

Evolutionary Constraints of Phosphorylation in Eukaryotes, Prokaryotes, and Mitochondria*[§]

Florian Gnad[‡], Francesca Forner[§], Dorota F. Zielinska[§], Ewan Birney[¶],
Jeremy Gunawardena[‡], and Matthias Mann[§]||

High accuracy mass spectrometry has proven to be a powerful technology for the large scale identification of serine/threonine/tyrosine phosphorylation in the living cell. However, despite many described phosphoproteomes, there has been no comparative study of the extent of phosphorylation and its evolutionary conservation in all domains of life. Here we analyze the results of phosphoproteomics studies performed with the same technology in a diverse set of organisms. For the most ancient organisms, the prokaryotes, only a few hundred proteins have been found to be phosphorylated. Applying the same technology to eukaryotic species resulted in the detection of thousands of phosphorylation events. Evolutionary analysis shows that prokaryotic phosphoproteins are preferentially conserved in all living organisms, whereas site specific phosphorylation is not. Eukaryotic phosphosites are generally more conserved than their non-phosphorylated counterparts (with similar structural constraints) throughout the eukaryotic domain. Yeast and *Caenorhabditis elegans* are two exceptions, indicating that the majority of phosphorylation events evolved after the divergence of higher eukaryotes from yeast and reflecting the unusually large number of nematode-specific kinases. Mitochondria present an interesting intermediate link between the prokaryotic and eukaryotic domains. Applying the same technology to this organelle yielded 174 phosphorylation sites mapped to 74 proteins. Thus, the mitochondrial phosphoproteome is similarly sparse as the prokaryotic phosphoproteomes. As expected from the endosymbiotic theory, phosphorylated as well as non-phosphorylated mitochondrial proteins are significantly conserved in prokaryotes. However, mitochondrial phosphorylation sites are not conserved throughout prokaryotes, consistent with the notion that serine/threonine phosphorylation in prokaryotes occurred relatively recently in evolution. Thus, the phosphoproteome reflects major events in the evolution of life. *Molecular & Cellular Proteomics* 9:2642–2653, 2010.

From the [‡]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, [§]Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany, and [¶]European Bioinformatics Institute (European Molecular Biology Laboratory-EBI), Hinxton, Cambridge CB10 1SD, United Kingdom

Received, July 4, 2010, and in revised form, August 4, 2010

Published, MCP Papers in Press, August 5, 2010, DOI 10.1074/mcp.M110.001594

Reversible protein phosphorylation on serines, threonines, and tyrosines plays a crucial role in regulating processes in all living organisms ranging from prokaryotes to eukaryotes (1). Traditionally, phosphorylation has been detected in single, purified proteins using *in vitro* assays. Recent advances in mass spectrometry (MS)-based proteomics now allow the identification of *in vivo* phosphorylation sites with high accuracy (2–7). On-line databases such as PhosphoSite (8), Phospho.ELM (9), and PHOSIDA¹ (10) have collected and organized thousands of identified phosphosites. These databases as well as dedicated analysis environments such as NetworkKIN (11, 12) offer and use contextual information including structural features, potential kinases, and conservation. They constitute resources that should allow the derivation of general patterns for phosphorylation events. Specifically, the recent availability of data for archaeal, prokaryotic, and diverse eukaryotic phosphoproteomes in these databases should enable investigation of the evolutionary history of this post-translational modification.

Prokaryotes have two separate classes of phosphorylation events. Apart from the canonical histidine/aspartate phosphorylation, which has been studied for decades, serine/threonine/tyrosine phosphorylation is also present and has recently become amenable to analysis by MS (13). Bacterial phosphoproteins are involved in protein synthesis, carbohydrate metabolism, and the phosphoenolpyruvate-dependent phosphotransferase system. Recent phosphoproteomics studies of *Bacillus subtilis*, *Escherichia coli*, and *Lactococcus lactis* described around 100 phosphorylation sites on serine, threonine, and tyrosine in each of these species (13–15). Bacterial phosphorylation sites can change in response to environmental conditions (16).

Interestingly, even archaea have serine/threonine and tyrosine phosphorylation. A recent study of *Halobacterium salinarum* described 75 serine/threonine/tyrosine phosphorylation sites on 62 proteins involved in a wide range of cellular processes including a variety of metabolic pathways (17).

Although only a few hundred phosphorylation events have been found in prokaryotic species, similar experimental conditions and effort have yielded the detection of thousands of phosphorylation events in eukaryotes ranging from yeast to

¹ The abbreviations used are: PHOSIDA, posttranslational modification database; IPI, International Protein Index; FDR, false discovery rate; PTM, post-translational modification; LTQ, linear trap quadrupole.

human (7, 18–21). Before the advent of large scale phosphoproteomics, serine/threonine/tyrosine phosphorylation has been estimated to affect one-third of all proteins (22). Recent large scale phosphoproteomics studies now suggest that more than half of all eukaryotic proteins are phosphorylated (23).

A key event in evolution was the endosymbiosis of prokaryotes that enabled the development of a much more complex type of life, the eukaryotic cell. Analyses of mitochondrial genes suggest that the α -proteobacterium *Rickettsia prowazekii* is the endosymbiotic precursor leading to modern mitochondria (24). Almost all of the mitochondrial genes have migrated to the nuclear genome during subsequent evolution, and it is predicted that 10–15% of eukaryotic nuclear genes of organisms encode mitochondrial proteins (25).

Thus, mitochondria with their unique evolutionary position between prokaryotes and eukaryotes form an interesting link for the evolutionary analysis of phosphorylation. Several studies investigated the mitochondrial phosphoproteome in different organisms using gel electrophoresis or specific enrichment methods coupled with mass spectrometry (26–28). Those studies established potential mitochondrial phosphoproteins. Three large scale studies based on affinity enrichment of phosphopeptides and mass spectrometry obtained direct experimental evidence of phosphorylation sites in mitochondria. Lee *et al.* (29) used a combination of different peptide enrichment strategies and found 80 phosphorylation sites of 48 different proteins from mouse liver. Very recently, a study by Deng *et al.* (30) characterized the murine cardiac mitochondrial mouse phosphoproteome, covering 236 phosphosites on 181 proteins. Investigating yeast, Reinders *et al.* (31) assigned 84 phosphorylation sites in 62 proteins.

To enable comparative analysis of phosphoproteomes between all domains of life and mitochondria, here we experimentally determined a high accuracy mitochondrial mouse phosphoproteome based on technology conditions similar to those applied to the identification of prokaryotic and eukaryotic phosphoproteomes. We then performed a detailed evolutionary study of the conservation of the identified phosphoproteins and phosphorylation sites in prokaryotes and in eukaryotes. This allowed an initial comparison of the phosphoproteomes of prokaryotes, mitochondria, and eukaryotes.

EXPERIMENTAL PROCEDURES

Cell Culture and Primary Cell Isolation—3T3-L1, brown preadipocytes, C2C12, and Hepa 1-6 cell lines were subcultured and differentiated in DMEM supplemented with 10% fetal bovine serum (Invitrogen) and antibiotics in 5% CO₂ at 37 °C. Stable isotope labeling by amino acids in cell culture was performed as described with L-[¹³C₆, ¹⁵N₂]lysine and L-[¹³C₆, ¹⁵N₄]arginine. 3T3-L1 preadipocytes were grown and differentiated as described previously (32). Brown preadipocytes were differentiated as described (33). Confluent C2C12 were differentiated into myotubes by reducing the percentage of serum to 2%. For the isolation of primary brown adipocytes, interscapular brown adipose tissue from 20 C57BL/6 newborn mice (1–2 days) was excised, immersed in Hank's buffered salt solution,

cleaned free of connective tissue under a binocular microscope, minced, and digested with 1 mg/ml collagenase A (Roche Applied Science) at 37 °C for 30 min. After digestion, the slurry was passed through 250- μ m-mesh opening fiber material (Sefar) and centrifuged at 500 \times *g* for 1 min. The floating adipocytes and the supernatant were discarded. The stromal vascular fraction was resuspended in DMEM supplemented with 10% fetal bovine serum (Invitrogen) and antibiotics and transferred into 7-cm-diameter Petri dishes. Preadipocytes were grown to confluence and differentiated as described (33).

