



# GPS-CCD Manual

Calpain Cleavage Detector

Version 1.0.0

11/11/2010

Author: Zexian Liu, Jian Ren & Yu Xue

Contact:

Zexian Liu, [lzx@mail.ustc.edu.cn](mailto:lzx@mail.ustc.edu.cn)

Dr. Jian Ren, [renjian.sysu@gmail.com](mailto:renjian.sysu@gmail.com)

Dr. Yu Xue, [xueyu@mail.hust.edu.cn](mailto:xueyu@mail.hust.edu.cn)

The software is only free for academic research.

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

Copyright (c) 2010. The CUCKOO Workgroup. All Rights Reserved.

# Index

STATEMENT .....	2
INTRODUCTION .....	3
DOWNLOAD & INSTALLATION .....	6
PREDICTION OF CALPAIN CLEAVAGE SITES .....	8
REFERENCES.....	15
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 researches. 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

Calpains constitute an important family of the  $\text{Ca}^{2+}$ -dependent cysteine proteases, which contain a nucleophilic cysteine in the catalytically active site (1-7). Calpains are widely expressed in mammals and conserved across eukaryotes (1-5,8-9). For instance, in budding yeast, at least one calpain-like protease, Rim13/Cpl1, has been identified, although its functions are still elusive (8-9). In humans, there are over 14 distinct members of the calpain superfamily, some of which are tissue specific. Calpain 1 ( $\mu$ -calpain, micromolar  $\text{Ca}^{2+}$ -requiring) and Calpain 2 (m-calpain, millimolar  $\text{Ca}^{2+}$ -requiring) are ubiquitously expressed and well characterized isoforms (1-2,4-5). Through spatial and temporal cleavage of a variety of substrates to change their conformation, function and stability (1-4),  $\text{Ca}^{2+}$ -activated calpains play an important role in numerous biological processes, including the regulation of gene expression, signal transduction, cell death/apoptosis, remodeling cytoskeletal attachments during cell fusion/motility and cell cycle progression (1-4,6,10-12). Moreover, calpain aberrancies are frequently implicated in a variety of diseases and cancers (5-7,13-14). Although many studies have tried to dissect the regulatory roles and molecular mechanisms of calpain-dependent cleavage, in fact our understanding of calpain is still fragmentary.

Identification of the site-specific calpain substrates is fundamental for dissecting the roles of calpain cleavage in numerous biological pathways. Besides the conventional experimental approaches with Edman N-terminal sequencing or mass spectrometry (MS) (12,15), a peptide library approach was also designed to investigate the sequence/structural specificities of calpains (16-18). Thus far, hundreds of calpain-cleaved proteins have been experimentally identified, including structural proteins, membrane receptors, and transcription factors (12,15-18). However, high-throughput technique for the identification of calpain substrates is still limited. Recently, besides time-consuming and labor-intensive experimental methods, the development of computational approaches has also promoted the discovery of the PTM sites (19). *In silico* prediction was shown to be able to rapidly generate useful information for further experimental verification. Although there have been ~170 databases and computational tools developed for various types of PTM analysis, especially phosphorylation (<http://www.biocuckoo.org/link.php>) (19), the computational prediction of calpain cleavage sites is still a great challenge. In a previous study (16), Tompa *et al.* collected 106 calpain cleavage sites in 49 substrates from the scientific literature, and determined the amino acid preferences around the cleavage bond, from P<sub>4</sub> (upstream) to P<sub>7</sub>' (downstream). They constructed a position-specific scoring matrix (PSSM), and observed that the preferred residues for  $\mu$ - calpain and m- calpain

recognition are Leu, Thr and Val in the P<sub>2</sub> position, and Lys, Tyr and Arg in the P<sub>1</sub> position (16). Based on this rationale, they designed the short peptide of TPLK|SPPPSR (“|” is the potentially cleaved position), which was experimentally verified to be a superior substrate of calpain (16). Recently, duVerle *et al.* constructed a CaMPDB database, which contains 104 experimental identified calpain targets with 267 sites (20). An online prediction of calpain cleavage sites with the support vector machines (SVMs) algorithm was also provided.

