HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2012 February 01.

Published in final edited form as:

Nat Methods. ; 8(8): 677–683. doi:10.1038/nmeth.1636.

A large-scale method to measure absolute protein phosphorylation stoichiometries

Ronghu Wu, **Wilhelm Haas**, **Noah Dephoure**, **Edward L. Huttlin**, **Bo Zhai**, **Mathew E. Sowa**, and **Steven P. Gygi***

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

Abstract

The underlying functional role of protein phosphorylation is impacted by its fractional stoichiometry. Thus, a comprehensive strategy to study phosphorylation dynamics should include an assessment of site stoichiometry. Here, we developed an integrated method that relies on phosphatase treatment and stable isotope labeling to determine the absolute stoichiometries of protein phosphorylation on a large-scale. This approach requires the measurement of only a single ratio relating phosphatase- and mock-treated samples. We applied the strategy to determine stoichiometries for 5,033 phosphorylation sites in *Saccharomyces cerevisiae*. Stoichiometries were determined from biological triplicate experiments with good reproducibility. We validated ten sites stoichiometries representing the full range of values with an absolute quantitative approach, showing excellent agreement. Using bioinformatics, we characterized the biological properties associated with phosphorylation sites with vastly differing absolute stoichiometries.

INTRODUCTION

Reversible protein phosphorylation plays an important role in biological systems and is involved in virtually every cellular function. Mass spectrometry (MS) -based proteomics can globally characterize this widespread post-translational modification and has been extensively reviewed[1–3]. In combination with stable isotope labeling, MS-based proteomics has produced large-scale data sets quantifying phosphorylation changes between cell states[1,4,5]. Current large-scale comparisons provide an informative view of the protein phosphorylation landscape[1,6–9]. However, their biological interpretation is complex. Notably, if the difference in protein abundance is known, a change in a site's phosphorylation levels results in a known relative stoichiometry change[10]. Yet, the absolute fractional occupancy remains unknown. For example, a two-fold down-regulation of site stoichiometry could result from either fractional occupancy changes of 0.2 → 0.1% or 100 → 50%, which likely represent fundamentally different cellular strategies. Therefore it

*Correspondence to: Steven Gygi (sgygi@hms.harvard.edu), Phone: (617) 432-3155, Fax: (617) 432-1144.

Note: Supplementary information is available.

AUTHOR CONTRIBUTIONS

S.P.G. and R.W. designed the research, R.W., W.H., N.D., E.L.H., B.Z., M.S. and S.P.G. participated in the data generation, analysis and interpretation, R.W. and S.P.G. wrote the manuscript and all authors edited it.

appears crucial to assess absolute phosphorylation site occupancy on a proteome scale in order to correctly and comprehensively understand its functional significance.

Conventionally, biochemical methods, such as Western blotting, have been used to measure phosphorylation stoichiometry. Phosphoproteins and non-phosphoproteins are separated physically *via* sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and their quantities are estimated using antibodies[11,12]. This method is time consuming and requires a phosphorylation-induced migration difference in the gel.[1] Protein phosphorylation stoichiometry can also be measured by MS[9,13,14]. For example, using a label-free approach, stoichiometry can be measured from the ratios of ion signals of phosphopeptides and the corresponding non-phosphopeptides[14–17]. An assumption of this method is that differences in the ionization and detection efficiencies of a peptide's phosphorylated and nonphosphorylated forms are negligible. To overcome this shortcoming, a recently reported method[18] determined the response ratios of phosphopeptides and non-phosphopeptides using synthetic peptide standards, and was applied to measure the stoichiometries of two tyrosine residues in the Lyn protein.

Another MS-based method, termed absolute quantification (AQUA)[13], is based on stable isotope dilution. Heavy versions of phosphorylated and non-phosphorylated peptides are synthesized and spiked into the sample in known quantities as internal standards. The phosphorylation stoichiometry can be obtained by measuring and comparing the absolute abundances of each peptide's phosphorylated and nonphosphorylated forms. Notably, using this method, it was demonstrated that S1526 from human separase is maintained at full occupancy until the metaphase-to-anaphase transition when it is partially dephosphorylated and activated, allowing the release of tethered sister chromatids. Knowing the stoichiometry of S1526 across the cell cycle proved vital to understanding its specialized mitotic function[19]. In addition to separase, this method has been adapted to measure occupancy in Akt[20].

Another elegant strategy was reported to measure phosphorylation stoichiometry by MS [21]. Phosphatase treatment was used in combination with specific and differential labeling of the N-termini of all peptides in a sample with either a D_5 or D_0 –propionyl group, then measuring the ratio of the abundance of the D_5/D_0 peptide pairs simultaneously using MALDI-MS. Stoichiometry was obtained based on the signal increase of the peptide from the dephosphorylation of the corresponding phosphopeptide. A similar strategy [22] measured the phosphorylation stoichiometries of different sites of protein Npr1 by protease dependent incorporation of ^{18}O or ^{16}O labeled peptides, followed by phosphopeptide enrichment, phosphatase treatment, and MALDI-MS analysis. This strategy has also been used in other reports[23–28].

However, all of the methods discussed above were used for small-scale studies, ranging from one to several proteins or a protein complex. With the development of MS-based proteomics methods, thousands of unique phosphorylation sites can be analyzed in one experiment, yet the underlying fractional occupancy of these sites remains unknown. Recently, a method [9] was reported to measure site stoichiometry at a large scale by obtaining a minimum of three different ratios representing protein, phosphopeptide, and

unmodified peptide changes based on stable isotope labeling with amino acids in cell culture (SILAC). But this approach is biased in that it can only be applied to sites after detecting a change in phosphopeptide levels.

In this study, we integrated phosphatase treatment and stable isotope labeling to determine site stoichiometries of protein phosphorylations on a large-scale. Importantly, the method can provide thousands of stoichiometry measurements for a single cellular condition. We measured basal stoichiometry levels for more than 5,000 events in exponentially-growing yeast in triplicate. The accuracy of this method was assessed for ten sites of varied stoichiometries using the AQUA strategy[13]. Bioinformatic analyses indicated that acidic sites were, on average, of higher stoichiometry than other motifs and that high stoichiometry sites were not more conserved across yeast species than low stoichiometry sites. Several biological and functional categories are statistically enriched for high or low stoichiometry.

