zha, X., Tan, Y., Hornbeck, P. V., Mastrangelo, A. J., Alessi, D. R., Polakiewicz, R. D., Comb, M. J. (2002) *J Biol Chem* **277**(42):39379-39387

32. Kim, S. C., Chen, Y., Mirza, S., Xu, Y., Lee, J., Liu, P., and Zhao, Y. (2006) *J Proteome Res* **5**(12), 3446-3452

33. Chen, Y., Kwon, S. W., Kim, S. C., and Zhao, Y. (2005) *J Proteome Res* **4**(3), 998-1005

34. Weiss, B. (2006) *J Bacteriol* **188** (3), 829-833

35. Santos, J. M., Freire, P., Vicente, M. and Arraiano, C. M. (1999) *Mol Microbiol* **32**(4), 789-798

36. Chandonia, J. M., Hon, G., Walker, N. S., Lo Conte, L., Koehl, P., Levitt, M., and Brenner, S. E. (2004) *Nucleic Acids Res* **32**(Database issue), D189-192

37. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J Mol Biol* **247**(4), 536-540

38. Suckau, D., Kohl, J., Karwath, G., Schneider, K., Casaretto, M., Bitter-Suermann, D., and Przybylski, M. (1990) *Proc Natl Acad Sci U S A* **87**(24), 9848-9852

39. Zhao, Y., and Chait, B. T. (1994) *Anal Chem* **66**(21), 3723-3726

40. Papac, D. I., Hoyes, J., and Tomer, K. B. (1994) *Protein Sci* **3**(9), 1485-1492

41. Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A. L., Appella, E., and Engelhard, V. H. (1992) *Science* **255**(5049), 1261-1263

42. Yip, T. T., Van de Water, J., Gershwin, M. E., Coppel, R. L., and Hutchens, T. W. (1996) *J Biol Chem* **271**(51), 32825-32833

43. Pandey, A., Podtelejnikov, A. V., Blagoev, B., Bustelo, X. R., Mann, M., and Lodish, H. F. (2000) *Proc Natl Acad Sci U S A* **97**(1), 179-184

44. Pandey, A., Fernandez, M. M., Steen, H., Blagoev, B., Nielsen, M. M., Roche, S., Mann, M., and Lodish, H. F. (2000) *J Biol Chem* **275**(49), 38633-38639

44. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) *Nat Biotechnol* **23**(1), 94-101

46. Pascal, M. C., Chippaux, M., Abou-Jaoude, A., Blaschkowski, H. P., and Knappe, J. (1981) *J Gen Microbiol* **124**(1), 35-42

47. Borthwick, A. C., Holms, W. H., and Nimmo, H. G. (1984) *FEBS Lett* **174**(1), 112-115

48. Thorsness, P. E., and Koshland, D. E., Jr. (1987) *J Biol Chem* **262**(22), 10422-10425

49. Heine, A., Luz, J. G., Wong, C. H., and Wilson, I. A. (2004) *J Mol Biol* **343**(4), 1019-1034

50. Neidhardt, F. (ed) (1996) *Escherichia coli and Salmonella Cellular and Molecular Biology*, Washington D.C.

51. Howitz, K. T., Bitterman, K. J., Cohen, H. Y., Lamming, D. W., Lavu, S., Wood, J. G., Zipkin, R. E., Chung, P., Kisielewski, A., Zhang, L. L., Scherer, B., and Sinclair, D. A. (2003) *Nature* **425**(6954), 191-196

52. Imlay, J. A., and Linn, S. (1987) *J Bacteriol* **169**(7), 2967-2976

53. Beyer, W., Imlay, J., and Fridovich, I. (1991) *Prog Nucleic Acid Res Mol Biol* **40**, 221-253

54. Karlberg, O., Canback, B., Kurland, C. G., and Andersson, S. G. (2000) *Yeast* **17**(3), 170-187

55. Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, C. D., and Marmorstein, R. (1999) *Nature* **401**(6748), 93-98

