Published in final edited form as:

J Proteome Res. 2008 April ; 7(4): 1675–1682. doi:10.1021/pr700696a.

Phosphoproteome Analysis of *Drosophila melanogaster* Embryos

Bo Zhai, Judit Villén, Sean A. Beausoleil, Julian Mintseris, and Steven P. Gygi*

Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115

Abstract

Protein phosphorylation is a key regulatory event in most cellular processes and development. Mass spectrometry-based proteomics provides a framework for the large-scale identification and characterization of phosphorylation sites. Here, we used a well-established phosphopeptide enrichment and identification strategy including the combination of strong cation exchange chromatography, immobilized metal affinity chromatography, and high-accuracy mass spectrometry instrumentation to study phosphorylation in developing *Drosophila* embryos. In total, 13 720 different phosphorylation sites were discovered from 2702 proteins with an estimated false-discovery rate (FDR) of 0.63% at the peptide level. Because of the large size of the data set, both novel and known phosphorylation motifs were extracted using the Motif-X algorithm, including those representative of potential ordered phosphorylation events.

Keywords

phosphoproteome; *Drosophila*; embryogenesis; SCX; IMAC; LC–MS/MS; signal transduction

Introduction

Drosophila melanogaster is one of the most studied organisms in all of biological research, particularly in developmental biology and genetics. Reasons include (i) its ease of growth in the laboratory, (ii) its relatively small size, (iii) mature larvae show polytene chromosomes in the salivary glands, (iv) *Drosophila* chromosomes consist of only three autosomes and one sex chromosome, and (v) its compact genome sequence was published in 2000.^{1,2} A vast array of cellular processes is also involved in the development of *Drosophila* embryos including cellularization, cell migration, cell division, apoptosis, and so forth.^{3–6} Phosphorylation has been shown to play a key role in each of these processes. A large-scale description of the phosphorylation state of *Drosophila* embryo will allow a deeper understanding of signal transduction pathways during development, and provide a defined starting point for future research.

The ability to catalog the precise sites of phosphorylation on a scale of thousands has been accomplished due primarily to the combination of three factors: (i) high mass accuracy precursor ion determination,⁷ (ii) optimized enrichment protocols for phosphopeptide

*Corresponding author: Department of Cell Biology, Harvard Medical School, 240 Longwood Ave., Boston, MA 02115. Phone, (617) 432-3155; fax, (617) 432-1144; steven_gygi@hms.harvard.edu.

Supporting Information Available: Complete list of phosphopeptides identified from *Drosophila melanogaster* w¹¹⁸ embryos (SI Table 1). Motifs extracted from this data set using the Motif-X algorithm (SI Table 2). Distribution of phosphorylated and nonphosphorylated peptides identified in each SCX fraction. (SI Figure 1). Solution charge state is a useful filtering criterion for SCX separations (SI Figure 2). Overlap of (A) phosphorylation sites and (B) phosphoproteins identified in the Bodenmiller et al. data set and this study (SI Figure 3). This material is available free of charge via the Internet at <http://pubs.acs.org>.

isolation,^{8–10} and (iii) enabling software for false-positive estimation and site localization.^{11,12} With these approaches, several large-scale studies have been reported.^{12–14} Recently, a trio of papers from Aebersold and colleagues has been published examining the phosphoproteome of *D. melanogaster*. They first performed a comparison of enrichment methods (IMAC, TiO₂, phosphoramidite chemistry).¹⁰ The phosphopeptides used in this comparison were derived from *Drosophila* Kc167 cells, and 887 different sites were reported by combining all methods. A second paper described in more detail the phosphoramidite chemistry optimization with 571 reported sites.¹⁵ A third report described 10 118 sites from Kc167 cells using a variety of enrichment techniques and peptide isoelectric focusing by free-flow electrophoresis.¹⁶

One long-term goal of our laboratory is the generation of phosphorylation databases for many model organisms and cell lines as a powerful tool to study phosphorylation in an evolutionary context. These model organisms and cell lines include *Saccharomyces cerevisiae*¹³ and *Schizosaccharomyces pombe*,³⁹ worm, fly, mouse,¹⁴ rat, and human cancer cell lines.^{12,17} In the current study, we analyzed phosphorylation occurring during *Drosophila* embryonic development. From 24 LC–MS/MS analyses, we identified 13 720 unique phosphorylation sites from 2702 proteins. This data set contained a defined false-discovery rate (0.63%) at the peptide level and a probability assessment for correct site localization.

