Phosphoproteomic Analysis of Protein Phosphorylation Networks in Tetrahymena thermophila, a Model Single-celled Organism**§

Miao Tian‡‡, Xiulan Chen¶¶, Qian Xiong¶¶, Jie Xiong‡, Chuanle Xiao**, Feng Ge‡‡‡, Fuquan Yang‡‡, and Wei Miao‡‡‡

Tetrahymena thermophila is a widely used unicellular eukaryotic model organism in biological research and contains more than 1000 protein kinases and phosphatases with specificity for Ser/Thr/Tyr residues. However, only a few dozen phosphorylation sites in T. thermophila are known, presenting a major obstacle to further understanding of the regulatory roles of reversible phosphorylation in this organism. In this study, we used high-accuracy mass-spectrometry-based proteomics to conduct global and site-specific phosphoproteome profiling of T. thermophila. In total, 1384 phosphopeptides and 2238 phosphorylation sites from 1008 T. thermophila proteins were identified through the combined use of peptide pre-fractionation, TiO2 enrichment, and two-dimensional LC-MS/MS analysis. The identified phosphoproteins are implicated in the regulation of various biological processes such as transport, gene expression, and mRNA metabolic process. Moreover, integrated analysis of the T. thermophila phosphoproteome and gene network revealed the potential biological functions of many previously unannotated proteins and predicted some putative kinase–substrate pairs. Our data provide the first global survey of phosphorylation in T. thermophila using a phosphoproteomic approach and suggest a wide-ranging regulatory scope of this modification. The provided dataset is a valuable resource for the future understanding of signaling pathways in this important model organism. Molecular & Cellular Proteomics 13: 10.1074/mcp.M112.026575, 503–519, 2014.

Tetrahymena thermophila is a species of free-living unicellular ciliated protozoa that is widely distributed in freshwater environments around the world (1). Since the onset of molecular biology, T. thermophila has been one of the most widely used eukaryotic model organisms (1). A long tradition of fundamental research on T. thermophila has contributed to numerous scientific breakthroughs such as the discovery of catalytic RNA (2), telomere structure (3), and telomerase activity (4). Genome sequencing and analysis of T. thermophila have revealed that it shares more orthologs with Homo sapiens and has many core biological processes that are broadly conserved across eukaryotes (including human) and are not found in the yeast Saccharomyces cerevisiae (5). Notably, 1069 protein kinases are predicted in the T. thermophila genome, corresponding to nearly 3.8% of the predicted proteome. The kinome size of T. thermophila is much larger than that of other model organisms such as S. cerevisiae, Drosophila melanogaster, and Mus musculus (5). Even compared with the kinome of H. sapiens, containing 518 protein kinases, the kinome of T. thermophila is significantly extended in several kinase classes such as the mitotic kinase families and multiple kinases that interact with the microtubule network (5).

Knowledge of the substrates of each of the T. thermophila kinases is essential for understanding their function. However, only a few dozen T. thermophila phosphoproteins have been identified to date, and little progress has been made in ascertaining precisely which sites are phosphorylated, which kinases/phosphatases are involved, and what cellular processes are targeted by this post-translational modification (6, 7). In eukaryotes, reversible protein phosphorylation at serine (S), threonine (T), and tyrosine (Y) residues is a key post-translational modification that plays vital roles in cell signaling and general regulation of protein activity (8, 9). Phosphorylation is catalyzed in a reversible fashion by specific protein kinases and phosphatases. In this process, protein kinases modify other proteins by chemically adding phosphate groups to amino acid residues of the proteins, and protein phosphatases remove the phosphate groups that have been attached by protein kinases (8, 9). The importance of S/T/Y kinases and phosphatases for cell physiology has been widely documented in eukaryotes ranging from yeast (10) to human (11), where 2% to 3% of the open reading frames in their genomes encode known or potential protein kinases and protein phosphatases (12). In recent years, the development of phosphopeptide enrichment strategies such as immobilized...
Phosphoproteomic Analysis of Tetrahymena thermophila

metal ion affinity chromatography, metal oxide affinity chromatography, and strong cation exchange chromatography (13, 14) in combination with mass spectrometry (MS)-based phosphoproteomic techniques has provided new opportunities to investigate protein phosphorylation globally. Large-scale phosphoproteome studies for yeast (15), worm (16), fly (17), mouse (18), and human (19) have been reported and provide novel insights into the extent and function of this post-translational modification (20–24).

To obtain a comprehensive understanding of in vivo phosphorylation events in T. thermophila, we enzymatically digested the entire T. thermophila proteome, enriched the phosphopeptides by using titanium dioxide (TiO₂), and analyzed the phosphopeptides with hybrid LTQ-Orbitrap MS. In total, 1384 phosphopeptides and 2238 phosphorylation sites from T. thermophila proteins were identified. These results provide the first and most extensive data on S/T/Y phosphorylation currently available on T. thermophila and novel insights into the range of functions regulated by S/T/Y phosphorylation in T. thermophila. Recently, we reported a comprehensive transcriptome analysis of T. thermophila, with the aim of refining its genome annotation (25, 26). In this study, integrated analysis of the T. thermophila phosphoproteome and transcriptome revealed the potential biological functions of many previously unannotated proteins. In addition to providing insights into S/T/Y phosphorylation in T. thermophila, these data will serve as an important resource for further investigation of these signal transduction pathways in T. thermophila, and in eukaryotes more broadly.

EXPERIMENTAL PROCEDURES

T. thermophila Cultures—Wild-type B2086 (II) and CU428 (VII) strains of T. thermophila were both obtained from the Tetrahymena Stock Center at Cornell University and maintained in Super Proteose Peptone medium (1% Proteose Peptone, 0.2% glucose, 0.1% yeast extract, 0.003% Sequestrene) (27) on a rotary shaker at 135 rpm at 30 °C. Both B2086 and CU428 have the same genetic background as strain SB210, whose macronuclear genome has been sequenced (5). Cells were grown under several conditions and harvested at different growth stages according to published protocols (28). A total of seven culture extracts in three different physiological/developmental stages were obtained: (1) vegetative growth—CU428 cells in logarithmic growth phase (~2 × 10⁶ cells/ml) and in stationary phase (~1 × 10⁶ cells/ml) were both harvested; (2) starvation—CU428 cells were starved for 3 and 15 h at ~2 × 10⁶ cells/ml in 10 mM Tris-HCl (pH 7.5); and (3) conjugation—equal volumes of CU428 and B2086 cells at ~2 × 10⁵ cells/ml were starved in 10 mM Tris (pH 7.5) for 18 h, mixed, and harvested after 0, 4, and 10 h. A Beckman-Coulter counter was used for cell counting.

Protein Extraction and In-solution Tryptic Digestion—Protein extraction and tryptic digestion were performed as described by Hou et al. (29), with some modifications. Briefly, after the seven cell cultures were harvested, equal numbers of cells were mixed and washed three times with cold PBS buffer. Cell pellets were resuspended in lysis buffer containing 8 M urea, 0.2 M sucrose, 10 mM Hepes (pH 7.4), 5 mM β-glycerophosphoric acid, 10 mM NaF, 1 mM Na₂VO₄, 10 mM Na₃P₂O₇, 1 × protease inhibitor mixture, and 1 × phosphatase inhibitor mixture (Thermo Fisher Scientific, Waltham, MA) and sonicated on ice for 20 min. All other chemicals, unless otherwise stated, were obtained from Sigma Chemical Co. (St. Louis, MO). Cellular debris was removed via centrifugation at 16,100 g for 30 min at 4 °C and the supernatant was collected. The protein concentration was determined via the Bradford assay. Protein extracts (3 mg) were subjected to disulfide reduction with 10 mM dithiothreitol (DTT) at 56 °C for 60 min and alkylated with 40 mM iodoacetamide at 25 °C for 45 min in the dark. The alkylation reaction was stopped by the addition of 40 mM DTT at 25 °C for 15 min. Protein digestion was carried out by adding endoproteinase Lys-C (Wako Biochemicals, Osaka, Japan) at a ratio of 1:100 (w/w, enzyme/total protein) and digesting at 37 °C for 6 h. Before further digestion with trypsin, the urea concentration in the sample was diluted to less than 1.6 M with 25 mM NH₄HCO₃. The diluted sample was digested with sequencing-grade modified trypsin (Promega, Madison, WI) at a ratio of 1:100 (w/w, enzyme/total protein) and incubated overnight at 37 °C. Tryptic digestion was quenched with 0.1% trifluoroacetic acid (TFA). The supernatant (peptides) was collected via centrifugation at 16,100 g for 10 min and divided into three aliquots, which were treated in parallel in subsequent experiments.

