GPS-ARM Manual

Prediction of APC/C Recognition Motif

Version 1.0
17/10/2011

Author: Zexian Liu, Jian Ren & Yu Xue

Contact:
Zexian Liu, lzx@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-ARM software is available from http://arm.biocuckoo.org

Index

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 the National Basic Research Program (973 project) (2010CB945400), Natural Science Foundation of China (90919001, 31071154, 30900835, 30830036, 91019020, 31171263), and Fundamental Research Funds for the Central Universities (HUST: 2010JC049, 2010ZD018, 2011TS085; SYSU: 11lgzd11).

Introduction

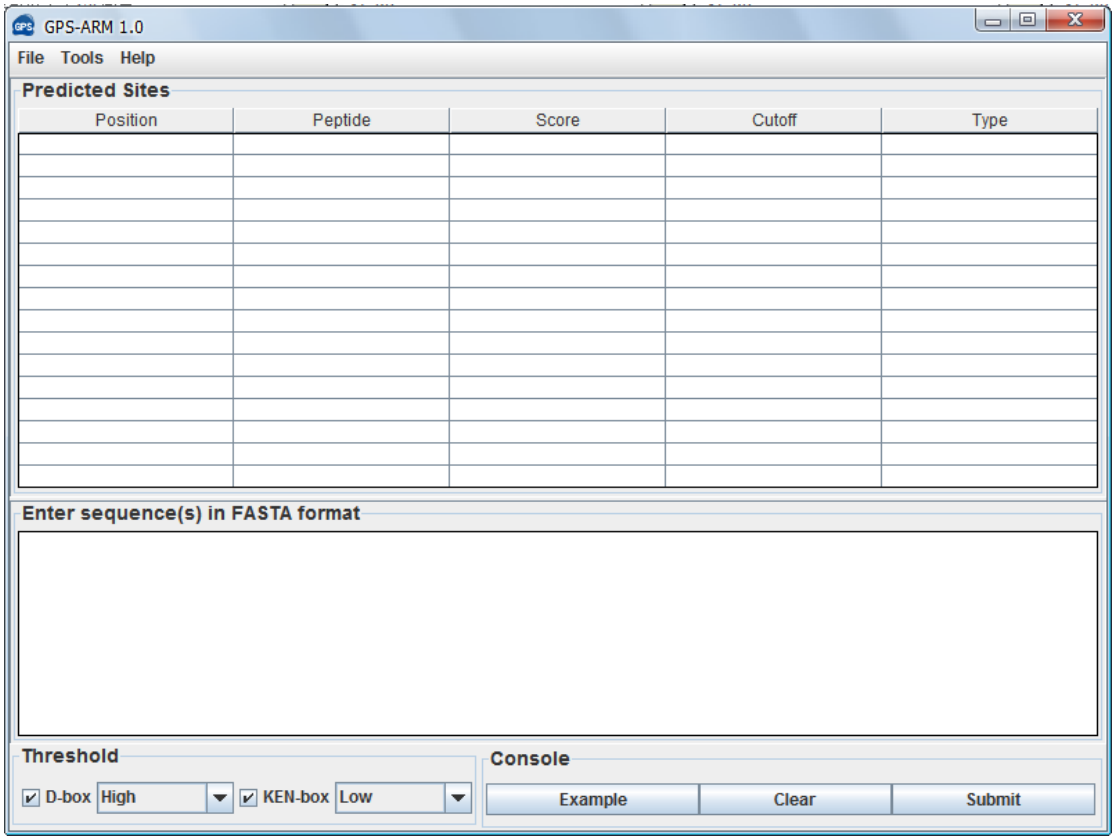
The 2001 Noble Prize in Physiology or Medicine was awarded to Leland Hartwell, Paul Nurse and Timothy Hunt for their seminal discoveries of key regulators of cyclin-dependent kinases (CDKs) in cell cycle and proliferation (1-2). Besides CDK-mediated phosphorylation, cell cycle proteins can also be precisely modulated by other mechanisms, such as ubiquitin-dependent degradation, mainly mediated by the Skp1-cullin-F box (SCF) and the APC/C (3-8). As a high-molecular-mass complex composed of 13 core subunits (3,5), APC/C was firstly identified as an E3 ligase for the degradation of mitotic cyclins (9). Beyond mitosis, APC/C-mediated degradation also plays an important role in regulating Rho GTPase activity (10-11), axon growth (12), cell adhesion (13) and glycolysis (14-15). In this regard, identification of APC/C-specific degradation substrates is fundamental for understanding the molecular mechanisms and regulatory roles of APC/C.

In 1991, Glotzer *et al*. firstly characterized an ennea-peptide (9aa) located at the N-terminus of cyclin B is responsible for its degradation during the mitotic exit (16). Further analyses revealed that the destruction box or D-box follows a minimal consensus of R*XX*L (where *X* is any amino acid), while two co-activators of APC/C, such as Cdh1 and Cdc20, can directly target and interact with the D-box (17-18). Recently, a structural analysis observed that a core APC/C subunit of Apc10 can also interact with the D-box and contribute recognition specificity together with Cdh1 (19). A second APC/C degron, the KEN-box motif with consensus of KEN, can be recognized by Cdh1 and Cdc20 (20-21). Although a number of non-canonical destruction signals were experimentally identified, such as the A-box (QRVL) of Aurora-B kinase (22), the G*X*EN motif in *Xenopus* chromokinesin Kid (Xkid) (23), the CRY-box in mammalian Cdc20 (24) and so on, the D-box and the KEN-box are regarded as the major APC/C recognition motifs (3-6,25).

Conventionally, experimental identification of APC/C targets with a site-directed mutagenesis strategy is time-consuming, labor-intensive and inefficient (16-19). Although many experimental efforts have been devoted for the past two decades, the number of known APC/C substrates is still quite limited. In contrast with the experimental approaches, computational prediction and analysis of the D-box and the KEN-box proteins can generate useful information for the further experimental manipulation. However, it can be expected that the prediction with two loosely defined motifs of R*XX*L and KEN will generate too much false positive hits, while more sophisticated approaches should be developed. For example, Michael *et al*. predicted 25 KEN-box proteins as potentially new APC/C targets, through an enrichment of cell cycle Gene Ontology (GO) terms combined with native disorder prediction and motif conservation information (26). Because the filters were quite

stringent and the number of predicted candidates is too small, further experimental validations still remain to be performed. In this regard, development a general and efficient predictor for D-boxes and KEN-boxes is an urgent demand.

