Phosphoproteome Analysis of E. coli Reveals Evolutionary Conservation of Bacterial Ser/Thr/Tyr Phosphorylation*

Boris Macek‡§, Florian Gnad‡§, Boumediene Soufi¶, Chanchal Kumar‡, Jesper V. Olsen‡, Ivan Mijakovic¶, and Matthias Mann‡¶

Protein phosphorylation on serine, threonine, and tyrosine (Ser/Thr/Tyr) is generally considered the major regulatory posttranslational modification in eukaryotic cells. Increasing evidence at the genome and proteome level shows that this modification is also present and functional in prokaryotes. We have recently reported the first in-depth phosphorylation site-resolved dataset from the model Gram-positive bacterium, Bacillus subtilis, showing that Ser/Thr/Tyr phosphorylation is also present on many essential bacterial proteins. To test whether this modification is common in Eubacteria, here we use a recently developed proteomics approach based on phosphopeptide enrichment and high accuracy MS to analyze the phosphoproteome of the model Gram-negative bacterium Escherichia coli. We report 81 phosphorylation sites on 79 E. coli proteins, with distribution of Ser/Thr/Tyr phosphorylation sites 68%/23%/9%. Despite their phylogenetic distance, phosphoproteomes of E. coli and B. subtilis show striking similarity in size, classes of phosphorylated proteins, and distribution of Ser/Thr/Tyr phosphorylation sites. By combining the two datasets, we created the largest phosphorylation site-resolved database of bacterial phosphoproteins to date (available at www.phosida.com) and used it to study evolutionary conservation of bacterial phosphoproteins and phosphorylation sites across the phylogenetic tree. We demonstrate that bacterial phosphoproteins and phosphorylated residues are significantly more conserved than their nonphosphorylated counterparts, with a number of potential phosphorylation sites conserved from Archaea to humans. Our results establish Ser/Thr/Tyr phosphorylation as a common posttranslational modification in Eubacteria, present since the onset of cellular life. Molecular & Cellular Proteomics 7:299–307, 2008.

The vital roles of protein phosphorylation on serine, threonine, and tyrosine (Ser/Thr/Tyr) in cell signaling and general regulation of protein activity are traditionally associated with eukaryotic cells (1). However, recent large scale genomic sequencing combined with the Global Ocean Sampling project (2) revealed that microbial eukaryotic-like protein kinases outnumber histidine kinases, traditionally considered the canonical microbial kinases (3). Likewise, our recent global and site-specific proteomics study of Ser/Thr/Tyr phosphorylation in Bacillus subtilis, a model Gram-positive bacterium, detected more than 100 phosphorylation events on 78 proteins, many of them classified as essential (4). Combined with previous two-dimensional gel studies using 32P labeling (5–7), this strongly suggests that Ser/Thr/Tyr protein phosphorylation is common in bacterial cell, although at much lower levels than in its eukaryotic counterpart, and that it may be important in the regulation of bacterial cellular processes.

Since the onset of molecular biology, Escherichia coli, a Gram-negative bacterium, has been one of the most used experimental cell systems; recently, it has become an invaluable host for heterologous gene cloning (8) and expression (9). Compared with other bacteria, studies of Ser/Thr/Tyr phosphorylation in E. coli commenced very early (7, 10). Two-dimensional gel experiments with protein extracts labeled with radioactive phosphorus suggested as many as 128 different Ser/Thr/Tyr phosphorylated E. coli protein spots. However most of them were only described in terms of pi and molecular weight (7) and, surprisingly, were never identified. Instead, other organisms, such as B. subtilis (5) and Corynebacterium glutamicum (6), became models for studying protein phosphorylation in bacteria; consequently, there is little current knowledge on Ser/Thr/Tyr phosphorylation in E. coli. Homologs of eukaryotic Ser/Thr kinases have so far been found in many bacterial species (11), but not in E. coli. In addition to a preliminary description of a protein kinase C-like activity (12), this organism possesses two well characterized Ser/Thr kinases: the isocitrate dehydrogenase kinase/phosphorylase (13) and the recently described YihE kinase (14). With respect to tyrosine phosphorylation, E. coli possesses two BY-kinases (15), Wzc and Etk (16), which are capable of auto- and substrate phosphorylation. Several E. coli proteins were also reported to autophosphorylate on Ser/Thr residues, including the molecular chaperone DnaK (17), an essential GTPase termed Era (18), and the nucleotide diphosphate...
Kinase (19). Presently, available data include 12 phosphorylation sites on 6 proteins in *E. coli*, obtained from individual biochemical studies (20).

Here we analyze the Ser/Thr/Tyr phosphoproteome of *E. coli* K12 using high accuracy MS in combination with phosphopeptide enrichment, and compare it with the phosphoproteome of *B. subtilis*. We report 81 phosphorylation sites on 79 *E. coli* proteins and show that the phosphoproteomes of these model bacteria have similar size and distribution of Ser/Thr/Tyr phosphorylation and share about 20% of phosphoproteins. We use these two datasets to study conservation of bacterial phosphoproteins and phosphorylation sites across the phylogenetic tree. Both phosphoproteomes show significantly higher conservation levels than a random protein population, with several phosphorylation sites being conserved from Archaea to humans, suggesting functional importance of this modification in prokaryotes as well as eukaryotes.

