Resource

Cell

Systematic Functional Prioritization of Protein Posttranslational Modifications

Pedro Beltrao,[1,3,*] Véronique Albanèse,[4] Lillian R. Kenner,[1,3] Danielle L. Swaney,[5] Alma Burlingame,[2,3] Judit Villén,[5] Wendell A. Lim,[1,3,6] James S. Fraser,[1,3] Judith Frydman,[4] and Nevan J. Krogan[1,3,7,*]

[1]Department of Cellular and Molecular Pharmacology
[2]Department of Pharmaceutical Chemistry
University of California, San Francisco, San Francisco, CA 94107, USA
[3]California Institute for Quantitative Biosciences, QB3, San Francisco, CA 94107, USA
[4]BioX Program, Biology Department, Clark Center, Stanford, CA 94305, USA
[5]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[6]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA
[7]J. David Gladstone Institute, San Francisco, CA 94158, USA
*Correspondence: pedro.beltrao@ucsf.edu (P.B.), krogan@cmp.ucsf.edu (N.J.K.)
http://dx.doi.org/10.1016/j.cell.2012.05.036

SUMMARY

Protein function is often regulated by posttranslational modifications (PTMs), and recent advances in mass spectrometry have resulted in an exponential increase in PTM identification. However, the functional significance of the vast majority of these modifications remains unknown. To address this problem, we compiled nearly 200,000 phosphorylation, acetylation, and ubiquitination sites from 11 eukaryotic species, including 2,500 newly identified ubiquitylation sites for *Saccharomyces cerevisiae*. We developed methods to prioritize the functional relevance of these PTMs by predicting those that likely participate in cross-regulatory events, regulate domain activity, or mediate protein-protein interactions. PTM conservation within domain families identifies regulatory "hot spots" that overlap with functionally important regions, a concept that we experimentally validated on the HSP70 domain family. Finally, our analysis of the evolution of PTM regulation highlights potential routes for neutral drift in regulatory interactions and suggests that only a fraction of modification sites are likely to have a significant biological role.

INTRODUCTION

The activity and localization of proteins inside the cell can be regulated by reversible posttranslational modifications (PTMs), including protein phosphorylation, acetylation, and ubiquitylation. How these modifications regulate protein function and how this regulation diverges during evolution is crucial for understanding signaling systems. Recent advances in mass spectrometry (MS) have increased the ability to identify PTMs with thousands of sites now routinely discovered per study (Choudhary and Mann, 2010). However, the functional characterization of these modifications is now rate limiting, a fact further complicated by the recent findings that they can be highly divergent across species (Beltrao et al., 2009; Holt et al., 2009; Landry et al., 2009; Tan et al., 2009). Despite the poor conservation within single proteins, the overall number of phosphosites per protein within different functional modules (i.e., protein complex or pathways) is conserved (Beltrao et al., 2009). This phenomenon could be explained by compensatory turnover of phosphorylation sites, similar to documented cases of compensatory turnover of transcription factor binding sites in promoter regions (Ludwig et al., 2000). The similarities in the evolutionary properties of transcriptional and posttranslational regulatory networks (Moses and Landry, 2010) lend credence to the idea that phenotypic diversity is primarily driven by changes in regulatory networks (Carroll, 2005).

Although phosphosites observed in high-throughput studies are, on average, poorly conserved, sites with a known function are more significantly constrained (Ba and Moses, 2010; Landry et al., 2009). These trends have led some to speculate that there is a substantial fraction of phosphorylation sites that are nonfunctional (Landry et al., 2009; Lienhard, 2008). Conservation of modification sites or regulatory interactions can be used to prioritize experimental validation (Tan et al., 2009) but do not provide a putative functional consequence for the modification. Therefore, developing approaches to dissect the functional importance of PTMs is currently the most significant bottleneck in proteomic studies of posttranslational regulation.

In this study, we experimentally determined 2,500 ubiquitylation sites for *S. cerevisiae* and compiled a list of nearly 200,000 modification sites across 11 eukaryotic species in order to develop predictors of PTM functional relevance. These data, as well as structural information, were used to identify modifications that might regulate protein-protein interactions, mediate domain activity or be part of cross-regulatory events between

different PTMs. We show that sites with predicted function are more likely to be conserved and that conservation of PTMs within domain families identifies important regulatory regions (termed here regulatory "hot spots"). We validate these approaches by experimentally characterizing regulatory hot spots within the HSP70 chaperone domain family and characterizing a phosphosite within Skp1 (part of the Skp1/Cullin/F-box E3 ligase) as likely regulating the interaction between Skp1 and the Met30 F-box protein. In summary, the resource developed in this study, which is accessible online (http://ptmfunc.com), can provide mechanistic functional annotations to PTMs and generate specific predictions for experimental validation. This analysis also allows for a better understanding of the evolution of posttranslational networks and suggests that only a fraction of PTMs is likely to have a regulatory role.

RESULTS

A Resource of Eukaryotic PTMs for Functional and Evolutionary Studies

In order to study the evolutionary properties and functional role of protein posttranslational regulation, we compiled previously published in vivo, MS-derived PTMs (Table S1 available online). We compiled a total of 153,478 phosphorylation sites for 11 eukaryotic species, retaining only sites that have high site localization probability (Experimental Procedures). The phosphorylation data set covers a broad evolutionary time scale with information for three fungi (*S. cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans*), two plant species (*Arabidopsis thaliana* and *Oryza sativa*), three mammals (*Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*) as well as *Xenopus laevis*, *Drosophila melanogaster*, and *Caenorhabditis elegans*. We also compiled 13,133 lysine acetylation sites (covering *H. sapiens*, *M. musculus*, and *Drosophila melanogaster*) and 22,000 human ubiquitylation sites (Emanuele et al., 2011; Kim et al., 2011; Wagner et al., 2011). In addition, we used a MS approach to experimentally determine 2,500 ubiquitylation sites in *S. cerevisiae* to facilitate comparative studies. Using a set of 12 different *S. cerevisiae* phosphoproteomics experiments, we estimate that the curated data sets should have less than 4% of false-positive sites (Table S2).

Previous studies have used sequence conservation to study the evolution of phosphosites (Ba and Moses, 2010; Holt et al., 2009; Landry et al., 2009). In this work, we used the compiled data to directly compare the phosphoproteomes across these 11 species and to evaluate the impact of data quality on the evolutionary observations. We selected one of the species with the highest coverage, the human data set, as reference and compared the data from all other species to it (Figure 1A). We aligned 1-to-1 orthologs of each species to *H. sapiens* proteins, and for each phosphosite, determined the conservation in the human protein of both the phospho-acceptor residue (i.e., sequence conservation) and the phosphorylation site (i.e., phosphosite conservation). In order to account for potential errors in MS phosphosite positional assignments, we considered a phosphosite conserved if the corresponding human ortholog was also phosphorylated within a window of ±2 alignment positions. Both residue and phosphosite conservation were found to be proportional to the divergence age away from *H. sapiens* (Figure 1A). Phosphosite conservation ranged from ~8%–18% for the distantly related plants and fungi to ~40% for the closely related mouse and rat. We then asked whether the observed value was higher than expected by randomly re-assigning the same number of phosphosites within each orthologous protein. As previously described by Landry and colleagues, we observed that the sequence conservation of the phospho-acceptor residue was not higher than expected by chance (Figure 1B) (Landry et al., 2009). However, the conservation of the phosphorylation sites was approximately two to three times higher than random (Figure 1B). This difference suggests that the conservation of the phosphorylation state is a better indicator of functional importance than sequence conservation of the phospho-acceptor residue. For all the following analysis, we used the conservation of the PTM state as the comparative metric.

In order to evaluate the generality of these evolutionary observations across different PTMs, we studied the conservation over random expectation of ubiquitylation and acetylation sites (Figure 1C). We compared the conservation of *S. cerevisiae* phosphorylation and ubiquitylation sites in *H. sapiens* over a random expectation calculated based on random sampling of a similar number of modification acceptor residues within the same proteins. Similarly, we compared the conservation of *D. melanogaster* phosphorylation and acetylation sites in human orthologs. The three modifications show a low level of evolutionary constraint, ranging from 1.3 to 2.2 times higher conservation than expected based on an equivalent random sample of PTM acceptor residues (Figure 1C). Protein acetylation shows a higher value of conservation over random when compared to phosphorylation, consistent with previous work (Weinert et al., 2011), whereas ubiquitylation appears to have a lower evolutionary constraint when compared to phosphorylation (Figure 1C).

Given that these data sets gathered so far are likely to be incomplete, the conservation values presented here are underestimated. We used 12 different *S. cerevisiae* phosphoproteomics experiments to evaluate the error in the conservation estimates by plotting the values of conservation versus coverage for each *S. cerevisiae* phosphoproteomic data set (Figure 1D). Extrapolating from the regression analysis, we estimate that, when corrected for coverage, ~10% of *H. sapiens* phosphosites would be conserved in *S. cerevisiae*. We also calculated the corrected conservation value for each data set independently and estimated the corrected median conservation value for *H. sapiens* phosphosites in *S. cerevisiae* as ~13% (Figure 1D). This value is higher but comparable to the observed 8% conservation measured with the complete data set. This suggests that, at least for the extensively studied phosphoproteomes of *S. cerevisiae* and *H. sapiens*, additional data are unlikely to dramatically change the conservation estimates.

Besides coverage (i.e., false negatives), data quality (i.e., false positives), and low phosphosite abundance are also important factors when estimating phosphoproteome conservation. We compared the conservation of *S. cerevisiae* phosphosites in *H. sapiens* across different data quality criteria, including the number of spectral counts, match to known kinase recognition motifs and information on dynamically regulated phosphosites

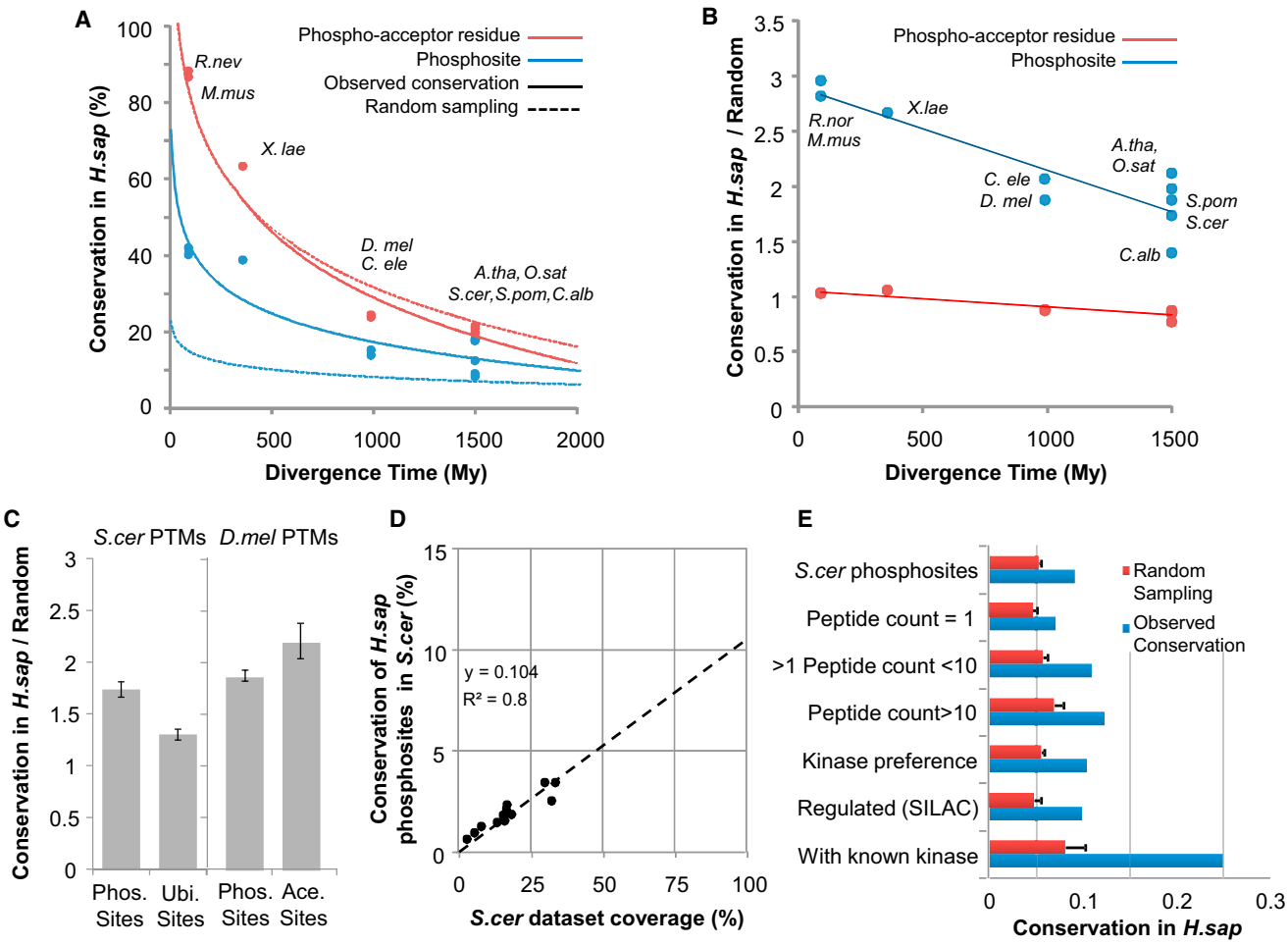

Figure 1. Evolutionary Properties of Eukaryotic Posttranslational Modification Sites

(A) We analyzed the conservation of ten phosphoproteomes against that of *H. sapiens* using protein alignments of 1-to-1 orthologs. For each species, we compared the conservation of the phosphoacceptor residues (i.e., sequence conservation) with the conservation of phosphorylation site based on the MS experimental evidence. A random expectation or null model for each case was determined based on the random shuffling of phosphosite positions within each human protein.

