

Phosphorylation sites of higher stoichiometry are more conserved

To the Editor: Phosphoproteomic screens by mass spectrometry have uncovered thousands of new phosphorylated residues (here referred to as sites) in various organisms. Follow-up studies are needed to assay functional effects of these sites. Knowing site stoichiometry and conservation, in addition to other information, may be useful to prioritize identified phosphorylation sites for functional characterization and to guide experiment design. Wu *et al.*¹ had determined stoichiometries of ~5,000 phosphorylation sites in *Saccharomyces cerevisiae* at mid-log phase of growth. Unexpectedly, they observed that the most conserved sites in their data are very-low-stoichiometry sites. They had found that high-stoichiometry sites are less conserved, on average, than low-stoichiometry ones and suggest that stoichiometry does not positively correlate with a site being biologically essential.

These observations by Wu *et al.*¹ could arise if the very-low-stoichiometry phosphorylation sites ($\leq 5\%$) were over-represented on high-abundance proteins, which are generally very conserved in sequence^{2,3}. We sorted phosphoproteins containing high-confidence phosphorylation sites reported in ref. 1 into three equal groups according to their cellular abundance⁴; ~55% of the very-low-stoichiometry (0–5% phosphorylation) sites were in the highest protein abundance group, compared to ~35% of the low-stoichiometry (5–20%), ~16% of moderate-stoichiometry (20–80%) and ~25% of high-stoichiometry (80–100%) sites. This over-representation ($P < 3.9 \times 10^{-42}$, χ^2 test, very-low-stoichiometry sites compared to remaining sites) likely partly arises because a very-low-stoichiometry phosphorylation site on a high-abundance protein has more phosphopeptides in a sample, which increases its mass spectrometry detection probability and need not imply that very-low-stoichiometry phosphorylation sites occur less frequently on low-abundance proteins.

To interpret whether phosphorylation sites of higher stoichiometry are more conserved, it helps to correct for the background conservation rate of residues, structural protein regions and proteins. Hence, we compared the conservation rate of phosphorylation sites to randomly selected phosphorylatable residues while preserving the number of each residue type from structured and unstructured protein regions as the phosphorylation sites from each phosphorylated protein. This allowed us to compute a measure we term 'relative divergence rate' for each group of

phosphorylated sites (**Supplementary Methods**). Relative divergence rates of 0.5 and 2.0, for example, respectively imply that 50 and 200 phosphorylated sites have mutated for every 100 phosphorylatable residues that have mutated. Relative divergence rates above and below 1 indicate phosphorylation sites are, respectively, less and more conserved than background residues that can be phosphorylated. This analysis revealed that high-stoichiometry phosphorylation sites were more conserved than sites of lower

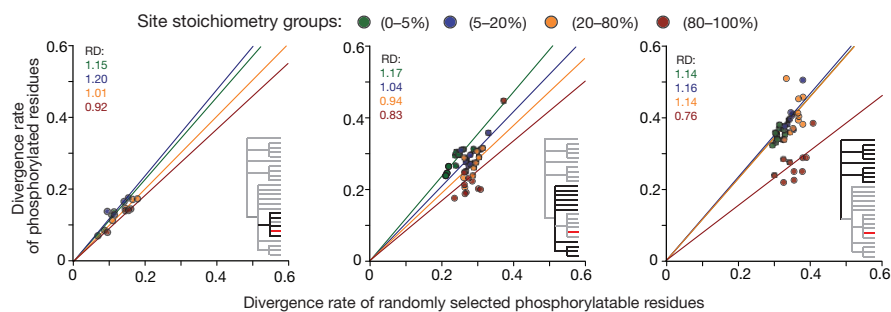


Figure 1 | High-stoichiometry sites were most conserved, and site stoichiometry correlated negatively with divergence rate. Every point on each plot represents the divergence rate of a set of phosphorylation sites (residues) versus a set of randomly selected phosphorylatable residues for each species compared in Wu *et al.*¹ except *S. pastorianus*. Line of best fit from the origin was computed for each site stoichiometry group to interpret the general divergence (or conservation) trend, with the gradient (slope) interpreted as the general relative divergence rate (RD). We divided 24 species analyzed by Wu *et al.*¹, into three groups based on their phylogenetic distance from *S. cerevisiae*. Cladograms (bottom right insets) highlight phylogenetic relationships of each species group (black) to *S. cerevisiae* (red).

stoichiometry (**Fig. 1**) in contrast to the results in Wu *et al.*¹. Also, phosphorylation-site stoichiometry generally correlated positively with site conservation, albeit in phylogenetically more related species (**Fig. 1**). These results imply high-stoichiometry sites are generally more essential and suggest that sites of lower stoichiometry at mid-log phase of growth tend to appear later in evolution⁵.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the Canadian Institutes of Health Research grant MOP-84324.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Chris Soon Heng Tan^{1–4} & Gary D Bader^{1–3}

¹Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

²Department of Molecular Genetics, University of Toronto, Toronto, Canada.

³Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada. ⁴Present address: Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria.

e-mail: ctan@cemmm.oew.ac.at

1. Wu, R. *et al. Nat. Methods* **8**, 677–683 (2011).
2. Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O. & Arnold, F.H. *Proc. Natl. Acad. Sci. USA* **102**, 14338–14343 (2005).
3. Pál, C., Papp, B. & Hurst, L.D. *Genetics* **158**, 927–931 (2001).
4. Ghaemmghami, S. *et al. Nature* **425**, 737–741 (2003).
5. Budovskaya, Y.V., Stephan, J.S., Deminoff, S.J. & Herman, P.K. *Proc. Natl. Acad. Sci. USA* **102**, 13933–13938 (2005).