

# PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins

Pablo Minguez<sup>1</sup>, Ivica Letunic<sup>2</sup>, Luca Parca<sup>1</sup>, Luz Garcia-Alonso<sup>3</sup>, Joaquin Dopazo<sup>3</sup>, Jaime Huerta-Cepas<sup>1</sup> and Peer Bork<sup>1,4,\*</sup>

<sup>1</sup>European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany, <sup>2</sup>Biobyte solutions GmbH, Bothestr 142, 69117 Heidelberg, Germany, <sup>3</sup>Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain and <sup>4</sup>Max-Delbrück-Centre for Molecular Medicine, Berlin-Buch, Germany

Received September 15, 2014; Revised October 13, 2014; Accepted October 16, 2014

## ABSTRACT

The post-translational regulation of proteins is mainly driven by two molecular events, their modification by several types of moieties and their interaction with other proteins. These two processes are interdependent and together are responsible for the function of the protein in a particular cell state. Several databases focus on the prediction and compilation of protein–protein interactions (PPIs) and no less on the collection and analysis of protein post-translational modifications (PTMs), however, there are no resources that concentrate on describing the regulatory role of PTMs in PPIs. We developed several methods based on residue co-evolution and proximity to predict the functional associations of pairs of PTMs that we apply to modifications in the same protein and between two interacting proteins. In order to make data available for understudied organisms, PTMcode v2 (<http://ptmcode.embl.de>) includes a new strategy to propagate PTMs from validated modified sites through orthologous proteins. The second release of PTMcode covers 19 eukaryotic species from which we collected more than 300 000 experimentally verified PTMs (>1 300 000 propagated) of 69 types extracting the post-translational regulation of >100 000 proteins and >100 000 interactions. In total, we report 8 million associations of PTMs regulating single proteins and over 9.4 million interplays tuning PPIs.

## INTRODUCTION

The complexity of the eukaryotic cell cannot be explained only by the number of genes and proteins but by their complex degree of regulation which includes several levels and

mechanisms. Thus, one of the biggest challenges of the new molecular biology is to understand these systems globally in order to extract the active part of the regulation landscape from one or correlated snapshots of a cell state. One important step of this regulation is performed after translation where protein function is defined by the interplay between protein–protein interactions (PPIs) and post-translational modifications (PTMs). These two mechanisms are interdependent since PPIs are described to be regulated by PTMs (1,2) and intermediate enzymes are also subject of modification.

PTMs are indeed an abundant (3) and widely spread (4) source of protein regulation. They are involved in a vast number of functions, from protein stabilizing factors (5) to regulators of molecular switches (6). Their outcome on the proteins depends on the type of the modification, several hundreds are described, and on their possible combinations, PTMs may not act alone but coupled to others by their co-operation or competition (7). A few well-studied examples of functions defined by these types of combinatorial patterns are described in particular protein families, such as the regulation of histones (8) or the regulation of the family of transcription factors FoxO (9). Some universal PTM combinations also have a known role, as phosphorylation coupled to ubiquitination in the protein degradation pathway (10), and there are several cases of PTM types modifying the same residue in a competitive manner (11). These and other evidences suggest the existence of a universal molecular barcode, dubbed the PTM code, that would encrypt the signal for the regulation of protein location and function including their interaction with other proteins.

Here we present the second release of the PTMcode database, a resource for known and predicted PTM functional associations. For this update we have tripled the number of compiled experimentally verified PTMs (up to ~300 000) and spread their signal through conserved sites in orthologous proteins of 19 eukaryotic species augmenting this

\*To whom correspondence should be addressed. Tel: +49 6221 387 8361; Fax: +49 6221 387 517; Email: [bork@embl.de](mailto:bork@embl.de)

number to 1.3 million. In total, we provide PTM functional annotation for more than 100 000 proteins. For each of them, PTMcode outlines its known and predicted post-translational regulatory landscape which includes the functional links between its modified residues and from this second release also the functional associations between PTMs that may regulate its interaction with other proteins. PTMcode v2 is accessible through the url <http://ptmcode.embl.de>.

## METHODS SUMMARY

### Relative Residue Conservation Score (rRCS) calculation

The rRCS measures the conservation of an amino acid within a multiple sequence alignment (MSA) from an orthologous group of proteins. It takes into account the occurrence of the amino acid in the exact position and the maximum branch distance between the species of the proteins with the amino acid conserved, for more details see (12). For this release we used orthologous groups from eggNOG 4.0 (13) and a species tree generated out of marker genes. ETE (14) and treeKO (15) python libraries were used to manipulate trees. DisEMBL algorithm was used for the calculation of protein disordered regions (16).

### Co-evolution algorithm

We use mutual information (MI; 17–19) to measure co-evolution of two modified residues. MI is calculated for every pair of PTMs in a protein using the MSA of its most ancient orthologous group. A background distribution of MI values is calculated from non-modified residues of the same protein, type of amino acid and located in similar protein regions (ordered or disordered). Pairs of PTMs with a MI value higher than 95% of the background distribution are selected as co-evolving. For the residues where MI cannot be calculated (for restriction of MI calculation see 12), we calculate the ratio of the conserved site and compared to the distribution of the non-modified sites with the same limitation taken as background, pairs with a ratio above 95% of the background distribution are selected as co-evolving.

