PTMcode: a database of known and predicted functional associations between post-translational modifications in proteins

Pablo Minguez[1], Ivica Letunic[2], Luca Parca[3] and Peer Bork[1,4,*]

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, [2]Biobyte solutions GmbH, Bothestrasse 142, 69126 Heidelberg, Germany, [3]Department of Biology, Centre for Molecular Bioinformatics, University of Rome 'Tor Vergata', Via della Ricerca Scientifica snc, 00133 Rome, Italy and [4]Max-Delbruck-Centre for Molecular Medicine, Berlin-Buch, Germany

ABSTRACT

Post-translational modifications (PTMs) are involved in the regulation and structural stabilization of eukaryotic proteins. The combination of individual PTM states is a key to modulate cellular functions as became evident in a few well-studied proteins. This combinatorial setting, dubbed the PTM code, has been proposed to be extended to whole proteomes in eukaryotes. Although we are still far from deciphering such a complex language, thousands of protein PTM sites are being mapped by high-throughput technologies, thus providing sufficient data for comparative analysis. PTMcode (http://ptmcode.embl.de) aims to compile known and predicted PTM associations to provide a framework that would enable hypothesis-driven experimental or computational analysis of various scales. In its first release, PTMcode provides PTM functional associations of 13 different PTM types within proteins in 8 eukaryotes. They are based on five evidence channels: a literature survey, residue co-evolution, structural proximity, PTMs at the same residue and location within PTM highly enriched protein regions (hotspots). PTMcode is presented as a protein-based searchable database with an interactive web interface providing the context of the co-regulation of nearly 75 000 residues in >10 000 proteins.

INTRODUCTION

Most eukaryotic proteins are targeted by a multitude of post-translational modifications (PTMs) that fine-tune their function as a rapid response to stimuli without involvement of genomic, transcriptomic or translational regulation. These PTMs are present in various types and combinations, and their on–off status can vary during the life time of a protein, thereby fine-tuning its function, localization and interaction with other molecules. Specific mechanisms have been associated to particular pairs of PTM types, like the competition for serine and threonine residues by phosphorylation and O-linked glycosylation (1,2) or the promotion of ubiquitination by phosphorylation that leads to protein degradation (3), and there are extensive studies describing the regulation by PTM interplay of individual proteins such as the tumor suppressor p53 (4) or the well known compilation of molecular switches that occur within histone tails (5), suggesting the existence of a 'PTM code' (6–9) based on the presence and the association of several PTMs that leads to perform a particular function (10). Most of the studies trying to decipher this molecular barcode are based on single proteins and few PTM types (4,11); however, thanks to the recent technological advances in the mass spectrometry-based detection methods (12), an increasing amount of data about protein modifications is becoming available, increasing the diversity and abundance of reported PTMs, although it is far from being complete. Yet, deciphering a potential PTM code remains a difficult challenge, as the collection of the parts lists, the PTMs repertoire, is only a first step, and individual PTMs have to be functionally associated, somewhat analogous to the delineation of proteomes whereby individual proteins are involved in complex protein–protein interactions.

Only recently, several studies have explored the association of several PTM types in whole proteomes based on, for instance, the study of acetylation status by the systematic perturbation of kinases (13), structural changes in

modified residues in response to *in-silico* perturbations of other modification sites (14), the competition of several PTMs for a residue (15), the presence of clusters of PTMs within the protein sequence (16) or the co-evolution of the modified residues, where pairs of particular PTM types were associated to specific protein localizations, functions and protein functional units such as short linear motifs and globular domains (17). The latter being the first attempt to characterize the interplay between a large number of PTM types in several eukaryotes. Based on these recent independent developments, one should be able to derive coherent functional associations between modified residues, thus adding value to information about individual modifications, stored in classical protein resources (18) or PTM specific databases (19–22), again analogous to protein sequence and protein-interaction databases. A functional association between two PTMs should be seen here as a broad concept that not only stands for a physical interaction or a competition (PTM crosstalk) but also describing more broad associations, i.e. PTMs that are not present in the protein at the same time but involved in the same protein function.

The PTMcode database combines results from several large-scale analyses that identify known and predicted functionally associated PTMs of 13 different types from 8 eukaryotes. As the first large-scale public database providing this kind of information, we believe that PTMcode will enable both computational and molecular biology laboratories to further PTM research in many ways and at various scales, ranging from individual mechanistic studies to global network analyses towards a global PTM code.