RESULTS

A method to assess absolute stoichiometry on a large scale

An overview of the method is shown in Fig 1. A cell lysate is first digested with endoproteinase lys-C. Two identical 0.5 mg peptide aliquots are subjected to either phosphatase treatment or a mock reaction. Following the reactions, peptides in the phosphatase-treated sample are chemically labeled by reductive dimethylation (ReDi) using deuteroformaldehyde to dimethylate free amines[29]. The mock reaction sample is chemically labeled using formaldehyde. The two aliquots are then mixed, which results in a 1:1 ratio for all peptides unaffected by phosphatase treatment. The mixed sample is then analyzed by deep MS sequencing to identify and quantify as many peptide species as possible. This involves separation by hydrophilic interaction liquid chromatography (HILIC), and then analysis of each fraction by LC-MS/MS techniques. Among the thousands of detected peptides will be those previously harboring phosphorylation sites. They will display and increase in the heavy partner intensity over the co-eluting light version directly representing the fraction of the peptide that was phosphorylated. Each of these peptides is termed an “occupancy determining peptide” (ODP), and they directly encode the absolute stoichiometry based on the ratio of heavy/light species ((1-1/Ratio) × 100%). A list of ODPs can be collected from previously published data sets of site-specific phosphorylation. The overlap of these published data sets with the phosphatase-treated one defines the number of sites for which stoichiometries can be determined.

Applying the method to define absolute stoichiometries

We evaluated the potential for this method by applying it defines the absolute stoichiometries of yeast growing through log phase with glucose as a carbon source. Separate yeast cultures (n = 3) were grown and protein was harvested. The method was performed as described in Fig. 1 with deep MS sequencing of 20 fractions for each experiment. The total numbers of peptides detected in each culture were 76,893; 77,439 and 81,012; respectively (Supplementary Table 1–4). The false-discovery rates are shown in parenthesis and were based on the target-decoy database approach[30]. In a separate

experiment, the phosphatase reaction performed at the peptide level was found to be > 99% effective (see Methods).

In order to use this method, a list of phosphorylation sites was needed. We chose to use localized sites determined in 5 published yeast phosphorylation studies[7,10,31–33]. The overlap between our phosphatase-treated data sets and the sequences surrounding these published sites was assessed. Due to the shotgun nature of LC-MS/MS, not every ODP was identified. In this proof-of-principle experiment, for 5,033 sites in the literature, we detected the nonphosphorylated peptide forms. Their H/L ratios are plotted in Figs. 2a, and 2b displays an example of the occupancy calculation from the measured ratio. The stoichiometry distribution is shown in Fig. 2c. Very high occupancy (> 90%) was not common, occurring at only about one in ten sites. The majority of sites showed low occupancy, with half displaying occupancies of < 30%. Some examples with different motifs are shown in Supplementary Table 5.

One limitation of this method stems from the fact that bottom-up proteomics methods measure ratios for individual peptides whose sequences can contain one or more phosphorylation sites. While the ratios observed for singly phosphorylated peptides represent stoichiometries for individual sites, peptides containing multiple phosphorylation sites reflect the extent of modification for all phosphorylation sites in any combination. Our MS-based method cannot distinguish the contribution of each site within these multiply modified peptides. For simplicity, we have assigned the stoichiometries calculated from the peptide ratio to each site in the peptide. Thus, these stoichiometries should be considered as maximum stoichiometries when more than one site is known on the peptide. 4,016 sites (80%) were identified from peptides displaying only a single site (Supplementary Table 6).

In order to assess the reproducibility of the method, biological triplicate experiments were performed. Stoichiometries were obtained for 3,843; 4,077 and 3,756 events, respectively. The Pearson correlations between replicates was strong, as shown for experiments 1 and 2, and 1 and 3 (Figs. 3a and 3b, Supplementary Figure 1). For 2,664 site stoichiometries measured in all experiments, the average standard deviation (SD) was 6.3%. The SD distribution is plotted in Supplementary Figure 1c. Using one SD, stoichiometries were accurate to ± 6%. Values could be made even more accurate by increasing the number of biological replicates. However, these measurements are accurate enough for site classification, allowing further biological characterization.

Validation of absolute site stoichiometries using AQUA

To validate our stoichiometry measurements, we used another independent method, the AQUA strategy[13], to accurately determine the stoichiometries of ten sites chosen to represent the full range of occupancy states (Fig. 3 and Table 1). Ten pairs of heavy phosphopeptides and corresponding non-phosphopeptides were then synthesized. Known amounts of heavy peptides were spiked into proteolyzed whole cell lysate and samples were analyzed by LC-MS/MS (see Methods). As an example of AQUA validation, we show the extracted chromatographic peaks of non-phosphopeptide (RIIEHSDVENENVK) and phosphopeptide (RIIEHS*DVENENVK) for site S618 of protein UBP1, a ubiquitin-specific protease that removes ubiquitin from conjugated proteins (Fig. 3a), and the identification of

the corresponding phosphopeptide by MS/MS (Fig. 3b). According to the H/L ratio and the amount of heavy peptides spiked into the sample, we calculated the amount of the peptide and phosphopeptide in the lysate to be 20.5 and 46.0 fmol, respectively. Thus, the stoichiometry of this site is 69.2%, which agrees with the values of 59%, 59% and 66%, respectively, in the biological triplicate experiments. The stoichiometries of sites S266 of IPP1 and T710 of DCP2 were very low based on the results of our large-scale experiments. This was also confirmed by AQUA, i.e. 1.8% and 2%, respectively. In addition, both acidic sites of S562 of YML093W and S379 of BFR2 showed full or nearly full occupancies (Fig. 4c, Table 1). The AQUA results were again consistent. The ten site stoichiometries quantified by AQUA strongly suggest that our large-scale method for determination of phosphorylation site occupancy is reliable.

Relationship between kinase motifs and site stoichiometry

We next examined the relationship between specific sequence motifs favored by kinases and the resulting site stoichiometries. Phosphorylation sites containing an acidic (caseine-kinase II-like) motif demonstrated higher occupancy levels than other motifs (Fig. 4a). Proline-directed sites, which are often modified by cyclin-dependent kinase (CDK) and mitogen-activated protein kinase (MAPK), have lower than average occupancies.

