



GPS-Polo Manual

Prediction for Polo-like Kinases

Version 1.0
03/17/2011

Author: Jun Cao, Jian Ren & Yu Xue

Contact:

Jun Cao, wtfcj@mail.ustc.edu.cn

Dr. Jian Ren, renjian.sysu@gmail.com

Dr. Yu Xue, xueyu@mail.hust.edu.cn

The software is only free for academic research.

The latest version of GPS-Polo software is available from <http://polo.biocuckoo.org>

Copyright (c) 2011. The CUCKOO Workgroup. All Rights Reserved.

Index

STATEMENT	2
INTRODUCTION	3
DOWNLOAD & INSTALLATION	7
PREDICTION OF PHOSPHO-BINDING AND PHOSPHORYLATION OF POLO-LIKE KINASES SITES.....	9
REFERENCES.....	16
RELEASE NOTE	17

Statement

1. Implementation.

The softwares of the CUCKOO Workgroup are implemented in JAVA (J2SE). Usually, both of online service and local stand-alone packages will be provided.

2. Availability.

Our softwares are freely available for academic reserches. For non-profit users, you can copy, distribute and use the softwares for your scientific studies. Our softwares are not free for commercial usage.

3. GPS.

Previously, we used the GPS to denote our Group-based Phosphorylation Scoring algorithm. Currently, we are developing an integrated computational platform for post-translational modifications (PTMs) of proteins. We re-denote the GPS as Group-based Prediction Systems. This software is an indispensable part of GPS.

4. Usage.

Our softwares are designed in an easy-to-use manner. Also, we invite you to read the manual before using the softwares.

5. Updation.

Our softwares 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 software websites. We wish you could cite the article if the software has been helpful for your work.

7. Acknowledgements.

The work of CUCKOO Workgroup is supported by grants from the National Basic Research Program (973 project) (2006CB933300, 2007CB947401, 2007CB914503, and 2010CB912103), Natural Science Foundation of China (90919001, 30700138, 30900835, 30830036, 30721002, 30871236, and 90913016), Chinese Academy of Sciences (KSCX1-YW-R65, KSCX2-YW-R-139, INFO-115-C01-SDB4-36), and National Science Foundation for Post-doctoral Scientists (20080430100).

Introduction

The reversible phosphorylation, as one of the most pivotal post-translational modification (PTM) of proteins, regulates multiple physiological processes, such as cell cycle progression, signal transduction, metabolism pathway(1). Furthermore, tremendous advances over the past years have greatly promoted our understanding that as a crucial PTM-based regulatory mechanism, much of phosphorylation-mediated signal transduction are achieved through a specific recognition process in which 'phosphoprotein-binding domains' (PPBDs) bind to short motifs containing phosphoserine/threonine/tyrosine (*pS/pT/pY*)(2). This feature not only offers a reversible means to regulate protein-protein interactions both spatially and temporally, but also plays a significant role in orchestrating the cell cycle progression(3). Previous considerable works have identified 10 PPBDs—SH2, 14-3-3, FHA, MH2, WD-40, WW, PTB, BRCT, C2, PBD—in the human proteome(2). Thereinto, PBD is the research focus in this work.

Since the direct relationship between mutant alleles of the *Drosophila melanogaster* locus *polo* with the abnormal spindle poles was first described over 20 years ago(4), more and more attention drawn on these serine-threonine kinases, Plks, which are highly conserved from yeast to humans. The most striking character of Plks is the presence of polo-box domain (PBD) in the C-terminal non-catalytic region, which is indispensable and sufficient for proper sub-cellular localization of Plks during multiple stages of mitosis(5). Previously extensive research indicated that there are two kinds of proposed, not mutually contradictory, models for PBD-mediated substrate binding(6). In the processive phosphorylation model, the PBD first identifies and binds to specific *pS/pT* on a substrate, which provides an excellent temporal platform for the active kinase domain of Plks to subsequently phosphorylate the same substrate at other sites. In the distributive phosphorylation model, the substrate to which the PBD binds is distinct from the substrate which active kinase domain phosphorylates. Moreover, the former substrate always functions as phosphorylated

scaffold or docking protein, in order to anchor the kinase domain near to the latter substrate which locates in the specific sub-cellular structure. To a certain extent, the interaction between PBD and the former substrate offers a spatial opportunity for the kinase domain to phosphorylate the latter one(6). Moreover, as mentioned earlier, the PBD specifically binds to phosphoserine/threonine (*pS/pT*) of short motif implied that another 'priming kinase' has to phosphorylate the site before PBD binding. Indeed, for precisely the same reason, study about the physical recognition between PBD and specific substrates has always been constrained by the identification of priming kinases. In addition, clinical evidence suggests that aberrant expression of Plk1 are closely associated with tumorigenesis(7). Accordingly, identification of substrates with exact sites is an urgent topic to get a better understanding of the exact molecular mechanism of Plks.