RESULTS

Available PTMs and their functional associations

We extracted all experimental validated PTMs available at UniProt (18), PHOSIDA (19), PhosphoSite (20), PhosphoELM (21), O-GlycBase (22), dbPTM (23) and HPRD (24) and performed a pre-processing task to avoid redundancy and non-matching modifications. First, protein Ids from the sources were converted to a reference Id, taken from the STRING database (25), and second, the modified residues were required to match the amino acids in a reference sequence, taken from the eggNOG database (26) which in turn, fetches the longest protein isoform from the Ensembl database (27). In total, we integrated 136258 experimentally determined, non-redundant PTMs of 13 different types in 25765 proteins of 8 different eukaryotes (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus*, *Gallus gallus*, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae*).

We combine five different channels for the extraction of known or predicted functional association between PTMs (Figure 1): (i) sites that co-evolve across many eukaryotes as described in (17); (ii) PTMs associated based on their proximity in the protein structure, as modified sites that tightly co-operate seems to be clustered (4,28); (iii) PTMs
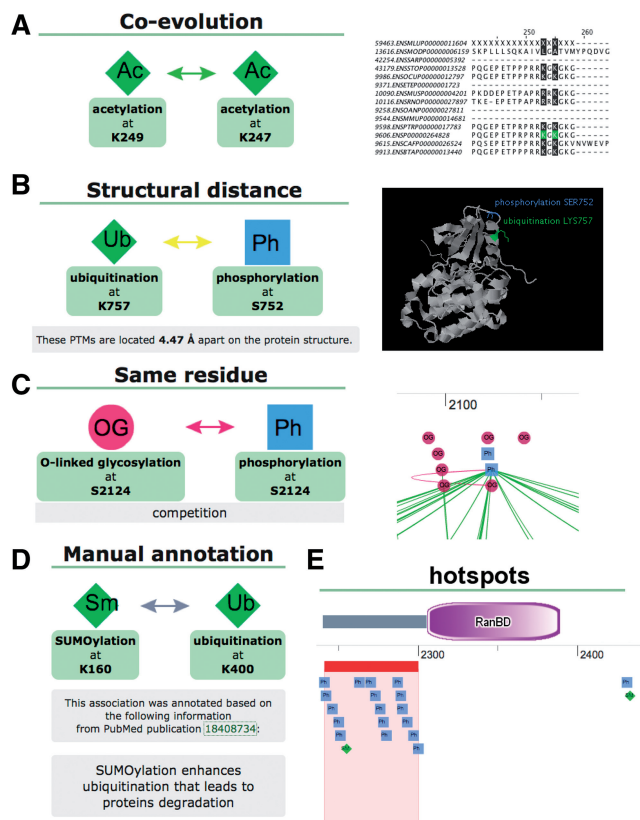

Figure 1. PTMcode integrates five types of evidence channels to collect known and predicted functional associated PTMs within the same protein, named in here as: (**A**) 'Co-evolution' where two modified residues are found to be significantly co-evolving across eukaryotes. The multiple sequence alignment in the right corresponds to the orthologous group of the protein modified by the two acetylated lysines shown. (**B**) 'Structural distance' where two modified residues are found closer than a threshold (based on the distance of known PTM interactions). (**C**) 'Same residue' where the two modifications target the exact same residue, for instance, serine 2124 of the mouse protein Bsn can be modified either by phosphorylation or O-linked glycosylation. (**D**) 'Manual annotation', derived from a literature survey in order to cover known associated PTMs. (**E**) 'PTM hotspots' that represent protein regions enriched in modifications, the figure shows a region of 60 amino acids where 18 of them can be modified.

known to modify the same residue in the protein sequence; (iv) a manual annotation survey to identify known PTM crosstalks in the literature; and (v) PTMs that are located within PTM 'hotspots', significant high-density modified regions within the protein sequence (16). From the four channels that assign pairwise associations (co-evolution, structural distance, PTMs modifying the same residue and manual annotation), PTMcode holds 401690 distinct functional associations describing the co-regulation of 74839 residues in 10410 proteins. In addition, 7400 residues have been extracted from 1635 computed regions with high PTM density. More details on the PTM associations content are in Table 1.

