GPS-MSP Manual

Prediction of protein methylation site

Version 1.0

20/11/2014

Author: Wankun Deng, Yu Xue.

Contact:

Wankun Deng: dengwankun@hust.edu.cn; Yu Xue: xueyu@hust.edu.cn

The software is only free for academic research.

The latest version of GPS-MSP software is available from http://msp.biocuckoo.org

Statement

1. Implementation. The soft wares 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 soft wares are freely available for academic researches. For non-profit users, you can copy, distribute and use the soft wares for your scientific studies. Our soft wares 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 soft wares are designed in an easy-to-use manner. Also, we invite you to read the manual before using the soft wares.
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).

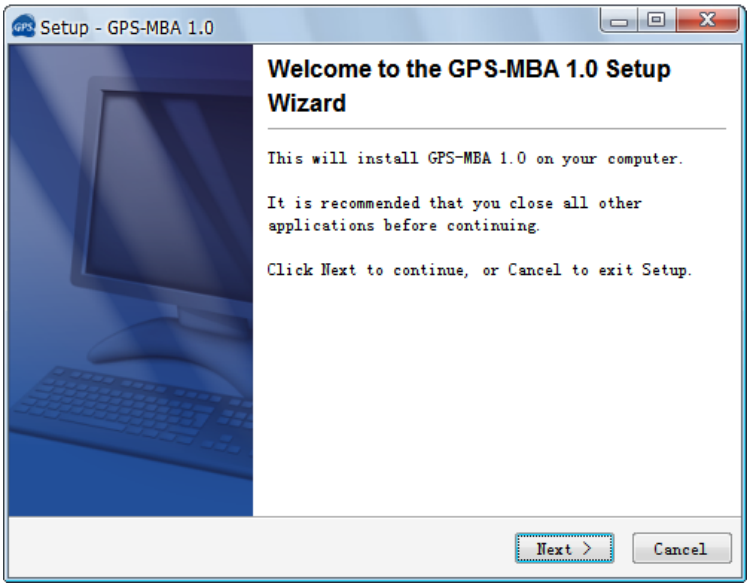
Introduction

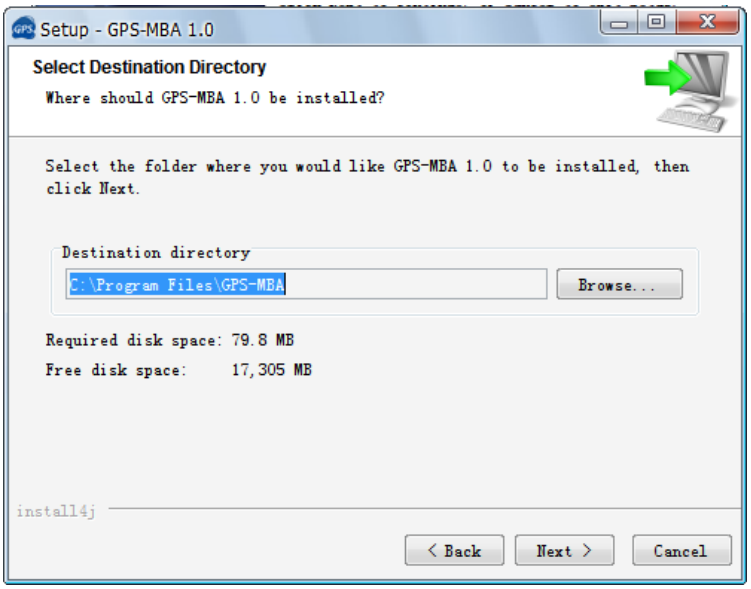
Protein methylation is an important post-translation modification (PTM), which was firstly discovered near half a century ago (1959), and was proved to be reversible later. Despite the long history of its discovery, the function and role it played in cellular remains to be elucidate. The foundation of function of protein methylation research is identification of protein methylated sites. The experimental identification of protein methylation sites is labor-extensive and expensive. Thus, a computational tool of prediction is needed. But all the prediction tools we can get now have some limitations. The detail methylation type is closely related to its function, but until now, all the prediction tools can't inform us anything about this. So, we developed MSP to fix this require.

