GPS-MBA Manual

Prediction of MHC Class II Epitopes

Version 1.0
17/11/2011

Author: Ruikun Cai, Zexian Liu, Jian Ren, Yanhong Zhou & 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-MBA software is available from http://mba.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

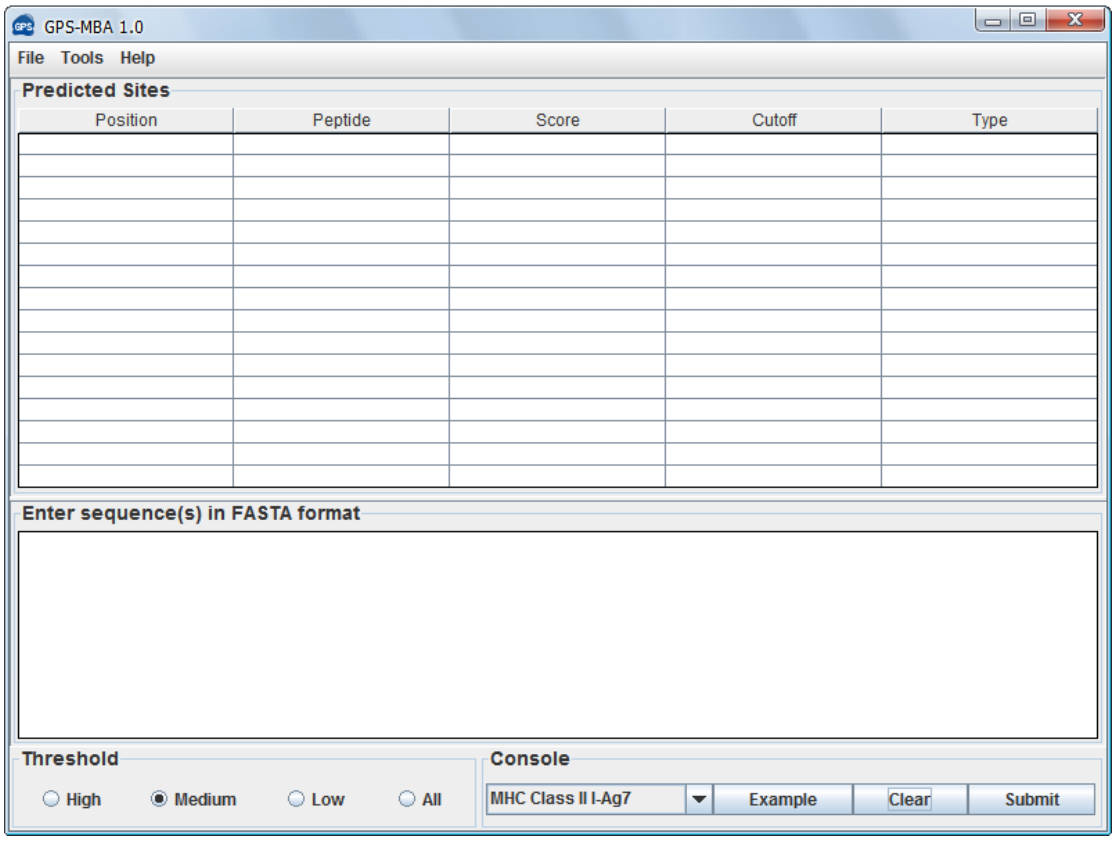
Type 1 diabetes (Diabetes mellitus type 1, T1D or T1DM) is a severe chronic autoimmune disease with a relapsing-remitting course that is characterized by the insidious loss of self-tolerance and progressive destruction of insulin-producing pancreatic β-cells in the islets of Langerhans, with the presence of overt hyperglycemia at the time of clinical diagnosis (1-7). The incidence and prevalence of T1D has dramatically increased worldwide over the past several decades, and the onset and development of T1D is believed to be controlled by both genetic and environmental factors (1-5,8). The cumulative analysis has revealed that a variety of immune cell types, including $CD4^+$, $CD8^+$ T cells, macrophages and dendritic cells (DCs) are involved in β-cell death, and $CD4^+$ T cells play the predominant role in the overall T1D pathology (1-2,8). Thus, the development of immunoregulatory therapeutic approaches has come to be an urgent demand for preventing, treating or even curing T1D (1-7).

Besides immunosuppressive drugs and antibody-based immunotherapies, antigen-based “tolerogenic” immunotherapy has attracted considerable attention as a third-generation approach, particularly for its highly selective targeting of aberrant T cells (1-6). It was demonstrated that the MHC class II haplotype, $I\text{-}A^{g7}$, is strongly linked to susceptibility to T1D in the non-obese diabetic (NOD) mouse (9-11). Similar linkage to the human HLA-DQ8 molecule, $I\text{-}A^{g7}$ is expressed by DCs to present β-cell epitopes from certain well-defined autoantigens, including insulin, glutamic acid decarboxylase (GAD) and insulinoma antigen 2 (IA-2) (1-6,8). These epitopes are usually composed of 10 to 30 amino acids, with a 9-amino acid core sequence for $I\text{-}A^{g7}$/HLA-DQ8 and T-cell receptor (TCR) binding (9-11). In this regard, identification of $I\text{-}A^{g7}$/HLA-DQ8 epitopes is fundamental for an understanding of the molecular mechanisms of T1D and the improved design of immunotherapeutic peptides. In 2009, the first-in-human beings Phase I clinical study reported that proinsulin peptide injection is both well tolerated and safe (12). Recently, a C-peptide deduced from the GAD 65 isoform has generated promising results in Phase II trials, and three Phase III trials are still ongoing (1-2,5).

As a complement to labor-intensive and time-consuming experimental assays, the *in silico* prediction of MHC-binding epitopes has emerged as an efficient approach to generate useful information for the purposes of biomedical design (13-14) (see also http://mba.biocuckoo.org/links.php). For example, the prediction results of SYFPEITHI (15) and BIMAS (16) were successfully used for the experimental identification of novel MHC class I epitopes derived from type 1 diabetes autoantigens (17-19). Since $I\text{-}A^{g7}$ is the only expressed MHC class II molecule in the NOD mouse (9-10), additional efforts have subsequently been expended on the prediction of $I\text{-}A^{g7}$ or HLA-DQ8 epitopes (20-23). In 2006, Rajapakse *et al*. developed the first online server of $PRED^{NOD}$ for the prediction of $I\text{-}A^{g7}$, and the two MHC class I molecules K^d and D^b binding peptides in the NOD mouse (20). They subsequently refined the

predictor using multi-objective evolutionary algorithms (MOEA) (23). Chang *et al*. used an expectation-maximization alignment algorithm to design computational programs for the prediction of I-A^{g7} (21) and HLA-DQ8 (22) epitopes, respectively. Furthermore, the two integrative tools of MHC2Pred (24-25) and RANKPEP(26) also include predictors for I-A^{g7} and HLA-DQ8, although they were developed for the comprehensive prediction of a variety of MHC class I and/or II binding peptides. Currently, although a number of computational studies have been performed, only MHC2Pred (24-25) and RANKPEP(26) are accessible over the internet.