In this work, we developed a novel software package of GPS-ARM for the prediction of potentially functional D-boxes and KEN-boxes in APC/C substrates. The experimental data was collected from the scientific literature, while s previously developed algorithm of GPS 2.2 was adopted for training and predicting. By extensive evaluations, the prediction performance of GPS-ARM is promising and much better than simple short motifs. With the powerful tool, we systematically analyzed the functional abundance and diversity of D-box and KEN-box proteins in *H. sapiens*. From the results, we observed that KEN-box proteins are more implicated in cell cycle and mitosis, while both D-box and KEN-box proteins can also regulate a variety of biological processes beyond cell cycle. Moreover, with additional co-localization information, we predicted hundreds of mitosis-specific D-box and KEN-box proteins in eukaryotes with high confidence. Taken together, the prediction and analysis results can be helpful for further experimental consideration, while the GPS-ARM can serve as a useful program for experimentalists. The online service and local packages of GPS-AMD 1.0 were implemented in JAVA and could be freely accessed for academic researches at: http://arm.biocuckoo.org.

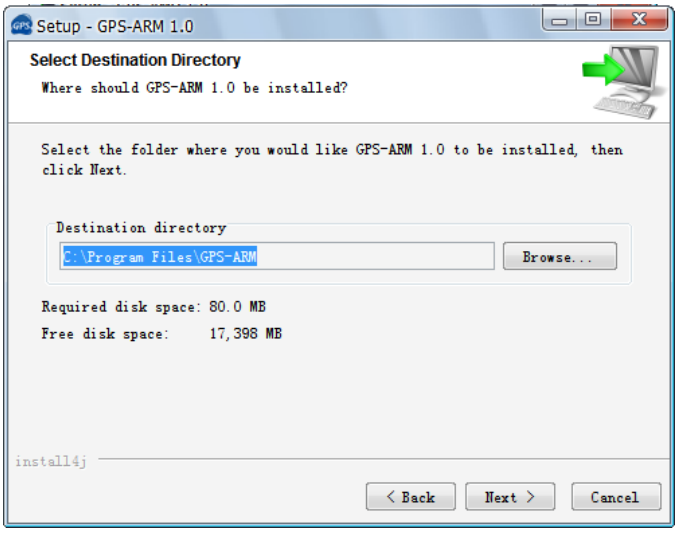


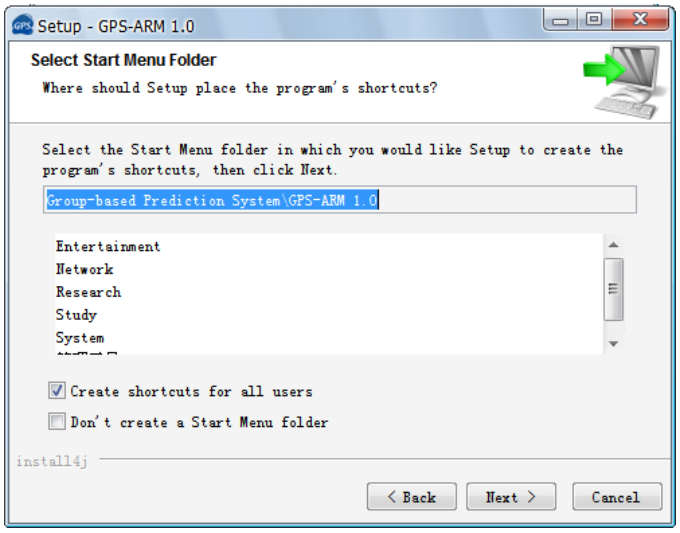
GPS-ARM 1.0 User Interface

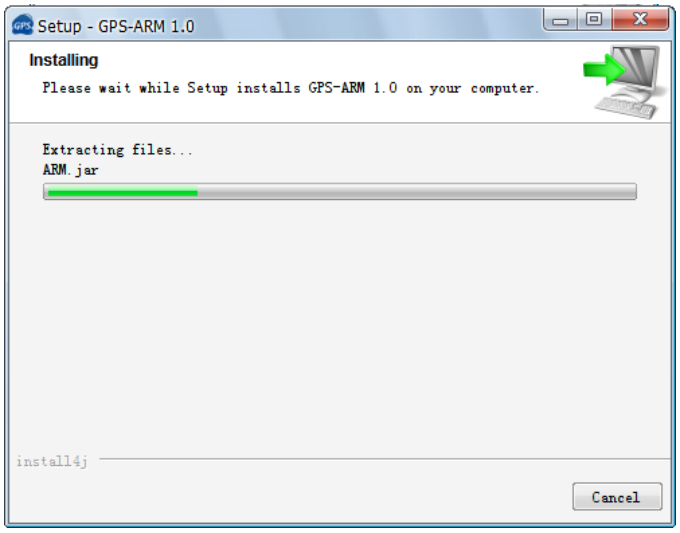
Download & Installation

The GPS-ARM 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://arm.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:







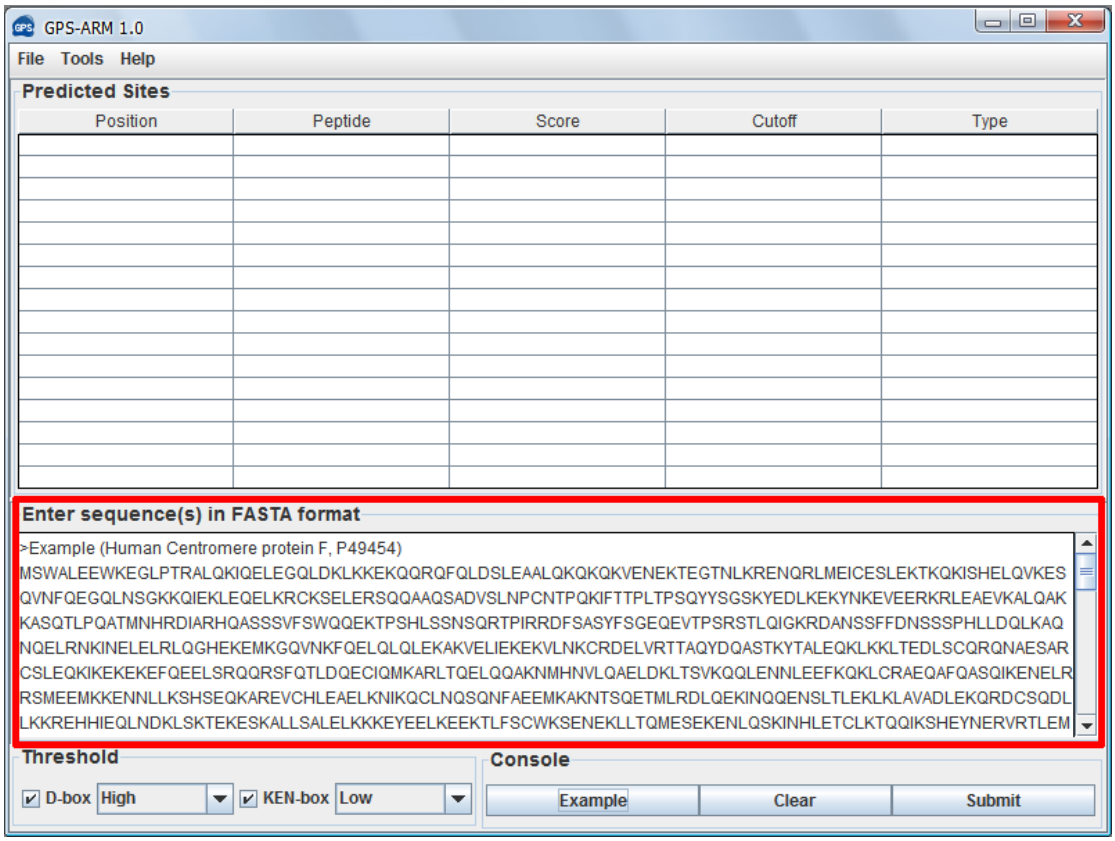
Finally, please click on the **Finish** button to complete the setup program.

Prediction of APC/C Recognition Motif

1. A single protein sequence in FASTA format

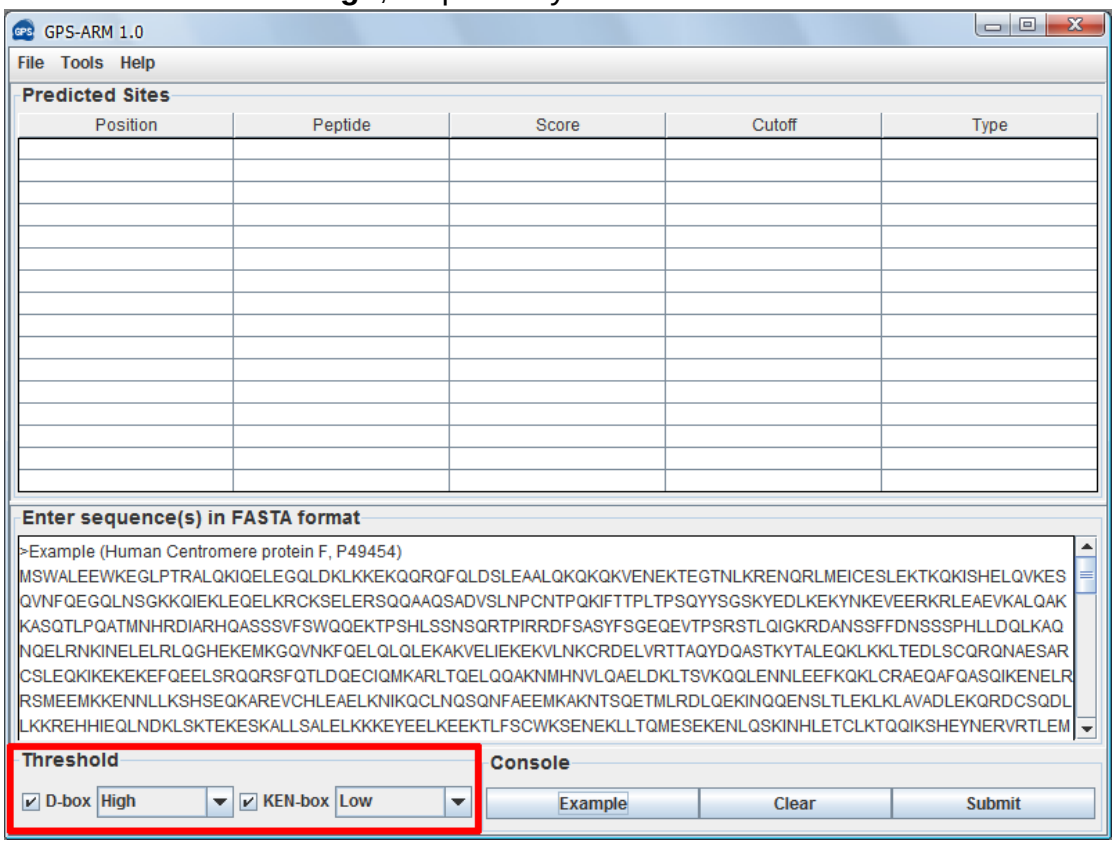
The following steps show you how to use the GPS-ARM 1.0 to predict D-box or KEN-box for a single protein sequence in FASTA format.