### Co-evolution algorithm for PPIs

The two proteins of the PPI are mapped to their most ancient orthologous groups. The MSAs from these two orthologous groups are pruned to keep only proteins of their common species. Thus, we guarantee a fair comparison between the conservation of the modified residues over the organisms present in both alignments. After this, the algorithm behaves as described in previous section.

## RESULTS

### Mapping PTMs from various sources into a single framework

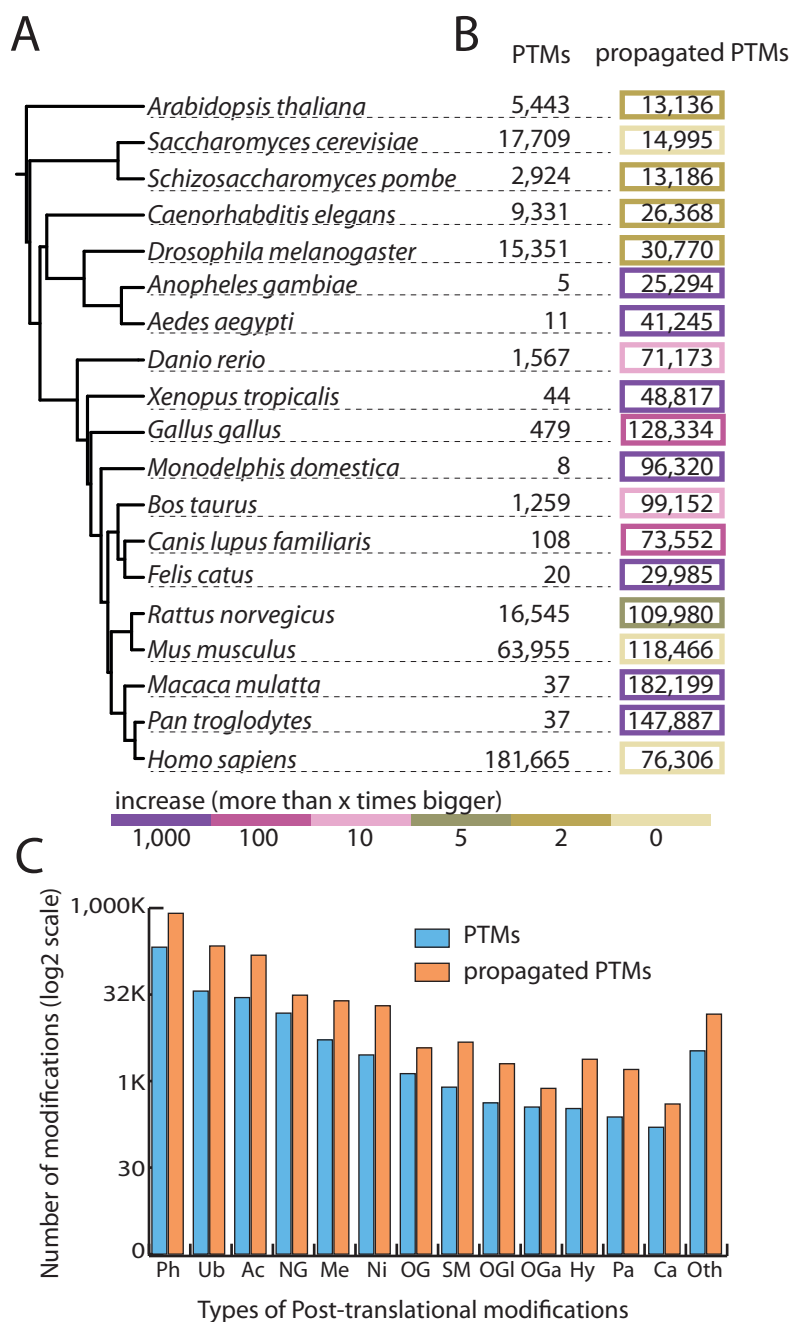
PTMcode v2 provides a highly curated data set of PTMs collected from six public databases: UniProt (20), PHOSIDA (21), PhosphoSite (22), PhosphoELM (23), dbPTM (24) and HPRD (25) and from nine high-throughput experiments reported in papers (11,26–33). Our

input data set has tripled the previous release numbers and now consists of 316 546 experimentally verified PTMs of 69 different types spread over 45 361 proteins from 19 eukaryotes (Figure 1A). Modified residues were mapped into reference protein sequences from the eggNOG 4.0 database and each source validated for every protein requiring all its PTMs matching the correct amino acid. Besides the extensive exercise of PTM compilation, the value of our data set relies on the framework built for its collection, mapping, annotation and finally its visualization and data retrieval. PTMcode uses the same protein repertoire and synonyms dictionary as the databases eggNOG (13) and STRING (34), a powerful tool that allows us to easily include orthologs and network neighborhood information. We use the protein orthologous groups provided by eggNOG 4.0 to calculate our rRCS (12, Methods Summary), that evaluates every PTM conservation. For a user tip, a rRCS >95 means that the modified residue is more conserved than the 95% of the same type of amino acids in the same type of region of the protein, still some caution must be taken when filtering the most conserved PTMs in a protein since fast-evolving PTMs may also have functional roles as shown in phosphorylation sites (35).