In this work, we developed a novel GPS-MBA software package for the prediction of I-A^{g7} and HLA-DQ8. The experimentally identified epitopes were obtained from the scientific literature, and a modified Gibbs sampling approach was adopted to determine the core nonamers in these epitopes. For training and prediction, a refined GPS algorithm (27-28) was used. By extensive evaluation and comparison, the prediction performance of GPS-MBA was shown to be highly promising and much better than the other tools currently in use. Moreover, by cross-evaluation using the HLA-DQ8 predictor in GPS-MBA to predict the I-A^{g7} epitopes and *vice versa*, the results show that I-A^{g7} and HLA-DQ8 recognize highly similar peptide profiles. With this powerful tool, we predicted potentially novel I-A^{g7} and HLA-DQ8 binding peptides from T1D-associated epitopes, which bind to other types of MHC molecules. All of the experimentally identified T1D antigens together with their epitopes were absorbed into TEDB 1.0. The *ab initio* predicted epitopes were also provided. Taken together, the prediction and analysis results are helpful for further experimental investigation, and the GPS-MBA can serve as a practically useful adjunct program for experimentalists. The online service and local packages of GPS-MBA 1.0 were implemented in JAVA and freely accessible for academic research purposes at: http://mba.biocuckoo.org.

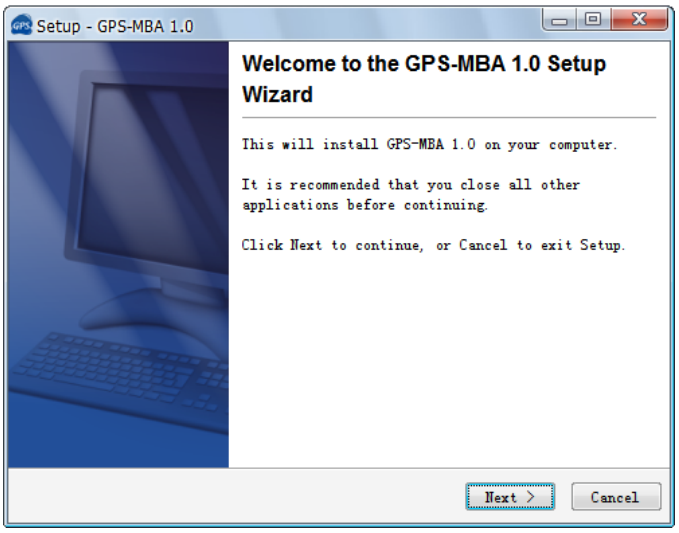


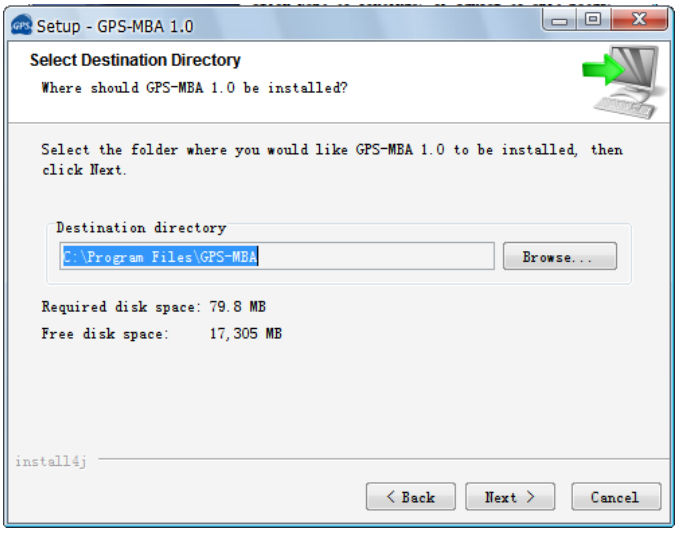
GPS-MBA 1.0 User Interface

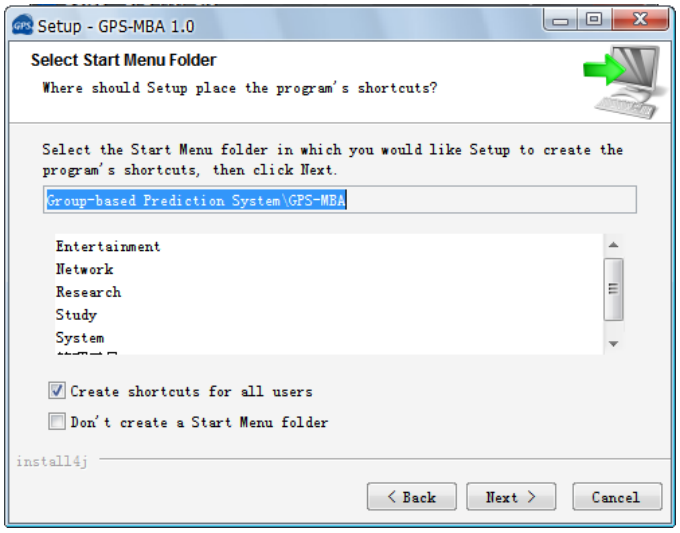
Download & Installation

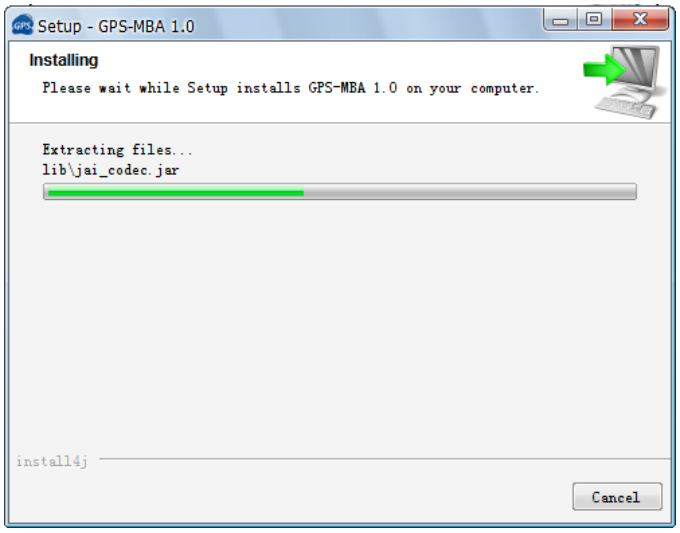
The GPS-MBA 1.0 was implemented in JAVA (J2SE), and could support three major Operating Systems (OS), including Windows, Linux/Unix or Mac OS X systems. Both of online web service and local stand-alone packages are available from: http://mba.biocuckoo.org/. We recommend that users could download the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