Peptide Fractionation and Phosphopeptide Enrichment—Peptides were fractionated and desalted using a self-packed reverse-phase C₁₈ solid-phase extraction column as described elsewhere (29), with slight modifications. Briefly, 50 mg of C₁₈ bead (40 to 60 μm, 120-Å pore size, SunChrom, Friedrichsdorf, Germany) slurry, suspended in 100% methanol, was loaded onto a 1.5-ml AGT Cleanert SPE column (~0.3-ml bed volume), washed with 5 ml of 100% acetonitrile (ACN), and equilibrated with 5 ml of 0.1% TFA. After that, the peptide sample was sequentially loaded onto three equilibrated reverse-phase C₁₈ solid-phase extraction columns. Peptides were eluted with a series of elution buffers (1.0 ml containing 0.1% TFA/50% ACN, in which the percentage of ACN (X) was 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, or 100%). The peptides containing flow-through that eluted with a given eluant were combined and lyophilized in a SpeedVac (Thermo Savant, Milford, MA). A total of 10 fractions were collected for further phosphopeptide enrichment. The phosphopeptide enrichment was performed as described elsewhere (29), with some modifications. Briefly, the lyophilized peptide fractions were dissolved in 200 μl of binding buffer (65% ACN, 2% TFA saturated with glutamic acid). Titanium dioxide (TiO₂) beads (Titansphere, 5 μm, GL Sciences, Torrance, CA, USA) were acidified with binding buffer to obtain a TiO₂ slurry at a final concentration of 10 mg/ml. 40 μl of TiO₂ slurry was added into each dissolved peptide fraction and shaken vigorously at room temperature for 60 min. The supernatant was discarded after brief centrifugation. The TiO₂ beads were then washed sequentially with 800 μl of binding buffer, washing buffer I (65% ACN/0.5% TFA), and washing buffer II (50% ACN/0.1% TFA). Finally, the phosphopeptides bound to TiO₂ beads were eluted with 200 μl of elution buffer I (300 mM NH₄OH/50% ACN) once and then eluted twice with 200 μl elution buffer II (500 mM NH₄OH/60% ACN). The flow-through fractions were combined and lyophilized to ~50 μl and acidified in 1% formic acid for further two-dimensional LC-MS/MS analysis.

Two-dimensional Nano-LC-MS/MS Analysis—Phosphopeptide samples were dissolved in 20 μl of 0.1% formic acid and pressure-loaded onto a 200-μm inner diameter biphasic silica capillary column with a 3-cm bed volume of reverse-phase C₁₈ resin (3-μm particles,

---

1 The abbreviations used are: TFA, trifluoroacetic acid; ACN, acetonitrile; PTM, post-translational modification; GO, Gene Ontology; CK, casin kinase; PIKK, phosphatidylinositol 3-kinase-related kinase; DNA-PK, DNA-dependent protein kinase; TGN, Tetrahymena Gene Network; MAPK, mitogen-activated protein kinase; ULK, uncoordinated-51-like kinase.
Phosphoproteomic Analysis of *Tetrahymena thermophila*

120-Å pore size, SunChrom, Germany) and a 3-cm bed volume of strong cation exchange resin (Luna, 5-μm particles, 100-Å pore size, Phenomenex, Torrance, CA). After loading, the peptides were analyzed on an Eksigent HPLC system (Dublin, CA) coupled with a linear ion trap LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as described by Hou et al. (29). The flow-rate used for two-dimensional LC was 300 nl/min. Nano-electrospray ionization was accomplished with a spray voltage of 2.1 kV and a heated capillary temperature of 230 °C. A cycle of one full-scan mass spectrum (350–1700 m/z) followed by seven data-dependent MS/MS spectra was repeated continuously throughout each salt step of the multidimensional separation. All MS/MS spectra were obtained using the following parameters: normalized collision energy, 35%; activation Q, 0.25; and activation time, 30 ms. Selected ions were dynamically excluded for 30 s. To improve the fragmentation of phosphopeptides, the multistage activation algorithm in the Xcalibur software was enabled for each MS/MS spectra using the neutral loss values of 97.97, 48.99, and 32.66 m/z units. The application of MS scan functions and HPLC solvent gradients was controlled by an XCalibur data system (Thermo Fisher Scientific, Waltham, MA).

**Phosphopptide Identification and Phosphosite Validation**—Peak lists for the database search were produced in the Mascot generic format using DTASuperCharge (version 2.0.07, SourceForge) with default parameters, and the derived peak lists were searched using the Mascot 2.2.04 search engine (Matrix Science, London, UK) against a real and false *T. thermophila* database including 24,725 protein entries. The total number of entries of real and false *T. thermophila* proteins was 49,450. The following search criteria were employed: full tryptic specificity was required; two missed cleavages were allowed; carboxymethylation was set as fixed modification; and oxidation of methionine, phosphorylation of S and T, and phosphorylation of Y were considered as variable modifications. Precursor ion mass tolerances were 10 ppm for all MS spectra acquired in the Orbitrap mass analyzer, and the fragment ion mass tolerance was 0.5 Da for all MS² spectra acquired in the LTQ. Mass spectra of identified phosphopeptides with an E value of 0.05 and a peptide score of >5 were further processed and validated with the MSQuant 2.0 software for post-translational modification (PTM) score analysis (19). The following filter criteria were used: The total threshold for the PTM score and peptide score was 57. The estimated false positive rate based on the decay database search was 1.32%. All fragmentation spectra of putative phosphopeptides identified were verified by means of manual inspection for the presence of at least three successive b- or y-ions. Simultaneously, the singly charged peptides containing Ser (P) and Thr (P) had to have at least one lost phosphoric acid (~98 Da). All fragmentation spectra of these potential phosphopeptides were manually verified using a method described by Mann and colleagues (20, 21). For phosphopeptides with multiple potential phosphorylation sites, the probability of phosphorylation at each site was calculated from the PTM scores as described by Olsen et al. (19). Phosphorylation sites that were occupied with a probability of >0.75 were reported as class I phosphorylation sites. For class II sites, the localization probability was between 0.75 and 0.25. Phosphorylation sites with a localization probability of <0.25 were discarded. All raw data have been deposited in the publicly accessible database PeptideAtlas and can be accessed with the identifier PASS001118.

**Analysis of Amino Acid Residues Flanking the Phosphorylation Sites**—To investigate the amino acid frequencies around each serine and threonine phosphorylation site identified, the 12 surrounding amino acids (6 amino acids N- and C-terminal of the phosphorylated residue) were retrieved to generate a list of “phospho-13-mers” using a custom Perl script. Phosphotyrosines were omitted because of the low number of sites identified. For *Plasmodium falciparum* and *H. sapiens*, the phospho-13-mer datasets were obtained from *P. falciparum* phosphoproteome results (30) and the PHOSIDA database (31), respectively. Background sets of “nonphospho-13-mers” (the 12 residues surrounding each serine/threonine that was not found to be phosphorylated) were retrieved from the proteome databases of *T. thermophila*, *P. falciparum*, and *H. sapiens* using a custom Perl script. The databases were downloaded from the *Tetrahymena* Genome Database (release 2008) (32, 33), PlasmodDB (Release 7.0), and the EMBL-EBI server (Human, 3.87). The amino acid frequencies flanking the phosphorylated serine and threonine in phospho-13-mer data and the amino acid frequencies flanking the nonphosphorylated serine and threonine in nonphospho-13-mer datasets of each organism were both calculated with custom Perl scripts. Peptides of fewer than 13 amino acids or containing rare amino acids were omitted. The relative occurrence of each amino acid flanking the phosphorylation site was calculated by log₂ (frequency of each flanking amino acid in phospho-13-mers)/(frequency of each flanking amino acid in non-phospho-13-mers). An intensity map was generated as previously described (30).

