A tale of two communities: the Scop3PTM system to provide context to protein PTMs

Pathmanaban Ramasamy [1,2,3,4], *Demet Turan* [1,2], *Natalia Tichshenko* [1,2], *Niels Hulstaert* [1,2], *Elien Vandermarliere* [1,2], *Wim Vranken* [3,4], *Lennart Martens* [1,2]

[1] VIB-UGent Center for Medical Biotechnology, VIB, Ghent 9000, Belgium
[2] Department of Biomolecular Medicine, Faculty of Health Sciences and Medicine, Ghent University, Ghent 9000, Belgium
[3] Interuniversity Institute of Bioinformatics in Brussels, ULB-VUB, 1050 Brussels, Belgium
[4] Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

pathmanaban.ramasamy@ugent.be
pathmanaban.ramasamy@vub.ac.be

Background

Post-translational modifications (PTMs) of proteins play an important role in various cellular processes. One of the key PTM is phosphorylation, which has already been studied extensively. With the advancements in high-throughput mass spectrometry (MS/MS) techniques, the amount of publicly available data on phosphorylation has increased dramatically over time. However, available resources on phosphorylation usually contain only sequence and phosphosite information, generally omitting structural information. Yet such structural information is particularly relevant in a crucial task: to differentiate between functional and non-functional phosphosites. We therefore developed Scop3P[1]: a database of public proteomics data-derived human phosphosites that are annotated with detailed, residue-level structural annotation based on state-of-the-art prediction tools. Moreover, these phosphosites are also directly mapped onto 3D protein structures when available in PDB.

Methods

Information about available human phosphosites were obtained from UniProtKB/Swiss-Prot. Next we re-processed public phopshoproteomics data from PRIDE. Phosphosites were then mapped onto their protein structure when available in the PDB using the SIFTS mapping. For each human protein - even when no structure is available - backbone dynamics, disordered propensity and early folding properties were predicted using DynaMine, DisOmine and EFoldMine respectively. Secondary structures were obtained from DSSP and solvent accessibility from PDBePISA. Human amino acid variations were retrieved from the Humsavar dataset from Swiss-Prot. Evolutionary conservation information for the phosphosites with structures were retrieved from AMINODE (Figure 1). Visualization of the protein structure is performed by NGL Viewer. Scop3P uses a relational database as data repository.

