GPS-TSP Manual

Prediction of Tyrosine Sulfation Sites

Version 1.0.0
11/27/2014

Author: Zhicheng Pan, Zexian Liu, Jian Ren & Yu Xue

Contact:
Zhicheng Pan, zhichengpan@hust.edu.cn
Zexian Liu, lzx.bioinfo@gmail.com
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-TSP software is available from http://tsp.biocuckoo.org

Index

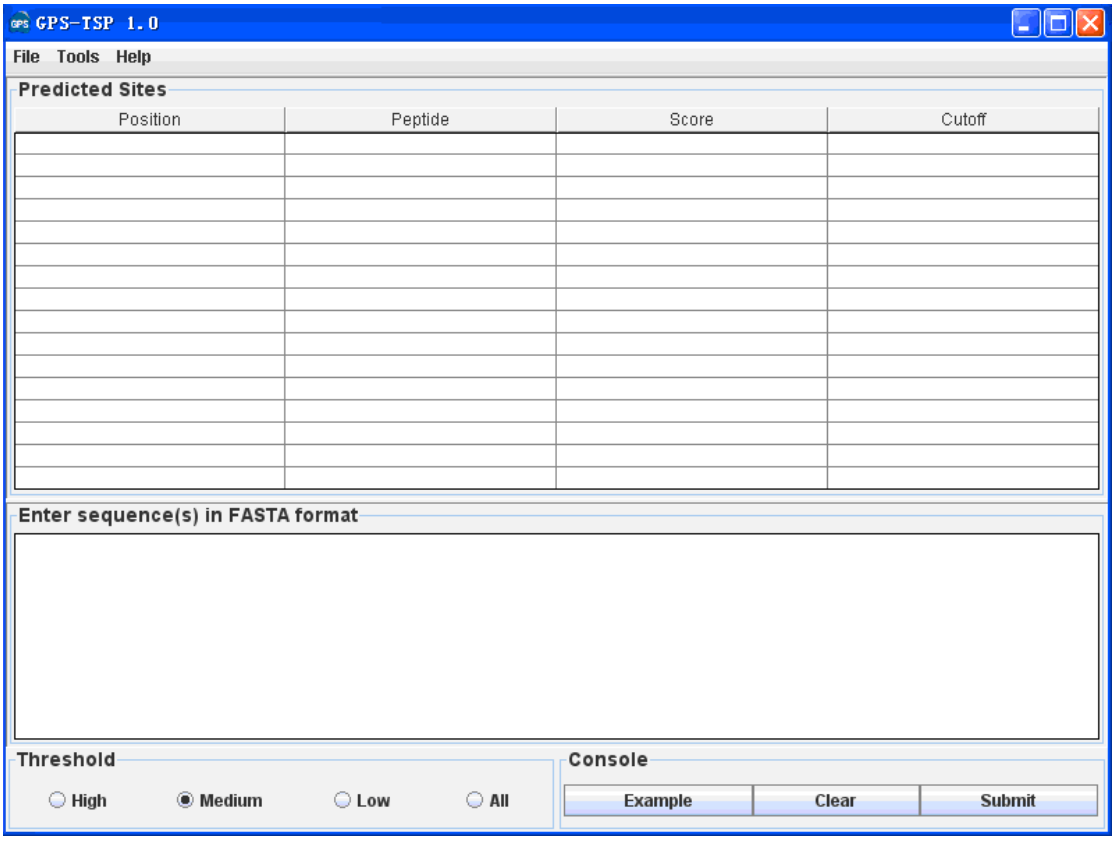
Statement

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

Tyrosine sulfation is a ubiquitous PTM that predominantly modifies trans-membrane and extracellular proteins in the secretory pathway (1-7), and plays an important role in regulating chemotaxis (4-6), inflammatory response (8), and cell adhesion (5). In animals, the sulfation is catalyzed by two closely related tyrosylprotein sulfotransferases (TPST-1 and TPST-2) (4,6), while a non-homologous AtTPST through convergent evolution has been identified in plants (9). In contrast with labor-intensive and time-consuming experimental assays, computational prediction of sulfation sites in proteins has become an efficient approach to generate useful information for further experimental verification. Previous studies suggested the short linear motif (SLM) around the sulfation site is informative, and raised several consensus determinants for the prediction (1-3,6). In 2002, Monigatti *et al*. presented the first online predictor of Sulfinator with four distinct Hidden Markov Models (HMMs) (10). With a Support Vector Machines (SVMs) classifier, Chang *et al*. developed SulfoSite for the prediction of sulfation sites (11). Recently, the algorithms of random forest (12) and nearest neighbor (13) were also adopted for predicting sulfation respectively, although the applicable tools were not released.

In this work, we manually collected 273 experimentally indentified protein sulfation sites in 171 unique proteins from scientific literature. A previously self-developed GPS (Group-based Prediction System) algorithm was employed with great improvement. We calculated the leave-one-out validation and 4-, 6-, 8-, 10-fold cross-validations to evaluate the prediction performance and system robustness. The leave-one-out validation result is accuracy (*Ac*) of 90.23%, sensitivity (*Sn*) of 89.60%, and specificity (*Sp*) of 90.36%. The online service and stand-alone packages of GPS-TSP 1.0 were implemented in JAVA 1.4.2 and freely available at: http://tsp.biocuckoo.org/.

