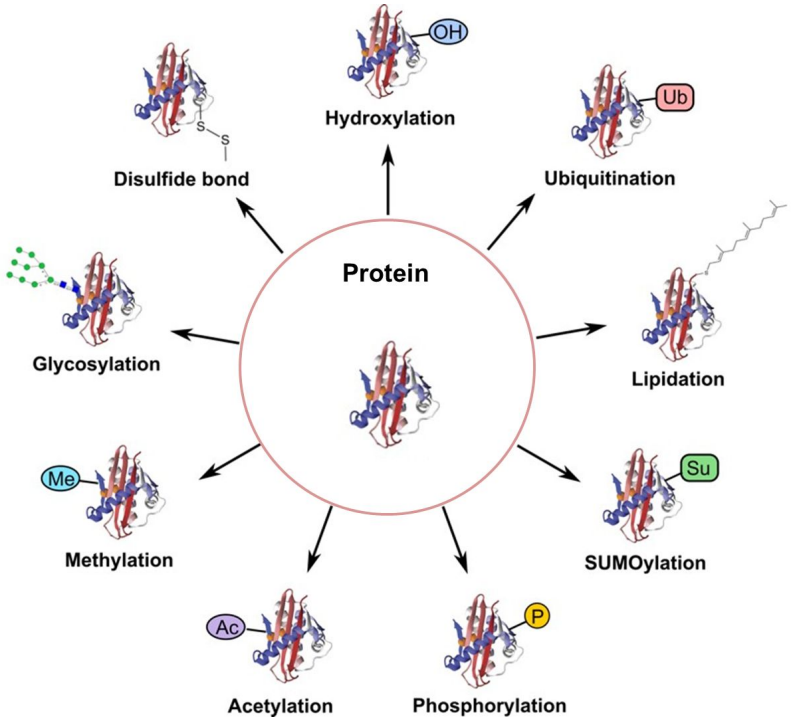
Post Translational Modification Prediction

Techniques and implementation

Why do PTMs matter?

- Allow for a greater range of functionality
- Play a role in
  - Localization in the cell
  - Stability (folding and degrading)
  - Enzymatic activity
- Particularly important in signalling pathways, relevant to cancer therapeutics
- Increase the complexity of the proteome
  - Act as an additional layer on top of sequence



What we need in PTM prediction

- Because our goal is to incorporate this function into our existing programs (EvoDesign, QUARK, etc), it should not hamper the function or accuracy of the other programs, meaning we need
  - Reliability
    - High accuracy of predictions
  - Versatility
    - Able to identify a wide variety of PTMs
  - Speed
    - Quick processing as to not hold up the pipeline

Computational and Structural Biotechnology Journal 15 (2017) 307–319

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Protein post-translational modifications: *In silico* prediction tools and molecular modeling

Martina Audagnotto *, Matteo Dal Peraro *

Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 4 December 2016
Received in revised form 17 March 2017
Accepted 21 March 2017
Available online 31 March 2017