**Kinase Prediction and Motif Identification**—For protein kinases and protein-docking motif prediction, the confirmed phosphopeptides were searched with the online algorithm ScanSite (34). ScanSite datasets used for comparative analysis were retrieved from large-scale phosphoproteomic studies in human liver (35), primary human multiple myeloma cells (36), HeLa nuclear protein (37), mouse liver (38), zebrafish embryo (39), *Trypanosoma brucei* (40), and *S. cerevisiae* (41). For those phosphorylation sites that we were unable to assign to a kinase recognition motif when using the ScanSite algorithm under low stringency, the Motif-X algorithm was employed to identify phosphorylation motifs (42). Phosphopeptide sequences for those phosphorylation sites were pre-aligned by a custom Perl script, and their lengths were adjusted to ±6 amino acids from the central position and submitted to the Motif-X algorithm as foreground. Because of the upload restrictions, a shortened version of the *T. thermophila* protein sequence dataset (10 Mb) was generated with a custom Perl script and submitted as background. The occurrence threshold was set at 1.8% of the input dataset at a minimum of four peptides. A p value of <10⁻⁶ was considered significant.

**T. thermophila Phosphoproteome Functional Annotation, Classification, and Enrichment Analysis**—Blast searching, mapping, and annotation of phosphoproteins were performed using Blast2GO software (43). Functional enrichment analysis of phosphoproteins was performed to determine the significantly enriched Gene Ontology (GO) terms and relevant proteins by using the BINGO 2.4.4 (44) plugin in the Cytoscape platform (45). Enrichment analysis of GO term assignment was performed in reference to the entire annotated *T. thermophila* proteome (containing 7199 proteins). The corrected p values were derived from a hypergeometric test followed by Benjamini and Hochberg false discovery rate correction. A corrected p value of <0.05 was regarded as significant.

**Pathway Analysis, Network Generation, and Module Partition**—For pathway analysis, KEGG pathway information on *T. thermophila* was extracted from the KEGG database and 88 phosphoproteins were assigned to 45 KEGG pathways. A phosphoprotein network was built on correlations (z-score, calculated with a context likelihood of relatedness algorithm) of pairwise phosphoprotein genes extracted from the published *T. thermophila* gene network (26). The correlation cut-off value that was used to establish the phosphoprotein network was the same as in a previous study (26). The Cytoscape software (version 2.6) was used for network visualization (45). The scalable unsupervised cluster tool Markov Cluster Algorithm was applied for module partition of phosphoprotein networks, with the same parameters as described by Enright et al. (46).
RESULTS AND DISCUSSION

Establishment of the T. thermophila Phosphoproteome—In order to obtain a large-scale dataset of phosphorylation sites and phosphoproteins in T. thermophila, a systematic investigation employing an MS-based high-throughput proteomic approach was performed. As depicted in Fig. 1, proteins extracted from mixed T. thermophila cell samples harvested at seven time points in three different physiological/developmental stages were analyzed via online two-dimensional (strong cation exchange/reverse-phase) nano-LC-MS/MS. After all results from three independent experiments had been combined, 1384 phosphopeptides from 1008 phosphoproteins were identified with a false discovery rate of 1.32%, which corresponded to 2238 phosphorylation sites (supplemental Table S1). Details of the identified phosphopeptides, including identifying search algorithm scores and PTM scores, are provided in supplemental Table S1, in which hyperlinks have been included leading to all the MS/MS spectra. Among the 2238 phosphorylation sites, phosphoserine had the greatest proportion (80.48%), followed by phosphothreonine (15.68%) and phosphotyrosine (3.84%) (Fig. 2A). Fig. 2B shows an example of the analysis of a phosphopeptide sequence and the assignment of the phosphorylation site. The MS² spectrum identified the sequence as LM(ox)ANTTIADG-GVPKNINPM(ox)LLPpSK and the site of phosphorylation at S-123 of HTA1 (TTHERM_00790790, T. thermophila histone H2A). This doubly charged peptide was detected at m/z 1295.64 and exhibited a clear loss of 98 Da in the MS² spectrum, the signature cleavage of a phosphoric acid molecule (H₃PO₄) from phosphorylated peptide ions. A series of y-ions and b-ions indicated an unambiguous phosphoserine site (PTM score = 187). As expected, some previously reported T. thermophila phosphoproteins and their phosphorylation sites were also identified in the present research. For example, three phosphorylation sites (S-43/S-45/T-47) of T. thermophila linker histone H1 were identified in this study, which is consistent with previous research (6). In agreement with previous predictions, the phosphorylation of S-231 in T. thermophila macronuclear phosphoprotein Nopp52 was also identified in this study (47).

To address the functional distribution of the 1008 identified phosphoproteins, we performed a functional annotation using the online automatic annotation tool Blast2GO (43). After Blast, mapping, and annotation, 41.17% (415/1008) of the
phosphoproteins were annotated with at least one GO term. The annotation rate of the phosphoproteome was significantly higher than that of the whole genome (29.12%). The GO term distribution of *T. thermophila* phosphoproteins in three categories is summarized and represented in Fig. 2C and supplemental Fig. S1. In the “biological process” category, the annotated phosphoproteins were predicted to participate in a wide variety of biological processes, including transport, gene expression, mRNA metabolic processes, and macromolecular biosynthesis, which corresponded to the pleiotropic nature of protein phosphorylation. In the “molecular function” category, we found that the most predominant function of phosphoproteins was binding, including ATP binding, protein binding, metal ion binding, and DNA binding. The second significant function was protein serine/threonine kinase activity (supplemental Fig. S1A). In addition, GO terms annotated in the “cellular component” category revealed a ubiquitous distribution of phosphoproteins in *T. thermophila* (supplemental Fig. S1B).

**Analysis of Amino Acid Residues Flanking the Phosphorylation Sites**—*T. thermophila* have a large and unique kinome in which nearly 60% of kinases cannot be assigned to any...
known kinase groups (5). In addition, the *T. thermophila* genome (AT content ~ 78%) has a strong bias to AT-rich codons, and amino acids encoded by these high-AT codons are found more frequently than other residues (48). Recent phosphoproteomic analysis of *P. falciparum*, a parasitic relative of *T. thermophila* that also has a high AT content in its genome, indicates that the AT-rich genome leads to a unique amino acid composition around the phosphorylation sites (30). The unique kinase and genome codon bias of *T. thermophila* prompted us to investigate whether *T. thermophila* has distinct phosphorylation motifs.

To determine the residue composition surrounding the phosphorylation sites, we compared the occurrence of each amino acid in phospho-13-mers among *T. thermophila*, *P. falciparum*, and *H. sapiens*. As shown in Fig. 3A, we found that the occurrences of amino acids encoded by AT-rich codons (e.g., N, I) were very high in both protists but were relatively lower in human. In contrast, the occurrence of P and R, encoded by GC-rich codons, was much higher in human than in the two protists (Fig. 3A). To further reveal the overrepresented residues in the phospho-13-mers, we compared the occurrence of each amino acid in phospho-13-mers with that in nonphospho-13-mers extracted from the whole proteome. We found that *T. thermophila* shared some conserved features in the phosphoproteome with *P. falciparum* and *H. sapiens*. For instance, residues surrounding the phosphorylation sites in the three organisms tended to be flexible residues rather than rigid residues (Fig. 3B). Amino acids within the phospho-13-mers in the three organisms were all significantly depleted in H, L, Y, F, I, C, and W and enriched in E, D, and R (Fig. 3B). However, Q, A, and V were specifically enriched in *T. thermophila* phosphopeptides relative to *H. sapiens* and *P. falciparum*, indicating distinct features of *T. thermophila*. Interestingly, most of the enriched residues (E, P, S, Q, G, R, A) in *T. thermophila* phosphopeptides were significantly overrepresented in regions of intrinsic sequence disorder (Fig. 3B), which is in line with previous observations (49).