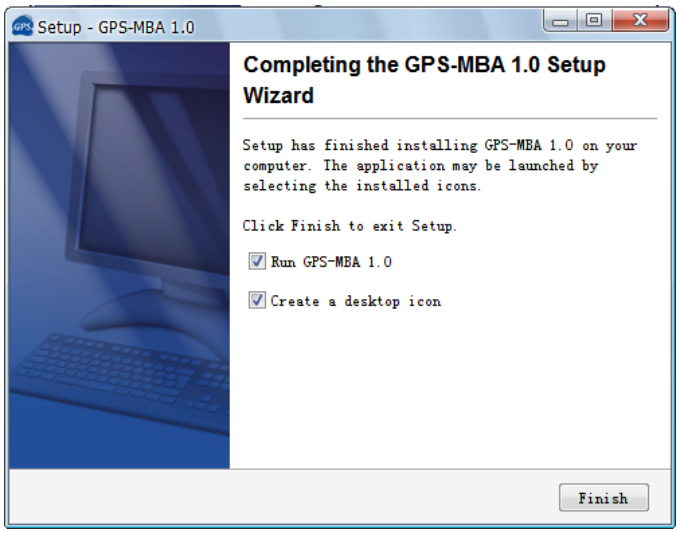
Please choose the proper package to download. After downloading, please double-click on the software package to begin installation, following the user prompts through the installation. And snapshots of the setup program for windows are shown below:











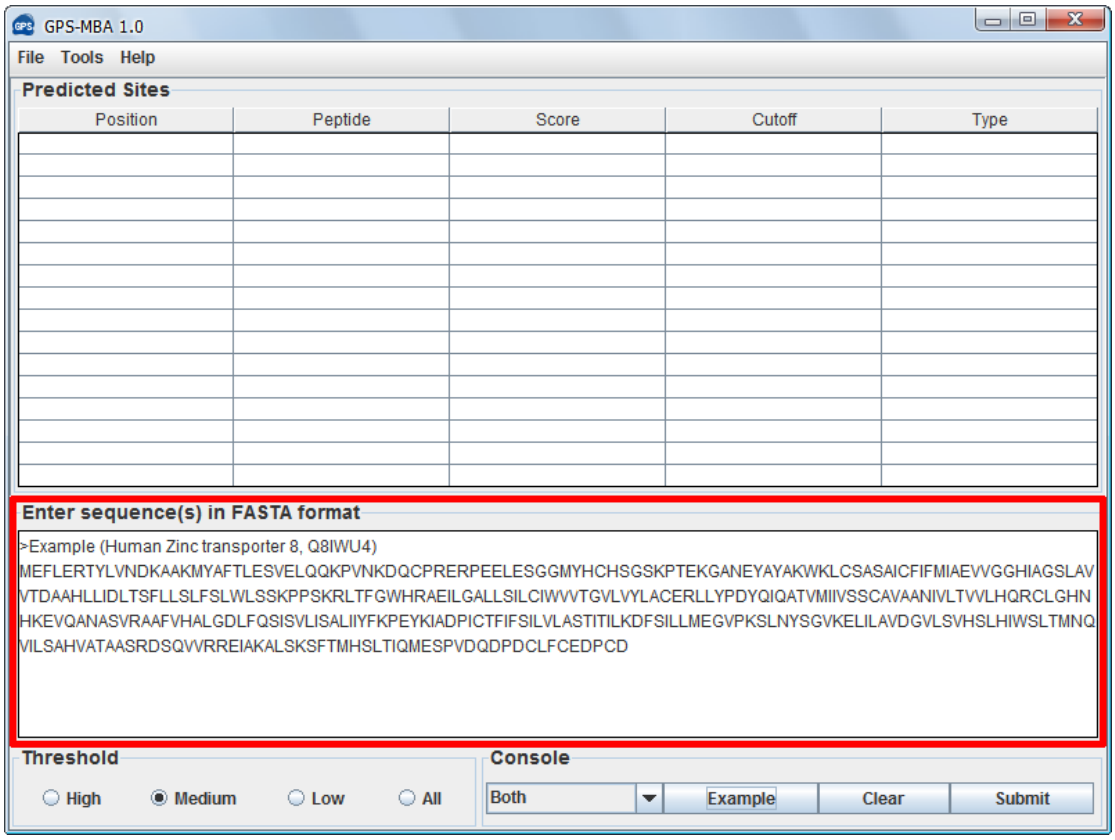
Finally, please click on the **Finish** button to complete the setup program.