56. Scarsdale, J. N., Radaev, S., Kazanina, G., Schirch, V., Wright, H. T. (2000) *J Mol Biol* **296**(1), 155-168

57. Singh, S. K., Matsuno, K., LaPorte, D. C., Banaszak, L. J. (2001) *J Biol Chem* **276**(28), 26154-26163

58. Wolan, D. W., Cheong, C. G., Greasley, S. E., Wilson, I. A. (2004) *Biochemistry 43*(5), 1171-1183.

59. Driscoll, R., Hudson, A., and Jackson, S. P. (2007) *Science* **315**(5812), 649-652

60. Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J., Gu, W., and Zhao, Y. (2007) *Mol Cell Proteomics* **6**(5), 812-819

Footnotes

This work was supported by The Robert A. Welch Foundation (I-1550 to Y.Z.) and NIH (CA107943 to Y.Z. and TCNP U45 RR020839 to H.Z.). We thank Marie-Alda Gilles-Gonzalez and Zhihong Zhang for helpful suggestions. We are grateful to Xiang-Jiao Yang at McGill University for critical reading the manuscript and helpful comments.

The raw MS/MS spectra of all the acetylated peptides and the corresponding acetylated proteins identified in this screen are available as supplemental information.

Figure Legends

Fig. 1. Synopsis of lysine acetylation proteomics. (A) Schematic representation of the sequential steps used for global profiling of lysine acetylation in *E. coli*. (B) Western blotting analysis of protein lysate from *E. coli* with anti-acetyllysine antibody from ImmunoChem Inc. Lane 1: 30 μg of whole cell lysate from *E. coli* probed with anti-acetyllysine antibody; Lane 2: competition with acetylated BSA (300 μg/ml). (C) An example of MS and MS/MS analysis of a lysine-acetylated peptide for peptide identification and mapping of acetyllysine site in *E. coli*. Left: Full MS spectrum at a retention time of 63.69 min. Right: MS/MS spectrum of m/z 1261.2, which identifies acetylated peptide YYQGTPSPVK*HPELTDMVIFR in isocitrate dehydrogenase, a metabolic protein in *E. coli*.

Fig. 2. (A) Pie chart of functionally annotated protein groups that are lysine acetylated. (B) Density map of lysine-acetylated peptides. The frequency of occurrence of amino acid residues surrounding sites of lysine acetylation was calculated, relative to the frequency of the residue within the entire *E. coli* genome, and schematically represented by a density map using a method previously described (11). Prevalence of specific amino acids at positions surrounding lysine acetylation sites is shown.

Fig. 3. Lysine acetylated proteins involved in glucose degradation and TCA cycle. Proteins identified as lysine-acetylated in the *E. coli* screen are underlined. Those identified in a screen of mammalian cells (11) are marked with an asterisk.

Fig. 4. Western blotting analysis of protein lysate with an anti-acetyllysine antibody from ImmunoChem Inc. The 40-μg protein lysate was resolved in 4-20% SDS-PAGE. The Western blotting analysis was carried out as previously described (11). Each of the gel bands that have significant differences between the two *E. coli* strains or among the four oxygen conditions is labeled with an arrow on the left. The molecular weight markers are labeled on the right. (A) Lysine acetylation profiles of two E. *coli* strains (Lanes 1 and 3: MG1655; Lanes 2 and 4: JW1106). Lanes 1-2: probed with anti-acetyllysine antibody and non-acetylated BSA (300 μg/2ml); Lanes 3-4: probed with anti-acetyllysine antibody and acetylated BSA (300 μg/2ml). (B) Impact of nutritional conditions (*E. coli* MG1655). Lane 5: control, and Lane 6: starvation. (C) Influences of oxygen conditions (*E. coli* MG1655). Lane 7: control (100% air), Lane 8: 25% air, Lane 9: 5% air, and Lane 10: 100% nitrogen.

Table 1. A list of metabolic proteins identified as acetylated in *E. coli*. Protein name, gi number, functional classification and number of acetylation sites identified are indicated.