**MATERIALS AND METHODS**

*Cell Culture and Lysis—* Wild-type *E. coli* cells (strain: K12 isolate: MG1665) were grown in 4 L of Luria-Bertani (LB) medium under vigorous shaking. At OD600 = 1, the cells were precipitated by centrifugation (5 min at 8000 × g) and resuspended in lysis buffer containing 50 mM Tris-HCl (pH 7.5), 5 mg/ml lysozyme, and 5 mM of each of the following phosphatase inhibitors: sodium fluoride, 2-glycerol phosphate, sodium vanadate, and sodium pyrophosphate (Sigma). Cell wall lysis was performed for 15 min at 37 °C, and cell membranes were disrupted by sonication. DnaE (100 µg/ml) (Sigma) was added to the lysate and incubated for additional 10 min at 37 °C. N-octylglucoside detergent (Sigma) was added to final concentration of 1% for more efficient extraction and solubilization of membrane proteins. Cellular debris was removed by centrifugation at 25,000 × g for 30 min. The crude protein extract was extensively dialyzed against deionized water and lyophilized.

*Protein Digestion and Phosphopeptide Enrichment—* The lyophilized crude protein extract (about 20 mg) was dissolved in denaturation solution (6 M urea, 2 M thiourea in 20 mM ammonium hydrogen carbonate), and prepared for phosphopeptide enrichment using strong cation exchange (SCX) chromatography, essentially as described previously (4). Briefly, proteins were reduced with dithiothreitol, alkylated with iodoacetamide, digested for 3 h with endoproteinase Lys-C (1/100 w/w) (Waco), diluted 4 times with 20 mM ammonium hydrogen carbonate, and further digested overnight with trypsin (1/100 w/w) (Promega). The resulting peptide mixture was divided in two parts that were analyzed separately as two technical replicates. The first stage of phosphopeptide enrichment was performed on a Resource S (GE Healthcare) SCX column as described previously (4). Ten SCX fractions were collected in each run and subjected separately to the second stage of phosphopeptide enrichment using TiO2 beads (GL Sciences).

Before enrichment, SCX fractions and TiO2 beads were pre-incubated with 2.5 dihydroxybenzoic acid (final concentration 5 g/L). About 10 mg of TiO2 beads were added to each SCX fraction and incubated batch-wise for 60 min at room temperature. After washing twice with 80% acetonitrile/0.2% trifluoroacetic acid, peptides were eluted from the beads with 0.5% ammonium hydroxide solution in 40% acetonitrile (pH 10.5), almost completely dried in a vacuum-centrifuge and resuspended in 10 µl of 1% trifluoroacetic acid/2% acetonitrile in water for LC-MS analysis.

*LC-MS/MS Analysis—* Liquid chromatography was performed on a 1100 nano-HPLC (Agilent) coupled to the LTQ-Orbitrap mass spectrometer (Thermo Fisher) as described previously (4). The LTQ-Orbitrap was operated in the positive ion mode, with the following acquisition cycle: a full scan recorded in the orbitrap analyzer at resolution R = 60000 was followed by MS/MS of the five most-intense peptide ions in the LTQ analyzer. Multistage activation was enabled in all MS/MS events to improve fragmentation spectra of phosphopeptides and “lock mass” option was enabled in all full scans to improve mass accuracy of precursor ions (21).

*Data Processing and Validation—* Raw MS spectra were processed using Raw2msms software (21) and peak lists were searched against concatenated forward and reversed NCBI *E. coli* K12 protein database using Mascot search engine (Matrix Science). The following search criteria were used: full trypsic specificity was required; 2 missed cleavages were allowed; carbamidomethylation was set as fixed modification; oxidation (M), N-acetylation (protein), phosphorylation (STY), (H) and (D) were set as variable modifications; initial allowed precursor maximum mass deviation (MMU) was 10 ppm and fragment MMU was 0.5 Da. All phosphopeptide spectra identified by Mascot were processed and validated using MSQuant software. Stringent acceptance criteria were used: besides good coverage of b- and y-ion series, only peptides of six or more amino acids were considered, and the maximum allowed mass deviation of precursor ion was 7 ppm. The probabilities for phosphorylation at each potential site were calculated from the posttranslational modification (PTM) scores, as described previously (4, 22). Phosphorylation sites that were occupied with probability > 75% are reported as identified (“Localization p value” = 1 in the supplemental Table I).

*Enrichment Analysis of Gene Ontology Categories—* The E. coli identifiers were mapped to their UniProt counterparts using “EcoGene” database (http://ecogene.org). The Gene Ontology (GO) terms associated with these UniProt identifiers were extracted from EBI GOA site, which were used to create reference GO ontology in Cytoscape (23) ontology format. The Cytoscape plugin BiNgo (24) was used to perform enrichment analysis for the GO categories. The analysis was done using “Hypergeometric test,” and we selected all GO terms that were significant with p < 0.05 after correcting for multiple term testing by “Benjamini and Hochberg false discovery rate.”

