

PHOSPHORYLATION SITE PREDICTION TOOL FOR CHANNEL PROTEINS USING STATISTICAL MACHINE LEARNING TECHNIQUES

Devasia Arun George¹
deva0009@ntu.edu.sg

Kwoh Chee Keong¹
aschkwoh@ntu.edu.sg

Nikhil Prasanna Jayan¹
nikh0008@ntu.edu.sg

¹ School of Computer Engineering, Nanyang Technological University, Singapore.

Keywords: protein phosphorylation, channel proteins, statistical machine learning, HMM, ANN, SVM.

Abstract

1 Introduction

Protein phosphorylation network acts like a gateway of information to key cellular processes. Most of the phosphorylation processes are reversible and highly kinase-specific in nature. Besides phosphorylation reactions also exhibit high substrate specificity. Hence identifying the kinase-specific phosphorylation sites in proteins will enable better understanding of molecular mechanisms. Phosphorylation in channel proteins modifies their functional properties and thereby regulates all downstream signaling processes. However they have not been well characterized in channel proteins. The authors believe that prediction of phosphorylation sites in channel proteins will enable better understanding of their diverse signal transduction pathways.

2 Method and Results

The authors are in progress of creating a tool that predicts the protein phosphorylation process in the channel proteins. The prediction tool incorporates secondary features of the channel proteins such as structural and conformation details, hydrophobicity, and the evolutionary relationship. Datasets were created from protein kinase A (PKA) and protein kinase C (PKC) specific phosphorylation sites in channel proteins. These datasets were trained using different statistical method such as HMM, ANN and SVM. A combined architecture is made using these well trained models which has the capability to predict the phosphorylation sites much efficiently and accurately.

3 Discussions

Though this prediction architecture is on channel proteins and only two kinases, authors believe that their predictions will generalize well for other protein families and protein kinases as well due to the close familial relationship.

References

1. Diella,F., Gould,C.M., Chica,C., Via,A. and Gibson,T. J. (2007) *Phospho.ELM: a database of phosphorylation sites—update 2008*. NucleicAcids Research, 1–5.
2. Huang,H.D., Lee,T.Y., Tzeng,S.W. and Horng,J.T. (2005) *KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites*. Nucleic Acids Research, 2005, Vol. 33, W226–W229.
3. Huang,H.D., Lee,T.Y., Tzeng,S.W. and Horng,J.T., Tsou,A.P., Huang,K.T. (2005) *Incorporating hidden Markov models for identifying protein kinase-specific phosphorylation sites*. Journal of Computational Chemistry, Vol. 26, 1032 – 1041.
4. Kim,J.H., Lee,J., Oh,B., Kimm,K., and Koh,I. (2004) *Prediction of phosphorylation sites using SVMs*. Bioinformatics, Vol. 20, 3179–3184.
5. Kreegipuu,A., Blom,N. and Brunak,S. (1999) *PhosphoBase, a database of phosphorylation sites: release 2.0*. Nucleic Acids Res., Vol. 27, 237–239.
6. Lee,T.Y., Huang,H.D., Hung,J.H., Huang H.Y., Yang,Y.S. and Wang,T.H. (2006) *dbPTM: an information repository of protein post-translational modification*. Nucleic Acids Research, Vol. 34, D622–D627.
7. Plewczyński,D., Tkacz,A., Godzik,A., and Rychlewski,L. (2004) *A support vector machine approach to the identification of phosphorylation sites*. Cellular & Molecular biology letters, Vol. 10, 73 – 89.