Cell biology

Motifs mapped for almost all human kinase enzymes

Sean J. Humphrey & Elise J. Needham

A computational resource can identify candidate protein targets for almost all members of a major class of kinase enzyme in humans, with implications for understanding cell signalling in health and disease. **See p.759**

Protein phosphorylation — the selective addition of phosphate groups to proteins — is a regulatory mechanism that is fundamental to life. Conversely, dysregulated phosphorylation has been implicated in conditions such as Alzheimer's disease, cancer and diabetes[1]. The enzymes that catalyse phosphorylation, known as kinases, are major targets for drugs[2], so understanding their regulatory roles could provide new therapeutic opportunities. Advances in our ability to identify and quantify phosphorylation using mass spectrometry has led to a rapid rise in the number of known phosphorylation sites in human proteins (collectively known as the phosphoproteome), from hundreds at the turn of the century to more than 100,000 today[3]. However, linking these sites with their associated kinases has been a laborious process[3]. On page 759, Johnson *et al.*[4] take a major step towards resolving this problem, describing a comprehensive resource that defines the potential substrates for almost all members of one major class of human kinase.

The phosphoproteome is highly complex, comprising tightly interconnected networks of hundreds of protein kinases and tens of thousands of their substrates. Together, they form cell-signalling networks that can function like microprocessors, by encoding, processing and integrating cellular information and regulating outputs in the form of myriad cellular processes, from gene expression to cell division[5]. Such capabilities are possible only because different kinases have different specificities for the many possible protein substrates.

The specificity of a kinase arises from many extrinsic and intrinsic factors. Extrinsic factors include whether the kinase and its substrate are expressed in the same cell type or in the same part of the cell, and interactions with other molecules, such as scaffolding proteins. Intrinsic factors arise from the biochemical and structural properties of the kinases and substrates. For example, the presence of electrically charged or bulky amino-acid residues in the vicinity of the phosphorylation site might promote or impede a kinase's ability to phosphorylate a given protein. Intrinsic specificity results in kinases having a preferred motif — an ideal sequence of residues that surround the phosphorylation site.

Johnson *et al.*[4] use a cell-free technique called positional scanning peptide array analysis[6] to determine the intrinsic substrate specificities of almost all kinases that target the amino acids serine and threonine — representing roughly 99% of the phosphorylation sites in human cells[7]. The approach involved screening a library of 303 human kinases to determine how capable each kinase is of transferring a phosphate group to the central serine or threonine residue of hundreds of different short strings of amino acids (Fig. 1). Remarkably, the authors find that almost two-thirds of phosphorylation sites could be assigned to one of a small handful of kinases.

The study also emphasizes the importance of 'negative selectivity' in defining substrate selection. Just as some kinases favour certain amino-acid residues at specific locations in their target, the converse — selection against such residues — also occurs. This had been known for some kinases[8], but the authors show that it is a general property driving much of the overall substrate selectivity of the 'kinome' (the complete repertoire of human kinases).

By systematically defining intrinsic substrate specificities for most human serine/threonine kinases, the authors present a wide-ranging comparison of these enzymes. The kinase superfamily had previously been classified into more than 100 families, mainly on the basis of sequence comparisons of their catalytic domains[9]. Intrinsic substrate specificity was assumed to follow these family groupings, given that the sequence (and, by extension, the structure) of the catalytic domain determines substrate accessibility. But Johnson and colleagues' findings lead the authors to reclassify these kinases into at least 38 motif-based classes. They find that kinases from disparate phylogenetic families converge on similar substrate sequence specificities, and they define key specificity-determining amino-acid residues in the kinase that are responsible for some (but not all) of the observed selectivity.

The increasing availability of structural models as a result of developments such as the computational tool AlphaFold, which predicts protein structures[10], will allow structural information on kinases and their putative substrates to be incorporated into predictions of specificity. Crucial to the accuracy of these predictions will be the inclusion of modifications to the enzymes themselves in models of kinase structure (and, indeed, all protein structures), because many of these enzymes are active only when their catalytic region itself is phosphorylated.

These newly defined kinase specificities could also be used to predict the functions of

