Molecular Systems Biology 3; Article number 139; doi:10.1038/msb4100182
Citation: ***Molecular Systems Biology*** **3:139**

www.molecularsystemsbiology.com

molecular systems biology

REPORT

PhosphoPep—a phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells

Bernd Bodenmiller[1,2], Johan Malmstrom[1], Bertran Gerrits[3], David Campbell[4], Henry Lam[4], Alexander Schmidt[1,5], Oliver Rinner[1], Lukas N Mueller[1,2], Paul T Shannon[4], Patrick G Pedrioli[6], Christian Panse[3], Hoo-Keun Lee[1], Ralph Schlapbach[3] and Ruedi Aebersold[1,4,7,*]

[1] Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland, [2] Zurich PhD Program in Molecular Life Sciences, Zurich, Switzerland, [3] Functional Genomics Center Zurich, UZH | ETH Zurich, Zurich, Switzerland, [4] Institute for Systems Biology, Seattle, WA, USA, [5] Competence Center for Systems Physiology and Metabolic Diseases, ETH Zurich, Zurich, Switzerland, [6] Institute of Biochemistry, ETH Zurich, Zurich, Switzerland and [7] Faculty of Science, University of Zurich, Zurich, Switzerland

* Corresponding author. Institute of Molecular Systems Biology, ETH Zurich, Wolfgang-Pauli-Street 16, Zurich 8093, Switzerland. Tel.: +41 44 633 31 70; Fax: +41 44 633 10 51; E-mail: aebersold@imsb.biol.ethz.ch

Received 6.6.07; accepted 29.8.07

The ability to analyze and understand the mechanisms by which cells process information is a key question of systems biology research. Such mechanisms critically depend on reversible phosphorylation of cellular proteins, a process that is catalyzed by protein kinases and phosphatases. Here, we present PhosphoPep, a database containing more than 10 000 unique high-confidence phosphorylation sites mapping to nearly 3500 gene models and 4600 distinct phosphoproteins of the *Drosophila melanogaster* Kc167 cell line. This constitutes the most comprehensive phosphorylation map of any single source to date. To enhance the utility of PhosphoPep, we also provide an array of software tools that allow users to browse through phosphorylation sites on single proteins or pathways, to easily integrate the data with other, external data types such as protein–protein interactions and to search the database via spectral matching. Finally, all data can be readily exported, for example, for targeted proteomics approaches and the data thus generated can be again validated using PhosphoPep, supporting iterative cycles of experimentation and analysis that are typical for systems biology research.

Molecular Systems Biology 16 October 2007; doi:10.1038/msb4100182
Subject Categories: proteomics; signal transduction
Keywords: data integration; *Drosophila*; interactive database; phosphoproteomics; systems biology

Introduction

It is the premise of systems biology that biological processes are studied as integrated systems consisting of multiple interacting elements and that the basis for the system's properties is the contextual information of the elements interactions. Operationally, biological systems are frequently represented as networks and their properties are studied by iterative cycles of targeted network perturbation followed by quantitative measurement of all the system's elements (Ideker *et al*, 2001).

Networks typically studied are transcriptional networks analyzed by gene expression arrays (Schena *et al*, 1995; Lipshutz *et al*, 1999) and CHIP on chip assays (Ren *et al*, 2000; Iyer *et al*, 2001), protein interaction networks analyzed by the yeast two-hybrid systems (Fields and Song, 1989; Uetz *et al*, 2000; Giot *et al*, 2003) or mass spectrometry of purified protein complexes (Rigaut *et al*, 1999; Gavin *et al*, 2002; Gingras *et al*, 2005; Ewing *et al*, 2007) and genetic interactions analyzed by synthetic lethal screens (Tong *et al*, 2001). Protein phosphorylation, a network of protein kinases and phosphatases and their respective cellular substrates, is a universal regulatory mechanism and plays a pivotal role in the control of most cellular process. Thus, the understanding of protein phosphorylation networks and their dynamic changes is of fundamental importance for systems biology (Hunter, 2000).

Recently, phosphoproteomics has become a robust technique for the analysis of protein phosphorylation networks. Typically, (phospho)protein samples are digested with a protease, and the peptides are analyzed by liquid-chromatography tandem mass spectrometry (LC-MS/MS) (Aebersold and Mann, 2003). As after the digestion of a proteome phosphopeptides are present at a low concentration, it is necessary

to specifically enrich them before analysis (Aebersold and Goodlett, 2001; Reinders and Sickmann, 2005). Recently, several phosphopeptide enrichment methods have been described and their performance has been compared (Bodenmiller *et al*, 2007a). They include affinity chromatography and phosphoramidate chemistry-based purification. The most commonly used affinity-based methods are immobilized metal affinity chromatography (IMAC) (Andersson and Porath, 1986) and titanium dioxide (TiO_2) (Pinkse *et al*, 2004; Larsen *et al*, 2005). As an alternative phosphoramidate chemistry (PAC), in which the phosphopeptides are covalently captured on an amino-modified solid phase (e.g. a dendrimer (Tao *et al*, 2005) or glass beads (Zhou *et al*, 2001; Bodenmiller *et al*, 2007b)) and are released by acid hydrolysis of the phosphoramidate bond (Zhou *et al*, 2001; Tao *et al*, 2005; Bodenmiller *et al*, 2007a, b) can be used.

Using the technologies described above, several large scale data sets on protein phosphorylation have recently been published (Ficarro *et al*, 2002; Beausoleil *et al*, 2004; Schwartz and Gygi, 2005; Olsen *et al*, 2006). However, a number of factors limit the usefulness of these data for systems biology research. First, the data sets are far from being complete. Second, false-positive and false-negative error rates are frequently unknown and spectra may not be accessible to independently assess the quality of peptide identification and assigned site of phosphorylation. Third, the data are mostly presented as lists of identified phosphopeptides, limiting their use for further experimentation or meta-analysis.

In this report, we describe PhosphoPep, a database for phosphopeptides and phosphoproteins from *Drosophila melanogaster* Kc167 cells and a suite of associated software tools as a resource for systems biology research in *D. melanogaster*. The small genome size, short generation time, the highly developed genetic tools that can be easily combined with biochemical analysis (Bier, 2005) and the high degree of conservation of signaling pathways between the fly and humans (Reiter *et al*, 2001) make *Drosophila* an ideal, but as yet largely unexplored species for systems biology. PhosphoPep contains over 10 000 high-confidence phosphorylation sites from 3472 gene models and 4583 distinct phosphoproteins, and therefore, is the as yet most completely mapped phosphoproteome of any single source.