(B) Ratio of observed conservation over random expectation.

(C) Ratio of conservation over random expectation for different PTMs. Error bars represent 1 SD.

(D) Predicted coverage and conservation in *H. sapiens* for 12 different previously reported phosphoproteomics experiments for *S. cerevisiae*.

(E) To access the impact of data quality on the conservation values we made use of different criteria to define subsets of *S. cerevisiae* phosphosites. For each subset we calculated the observed conservation in *H. sapiens* as well as the expected value based on random sampling. Error bars represent 1 SD.

(Figure 1E). The conservation of different classes of phosphorylation sites (Figure 1E, blue bars) was compared to an equivalent random sample (Figure 1E, red bars). To determine statistical significance of the results, the ratios of conserved over expected values for the different phosphosite groups were compared using a Mann-Whitney ranked test. We assumed that spectra and/or peptide count for each phosphosite is correlated with data quality and/or phosphosite abundance and observed that phosphosites supported by multiple spectra/peptides (Figure 1E, peptide count >1) are more likely to be conserved than those observed only once (Figure 1E, peptide count = 1, p value < 10^{-8}). Additionally, sites that represent well-matched kinase-recognition motifs (Figure 1E, "Kinase preference") or are known to be regulated (Figure 1E, "Regulated"), as measured in quantitative MS studies (Holt et al., 2009; Huber et al., 2009; Soufi et al., 2009), are moderately more conserved than average sites and more highly conserved than expected by chance (Figure 1E, "*S.cer* phosphosites," p value < 10^{-9}). Finally, sites that are known to be functionally important (Ba and Moses, 2010) or have described in vivo kinase regulators (http://www.phosphogrid.org) (Stark et al., 2010) are more than three times more conserved than average sites (Figure 1E, "With known kinases" versus "*S.cer* phosphosites," p value < 10^{-16}). These results imply that higher phosphosite functionality, quality, and/or abundance are correlated with conservation and support previous observations made with sequence analysis (Landry et al., 2009). It is likely that low abundance and/or nonfunctional phosphosites, with low conservation, dominate

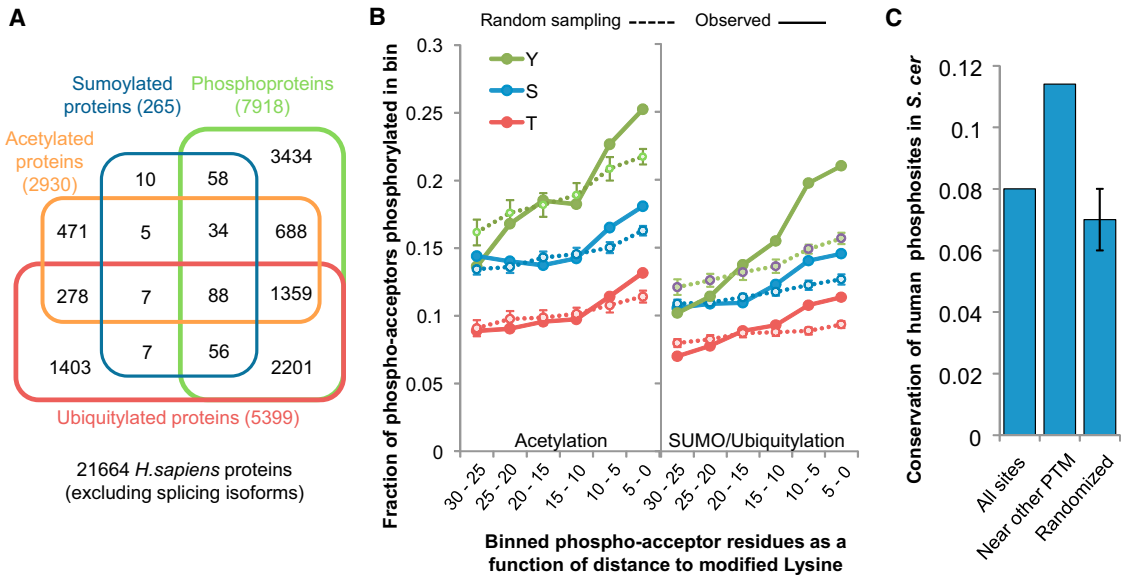

Figure 2. Association of Protein Phosphorylation with Lysine Posttranslational Modifications

(A) A Venn diagram representing the overlap of the different lysine modified proteins with the human phosphoproteome. Whereas 33% of the human proteins are phosphorylated, 71% of the acetylated proteins, 69% of the ubiquitylated proteins, and 89% of the sumoylated proteins are also phosphorylated.

(B) The fraction of the phosphoacceptor residues was plotted as function of the distance to modified lysine residues. Observed values were compared with expected values based on random sampling. Error bars represent 1 SD.

(C) Human phosphorylation sites that are near an acetylated lysine residue (<15 amino acid distance) were more likely conserved in *S. cerevisiae* than average sites or an equivalent random sample. Error bars represent 1 SD.

See also Figure S1 and Table S3.

the overall measured divergence. These results further underscore the need to devise methods to assign functional roles to PTM sites.

Phosphorylation Coupled to Other Classes of Modifications

On average, ~75% of known phosphorylation sites, 40% of acetylation sites and 45% of ubiquitylation sites occur outside known PFAM globular domains (Table S1). It has become increasingly apparent that phosphosites within unstructured regions not only recruit phospho-binding domains but also often regulate other PTMs or localization signals (Hunter, 2007). This complex interplay between PTMs has been previously described in histone tails where they have been proposed to form a code to be read by different effectors and control gene chromatin states (Strahl and Allis, 2000). More recently, examples of cross-regulation between adjacent PTMs have been observed in several other proteins suggesting this to be a universal mode of protein regulation (Hunter, 2007). Examples include the promotion of sumoylation by a priming phosphorylation in several transcription factors (Yang and Grégoire, 2006), the cross-inhibition between adjacent phosphorylation and methylation sites in DNMT1 (Estève et al., 2011) and the positive role of lysine acetylation on the phosphorylation of Cdc6 (Paolinelli et al., 2009).

We hypothesized that it is possible to assign a functional role to PTM sites by searching for the co-occurrence of different modifications within the same protein. For the human proteome, using the information on lysine acetylation, ubiquitylation and sumoylation, we observed a significant overlap between proteins containing these lysine modifications and the phosphoproteome (Figure 2A). Though 36% of all proteins are phosphoproteins, >69% of proteins containing any of these lysine modifications are also phosphorylated (Figure 2A). This enrichment is highly significant (p value $< 1 \times 10^{-70}$, with a Fisher's exact test) and not merely due to MS detection bias for abundant proteins (Table S3). We next asked whether these PTMs tend to cluster within the protein sequence (Figure 2B). Given the small number of currently characterized sumoylation sites, we grouped these together with ubiquitylation for the analysis. We binned phospho-acceptor residues (i.e., serine, threonine, and tyrosine) according to their smallest distance to a modified lysine residue. In each distance bin, we then calculated the fraction of acceptor residues that is phosphorylated and compared this observed value with a random expectation by randomly reassigning the same number of phosphorylation sites within each protein. We observed that, on average, phospho-acceptors near modified lysines are preferentially phosphorylated when compared to more distant residues or an equivalent random sample of sites. These results show that the different PTMs tend to cluster within protein sequences. This result is not merely due to preferential accumulation of PTMs in unstructured regions (Figure S1) and was also observed using phosphorylation and lysine acetylation data for mouse and phosphorylation and ubiquitylation data for *S. cerevisiae* (Figure S1). If phosphorylation sites near other PTMs are more likely to be functionally relevant, then we assumed that these should also show higher conservation. We tested this by comparing the conservation, in

S. cerevisiae, of all human phosphorylation sites with those within 15 amino acids of another PTM. The conservation of phosphorylation sites near modified lysines were higher than for average phosphosites and higher than an equivalent random sample (p value < 10^{-7}) (Figure 2C).

Regulation of Protein-Protein Interactions

For the PTM sites that occur within structured regions, we can make use of the growing structural knowledge deposited in the PDB (http://www.pdb.org) to assign putative functional roles for the protein modifications, especially those found within protein interfaces, since they may be involved in the regulation of protein-protein interactions. For human or *S. cerevisiae* protein-pairs that are known to physically interact, we used X-ray structures, homology models or docking solutions to define the most likely interface regions (Experimental Procedures). Using these models, we identified 870 phosphorylation, 632 ubiquitylation, and 263 acetylation sites at putative interface residues that can potentially regulate protein-protein interactions (available at http://ptmfunc.com). To expand the number of putative interface residues, we made use of the 3DID database of domain-domain interactions (Stein et al., 2011). For each domain family, as annotated in PFAM (http://pfam.sanger.ac.uk), 3DID contains annotations of what residues have been shown to participate in physical interactions in X-ray structures. We have used these annotations to assign interaction residues for PFAM domains in the 11 proteomes (Experimental Procedures) and identified 3,968 phosphorylation, 1,802 ubiquitylation, and 1,691 acetylation sites that potentially regulate protein-protein interactions (http://ptmfunc.com). Using either of these definitions, we observed that *S. cerevisiae* phosphosites at interface residues are approximately two to three times more likely to be conserved in *H. sapiens* than average phosphosites (Figure 3A, "Interface residue" or "PFAM interaction residue" versus "All phosphosites," p value < 10^{-14}). It is known that globular domain regions are easier to align than the unstructured regions where most phosphosites occur. However, the higher conservation of phosphosites at interface residues is not merely due to alignment issues since phosphosites that occur within PFAM domains are not more conserved than average sites (Figure 3A). We confirmed these evolutionary trends using interface models for human protein-protein interactions (Figure S2A).

In order to test the generality of some of these observations across different posttranslational modifications, we compared the conservation (over random expectation) of all acetylation, ubiquitylation and phosphorylation sites with those occurring at predicted interface residues (Figure 3B). In line with the observations made for protein phosphorylation, lysine acetylation at interface residues is more likely to be conserved (Figure 3B, p value < 10^{-5}), however, ubiquitylation at interface residues shows a similar level of constraint when compared to average ubiquitylation sites. These results suggest that phosphorylation and acetylation but not ubiquitylation sites at interface residues are more likely to be functionally important than average sites, suggesting that these PTMs are commonly used by the cell to reversibly regulate the binding affinity of protein interactions.

The analysis of protein-protein interfaces creates specific predictions for the functional role of PTMs. For example, several alpha subunits of the proteasome are phosphorylated at interface regions (Figures S2B and S2C). Serine 13 and tyrosine 5 of Pre8 (the *S. cerevisiae* alpha 2) are phosphorylated in yeast and human, respectively, and could potentially regulate the interactions with Pre9 (alpha 3). The N terminus of alpha 5 is also phosphorylated in 7 of the 11 species. This N-terminal region has been shown to be important for proteasome activity (Groll et al., 2000) indicating that these N-terminal phosphorylations might regulate the interactions between alpha subunits or the activity of the proteasome (Figures S2B and S2C). Similarly, we predicted that a phosphosite at position S162 in the *S. cerevisiae* Skp1 could regulate the interaction with Met30 (Figure 3C). Skp1 is a highly conserved protein that is part of the Skp1/Cullin/F-box (SCF) multisubunit E3 ubiquitin ligase complex (Petroski and Deshaies, 2005). Skp1 interacts with different F-box domain containing proteins that can modulate the ubiquitylation substrate specificity (Petroski and Deshaies, 2005). In *S. cerevisiae*, Skp1 can interact with the Met30 F-box protein to regulate proteins involved in sulfur metabolism (Jonkers and Rep, 2009) an interaction that is known to be regulated under different stress conditions (Jonkers and Rep, 2009). We postulated that the highly conserved phosphorylation site in Skp1 might regulate the interaction with Met30 and/or other F-box proteins (Figure 3C). We note that given the position of the residue at the end of the helix, it would have to partially unwind to adopt a coil conformation that can access the kinase active site.

To experimentally probe the dependency of the Skp1:Met30 interaction on the phosphorylation status of S162, we used a protein complementation assay (Ear and Michnick, 2009; Michnick et al., 2010) that reports on the strength of the protein-protein interaction in vivo. We fused Skp1 and Met30 to two fragments of the yeast cytosine deaminase and transformed the constructs into a strain that lacks the endogenous enzyme. Skp1 and Met30 interact directly in vivo allowing the two fragments to reconstitute cytosine deaminase activity. Reconstitution of enzyme activity permits growth on media lacking uracil (-Ura) and leads to death on media containing 5-fluorocytosine (+5-FC) (Figure 3D). To test the idea that phosphorylation reversibly regulates the assembly of this interaction, we mutated S162 to alanine (S162A) or the phosphomimetic aspartic acid (S162D). The S162A mutant, similar to wild-type, supported growth on –Ura media, which selects positively for interacting proteins, and grew poorly on +5-FC media, which counter-selects for interacting proteins, indicating that the unphosphorylated state binds Met30 (Figure 3D). In contrast, the S162D mutant grew better on +5-FC than on –Ura media, indicating that the phosphorylated state binds Met30 weaker than the unphosphorylated state (Figure 3D). To validate this result, Flag-tagged Skp1-S162A and Skp1-S162D were immunoprecipitated in the presence of Met30-Myc, and the Met30:Skp1 interaction was monitored using an α-Myc antibody. We found that the Skp1:Met30 interaction is impaired in the phosphomimetic mutant, but not in the alanine mutant (Figure 3F), suggesting that the phosphorylation of S162 acts as a reversible switch for Met30 affinity.