For each site with occupancy, we also applied secondary structure prediction algorithms [34] to classify events according to likely structure. Most sites (91%) were predicted to be in disordered regions (Fig. 4b). However, sites predicted to be in ordered regions demonstrated a strong preference for low stoichiometry. As we have shown previously[7], phosphorylation is overwhelmingly detected in protein regions predicted or known to be disordered. Yet, when phosphorylation does occur in ordered regions, we found that it is almost always of low occupancy in exponentially-growing yeast. This suggests two things: i) the frequency of high occupancy sites serving structural roles is low, and ii) phosphorylation events occurring in ordered regions are nearly always activating in nature. Indeed, the best example of this phenomenon is that phosphorylation within the kinase domains is often required for kinase activation[35].

Correlating protein function to stoichiometry levels

In order to obtain functional insights into site stoichiometry, we separated phosphoproteins into four groups based on their highest stoichiometries (low (< 10%), medium (10–30%), high (30–75%), and very high (75–100%)) to examine differences in biological processes and cellular compartment enrichment using DAVID[36]. P-values were z-transformed and then clustered (Figs. 4c and 4d). Biological processes enriched in high stoichiometry sites included “Chromatin silence” and “cytokinesis during cell cycle.” Full versions of these figures are shown in Supplementary Fig. 2.

We next examined site conservation across 25 fungal proteomes for thousands of sites (Fig. 5, Supplementary Fig. 3). Homologs from 25 fungal genomes from the same family (Saccharomycetaceae) were identified using a TBlastN search and then aligned. If a homolog was found, conservation was determined based on identifying the same phospho-acceptor residue at the same position in both species (see Methods). To facilitate this

analysis, only singly-phosphorylated sites with high confidence localizations were selected (n=3,443), and we binned stoichiometries into three categories of low, medium, and high occupancies. Site residue conservation was generally localized to only a few very closely related species. However, ~10% of sites were highly conserved. Previously, we had found a similarly poor conservation of 541 Cdk1-phosphorylated residues across fungal species[7]. Notably, high occupancy site residues were actually less conserved, on average, than low occupancy ones. Strikingly, most sites are poorly conserved even within budding yeast species, and high occupancy sites were not found on more conserved residues. Indeed, the most conserved site residues harbored very low stoichiometry sites. These events are also the ones predicted to be in ordered regions (data not shown). A website has been created for visualizing these conservation data and the alignments (http://gygi.med.harvard.edu/pubs/occupancy_evolution).

DISCUSSION

Our method of determining absolute site stoichiometry provides unbiased insight into this functionally important protein property for thousands of basally phosphorylated sites. Notably, the approach requires only the measurement of a single ratio in a single experiment for ODPs. The stoichiometry relationship is completely straight-forward to determine for singly-phosphorylated peptides. Determining stoichiometries for multiply-phosphorylated peptides is more complex and may not accurately reflect values at individual sites, but do still represent occupancy levels encompassing the sequence region shown. They should be considered to represent the maximum potential occupancy at each site. Of the 5,033 stoichiometry measurements collected, 1,017 were derived from peptides where multiple sites were present.

Our stoichiometry measurements were averaged from three separate yeast cultures. The standard deviation of these measurements was found to be ~6%. This implies that the method is not suitable to distinguish small differences in stoichiometry, but is capable of classifying sites into at least 5–10 categories of increasing stoichiometry. We also validated the absolute occupancy state for ten sites by a highly quantitative method (AQUA[13]). Even though sites were chosen to represent the full range of occupancy states, the agreement between the AQUA stoichiometries and the phosphatase-based method was strong.

A special stoichiometry case is full occupancy, which predicts constitutive kinase activity and/or little phosphatase activity toward the site. Many cyclin-dependent kinase sites have been shown to achieve full occupancy under mitotic conditions[9,19]. However, we collected our measurements from asynchronous yeast where the vast majority of cells would not be in M phase. Thus, most targets of CDK (mitotic sites) would be expected to be detected at low stoichiometry. Surprisingly, we found that sites determined in that study to be substrates for Cdk1 and overlapping with ours (n = 95) contained a similar distribution of stoichiometries as the entire phosphoproteome with many high and low stoichiometry events (mean =40.6%; Supplementary Fig. 4; Supplementary Table 7). This indicates that many of these “mitotic” sites are actually phosphorylated during other phases of the cell cycle. They likely have additional, non-mitotic roles which are then overwhelmed or repurposed during mitosis.

During exponential growth in asynchronous cultures, we found that ~10% of sites demonstrated full or almost full occupancy (> 90%) (Fig. 2c). This finding has important implications. It suggests that kinase pathways are generally inactive or less active in their default state and that most phosphorylation events function by influencing only a fraction of the available protein molecules. Maintaining sites at full occupancy thus requires considerable energy investment. Casein kinase II (CKII) is a constitutively active and essential yeast kinase with hundreds of substrates[37]. Strong casein kinase II motifs contain acidic residues at the +1 and +3 positions (e.g., SDxE). This acidic motif encompassed higher stoichiometries, on average, than basophilic or proline-directed motifs, demonstrating that strong or even constitutive pathway activation can result in high basal stoichiometries.

Only one other method has been reported that can assess absolute site stoichiometry on a large or global scale[9]. The authors determined occupancy by obtaining a minimum of three different SILAC ratios representing protein, phosphopeptide, and unmodified peptide changes. This is accomplished in separate proteome- and phosphorylation-based analyses. Using these three ratios from two experiments, a fractional occupancy level can be calculated. Importantly, there is a requirement that a substantial change in phosphopeptide amount occurs. In addition, because of the complex math involved, even slight errors in measurement can cause wildly changing stoichiometry calculations, and there are many undefined possibilities among the three ratios. For example, a change of any amount at the protein level with no corresponding change at the phosphopeptide level results in a change in occupancy, but is undefined by their method.