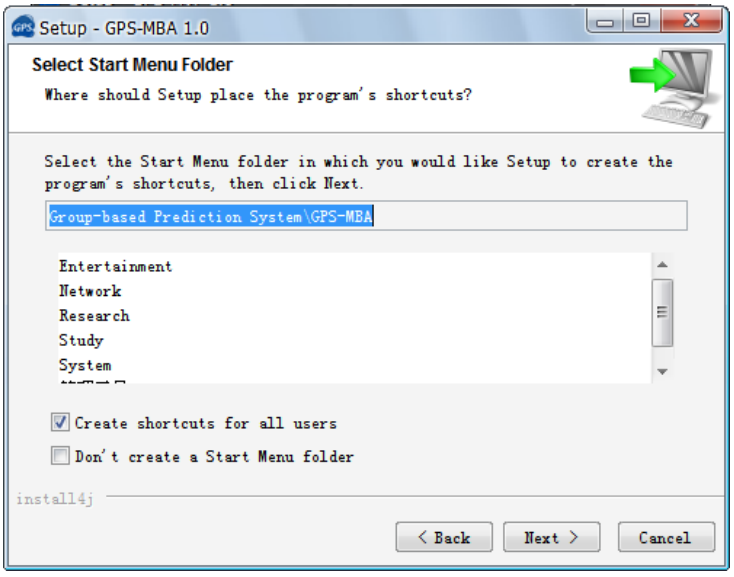
Download & Installation

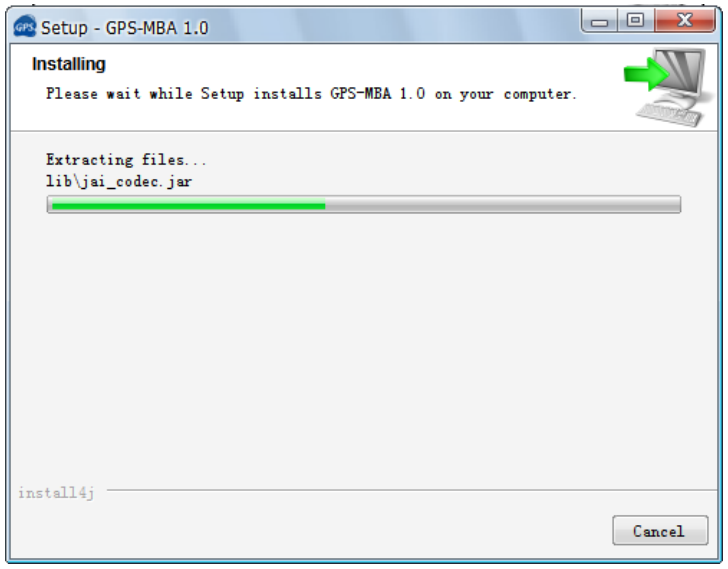
The GPS-MSP 1.0 was implemented in JAVA (J2SE), and could support threemajor Operating Systems (OS), including Windows, Linux/Unix or Mac OS Xsystems. Both of online web service and local stand-alone packages areavailable from: http://msp.biocuckoo.org/. We recommend that users coulddownload the latest release.

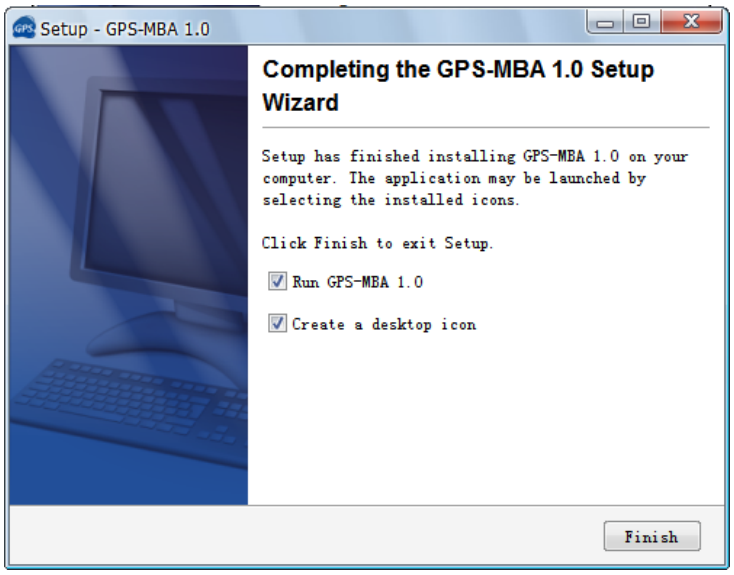
Please choose the proper package to download. After downloading, pleasedouble-click on the software package to begin installation, following the userprompts through the installation. And snapshots of the setup program forwindows are shown below:











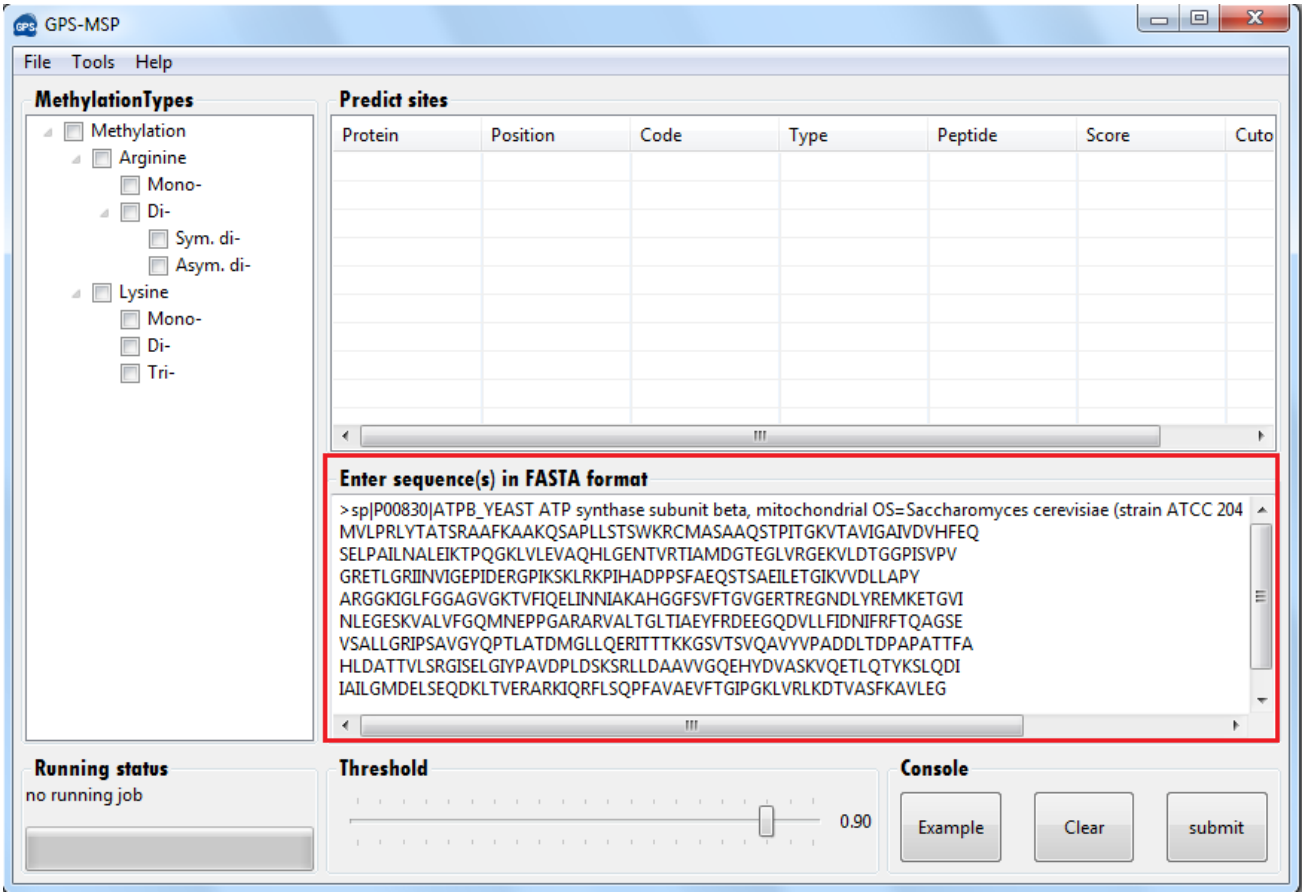
Finally, please click on the Finish button to complete the setup program.