Furthermore, the relative occurrence of each amino acid residue flanking the phosphorylation sites was calculated and visualized as an intensity map (Fig. 3C). We found that *T. thermophila* shared some conserved features with *P. falciparum* and *H. sapiens* in the properties of amino acid residues surrounding the phosphorylation sites. P at position +1 and R at position −3 for the serine phosphorylation site were overrepresented in *T. thermophila*, as well as in *H. sapiens* and *P. falciparum*. Although E, D, and S were not as enriched as in human and *P. falciparum*, a moderate enrichment of E at positions +2 and +3 and D at positions +1 and +2 for S was found in *T. thermophila*. In addition to those evolutionarily conserved features, the intensity map also shows distinct features of *T. thermophila* phosphorylation motifs. For example, in *T. thermophila*, G and Q at position +1 for the serine phosphorylation site were found to be moderately enriched, which was not observed in human, and only Q appeared to be slightly enriched in *P. falciparum* (Fig. 3C). Moreover, unlike in human and *P. falciparum*, S at positions +1 and −1 in the *T. thermophila* threonine phosphorylation site were significantly enriched, and L at position −3 for both serine and threonine phosphorylation sites was specifically enriched, revealing the unique features of phosphorylation motifs in *T. thermophila* (Fig. 3C).

**Kinase Prediction and Motif Identification**—To predict the upstream kinases responsible for phosphorylation in the *T. thermophila* phosphoproteome, we analyzed the phosphorylated sites in *T. thermophila* with an online searching algorithm, Scansite, which was developed for identifying 65 kinase/binding domain recognition motifs by using the position-specific scoring matrices derived from experimental data (34). In total, 992 unique phosphorylation sites in *T. thermophila* were predicted to be associated with kinase consensus motifs by Scansite at low stringency, corresponding to the top 5% of all scored sites (supplemental Table S2A). Within the top 1% of Scansite hit scores, 603 of phosphorylation motifs were identified. Most of the phosphorylation motifs were classified into the basophilic serine/threonine kinase group (26.4%), followed by the DNA damage kinase group (22.9%) and the acidophilic kinase group (18.9%) (supplemental Table S2A). To test whether this phylogenetically remote single-celled model organism contains unique features in motif occurrence, we performed a comparative analysis of Scansite results with results from two other representative studies on the parasitic protozoa *T. brucei* (40) and human (35).

As shown in Fig. 4, one of the most abundant phosphorylation motifs in *T. thermophila* and the other two organisms was an acidophilic serine/threonine kinase, casein kinase 2 (CK2). This is consistent with the notion that the evolutionarily conserved CK2 is a constitutively activated kinase responsible for the phosphorylation of hundreds of substrates that functions in a wide variety of cellular processes (50). Another predominant phosphorylation motif in the three organisms was the cyclin-dependent kinase 5 recognition motif. It is well established that the pleiotropic cyclin-dependent kinase 5 is one of the key factors in regulating mitochondria fission during neuronal cell apoptosis (51). Therefore, those putative cyclin-dependent kinase 5 motifs containing phosphoproteins identified in this study might provide a point of departure for more in-depth studies for understanding the regulatory roles of proteins in the cyclin-dependent kinase 5 family in *T. thermophila*.

Notably, the occurrence of phosphorylation motifs recognized by two kinase members of the phosphatidylinositol 3-kinase-related kinase (PIKK) family, ataxia-telangiectasia mutated and DNA-dependent protein kinase (DNA-PK), were uniquely high in *T. thermophila*. Ataxia-telangiectasia mutated/ ATR and DNA-PKs are important regulators of genome stability and specifically recognize the phosphorylated serine or threonine residues followed by glutamine (the pS/pQ motif) (52, 53). As expected, the high occurrence of the PIKK rec-
Phosphoproteomic Analysis of *Tetrahymena thermophila*

**Fig. 3.** *Comparative analysis of *T. thermophila* phosphorylation sites with *P. falciparum* and *H. sapiens.* A, frequency of each amino acid flanking phosphoserine and phosphothreonine within pre-aligned “phospho-13-mers” of *T. thermophila*, *P. falciparum*, and *H. sapiens*. Phosphopeptides that had fewer than 13 amino acids were omitted. Amino acids on the x-axis are arranged according to the flexibility scale. B, analysis of enrichment and depletion of amino acids in phospho-13-mers by comparing the frequency of each amino acid described in A with the frequency of each amino acid flanking serine and threonine within the pre-aligned “nonphospho-13-mers.” The positive log2 score indicates the enrichment of this amino acid, and the negative log2 score shows depletion. Amino acids on the x-axis are arranged according to the flexibility scale. C, intensity maps of three organisms show the enrichment and depletion of each residue in specific positions of phospho-13-mers versus nonphospho-13-mers. The color is plotted according to the log2 score of the ratio of each residue’s frequency within phospho-13-mers versus nonphospho-13-mers. Red indicates enrichment, and gray indicates depletion. Amino acids on the y-axes are arranged according to the flexibility scale.

**Fig. 4.** *Comparative analysis of motif occurrence of *T. thermophila* with *T. brucei* and *H. sapiens.* Phosphorylation motif recognized by acidophilic kinase casein kinase 2 and basophilic kinase cyclin-dependent kinase 5 were predominant in all three organisms. Many phosphorylation sites in these three organisms were predicted to be the potential 14–3–3 binding domain recognition site. Notably, the percentage of DNA damage group kinase recognition motif was uniquely high in *T. thermophila*. The occurrence of calmodulin-dependent protein kinase 2 recognition motif was also high in *T. thermophila*. Motif occurrence numbers used for comparative analysis were retrieved from Scansite database in the literature at medium stringency. Kinase group classification was based on the Scansite algorithm (34).

Oligonucleotide motif is consistent with the observation of enriched Q at position +1 for phosphoserine in the intensity map (Fig. 3C).

Four putative PIKKs have been identified in the *T. thermophila* proteome, and functional studies support the notion that PIKK family members play vital roles in the maintenance of the *T. thermophila* life cycle (5, 54). For example, the disruption of *T. thermophila* ATR kinase (also called ATR1) results in pleiotropic defects in growth and conjugation, such as decreased cell size, micronuclear chromosome instability, and defects in meiotic micronuclear elongation (54). Although direct experimental evidence supporting the important roles of DNA-PK in *T. thermophila* is lacking, studies of this evolutionarily conserved protein kinase in other model organisms, such as yeast (55) and mammals (56), have revealed that DNA-PK is a key factor in response to DNA damage and functions as a caretaker to maintain telomere stability (56, 57). *T. thermophila*, as well as other ciliates, undergoes genome reorganization during macronuclear formation. Chromosomal breakage occurs at hundreds of specific sites in *T. thermophila*, along with DNA deletion, rejoining, and telomere addition (58, 59). It is likely that PIKK family members, which are involved in some critical cellular events such as telomere maintenance and DNA damage repair (60, 61), play important roles during the *T. thermophila* life cycle. In addition to ATM/ATR and DNA-PK recognition motifs, the calmodulin-dependent kinase 2 phosphorylation motif was more common in the *T. thermophila* phosphoproteome than in the other organisms. One possible reason for the high abundance of the calmodulin-dependent kinase 2 motif in ciliate *Tetrahymena* is that calmodulin-dependent kinase 2 might play an important role in ciliary movement (62).

The expansion of some phosphorylation motifs observed in *T. thermophila* prompted us to test whether there is a correlation between the expansions/contractions of the kinase domain and the relative counts of the corresponding phosphorylation motifs. We performed a Pearson correlation analysis comparing kinase numbers and motif numbers in *T. thermophila* (5) and human (11, 35) (supplemental Table S2C). Our results showed that there was no significant correlation between the relative number of kinase families and the relative number of their equivalent motifs: for *T. thermophila*, the Pearson correlation coefficient and *p* value were −0.08 and 0.78, respectively; for *H. sapiens*, the Pearson correlation coefficient and *p* value were 0.37 and 0.22, respectively.
Interestingly, we observed different kinase-motif patterns (supplemental Table S2C), suggesting that different kinase families may use different mechanisms to regulate their substrates. For example, the motif numbers of CK2 and CK1 are 82 and 11, respectively, whereas their kinase numbers are 2 and 19, respectively. It is tempting to speculate that CK2 tends to be pleiotropic and constitutively activated (50) (pattern 1), whereas CK1 may regulate diverse signaling pathways through gene duplication and functional specification (pattern 2) (supplemental Table S2C). Furthermore, by comparing the kinase-motif patterns of *T. thermophila* and human, we found that different organisms contain different kinase-motif patterns in specific protein kinase families. For example, there was an apparent expansion of Akt family protein kinases in *T. thermophila*, and the Akt family kinase-motif pattern in *T. thermophila* resembled pattern 2, whereas the human Akt family kinase-motif pattern resembled pattern 1. These results could help us understand kinase expansions/contractions better through comparisons of the kinase-motif patterns of specific protein kinase families in different organisms. Overall, the identification of kinase recognition motifs in phosphoproteins provides a rich resource that can be used to examine the kinase-mediated signaling networks in *T. thermophila*.