A major question is how stoichiometry as a biological property might influence protein function or regulation. We analyzed differences in the enrichment of gene ontology categories for low, medium, high, and very high stoichiometries. This provides hints at where high and low occupancy may be important for biological function. Phosphoproteins with reported localization to the cytoplasm, ribosome and mitochondria generally showed lower stoichiometries while proteins in the nucleus and cellular buds were enriched in high phosphorylation occupancies. Nuclear and budding proteins are enriched in regulatory proteins suggesting fine regulatory control. Proteins in many biological processes such as “cytokinesis during cell cycle” and “chromatin silence” were enriched with high phosphorylation stoichiometries, but site stoichiometries of proteins related to, for example, “mitochondrial organization” and “RNA transport” were enriched in low stoichiometry sites. Since high stoichiometry sites are rare, their association with biological processes reflect their important role in regulating these events. Examples of very high stoichiometry sites are shown in Supplementary Table 8.

Phosphorylation occurs most commonly on serine, threonine, and tyrosine residues in yeast. We examined whether high site stoichiometry correlated with residue conservation across related yeast species. We chose 25 fungal species with sequenced genomes all from the same family, Saccharomycetaceae, which reproduce based on “budding”. Our results suggest that stoichiometry does not positively correlate with a site’s biological essentialness. The strong lack of conservation in general and even within very closely related species implies that phosphorylation networks are set up in such a way (perhaps through multiple

phosphorylation events per protein[8]) that extensive evolutionary freedom even at previous sites of regulation is permissible.

METHODS

Methods and any associated references are available in the online version of the paper

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported in part by NIH grants (HG3456) to S.P.G. We thank all members of the Gygi lab for help, especially R.A. Everley for his help on instrumentation, and L. Ting for her critical reading.

References

1. Grimsrud PA, Swaney DL, Wenger CD, Beauchene NA, Coon JJ. Phosphoproteomics for the Masses. ACS Chem. Biol. 2010; 5:105–119. [PubMed: 20047291]
2. Macek B, Mann M, Olsen JV. Global and Site-Specific Quantitative Phosphoproteomics: Principles and Applications. Annu. Rev. Pharmacol. Toxicol. 2009; 49:199–221. [PubMed: 18834307]
3. Thingholm TE, Jensen ON, Larsen MR. Analytical strategies for phosphoproteomics. Proteomics. 2009; 9:1451–1468. [PubMed: 19235172]
4. Boersema PJ, Mohammed S, Heck AJR. Phosphopeptide fragmentation and analysis by mass spectrometry. J. Mass Spectrom. 2009; 44:861–878. [PubMed: 19504542]
5. Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. Nature Methods. 2007; 4:798–806. [PubMed: 17901869]
6. Choudhary C, et al. Mislocalized Activation of Oncogenic RTKs Switches Downstream Signaling Outcomes. Mol. Cell. 2009; 36:326–339. [PubMed: 19854140]
7. Holt LJ, et al. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science. 2009; 325:1682–1686. [PubMed: 19779198]
8. Huttlin EL, et al. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. Cell. 2010; 143:1174–1189. [PubMed: 21183079]
9. Olsen JV, et al. Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. Sci. Signal. 2010; 3:15.
10. Wu R, et al. Correct Interpretation of Comprehensive Phosphorylation Dynamics Requires Normalization by Protein Expression Changes. Mol. Cell. Proteomics. 2011 (in press).
11. Cooper JA, Hunter T. Identification and characterization of cellular targets for tyrosine protein kinases. J. Biol. Chem. 1983; 258:1108–1115. [PubMed: 6571834]
12. Stukenberg PT, et al. Systematic identification of mitotic phosphoproteins. Curr. Biol. 1997; 7:338–348. [PubMed: 9115395]
13. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:6940–6945. [PubMed: 12771378]
14. Steen H, Jebanathirajah JA, Springer M, Kirschner MW. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:3948–3953. [PubMed: 15741271]
15. Carr SA, Huddleston MJ, Annan RS. Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. Anal. Biochem. 1996; 239:180–192. [PubMed: 8811904]
16. Guo L, et al. Studies of ligand-induced site-specific phosphorylation of epidermal growth factor receptor. J. Am. Soc. Mass Spectrom. 2003; 14:1022–1031. [PubMed: 12954170]