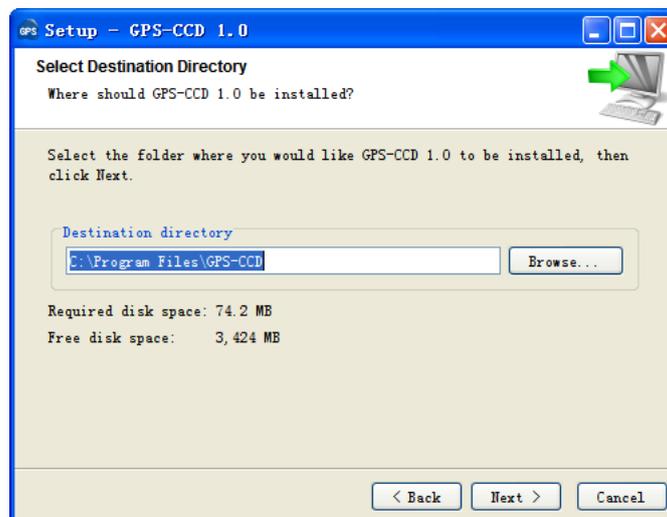
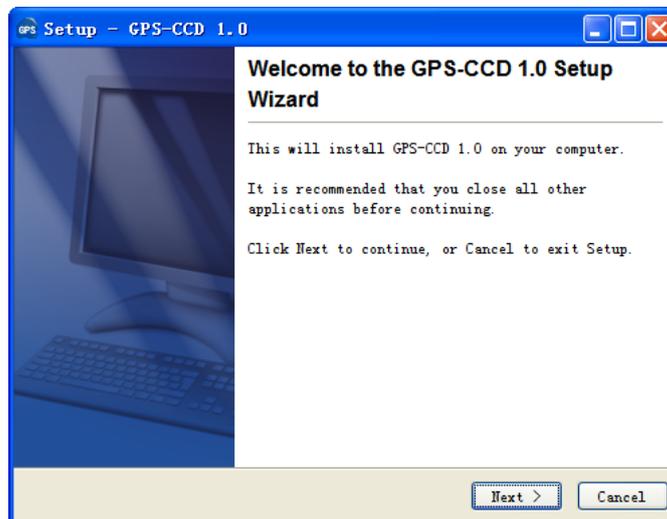
In this work, we collected 368 experimentally verified calpain cleavage sites in 130 proteins (Supplementary Table S1). With a previously developed algorithm of GPS (Group-based Prediction System) (21), we developed a novel software package of GPS-CCD (Calpain Cleavage Detector) for the prediction of calpain cleavage sites. The leave-one-out validation and 4-, 6-, 8-, 10-fold cross-validations were performed to evaluate the performance of the prediction system. By comparison, the GPS 2.0 algorithm was employed for its outstanding prediction performance, with an accuracy 89.98%, sensitivity 60.87% and specificity 90.07%. Furthermore, there are many proteins experimentally identified as calpain substrates for which the exact cleavage sites have not been verified, and we collected 196 such proteins from PubMed (Supplementary Table S2). As an application, we predicted potential calpain cleavage sites for these targets (Supplementary Table S2). These prediction results might be a useful resource for further experimental investigation. Finally, the online service and local packages of GPS-CCD 1.0 were implemented in JAVA 1.5 (J2SE 5.0) and are freely available for academic researchers at: <http://ccd.biocuckoo.org/>.