Methods

Fly Embryo Lysate Preparation

The 0–24 h old *w¹¹¹⁸* *D. melanogaster* embryos were collected in a population cage, dechorionated with 50% bleach, washed, dounce-homogenized in lysis buffer [50 mM Tris (pH 8.1)/75 mM NaCl/8 M urea/10 mM sodium pyrophosphate/1 mM sodium fluoride/1 mM β -glycerophosphate/1 mM sodium orthovanadate/1 tablet complete Mini protease inhibitor mixture (Roche) per 10 mL], and further lysed by sonication. Supernatant was collected by centrifugation at 13 000 rpm at 4 °C for 15 min. Protein concentration was measured by Bio-Rad Protein Assay (Bio-Rad).

In-Solution Trypsin Digestion

Disulfide bonds were reduced with 2.5 mM DTT for 25 min at 60 °C, and then the free sulfhydryl groups were alkylated with 7 mM iodoacetamide at room temperature in the dark for 30 min. The alkylation reaction was quenched by addition of DTT to 2.5 mM and incubation for 15 min at room temperature. Lysate was diluted 8-fold into 25 mM Tris (pH 8.1) and 1 mM CaCl₂, and sequencing grade trypsin (Promega, Madison, WI) was added (~5 ng/ μ L). Following a 15-h incubation at 37 °C, TFA was added to 0.4% to stop digestion, and pH was verified at ~2. The digest was centrifuged at 3200 rpm to remove insoluble material and then desalted with a 500-mg tC18 SepPak cartridge (Waters, Milford, MA). Eluted peptides were lyophilized and stored at –20 °C.

Strong Cation Exchange (SCX) Chromatography

Ten milligrams of peptides was dissolved in 400 μ L of SCX buffer A (5 mM KH₂PO₄, pH 2.65/30% acetonitrile). Preparative separations were carried out on a 9.4 $\times$ 200 mm column packed with polysulfoethyl aspartamide (PolyLC, Columbia, MD) material (5- μ m particle size; 300-Å pore), using a Surveyor pump operating at 2 mL/min and a PDA detector (Thermo Fisher, San Jose, CA). Three minutes of isocratic buffer A were followed by a linear gradient from 0% to 25% buffer B (5 mM KH₂PO₄, pH 2.65/30% acetonitrile/350 mM KCl) over 35 min and then several washing steps with 100% buffer B and 100% buffer C (50 mM KH₂PO₄, pH 7.5/500 mM KCl). A total of 12 fractions (~4-min intervals) were

collected. All fractions were lyophilized and desalted with 100-mg tC18 SepPak cartridges (Waters, Milford, MA). Eluted peptides were lyophilized and stored at -20°C .

Immobilized Metal-Affinity Chromatography (IMAC)

Each SCX fraction sample was dissolved in 100 μL of wash/equilibrate buffer (25 mM formic acid/40% acetonitrile) to which 15 μL of pre-equilibrated PHOS-Select Iron Affinity Gel (Sigma) slurry (liquid/resin = 1:1) were added. After 60-min incubation at room temperature with vigorous shaking, the supernatant was removed. The resin with phosphopeptides was then washed three times with 200 μL of wash/equilibrate buffer. Phosphopeptides were eluted three times with 70 μL of 50 mM $\text{KH}_2\text{PO}_4/\text{NH}_3$, pH 10.0, after incubating 5 min at room temperature. Elutes were acidified with 20 μL of 5% formic acid/50% acetonitrile, lyophilized, and afterward desalted with C₁₈ Empore Disks (3M, Minneapolis, MN) using StageTips.¹⁸

Mass Spectrometry

LC-MS/MS analyses were conducted in an LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher, San Jose, CA). Enriched phosphopeptides were reconstituted in 10 μL of 5% acetonitrile/5% formic acid. A total of 0.5 μL of peptide mixture was loaded (7 min) by a Famos autosampler (LC Packings, San Francisco, CA) onto a 125 μm (i.d.) $\times$ 18 cm fused silica microcapillary column in-house-packed with C₁₈ reverse-phase resin (Magic C18AQ; 5- μm particles; 200- $\text{\AA}$ pore size; Michrom Bioresources, Auburn, CA), and separated with an Agilent 1100 series binary pump with in-line flow splitter across a 35-min linear gradient ranging from 6% to 28% acetonitrile in 0.125% formic acid. The LTQ-Orbitrap was operated in the data-dependent mode using the TOP10 strategy.⁷ For each cycle, one full MS scan [375–1800 m/z ; acquired in the orbitrap at 6×10^4 resolution setting and automatic gain control (AGC) target of 10^6] was followed by 10 data-dependent MS/MS spectra (AGC target, 5000; threshold 3000) in the linear ion trap from the 10 most abundant ions. Selected ions were dynamically excluded for 30 s. Charge-state screening was used to reject singly charged ions. Duplicate runs were performed for each SCX fraction sample.

