GPS-Lipid Manual

Prediction of Protein Lipid Modifications

Version 1.0

26/06/2014

Author: Jian Ren & Yu Xue

Contact: Dr. Jian Ren, renjian.sysu@gmail.com; Dr. Yu Xue, xueyu@hust.edu.cn

The software is only free for academic research.

The latest version of GPS-LipidDB software is available from

http://lipid.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 National Basic Research Program (973 project) [2013CB933902, 2012CB911201, 2012CB910101]; National Natural Science Foundation of China [31171263, 81272578]; the Guangdong Natural Science Funds for Distinguished Young Scholar [S20120011335]; Zhujiang Nova Program of Guangzhou [2011J2200042]; Program of International S&T Cooperation [2014DFB30020].

Introduction

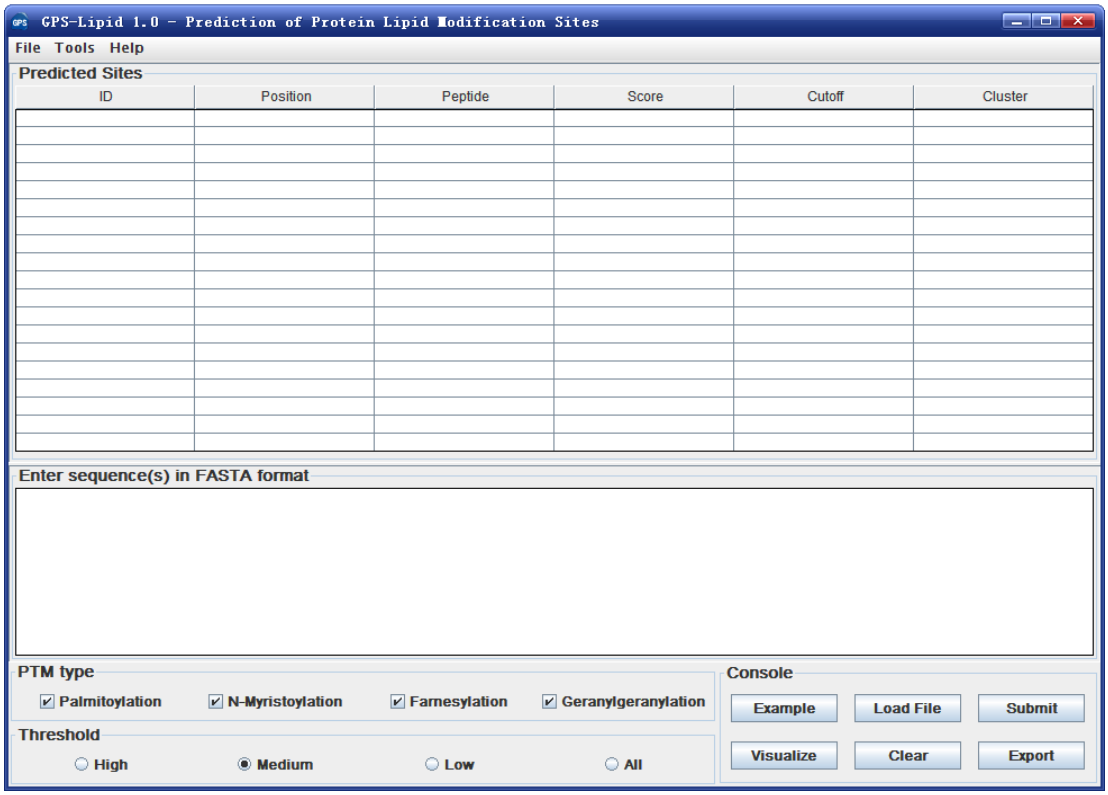
Most proteins in eucaryote cells are post-transcriptionally modified by a wide range of chemical groups. Among these modifications, the attachement of lipid groups to certain amino acids is a key modification that orchestrates the cellular protein's trafficking (1,2), signaling (3,4) and membrane association (5). With the rapid development of testing techniques, several lipid modifications, such as palmitoylation, prenylation and N-myristoylation, are now extensively studied.

Palmitoylation is a reversible post-translational modification that involves the addition of a 16-carbon chain palmitic acid on cysteine (6). As the palmityl group can effectively increase the hydrophobicity, the palmitoylation process can dynamically regulate the membrane targeting of cellular proteins (1,2). Particularly, the palmitoylation of small GTPase N-Ras and H-Ras has been extensively illustrated (7), which indicates a crucial role in modulating the signal transduction pathway (8,9). Another important type of lipid modification is prenylation. This process involves the addition of a 15-carbon farnesyl group or a 20-carbon geranylgeranyl group to a C-terminal cysteine that conform to a consensus **CAAX** motif (10). Typically, the farnesylation is catalyzed by protein farnesyltransferase (FTase) (11), whereas the geranylgeranylation is performed by protein geranylgeranyltransferase type I (GGTase-I) (12,13). However, in case of Rab proteins, the geranylgeranyltransferase type II (GGTase-II) which recognized a C-terminal **CC/CXC** motif is found to catalyze the geranylgeranylation process (10,14). In addition to the above complex mechanism, prenylation also tightly controls the signaling activities in cell by modifying small GPTase Ras (15,16), Rho (17,18) and Rab (18,19). As for the last type of lipid modification, Protein N-myristoylation is catalyzed by the N-myristoyl transferase (NMT), which recognizes a **MGXXXS/T** signature at N-terminus (20,21). It is reported that N-myristoylation can promote weak protein-protein and protein-membrane interaction (22). Moreover, with an essential role in regulating the relocation of signaling proteins, N-myristoylation functions widely in a variety of signal transduction pathways (23).

In view of the above essential physiological functions, the irregular lipid modification may lead to all sorts of diseases. As reported in published literatures (24,25), the abnormal lipid modification level is highly correlated with oncogenesis. Besides, the elevated levels of lipid modification may also linked with other severe diseases. For example, the overexpression of palmitoyl acyltransferases (PATs) may implicate in schizophrenia (26) and Huntington's disease (27). Also, the N-myristoylation is observed to mediate the viral Infectivity and eukaryotic infections (28). Taken together, the research on lipid modification will be particularly important for the diagnosis and treatment of diseases. However, due to the limitations of

integrative bioinformatics resources, the overall investigations that focusing on the co-regulation of lipid modifications are seldom performed. This deficiencies may grievously hamper the development of effective therapies for disorders related to lipid modifications.

In this work, we present GPS-Lipid, which is a comprehensive predictor for protein lipid modification sites. From the literatures published before November, 2014, we manually collected **737** S-palmitoylation sites in **361** proteins, **106** S-farnesylation sites in **97** proteins, **95** S- geranylgeranylation sites in **70** proteins and **283** N-myristoylation sites in **281** proteins. To obtain a high-performance predictor, we applied the Particle Swarm Optimization with an aging leader and challengers (ALC-PSO) (29) in our previous Group-based Prediction System (GPS) algorithm for model training and predicting. For convenience, an online service was developed using PHP + JavaScript, and is freely available at http://lipid.biocuckoo.org.