PTMcode also offers functional annotation of PTMs. First, they are mapped into highly curated protein domains and unstructured regions provided by the SMART database (36), another resource of reference that allows us to show the modifications in the functional context of the protein. We also provide when available the responsible enzymes for the modifications and a simple description of their potential role as regulators or structural stabilizers depending on the nature of the PTM. As a novelty in this new release we display the number of Pubmed articles in which a PTM has been originally described (extracted directly from the annotation in the sources). This information represents an additional tip for the functionality assessment of a modified site. Indeed, the PTMs described in more Pubmed articles are generally more conserved and have a higher co-evolution rate than those described in few articles (Supplementary Figure S1) which suggests a higher probability to be functional. Finally, PTMcode's main contribution to the field is to provide known and predicted functional associations between PTMs. All these features together make PTMcode an integrative framework for the study of protein post-translational regulation.

### Propagating PTM sites to orthologs in other organisms

The new technical advances on the identification of protein-modified sites (37) have drastically increased the availability of PTMs. However, we are still far from having complete PTM repertoires (3) partly due to the condition-specificity of most PTMs (28). The cascade of data production has just started in this field and there are still many organisms with no available high-throughput studies. In addition, although hundreds of different PTM types are described, only a few have been subject of proteome-wide screenings (e.g. phosphorylation, acetylation or glycosylation). Based on these two needs, to cover more species and to amplify the number of modifications of understudied PTM types, PTMcode v2 spreads the signal from experimental validated PTMs to



**Figure 1.** The PTMcode data set. We collected experimentally verified PTMs from 19 eukaryotes (A) and spread their annotation through conserved sites in orthologs tagging the new annotated sites as ‘propagated PTMs’. (B) Shows the increase on the size of available post-translational information per species. Organisms with fewer validated PTMs are the ones showing a higher increase as shown by the color code. (C) Shows the increment of the most abundant PTM types.

the conserved sites in orthologs from other species, we tag them as ‘propagated PTMs’ (Figure 1B and C).

Indeed, conservation has been widely used as a proxy for functionality (38–40) so it is fair to consider those conserved residues as potentially modified residues with the same constraints that are applied to the use of conservation to the functional assessment of experimentally verified PTMs (35). For example, for a total of 69 875 phosphoserines that could be propagated to human proteins, we ob-

tained 15 914 PTMs matching already known human phosphorylations. This represents a 22.7% of overlap (still a big underestimation since specially the other species phosphoproteome is far from being complete) while the random expectation of hitting the correct residue is 15% taking into account the different distribution of phosphorylations in ordered and disordered regions (8038 known phosphoserines in 172 678 serines in ordered regions and 60 095 known phosphoserines in 577 025 serines in disordered).

Thus, although propagated PTMs should be taken with much caution and used only for exploratory analyses, they exhibit a reasonable overlap with known PTMs as a good indicator for their reliability. By this simple exercise we obtained 1 347 165 sites with a modification signal propagated from verified PTMs allowing us to produce post-translational information for species that otherwise count with almost none experimental data (Figure 2).

The aim is to bring the number of PTMs of organisms with no large-scale PTM surveys closer to the reality, thus, for *Pan troglodytes* from which we only could map 10 experimentally verified phosphorylations, we are able to report almost 100 000 propagated phosphorylations, definitely a closer number to the 122 000 validated phosphosites that we have in human. The percentage of increase in the number of PTMs differs among the PTM types, while types with no high-throughput screenings, such as malonylation, nitration, glycation or neddylation, have a high increment (12.1%, 9.8%, 9% and 9.1%, respectively), the most studied PTM types, such as phosphorylation, ubiquitination and acetylation, have an increase of 3.8%, 6% and 5.4%.

### Predicting functional links between PTMs within proteins

The main focus of the PTMcode database is the study of PTMs that are functionally linked. Several types of associations are possible (41): PTMs mutually exclusive that compete for the same residue (11), PTMs that are required for others to take place in a signaling cascade (42), PTMs that are close enough to influence each other (1) or even modifications involved in concerted allosteric conformational changes (43). In order to cover this broad spectrum of possible scenarios PTMcode counts on five evidence channels that explore different properties of the coupled PTMs: (i) the 'co-evolution' channel (Figure 3A) that extracts pairs of PTMs with a similar evolutionary history (see Methods Summary); (ii) the 'structural distance' channel (Figure 3B) that reports close modified residues in 3D structures; (iii) the 'competition' channel which highlights residues modified by several PTM types; (iv) the 'manual annotation' channel that offers associated PTM pairs described in the literature; and (v) finally, the 'hotspots' channel that calculates high-density modified protein regions.

We applied our previously described five methodologies (44) to the new data set which includes the propagated PTMs. Thus, PTMcode v2 has now more than 1.2 million functional links of PTMs within the same protein formed of 205 571 PTMs in 21 713 proteins (~8 million pairs of 1.2 million PTMs in 100K proteins including the propagated PTMs) representing an increase of ~3.2 times of reported PTM associations (20 times including propagated) compared to previous release.