Database Search, Data Filtering, and Site Localization

MS/MS spectra collected from the 24 runs were searched using the Sequest algorithm with the target-decoy database searching strategy¹¹ against a composite database containing the *D. melanogaster* protein database combining euchromatic (version 4.3)¹⁹ and heterochromatic (version 3.1)²⁰ sequences and their reversed complements. Parameters included tryptic specificity, a mass tolerance of ± 100 ppm, up to 3 miscleavage sites, a static modification of 57.0215 Da (carboxyamidomethylation) on cysteine, and dynamic modifications of 79.9663 Da (phosphorylation) on serine, threonine, and tyrosine and 15.9949 Da (oxidation) on methionine. Up to six phosphorylation sites were allowed per peptide. Results were analyzed as described^{12–14,17} including determining scoring and mass tolerance thresholds using decoy matches as a guide. The final data set (all 24 analyses) contained 36 203 phosphopeptides with an estimated 0.63% false-discovery rate (229 decoy matches). All matched phosphopeptides and corresponding spectra are provided in Supporting Information Table 1. The probability of correct phosphorylation site localization was determined for every site in each peptide using the Ascore algorithm.¹² A mass window setting of 100 m/z units and a fragment ion tolerance of ± 0.6 m/z units were used. Sites with Ascores ≥ 13 ($P \leq 0.05$) were considered confidently localized. Counting unique sites was complicated by the fact that some phosphopeptides contained sites that were not localized with high certainty (Ascore < 13 ; $P > 0.05$). For peptides with Ascore < 13 , we were careful to never allow an ambiguous site to count for more than one site, regardless of the number of MS/MS spectra or potential site localizations for this peptide. A conservative approach as

well was applied such that different charge states, oxidized methionines, miscleaved versions, and ragged ends did not add identifications to our nonredundant numbers.

Classification of Phosphorylation Sites by Kinase Specificities

Centered 13-mer sequences were assigned to general motif classes (Acidophilic, Basophilic, Proline-directed, or Others), following sequential assignment as described.¹⁴

Motif Analysis

Phosphopeptide sequences were submitted to the Motif-X algorithm (motif-x.med.harvard.edu).²¹ The *D. melanogaster* protein database was used as a background. Only those sites with Ascore values of at least 13 were used. For single phosphorylation motif, sequences were centered on each phosphorylation site and extended to 13 aa (± 6 residues). Sites which could not be extended because of N- or C-termini were excluded by the Motif-X algorithm.

Degenerate motifs were also extracted by allowing for conservative amino acid substitutions at various positions except central one as follow: [AG], [DE], [FWY], [ILMV], [KR], [NQ], [ST].

Multiple phosphorylation motif discovery was carried out as follows. For double phosphorylation motifs, the foreground was created by mapping all phosphopeptides to the *D. melanogaster* protein database and extracting 13 amino acids long sequence, centered on one of the two phosphorylatable residues, and keeping the other phosphorylated. Phosphorylated residues were mapped to B, X, and Z, as described previously.¹⁴ Thus, only two phosphorylation events were considered if they are separated by 5 amino acids or fewer. For the background, all 13-mers from the protein database that were centered on an S, T, or Y were extracted while making sure that the 13-mer contained at least two phosphorylatable residues, including the central residue. For every background sequence, one off-center residue was called as phosphorylated. If there were more than one to choose from, it was picked randomly. Thus, both in the foreground and background, the 13-mer were centered on a potentially phosphorylated residue, with one other residue in the sequence already phosphorylation. The significance was then calculated as described for single phosphorylation motifs. For triple phosphorylation motifs, the procedure was repeated, increasing the sequence length to 17, centering on each of the three phosphorylated residues in the foreground, and requiring at least two off-center phosphorylated residues in the background. As a result, both foreground and background contained sequences centered on a potentially phosphorylated residue with two other residues in the 17-mer already phosphorylated.