To support further experimentation and analysis of the phosphorylation data, we added to the PhosphoPep database a number of software tools. First, we implemented a search function to detect the sites of phosphorylation on individual proteins and to place phosphoproteins within cellular pathways as defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa *et al*, 2006). Such pathways, along with the identified phosphoproteins can be interrogated by a pathway viewer and exported to Cytoscape (Shannon *et al*, 2003), a software tool, which supports the integration of the data from PhosphoPep and other databases. Second, we added utilities for the use of the phosphopeptide data for targeted proteomics experiments. In a typical experiment of this type, the known phosphorylation sites of a protein or set of proteins are detected and quantified in extracts representing different cellular conditions via targeted mass spectrometry experiments such as MRM (Gerber *et al*, 2003; Domon and Aebersold, 2006; Picotti *et al*, 2007; Stahl-Zeng *et al*, 2007; Wolf-Yadlin *et al*, 2007). Third, we made the data in PhosphoPep searchable by spectral matching through SpectraST (Lam *et al*, 2007). Specifically, for each distinct phosphopeptide ion identified in this study, all corresponding MS2 spectra were collapsed into a single consensus spectrum. Unknown query spectra can then be identified by spectral searching against the library of phosphopeptide consensus spectra.

Collectively, PhosphoPep and the associated software tools and data mining utilities support the use of the data for diverse types of studies, from the analysis of the state of phosphorylation of a single protein to the detection of quantitative changes in the state of phosphorylation of whole signaling pathways at different cellular states and has been designed to enable the iterative cycles of experimentation and analysis that are typical for systems biology research.

Results and discussion

Strategy

To generate an extensive phosphopeptide map of *D. melanogaster* KC 167 cells, we first performed a large-scale phosphorylation site mapping project as described in the Supplementary information and Supplementary Figure S1. Briefly, as the phosphoproteome strongly depends on the cellular state, we performed tryptic digestion of protein extracts from *D. melanogaster* Kc167 cells grown under various conditions: nutrient-rich medium; nutrient-depleted medium; medium supplemented with insulin (a growth inducer); medium supplemented with rapamycin (a growth inhibitor); and medium containing Calyculin A, an inhibitor of protein phosphatase 1 and protein phosphatase 2A. The combined peptide sample was separated by peptide isoelectric focusing (IEF) in a free-flow electrophoresis (FFE) instrument (Malmstrom *et al*, 2006). From each fraction phosphopeptides were isolated using three different phosphopeptide isolation methods (IMAC, TiO2 and PAC) to maximize coverage of the phosphoproteome (Bodenmiller *et al*, 2007a). Each phosphopeptide fraction was then subjected to LC-MS/MS using a high mass accuracy tandem mass spectrometer. The generated LC-MS/MS data were searched against a protein (decoy) database and the identified phosphorylation sites were validated using the PeptideProphet software tool (Keller *et al*, 2002) or the target-decoy search strategy (Elias and Gygi, 2007). The resulting combined data set consisting of 10 118 high-confidence phosphorylation sites from 3472 gene models and 4583 distinct phosphoproteins was incorporated into the PhosphoPep database.

Assignment of fragment ion spectra to phosphopeptide sequences

The fragment ion spectra obtained in this study were assigned to (phospho)peptide sequences using the sequence database search tool Sequest (Eng *et al*, 1994) and were investigated for two forms of errors in the data set: first, the miss-assignment of the fragment ion spectrum to a peptide sequence (Keller *et al*, 2002; Elias and Gygi, 2007) and second, the miss-assignment of the phospho-amino acid in an otherwise correctly identified phosphopeptide (Beausoleil *et al*, 2006).

When assessing the first type of error using the statistical tool PeptideProphet (Keller *et al*, 2002) or a decoy database (DD) (Elias and Gygi, 2007), we found that at a PeptideProphet probability score cut off value of 0.8 approximately 2.6% (1.8% DD), at a cut off of 0.9 1.5% (0.8% DD) and at a cut off of 0.99 approximately 0.2% (0% DD) of all identifications were false-positive assignments. Based on these results, we decided to upload all phosphopeptides with a PeptideProphet probability score greater than 0.8 into PhosphoPep.

To assess the second type of error, the miss-assignment of the phospho-amino acid in a correctly identified phosphopeptide we used the dCn score computed by Sequest (Eng *et al*, 1994) as described in the Supplementary information and Supplementary Figure S2. We found that a dCn value greater than 0.1 corresponds to >90% certainty in phosphorylation site assignment. Overall, the application of a dCn threshold of 0.1 yielded 10 118 distinct phosphorylation sites (PeptideProphet probability score >0.9) or 12 756 phosphorylation sites (PeptideProphet probability score >0.8). Without any dCn filter PhosphoPep contains 12 596 (PeptideProphet probability score >0.9) or 16 608 phosphorylation sites (PeptideProphet probability score >0.8).

Structural and functional properties of the identified phosphopeptides

We next analyzed the structural and functional properties, namely the distribution and number of phosphorylated residues per phosphopeptide, the molecular functions, and the biological processes and the pathways that are associated with the identified phosphoproteins along with their predicted abundance.

Distribution of phosphorylated amino acids

We found that 78% of the identified phospho-amino acids were phosphorylated on a serine, 19% on a threonine and 3% on a tyrosine. Furthermore, nearly 87% of all peptides were phosphorylated at one site, 10% at two sites and 3% at three sites. These results are slightly different from the so far assumed distribution of phospho-amino acids (Hunter and Sefton, 1980) (89% serine, 10% threonine and 1% tyrosine) and other large-scale data sets (Olsen *et al*, 2006).

Molecular function and biological processes

To derive the molecular functions and biological processes of the identified phosphoproteins, we used 'panther' ontology (PO) (Mi *et al*, 2007). We also investigated whether some molecular functions or biological processes were enriched or depleted in the phosphoprotein data set compared to an external (proteome predicted from the FlyBase (r4.3) sequence database) and an internal reference (proteins identified from the peptide sample before the phosphopeptide enrichment).

For both the molecular function (Figure 1A) and the biological processes (Figure 1B), all possible PO annotations were identified from the phosphoprotein data set. However, for many processes and functions, biases were visible compared to the external reference. Many of these biases can be explained by proteomics workflows, in which low-abundant, small or membrane proteins are often underrepresented (Brunner *et al*, 2007). This is also reflected in the comparison between the internal and external reference. We therefore also contrasted the phosphoprotein data set to the internal reference detecting differences between the two proteomic data sets (Figure 1A and B).

In regards to the molecular functions and biological processes, enrichment for phosphoproteins (compared to the internal reference) involved in regulatory processes was apparent, in particular for kinases, transcription factors, ion channels (Figure 1A) or developmental processes (Figure 1B). In contrast, in the categories metabolism (lyases, isomerases and synthases) or metabolic processes (sulfur, coenzyme, carbohydrate and other metabolism) phosphoproteins were depleted (Figure 1B). The overrepresentation of kinases, transcription factors and ion channels compared to the internal reference is expected as these classes of proteins are known to be highly regulated by protein phosphorylation (Hunter, 2000). In addition, the enrichment of phosphoproteins in developmental processes indicates that these processes are highly regulated by protein phosphorylation as well.

Pathway association and abundance of identified phosphoproteins

We next investigated the depth of phosphoproteome coverage achieved by the data set. Of 118 PO pathways (Mi *et al*, 2007) (from the FlyBase database (r4.3) (Grumbling and Strelets, 2006)) 98 were represented by the phosphoproteome data set. Most of the pathways to which no phosphoprotein could be assigned (15 of the 20) consisted of equal to or less then three proteins, thus reducing the likelihood of their detection.