The 'co-evolution' channel is the one that provides the majority of our predictions. In a previous work (12) we showed that co-evolution of PTM pairs can be used as a proxy for their functional association. Indeed, co-evolving PTM pairs were shown to be associated with protein short-linear motifs and globular domains as well as to be closer in sequence and space. Moreover, sets of proteins with particular types of co-evolving PTMs were enriched in certain functions, locations and PPI clusters compared to proteins with

non-co-evolving PTMs of the same type. On the other hand, co-evolution is also the evidence channel harder to interpret in terms of mechanism as it may point to very broad functional associations. In this release we have fine-tuned our co-evolution algorithm taking extra controls as we extract pairs of PTMs with a higher co-evolution rate than pairs of non-modified residues in the same protein with the same type of amino acid and placed on similar protein regions (Methods Summary) while before we used the random expectation as background. For instance, a protein ubiquitinated in a disordered region and with a phosphoserine in an ordered region would have as background distribution the co-evolution values of the pairs formed by all non-modified lysines in disordered regions and non-modified serines in ordered regions.

Thus, in PTMcode v2 we report over 1.2 million co-evolving PTM pairs within more than 20 000 proteins (almost 8 million interplays in ~98 000 proteins if we include the propagated PTMs). Co-evolving PTM pairs with a co-evolution score higher than 95% of the background distribution are reported within the 'co-evolution' channel and they may be further explored using the Jalview plugin (45), that visualizes the protein sequence alignment of the orthologous group used for the calculation. The common species where the residues are conserved are also reported in the co-evolution pop-up window.

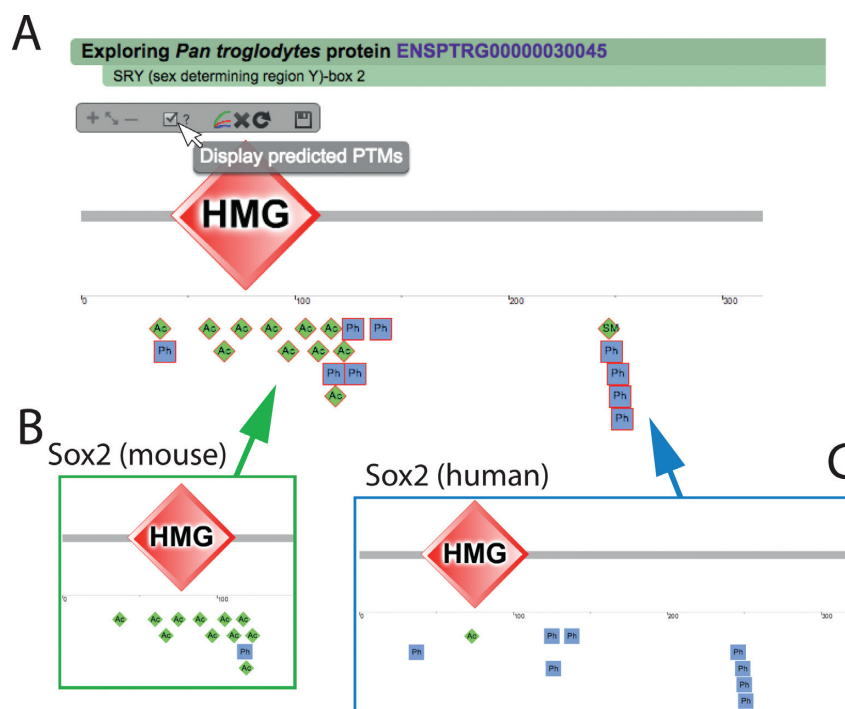
### Predicting functional links between PTMs in interacting proteins

Probably the most ambitious aim for this second PTMcode release is to extract associations between PTMs placed on interacting proteins. These functional links would be candidates to regulate the interaction although several other indirect associations are possible. Several computational approaches have shown enrichments in PTMs clusters in proteins complexes (46) and a higher number of interaction partners in modified proteins compared to non-modified (47). PPIs are indeed subject of this type of regulation mostly due to the activity of PTMs located in protein interfaces (1) although allosteric regulation has also been reported (48). PTMcode v2 provides two channels for the extraction of associated PTMs in the regulation of two interacting proteins, a 'co-evolution' and a 'structural distance' channel.

We adapted our co-evolution algorithm to measure the interplay between two modified residues from different proteins (see Methods Summary) and applied it to PTMs in physically interacting proteins taken from the STRING 9.1 database (34) with a score over 700. In total, PTMcode v2 provides over 3.6 million predicted PTM associated pairs (~9.2 million considering propagated PTMs) through the 'co-evolution' channel covering ~11 000 proteins (~31 000 including propagated PTMs) in ~44 000 PPIs (~102 000 adding the propagated PTMs).