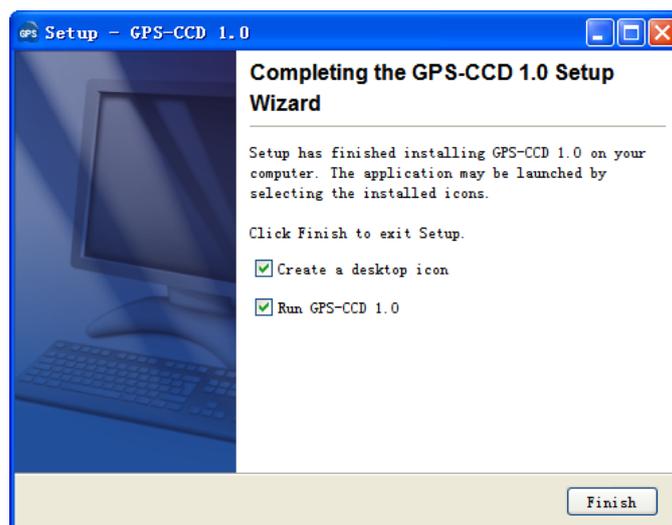
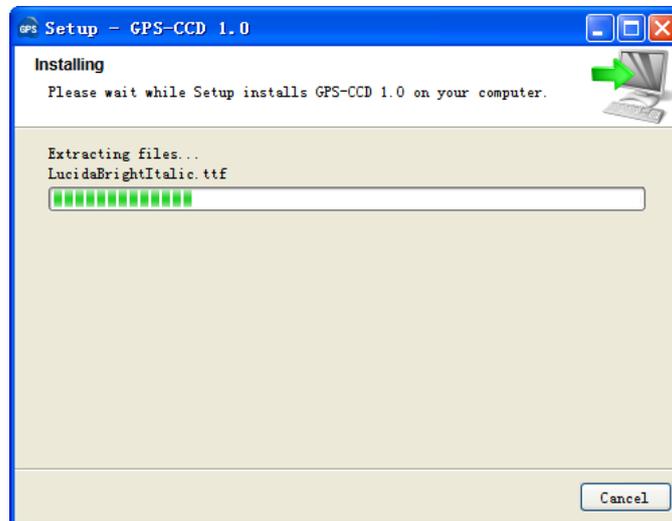
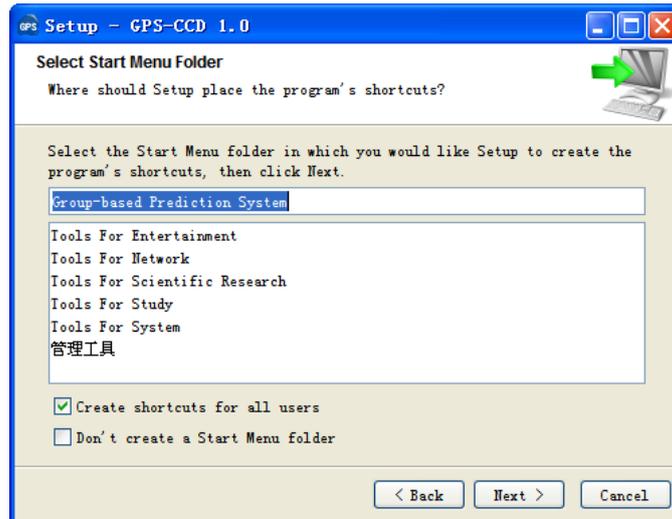


## Download & Installation

The GPS-CCD 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://ccd.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.

# Prediction of Calpain Cleavage Sites

## 1. A single protein sequence in FASTA format

The following steps show you how to use the GPS-CCD 1.0 to predict calpain cleavage sites for a single protein sequence in FASTA format.

(1) Firstly, please use “Ctrl+C & Ctrl+V” (Windows & Linux/Unix) or “Command+C & Command+V” (Mac) to copy and paste your sequence into the text form of GPS-CCD 1.0

Position	Peptide	Score	Cutoff
25	AARGDVQEV <b>R</b>  RL LH	0.891	0.654
47	LNRFGKTAL <b>Q</b>  VMMF	0.78	0.654
64	AIALELLK <b>Q</b> G ASPN	0.761	0.654
73	GASPNVQDT <b>S</b>  GTSP	0.84	0.654
74	ASPNVQDT <b>S</b> G TSPV	0.837	0.654
80	DTSGTSPV <b>H</b> D AART	0.688	0.654
127	GHTAVVS <b>F</b> LA AESD	0.91	0.654

Enter sequence(s) in FASTA format

```
>Example (human CDKN2D, P55273)
MLLEEVRA GDRLSGAAARGDVQEVRRLLHRELVHPDALNRFGKTALQVMMFGSTAI ALEL
LKQGASPNVQDTSGTSPVHDAARTGFLDTLKV LVEHGADVNP DGTGALPIHLAVQEGHT
AVVSFLAAESDLHRRDARGLTPELALQRGAQDLVDILQGHMVAPL|
```

Threshold:  High  Medium  Low  All

Console:

Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

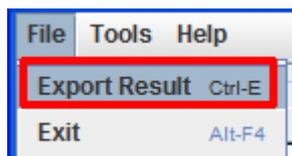


the “**Select All**” and “**Copy Selected**” to copy the selected results into Clipboard. Then please copy the results into a file, eg., an EXCEL file for further consideration. Also, you can choose “**Export Prediction**” to export the prediction results into a tab-delimited text file.

The screenshot shows the GPS-CCD 1.0 application window. At the top, there is a menu bar with 'File', 'Tools', and 'About'. Below the menu bar is a section titled 'Predicted sites' containing a table with four columns: 'Position', 'Peptide', 'Score', and 'Cutoff'. The table contains seven rows of data. A context menu is open over the row with Position 73, showing three options: 'Select All', 'Copy Selected', and 'Export Prediction'. Below the table is a text input area labeled 'Enter sequence(s) in FASTA format' with an example sequence. At the bottom, there is a 'Threshold' section with radio buttons for 'High', 'Medium', 'Low', and 'All', and a 'Console' section with 'Example', 'Clear Form', and 'Submit' buttons.