Prediction of MHC class II I-Ag7 and MHC class II DQ-8 Epitopes

1. A single protein sequence in FASTA format

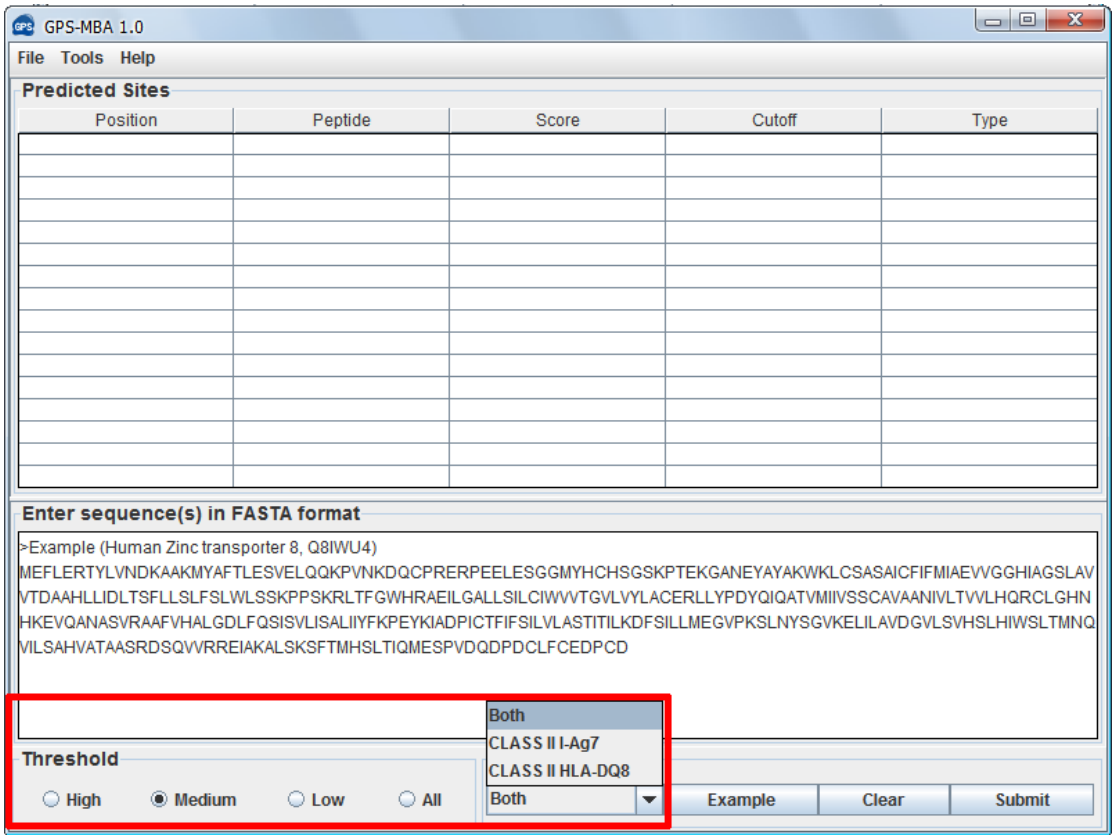
The following steps show you how to use the GPS-MBA 1.0 to predict MHC Class II Epitopes for a single protein sequence in FASTA format.

(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-MBA 1.0

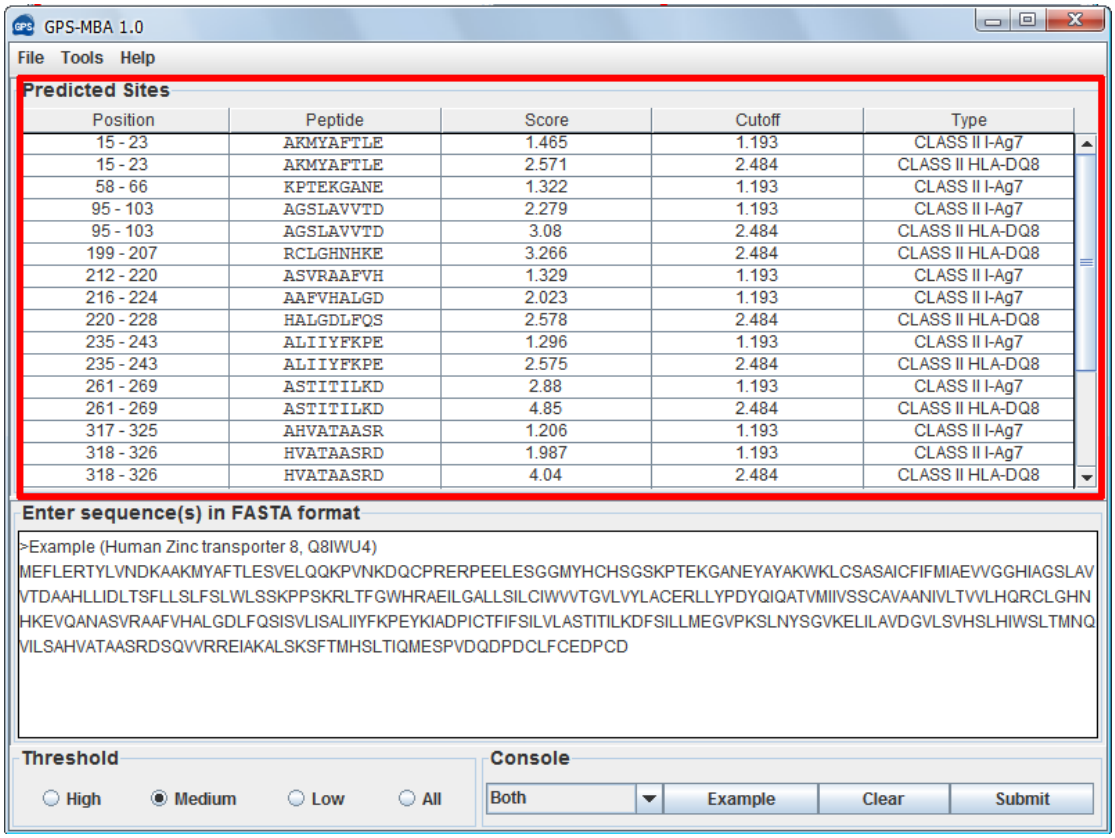


Note: for a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

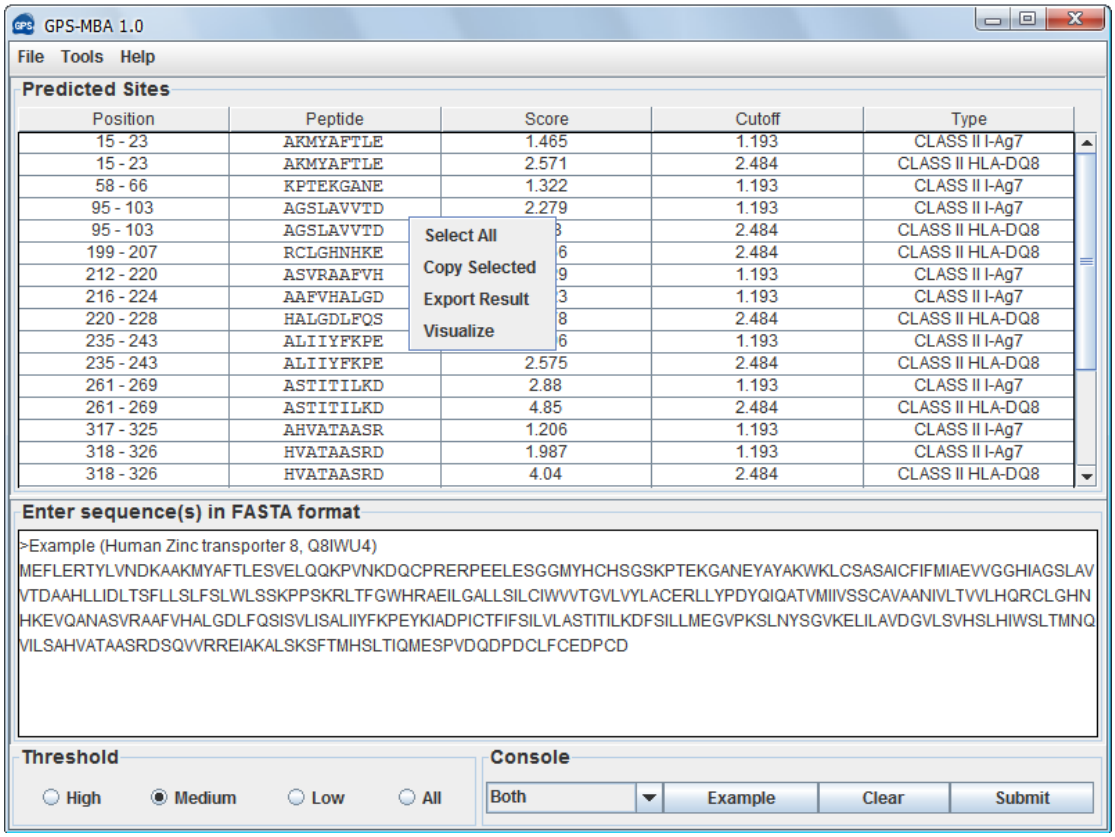
(2) Choose a **Threshold** and **Type** that you need, the default cut-off is **Medium** and **Both**.



