

# **Systematic profiling of the bacterial phosphoproteome reveals bacterium-specific features of phosphorylation**

Miao-Hsia Lin,<sup>1</sup> Naoyuki Sugiyama,<sup>1</sup> Yasushi Ishihama<sup>1\*</sup>

<sup>1</sup>Department of Molecular and Cellular Bioanalysis, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan.

\*Corresponding author. E-mail: yishihama@pharm.kyoto-u.ac.jp

## **Abstract**

Protein phosphorylation is a crucial posttranslational modification for regulating cellular processes in bacteria; however, it has not been extensively studied because of technical difficulties in the enrichment of phosphopeptides. We devised an enrichment protocol that enabled the identification of >1,000 phosphopeptides from a single bacterial sample. We discovered three high-confidence serine and threonine phosphorylation motifs, as well as 29 other motifs at various levels of confidence, from three distinct bacterial phosphoproteomes. We found that the proline-directed and basophilic phosphorylation motifs that are commonly enriched in eukaryotes were not observed in bacteria. Unlike eukaryotes, bacteria had a low occurrence of both phosphorylation and acetylation within N-terminal phosphopeptides. Because infection of host cells by bacterial pathogens is often accompanied by kinase-mediated phosphorylation events, the differences in phosphorylation preferences between bacteria and eukaryotes revealed by this study could be useful in identifying bacterial-specific targets for future therapies.

## Introduction

The phosphorylation of serine, threonine, and tyrosine in proteins is as important for bacteria as it is for other living organisms (1). Accumulating evidence suggests that protein phosphorylation serves as a fundamental regulatory mechanism in nearly all physiological processes in bacteria, especially in key steps of the infection process, such as adhesion to the host, perturbation of host signaling cascades, and impairment of host defense mechanisms (2, 3). Because of the close link between the dysregulation of phosphorylation and various human pathologies, phosphorylation mechanisms are expected to be promising targets for new drugs able to overcome antibiotic resistance (4, 5). Phosphoproteomic analysis has emerged as a promising tool for studying variations in protein phosphorylation in eukaryotic cells (6); however, the much lower degree of protein phosphorylation in bacteria (about 80 times less than that in eukaryotes) has severely hampered progress (7).

In 2007, Macek *et al.* were the first to evaluate the site-specific bacterial phosphoproteome in *Bacillus subtilis* with strong cation exchange chromatography and titanium dioxide (TiO<sub>2</sub>) enrichment coupled with high-accuracy mass spectrometry to identify phosphopeptides in the low-femtomole range (8). This approach has been applied to a number of bacteria, including *Escherichia coli* (9, 10), *Lactococcus lactis* (11), *Halobacterium salinarum* (12), and

pathogenic *Klebsiella pneumoniae* (13). The method of Macek *et al.* has become a standard procedure in bacterial phosphoproteomics, and on average, around 100 phosphorylation sites can be detected in each organism with 10 mg of starting protein under a given growth condition. Although the identified phosphorylated residues are better conserved than random residues, the phosphorylation sites in published phosphoproteome datasets have extremely low overlap, even within the same bacterial species (7, 9, 10). Therefore, the published phosphoproteomes are certainly incomplete, and may be missing many low-occupancy sites because of the technical limitations of current approaches.

Here, we present a protocol for the highly sensitive enrichment of phosphopeptides, together with its application for studying the phosphoproteomes of three distinct bacterial species, *E. coli*, *B. subtilis*, and *K. pneumoniae*, with enhanced phosphopeptide identification and coverage (14-16). The resulting large-scale bacterial phosphoproteome resource has enabled us to uncover various bacterium-specific phosphorylation features, including high-confidence phosphorylation motifs and the notably low coincidence of N-terminal acetylation and phosphorylation. In addition, neither proline-directed nor eukaryote-type basophilic phosphorylation motifs were enriched in bacteria. Given these bacterium-specific phosphorylation signatures, bacteria might have alternative kinase-recognition mechanisms that could be used as targets for the development of new therapeutic drugs.

## Results

### **Design and development of a highly sensitive approach to analyze the bacterial phosphoproteome**

The extremely low degree of protein phosphorylation and the presence of many anionic biomolecules, such as phospholipids, peptidoglycans, and lipid A, present a substantial technical challenge to the analysis of bacterial phosphoproteomes (17-22). To address this, we first developed a large-scale aliphatic hydroxy acid–modified metal oxide chromatography (HAMMOC) protocol coupled with liquid chromatography with tandem mass spectrometry (LC-MS/MS) to accommodate an up to five-fold increase in the amount of sample by increasing the amount of TiO<sub>2</sub> as well as the tip size. We also examined three different methods of pretreating the samples, including ultrafiltration, precipitation with acetone, and precipitation with methanol and chloroform. All of these methods increased the number of phosphopeptides that we identified compared with that in the absence of pretreatment (about 70), and precipitation with methanol and chloroform provided more than 300 phosphopeptides (Fig. 1A). Some biomolecules derived from *E. coli* interfered with and markedly suppressed the HAMMOC-based enrichment, because the lysate from untreated *E. coli* showed the highest intensity in total ion current (TIC) chromatograms, but the lowest number of identified phosphopeptides (fig. S1).

Because precipitation by methanol and chloroform was the most effective of the three procedures for removing interfering components, it was used in subsequent experiments. To refine the enrichment conditions for bacteria, a series of sample amounts ranging from 500 to 1,000  $\mu\text{g}$  was evaluated. The number of observed phosphopeptides increased with increasing sample size, which was expected given that the maximum capacity of large-scale HAMMOC is approximately 800 to 900  $\mu\text{g}$  (Fig. 1B). Accordingly, we suggest that our large-scale HAMMOC procedure is applicable to bacterial phosphoproteome analysis without sample fractionation (8-10), and the small sample size requirement [20 times less than that required for conventional procedures (>10 mg)] is particularly advantageous for clinical purposes.

### **Characterization of the phosphoproteome map of *E. coli* as a model gram-negative organism**

Our developed strategy yielded the most comprehensive *E. coli* phosphoproteome to date; a total of 1,212 phosphopeptides in 392 phosphoproteins with 1,088 phosphosites (Fig. 2A and table S1). After assessing phosphosite localization by site localization probability based on the posttranslational modification (PTM) score (23-25) and the Mascot delta score (23-26), and by the presence of site-determining ions (27), 766 phosphosites were retained as high-confidence sites (table S1). In comparison with previous studies, in which about 150 phosphopeptides were determined from five growth conditions, we obtained about eight-fold

more identified phosphopeptides, with a similar distribution of phosphosites; that is, 69.5% phosphorylated serine (pSer), 21.8% phosphorylated threonine (pThr), and 7.7% phosphorylated tyrosine (pTyr) (table. S1) (9, 10). Furthermore, our results showed large overlaps with published *E. coli* phosphoproteomes at the levels of both phosphopeptide and phosphoprotein (Fig. 2, A and B). For example, all of the reported phosphorylated ribosomal proteins in previous phosphoproteomes were included in our dataset (Fig. 2C) (9, 10). The high overlap between the previously reported *E. coli* phosphoproteomes and ours supports the validity of our developed approach.

In the Ecogene database, about 7% of the genome of *E. coli* K12 strain is annotated as essential genes, and our results suggest that almost 30% of the proteins encoded by essential genes were phosphorylated (Fig. 3A), supporting a major role for protein phosphorylation in bacteria. Similar to other bacteria, *E. coli* contained phosphoproteins that are widely localized from the cell wall to the cytosol, suggesting that our method has no bias with respect to protein localization (Fig. 3B). Based on functional analysis by the bioinformatics resource DAVID, protein phosphorylation participates extensively in various cellular functions in *E. coli*, including central carbon metabolism and housekeeping processes, such as DNA metabolism, transcription, and translation (Fig. 3B). For example, the phosphoproteins that we identified in *E. coli* were overrepresented in the glycolysis pathway and among ribosomal

proteins (Fig. 3, C and D). Almost all of the proteins that participate in the glycolysis and gluconeogenesis pathways were phosphorylated in *E. coli*.

We detected some phosphorylated histidine (pHis) and phosphorylated aspartate (pAsp) residues, although they were labile under acidic conditions (fig. S2). Among them, pHis<sup>15</sup> of phosphocarrier protein HPr (PtsH) was previously reported (28), and phosphorylation at this site is essential for the enzymatic activity of PtsH. Multidrug resistance protein MdtB, a component of the heterotrimeric efflux transporter complex, was phosphorylated on Asp<sup>649</sup>, whereas galactitol-specific phosphotransferase enzyme IIB component protein (GatB), which forms part of the phosphotransferase system, was phosphorylated on His<sup>30</sup>. Protein phosphorylation at these sites might modulate the biological functions of the modified proteins, especially for MdtB, which forms a tripartite complex with MdtA and MdtC and confers resistance to the novobiocin and deoxycholate (29). Indeed, many pSer- or pThr-containing phosphoproteins were also observed among phosphotransferases or two-component systems that were previously thought to be phosphorylated only on histidine, cysteine, or aspartate (fig. S2).