Isolation of Mitochondria—Cells were harvested with trypsin (Invitrogen) and diluted with DMEM supplemented with protease inhibitors (Roche Applied Science) and a 5 mM concentration of each of the following phosphatase inhibitors: sodium fluoride, 2-glycerol phosphate, sodium vanadate, and sodium pyrophosphate (Sigma). Cells were centrifuged at 1000 \times *g* for 10 min. Cells were then resuspended with SEH buffer (250 mM sucrose, 10 mM Hepes, pH 7.4, 0.1 mM EGTA) supplemented with protease and phosphatase inhibitors and washed twice. The cell suspensions were homogenized in a 7-ml Dounce homogenizer. Membrane disruption was checked with trypan blue staining. The tissue homogenate was centrifuged twice at 800 \times *g* for 10 min at 4 °C. The supernatant was centrifuged at 10,000 \times *g* for 10 min at 4 °C. The crude mitochondrial pellet was resuspended in SEH buffer supplemented with protease inhibitors (Roche Applied Science). The suspension was further centrifuged at 7000 \times *g* for 10 min at 4 °C and purified with Percoll gradients as described (34). 3T3-L1 cell mitochondria were purified by means of protease treatment of crude mitochondrial fractions with trypsin as described previously (35).

Phosphopeptide Enrichment—Mitochondrion-enriched fractions were resuspended in SEH buffer supplemented with inhibitors. 500 μ g of mitochondrial proteins were dissolved in 6 M urea, 2 M thiourea, 20 mM Hepes, pH 7.4; reduced with dithiothreitol; alkylated with iodoacetamide; digested for 3 h with endoproteinase Lys-C (1:50, w/w) (Waco); diluted 4 times with 10 mM ammonium hydrogen carbonate; and further digested overnight with trypsin (1:50, w/w) (Promega). Trypsin cleaves peptide chains at the carboxyl side lysine or arginine except when either is followed by proline. Lys-C cleaves at lysine residues. The resulting peptide mixture was captured on TiO₂ beads (GL Sciences). Briefly, TiO₂ beads were preincubated with 2,5-dihydroxybenzoic acid (final concentration, 5 g/liter). About 10 mg of TiO₂ beads were added to each sample and incubated 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 μ l of 1% trifluoroacetic acid, 2% acetonitrile in water for LC-MS analysis.

Mass Spectrometry—Liquid chromatography was performed on a 1100 nano-HPLC system (Agilent) coupled to an LTQ-Orbitrap mass spectrometer (Thermo Fisher). Peptides (5 μ l) were eluted over 140-min water/acetonitrile gradients. The LTQ-Orbitrap was operated in the positive ion mode with the following acquisition parameters. A full scan recorded in the Orbitrap analyzer (resolution, 60,000) 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. Raw MS spectra were processed in Quant.exe, the first module of our in-house built software MaxQuant (Version 1.0.13.13) (36). The derived peak list was searched with the MASCOT search engine (Matrix Science, London, UK) (Version 2.1.0). The following search criteria were used: tryptic specificity was required; carbamidomethylation was set as a fixed modification; oxidation (Met), *N*-acetylation (protein), and phosphorylation (Ser/Thr/Tyr) were set as variable modifications. We performed the MASCOT search against the International

Protein Index (IPI) mouse database Version 3.37 containing 51,292 proteins to which 175 commonly observed contaminants and all the reversed sequences had been added (37). Maximally two missed cleavages were allowed. Initial mass tolerance for precursor and fragments ions was 7 ppm and 0.5 Thompson, respectively. All spectra and all sequence assignments made by MASCOT were imported into MaxQuant.

The derived peptides and their assigned proteins were further processed in Identify.exe, the second module of MaxQuant. The posterior error probability and false discovery rate (FDR) were used for statistical evaluation. The FDR is derived from the number of identifications from reversed protein sequences. All phosphopeptide identifications suggested by MASCOT were filtered in MaxQuant by applying thresholds on MASCOT score, peptide length, and mass error. We accepted peptides based on the criteria that the number of forward hits in the database was at least 100-fold higher than the number of reversed database hits (incorrect peptide sequences), which gives an estimated FDR of less than 1%. To achieve highly reliable identifications, the following criteria were used: maximal peptide posterior error probability of 0.1, maximal peptide FDR of 0.01, and minimal peptide length of 6.

In case the identified peptides of one protein included all peptides of another protein, these proteins (e.g. homologs and isoforms) were combined by MaxQuant and reported as one protein group. Phosphorylation sites were made non-redundant with regard to their surrounding sequence. The PTM score was used for assignment of the phosphorylation sites as described Ref. 18. Phosphosites fulfilling the following two criteria are defined as unambiguously identified sites: 1) their localization probability for the assignment is at least 0.75, and 2) the PTM score difference from the second possible localization assignment is 5 or higher. These cutoffs proved to yield highly accurate results in previous studies (18). The localization probability of such class I sites is almost always higher than 0.98. Phosphorylated peptides were uploaded to PHOSIDA (10). Finally, to further reduce the probability of including non-mitochondrial proteins in the data set, we retained only those proteins that were also defined as mitochondrial in the MitoCarta repository or in the gene ontology annotation.

Conservation Analysis—To derive homologous proteins between mouse and prokaryotic species, we used BLASTP (38) (supplemental Fig. 1). To check the phosphorylation site conservation, we generated global protein alignments between homologous protein pairs via Needle (39). To derive orthologs between eukaryotic species, we checked the phylogenetic relationships and conservation of triplets encoding phosphosites in comparison with triplets that encode non-phosphorylated serines or threonines localized on exposed loop structures of phosphorylated proteins throughout 37 eukaryotes on the basis of cDNA alignments as provided by Ensembl.

Our previous large scale phosphoproteomics study demonstrated that phosphorylation sites are predominantly located in non-regularly structured regions on the protein surface (10). Therefore, the surrounding sequence regions may diverge to such an extent that the structural effect (fast sequence evolution in loop regions) effectively competes with the constraining pressure of function (slow sequence evolution of kinase substrate motifs). To correctly assess the degree of conservation of phosphosites, it is therefore important to take the structural effect into account. We did this by choosing only sites located in loop regions according to SABLE (40) predictions for the comparison set, which should isolate the functional, evolutionary constraints on the phosphosite itself.

Bootstrapping—We created a bootstrap distribution for a given species from 10,000 sets of randomly selected human proteins annotated as “known” in Ensembl. Each random set contained as

many proteins as the given phosphoset. For each set, we calculated the proportion of orthologs in the chosen species, and the histogram of these 10,000 proportions provided the bootstrap distribution. The resulting histograms reflecting the distribution of logarithmized proportions of orthologs illustrate the significant difference between the set of interest (phosphoset) and randomly selected sets.

Next, we created a bootstrap distribution for a given species from 10,000 sets of serines, threonines, and tyrosines that were randomly selected from phosphorylated proteins that had an ortholog. Only those residues were included in the analysis that showed the same predicted structural constraints as phosphorylated residues (high accessibility and localization in loops). For each set, we calculated the proportion of conserved residues in the chosen species, and the histogram of these 10,000 proportions provided the bootstrap distribution. Resulting histograms were illustrated using MatLab (MathWorks).