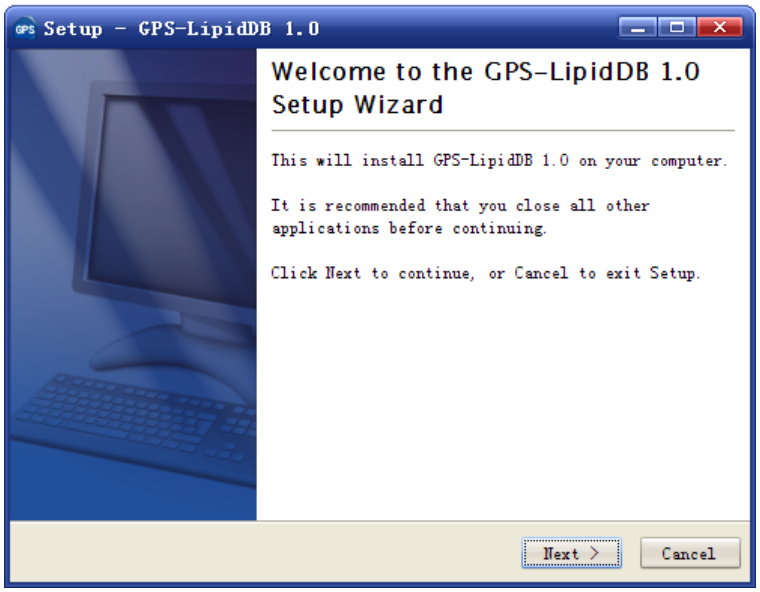


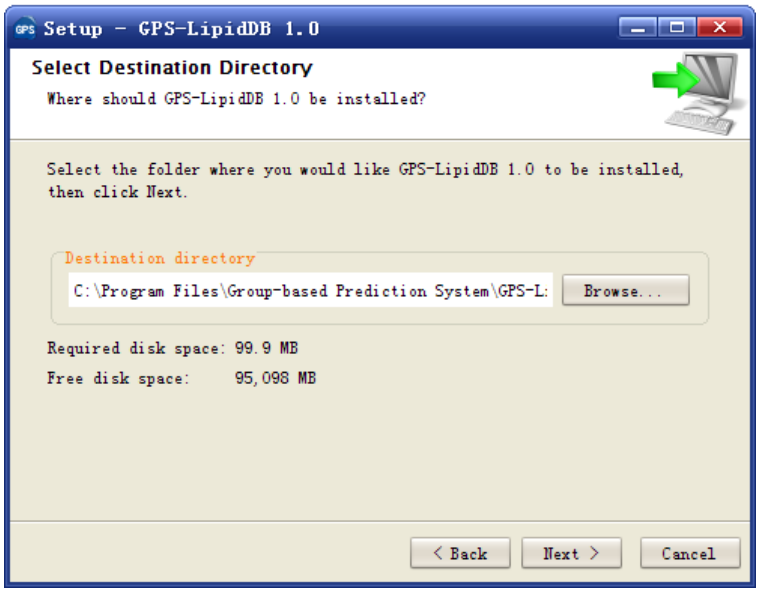
GPS-Lipid v1.0 Main Interface

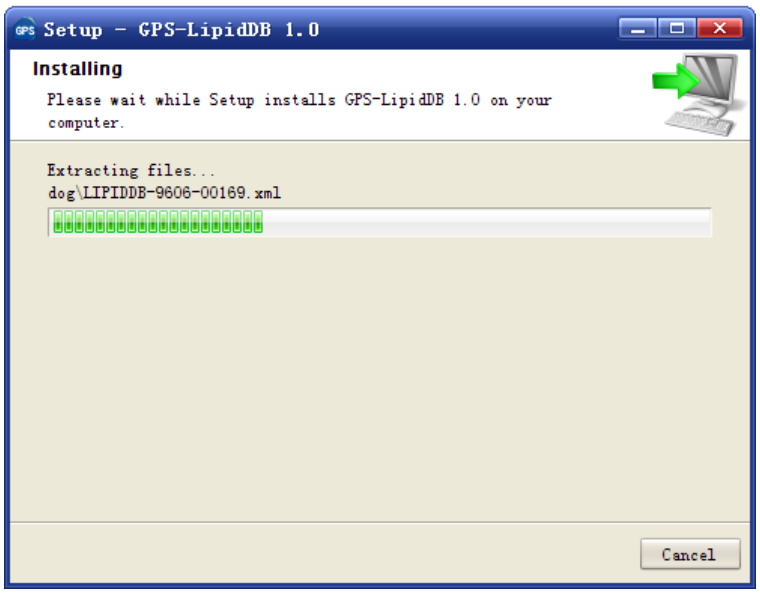
Download & Installation

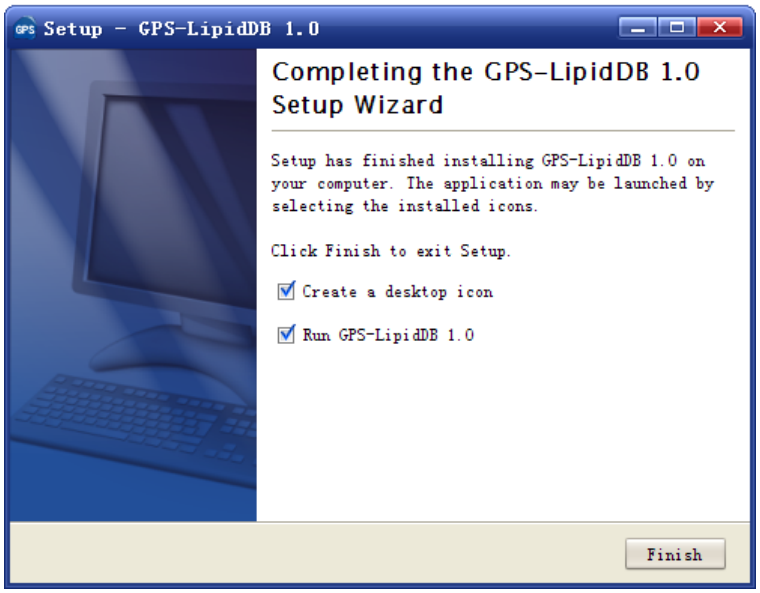
The local packages of GPS-Lipid 1.0 database were implemented in JAVA, and could be installed on Windows, Mac OS X or Unix/Linux systems. The latest distributions of GPS-Lipid could be found at http://lipid.biocuckoo.org/download.php. We recommend that users could download the latest release.

After downloading, please double-click on the install package to begin installation. Follow the user prompts through the installation. And snapshots of the setup program are shown below:









Click on the **Finish** button to complete the setup program.