Hansen *et al.* introduced immunoprecipitation of pTyr-containing proteins to study phosphotyrosine proteomics in *E. coli* in the stationary phase (30). The sequences

surrounding pTyr in our study varied greatly from those found by Hansen *et al.*, especially in the case of regions preceding the pTyr residue (Fig. 3E). These differences might be because of differences in sample state or the enrichment approaches used. Consequently, the combined results of these two studies provide the most comprehensive bacterial phosphotyrosine proteome that is currently available.

### **Construction of other model bacterial phosphoproteome maps**

Because the phosphorylation content and cellular components among different bacterial species are distinct, we also applied our strategy to another gram-negative bacterium, *K. pneumoniae*, as well as to a gram-positive model bacterium, *B. subtilis*. For *B. subtilis*, 441 phosphopeptides were observed in 175 phosphoproteins in 11 LC-MS/MS runs. Among 339 identified phosphosites, 226 were confidently identified, with a distribution of 74.8% pSer, 17.7% pThr, and 7.1% pTyr (table S2). On the other hand, for *K. pneumoniae*, we identified 663 phosphopeptides containing 559 phosphosites, of which 388 were confidently identified. The phosphosite distribution was 72.9% pSer, 13.7% pThr, and 12.9% pTyr in 286 phosphoproteins (table S3). These phosphorylation events, as expected, were mostly associated with central carbon metabolism (Fig. 3C). The large numbers of phosphopeptides discovered with our strategy should be helpful in developing a global view of the bacterial phosphoproteome and in elucidating the roles of phosphoregulation in bacteria (8, 13). These



results further confirm the feasibility of our strategy for the broad identification of the phosphoproteome by coupling sample pretreatment and a large-scale HAMMOC protocol (8, 9, 13).

To confirm the validity of our earlier results, we further analyzed the reproducibility of our protocol. We found that duplicate sample preparations generated similar numbers (Fig. 1B) and that triplicate technical replicates of a given preparation gave a median RSD of 18.2% for the peak area of 224 phosphopeptides (table S5). The scatter plots from duplicate LCMS runs (fig. S4A) and of the triplicate technical replicates (fig. S4, B to D) showed a similar distribution, suggesting that our protocol is reproducible. We found that the removal of some interfering biomolecules by protein precipitation improved the identification of phosphopeptides; however, the efficiency of this method varied among different species (fig. S5A), and the underlying influencing factors are still not clear. Different cell components such as the high lipopolysaccharide (LPS) content in gram-negative bacteria, the high peptidoglycan content in *B. subtilis*, and the acidic capsule in *K. pneumoniae* might all affect the efficiency of phosphopeptide enrichment. Therefore, the method of sample pretreatment could be further optimized or modified to improve the phosphoproteomic analysis of other bacteria. In addition, we noted that the lower coverage of the phosphoproteomes of *K. pneumoniae* and *B. subtilis* compared with that of *E. coli* might not be a result of the reduced

number of LC-MS/MS runs (26 runs in *E. coli*, 15 runs in *K. pneumoniae*, and 11 runs in *B. subtilis*), because the cumulative number of unique phosphopeptides had reached saturation (fig. S5B), and the total number of unique phosphopeptides in *E. coli* is much greater than that in other two bacteria. Note also that different culture conditions and different nutrients may provide an extended coverage of bacterial phosphoproteomic maps (10, 30-32).

### **Identification of distinct bacterial phosphorylation features**

Because of the large in vivo bacterial phosphoproteome dataset that we obtained, we were able to start addressing kinase-substrate preferences. The sequence surrounding the phosphorylation site usually plays a critical role in kinase-substrate specificity, so this information would be useful for the prediction of phosphorylation sites (33-35) or for designing new drugs to target bacterial kinases (4, 36). So far, no high-confidence motif for pSer, pThr, or pTyr in bacteria has been reported, although five weak pTyr motifs in *E. coli* O157 and one pThr motif in *M. tuberculosis* have been identified at lower confidence levels (30, 37). Here, we found three previously uncharacterized high-confidence motifs. We also found a further 29 weak motifs including pSer, pThr, and pTyr (table S4).

Our results suggest that bacterial phosphorylation motifs are quite distinct from those in eukaryotes; for example, the proline-directed motif and the typical eukaryote-type basophilic

motifs that are usually enriched in eukaryotes were not observed in bacteria with either strong or weak confidence (Fig. 4 and table S4). Instead, bacteria have one linear sequence motif and position-directed motifs in which the phospho-acceptor residues are near the N terminus or C terminus of the protein. Acidic motifs, the common phosphorylation motif in eukaryotes, were also found in bacteria at a lower confidence level, suggesting that some phosphorylation mechanisms or kinases may be common to both bacteria and eukaryotes.

It is of interest to know whether the suggested position-directed motifs found in this study are evolutionarily conserved. Across 11 evolutionarily divergent species, the occurrence of phosphorylation in the second amino acid position in the N terminus has gradually decreased during evolution, accounting for 6.71, 6.12, and 4.64% of the identified phosphoproteins in *B. subtilis*, *E. coli*, and *K. pneumoniae*, respectively, upon normalization to the content of Ser<sup>2</sup> and Thr<sup>2</sup> at the proteome level (Fig. 5), whereas the corresponding value in mammals is ~2%. In addition, up to 95.8% of the N-terminal phosphopeptides in *E. coli* and 100% of those in *B. subtilis* and *K. pneumoniae* were found in proteins whose first methionine was removed (tables S1, S2, and S3), suggesting that the phosphorylation preference is position-dependent. Moreover, in comparison to that in eukaryotes, the prevalence of the co-occurrence of N-terminal phosphorylation and acetylation at the penultimate position was low in bacteria, amounting to 0, 7.7, and 8.3% of N-terminal acetylation in N-terminal phosphoproteins in *B.*

*subtilis*, *K. pneumoniae*, and *E. coli*, respectively; the sum of N-terminal acetylation together with N-terminal phosphorylation accounts for 0.36 and 0.51% of the total phosphoproteins in *K. pneumoniae* and *E. coli*, respectively (Fig. 5) (12, 38-42). The trend appears to be correlated with evolutionary progression, in that there is a large increase in co-occurrence of N-terminal phosphorylation and acetylation from 8.3 to more than 75% at the branch point between eukaryotes and bacteria (Fig. 5). Also, N-terminal phosphorylation in eukaryotes highly favors acidic residues in the antepenultimate position (Asp<sup>3</sup> or Glu<sup>3</sup>), with ~50% or greater prevalence in eukaryotes, but only ~12.5% prevalence in *E. coli* and *B. subtilis* (Fig. 5). Furthermore, the N-terminal phosphoproteins that contain Asp<sup>3</sup> or Glu<sup>3</sup> are all acetylated at the N-terminus in eukaryotes, with the only exception being *C. elegans* (42), which has a slightly lower coincidence rate of 96% (fig. S3). In summary, bacteria appear to have distinct N-terminal phosphorylation preferences, including a high incidence of N-terminal phosphorylation at second amino acid position without the first methionine, a low co-occurrence of N-terminal acetylation with N-terminal phosphorylation, and a reduced prevalence for Glu<sup>3</sup> or Asp<sup>3</sup> at N-terminal phosphorylation sites.

## Discussion

We addressed the challenging task of comprehensively determining the site-specific phosphoproteome in bacteria. In general, extensive fractionation and labor-intensive

manipulation are required in the field of bacterial phosphoproteomics; however, our newly developed protocol is simpler and has the lowest requirement for sample size; these features are advantageous for systematic studies of bacterial phosphoregulation. To determine the localizations of the phosphosites, we used the site-determining ion combination (SIDIC) approach (27), which is based on the presence of a series of site-determining ions in the MS/MS spectra, which showed as high as a 97% overlap of high-confidence phosphosites with other probability-based algorithms such as PhosClac (Class I) (23), MaxQuant (Class I) (24, 25), and Mascot delta score ( $\geq 9$ ) (26) in the *E. coli* dataset (table S1.). Based on the phosphoproteome maps derived from these three bacteria, the identified phosphoproteins were overrepresented in the glycolysis and gluconeogenesis pathways and among ribosomal proteins. In terms of protein synthesis, of the 57 proteins in the 70S ribosome, we found that 28 were phosphorylated. Phosphorylation of these ribosomal proteins might be involved in the control of protein translation through conformational changes, protein-protein interactions, or alterations in catalytic activity. For example, the phosphosite pSer<sup>35</sup> of the S3 phosphoprotein, which is located in the mRNA-binding avenue of the 30S subunit, is proposed to be involved in intermolecular interactions, whereas pSer<sup>119</sup> and pSer<sup>122</sup> are located in the  $\alpha$ -helix of domain II, which may direct mRNA binding (43-45). In the large subunit, L19 is involved in intersubunit bridge formation between the small and large subunits, which is essential for decoding of mRNA code to protein sequence; therefore,

phosphorylation may have important roles in the processes of subunit assembly and decoding (46, 47). The precise mechanisms of ribosomal regulation remain unclear, but our results provide some clues to guide further investigations with genetic or biochemical methods.