*Evolutionary Analysis—* Homologous proteins to *E. coli* and *B. subtilis* proteins across 70 species from *Shigella flexneri* to human were determined via BLASTP (25). Homology search was performed on protein databases of 9 archaea, 53 bacteria, and 8 eukaryotes (for analyzed species, see supplemental Table IV). Databases were retrieved from Swissprot (http://expasy.org) in case of *Archaea* and *Eubacteria*, and SGD yeast database (26), FlyBase (27), and IPI (28) in case of eukaryotes. BLAST alignments were classified to be significant if the corresponding BLAST E values were lower than 1E-5. To distinguish between paralogs and orthologs, a two-directional BLASTP approach was used (29). Global alignments between homologous proteins were generated by the Needle software (30), based on the Needleman-Wunsch algorithm (30, 31).

*Data Presentation in the PHOSIDA Database—* Identified phosphoproteins, phosphorylation sites, and homology data were uploaded to the in-house curated PHOSIDA database (www.phosida.com) (22). In the evolutionary section of PHOSIDA, phosphorylated proteins showing two-directional homology between species are marked in green, proteins with one-directional homology are marked in blue, and pro-
teins that could not be aligned are marked in red. Integration of global alignments and their corresponding proportions of identities is also provided to enable users to define more conservative criteria for homology. In addition, PHOSIDA provides information on phosphor-ylation site conservation between species.

RESULTS

Our laboratory has developed a high-accuracy, high-sensitivity method to study the phosphoproteome of any species. We have previously applied this approach, based on enrichment and direct analysis of phosphopeptides rather than whole phosphoproteins, to both eukaryotic (22, 32) and prokaryotic (4) Ser/Thr/Tyr protein phosphorylation. In this study, we used this generic peptide-centric approach for analysis of Ser/Thr/Tyr protein phosphorylation in Escherichia coli at the phosphorylation site level (Fig. 1).

Briefly, protein extract from E. coli cells grown in LB medium was denatured, alkylated, and digested in-solution with a combination of endoproteinase Lys-C and trypsin. Resulting peptide mixture was subjected to two stages of phosphopeptide enrichment, using SCX and TiO₂ chromatography, and enriched samples were analyzed in duplicate using nano-LC-MS/MS on a high-accuracy LTQ-Orbitrap mass spectrometer. Peptide spectra were searched in sequence databases and validated using very stringent acceptance criteria (see “Materials and Methods”). Because several enzymes used in sample preparation originated from eukaryotic organisms (such as Bos bovis and Sus scrofa), we searched the detected E. coli phosphopeptide sequences, including possible amino acid substitutions that would result in very similar or identical mass (such as I/L, GG/N, and K/Q), against the complete nrNCBI database (version from September 2007). The search resulted in only one potential conflict: in addition to E. coli, the sequence of the glucose-6-phosphate isomerase peptide NRSNTPILVDGKDVMPEVNAVLEK is present in several eukaryotic organisms, including human. However, given the initial amount of E. coli protein extract used in the experiment (20 mg), the possibility of contamination with eukaryotic glucose-6-phosphate isomerase is extremely low.

Annotated spectra of all detected phosphopeptides are presented in the supplemental Fig. 1.