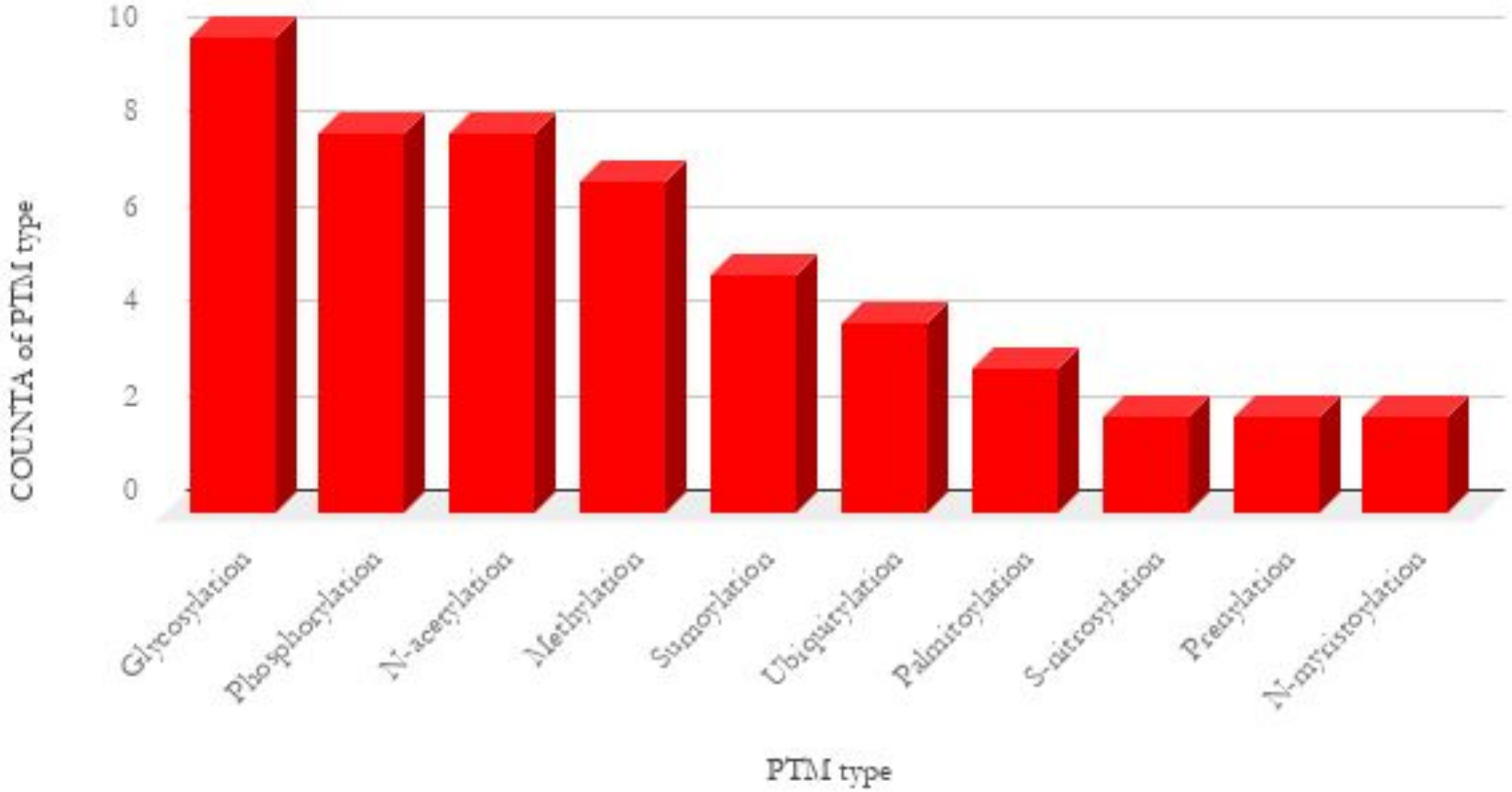
ABSTRACT

Post-translational modifications (PTMs) occur in almost all proteins and play an important role in numerous biological processes by significantly affecting proteins' structure and dynamics. Several computational approaches have been developed to study PTMs (*e.g.*, phosphorylation, sumoylation or palmitoylation) showing the importance of these techniques in predicting modified sites that can be further investigated with experimental approaches. In this review, we summarize some of the available online platforms and their contribution in the study of PTMs. Moreover, we discuss the emerging capabilities of molecular modeling and simulation that are able to complement these bioinformatics methods, providing deeper molecular insights into the biological function of post-translational modified proteins.

Overview of existing webservers

- Methods
  - Support vector machine
    - A supervised learning technique that constructs hyperplanes through data to categorize new examples
  - Random forest
    - Aggregates the results of multiple decision trees, each made with a subset of the overall training data to reduce noise
  - Neuronal network
- Variety of techniques including
  - Physicochemical properties
  - Evolutionary data
  - Protein-protein interaction data
  - Identification of motifs
  - Hydrophobicity

Number of webservers by PTM type



Number of PTM prediction webservers released by year

RSC Publishing MOLECULAR OMICS
Article | Journal | Search | Submit

Mol Omics. 2018 Jun 1; 14(3): 197–209.
Published online 2018 Jun 7. doi: 10.1039/c8mo00027a
PMCID: PMC6115748
PMID: 29876573

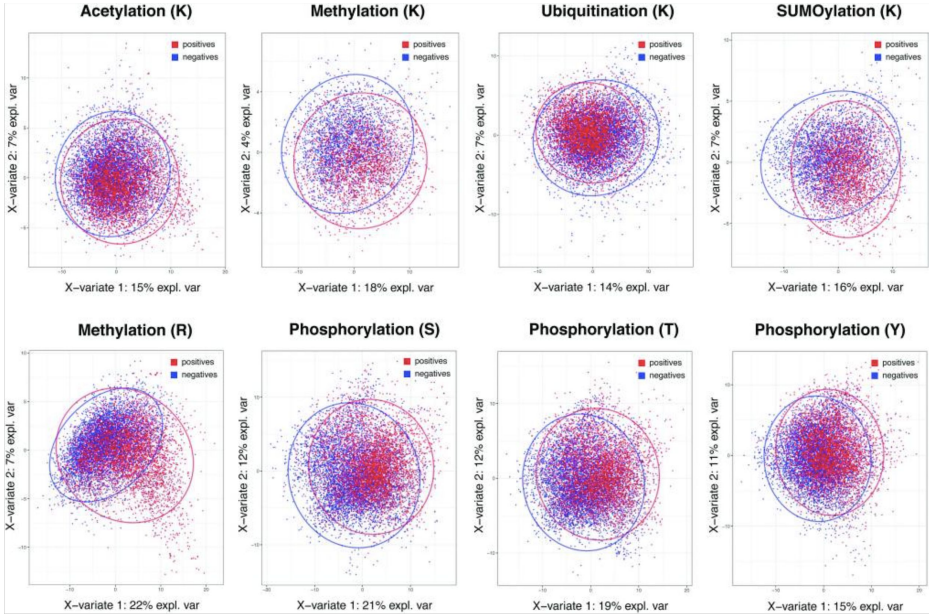
PTMscape: an open source tool to predict generic post-translational modifications and map modification crosstalk in protein domains and biological processes‡