The largest bacterial phosphoproteome dataset that could be assembled thus far enabled us to identify differences in phosphorylation features, including the characteristics of linear sequence and position preference, other than residue conservation, between evolutionary domains; although detailed phosphoproteome maps of diverse organisms will be needed to fully explore these issues. Accordingly, we can propose two alternative mechanisms of kinase specificity based on our findings: the first is the surrounding sequence and the second is position-based preference. In addition to exhibiting positional preference, bacteria also showed an opposing trend in the co-occurrence of N-terminal acetylation with Glu<sup>3</sup> or Asp<sup>3</sup> from that in eukaryotes. This low co-occurrence of N-terminal acetylation may at least in part be a result of the lower incidence of N-terminal acetylation in bacteria (48), whereas the preference for acidic residues in the antepenultimate position may be a result of differences between bacterial and eukaryotic kinases. Although we improved the sensitivity of the detection of the bacterial phosphoproteome in this study, it is important to note that further improvement in detection sensitivity might uncover other bacterial phosphorylation preferences, such as a higher degree of N-terminal acetylation or the presence of a

proline-directed motif, similar to those observed in eukaryotes. Extensive genetic and biochemical experiments will be required to establish in detail the mechanisms of molecular regulation, as well as to discover species-specific features that might be applied to the design of new therapeutic agents to overcome antibiotic resistance.

## **Materials and Methods**

### **Cell culture and lysate preparation**

*E. coli* strain BW25113, *B. subtilis* strain 168, and *K. pneumoniae* strain NTUH-K2044 were each grown in LB medium (Nacalai Tesque) with vigorous shaking at 37°C. Cells were collected by centrifugation at mid-log phase. Cell pellets were resuspended in ice-cold phase transfer surfactant (PTSF) buffer containing 100 mM Tris-Cl (pH 9.0), 12 mM sodium deoxycholate (Wako), 12 mM sodium lauroyl sarcosinate (Wako), and phosphatase inhibitor cocktail (Sigma) (49, 50). Cells were first heated at 95°C for 5 min to inactivate endogenous enzymes, the cell wall and cell membrane were then disrupted by sonication, and cellular debris and unbroken cells were removed by centrifugation at 30,140g for 30 min at 4°C. To remove interfering cellular components, the cell extract was subjected to methanol and chloroform precipitation, acetone precipitation, or ultrafiltration before undergoing digestion by trypsin.

### **Methanol and chloroform precipitation**

To the cell lysate was added four volumes of methanol (Wako), followed by an equal volume of chloroform (Wako), with thorough mixing. Three volumes of ultrapure water were then added. The sample was mixed well and centrifuged at 12,000g for 5 min. The upper aqueous phase was carefully removed, and the white protein precipitate that had formed between the upper and lower phases was collected. The precipitate was taken up in three volumes of methanol and mixed well. Proteins were precipitated by centrifugation at 12,500g for 5 min, and the supernatant was completely removed as described previously (51).

### **Acetone precipitation**

Cell lysates were thoroughly mixed with 5 volumes of ice-cold acetone (Wako), stored overnight at -20°C, and then centrifuged at 12,500g for 5 min. The supernatant was completely removed and the pellet was collected for trypsin digestion as described previously (52).

### **Ultrafiltration**

Cell lysates were placed into an Amicon Ultra-15 device (10-kD, Merck Millipore) and centrifuged at 5000g at 4°C until the volume of the concentrate reached about 2 ml. Then, 10



ml of lysis buffer was added and the mixture was centrifuged. This step was repeated five times before trypsin digestion was performed (53).

### **Protein digestion**

Protein crude extracts without pretreatment were directly reduced with 10 mM dithiothreitol (DTT, Wako) at 37°C for 30 min and then alkylated with 50 mM iodoacetamide (Wako) at 37°C in the dark for 30 min. Alkylated proteins were digested with Lys-C (1:50, w/w, Wako) for 3 hours followed by overnight digestion with trypsin at 37°C (1:50, w/w, Promega). The surfactants were removed by adding organic solvent as described previously (49). Tryptic peptides were dissolved in 5% acetonitrile (ACN, Wako) and 0.1% trifluoroacetic acid (TFA, Wako) for desalting with a styrene-divinylbenzene copolymer (SDB-XC) solid-phase extraction cartridge (Empore, 3M). In the case of pretreated samples, protein extracts were dissolved and denatured in 8 M urea (Invitrogen), which was followed by protein reduction and alkylation at room temperature. The alkylated protein was digested with Lys-C for 3 hours and then diluted four times with 50 mM ammonium bicarbonate (Wako) for overnight digestion with trypsin at room temperature. The resulting peptide mixture was acidified with TFA and desalted with an SDB-XC solid-phase extraction cartridge.

### **Phosphopeptide enrichment**

Conventional HAMMOC with titanium dioxide (Titansphere, GL Sciences, Cat. 5020-75010) was performed with a 10- $\mu$ l C8-StageTip as described previously (50). For large-scale HAMMOC, a homemade 200- $\mu$ l C8-StageTip was used; due to backpressure, the flow rate was ~10 to 30  $\mu$ l/min. The amount of TiO<sub>2</sub> beads was increased to 1.5 mg, and the sample amount was increased to 500  $\mu$ g for the first trial. For stepwise enrichment, the eluent of the large-scale HAMMOC was acidified with TFA and desalted with an SDB-XC StageTip, which was followed by conventional HAMMOC enrichment. For phosphopeptide elution, the sequential elution approach was adapted with both 0.5 and 5% piperidine (Wako). After desalting with an SDB-XC StageTip, phosphopeptides were resuspended in 0.5% TFA, 4% ACN and then were subjected to nanoscale liquid chromatography (nanoLC)–MS/MS analysis.

### **LC-MS/MS analysis**

The TripleTOF 5600 MS system was coupled with an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) with an HTC-PAL autosampler (CTC Analytics). Peptide mixtures were loaded onto a 0.1 mm id $\times$  150 mm long fused-silica capillary column packed with C18 material (3  $\mu$ m, Dr. Maisch GmbH, Amerbuch, Germany). The injection volume for peptide samples was 5  $\mu$ l, and the flow rate was 500 nl/min. The mobile phases consisted of solution A of 0.5% acetic acid (Wako) and solution B of 0.5% acetic acid and 80%

acetonitrile. A three-step linear gradient of 5 to 10% B for 5 min, 10 to 40% B for 60 min, 40 to 100% B for 5 min, and 100% B for 10 min was used. The MS instrument was operated in the positive ion mode, with an ion-spray voltage of 2.3 kV and an interface heater temperature of 150°C. Data were acquired from one full MS scan ( $m/z$  300 to 1500) for 250 ms, which was followed by high-sensitivity MS/MS scans from the top seven most abundant precursor ions, each with a 100-ms accumulation time.

### **Data analysis**

The peak list of each raw MS spectrum was generated as previously described (27). Peptide identification was performed by Mascot ver 2.4 (Matrix Science) against a composite target-decoy protein sequence database containing 4,316 proteins for *E. coli*, 5,262 for *K. pneumoniae*, and 4,188 for *B. subtilis*. The search criteria used in this study were as follows: trypsin specificity allowing for up to two missed cleavage events; fixed modification of carbamidomethyl (Cys); and variable modifications of oxidation (Met), acetylation (protein N-term), and phosphorylation (Ser and Thr), (Tyr), (Asp), and (His). The precursor mass tolerance was set at 20 ppm, and the fragment ion tolerance was set at 0.1 Daltons. Data from each nanoLC–MS/MS run were searched individually. Peptides were considered identified if the Mascot score yielded a confidence limit >95% based on the significance threshold ( $P < 0.05$ ) and if at least three successive y- or b-ions with an additional two or more y-, b-, or

precursor-origin neutral loss ions were observed, based on the error-tolerant peptide sequence tag concept (54). These criteria gave a false positive rate of <1% for phosphopeptide identification, as evaluated by the target-decoy strategy. For those phosphopeptides identified in more than one run, only the phosphopeptides with the highest score were reported. The reliability of phosphosite localization was assessed by the site-determining ion combination (SIDIC) method (27). To confirm the SIDIC assessment, the site localization probability for all *E. coli* phosphosites was calculated by PhosCalc version 1.2 (23), MaxQuant (24, 25), and Mascot delta score (MD-score) (26). Phosphosites with PTM score-based probability by PhosCalc or MaxQuant were grouped into Class I (phosphosite probability  $P > 0.75$ ), Class II ( $0.5 < P \leq 0.75$ ), and Class III ( $P \leq 0.5$ ) (24). Class I phosphosites were accepted automatically as unambiguous sites (24). For MD-score, those phosphosites with an MD-score of  $\geq 9$  (99% accuracy) were considered as unambiguous sites.

### **Motif-X analysis**

All confirmed phosphosites were used as input for Motif-x software to determine motifs for these phosphorylated substrates with the proteome as background (55). Parameters for the Motif-x analyses were as follows: foreground central residue = S or T or Y (depending on the analysis), width = 13, the number of occurrences was 10, and the statistical significance was 0.00001, unless otherwise stated.