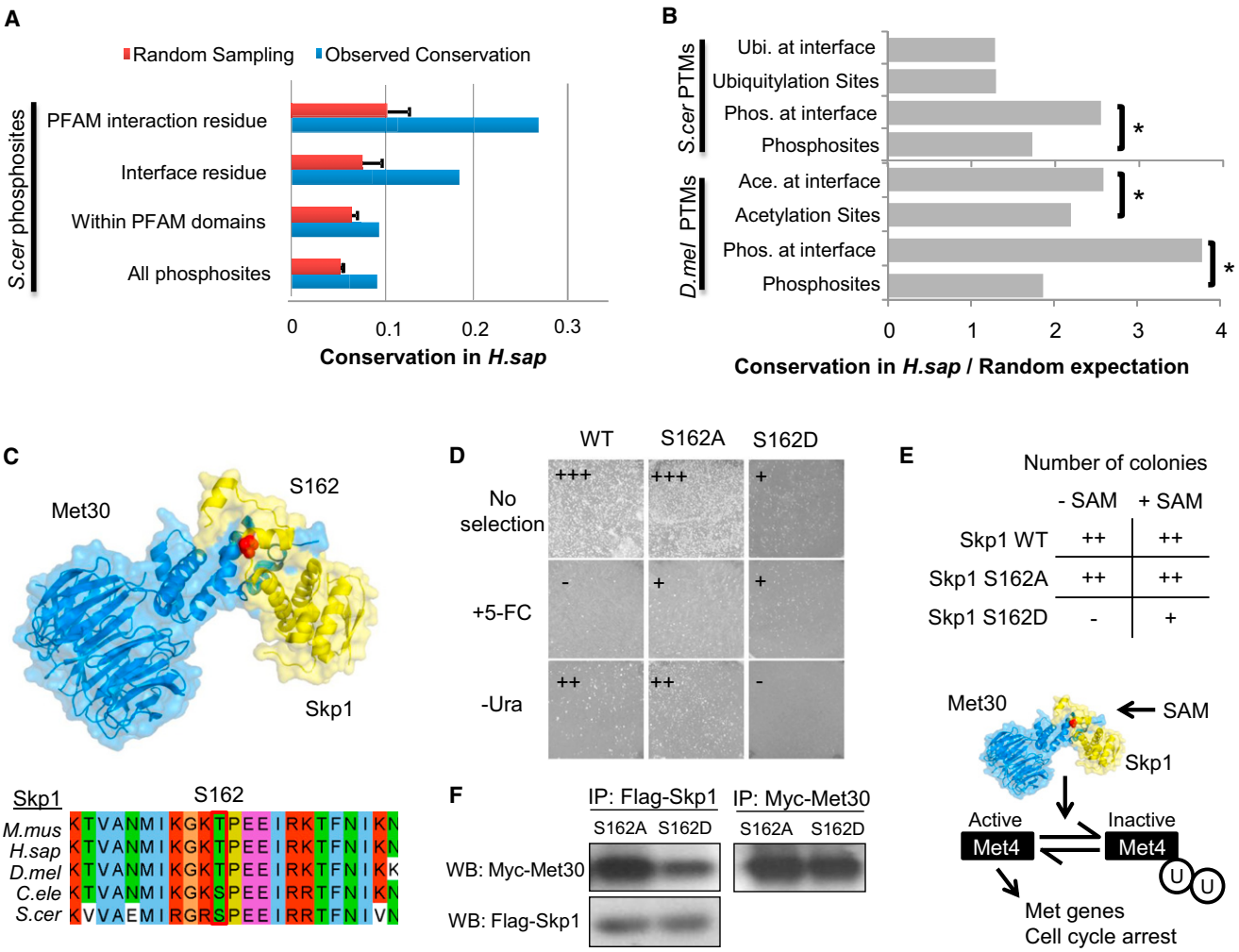

Figure 3. Regulation of Interface Residues by PTMs

(A) The conservation of all *S. cerevisiae* phosphosites in *H. sapiens* was compared with the conservation of phosphosites within PFAM domain interface residues (from the 3DID database) and at putative interface residues (from interaction models). Error bars represent 1 SD.

(B) We compared the conservation over random expectation of different PTMs at PFAM interface residues (from the 3DID database). The ratios of conservation over random expectation were compared using a Mann-Whitney rank test. *p value < 0.05.

(C) The Skp1:Met30 interaction model. The S162 position of the *S. cerevisiae* Skp1 was found to be phosphorylated in *S. cerevisiae*, *H. sapiens*, *M. musculus*, and *C. elegans*.

(D) A protein complementation assay of cytosine deaminase activity reports on the Skp1:Met30 interaction. The phosphomimetic mutation reversed the growth pattern on selective media reporting on interaction strength by growing on +5-FC plates and without any observable growth on –Ura plates suggesting that the in vivo affinity of the Skp1:Met30 complex is reduced by phosphorylation.

(E) Skp1 wild-type and mutants (S162A and S162D) were plated in the presence or absence of SAM. Overexpression of the S162D mutant resulted in poor growth, a phenotype that was relieved in the presence of SAM.

(F) Skp1 S162D mutation shows a significant decrease in bound Met30 when compared to the S162A mutant by coimmunoprecipitation.

See also Figure S2.

The Skp1:Met30 interaction is required to keep the Met4 transcription factor inactivated via ubiquitylation (Kaiser et al., 2006). The activation of Met4 regulates genes involved in the biosynthesis of sulfur-containing amino acids and glutathione metabolism but it also results in cell cycle arrest (Aghajan et al., 2010). During our interaction studies, we observed that overexpression of Skp1 S162D resulted in poor growth, a phenotype that was not observed with the overexpression of Skp1 WT or S162A mutant (Figure 3E). These results suggest that Skp1 S162D impairs the interaction with Met30 resulting in an activation of Met4 and cell cycle arrest. The Met4 inactivation by Skp1:Met30 is known to be promoted by SAM (Kaiser et al., 2006). Consistent with the hypothesis that Skp1 S162D overexpression results in Met4 activation and cell cycle arrest, growth in the presence of SAM relieves the impaired growth (Figure 3E), presumably by further activating the available pool of Skp1:Met30 and/or relieving independently a cell cycle block. These collective results strongly suggest that the

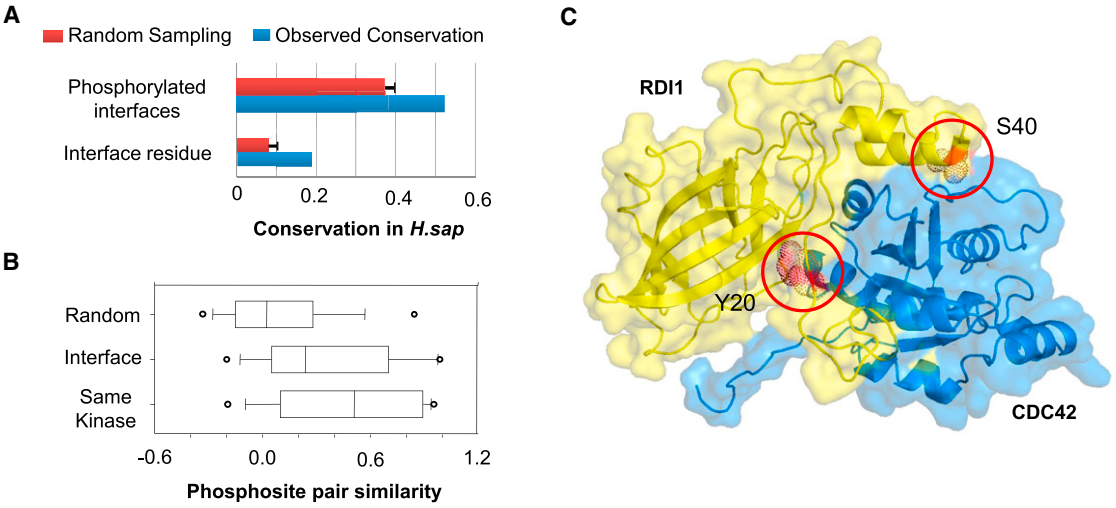

Figure 4. Conservation of Interface Phosphorylation Can Be Achieved by Regulation of Different Positions

(A) We compared the conservation of all *S. cerevisiae* interface phosphorylation sites in *H. sapiens* (Interface residues) with the conservation of the phosphorylation of *S. cerevisiae* interfaces in *H. sapiens* without regard to the actual phosphosite position (Phosphorylated interfaces). Error bars represent 1 SD.

(B) A metric of phosphosite similarity (Experimental Procedures) was used to compare phosphosite pairs found at the same interfaces in the two different species (*S. cerevisiae* and *H. sapiens*) with random phosphosite pairs and pairs known to be regulated by the same kinases. Open circles represent the top and bottom fifth percentiles.

(C) The model of the *S. cerevisiae* Cdc42p:Rdi1p interface was annotated with currently known phosphorylation data. The Rdi1p Y20 and S40 positions refer to the *S. cerevisiae* protein sequence positions. The S40 position is currently known to be phosphorylated in *S. cerevisiae* and *C. albicans* but not in human. Conversely the Y20 position is known to be phosphorylated in human but not in fungi.

phosphorylation of Skp1 on residue S162 has the potential to reversibly alter the binding affinity of Skp1 with Met30 and regulate the function of Skp1:Met30.

Next, we used the interface models to differentiate between the conservation of phosphorylation sites at interface residues from the conservation of the predicted function (i.e., regulation of the interaction). Intuitively, one can imagine that the phosphorylation of an interface might be conserved despite the divergence of the actual phosphosite position. We observed that >50% of the interfaces that are phosphorylated in *S. cerevisiae* are also phosphorylated in *H. sapiens* despite only ~18% of the interface phosphorylation sites showing positional conservation (Figure 4A). However, given that the current phosphoproteomes are likely to be incomplete, we cannot rule out that some of the observed positional divergence is not due to a coverage issue. A similar trend is observed using human interface models (Figure S2A). If the conservation of function with divergence of phosphosite position is mostly the product of a neutral variation, we might expect to observe a conservation of the kinase recognition for the phosphosites at the same interface. To study this issue, we devised a metric of phosphosite similarity based on the models of binding preferences of 63 *S. cerevisiae* kinases and calculated the similarity of *S. cerevisiae* interface phosphosites with human phosphosites at the same interface (Experimental Procedures). We then compared these scores with the similarity scores for random pairs of phosphosites and sites known to be regulated by the same kinases (Figure 4B). The median phosphosite similarity for interface phosphosites is higher than for random pairs (p value $< 2 \times 10^{-16}$ with a Kolmogorov-Smirnov test) suggesting that a significant fraction of phosphosites observed at the same interface in different species are phosphorylated by kinases of similar specificity.

An example of conserved phosphorylation of an interface at different positions is shown in Figure 4C for the interaction between the *S. cerevisiae* Rho family GTPase Cdc42p and the Rho inhibitor Rdi1p. Rdi1p is phosphorylated at the S40 position in *S. cerevisiae*. Although the S40 equivalent position is phosphorylated in the *C. albicans* ortholog, it is currently not know to be phosphorylated in human. Instead, the Rdi1p Y20 position is phosphorylated in the human ortholog (Figure 4C), but it is currently not know to be regulated in fungi. Regulation of Rho-inhibitor interactions by phosphorylation has been previously described as an important mechanism for the control of the function of Rho proteins (DerMardirossian et al., 2004). Our analysis suggests that the phospho-regulation of the Cdc42:Rdi1 might be highly conserved but achieved by the phosphorylation of different positions in different species.