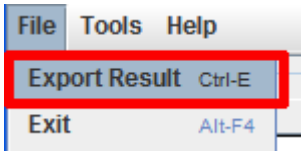
(3) Click on the **Submit** button, then the predicted MHC class II I-Ag7 and MHC class II DQ-8 MHC Class II Epitopes will be shown.



(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 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.

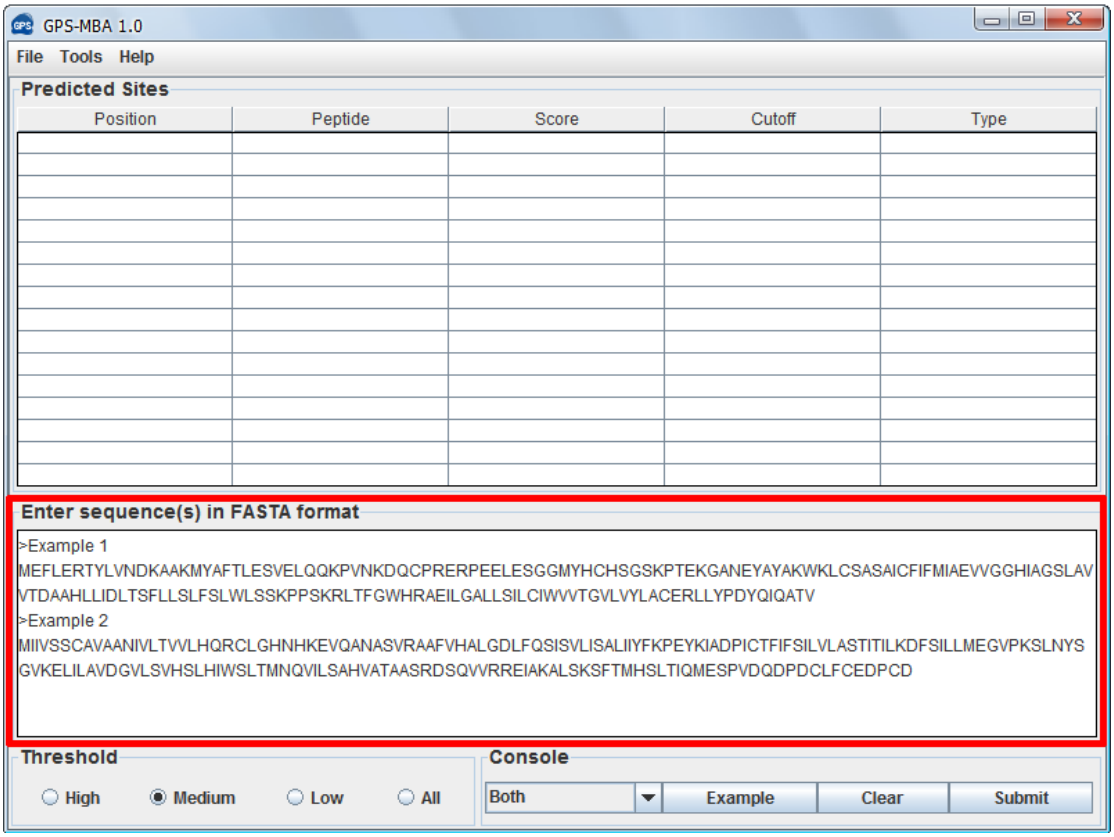


2. Multiple protein sequences in FASTA format

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

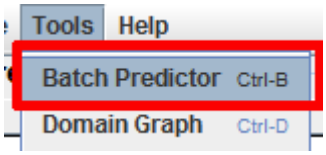
A. Input the sequences into text form directly. (Num. of Seq ≤ 2,000)

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.