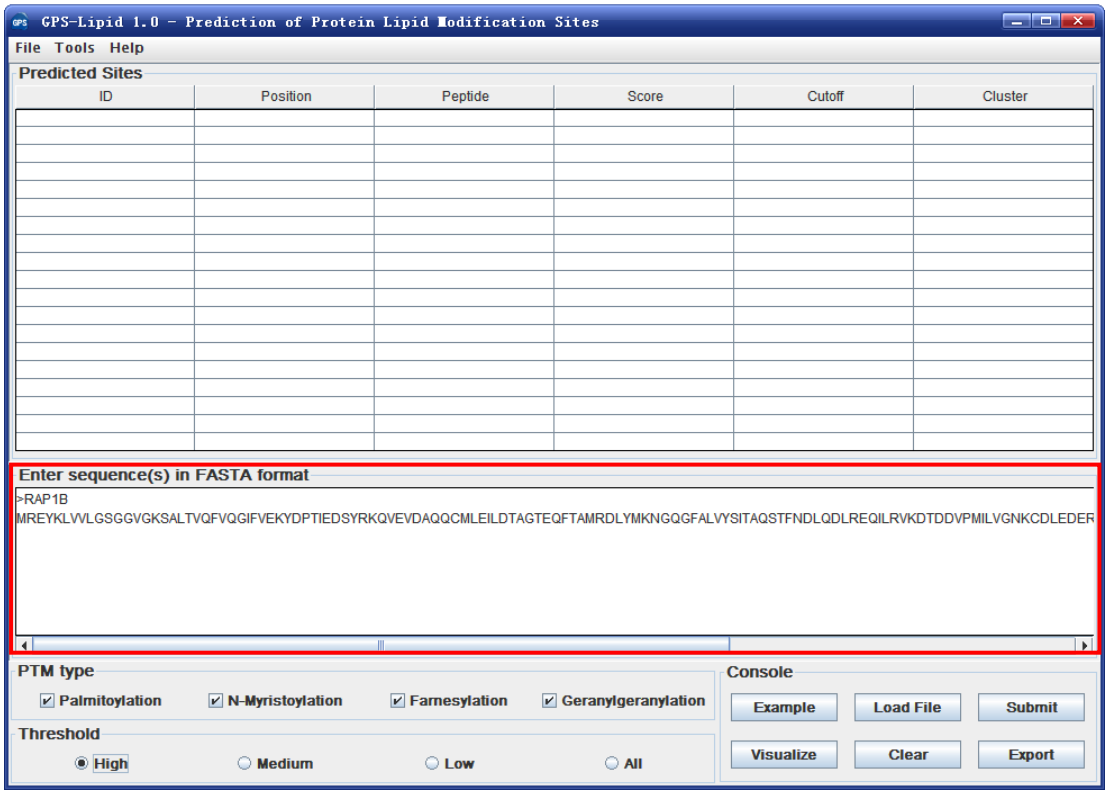
The usage of GPS-Lipid

In GPS-Lipid 1.0, we provide a predictor for lipid modification sites prediction. With Group-based Prediction System (GPS) algorithm, the prediction models for S-Palmitoylation, N-Myristoylation, S-Farnesylation and S- Geranylgeranylation were trained.

A single protein sequence in FASTA format

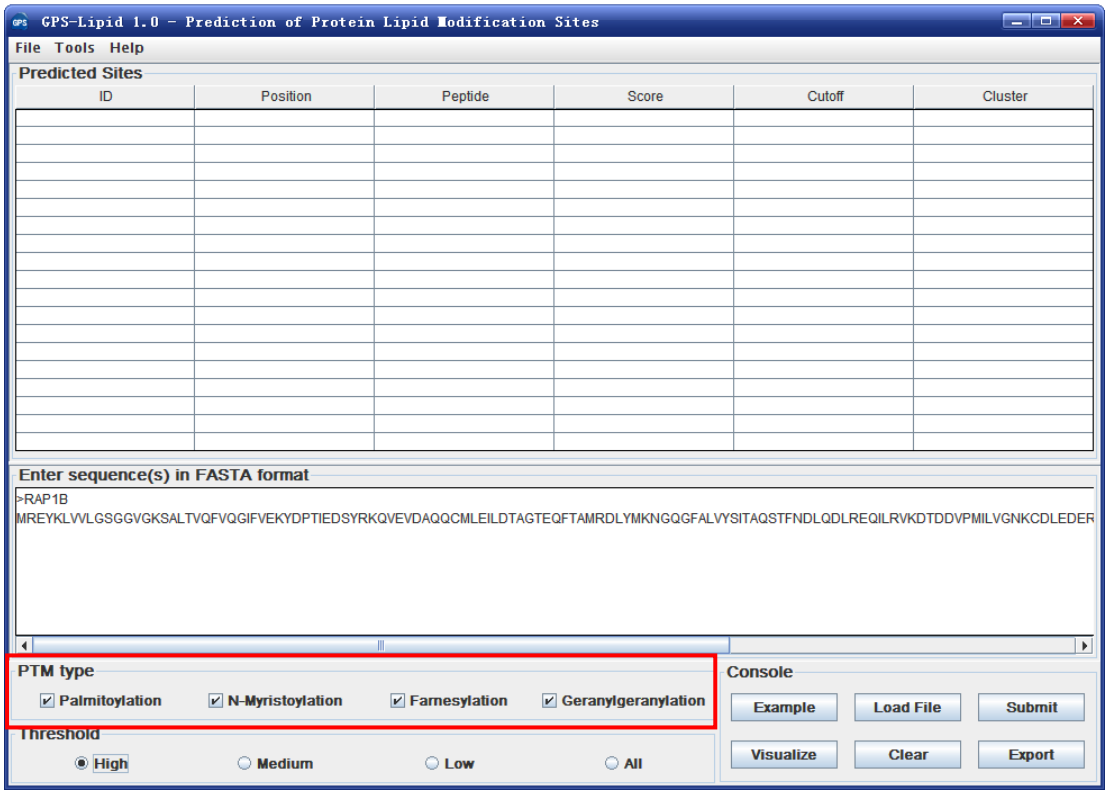
The following steps show you how to use the GPS-Lipid 1.0 to predict lipid modification sites for a single protein sequence in FASTA format.

(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 CSS-Palm 4.0Firstly, 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-Lipid.

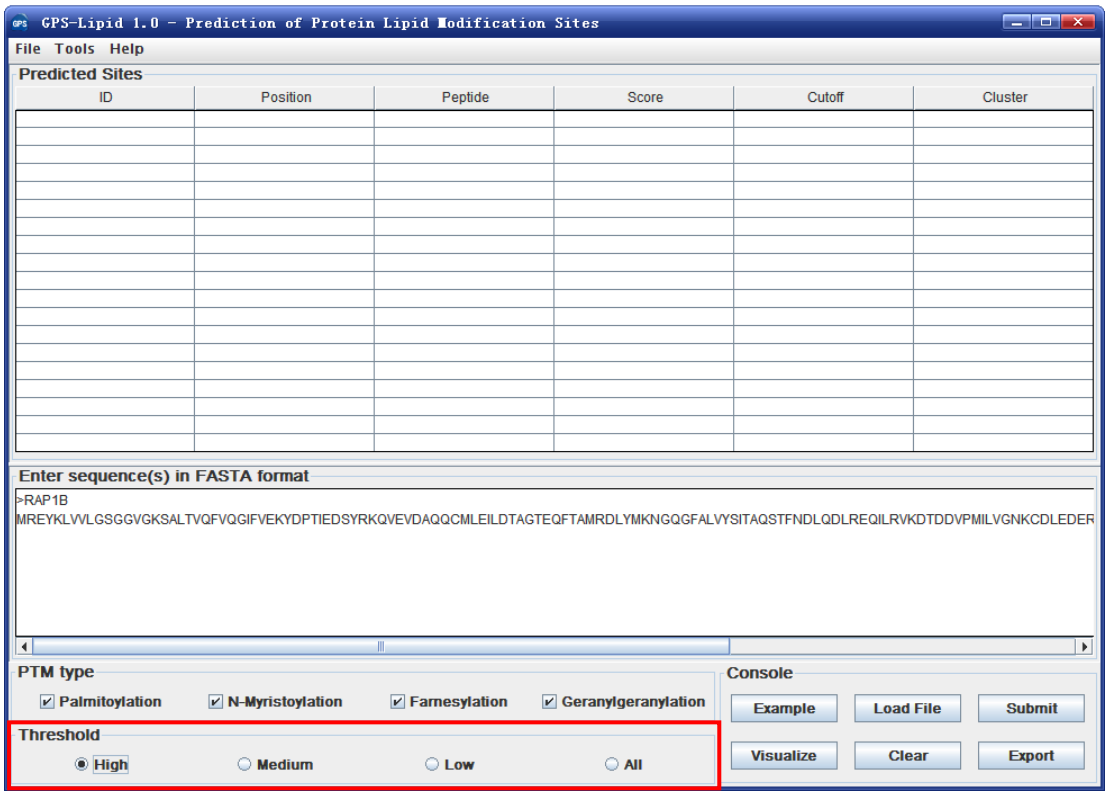


Note: For a single protein, the sequence without a name in raw format is also OK. However, for multiple sequences, the name of each protein should be presented.