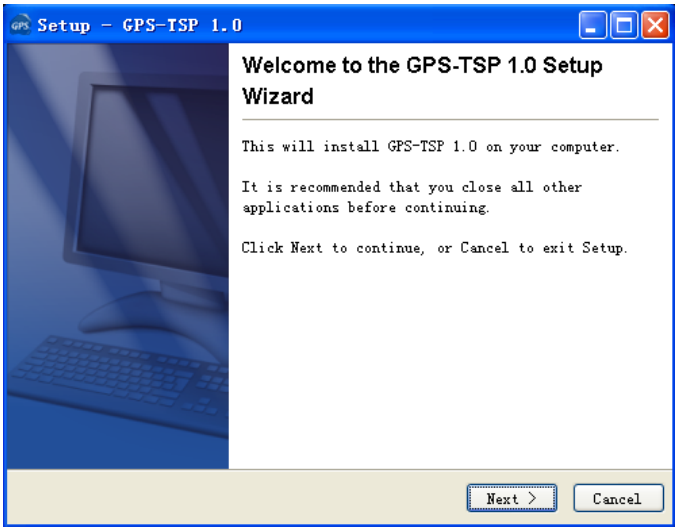


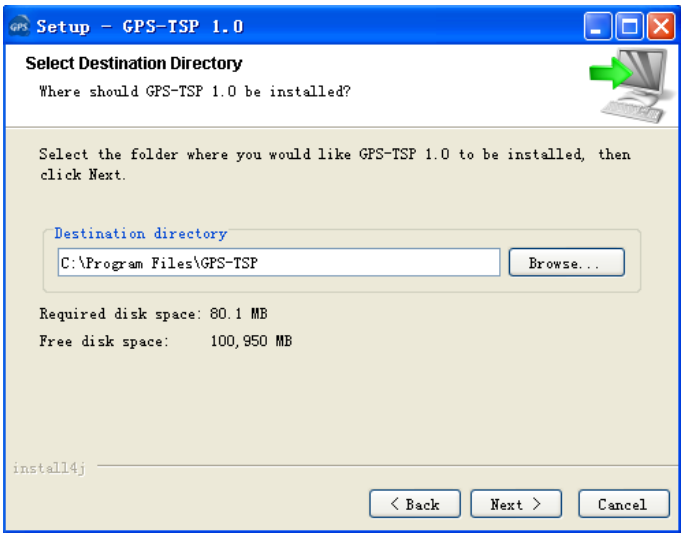
GPS-TSP 1.0 User Interface

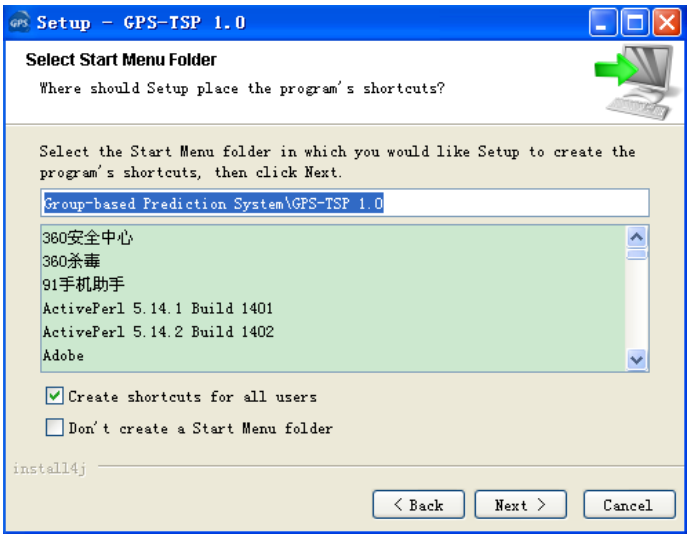
Download & Installation

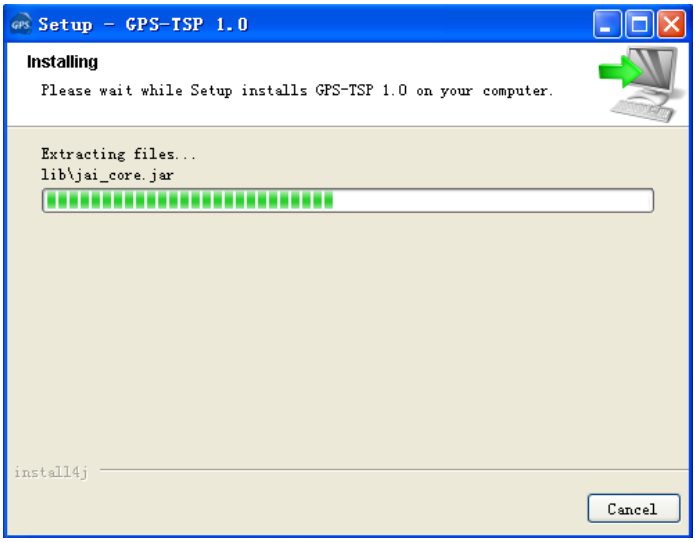
The GPS-TSP 1.0 was implemented in JAVA (J2SE), and could support three major Operating Systems (OS), including Windows, Linux/Unix or Mac OS X systems. Both of online web service and local stand-alone packages are available from: http://tsp.biocuckoo.org/ . We recommend that users could download the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

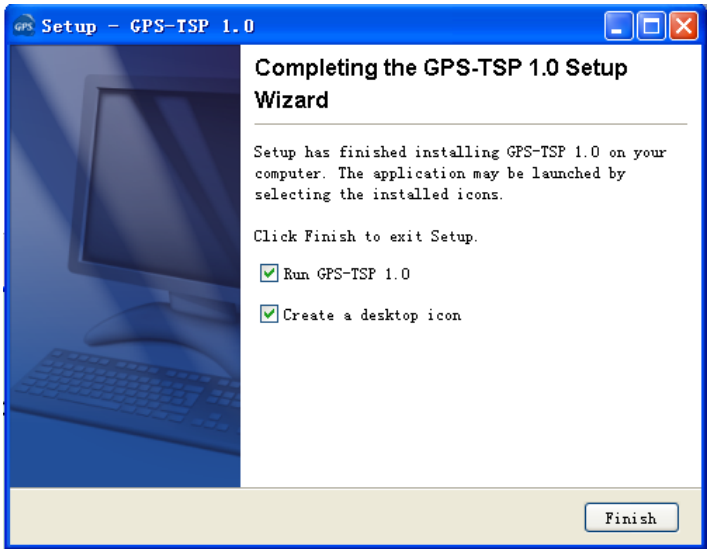
Please choose the proper package to download. After downloading, please double-click on the software package to begin installation, following the user prompts through the installation. And snapshots of the setup program for windows are shown below:











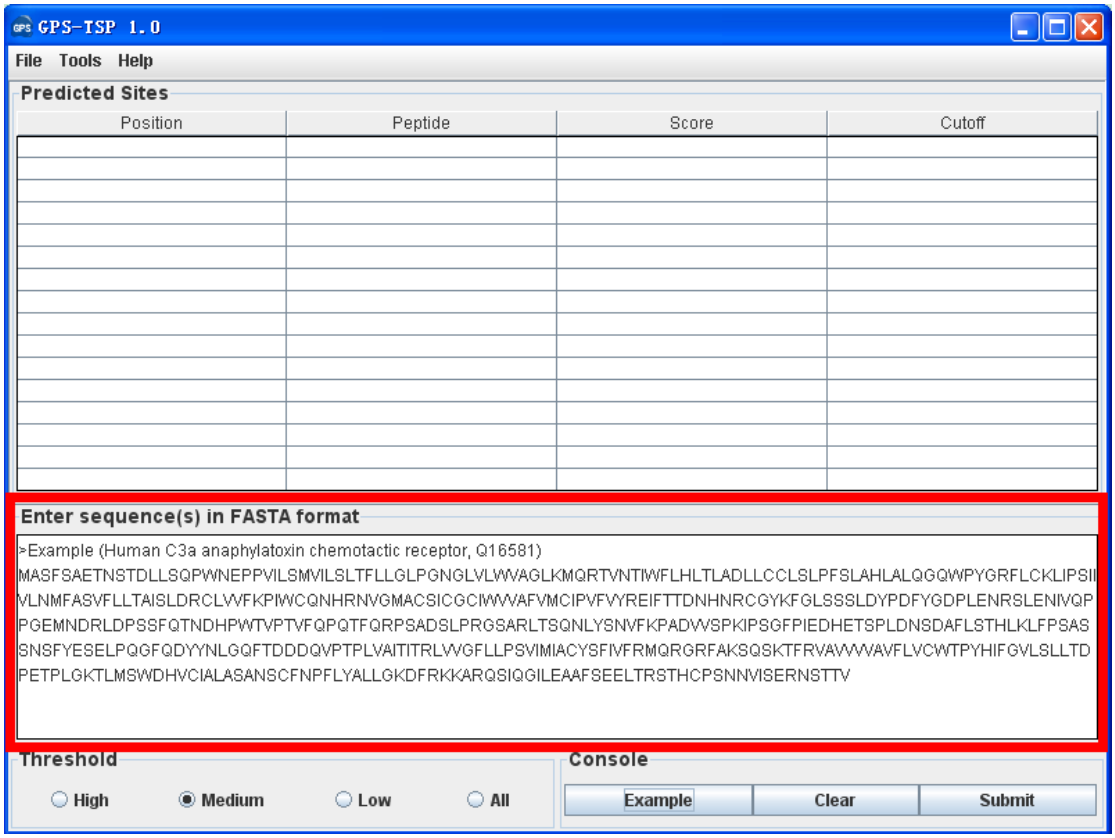
Finally, please click on the **Finish** button to complete the setup program.

Prediction of Tyrosine Sulfation Sites

1. A single protein sequence in FASTA format

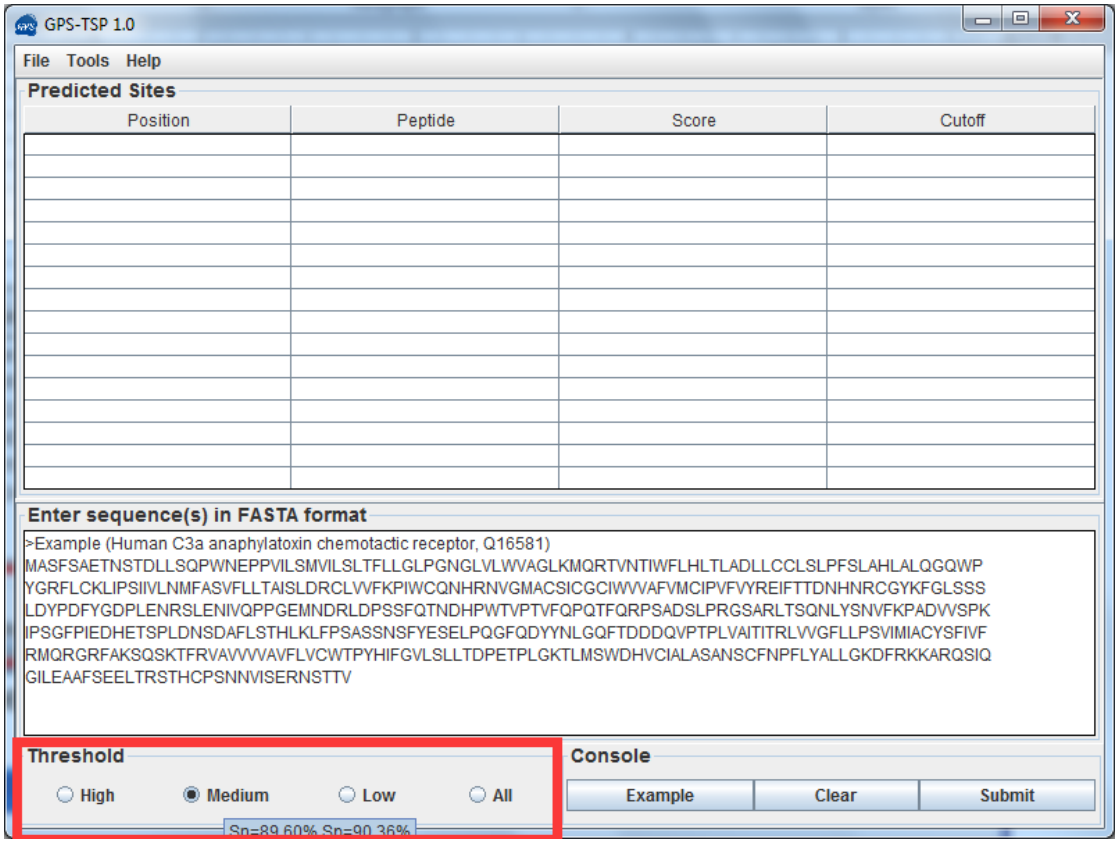
The following steps show you how to use the GPS-TSP 1.0 to predict tyrosine sulfation sites for a single protein sequence in FASTA format.