Phosphoproteome of E. coli—Using this approach, we detected 105 phosphopeptides from 79 E. coli proteins with an estimated false positive rate of less than 1%. Furthermore, the precise site of phosphorylation in the fragmented phosphopeptide was determined with probability higher than 75% for 81 phosphorylation sites (Table I). Very high mass accuracy was crucial to reaching these values, and we achieved an average absolute mass accuracy of the phosphopeptides of 1.03 ppm, despite their low abundance. A representative MS/MS spectrum of a detected phosphopeptide is presented in Fig. 1b. Details about identified phosphopeptides are listed in supplemental Table I. In the identified phosphopeptides, a total of 55 serines, 19 threonines, and 7 tyrosines were found to be phosphorylated, yielding a Ser/Thr/Tyr phosphorylation ratio of 67.9%/23.5%/8.6%. Information on the classes of detected phosphoproteins was obtained using GO analysis and is summarized in supplemental Table II. Detected phosphoproteins were significantly over-represented in the main pathways of carbohydrate metabolism (p = 1.2 E-5), protein biosynthesis (p = 9.6 E-5), and phosphoenolpyruvate-dependent phosphotransferase system (PTS) (p = 1.3 E-4). Although proteins from all parts of bacterial cell were detected, membrane proteins were under-represented (p = 4.9 E-7). No special attempt was made to target this protein population.
| Gene name | Phosphopeptide sequence | Gene name | Phosphopeptide sequence |
|-----------|-------------------------|-----------|-------------------------|
| acnB      | GFPLAYVGDVGp[TGSS]RK   | mdh       | GKPQGVEEVPIVHGHSVTLPPLOVQVGpSFTEQEVALTK |
| acnB      | LNKEPIEYLNpSNIVLVL   | pSGLTSAEQONALEGMQTLK  |
| adk       | YGIPOQPStGDMLR       | mdh       | TQLPp[SGS][ELS]DIAVPGVGAVDLHPIAVT      |
| ahpC      | AWHpSSSETIKAk          | menF      | RCGGGKApTLR          |
| ahpC      | DaPSDRLRK            | metK      | pSLEQAVMEEIIKPKPAELWTSATK |
| ampC      | KQVPVTQTLFELGpSVSK   | ndk       | HRDLDAPTPANALAGTLR |
| asd       | DFDAIRPVFFSTpSQLQAAPSFGGTTGLQDLAFDLEALK | pck | GDVAVFGp[SGT]JK |
| asnS      | KFENPvYGVDLSSEHER      | pgi       | NRp[SNT]PIVLDGK |
| cmoA      | pSHRDTLFSAPIAR        | pgi       | NRNPpTPIVLDGKDVMDPVEAVLKV  |
| crr       | Lp[SGS]VTGEPpVpIR     | pgk       | ALKEPARPMVAVGpSFK  |
| cyaA      | NONLPpSVSASpODIGVLpTR | pgk       | VSTLpTVLDLSK  |
| cyaA      | RNNLPp[SVS]ASpODIGVLpTR | pgk       | VSTLpTVLDLSK  |
| dacA      | ASVLPpSSEHLAApLQ      | pgm       | GGPLADGTVTPpSHNPPEDGpGK |
| dacC      | Gp[SS]VMFLKPGqvSVASLNK | pheS     | pSHLAEVLSAK  |
| deoB      | DTPpGSHWIAEVPGvLpFEGyFSDhENSPFOELLDK | pheS | pSHLAEVLSAK  |
| dnaM      | QLEAINDFGgettEAvpppTLPVpVDpSGK | pnp      | EGLVPpSOIApDk |
| dnaK      | pSLQGFNLDpNAPpR       | ppa       | AQIAHFFHpYK  |
| dfs       | SpYpLPiHNQVDpHLK      | pps       | GGRpTCAAIApR  |
| eno       | p[SGETEaD]ATADAVGtaAQGQk | pta     | KPVNDpLSR  |
| fruB      | MFpLSpSVQHDpGEK       | ptsA      | p[T]pLEVELHAQPpR  |
| fruB      | EQOp[STST]FGNIApHGGTTTDTTR | ptsH     | MFFFQEVTTAPpNGLHpTPAAQpFK  |
| fruB      | EQOTSTLFNIApHGGpTT[T]DTTR | psl  | VLGFTIDAGpGpTSHTSIMAR |
| fruB      | EQOTSTLFNIApHGGpTTTdRvDvLk | pyKpFk | LNPpSHGpDYAHpGQpR  |
| fruB      | QLTHLVp[SDDS]IAEQLK   | rplL      | DLVPpSAPpLAALK |
| fusA      | YDEAPPpSNAApAVIAR      | rplP      | GLAQGDpVSFpGSFLK  |
| fusA      | VVEpETEEPNTGpVDIgLpSR | rplS      | QDVpPpFPpGDTpVEVK  |
| galR      | GypSVSHR               | rpmB      | pSHALNATpK  |
| gatA      | pSSAIApLlRPTpNK        | rpsG      | p[SGS]p[ELAFApVALpENVRpTEVK  |
| gatB      | VDRpSFGpDIPLHVpGMpFISpGIEALQpNK | rxsG     | DGSpDFqDFTGApTTpR  |
| gatY      | SVMpADSHLPFpAQNPipLR   | rxsG      | KDGpDFDFTGApTTpR  |
| gatY      | LENpROWNLpPLpLVHGApSpGLSTK | secB     | DipSFApNpAPHpVFQpK  |
| gatY      | QWNPNLpLVHGApSpGLSTK  | serA      | SRpTHLpTEDVpINAAEK |
| gatZ      | pTAyWEyLVR            | talB      | EHYGpETVpMGAPpSFR  |
| glmM      | AEApGpISpSHNFPpDYpNGIK | talB     | LYQpPDpATTpNpSLpNApAQpPEYpR  |
| glnE      | LGEApFNp[p]SP[p]DIpLpFApPHEpGCTQpGpGR | tnaA | NIFpGpYQpYTIPpHTQpQR  |
| glyS      | SPpDLVp[SPp]DpRNpASTL| tpx       | SQVpHFQpGpNPpTVpANpSpQpAGpSK  |
| gpmM      | QMGpMpSEpVGHpNLpAGR  | tsp       | AEITAtppSLV  |
| groS      | VGDpIVFpNDGPpYpGK   | tufA      | AIDKFPpLLIPDEpVFSpSQR |
| guaA      | pSHpHNpVGGLpK          | tufA      | ARGIPp[TINTS]pHVEpYDPpTR |
| hcaD      | WIQpSgpKpTFpDk         | tufA      | FpESEpVLpS RK  |
| hinT      | Dip[SPp]QpAPT[p]pILIpNLpIPTVDpNSvAEShEOALpQR | tufA     | GITpNP[p]pHEpYDPpTR  |
| hipA      | EEnpDFpRpVpAQGQEK     | tufA      | GitpINTSHpVEpYDPpTR |
| hisJ      | LGpFgpGpTMpGLRk       | tufA      | GYpRpQFpYpFR  |
| htpG      | FApSTHTpDSSApAQTVSPLEDPYpVR | tufA | ELpSpQYDPFGpPDpTPpR |
| htpG      | pSFpQLEpLSGSDpAQK      | tufA      | ELpSqpDFpGpDGp[p]pTPpRGSpALK  |
| hybD      | ISpFLpQGLAVLpSALR     | udk       | MKEp[TNT]pTEpVK  |
| icd       | pSLpNLVApLR           | uspG      | NPpSISpTLLGlSpNSApSQR  |
| icd       | GpLpTTPpVGpGIpSpNLpVApLR | ybpB    | LLLApTQpLR  |
| kdgR      | ANADLpKpOqDpSpSVS[Vp]VLk | ycfH | pSTFLpNALTk  |
| kdsD      | pSHpVLpQpGFpDQpAQK   | ydcY      | SHLPpDEpAR  |
| lysS      | TkpGpTELpSIHCTELR     | ydpE      | FTNPpYAHpGpSVpSQR  |
| lysS      | pSEpQAQGpDApAVDLpNLNELk | ygfB    | NpAQpGpYDEDEpDpDEpQpMSLPpEPIpYpVR  |
| lysU      | TQpGpTELpSIHCTELR     | yijL      | LTPpYEpLpSDALpAGEpIVpSpSAGpGQpEK  |