Posttranslational Hot Spots within Domain Families

We show above that PTMs with putative functional annotations are more likely to be conserved across species than average sites. We hypothesized that we could use conservation to identify regions within domain families with high regulatory potential. Ten domain families that are extensively phosphorylated across the 11 species with available phosphorylation data were initially selected for this analysis (Table S4). For each domain family, we selected a representative sequence/structure from the PDB (Table S4), then aligned each domain from the 11 species to the representative sequences/structures and mapped to them

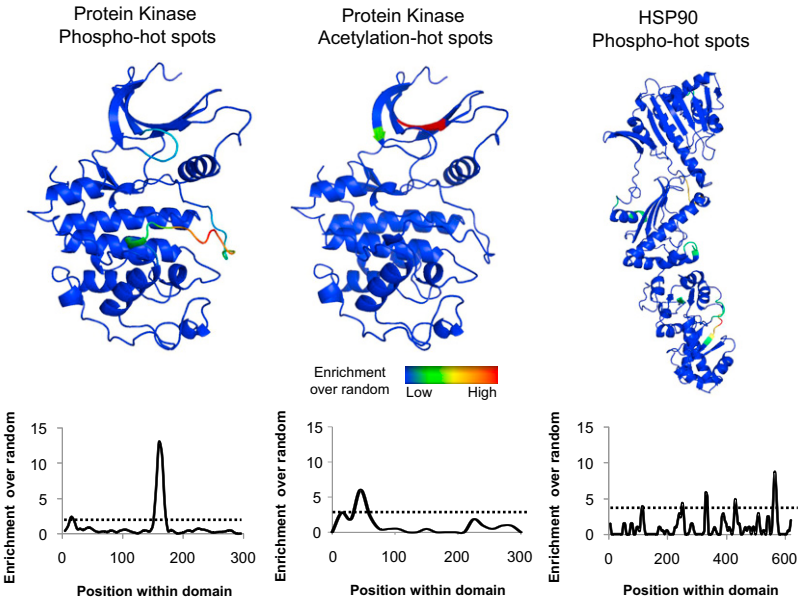

Figure 5. Phosphorylation Enrichment Analysis Identifies Regulatory Hot Spots

For each domain family under analysis, we selected a representative structure from the PDB database (Table S4). Phosphorylation and acetylation data was mapped to the representative sequence/structure using sequence alignments and random sampling was used to calculate the enrichment over random. This value was plotted along the domain sequence position of the representative structure as a moving average and a cutoff of p value < 0.005 was used to identify significant enrichment (dotted line). See also Figure S3.

all phosphorylation sites. Putative regulatory regions were identified by calculating the enrichment of phosphosites over random expectation. Significantly enriched domain regions, or regulatory hot spots, were determined based on random sampling with a p value cutoff of 0.005 or less (Experimental Procedures). We hypothesize that phosphorylation of residues within these regulatory hot spots are more likely to regulate domain function. A similar analysis for lysine acetylation was performed for the protein kinase domain.

In Figure 5, we show the sequence and structural mapping for the enrichment of phosphorylation or acetylation sites for two example domains (protein kinase and HSP90). As expected, the most significantly phosphorylation enriched region for the kinase family is the activation loop region (Nolen et al., 2004). A second hot spot of phospho-regulation was observed within the ''glycine-rich'' loop that contributes to ATP binding and has been described to activate or inhibit the activity of kinases, in particular, CDKs (Narayanan and Jacobson, 2009). Interestingly, there is no significant enrichment of acetylation sites within the activation loop of kinases but instead these are preferentially observed within the N-terminal lobe region. This enrichment is primarily due to a catalytic lysine residue that is often observed to be acetylated, a modification that has been previously shown to be important for the regulation of kinase activity (Choudhary et al., 2009).

The HSP90 domain family is a highly conserved dimeric heat-shock protein family that facilitates the folding of client proteins involved in a multitude of biological functions (Taipale et al., 2010). We identified 145 phosphorylation sites within members of the HSP90 domain that were preferentially enriched in the C-terminal region (Figure 5). The strongest enrichment segment corresponds to the residues 600–610 of the yeast HSP90 (*HSC82*) sequence that projects from the C-terminal region and forms contacts with the equivalent segment of the opposing dimer (Ali et al., 2006). Phosphorylation of this region is therefore likely to regulate HSP90 function. It has been shown that the Ppt1 phosphatase binds to the HSP90 C-terminal region and that the disruption of this interaction results in hyperphosphorylation and misregulation of HSP90 (Wandinger et al., 2006). Consistent with these ideas, Soroka and colleagues validated this prediction by demonstrating that the 600–610 region of the yeast HSP90 is in fact regulated by Ppt1 and that phosphorylation of this region has the potential to regulate HSP90 function (Soroka et al., 2012).

We believe that this enrichment approach can be used to study the regulatory potential of different domain families and we provide additional examples in Supplemental Information (Figure S3).

Regulation of the HSP70 Domain Family by Protein Phosphorylation

The results above strongly suggest that our statistical enrichment analysis can highlight functionally important sites subject to regulation by PTMs. In order to further validate this approach, we studied in more detail the regulation of the heat shock 70 kDa (HSP70) domain family. The HSP70 is a highly conserved chaperone that folds client proteins through an ATP-dependent cycle of binding and release (Kampinga and Craig, 2010). HSP70 proteins are constituted of two domains, an N-terminal nucleotide binding domain (NBD) and a C-terminal substrate binding domain (SBD) (Figure 6A). Although the HSP70 family has been extensively studied and is implicated in a myriad of cellular functions (Kampinga and Craig, 2010), its regulation by protein phosphorylation has not been previously explored.

We identified 313 phosphosites within HSP70 proteins across the 11 species and our enrichment analysis highlighted two significant hot spots (Figure 6A). Strikingly, both of these mapped to functionally and structurally important regions, one near the nucleotide binding pocket (Region 1) and the second near the entrance to the peptide binding groove (Region 2). The two regions were then used to predict the corresponding regulatory phosphosites in *SSA1*, an abundant cytosolic HSP70 in the budding yeast. *SSA1* has been involved in multiple cellular functions, including binding to polysomes and nascent chains, and assisting the refolding of newly made and stress-denatured polypeptides, as well as prevention of protein aggregation, the posttranslational translocation of newly synthesized secreted proteins into the endoplasmid reticulum (ER) and mitochondria,

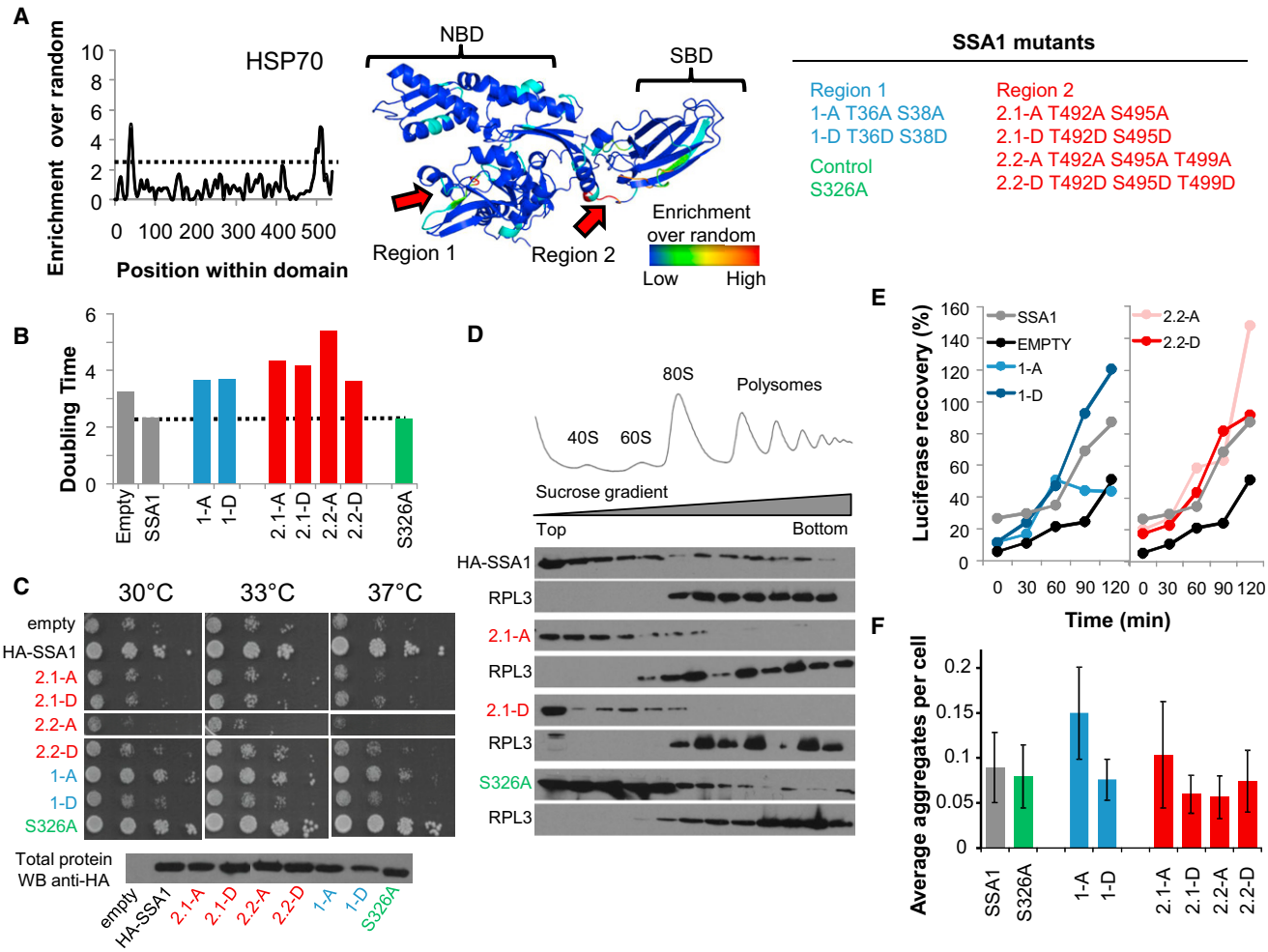

Figure 6. Phosphorylation Hot Spots within the HSP70 Domain Family

(A) A cutoff of p value < 0.005 (dotted line) was used to identify two regions that are significantly enriched for phosphosites in the HSP70 family.

(B) Phosphorylation hot spot mutants do not complement SSA1 functions as measured by growth in liquid media.

(C) Ten-fold dilution series of Ssa1 mutants on –URA plates and incubated at 30°C, 33°C, and 37°C for 2 days. The protein abundance of the different mutants was compared to WT Ssa1 by western blot.

(D) The association of the indicated *SSA1* mutants with polysomes was examined by immunoblot analysis. Ribosomal profiles (top) were determined by OD 254 nm and confirmed by immunoblot analysis of the ribosomal proteins Rpl3p.

(E) Recovery of luciferase activity is expressed as a percentage of activity before heat treatment and is an average of 2 experiments.

(F) Percentage of cells with multiple and single ubc9-2-GFP aggregates after heat shocked at 37°C for 30 min. Errors bars quantify the standard deviations of, at least, four technical replicates.

See also Figure S4.

and degradation of misfolded proteins (Albanèse et al., 2006; Horton et al., 2001). To test the functional relevance of the predicted sites, *SSA1* constructs were designed with alanine or phosphomimetic mutations of residues that were known to be phosphorylated and within these hot spot regions. Two closely spaced phosphorylated threonines were mutated in Region 1 (T36, T38) and three phosphosites were mutated in Region 2 (T492, S495, T499) (Figure 6A). We also mutated serine 326, a position known to be phosphorylated but outside the hot spot regions to serve as a control. Since the cytosol of yeast contains four nearly identical SSA homologs (*SSA1–4*) the different *SSA1* mutants were studied in two yeast strains engineered to lack cytosolic Hsp70 function: (1) a strain lacking *SSA2–4* and containing a single copy of *SSA1* with a temperature sensitive point mutation renders it inactive above 37°C (ssa1-45) and (2) a strain lacking both *SSA1* and *SSA2*, but containing functional copies of the less abundant *SSA3* and *SSA4*. Similar results were obtained in both types of cells.

Growth of the wild-type and mutant strains were measured in liquid culture (Figure 6B) or using serial spot dilution assays (Figure 6C). Though the control phosphorylation mutant behaved like wild-type, none of the phosphorylation mutants in Regions 1 and 2 were able to fully complement the growth even under nonstress conditions of 30°C, indicating that the regulatory hot

spot phosphorylation sites are important for *SSA1* function. In addition, we performed serial spot dilution assays on the single alanine and phosphomimetic mutants of both regions (Figure S4A). With the potential exception of S38D, all single mutants displayed a growth defect under heat shock conditions that is not observed when the WT Ssa1 is expressed or when a control mutation T326A is introduced (Figure S4A). Importantly, the protein abundance of the mutants was comparable to WT Ssa1 (Figure 6C). However, no dramatic differences were observed between the alanine or phosphomimetic mutants suggesting that either the phosphorylation cycle is important for the function of Ssa1 or alternatively, the phosphorylation state of region 1 and region 2 could distinctly affect the function of Ssa1 in the multiple distinct cellular tasks required for cell growth. To obtain further insight we explored the effect of the phosphorylation mutants on a small subset of assays reporting on different Hsp70/SSA functions, namely: association with polysomes (Figure 6D); refolding of firefly luciferase following heat stress (Figure 6E) and prevention of misfolded protein aggregation (Figure 6F). We examined the association of Ssa1 with polysomes by fractionation extracts on 7%–47% sucrose gradients followed by western blot analysis for the presence of WT or mutant Ssa1 as well as the ribosomal protein Rpl3 (Figure 6D). Both Ssa1 WT and the control mutant associated with polysomes as previously reported (Albanèse et al., 2006). However, Ssa1 mutated in Region 2 were defective in binding to polysomes. The Ssa1 mutants of Region 1 show similar to wild-type association with polysomes (Figure S4B). In addition we observed that a single phosphomimetic mutation in region 2 (S495D) is sufficient to disrupt the association with polysomes (Figure S4C), a phenotype not observed for the equivalent alanine mutant (S495A) (Figure S4C). These data suggest that the regulatory hot spot we identified in Region 2 may be involved in binding to nascent chains or in mediating interactions between the translational cofactors Sis1 and Pab1 to the ribosome (Horton et al., 2001).

