GPS-PBS Manual

Group-based Prediction System for PBS

7/5/2020

Author: Yaping Guo, Wanshan Ning and Yu Xue*

Contact: Yu Xue, xueyu@hust.edu.cn

The websever is only free for academic research.

The latest version of GPS-PBS is available from http://pbs.biocuckoo.cn/.

Index

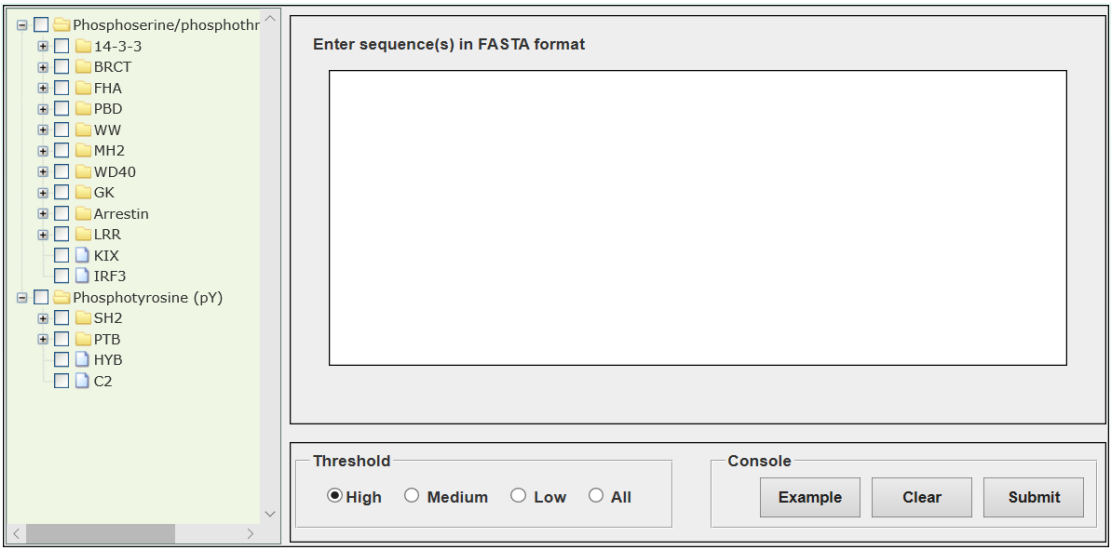
Statement

1. **Implementation**. The online service of the CUCKOO Workgroup are implemented in in PHP and JavaScript and will be provided.
2. **Availability**. Our online service are freely available for academic researches.
3. **GPS**. Previously, we used the GPS to denote our Group-based Phosphorylation Scoring algorithm. Currently, we re-denote the GPS as Group-based Prediction Systems. This online service is an indispensable part of GPS.
4. **Usage**. Our online service are designed in an easy-to-use manner. Also, we invite you to read the manual before using the online service.
5. **Updation**. Our online service will be updated routinely based on users' suggestions and advices. Thus, your feedback is greatly important for our future updation. Please do not hesitate to contact with us if you have any concerns.
6. **Citation**. Usually, the latest published articles will be shown on the websites. We wish you could cite the article if the online service has been helpful for your work.
7. **Acknowledgements.** This study was funded by Special Project on Precision Medicine under the National Key R&D Program (2017YFC0906600 and 2018YFC0910500), the Natural Science Foundation of China (31930021, 31970633, 31671360, and 81701567), the Fundamental Research Funds for the Central Universities (2017KFXKJC001 and 2019kfyRCPY043), Changjiang Scholars Program of China, and the program for HUST Academic Frontier Youth Team.

Introduction

In eukaryotes, protein phosphorylation is by far the most important and widespread post-translational modification (PTM) that mainly occurs on specific serine (S), threonine (T) or tyrosine (Y) residues in protein substrates, and orchestrates a variety of biological processes including signaling transduction, cell cycle/proliferation, autophagy and metabolism. Importantly, numerous proteins containing phosphoprotein-binding domains (PPBDs) can recognize and bind phosphoserine (pS), phosphothreonine (pT) or phosphotyrosine (pY) residues in specific substrates as "readers", which dictate the phosphorylation signaling events delivered from "writers", namely, protein kinases (PKs), and accurately propagate signals into downstream pathways. Dysregulation of normal interactions between PPBDs and p-sites is frequently associated with human diseases such as cancer and neurodegenerative disorders. Thus, the identification of PPBD-specific binding p-sites (PBSs) is fundamental for revealing dynamic phosphorylation signaling networks.

In this work, we manually collected 4458 experimentally identified PBSs in 950 PPBD-binding proteins (PPBPs) that interact with 268 PPBD-containing proteins (PPCPs) from 12 eukaryotic species. We classified these known PBSs into a hierarchical structure with three levels, including group, family and single PPBD cluster, based on the annotations of PPCPs. With a hypothesis that PPBDs in the same family/cluster might recognize similar sequence motifs in substrates, we considerably improved our previously developed group-based prediction system (GPS) algorithm, and adopted a deep learning plus transfer learning for model training. Finally, we developed a new online service named **GPS-PBS**, which implemented **138** predictors for 122 PPBD clusters belonged to 2 groups and 16 families.