A comparison of the codon bias distribution (Duret and Mouchiroud, 1999) of the complete predicted *D. melanogaster* proteome (from the FlyBase database (r4.3)) with that of the identified phosphoproteins showed similar curves, indicating that proteins from all levels of abundance were identified (Figure 1C). Overall, these data indicate that the phosphoprotein data set reached a considerable depth of the analysis of the phosphoproteome of Kc167 cells. This finding is further strengthened by the observation that we detected proteins mapping to over 50% of so far ~6200 gene models in *D. melanogaster* Kc167 cells for which a protein was detectable (Brunner *et al*, 2007).

For systems biology-based signaling research, such an in-depth coverage of phosphorylation sites is highly beneficial and strengthens the use of *D. melanogaster* Kc167 cells as a model organism for systems biology.

PhosphoPep—a database and associated utilities for systems biology signaling research

To increase the utility of the phosphopeptide data set described above, we organized the data in a publicly accessible relational database, PhosphoPep, and added functions supporting data mining and meta-analysis. The following sections describe the database and the added functions.

The PhosphoPep database

The consolidated *D. melanogaster* Kc167 cell phosphopeptide data set was uploaded to PhosphoPep, which is publicly

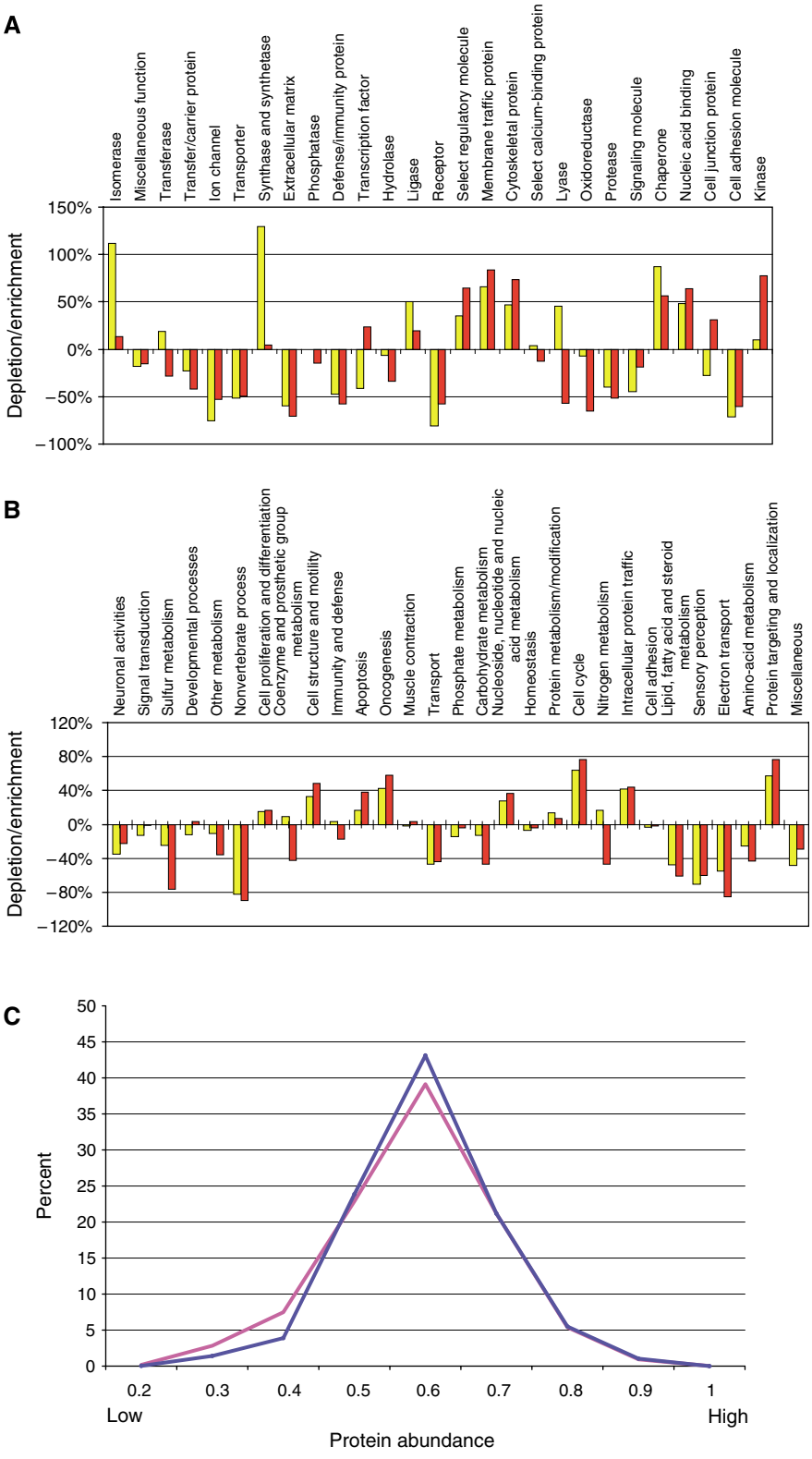


Figure 1 Phosphoprotein properties. (**A**) Depletion/enrichment of molecular functions derived from 'panther' ontology (Mi *et al*, 2007) of the corresponding phosphoproteins (red) and the proteins identified from the separated peptides before enrichment (yellow) relative to the FlyBase database (0%) is shown. (**B**) Depletion/enrichment of biological functions derived from 'panther' ontology (Mi *et al*, 2007) of the corresponding phosphoproteins (red) and the proteins identified from the separated peptides before enrichment (yellow) compared to the FlyBase database (0%) is shown. (**C**) A comparison of the predicted phosphoprotein abundance (blue) with the predicted abundance (Duret and Mouchiroud, 1999) of all proteins of the used FlyBase database (pink) is shown. The scale ranges from 0 (low abundance) to 1 (highly abundant). Proteins for which no molecular function or biological process could be assigned were omitted for (A) and (B). χ^2 test results for (A) and (B) are shown in Supplementary Table II.

accessible (www.phosphopep.org). PhosphoPep is a derivative of the UniPep (Zhang *et al*, 2006) and PeptideAtlas (Desiere *et al*, 2005) databases, connected to the Systems Biology Experiment Analysis Management System (SBEAMS; http://www.sbeams.org), a tool to collect, store and access different data types. All peptides were parsed and loaded into a relational database using SQL (structured query language). Access to the phosphorylation sites and the database is provided by a cgi web interface.

We designed a '*Search interface*' that allows users to query the data using different parameters (Figure 2A). These include searches for single proteins (using the gene ID, protein name, gene symbol, swiss-prot/FlyBase accession number or amino-acid sequence) or searches for a set of proteins (identified proteins search, bulk search and pathway search) at a user-defined PeptideProphet probability score. When a search is executed, a list of all proteins that match the search criteria is shown. Each listing contains a link to view a detailed record for the respective phosphoprotein entry, called 'protein information page'. On that page for each protein in the PhosphoPep database, four different types of information (Figure 2B) are displayed.