Prediction of Protein Methylated Sites

1. A single protein sequence in FASTA format

The following steps show you how to use the GPS-MSP 1.0 to predict methylated 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-MSP 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) Secondly, choose a methylation type you need to predict on the left side:

The checkbox marked by red numbers means:

1: this checkbox can’t be checked;

2: predict methylated sites on arginine regardless detail methylation type;

3: predict mono-methylated arginine;

4: predict di-methylated arginine, regardless the detail methylation type (symmetry di-methylation and asymmetry di-methylation);

5: predict symmetry di-methylation arginine site;

6: predict asymmetry di-methylation arginine site;

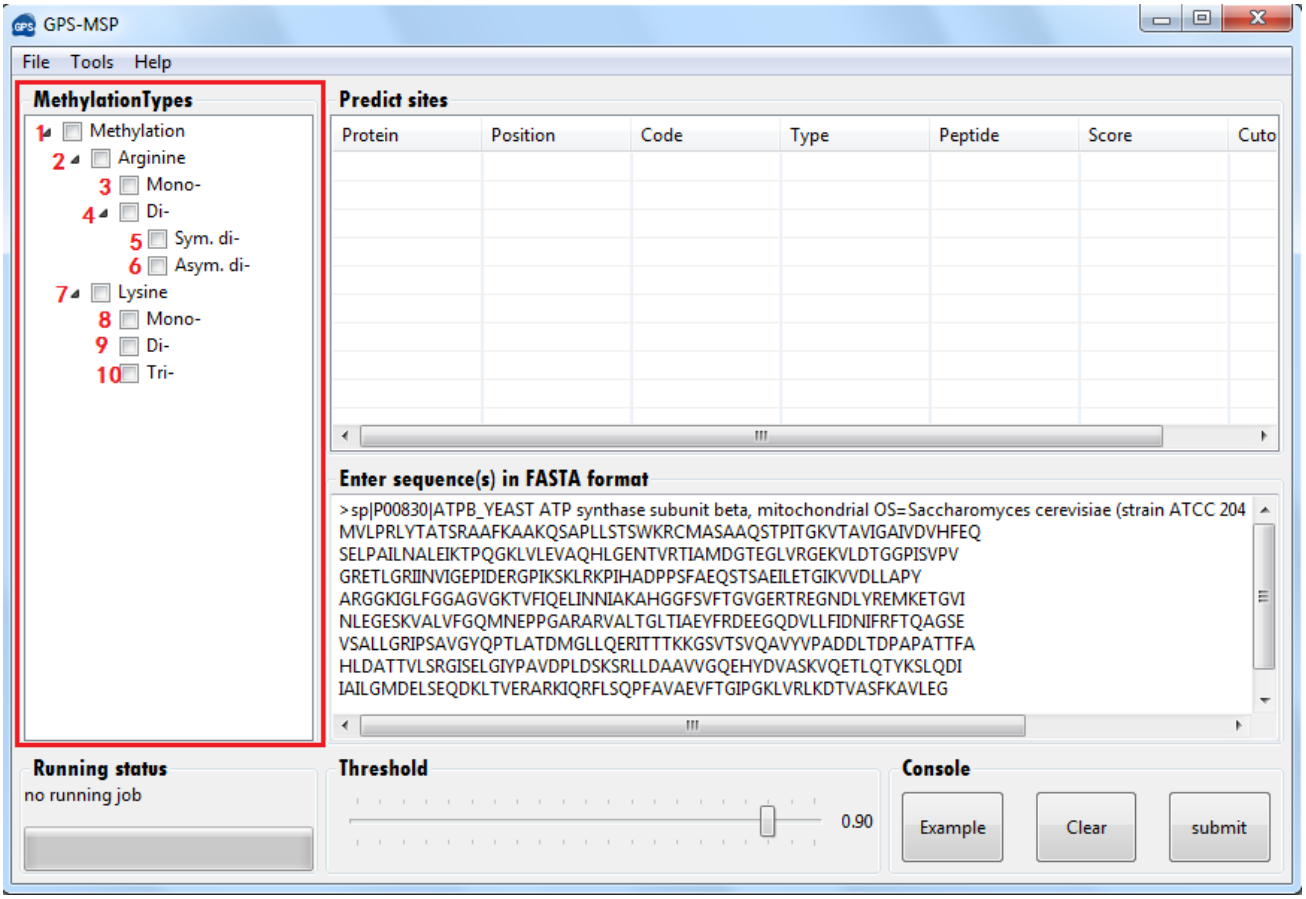
7: predict methylated lysine regardless detail methylation type;

8: predict mono-methylated lysine;

9: predict di-methylated lysine;

10: predict tri-methylated lysine.