## Supplementary Materials

Fig. S1. Total ion current (TIC) chromatograms of samples with different pretreatments.

Fig. S2. Coexistence of serine, threonine, and tyrosine phosphorylation with aspartate and histidine phosphorylation in the PTS and two-component systems.

Fig. S3. Patterns of position-directed phosphorylation from prokaryotes to mammals.

Fig. S4. Enrichment reproducibility in technical and biological replicates.

Fig. S5. Comparison of enrichment efficiency among three bacterial strains.

Table S1. List of identified phosphopeptides in *E. coli*.

Table S2. List of identified phosphopeptides in *B. subtilis*.

Table S3. List of identified phosphopeptides in *K. pneumoniae*.

Table S4. Phosphorylation motifs found in bacteria.

Table S5. Reproducibility of technical replicates.

## References and Notes

1. J. Deutscher, M. H. Saier, Jr., Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established. *Journal of molecular microbiology and biotechnology* **9**, 125-131 (2005).
2. A. J. Cozzone, Role of protein phosphorylation on serine/threonine and tyrosine in the virulence of bacterial pathogens. *Journal of molecular microbiology and biotechnology* **9**, 198-213 (2005).
3. R. Ge, W. Shan, Bacterial phosphoproteomic analysis reveals the correlation between protein phosphorylation and bacterial pathogenicity. *Genomics, proteomics & bioinformatics* **9**, 119-127 (2011).
4. A. J. Cozzone, An insight into future antibacterial therapy. *Emerging Microbes and Infections* **1**, (2012).
5. V. N. Danilenko, D. I. Osolodkin, S. A. Lakatos, M. N. Preobrazhenskaya, A. A. Shtil, Bacterial eukaryotic type serine-threonine protein kinases: from structural biology to targeted anti-infective drug design. *Current topics in medicinal chemistry* **11**, 1352-1369 (2011).
6. M. Mann, S. E. Ong, M. Gronborg, H. Steen, O. N. Jensen, A. Pandey, Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends in biotechnology* **20**, 261-268 (2002).
7. I. Mijakovic, B. Macek, Impact of phosphoproteomics on studies of bacterial physiology. *FEMS microbiology reviews* **36**, 877-892 (2012).
8. B. Macek, I. Mijakovic, J. V. Olsen, F. Gnad, C. Kumar, P. R. Jensen, M. Mann, The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*.

*Molecular & cellular proteomics : MCP* **6**, 697-707 (2007).

9. B. Macek, F. Gnad, B. Soufi, C. Kumar, J. V. Olsen, I. Mijakovic, M. Mann, Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Molecular & cellular proteomics : MCP* **7**, 299-307 (2008).
10. N. C. Soares, P. Spat, K. Krug, B. Macek, Global dynamics of the *Escherichia coli* proteome and phosphoproteome during growth in minimal medium. *Journal of proteome research* **12**, 2611-2621 (2013).
11. B. Soufi, F. Gnad, P. R. Jensen, D. Petranovic, M. Mann, I. Mijakovic, B. Macek, The Ser/Thr/Tyr phosphoproteome of *Lactococcus lactis* IL1403 reveals multiply phosphorylated proteins. *Proteomics* **8**, 3486-3493 (2008).
12. M. Aivaliotis, B. Macek, F. Gnad, P. Reichelt, M. Mann, D. Oesterhelt, Ser/Thr/Tyr protein phosphorylation in the archaeon *Halobacterium salinarum*--a representative of the third domain of life. *PloS one* **4**, e4777 (2009).
13. M. H. Lin, T. L. Hsu, S. Y. Lin, Y. J. Pan, J. T. Jan, J. T. Wang, K. H. Khoo, S. H. Wu, Phosphoproteomics of *Klebsiella pneumoniae* NTUH-K2044 reveals a tight link between tyrosine phosphorylation and virulence. *Molecular & cellular proteomics : MCP* **8**, 2613-2623 (2009).
14. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2**, 1896-1906 (2007).
15. Y. Kyono, N. Sugiyama, K. Imami, M. Tomita, Y. Ishihama, Successive and selective release of phosphorylated peptides captured by hydroxy acid-modified metal oxide chromatography. *Journal of proteome research* **7**, 4585-4593 (2008).
16. N. Sugiyama, T. Masuda, K. Shinoda, A. Nakamura, M. Tomita, Y. Ishihama, Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Molecular & cellular proteomics : MCP* **6**, 1103-1109 (2007).
17. J. Zhu, F. Wang, K. Cheng, J. Dong, D. Sun, R. Chen, L. Wang, M. Ye, H. Zou, A simple integrated system for rapid analysis of sialic-acid-containing N-glycopeptides from human serum. *Proteomics* **13**, 1306-1313 (2013).
18. X. Zhao, C. Ma, H. Han, J. Jiang, F. Tian, J. Wang, W. Ying, X. Qian, Comparison and optimization of strategies for a more profound profiling of the sialylated N-glycoproteomics in human plasma using metal oxide enrichment. *Analytical and bioanalytical chemistry* **405**, 5519-5529 (2013).
19. H. Li, X. Shi, L. Qiao, X. Lu, G. Xu, Synthesis of a new type of echinus-like Fe<sub>3</sub>O<sub>4</sub>@TiO<sub>2</sub> core-shell-structured microspheres and their applications in selectively enriching phosphopeptides and removing phospholipids. *Journal of chromatography*.

- A **1275**, 9-16 (2013).
20. H. J. Y. a. K. Håkansson, Phosphate-containing Metabolite Enrichment with TiO<sub>2</sub> Micro-tips. *BULLETIN OF THE KOREAN CHEMICAL SOCIETY* **33**, 2475 (2012).
  21. M. Simundic, B. Drasler, V. Sustar, J. Zupanc, R. Stukelj, D. Makovec, D. Erdogmus, H. Hagerstrand, D. Drobne, V. Kralj-Iglic, Effect of engineered TiO<sub>2</sub> and ZnO nanoparticles on erythrocytes, platelet-rich plasma and giant unilamellar phospholipid vesicles. *BMC veterinary research* **9**, 7 (2013).
  22. N. J. Cho, C. W. Frank, Fabrication of a planar zwitterionic lipid bilayer on titanium oxide. *Langmuir : the ACS journal of surfaces and colloids* **26**, 15706-15710 (2010).
  23. D. MacLean, M. A. Burrell, D. J. Studholme, A. M. Jones, PhosCalc: a tool for evaluating the sites of peptide phosphorylation from mass spectrometer data. *BMC research notes* **1**, 30 (2008).
  24. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648 (2006).
  25. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).
  26. M. M. Savitski, S. Lemeer, M. Boesche, M. Lang, T. Mathieson, M. Bantscheff, B. Kuster, Confident phosphorylation site localization using the Mascot Delta Score. *Molecular & cellular proteomics : MCP* **10**, M110 003830 (2011).
  27. H. Nakagami, N. Sugiyama, K. Mochida, A. Daudi, Y. Yoshida, T. Toyoda, M. Tomita, Y. Ishihama, K. Shirasu, Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant physiology* **153**, 1161-1174 (2010).
  28. A. A. van Dijk, L. C. de Lange, W. W. Bachovchin, G. T. Robillard, Effect of phosphorylation on hydrogen-bonding interactions of the active site histidine of the phosphocarrier protein HPr of the phosphoenolpyruvate-dependent phosphotransferase system determined by <sup>15</sup>N NMR spectroscopy. *Biochemistry* **29**, 8164-8171 (1990).
  29. N. Baranova, H. Nikaido, The baeSR two-component regulatory system activates transcription of the yegMNOB (mdtABCD) transporter gene cluster in Escherichia coli and increases its resistance to novobiocin and deoxycholate. *Journal of bacteriology* **184**, 4168-4176 (2002).
  30. A. M. Hansen, R. Chaerkady, J. Sharma, J. J. Diaz-Mejia, N. Tyagi, S. Renuse, H. K. Jacob, S. M. Pinto, N. A. Sahasrabudhe, M. S. Kim, B. Delanghe, N. Srinivasan, A. Emili, J. B. Kaper, A. Pandey, The Escherichia coli phosphotyrosine proteome relates to core pathways and virulence. *PLoS pathogens* **9**, e1003403 (2013).
  31. Y. Qu, S. Wu, R. Zhao, E. Zink, D. J. Orton, R. J. Moore, D. Meng, T. R. Clauss, J. T.