The number of occurrence and significance used for Motif-X analysis were indicated in the Supporting Information Table 2.

Results and Discussion

Figure 1 shows the strategy used for the phosphorylation analysis of *Drosophila* embryos. Lysate was collected from 0 to 24 h old *w¹¹⁸* embryos. Ten milligrams of proteins was subjected to in-solution digestion and the resulting peptides were enriched for phosphopeptides by two steps.¹⁴ First, SCX chromatography was performed. Twelve fractions were collected and subjected to further phosphopeptide enrichment using IMAC. Each fraction was analyzed in duplicate (24 analyses in total) by LC-MS/MS using an LTQ Orbitrap hybrid mass spectrometer. In total, 177 898 MS/MS spectra were acquired. Spectra were searched using the Sequest algorithm²² with the target-decoy database approach¹¹ against the *D. melanogaster* protein database and its reversed complement. With the use of

decoy hits as a guide, filtering criteria for mass deviation, solution charge state, XCorr, and dCn' were applied to achieve a 0.63% false-discovery rate (Figure 1B). The final list contained 36 203 phosphorylated peptides corresponding to 2702 proteins (Supporting Information Table 1). Each phosphopeptide was then passed to the Ascore algorithm where the probability of correct site localization was determined.¹² All spectra are provided via hyperlink with computer-assisted annotation in Supporting Information Table 1.

The number of phosphopeptides detected in each run is shown in Figure 2A. On average, more than 1500 phosphopeptides were detected in every 1-h analysis. Because each fraction was analyzed in duplicate, an assessment of the reproducibility was possible. The correlation between results for duplicate analyses for the 12 SCX fractions was exceptional at every level examined including phosphopeptides identified (Figure 2A), solution charge state (Figure 2C), and number of phosphorylation events per peptide (Figure 2D). Despite this impressive agreement, examination of the overlapping information content between duplicates showed an average of $41.3 \pm 6.8\%$ increase in phosphopeptides detected due to a replicate analysis, demonstrating the shotgun nature of these experiments. Figure 2B shows an example of the overlap in phosphopeptide identifications in duplicate samples from fraction 3. As shown previously,^{7,23} these data argue that replicate analyses are the most important parameter available for increasing sample sensitivity.

SCX chromatography separates peptides mainly based on solution charge.¹⁷ At pH 2.7, phosphate groups still maintain a negative charge which contributes to the overall solution charge state of each tryptic peptide. Figure 2C shows the distribution of phosphopeptides in each fraction by predicted solution charge. Phosphopeptides with <1 charge were poorly retained and elute near the void volume of the column. The majority of phosphopeptides detected in the analysis contained a net charge of either one or two. However, many other charge states were enriched by the IMAC step and detected in fractions appropriate to their charge. We note that solution charge was often a powerful constraint for reducing false positives (Supporting Information Figure 2). For example, all correctly assigned phosphopeptides in fraction 2 contained either 0 or +1 net charge, while phosphopeptides in fraction 7 contained net charges between -1 and +3.

Often multiple phosphorylation events occur within a short amino acid sequence stretch of protein and therefore are contained concurrently within the same tryptic peptide. Figure 2D shows the distribution of peptides in each SCX fraction with up to 6 sites of phosphorylation. Most phosphopeptides contained three or fewer sites. Approximately 68% of all detected phosphopeptides were multiply phosphorylated. Notably, 195 and 35 peptides were identified containing 5 and 6 sites, respectively.

Site localization in phosphorylation analysis by mass spectrometry is often performed by tedious manual examination of MS/MS spectra. When multiple serines, threonines, or tyrosines are present in the sequence, the detection of specific fragment ions, termed site-determining ions, can distinguish between potential site locations. This process has been automated by the Ascore algorithm,¹² which computes the probability that a difference in detected site-determining ions between two potential site locations occurred due to random chance. The score is based on the cumulative binomial distribution and is returned as $-10 \log(P)$. The Ascore distribution of all returned sites revealed that 10 038 (73.2%) sites had Ascore values ≥ 13 ($P \leq 0.05$) and were thus localized with greater than 95% certainty in this analysis (Figure 3A).