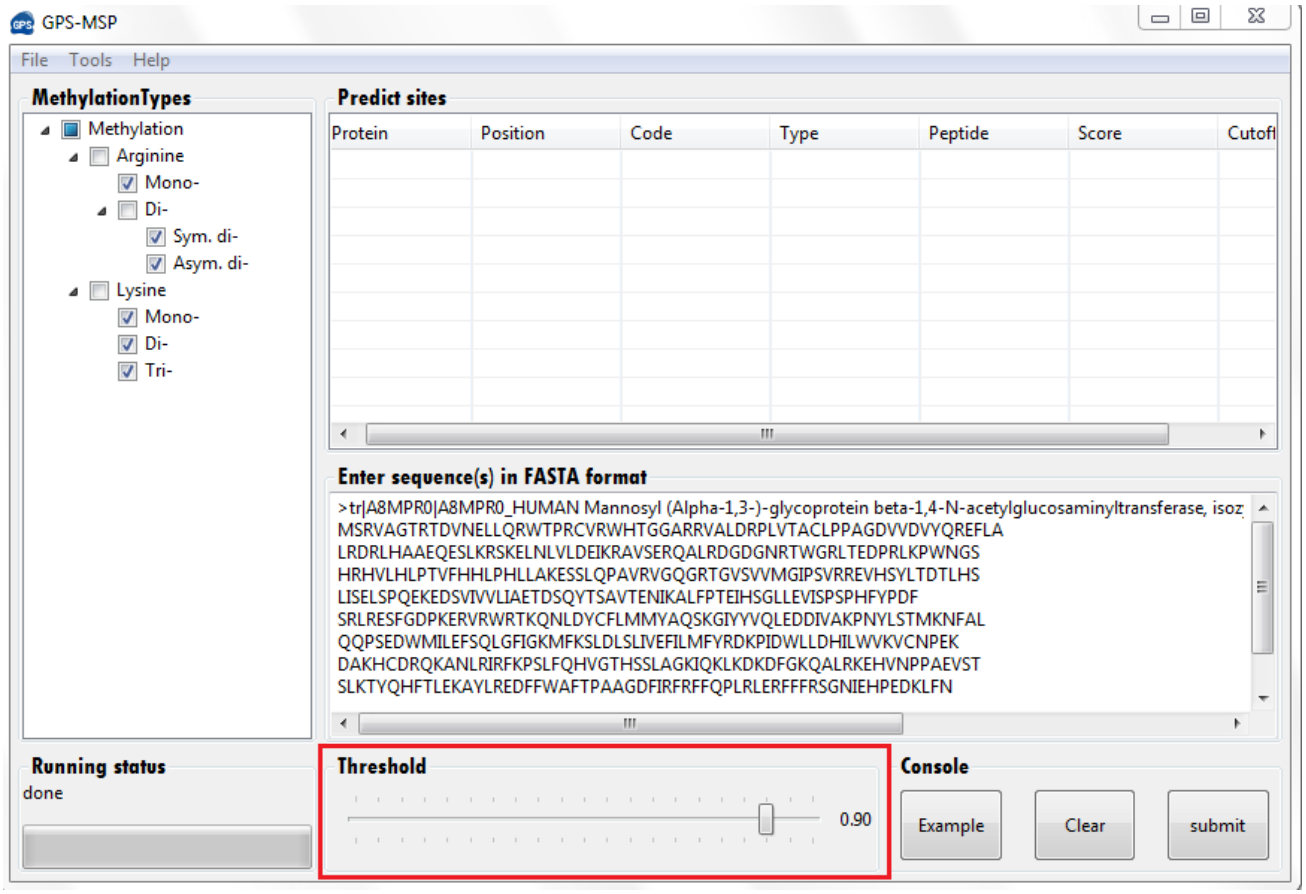
All these checkbox (except 1) can be check alone or in combination.



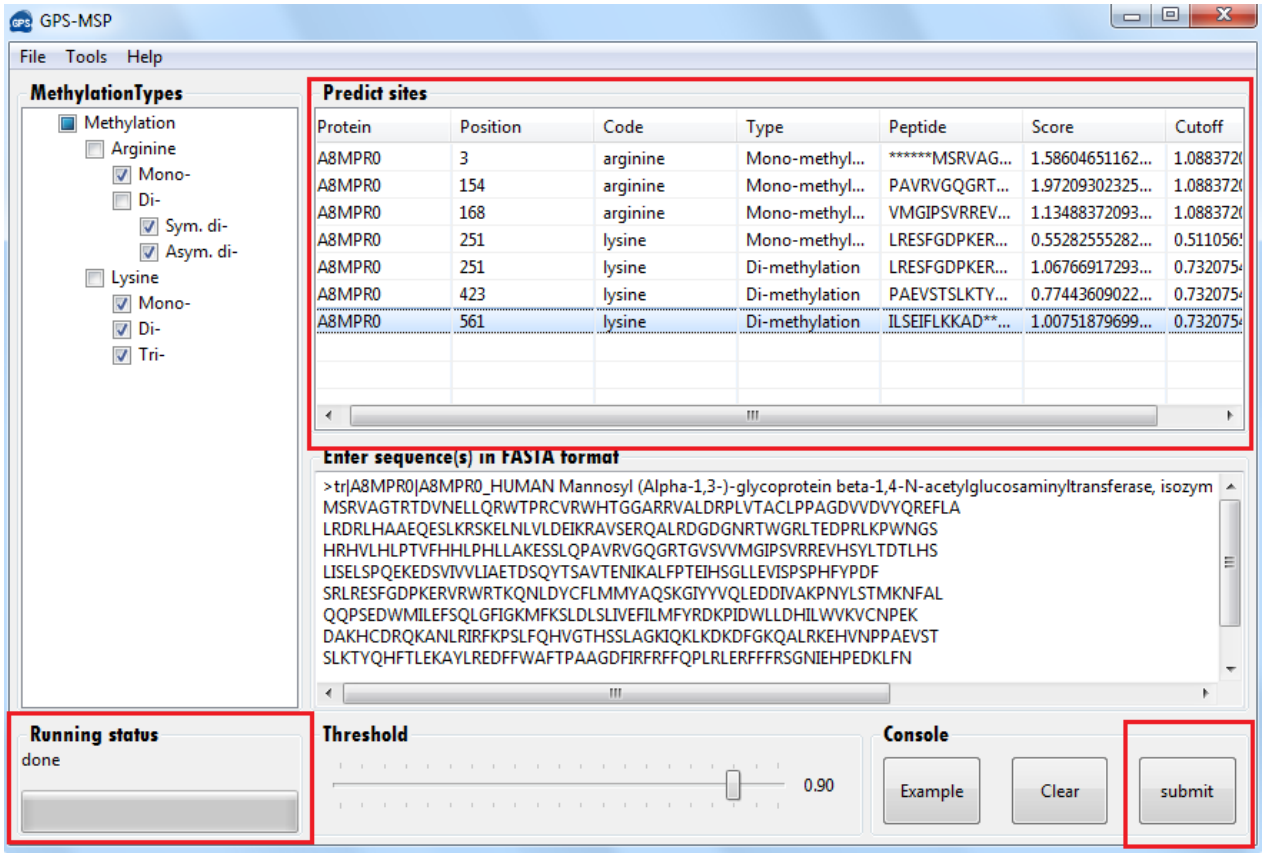
If no methylation type has been chose, program will do nothing but pop out a warning window.