(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-TSP 1.0

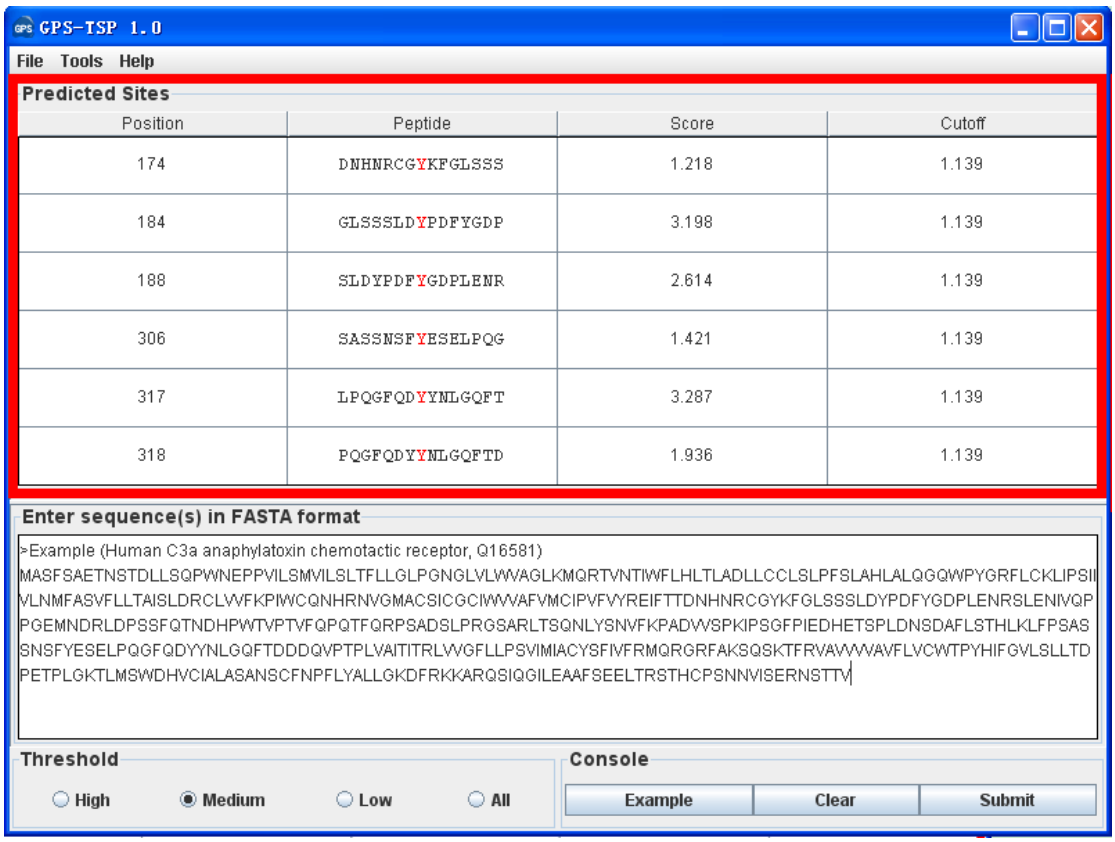


Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

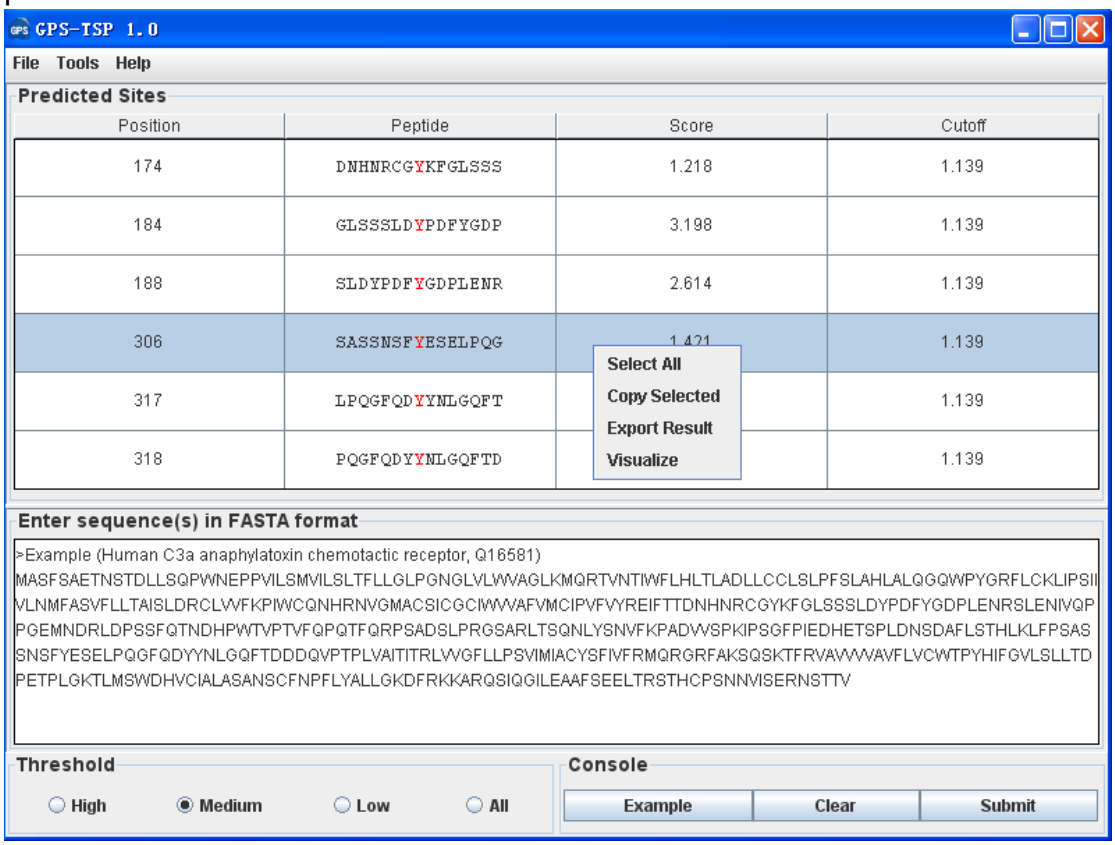
(2) Choose a **Threshold** that you need, the default cut-off is **Medium**.