Conventional experimental identification with a site-directed mutagenesis strategy is time-consuming and of low-throughput. Remarkably, landmark studies on Plks, accomplished by Elia *et al.*, identified PBD as a specific *pS/pT* binding domain and determined its optimal binding motif with the application of peptide-library screens(8). In 2007, Lowery and colleagues performed a large-scale identification of PBD interactome with mass spectrometry (MS) and detected 622 proteins(9). Another significant breakthrough was executed by Santamaria *et al.*, which identified 358 unique Plk1-dependent phosphorylation sites and over 100 sites were validated in vitro using peptide arrays(10). Currently, the convenient computational approaches have aroused wide concern. A series of databases and predictors have been springing up, which had made a great contribution to promote the development of PTM research(11). Despite the availability of these existing resources, the special computational tools developed for Plks analysis is still lacking.

In this work, we manually collected 56 phospho-binding sites in 47 distinct substrates and another 275 experimentally verified phosphorylation sites in 124 unique proteins from scientific literatures and public databases. Based on our previous works(12), the GPS 2.1 algorithm was applied to develop GPS-Polo 1.0 for prediction of Plks. The leave-one-out validation and 4-, 6-, 8- and 10-fold

cross-validation were calculated to evaluate the prediction performance and system robustness. The performance of the GPS-Polo 1.0 for phospho-binding, displayed a potential accuracy of 95.27%, a sensitivity of 82.14% and a specificity of 95.41%. On the other side, for phosphorylation, it also demonstrated a promising accuracy of 84.55%, a sensitivity of 53.62% and a specificity of 85.03%. As a demonstration application, we collected hundreds of phosphorylated proteins from PubMed and extant databases, which the particular Plks related *pS/pT* sites had to be experimentally verified. Through systematically comparison of the GO terms of the substrates between Plks phospho-binding and phosphorylation, we observed that the former prefers to attack mitosis. To further validate the reliability of the GPS-Polo 1.0, we randomly selected Mis18B as experimental candidate after a rigorous screening. Our biochemical experiments demonstrated that Plk1 interacts with Mis18B *in vivo* and *in vitro*. *pT14* and *pS48* of hMis18B are identified as the major PBD binding sites, which is consistent with the prediction of GPS-Polo 1.0. Furthermore, this binding association plays a significant role in maintaining the stability of hMis18B. Therefore, we speculated that the interaction between Plk1 and Mis18B is perhaps relevant to the CENP-A recruitment into subkinetochore chromatin, which is closely related to the chromosome segregation in eukaryotic cells. Taken together, our prediction and analysis might be of use for further experimental investigations. Finally, the online service and local packages of GPS-Polo 1.0 were implemented in JAVA 1.4.2 and freely available at: <http://polo.biocuckoo.org/> .

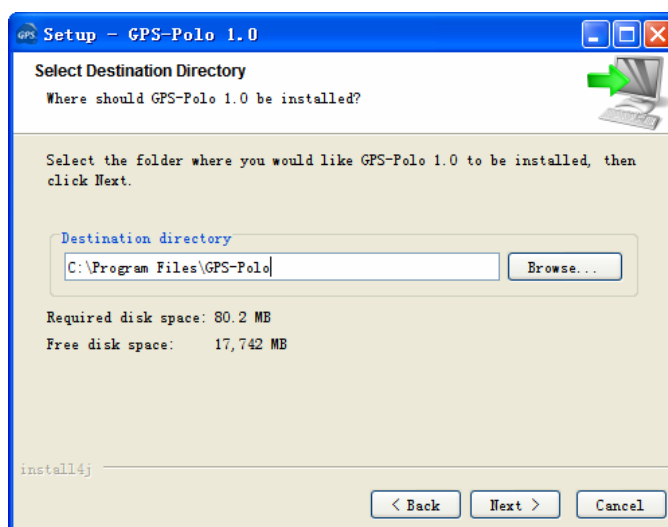
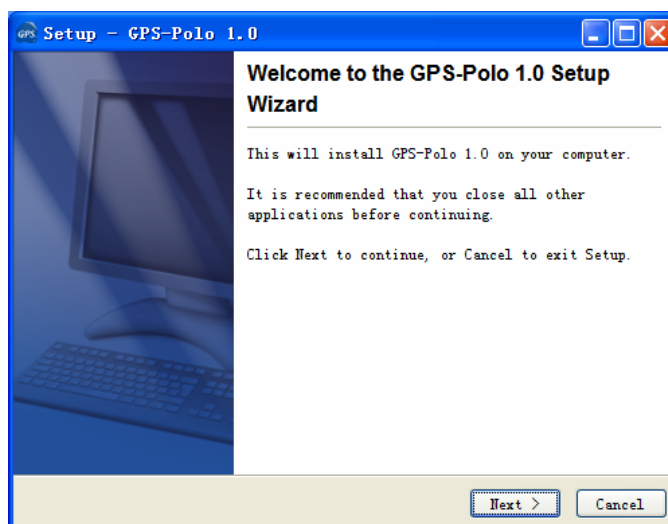
[illegible]