In order to assess the accuracy of our co-evolution algorithm catching pairs of associated PTMs within interacting proteins, we compared the results in the PPI data set with the co-evolution scores calculated from pairs of PTMs placed on experimentally proved non-interacting proteins extracted from the Negatome database (49). Although we





**Figure 2.** Propagated PTMs. Users can choose to display the propagated PTMs by clicking in the checkbox in the display menu (A). Here, we show the protein Sox2 in *Pan troglodytes* (A) which had no PTMs experimentally validated while several are predicted from orthologs, mainly acetylations from the mouse copy of the protein (B) and phosphorylations from human (C). Those PTMs were also propagated between mouse and human (not displayed in the figure).

used the most stringent data set available, Negatome cannot guarantee that the protein pairs do not have another functional association rather than their binding. Still, we found that PPIs show a higher rate of PTMs pairs selected as co-evolving (with our usual threshold of 95) than the non-interacting proteins, compared using a fisher test ( $P$ -value  $< 0.001$ ).

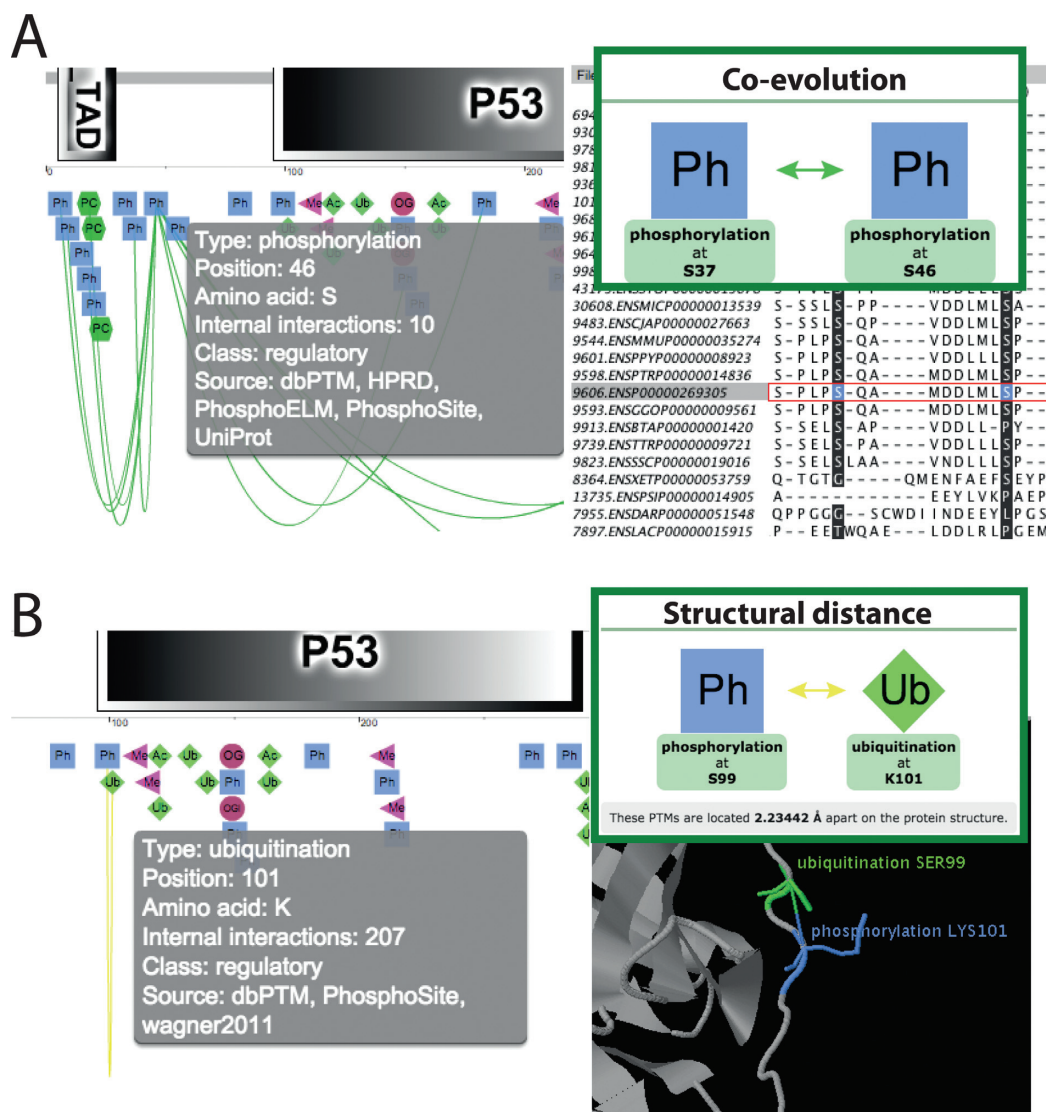
The nature of this ‘co-evolution’ channel allows us to catch not only PTMs that show a direct regulation but also those that would have more broadly defined interplays, such as PTMs, contributing to the same functionality not necessarily at the same stage. Thus, as in the case of co-evolving pairs of PTMs within the same protein users should interpret them from a broad perspective.