(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-ARM 1.0

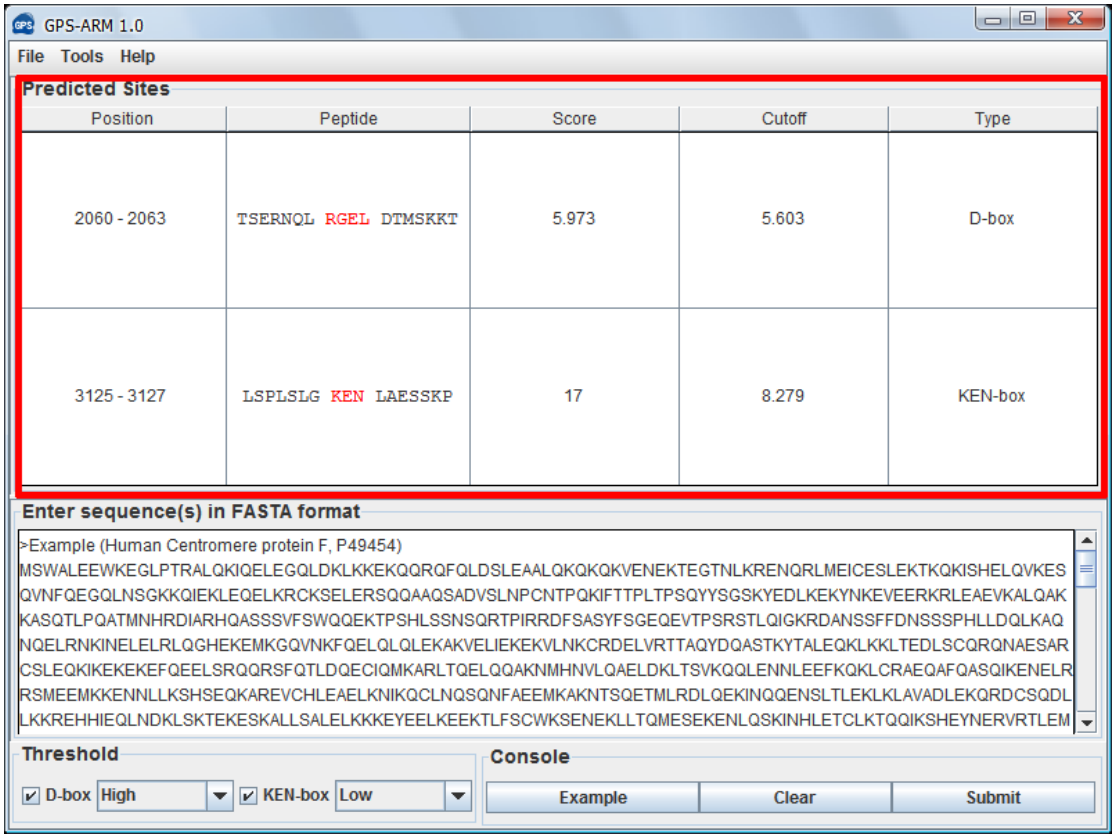


Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

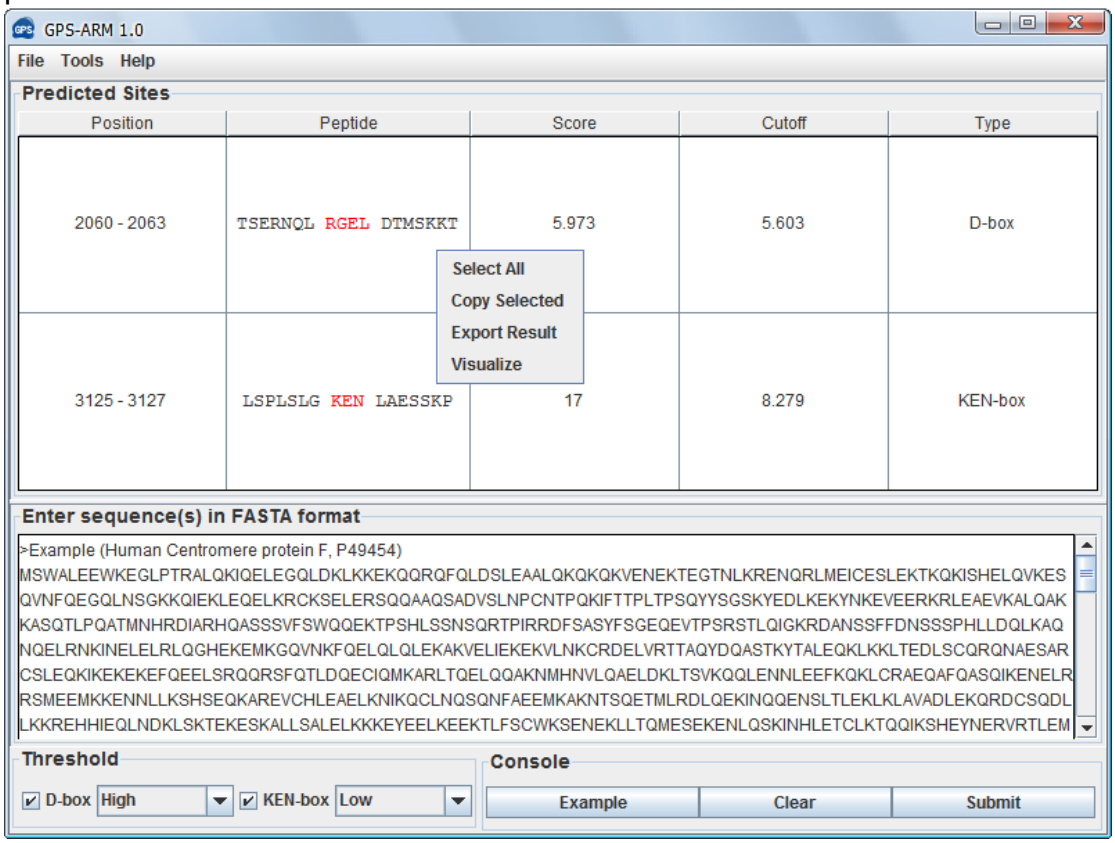
(2) Choose a **Threshold** that you need, the default cut-off for D-box and KEN-box is **Low** and **High**, respectively.



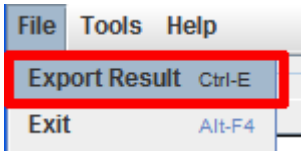
(3) Click on the **Submit** button, and then the predicted KEN-box or D-box will be shown.