Hsp70 also assists the refolding of heat-denatured polypeptides, a function that can be monitored by following the recovery of luciferase enzymatic activity following heat stress (Experimental Procedures). As expected, the cells containing wild-type Ssa1 showed robust recovery of luciferase activity, whereas little recovery was observed in the SSA-defective cells transformed with the vector control (Figure 6E). Most phosphorylation mutants performed similar to wild-type in this assay; however, a phosphorylation incompetent variant in Region 1 (i.e., the nucleotide binding site) was significantly impaired in assisting the recovery of stress-denatured luciferase when compared to WT or a phosphomimetic mutant. Since Hsp70 is also implicated in the prevention of aggregation following misfolding, we also tested the effect of the phosphorylation mutant to prevent the aggregation of a temperature-sensitive (ts) mutant of Ubc9 under nonpermissive conditions (Figure 6F). Similar to what was observed for the recovery of stress-denatured luciferase, the phosphorylation incompetent Ssa1 mutant (i.e., alanine mutant) in Region 1 was impaired in the ability to prevent aggregate formation (Figure 6F) when compared to WT or the phosphomimetic mutant. The phosphomimetic mutant appears to have a similar to WT luciferase recovery and aggregation prevention capacity.

Taken together, these results indicate that the two conserved phosphorylation hot spots in Hsp70 are functionally relevant, and distinctly affect various Hsp70-dependent activities. Because phosphorylation of Hsp70 had been previously observed, our approach provides evidence that such phosphorylation is important for Hsp70 regulation. Given the many distinct functions of Hsp70 during the life cycle of the cell our results open the way to future studies dissecting the precise contribution of regulation of each region to overall Hsp70 function as well as the modulation of its activity under various growth conditions.

DISCUSSION

We have compiled a resource of nearly 200,000 PTMs covering 11 eukaryotic species and developed approaches to annotate PTMs that are more likely to cross-regulate each other or to regulate protein-interfaces or domain activity. To make this resource easily available to others, these data are available through a website (http://ptmfunc.com) that contains known PTMs, spectral counts, information on conditional regulation, conservation and putative functional assignments. Using these methods, we have identified a phosphorylation site within Skp1 that is likely to reversibly alter the binding affinity of Skp1 to Met30. Given the position in the crystal structure, it is possible that the phosphorylation of Skp1 at S162 acts by sterically or electrostatically repulsing Met30. However, given that Skp1 interacts with other F-box proteins, it is also possible that the phosphorylation increases the affinity for another protein and titrates Skp1 away from Met30. Based on the assumption that conserved sites are more likely to be functionally relevant, we have identified regions within domain families that show significant enrichment of PTM sites across the 11 species analyzed here (regulatory hot spots). We have experimentally characterized two such regions within the HSP70 chaperone domain family showing that they affect Hsp70 function and provide additional examples for future studies. Putative functional annotations for 8,776 phosphosites from these 11 species are available through our website.

Besides the functional prioritization, this resource can also be used to study the evolution of posttranslational regulation. Past work on the evolution of cellular interaction networks has shown that, whereas protein complex membership diverges slowly and mostly through subunit duplication (Pereira-Leal et al., 2007; van Dam and Snel, 2008), cellular interactions of broad specificity such as protein interactions mediated by small peptide ''linear motifs'' diverge quickly (Beltrao and Serrano, 2007; Neduva and Russell, 2005). Our analysis of 11 partial phosphoproteomes further validates previous observations regarding the fast divergence of kinase-substrate interactions (Landry et al., 2009). In addition, we show that lysine modifications are equally poorly constrained when compared to an equivalent random sample of lysine acceptor residues.

Although regulatory interactions diverge quickly, it is possible for these changes to be neutral in respect to phenotype, much in the same way that mutations within open reading frames can be neutral to the coding sequence. Examples include the conservation of the mating-type logic, regulation of the origin of replication

complex and ribosome transcriptional regulation, despite the divergence of the underlying interactions in some fungal species (Lavoie et al., 2010; Moses et al., 2007; Tsong et al., 2006). The existence of equivalent or "neutral networks" as described by Andreas Wagner, among others, may in fact be important for the exploration of novel phenotypes (Wagner, 2008). We observed that PTMs that are known to be regulated in vivo or are predicted to have a function are more likely to be conserved across species than average sites. One explanation for these results would be that there is a significant fraction of PTMs that serve no functional role but result simply from the constant evolutionary turnover of regulatory interactions. We also observed several examples of conservation of the phosphorylation of a protein-protein interface despite the divergence of the actual phosphosite position. Although we cannot rule out that this observation is due to incomplete phosphoproteome coverage, it provides specific examples of possible neutral variation of PTM regulation for future studies.

The development of PTM functional classifiers will improve our fundamental understanding of cellular signaling but could also, in the long term, have practical biomedical applications. This is particularly relevant for the study of disease since it has been recently shown that disease causing mutations are significantly associated with PTM sites (Li et al., 2010). This resource can aid in the understanding of how mutations associated with PTMs result in disease.

EXPERIMENTAL PROCEDURES

Posttranslational Modification Sites and Computational Methods

All of the sites compiled are provided in a searchable website (http://ptmfunc.com). Protein sequences, protein identifiers and ortholog assignments were obtained from the Inparanoid database (http://inparanoid.sbc.su.se, version 7). For the comparative analysis we considered only 1-to-1 orthologs with Inparanoid confidence scores greater than 90%. The total number of human to species ortholog pairs used in this studied are listed in Table S5. Protein sequence alignments were done with MUSCLE version 3.6 (Edgar, 2004). Additional information on the computational methods is provided in the Extended Experimental Procedures.

Immunoprecipitation, Mass Spectrometry, and Ubiquitylation Site Identification

S. cerevisiae Sub592 (containing a HisTag modified ubiquitin) and Sub62 were grown separately in YPD and harvested during mid-log phase (OD 600 ~1.0). Protein extract from Sub592 cells (~40 mg) was enriched for ubiquitylated proteins via HisTag. Sub62 proteins and half of the ubiquitin-enriched Sub592 protein were digested overnight with trypsin, while the remaining half was digested with ArgC. After enzymatic digestion, all three samples were desalted and enriched for diGly containing peptides using a polyclonal antibody as previously described (Cell Signaling, Technology, Danvers, MA) (Kim et al., 2011) and analyzed in an Orbitrap Velos mass spectrometer. Raw files were searched with Sequest against the target-decoy *S. cerevisiae* protein sequence database. Peptide spectral matches were filtered to a 1% false-discovery rate at the peptide and protein level and diGly sites were localized using a version of the Ascore algorithm that can accept any post-translational modification (Ascore > 13) (Beausoleil et al., 2006). See Extended Experimental Procedures for more details.

Protein Complementation Assays and Co-IP

Skp1 and Met30 were fused to fragments F1 and F2, respectively, of a split cytosine deaminase by gap-repair cloning. Point mutants were constructed using PCR with site directed oligonucleotides. Protein complementation was assayed as previously described (Ear and Michnick, 2009; Michnick et al., 2010). For Co-IP experiments, yeast cells expressing endogenous Myc-tagged Met30 were transformed with a plasmid expressing a Flag-tagged SKP1 S162A or S162D under control of the Gal promoter and selected for Leucine auxotrophy. Detailed information is available in Extended Experimental Procedures.

DNA, Yeast Strains

S. cerevisiae strains used were as follows: the ssa1 temperature sensitive strain (mat alpha leu2 trp1 ura3 ade2 his3 lys2, ssa1-45BKD, ssa2::LEU2, ssa3::TRP1, ssa4::LYS2) and was a gift from Betty Craig. The Δssa1::KanMX4 Δssa2::NAT was generated by direct replacement of the *SSA2* coding region with the NatMX4 cassette in the single deletion strain Δssa1::KanMX4 and confirmed by PCR.

Drop Test Assay

Cells were grown overnight in selective medium and then diluted to OD 600 nm O.4. Cells were grown for another 3 hr and then diluted to OD 600 nm of 0.1. This sample was then subjected to 10-fold serial dilutions. Ten microliters of each dilution was then spotted onto –URA plates and allowed to grow at 30°C, 33°C, and 37°C for 2 days.

Lysate Preparation and Ribosome Fractionation

Yeast (200 ml) in exponential growth was treated with 100 μg/ml of cycloheximide, harvested, washed with cold water, resuspended, and frozen as drops in liquid nitrogen. The cell lysate was loaded on a 12 ml 7%–47% sucrose gradient and centrifuged for 150 min at 39,000 rpm at 4°C. Fractions were collected using a UA/6 detector. The fractions were TCA precipitated and separated by SDS-PAGE and subjected to immunoblot analysis. The detailed protocol is available in Extended Experimental Procedures.

Luciferase In Vivo Refolding Assay

The ssa1-45 ts cells were transformed with firefly luciferase and a plasmid driving the expression of the wild-type *SSA1* or the phosphomutants. After growth at 30°C, the cells were shifted to 44°C for 1 hr, which causes the heat-induced denaturation of luciferase. Cycloheximide was added to 10 mg/ml 15 min before the end of the heat shock to prevent further expression of luciferase. Cells were then transferred to 30°C to recover. At different time points during the recovery, aliquots were taken, centrifuged, and frozen in liquid nitrogen. The luciferase activity was measured and recovery is expressed as a percentage of the activity before heat shock treatments.

Microscopy and Aggregation Assay

The ssa1-45 ts strain was transformed with the different *SSA1* mutants as well as with the Gal-Ubc9-2-GFP construct. Cells were grown overnight at 30°C and then diluted to OD 600 nm 0.3 and induced with 2% galactose for 6 hr. Cells were then shifted to 37°C for 30 min to induce the misfolding of Ubc9-2-GFP. The formation of Ubc9-2-GFP puncta was then examined by fluorescence microscopy.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.05.036.

ACKNOWLEDGMENTS

We thank Stephen Michnick and Peter Kaiser for strains and plasmids, Betty Craig for the ssa1-45 ts cells, and Stéphanie Escusa and Jonathan Warner for reagents. This work was supported by grants from the National Institutes of Health (AI090935, GM082250, GM084448, GM084279, GM081879 [N.J.K.]); GM55040, GM062583, GM081879, PN2 EY016546 [W.A.L.]; GM56433 [J.F.]; DP5 OD009180 [J.S.F.]; and RR01614 [A.B.]), the Howard Hughes Medical Institute (W.A.L.), and the Packard Foundation

(W.A.L.). P.B. is supported by the Human Frontier Science Program, J.S.F. is a QB3@UCSF Fellow, and N.J.K. is a Searle Scholar and a Keck Young Investigator.

Received: August 19, 2011
Revised: March 21, 2012
Accepted: May 18, 2012
Published: July 19, 2012

REFERENCES

Aghajan, M., Jonai, N., Flick, K., Fu, F., Luo, M., Cai, X., Ouni, I., Pierce, N., Tang, X., Lomenick, B., et al. (2010). Chemical genetics screen for enhancers of rapamycin identifies a specific inhibitor of an SCF family E3 ubiquitin ligase. Nat. Biotechnol. *28*, 738–742.

Albanèse, V., Yam, A.Y., Baughman, J., Parnot, C., and Frydman, J. (2006). Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. Cell *124*, 75–88.

Ali, M.M., Roe, S.M., Vaughan, C.K., Meyer, P., Panaretou, B., Piper, P.W., Prodromou, C., and Pearl, L.H. (2006). Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature *440*, 1013–1017.

Ba, A.N., and Moses, A.M. (2010). Evolution of characterized phosphorylation sites in budding yeast. Mol. Biol. Evol. *27*, 2027–2037.

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

Beltrao, P., and Serrano, L. (2007). Specificity and evolvability in eukaryotic protein interaction networks. PLoS Comput. Biol. *3*, e25.

Beltrao, P., Trinidad, J.C., Fiedler, D., Roguev, A., Lim, W.A., Shokat, K.M., Burlingame, A.L., and Krogan, N.J. (2009). Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. PLoS Biol. *7*, e1000134.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. PLoS Biol. *3*, e245.

Choudhary, C., and Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. Nat. Rev. Mol. Cell Biol. *11*, 427–439.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834–840.

DerMardirossian, C., Schnelzer, A., and Bokoch, G.M. (2004). Phosphorylation of RhoGDI by Pak1 mediates dissociation of Rac GTPase. Mol. Cell *15*, 117–127.

Ear, P.H., and Michnick, S.W. (2009). A general life-death selection strategy for dissecting protein functions. Nat. Methods *6*, 813–816.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Emanuele, M.J., Elia, A.E., Xu, Q., Thoma, C.R., Izhar, L., Leng, Y., Guo, A., Chen, Y.N., Rush, J., Hsu, P.W., et al. (2011). Global identification of modular cullin-RING ligase substrates. Cell *147*, 459–474.

Estève, P.O., Chang, Y., Samaranayake, M., Upadhyay, A.K., Horton, J.R., Feehery, G.R., Cheng, X., and Pradhan, S. (2011). A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat. Struct. Mol. Biol. *18*, 42–48.

Groll, M., Bajorek, M., Köhler, A., Moroder, L., Rubin, D.M., Huber, R., Glickman, M.H., and Finley, D. (2000). A gated channel into the proteasome core particle. Nat. Struct. Biol. *7*, 1062–1067.

Holt, L.J., Tuch, B.B., Villén, J., Johnson, A.D., Gygi, S.P., and Morgan, D.O. (2009). Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science *325*, 1682–1686.

Horton, L.E., James, P., Craig, E.A., and Hensold, J.O. (2001). The yeast hsp70 homologue Ssa is required for translation and interacts with Sis1 and Pab1 on translating ribosomes. J. Biol. Chem. *276*, 14426–14433.