- Aldrich, M. S. Lipton, L. Pasa-Tolic, Automated immobilized metal affinity chromatography system for enrichment of Escherichia coli phosphoproteome. *Electrophoresis* **34**, 1619-1626 (2013).
32. D. Esser, T. K. Pham, J. Reimann, S. V. Albers, B. Siebers, P. C. Wright, Change of carbon source causes dramatic effects in the phospho-proteome of the archaeon *Sulfolobus solfataricus*. *Journal of proteome research* **11**, 4823-4833 (2012).
33. N. Blom, T. Sicheritz-Ponten, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633-1649 (2004).
34. N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology* **294**, 1351-1362 (1999).
35. L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, A. K. Dunker, The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research* **32**, 1037-1049 (2004).
36. C. Cousin, A. Derouiche, L. Shi, Y. Pagot, S. Poncet, I. Mijakovic, Protein-serine/threonine/tyrosine kinases in bacterial signaling and regulation. *FEMS microbiology letters* **346**, 11-19 (2013).
37. S. Pristic, S. Dankwa, D. Schwartz, M. F. Chou, J. W. Locasale, C. M. Kang, G. Bemis, G. M. Church, H. Steen, R. N. Husson, Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 7521-7526 (2010).
38. S. Lemeer, M. W. Pinkse, S. Mohammed, B. van Breukelen, J. den Hertog, M. Slijper, A. J. Heck, Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *Journal of proteome research* **7**, 1555-1564 (2008).
39. F. Gnäd, L. M. de Godoy, J. Cox, N. Neuhauser, S. Ren, J. V. Olsen, M. Mann, High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics* **9**, 4642-4652 (2009).
40. O. Hekmat, S. Munk, L. Fogh, R. Yadav, C. Francavilla, H. Horn, S. O. Wurtz, A. S. Schrohl, B. Damsgaard, M. U. Romer, K. C. Belling, N. F. Jensen, I. Gromova, D. B. Bekker-Jensen, J. M. Moreira, L. J. Jensen, R. Gupta, U. Lademann, N. Brunner, J. V. Olsen, J. Stenvang, TIMP-1 increases expression and phosphorylation of proteins associated with drug resistance in breast cancer cells. *Journal of proteome research* **12**, 4136-4151 (2013).
41. M. Hilger, T. Bonaldi, F. Gnäd, M. Mann, Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. *Molecular & cellular proteomics : MCP* **8**, 1908-1920 (2009).



42. D. F. Zielinska, F. Gnad, M. Jedrusik-Bode, J. R. Wisniewski, M. Mann, Caenorhabditis elegans has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J Proteome Res* **8**, 4039-4049 (2009).
43. O. A. Dontsova, K. V. Rosen, S. L. Bogdanova, E. A. Skripkin, A. M. Kopylov, A. A. Bogdanov, Identification of the Escherichia coli 30S ribosomal subunit protein neighboring mRNA during initiation of translation. *Biochimie* **74**, 363-371 (1992).
44. J. Frank, J. Zhu, P. Penczek, Y. Li, S. Srivastava, A. Verschoor, M. Radermacher, R. Grassucci, R. K. Lata, R. K. Agrawal, A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature* **376**, 441-444 (1995).
45. G. Z. Yusupova, M. M. Yusupov, J. H. Cate, H. F. Noller, The path of messenger RNA through the ribosome. *Cell* **106**, 233-241 (2001).
46. S. Maisnier-Patin, W. Paulander, A. Pennhag, D. I. Andersson, Compensatory evolution reveals functional interactions between ribosomal proteins S12, L14 and L19. *Journal of molecular biology* **366**, 207-215 (2007).
47. B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, J. H. Cate, Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827-834 (2005).
48. J. Soppa, Protein acetylation in archaea, bacteria, and eukaryotes. *Archaea* **2010**, (2010).
49. T. Masuda, M. Tomita, Y. Ishihama, Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *Journal of proteome research* **7**, 731-740 (2008).
50. T. Masuda, N. Sugiyama, M. Tomita, Y. Ishihama, Microscale phosphoproteome analysis of 10,000 cells from human cancer cell lines. *Analytical chemistry* **83**, 7698-7703 (2011).
51. D. Wessel, U. I. Flugge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* **138**, 141-143 (1984).
52. A. M. Crowell, M. J. Wall, A. A. Doucette, Maximizing recovery of water-soluble proteins through acetone precipitation. *Analytica chimica acta* **796**, 48-54 (2013).
53. L. Jiang, L. He, M. Fountoulakis, Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *Journal of chromatography. A* **1023**, 317-320 (2004).
54. M. Mann, M. Wilm, Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical chemistry* **66**, 4390-4399 (1994).
55. D. Schwartz, S. P. Gygi, An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nature biotechnology* **23**, 1391-1398 (2005).

56. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127-128 (2007).

**Acknowledgments:** We thank M. Itaya and K. Nakahigashi (Keio University) for providing *B. subtilis* strain 168 and *E. coli* strain BW25113, respectively, and for related information. We also thank all the members of our lab for stimulating discussions. **Funding:** This research was funded by the Japan Society for the Promotion of Science (24/02078, 26670015, 24116513, and 24241062). M.-H.L. is a JSPS postdoctoral fellow. **Author contributions:** M.-H.L. and Y.I. designed the research; M.-H.L. performed research; M.-H.L., N.S., and Y.I. analyzed data; and M.-H.L. and Y.I. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The MS proteomics data have been deposited with the ProteomeXchange Consortium through the PRIDE partner repository with the dataset identifiers PXD001264 (*E. coli*), PXD001298 (*B. subtilis*), and PXD001263 (*K. pneumoniae*).

**Fig. 1. Comparison of phosphopeptide enrichment efficiencies.** (A) Samples of *E. coli* were subjected to conventional or large-scale HAMMOC without pretreatment or else were pretreated by acetone precipitation, ultrafiltration, or precipitation with methanol and chloroform (MeOH/CHCl<sub>3</sub>) before being subjected to large-scale HAMMOC. Bars represent the number of identified phosphopeptides by duplicate LCMS injections from the same enriched samples. Tryptic peptides from lysates (500 µg) were used for large-scale HAMMOC, whereas 100 µg of sample was used for conventional HAMMOC. (B) Sample capacity test of large-scale HAMMOC. The amount of peptide used for the analysis ranged from 500 to 1,000 µg, as indicated. Paired bars show the numbers of identified phosphopeptides and represent duplicate samples and handling, including protein precipitation, trypsin digestion, phosphopeptide enrichment, and LC-MS/MS measurement.

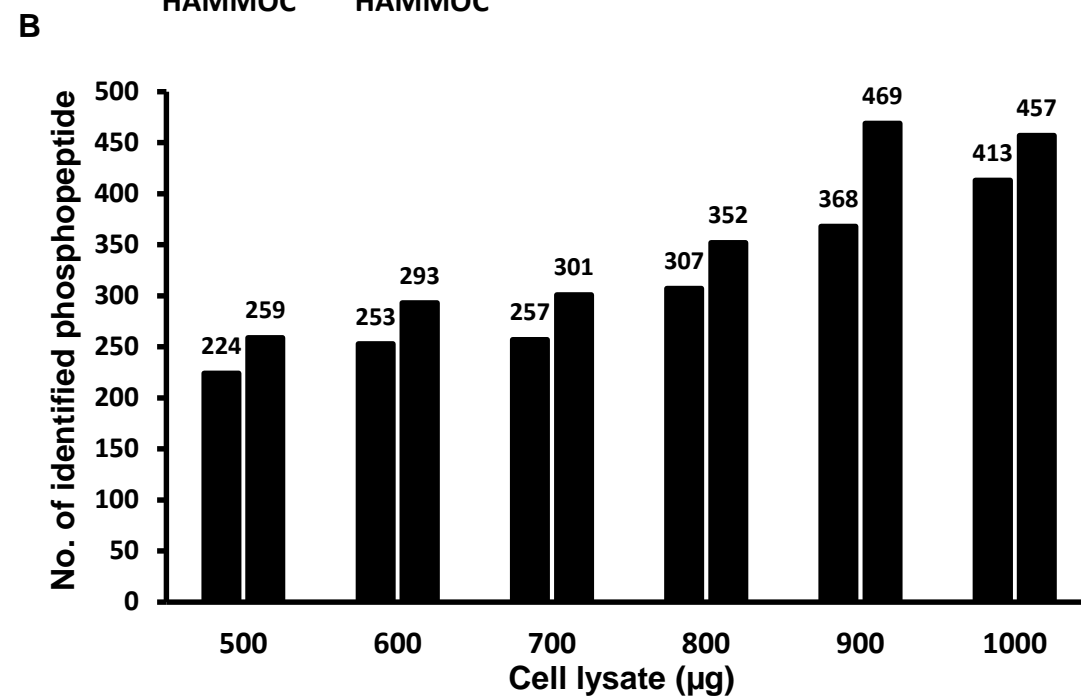
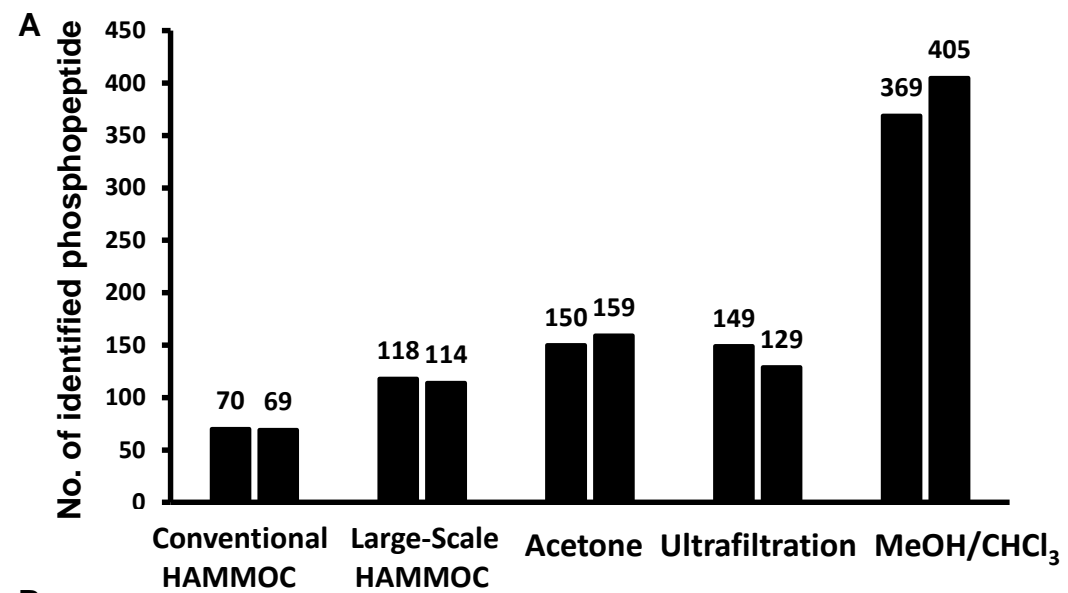
**Fig. 2. Overlap of *E. coli* phosphoproteomes.** (A to C) Analysis of the overlap in the numbers of (A) phosphopeptides, (B) phosphoproteins, and (C) phosphorylated ribosomal proteins identified in *E. coli* in the current study and those in published *E. coli* phosphoproteomes (9, 10).