(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.



Again, you can also click the "**Export Result**" in **File** menu to export the results.

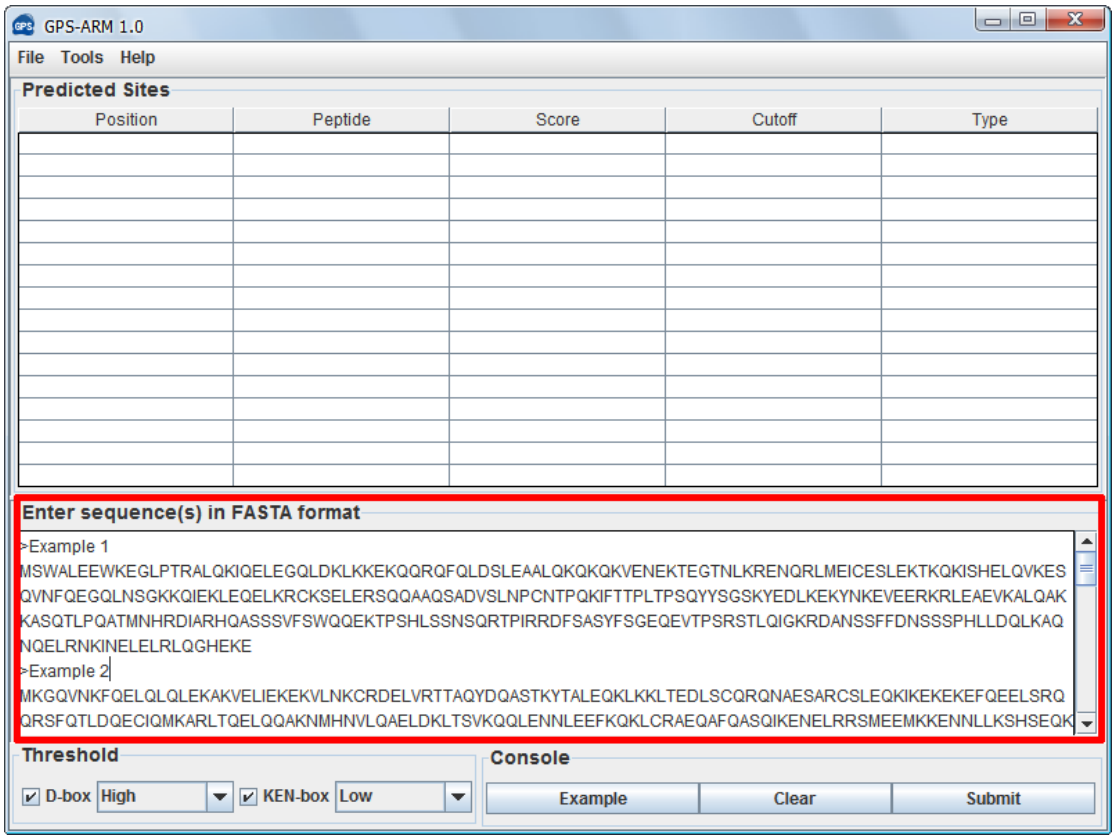


2. Multiple protein sequences in FASTA format

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

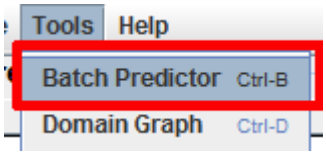
A. Input the sequences into text form directly. (Num. of Seq ≤ 2,000)

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-ARM 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.