The first section, '*Protein info*,' indicates the protein database ID, the protein name (including synonyms), and a protein summary. The '*Protein info*' section also contains three links represented by symbols. The first link queries the protein sequence for potential kinase motives using the Scansite (Obenauer *et al*, 2003) algorithm. The second link displays all KEGG pathways in which the respective phosphoprotein is represented and the third link allows exporting the phosphoprotein to the Cytoscape software (see '*Pathway search, pathway building and data integration*'). Additionally, the '*Protein info*' section categorizes the subcellular location of the proteins into cell surface, secreted, transmembrane or intracellular (Nielsen *et al*, 1997; Krogh *et al*, 2001).

The second section displays the '*Observed phosphopeptides*'. For every protein, all phosphopeptides identified in the data set are shown. To allow the user to assess the quality of the phosphopeptide assignment, the PeptideProphet (Keller *et al*, 2002) score is given as well as the number of tryptic ends, the mass of the phosphopeptide, the dCn value (Eng *et al*, 1994), a link to the MS2 consensus spectrum and a link to export the consensus spectrum ion values for targeted proteomic approaches (See consensus spectra section below). In addition unambiguously assigned phosphorylation sites ($dCn > 0.1$) are highlighted in red and ambiguous sites ($dCn < 0.1$) are highlighted in yellow. Finally, for each phosphopeptide, it is indicated if it maps to a single protein or to several, an important aspect for quantitative targeted proteomics experiments.

In the third section, '*Protein/Peptide sequence*', the whole sequence of the respective phosphoprotein is shown with the identified phosphopeptides, the site(s) of phosphorylation and transmembrane regions, which are highlighted to give a general overview.

In the forth section '*Protein/Peptide map*', the phosphopeptides and the phosphorylation sites are shown according to their position in the protein sequence, thereby giving an indication of the general protein topology.

Pathway search, pathway building and data integration

To build pathways and query the phosphorylation state of the constituent proteins, we placed a protein or proteins contained in PhosphoPep within pathways retrieved from KEGG (Kanehisa *et al*, 2006) ('*Pathway view*', Figure 2A). Proteins can be placed into '*Pathway view*' from both the '*Search interface*' as well as from the '*Protein information*' page of a given protein. '*Pathway view*' also retrieves from PhosphoPep and displays all other identified phosphoproteins of a particular pathway. A '*Bulk search*' option allows placing all of the proteins within their respective pathways. Finally, each pathway can readily be exported, annotated with the relevant phosphoprotein information to 'Cytoscape' (Shannon *et al*, 2003). Cytoscape is a generic visualization tool to integrate and visualize different data types. In this case, the phosphoprotein information contained in PhosphoPep can be complemented with additional data types, such as biomolecular interaction networks, accessible through the web. To facilitate the retrieval of relevant information, 'Cytoscape' is automatically linked to 'Gaggle' (Shannon *et al*, 2006). Gaggle is an informatics-working environment in which information from different web resources can be retrieved and imported into the Cytoscape environment.

Consensus spectra: a searchable fragment ion representation of the phosphoproteome

The analysis of proteomic data sets carries a large computational overhead. This is particularly true for spectra of phosphopeptides, due to their particular fragmentation characteristics and increased peptide search space in database searching. Furthermore, targeted proteomic workflows are emerging in which sets of specific analytes, for example, the phosphorylation sites on proteins constituting a signaling pathway are analyzed under varying cellular conditions (Domon and Aebersold, 2006; Wolf-Yadlin *et al*, 2007). To support the rapid (Supplementary Figure S3A), highly sensitive (Supplementary Figure S3B and Supplementary Table I) and reliable identification of phosphopeptides in future experiments and targeted mass spectrometry by MRM, we built a searchable consensus spectral library of most identified peptides in PhosphoPep, and made them available in a searchable and downloadable form (Figure 2A).

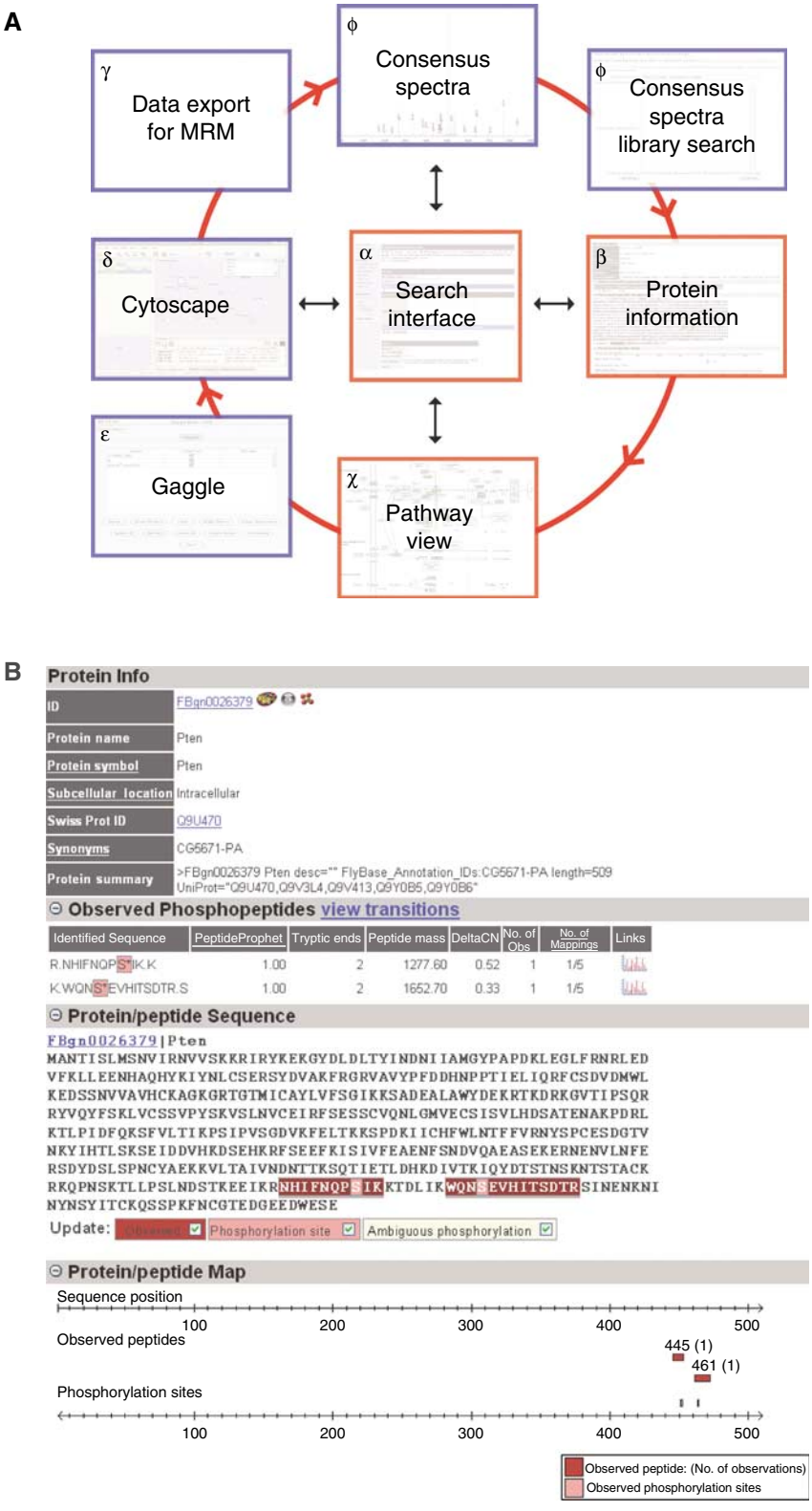
By using the spectral matching search tool SpectraST (Lam *et al*, 2007), both as a web interface in PhosphoPep, and as a stand-alone application released as part of the TPP suite of software (Keller *et al*, 2005), spectra can be searched against the phosphopeptide consensus library (see also Supplementary information).

