Phosphorylation at the Interface

Fred P. Davis[1,*]
[1]Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Dr., Ashburn, VA 20147, USA
*Correspondence: davisf@janelia.hhmi.org
DOI 10.1016/j.str.2011.11.006

Proteomic studies have identified thousands of eukaryotic phosphorylation sites (phosphosites), but few are functionally characterized. Nishi et al., in this issue of *Structure*, characterize phosphosites at protein-protein interfaces and estimate the effect of their phosphorylation on interaction affinity, by combining proteomics data with protein structures.

Since its discovery nearly 60 years ago, protein phosphorylation has been revealed as a widespread signal transduction mechanism that contributes to disease when misregulated and can serve as a valuable therapeutic target (Pawson and Scott, 2005). The expanding toolbox of modern biology has been used to systematically identify kinases and phosphatases in whole genomes and characterize the regulatory networks they form with their substrates. For example, mass spectrometry proteomics suggests that nearly a third of human proteins are phosphorylated. However, the functional consequences—if any (Landry et al., 2009)—of most phosphorylation events are unknown. Structure is one way to address this functional gap. In this issue, Nishi et al. (2011) study the effect of phosphorylation on protein-protein interactions using structural bioinformatics to combine the large-scale coverage of mass spectrometry proteomics with the structural insight of atomic resolution protein structures (Figure 1).

What did they do? They began with a list of published human phosphosites either observed by traditional low-throughput experiments or high-throughput proteomics and filtered by the GPS prediction algorithm (Xue et al., 2008). Next, they map these sites onto quaternary protein structures from the Protein DataBank verified by the PISA algorithm (Krissinel and Henrick, 2007). This three-dimensional (3D) phosphosite dataset serves as the basis for a comprehensive series of bioinformatics experiments that characterize the composition of phosphosites, their location relative to protein interfaces, their evolutionary history, and their functional properties. Lastly, they estimate the contribution of interface phosphosites to complex stability using the FoldX program for computational alanine scanning and to model phosphorylation of the site (Sánchez et al., 2008). Previous work has mapped phosphosites onto protein structures (e.g., Zanzoni et al., 2011), but this study maps this information onto protein-protein interfaces and then uses this to investigate the energetic effects of phosphorylation on protein interactions.

What did they find? Several of the results are particularly interesting because they are uniquely accessible to the authors' structure-based approach. First, phosphosites are ~1.5× enriched at interfaces relative to the rest of the protein surface for hetero-oligomers, but not most homo-oligomers. Second, interface phosphosites are twice as likely to be a predicted binding "hot spot" (contributing >2 kcal/mol to complex stability) than the rest of the interface in hetero-oligomers, but not in homo-oligomers. Next, in most cases, phosphorylation is predicted to minimally affect complex stability (<1 kcal/mol). In approximately a third of cases, phosphorylation is predicted to destabilize the complex, and, in only a handful of cases, to stabilize the complex by more than 2 kcal/mol. Finally, their evolutionary and functional findings are also informed by 3D structure. They find that interface phosphosites are, on average, more conserved than the rest of the interface (for hetero- but not homo-oligomers) and occur more often than expected in proteins with particular functions, including involvement in signaling pathways.

What are the limitations? One of the main challenges of large-scale bioinformatics is estimating the accuracy of individual predictions, especially when multiple steps can each introduce different kinds of errors. Many of Nishi's results are largely descriptive, and except for global biases in the initial list of phosphosites—such as those introduced by the proteomic experiments or the GPS algorithm—should be largely free of error. Throughout the rest of their analyses, the most significant source of error is likely the prediction of protein stability from structure, which remains a difficult problem. The methods that they employ, PISA and FoldX, have both been rigorously evaluated and represent state-of-the-art methods in terms of their accuracy but nonetheless have error bars. For example, FoldX has been benchmarked for its ability to predict experimentally measured changes in free energy ($\Delta\Delta G$) upon mutation of protein complexes (correlation coefficient $R = 0.8$) and specifically phosphopeptide-mediated interactions ($R = 0.72$) (Guerois et al., 2002; Sánchez et al., 2008). Finally, by design, this study almost exclusively considers phosphosites in structured regions of proteins and largely ignores those in unstructured regions. Nevertheless, this study provides an intriguing proteome-wide 3D view of the effects of phosphorylation on protein-protein interactions.