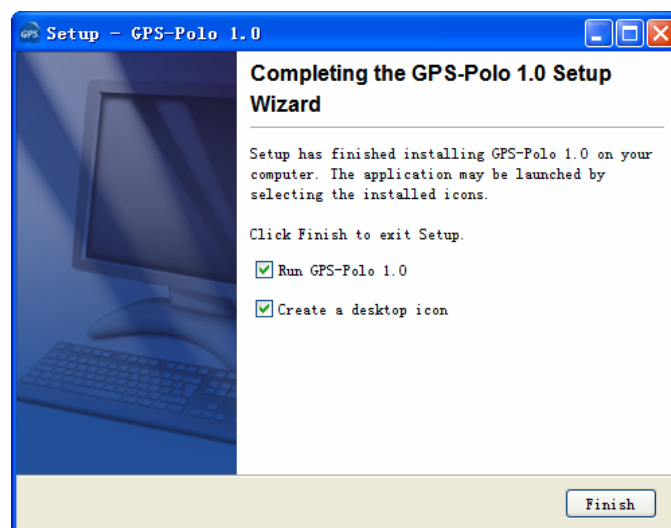
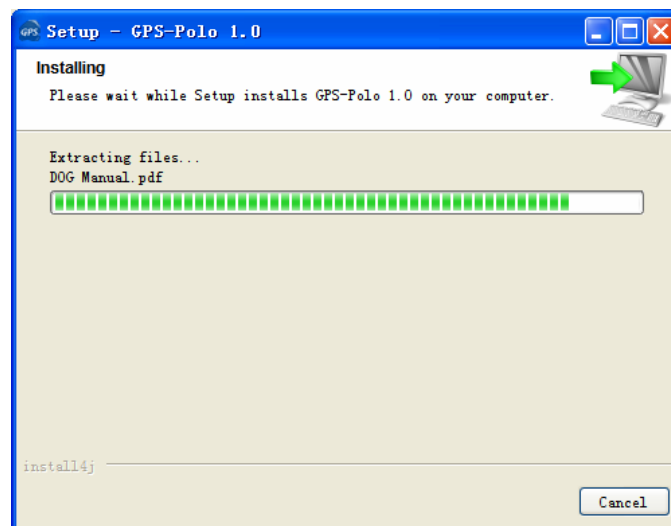
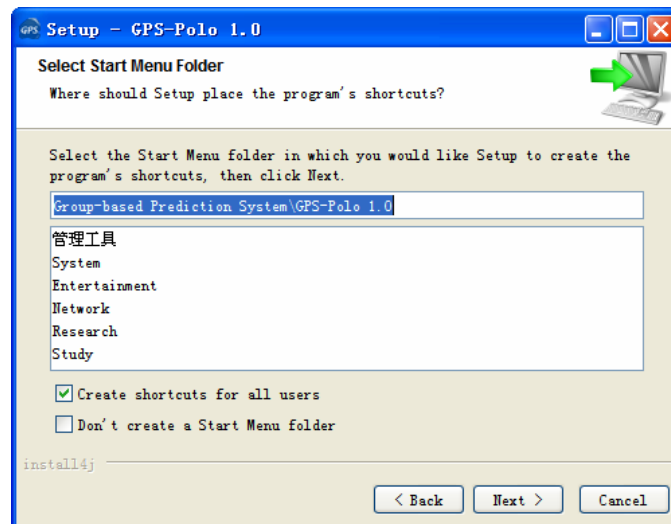
GPS-Polo 1.0 User Interface

Download & Installation

The GPS-Polo 1.0 was implemented in JAVA (J2SE), and could support three major Operating Systems (OS), including Windows, Linux/Unix or Mac OS X systems. Both of online web service and local stand-alone packages are available from: <http://polo.biocuckoo.org/>. We recommend that users could download the latest release.

Please choose the proper package to download. After downloading, please double-click on the software package to begin installation, following the user prompts through the installation. And snapshots of the setup program for windows are shown below:





Finally, please click on the **Finish** button to complete the setup program.

(2) Choose a **Threshold** that you need, the default cut-off is **Medium**.

The screenshot shows the GPS-Polo 1.0 web interface. The 'Predicted Sites' table is empty. Below it, the 'Enter sequence(s) in FASTA format' section contains an example sequence. The 'Threshold' section has four radio buttons: High, Medium (selected), Low, and All. The 'Console' section has a dropdown menu set to 'Both' and buttons for 'Example', 'Clear', and 'Submit'.

Position	Peptide	Score	Cutoff	Type

Enter sequence(s) in FASTA format

```
>Example(Human Mis18-beta, Q43482)
MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMIEWDTQVVKGSSPLGPAGLGAEPPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWD
LSRSLGAVVFSRVTNVNVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAE
LKEKIVLTHNRLKSLMKILSEVTPDQSKPEN
```

Threshold

☐ High ☒ Medium ☐ Low ☐ All

Console

Both

(3) Click on the **Submit** button, then the predicted phospho-binding and phosphorylation sites will be shown.