Huber, A., Bodenmiller, B., Uotila, A., Stahl, M., Wanka, S., Gerrits, B., Aebersold, R., and Loewith, R. (2009). Characterization of the rapamycin-sensitive phosphoproteome reveals that Sch9 is a central coordinator of protein synthesis. Genes Dev. *23*, 1929–1943.

Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol. Cell *28*, 730–738.

Jonkers, W., and Rep, M. (2009). Lessons from fungal F-box proteins. Eukaryot. Cell *8*, 677–695.

Kaiser, P., Su, N.Y., Yen, J.L., Ouni, I., and Flick, K. (2006). The yeast ubiquitin ligase SCFMet30: connecting environmental and intracellular conditions to cell division. Cell Div. *1*, 16.

Kampinga, H.H., and Craig, E.A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat. Rev. Mol. Cell Biol. *11*, 579–592.

Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J., et al. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol. Cell *44*, 325–340.

Landry, C.R., Levy, E.D., and Michnick, S.W. (2009). Weak functional constraints on phosphoproteomes. Trends Genet. *25*, 193–197.

Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. PLoS Biol. *8*, e1000329.

Li, S., Iakoucheva, L.M., Mooney, S.D., and Radivojac, P. (2010). Loss of post-translational modification sites in disease. Pac. Symp. Biocomput., 337–347.

Lienhard, G.E. (2008). Non-functional phosphorylations? Trends Biochem. Sci. *33*, 351–352.

Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature *403*, 564–567.

Michnick, S.W., Ear, P.H., Landry, C., Malleshaiah, M.K., and Messier, V. (2010). A toolkit of protein-fragment complementation assays for studying and dissecting large-scale and dynamic protein-protein interactions in living cells. Methods Enzymol. *470*, 335–368.

Moses, A.M., and Landry, C.R. (2010). Moving from transcriptional to phospho-evolution: generalizing regulatory evolution? Trends Genet. *26*, 462–467.

Moses, A.M., Liku, M.E., Li, J.J., and Durbin, R. (2007). Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. Proc. Natl. Acad. Sci. USA *104*, 17713–17718.

Narayanan, A., and Jacobson, M.P. (2009). Computational studies of protein regulation by post-translational phosphorylation. Curr. Opin. Struct. Biol. *19*, 156–163.

Neduva, V., and Russell, R.B. (2005). Linear motifs: evolutionary interaction switches. FEBS Lett. *579*, 3342–3345.

Nolen, B., Taylor, S., and Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. Mol. Cell *15*, 661–675.

Paolinelli, R., Mendoza-Maldonado, R., Cereseto, A., and Giacca, M. (2009). Acetylation by GCN5 regulates CDC6 phosphorylation in the S phase of the cell cycle. Nat. Struct. Mol. Biol. *16*, 412–420.

Pereira-Leal, J.B., Levy, E.D., Kamp, C., and Teichmann, S.A. (2007). Evolution of protein complexes by duplication of homomeric interactions. Genome Biol. *8*, R51.

Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. Nat. Rev. Mol. Cell Biol. *6*, 9–20.

Soroka, J., Wandinger, S.K., Mäusbacher, N., Schreiber, T., Richter, K., Daub, H., and Buchner, J. (2012). Conformational switching of the molecular chaperone Hsp90 via regulated phosphorylation. Mol. Cell *45*, 517–528.

Soufi, B., Kelstrup, C.D., Stoehr, G., Fröhlich, F., Walther, T.C., and Olsen, J.V. (2009). Global analysis of the yeast osmotic stress response by quantitative proteomics. Mol. Biosyst. *5*, 1337–1346.

Stark, C., Su, T.C., Breitkreutz, A., Lourenco, P., Dahabieh, M., Breitkreutz, B.J., Tyers, M., and Sadowski, I. (2010). PhosphoGRID: a database of

experimentally verified in vivo protein phosphorylation sites from the budding yeast Saccharomyces cerevisiae. Database (Oxford) *2010*, bap026.

Stein, A., Céol, A., and Aloy, P. (2011). 3did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. *39* (Database issue), D718–D723.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat. Rev. Mol. Cell Biol. *11*, 515–528.

Tan, C.S., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M.O., Jørgensen, C., Bader, G.D., Aebersold, R., Pawson, T., and Linding, R. (2009). Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. Sci. Signal. *2*, ra39.

Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. Nature *443*, 415–420.

van Dam, T.J., and Snel, B. (2008). Protein complex evolution does not involve extensive network rewiring. PLoS Comput. Biol. *4*, e1000132.

Wagner, A. (2008). Neutralism and selectionism: a network-based reconciliation. Nat. Rev. Genet. *9*, 965–974.

Wagner, S.A., Beli, P., Weinert, B.T., Nielsen, M.L., Cox, J., Mann, M., and Choudhary, C. (2011). A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol. Cell. Proteomics *10*, M111.013284.

Wandinger, S.K., Suhre, M.H., Wegele, H., and Buchner, J. (2006). The phosphatase Ppt1 is a dedicated regulator of the molecular chaperone Hsp90. EMBO J. *25*, 367–376.

Weinert, B.T., Wagner, S.A., Horn, H., Henriksen, P., Liu, W.R., Olsen, J.V., Jensen, L.J., and Choudhary, C. (2011). Proteome-wide mapping of the Drosophila acetylome demonstrates a high degree of conservation of lysine acetylation. Sci. Signal. *4*, ra48.

Yang, X.J., and Grégoire, S. (2006). A recurrent phospho-sumoyl switch in transcriptional repression and beyond. Mol. Cell *23*, 779–786.

Cell

EXTENDED EXPERIMENTAL PROCEDURES

Posttranslational Modification Sites, Protein Disorder, Genomes, and Ortholog Assignments

From each publication we obtained PTM sites that had high localization probability and obtained, whenever possible, counts for the number of times the PTM, modified-peptide or mass-spectra was observed in the experiment. All of the sites compiled are provided in a searchable website. Protein sequences, protein identifiers and ortholog assignments were obtained from the Inparanoid database (http://inparanoid.sbc.su.se, version 7). We note that for genes having multiple splicing isoforms, Inparanoid retains only the longest form. For the comparative analysis we considered only 1-to-1 orthologs with Inparanoid confidence scores greater than 90%. One-to-many ortholog assignments were discarded since it has been shown that gene duplication influences the evolution of protein phosphorylation (Amoutzias et al., 2010). Protein disorder predictions were obtained using disEMBL (Linding et al., 2003). The total number of human to species ortholog pairs used in this studied are listed in Table S5.

Immunoprecipitation, Mass Spectrometry, and Ubiquitylation Site Identification

S. cerevisiae Sub592 (containing a HisTag modified ubiquitin) and Sub62 were grown separately in YPD and harvested during mid-log phase (OD600 ~1.0). Harvested cells were resuspended in lysis buffer containing 8 M urea, 300 mM NaCl, 50 mM Tris pH 8.2, 50 mM NaF, 50 mM sodium β-glycerophosphate, 10 mM sodium pyrophosphate, 1mM sodium orthovanadate, and 1 tablet mini protease inhibitor (Roche Diagnostics, Indianapolis, IN) and lysed by bead beating. Protein extract from Sub592 cells (~40mg) was enriched for ubiquitylated proteins via HisTag purification on a Co^{2+} sepharose resin (Talon Superflow Resin, Clontech Laboratories, Inc., Mountain View, CA). All samples were reduced, alkylated, and diluted 5-fold with 50 mM Tris, pH 8.2. Sub62 proteins and half of the ubiquitin enriched Sub592 protein were digested overnight with trypsin, while the remaining half was supplemented with 10 mM $CaCl_2$ and digested with ArgC. After enzymatic digestion, all three samples were desalted and enriched for diGly containing peptides using a polyclonal antibody as previously described (Cell Signaling, Technology, Danvers, MA) (Kim et al., 2011). Peptides were separated over a linear gradient from ~8%–30% acetonitrile in 0.125% formic acid and injected into an Orbitrap Velos mass spectrometer. We analyzed each sample once using CAD MS/MS spectra collected in the linear ion trap (120min analysis) and once with HCD MS/MS collected in the orbitrap (180min analysis). Raw files were searched with Sequest against the target-decoy (Elias and Gygi, 2007) *S. cerevisiae* protein sequence database (downloaded from SGD Jan 2011). Search parameters included a static modification on cysteine residues (57.02146 Da) and variable modifications of methionine oxidation (15.99491 Da) and diGly on lysine (114.04293 Da). Peptide spectral matches were filtered to a 1% false-discovery rate at the peptide and protein level and diGly sites were localized using a version of the Ascore algorithm that can accept any posttranslational modification (Ascore > 13) (Beausoleil et al., 2006).

Alignments and Conservation Metrics

Protein sequence alignments were done with muscle version 3.6 (Edgar, 2004) using standard alignments options. We defined two different conservation metrics: PTM acceptor residue conservation and site (or state) conservation. We consider that the acceptor residue (lysine for ubiquitylation and acetylation and serine, threonine or tyrosine for phosphorylation) is conserved when the aligned residue, opposite to the PTM site under analysis, is identical. We consider that the site (or state) is conserved when we find experimental evidence supporting the conservation of the PTM site under analysis. To account for errors in the PTM positional assignments we define a PTM site to be conserved if the aligned peptide is also modified within a window of ±2 alignment positions. In order to estimate a random expectation (i.e., null model) for the conservation of acceptor residues and PTM sites from species A to species B, we randomly shuffled the PTM sites of each protein in species A. When determining the fraction of phosphoacceptor residues phosphorylated as a function of the distance to a modified lysine we define the distance as the number of amino-acids to the closest modified lysine residue. For each fixed distance N from a modified lysine we counted the total number of phosphoacceptor residues (S/T/Y) and phosphorylation sites within a window of 10 amino-acids centered on N.

Phosphosite Kinase Preference and Phosphosite Similarity

We used the kinase recognition matrices from the study by Mok et al. (2010) to define a set of *S. cerevisiae* phosphosites that match a kinase substrate preference. We assumed that sites that had a matrix score above 2 standard deviations away from the mean score for any given kinase matched that kinase's substrate recognition preference. Using these kinase recognition matrices we also devised a phosphosite similarity metric. For each phosphosite, we calculated the predicted z-scored normalized recognition score for the 63 kinases studied by Mok and colleagues. These 63 scores constitute a vector that describes how likely it is that different types of kinases will phosphorylate a given site. We assume that the correlation score for a pair of vectors will be proportional to the similarity of the corresponding phosphosites. We benchmarked this approach by comparing the similarity of random pairs of phosphosites with pairs that are known to be phosphorylated by the same kinase based on the Phosphogrid database (http://www.phosphogrid.org) (Stark et al., 2010) (Figure 4B).

Protein Interface Models and Interface Residue Assignments

We used interface models from x-ray structures and homology models for *S. cerevisiae* and *H. sapiens* and complemented these with docking solutions for *S. cerevisiae* from the work of Mosca and colleagues (Mosca et al., 2009). Homology models for protein-pairs

that are known to interact were obtained from the GWIDD database (Kundrotas et al., 2010) with a cut-off of 25% sequence identity between the modeled proteins and template structures. Interface residues for these structure models were assigned using PSAIA (Mihel et al., 2008). Additional interface residues were predicted by homology using the 3DID database of inter-domain contacts for PFAM domains (Stein et al., 2011). 3DID maintains a list of inter-domain interface contact residues. We used HMMER (http://www.hmmer.org, version 3.0) to annotate the 11 proteomes with PFAM domains and used the 3DID annotations to predict PFAM domain residues that are more likely to participate in protein-protein interactions. For the 3DID annotations (Global Interface file from 13 of March 2011) we considered only PFAM domain residues that were annotated in 3DID to participate in interactions with, at least, 3 domains or linear motifs.

Protein Complementation Assays

Skp1 and Met30 were fused to fragments F1 and F2, respectively, of a split cytosine deaminase by gap-repair cloning. Point mutants were constructed using PCR with site directed oligonucleotides. Protein complementation was assayed as previously described (Ear and Michnick, 2009; Michnick et al., 2010). Briefly, a direct protein-protein interaction between proteins fused to the two fragments of the split cytosine deaminase allows it to fold and become enzymatically active. The folded cytosine deaminase converts cytosine to uracil, permitting growth on –Ura media, but also converts 5-FC to 5-FUTP, which kills cells on media containing 5-FC. In contrast, when the cytosine deaminase fragments are fused to proteins that do not interact, growth is no longer supported on –Ura media, but the strain is viable on media containing 5-FC. To assay the interaction, 300ng of Met30 and 300ng of Skp1 (WT, S162A, or S162D) plasmids were transformed into 25uL of log phase cells using the LiAc/PEG transformation procedure. Equal volumes were plated on all three media types and transformants were grown for 3 days (-His-Leu, +5-FC) or 5 days (-Ura).

