GPS: A computational protocol for kinase-specific phosphorylation site prediction

Yu Xue
Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Sciences, and the University of Science and Technology of China, Hefei, China 230027

Fengfeng Zhou
Computational Systems Biology Laboratory, Department of Biochemical and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

Ying Xu
Computational Systems Biology Laboratory, Department of Biochemical and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

Xuebiao Yao
Laboratory of Cellular Dynamics, Hefei National Laboratory for Physical Sciences, and the University of Science and Technology of China, Hefei, China 230027

Method Article

Keywords: phosphorylation, post-translational modification (PTM), protein PTM site prediction, GPS, FASTA, Aurora-B, kinase

Posted Date: May 11th, 2007

DOI: https://doi.org/10.1038/nprot.2007.219

Abstract

Introduction

Protein phosphorylation, one of the most ubiquitous post-translational modifications \(PTM), is catalyzed by protein kinases \(PKs). Each PK only modifies a specific set of substrates to ensure signaling fidelity, and defects of PK functions often induce a variety of diseases, including cancers1. Thus, identification of PK-specific phosphorylation sites is essential for molecular delineation of signaling cascade of physiology and potential intervention of pathology of diseases. This motivated the development of the Group-based Phosphorylation Scoring \(GPS) algorithm[1-3]. GPS algorithm is based on the hypothesis that the pattern of phosphorylation sites of a specific PK might be compromised by heterogeneity of multiple structural determinants with different features. In this regard, we partition the known phosphorylation sites of each PK into several groups, and predict the query peptide as the phosphorylation site if it is significantly similar in sequence to the known phosphorylation sites in at least one group. The details of the algorithm can be found in our previously published works [2, 3]. The current version 1.10 of GPS web service provides the prediction of phosphorylation sites for 71 PK groups, including 216 unique PKs, and is freely available at "http://bioinformatics.lcd-ustc.org/gps_web":http://bioinformatics.lcd-ustc.org/gps_web. The following protocol describes how to use GPS to predict the PK-specific phosphorylation sites, and to choose potentially interesting candidates for further consideration.

Reagents

A single protein sequence or multiple protein sequences to be predicted are required in raw or FASTA format.

Equipment

1. A personal computer with a typical operating system \(Windows, Unix/Linux or Apple Mac OSX). 2. A web browser program to connect the GPS web site. The supported programs include Netscape, Firefox, Mozilla or Microsoft Internet Explorer.

Procedure

How to predict the PK-specific phosphorylation sites in protein sequences: 1. Open the prediction page at "http://bioinformatics.lcd-ustc.org/gps_web/predict.php":http://bioinformatics.lcd-ustc.org/gps_web/predict.php . 2. One or more protein sequences can be input into the text box of the prediction page of GPS web service, as shown in **Figure 1**. The sequence data can be in raw or FASTA format. 3. Choose the protein kinases \(PKs) for prediction. 4. Select a proper cut-off value for each PK, or the default parameter will be used. 5. Click on the “Submit” button and wait for results. **How to interpret the results**: 6. If a single protein sequence is input in the raw sequence format, i.e. only the amino acid

sequence, it will be named as ProteinWithNoName. 7. If protein sequences are provided in the FASTA format, the results will be separated by the description line, starting with “>”, of each sequence in FASTA format. 8. The prediction results are organized in tab delimited format. For each predicted phosphorylation site, the position, kinase, flanking peptide, GPS score and Cut-off value are presented. A higher score indicates that the peptide is more probable to be a real phosphorylation site. 9. The prediction results are downloadable as a text file, as in **Figure 2**, and can be easily accessed with other automatic analyzing programs.

Timing

Approximately, 15~20 seconds per protein sequence using the aforementioned personal computer.

Critical Steps

Step 3: Choosing protein kinases \(PKs) for prediction is the most important step in this procedure. The phosphorylation sites on a protein can be mapped through large-scale phosphoproteome scanning with mass spectrometry or mutagenesis of the potential sites. And the predicted scores on these phosphorylation sites for all PK groups might be useful to infer which PK mediated the phosphorylation. It would be productive to consider the spatio-temporal profile of a PK and its candidate substrates as PK can only phosphorylate proteins proximally co-localized. Thus, an additional layer of fidelity check is to ascertain if a PK is co-distributed with the predicted protein. Conversely, our GPS is useful to predict a cognate kinase when a protein phosphorylation is experimentally confirmed but the mediating PK is unknown. Many other factors can also contribute to the specificity of PK recognition _in vivo_ [4]. These factors include co-complex of PKs with their substrates, interacting through modular docking sites, and phosphopeptide-binding mechanisms, etc. Thus, the prediction results of the GPS web service can be further refined using these factors. **Step 4**: Choosing a proper cut-off value is another essential step. In the GPS web service, lower threshold means lower specificity \(Sp) but higher sensitivity \(Sn), and higher threshold is vice versa. The default cut-off value stands for the balance between Sn and Sp. If a user wants to explore all the potential phosphorylation sites on a protein for further experimental investigation, a low threshold should be chosen to promise high sensitivity. While the predicted phosphorylation sites with a higher threshold are more likely to be true positives, since they are more similar to the known sites.