(3) Set up threshold:

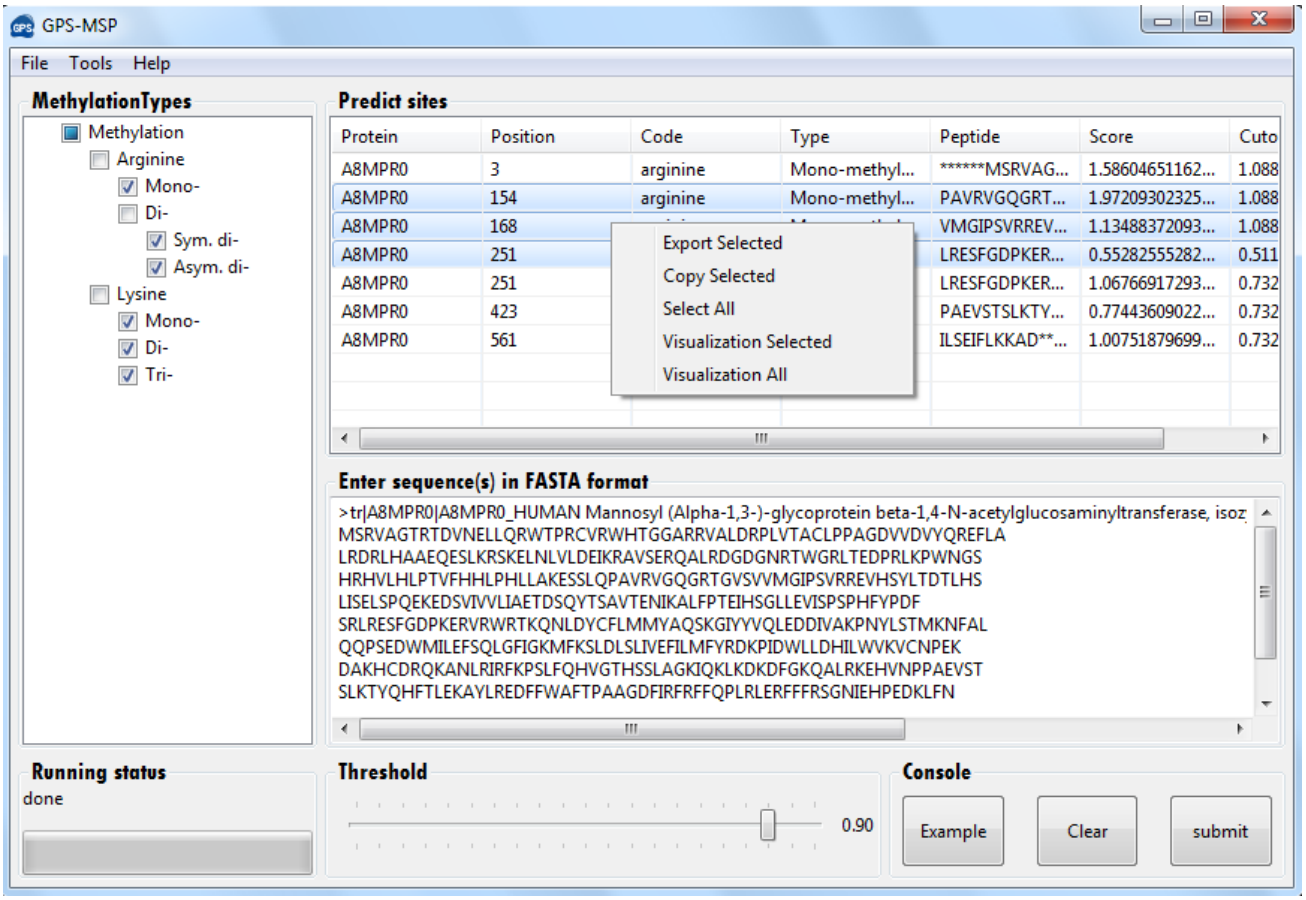
A threshold should be selected; it determines specificity and sensitivity of predict results. The value there means specificity you want to reach. It should be mentioned that as training sample is limited, the exact specificity of predict results is not equals to the one you chose, but it is close to it. We use score of relative LOO validation as the cutoff.



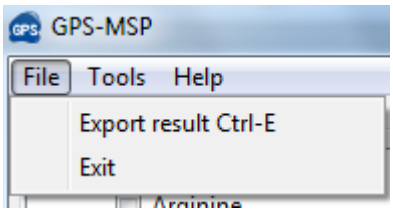
(4) click submit button to start predict, while running ,status will be shown at “Running status ” section. After a while, result will be present in “Prediction” section:



(5) 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 Result” to export the prediction results into a tab-delimited text file.