17. Steen JAJ, et al. Different phosphorylation states of the anaphase promoting complex in response to antimitotic drugs: A quantitative proteomic analysis. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:6069–6074. [PubMed: 18420821]
18. Jin LL, et al. Measurement of Protein Phosphorylation Stoichiometry by Selected Reaction Monitoring Mass Spectrometry. J. Proteome Res. 2010; 9:2752–2761. [PubMed: 20205385]
19. Stemmann O, Zou H, Gerber SA, Gygi SP, Kirschner MW. Dual inhibition of sister chromatid separation at metaphase. Cell. 2001; 107:715–726. [PubMed: 11747808]
20. Atrih A, et al. Stoichiometric Quantification of Akt Phosphorylation Using LC-MS/MS. J. Proteome Res. 2010; 9:743–751. [PubMed: 19902931]
21. Zhang XL, Jin QK, Carr SA, Annan RS. N-Terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. Rapid Commun. Mass Spectrom. 2002; 16:2325–2332. [PubMed: 12478578]
22. Bonenfant D, et al. Quantitation of changes in protein phosphorylation: A simple method based on stable isotope labeling and mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:880–885. [PubMed: 12540831]
23. Domanski D, Murphy LC, Borchers CH. Assay Development for the Determination of Phosphorylation Stoichiometry Using Multiple Reaction Monitoring Methods with and without Phosphatase Treatment: Application to Breast Cancer Signaling Pathways. Anal. Chem. 2010; 82:5610–5620. [PubMed: 20524616]
24. Hegeman AD, Harms AC, Sussman MR, Bunner AE, Harper JF. An isotope labeling strategy for quantifying the degree of phosphorylation at multiple sites in proteins. J. Am. Soc. Mass Spectrom. 2004; 15:647–653. [PubMed: 15121193]
25. Johnson H, Eyers CE, Eyers PA, Beynon RJ, Gaskell SJ. Rigorous Determination of the Stoichiometry of Protein Phosphorylation Using Mass Spectrometry. J. Am. Soc. Mass Spectrom. 2009; 20:2211–2220. [PubMed: 19783156]
26. Kanshin E, et al. The stoichiometry of protein phosphorylation in adipocyte lipid droplets: Analysis by N-terminal isotope tagging and enzymatic dephosphorylation. Proteomics. 2009; 9:5067–5077. [PubMed: 19921680]
27. Pflieger D, et al. Quantitative proteomic analysis of protein complexes. Mol. Cell. Proteomics. 2008; 7:326–346. [PubMed: 17956857]
28. Previs MJ, et al. Quantification of protein phosphorylation by liquid chromatography-mass spectrometry. Anal. Chem. 2008; 80:5864–5872. [PubMed: 18605695]
29. Broberg A. High-performance liquid chromatography/electrospray ionization ion-trap mass spectrometry for analysis of oligosaccharides derivatized by reductive amination and N,N-dimethylation. Carbohydr. Res. 2007; 342:1462–1469. [PubMed: 17532306]
30. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods. 2007; 4:207–214. [PubMed: 17327847]
31. Albuquerque CP, et al. A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol. Cell. Proteomics. 2008; 7:1389–1396. [PubMed: 18407956]
32. Gnad F, et al. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. Proteomics. 2009; 9:4642–4652. [PubMed: 19795423]
33. Li X, et al. Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. J. Proteome Res. 2007; 6:1190–1197. [PubMed: 17330950]
34. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics. 2006; 7:17. [PubMed: 16409635]
35. Taylor SS, Knighton DR, Zheng JH, Teneyck LF, Sowadski JM. Structural framework for the protein-kinase family. Annu. Rev. Cell Biol. 1992; 8:429–462. [PubMed: 1335745]
36. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2009; 4:44–57. [PubMed: 19131956]
37. Guerra B, Issinger OG. Protein kinase CK2 in human diseases. Curr. Med. Chem. 2008; 15:1870–1886. [PubMed: 18691045]
38. Haas W, et al. Optimization and use of peptide mass measurement accuracy in shotgun proteomics. Mol. Cell. Proteomics. 2006; 5:1326–1337. [PubMed: 16635985]

39. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. J. Am. Soc. Mass Spectrom. 1994; 5:976–989. [PubMed: 24226387]

40. Bakalarski CE, et al. The Impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. J. Proteome Res. 2008; 7:4756–4765. [PubMed: 18798661]

41. Villen J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat. Protoc. 2008; 3:1630–1638. [PubMed: 18833199]

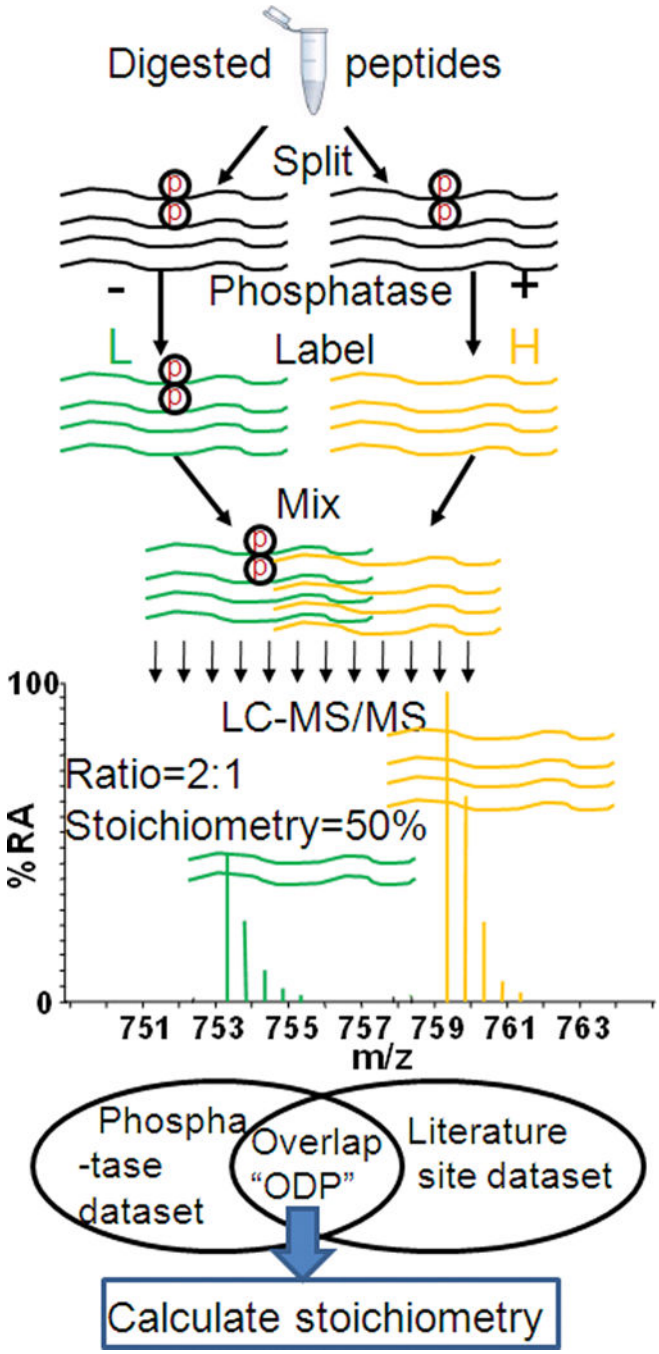


Figure 1.

Principle of the method for phosphatase-based, absolute stoichiometry measurements. Two identical aliquots of a proteolyzed protein lysate are either mock- or phosphatase-treated followed by differential chemical labeling with stable isotopes. After mixing, fractional occupancy is encoded in a single ratio comparing the nonphosphorylated form of a peptide with and without phosphatase treatment. Peptide sequences obtained here are then examined against a database of known sites from the literature. We term these overlapping,

nonphosphorylated forms “occupancy-determining peptides” (ODPs). Based on the ratio of the ODPs, stoichiometries are calculated.