**Phosphoproteome of Escherichia coli**

*Detected phosphopeptides from the digest of total E. coli cell lysate.*

Nonredundant gene symbols, rather than protein names, were used for clarity.
and this finding may simply reflect the commonly observed undersampling of membrane proteins because of their biophysical properties, rather than their lower phosphorylation levels.

*E. coli* (strain K12 isolate MG1665) was recently the subject of a global study where genetic footprinting identified 620 genes as essential for growth in aerobic conditions in a rich medium (33). Of 79 phosphoproteins detected in our study, 21 were classified as essential, 45 as dispensable, and for 12 there was no information on essentiality (supplemental Table I). Therefore, at least 27% of detected *E. coli* phosphoproteins are classified as essential compared with about 17% at the genome level. This presents a highly significant over-representation in essentiality of *E. coli* phosphoproteins.

To check whether detected phosphorylation sites in *E. coli* conform to known target sequences of eukaryotic protein kinases, we matched their surrounding motifs with 33 known kinases. We matched their surrounding motifs with 33 known kinases and determined their interspecial homologs. To distinguish between paralogs and orthologs, we aligned the best-scoring homologous protein from every species with all proteins of the tested proteome (*E. coli* or *B. subtilis*). We defined orthologs as proteins that showed the highest homology to each other in both directions. To measure the phosphoproteome conservation, we counted the number of phosphoprotein orthologs in every analyzed species and reported it as the percentage of the total number of proteins in the tested phosphoproteome. For example, if in a certain species an ortholog was found for every detected *E. coli* phosphoprotein, the evolutionary conservation of *E. coli* phosphoproteome in that species was reported as 100%. Phosphoproteome conservation in domains *Archaea*, *Eubacteria*, and *Eukarya* was reported as an average conservation in all analyzed species within a domain. Conservation at the proteome level was assessed and reported in the same manner.

This approach showed that the phosphoproteomes of *E. coli* and *B. subtilis* are highly conserved throughout all analyzed species. Details about the conservation of the *E. coli* phosphoproteome are summarized in the supplemental Table IVa. More than a half of *E. coli* phosphoproteins were found to be conserved in other bacterial organisms, in contrast to about 25% of the nonphosphorylated proteins (Fig. 2a). As expected, the highest conservation was found within *γ-Proteobacteria* and the highest degree of homology in the closely related *Shigella flexneri*, where 91% of the identified phosphoproteins had a homolog. Nearly 40% of *E. coli* phosphoproteins were also conserved in eukaryotes, in contrast to around 14% of the nonphosphorylated proteins. Unlike in prokaryotes, the proportion of conserved phosphoproteins was very constant throughout eukaryotic species. The extent of homology in *Archaeb* was comparable with that of eukaryotes: 31% of *E. coli* phosphoproteins were conserved at the phosphoproteome, compared with only 13% at the proteome level.