To perform a comparative analysis of Scansite results among different organisms, we collected the Scansite data from phosphoproteomic studies of other organisms. Supplemental Table S2D lists the fraction of phosphosites in other organisms that are assigned to a kinase motif in Scansite at low, medium, and high stringency. As shown in supplemental Table S2D, the fraction of phosphosites that are not assigned to a kinase motif in *Tetrahymena* is significantly higher than that of the other organisms. Interestingly, by comparing the Scansite results (medium stringency) for different organisms, we found a clear trend in which the percentage of phosphorylation sites recognized by Scansite increased from single-celled yeast to multicellular higher eukaryotes. The result is not surprising in the sense that the Scansite algorithm was developed based on the optimal phosphorylation sites for particular human protein Ser/Thr kinases (63, 64), and this trend might also indicate evolutionary alterations in eukaryotic protein kinase recognition motifs.

In addition to this, 630 (approximately 60%) of the putative kinases in the *Tetrahymena* kinase were reported to be ciliate-specific kinases, whereas the Scansite algorithm recognized only a limited number of evolutionarily conserved and well-defined kinase motifs. So, we assume that *T. thermophila* may contain more ciliate-specific phosphorylation motifs. Also, the intensity map showed that *T. thermophila* phosphoproteins may contain species-specific phosphorylation motifs (Fig. 3C). We therefore employed another online algorithm, Motif-X, to find the unique phosphorylation motifs in *T. thermophila* in the remaining 1246 phosphorylation sites. In total, seven motifs were identified by the Motif-X algorithm with stringent criteria (p < 0.000001), including four phosphoserine motifs, two phosphothreonine motifs, and a phosphotyrosine motif (Fig. 5, supplemental Table S2E). As described in a previous study, the NIMA-related protein kinase family has undergone a large expansion in *T. thermophila* (5). Consistent with this, we found that the most predominant phosphorylation motif identified by the Motif-X algorithm was the NIMA-related protein kinase recognition motif [LXXpS]. A rarely reported motif [pSG] and three consecutive serine or threonine motifs ([pSS], [pTS], and [pST]) with relatively high occurrence were also identified by Motif-X; however, their upstream protein kinases are still unknown. The high occurrence of those unique or rarely reported phosphorylation motifs might be consistent with the unique features of the *T. thermophila* kinome. It is widely understood that a protein kinase usually recognizes certain motifs on its substrate (42). Therefore, the phosphorylation motifs identified in this study might shed light on the substrate-recognition properties of *T. thermophila* protein kinases and will be a valuable resource for further investigations on the biological function of the corresponding protein kinases.

**Pathway Analysis of Phosphoproteins**—To determine whether the identified phosphoproteins play a role in any biological/signaling pathways, the identified phosphoproteins were searched against the KEGG database. 88 of the submitted 1008 phosphoproteins were annotated in the KEGG database and involved in several pathways (supplemental Table S3). For example, 19 of the 88 phosphoproteins were assigned to ribosome-related pathways, 10 of which were assigned to "ribosome biogenesis in eukaryotes" (tet03008), including 9 KEGG homologues (RIO2, NMD3, XRN1/2, BMS1, NUG2, UTP18, NOP56, and MDN1), and the other 9 phosphoproteins were annotated as ribosome components (tet03010), including 5 small subunit ribosomal proteins (S4e, S9e, S10e, S15e, and S17e) and 4 large subunit ribosomal proteins (L1, L2, L24e, and L10e). As an important mechanism for translational regulation, the phosphorylation of ribosomal proteins has been intensively investigated (65, 66). For example, phosphorylation of ribosomal protein S6, a component of the eukaryotic 40S ribosomal subunit, was reported to facilitate protein synthesis (67). Moreover, phosphorylation of ribosomal protein L13a, a component of the eukaryotic 60S ribosomal subunit, was reported to be responsible for the translational silencing of a specific protein (67). It is therefore tempting to speculate that the phosphorylated ribosomal proteins in our dataset might be involved in translational regulation or protein synthesis.

In addition to the ribosome-related pathways, quite a few phosphoproteins were involved in mRNA metabolism. For example, eight factors in spliceosome components (tet03040) were observed to be phosphorylated: SNW1 (SKIP), DDX46 (Prp5), UAP56, DHX8 (Prp22), SF3A3, SF3B1, U2AF2, and RBM8A (Y14). Studies in other model eukaryotes indicate that the phosphorylation of those evolutionarily conserved spliceosome components is essential for modulating spliceo-
some assembly and pre-mRNA splicing (68, 69). Therefore, the identification of phosphorylation sites in the *T. thermophila* spliceosome is a fundamental step for further elucidating the role of phosphorylation in the splicing process. Moreover, some phosphoproteins are assigned to pathways that relate to mRNA metabolism, such as the RNA transport pathway (tet03013), the RNA degradation pathway (tet03018), and the mRNA surveillance pathway (tet03015) (supplemental Table S3). Furthermore, significant numbers of phosphoproteins are also implicated in macromolecular metabolism processes, indicating that protein phosphorylation regulates diverse cellular functions in *T. thermophila*.

Analysis of Phosphorylated Protein Kinases—As protein kinases reside at the nodes of phosphorylation-based signal transduction pathways, the catalytic activity of many protein kinases is also regulated by phosphorylation or autophosphorylation (70). In this study, 145 phosphorylation sites in 72 protein kinases were identified (supplemental Table S4). According to the previous classification of *Tetrahymena* protein kinases (5), 35 of the 72 phosphorylated protein kinases identified in our study are assigned to the known protein kinase groups, including CMGC, AGC, CAMK, CK1, and TKL. However, the remaining phosphorylated protein kinases (~50%) could not be categorized into any known kinase groups and appeared to be unique kinases in *T. thermophila*.

Previous studies have shown that phosphorylation on the activation loop of certain protein kinase families is the most common way to maintain kinase activity (70). We therefore investigated the localization of phosphorylation sites on protein kinases via sequence alignment analysis. We found 38 phosphorylation events on the activation loops of 19 protein kinases (Fig. 6).

![Phosphorylation motif extraction](Fig. 5)

**Fig. 5.** *Results of phosphorylation motif extraction with Motif-x algorithm.* Phosphosites that were not recognized by the Scansite program were further analyzed with the Motif-X algorithm. A, NEK6 recognition motif. B, proline-directed motif resembling the ERK1/2 recognized motif. D-F, consecutive serine/threonine motif. G, phospho-tyrosine motif. The complete information generated in the Motif-X analysis is listed in supplemental Table S2F.
nine and tyrosine phosphorylation of a TXY motif in a conserved activation loop (71, 72).

Tyrosine phosphorylation sites in the activation loops of two putative glycogen synthase kinase 3 β (GSK-3β) protein kinases were also observed in this study and showed high sequence similarity to the phosphorylation sites in human glycogen synthase kinase 3 β (Fig. 6). It has been reported that human glycogen synthase kinase 3 β is activated by intramolecular autophosphorylation of tyrosine residue Y216 (211EPNVSpYICSR220) in its activation loop (73). It is therefore likely that the tyrosine phosphorylation sites of the two T. thermophila glycogen synthase kinase 3 β kinases may also be phosphorylated via autophosphorylation. In addition to the known protein kinase families, phosphorylation sites were also observed in the activation loops of four Tetrahymania-specific protein kinases (indicated by asterisks in Fig. 6), thus providing clues regarding the regulatory mechanisms of phosphorylation in these poorly understood protein kinases.