The ‘structural distance’ channel aims to collect the PTM pairs that may be in physical contact (if both are present at the same time) by means of their proximity. This does not exclude that distant PTMs in protein interfaces may be co-regulating the interaction. Here, we measure the distance of every two modified residues in protein interfaces that are available over Protein Data Bank (PDB) complex structures (50). We classify as ‘close enough’ the PTM pairs with residues below a separation of 4.69Å, a threshold extracted from manually curated cross-talking modifications described in our first release (44). A total of 65 000 possible cross-talking PTM pairs (~79 000 if we add propagated PTMs) are reported. The details of these predicted physical associations can be visualized and further analyzed by the Jmol plugin integrated in PTMcode.

### Visualizing PTMs associated with PPIs

PTMcode is a protein-oriented database in which users can search for their favorite protein and get information about its PTMs including their conservation assessment, an overview of the whole protein functional context and the functional associations between them (known and predicted). From this second release it also includes the analysis of the regulation of the PPIs in which the protein is involved. The entry point to the database is either a browser facility to obtain all associations of two particular PTM types or a powerful search engine that now includes the option to search for interacting proteins and jump directly to our ‘PPI view’ mode. Still from a single protein search, users are guided to the ‘single protein view’ mode that now also includes, if PPIs are available for that particular protein, basic information about how many interaction partners it has and the predictions for the PTM associations that may regulate them.

As the main novelty in visualization we have implemented a flash-based network view integrated into our interactive graphic interface where users can explore the network neighborhood centered on their favorite protein and jump to the ‘PPI view’ (Figure 4). There, the two proteins are shown facing each other and every pair of known or predicted associated PTMs may be further analyzed clicking on the channels pop-ups where extra information about every source of evidence is displayed.



**Figure 3.** Co-evolution and structural distance channels. The majority of the functional associations between PTM pairs are found by these two types of evidences. (A) The ‘co-evolution’ channel extracts pairs of PTMs that have a common evolutionary history. The protein MSA used to calculate the co-evolution score is displayed using the Jalview plugin where the two modified residues are highlighted for their closer analysis. (B) The ‘structural distance’ channels extracts PTM pairs that are close in PDB protein structures. The Jmol plugin offers further possibilities for their analysis.