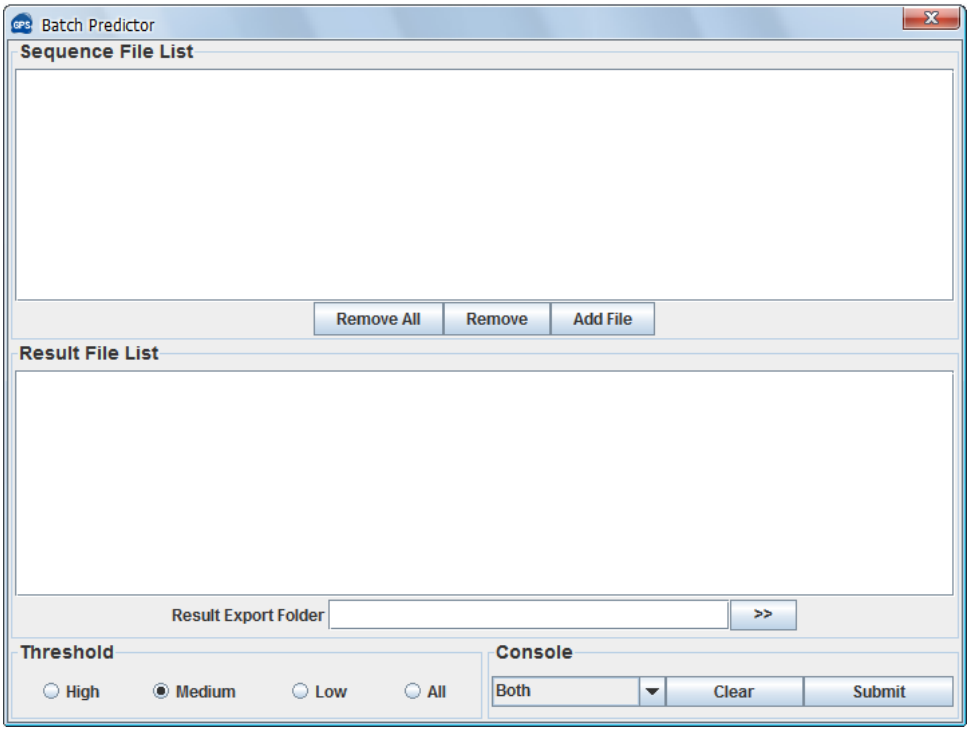


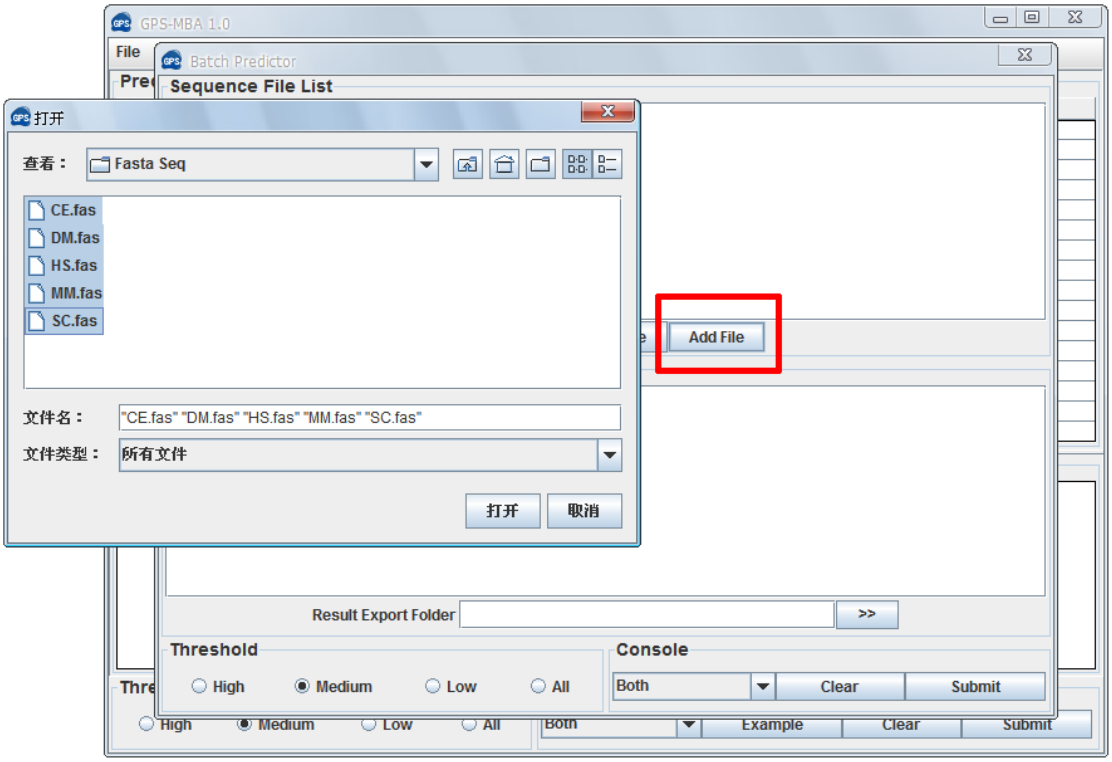
The following steps show you how to use it:

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

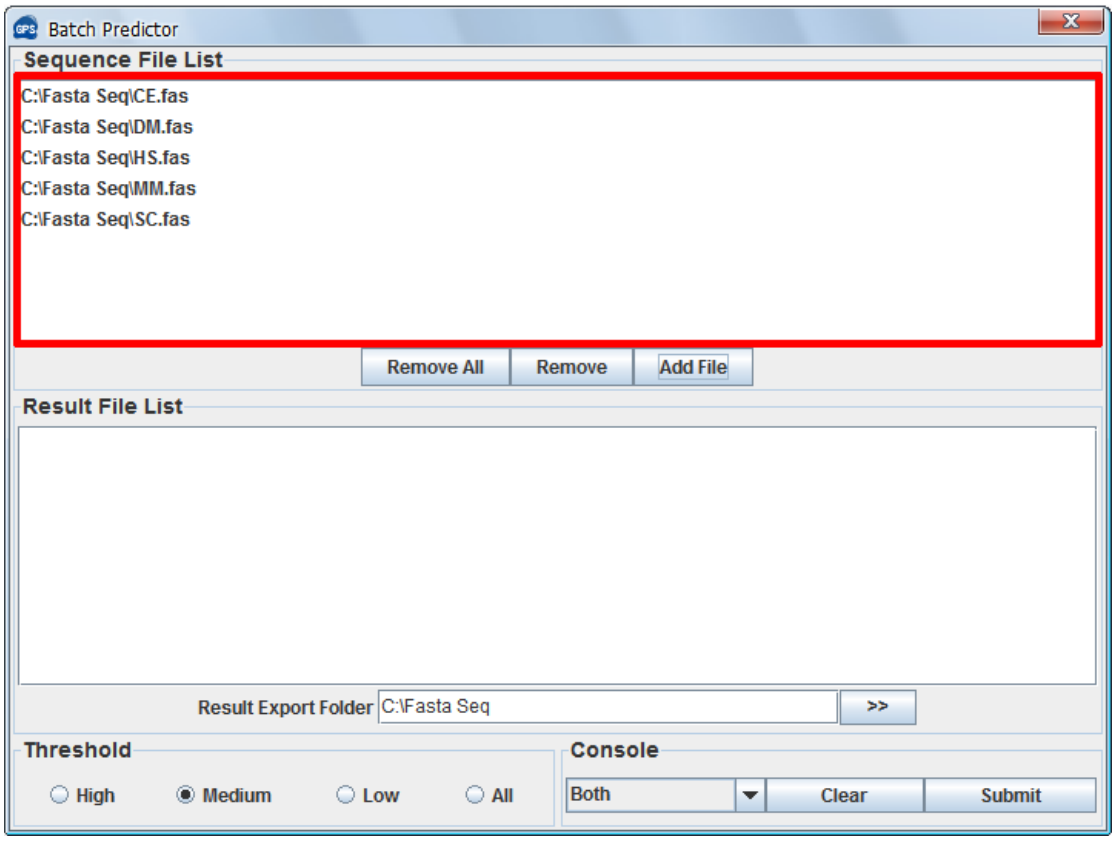
>protein1
XXXXXXXXXXXXX
XXXXXXXX
>protein2
XXXXXXXXXXXXXXX…
>protein3
XXXXXXXXXXX
…
Most importantly, the name of each protein should be presented.