Details about the conservation of the phosphoproteome of *B. subtilis* are summarized in the supplemental Table IVb. The
Evolutionary Conservation of Bacterial Phosphorylation Sites—
The simplest and most common way to analyze conservation of protein sequences is direct application of the BLAST algorithm. However, BLAST presents a heuristic method to find significant local, rather than global, alignments. Because phosphorylation sites are mainly located in rapidly evolving loop regions, we first checked evolutionary conservation at the protein level. To identify orthologs and distinguish between orthologs and paralogs, we used a two-directional BLAST approach. To cover the entire protein sequence alignments of orthologs, we then applied Needle (30), a program based on the Needleman-Wunsch algorithm (31), to study the evolutionary conservation at the phosphorylation site level.

Phosphorylation sites in *E. coli* and *B. subtilis* were on average more conserved than their nonphosphorylated counterparts. This observation is most evident in *Archaea*: in the *E. coli* dataset, 44% of phosphorylated serines were conserved, in comparison to 23% of nonphosphorylated serines (Fig. 3a). Accordingly, 54% phosphothreonines were conserved in comparison to 28% of their nonphosphorylated counterparts (Fig. 3b). In the *B. subtilis* dataset, 32% of phosphorylated serines were conserved, in comparison to 24% of nonphosphorylated serines (Fig. 4a) and 62% of phosphothreonines were conserved in comparison to 29% of nonphosphorylated threonines (Fig. 4b). Because of their low number, we could not draw any statistically significant conclusions about phosphorylated tyrosines.

Conservation of *E. coli* phosphorylation sites within eukaryotic species shows the same tendency as in the other domains. Interestingly, yeast presents an extreme outlier (phosphoserine, 19%; non-phosphoserine, 35%; phosphothreonine, 33%; non-phosphothreonine, 41%). If yeast is excluded, the conservation of phosphorylation sites in eukaryotes is clearly higher than the one of nonphosphorylated serines and threonines (phosphoserine, 39%; non-phosphoserine, 35%; non-phosphothreonine, 33%);
phosphothreonine, 47%; non-phosphothreonine, 41%). In B. subtilis, 44% of phosphoserines are conserved in eu-
aryotes (including yeast) in comparison to 30% of nonphos-
phorylated serines. However, the conservation pattern of
phosphothreonines was similar to the one of their nonphos-
phorylated counterparts (phosphothreonine, 29%; non-phos-
phorylated threonine, 32%). Conservation of
E. coli and B. subtilis phosphorylation sites in other bacteria was found to be quite
variable because of phylogenetic diversity within the domain.

Interestingly, phosphorylated residues from the following nine
proteins were found to be conserved from Archaea to humans:
cysteinyl t-RNA synthetase (S270 in B. subtilis), phosphoglu-
comutase (S100 in B. subtilis; S146 in E. coli), nucleoside
diphosphate kinase (T92 in B. subtilis; T93 in E. coli), pyruvate
kinase (S36 in B. subtilis and E. coli), enolase (Y281 in B. sub-
tilis), predicted GTP-binding protein (S16 in E. coli), α-3
phosphoglycerate dehydrogenase (T63 in E. coli), phosphogluco-
samine mutase (S100 in B. subtilis; S102 in E. coli), and elongation
factor Ei-Tu (T62 in E. coli). Note that observed conservation of
these residues does not mean that they are actually phospho-
rylated in all species. However, because the roles of several of
these phosphorylation events have been shown to be essential
for function (34–36), it is likely that they are also present in many
other organisms where the phosphorylation site is conserved.

DISCUSSION

Traditionally, PTMs on proteins are associated almost ex-
clusively with eukaryotic rather than prokaryotic cells. How-
ever, a growing number of bacterial PTMs, partly uncovered
through the increasing power of MS-based proteomics, lately
received increased attention of the research community. Such
modifications may be involved in hitherto unknown or undera-
preciated regulatory mechanisms of bacterial metabolic and
signaling networks and perhaps present potential targets for
new drugs. Here we have used powerful MS-based technol-
yogy, previously used to detect several thousand phosphoryl-
ation events in mammalian cells (22) and about 100 phos-
phorylation sites in the Gram-positive bacterial model B. subtilis
(4), to study the phosphoproteome of E. coli. By analyzing two
distant bacterial phosphoproteomes, we reasoned that we
could determine the generality of Ser/Thr/Tyr phosphorylation
in bacterial and acquire sufficient data to study its conserva-
tion in taxonomically distant species.

Several insights obtained from the phosphoproteome of
B. subtilis are now confirmed in the Gram-negative E. coli. We
find that, in bacterial cells growing on LB medium, there are at
least 80 proteins phosphorylated on Ser/Thr/Tyr residues. The
exact number of phosphoproteins will vary between species
and it is at present still unknown, but two-dimensional gel
studies using radioactive detection and performed on the two
microorganisms estimate it at not more than 150 (5, 7). In
B. subtilis and E. coli, this presents about 5% of the anno-
tated genes, which is a significantly lower number than in any
eukaryotic organism, including yeast (32).