Integrated Analysis of T. thermophila Phosphoproteome and Transcriptome—In this study, ~60% of the identified phosphoproteins could not be assigned to any GO terms, and nearly 90% of them could not be annotated in the KEGG database. Therefore, we performed an integrated analysis of the phosphoproteome and transcriptome to reveal the functional roles of these unannotated phosphoproteins in T. thermophila.

It is well established that gene expression network-based gene function prediction is an effective way to refine genome annotation (74), and two or more proteins whose genes have similar expression profiles in the gene expression network generally tend to interact physically or function in related processes (75, 76). We recently built the Tetrahymena Gene Network (TGN) based on gene expression data from 67 expression microarrays (26). Genes linked in TGN share similar mRNA expression patterns (26). To explore the potential biological functions of the unannotated phosphoproteins identified in this study, we performed an integrated analysis of phosphoproteins based on the established TGN. First, a “phosphoprotein-gene network” containing 915 of the 1008 phosphoproteins was generated by extracting the gene interaction information from TGN. Phosphoprotein genes that are not included in the network were filtered out during the TGN construction process because their expression profiles showed low similarity to others (26).

The expression correlation between two linked genes was represented as the network distance (Fig. 7). To better interpret the functional implications of phosphoproteins from the generated phosphoprotein-gene network, a widely used unsupervised cluster algorithm for networks, MCL (46), was applied to partition the phosphoprotein-gene network into highly interconnected modules. As shown in Fig. 7, several functional modules were classified, and genes in the same functional module are densely connected. It is well established that genes in the same module are more likely to function in related biological processes (77). Therefore, we used the four largest modules for further analysis (Fig. 7). In total, these four modules contained 531 phosphoproteins, 294 of which lacked GO annotations. We further performed GO enrichment analysis of each module to infer their potential functions and biological processes (Table I and
null
involved in and further refined the genome annotation of the biological processes that the phosphoproteins might be phosphoproteins with TGN provided novel insights into protein phosphorylation. The clustering analysis of these proteins (Fig. 7). Therefore, these phosphoproteins with each other, and no significant GO terms were enriched in part in processes that are related to ribosomal biogenesis. suggesting that phosphoproteins in this module might take to the ribosomal biogenesis pathway in the KEGG database, in this module participate in biological processes that are unique percentage of phosphoproteins in this module. In addition, genes in module 1 are also less informative in the KEGG database. So, it is possible that phosphoproteins clustered in this module participate in biological processes that are unique in T. thermophila. In module 4, 6 of 10 proteins were assigned to the ribosomal biogenesis pathway in the KEGG database, suggesting that phosphoproteins in this module might take part in processes that are related to ribosomal biogenesis.

As expected, phosphoproteins that were not assigned to the four modules mentioned above are sparsely connected with each other, and no significant GO terms were enriched in these proteins (Fig. 7). Therefore, these phosphoproteins might function in various aspects of cellular biological processes, which is in agreement with the diverse functions of protein phosphorylation. The clustering analysis of T. thermophila phosphoproteins with TGN provided novel insights into the biological processes that the phosphoproteins might be involved in and further refined the genome annotation of T. thermophila.

Profiling of Phosphorylation Pathways in T. thermophila— One of the most challenging problems in phosphoproteome studies is how to match the phosphoproteome to the kinome accurately. Up to now, kinase recognition motif-based prediction has been the most widely used method to predict kinase-substrate relationships (34, 81). But it is still infeasible to match a phosphorylation site to a specific protein kinase, as some protein kinases share the same or similar recognition motifs. As indicated by Linding et al., the reliability of motif-based phosphorylation network prediction is greatly influenced by lacking contextual information, such as expression data and cellular localization data (82). Previous reports suggest that genes with similar expression profiles are more likely to encode interacting proteins (75). In addition, by using gene co-expression data along with other datasets, Linding et al. improved the accuracy of in vivo phosphorylation network prediction (82). Therefore, we performed an integrated analysis of the phosphoproteomics data and the TGN data to test whether it is feasible to improve the accuracy of motif-based kinase-substrate predictions and infer the potential upstream protein kinases of the identified T. thermophila phosphorylated proteins. We speculate that if a phosphoprotein contains a recognition motif for a specific protein kinase, and if they are linked in TGN (with similar expression patterns), they could be regarded as a potential kinase-substrate pair in the T. thermophila signaling network. Based on this speculation, we identified 1656 putative kinase-substrate pairs in TGN among the well-defined protein kinase families. Through further comparison with motif-based predictions (derived from ScanSite data, Motif-X data, and manually identified MEK recognition motifs), we found that 150 of the TGN predictions overlapped with the motif-based predictions (supplemental Table S6A). That is, on average, about 9% of the phosphoproteins contained motifs that could be recognized by the connected protein kinases in TGN (supplemental Table S6B). These 150 overlapped predictions might help us to better understand the kinase-substrate interactions in T. thermophila. To examine the usefulness of this integrated prediction method, we first tried to compare the 150 overlapped kinase-substrate interaction predictions with well-defined kinase-substrate interactions. Because only a few kinase-substrate relationships have been clearly defined in T. thermohila, it was not feasible to evaluate our method by using established kinase-substrate interactions. We therefore tried to find evolutionarily conserved kinase-substrate interactions based on our results (supplementary Table S6). We found a MAPK/ERK recognition motif containing phosphoprotein (TThERM_00079330) linked with one of the T. thermophila MAPKs, MPK2 (TThERM_00760190), which was also a phosphoprotein and identified with a dual phosphorylation motif in its activation loop (Fig. 6). Then, we further investigated the genes linked with MPK2 in TGN. We found that a mitogen-activated protein kinase kinase homologue shared a similar mRNA expression pattern with MPK2 (Fig. 8), which is consistent with the notion that MEK was the upstream activator of MAPK in the evolutionarily conserved MAPK pathway and responsible for the dual phosphorylation of Thr and Tyr residues in the activation loop of MAPKs (83). Interestingly, a Ras family protein (TThERM 00037380) that was reported to be an indirect upstream activator of the MAPKs (83) and a MAPK/ERK phosphorylation motif containing phosphoprotein (TThERM_00389880) were found to have expression patterns similar to that of another MEK recognition motif containing a MAPK family protein (TThERM_00760270) (Fig. 6).

In addition to phosphorylation sites with known recognition motifs, most of the phosphoproteins in T. thermophila contain unknown motifs. In this situation, the potential kinase-substrate interactions can also be predicted by incorporating gene co-expression network analysis with a literature-mining strategy. For instance, several uncoordinated-51-like kinases (ULKs) were identified in our phosphoproteomic dataset without any known phosphorylation motifs. Previous research indicated that ULK is a crucial protein kinase in the autophagy pathway and responsible for autophagosome formation; it was reported to be phosphorylated and inhibited by mTOR.
In Tetrahymena thermophila, the autophagosome can be observed in cells after starvation (85). 52 putative ULKs were found in the T. thermophila kinome, which is 10 times more than in human (5, 11), indicating the important roles of the ULKs in T. thermophila. We identified nine phosphorylated ULKs in this study. The gene partners of these nine ULKs in TGN were investigated, and many of them were found to contain predicted protein kinase domains (data not shown). Among them was the mTOR homolog (TTHERM_00147610), which may be a potential inhibitor of ULK. Therefore, the identification of phosphorylated ULK homologues and their associated protein kinases in TGN might provide some clues regarding the regulatory mechanism of the autophagy pathway in T. thermophila. Therefore, the integrated analysis of phosphoproteomics data with TGN data should provide a point of departure for more in-depth studies on the physiological functions underlying phosphorylation and facilitate the elucidation of entire signaling networks in the specific context of T. thermophila.

