MedCrave
Step into the World of Research

Journal of Nanomedicine Research

Review Article

Open Access

Computational prediction of proteins sumoylation: a review on the methods and databases

Abstract

Protein is one of the biological macromolecules, which plays vital roles in the cell. There are numerous post-transitional modifications (PTMs) that strongly affect proteins and their functionality. A PTM occurs when a chemical functional group is being added on or removed from a specific amino acid. A PTM may consist of both enzymatic and non-enzymatic changes. Recent studies have introduced more than 500 different PTM types. SUMOylation is one of the most important PTM; disruption in SUMOylation process affects the cell function and one of the consequences of this change is cell morphology disorder and leads to a variety of sever diseases such as Alzheimer's disease and Parkinson's disease. In this paper we have reviewed the current state-of-the-art in silico methods to predict SUMOylation as well as related databases.

Keywords: Post-transitional modification, PTM, SUMOylation, Predicted, Databases, Bioinformatics, Algorithms

Volume 3 Issue 5 - 2016

Shahin Ramazi,[1] Javad Zahiri,[1] Seyed Shahriar Arab,[2] Yasaman Parandian[1]

[1]Department of Biophysics, Bioinformatics and Computational Omics Lab (BioCOOL), Iran

[2]Department of Biophysics, Tarbiat Modares University, Iran

Correspondence: Javad Zahiri, Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran Email zahiri@madares.ac,ir

Received: March 04, 2016 | **Published:** May 20, 2016

Introduction

Being happened in the nucleus, cytoplasm and organelles of cells; PTM is considered as one of the most important processes in protein functionality .[1] Generally, mass spectrometry data are applied to identify the PTMs and their related sites .[2] However, the experimental data of PTMs are very limited due to the sophistication and high expenses of the experiments. Recently, the practice of applying the computational methods to predict the protein PTM has interested many researchers .[3]

Post-translational modification by the Small Ubiquitin-like Modifier (SUMO) proteins, a process termed SUMOylation, is involved in many fundamental cellular processes. SUMOylation is a eukaryotic post-translational modification, which consists of a reversible attachment of members of the Small Ubiquitin-like Modifier (SUMO) protein family on a protein substrate resulting in the dynamic regulation of its biochemical properties .[4] Proteins involved in many fundamental cellular processes like DNA repair, transcriptional control, chromatin organization, macromolecular assembly and signal transduction are SUMOylated.[5] SUMOylation is also discovered to be involved in various diseases and disorders especially neural ones, such as Alzheimer's disease and Parkinson's disease.[6-8]

The size of SUMO proteins is almost 10 kDa and three-dimensional structures of these proteins are similar to ubiquitin proteins. Interestingly, SUMO proteins have different distribution of surface charge and have less than 20% amino acid sequence similarity.[10] During SUMOylation, a SUMO protein is attached to the target protein, which has an acceptor lysine, through three enzymes and then makes the modification. Finally, SUMO is detached by sumo-specific protease.[4] (Figure 2). SUMO proteins have been discovered in a wide range of eukaryotic organisms.[9] SUMO family has four isoforms in human, one isoform in yeast and eight isoforms in plants .[10,11] However, in most vertebrates family SUMO has three isoforms that are known as SUMO1 namely (sentrin، PIC1، GMP1، Ubl1، Smt3c) and SUMO2 namely (sentrin-2, Smt3b) and SUMO3 namely (sentrin-3, Smt3a).[11-15] Figure 3 shows the number of papers that have reported experimentally verified SUMOylation in different years, these information are based on the PubMed IDs reported by dbPTM .[16] recently published database about PTM data.

(A) 2MW5: Solution Structure of Human Small Ubiquitin like Modifier protein-1 (SUMO-1) in Homo sapiens. (B)1L2N: Smt3 Solution Structure in Saccharomyces cerevisiae. (C) 1WZ0: Solution Structure of Human SUMO-2 (SMT3B), a Ubiquitin-like Protein in Homo sapiens. (D) 11U4A: Solution Structure of Human Small Ubiquitin like Modifier protein-3 (SUMO-3 C47S)) in Homo sapiens.