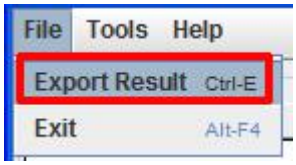
(3) Click on the **Submit** button, then the predicted tyrosine sulfation sites will be shown.



(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, e.g., 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 Prediction**" in **File** menu to export the results.

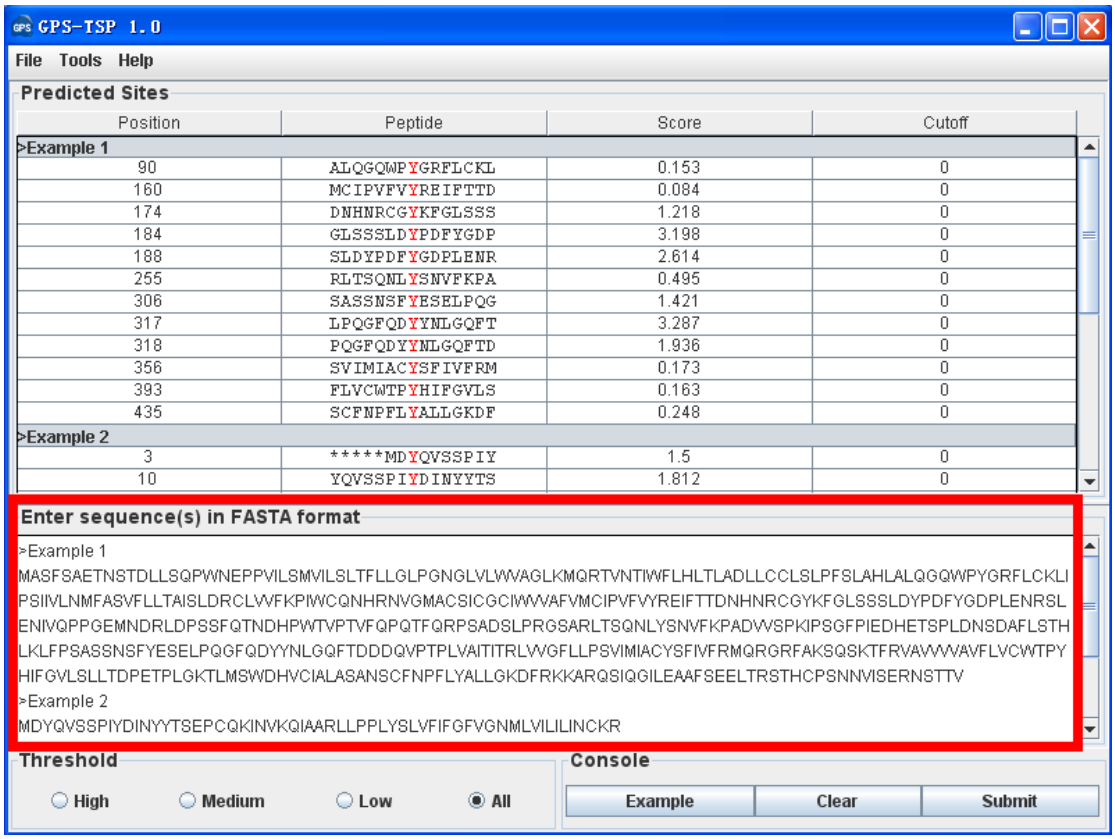


2. Multiple protein sequences in FASTA format

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

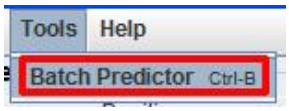
A. Input the sequences into text form directly. (Num. of Seq ≤ 2,000)

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-TSP 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.