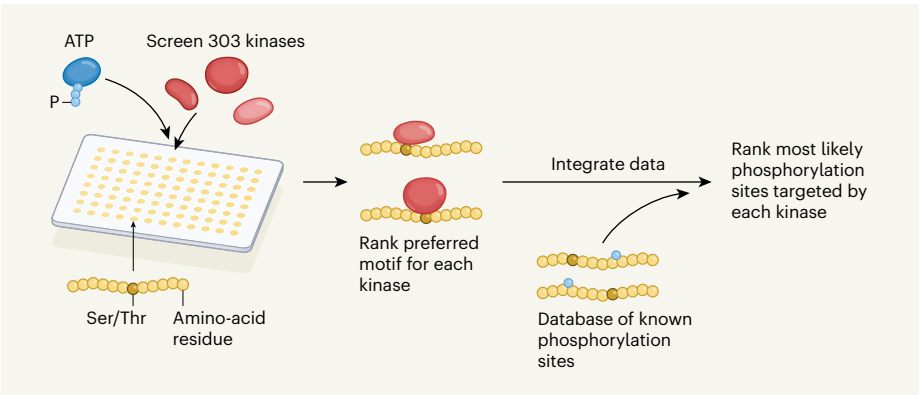

Figure 1 | Large-scale analysis of kinase-enzyme substrate specificities. To investigate the specificity of kinase enzymes for their protein substrates, Johnson *et al.*[4] generated arrays of peptide substrates — short strings of amino acids that included serine (Ser) or threonine (Thr) at their centres. They then screened 303 kinases for their effectiveness in phosphorylating Ser or Thr residues, using a phosphate group (P) donated by an ATP molecule. This allowed the authors to discover the sequence of amino-acid residues most preferred by each kinase (its preferred motif). They then combined these data with those of a repository of protein-phosphorylation sites previously identified by mass spectrometry[15] to predict and rank putative phosphorylation sites for each kinase.

the more than 100 'dark' kinases — those without any reported substrates[3]. Assembling the potential substrates of each kinase and integrating certain features (such as the tissues in which they are expressed, their function and how they are regulated) into various biological contexts could reveal the biological roles of these under-studied enzymes. This possibility should motivate researchers to venture 'into the dark', to better characterize these elusive proteins. That would be particularly beneficial because biomedical research tends to focus on certain 'favoured' proteins, even though these are not necessarily more important than any others, either for biological or medical research[11].

Today, mass spectrometry can measure tens of thousands of changes to proteins in minuscule biological samples[12]. Yet, despite the resulting deluge of information, researchers have often interpreted changes to cell-signalling networks on the basis of the responses of as few as 5% of phosphorylation sites whose kinases have been identified[3]. To deal with this shortcoming, numerous studies (for example, refs 13, 14) have set out to predict substrates for some (but not all) kinases, using the limited available experimental data. By contrast, Johnson *et al.* used their experimentally derived data to computationally rank the kinases most likely to act on most of the human phosphoproteome (Fig. 1). Researchers can now use these predictions when analysing cell-signalling data sets, to obtain a broader view of the potential network activity occurring in biological samples.

As biomedical fields continue to generate profiles of biological molecules at an accelerating rate and scale, advances in methods to contextualize these data are crucial. Johnson and colleagues' database will enable researchers to predict regulatory mediators of a molecule of interest in various biological contexts. The approach could be extended to analysis of tyrosine kinases (the largest kinase family not included here), or to other types of modification such as the addition of ubiquitin or acetyl groups, or even to the interplay between modifications, providing a more holistic view of cellular signalling.

Although intrinsic aspects alone can resolve potential substrates of the kinome remarkably well, kinase specificity is also influenced by other factors. Ideally, these factors — which include protein–protein interactions, tissue- and cell-specific expression and cellular localization — could be incorporated into topological models of cell-signalling networks. However, this would require huge amounts of data on many cell types and biological contexts. Fortunately, mass-spectrometry-based proteomics is rising to meet this challenge. Continued improvements in how we map regulatory kinases will enhance our ability to interpret the language of cell signalling.

Sean J. Humphrey is at the Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Victoria 3052, Australia. **Elise J. Needham** is in the BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 0BA, UK.
e-mail: sean.humphrey@mcri.edu.au

1. Cohen, P. *Nature Rev. Drug Discov.* **1**, 309–315 (2002).
2. Attwood, M. M., Fabbro, D., Sokolov, A. V., Knapp, S. & Schiöth, H. B. *Nature Rev. Drug Discov.* **20**, 839–861 (2021).
3. Needham, E. J., Parker, B. L., Burykin, T., James, D. E. & Humphrey, S. J. *Sci. Signal.* **12**, eaau8645 (2019).
4. Johnson, J. L. *et al. Nature* **613**, 759–766 (2023).
5. Humphrey, S. J., James, D. E. & Mann, M. *Trends Endocrinol. Metab.* **26**, 676–687 (2015).
6. Hutti, J. E. *et al. Nature Methods* **1**, 27–29 (2004).
7. Sharma, K. *et al. Cell Rep.* **8**, 1583–1594 (2014).
8. Zhu, G. *et al. J. Biol. Chem.* **280**, 10743–10748 (2005).
9. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. *Science* **298**, 1912–1934 (2002).
10. Jumper, J. *et al. Nature* **596**, 583–589 (2021).
11. Edwards, A. M. *et al. Nature* **470**, 163–165 (2011).
12. Aebersold, R. & Mann, M. *Nature* **537**, 347–355 (2016).
13. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
14. Linding, R. *et al. Nucleic Acids Res.* **36**, D695–D699 (2008).
15. Ochoa, D. *et al. Nature Biotechnol.* **38**, 365–373 (2020).

The authors declare no competing interests.
This article was published online on 11 January 2023.