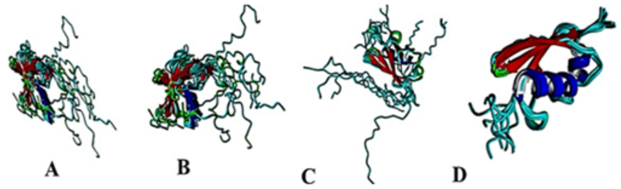

Figure 1 Schematic drawing of the three Sumo proteins using software YASARA.

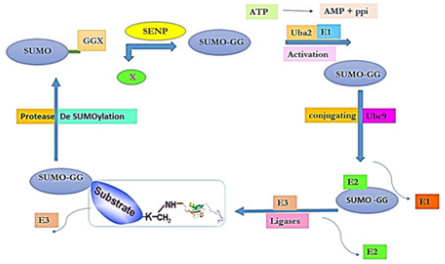

Figure 2 Schematic drawing of path SUMOylation protein.

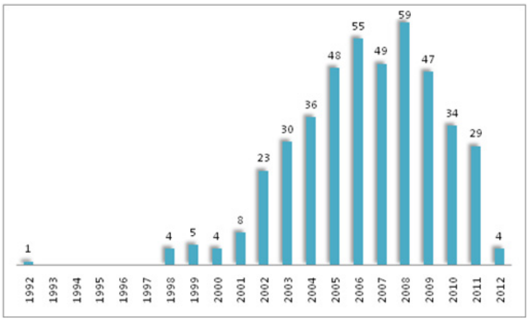

Figure 3 Schematic drawing of the number of SUMOylation articles in different years.

The SUMOylation prediction

Almost computational methods for SUMOylation prediction use sequence information in the neighborhood of the lysine amino acid. Specially considering sequence motifs that are recognized by SUMO Although there are many lysine residues in a protein, but few of these residues in certain motifs are SUMOylation site.[17] Many SUMOylation sites contain a consensus sequence motif of WKXE, in which W represents aliphatic amino acids such as I, V, L, A, P or M; X represents each amino acid and E represents glutamic acid. However, the experimental data analysis demonstrates that nearly 23% of SUMOylation sites do not follow SUMOylation of this consensus motif [18,19]. In addition to the consensus motif, other SUMOylation motifs are reported such as SUMOylation negatively charged amino acid motif (NDSM: WKXE (D/E), SUMOylation dependent phosphorylation motif (PDSM: WKXEXXSP) and SUMO-Style motif (WKXEP) [20,21] Generally, a computational method uses appropriate features of the potential SUMOylation sites considering the experimentally validated data to train a model for SUMOylation prediction. So, thee availability of the valid databases is crucial to construct accurate models. Then main databases for SUMOylation have been reviewed in Table 1 & 2.

Table 1 Information on database

Database	A Brief Description
Swiss-Prot/Uniprot database	This database is one of the largest experimental sources for a variety of post-translational modifications of proteins. (www.ebi.ac.uk/uniprot/)
Tr EMBL database	This database contains tools and extensive educational tools for both researchers and scholars. It also provides data about the different types of proteins PTMs and their related changes.
DbPTM database	This DB provides data on post-translational modifications of proteins. Using this database, protein-protein interactions and their specific protein binding positions with the domain could be identified. (http://dbPTM.mbc.nctu.edu.tw/)
HPRD database	This database is considered as a reference database. It contains data about the human proteins as well as protein PTM data. (http://www.ebi.ac.uk/RESID/)
Phospho site plus database	Currently, the database contains information on a variety of protein post-translational modifications such as acetylation, methylation, SUMOylation and O- glycosylation. (http://www.phosphorylation.biochem.vt.edu/)