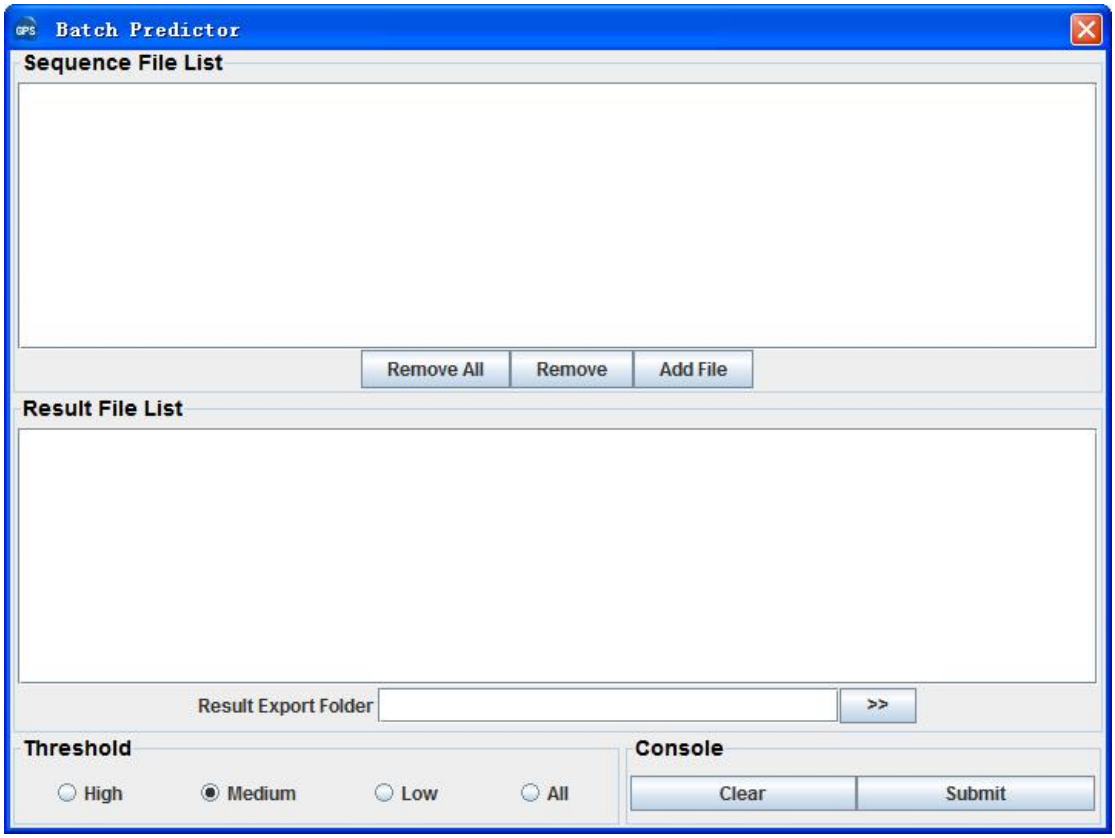


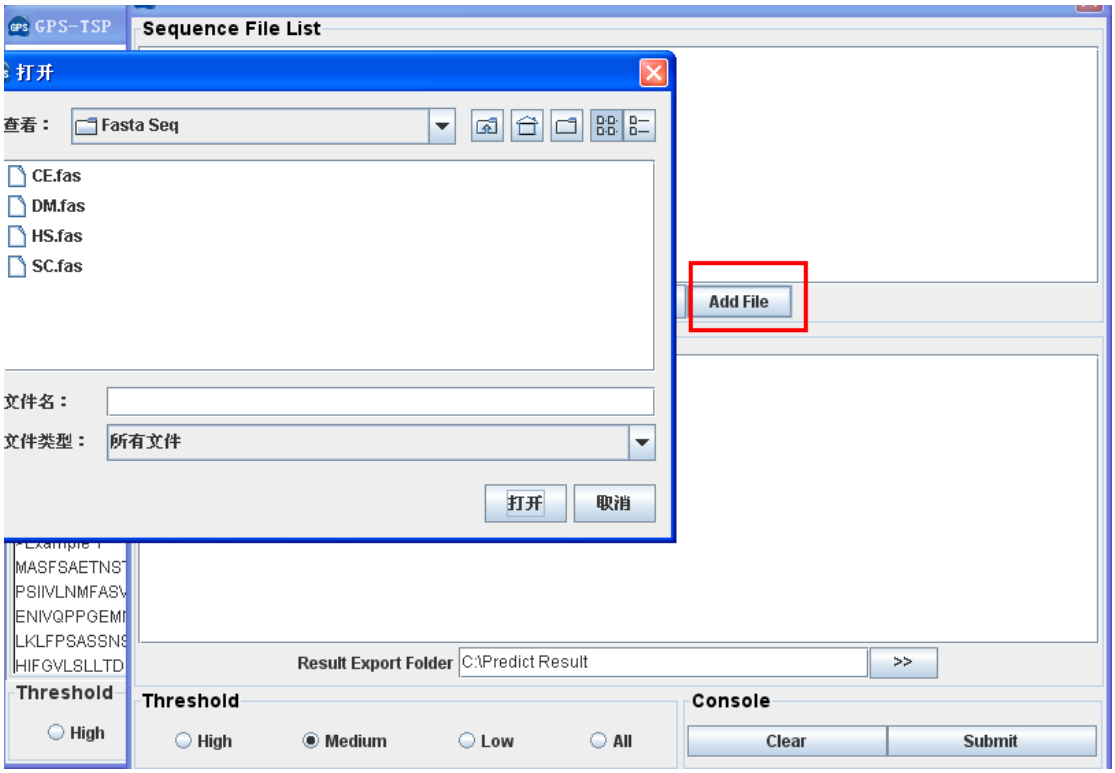
The following steps show you how to use it:

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

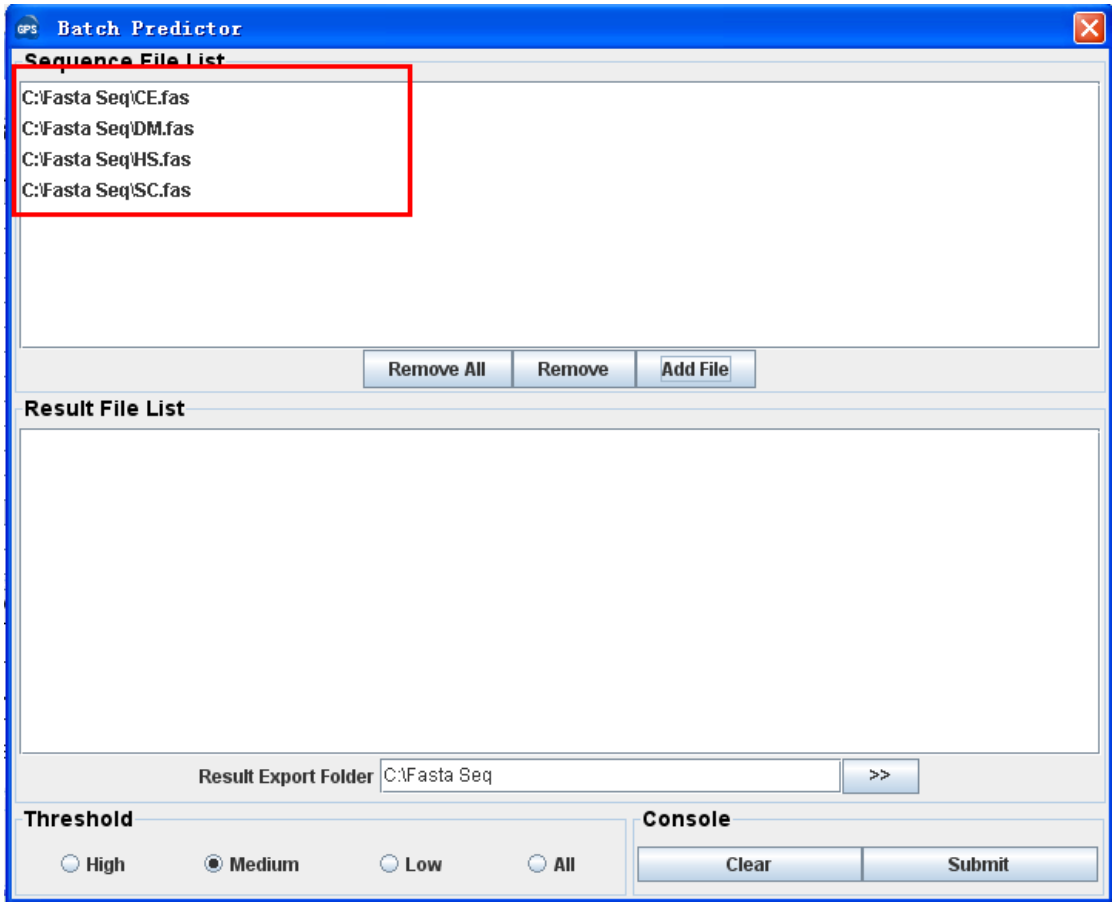
>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXXX
...
Most importantly, the name of each protein should be presented.

(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.