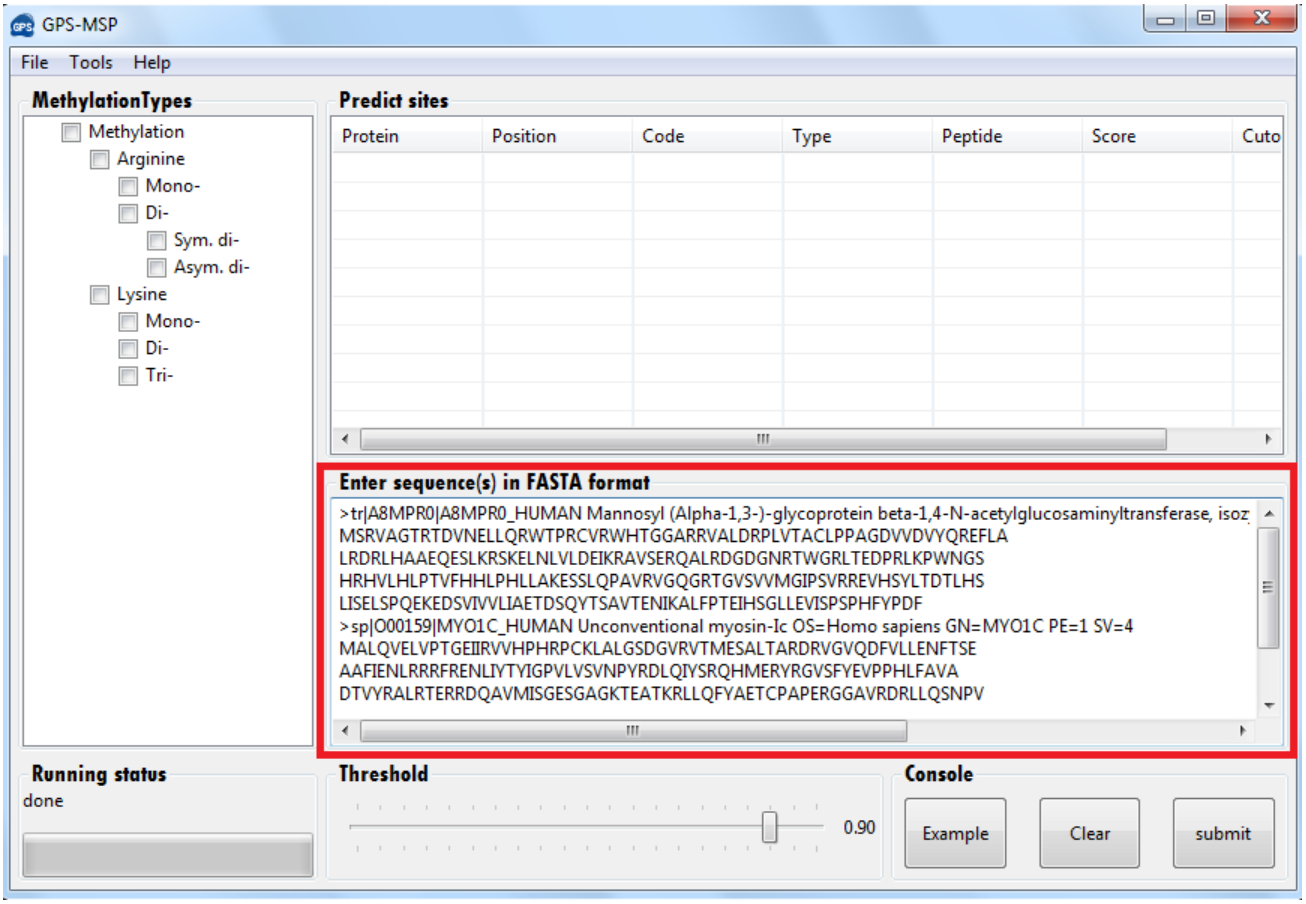
Again, you can also click the “Export Prediction” in File menu to export the results.



2. Multiple protein sequences in FASTA format

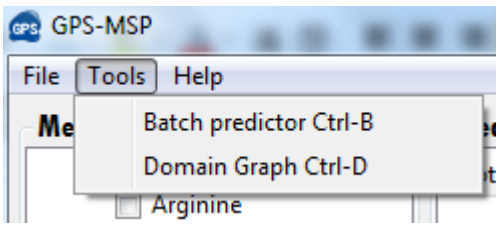
For multiple protein sequences, there are two ways to use the GPS-MSP 1.0.

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-MBA 1.0 for prediction.