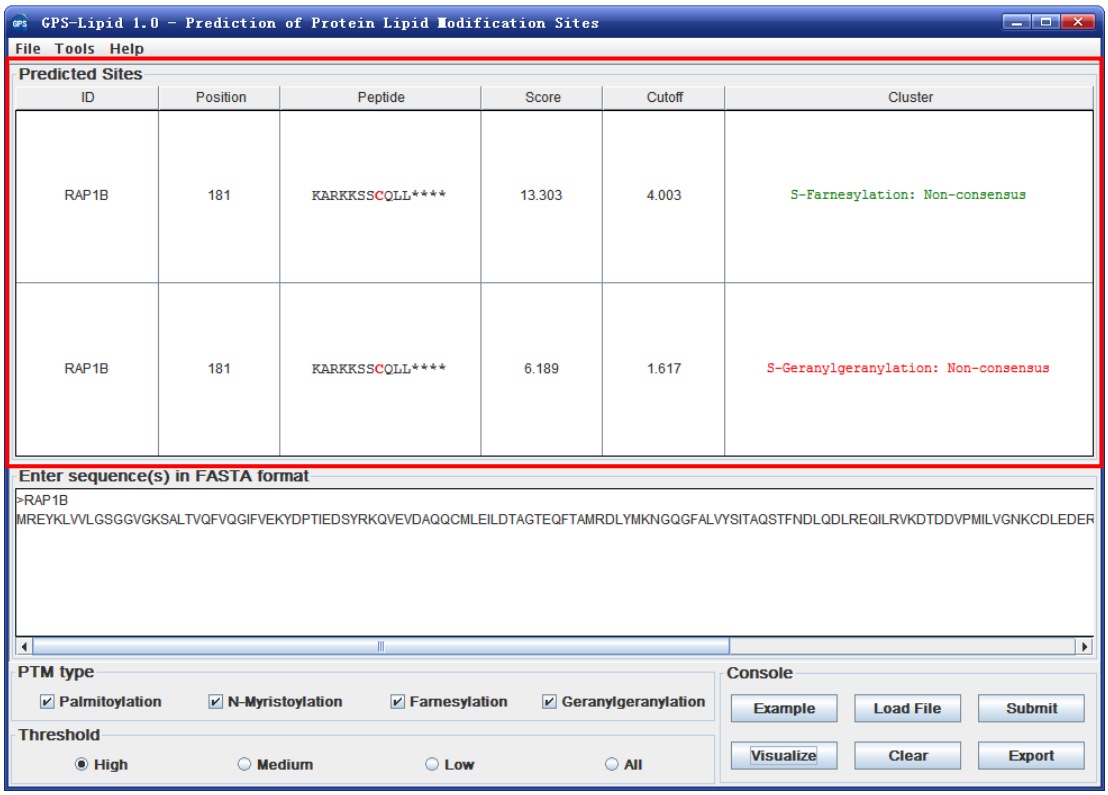
(2) Choose the **modification types** that you want to predict. There are four major types of lipid modifications provided in GPS-Lipid. By default, GPS-Lipid will predict all four types of modification sites.



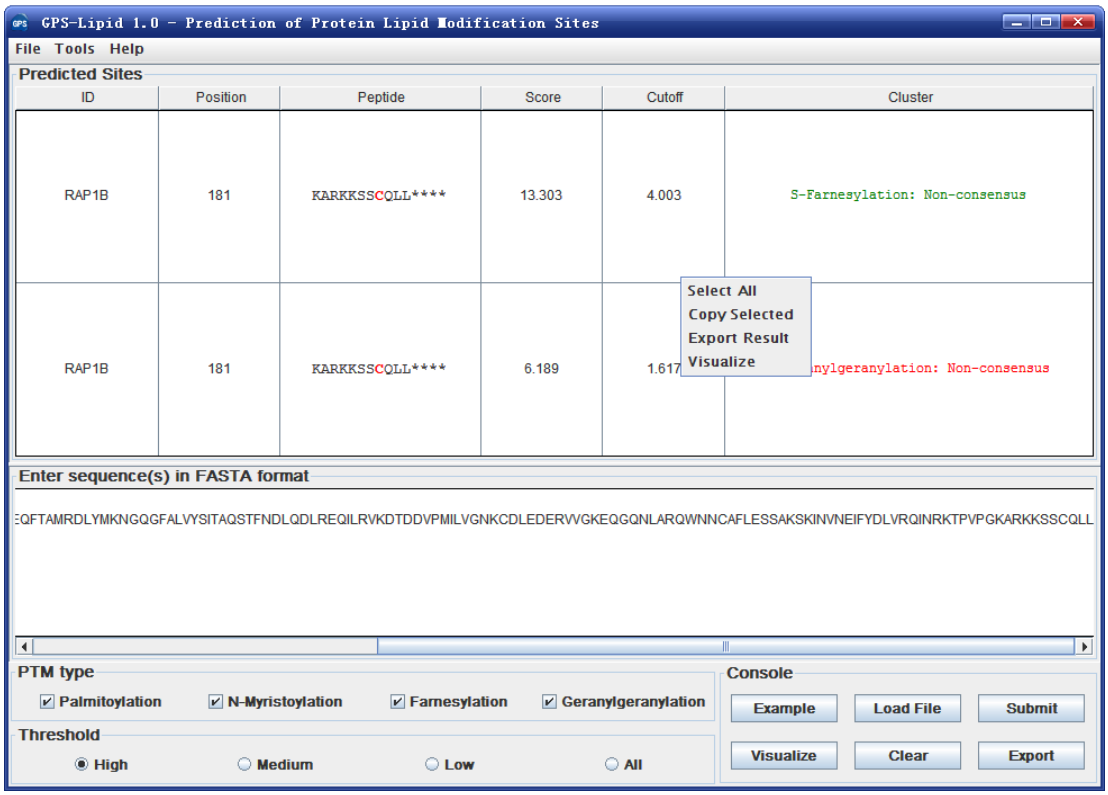
(3) Choose a **Threshold** that you need, the default cut-off is Medium.



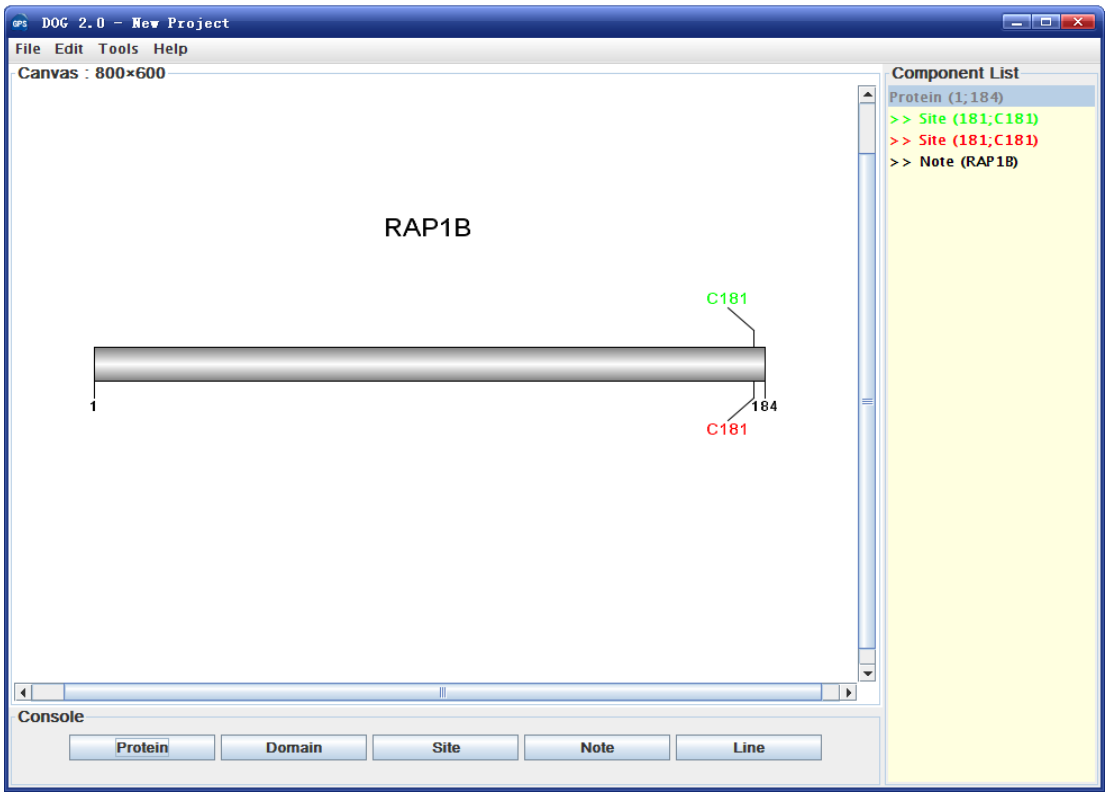
(4) Click on the **Submit** button, then the predicted lipid modification sites will be shown. To distinguish different types of lipid modifications, we colored the results based on their predicted PTM types. As shown below, the S-Palmitoylation is colored with blue, the N-Myristoylation is in orange, the S-Farnesylation is colored with green, and the S-geranylgeranylation is in red.