To support MRM-based targeted proteomic experiments, we provide a download function for consensus spectra representing a specific phosphopeptide (Domon and Aebersold, 2006; Picotti *et al*, 2007; Stahl-Zeng *et al*, 2007; Wolf-Yadlin *et al*, 2007). Such spectra can be a useful start for the optimization of precursor ion to fragment ion transitions for MRM experiments, for example by performing MRM-triggered MS2 experiments searchable against the phosphopeptide consensus spectra library (Lam *et al*, 2007).

Overall, these functionalities are highly useful for researchers focused on single proteins and especially for systems biologists who wish to conduct iterative cycles of experimentation and analysis on differentially perturbed cell states.

Assessment of the identified phosphoproteome

There is no 'gold standard' phosphoproteome data set that could be used to assess the extent to which the Kc167 phosphoproteome has been mapped out. To further investigate the achieved phosphoproteome coverage, we compared the phosphorylation sites from our data set that matched the highly conserved (Oldham *et al*, 2000; Garofalo, 2002) and clinically relevant insulin/TOR pathway with the already known sites in *D. melanogaster*.



The results are shown in Figure 3. Of the 15 pathway members, 6 (dAKT1, CHICO, dFOXO, dTSC2, dS6K and d4E-BP) have been known to be phosphorylated in *D. melanogaster*. In our data set, we found all 15 members to be phosphorylated. Furthermore, for the proteins for which phosphorylation sites have been published previously, we were able to identify multiple new sites. The most prominent example is the insulin receptor substrate, CHICO, for which the number of known phosphorylation sites increased from 2 to 20. For dFOXO and d4E-BP, we identified all, and for dS6K, we identified one already known phosphorylation sites. For dAKT1, CHICO and dTSC2, the already known sites were not found in our experiments, indicating that in spite of the high number of sites identified in this study the KC167 phosphoproteome is likely not complete at this time (see Supplementary information).

This example shows that we have reached a depth in phosphoproteome coverage that is suitable for systems biology signaling research in *D. melanogaster* and, due to a myriad of orthologous sites (Reiter *et al*, 2001), also in other species.

Materials and methods

All chemicals, if not otherwise mentioned, were bought with the highest available purity from Sigma-Aldrich, Taufkirchen, Germany.

Cell culture, lysis and protein digestion

D. melanogaster Kc167 cells were grown in Schneiders *Drosophila* medium (Invitrogen) supplemented with 10% fetal calf serum, 100 U penicillin (Invitrogen) and 100 μg/ml streptomycin (Invitrogen, Auckland, New Zealand) in an incubator at 25°C. To increase the number of mapped phosphorylation sites, different batches of cells were pooled. Cells were either grown in rich medium, or were serum-starved, or were treated for 30 min with 100 nM Rapamycin (LClabs, Woburn, MA, USA) in rich medium, or were treated for 30 min with 100 nM insulin (serum starved), or were treated for 30 min with 100 nM Calyculin A

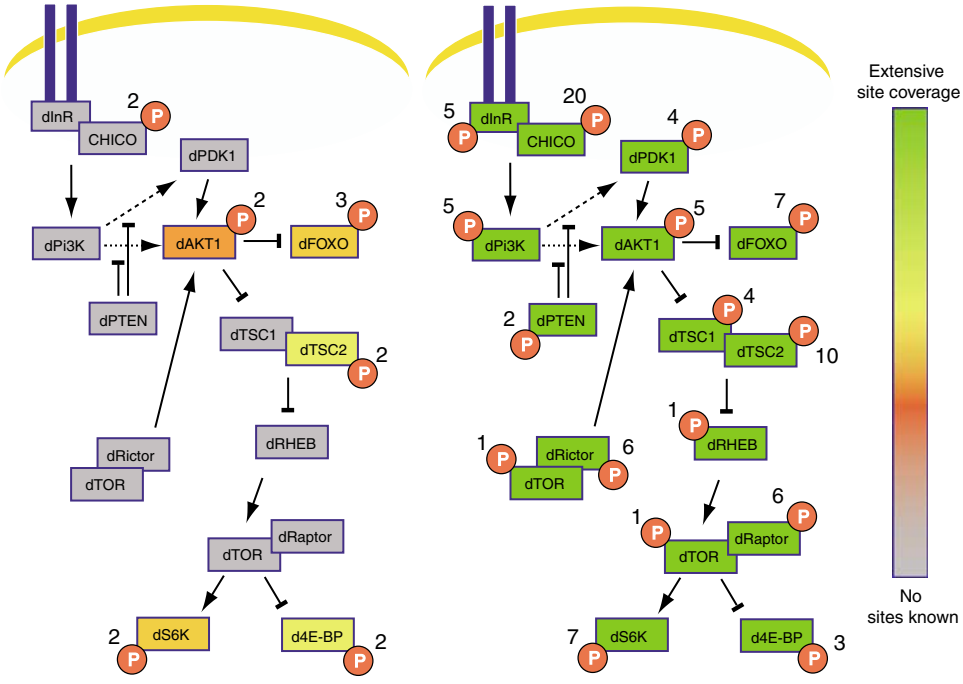


Figure 3 Proteins involved in the target of rapamycin (TOR) and insulin signaling. To demonstrate the usefulness of our database, we compared the already known phosphoproteins (left) with our identified phosphoproteins (right). As can be seen, compared to the literature in which only 6 out of the 15 proteins were found to be phosphorylated, we extended the phosphorylation map to all proteins of the pathway (Hay and Sonenberg, 2004; Oldham and Hafen, 2003) (phosphorylation sites are depicted by the P in a red circle, the number assigns the number of distinct phosphorylation sites). The number of identified phosphorylations ranged from 1 to 20 (CHICO). Peptides with $P > 0.8$ and a defined phosphorylation site (dCn>0.1) were considered.