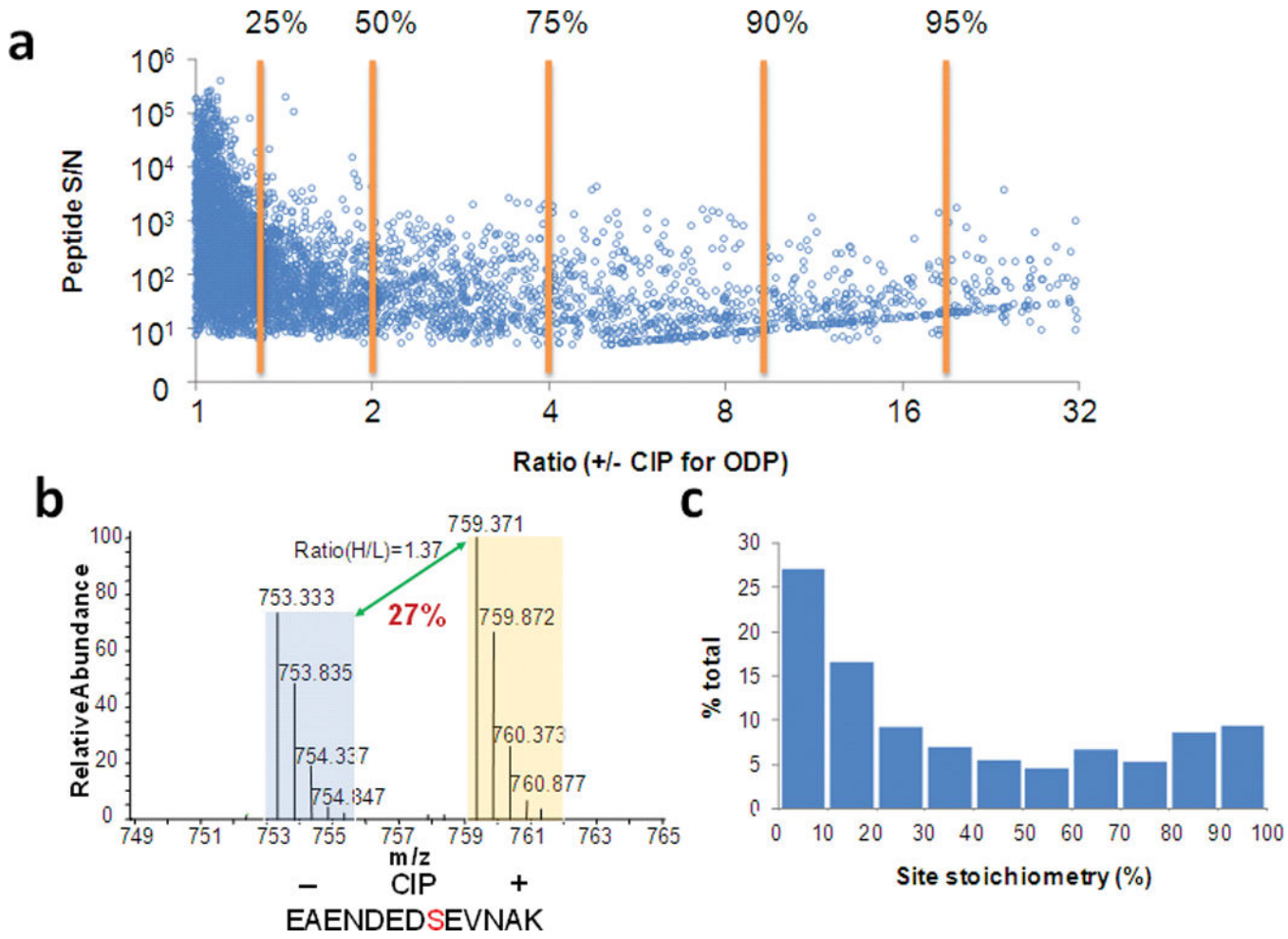
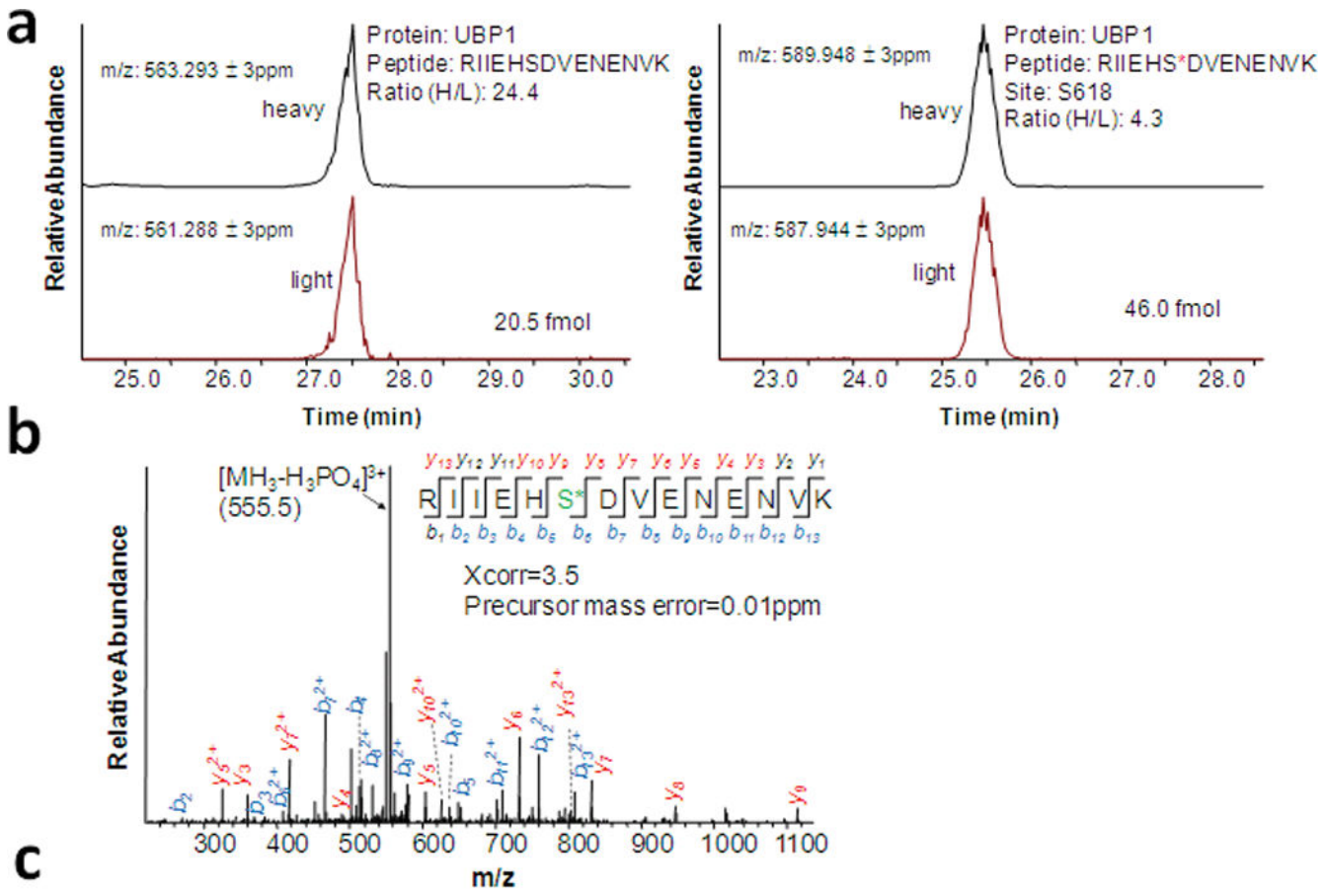