Most bacterial phosphorylation sites are on serine (70%)
and threonine (20%), and a lesser number are on tyrosines
(10%). Several proteins are multiply phosphorylated; how-
ever, this population is particularly scarce compared with
eukaryotic cell, where multiple phosphorylated proteins are
common and often serve as molecular “switchboards” for
integration of signals. The virtual absence of these proteins in
analyzed bacteria may not only reflect the much smaller num-
ber of phosphorylation events overall but also the fundamen-
tal difference in the role of Ser/Thr/Tyr in eukaryotic and
prokaryotic cell: in the former it is primarily involved in signal
transduction, whereas in the latter our data point to other
cellular functions. Among bacterial phosphoproteins, there is
a significant over-representation of enzymes involved in car-
bon metabolism (i.e. glycolysis, sugar transport) and essential
proteins, which further supports the emerging view of Ser/Thr/
Tyr phosphorylation as an important regulatory mechanism in
the bacterial cell. One potential mechanism for such a regu-
lation has already been described: phosphorylation of iso-
citrate dehydrogenase (icd) in E. coli occurs at Ser113, the
residue involved in binding of the substrate (isocitrate), and
therefore abolishes the enzyme action (37). Another such ex-
ample may come directly from our study: the Thr93 residue of
the essential protein nucleoside diphosphate kinase (ndk),

![Fig. 4. The evolutionary conservation of the phosphorylation sites from the phosphoproteome of B. subtilis. a, conservation of phosphoserine. b, conservation of phosphothreonine.](image-url)
found to be phosphorylated in both *E. coli* and *B. subtilis* and conserved from Archaea to humans, serves as an ATP-binding site. It is therefore reasonable to speculate that its phosphorylation will have a dramatic impact on the enzyme function.

Several important practical considerations are invoked by the results presented here. With more than 100 detected phosphopeptides in *E. coli*, it is somewhat surprising that only a few protein kinases capable of phosphorylating proteins at Ser/Thr/Tyr residues have been identified in this organism. This apparent lack of kinases lead to a widespread notion that this type of protein phosphorylation is practically absent in *E. coli* and bacteria in general. Therefore, it is widely assumed that a given eukaryotic protein, synthesized in *E. coli*, would purify in a homogenous nonphosphorylated form. However, with the potential of *E. coli* to phosphorylate proteins at Ser/Thr/Tyr residues demonstrated in this study, this assumption should be reexamined when considering *E. coli* and other bacteria as heterologous expression hosts.

The growing dataset of bacterial Ser/Thr/Tyr phosphorylation sites and the lack of similarity between target sequences of prokaryotic and eukaryotic protein kinases clearly indicates the existence of protein kinases specific for the bacterial kingdom. Currently available protein phosphorylation predictors are based on eukaryotic phosphorylation data (38); and although there are instances where their predictions were successful in bacteria (39), they remain of limited use for these systems. Therefore, the need for bacterial-specific phospho-predictors is evident and our data may provide a basis for training such predictors.

The high degree of conservation of proteins that are phosphorylated in *E. coli* or *B. subtilis* suggests their early presence in the course of evolution. In addition, it reflects the general constraints on these proteins, which are involved in major pathways such as carbon metabolism. However, a significant number of phosphorylation sites differ in the phosphoproteomes of the Gram-positive and Gram-negative models, and the extent of overall homology is not uniform because of the high diversity within bacteria. This indicates that protein phosphorylation kept on evolving as bacteria developed their particular lifestyles to conquer new niches. Traces of this evolutionary process can also be seen in the versatility of bacterial-specific kinase families, which evolved from nonkinase ancestors, such as the BY-kinases (15), the bifunctional Ser/Thr kinase/phosphorylases, and the tyrosine kinases with a guanidino-phosphotransferase domain (40). One important evolutionary factor that has also to be considered when examining the conservation of phosphoproteins is gene loss in parasites, which is mainly caused by the availability of the essential proteins from the host (41, 42). An interesting example is *Buchnera aphidicola*, an intracellular symbiont of the greenbug aphid (43). Although closely related to *E. coli*, only 13% of the entire *E. coli* proteome has an ortholog in this bacterium. However, 41.25% of the measured phosphoproteins in *E. coli* are present in *B. aphidicola*, clearly showing a higher conservation of the phosphoproteome compared with the entire proteome. The higher conservation of the proteins present in phosphoproteomes of *E. coli* and *B. subtilis* is evident throughout all selected species. The *E. coli* phosphoproteome appears to be even more conserved; however, this is related to the fact that Gram-negative bacteria outnumber *Firmicutes* in the set of selected representative bacteria.

Therefore, for example, the phosphoproteins that are required for the spore cortex synthesis appear as poorly conserved in the analyzed bacterial set. In addition, the proportion of hypothetical or unknown proteins with low conservation is much higher in the *B. subtilis* set of phosphoproteins so the conservation of *B. subtilis* proteins in general appears lower in Archaea and eukaryotes in comparison to *E. coli*.