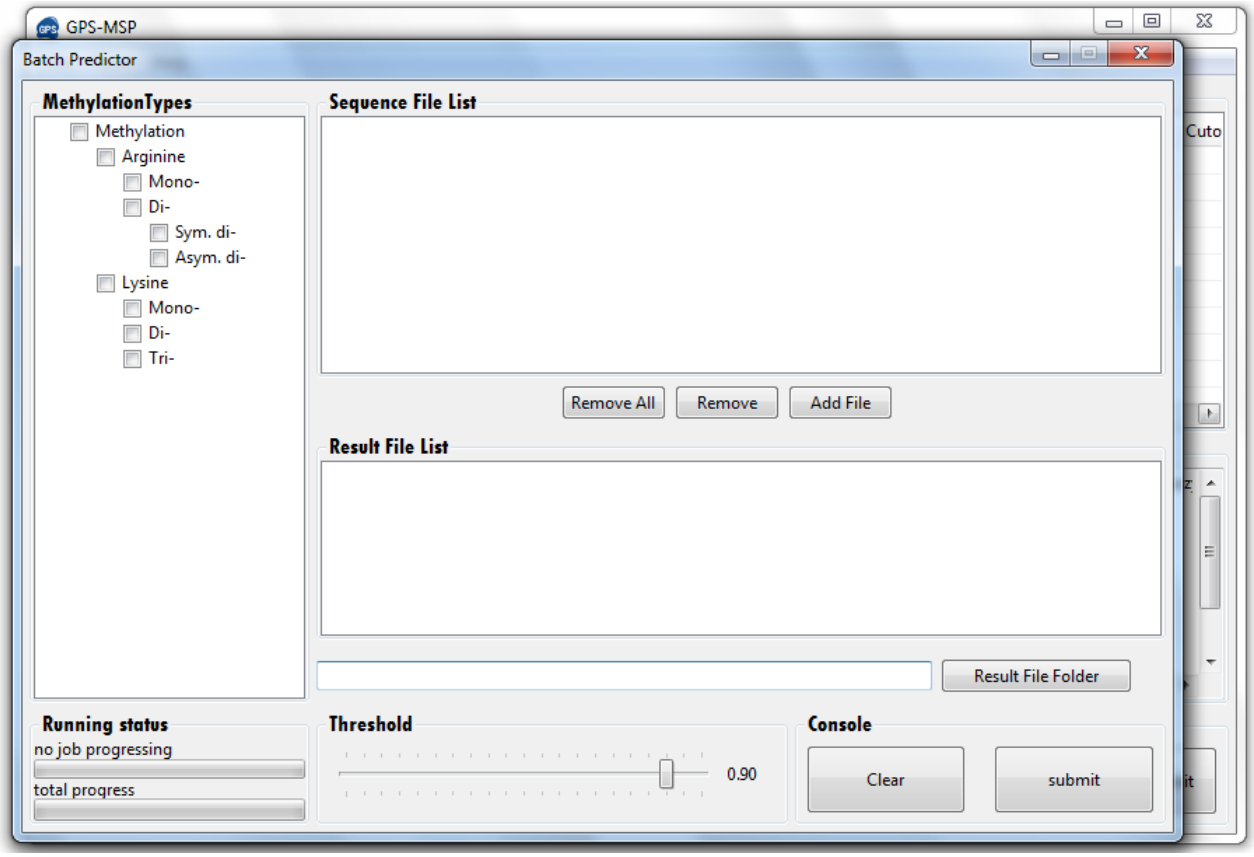


B. Use Batch Predictor tool.

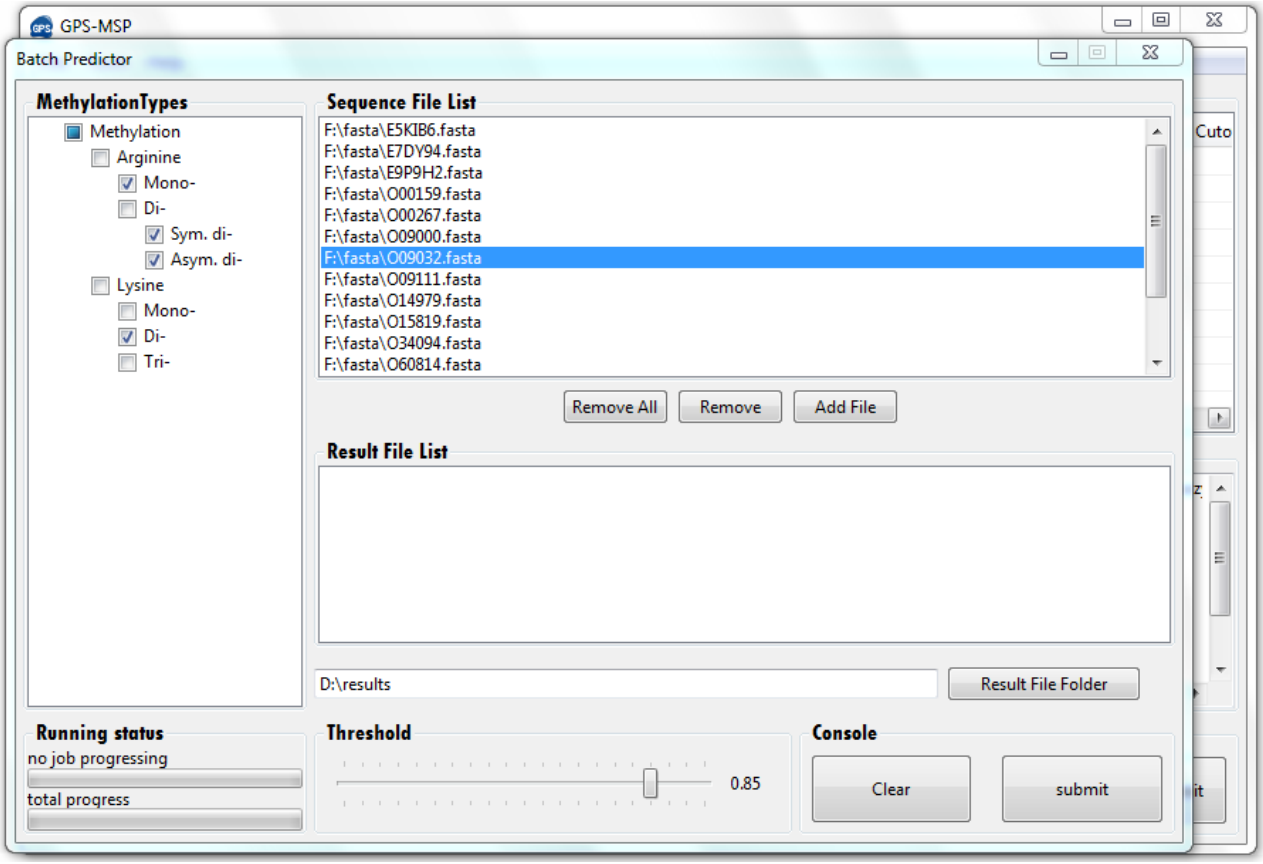
If the number of protein sequences is very large, eg., yeast or human proteome, please use the Batch Predictor. Please click on the “Batch Predictor” button in the Tools menu or hit “Ctrl + B” (windows and Linux/Unix) and “command + B” (Mac).