Figure 2 (**A**) Design of the PhosphoPep database. By using the '*Search interface*' (α) PhosphoPep can be interrogated for single proteins, a set of proteins or pathways. For each protein, several types of information including the observed phosphopeptides is shown in the '*Protein information*' page (see panel B and β. Single proteins or a set of proteins can be placed into their pathways (χ). From this '*Pathway view*' all phosphoproteins can be exported to Cytoscape (Shannon *et al*, 2003) (δ). This software tool allows integrating data from PhosphoPep with external data such as protein–protein interaction networks (ε). For most phosphopeptides, consensus MS2 spectra (φ) are given which can be exported for targeted proteomics experiments such as multiple reaction monitoring (Domon and Aebersold, 2006) (γ). As we supply an online spectral matching search tool, results generated by such experiments can be validated using PhosphoPep. (**B**) Representative output of the PhosphoPep database. The PhosphoPep (www.phosphopep.org) database contains more than 10 000 phosphorylation sites from nearly 3500 gene models and nearly 5800 phosphoproteins derived from the FlyBase (Grumbling and Strelets, 2006) nonredundant database (r4.3). For each phosphoprotein, the phosphopeptide sequence, the protein annotation and the predicted subcellular location is shown. Furthermore, additional information for each phosphopeptide is given: The probability, the number of tryptic ends, the dCn value, the mass, how often it was observed and to how many gene models and transcripts it maps. The phosphopeptides are represented in both the protein sequence and in a graphical representation, the protein map. Finally, a link to the '*Pathway view*', to the 'Cytoscape export' function and to http://scansite.mit.edu/ (Obenauer *et al*, 2003) is given as represented by the three symbols besides the FlyBase gene entry.

(rich medium). Then the cells were washed with ice-cold phosphate-buffered saline and resuspended in ice-cold lysis buffer containing 10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol and a protease inhibitor mix (Roche, Basel, Switzerland). To preserve protein phosphorylation, several phosphatase inhibitors were added to a final concentration of 20 nM calyculin A, 200 nM okadaic acid, 4.8 μm cypermethrin (all bought from Merck KGaA, Darmstadt, Germany), 2 mM vanadate, 10 mM sodium pyrophosphate, 10 mM NaF and 5 mM EDTA. After 10 min incubation on ice, cells were lysed by douncing. Cell debris and nuclei were removed by centrifugation for 10 min at 4°C using 5500 *g*. Then the cytoplasmic and membrane fraction were separated by ultracentrifugation at 100 000 *g* for 60 min at 4°C. The proteins of the cytosolic fraction (supernatant) were subjected to acetone precipitation. The protein pellets were resolubilized in 3 mM EDTA, 20 mM Tris–HCl, pH 8.3, and 8 M urea. The disulfide bonds of the proteins were reduced with tris (2-carboxyethyl) phosphine at a final concentration of 12.5 mM at 37°C for 1 h. The produced free thiols were alkylated with 40 mM iodoacetamide at room temperature for 1 h. The solution was diluted with 20 mM Tris–HCl (pH 8.3) to a final concentration of 1.0 M urea and digested with sequencing-grade modified trypsin (Promega, Madison, WI) at 20 μg per mg of protein overnight at 37°C. Peptides were desalted on a C18 Sep-Pak cartridge (Waters, Milford, MA) and dried in a speedvac. Finally, 280 mg of peptides were separated by IEF using FFE.

Peptide separation

FFE-Weber reagent basic kit (Prolyte 1, Prolyte 2, Prolyte 3 and Prolyte 4–7 and *pI* markers) were purchased from FFE-Weber Inc. (now BD-Diagnostics, NJ, USA). Hydroxyisobutyric acid, DL-2-aminobutyric acid, nicotinamide, glycyl-glycine and ethanolamine were purchased from Sigma-Aldrich (Steinheim, Germany), AMPSO and HEPES from Roth (Karslruhe, Germany) and TAPS from ACROS (NJ, USA).

Free-flow electrophoresis

IEF was performed using an FFE instrument, type prometheus from FFE Weber Inc. (now BD-Diagnostics, PAS). For a detailed description of the experimental procedure, please see Malmstrom *et al* (2006). The digested peptides were diluted in separation media containing 8 M Urea and 250 mM Mannitol and 20% ProLyte solution at a concentration of 10 mg/ml. This sample was loaded continuously for 1 h at 1 ml/h. Total collection time was 24 h and the volume of each collected fraction was about 25–50 ml. A Thermo Orion needle tip micro pH electrode (Thermo Electron Corporation, Beverly, MA) was used to measure the pH value of each fraction. Peptides from the FFE fractions 18–60 were purified on a C18 Sep-Pak cartridge (Waters Corporation, Milford, MA, USA).

After purification, the eluted peptides where split into three fractions (one fraction was used for phosphopeptide isolation using PAC, one for TiO_2 and one for IMAC) and dried down and used for phosphopeptide isolation.

Phosphopeptide isolation

The phosphopeptides were isolated using PAC, IMAC and TiO_2 as described by Bodenmiller *et al* (2007a, b).

MS analysis

The majority of samples were analyzed on a hybrid LTQ-Orbitrap mass spectrometer (ThermoFischer Scientific, Bremen, Germany) interfaced with a nanoelectrospray ion source. Chromatographic separation of peptides was achieved on an Eksigent nano LC system (Eksigent Technologies, Dublin, CA, USA), equipped with a 11 cm fused silica emitter, 75 μm inner diameter (BGB Analytik, Böckten, Switzerland), packed in-house with a Magic C18 AQ 3 μm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded from a cooled (4°C) Spark Holland auto sampler and separated using ACN/water solvent system containing 0.1% formic acid with a flow rate of 200 nl/min. Peptide mixtures were separated with a gradient from 3 to 35% ACN in 90 min.

Up to five data-dependent MS2 spectra were acquired in the linear ion trap for each FT-MS spectral acquisition range, the latter acquired at 60 000 FWHM nominal resolution settings with an overall cycle time of approximately 1 s. Charge state screening was employed to select for ions with two charges and rejecting ion with one or undetermined charge state. The same sample was injected a second time with the same setting besides the charge state screening, which was then set to three and higher (excluding 1, 2 and undetermined charge state). For injection control, the automatic gain control was set to 5e5 and 1e4 for full FTMS and linear ion trap MS2, respectively. The instrument was calibrated externally according to manufacturers instructions. The samples were acquired using internal lock mass calibration on m/z 429.088735 and 445.120025.