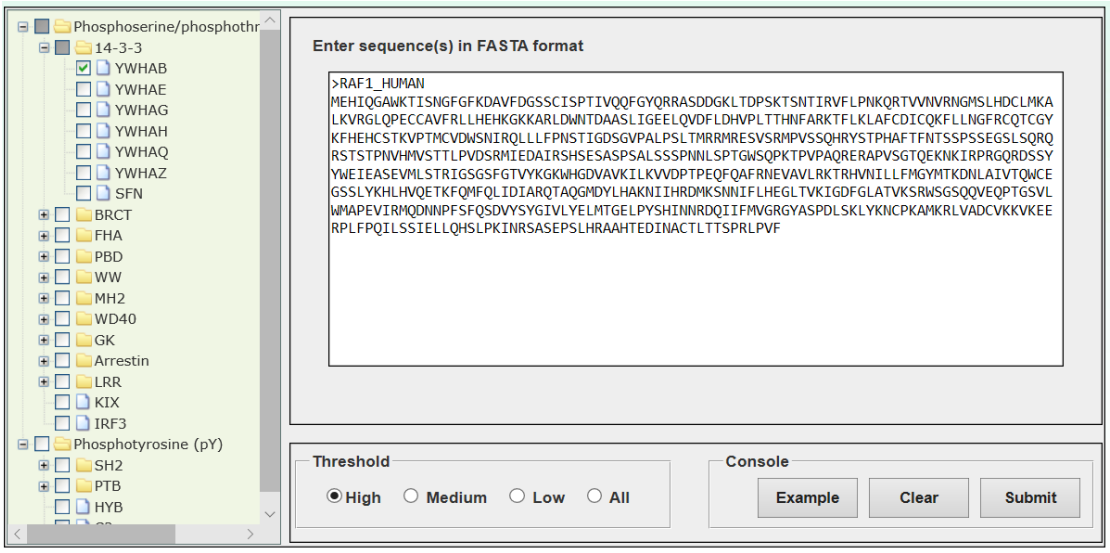
GPS-PBS User Interface

Prediction of PBS

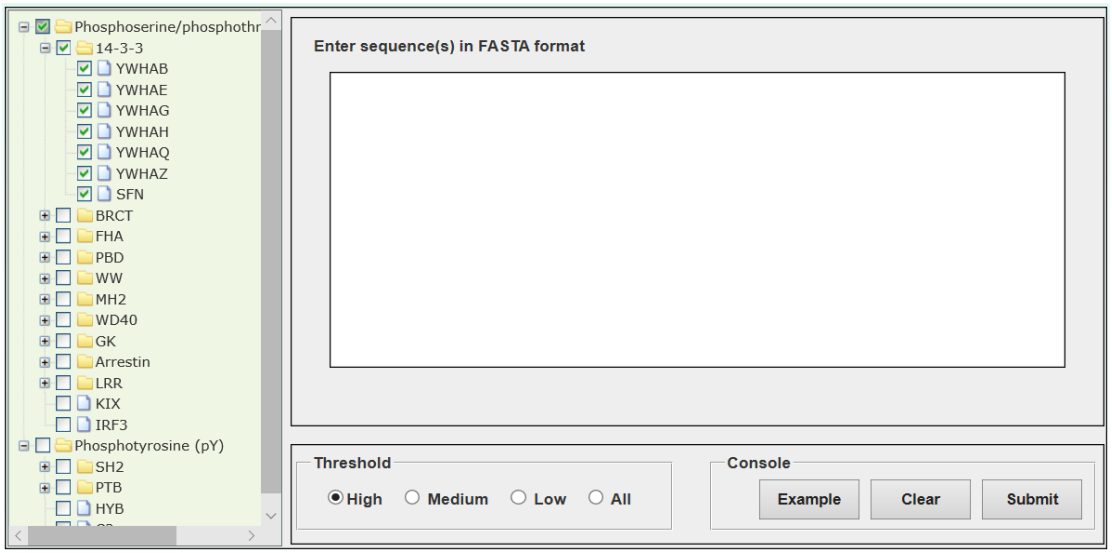
For convenience, the GPS-PBS allows users to input their protein sequences into the "TEXT form" for prediction. One or more protein sequences should be prepared in **FATSA** format as below:

>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXXX

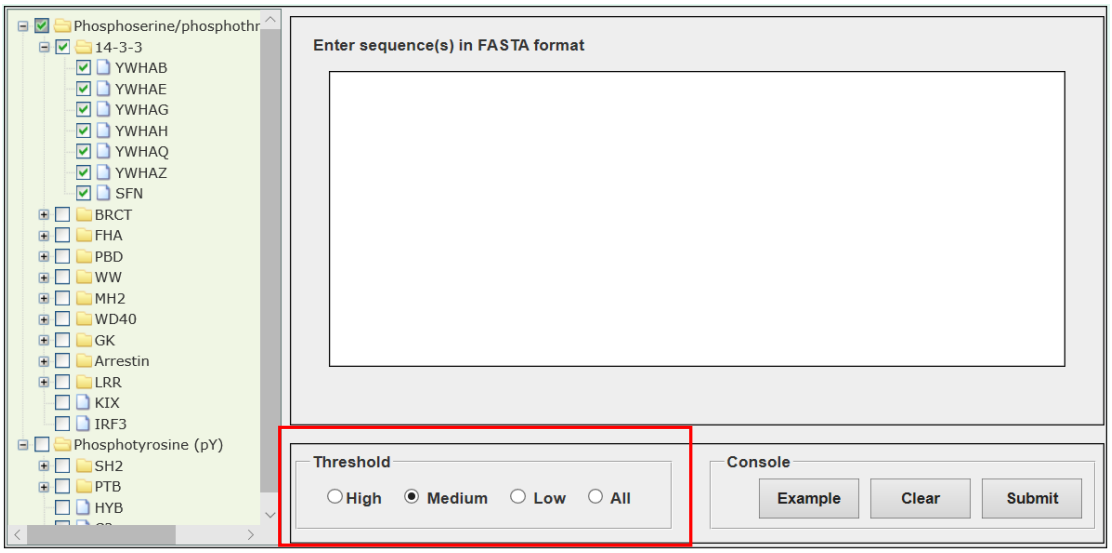
Please note: All irregular words, including non-amino acid word (eg, number) and blank, will be removed automatically. As an instance, we put **human RAF1** protein sequence as an example for GPS-PBS. Users could click on the "Example" button to access the example.



Choose one or more PPBDs from the **PPBD Hierarchy Tree**



Choose a **Threshold** what you need, default is **High**.

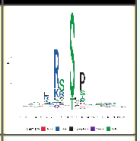
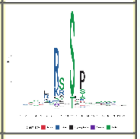
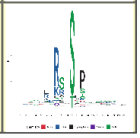


Click on the **Submit** button, then the predicted PBSs will be shown.

※ GPS-PBS Web Service

Download Return

Result has 3 items!

ID	Position	Code	PPBD	Peptide	Score	Cutoff	Source	Logo
RAF1_HUMAN	257	S	14-3-3/YWHAB	SLSQRQRSTSTPNVH	0.1863	0.1041	Pred.	
RAF1_HUMAN	259	S	14-3-3/YWHAB	SQRQRSTSTPNVHMV	0.4709	0.1041	Exp.	
RAF1_HUMAN	621	S	14-3-3/YWHAB	PKINRSASEPSLHRA	0.4708	0.1041	Exp.	

Total 1 Pages

Release Note

1. Jan. 1st, 2008, the online service and the local stand-alone packages of GPS 2.0 were released. The stand-alone software of GPS 2.0 could support Windows Operating Systems.
2. Jan. 29th, 2008, a bug was found that the version 2.0 couldn't be used under non-English Operating Systems. We fixed the bug and released the version 2.0.1 beta version. We thank Dr. Miguel Angel Sanchez (Malaga, Spain) and Dr. Gilles Vachon (Universite J. Fourier, France) to send us feedbacks.
3. Apr. 13th, 2008, The GPS 2.0.1 was released, with online service and local packages, to support three major Operating Systems, including Windows, Linux/Unix and Mac. Also, the GPS 2.0.1 manual was updated and included in the packages.
4. Mar. 1st, 2009, The GPS 2.1 was released, with online service and local packages and support three major Operating Systems including Windows, Linux/Unix and Mac. In this version, the newly developed motif length selection (MLS) method was introduced and the robustness of the prediction system was greatly improved.
5. Jul. 21st, 2009, The GPS 2.1.1 was released, with online service and local packages and support three major Operating Systems including Windows, Linux/Unix and Mac.
6. Sep. 13th, 2012, The GPS 2.1.2 was released, with online service and local packages and support three major Operating Systems including Windows, Linux/Unix and Mac.
7. May. 14th, 2013, The GPS-Polo was released with local packages and support three major Operating Systems including Windows, Linux/Unix and Mac.
8. Dec. 14th, 2014, The GPS 3.0 was released, with online service and local packages and support three major Operating Systems including Windows, Linux/Unix and Mac. In this version, *k-means* clustering, peptide selection (PS), and weight training (WT) procedures were added to enhance the prediction performance.
9. Jul. 20th, 2019, The GPS 5.0 was released, with online service and local packages and support three major Operating Systems including Windows, Linux/Unix and Mac. In this version, two novel methods of position weight determination (PWD) and scoring matrix optimization (SMO) were developed to improve the performance for kinase-specific phosphorylation sites prediction.