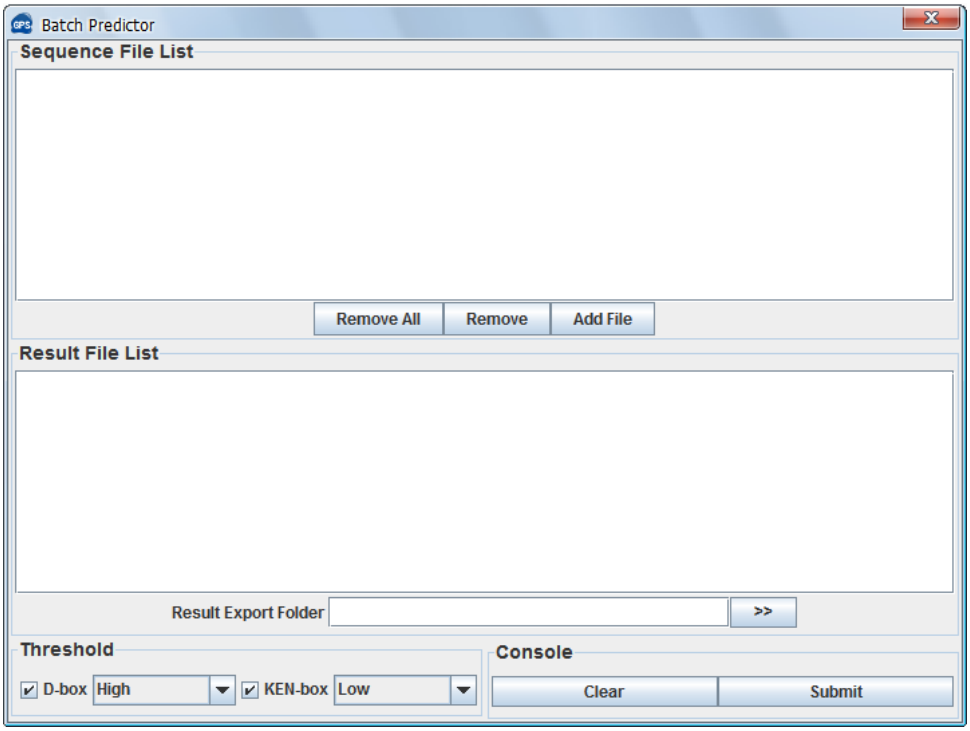


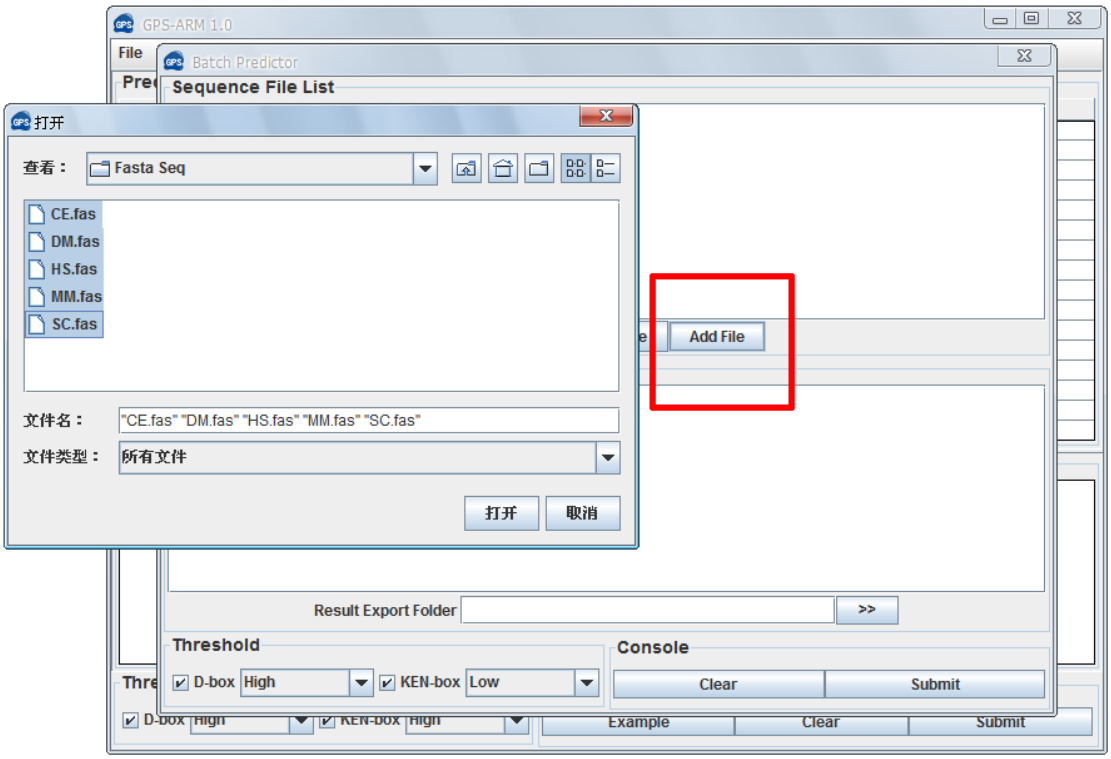
The following steps show you how to use it:

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

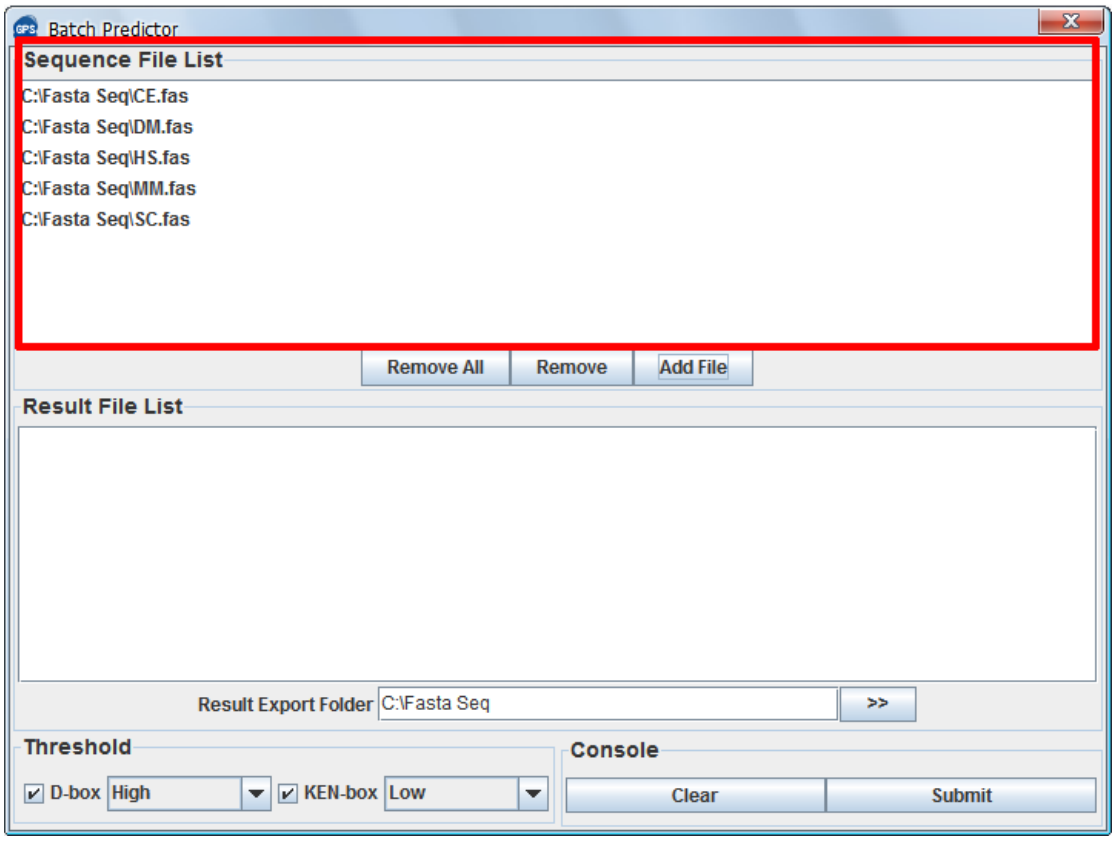
>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXX
...
Most importantly, the name of each protein should be presented.

(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.