For some pre-experiments and re-measurements, a hybrid LTQ-FTICR mass spectrometer (Thermo, San Jose, CA) interfaced with a nanoelectrospray ion source was used. Chromatographic separation of peptides was achieved on an Agilent Series 1100 LC system (Agilent Technologies, Waldbronn, Germany), equipped with an 11 cm fused silica emitter, 150 μm inner diameter (BGB Analytik, Böckten, Switzerland), packed in-house with a Magic C18 AQ 5 μm resin (Michrom BioResources, Auburn, CA, USA). Peptides were loaded from a cooled (4°C) Agilent auto sampler and separated with a linear gradient of ACN/water, containing 0.15% formic acid, with a flow rate of 1.2 μl/min. Peptide mixtures were separated with a gradient from 2 to 30% ACN in 90 min. Three MS2 spectra were acquired in the linear ion trap per each FT-MS scan, the latter acquired at 100 000 FWHM nominal resolution settings with an overall cycle time of approximately 1 s. Charge state screening was employed to select for ions with at least two charges and rejecting ions with undetermined charge state. For each peptide sample, a standard data-dependent acquisition method on the three most intense ions per MS-scan was used and a threshold of 200 ion counts was used for triggering an MS2 attempt.

Data analysis

The MS2 data were searched against the FlyBase (Release 4.3) (Grumbling and Strelets, 2006) nonredundant database containing 19 465 proteins using SORCERER-SEQUEST(TM) v3.0.3, which was run on the SageN Sorcerer2 (Thermo Electron, San Jose, CA, USA). For the *in silico* digest, trypsin was defined as protease, cleaving after K and R (if followed by P the cleavage was not allowed). Two missed cleavages and one nontryptic terminus were allowed for the peptides that had a maximum mass of 6000 Da. The precursor ion tolerance was set to 5 p.p.m. and the fragment ion tolerance was set to 0.8 Da. Before searching using Sequest, the neutral loss peaks were removed and indicated as described previously (Bodenmiller *et al*, 2007b). Then data were searched (for IMAC and TiO_2) allowing phosphorylation (+79.9663 Da) of serine, threonine and tyrosine as a variable modification and carboxyamidomethylation of cysteine (+57.0214 Da) residues as a fixed modification. For PAC, in addition to the just mentioned modifications, the methylation (+14.0156 Da) of all carboxylate groups as a static modification was also defined. In the end, the search results obtained by Sequest were subjected to statistical filtering using PeptideProphet (V3.0) (Keller *et al*, 2002) and ProteinProphet(V3.0) (Keller *et al*, 2002). Proteins identified that way were used for the analysis in Figure 1A and B. The proteins were queried using the 'panther classification system' (Mi *et al*, 2007) http://www.pantherdb.org/ by using the batch search. FlyBase (r4.3) was used as reference (Grumbling and Strelets, 2006) (0% depletion/enrichment). Significance of the biases was determined using a χ^2 test.

If the same analysis is carried out using all proteins from PhosphoPep (PeptideProphet $P>0.9$; in the construction of PhosphoPep each peptide identified using PeptideProphet (with $P>0.8$) was mapped against each possible protein derived from the FlyBase database (r4.3)) basically the same biases (with similar significances) as shown in Figure 1A and B were visible if queried using the 'panther classification system' (Mi *et al*, 2007) http://www.pantherdb.org/ by using the batch search.

To determine the certainty of the assignment of a phosphate group to a hydroxyamino acid, the dCn was used as it has been shown recently that it directly correlates with the certainty of phosphorylation

site assignment (Beausoleil *et al*, 2006). To estimate a dCn cut-off to consider a site well assigned (>90% certainty), the following assumption was made: as many of our phosphopeptides were sequenced more than once, an uncertainty in the phosphorylation site assignment will result in several 'versions' of a phosphopeptide, namely that the amino-acid sequence is identical but that the site of phosphorylation is different. After consolidation of the phosphopeptides using the computer program 'Phosphogigolo' (Bodenmiller *et al*, 2007b), we computed for a given dCn value the percentage of peptides that have the same amino-acid sequence (ignoring the phosphate group and the fact that a peptide can exist in two phosphorylation states with a high certainty of phosphorylation site assignment). Finally, the 'percent' ambiguous was computed by 2 × (percentage of redundant 'stripped' peptide entries) (Supplementary Figure S2).

Decoy database search strategy

The decoy database was designed in the following way: FlyBase database (r4.3) was *in silico* digested using trypsin. Then the amino acids of these peptides were scrambled except for the c-terminal lysine or arginine. Proteins were reconstructed by the scrambled peptides and the label Rev_ was added to the protein names. This resulted in a protein database with half the proteins being original and the other half concatenated from the scrambled peptides. This decoy protein database gives rise to peptides with approximately the same length distribution as the original database. The false-positive rate was estimated as described by Elias and Gygi (2007).

Creation of the consensus spectral library

The PeptideProphet-processed SEQUEST search result from all LC-MS/MS runs performed on either a LTQ-Orbitrap or LTQ-FT mass spectrometer was screened for spectra that are identified above a probability threshold of 0.9 and a dCn value of 0.1. A total of over 170 000 confidently identified spectra mapping to about 33 000 distinct peptide ions were collected. The spectra identified to the same peptide ion (replicates) were then grouped, and collapsed into a single consensus spectrum. The corresponding peaks in the replicates are m/z-aligned, and only peaks that are present in a majority of the replicates are included in the consensus, making no assumption about the possible identities of the fragments. The consensus intensity of each peak is calculated as the average of the peak intensities in the replicates, weighted by a measure of the varying spectral quality of the replicates. For peptide ions for which only a single observation is made, the raw spectrum is included after simple noise reduction. All the resulting spectra are then annotated and indexed for fast searching. The details of the consensus spectrum building algorithm, as well as the software to perform it, will be provided in a future publication.

For the comparison of SpectraST and the Sequest database search algorithms in regards of search speed, two test data sets were used. For the LTQ-Orbitrap, a randomly chosen data set with 10 166 spectra and for the LTQ data set randomly chosen 27 556 spectra were used. SpectraST was run on a single processor while SORCERER-SEQUEST(TM) v3.0.3, which was run on the SageN Sorcerer2. For the database search, a 5 p.p.m. parental mass tolerance was used for the Orbitrap data set and 3 Da for the LTQ data set.

The sensitivity and error curves were determined using the PeptideProphet (Keller *et al*, 2002) (Supplementary Figure S3).

For the comparison of between SpectraST and the Sequest database search algorithms in regards of achieved sensitivity/identifications three randomly chosen test data set were used for each IMAC, TiO_2 and PAC. After database search (SpectraST was run on a single processor, SORCERER-SEQUEST(TM) v3.0.3 was run on the SageN Sorcerer2) the sensitivity and error curves were determined using the PeptideProphet (Keller *et al*, 2002) (Supplementary Table I).

Determination of protein abundance based on codon bias

As described previously (Duret and Mouchiroud, 1999) for all proteins, the abundance, ranging from 1 (highly abundant) to 0 (very low abundant), was calculated (Figure 1C).