Anticipated Results

The prediction page of GPS is shown in **Figure 1**. Here we use the prediction of Aurora-B specific phosphorylation sites on MCAK of Xenopus \(shown in **Figure 2**) as an example. The predicted phospho-serine/threonine residues are highly consistent with experimental observations, as shown in **Figure 3**.

References

1. Parsons, D.W. et al. Colorectal cancer: mutations in a signalling pathway. _Nature_ **436**, 792 \(2005). 2. Xue, Y. et al. GPS: a comprehensive www server for phosphorylation sites prediction. _Nucleic Acids Res_ **33**, W184-187 \(2005). 3. Zhou, F.F., Xue, Y., Chen, G.L. & Yao, X. GPS: a novel group-based phosphorylation predicting and scoring method. _Biochem Biophys Res Commun_ **325**, 1443-1448 \(2004). 4. Biondi, R.M. & Nebreda, A.R. Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. _Biochem J_ **372**, 1-13 \(2003).

Acknowledgements

Y. Xue's and X. Yao's work is supported by Chinese Natural Science Foundation \(39925018, 30270654 and 30270293), Chinese Academy of Science \(KSCX2-2-01), Chinese 973 project \(2002CB713700), Chinese Minister of Education \(20020358051), American Cancer Society \(RPG-99-173-01), National Institutes of Health \(DK56292; CA92080) and a Distinguished Scholar award from the Georgia Cancer Coalition. F. Zhou's and Y. Xu's work is supported in part by the National Science Foundation \(NSF/DBI-0354771, NSF/ITR-IIS-0407204, NSF/DBI-0542119, NSF/CCF0621700) and a Distinguished Scholar award from the Georgia Cancer Coalition.