Figure 2.

Absolute site stoichiometries for 5,033 events in exponentially-growing yeast. Three biological yeast replicates were grown to mid-log phase. Lysates were proteolyzed with endoproteinase lys-C. An identical aliquot of each proteolyzed lysate was either mock- or phosphatase-treated as in Fig. 1. Occupancy-determining peptides (5,033) were identified based on the overlap with sites in five published studies[7,10,31–33]. (**a**) Scatter plot of all ODP ratios. Ratios directly encode fractional occupancies (orange lines). (**b**) Example of an ODP from protein Bbc1. Phosphatase treatment increased the amount of the heavy version of the ODP by the amount that existed in the phosphorylated form (27%). (**c**) Site stoichiometry distribution for 5,033 events from WT yeast undergoing exponential growth.



c

Protein	Site	Peptide	Stoichiometry (%) Expt #1	Expt #2	Expt #3	Validation (AQUA)
UBP1	618	RIIEHS*DVENENVK	59	59	66	69.2
BFR2	379	DSVDDNENS*DDGLDIPK	97	99	97	99.8
UBP1	638	DNEELQEIDNVS*LDEPK	75	76	ND	61.5
IPP1	266	AASDAIPPAS*PK	1	1	1	1.8
NUP145	689	AYEPDLSDADFEGIEAS*PK	68	73	71	78.8
DCP2	710	RGET*FASLANDK	ND	1	ND	2.0
YML093W	562	LFES*DEEETNGENIQINK	100	ND	99	99.5
MOT2	92	YDDENVRYVTLS*PEELK	4	14	13	3.5
GNP1	45	QIGS*IEPENEVEYFEK	1	9	10	8.2
TAT1	84	RQLPPDRNSELES*QEK	45	ND	ND	41.4

Figure 3.

Validation of ten sites with varied stoichiometries by AQUA. Synthetic peptides were generated representing heavy phosphorylated and nonphosphorylated versions of the peptide. The synthetic peptides were spiked into proteolyzed lysates and separated by LC-MS/MS techniques (see Methods). Using the heavy peptides as internal standards, the light versions were quantified and stoichiometries were calculated. (**a**) An example of the measured amount of peptide (RIIEHSDVENENVK) and phosphopeptide

(RIIEHS*DVENENVK) by AQUA. The calculated stoichiometry was 69.2%. (**b**) An example of phosphopeptide identification (RIIEHS*DVENENVK) by MS/MS.