**Fig. 3. Overview of the *E. coli* phosphoproteome.** (A) The percentage of identified phosphoproteins assigned to essential genes according to the Ecogene Database annotation. (B) Gene ontology (GO) distribution of identified phosphoproteins with respect to biological processes and cellular components with the DAVID Functional Annotation tools. The percentages indicate those proteins that belong to distinct categories. (C and D) Phosphoproteins participate widely in housekeeping pathways based on the KEGG database, including (C) the glycolysis and gluconeogenesis pathways and (D) ribosomal proteins. Blue circles in (C) represent phosphoproteins identified in *E. coli*, *K. pneumoniae*, or *B. subtilis*. White circles or boxes in (C) and (D) indicate unphosphorylated proteins in distinct pathways. The colored boxes in (D) indicate the phosphorylated ribosomal proteins that were observed in this study. (E) Frequency distribution analyzed by Sequence Logo Generator (<http://www.phosphosite.org/>) (56) of the general representation of residues surrounding the phosphorylated tyrosine residue based on data from a previous study (30) and from the current study.

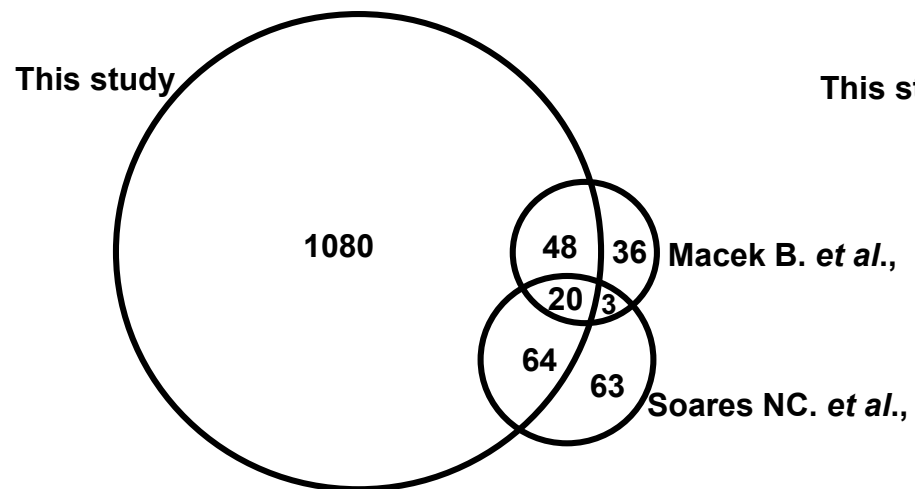
**Fig. 4. Statistically significantly enriched phosphorylation motifs of bacterial phosphoproteomes obtained by Motif-x analysis.** Motif-x software was used to represent the N- or C-terminal position, but not any particular amino acid, with the three proteomes as background. Three motifs are overrepresented in the large-scale phosphoproteomic maps. **(A)** The linear sequence with lysine (K) at the -1 position relative to serine (S, left) and threonine (T, right) is preferred in bacteria. **(B)** Three position-directed motifs in which a phosphosite near the N or C terminus of the protein is statistically significantly enriched. The parameters of the Motif-x algorithm were occurrences = 10 and significance = 0.00001.

**Fig. 5. Co-occurrence of N-terminal phosphorylation with N-terminal acetylation, and analysis of adjacent acidic residues in the antepenultimate position of 11 evolutionarily divergent organisms.** The phosphoproteomes of evolutionarily divergent organisms were collected from the literature and our own datasets (12, 38-42). Bars represent the numbers of pSer<sup>2</sup>- or pThr<sup>2</sup>-containing proteins divided by the total number of phosphoproteins. The white portion of the bar shows the proportion of N-terminal phosphoproteins that underwent N-terminal acetylation, whereas the black portion indicates the proportion of N-terminal phosphoproteins that were not acetylated. Each bar was normalized by the relative value of the Ser<sup>2</sup> and Thr<sup>2</sup> content in the total proteins of each indicated organism to the Ser<sup>2</sup> and Thr<sup>2</sup>

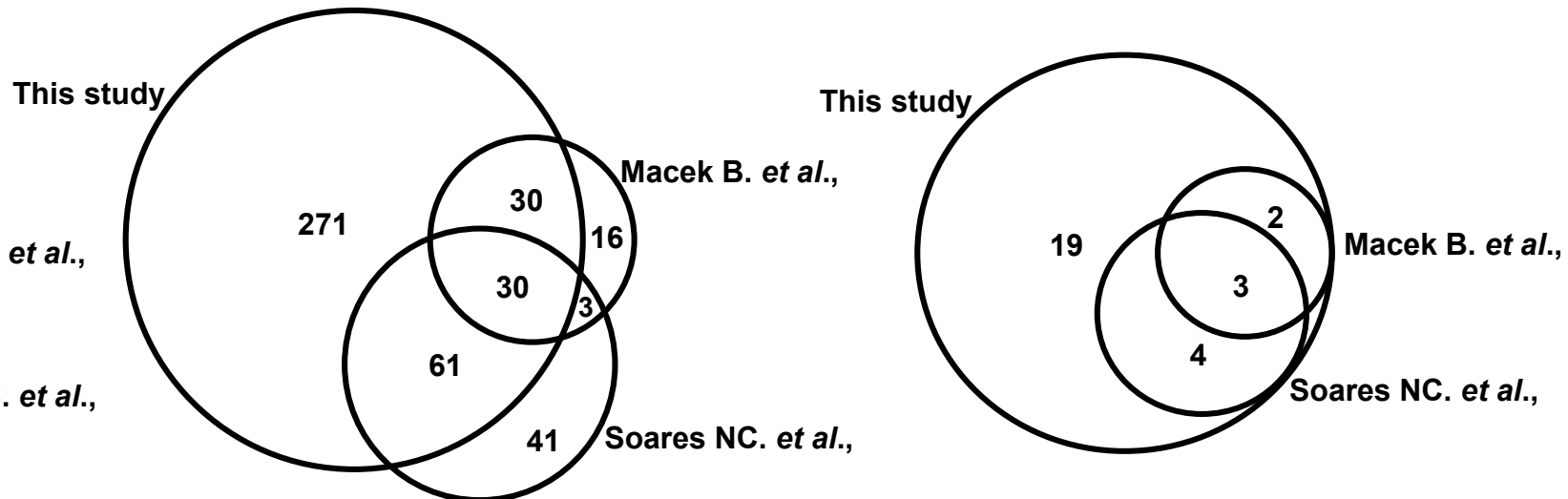
content of *E. coli*. The gray curve represents the number of N-terminal phosphoproteins that contain Asp<sup>3</sup> or Glu<sup>3</sup> as a percentage of the total number of N-terminal phosphoproteins. The numbers in parentheses below the x-axis denote the total number of proteins in each organism. The phylogenetic tree was constructed with an open tool, the Interactive Tree Of Life (iTOL, <http://itol.embl.de/index.shtml>) (56).