Significant redundancy exists in our data set due primarily to the collection of replicate analyses for each fraction (Figure 2B) but also to other factors including (i) detection of singly and multiply phosphorylated forms of the same peptide, (ii) missed cleavages by

trypsin for lysines or arginines adjacent to phosphorylated residues, and (iii) detection of so-called ragged ends where trypsin cleaved a fraction of the time at each of two adjacent basic residues. From 36 203 total identified phosphopeptides, 13 720 nonredundant sites of phosphorylation could be assigned.

In the 13 720 nonredundant phosphorylation sites, 97% were serine and threonine (10 799 and 2536, respectively) and only 3% were tyrosine (385). Ser/Thr protein kinases can be divided into 4 general classes based on substrate sequence specificity:¹⁴ acidophilic, basophilic, proline-directed, and others (see Methods). This simplified categorization is useful as kinases in different pathways often utilize the same general motif (e.g., basophilic kinases all phosphorylate Rxx[st]). The first three categories (Figure 3B) constituted over 80% of the data set. The most abundant class of sites was acidophilic (35.8%) followed by proline-directed (26.3%), basophilic (19%), and other (18.9%).

Amino acid sequences neighboring serine, threonine, or tyrosine often define kinase specificity. Because of the large size of the data set, we were able to extract both novel and known phosphorylation-specific motifs. Only confidently localized phosphorylation sites were subjected to motif analysis. Sequences were centered on the phosphorylated site and extended 6 amino acids on each side. Motifs were extracted using the Motif-X algorithm.²¹ In total, 76 serine motifs and 20 threonine motifs ($P < 10^{-6}$) were found. The complete set of motifs is listed in Supporting Information Table 2. Notably, 3 tyrosine motifs ($P < 10^{-3}$) were identified from 161 localized phosphotyrosines in the data set (Supporting Information Table 2).

Besides motifs generated using all 20 amino acids, degenerate motifs were also extracted by allowing conservative amino acid substitutions at various positions (see Methods). As shown in Figure 4, three nondegenerate acidic motifs (Figure 4A) could be grouped into one degenerate motif (Figure 4B). In addition, the acidic residue at +4 became significant. Moreover, the degenerate analysis often showed a preference of one amino acid over another. As shown in Figure 5A, the PKA substrate motif RRx[s/t] preferred arginine over lysine at positions −1 and −2. Only RRx[s/t] was extracted in the nondegenerate analysis, but not KKx[s/t] (Supporting Information Table 2). Additional examples are present in Figure 5.

In this study, nearly 70% of the phosphopeptides identified were found to be multiply phosphorylated. This allowed motifs involving more than one phosphorylated residue to be detected suggestive of ordered phosphorylation events. Similar to another study,¹⁴ we found more acidic and proline-directed motifs than basic motifs. Furthermore, most double phosphorylation motifs could be deconvoluted into two already-identified single phosphorylation motifs. Within acidic motifs, phosphorylated serine and threonine functioned as an acidic amino acid residue replacing glutamate and aspartate in already-detected motifs (Figure 4C). Figure 5 shows examples of logo-like motif representations for both standard and degenerate motif analysis.

As an example of the potential impact of these data, we overlaid the identified sites onto the Salvador—Warts—Hippo (SWH) pathway (Figure 6). The SWH pathway controls organ size by modulating cell growth, proliferation, and apoptosis.^{24–26} Many components of this pathway are conserved from yeast to human and have been implicated in tumor formation. The core components of the SWH pathway are two serine/threonine kinases, Hippo and Warts. Hippo (Hpo) is a Ste-20 family protein kinase.^{27–31} When Hpo is activated by phosphorylation, it can phosphorylate Warts as well as Salvador and Mats. Warts (Wts) encodes a kinase of the Nuclear Dbf-2-related (NDR) family.^{32,33} Its phosphorylation and activation in turn phosphorylates downstream transcriptional coactivator Yorkie. Yorkie

(Yki) phosphorylation downregulates its putative transcriptional targets including the antiapoptotic molecules DIAP1, Cyclin E, and Bantam,^{34–36} leading to an increase in cell number and tissue size. Although new members of this pathway are emerging, the activation process is not yet understood, and sites of regulatory phosphorylation are mostly unknown. The only known sites are S920 and T1083 of Wts³⁷ and T195 of Hpo.³⁸ In this large-scale study, we identified phosphorylation sites affecting most major players including Fat, Hpo, Salvador, and Wts. As the first physiological substrate identified for any NDR family kinase, molecular studies of Yki phosphorylation by Wts could provide general insights into substrate specificity for this family of protein kinases, which remains poorly understood at present.²⁵ Actually, in our analysis, Yki in particular was found to be highly phosphorylated (10 different sites). Fat was also phosphorylated at 5 sites, all within the intracellular domain. These sites represent a defined starting point for further studies of protein function within this pathway.