Table 2 Information on predictions SUMOylation articles

Predictor	Training Feature	AC	SN	SP	MCC	AUC
SUMO plot .[22]	Only sequence	90%	80%	93%	48%	-
SUMOSP1.[19]	Only sequence	92.71%	83.60%	93.08%	50.12%	73%
Boshuliuetal .[23]	Sequence and physicochemical properties	89.18%	_	_	_	_
Find SUMO .[24]	Only sequence	87.40%	86.40%	87.50%		_
SUMO pre.[10]	Only sequence	97.71%	73.96%	97.67%	63.64%	_
SUMOSP2 .[25]	Sequence and physicochemical properties	_	96.69%	88.17%	_	_
SUMO tr .[26]	Sequence and 3D structure and hydrophobicity	85%	95%	75%	68%	85%
See SUMO.[21]	Sequence and physicochemical properties	97.68%	67.57%	99.79%	67.86%	92%
SUMO hydro .[27]	Sequence and physicochemical properties	58.30%	94.40%	93.30%	41.90%	-
SUMO hunt.[28]	Only sequence	85%	95%	75%	68%	-
GPS – SUMO.[7]	Only sequence Hydrophobicity	-	-	-	-	-

A review of articles from 2004 to 2015, predictions SUMOylation.

Assessing the Performance of the Sumoylation Prediction Methods

There are different assessment measures based the four following basic parameters:

A. "True positive" (TP): the experimentally validated SUMOylation sites that have been correctly predicted by the prediction method.

B. "True negative" (TN): the non-SUMOylation sites that have been correctly predicted.

C. "False positive" (FP): the non-SUMOylation sites that have been incorrectly predicted as SUMOylation sites.

D. "False negative" (FN): the experimentally validated SUMOylation sites that have been incorrectly predicted non-SUMOylation sites.

The most important assessment measures based on the above-mentioned parameters have been described in the following.

a. **Sensitivity:** sensitivity indicates the percentage ofSUMOylation sites that have been predicted correctly.

$$Sensitivity = \frac{TP}{TP+FN} \quad \text{(Formula 1-1)}$$

b. **Specificity:** shows the percentage of non-SUMOylation sites that have been predicted correctly as non-SUMOylation.

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad \text{(Formula 1-2)}$$

c. **Accuracy:** The percentage of the correct prediction.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \quad \text{(Formula 1-3)}$$

d. **Matthews correlation coefficient:** Matthews correlation coefficient can be calculated using the formula defines.[2]

$$MCC = \frac{(TP)(TN)-(FP)(FN)}{\sqrt{[TP+FP][TP+FN][TN+FP][TN+FN]}} \quad \text{(Formula 1-4)}$$

e. **Area Under Curve**: (AUC or "Area Under Curve"), is another measure of classification accuracy, the closer the AUC to one the more accurate the classification.

Conclusion

SUMOylation is one the most important PTM type, which a disruption in this PTM can lead to various diseases such as type 1 diabetes, Parkinson's disease, Alzheimer's disease, heart disease, cancer and brain failure. Considering the cost and limitations of the experimental methods, in the recent years, many studies devoted to computational detection of SUMOylation. In this paper, the databases of experimentally verified SUMOylation and the computational methods for the prediction of SUMOylation have been reviewed. While there are promising methods for the SUMOylation prediction, but considering the limited experimental SUMOylation data, there are considerable rooms to improve the SUMOylation prediction tools.

Acknowledgments

None.

Conflicts of interest

None.

References

1. Nickchi P, Jafari M, Kalantari S PEIMAN 1.0: Post–translational modification Enrichment, Integration and Matching ANalysis. *Database 2015: bav037.* 2015

2. Lothrop AP, Torres MP, Fuchs SM Deciphering post–translational modification codes. *FEBS letters.* 2013;587(8):1247–1257.

3. Kamath KS, Vasavada MS, Srivastava S Proteomic databases and tools to decipher post–translational modifications. *J proteomics.* 2011;75(1):127–144.

4. Song J, Durrin LK, Wilkinson TA et al. Identification of a SUMO–binding motif that recognizes SUMO–modified proteins. *Proc Natl Acad Sci.* 2004;101(40):14373–14378.5.Lomelí H, Vázquez M Emerging roles of the SUMO pathway in development. *Cell Mol Life Sci.* 2011;68(24):4045–4064.5.

5. Lu L, Shi XH, Li SJ, Xie ZQ, Feng YL, et al. Protein sumoylation sites prediction based on two–stage feature selection. *Mol Divers.* 2010;14(1):81–86.6.