In addition to proteome-wide observations, the authors' predictions may also be useful for individual protein complexes. For example, they predict that phosphorylation of two members of the SMAD transcription factor family increases their homotrimer complex stability, in agreement with published experiments. As more experimental data characterizing individual phosphosites becomes available, further comparison will be useful for evaluating the utility of the authors' individual predictions for experimentalists.

What's next? Besides the analyses they perform, the dataset that the authors have

assembled is an enabling resource for other researchers to quickly ask new questions. For instance, a recent study suggests that nonsynonymous single nucleotide polymorphisms may affect nearly 10% of experimentally observed phosphosites by mutating either the phosphorylated residue or neighboring sites that are recognized by the regulatory kinase or phosphatase (Ren et al., 2010). How often do these polymorphisms reside at protein interfaces and do they affect complex stability? From yeast to humans, proteomic tyrosine content has decreased as the number of tyrosine kinases encoded in the genome has increased (Tan et al., 2009). Is this trend reflected in the tyrosine composition and phosphosite enrichment at protein interfaces of different species? Lastly, Nishi et al.'s approach is generally applicable to other posttranslational modifications. For example, extending FoldX or using forcefields (e.g., CHARMM or AMBER) that have been previously parameterized for other covalent modifications could expand their approach to other widespread modifications such as lysine acetylation (Choudhary et al., 2009).

The study by Nishi et al. (2011) elegantly combines existing structural and biochemical data to characterize the effect of phosphorylation on protein interactions. It is exciting both for the breadth of analyses and for the future work that it enables. The coverage and accuracy of their approach will continue to improve as more phosphosites are identified, more protein structures are determined, and forcefields are further refined. Together with other experimental and computational efforts, structural bioinformatics will continue to contribute toward characterizing the structure, function, and evolution of posttranslational protein modifications.

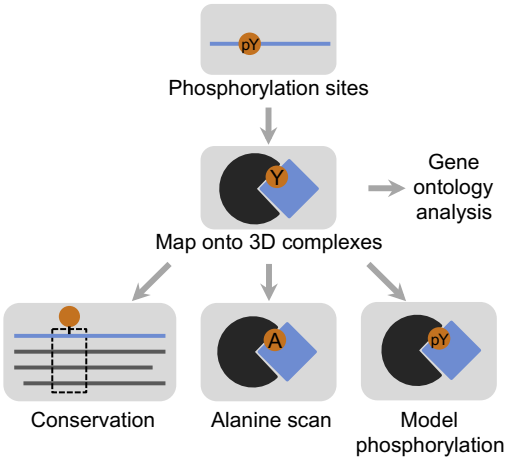

Figure 1. Characterizing Interface Phosphosites
Nishi et al. (2011) map published protein phosphorylation sites onto the 3D structures of protein complexes and characterize their structure, function, and evolution.

REFERENCES

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Science *325*, 834–840.

Guerois, R., Nielsen, J.E., and Serrano, L. (2002). J. Mol. Biol. *320*, 369–387.

Krissinel, E., and Henrick, K. (2007). J. Mol. Biol. *372*, 774–797.

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

Nishi, H., Hashimoto, K., and Pachenko, A.R. (2011). Structure *19*, this issue, 1807–1815.

Pawson, T., and Scott, J.D. (2005). Trends Biochem Sci. *30*, 286–290.

Ren, J., Jiang, C., Gao, X., Liu, Z., Yuan, Z., Jin, C., Wen, L., Zhang, Z., Xue, Y., and Yao, X. (2010). Mol. Cell. Proteomics *9*, 623–634.

Sánchez, I.E., Beltrao, P., Stricher, F., Schymkowitz, J., Ferkinghoff-Borg, J., Rousseau, F., and Serrano, L. (2008). PLoS Comput. Biol. *4*, e1000052.

Tan, C.S.H., Pasculescu, A., Lim, W.A., Pawson, T., Bader, G.D., and Linding, R. (2009). Science *325*, 1686–1688.

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., and Yao, X. (2008). Mol. Cell. Proteomics *7*, 1598–1608.

Zanzoni, A., Carbajo, D., Diella, F., Gherardini, P.F., Tramontano, A., Helmer-Citterich, M., and Via, A. (2011). Nucleic Acids Res. *39* (Database issue), D268–D271.