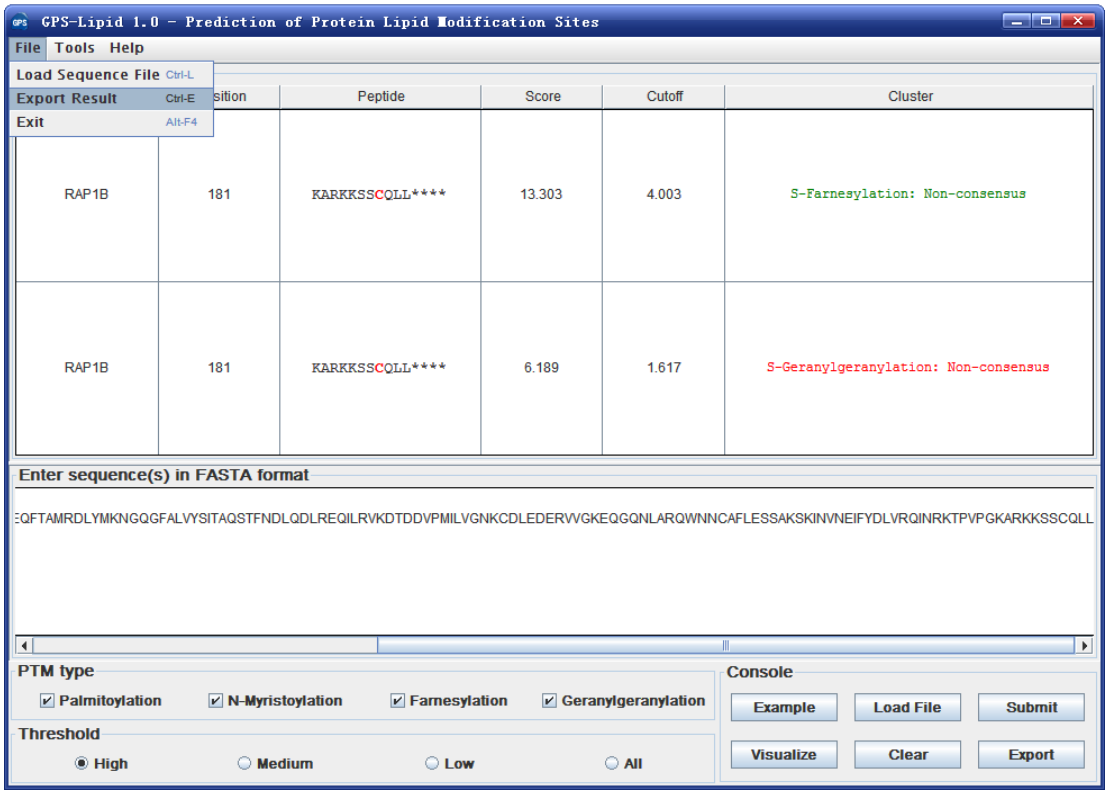
(5) Then please click on the **RIGHT** button in the prediction form. You can use the "**Select All**" and "**Copy Selected**" to copy the selected results into Clipboard. Then please copy the results into a file, eg., an EXCEL file for further consideration. Also, you can choose "**Export Prediction**" to export the prediction results into a tab-delimited text file.



To visualize the predicted site please click on the "**Visualize**" button. Again, different types of modification sites will be marked in different color.



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

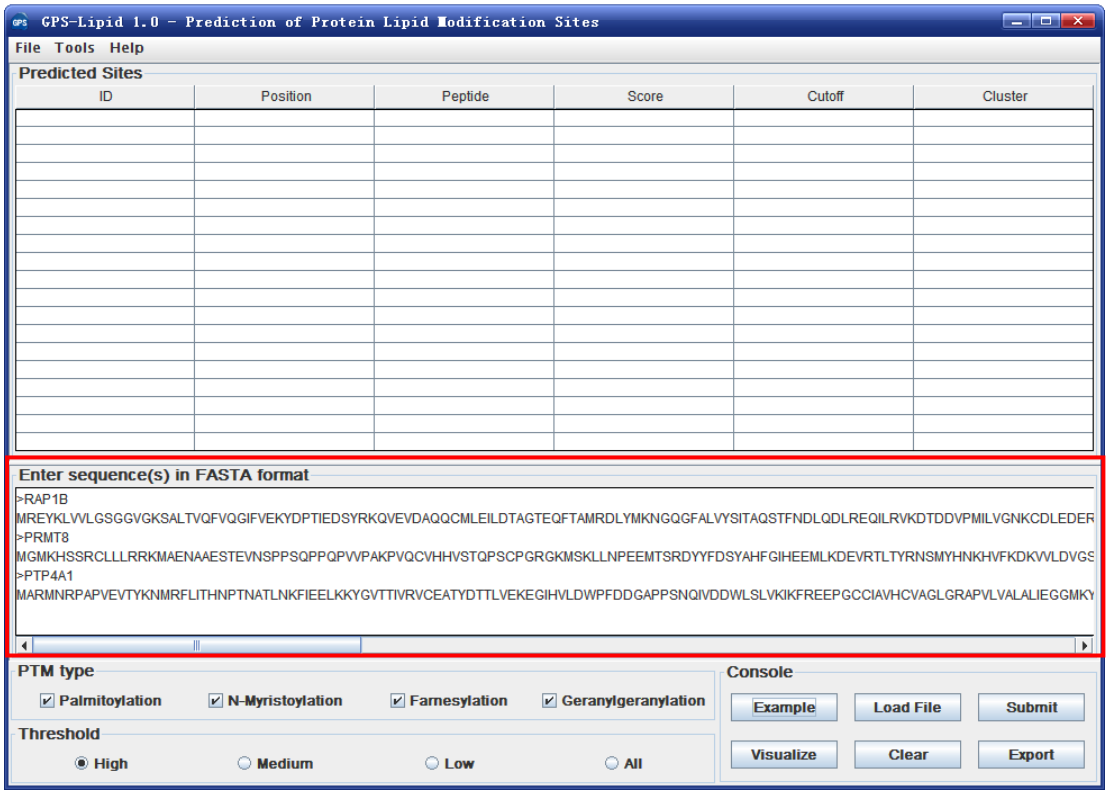

GPS-Lipid 1.0 - Prediction of Protein Lipid Modification Sites
File
Tools
Help
Load Sequence File Ctrl-L
Export Result Ctrl-E
Exit Alt-F4
Peptide
Score
Cutoff
Cluster
RAP1B
181
KARKKSSCQLL****
13.303
4.003
S-Farnesylation: Non-consensus
RAP1B
181
KARKKSSCQLL****
6.189
1.617
S-Geranylgeranylation: Non-consensus
Enter sequence(s) in FASTA format
EQFTAMRDLYMKNGQGFALVYSITAQSTFNDLQDLREQILRVKDTDDVPMILVGNKCDLEDERVVGKEQGQNLARQWNNCAFLESSAKSKINVNEIFYDLVRQINRKTPVPGKARKKSSCQLL
PTM type
Palmitoylation
N-Myristoylation
Farnesylation
Geranylgeranylation
Threshold
High
Medium
Low
All
Console
Example
Load File
Submit
Visualize
Clear
Export