PTMs in the context of protein domains and structures

PTMcode is not a resource aiming at PTM collections, but instead for exploration, retrieval and analysis of predicted

Table 1. Number of functional associations that are predicted by each of the evidence channels and those that at least have two different evidences, only considering the four types of evidences that assign concrete pairwise association (co-evolution, structural distance, same residue and manual annotation)

	Co-evolution	Structural distance	Same residue	Manual annotation
Co-evolution	399 707			
Structural distance	1101	2466		
Same residue	0	0	576	
Manual annotation	13	6	0	57

and known functional association between PTMs within proteins. Yet, we map PTMs onto curated protein domains and unstructured regions as defined by the database SMART (29). PTMs become then part of the functional annotation, and their associations can be seen as part of the whole regulation of protein domains. We classify PTMs into three categories: 'regulatory' (those involved in regulation of protein function), 'stabilizing' (those that are not involved in regulation of function but required for conformational purposes) and 'uncharacterized' (those with unknown or unclear function, as in the case of C-linked glycosylations). In addition, we added, when applicable, a view of the three dimensional structure of the domain or the entire protein, highlighting any two modified residues that are predicted to be functionally associated by any of our five source channels. The structure is displayed using the popular java viewer Jmol (http://www.jmol.org) in an interacting way and can be explored by using all Jmol features. Information about modifying enzymes (such as protein kinases) is also provided by extracting it from several sources (19–21,24).

PTMcode also supplies a score for each PTM, the relative Residue Conservation Score (rRCS), that measures the conservation of the modified residue over the oldest eukaryotic orthologous group where the protein in present and takes into account both the conservation of the residue within the orthologous proteins and the evolutionary distance between the species with the conserved residue; for full details on rRCS algorithm and performance see (17). An rRCS >95 means that the modified residue is more conserved than the 95% of the same amino acids within the same type of protein region. The rRCS can be used as a proxy to hint at PTM functionality in the absence of other data. Conservation has been used before for the same purpose (30–32), although caution is required in its interpretation (33). Yet, several other PTM databases provide simple information about protein and residue conservation to be used as a filter for functional sites (19,20).

Co-evolution

We used the co-evolution of two modified residues to predict their functional association as described in (17). This strategy showed already that co-evolving pairs of different PTM types can be specifically linked in proteins with certain functionalities, localizations and can even potentially co-regulate protein interactions through their association to particular protein domains and short linear motifs (17). The functional associations provided by this prediction channel should be considered from a broad perspective, as they can range from physical interactions (as seen by the overlap with pairs of PTMs found close in the protein structure) to their participation in the same protein functionality although not necessarily at the same stage. The species where both amino acids are conserved over the protein orthologous groups are shown in the co-evolution pop-up window in which the protein alignment with the respective columns highlighted can be visualized using Jalview (34).

Structural distance

A straightforward mechanism of two PTMs to be associated is based on their proximity (4,28), measured here using the 3D structure of the protein. If they are close enough, they could be either competing for the same space, i.e. methylation inhibiting the phosphorylation of adjacent serines (35), or co-operating in the regulation of the same protein region [i.e. the highly modified cassette of amino acids in p53 (4)]. We mapped PTM residues to three-dimensional structures of proteins from the Protein Data Bank (36) and calculated the spatial distance between pairs of modified residues. To delineate a first estimation for an appropriate distance to conclude physical interaction, we measured the average distance for 12 pairs of associated modifications reported in the literature to physically interact. Thus, modified residues closer than 4.69 Å are predicted to either be physically in contact or being mutually exclusive competing for the same protein niche; their conformation can be visualized using the Jmol plugin.

PTMs modifying the same residue

The simplest evidence for a direct crosstalk of two PTMs is a modification of the same residue in the protein sequence, which would reflect either that they compete for the same amino acid (mutually exclusive PTMs) or that they co-operate for the same function if the modifications happen sequentially in time. Two well-known associations between PTM types are described to follow a competition strategy. Phosphorylation and O-linked glycosylation modify serine and threonine amino acids and constitute molecular switches that co-regulate protein function and localization within the so-called yin-yang sites (1,2). The promiscuous amino acid lysine can be acetylated, SUMOylated, ubiquitinated and methylated, and it has been described to be co-regulated by several PTM types at the same position, for example, during the regulation of histone tails (5).