In summary, we have presented a comprehensive phosphoproteome analysis of the protozoan T. thermophila. Many phosphoproteins were reliably identified, and functional inferences were made. The phosphoproteins identified are implicated in the regulation of various biological processes. Further integrated analysis of the phosphoproteome and TGN revealed the potential biological functions of many unannotated proteins. Our dataset provides an overview of phosphorylation in T. thermophila and is a valuable resource for the future understanding of signaling pathways in this single-celled eukaryote model organism. Because phosphorylation in signal transduction pathways is a dynamic process, it is now important to further characterize the biological functions underlying phosphorylation and elucidate the complex cell signaling networks in response to

Fig. 8. Expression profiles of MAPKs and putative upstream activators and downstream substrates. Gene expression data were retrieved from the Tetrahymena Functional Genomics Database (TetraFGD). Gene expression profiles are represented as physiological/developmental stages (x-axes) and log_{10} scales of the gene expression level (y-axes). A, expression profiles of a T. thermophila MAPK family protein, MPK2 (TTHERM_00760190), with its putative upstream activator (MEK homolog, THERM_00697080) and its putative downstream substrate (function-unknown protein, THERM_00759330). B, expression profiles of a T. thermophila MAPK family protein (TTHERM_00760270) with its putative indirect upstream activator (Ras family protein, THERM_00037380) and its putative substrate (function-unknown protein, THERM_00389880).
changing physiological conditions in this biologically important organism.

Acknowledgments—We thank Prof. Ronald Pearlman (York University) for his critical comments on the manuscript.

* This work was supported by grants from the Projects of International Cooperation and Exchanges Ministry of Science and Technology of China (No. 2013DFG32390), the National Natural Science Foundation of China (No. 91231118), and the Knowledge Innovation Program of CAS (KSCX2-EW-G-6-4-6) to W.M.; the Hundred Talents Program of the Chinese Academy of Sciences to F.G.; and the National Basic Research Program of China (973) (Nos. 2010CB833703 and 2012CB966803) to F.Y.

† This article contains supplemental material.

‡‡ These authors contributed to this work equally.

To whom correspondence should be addressed: Prof. Wei Miao, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China. Tel./Fax: 86-27-68780500; E-mail: miaowei@ihb.ac.cn; Prof. Fuquan Yang, Institute of Biophysics, Chinese Academy of Sciences, Wuhan, Hubei 430072, China. Tel./Fax: 86-27-68780050; E-mail: miaowei@ihb.ac.cn.

††† These authors contributed to this work equally.

We thank Prof. Ronald Pearlman (York University) for his critical comments on the manuscript.

REFERENCES

1. Collins, K., and Gorovsky, M. A. (2005) Tetrahymena thermophila. Curr. Biol. 15, R817–R818
2. Latham, J. A., and Cech, T. R. (1989) Defining the inside and outside of a catalytic RNA molecule. Science 245, 276–282
3. Blackburn, E. H., and Gall, J. G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J. Mol. Biol. 120, 33–53
4. Greider, C. W., and Blackburn, E. H. (1985) Identification of a specific telomere terminal transferase-activity in Tetrahymena extracts. Cell 43, 405–413
5. Eisen, J. A., Coyne, R. S., Wu, M., Wu, D. Y., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q. H., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K., Garg, J., Pearlman, R. E., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkewitz, A. P., Asai, D. J., Wing, E. F., Wang, Y. F., Cai, H., Collins, K., Stewart, A., Lee, S. R., Wilamowska, K., Weinberg, Z., Ruzzo, W. L., Wiwa, D. G. (2008) The model bacterium Bacillus subtilis. Mol. Cell. Proteomics 6, 697–707
7. Song, X. Y., Gjoneska, E., Ren, Q. H., Taverna, S. D., Allis, C. D., and Gorovsky, M. A. (2007) Phosphorylation of the SQ H2A.X motif is required for proper meiosis and mitosis in Tetrahymena thermophila. Mol. Cell. Proteomics 5, 1933–1939
8. Song, Y. Y., Gjoneska, E., Ren, Q. H., Taverna, S. D., Allis, C. D., and Gorovsky, M. A. (2007) Phosphorylation of the SQ H2AX motif is required for proper meiosis and mitosis in Tetrahymena thermophila. Mol. Cell. Proteomics 5, 1933–1939
9. Song, Y. Y., Gjoneska, E., Ren, Q. H., Taverna, S. D., Allis, C. D., and Gorovsky, M. A. (2007) Phosphorylation of the SQ H2AX motif is required for proper meiosis and mitosis in Tetrahymena thermophila. Mol. Cell. Proteomics 5, 1933–1939
10. Huang, P. H., and White, F. M. (2008) Phosphoproteomics: unraveling the signaling web. Mol. Cell 31, 777–781
11. Ozu, N., Arumugam, V., Timm, W., Haseley, N., Steen, H., and Steen, J. A. (2010) Phosphoproteomics. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 259–277
12. Placet, J., Degvan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCarty, R. R., Schmidt, M. C., Rachidi, N., Lee, S. J., Mah, A. S., Meng, L., Stark, M. J., Stern, D. F., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P. F., and Snyder, M. (2005) Global analysis of protein phosphorylation in yeast. Nature 438, 679–684
13. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. Science 298, 1912–1934
14. Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Haritarun, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Chen, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Biney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Cherwitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J. L., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C. K., O’Farrell, P. H., Pickeral, O. K., Shue, C., Voellh, L. B., Zhang, J., Zhao, Q., Zheng, X. H., and Lewis, S. (2000) Comparative genomics of the metazoan proteomes. Science 287, 2025–2034
15. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Tian, R., Wan, D., Zou, H., and Gu, J. (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics 8, 1346–1361
16. Thingholm, T. E., Jensen, O. N., and Larsen, M. R. (2009) Analytical strategies for phosphoproteomics. Proteomics 9, 1451–1468
17. 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
18. Pan, C., Gnad, F., Olsen, J. V., and Mann, M. (2008) Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphopeptide inhibitors. Proteomics 8, 4534–4546
19. 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
20. Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., and Mann, M. (2008) Phosphoproteome analysis of E-coli reveals evolution and conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol. Cell. Proteomics 7, 299–306
21. Macek, B., Mijakovic, I., Olsen, J. V., Gnad, F., Kumar, C., Jensen, P. R., and Mann, M. (2007) The serine/threonine/tyrso phosphate phosphoproteome of the model bacterium Bacillus subtilis. Mol. Cell. Proteomics 6, 697–707
22. Swaney, D. L., Weng, C. D., Thomson, J. A., and Coon, J. J. (2009) Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. 106, 995–1000
23. Villen, 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
24. Weintz, G., Olsen, J. V., Frühauft, K., Niedzielska, M., Amić, I., Jantsch, J., Mages, J., Frech, C., Dolken, L., Mann, M., and Lang, R. (2010) The phosphoproteome of toll-like receptor-activated macrophages. Mol. Cell. Proteomics 9, 1–16
25. Xiong, J., Lu, X., Zhou, Z., Chang, Y., Yuan, D., Tian, M., Wang, L., Fu, C., Orias, E., and Miao, W. (2012) Transcriptome analysis of the model protistan Tetrahymena pyriformis. Methods Cell Biol. 91, 311–327
26. Miao, W., Xiong, J., Bowen, J., Wang, L., Wu, Y. F., Fu, C., Pearlman, R. E., and Miao, W. (2011) Phosphoproteomic Analysis of Tetrahymena thermophila
Phosphoproteomic Analysis of *Tetrahymena thermophila*

cycle. *PLoS One* 4, e4429

29. Hou, J., Cui, Z., Xie, Z., Xue, P., Wu, P., Chen, X., Li, J., Cai, T., and Yang, F. (2009) Phosphoproteome analysis of rat L6 myotubes using reversed-phase C18 prefractionation and titanium dioxide enrichment. *J. Proteome Res.* 8, 777–788

30. Treeck, M., Sanders, J. L., Elias, J. E., and Boothroyd, J. C. (2011) The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites’ boundaries. *Cell Host Microbe* 10, 410–419

31. Gnädig, F., Gunawardena, J., and Mann, M. (2010) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.* 39, D253–D260

32. Stover, N. A., Krieger, C. J., Binkley, G., Dong, Q., Fisk, D. G., Nash, R., Sutharan, A., Weng, S., and Cherry, J. M. (2006) *Tetrahymena Ge...* *Mol. Cell. Proteomics* 5, 3635–3641