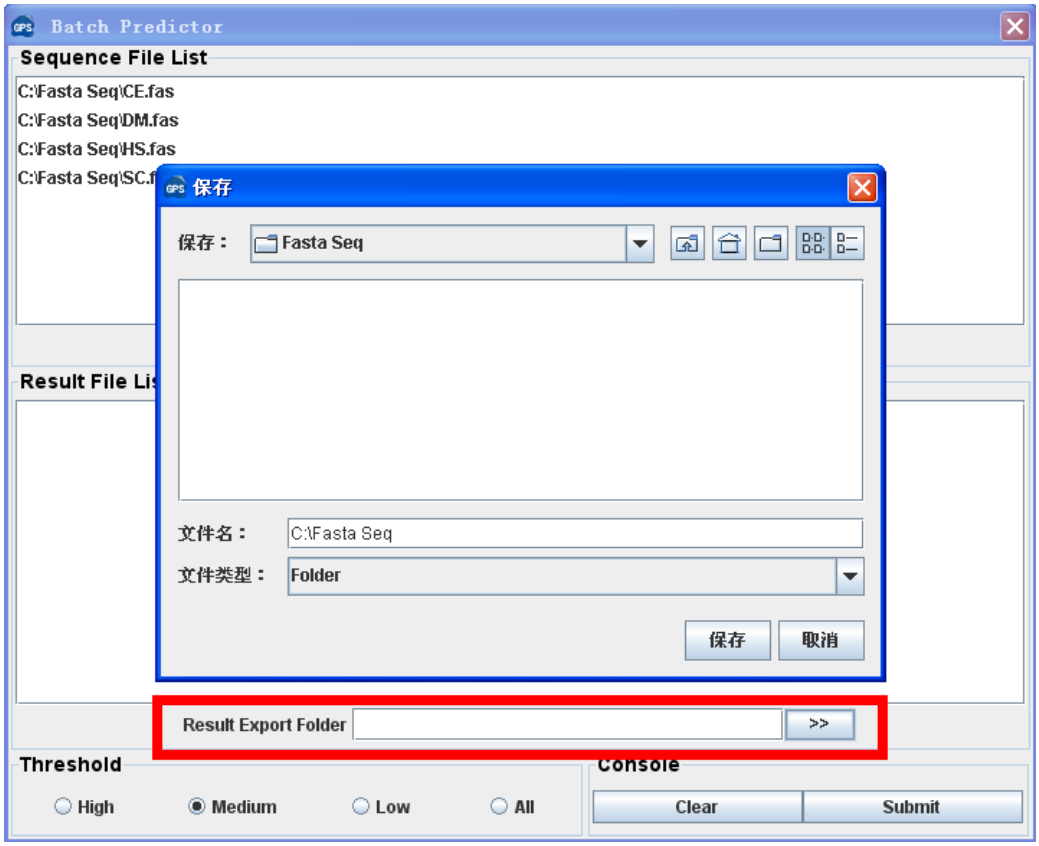




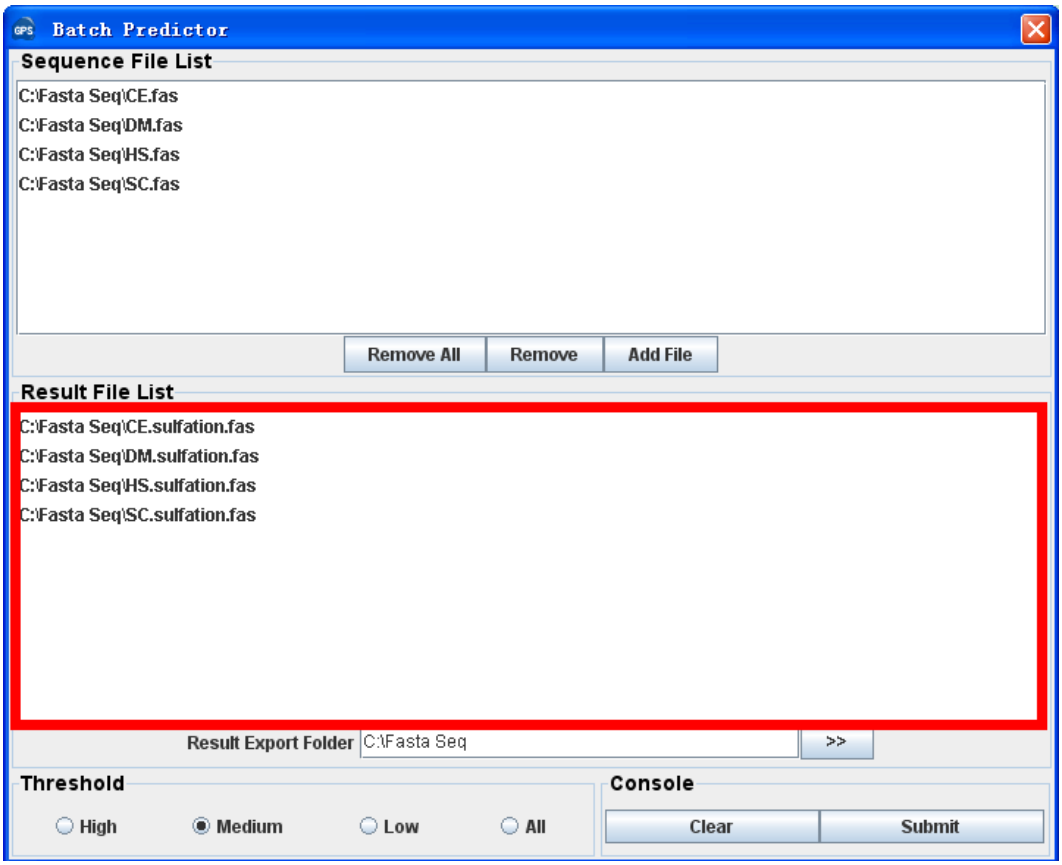
Then the names of added files will be shown in the **Sequence File List**.



(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. Bundgaard, J.R., Vuust, J. and Rehfeld, J.F. (1995) Tyrosine O-sulfation promotes proteolytic processing of progastrin. *The EMBO journal*, **14**, 3073-3079.
2. Bundgaard, J.R., Vuust, J. and Rehfeld, J.F. (1997) New consensus features for tyrosine O-sulfation determined by mutational analysis. *The Journal of biological chemistry*, **272**, 21700-21705.
3. Nicholas, H.B., Jr., Chan, S.S. and Rosenquist, G.L. (1999) Reevaluation of the determinants of tyrosine sulfation. *Endocrine*, **11**, 285-292.
4. Walsh, G. and Jefferis, R. (2006) Post-translational modifications in the context of therapeutic proteins. *Nature biotechnology*, **24**, 1241-1252.
5. Stone, M.J., Chuang, S., Hou, X., Shoham, M. and Zhu, J.Z. (2009) Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *New biotechnology*, **25**, 299-317.
6. Monigatti, F., Hekking, B. and Steen, H. (2006) Protein sulfation analysis--A primer. *Biochimica et biophysica acta*, **1764**, 1904-1913.
7. Kehoe, J.W. and Bertozzi, C.R. (2000) Tyrosine sulfation: a modulator of extracellular protein-protein interactions. *Chemistry & biology*, **7**, R57-61.
8. Gao, J., Choe, H., Bota, D., Wright, P.L., Gerard, C. and Gerard, N.P. (2003) Sulfation of tyrosine 174 in the human C3a receptor is essential for binding of C3a anaphylatoxin. *The Journal of biological chemistry*, **278**, 37902-37908.
9. Komori, R., Amano, Y., Ogawa-Ohnishi, M. and Matsubayashi, Y. (2009) Identification of tyrosylprotein sulfotransferase in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 15067-15072.
10. Monigatti, F., Gasteiger, E., Bairoch, A. and Jung, E. (2002) The Sulfinator: predicting tyrosine sulfation sites in protein sequences. *Bioinformatics*, **18**, 769-770.
11. Chang, W.C., Lee, T.Y., Shien, D.M., Hsu, J.B., Horng, J.T., Hsu, P.C., Wang, T.Y., Huang, H.D. and Pan, R.L. (2009) Incorporating support vector machine for identifying protein tyrosine sulfation sites. *Journal of computational chemistry*, **30**, 2526-2537.
12. Yang, Z.R. (2009) Predicting sulfotyrosine sites using the random forest algorithm with significantly improved prediction accuracy. *BMC bioinformatics*, **10**, 361.
13. Niu, S., Huang, T., Feng, K., Cai, Y. and Li, Y. (2010) Prediction of tyrosine sulfation with mRMR feature selection and analysis. *Journal of proteome research*, **9**, 6490-6497.

Release Note

1. April 12, 2012, the online service and the local stand-alone packages of GPS-TSP 1.0 were released.