We identified 576 residues regulated by this channel, mostly between the above reported cases but also between other pairs of PTM types [i.e. 10 instances of hydroxylated and O-linked glycosylated lysines that happen sequentially in time in collagen proteins (37) or 7 instances between phosphorylation and sulfation].

PTMcode provides a tentative annotation for the pairs of associated PTMs predicted by this channel, classifying them as 'competing', 'co-operating' or 'uncharacterized' associations.

Manual annotation

PTMcode not only holds predicted associations but also PTM sites that are reported in the literature to crosstalk. They were extracted using the literature review about PTM types interplay in Minguez *et al.* (17) and are now introduced in the database after their mapping into the correct protein sequences. Links to the scientific articles through PubMed and a short description on the mechanism of action are provided for the 57 associations found using this channel.

Hotspots

PTMs can also be part of regulatory hotspots, small regions in the protein sequence that are enriched in modifications (4). Such regions have been recently defined (16), and rules have been established and benchmarked. For example, modified lysines are more probably located within a distance of 15 amino acids to a phosphorylated residue therefore forming hotspot regions where PTMs tend to cluster. According to the established rules, for each of the modified residues in a protein, we define a window of 31 amino acids (15 downstream and 15 upstream), count the number of modifications there and compared them using a Fisher exact test to the number of modifications in the whole protein. All resultant *P*-values were adjusted by False Discovery Rate, and overlapping regions were collapsed to give a total of 1635 hotspots that are visible and explorable within the PTMcode web interface.

PTMcode web interface. Query, results and availability

PTMcode is accessible through the url http://ptmcode. embl.de. The web interface provides a browser to access all proteins that have some known or predicted functional associations between PTMs (Figure 2) and a search engine where the user can introduce a protein sequence or any protein id. In addition, the user can restrict the search to a particular residue or protein region. PTMcode is a protein-oriented database, as one of the major motivation for the resource is to help experts of particular proteins or research fields to explore functional hypotheses. A flash-based graphic interface enables intuitive interactivity. A single or many functional associations can be explored, and supporting information (alignment, structure etc) for each of the five channels can be easily called, all within the context of the protein domain architecture with links to the respective SMART entries. Figure 3 shows an overview of the predicted functional co-regulation by several post-translational modifications of a protein within the PTMcode environment.
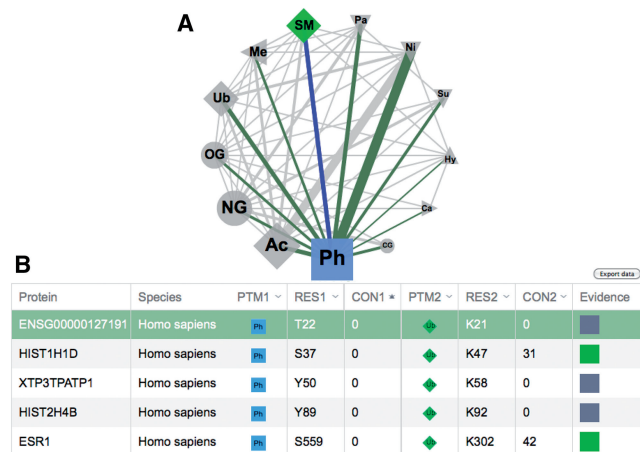

Figure 2. The PTMcode database uses in its first release 13 different types of PTMs that are abbreviated in a two letter code as: Ph (phosphorylation), NG (N-linked glycosylation), Ac (acetylation), OG (O-linked glycosylation), Ub (ubiquitination), Me (methylation), SM (SUMOylation), Hy (hydroxylation), Ca (carboxylation), Pa (palmitoylation), Su (sulfation), Ni (nitrosylation) and CG (C-linked glycosylation). The known and predicted functional associations between two of these types of modifications are accessible by selecting any two PTM types in an interactive network that offers all possible connections (**A**). These connections are constrained by the fact that both modifications at least should happen in the same protein. The PTM types are represented here by symbols, the size of which corresponds to their relative abundance in the database. The link widths represent the number of proteins modified by the two respective PTM types normalized by the total number of proteins harbouring the less abundant PTM type. The width thus indicates relative coverage of the particular pair-wise functional associations. For instance, based on the link widths, we see that a large proportion of proteins that are nitrosylated are also phosphorylated, while the proportion of proteins that are both hydroxylated and phosphorylated is smaller. Upon selection in the WEB interface, two types of modifications are activated, and all the pairs of known and predicted functional association in all proteins are shown in a table (**B**) where each of the entries can be further explored. These tables are available for download as text files.