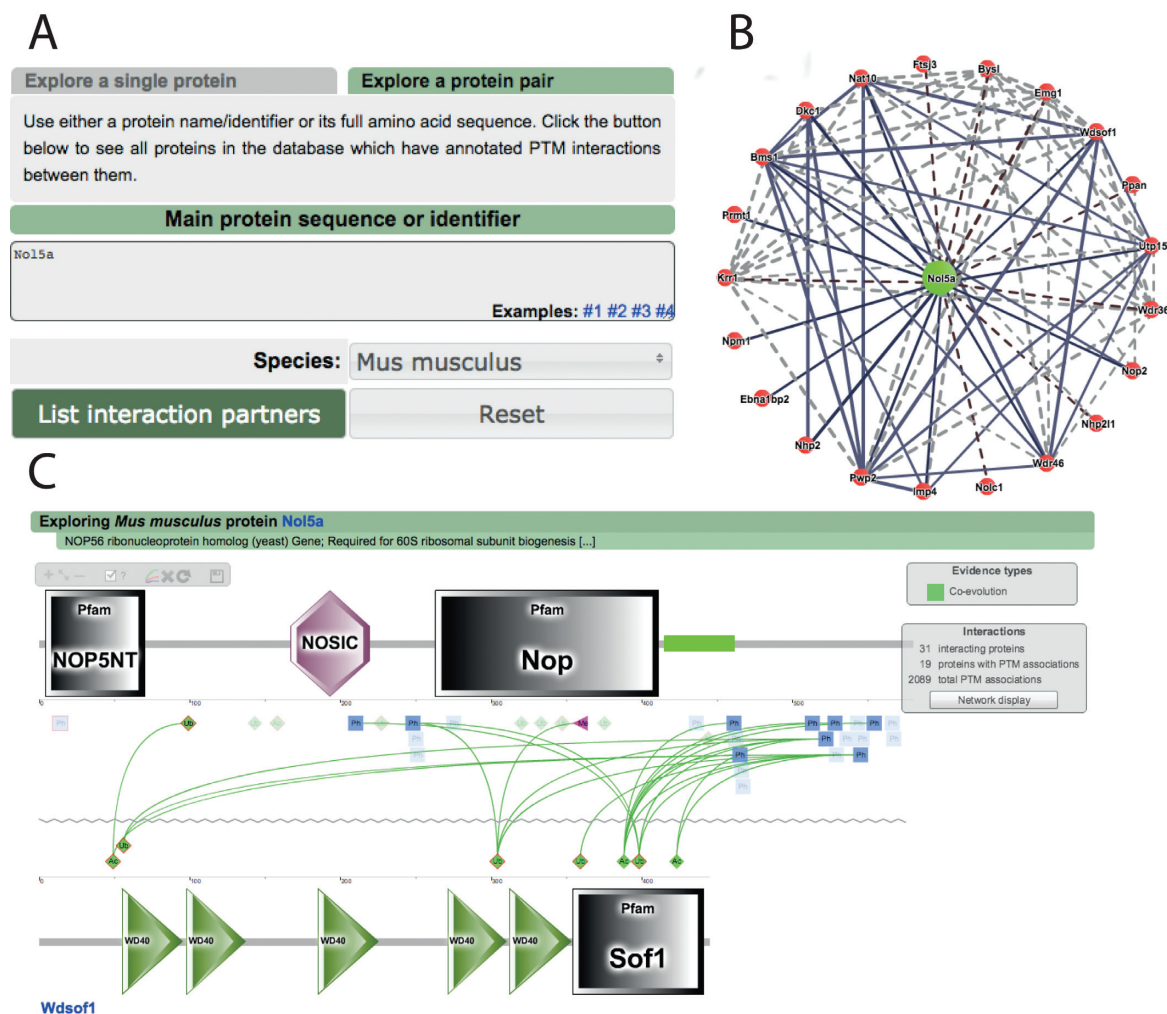
## CONCLUSIONS

The PTMcode database is imbued by systems biology philosophy in the sense of being a resource that aims to provide a global picture of the post-translational regulatory landscape of eukaryotic proteins. We offer a unique environment where PTMs are shown as active players into the whole functional context of the protein, from its domain architecture to its interaction network neighborhood. Within this framework, in addition to the extra value given to the data taken from other resources, the main contribution of PTMcode to the field is to supply known and predicted functional associations between PTMs. Those associations are extracted using in-house implemented methodologies that measure (i) the co-evolution between two protein residues and (ii) different types of PTMs proximity (close in space or part of highly modified protein regions).

In this second release we have tripled the size of our PTM collection and implemented a simple algorithm for the propagation of the modifications through orthologs making available for the first time a predicted post-translational regulatory scheme for thousands of proteins that had no information available.

Another major update is the analysis of associated PTM pairs in interacting proteins which may contain tips for the regulation of PPIs. This places PTMcode as a bridge of information between databases of PPIs (34,51) and those dedicated to collect and analyze PTMs (21,22).

As the chosen name for the database suggests our final aim is to contribute to the understanding of the so-called PTM code, a molecular barcode composed by function-specific combinations of PTMs. This challenge entails significant limitations. First, we are far from having a complete PTM repertoire even for a single organism and many PTM



**Figure 4.** Visualization of functional associations between PTMs in interacting proteins. Users can explore the regulation of the interactions where their favorite protein is involved (A). We implemented a network viewer where all its interactions are displayed (B). Every partner in the network opens the 'PPI view', a framework to explore the functional associations found between PTMs from both proteins (C).

types are still vastly understudied so their functional impact might be being underestimated. On top of that, both PTMs and PPIs are indeed dependent on particular cell states while they are normally provided as a collage with no condition specificity information. These and others are the challenges for future studies, the new release for our PTM-code database that we report herein comes to add a step forward into this objective since it produces new knowledge out of the collected data feeding the scientific community with more information and more hypothesis.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

We thank Yan Yuan for all his help and support on all technical and infrastructure issues we encountered during this project.

## FUNDING

EMBL. Funding for open access charge: EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

*Conflict of interest statement.* None declared.

## REFERENCES

- Beltrao, P., Albanese, 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.
- Li, X., Foley, E.A., Kawashima, S.A., Molloy, K.R., Li, Y., Chait, B.T. and Kapoor, T.M. (2013) Examining post-translational modification-mediated protein-protein interactions using a chemical proteomics approach. *Protein Sci.*, **22**, 287–95.
- Khoury, G.A., Baliban, R.C. and Floudas, C.A. (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.*, **1**, 90.
- Beltrao, P., Bork, P., Krogan, N.J. and van Noort, V. (2013) Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.*, **9**, 714.
- Yamauchi, M. and Shiiba, M. (2008) Lysine hydroxylation and cross-linking of collagen. *Methods Mol. Biol.*, **446**, 95–108.