Multiple protein sequences in FASTA format

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

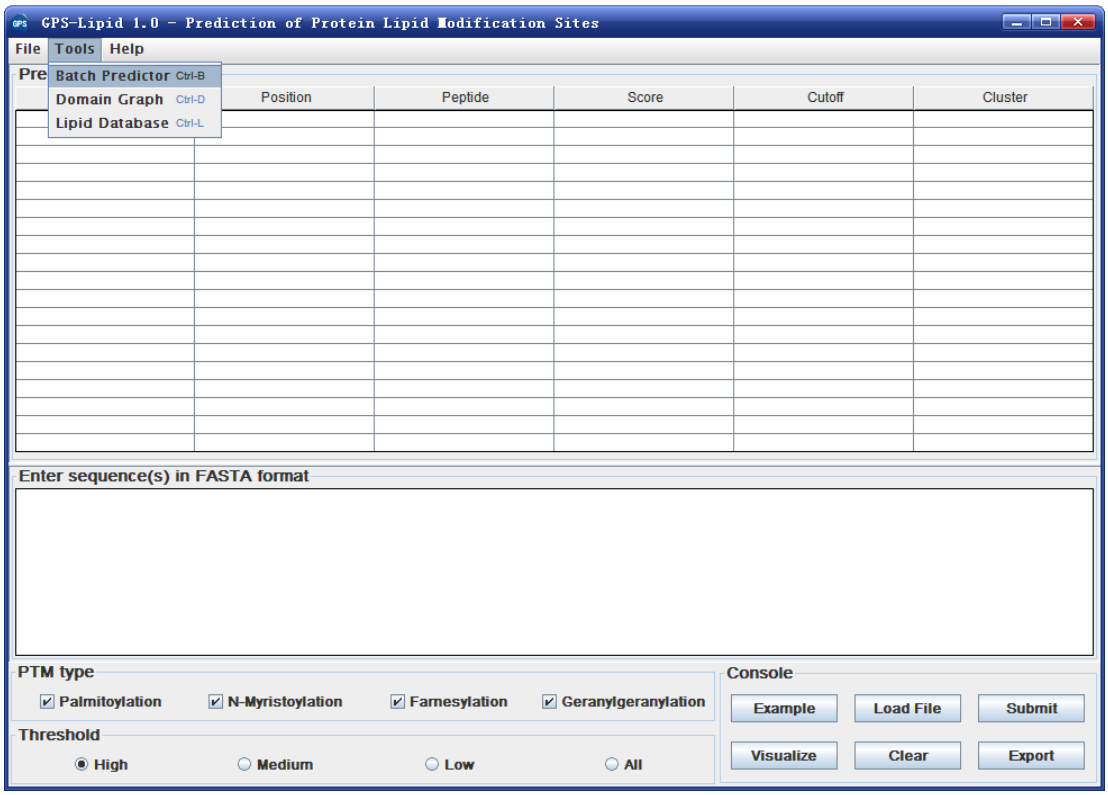
A. Input the sequences into text form directly. (Num. of Seq ⩽ 2,000)

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-Lipid 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.

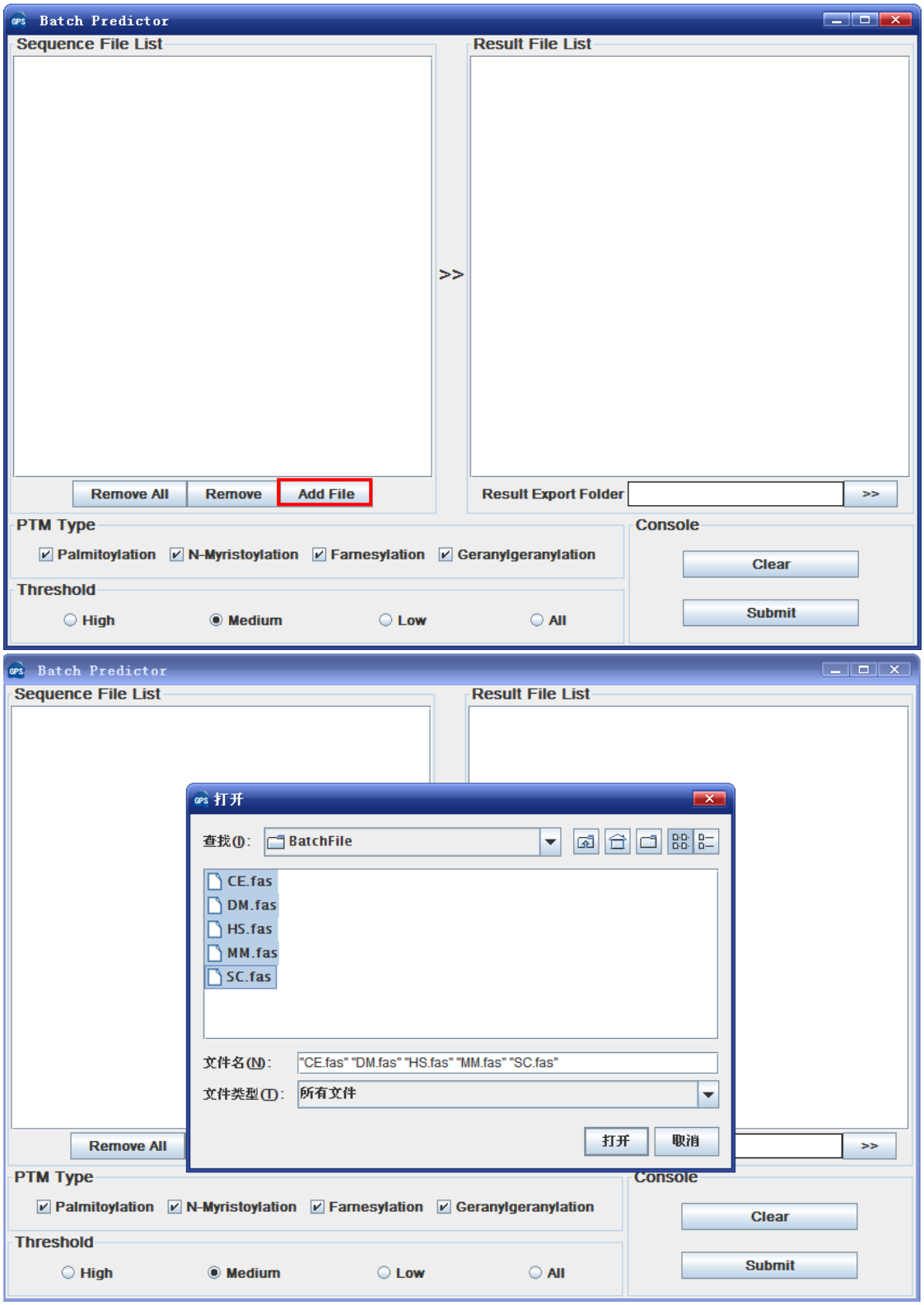


The following steps show you how to use it:

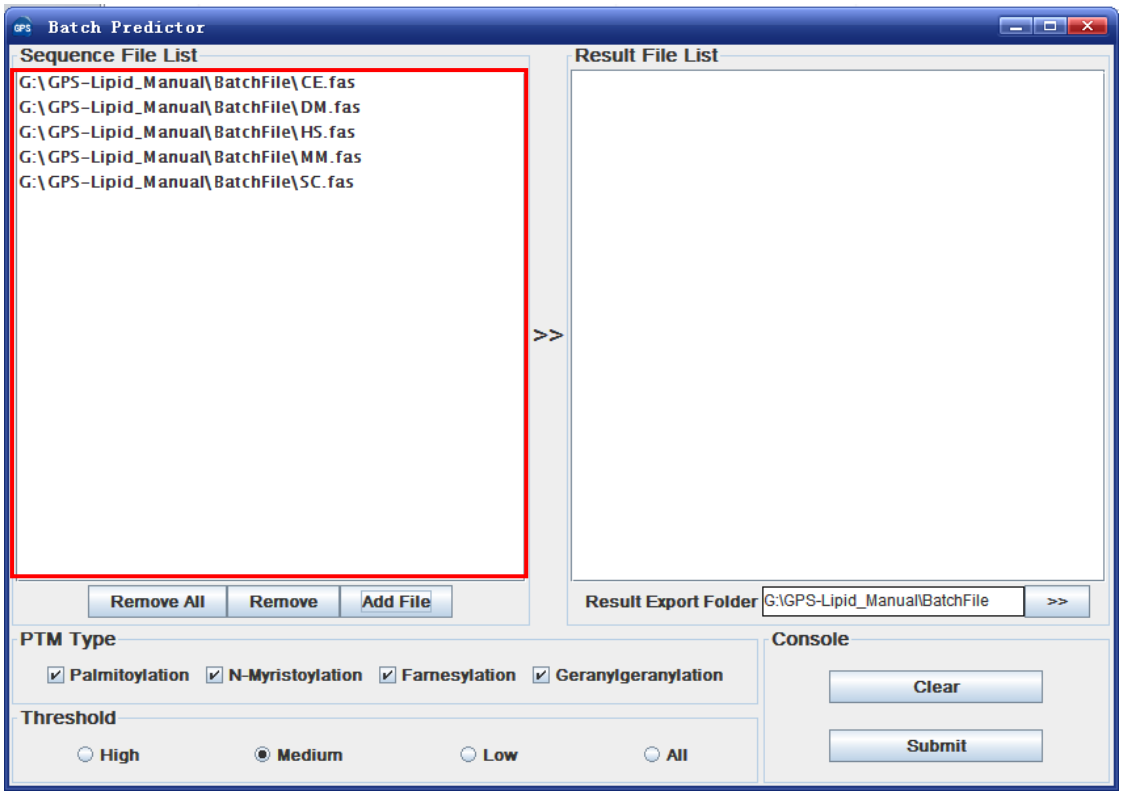
(1) Put protein sequences into one or several files (eg., SC.fas, CE.fas, and etc) with FATSA format as below:

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

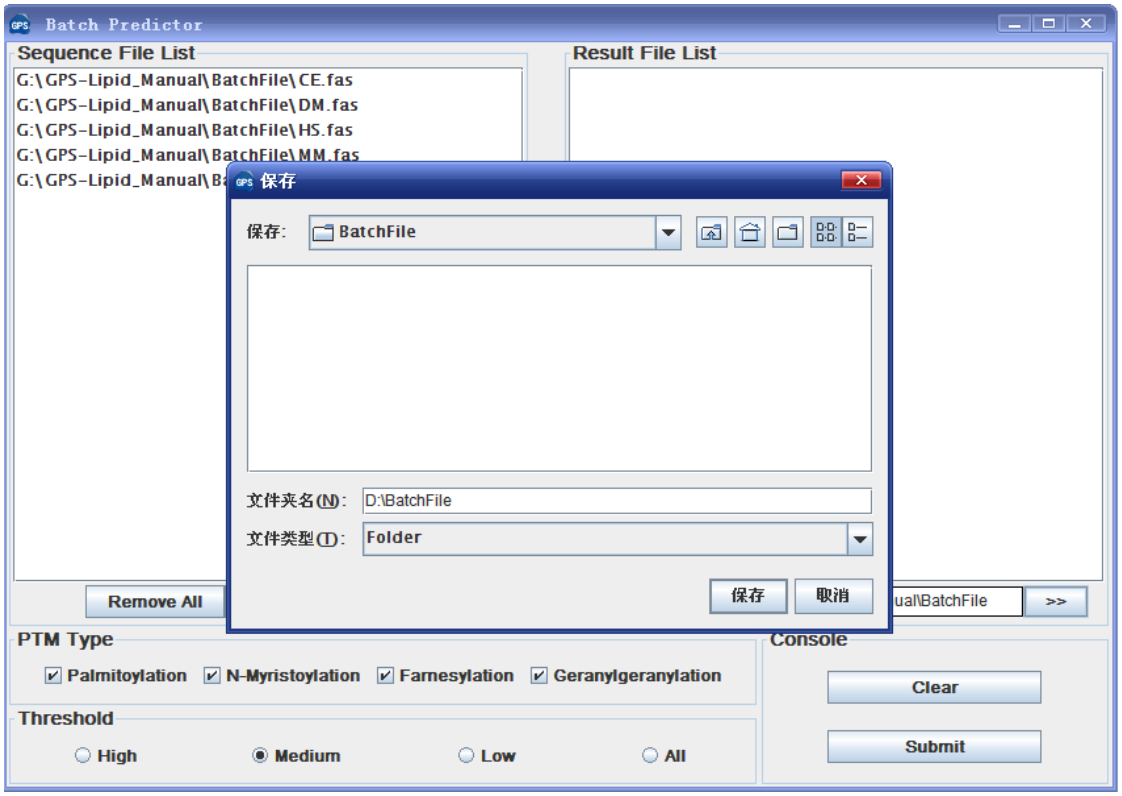
(2) Click on the **Batch Predictor** button and then click on the **Add File** button and add one or more protein sequence files in your hard disk.