RESULTS

The basis of our evolutionary studies is high accuracy phosphoproteomes that have been published during the last 3 years and that are deposited in PHOSIDA (10). The data were acquired using the technology described in detail in Olsen *et al.* (18). Briefly, cellular proteomes are digested to peptides, and phosphopeptides are enriched using a TiO₂ metal affinity matrix in the presence of 2,5-dihydrobenzoic acid (41) and measured by tandem mass spectrometry on a linear ion trap Orbitrap mass spectrometer using multistage activation (42). Phosphopeptide assignment including localization of the phosphogroups to particular amino acids is performed in MaxQuant (36). Mass accuracies are in the ppb range, and the false discovery rate of each data set is lower than 1%. Note that the high accuracy of detection of phosphorylation sites refers to the minimization of false positives. The percentage of false negatives, on the other hand, is not known in phosphoproteomics data sets. Site-specific localization assignments of the phosphogroup peptides from MS/MS fragmentation spectra were determined in MaxQuant, which contains an algorithm to predict phosphorylation site localization (18, 43). We used a threshold probability of 75%; however, median localization scores on phosphopeptides were typically about 98% (44). Eleven high quality phosphorylation sets from nine species that are available in PHOSIDA were used in our evolutionary analysis. Together they comprise 39,574 phosphorylation sites from *E. coli* (14), *B. subtilis* (13), *L. lactis* (15), *H. salinarum* (17), *Saccharomyces cerevisiae* (46), *Caenorhabditis elegans* (44), *Drosophila melanogaster* (47), *Mus musculus* (48, 49) and *Homo sapiens* (18, 50).

The generation of each phosphoproteome with the same generic work flow ensured that analysis on the phosphoproteome was performed to a comparable depth for each data set. Because no mitochondrial phosphorylation data set was available that was acquired in the same way, we performed a high accuracy large scale mass spectrometry study of the mitochondrial phosphoproteome as described below. This extends the compendium of evolutionary data for these phos-

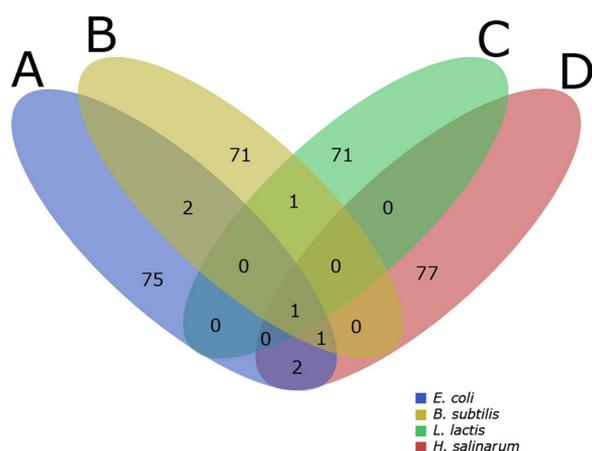


FIG. 1. **Overlap of prokaryotic phosphorylation sites.** The overlap between phosphorylation sites of *E. coli* (A), *B. subtilis* (B), *L. lactis* (C), and *H. salinarum* (D) is very low.

phosphoproteomes and in particular allowed us to analyze the mitochondrial phosphoproteome, the link between the domains of life, including its conservation.

Conservation of Prokaryote and Eukaryote Phosphoproteomes—The size of the phosphoproteomes of the four prokaryotes analyzed here are remarkably similar. They range from 73 sites for *L. lactis* to 81 sites for *E. coli* (www.phosida.com). Note that these numbers do not reflect the true size of the prokaryote phosphoproteome but rather the extent that can be readily probed with current technology. Nevertheless, the fact that the same technology results in about the same size phosphoproteome clearly indicates that the phosphoproteome is similarly sparse across different prokaryotic species.

Furthermore, the prokaryotic phosphoproteomes hardly overlap with each other. We found that a considerable proportion of proteins including elongation factors and glycolytic enzymes are phosphorylated in prokaryotes and highly conserved throughout the domain. However, these proteins, although conserved in prokaryotes, are not phosphorylated in all prokaryotic species, and if a given phosphoprotein occurs in more than one species, the phosphorylation sites do not overlap (Fig. 1).

In eukaryotes, phosphorylation regulates many key processes including cell growth, proliferation, differentiation, and immune response (51, 52), many of which do not occur in prokaryotes. Reflecting the fundamentally different biological roles of phosphorylation in eukaryotes and prokaryotes, the phosphoproteomes obtained with current technology are vastly larger. Using the technology described here they range from more than 3500 sites for yeast (46) to more than 20,000 sites for mammalian species such as human (23).

In previous bioinformatics analyses of the eukaryotic phosphoproteomes, we determined that phosphorylation sites, especially those on serine and threonine, tend to be located in fast evolving loop and hinge regions of proteins. We consid-

ered the PHOSIDA serine/threonine/tyrosine phosphoproteomes of mouse and other eukaryotes including human, fly, worm, and yeast, representing the most comprehensive evolutionary study on eukaryotic phosphoproteomes so far. The results on the protein level showed commonality between all eukaryotes. Within the eukaryotic domain, identified phosphoproteins have more orthologs than non-phosphorylated proteins (supplemental Table 1). This held true in a large variety of experimental systems throughout the eukaryotic domain. For each analyzed species, based on a bootstrap approach, the proportion of orthologous phosphoproteins to all phosphoproteins was clearly outside the distribution, and the resulting χ^2 values indicated extremely high statistical significance (see “Experimental Procedures”).

Analysis of phosphoproteome conservation at the level of individual sites is potentially much more informative than analysis of entire substrate proteins. This is because secondary effects such as protein abundance cannot skew comparison between sites on orthologous proteins. More importantly, individual phosphorylation sites are specific substrates of one or more kinases and phosphatases, and they mediate the functionality of this post-translational modification.

To study the conservation of phosphorylation that occurs throughout the entire mouse cell and other eukaryotic cells at the site level, we created bootstrap distributions by repeated random selections of non-phosphorylated serines, threonines, and tyrosines from proteins of each species from the Ensembl database. For these thousands of random sets, the proportion of conserved residues in the given species was derived, and the histogram of the proportions of conserved sites then provided the bootstrap distribution. The χ^2 test assigns a p value to the difference between the proportion of conserved and non-conserved phosphosites and the proportion of conserved and non-conserved counterparts (non-phosphorylated Ser/Thr/Tyr). In some cases, the conservation of phosphothreonines differs from the conservation of phosphoserines, probably because of the roughly 10-fold lower number of phosphothreonines in the data sets. For this reason, only the analysis on serine residues is discussed in more detail. For none of the investigated organisms was it possible to find a significant pattern regarding the conservation of highly accessible phosphotyrosines in alignments of orthologous proteins because of the low number of these residues present in the data sets.

We found that the proportion of conserved phosphorylation sites was clearly outside the bootstrap distribution, and the calculated χ^2 values were higher than 6, which corresponds to p values lower than 0.01, for all analyzed species except for yeast and worm (supplemental Table 2). The yeast conservation study showed that ~13% of phosphorylated serines were conserved in human orthologs, the same number as for non-phosphorylated serines. The same trend was observed in worm. In these two species, the p values was not statistically significant, which indicates that

phosphorylated and non-phosphorylated residues have the same degree of conservation.

Interestingly, in fly, which has roughly the same phylogenetic distance to mammals as worm, the phosphorylated serines were more conserved than non-phosphorylated serines throughout higher eukaryotes (e.g. 20% of fly phosphoserines and 16% of non-phosphorylated fly serines were conserved in human) (supplemental Fig. 2). The same pattern was observed for all mammalian phosphoproteomes. Note that conservation at the residue level can be very high between mammalian species in general. For example, 92% of human serines are conserved in proteins that are orthologous in chimpanzee and humans. However, even in this case, our study of human phosphosites (18) found that the conservation of phosphorylated residues is even higher; for example, 97% of human phosphoserines are conserved in chimpanzee alignments (supplemental Fig. 3). The different mouse and other mammalian phosphodata sets resulted in similar numbers.

Together, phosphorylation sites of fly, mouse, and human are more conserved than non-phosphosites of the same proteins with the same structural features throughout higher eukaryotes. In contrast, yeast and worm phosphorylation sites are not highly conserved with respect to higher eukaryotes. The bootstrapping approach and the χ^2 test verified the statistical significance of these observations. They agree with recent findings that many of the known human kinases evolved after the divergence between yeast and higher eukaryotes (54). Therefore, there are a considerable number of yeast-specific kinases that are not present in higher eukaryotes and vice versa. The worm kinome also differs significantly from kinomes of other eukaryotes. It is 2 times larger than the fly kinome, and half of the worm kinases are nematode-specific. This shows that the majority of phosphorylation events have evolved after the divergence of higher eukaryotes from yeast, which is in concordance with the unusually large number of nematode-specific kinases.