(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.

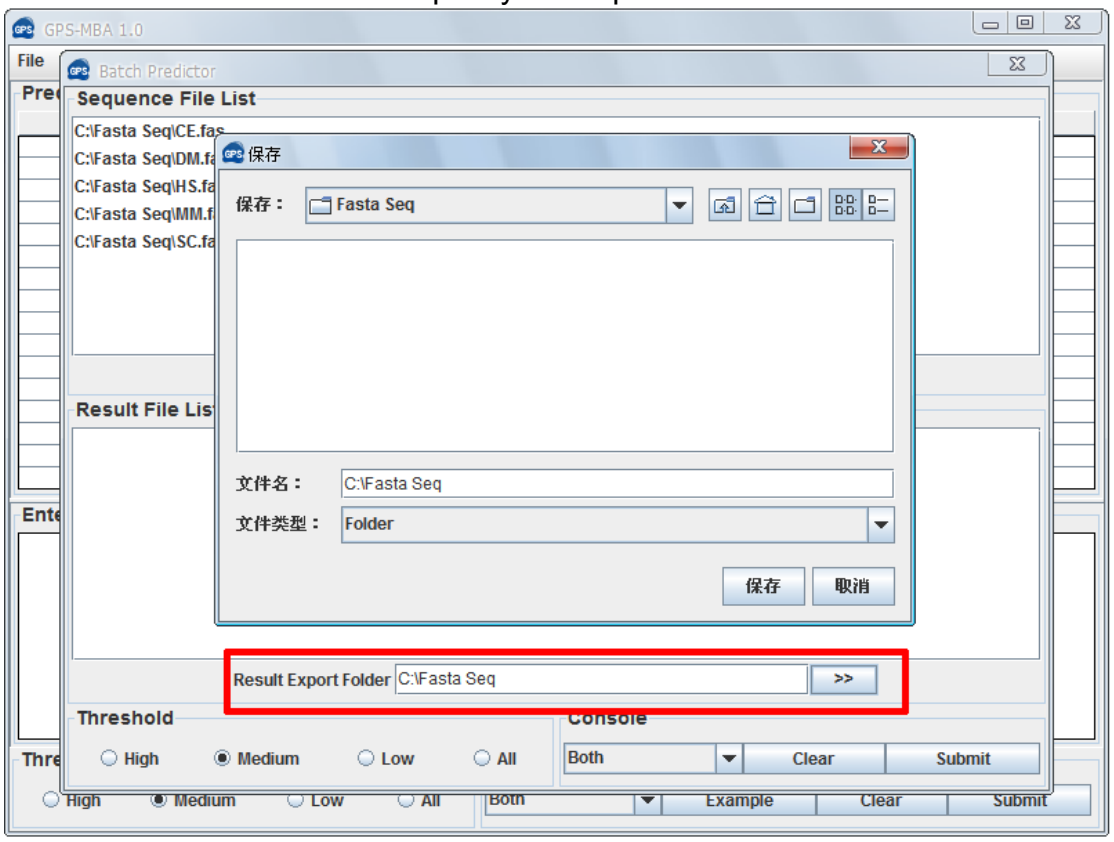




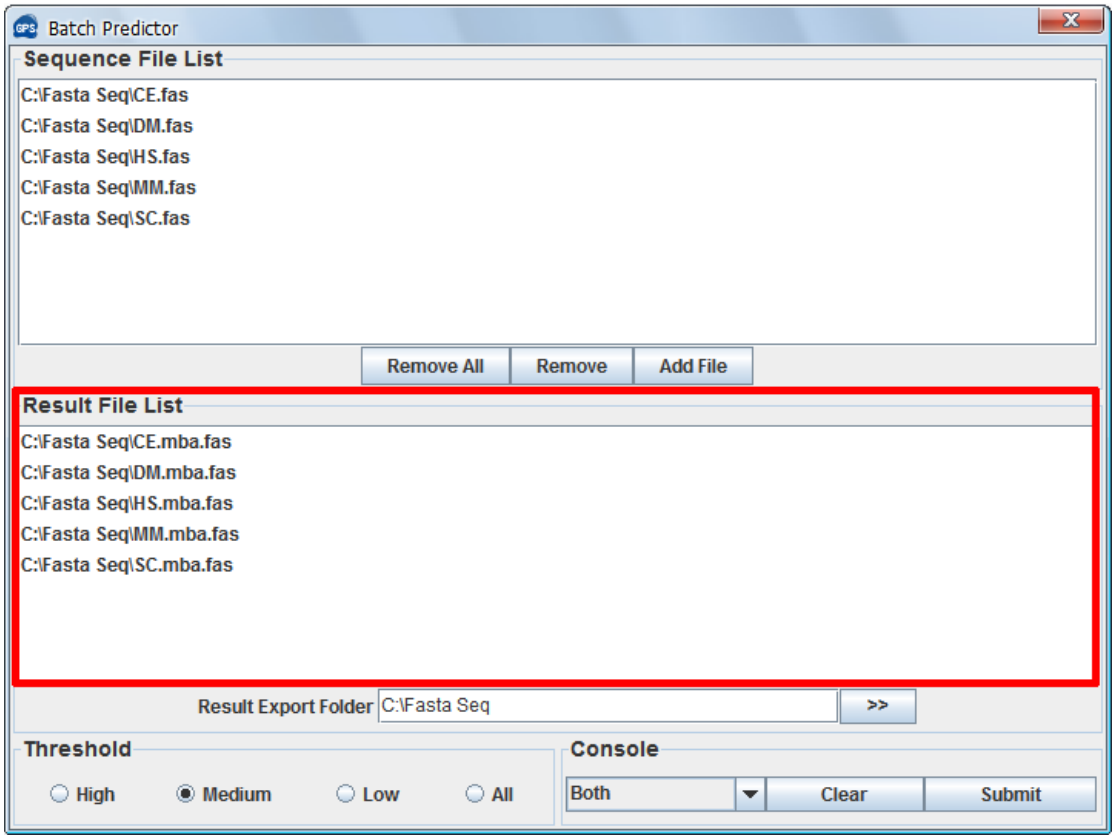
Then the names of added files will be shown in the **Sequence File List**.