Position	Peptide	Score	Cutoff
25	AARGDVQEVRI RLRH	0.891	0.654
47	LNRFGKTALQ VMMF	0.78	0.654
64	AIALELLKQGI ASPN	0.761	0.654
73	GASPNVQDTS GTSP		0.654
74	ASPNVQDTSG TSPV		0.654
80	DTSGTSPVHD AART	0.688	0.654
127	GHTAVVSFLA AESD	0.91	0.654

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-CCD 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-CCD 1.0 for prediction.

Position	Peptide	Score	Cutoff
2	*****MA SGNARIG	0.7231	0.6377
7	*MASGNAR IGKPAPD	0.6731	0.6377
182	GWKPGSDT IKPNVDD	0.65	0.6377
193	NVDDSKEY FSKHN**	0.6885	0.6377
197	SKEYFSKH N*****	0.8462	0.6377

Enter sequence(s) in FASTA format

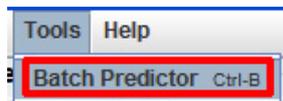
```
>Example 1
MASGNARIGKPAPDFKATAVVDGAFKEVKLSDYKGYWVLFYPLDFTFVCPTEIIAFSNRAEDFRKLGCEVLGVSVD SQFTHLAWINTPRKEGGLGPL
NIPLLADVTRRLSEGYVLTDEGIAYRGLFIIDGKGVLRQITVNDLPVGRSVDEALRLVQAFQYTDHGEVCPAGWKPGSDTIKPNVDDSKEYFSKH
>Example 1
MASGNARIGKPAPDFKATAVVDGAFKEVKLSDYKGYWVLFYPLDFTFVCPTEIIAFSNRAEDFRKLGCEVLGVSVD SQFTHLAWINTPRKEGGLGPL
NIPLLADVTRRLSEGYVLTDEGIAYRGLFIIDGKGVLRQITVNDLPVGRSVDEALRLVQAFQYTDHGEVCPAGWKPGSDTIKPNVDDSKEYFSKH
>Example 1
MASGNARIGKPAPDFKATAVVDGAFKEVKLSDYKGYWVLFYPLDFTFVCPTEIIAFSNRAEDFRKLGCEVLGVSVD SQFTHLAWINTPRKEGGLGPL
```

Threshold:  High  Medium  Low  All

Console:

### 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.



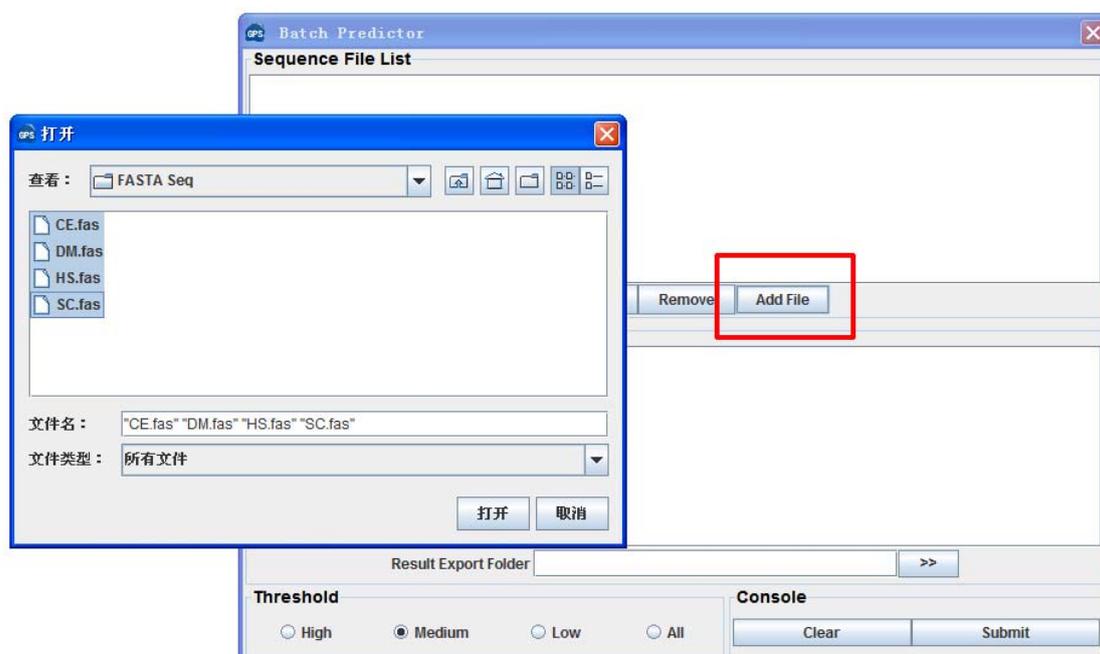
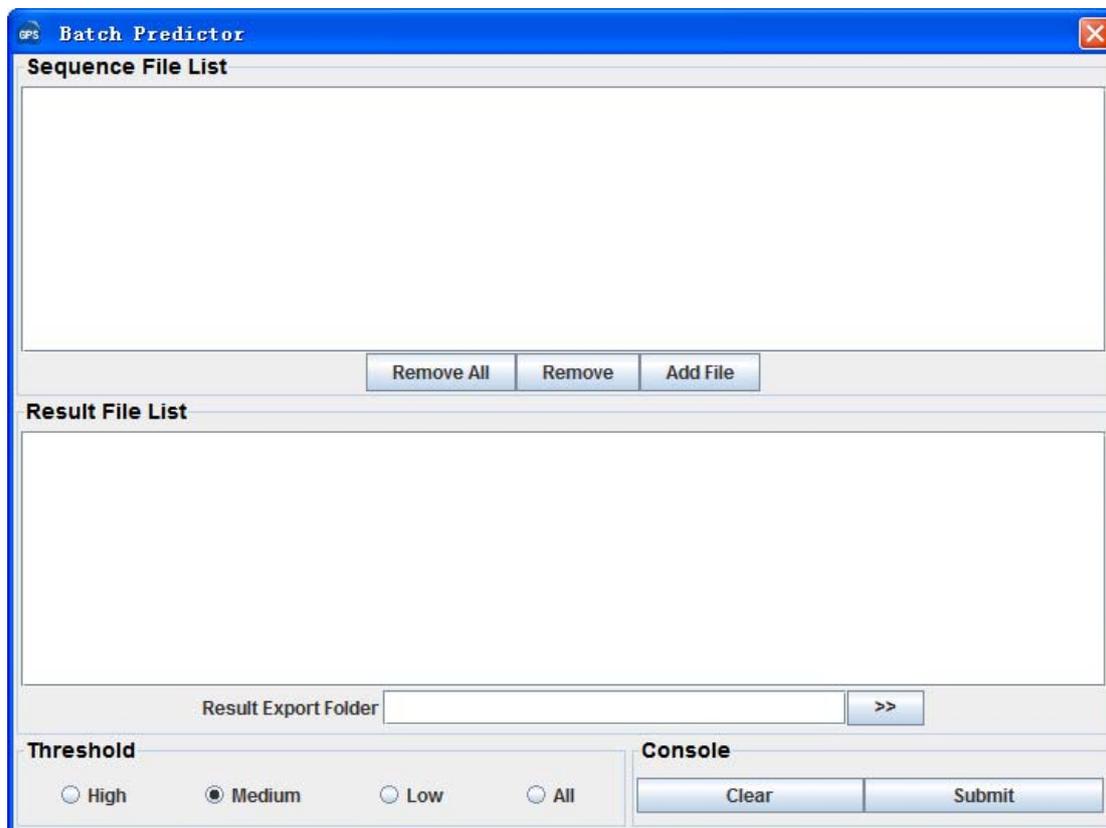
The following steps show you how to use it:

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

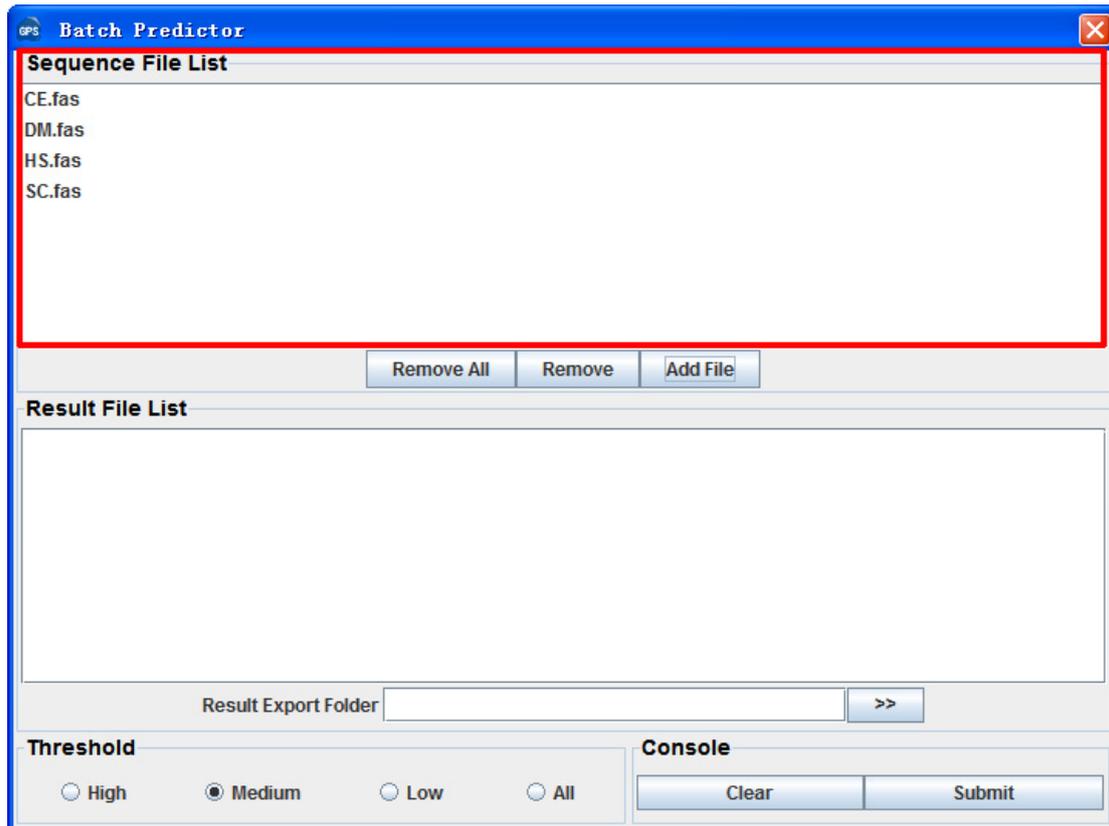
```
>protein1
XXXXXXXXXXXXXXXXX
XXXXXXXXXX
>protein2
XXXXXXXXXXXXXXXXX...
>protein3
XXXXXXXXXXXXXXXXX
...
```

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

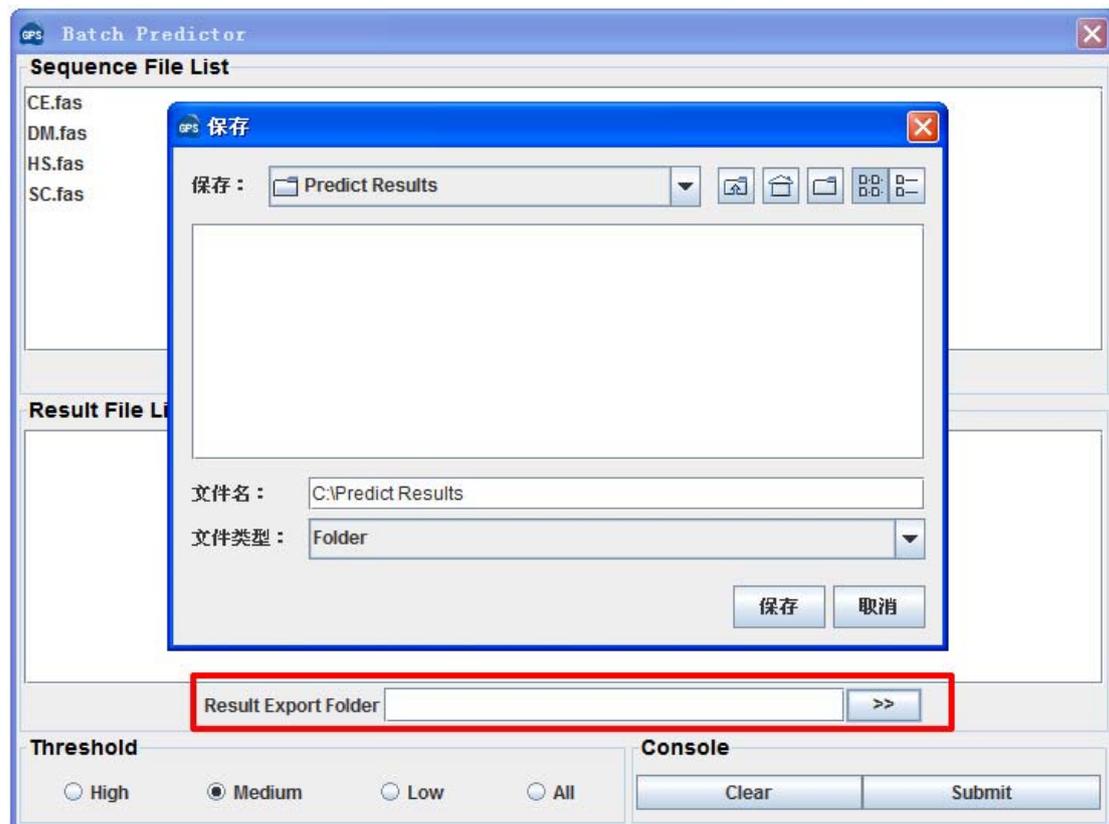
(2) Click on the **Batch Predictor** button and then click on the **Add File** button and add one or more protein sequence files in your hard disk.