6. Zhao Q, Xie Y, Zheng Y et al. GPS–SUMO: a tool for the prediction of sumoylation sites and SUMO–interaction motifs. *Nucleic Acids Res.* 2014;24:325–330.7.

7. Müller S, Ledl A, Schmidt D SUMO: a regulator of gene expression and genome integrity. *Oncogene.* 2004;23(11):1998–2008.8.

8. Seeler JS, Dejean A SUMO: of branched proteins and nuclear bodies. *Oncogene.* 2001;20(49):7243–7247.

9. Xu J, He Y, Qiang B et al. A novel method for high accuracy sumoylation site prediction from protein sequences. *BMC bioinformatics.* 2008;9:1.

10. Hong Y, Rogers R, Matunis MJ et al. Regulation of heat shock transcription factor 1 by stress–induced SUMO–1 modification. *J Biol Chem.* 2001;276(43):40263–40267.

11. Saitoh H, Hinchey J Functional heterogeneity of small ubiquitin–related protein modifiers SUMO–1 versus SUMO–2/3. *J Biol Chem.* 2000;275(9):6252–6258.

12. Tatham MH, Kim S, Yu B, Jaffray E et al. Role of an N–terminal site of Ubc9 in SUMO–1,–2, and–3 binding and conjugation. *Biochemistry.* 2003;42(33):9959–9969.

13. Guo D, Li M, Zhang Y et al. A functional variant of SUMO4, a new IκBα modifier, is associated with type 1 diabetes. *Nat Genet.* 2004;36(8):837–841.

14. Melchior F SUMO–nonclassical ubiquitin. *Annual review of cell and developmental biology.* 2000;16(1):591–626.

15. Huang KY, Su MG, Kao HJ et al. dbPTM 2016: 10–year anniversary of a resource for post–translational modification of proteins. *Nucleic Acids Res.* 2016;44(D1):D435–D446.

16. Hay RT SUMO: a history of modification. *Mol Cell.* 2005;18(1):1–12.

17. Mann M, Jensen ON Proteomic analysis of post–translational modifications. *Nat Biotechnol.* 2003;21(3):255–261.

18. Xue Y, Zhou F, Fu C et al. SUMOsp: a web server for sumoylation siteprediction. *Nucleic Acids Res.* 2006;34(suppl 2):W254–W257.

19. Martin S, Wilkinson KA, Nishimune A et al. Emerging extranuclear roles of protein SUMOylation in neuronal function and dysfunction. *Nat Rev Neurosci.* 2007;8(12):948–959.

20. Teng S, Luo H, Wang L Predicting protein sumoylation sites from sequence features. *Amino Acids.* 2012;43(1):447–455.

21. Yu Xue, Fengfeng Z, Hualei L et al. SUMO Substrates and Sites Prediction Combining Pattern Recognition and Phylogenetic Conservation. *arXiv preprint q–bio/040901*. 2004

22. Liu B, Li S, Wang Y et al. Predicting the protein SUMO modification sites based on Properties Sequential Forward Selection (PSFS). *Biochem Biophys Res Commun.* 2007;358(1):136–139.

23. Friedline CJ, Zhang X, Zehner ZE et al. FindSUMO: a PSSM–basedmethod for sumoylation site prediction. Advanced Intelligent Computing Theories and Applications. *With Aspects of Artificial Intelligence pp.* 2008;1004–1011.

24. Ren J, Gao X, Jin C, Zhu M, Wang X Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *Proteomics.* 2009;9(12):3409–3412.

25. Yavuz AS, Sezerman U SUMOtr: SUMOylation site prediction based on 3D structure and hydrophobicity. in *2010 5th Int Symp Heal Informatics Bioinforma.* 2010

26. Chen YZ, Chen Z, Gong YA, Ying G SUMOhydro: a novel method for the prediction of sumoylation sites based on hydrophobic properties. *PLoS One.* 2012;7(6):e39195.

27. Ijaz A SUMOhunt: Combining spatial staging between lysine and SUMO with random forests to predict sumoylation. *ISRN Bioinform.* 2013

28. Zahiri J, Bozorgmehr JH, Masoudi–Nejad A Computational prediction of protein–protein interaction networks: algo–rithms and resources. *Curr Genomics.* 2013;14(6):397–414.