The screenshot shows the GPS-Polo 1.0 web interface after clicking the 'Submit' button. The 'Predicted Sites' table now contains five rows of results. The 'Threshold' section remains the same, with 'Medium' selected. The 'Console' section is also the same.

Position	Peptide	Score	Cutoff	Type
14	RHRSRCATPPRGDFC	6.446	5.268	Phospho-binding
48	TQVVKGSSPLGPAGL	6.75	5.268	Phospho-binding
183	TKAIVNASEMDIQNV	3.399	2.793	Phosphorylation
221	MKILSEVTPDQSKPE	6.911	5.268	Phospho-binding
225	SEVTPDQSKPEN***	4.199	2.793	Phosphorylation

Enter sequence(s) in FASTA format

```
>Example(Human Mis18-beta, Q43482)
MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMIEWDTQVVKGSSPLGPAGLGAEPPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWD
LSRSLGAVVFSRVTNVNVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAE
LKEKIVLTHNRLKSLMKILSEVTPDQSKPEN
```

Threshold

☐ High ☒ Medium ☐ Low ☐ All

Console

Both

(4) Then please click on the **RIGHT** button in the prediction form. You can use the “**Select All**” and “**Copy Selected**” to copy the selected results into Clipboard. Then please copy the results into a file, eg., an EXCEL file for further consideration. Also, you can choose “**Export Prediction**” to export the prediction results into a tab-delimited text file.

Position	Peptide	Score	Cutoff	Type
14	RHRSRCATPPRGDFC	6.446	5.268	Phospho-binding
48	TQVVKGSPLGPAGL	6.75	5.268	Phospho-binding
183	TKAIVNASEMDIQNV	3.399	2.793	Phosphorylation
221	MKILSEVTPDQSKPE	6.911	5.268	Phospho-binding
225	SEVTPDQSKPEN***	4.199	2.793	Phosphorylation

Enter sequence(s) in FASTA format

```
>Example(Human Mis18-beta, O43482)
MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMIEWDTQVVKGSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWD
LSRSLGAVVFSRVTNNVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAE
LKEKIVLTHNRLKSLMKILSEVTPDQSKPEN
```

Threshold: ☐ High ☒ Medium ☐ Low ☐ All

Console:

Again, you can also click the “**Export Prediction**” in **File** menu to export the results.