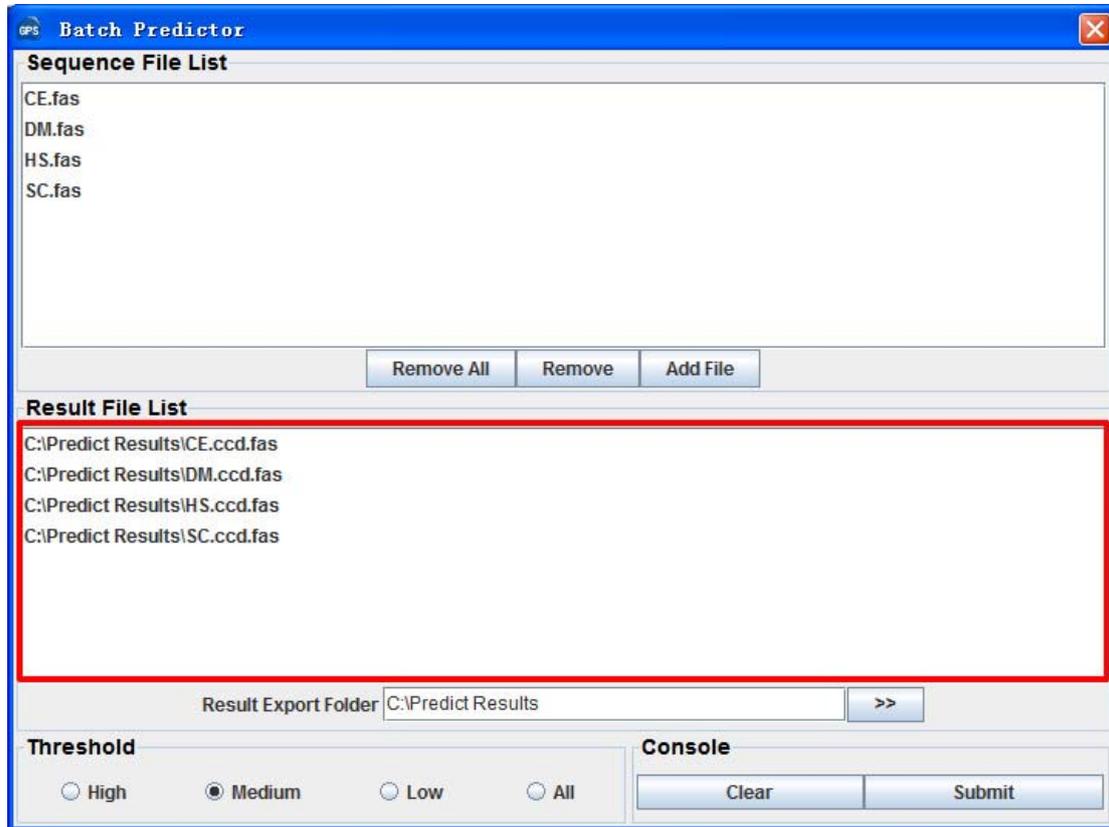
Then the names of added files will be shown in the **Sequence File List**.