SKP1-Met30 Co-IP

Yeast cells expressing endogenous Myc-tagged Met30 (Aghajan et al., 2010) were transformed with a plasmid expressing a Flag-tagged SKP1 S162A or S162D under control of the Gal promoter and selected for Leucine auxotrophy. Cells were grown in –Leu SC liquid media containing 2% raffinose and 2% galactose and harvested at an OD of 1.0. Cell pellets where resuspended in buffer containing 20mM HEPES pH 7.5, 10% glycerol, 200mM NaCl, 1mM EDTA, 1mM DTT, 1mM NaF, and protease inhibitors (Sigma). Pellets were broken by bead beating with glass beads 5 X 1 min with 2 min on ice between beatings. Whole cell lysate was centrifuged at 14,000 rpm for 20 min, and supernatant was collected and immunopurified on Anti-Flag Magnetic beads (Sigma), which were washed and incubated according to manufacturer's instructions. Co-immunopurified proteins where washed and resuspended in 3M SDS loading Buffer, run on an SDS-PAGE gel and transferred to a PDVF membrane. The membrane was probed with either primary Anti-c-*myc* produced in chickens (Invitrogen) with Anti-Chicken HRP conjugate (Promega) secondary or with Anti-Flag produced in mouse (Sigma) primary with Goat anti Mouse HRP Light Chain Specific (Jackson ImmunoResearch) secondary. Immunoblots were processed using ECL detection.

PTM Hot Spots for Domain Families

Ten domain families that are heavily phosphorylated in most species studied were selected for analysis (Table S4). For each domain family, we selected a representative structure and transferred the phosphosites, occurring in all instances of this domain across the 11 species, using protein sequence alignments. We then used a sliding window of 10 amino acids and random sampling to identify regions, within each domain family, that are enriched for phosphorylation. Any 10 amino-acid peptide with a significant (p value < 0.005) enrichment of phosphorylation when compared to random was defined a potential regulatory hot spot. The same analysis was performed using lysine acetylation data for the protein kinase domain.

DNA, Yeast Strains

The *SSA1* ORF was amplified by PCR and then cloned into a CEN gateway plasmid containing a GPD promoter. All the mutants were then generated by Quick Change mutagenesis according to the manufacturer's instructions All the plasmids were confirmed by sequencing. *S. cerevisiae* strains were as follows: the ssa1 temperature sensitive strain (mat alpha leu2 trp1 ura3 ade2 his3 lys2, ssa1-45BKD, ssa2::LEU2, ssa3::TRP1, ssa4::LYS2) and was a gift from Betty Craig. The Δssa1::KanMX4 Δssa2::NAT was generated by direct replacement of the *SSA2* coding region with the NatMX4 cassette in the single deletion strain Δssa1::KanMX4. The correct insertion of the marker was confirmed by PCR from genomic DNA. Standard yeast protocols were used for yeast transformations.

Drop Test Assay

Cells were grown overnight in selective medium and then diluted to OD 600 nm O.4. Cells were grown for another 3 hr and then diluted to OD 600 nm of 0.1. This sample was then subjected to 10-fold serial dilutions. 10μl of each dilution was then spotted onto –URA plates and allowed to grow at 30°C, 33°C and 37°C for 2 days.

Lysate Preparation and Ribosome Fractionation

200 ml yeast in exponential growth was treated with 100μg/ml of cycloheximide, harvested, washed with cold water, resuspended in 1ml of Buffer A (20 mM HEPES pH = 7.5, 50 mM KCl, 10 mM MgCl2, 1% Triton X-100, 1mM DTT, 100μg/ml cycloheximide and protease inhibitor cocktail) and frozen as drops in liquid nitrogen. The cells were then ground using a grinder (MM301, Retsch).

0.5 ml of lysis buffer A was added to the powder and the lysates were clarified by centrifugation at 14,000 g for 10 min. 20 OD of lysate was loaded on a 12 ml 7%–47% sucrose gradient in buffer A without Triton X-100 and centrifuged in a SW41 rotor (Beckamn Coulter) for 150 min at 39,000 rpm at 4°C. Fractions were collected using a UA/6 detector (ISCO, Inc). The fractions were TCA precipitated and separated by SDS-PAGE and subjected to immunoblot analysis. The Rpl3 antibody was a gift from Jonathan Warner. The monoclonal HA antibody was from Covance.

Luciferase In Vivo Refolding Assay

The ssa1-45 ts cells were transformed with firefly luciferase, which is labile above 37°C, and a plasmid driving the expression of the wild-type *SSA1* or the phosphomutants. After growth at 30°C, the cells were shifted to 44°C for 1 hr, which causes the heat-induced denaturation of luciferase. Cycloheximide was added to 10mg/ml 15 min before the end of the heat shock to prevent further expression of luciferase. Cells were then transferred to 30°C to recover. At different time point during the recovery, aliquots were taken, centrifuged and frozen in liquid nitrogen. The luciferase activity was measured using the Promega luciferase kit. Recovery is expressed as a percentage of the activity before heat shock treatments.

Microscopy and Aggregation Assay

The ssa1-45 ts strain was transformed with the different *SSA1* mutants as well as with the Gal-Ubc9-2-GFP construct. Cells were grown overnight at 30°C and then diluted to OD 600 nm 0.3 and induced with 2% galactose for 6 hr. Cells were then shifted to 37°C for 30 min to induce the misfolding of Ubc9-2-GFP. The formation of Ubc9-2-GFP puncta was then examined by fluorescence microscopy using a microscope (Axiovert; Carl Zeiss, Inc) equipped with a 100x NA 1.3 oil immersion objective lens and a digital camera controlled with AxioVision software. Images were then prepared using Photoshop.

SUPPLEMENTAL REFERENCES

Aghajan, M., Jonai, N., Flick, K., Fu, F., Luo, M., Cai, X., Ouni, I., Pierce, N., Tang, X., Lomenick, B., et al. (2010). Chemical genetics screen for enhancers of rapamycin identifies a specific inhibitor of an SCF family E3 ubiquitin ligase. Nat. Biotechnol. *28*, 738–742.

Amoutzias, G.D., He, Y., Gordon, J., Mossialos, D., Oliver, S.G., and Van de Peer, Y. (2010). Posttranslational regulation impacts the fate of duplicated genes. Proc. Natl. Acad. Sci. USA *107*, 2967–2971.

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

Ear, P.H., and Michnick, S.W. (2009). A general life-death selection strategy for dissecting protein functions. Nat. Methods *6*, 813–816.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods *4*, 207–214.

Kundrotas, P.J., Zhu, Z., and Vakser, I.A. (2010). GWIDD: Genome-wide protein docking database. Nucleic Acids Res. *38* (*Database issue*), D513–D517.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. Structure *11*, 1453–1459.

Michnick, S.W., Ear, P.H., Landry, C., Malleshaiah, M.K., and Messier, V. (2010). A toolkit of protein-fragment complementation assays for studying and dissecting large-scale and dynamic protein-protein interactions in living cells. Methods Enzymol. *470*, 335–368.

Mihel, J., Sikić, M., Tomić, S., Jeren, B., and Vlahovicek, K. (2008). PSAIA - protein structure and interaction analyzer. BMC Struct. Biol. *8*, 21.

Mok, J., Kim, P.M., Lam, H.Y., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M., et al. (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci. Signal. *3*, ra12.

Mosca, R., Pons, C., Fernández-Recio, J., and Aloy, P. (2009). Pushing structural information into the yeast interactome by high-throughput protein docking experiments. PLoS Comput. Biol. *5*, e1000490.

Stark, C., Su, T.C., Breitkreutz, A., Lourenco, P., Dahabieh, M., Breitkreutz, B.J., Tyers, M., and Sadowski, I. (2010). PhosphoGRID: a database of experimentally verified in vivo protein phosphorylation sites from the budding yeast Saccharomyces cerevisiae. Database (Oxford) *2010*, bap026.

Stein, A., Céol, A., and Aloy, P. (2011). 3did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. *39* (*Database issue*), D718–D723.

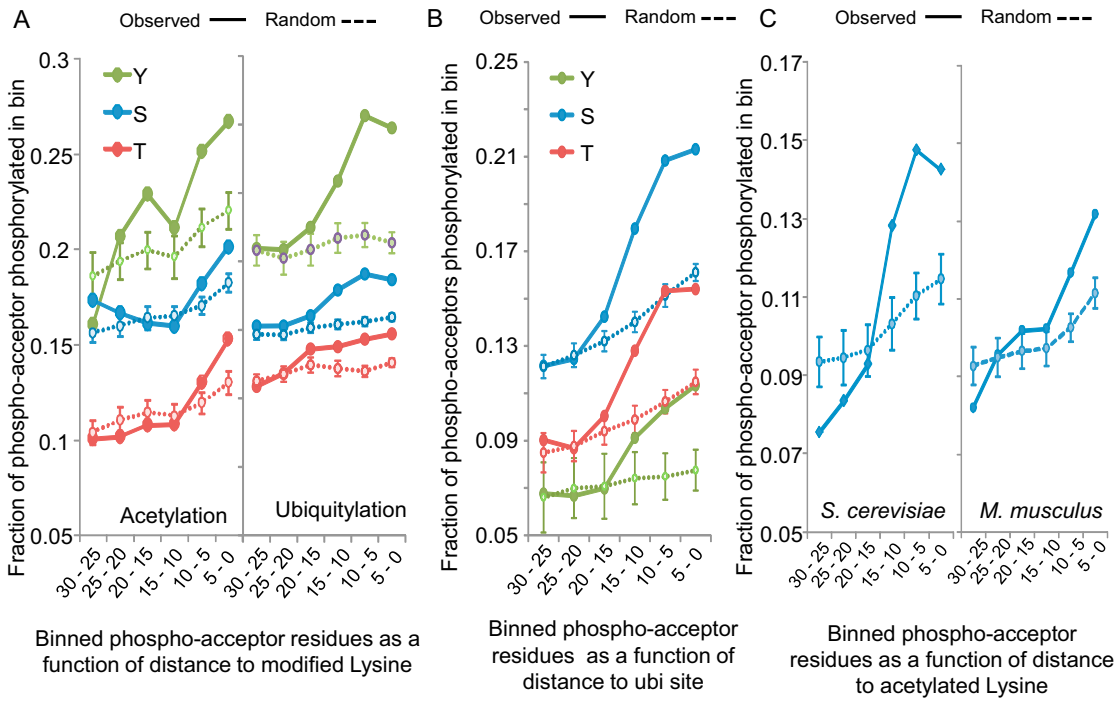

Figure S1. Clustering of Lysine Modifications and Phosphorylation Sites, Related to Figure 2

Phosphoacceptor residues—serine (S), threonine (T) and tyrosine (Y) —were found to be preferentially phosphorylated closer to lysines that were modified, when compared to random (Figure 2, main text).

(A) In order to rule out that this was due to a preferential accessibility within unstructured regions of proteins we repeated the analysis excluding all PFAM domain sequences. Solid lines represent observed fraction of phosphorylated residues, and the dotted lines represent the average from random samples of a similar number of phosphosites.

(B) In order to show that the observed clustering of phosphorylation sites with ubiquitylation sites seen in human proteins is conserved in other species we performed a similar analysis in *S. cerevisiae*.

(C) In order to show that the observed clustering of phosphorylation sites with acetylation sites seen in human proteins is conserved in other species we performed a similar analysis in *S. cerevisiae* and *M. musculus*. Acetylation data for these species was obtained from the Global Proteome Machine Database (http://www.thegpm.org). As observed in human, the fraction phosphoacceptor residues that are phosphorylated increases near acetylation sites and is higher than expected by chance when within 15 residues.

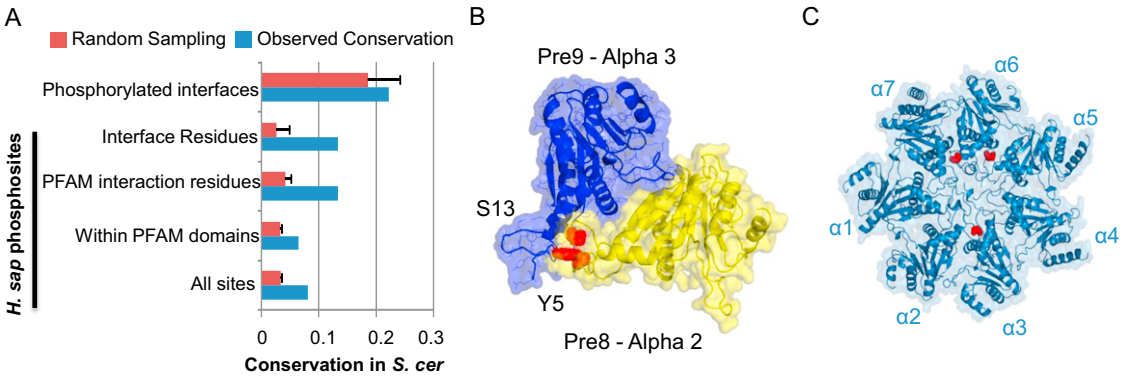

Figure S2. Phosphosites at Human Interface Residues and at Proteasome Subunit Interfaces, Related to Figure 3

(A) We used models of interfaces for human protein-interactions either from x-ray structures or homology models to define interface residues. Additionally, we used the 3DID database to annotate all human PFAM domain residues that might participate in interactions (PFAM interaction residues). We observed that human phosphorylation sites found at interface residues were found to be more conserved that average sites or sites found within PFAM domains. Additionally, the conservation of the phosphorylation of an interface was twice that of the positional conservation of average phosphosites. The observed conservation of all categories was found to be significantly higher than expected based on a random sampling of an equivalent number of phosphorylation sites with exception for the conservation at the level of the interface. The lack of significance could be, in part, due to the low total number of human interfaces that are currently available for analysis.

(B and C) Conserved N-terminal phosphorylation sites in the *S. cerevisiae* alpha-proteasome subunits might regulate proteasome assembly or activity.