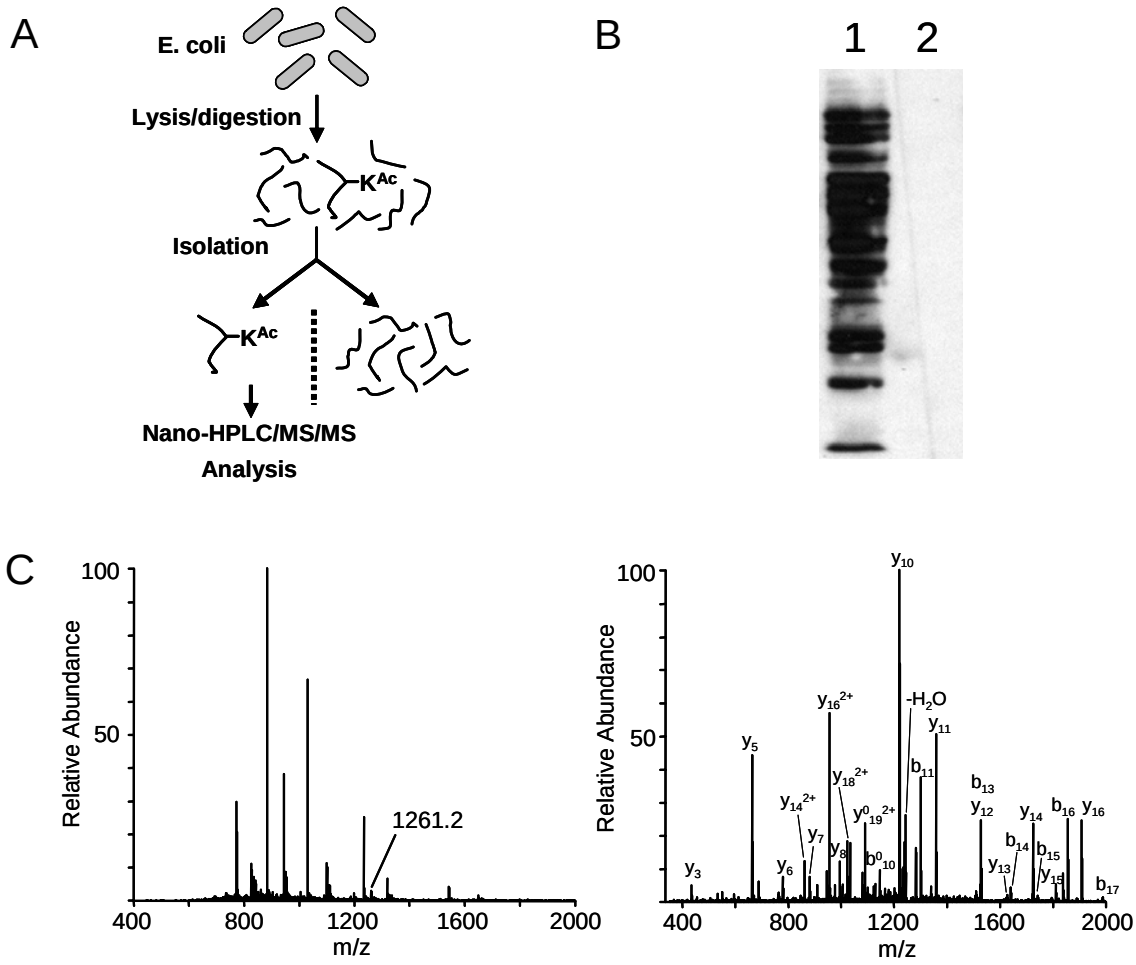

Figure 1

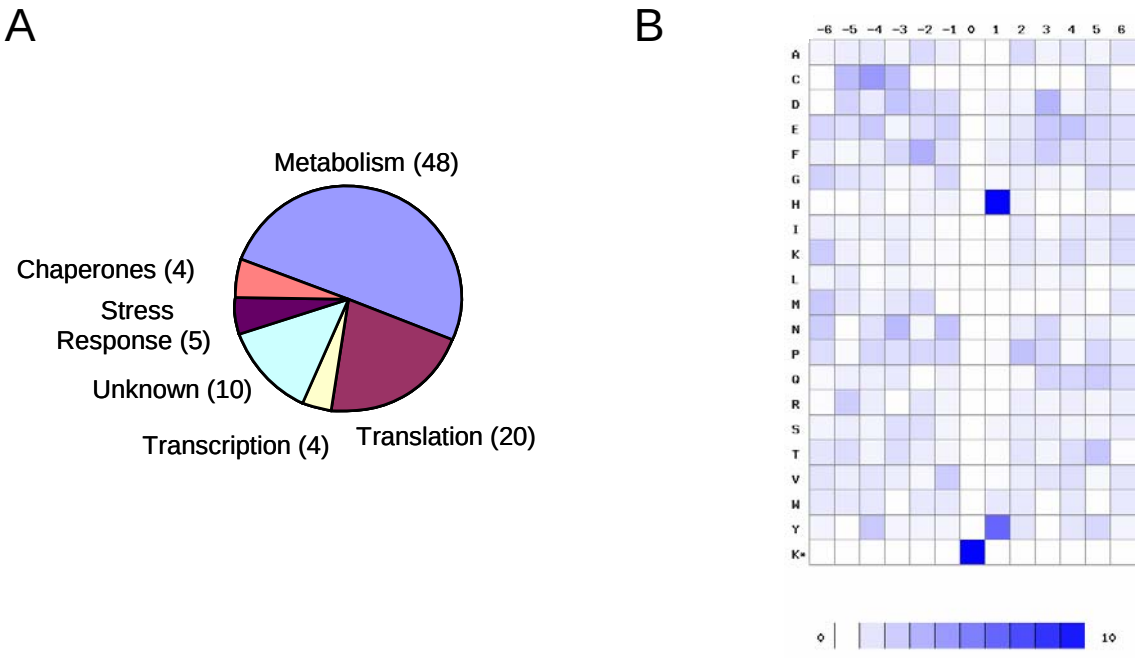

Figure 2

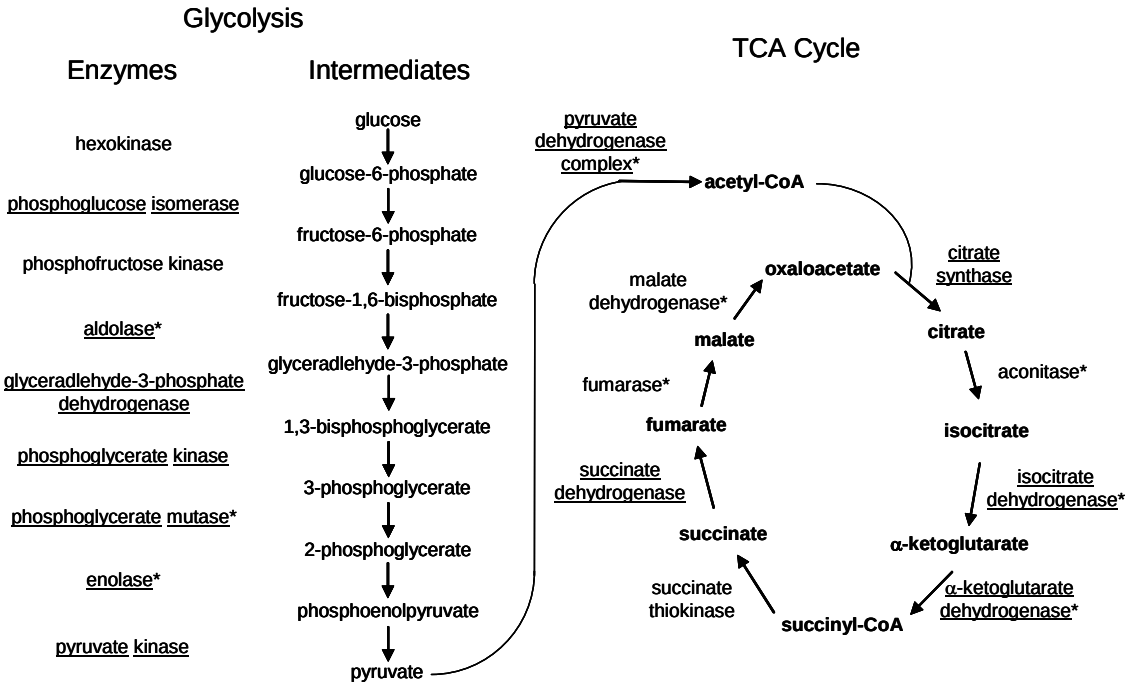

Figure 3

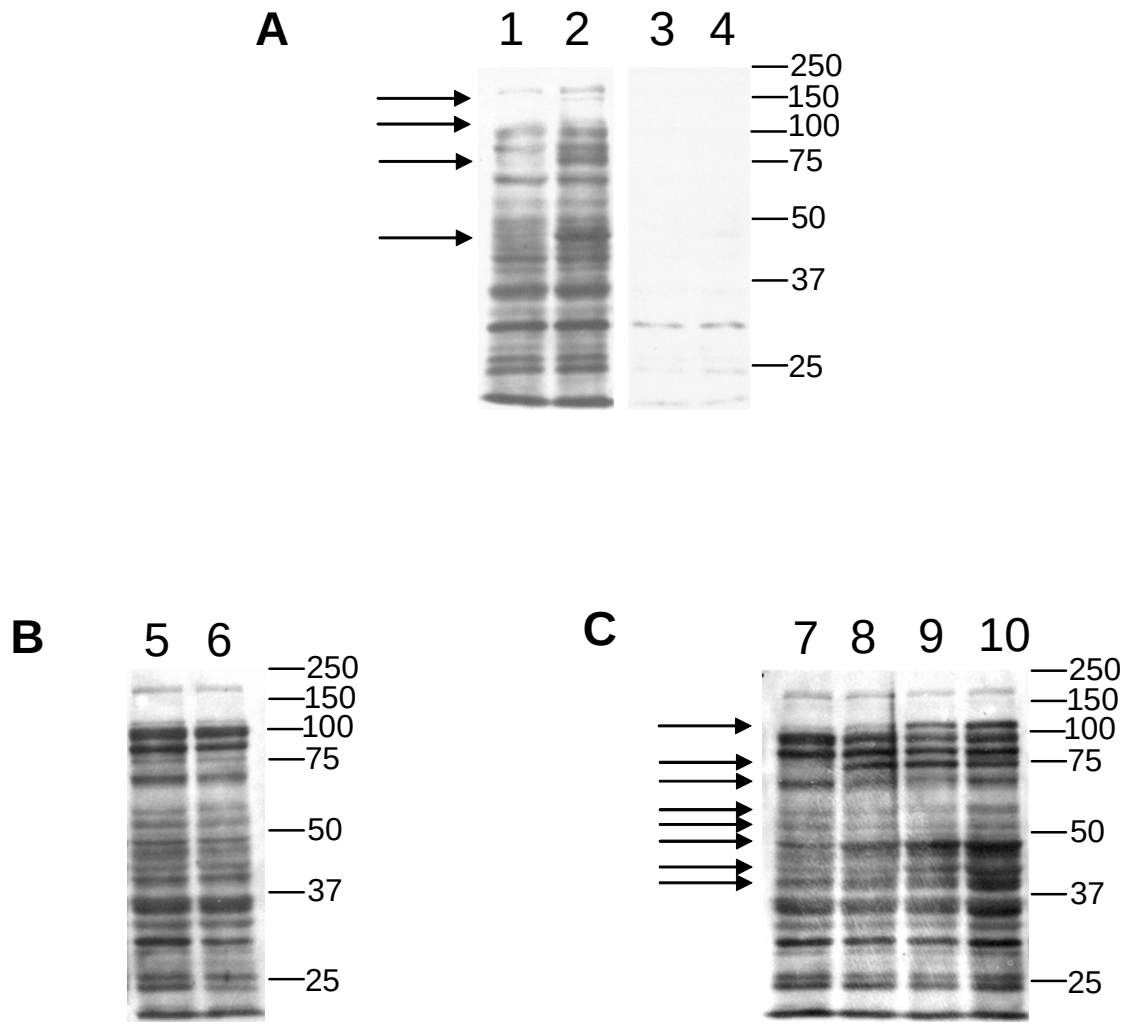

Figure 4

Protein Name	gi	Functional Group	# Sites
citrate synthase	gi|16128695	TCA cycle	1
isocitrate dehydrogenase	gi|2618886	TCA cycle	1
pyruvate dehydrogenase complex, dehydrogenase component	gi|83584560	TCA cycle	1
dihydrolipoamide dehydrogenase	gi|26106453	TCA cycle	1
dihydrolipoamide acetyltransferase	gi|26246694	TCA cycle	1
dihydrolipoamide acyltransferase (E2)	gi|75258892	TCA cycle	1
succinate dehydrogenase catalytic subunit	gi|26246691	TCA cycle	1
phosphoglucose isomerase	gi|9664438	glycolysis	2
pgi	gi|68304110	glycolysis	1
fructose-bisphosphate aldolase class II	gi|110643069	glycolysis	1