Recently, the Aebersold group published a large-scale identification of *D. melanogaster* phosphorylation sites (10 118).¹⁶ Although that analysis and the one described here report similar numbers of phosphorylated peptides, the overlap between these two data sets was low (Supporting Information Figure 3). Several substantial dissimilarities between the two reports, both in the methods of sample handling and data analysis could have contributed to the differences. Sample-attributed differences include the samples themselves (Kc167 cells vs embryos), the use of external stimulation to boost phosphorylation (e.g., rapamycin, Calyculin A, insulin vs none), the degree of fractionation (>250 total samples vs 24), and the phosphopeptide enrichment strategies used (three different methods vs IMAC alone). A further data-attributable discrepancy is their use of SEQUEST's dCn score alone to localize ambiguous phosphorylation sites to specific residues versus the probabilistic Ascore. Together, these differences are likely reasons for large disparities in the number of phosphoproteins identified (4583 versus only 2702 here) and the fraction of multiply phosphorylated peptides detected (13% versus 68% here). Although these two analyses paint contrasting pictures of the fly proteome, they both emphasize that a combination of multiple analysis strategies will likely be needed to completely reveal the complexity of the phosphoproteome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by NIH grants HG3616 and HG3456 (to S.P.G.), and a postdoctoral fellowship from the Spanish Ministry of Education and Science (to J.V.). We thank Tomer Avidor-Reiss and Jayachandran Gopalakrishnan for help preparing *Drosophila melanogaster* embryos.

References

1. Celniker SE, Rubin GM. The *Drosophila melanogaster* genome. *Ann Rev Genomics Hum Genet.* 2003; 4(1):89–117. [PubMed: 14527298]
2. Adams MD. The genome sequence of *Drosophila melanogaster*. *Science.* 2000; 218:5–2195.
3. Peter Bangs KW. Regulation and execution of apoptosis during *Drosophila* development. *Dev Dyn.* 2000; 218(1):68–79. [PubMed: 10822260]
4. Vidal M, Cagan RL. *Drosophila* models for cancer research. *Curr Opin Genet Dev.* 2006; 16(1):10–16. [PubMed: 16359857]
5. Hay BA, Guo M. Caspase-dependent cell death in *Drosophila*. *Ann Rev Cell Dev Biol.* 2006; 22(1):623–650. [PubMed: 16842034]

6. Montell DJ. The social lives of migrating cells in *Drosophila*. *Curr Opin Genet Dev*. 2006; 16(4): 374–383. [PubMed: 16797177]
7. Haas W, Faherty BK, Gerber SA, Elias JE, Beausoleil SA, Bakalarski CE, Li X, Villen J, Gygi SP. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. *Mol Cell Proteomics*. 2006; 5(7):1326–1337. [PubMed: 16635985]
8. Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol*. 2002; 20(3):301–305. [PubMed: 11875433]
9. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJD. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics*. 2005; 4(7):873–886. [PubMed: 15858219]
10. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods*. 2007; 4(3):231–237. [PubMed: 17293869]
11. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*. 2007; 4(3):207–214. [PubMed: 17327847]
12. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*. 2006; 24(10): 1285–1292. [PubMed: 16964243]
13. Li X, Gerber SA, Rudner AD, Beausoleil SA, Haas W, Villen J, Elias JE, Gygi SP. Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*. *J Proteome Res*. 2007; 6(3):1190–1197. [PubMed: 17330950]
14. Villen J, Beausoleil SA, Gerber SA, Gygi SP. Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci USA*. 2007; 104(5):1488–1493. [PubMed: 17242355]
15. Bodenmiller B, Mueller LN, Pedrioli PGA, Pflieger D, Junger MA, Eng JK, Aebersold R, Tao WA. An integrated chemical, mass spectrometric and computational strategy for (quantitative) phosphoproteomics: application to *Drosophila melanogaster* Kc167 cells. *Mol Biosyst*. 2007; 3(4): 275–286. [PubMed: 17372656]
16. Bodenmiller B, Malmstrom J, Gerrits B, Campbell D, Lam H, Schmidt A, Rinner O, Mueller LN, Shannon PT, Pedrioli PG, Panse C, Lee HK, Schlapbach R, Aebersold R. PhosphoPep—a phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells. *Mol Syst Biol*. 2007; 3:139. [PubMed: 17940529]
17. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA*. 2004; 101(33):12130–12135. [PubMed: 15302935]
18. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*. 2003; 75(3):663–670. [PubMed: 12585499]
19. Celniker S, Wheeler D, Kronmiller B, Carlson J, Halpern A, Patel S, Adams M, Champe M, Dugan S, Frise E, Hodgson A, George R, Hoskins R, Lavery T, Muzny D, Nelson C, Pacleb J, Park S, Pfeiffer B, Richards S, Sodergren E, Svirkas R, Tabor P, Wan K, Stapleton M, Sutton G, Venter C, Weinstock G, Scherer S, Myers E, Gibbs R, Rubin G. Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biology*. 2002; 3(12):research0079. [PubMed: 12537568]
20. Hoskins R, Smith C, Carlson J, Carvalho AB, Halpern A, Kaminker J, Kennedy C, Mungall C, Sullivan B, Sutton G, Yasuhara J, Wakimoto B, Myers E, Celniker S, Rubin G, Karpen G. Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biology*. 2002; 3(12):research0085. [PubMed: 12537574]
21. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol*. 2005; 23(11):1391–1398. [PubMed: 16273072]
22. Eng JK, McCormack AL, Yates JR III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*. 1994; 5(11): 976–989.

23. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods*. 2005; 2(9):667–675. [PubMed: 16118637]
24. Harvey K, Tapon N. The Salvador-Warts-Hippo pathway—an emerging tumour-suppressor network. *Nat Rev Cancer*. 2007; 7(3):182–191. [PubMed: 17318211]
25. Pan D. Hippo signaling in organ size control. *Genes Dev*. 2007; 21(8):886–897. [PubMed: 17437995]
26. Saucedo LJ, Edgar BA. Filling out the Hippo pathway. *Nat Rev Mol Cell Biol*. 2007; 8(8):613–621. [PubMed: 17622252]
27. Wu S, Huang J, Dong J, Pan D. Hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with Salvador and Warts. *Cell*. 2003; 114(4):445–456. [PubMed: 12941273]
28. Harvey KF, Pflieger CM, Hariharan IK. The *Drosophila* Mst ortholog, Hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell*. 2003; 114(4):457–467. [PubMed: 12941274]
29. Udan RS, Kango-Singh M, Nolo R, Tao C, Halder G. Hippo promotes proliferation arrest and apoptosis in the Salvador/Warts pathway. *Nat Cell Biol*. 2003; 5(10):914–920. [PubMed: 14502294]
30. Pantalacci S, Tapon N, Leopold P. The Salvador partner Hippo promotes apoptosis and cell-cycle exit in *Drosophila*. *Nat Cell Biol*. 2003; 5(10):921–927. [PubMed: 14502295]
31. Jia J, Zhang W, Wang B, Trinko R, Jiang J. The *Drosophila* Ste20 family kinase dMST functions as a tumor suppressor by restricting cell proliferation and promoting apoptosis. *Genes Dev*. 2003; 17(20):2514–2519. [PubMed: 14561774]
32. Justice RW, Zilian O, Woods DF, Noll M, Bryant PJ. The *Drosophila* tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes Dev*. 1995; 9(5):534–546. [PubMed: 7698644]
33. Xu T, Wang W, Zhang S, Stewart R, Yu W. Identifying tumor suppressors in genetic mosaics: the *Drosophila* lats gene encodes a putative protein kinase. *Development*. 1995; 121(4):1053–1063. [PubMed: 7743921]
34. Thompson BJ, Cohen SM. The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in *Drosophila*. *Cell*. 2006; 126(4):767–774. [PubMed: 16923395]
35. Nolo R, Morrison CM, Tao C, Zhang X, Halder G. The bantam microRNA is a target of the Hippo tumor-suppressor pathway. *Curr Biol*. 2006; 16(19):1895–1904. [PubMed: 16949821]
36. Huang J, Wu S, Barrera J, Matthews K, Pan D. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* homolog of YAP. *Cell*. 2005; 122(3):421–434. [PubMed: 16096061]
37. Emoto K, Parrish JZ, Jan LY, Jan Y-N. The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance. *Nature*. 2006; 443(7108):210–213. [PubMed: 16906135]
38. Colombani J, Polesello C, Josue F, Tapon N. Dmp53 activates the Hippo pathway to promote cell death in response to DNA damage. *Curr Biol*. 2006; 16(14):1453–1458. [PubMed: 16860746]
39. Wilson-Grady JT, Villén J, Gygi SP. Phosphoproteome analysis of fission yeast. *J Proteome Res*. 2008; 7(13):1088–1097. [PubMed: 18257517]

Figure 1.