33. Stover, N. A., Krieger, C. J., Binkley, G., Dong, Q., Sutharan, A., Weng, S., and Cherry, J. M. (2007) *Tetrahymena* genome database (TGD): a resource for comparative studies with a model protist. *J. Eukaryot. Microbiol.* 54, 54s

34. Obenauber, J. C., Cantley, L. C., and Yaffe, M. B. (2003) Scansite 2.0: predicting protein interactions using kinase-mediated cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31, 3635–3641

35. Han, G., Ye, M., Liu, H., Song, C., Sun, D., Wu, Y., Jiang, X., Chen, R., Wang, C., Wang, L., and Zou, H. (2010) Phosphoproteome analysis of human liver tissue by long-gradient nanoflow LC coupled with multiple stage MS analysis. *Electrophoresis* 31, 1080–1089

36. Ge, F., Xiao, C. L., Yin, X. F., Lu, C. H., Zeng, H. L., and He, Q. Y. (2010) Phosphoproteomic analysis of primary human myeloma cells. *J. Proteomics* 73, 1381–1390

37. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, 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

38. Tan, F., Zhang, Y., Mi, W., Wang, J., Wei, J., Cai, Y., and Qian, X. (2008) Enrichment of phosphopeptides by Fe3+–immobilized magnetic nanoparticles for phosphoproteome analysis of the plasma membrane of mouse liver. *J. Proteome Res.* 7, 1078–1087

39. Lemeer, S., Pinkse, M. W., Mohammed, S., van Breukelen, B., den Hertog, J., Sliper, M., and Heck, A. J. (2008) Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *J. Proteome Res.* 7, 1555–1564

40. Nett, I. R., Martin, D. M., Miranda-Saavedra, D., Lamont, D., Barber, J. D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. * Genome Res.* 13, 2498–2504

41. Bork, P., Kondo, T., and Kondor, R. (2000) Cytoscape: a software tool for displaying biological networks. *Bioinformatics* 16, 367–368

42. Maier, A., Schick, M., Hellwig, J., and Mohr, C. (2007) miRanda: a web server for predicting microRNA specificity. *Nucleic Acids Res.* 35, 3635–3638

43. Sethuraman, A., Weng, S., and Cherry, J. M. (2007) Tetrahymena genome database (TGD): a resource for comparative studies with a model protist. *Nucleic Acids Res.* 35, D500–D503

44. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, 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

45. Tan, F., Zhang, Y., Mi, W., Wang, J., Wei, J., Cai, Y., and Qian, X. (2008) Enrichment of phosphopeptides by Fe3+–immobilized magnetic nanoparticles for phosphoproteome analysis of the plasma membrane of mouse liver. *J. Proteome Res.* 7, 1078–1087

46. Nemeer, S., Pinkse, M. W., Mohammed, S., van Breukelen, B., den Hertog, J., Sliper, M., and Heck, A. J. (2008) Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *J. Proteome Res.* 7, 1555–1564

47. Nett, I. R., Martin, D. M., Miranda-Saavedra, D., Lamont, D., Barber, J. D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. * Genome Res.* 13, 2498–2504

48. Bork, P., Kondo, T., and Kondor, R. (2000) Cytoscape: a software tool for displaying biological networks. *Bioinformatics* 16, 367–368

49. Maier, A., Schick, M., Hellwig, J., and Mohr, C. (2007) miRanda: a web server for predicting microRNA specificity. *Nucleic Acids Res.* 35, 3635–3638

50. Sethuraman, A., Weng, S., and Cherry, J. M. (2007) Tetrahymena genome database (TGD): a resource for comparative studies with a model protist. *Nucleic Acids Res.* 35, D500–D503

51. Meier, K., Suppanz, I. E., Linger, P., Planchamp, V., Goricke, B., Fichtner, L., Braus, G. H., Dietz, G. P., Jakobs, S., Bahr, M., and Weishaupt, J. H. (2007) Cyclin-dependent kinase 5 is an upstream regulator of mitochondrial fission during neuronal apoptosis. *Cell Death Differ.* 14, 651–661

52. Shiloh, Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* 3, 155–168

53. Meier, K., Suppanz, I. E., Linger, P., Planchamp, V., Goricke, B., Fichtner, L., Braus, G. H., Dietz, G. P., Jakobs, S., Bahr, M., and Weishaupt, J. H. (2007) Cyclin-dependent kinase 5 is an upstream regulator of mitochondrial fission during neuronal apoptosis. *Cell Death Differ.* 14, 651–661

54. Shiloh, Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* 3, 155–168

55. Meier, K., Suppanz, I. E., Linger, P., Planchamp, V., Goricke, B., Fichtner, L., Braus, G. H., Dietz, G. P., Jakobs, S., Bahr, M., and Weishaupt, J. H. (2007) Cyclin-dependent kinase 5 is an upstream regulator of mitochondrial fission during neuronal apoptosis. *Cell Death Differ.* 14, 651–661

56. Shiloh, Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* 3, 155–168

57. Meier, K., Suppanz, I. E., Linger, P., Planchamp, V., Goricke, B., Fichtner, L., Braus, G. H., Dietz, G. P., Jakobs, S., Bahr, M., and Weishaupt, J. H. (2007) Cyclin-dependent kinase 5 is an upstream regulator of mitochondrial fission during neuronal apoptosis. *Cell Death Differ.* 14, 651–661

58. Shiloh, Y. (2003) ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* 3, 155–168

59. Meier, K., Suppanz, I. E., Linger, P., Planchamp, V., Goricke, B., Fichtner, L., Braus, G. H., Dietz, G. P., Jakobs, S., Bahr, M., and Weishaupt, J. H. (2007) Cyclin-dependent kinase 5 is an upstream regulator of mitochondrial fission during neuronal apoptosis. *Cell Death Differ.* 14, 651–661
controlling activity through activation segment conformation. *Mol. Cell* **15**, 661–675
71. Seger, R., and Krebs, E. G. (1995) The MAPK signaling cascade. *FASEB J.* **9**, 726–735
72. Miyata, Y., and Nishida, E. (1999) Distantly related cousins of MAP kinase: biochemical properties and possible physiological functions. *Biochem. Biophys. Res. Commun.* **266**, 291–295
73. Doble, B. W., and Woodgett, J. R. (2003) GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **116**, 1175–1186
74. Sharan, R., Ulitsky, I., and Shamir, R. (2007) Network-based prediction of protein function. *Mol. Syst. Biol.* **3**, 1–13
75. Ge, H., Liu, Z. H., Church, G. M., and Vidal, M. (2001) Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. *Nat. Genet.* **29**, 482–486
76. Hahn, A., Rahnenfuhrer, J. R., Talwar, P., and Lengauer, T. (2005) Confirmation of human protein interaction data by human expression data. *BMC Bioinformatics* **6**, 1–11
77. Girvan, M., and Newman, M. E. (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826
78. Martin, K. C., and Ephrussi, A. (2009) mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730
79. Dienstbier, M., Boehl, F., Li, X., and Bullock, S. L. (2009) Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes Dev.* **23**, 1546–1558
80. Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P., and Krause, H. M. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187
81. Hjerrild, M., Stensballe, A., Rasmussen, T. E., Kofoed, C. B., Blom, N., Sicheritz-Ponten, T., Larsen, M. R., Brunak, S., Jensen, O. N., and Gammeltoft, S. (2004) Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.* **3**, 426–433
82. Linding, R., Jensen, L. J., Ostheimer, G. J., van Vugt, M. A. T. M., Jorgensen, 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
83. Hommes, D. W., Peppelenbosch, M. P., and van Deventer, S. J. H. (2003) Mitogen activated protein (MAP) kinase signal transduction pathways and novel anti-inflammatory targets. *Gut* **52**, 144–151
84. Jung, C. H., Jun, C. B., Ro, S. H., Kim, Y. M., Otto, N. M., Cao, J., Kundu, M., and Kim, D. H. (2009) ULK-Atg13-FIP200 complexes mediate mTOR signaling to the autophagy machinery. *Mol. Biol. Cell* **20**, 1992–2003
85. Gjermansen, M., Nilsson, M., Yang, L., and Tolker-Nielsen, T. (1984) Characterization of starvation-induced dispersion in Pseudomonas putida biofilms: genetic elements and molecular mechanisms. *Mol. Microbiol.* **75**, 815–826