(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. Croall, D.E. and Ersfeld, K. (2007) The calpains: modular designs and functional diversity. *Genome Biol*, **8**, 218.
2. Franco, S.J. and Huttenlocher, A. (2005) Regulating cell migration: calpains make the cut. *J Cell Sci*, **118**, 3829-3838.
3. Glading, A., Lauffenburger, D.A. and Wells, A. (2002) Cutting to the chase: calpain proteases in cell motility. *Trends Cell Biol*, **12**, 46-54.
4. Goll, D.E., Thompson, V.F., Li, H., Wei, W. and Cong, J. (2003) The calpain system. *Physiol Rev*, **83**, 731-801.
5. Huang, Y. and Wang, K.K. (2001) The calpain family and human disease. *Trends Mol Med*, **7**, 355-362.
6. Liu, J., Liu, M.C. and Wang, K.K. (2008) Calpain in the CNS: from synaptic function to neurotoxicity. *Sci Signal*, **1**, re1.
7. Zatz, M. and Starling, A. (2005) Calpains and disease. *N Engl J Med*, **352**, 2413-2423.
8. Futai, E., Maeda, T., Sorimachi, H., Kitamoto, K., Ishiura, S. and Suzuki, K. (1999) The protease activity of a calpain-like cysteine protease in *Saccharomyces cerevisiae* is required for alkaline adaptation and sporulation. *Mol Gen Genet*, **260**, 559-568.
9. Hayashi, M., Fukuzawa, T., Sorimachi, H. and Maeda, T. (2005) Constitutive activation of the pH-responsive Rim101 pathway in yeast mutants defective in late steps of the MVB/ESCRT pathway. *Mol Cell Biol*, **25**, 9478-9490.
10. Squier, M.K., Sehnert, A.J., Sellins, K.S., Malkinson, A.M., Takano, E. and Cohen, J.J. (1999) Calpain and calpastatin regulate neutrophil apoptosis. *J Cell Physiol*, **178**, 311-319.
11. Tan, Y., Wu, C., De Veyra, T. and Greer, P.A. (2006) Ubiquitous calpains promote both apoptosis and survival signals in response to different cell death stimuli. *J Biol Chem*, **281**, 17689-17698.
12. Yousefi, S., Perozzo, R., Schmid, I., Ziemiecki, A., Schaffner, T., Scapozza, L., Brunner, T. and Simon, H.U. (2006) Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis. *Nat Cell Biol*, **8**, 1124-1132.
13. Arrington, D.D., Van Vleet, T.R. and Schnellmann, R.G. (2006) Calpain 10: a mitochondrial calpain and its role in calcium-induced mitochondrial dysfunction. *Am J Physiol Cell Physiol*, **291**, C1159-1171.
14. Williams, A., Sarkar, S., Cuddon, P., Ttofi, E.K., Saiki, S., Siddiqi, F.H., Jahreiss, L., Fleming, A., Pask, D., Goldsmith, P. *et al.* (2008) Novel targets for Huntington's disease in an mTOR-independent autophagy pathway. *Nat Chem Biol*, **4**, 295-305.
15. Ottens, A.K., Golden, E.C., Bustamante, L., Hayes, R.L., Denslow, N.D. and Wang, K.K. (2008) Proteolysis of multiple myelin basic protein isoforms after neurotrauma: characterization by mass spectrometry. *J Neurochem*, **104**, 1404-1414.
16. Tompa, P., Buzder-Lantos, P., Tantos, A., Farkas, A., Szilagy, A., Banoczi, Z., Hudecz, F. and Friedrich, P. (2004) On the sequential determinants of calpain cleavage. *J Biol Chem*, **279**, 20775-20785.
17. Cuerrier, D., Moldoveanu, T. and Davies, P.L. (2005) Determination of peptide substrate

- specificity for mu-calpain by a peptide library-based approach: the importance of primed side interactions. *J Biol Chem*, **280**, 40632-40641.
18. Thomas, D.A., Francis, P., Smith, C., Ratcliffe, S., Ede, N.J., Kay, C., Wayne, G., Martin, S.L., Moore, K., Amour, A. *et al.* (2006) A broad-spectrum fluorescence-based peptide library for the rapid identification of protease substrates. *Proteomics*, **6**, 2112-2120.
  19. 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*, **11**, 485-496.
  20. duVerle, D., Takigawa, I., Ono, Y., Sorimachi, H. and Mamitsuka, H. (2010) CaMPDB: a resource for calpain and modulatory proteolysis. *Genome Inform*, **22**, 202-213.
  21. Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L. and Yao, X. (2008) GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics*, **7**, 1598-1608.

## Release Note

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