Schematic illustration of the strategy for large-scale phosphorylation site identification from *Drosophila* embryos. (A) The 0–24 h old *D. melanogaster w¹¹⁸* embryos were lysed and directly digested with trypsin. Tryptic peptides were desalted and then separated by SCX chromatography. Phosphopeptides from 12 SCX fractions were further enriched by IMAC and then analyzed by LC–MS/MS techniques. (B) MS/MS spectra from 24 analyses (duplicates for each sample) were searched against a composite target-decoy *Drosophila* protein database.¹¹ Mass deviation, XCorr, dCn', and solution charge state were used to filter correct from incorrect matches, maintaining <1% false-discovery rate (FDR). In total, 36 203 phosphopeptides (16 822 unique phosphopeptides) and 13 720 nonredundant phosphorylation sites were identified at a FDR of 0.63% (229 decoy matches). High-certainty localization (Ascore ≥ 13 ; $P \leq 0.05$)¹² was found for 10 038 sites. Finally, phosphorylation motifs (standard, degenerate, and multiply phosphorylated) were extracted from the data set with the Motif-X algorithm.²¹

Figure 2.

Distributions of phosphopeptides and their properties across 12 SCX fractions. (A) The number of phosphopeptides identified from duplicate analyses of each fraction. While similar numbers of peptides were identified in each replicate, an average of $41.3 \pm 6.8\%$ more peptides could be attributed solely to analyzing each fraction twice. (B) Venn diagram depicting the extent of overlap for the phosphopeptides identified in duplicate analyses of fraction 3. Numbers in parentheses indicate the percentage of either replicate that lies outside the overlap region. (C) Nonredundant phosphopeptides in each fraction (and replicate) with calculated solution charge states between -1 and $+4$. SCX separates phosphopeptides based primarily on solution charge. (D) Nonredundant phosphopeptides in each fraction containing 1 (1P) to 6 (6P) phosphorylation sites. (a and b correspond to duplicate LC-MS/MS runs of each SCX fraction).

Figure 3.

(A) Ascore distribution for all identified sites from 36 203 peptides. Most sites could be localized with near ($P \leq 0.01$) or high ($P \leq 0.05$) certainty. (B) Classification of phosphorylation events into 4 general sequence categories based on kinase specificities.

Figure 4.

Logo-like representations of acidic phosphorylation motifs identified by the Motif-X algorithm.²¹ (A) Examples of motifs extracted utilizing all 20 amino acids. Note that several permutations of a similar motif are identified. (B) Degenerate motif for casein-kinase II-like phosphorylation. Note that all three motifs in panel A are represented in this single motif. The residue at position +4 is also now significant. (C) Acidic, double phosphorylation motif identified using both degenerate analysis and considering multiple phosphorylation events. Phosphorylated serine and threonine (denoted as B and X, respectively) function as an acidic residue at position +1 in this motif.

Figure 5.

Examples of motifs (represented by logo-like representations) identified in this study. The phosphorylated serine, threonine, or tyrosine are centered. (A) PKA-like substrate motifs showing a preference for arginine over lysine in degenerate analyses. (B) Novel threonine single phosphorylation motif. (C) Double phosphorylation motif (B and X represent phosphoserine and threonine, respectively) suggestive of ordered phosphorylation. (D) Examples of tyrosine single phosphorylation motifs.

Figure 6.

Overlay example of phosphorylation sites from this study on a developmental pathway. The Salvador–Warts–Hippo (SWH) pathway controls organ size by modulating cell growth, proliferation, and apoptosis. Many of the genetic and biochemical interactions are known, but post-translational modifications such as phosphorylation affecting core components are almost entirely lacking. Phosphorylation events detected in this study are overlaid onto a representation of the pathway.²⁴ Yorkie (Yki) is hyperphosphorylated in *Drosophila* embryos. The arrows in the figure represent what is known about the pathway from the literature where a given protein acts to stimulate or inhibit the function of the pathway. They do not indicate the direct kinase–substrate relationship.