6. Van Roey, K., Dinkel, H., Weatheritt, R.J., Gibson, T.J. and Davey, N.E. (2013) The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.*, **6**, rs7.
7. Hunter, T. (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell*, **28**, 730–738.
8. Latham, J.A. and Dent, S.Y.R. (2007) Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.*, **14**, 1017–1024.
9. Calnan, D.R. and Brunet, A. (2008) The FoxO code. *Oncogene*, **27**, 2276–2288.
10. Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J. and Villén, J. (2013) Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods*, **10**, 676–682.
11. Danielsen, J.M.R., Sylvestersen, K.B., Bekker-Jensen, S., Szklarczyk, D., Poulsen, J.W., Horn, H., Jensen, L.J., Møllgaard, N. and Nielsen, M.L. (2011) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol. Cell. Proteom.*, **10**, doi:10.1074/mcp.M110.003590.
12. Minguez, P., Parca, L., Diella, F., Mende, D., Kumar, R., Helmer-Citterich, M., Gavin, A., van Noort, V. and Bork, P. (2012) Deciphering a global network of functionally associated post-translational modifications. *Mol. Syst. Biol.*, **8**, 599.
13. Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldón, T., Rattei, T., Creevey, C., Kuhn, M. *et al.* (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.*, **42**, D231–D239.
14. Huerta-Cepas, J., Dopazo, J. and Gabaldón, T. (2010) ETE: a python Environment for Tree Exploration. *BMC Bioinform.*, **11**, 24.
15. Marcet-Houben, M. and Gabaldón, T. (2011) treeKO a duplication aware algorithm for the comparison of phylogenetic trees. *Nucleic Acids Res.*, **39**, e66.
16. 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.
17. Yip, K.Y., Patel, P., Kim, P.M., Engelman, D.M., McDermott, D. and Gerstein, M. (2008) An integrated system for studying residue coevolution in proteins. *Bioinformatics*, **24**, 290–292.
18. Simonetti, F.L., Teppa, E., Chernomoretz, A., Nielsen, M. and Buslje, C. (2013) MISTIC: mutual information server to infer coevolution. *Nucleic Acids Res.*, **41**, W8–W14.
19. Korber, B., Farber, R., Wolpert, D. and Lapedes, A.S. (1993) Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl. Acad. Sci.*, **90**, 7176–7180.
20. The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
21. Gnad, F., Gunawardena, J. and Mann, M. (2010) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.*, **39**, D253–D260.
22. 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.
23. Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J. and Diella, F. (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, D261–D267.
24. Lu, C.-T., Huang, K.-Y., Su, M.-G., Lee, T.-Y., Bretaña, N.A., Chang, W.-C., Chen, Y.-J., Chen, Y.-J. and Huang, H.-D. (2012) dbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. *Nucleic Acids Res.*, **41**, D295–D305.
25. 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.
26. 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.
27. Henriksen, P., Wagner, S.A., Weinert, B.T., Sharma, S., Bacinskaja, G., Rehman, M., Juffer, A.H., Walther, T.C., Lisby, M. and Choudhary, C. (2012) Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics*, **11**, 1510–1522.
28. Lundby, A., Secher, A., Lage, K., Nordsborg, N.B., Dmytriiev, A., Lundby, C. and Olsen, J.V. (2012) Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat. Commun.*, **3**, 876.
29. Matic, I., Schimmel, J., Hendriks, I.A., van Santen, M.A., van de Rijke, F., van Dam, H., Gnad, F., Mann, M. and Vertegaal, A.C.O. (2010) Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif. *Mol. Cell*, **39**, 641–652.
30. Murray, C.I., Kane, L.A., Uhrigshardt, H., Wang, S.-B. and Van Eyk, J.E. (2011) Site-mapping of in vitro S-nitrosation in cardiac mitochondria: implications for cardioprotection. *Mol. Cell. Proteomics*, **10**, doi:10.1074/mcp.M110.004721.
31. 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**, doi:10.1074/mcp.M111.013284.
32. 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–ra48.
33. Zielinska, D.F., Gnad, F., Schropp, K., Wiśniewski, J.R. and Mann, M. (2012) Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell*, **46**, 542–548.
34. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.
35. Tan, C.S.H., 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.
36. Letunic, I., Doerks, T. and Bork, P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.*, **40**, D302–D305.
37. Choudhary, C. and Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.*, **11**, 427–439.
38. 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.
39. Landry, C.R., Levy, E.D. and Michnick, S.W. (2009) Weak functional constraints on phosphoproteomes. *Trends Genet.*, **25**, 193–197.
40. Chen, S.C.-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.
41. Venne, A.S., Kollipara, L. and Zahedi, R.P. (2014) The next level of complexity: crosstalk of posttranslational modifications. *Proteomics*, **14**, 513–524.
42. Lyons, N.A., Fonslow, B.R., Diedrich, J.K., Yates, J.R. and Morgan, D.O. (2013) Sequential primed kinases create a damage-responsive phosphodegron on Eco1. *Nat. Struct. Mol. Biol.*, **20**, 194–201.
43. Nussinov, R., Tsai, C.-J., Xin, F. and Radivojac, P. (2012) Allosteric post-translational modification codes. *Trends Biochem. Sci.*, **37**, 447–455.
44. Minguez, P., Letunic, I., Parca, L. and Bork, P. (2013) PTMcode: a database of known and predicted functional associations between post-translational modifications in proteins. *Nucleic Acids Res.*, **41**, D306–D311.
45. 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.
46. Woodsmith, J., Kamburov, A. and Stelzl, U. (2013) Dual coordination of post translational modifications in Human Protein Networks. *PLoS Comput. Biol.*, **9**, e1002933.



47. Yachie,N., Saito,R., Sugiyama,N., Tomita,M. and Ishihama,Y. (2011) Integrative features of the yeast phosphoproteome and protein–protein interaction map. *PLoS Comput. Biol.*, **7**, e1001064.
48. Nussinov,R., Ma,B., Tsai,C.-J. and Csermely,P. (2013) Allosteric conformational barcodes direct signaling in the cell. *Structure*, **21**, 1509–1521.
49. Blohm,P., Frishman,G., Smialowski,P., Goebels,F., Wachinger,B., Ruepp,A. and Frishman,D. (2014) Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res.*, **42**, D396–D400.
50. Berman,H.M., Kleywegt,G.J., Nakamura,H. and Markley,J.L. (2012) The future of the protein data bank. *Biopolymers*, **99**, 218–222.
51. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.