(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 **Result Export Fold,** and the name of result files will be shown in the **Result File List**.



References

1. Fierabracci, A. (2011) Peptide immunotherapies in Type 1 diabetes: lessons from animal models. *Curr Med Chem*, **18**, 577-586.
2. Luo, X., Herold, K.C. and Miller, S.D. (2010) Immunotherapy of type 1 diabetes: where are we and where should we be going? *Immunity*, **32**, 488-499.
3. Bluestone, J.A., Herold, K. and Eisenbarth, G. (2010) Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*, **464**, 1293-1300.
4. Sherr, J., Sosenko, J., Skyler, J.S. and Herold, K.C. (2008) Prevention of type 1 diabetes: the time has come. *Nat Clin Pract Endocrinol Metab*, **4**, 334-343.
5. Waldron-Lynch, F. and Herold, K.C. (2011) Immunomodulatory therapy to preserve pancreatic beta-cell function in type 1 diabetes. *Nat Rev Drug Discov*, **10**, 439-452.
6. Panagiotopoulos, C., Trudeau, J.D. and Tan, R. (2004) T-cell epitopes in type 1 diabetes. *Current diabetes reports*, **4**, 87-94.
7. von Herrath, M., Sanda, S. and Herold, K. (2007) Type 1 diabetes as a relapsing-remitting disease? *Nat Rev Immunol*, **7**, 988-994.
8. Lehuen, A., Diana, J., Zaccone, P. and Cooke, A. (2010) Immune cell crosstalk in type 1 diabetes. *Nat Rev Immunol*, **10**, 501-513.
9. Reizis, B., Eisenstein, M., Bockova, J., KonenWaisman, S., Mor, F., Elias, D. and Cohen, I.R. (1997) Molecular characterization of the diabetes-associated mouse MHC class II protein, I-A(g7). *International Immunology*, **9**, 43-51.
10. Harrison, L.C., Honeyman, M.C., Trembleau, S., Gregori, S., Gallazzi, F., Augstein, P., Brusic, V., Hammer, J. and Adorini, L. (1997) A peptide-binding motif for I-A(g7), the class II major histocompatibility complex (MHC) molecule of NOD and Biozzi AB/H mice. *J Exp Med*, **185**, 1013-1021.
11. Suri, A., Walters, J.J., Gross, M.L. and Unanue, E.R. (2005) Natural peptides selected by diabetogenic DQ8 and murine I-A(g7)molecules show common sequence specificity. *Journal of Clinical Investigation*, **115**, 2268-2276.
12. Thrower, S.L., James, L., Hall, W., Green, K.M., Arif, S., Allen, J.S., Van-Krinks, C., Lozanoska-Ochser, B., Marquesini, L., Brown, S. *et al*. (2009) Proinsulin peptide immunotherapy in type 1 diabetes: report of a first-in-man Phase I safety study. *Clin Exp Immunol*, **155**, 156-165.
13. Rudensky, A., Preston-Hurlburt, P., Hong, S.C., Barlow, A. and Janeway, C.A., Jr. (1991) Sequence analysis of peptides bound to MHC class II molecules. *Nature*, **353**, 622-627.
14. Sette, A., Buus, S., Appella, E., Smith, J.A., Chesnut, R., Miles, C., Colon, S.M. and Grey, H.M. (1989) Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc Natl Acad Sci U S A*, **86**, 3296-3300.
15. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, **50**, 213-219.
16. Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol*, **152**, 163-175.

17. Panagiotopoulos, C., Qin, H., Tan, R. and Verchere, C.B. (2003) Identification of a beta-cell-specific HLA class I restricted epitope in type 1 diabetes. *Diabetes*, **52**, 2647-2651.
18. Hassainya, Y., Garcia-Pons, F., Kratzer, R., Lindo, V., Greer, F., Lemonnier, F.A., Niedermann, G. and van Endert, P.M. (2005) Identification of naturally processed HLA-A2--restricted proinsulin epitopes by reverse immunology. *Diabetes*, **54**, 2053-2059.
19. Standifer, N.E., Ouyang, Q., Panagiotopoulos, C., Verchere, C.B., Tan, R., Greenbaum, C.J., Pihoker, C. and Nepom, G.T. (2006) Identification of Novel HLA-A*0201-restricted epitopes in recent-onset type 1 diabetic subjects and antibody-positive relatives. *Diabetes*, **55**, 3061-3067.
20. Rajapakse, M., Zhang, G.L., Srinivasan, K.N., Schmidt, B., Petrovsky, N. and Brusic, V. (2006) PREDNOD, a prediction server for peptide binding to the H-2g7 haplotype of the non-obese diabetic mouse. *Autoimmunity*, **39**, 645-650.
21. Chang, K.Y., Suri, A. and Unanue, E.R. (2007) Predicting peptides bound to I-Ag7 class II histocompatibility molecules using a novel expectation-maximization alignment algorithm. *Proteomics*, **7**, 367-377.
22. Chang, K.Y. and Unanue, E.R. (2009) Prediction of HLA-DQ8beta cell peptidome using a computational program and its relationship to autoreactive T cells. *Int Immunol*, **21**, 705-713.
23. Rajapakse, M., Schmidt, B., Feng, L. and Brusic, V. (2007) Predicting peptides binding to MHC class II molecules using multi-objective evolutionary algorithms. *BMC bioinformatics*, **8**, 459.
24. Bhasin, M. and Raghava, G.P.S. (2004) SVM based method for predicting HLA-DRB1(*)0401 binding peptides in an antigen sequence. *Bioinformatics*, **20**, 421-423.
25. Lata, S., Bhasin, M. and Raghava, G.P. (2007) Application of machine learning techniques in predicting MHC binders. *Methods Mol Biol*, **409**, 201-215.
26. Reche, P.A. and Reinherz, E.L. (2007) Prediction of peptide-MHC binding using profiles. *Methods Mol Biol*, **409**, 185-200.
27. Liu, Z., Cao, J., Gao, X., Ma, Q., Ren, J. and Xue, Y. (2011) GPS-CCD: a novel computational program for the prediction of calpain cleavage sites. *PLoS One*, **6**, e19001.
28. Xue, Y., Liu, Z., Gao, X., Jin, C., Wen, L., Yao, X. and Ren, J. (2010) GPS-SNO: computational prediction of protein S-nitrosylation sites with a modified GPS algorithm. *PLoS One*, **5**, e11290.

Release Note

1. November 17, 2011, the online service and the local stand-alone packages of GPS-MBA 1.0 were released.