Characterization of Mitochondrial Phosphoproteome—To obtain a map of mitochondrial phosphorylated proteins across different cell types in basal conditions, we isolated and purified mitochondria from different mouse cell lines (C2C12, Hepa 1-6, 3T3-L1, and brown adipose tissue) and from mouse primary brown adipocytes. From these mitochondrial fractions, we enriched phosphorylated peptides using TiO_2 chromatography following the work flow of Olsen *et al.* (18). Peptides were analyzed via nano-LC-MS/MS on the high mass accuracy LTQ-Orbitrap. Raw data were processed with in-house developed MaxQuant software (36).

To further distinguish between mitochondrial proteins and the contaminants resulting from the biochemical preparation, mitochondrial localization of these proteins was manually verified against the MitoCarta database (55) and gene ontology localization description. In total, we identified and confidently assigned 174 phosphorylation sites on 74 mitochondrial pro-

teins with an estimated false positive rate of less than 1% for phosphopeptide identification and median phosphorylation localization probability greater than 99.9% (Table I and supplemental Table 3). The median MASCOT score and the median PTM score were 47.36 and 147.94, respectively. The median absolute mass deviation was 0.29 ppm. Further parameters for quality assurance in MaxQuant are described by Cox and Mann (36). The MitoCarta database comprises around 1100 mitochondrial proteins. Thus, the proportion of phosphorylated to all mitochondrial proteins is around 7%, which is similar to that of bacterial phosphoproteomes. In *E. coli*, for example, around 100 of 4000 (3%) proteins are known to be phosphorylated. In comparison, more than half of all proteins in eukaryotic cells are estimated to be phosphorylated.

Around 40% of the identified mitochondrial phosphosites were identified in at least two cell types (supplemental Fig. 4); this is a relatively high value given the comparatively low number of phosphorylation sites. The phosphorylation distribution on serines, threonines, and tyrosines was 89.7, 9.2, and 1.1%, respectively (Fig. 2A). This distribution is similar to the distributions we previously measured for eukaryotic and prokaryotic cells (18, 20). Ptptra (receptor-type tyrosine-protein phosphatase α precursor) and Pgrmc1 (receptor-type tyrosine-protein phosphatase α) were found to be phosphorylated on tyrosine residues. Overall, 80 of 174 phosphorylation sites are novel according to the Swiss-Prot database. For example, elongation factor Ts (Ser-269), pyruvate carboxylase (Ser-22, Ser-143, and Ser-1033), and long chain-specific acyl-CoA dehydrogenase (Ser-55) are not annotated to be phosphorylated on these sites in the Swiss-Prot database. Overall, 11 of 84 mitochondrial phosphorylation sites identified by Lee *et al.* (29) overlap with our set.

Gene ontology analysis shows that the mitochondrial phosphoproteome is enriched for proteins that are involved in protein binding, transporter activity, ATP binding, nucleotide binding, and ribonucleotide binding (Fig. 2B) compared with the entire mouse proteome. Supplemental Table 4 lists all overrepresented gene ontology categories along with the corresponding mitochondrial phosphoproteins based on the DAVID analysis tool (56). Furthermore, we checked the surrounding sequences of the identified mitochondrial phosphosites to derive the putative corresponding kinases. The phosphorylation sites match significantly with motifs of various kinases including casein kinase II, calcium/calmodulin-dependent protein kinase type II, and protein kinase D (supplemental Table 5). The wide spectrum of corresponding kinases is also reflected in the overall position-specific amino acid frequency as illustrated in Fig. 2C (57).

Evolutionary Conservation of Mitochondrial Phosphoproteome in Prokaryotes—Mitochondria and bacteria probably share an α -proteobacterial ancestor and are therefore evolutionarily related. Phylogenetic analyses suggest that the closest relative of the mitochondrial ancestor is *Rickettsia*, an

TABLE I
Mitochondrial mouse phosphorylation sites

IPI accession no.	UniProt accession no.	Gene symbol	Phosphosites
IPI00108147	P70335	<i>Rock1</i>	Ser-1102, Ser-1105
IPI00108685	P18052	<i>Ptpra</i>	Tyr-825
IPI00109033	Q3UJK3	<i>Fam54b</i>	Ser-100, Ser-235
IPI00111770	Q06185	<i>Atp5i</i>	Ser-68
IPI00112339	Q9ERG0	<i>Lima1</i>	Ser-367, Ser-372, Ser-374, Ser-488, Ser-580, Ser-581, Ser-615, Ser-617, Ser-620
IPI00113052	Q9CZR8	<i>Tsfn</i>	Ser-269
IPI00114209	P26443	<i>Glud1</i>	Ser-128, Thr-410
IPI00114710	Q05920	<i>Pc</i>	Ser-22, Ser-143, Ser-1033
IPI00116138	Q99KK9	<i>Hars2</i>	Ser-66
IPI00116414	Q99L88	<i>Sntb1</i>	Ser-86, Ser-218
IPI00117689	O54724	<i>Ptrf</i>	Ser-21, Ser-42, Ser-169, Ser-368, Ser-389, Ser-391
IPI00119114	P51174	<i>Acadl</i>	Ser-55
IPI00119320	Q91W06	<i>Tc1rg</i>	Ser-693
IPI00120715	Q9DCC8	<i>Tomm20</i>	Ser-135, Ser-138
IPI00121190	Q01279	<i>Egfr</i>	Thr-695
IPI00122547	Q60932	<i>Vdac2</i>	Ser-116
IPI00122549	Q60932	<i>Vdac1</i>	Ser-117, Thr-120
IPI00123183	Q02013	<i>Aqp1</i>	Ser-262
IPI00123410	A2APH4	<i>Usp24</i>	Ser-2044
IPI00123709	Q9WTQ5	<i>Akap12</i>	Ser-631
IPI00123891	P97315	<i>Csrp1</i>	Ser-192
IPI00126253	Q924Z4	<i>Lass2</i>	Ser-341, Ser-346, Ser-348, Ser-349
IPI00126634	Q91VA6	<i>Poldip2</i>	Thr-292
IPI00126939	Q8C064	<i>Prkcdpb</i>	Ser-165, Ser-166, Ser-188, Ser-194
IPI00128522	P14602	<i>Hspb1</i>	Ser-86, Ser-87
IPI00130280	Q03265	<i>Atp5a1</i>	Ser-53, Ser-65, Ser-76, Ser-451
IPI00130444	Q9Z1T1	<i>Ap3b1</i>	Ser-276
IPI00132076	O88587	<i>Comt</i>	Ser-260, Ser-261
IPI00132478	Q9CQF4		Ser-106, Ser-116
IPI00133440	P67778	<i>Phb</i>	Ser-151
IPI00133608	Q9CRD0	<i>Ociad</i>	Ser-108, Ser-147
IPI00135660	Q63918	<i>Sdpr</i>	Ser-203, Ser-204, Ser-218, Ser-359, Ser-363
IPI00137194	P53986	<i>Slc16a1</i>	Ser-210, Ser-213, Ser-461, Ser-462, Ser-477, Ser-482, Ser-491
IPI00138190	P55288	<i>Cdh11</i>	Ser-714, Ser-788
IPI00139795	P99027	<i>Rplp2</i>	Ser-79, Ser-86, Ser-102, Ser-105
IPI00153842	Q9JLR9	<i>Higd1a</i>	Ser-8
IPI00169862	Q8K1Z0	<i>Coq9</i>	Thr-78, Ser-81
IPI00222496	Q922R8	<i>Pdia6</i>	Ser-433
IPI00228826	P54310	<i>Lipe</i>	Ser-77, Ser-600, Ser-602, Ser-694
IPI00229548	P51912	<i>Slc1a5</i>	Ser-509, Ser-521
IPI00230283	Q8K0D5	<i>Gfm1</i>	Ser-92
IPI00272690	P09470	<i>Ace</i>	Ser-1305
IPI00308161	Q8BGV8	<i>Smcr7l</i>	Ser-55, Thr-58, Ser-59
IPI00308885	P63038	<i>Hspd1</i>	Ser-70
IPI00315135	Q9CPQ3	<i>Tomm22</i>	Ser-15, Ser-45
IPI00317684	Q8CCJ3-3	<i>Kiaa0776</i>	Ser-458
IPI00318935	Q9D8T7	<i>Slirp</i>	Thr-104, Ser-105
IPI00319973	O55022	<i>Pgrmc1</i>	Tyr-180, Ser-181
IPI00321499	P59017	<i>Bcl2l13</i>	Ser-387, Thr-389
IPI00331555	P50136	<i>Bckdha</i>	Ser-338, Thr-339, Ser-348
IPI00337893	P35486	<i>Pdha1</i>	Thr-231, Ser-232, Ser-293, Ser-295, Ser-300
IPI00344090	Q5ND52	<i>Rnmtl1</i>	Ser-42
IPI00352475	Q3UHX2	<i>Pdap1</i>	Ser-60, Ser-63
IPI00356904	Q3UY58	<i>Farp1</i>	Thr-24, Ser-373, Ser-892
IPI00378438	Q3U0Q9	<i>Tns1</i>	Ser-790, Ser-792, Thr-1343, Ser-1346, Ser-1468, Ser-1470, Ser-1535