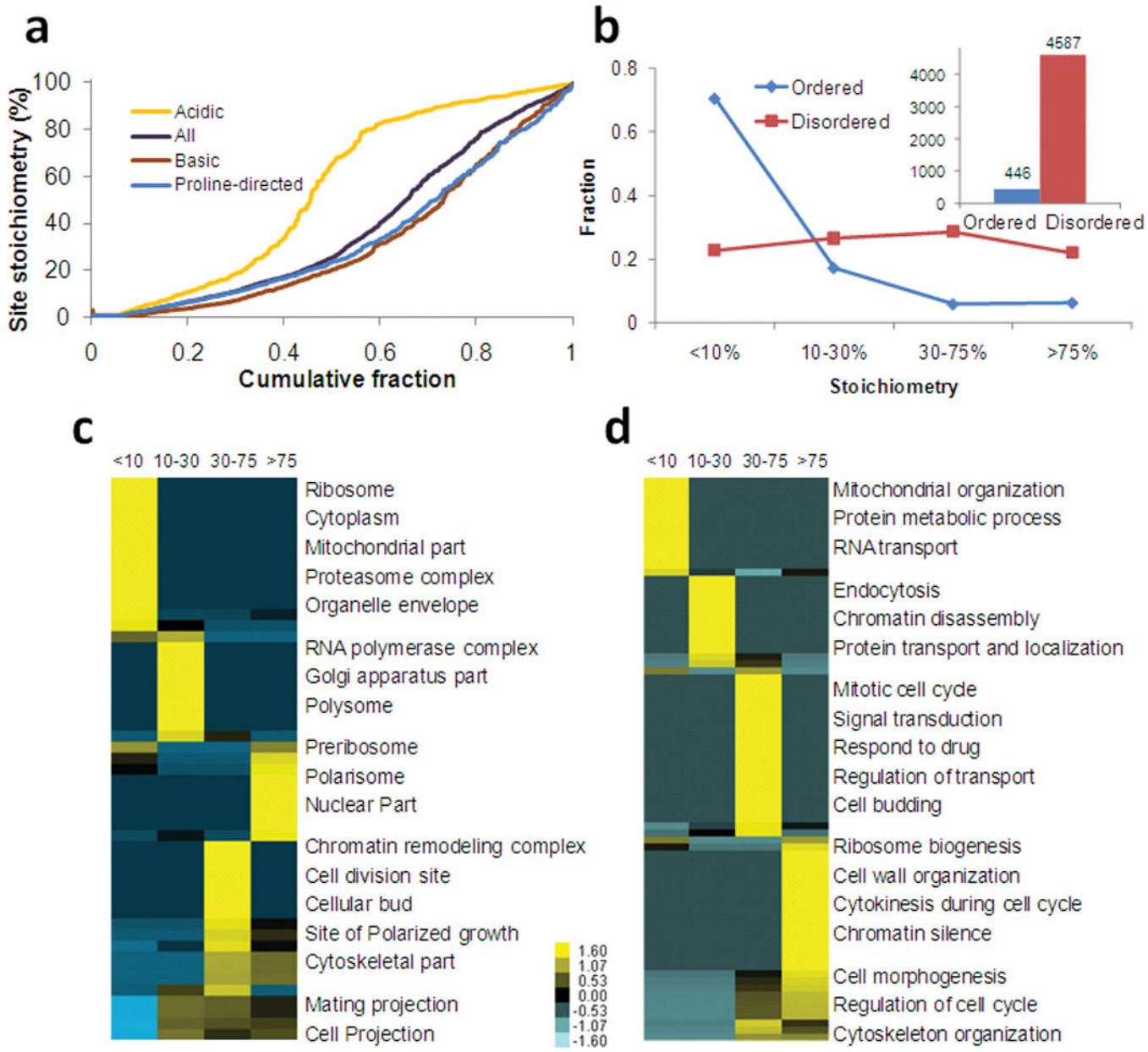


Figure 4.

Bioinformatic analyses of site stoichiometry with respect to kinase motifs and gene ontology. (**a**) Phosphorylation events containing an acidic (i.e., casein kinase II-like) motif, but not basic or proline-directed ones, have higher absolute site stoichiometries on average. (**b**) The number of phosphorylation events in secondary structure predicted to be ordered is inversely proportional to absolute site stoichiometry. The inset shows that most sites are predicted to be in disordered regions. Clustering of the phosphoproteins based on their enrichment in specific biological processes (**c**) and cell compartments (**d**). Phosphoproteins were grouped into four classes according to their highest stoichiometry site and examined for enrichment by the DAVID software[36]. Categories without a P-value were assigned a conservative value of 1. The P-values were log transformed and then z-transformed.

Phosphoproteins were then grouped based on their z-scores via hierarchical clustering. Full versions of these figures are found in Supplementary Fig. 2.

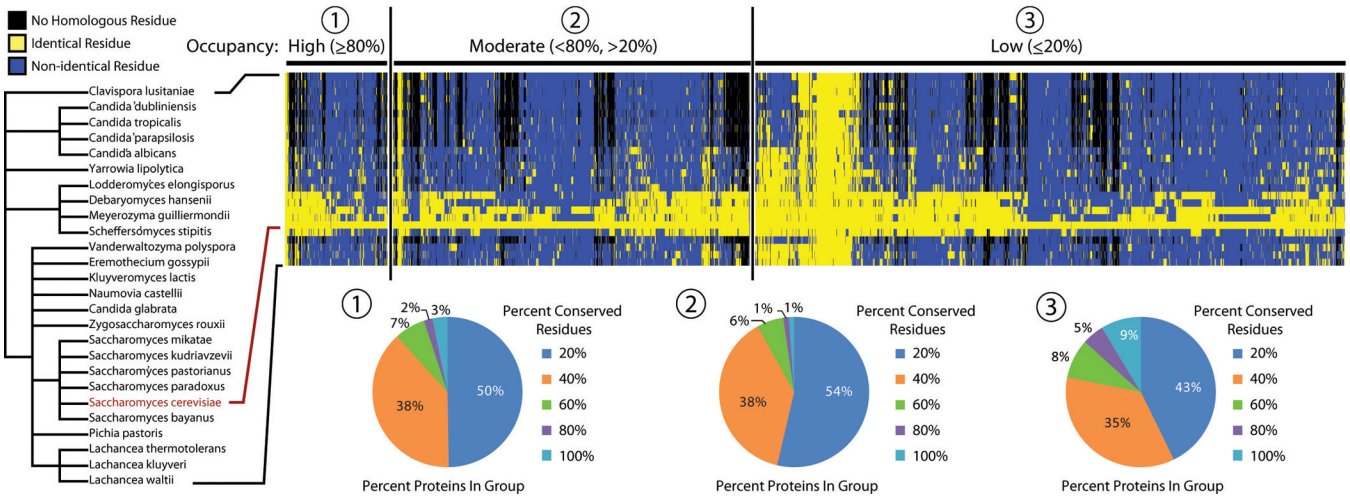


Figure 5.

Evolutionary conservation of the site residues across 25 yeast species. Sites were subgrouped by fractional occupancy into “high,” “medium,” and “low” sets and then clustered based on overall conservation levels. Each column represents a single site residue. If a homolog was identified and the site residue was conserved, the corresponding cell is yellow otherwise it is blue. If no homolog was identified, the cell is black. Most sites were conserved only within the closest neighboring species.

Table 1

Site stoichiometries validated by the AQUA strategy

Protein	Site	Peptide	Stoichiometry (%)			Validation (AQUA)
			Expt #1	Expt #2	Expt #3	
UBP1	618	RIIEHS*DVENENVK	59	59	66	69.2
BFR2	379	DSVDDNENS*DDGLDIPK	97	99	97	99.8
UBP1	638	DNEELQEIDNVS*LDEPK	75	76	ND	61.5
IPP1	266	AASDAIPPAS*PK	1	1	1	1.8
NUP145	689	AYEPDLSDADFEGIEAS*PK	68	73	71	78.8
DCP2	710	RGET*FASLANDK	ND	1	ND	2.0
YML093W	562	LFES*DEEETNGENIQINK	100	ND	99	99.5
MOT2	92	YDDENVRYVTLS*PEELK	4	14	13	3.5
GNP1	45	QIGS*IEPENEVEYFEK	1	9	10	8.2
TAT1	84	RQLPPDRNSELES*QEK	45	ND	ND	41.4

“ND” denotes “not detected” in that analysis.