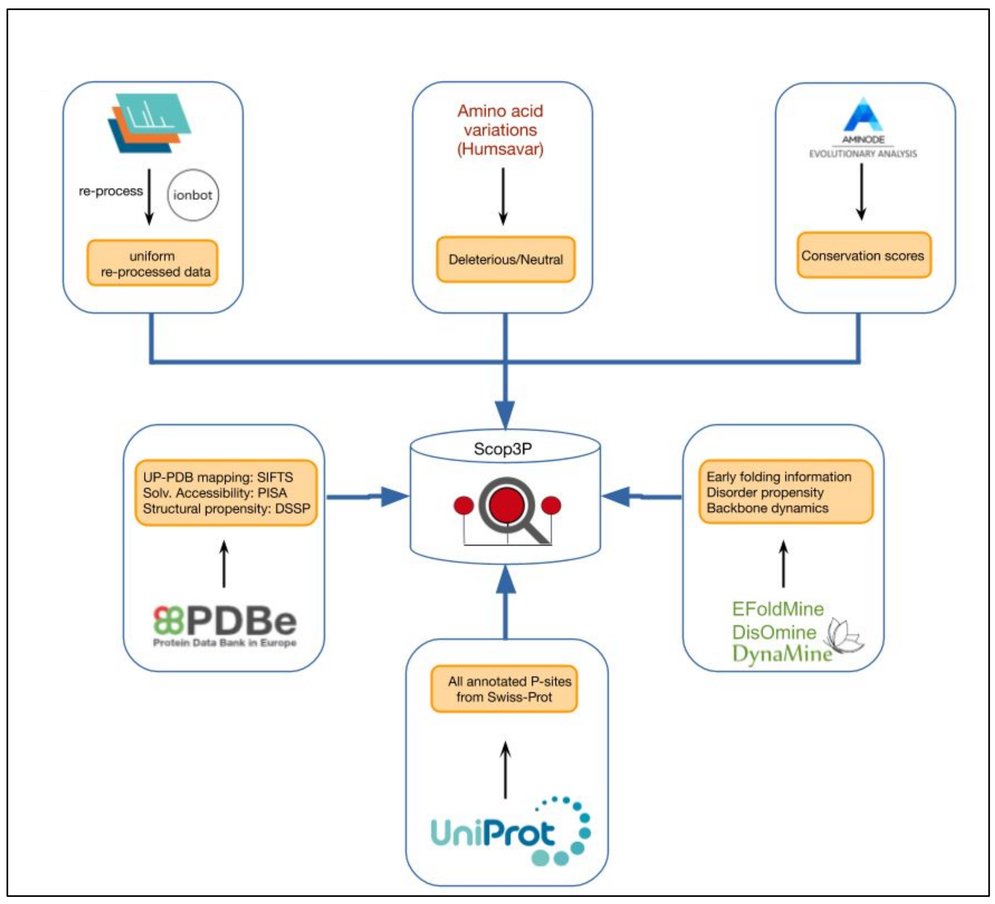


Figure 1: Different data sources integrated into Scop3P

Database content

In total, our database contains 108,130 P-sites from 14,261 phosphoproteins (Human proteome contains 20,408 proteins) that contain 82,116 unique phospho instances (P-sites) from Swiss-Prot (14,019) and PRIDE (68,097). 10,044 of these unique phospho instances were mapped onto 21,937 different PDB structures in which 4,399 P-sites has more than one structure in Scop3P. Every P-site is also annotated with frequency of phosphorylation as observed in different phospho proteomics experiments.