TABLE I—continued

IPI accession no.	UniProt accession no.	Gene symbol	Phosphosites
IPI00379441	XP_485703	<i>EG433968</i>	Thr-123
IPI00387392	Q9CQS4	<i>Slc25a46</i>	Thr-45
IPI00387480	Q8CGN5	<i>Plin</i>	Ser-130, Ser-174, Ser-384, Ser-410
IPI00408119	P27546	<i>Map4</i>	Ser-475, Ser-506, Ser-517, Ser-901, Ser-1046
IPI00453589	O70404	<i>Vamp8</i>	Ser-5
IPI00453688	Q8C0T5-1	<i>Sipa111</i>	Ser-1528
IPI00459443	P58871	<i>Tnks1bp1</i>	Thr-533, Ser-807, Ser-974, Ser-1370, Ser-1375, Ser-1611, Ser-1612, Ser-1657
IPI00480432	Q6ZQ33	<i>Rab11fip5</i>	Ser-307
IPI00551411	Q8BX70	<i>Vps13c</i>	Ser-871, Ser-873
IPI00556700	Q2YDW1	<i>Eif3j</i>	Ser-14, Ser-16
IPI00620800	Q3UM62	<i>Fam38a</i>	Thr-632, Ser-634
IPI00623114	Q6ZQ33	<i>Fat1</i>	Ser-4483
IPI00649184	Q9WTI7	<i>Myo1c</i>	Ser-408
IPI00649326	Q9JMH9	<i>Myo18a</i>	Ser-1950, Ser-1978, Ser-1982, Ser-2049, Ser-2051
IPI00678532	Q3UJU9	<i>Fam82c</i>	Ser-44, Ser-46
IPI00751137	Q3TSX8	<i>Tomm70a</i>	Ser-94, Ser-99
IPI00754876	Q3UFQ5	<i>Ttc7b</i>	Ser-732, Ser-733
IPI00757909	Q9D0L7	<i>Armc10</i>	Ser-41, Ser-43
IPI00848443	Q5SWU9	<i>Acaca</i>	Ser-63, Ser-647

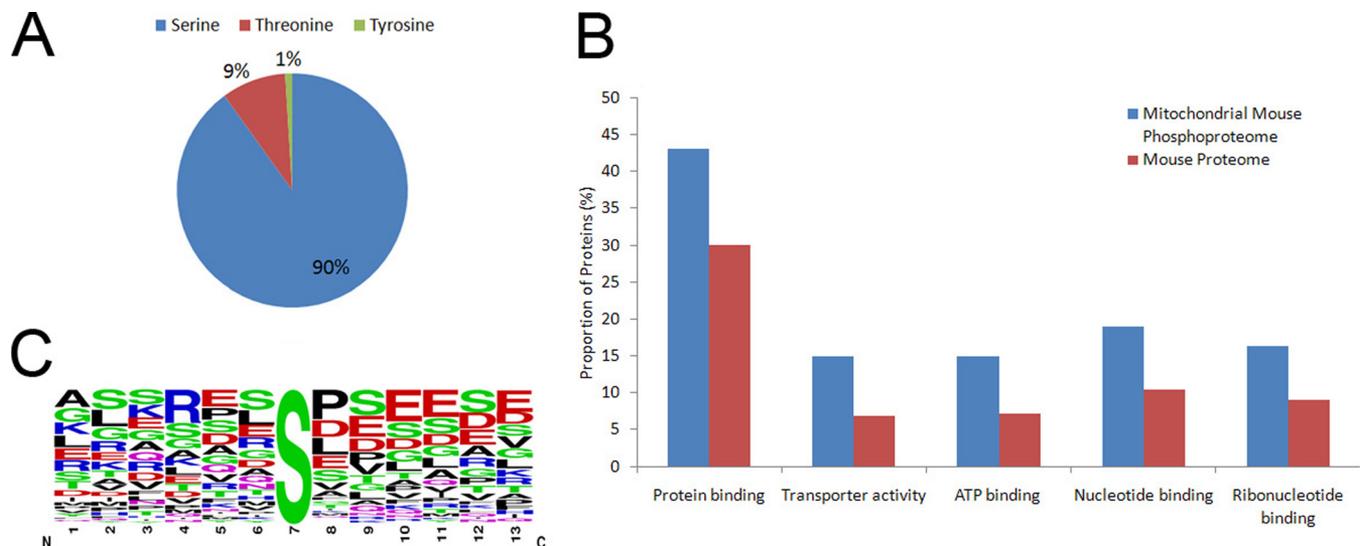


FIG. 2. Characterization of mitochondrial mouse phosphoproteome. A, serine/threonine/tyrosine distribution. B, molecular functions that are enriched in the mitochondrial phosphoproteome compared with the entire mouse proteome. C, relative position-specific amino acid frequency around phosphorylated sites.

obligate intracellular parasitic α -proteobacterium that could have initiated the endosymbiotic event (24). Our study provides independent evidence for the endosymbiotic hypothesis through the comparison of the conservation of mitochondrial versus non-mitochondrial proteins in prokaryotes (supplemental Fig. 5 and supplemental Table 6) (see “Experimental Procedures”). Mitochondrial mouse proteins are more highly conserved in prokaryotes than mouse proteins that are located in other organelles (Fig. 3). On average, around 35% of the mitochondrial proteins (MitoCarta data set) are conserved in any prokaryote in comparison with 13% of the non-mitochondrial proteins. The same pattern

was observed for phosphorylated proteins. Overall, 28% of the phosphorylated mitochondrial proteins (from our set as well as from the Lee *et al.* (29) data set) are conserved in a prokaryotic species compared with 20% of the phosphorylated non-mitochondrial proteins. These observations are in concordance with the prokaryotic character of mitochondria expected from the endosymbiotic hypothesis.