Nine specific phosphorylation sites from this study are conserved from Archaea to humans. They are in cysteinyl t-RNA synthetase, phosphoglucosamine mutase, nucleoside diphosphate kinase, pyruvate kinase, enolase, predicted GTP-binding protein, d-3 phosphoglycerate dehydrogenase, phosphoglucomutase and elongation factor Ef-Tu. This unusually high conservation suggests that the phosphorylation of these sites may be also conserved in all life forms, although this has still to be proven. Nevertheless, the high conservation suggests that phosphorylation is in each case indispensable for enzyme activity, which is in turn essential for the cell. Good examples are phosphoglucosamutases and phosphoglucomutase mutases, where phosphorylation of a key serine residue is a part of the reaction mechanism in which the phosphate is being transferred to the reaction product (34, 35). Another example is the phosphorylation of the elongation factor Ef-Tu, which occurs after its binding to the A-site of the ribosome (36); this phosphorylation triggers the release of Ef-Tu from the ribosome and therefore occurs in a cyclic fashion to optimize the translation rate. These highly conserved essential phosphorylation events can hardly be considered regulatory events in themselves, but they illustrate how the earliest life forms immediately used protein phosphorylation in the housekeeping of their cells.

* This study was supported by the Max Planck Society (to M. M.), the EU Framework 6 “Interaction proteome” grant (to M. M., B. M.), and the Danish Natural Science Research Council (to I. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

§ These authors have equally contributed to this work.

∫ To whom correspondence should be addressed. Tel.: 49-89-8578-2557; Fax: 49-89-8578-2219; E-mail: mmann@biochem.mpg.de.

REFERENCES

1. Pawson, T., and Scott, J. D. (2005) Protein phosphorylation in signaling: 50 years and counting. Trends Biochem. Sci. 30, 286–290

2. Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W., Jaroszewski, L., Clepik, P., Miller, C. S., Li, H., Mashiya, S. T.,
22. Olsen, J. V., de Godoy, L. M., Li, G., Mac, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol. Cell. Proteomics 4, 1020–2021.

23. Olsen, J. V., Blagoev, B., Grudzien, D., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635–648.

24. Maere, S., Heymans, K., and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448–3449.

25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

26. Sood, P., Lerner, C. G., Shimamoto, T., Lu, Q., and Inouye, M. (1994) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

27. Grumbling, G., and Strelets, V. (2006) FlyBase: anatomical data, images and queries. Nucleic Acids Res. 34, D484–D488.

28. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

29. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.

30. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

31. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

32. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faereng, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pearomine signaling pathway. Mol. Cell. Proteomics 4, 310–327.

33. Gundersen, H. M., Balazsi, G., Ravasz, E., Daugherty, D. M., Somera, A. L., Krypides, N. C., Anderson, I., Gelfand, M. S., Kocivar, C., and Kyrpides, N. C. (2005) Construction and characterization of a novel two-plasmid system for mapping essential genes in Escherichia coli. J. Mol. Biol. 348, 1–8.

34. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

35. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.

36. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

37. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faereng, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pearomine signaling pathway. Mol. Cell. Proteomics 4, 310–327.

38. Gundersen, H. M., Balazsi, G., Ravasz, E., Daugherty, D. M., Somera, A. L., Krypides, N. C., Anderson, I., Gelfand, M. S., Kocivar, C., and Kyrpides, N. C. (2005) Construction and characterization of a novel two-plasmid system for mapping essential genes in Escherichia coli. J. Mol. Biol. 348, 1–8.

39. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

40. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.

41. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

42. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faereng, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pearomine signaling pathway. Mol. Cell. Proteomics 4, 310–327.

43. Gundersen, H. M., Balazsi, G., Ravasz, E., Daugherty, D. M., Somera, A. L., Krypides, N. C., Anderson, I., Gelfand, M. S., Kocivar, C., and Kyrpides, N. C. (2005) Construction and characterization of a novel two-plasmid system for mapping essential genes in Escherichia coli. J. Mol. Biol. 348, 1–8.

44. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

45. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.

46. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

47. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faereng, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pearomine signaling pathway. Mol. Cell. Proteomics 4, 310–327.

48. Gundersen, H. M., Balazsi, G., Ravasz, E., Daugherty, D. M., Somera, A. L., Krypides, N. C., Anderson, I., Gelfand, M. S., Kocivar, C., and Kyrpides, N. C. (2005) Construction and characterization of a novel two-plasmid system for mapping essential genes in Escherichia coli. J. Mol. Biol. 348, 1–8.

49. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

50. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.

51. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

52. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faereng, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pearomine signaling pathway. Mol. Cell. Proteomics 4, 310–327.

53. Gundersen, H. M., Balazsi, G., Ravasz, E., Daugherty, D. M., Somera, A. L., Krypides, N. C., Anderson, I., Gelfand, M. S., Kocivar, C., and Kyrpides, N. C. (2005) Construction and characterization of a novel two-plasmid system for mapping essential genes in Escherichia coli. J. Mol. Biol. 348, 1–8.

54. Kersey, P. J., Duarte, J., Williams, A., Remm, M., and Sonnhammer, E. L. (2005) InParanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

55. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276–277.

56. Needleman, S. B., and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

57. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faereng, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pearomine signaling pathway. Mol. Cell. Proteomics 4, 310–327.