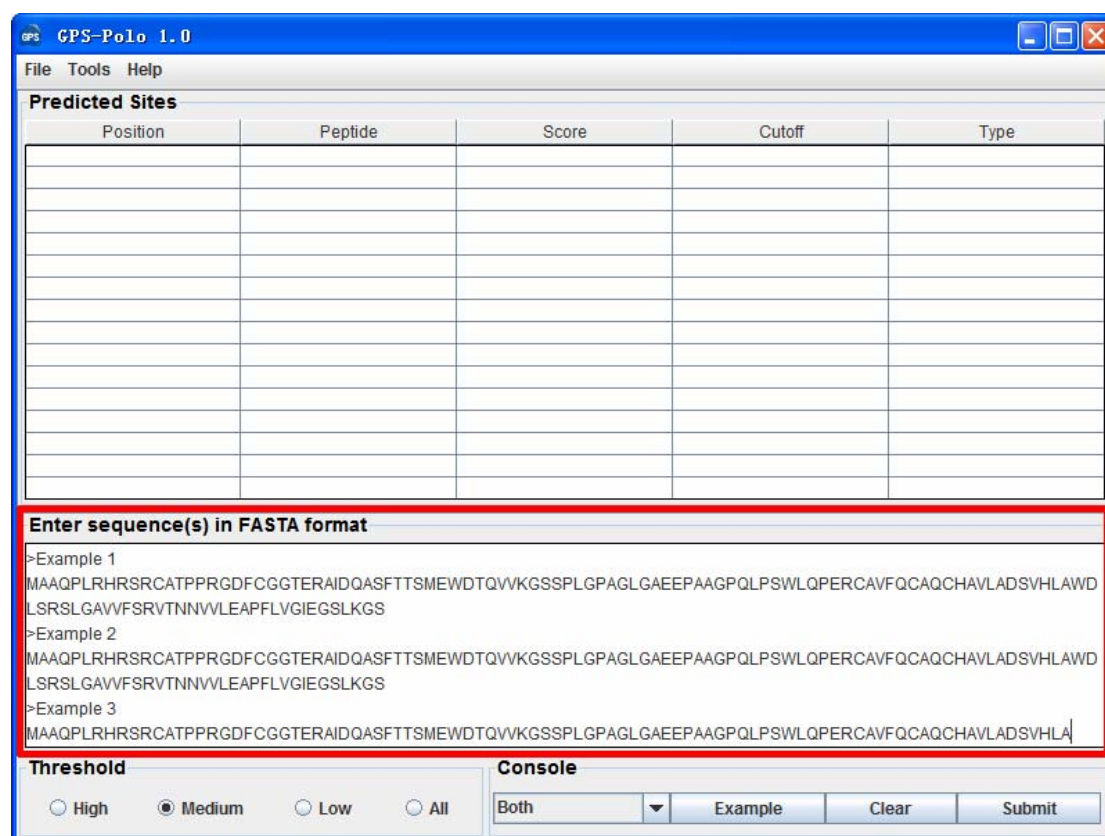


2. Multiple protein sequences in FASTA format

For multiple protein sequences, there are two ways to use the GPS-Polo 1.0.

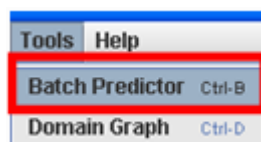
A. Input the sequences into text form directly. (Num. of Seq ≤ 2,000)

If the number of total protein sequences is not greater than 2,000, you can just use “Ctrl+C & Ctrl+V” (Windows & Linux/Unix) or “Command+C & Command+V” (Mac) to copy and paste your sequences into the text form of GPS-Polo 1.0 for prediction.



B. Use Batch Predictor tool.

If the number of protein sequences is very large, eg., yeast or human proteome, please use the **Batch Predictor**. Please click on the “**Batch Predictor**” button in the **Tools** menu.



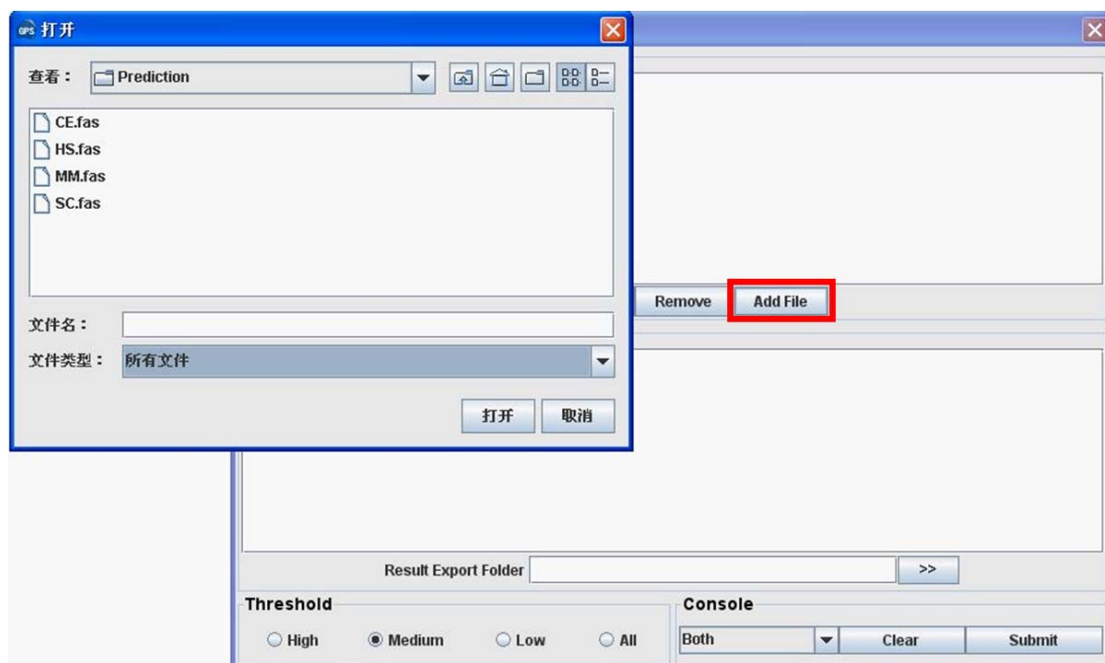
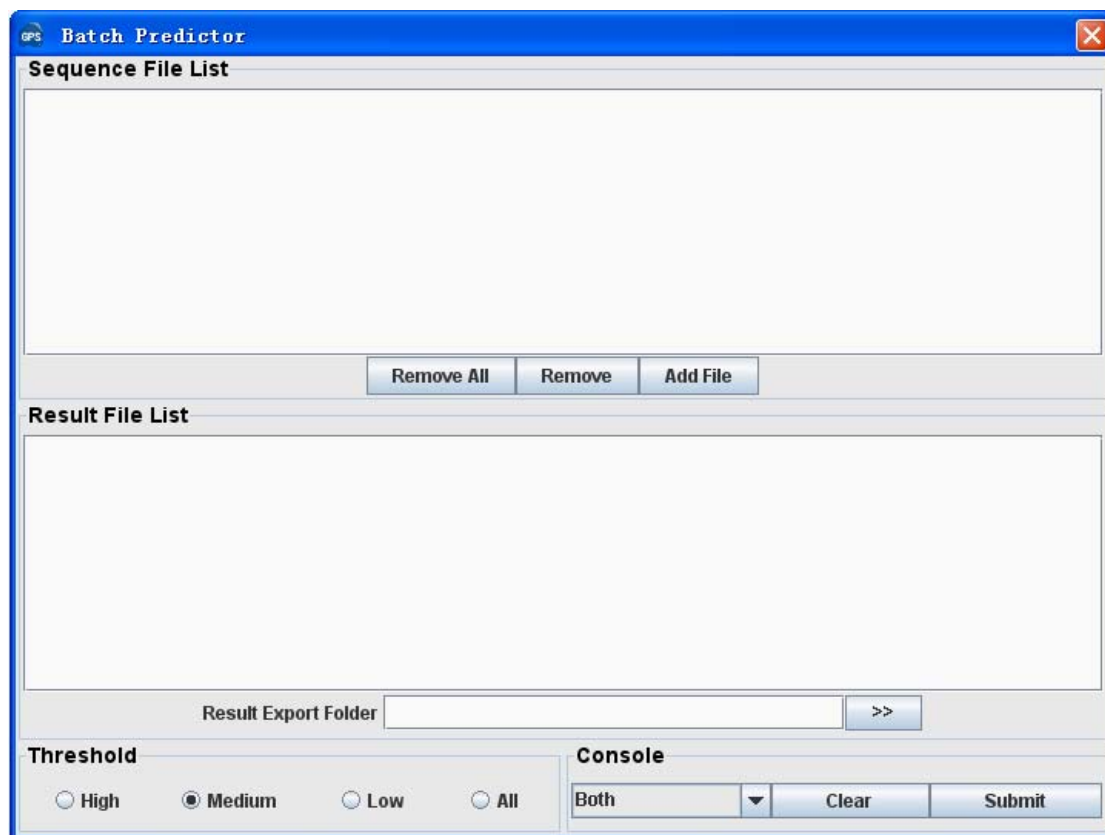
The following steps show you how to use it:

(1) Put protein sequences into one or several files (eg., SC.fas, CE.fas, and etc) with FATS format as below:

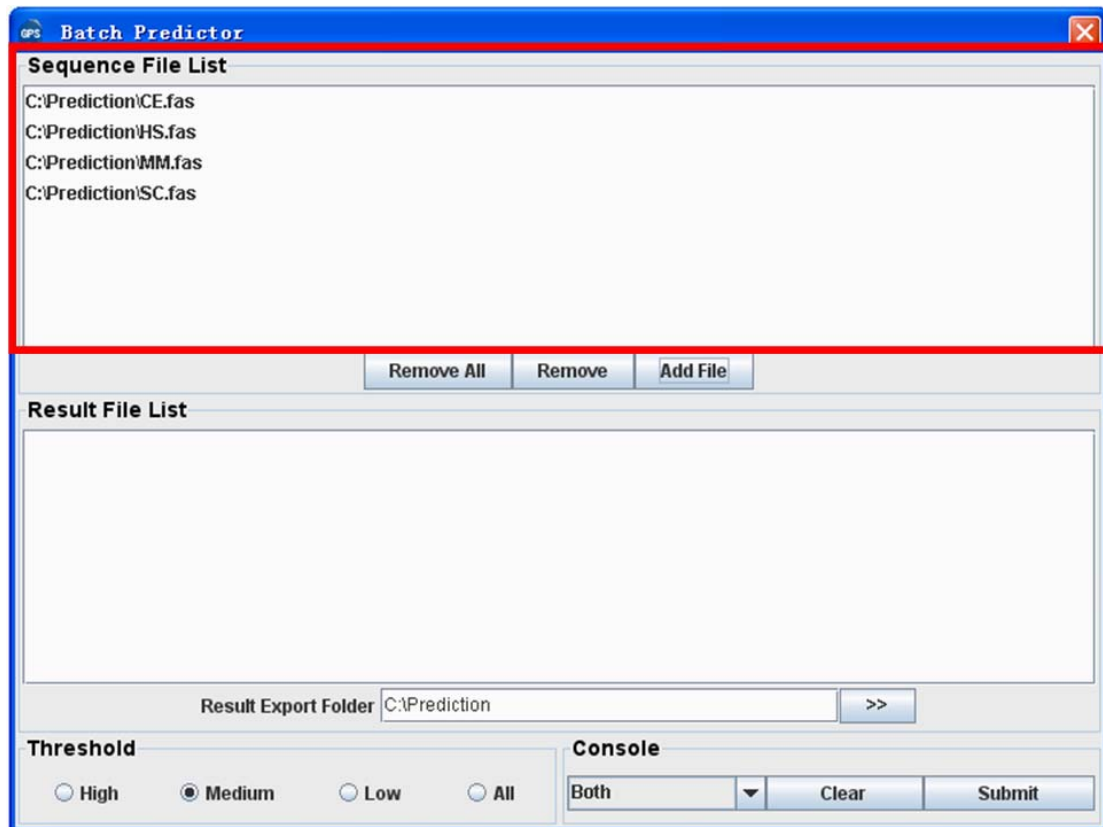
```
>protein1
XXXXXXXXXXXXXXXXX
XXXXXXXXXX
>protein2
XXXXXXXXXXXXXXXXXXXX...
>protein3
XXXXXXXXXXXXXXXXX
...
```

Most importantly, the name of each protein should be presented.