However, the mitochondrial and the prokaryotic phosphoproteomes of *E. coli* (14), *B. subtilis* (13), and *L. lactis* (15) hardly overlap. Elongation factor G and elongation factor Ts were found to be phosphorylated in mitochondria as well as in all investigated bacteria. However, phosphorylation events on

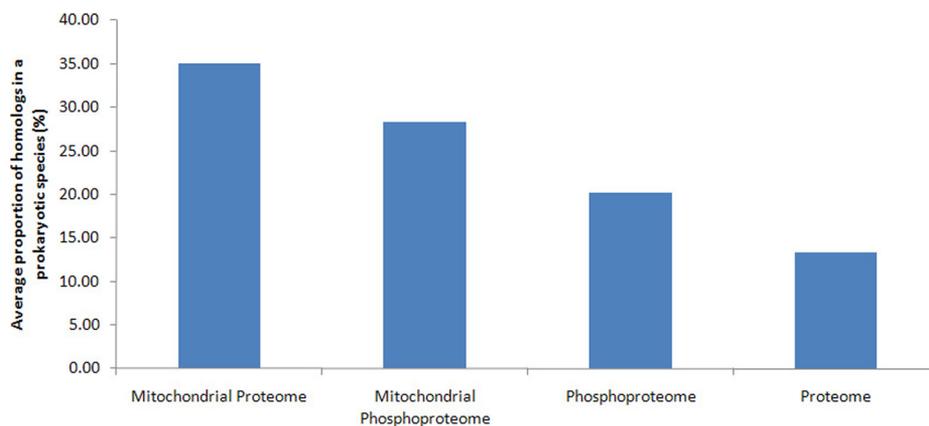
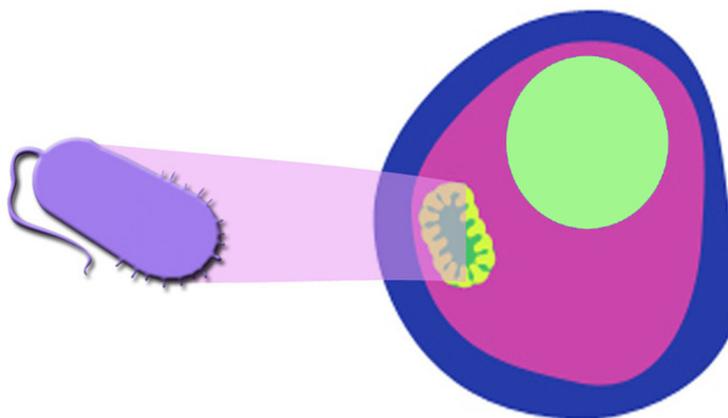


FIG. 3. Proportion of mouse proteins that are orthologous to prokaryotic proteins. Phosphorylated (our data set) and non-phosphorylated (MitoCarta data set) mitochondrial mouse proteins are more highly conserved in prokaryotes than phosphorylated (PHOSIDA data set) and non-phosphorylated (Swiss-Prot Database) proteins, which are located in other compartments of the cell. These observations are in concordance with the endosymbiotic scenario as illustrated in the lower panel. As a measure for conservation, we used the average proportion of orthologs in 62 prokaryotes.



these elongation factors were identified on different residues of the proteins. Other homologous proteins that were phosphorylated in mitochondria and in at least one of the selected bacteria were histidyl-tRNA synthetase, 60-kDa heat shock protein, and a “protein similar to nucleoside-diphosphate kinase.” This protein is the only one with a phosphorylation site (threonine 123) that was found to be phosphorylated in mitochondria as well as in bacteria (serine 123 in *B. subtilis*). These findings are in agreement with a model that regulation of mitochondrial proteins via phosphorylation evolved after the endosymbiotic event.

Evolutionary Conservation of Mitochondrial Phosphoproteome in Eukaryotic Domain—In the eukaryotic evolutionary study, we studied the conservation of the mitochondrial phosphoproteome in 36 eukaryotic species that are contained in the Ensembl database and that span the eukaryotic domain. Based on χ^2 statistics, mitochondrial mouse phosphoproteins derived from our study and the one by Lee *et al.* (29) have significantly more homologs in the eukaryotic domain than other mouse proteins (supplemental Table 7). Even for yeast, the species most evolutionarily distant from mouse, this was the case.

Previous studies have already shown that non-mitochondrial phosphoproteins of other eukaryotes are also significantly conserved in a defined set of a few species (10). To check in more detail whether the high proportion of or-

thologs is a common feature of different phosphoproteomes in more detail, we studied the conservation of non-mitochondrial proteins of different phosphoproteomes in 36 eukaryotic species.

To study the conservation of mitochondrial phosphorylation at the site level, we derived the conservation of phosphorylated serines/threonines and non-phosphorylated serines/threonines of the same mitochondrial phosphoproteins in orthologous proteins of different eukaryotic species that are annotated in the Ensembl database. Because the identification of eukaryotic phosphoproteomes is far from being complete, we restricted our analysis to amino acid conservation, which means that in most of the cases there is no experimental evidence for phosphorylation of sites that are phosphorylated in mouse and conserved in another species. We found that mitochondrial phosphosites tend to be more conserved than non-phosphorylated serines/threonines in other eukaryotes (supplemental Table 8). Although the high conservation of mitochondrial phosphosites is evident in all eukaryotic species, the significance decreases for lower eukaryotes, probably because of the low number of homologous proteins in species such as *Ciona intestinalis*.

DISCUSSION

Despite a large number of phosphoproteome studies of diverse organisms, our knowledge of each of them is not

comprehensive. The huge and continuing increase of measured mammalian phosphorylation sites makes it clear that the identification of the whole phosphoproteome is far from being complete. In fact, although previously one-third of all mammalian proteins were estimated to be phosphorylated, recent developments even suggest that the majority of all proteins are phosphorylated at least under some circumstances. As a practical matter, all evolutionary studies on phosphorylation are limited to the currently identified phosphoproteins. However, the number of known phosphosites turns out to very similar for different prokaryotic species and for mitochondria, whereas phosphoproteomes for eukaryotes range from over 3000 to more than 20,000 sites. This argues that large scale comparisons of phosphoproteomes can usefully be undertaken at this time. To this end, we took advantage of the recent availability of high resolution phosphorylation data and the PHOSIDA environment, which integrates biological context to quantify constraints of phosphorylation on a proteome-wide scale. We focused on phosphosets from PHOSIDA because it contains phosphoproteomes of a wide range of organisms all obtained with high accuracy and analyzed with the same stringent criteria. Other phosphorylation databases such as PhosphoSite and Phospho.ELM focus on mammalian phosphoproteomes only. Although they are more comprehensive than PHOSIDA, they also contain low resolution data sets with difficult to control overall false positive rates. More importantly, those data sets were obtained with widely diverging experimental work flows precluding comparative analysis. The data sets derived from PHOSIDA, although analyzed in the same way, are derived from different investigations. To exclude that this could cause any bias, we compared changing and non-changing subsets of the phosphoproteome, which showed that the results from our evolutionary study are valid for the whole phosphoproteome. For example, human phosphorylation sites that are unaffected by the stimulation of EGF show nearly the same conservation patterns as all other human phosphosites. Furthermore, our phosphoproteomes almost always contain the “basally phosphorylated peptide” for each phosphopeptide that is found to change upon a stimulus. This latter observation also argues that we have already covered a sizable proportion of the stimulus-specific phosphoproteome. We conclude that the general trends regarding the conservation of phosphorylated residues should already be contained in our analysis.

For all analyzed prokaryotes, the phosphoproteome is very sparse. Prokaryotic phosphorylation sites are hardly conserved, and only one site is identical in all four investigated prokaryotic phosphoproteomes. This finding is compatible with a model in which site-specific phosphorylation co-evolved with the adaptation of the bacterial species to their present-day ecological niches (15). The presence of phosphorylated prokaryotic proteins might also partially result from gene transfers as many eukaryote-like kinases have been found in prokaryotes by metagenomics projects (45). This

suggests that regulation via phosphorylation was a relatively recent development that occurred late in prokaryotic evolution. Future studies will likely lead to the identification of larger prokaryotic phosphorylation data sets, but it seems unlikely that these studies will result in numbers comparable with eukaryotic phosphoproteomes or to a large overlap between prokaryotic phosphoproteomes.

In contrast to the mitochondrial and prokaryotic phosphoproteomes, which comprise a few hundred phosphosites, the application of high accuracy mass spectrometry to eukaryotic cells yields the identification of thousands of phosphorylation sites. A combined analysis of the phosphoproteomes of human, mouse, and fly demonstrates that phosphorylated serines and threonines are more highly conserved than non-phosphorylated serines and threonines that are also localized in loops or turns on the protein surface. In contrast, yeast and worm phosphorylation sites are not highly conserved with respect to higher eukaryotes. These observations are in concordance with recent findings that many of the known human kinases evolved after the divergence between yeast and higher eukaryotes (54). Therefore, there are a considerable number of yeast-specific kinases that are not present in higher eukaryotes and vice versa. The worm kinome also differs significantly from kinomes of other eukaryotes. It is 2 times larger than the fly kinome, and half of the worm kinases are nematode-specific.