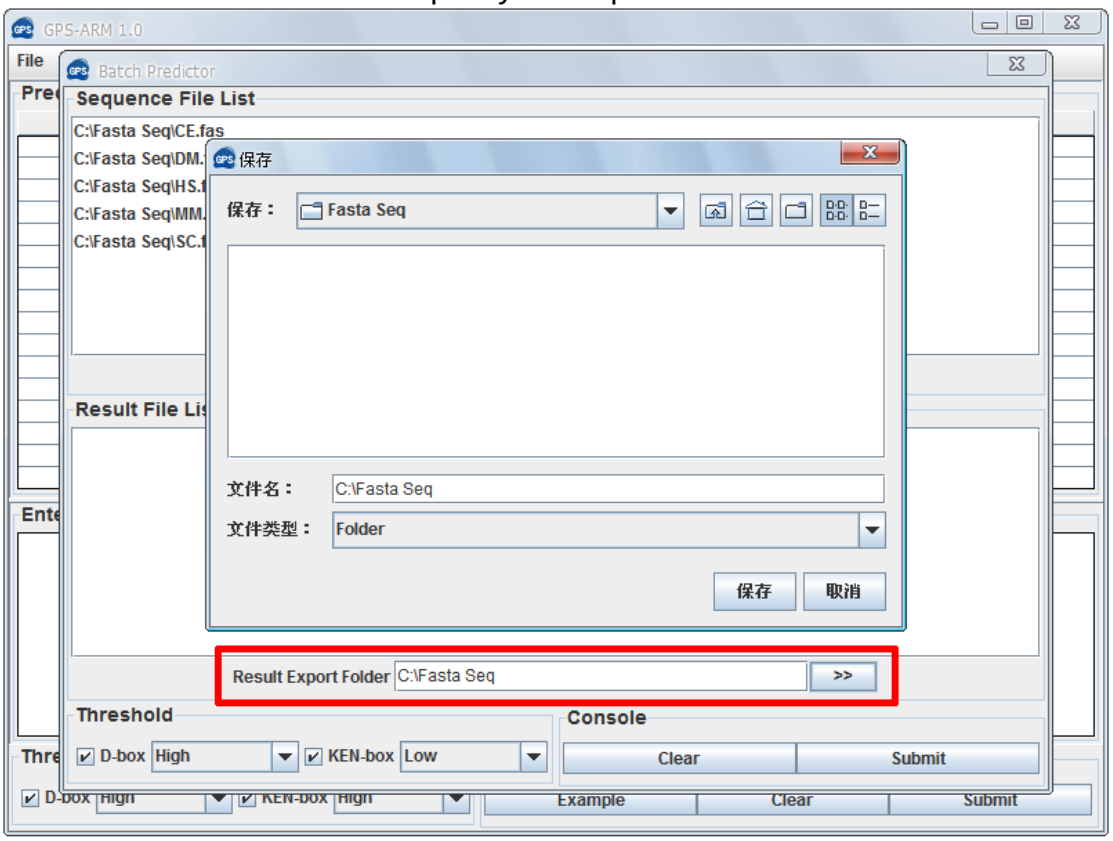




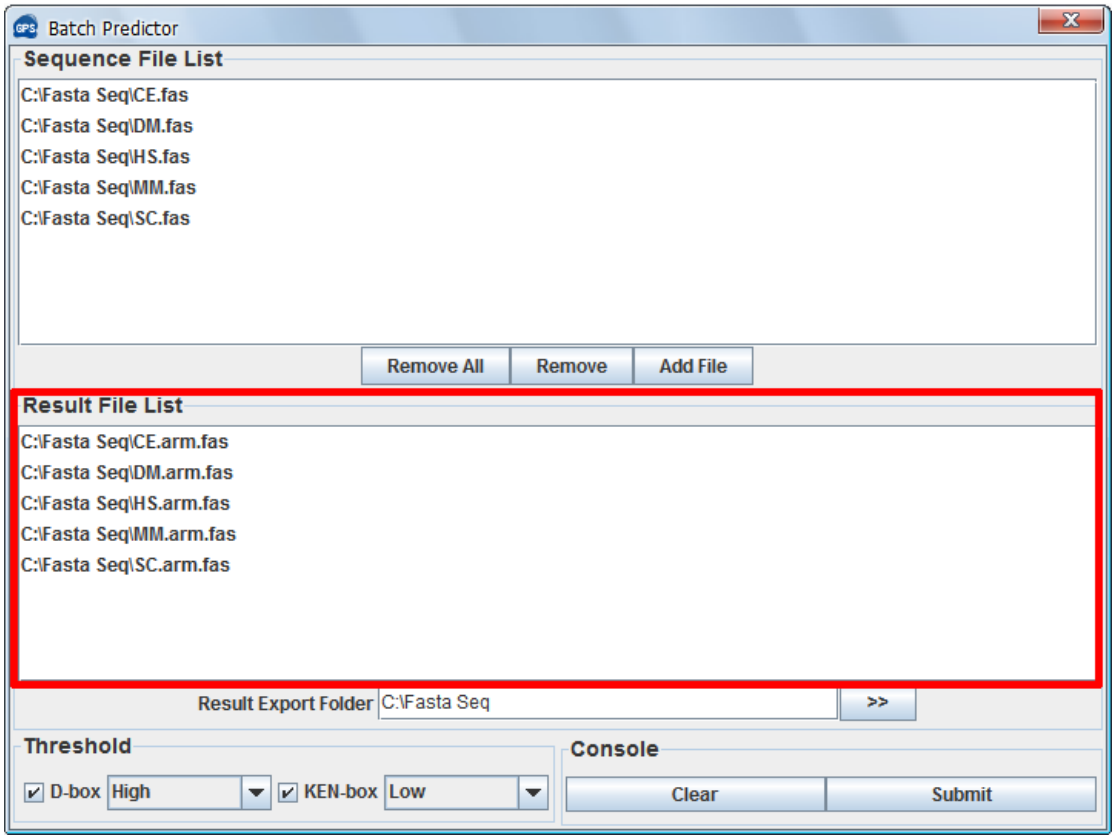
Then the names of added files will be shown in the **Sequence File List**.