Conclusion and future plans

PTMcode is a unique database in that it goes beyond mere compilation of PTMs, which is already covered by useful resources either detailed for the single PTM type (21) or for various PTM types (19,20,23). The aim is to put these PTMs into context, and focus of the first PTMcode release is to capture known and predicted functional associations between PTMs within proteins. Other databases start to implement functional context of the individual PTMs providing for instance, conservation information, mapping PTMs into proteins domains or 3D structures. Ptmfunc (16) for example provides information about the regulation of binding interfaces, protein domains and even the presence within a hotspot, although it lacks of a graphical display of the protein. PTMcode goes a step further and is PTM association-centric, aiming at a description of a co-regulation landscape of a protein by means of its modified sites. In the future, we plan to develop PTMcode further by introducing more PTMs, more accurate prediction methods and to extend to associations between proteins in order to prepare the ground for deciphering the global PTM code.
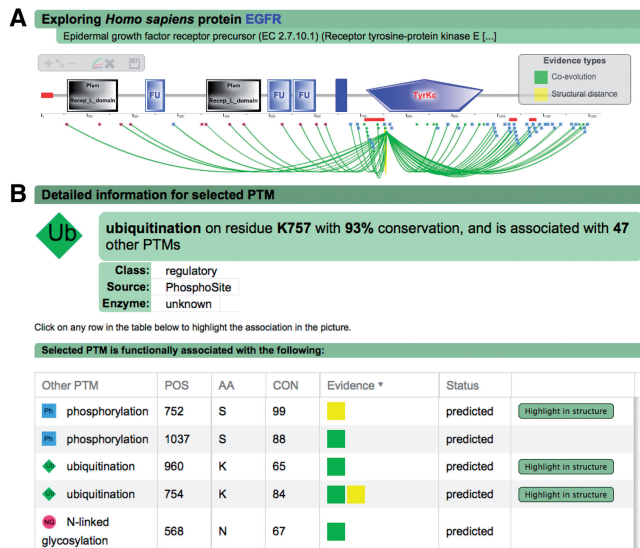
Figure 3. PTMcode offers the exploration of post-translational regulation within thousands of proteins. (**A**) Interactive graphical display of functional associations between PTMs within the human EGF receptor (EGFR). The protein is represented by the grey line at the top with globular domains and unstructured regions taken from the database SMART. Below this, PTM hotspots are shown as red lines (in EGFR, three hotspots were identified). At the bottom, the different PTMs are mapped and by clicking at one, the functional associations with any other PTM are shown by arches coloured according to the five evidence channels. All functional associations for a particular PTM can be further explored from a table (**B**) that is interactively displayed. The conservation score for the PTMs that are predicted to be associated is listed in the table together with all the evidences that support the prediction. Clicking on each of the evidences will show a pop-up with more detailed information, for example, alignments or mapping of PTM pairs onto three dimensional structures.

ACKNOWLEDGEMENTS

Conflict of interest statement: None declared.