Ginny X. H. Li,[a] Christine Vogel,[b] and Hyungwon Choi[a,c]

▸ Author information ▸ Article notes ▸ Copyright and License information Disclaimer

PTMscape - an R package for PTM annotation

- Uses a linear SVM model to predict PTM sites
  - Unlike other programs, has support for a wide range of PTM
- Also allows for positive and negative crosstalk (ie interference) analysis
- However, requires training data be input on every run
  - Without this data, the package isn't very useful
- Also, not very accurate



D298–D308 *Nucleic Acids Research, 2019, Vol. 47, Database issue* *Published online 10 November 2018*
doi: 10.1093/nar/gky1074

dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications

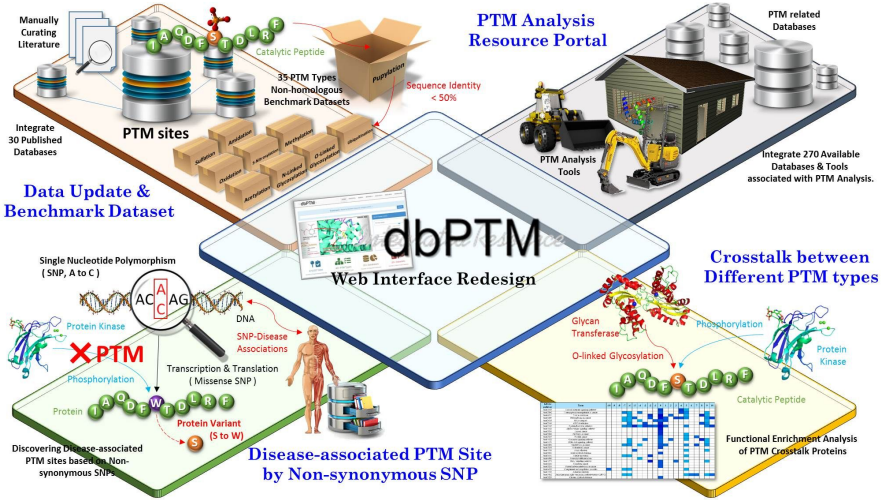
Kai-Yao Huang[1,2,3,†], Tzong-Yi Lee[1,2,3,*,†], Hui-Ju Kao[4], Chen-Tse Ma[4], Chao-Chun Lee[4], Tsai-Hsuan Lin[4], Wen-Chi Chang[5] and Hsien-Da Huang[1,2,3,*]

[1]Warshel Institute for Computational Biology, The Chinese University of Hong Kong, Shenzhen 518172, China, [2]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China, [3]School of Life and Health Science, The Chinese University of Hong Kong, Shenzhen 518172, China, [4]Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 32003, Taiwan and [5]Institute of Tropical Plant Sciences, College of Biosciences and Biotechnology, National Cheng Kung University, Tainan 70101, Taiwan

Received September 15, 2018; Revised October 18, 2018; Editorial Decision October 18, 2018; Accepted October 19, 2018

dbPTM - a database of PTM sites

- Has been active for the last 10 years
- Has data aggregated from other databases, papers, etc
- Holds 900,000+ experimental PTM sites of over 130 types of PTM
- Uses a standardized format across the entire data set



How do we incorporate this data?

- Treat residues with PTMs as unique amino acids
  - Would require
    - A new, FASTA-like format that allows for the notation of PTMs
    - Development of new sequence alignment tools
    - New physics and knowledge based force fields
    - Use of eukaryotic cells for experimental verification
- Avoid use of regions that would contain PTMs
  - Increases likelihood of similar folding in pro- and eukaryotes
  - Would either
    - Reduce functionality of designs or
    - Require research into ways to mimic PTM function with the current set of residues

Next steps

- Compare overall reliability of different methods to determine which to use
- Begin to develop code to implement this
- Apply feedback/suggestions to next year’s UROP program
- Come to the UROP symposium on Wednesday!