(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. (2001) Milestones in cell division. *Nat Cell Biol*, **3**, E265.
2. Nasmyth, K. (2001) A prize for proliferation. *Cell*, **107**, 689-701.
3. Barford, D. (2011) Structure, function and mechanism of the anaphase promoting complex (APC/C). *Q Rev Biophys*, **44**, 153-190.
4. Thornton, B.R. and Toczyski, D.P. (2006) Precise destruction: an emerging picture of the APC. *Genes Dev*, **20**, 3069-3078.
5. Peters, J.M. (2006) The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat Rev Mol Cell Biol*, **7**, 644-656.
6. Fang, G., Yu, H. and Kirschner, M.W. (1999) Control of mitotic transitions by the anaphase-promoting complex. *Philos Trans R Soc Lond B Biol Sci*, **354**, 1583-1590.
7. Pagano, M. (1997) Cell cycle regulation by the ubiquitin pathway. *FASEB J*, **11**, 1067-1075.
8. Peters, J.M. (1998) SCF and APC: the Yin and Yang of cell cycle regulated proteolysis. *Curr Opin Cell Biol*, **10**, 759-768.
9. Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F.C., Ruderman, J.V. and Hershko, A. (1995) The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol Biol Cell*, **6**, 185-197.
10. Liot, C., Seguin, L., Siret, A., Crouin, C., Schmidt, S. and Bertoglio, J. (2011) APC Mediates Degradation of the Oncogenic Rho-GEF Ect2 after Mitosis. *PLoS One*, **6**, e23676.
11. Naoe, H., Araki, K., Nagano, O., Kobayashi, Y., Ishizawa, J., Chiyoda, T., Shimizu, T., Yamamura, K., Sasaki, Y., Saya, H. *et al.* (2010) The anaphase-promoting complex/cyclosome activator Cdh1 modulates Rho GTPase by targeting p190 RhoGAP for degradation. *Mol Cell Biol*, **30**, 3994-4005.
12. Kim, A.H. and Bonni, A. (2007) Thinking within the D box: initial identification of Cdh1-APC substrates in the nervous system. *Mol Cell Neurosci*, **34**, 281-287.
13. Silies, M. and Klambt, C. (2010) APC/C(Fzr/Cdh1)-dependent regulation of cell adhesion controls glial migration in the Drosophila PNS. *Nat Neurosci*, **13**, 1357-1364.
14. Tudzarova, S., Colombo, S.L., Stoeber, K., Carcamo, S., Williams, G.H. and Moncada, S. (2011) Two ubiquitin ligases, APC/C-Cdh1 and SKP1-CUL1-F (SCF)-beta-TrCP, sequentially regulate glycolysis during the cell cycle. *Proc Natl Acad Sci U S A*, **108**, 5278-5283.
15. Colombo, S.L., Palacios-Callender, M., Frakich, N., De Leon, J., Schmitt, C.A., Boorn, L., Davis, N. and Moncada, S. (2010) Anaphase-promoting complex/cyclosome-Cdh1 coordinates glycolysis and glutaminolysis with transition to S phase in human T lymphocytes. *Proc Natl Acad Sci U S A*, **107**, 18868-18873.
16. Glotzer, M., Murray, A.W. and Kirschner, M.W. (1991) Cyclin is degraded by the ubiquitin pathway. *Nature*, **349**, 132-138.
17. Owens, T.J. and Hoyt, M.A. (2005) The D box asserts itself. *Mol Cell*, **18**, 611-612.
18. King, R.W., Glotzer, M. and Kirschner, M.W. (1996) Mutagenic analysis of the destruction signal of mitotic cyclins and structural characterization of ubiquitinated intermediates. *Mol Biol Cell*, **7**, 1343-1357.
19. da Fonseca, P.C., Kong, E.H., Zhang, Z., Schreiber, A., Williams, M.A., Morris, E.P. and

Barford, D. (2011) Structures of APC/C(Cdh1) with substrates identify Cdh1 and Apc10 as the D-box co-receptor. *Nature*, **470**, 274-278.

20. Gurden, M.D., Holland, A.J., van Zon, W., Tighe, A., Vergnolle, M.A., Andres, D.A., Spielmann, H.P., Malumbres, M., Wolthuis, R.M., Cleveland, D.W. *et al.* (2010) Cdc20 is required for the post-anaphase, KEN-dependent degradation of centromere protein F. *J Cell Sci*, **123**, 321-330.
21. Pfleger, C.M. and Kirschner, M.W. (2000) The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev*, **14**, 655-665.
22. Nguyen, H.G., Chinnappan, D., Urano, T. and Ravid, K. (2005) Mechanism of Aurora-B degradation and its dependency on intact KEN and A-boxes: identification of an aneuploidy-promoting property. *Mol Cell Biol*, **25**, 4977-4992.
23. Castro, A., Vigneron, S., Bernis, C., Labbe, J.C. and Lorca, T. (2003) Xkid is degraded in a D-box, KEN-box, and A-box-independent pathway. *Mol Cell Biol*, **23**, 4126-4138.
24. Reis, A., Levasseur, M., Chang, H.Y., Elliott, D.J. and Jones, K.T. (2006) The CRY box: a second APCcdh1-dependent degron in mammalian cdc20. *EMBO Rep*, **7**, 1040-1045.
25. Pfleger, C.M., Lee, E. and Kirschner, M.W. (2001) Substrate recognition by the Cdc20 and Cdh1 components of the anaphase-promoting complex. *Genes Dev*, **15**, 2396-2407.
26. Michael, S., Trave, G., Ramu, C., Chica, C. and Gibson, T.J. (2008) Discovery of candidate KEN-box motifs using cell cycle keyword enrichment combined with native disorder prediction and motif conservation. *Bioinformatics*, **24**, 453-457.

Release Note

1. December 26, 2011, the online service and the local stand-alone packages of GPS-AMD 1.0 were released.
2. July 20, 2011, the online service and the local stand-alone packages of GPS-AMD 1.0 were revised.
3. September 20, 2011, the online service and the local stand-alone packages of GPS-AMD 1.0 were revised.
4. October 17, 2011, the software was renamed as GPS-ARM for “APC/C Recognition Motif”. The online service and the local stand-alone packages of GPS-ARM 1.0 were released.