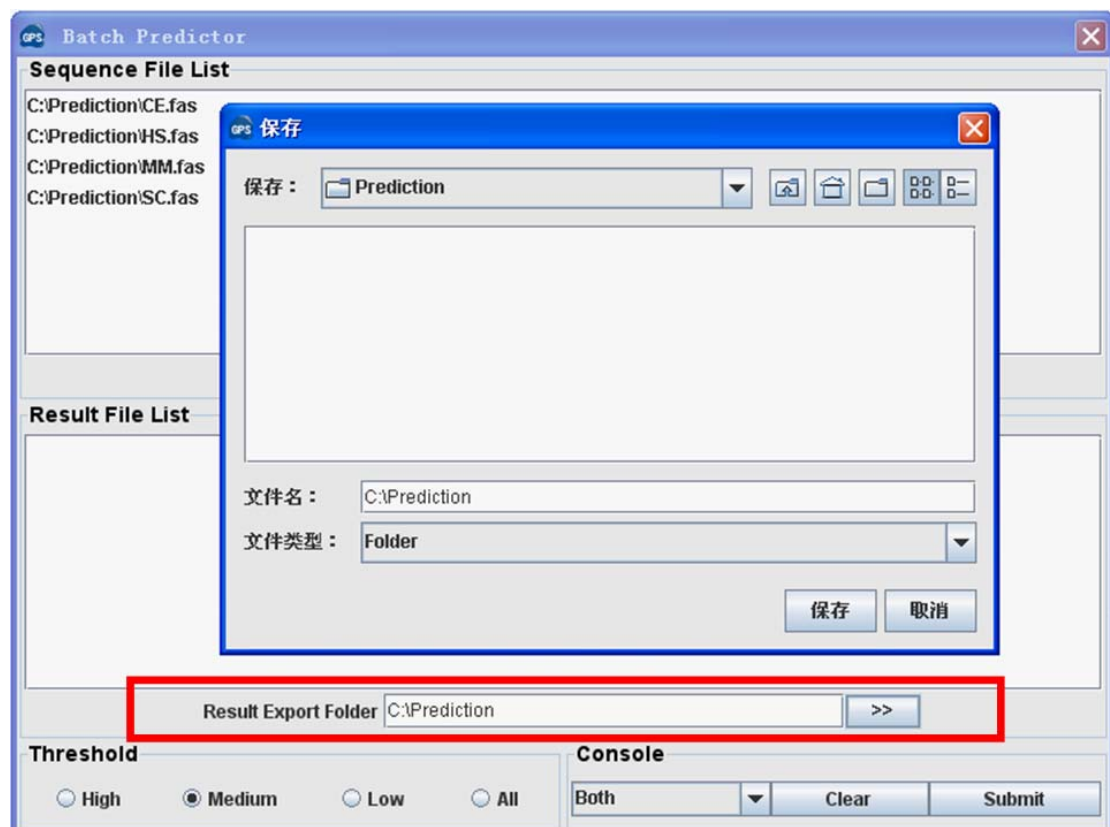
(2) Click on the **Batch Predictor** button and then click on the **Add File** button and add one or more protein sequence files in your hard disk.