In this study, we were particularly interested in the evolution of phosphorylation in the mitochondrion, the organelle that presents the phylogenetic link between prokaryotes and eukaryotes according to the endosymbiotic theory. Phosphorylated peptides from mitochondria were enriched from different mouse cell types (myotubes, hepatocytes, and adipocytes). The experiments were performed in cell lines (3T3-L1, C2C12, Hepa 1-6, and brown preadipocytes) and tissue samples (brown adipose tissue). These cells play a pivotal role in essential body functions such as energy expenditure, management, and storage, which are associated with main mitochondrial processes (e.g. pyruvate/acetyl-CoA utilization, oxidative phosphorylation, and fatty acid oxidation). The phosphoproteome of mouse mitochondria represents a rich source of novel roles of protein phosphorylation in this organelle and forms the basis for comparison of mitochondria with prokaryotes and eukaryotes.

Mitochondria from the different cell types were enriched by biochemical approaches. Proteins identified as phosphorylated by MS were further filtered based on validated mitochondria repositories and annotations (MitoCarta and gene ontology). This was done because all biochemical methods have some limitations, making it impossible to purify any organellar fraction to complete homogeneity. Without this filtering, phosphorylation sites originating from relatively minor cytosolic contaminations of the mitochondrial fraction would otherwise dominate the phosphoset, especially when adding the phosphoproteome from several cell types.

The measured mitochondrial phosphoproteome consists of 174 phosphorylation sites on 74 mitochondrial proteins. Although mitochondria are part of eukaryotic cells, our analysis thus clearly shows that in terms of the size of their phosphoproteome they are similar to prokaryotes. The phosphoproteome is enriched for essential proteins that are involved in binding and transport. The comparatively sparse mitochondrial phosphoproteome underlines the difficulty to identify phosphorylation sites in this organelle. Thus, the overlaps between different identified mitochondrial phosphoproteomes are not expected to be high. In fact, only 13% of our phosphosites overlap with data from a previous mitochondrial study by Lee *et al.* (29). In comparison, the overlap between our data sets derived from different cell types is relatively high (around 40%). This motivated us to confirm our conclusions with other mitochondrial data sets. We found that mitochondrial proteins are indeed relatively more conserved in prokaryotes than the non-mitochondrial eukaryotic proteome (Fig. 3). This is consistent with the endosymbiotic theory. However, the overlap between mitochondrial and bacterial phosphoproteomes is very low, which is in agreement with a model in which regulation of mitochondrial proteins via phosphorylation may have evolved after the endosymbiotic event. This finding is also in concordance with the fact that prokaryotic phosphoproteomes themselves have a very low overlap and with the idea that regulation by phosphorylation is a relatively recent introduction during evolution.

Mitochondrial phosphoproteins also proved to have significantly more orthologs in the eukaryotic domain. This was expected because of their essential roles in the eukaryotic cell. This pattern has also been reported for other phosphoproteomes.

Within the eukaryotic domain, we found that mitochondrial as well as non-mitochondrial phosphoproteins have more orthologs than non-phosphorylated proteins. This held true in a large variety of experimental systems; for example, for the analysis of mammalian phosphorylation conservation, we used data obtained from measurements of two different human cell lines as well as a mouse cell line and a mouse tissue. Our analysis of the global alignments of orthologs in 37 eukaryotes shows that mitochondrial mouse phosphorylation sites are more conserved than non-phosphosites of the same proteins with the same structural features throughout higher eukaryotes. The significance of this observation decreases with more distantly related lower eukaryotes.

The overall picture that emerges from our evolutionary analysis of serine/threonine/tyrosine phosphorylation contains several major discontinuities. One is between bacteria and eukaryotes: bacteria have only a few percent of the phosphorylation events that eukaryotes have. In support of the endosymbiotic hypothesis of mitochondrial origin, we found this organelle to be sparsely phosphorylated, just like bacteria. The second discontinuity of the phosphoproteome is between yeast and other eukaryotes. Here our phosphoproteomics

analysis parallels the notion that much of the signaling function of phosphorylation involves cell-cell communication. This is evident in the high conservation of phosphoproteins and phosphosites within metazoans but low conservation in a single cell organism. Furthermore, the worm phosphoproteome is poorly conserved throughout the eukaryotic domain, which is in concordance with its distinct kinome evolution and overrepresentation of phosphorylation events on substrates that are involved in the sex determination and development of *C. elegans* (44).

Taken together, our analysis establishes the phylogenetic study of the mitochondrial and non-mitochondrial phosphoproteome as a fruitful adjunct to the evolutionary study of the kinase-encoding genes. The recent increase in the number of studies on the evolution of phosphorylation and its regulation shows the current interest in this field. For example, a very recent comparison of phosphorylation patterns across yeast species showed that kinase-substrate interactions change 2 orders of magnitude more slowly than transcription factor-promoter interactions and that protein kinases are important for the phenotypic diversity (53).

In the future, the ever increasing quality and depth of proteomics studies will help to identify more complete phosphoproteomes and will also provide important quantitative information, for example on phosphorylation site stoichiometry. Combined with functional studies at the site level, which is still a major bottleneck, as well as more in-depth knowledge of kinase-substrate relationships, this will allow the derivation of a more detailed picture of the evolution of phosphorylation and its role in the cell. Nevertheless, the major features of the evolution of the kinome (54) and the phosphoproteome are already clear.

Acknowledgments—We thank Johannes Graumann, Juergen Cox, Albert Vilella Bertran, and Sudhakaran Prabhakaran for fruitful discussions.

* This work was supported, in whole or in part, by National Institutes of Health Grant R01 GM081578-02 on “Complex dynamics in multisite phosphorylation.” This work was also supported by the Max Planck Society and the HepatoSys Network.

§ This article contains [supplemental Figs. 1–5 and Tables 1–8](#).

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

REFERENCES

- Hunter, T. (2000) Signaling—2000 and beyond. *Cell* **100**, 113–127
- Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
- Salomon, A. R., Ficarro, S. B., Brill, L. M., Brinker, A., Phung, Q. T., Ericson, C., Sauer, K., Brock, A., Horn, D. M., Schultz, P. G., and Peters, E. C. (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 443–448
- Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., Down, T., Eyra, E., Fernandez-Suarez, X. M., Gane, P., Gibbins, B., Gilbert, J., Hammond, M., Hotz, H. R., Iyer, V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Keenan, S., Lehvaslaiho, H., McVicker, G., Melsopp, C., Meidl, P., Mongin, E.,