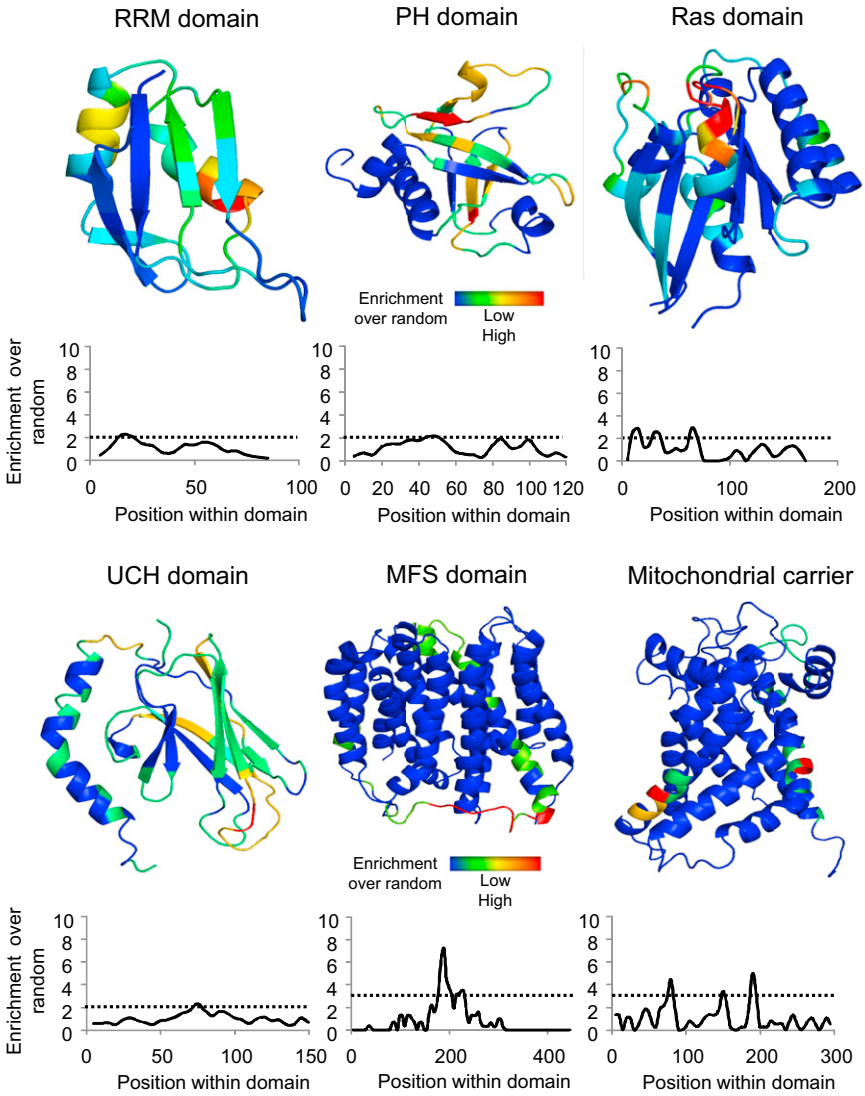

Figure S3. Domain Family Phosphopeptide Enrichment Analysis for RRM, PH, Ras, UCH, MSF, and Mitochondrial Carrier Domains, Related to Figure 5

All domain instances were aligned to a representative structure sequence and the phosphopeptide positions were transferred by homology. Using a moving window we calculated for each position within the domain family the enrichment over a random distribution of all phosphorylation sites. The significance was accessed by randomly sampling the same number of positions within the domain. The dotted line corresponds, in each case, to a p value cut-off of 0.005. Positions that had an enrichment value above the cut-off were considered to be significantly enriched for phosphorylation and we denominate these regulatory "hot spots." Phosphosites from the 11 species found within these regions are provided at ptmfunc.com.

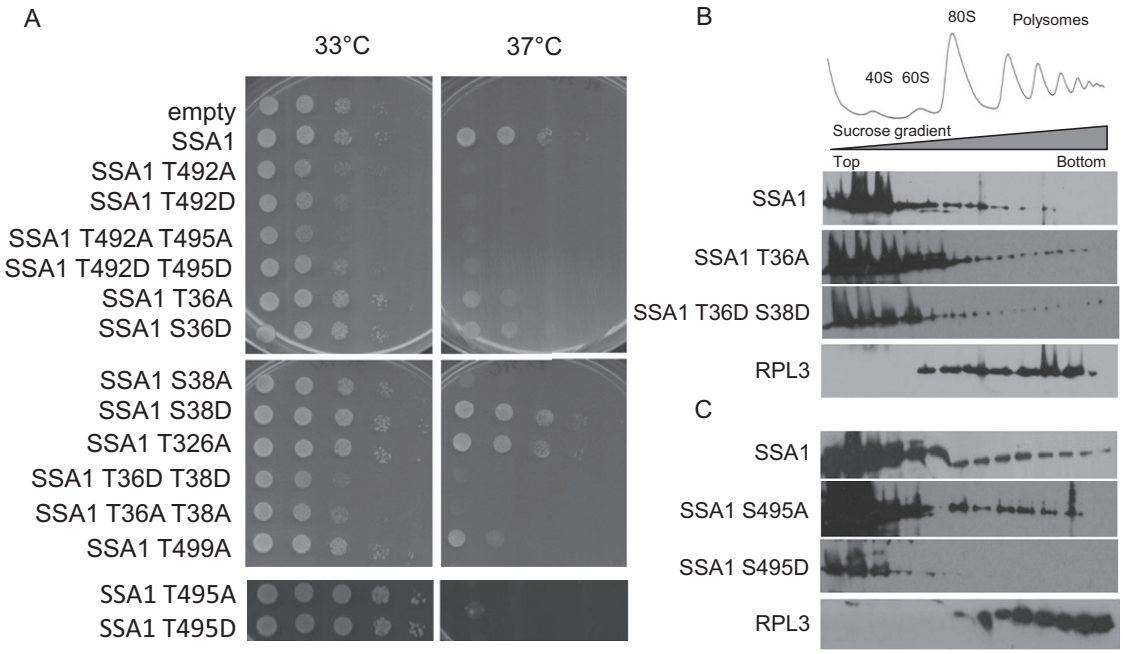

Figure S4. Phosphorylation Hot Spots within the HSP70 Domain Family, Related to Figure 6

(A) All alanine and phosphomimetic mutants of the Regions 1 and 2 described in the main text were analyzed for growth defects under heat-shock and in a strain deficient in SSA function (ssa1-45Δssa2Δssa3Δssa4). All mutants, with the potential exception of S38D, displayed a growth defect under heat shock conditions that is not observed when the WT *SSA1* is expressed or when a control mutation T326A is introduced.

(B and C) The association of the indicated *SSA1* mutants with polysomes was examined by immunoblot analysis following the fractionation of cell lysates by sedimentation on sucrose gradients. Ribosomal profiles (top) were determined by OD 254 nm and confirmed by immunoblot analysis of the ribosomal proteins Rpl3p. (B) Unlike region 2 mutants, region 1 mutants, including a double phosphomimetic mutant, displayed a wild-type like association with polysomes. (C) A phosphomimetic mutation in position S495 is sufficient to disrupt the association of *SSA1* with polysomes. The same phenotype is not observed with an alanine mutant.

Table S1. Phosphorylation, Acetylation, and Ubiquitylation Sites Used for the Analysis, Related to Results

The data can be accessed via a website (http://ptmfunc.com). For each species we used PFAM annotations to count the number of PTMs that are found within globular protein domains. On average, ~75% of the phosphosites ~40% of acetylation and ~45% of ubiquitylation sites are found outside PFAM domains.

Species	PTM type	PTM total	Within PFAM domains	% Outside PFAM domain
H. sapiens	Phosphorylation	31165	11726	62.4
	Acetylation	8042	4604	42.8
	Ubiquitylation	22057	11079	49.7
M. musculus	Phosphorylation	24921	6825	72.6
	Acetylation	3384	2298	32.1
R. norvegicus	Phosphorylation	1885	913	51.6
X. laevis	Phosphorylation	470	149	68.3
C. elegans	Phosphorylation	6715	1074	84.0
D. melanogaster	Phosphorylation	17535	2081	88.1
	Acetylation	1707	858	49.7
S. pombe	Phosphorylation	2540	636	75.0
S. cerevisiae	Phosphorylation	15144	3747	75.3
	Acetylation	657	433	34.1
	Ubiquitylation	2499	1426	42.9
C. albicans	Phosphorylation	2910	532	81.7
A. thaliana	Phosphorylation	4527	648	85.7
O. sativa	Phosphorylation	3140	633	79.8

Table S2. Estimation of False Discovery Rate for the *S. cerevisiae* Phosphosite Data Set after Compilation of 12 Different Reported Experiments, Related to Results

Although individual proteomic experiments report false discovery rates (FDR) of peptide identification on the order of 2% or less, the accumulation of different independent datasets results in an increase of FDR. We collected 12 different phosphoproteomics studies for *S. cerevisiae* and used these to estimate an upper bound for this increase. Assuming a rate of 2% peptide FDR for each individual dataset and that no false-positive site is identified more than once, we estimate that the combined *S. cerevisiae* phosphoproteome has, at most, an FDR of ~4%. This suggests that even for species that have been extensively studied such as *S. cerevisiae* and *H. sapiens*, the fraction of incorrectly identified PTMs is likely to be low. For each study we estimated the total number of false positive phosphosites assuming a false-discovery rate (FDR) of 2%. We assumed, conservatively, that no false-positive phosphosite is observed more than once. Under these assumptions the total expected number of false positives would be equivalent to 4% of the compiled non-redundant *S. cerevisiae* phosphosites (836 out of 20658).

Pubmed ID	Phosphosites obtained from study	Projected false-positives (assuming 2% FDR)
19823750	2876	57.52
17563356	6489	129.78
19684113	4000	80
17287358	1154	23.08
19795423	3010	60.2
20377248	2526	50.52
15665377	591	11.82
19547744	3435	68.7
17330950	1386	27.72
21177495	3540	70.8
21298081	6071	121.42
19779198	6744	134.88
Sum	**41822**	**836.44**
Sum non-redundant	**20658**	**836.44**

Table S3 - Overlap of Different Posttranslational Modifications within the Same Proteins Does Not Depend on Protein Abundance, Related to Figure 2

We observed that proteins with different lysine modifications (acetylation, ubiquitylation and sumoylation) are also very likely to be phosphorylated. One possible explanation for this would be an experimental identification bias in MS experiment. Since MS experiments preferentially identify highly abundant proteins then the overlap could be due this bias. In order to control for this we used protein abundance estimates for human proteins (http://pax-db.org/). We excluded all human proteins with estimated abundance over half of the median such that there was no significant difference in the abundance levels of phosphorylated versus non-phosphorylated proteins (p-value=0.44 with a KS ranked test). After controlling for abundance we still see a very significant enrichment of phosphoproteins among the lysine modified proteins over random (using a Fisher's exact test).

Protein Subset	Total	Phosphoproteins (% from total)	Enrichment over random	p-value for enrichment
All proteins	3522	1526 (43%)	-	-
Ubiquitylated	816	539 (66%)	1.5	$<10^{-40}$
Acetylated	306	237 (77%)	1.8	$<10^{-33}$
Sumoylated	46	43 (93%)	2.1	$<10^{-12}$

Table S4. Domain Families Selected for Phosphopeptide Enrichment Analysis, Related to Results

We used PFAM annotations and the compilation of known phosphorylation sites for the 11 species considered in this study to count the total number of phosphorylation sites found for each PFAM domain. Among the domain families with higher number of total phosphosites across all the species we selected 10 that had an available representative structure deposited in the PDB for analysis. For each of these PFAM domains we provide here the total number of domain instances annotated across the 11 species as well as the total number of phosphosites found for each and the representative structure used for the enrichment analysis.

PFAM id	Domain name	Total phosphosites	Total number of domains	PDB ID of structure used
Pkinase	Protein kinase	1273	4269	1QMZ
Pkinase_Tyr	Protein tyrosine kinase	495	1427	1M14
HSP70	Heat shock proteins, Hsp70	313	195	1YUW
RRM_1	RNA recognition motif	253	2438	3BS9
UCH	Ubiquitin carboxyl-terminal hydrolase	200	390	3H0X
Ras	Ras domain	190	821	1EK0
HSP90	Heat shock proteins, Hsp90	145	78	2IOP
PH	Pleckstrin homology domain	144	856	1MAI
MFS_1	Major Facilitator Superfamily (MFS) transporters	120	723	2GFP
Mito_carr	Mitochondrial carrier	119	1434	2C3E

Table S5. Total 1-to-1 Orthoproteins and Phosphosites Used in Species to Human Comparative Analysis, Related to Experimental Procedures

We counted the total number of human to species 1-to-1 orthologs with an inparanoid score greater than 90%. We also detailed how many of these are phosphoproteins and how many phosphosites in total were used for the comparative studies throughout the paper.

Species	Total number of orthologs	Orthologs that are phosphoproteins	Number of phosphosites in orthologs
M.musculus	15982	5923	31317
R. norvegicus	14940	1059	2547
X.laevis	8261	279	394
C.elegans	4489	1445	4309
D.melanogaster	5346	2279	9264
S.pombe	2225	597	1382
S.cerevisiae	1992	1234	6413
C.albicans	2129	526	1481
A.thaliana	2888	476	1036
O.sativa	2872	464	989