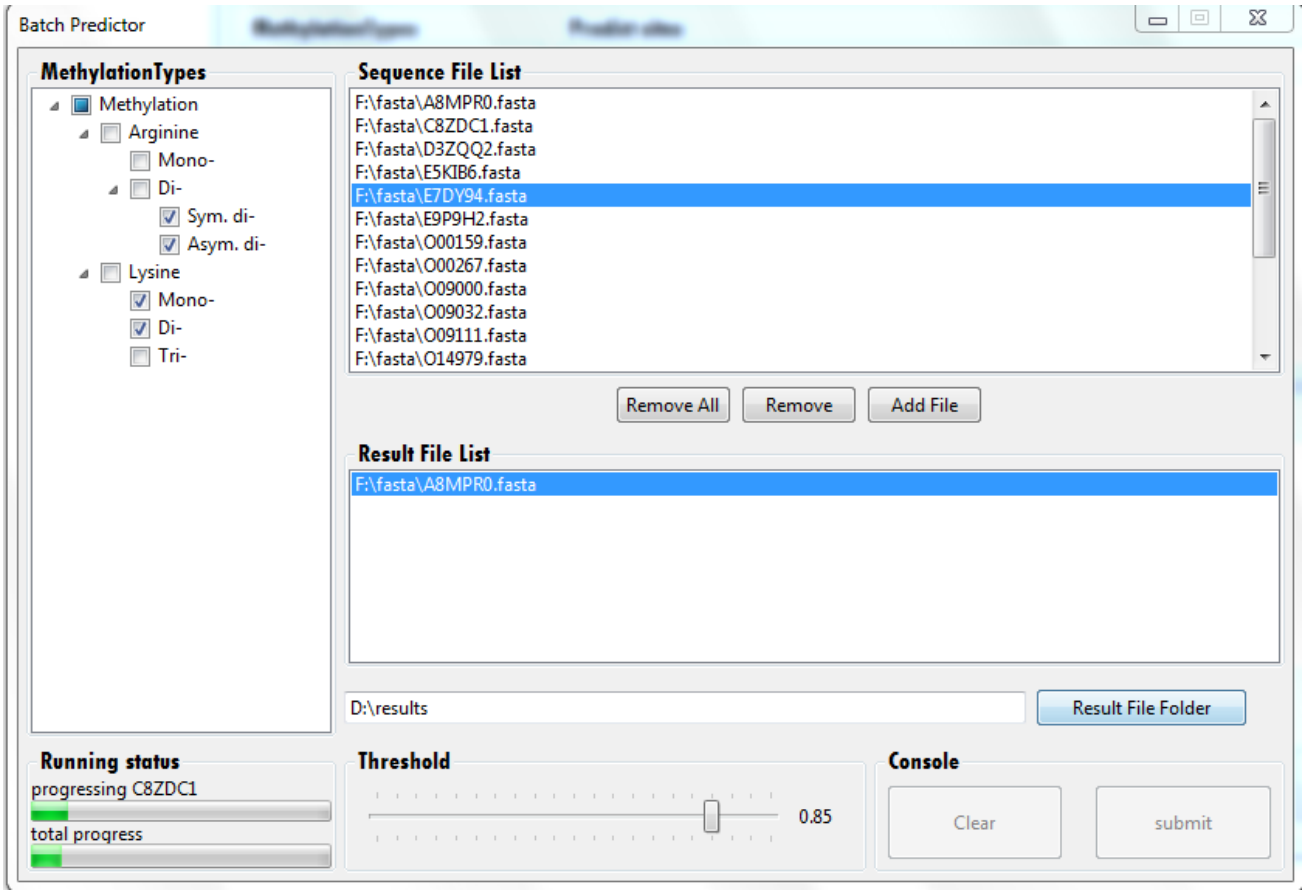
It should be emphasized that in batch predictor, only one sequence should be in a sequence file, and once you chose all parameters and the results folder, hit submit to start predict. The results will be output as a text file which has the same file name as relative sequence file and whose suffix is “.gps”.



User interface of batch predictor



Parameter selection and sequence selection



In progressing

C8ZDC1.gps

Position	Code	Type	Peptide	Score	Cutoff↓
27	lysine	Di-methylation	VLPPKEIVKSNTSSKKA	0.7894736842105263	0.6037735849056604↓
46	lysine	Mono-methylation	PPPSADPSKARKNRPRP	0.3808353808353808	0.3538083538083538↓
49	lysine	Mono-methylation	SADPSKARKNRPRPSGN	1.393120393120393	0.3538083538083538↓
49	lysine	Di-methylation	SADPSKARKNRPRPSGN	1.1090225563909775	0.6037735849056604↓
51	arginine	Asymmetry-di-methylation	DPSKARKNRPRPSGNEG	1.36	1.1885714285714286↓
64	lysine	Di-methylation	GNEGAIRDKTAGRRNNR	1.112781954887218	0.6037735849056604↓
68	arginine	Asymmetry-di-methylation	AIRDKTAGRRNNRSKDV	1.3714285714285714	1.1885714285714286↓
72	arginine	Asymmetry-di-methylation	KTAGRRNNRSKDVTDSA	1.262857142857143	1.1885714285714286↓
74	lysine	Di-methylation	AGRRNNRSKDVTDSATT	1.0827067669172932	0.6037735849056604↓
83	lysine	Di-methylation	DVTDSATTKKSNTRRAT	0.6090225563909775	0.6037735849056604↓
84	lysine	Mono-methylation	VTDSATTKKSNTRRATD	0.3955773955773956	0.3538083538083538↓
84	lysine	Di-methylation	VTDSATTKKSNTRRATD	1.9736842105263157	0.6037735849056604↓
105	lysine	Di-methylation	TGKTDTKKKVNQGWGDD	1.2593984962406015	0.6037735849056604↓
114	lysine	Mono-methylation	VNQGWGDDKKELSAEKE	0.35626535626535627	0.3538083538083538↓
115	lysine	Mono-methylation	NQGWGDDKKELSAEKEA	0.39803439803439805	0.3538083538083538↓
220	lysine	Di-methylation	TFVESNTRKNFGDRNNN	0.7593984962406015	0.6037735849056604↓
230	arginine	Symmetry-di-methylation	FGDRNNNSRNNFNNRRG	1.7430555555555556	1.2083333333333333↓
230	arginine	Asymmetry-di-methylation	FGDRNNNSRNNFNNRRG	1.4857142857142858	1.1885714285714286↓
236	arginine	Asymmetry-di-methylation	NSRNNFNNRRGGRGARK	1.9657142857142857	1.1885714285714286↓
237	arginine	Symmetry-di-methylation	SRNNFNNRRGGRGARKG	2.75	1.2083333333333333↓
237	arginine	Asymmetry-di-methylation	SRNNFNNRRGGRGARKG	3.5542857142857143	1.1885714285714286↓
240	arginine	Symmetry-di-methylation	NFNNRRGGRGARKGNNT	2.6805555555555554	1.2083333333333333↓
240	arginine	Asymmetry-di-methylation	NFNNRRGGRGARKGNNT	3.3085714285714287	1.1885714285714286↓
243	arginine	Asymmetry-di-methylation	NRRGGRGARKGNNTANA	2.4457142857142857	1.1885714285714286↓
244	lysine	Mono-methylation	RRGGRGARKGNNTANAT	0.44471744471744473	0.3538083538083538↓

Result file format