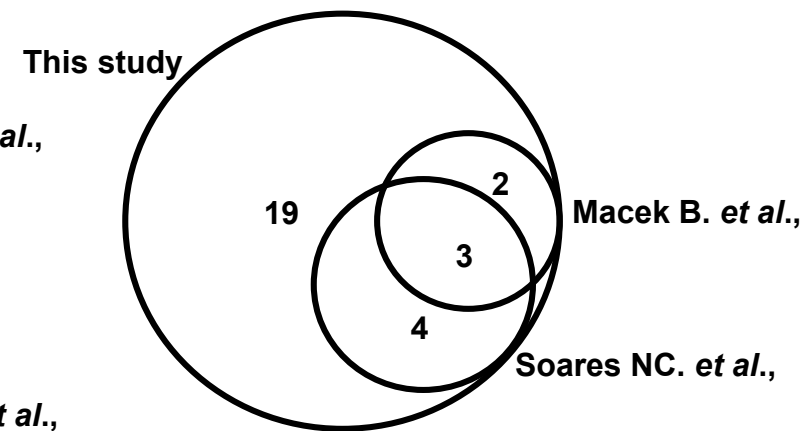
**A Phosphopeptides**

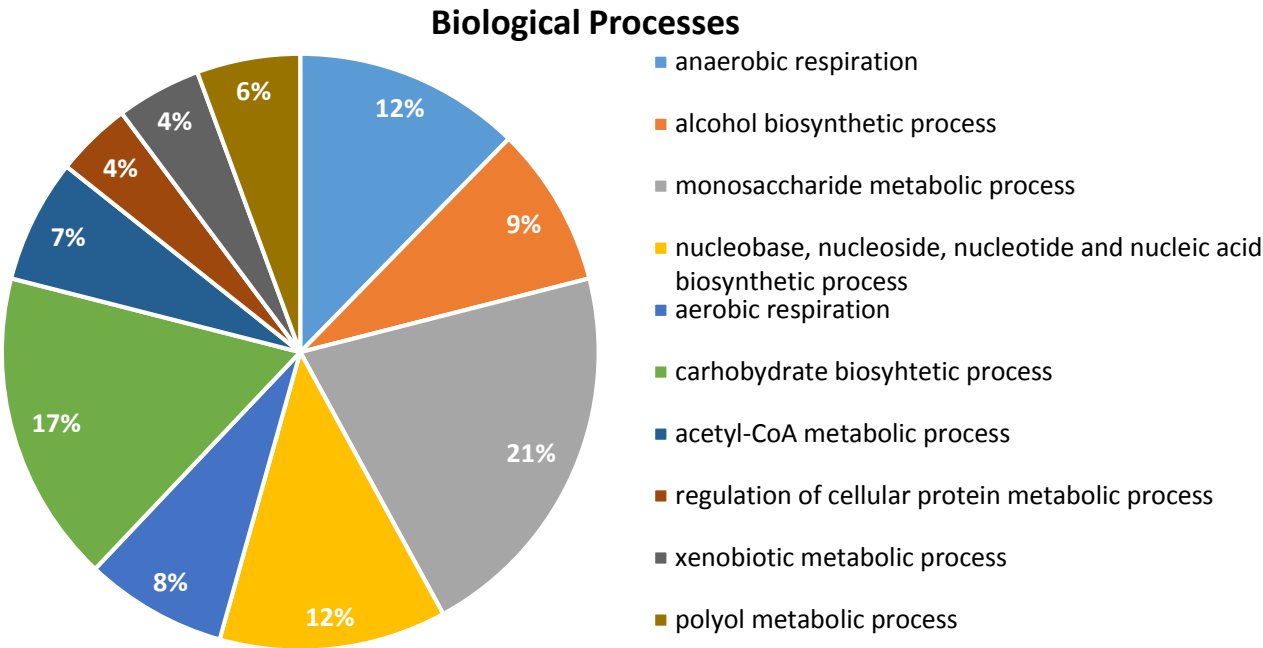
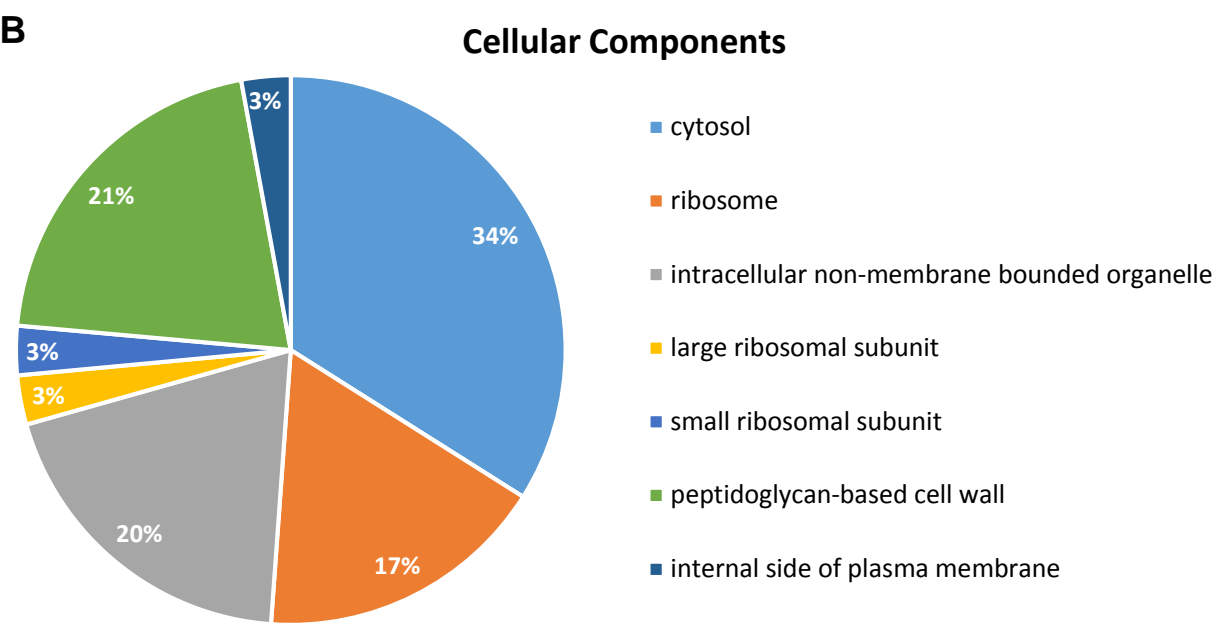
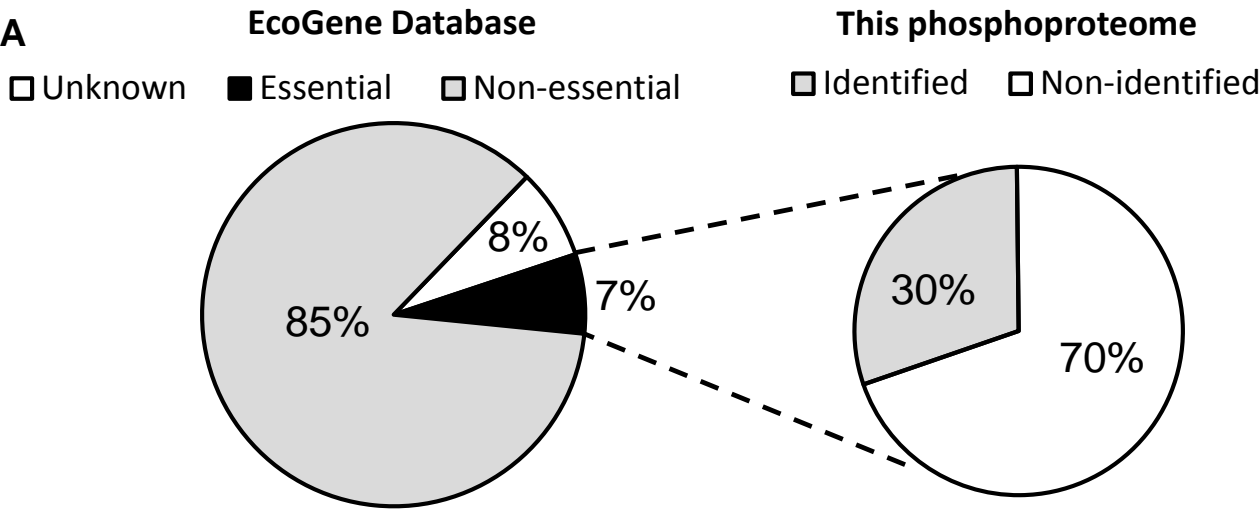