Figures

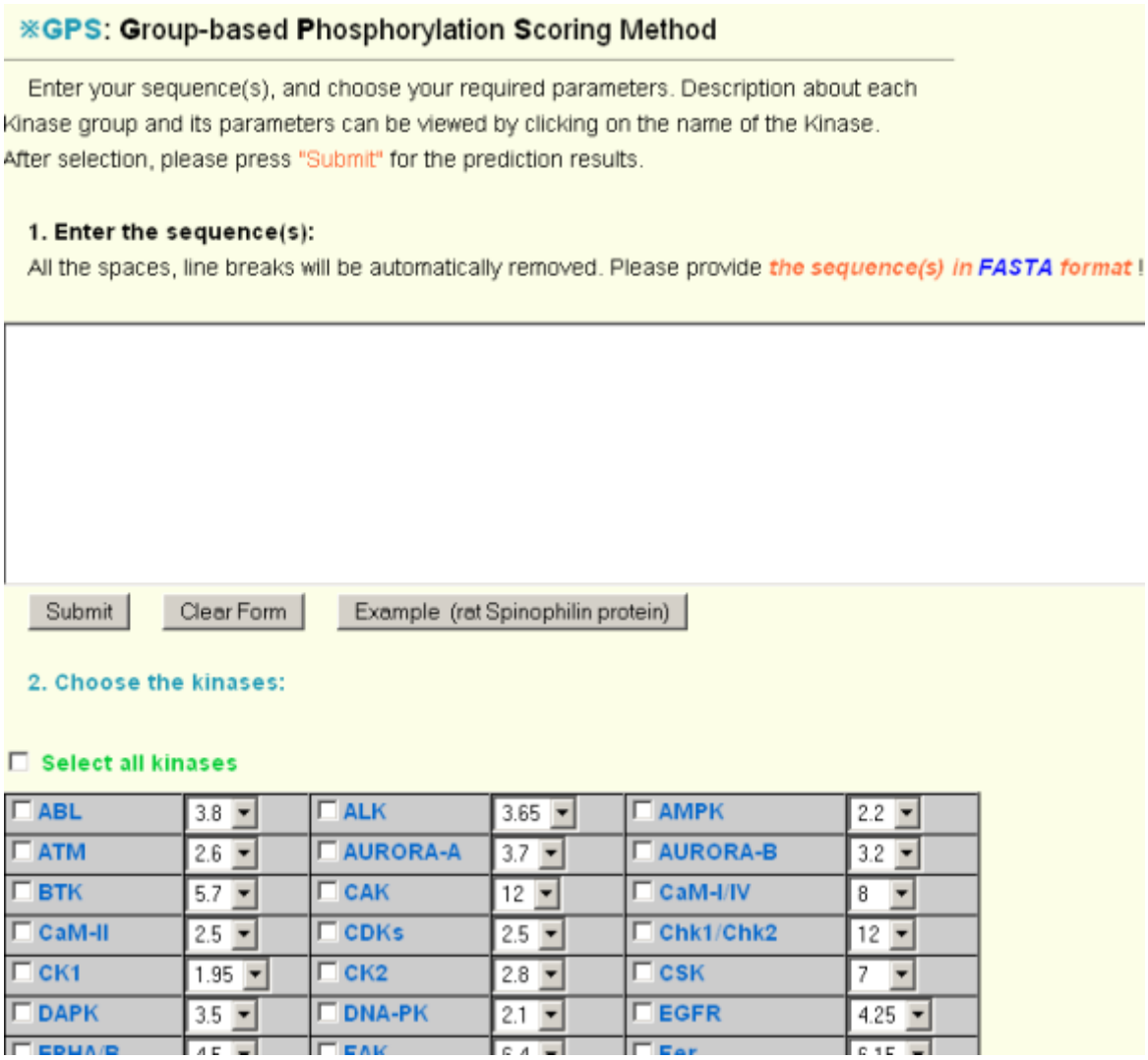


Figure 1

The prediction page of GPS "http://bioinformatics.lcd-ustc.org/gps_web/predict.php":http://bioinformatics.lcd-ustc.org/gps_web/predict.php

KCM1/MCAK; Kinesin central motor 1 (Q91636); *Xenopus laevis*

※GPS: Group-based Phosphorylation Scoring Method

Go back to GPS main page

Predicted phosphorylation sites:

Download the TAB-delimited data file from **here**.

>KCM1/MCAK

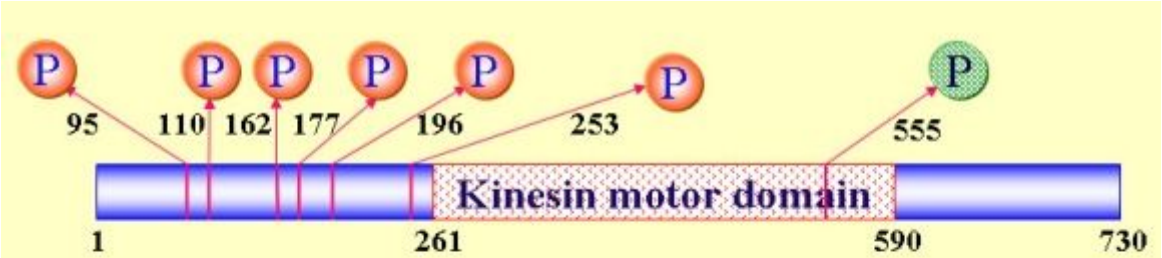
Position	Kinase	Peptide	GPS Score	Cutoff Score
95	Aurora-B	KRKTISK	6.174	3.2
110	Aurora-B	AKNSLLS	3.783	3.2
162	Aurora-B	SRSTKVS	5.174	3.2
177	Aurora-B	TRISEIV	4.652	3.2
196	Aurora-B	RRKSNIV	7.609	3.2
253	Aurora-B	HRISMAD	4.217	3.2
555	Aurora-B	LRDSFIG	3.478	3.2

Download the TAB-delimited data file from **here**.

Go back to GPS main page

Figure 2

The kinase Aurora-B specific phosphorylation sites on Xenopus MCAK (Q91636) is predicted.



KCM1/MCAK; Kinesin central motor 1 (Q91636); Xenopus laevis

MCAK	PMID	Phosphorylation sites
Lan *et al.*, 2004	14972678	S161, **T162**, **S177**, **S196**, T229, and **S253**
Ohi *et al.*, 2004	15064354	**T95**, **S110**, S161, **S177**, and **S196**
GPS	15555589, 15980451	**T95**, **S110**, **T162**, **S177**, **S196**, **S253**, and S555

Figure 3

Comparison of GPS prediction results of kinase Aurora-B with experimental results. For Xenopus MCAK (Swissprot Accession No.: Q91636), GPS predicts seven sites as positive (T95, S110, S161, S177, S196, S253, and S555), of which six sites (T95, S110, S161, S177, S196, and S253) were experimentally verified as phosphorylation sites of Aurora-B.