REFERENCES

1. Zeidan,Q. and Hart,G.W. (2010) The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *J. Cell Sci.*, **123**, 13–22.
2. Butt,A.M., Khan,I.B., Hussain,M., Idress,M., Lu,J. and Tong,Y. (2011) Role of post translational modifications and novel crosstalk between phosphorylation and O-beta-GlcNAc modifications in human claudin-1, -3 and -4. *Mol. Biol. Rep.*, **39**, 1359–1369.
3. Vodermaier,H.C. (2004) APC/C and SCF: controlling each other and the cell cycle. *Curr. Biol.*, **14**, R787–R796.
4. Brooks,C.L. and Gu,W. (2003) Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr. Opin. Cell Biol.*, **15**, 164–171.
5. Latham,J.A. and Dent,S.Y.R. (2007) Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.*, **14**, 1017–1024.
6. Hunter,T. (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell*, **28**, 730–738.
7. Yang,X.J. and Seto,E. (2009) Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol. Cell*, **31**, 449–461.
8. Creixell,P. and Linding,R. (2012) Cells, shared memory and breaking the PTM code. *Mol. Syst. Biol.*, **8**, 598.
9. Benayoun,B.A. and Veitia,R.A. (2009) A post-translational modification code for transcription factors: sorting through a sea of signals. *Trends Cell Biol.*, **19**, 189–197.
10. Seet,B.T., Dikic,I., Zhou,M.M. and Pawson,T. (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.*, **7**, 473–483.
11. 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.
12. Choudhary,C. and Mann,M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.*, **11**, 427–439.
13. van Noort,V., Seebacher,J., Bader,S., Mohammed,S., Vonkova,I., Betts,M.J., Kühner,S., Kumar,R., Maier,T., O'Flaherty,M. *et al.* (2012) Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol. Syst. Biol.*, **8**, 571.
14. Lu,Z., Cheng,Z., Zhao,Y. and Volchenboum,S.L. (2011) Bioinformatic analysis and post-translational modification crosstalk prediction of lysine acetylation. *PLoS One*, **6**, e28228.
15. Danielsen,J.M., Sylvestersen,K.B., Bekker-Jensen,S., Szklarczyk,D., Poulsen,J.W., Horn,H., Jensen,L.J., Mailand,N. and Nielsen,M.L. (2011) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. Proteomics*, **10**, M110.003590.
16. Beltrao,P., Albanèse,V., Kenner,L.R., Swaney,D.L., Burlingame,A., Villén,J., Lim,W.A., Fraser,J.S., Frydman,J. and Krogan,N.J. (2012) Systematic functional prioritization of protein posttranslational modifications. *Cell*, **150**, 413–425.
17. Minguez,P., Parca,L., Diella,F., Mende,D.R., Kumar,R., Helmer-Citterich,M., Gavin,A.C., van Noort,V. and Bork,P. (2012) Deciphering a global network of functionally associated post-translational modifications. *Mol. Syst. Biol.*, **8**, 599.
18. The UniProt Consortium. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **40**, D71–D75.
19. Gnad,F., Gunawardena,J. and Mann,M. (2010) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.*, **39**, D253–D260.
20. Hornbeck,P.V., Kornhauser,J.M., Tkachev,S., Zhang,B., Skrzypek,E., Murray,B., Latham,V. and Sullivan,M. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **40**, D261–D270.
21. Dinkel,H., Chica,C., Via,A., Gould,C.M., Jensen,L.J., Gibson,T.J. and Diella,F. (2010) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
22. Gupta,R., Birch,H., Rapacki,K., Brunak,S. and Hansen,J.E. (1999) O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res.*, **27**, 370–372.
23. Lee,T.Y., Huang,H.D., Hung,J.H., Huang,H.Y., Yang,Y.S. and Wang,T.H. (2006) dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res.*, **34**, D622–D627.
24. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
25. Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguez,P., Doerks,T., Stark,M., Muller,J., Bork,P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.

26. Muller,J., Szklarczyk,D., Julien,P., Letunic,I., Roth,A., Kuhn,M., Powell,S., von Mering,C., Doerks,T., Jensen,L.J. *et al.* (2010) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.*, **38**, D190–D195.

27. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.

28. Christensen,B., Nielsen,M.S., Haselmann,K.F., Petersen,T.E. and Sørensen,E.S. (2005) Post-translationally modified residues of native human osteopontin are located in clusters: identification of 36 phosphorylation and five O-glycosylation sites and their biological implications. *Biochem. J.*, **390**, 285–292.

29. Letunic,I., Doerks,T. and Bork,P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.*, **40**, D302–D305.

30. Boekhorst,J., van Breukelen,B., Heck,A. and Snel,B. (2008) Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. *Genome Biol.*, **9**, R144.

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

32. Chen,S.C., Chen,F.C. and Li,W.H. (2010) Phosphorylated and nonphosphorylated serine and threonine residues evolve at different rates in mammals. *Mol. Biol. Evol.*, **27**, 2548–2554.

33. 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.

34. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.

35. Zhang,K., Lin,W., Latham,J.A., Riefler,G.M., Schumacher,J.M., Chan,C., Tatchell,K., Hawke,D.H., Kobayashi,R. and Dent,S.Y.R. (2005) The Set1 methyltransferase opposes Ipl1 aurora kinase functions in chromosome segregation. *Cell*, **122**, 723–734.

36. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

37. Oikarinen,A., Anttinen,H. and Kivirikko,K.I. (1976) Hydroxylation of lysine and glycosylation of hydroxylysine during collagen biosynthesis in isolated chick-embryo cartilage cells. *Biochem. J.*, **156**, 545–551.