**B Phosphoproteins**

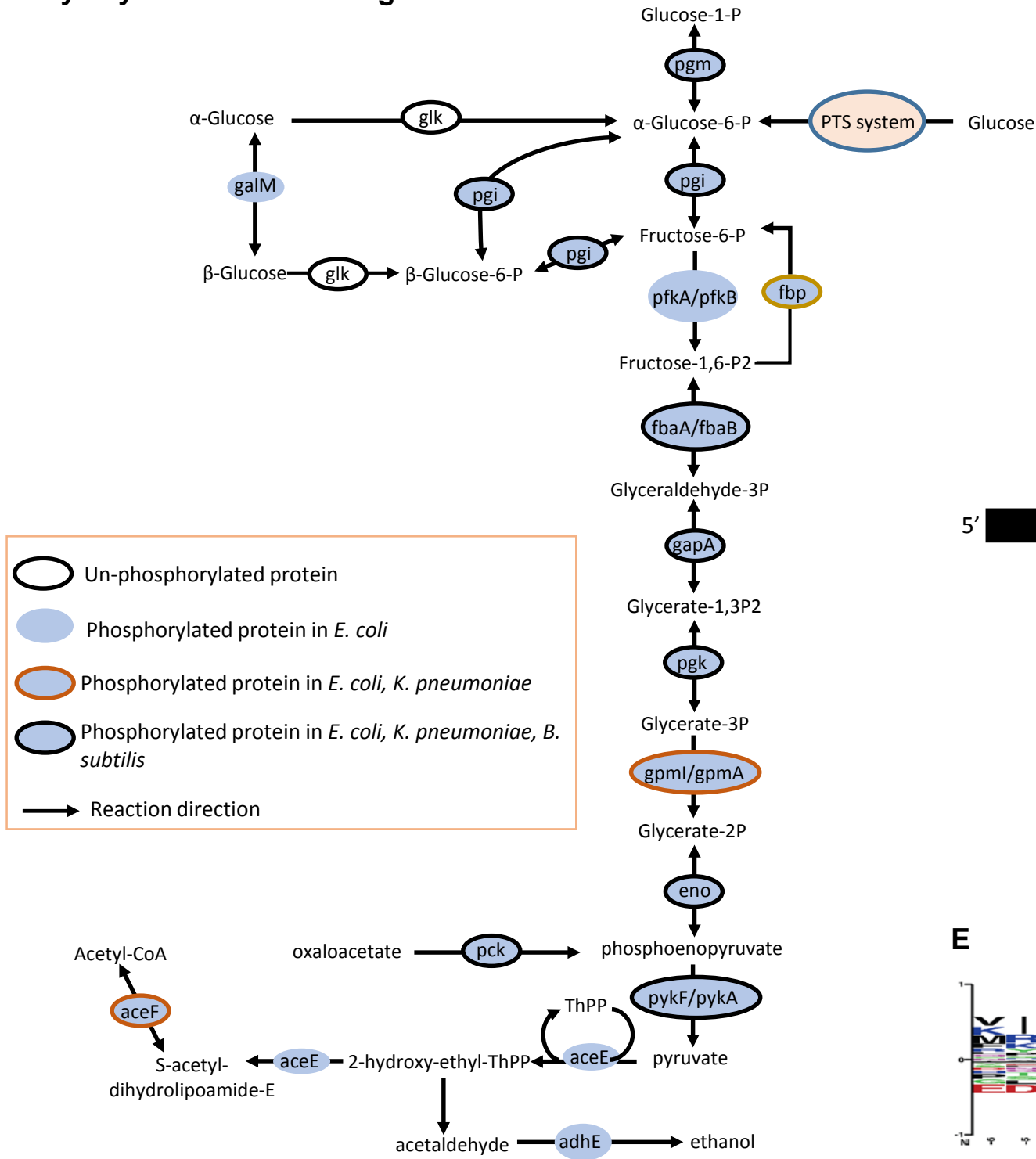


**C Phosphorylated ribosomal proteins**

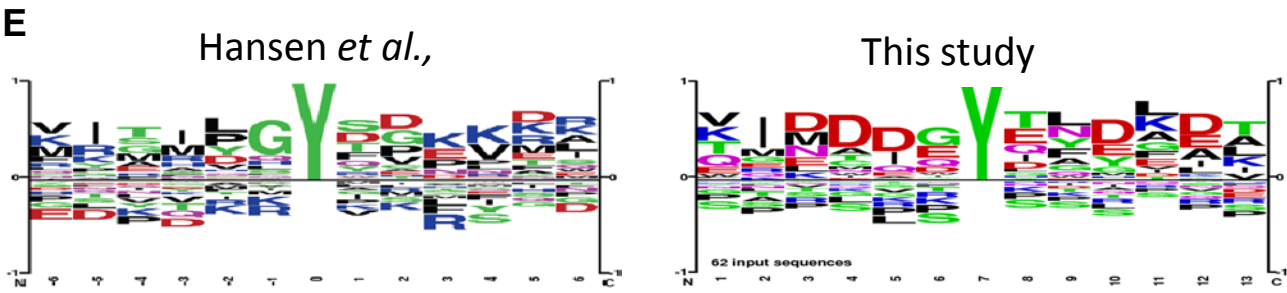
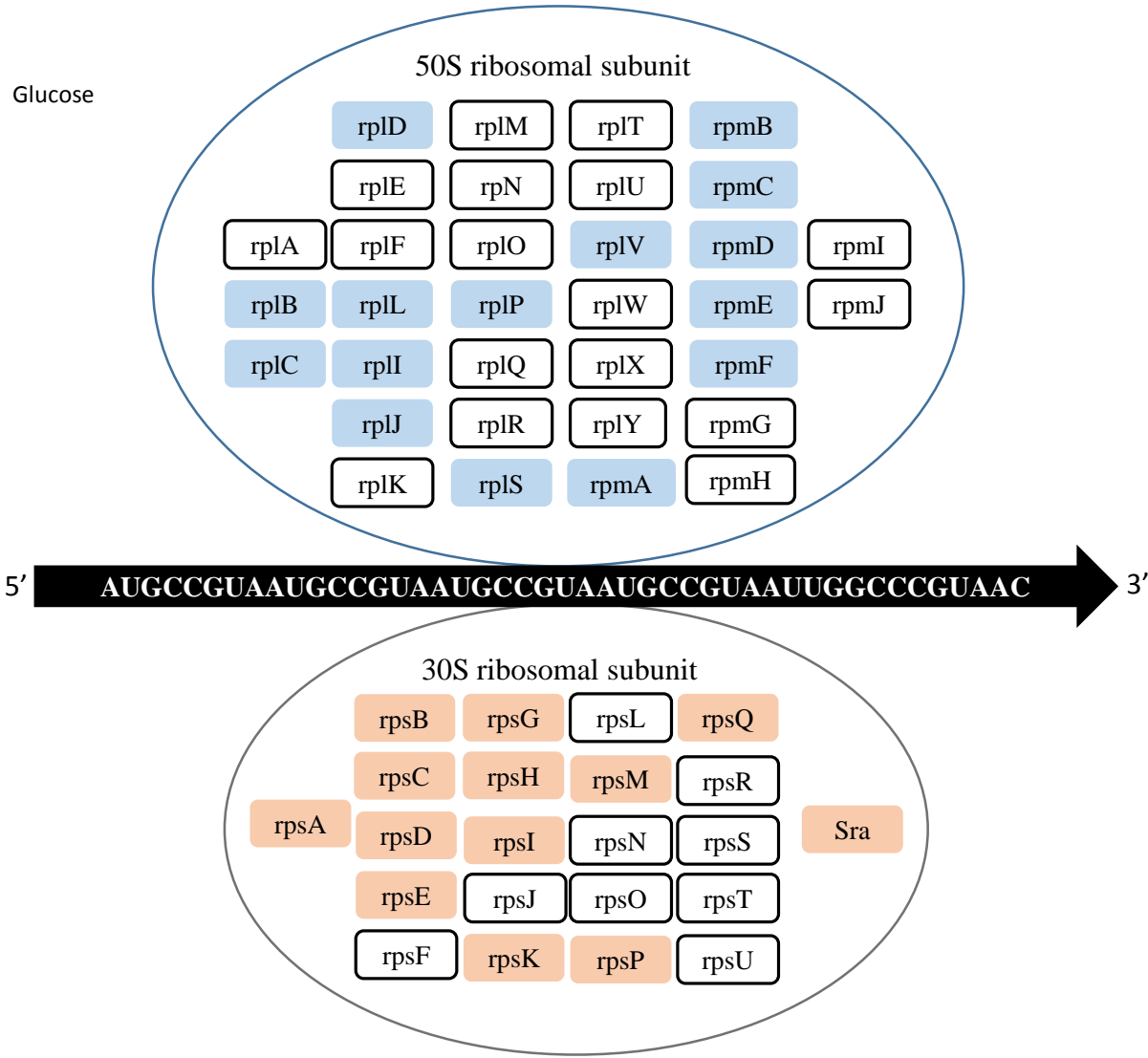




**C Glycolysis and Gluconeogenesis**

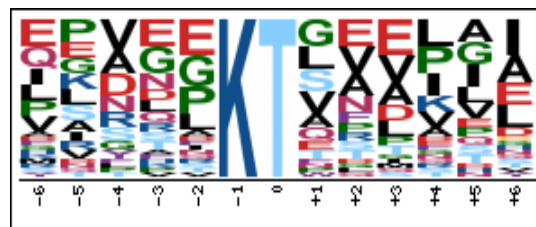
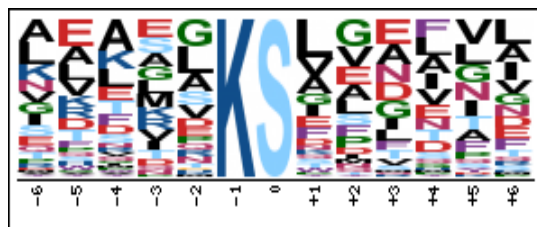


**D Ribosome**





**A**



**B**