- Pettett, R., Potter, S., Proctor, G., Rae, M., Searle, S., Slater, G., Smedley, D., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Storey, R., Ureta-Vidal, A., Woodwark, K. C., Cameron, G., Durbin, R., Cox, A., Hubbard, T., and Clamp, M. (2004) An overview of Ensembl. *Genome Res.* **14**, 925–928
5. Nita-Lazar, A., Saito-Benz, H., and White, F. M. (2008) Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics* **8**, 4433–4443
6. Mann, M., and Kelleher, N. L. (2008) Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18132–18138
7. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12130–12135
8. Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E., and Zhang, B. (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561
9. Diella, F., Cameron, S., Gemünd, C., Linding, R., Via, A., Kuster, B., Sicheritz-Pontén, T., Blom, N., and Gibson, T. J. (2004) PhosphoELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5**, 79
10. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroschi, M., and Mann, M. (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8**, R250
11. Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M. A., Jørgensen, C., Miron, I. M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin, J., Park, J. G., Samson, L. D., Woodgett, J. R., Russell, R. B., Bork, P., Yaffe, M. B., and Pawson, T. (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426
12. Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M. B., and Pawson, T. (2008) NetworkKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699
13. Macek, B., Mijakovic, I., Olsen, J. V., Gnad, F., Kumar, C., Jensen, P. R., and Mann, M. (2007) The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol. Cell. Proteomics* **6**, 697–707
14. Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., and Mann, M. (2008) Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* **7**, 299–307
15. Soufi, B., Gnad, F., Jensen, P. R., Petranovic, D., Mann, M., Mijakovic, I., and Macek, B. (2008) The Ser/Thr/Tyr phosphoproteome of *Lactococcus lactis* IL1403 reveals multiply phosphorylated proteins. *Proteomics* **8**, 3486–3493
16. Eymann, C., Becher, D., Bernhardt, J., Gronau, K., Klutzny, A., and Hecker, M. (2007) Dynamics of protein phosphorylation on Ser/Thr/Tyr in *Bacillus subtilis*. *Proteomics* **7**, 3509–3526
17. Aivaliotis, M., Macek, B., Gnad, F., Reichelt, P., Mann, M., and Oesterhelt, D. (2009) Ser/Thr/Tyr protein phosphorylation in the archaeon *Halobacterium salinarum*—a representative of the third domain of life. *PLoS One* **4**, e4777
18. Olsen, J. V., Blagoev, B., Gnad, F., 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
19. Villén, J., Beausoleil, S. A., Gerber, S. A., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1488–1493
20. Macek, B., Mann, M., and Olsen, J. V. (2009) Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu. Rev. Pharmacol. Toxicol.* **49**, 199–221
21. Bodenmiller, B., Malmstrom, J., Gerrits, B., Campbell, D., Lam, H., Schmidt, A., Rinner, O., Mueller, L. N., Shannon, P. T., Pedrioli, P. G., Panse, C., Lee, H. K., Schlapbach, R., and Aebersold, R. (2007) PhosphoPep—a phosphoproteome resource for systems biology research in *Drosophila Kc167* cells. *Mol. Syst. Biol.* **3**, 139
22. Cohen, P. (2001) The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur. J. Biochem.* **268**, 5001–5010
23. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3
24. Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C., Podowski, R. M., Näslund, A. K., Eriksson, A. S., Winkler, H. H., and Kurland, C. G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140
25. Neupert, W., and Herrmann, J. M. (2007) Translocation of proteins into mitochondria. *Annu. Rev. Biochem.* **76**, 723–749
26. Thomson, M. (2002) Evidence of undiscovered cell regulatory mechanisms: phosphoproteins and protein kinases in mitochondria. *Cell. Mol. Life Sci.* **59**, 213–219
27. Lewandrowski, U., Sickmann, A., Cesaro, L., Brunati, A. M., Toninello, A., and Salvi, M. (2008) Identification of new tyrosine phosphorylated proteins in rat brain mitochondria. *FEBS Lett.* **582**, 1104–1110
28. Hopper, R. K., Carroll, S., Aponte, A. M., Johnson, D. T., French, S., Shen, R. F., Witzmann, F. A., Harris, R. A., and Balaban, R. S. (2006) Mitochondrial matrix phosphoproteome: effect of extra mitochondrial calcium. *Biochemistry* **45**, 2524–2536
29. Lee, J., Xu, Y., Chen, Y., Sprung, R., Kim, S. C., Xie, S., and Zhao, Y. (2007) Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. *Mol. Cell. Proteomics* **6**, 669–676
30. Deng, N., Zhang, J., Zong, C., Wang, Y., Lu, H., Yang, P., Wang, W., Young, G. W., Wang, Y., Korge, P., Lotz, C., Doran, P., Liem, D. A., Apweiler, R., Weiss, J. N., Duan, H., and Ping, P. (September 7, 2010) Phosphoproteome analysis reveals regulatory sites in major pathways of cardiac mitochondria. *Mol. Cell. Proteomics* **10**, 1074/mcp.M110.000117
31. Reinders, J., Wagner, K., Zahedi, R. P., Stojanovski, D., Eylich, B., van der Laan, M., Rehling, P., Sickmann, A., Pfanner, N., and Meisinger, C. (2007) Profiling phosphoproteins of yeast mitochondria reveals a role of phosphorylation in assembly of the ATP synthase. *Mol. Cell. Proteomics* **6**, 1896–1906
32. Kratchmarova, I., Kalume, D. E., Blagoev, B., Scherer, P. E., Podtelejnikov, A. V., Molina, H., Bickel, P. E., Andersen, J. S., Fernandez, M. M., Bunkenborg, J., Roepstorff, P., Kristiansen, K., Lodish, H. F., Mann, M., and Pandey, A. (2002) A proteomic approach for identification of secreted proteins during the differentiation of 3T3-L1 preadipocytes to adipocytes. *Mol. Cell. Proteomics* **1**, 213–222
33. Fasshauer, M., Klein, J., Kriauciunas, K. M., Ueki, K., Benito, M., and Kahn, C. R. (2001) Essential role of insulin receptor substrate 1 in differentiation of brown adipocytes. *Mol. Cell. Biol.* **21**, 319–329
34. Forner, F., Foster, L. J., Campanaro, S., Valle, G., and Mann, M. (2006) Quantitative proteomic comparison of rat mitochondria from muscle, heart, and liver. *Mol. Cell. Proteomics* **5**, 608–619
35. Forner, F., Arriaga, E. A., and Mann, M. (2006) Mild protease treatment as a small-scale biochemical method for mitochondria purification and proteomic mapping of cytoplasm-exposed mitochondrial proteins. *J. Proteome Res.* **5**, 3277–3287
36. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
37. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2**, 667–675
38. 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
39. 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
40. Wagner, M., Adamczak, R., Porollo, A., and Meller, J. (2005) Linear regression models for solvent accessibility prediction in proteins. *J. Comput. Biol.* **12**, 355–369
41. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jørgensen, T. J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **4**, 873–886
42. Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., and Coon, J. J. (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal. Chem.* **76**, 3590–3598

43. Olsen, J. V., and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13417–13422
44. Zielinska, D. F., Gnad, F., Jedrusik-Bode, M., Wiñeniewski, J. R., and Mann, M. (2009) *C. elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. *J. Proteome Res.* **8**, 4039–4049
45. Yoosseph, 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., Cieplak, P., Miller, C. S., Li, H., Mashiyama, S. T., Joachimiak, M. P., van Belle, C., Chandonia, J. M., Soergel, D. A., Zhai, Y., Natarajan, K., Lee, S., Raphael, B. J., Bafna, V., Friedman, R., Brenner, S. E., Godzik, A., Eisenberg, D., Dixon, J. E., Taylor, S. S., Strausberg, R. L., Frazier, M., and Venter, J. C. (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16
46. Gnad, F., de Godoy, L. M., Cox, J., Neuhauser, N., Ren, S., Olsen, J. V., and Mann, M. (2009) High accuracy identification and bioinformatic analysis of in-vivo protein phosphorylation sites in yeast. *Proteomics* **9**, 4642–4652
47. Hilger, M., Bonaldi, T., Gnad, F., and Mann, M. (2009) Systems-wide analysis of a phosphatase knock down by quantitative proteomics and phosphoproteomics. *Mol. Cell. Proteomics* **8**, 1908–1920
48. Pan, C., Gnad, F., Olsen, J. V., and Mann, M. (2008) Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphatase inhibitors. *Proteomics* **8**, 4534–4546
49. Zanivan, S., Gnad, F., Wickström, S. A., Geiger, T., Macek, B., Cox, J., Fässler, R., and Mann, M. (2008) Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J. Proteome Res.* **7**, 5314–5326
50. Daub, H., Olsen, J. V., Bairlein, M., Gnad, F., Oppermann, F. S., Körner, R., Greff, Z., Kéri, G., Stemmann, O., and Mann, M. (2008) Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol. Cell* **31**, 438–448
51. Schlessinger, J. (2000) Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211–225
52. Pawson, T., and Scott, J. D. (2005) Protein phosphorylation in signaling—50 years and counting. *Trends Biochem. Sci.* **30**, 286–290
53. Beltrao, P., Trinidad, J. C., Fiedler, D., Roguev, A., Lim, W. A., Shokat, K. M., Burlingame, A. L., and Krogan, N. J. (2009) Evolution of phosphorylation: comparison of phosphorylation patterns across yeast species. *PLoS Biol.* **7**, e1000134
54. Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520
55. Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A., and Mootha, V. K. (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123
56. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57
57. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190