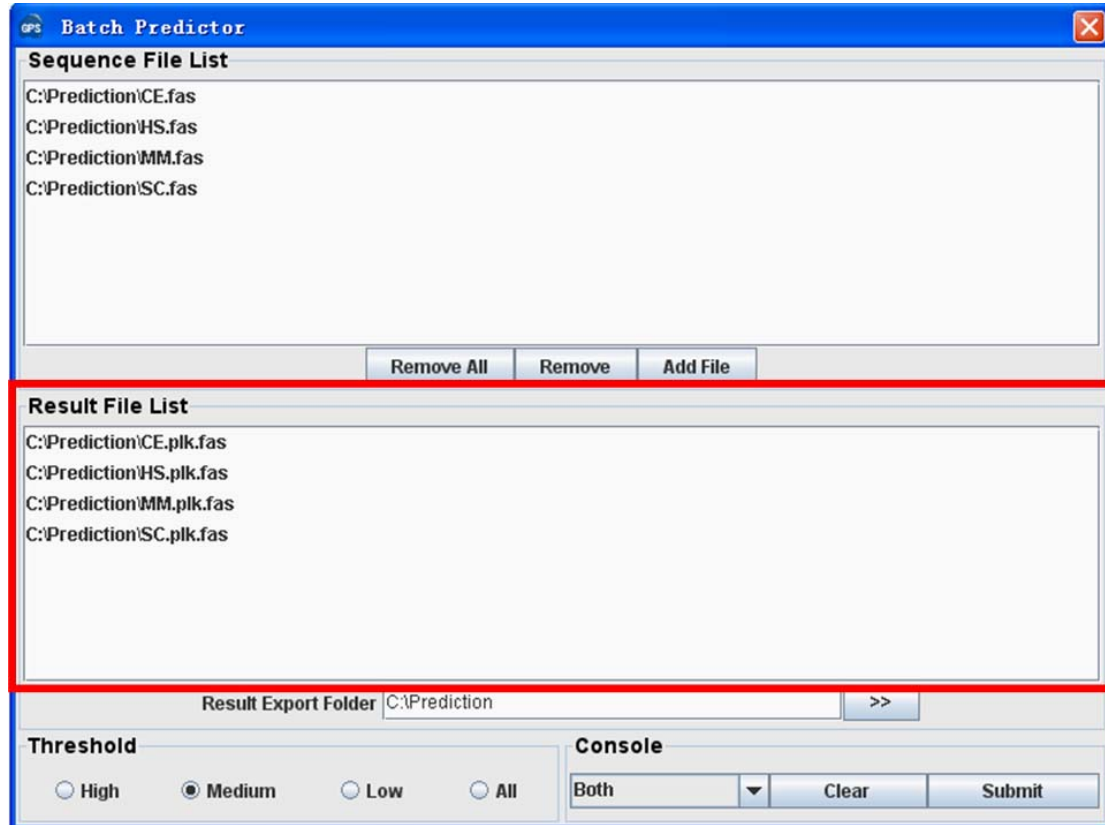
Then the names of added files will be shown in the **Sequence File List**.



(3) The output directory of prediction results should also be defined. Please click on the >> button to specify the export fold.



(4) Please choose a proper threshold before prediction. Then please click on the **Submit** button, then the **Batch Predictor** begin to process all of the sequence files that have been added to the list. The result of prediction will be export to the **Prediction Export Fold**, and the name of result files will be shown in the **Prediction File List**.



References

1. Cohen, P. (2002) The origins of protein phosphorylation. *Nat Cell Biol*, **4**, E127-130.
2. Gong, W., Zhou, D., Ren, Y., Wang, Y., Zuo, Z., Shen, Y., Xiao, F., Zhu, Q., Hong, A., Zhou, X. *et al.* (2008) PepCyber:P~PEP: a database of human protein protein interactions mediated by phosphoprotein-binding domains. *Nucleic Acids Res*, **36**, D679-683.
3. Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell*, **116**, 191-203.
4. Sunkel, C.E. and Glover, D.M. (1988) polo, a mitotic mutant of *Drosophila* displaying abnormal spindle poles. *J Cell Sci*, **89** (Pt 1), 25-38.
5. van Vugt, M.A. and Medema, R.H. (2005) Getting in and out of mitosis with Polo-like kinase-1. *Oncogene*, **24**, 2844-2859.
6. Lowery, D.M., Lim, D. and Yaffe, M.B. (2005) Structure and function of Polo-like kinases. *Oncogene*, **24**, 248-259.
7. Lu, L.Y. and Yu, X. (2009) The balance of Polo-like kinase 1 in tumorigenesis. *Cell Div*, **4**, 4.
8. Elia, A.E., Cantley, L.C. and Yaffe, M.B. (2003) Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science*, **299**, 1228-1231.
9. Lowery, D.M., Clauser, K.R., Hjerrild, M., Lim, D., Alexander, J., Kishi, K., Ong, S.E., Gammeltoft, S., Carr, S.A. and Yaffe, M.B. (2007) Proteomic screen defines the Polo-box domain interactome and identifies Rock2 as a Plk1 substrate. *EMBO J*, **26**, 2262-2273.
10. Santamaria, A., Wang, B., Elowe, S., Malik, R., Zhang, F., Bauer, M., Schmidt, A., Sillje, H.H., Koerner, R. and Nigg, E.A. (2010) The Plk1-dependent phosphoproteome of the early mitotic spindle. *Mol Cell Proteomics*.
11. Xue, Y., Gao, X., Cao, J., Liu, Z., Jin, C., Wen, L., Yao, X. and Ren, J. (2010) A Summary of Computational Resources for Protein Phosphorylation. *Curr Protein Pept Sci*.
12. Xue, Y., Liu, Z., Cao, J., Ma, Q., Gao, X., Wang, Q., Jin, C., Zhou, Y., Wen, L. and Ren, J. (2010) GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Eng Des Sel*.

Release Note

1. Sep. 28th, 2009, the online service and the local stand-alone packages of GPS-Polo 1.0 were released.
2. Nov. 12th, 2010, GPS-Polo 1.0 was renewed.