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank P Picotti for proof reading of the manuscript and the whole FGCZ team for the support and fruitful discussions. We also thank Massimo Merlini for advice on the statistical analysis of our data. We also thank Hui Zhang and Pat Moss for the development of the Unipep database. This project has been funded in part by ETH Zurich, and with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, under contract No. N01-HV-28179 and by the Center for Model Organism Proteomics of SystemsX.ch the Swiss initiative for systems biology. Work at the FGCZ has been supported by the University Research Priority Program Systems Biology and Functional Genomics of the University of Zurich. AS and RA were supported in part by a grant from F Hoffmann-La Roche Ltd (Basel, Switzerland) provided to the Competence Center for Systems Physiology and Metabolic Disease. JM is the recipient of a postdoctoral fellowship from the Swedish Society for Medical Research (SSMF). OR was supported by fellowships of the Roche Research Foundation and the Deutsche Forschungsgemeinschaft (DFG). BG is supported by Bonizzi-Theler Foundation. BB is the recipient of a fellowship by the Boehringer Ingelheim Fonds.

Data availability

All data presented in this study are available from PhosphoPep (www.phosphopep.org).

Author contributions

BB coordinated the project, conducted most of the experimental work, data analysis, and was also responsible for ideas and concepts and wrote the core of the manuscript. JM is responsible for FFE separation of peptides, idea and concept and wrote part of the core of the manuscript. BG did the LC-MS/MS measurements on the LTQ-Orbitrap, performed sequence database searches and data compilation. DC designed and programmed the PhosphoPep database. HL developed SpectraST and produced and validated the consensus spectral library, wrote part of the manuscript. AS contributed data measured on a LTQ-FT-ICR. LNM contributed to bioinformatics analysis of data. OR contributed to bioinformatics analysis of data, wrote part of the core of the manuscript. PTS contributed to bioinformatics analysis of data. PP contributed to bioinformatics analysis of data. CP contributed to bioinformatics analysis of data. HKL conducted part of LC-MS/MS measurements. RA carried senior authorship responsibility, coordinated the project, wrote the manuscript and was responsible for ideas and concept.

References

Aebersold R, Goodlett DR (2001) Mass spectrometry in proteomics. *Chem Rev* **101:** 269–295

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422:** 198–207

Andersson L, Porath J (1986) Isolation of phosphoproteins by immobilized metal (Fe^{3+}) affinity-chromatography. *Anal Biochem* **154:** 250–254

Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li JX, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA* **101:** 12130–12135

Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24:** 1285–1292

Bier E (2005) Drosophila, the golden bug, emerges as a tool for human genetics. *Nat Rev Genet* **6:** 9–23

Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007a) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Meth* **4:** 231–237

Bodenmiller B, Mueller LN, Pedrioli PG, Pflieger D, Junger MA, Eng JK, Aebersold R, Tao WA (2007b) An integrated chemical, mass spectrometric and computational strategy for (quantitative) phosphoproteomics: application to *Drosophila melanogaster* Kc167 cells. *Mol Biosyst* **3:** 275–286

Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de Lichtenberg U, Rinner O, Lee H, Pedrioli PG, Malmstrom J, Koehler K, Schrimpf S, Krijgsveld J, Kregenow F, Heck AJ, Hafen E, Schlapbach R *et al* (2007) A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat Biotechnol* **25:** 576–583

Desiere F, Deutsch EW, Nesvizhskii AI, Mallick P, King NL, Eng JK, Aderem A, Boyle R, Brunner E, Donohoe S, Fausto N, Hafen E, Hood L, Katze MG, Kennedy KA, Kregenow F, Lee H, Lin B, Martin D, Ranish JA *et al* (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol* **6:** R9

Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* **312:** 212–217

Duret L, Mouchiroud D (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci USA* **96:** 4482–4487

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Meth* **4:** 207–214

Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectr* **5:** 976–989

Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y *et al* (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3:** 89

Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* **20:** 301–305

Fields S, Song O (1989) A novel genetic system to detect protein–protein interactions. *Nature* **340:** 245–246

Garofalo RS (2002) Genetic analysis of insulin signaling in *Drosophila*. *Trends Endocrinol Metab* **13:** 156–162

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T *et al* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141–147

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* **100:** 6940–6945

Gingras AC, Aebersold R, Raught B (2005) Advances in protein complex analysis using mass spectrometry. *J Physiol* **563:** 11–21

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M *et al* (2003) A protein interaction map of *Drosophila melanogaster*. *Science* **302:** 1727–1736

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

Hay N, Sonenberg N (2004) Upstream and downstream of mTOR. *Genes Dev* **18:** 1926–1945

Hunter T (2000) Signaling—2000 and beyond. *Cell* **100:** 113–127

Hunter T, Sefton BM (1980) Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci USA* **77:** 1311–1315

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292:** 929–934

Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409:** 533–538

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34:** D354–D357

Keller A, Eng J, Zhang N, Li XJ, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **1:** 2005.0017

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74:** 5383–5392

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305:** 567–580

Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **7:** 655–667

Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* **4:** 873–886

Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet* **21:** 20–24

Malmstrom J, Lee H, Nesvizhskii AI, Shteynberg D, Mohanty S, Brunner E, Ye M, Weber G, Eckerskorn C, Aebersold R (2006) Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. *J Proteome Res* **5:** 2241–2249

Mi H, Guo N, Kejariwal A, Thomas PD (2007) PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res* **35:** D247–D252

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8:** 581–599

Obenauer JC, Cantley LC, Yaffe MB (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31:** 3635–3641

Oldham S, Bohni R, Stocker H, Brogiolo W, Hafen E (2000) Genetic control of size in *Drosophila*. *Philos Trans R Soc Lond* **355:** 945–952

Oldham S, Hafen E (2003) Insulin/IGF and target of rapamycin signaling: a TOR de force in growth control. *Trends Cell Biol* **13:** 79–85

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127:** 635–648

Picotti P, Aebersold R, Domon B (2007) The Implications of proteolytic background for shotgun proteomics. *Mol Cell Proteomics* **6:** 1589–1598

Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **76:** 3935–3943

Reinders J, Sickmann A (2005) State-of-the-art in phosphoproteomics. *Proteomics* **5:** 4052–4061

Reiter LT, Potocki L, Chien S, Gribskov M, Bier E (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* **11:** 1114–1125

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP,

Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17:** 1030–1032

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:** 467–470

Schwartz D, Gygi SP (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **23:** 1391–1398

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498–2504

Shannon PT, Reiss DJ, Bonneau R, Baliga NS (2006) The gaggle: an open-source software system for integrating bioinformatics software and data sources. *BMC bioinformatics* **7:** 176

Stahl-Zeng J, Lange V, Ossola R, Aebersold R, Domon B (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 28 August 2007 [E-pub ahead of print]

Tao WA, Wollscheid B, O'Brien R, Eng JK, Li XJ, Bodenmiller B, Watts JD, Hood L, Aebersold R (2005) Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat Meth* **2:** 591–598

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364–2368

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403:** 623–627

Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci USA* **104:** 5860–5865

Zhang H, Loriaux P, Eng J, Campbell D, Keller A, Moss P, Bonneau R, Zhang N, Zhou Y, Wollscheid B, Cooke K, Yi EC, Lee H, Peskind ER, Zhang J, Smith RD, Aebersold R (2006) UniPep, a database for human N-linked glycosites: a resource for biomarker discovery. *Genome Biol* **7:** R73

Zhou HL, Watts JD, Aebersold R (2001) A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* **19:** 375–378