

## PhoglyStruct: Prediction of phosphoglycerylated lysine residues using structural properties of amino acids

Abel Chandra<sup>5, \*, †</sup>, Alok Sharma<sup>1,2, 3,5,9 \*, †</sup>, Abdollah Dehzangi<sup>4</sup>, Shoba Ranganathan<sup>6</sup>, Anjeela Jokhan<sup>7</sup>, Kuo-Chen Chou<sup>8,&</sup>, Tatsuhiko Tsunoda<sup>2, 3, 9,&</sup>

<sup>†</sup>Correspondence: [abelavit@gmail.com](mailto:abelavit@gmail.com); [alok.sharma@griffith.edu.au](mailto:alok.sharma@griffith.edu.au)

<sup>\*</sup>Equal Contributors

<sup>&</sup>Last Authors

<sup>1</sup>Institute for Integrated and Intelligent Systems, Griffith University, Brisbane, QLD-4111, Australia.

<sup>2</sup>Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

<sup>3</sup>Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan

<sup>4</sup>Department of Computer Science, Morgan State University, Baltimore, Maryland, USA

<sup>5</sup>School of Engineering and Physics, Faculty of Science, Technology and Environment, University of the South Pacific, Suva, Fiji Islands

<sup>6</sup>Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia

<sup>7</sup>Faculty of Science, Technology and Environment, University of the South Pacific, Suva, Fiji Islands

<sup>8</sup>The Gordon Life Science Institute, Boston, MA 02478, USA

<sup>9</sup>CREST, JST, Tokyo 113-8510, Japan

### Supplement 2: Comparison of PhoglyStruct with simpler set of features

The table below shows the comparison of PhoglyStruct with simpler set of features for the same 10-fold cross-validation set. Simpler\_Features are features for each lysine taking into account the occurrence of amino acids in the  $\pm 2$  residue window. The matrix representing each lysine is of the size 20x5 where 20 is the different types of amino acid in the genome and 5 is the 2 upstream, the lysine and 2 downstream amino acid. Out of the 20 amino acids, a value of 1 is given to the amino acid present at that position otherwise the rest are given a value of 0. Therefore, the feature representing each lysine is a 100-dimensional vector. As it can be seen, PhoglyStruct, which uses the structural properties of amino acids, outperforms this simpler set of features.

Method	Sensitivity	Specificity	G-Mean	Accuracy	MCC	F-Measure	AUC
Simpler_Features	0.6470	<b>0.7490</b>	0.6887	0.7233	0.3612	0.5326	0.6982
PhoglyStruct	<b>0.8726</b>	0.7331	<b>0.7976</b>	<b>0.7678</b>	<b>0.5343</b>	<b>0.6513</b>	<b>0.8034</b>