dehydrin (fructose-bisphosphate aldolase class I)	gi|1658028	glycolysis	2
enolase	gi|563868	glycolysis	1
phosphoglycerate kinase	gi|26249339	glycolysis	1
phosphoglyceromutase	gi|13360242	glycolysis	3
glyceraldehyde-3-phosphate dehydrogenase	gi|37699654	glycolysis	4
pyruvate kinase	gi|26248120	glycolysis	1
pyruvate kinase I	gi|147276	glycolysis	2
pyruvate formate lyase subunit	gi|16130504	anerobic glycolysis	3
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	6
anaerobic class I fumarate hydratase	gi|146048	anaerobic glycolysis	1
phosphoenolpyruvate carboxykinase	gi|147113	gluconeogenesis	2
transketolase	gi|75227108	carbohydrate metabolism	1
NADP-dependent malic enzyme	gi|26109236	carbohydrate metabolism	1
transketolase 2, thiamin-binding	gi|16130390	carbohydrate metabolism	1
phosphopentomutase	gi|1790843	carbohydrate metabolism	1
NAD-dependent aldehyde dehydrogenase	gi|75515146	carbohydrate metabolism	1
mannitol-1-phosphate dehydrogenase	gi|46095211	carbohydrate metabolism	1
6-phosphogluconolactonase	gi|16128735	carbohydrate metabolism	1
phosphorylase,maltodextrin	gi|224195	carbohydrate metabolism	1
mannose-6-phosphate isomerase	gi|1742663	carbohydrate metabolism	1
fused mannose-specific PTS enzymes: IIA component/IIB component	gi|1788120	carbohydrate metabolism	2
nucleoside-diphosphate-sugar epimerase	gi|75229016	carbohydrate metabolism	1
ADP-heptose synthase	gi|26249631	lipopolysaccharide biosynthesis	1
glutamate decarboxylase isozyme	gi|15804061	amino acid metabolism	3
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	6
tryptophanase	gi|41936	amino acid metabolism	5
3-deoxy-D-arabino-heptulosonate-7-phosphate synthase	gi|1651339	amino acid metabolism	1
glutaminase	gi|26246500	amino acid metabolism	1
aminoacyl-histidine dipeptidase	gi|26246281	protein metabolism	1
S-adenosylmethionine synthetase II	gi|146851	coenzyme metabolism	1
purine nucleoside phosphorylase	gi|15804956	nucleotide salvage pathway	1
adenylate kinase	gi|1773156	nucleotide metabolism	1
IMP dehydrogenase	gi|146275	nucleotide metabolism	2
IMP cyclohydrolase	gi|26250778	nucleotide metabolism	1
ribonucleotide reductase, alpha subunit	gi|75259700	nucleotide metabolism	1
deoxyribosephosphate aldolase	gi|537221	nucleotide metabolism	1
glycinamide ribonucleotide transformylase	gi|16129802	purine biosynthesis	1
adenylosuccinate lyase	gi|145203	purine biosynthesis	2

Table 1