Web interface: https://iomics.ugent.be/scop3p

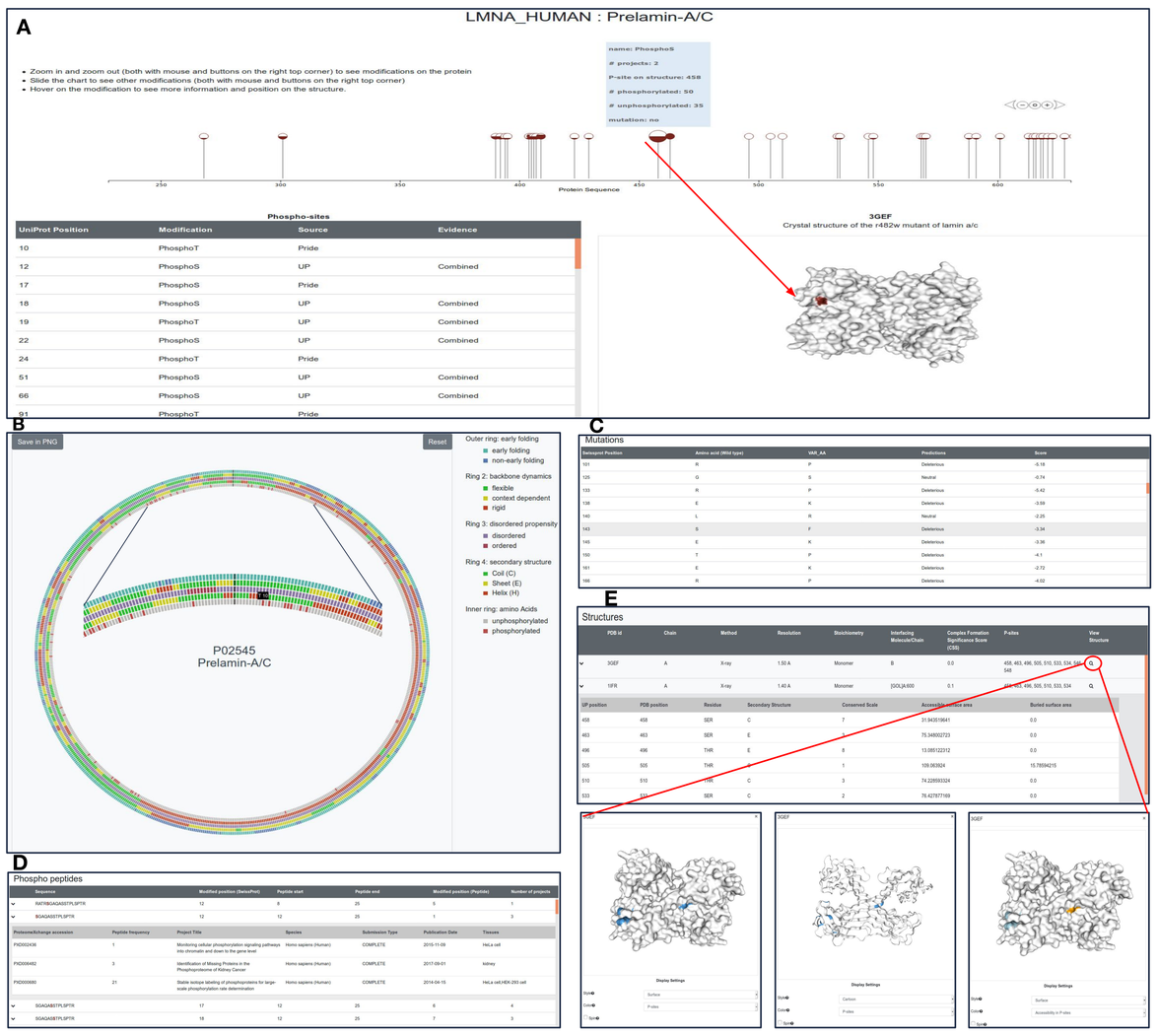
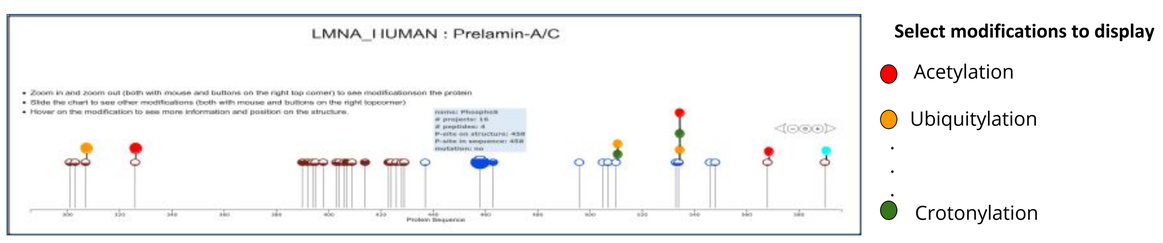


Figure 2: For every phosphoprotein, Scop3P provides a quick overview (A), residue level biophysical predictions (B), effect of amino acid variations (C), all identified phospho peptides from different experiments (D), and P-sites mapped onto one or more protein structures (E).

Scop3P is being extended to Scop3PTM



Scop3P	Scop3PTM
Projects: **36**	Projects: **539**
Processed spectra: **60.2 million**	Processed spectra: **~1 billion**
Modifications: **1 (Phospho)**	Modifications: **117 + artefacts**

Conclusion

- Scop3PTM integrates information from different knowledge bases and shows how re-analysis of large scale public proteomics data sets can add an additional level of significance and confidence to the PTM-sites.
- Scop3PTM system will provide a unique and powerful resource to understand the impact of PTM-sites on human protein structure-function relationship.
- Early folding, disordered propensity and backbone dynamics data will provide valuable added information for researchers seeking to understand whether any PTM sites or related residues in close proximity are crucial for protein folding and stability.

Reference: [1] https://doi.org/10.1021/acs.jproteome.0c00306

VIB-UGENT CENTER FOR MEDICAL BIOTECHNOLOGY · GHENT UNIVERSITY · VRIJE UNIVERSITEIT BRUSSEL · Compomics · B2 · SCIENCE MEETS LIFE