

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

Predicting protein phosphorylation sites

ArticleInfo		
ArticleID	:	3556
ArticleDOI	:	10.1186/gb-2000-1-1-reports022
ArticleCitationID	:	reports022
ArticleSequenceNumber	:	47
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2000-1-18 Received : 2000-1-18 OnlineDate : 2000-4-27
ArticleCopyright	:	BioMed Central Ltd2000
ArticleGrants	:	
ArticleContext	:	130591111

Rachel Brem

Abstract

The design and testing of a neural network to predict phosphorylation sites on proteins is described.

Significance and context

A common way of regulating a protein's activity is to add or take away phosphate groups at specified phosphorylation sites. Here, Blom *et al.* design neural networks to predict phosphorylation sites given only a protein's sequence or structure. Up to now, biochemists have predicted phosphorylation sites by a simple sequence comparison of a new protein against known sites. Blom *et al.* argue that a neural network can do better because it can keep track of residue correlations - for example, it might store the fact that among most serine phosphorylation sites, residue 6 is always a proline when residue 10 is an alanine. The new tools correctly identify 50-90% of known phosphorylation sites in their training-set database. The authors also use their networks to make new predictions, which remain to be tested. If the method proves to be reliable and accurate, it could be valuable for predicting the functions of new proteins and may be more sensitive than current sequence comparison methods.

Key results

The authors built one network to predict tyrosine phosphorylation sites, one for serines and one for threonines. They trained each network as follows. First they made a list of all proteins known (from experiment) to be phosphorylated at the relevant residue. For each protein, they identified the peptide of between nine and eleven residues that included, for example, the phosphotyrosine; these peptides served as positive controls. Then Blom *et al.* assumed that all other tyrosines in the proteins were not part of phosphorylation sites, so the peptides that included these tyrosines were used as negative controls. After the authors had trained the neural networks on groups of such peptides for phosphorylated tyrosine, serine and threonine, the networks predicted 52% of known threonine phosphorylation sites, 86% of known serine sites, and 68% of known tyrosine sites in a test set of data. The authors also used serine networks to predict threonine phosphorylation sites in a test set of data, and correctly identified 81% of the known sites; when they tried the reverse experiment, predicting serine sites using threonine networks, the score was only 54%. They also predicted phosphorylation sites on the transcriptional adaptor p300/CBP, which remain to be tested by experiment. In the last section of the paper, the authors

trained another set of neural networks using predicted three-dimensional structures of phosphopeptides. The results were less accurate than those obtained using the sequences.

Links

Users can make their own predictions with the neural networks at [NetPhos](#). The experimental controls came from [PhosphoBase](#).

Reporter's comments

It is hard to evaluate exactly how good the authors' neural networks are, as we do not know exactly what is in the test set that produced the prediction scores given in the paper. But they seem to be useful tools, especially for experimentalists who are planning to confirm the predictions on new proteins. There is one major unexplained puzzle in this paper: most enzymes that phosphorylate serines also work on threonines, but the authors' serine and threonine networks perform differently on each other's data. Apparently, sequences around threonines in phosphoproteins are quite different from those around serines, as they produced quite different networks. Do these differences mean anything biologically, or is it just random chance?

Table of links

[Journal of Molecular Biology](#)

[NetPhos](#)

[PhosphoBase](#)

References

1. Blom N, Gammeltoft S, Brunak S: Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 1999, 294: 1351-1362. 0022-2836