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

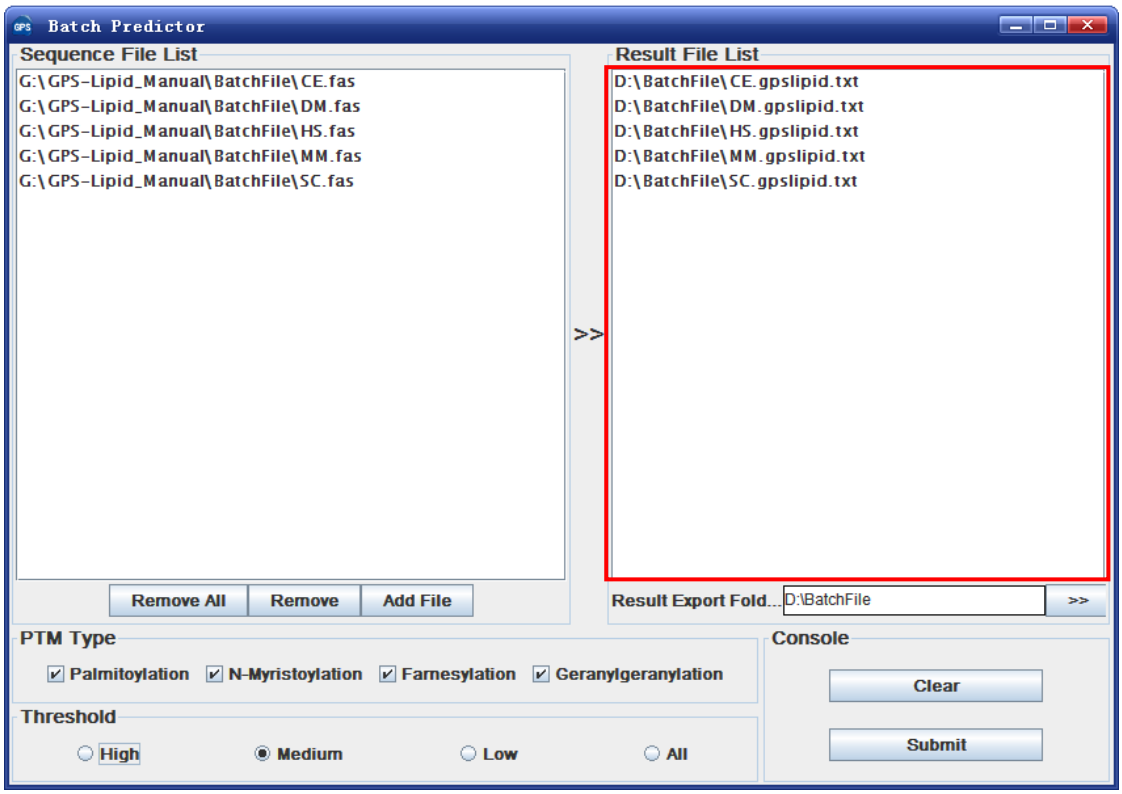


(3) The output directory of prediction results should also be defined. Please click on the >> button to specify the export fold.



(4) Please choose a proper modification type and threshold before prediction. Then

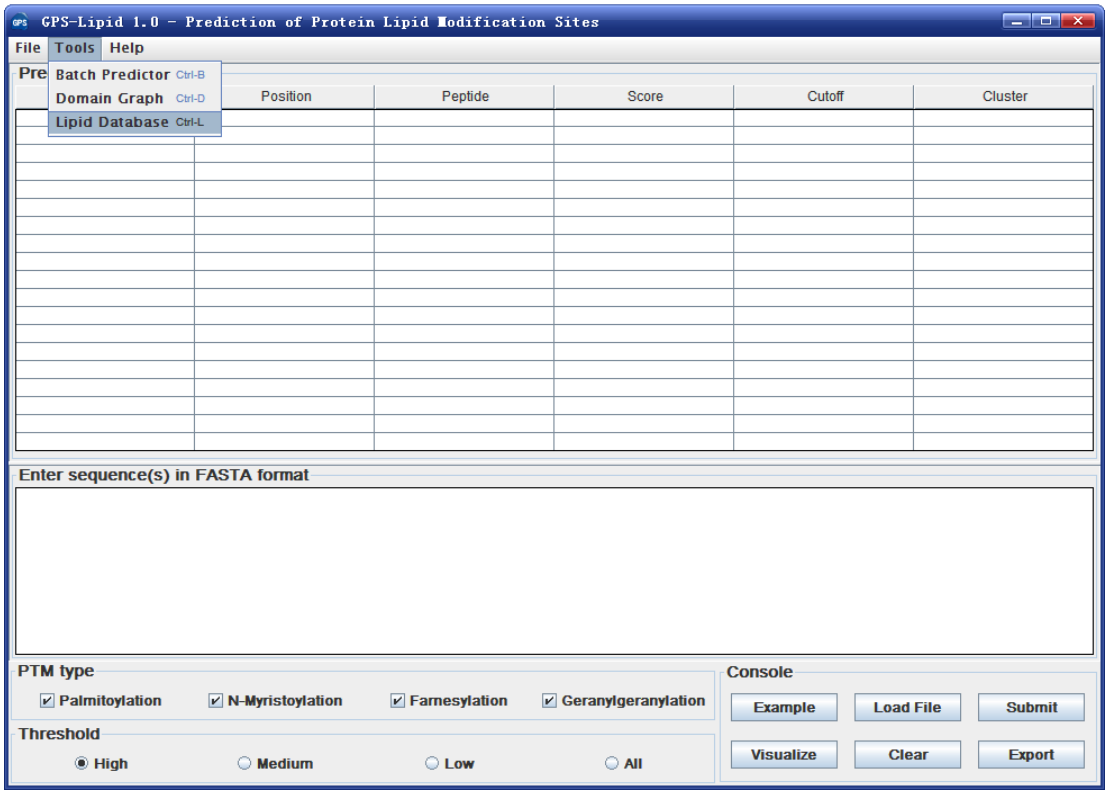
please click on the **Submit** button, after that, the **Batch Predictor** begin to process all of the sequence files that have been added to the list. The result of prediction will be export to the **Result Export Fold,** and the name of result files will be shown in the **Result File List**.



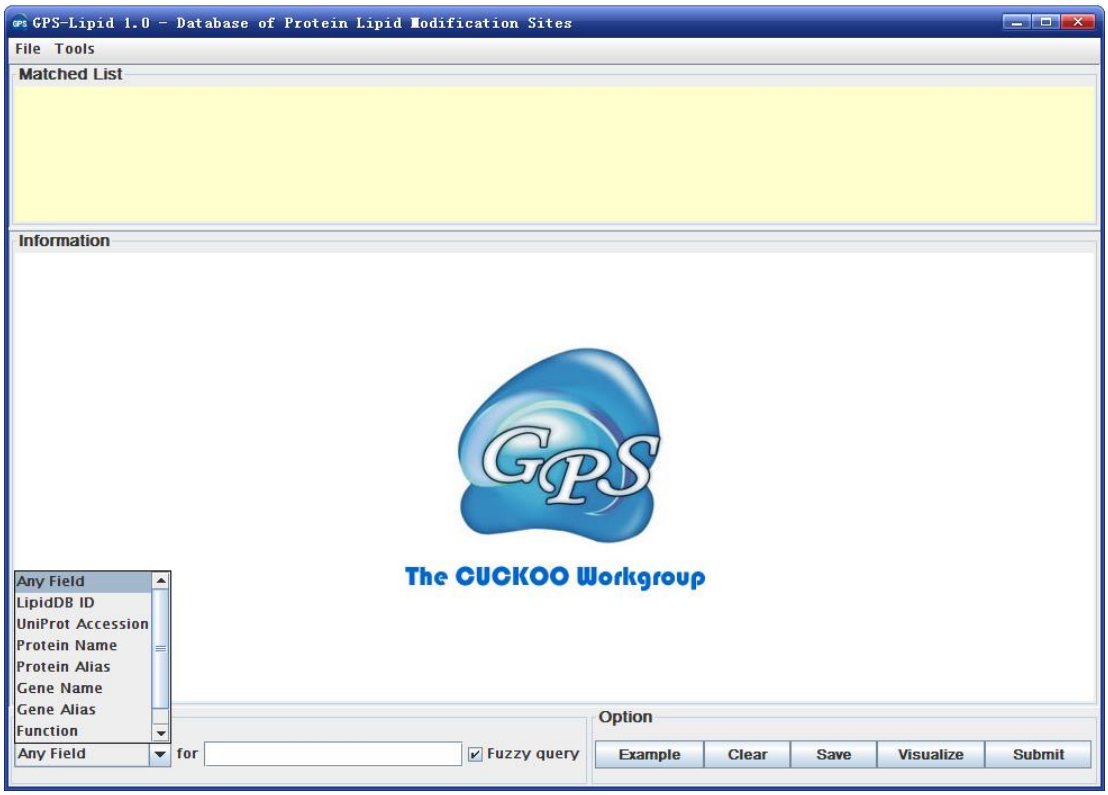
The usage of Lipid database

Search

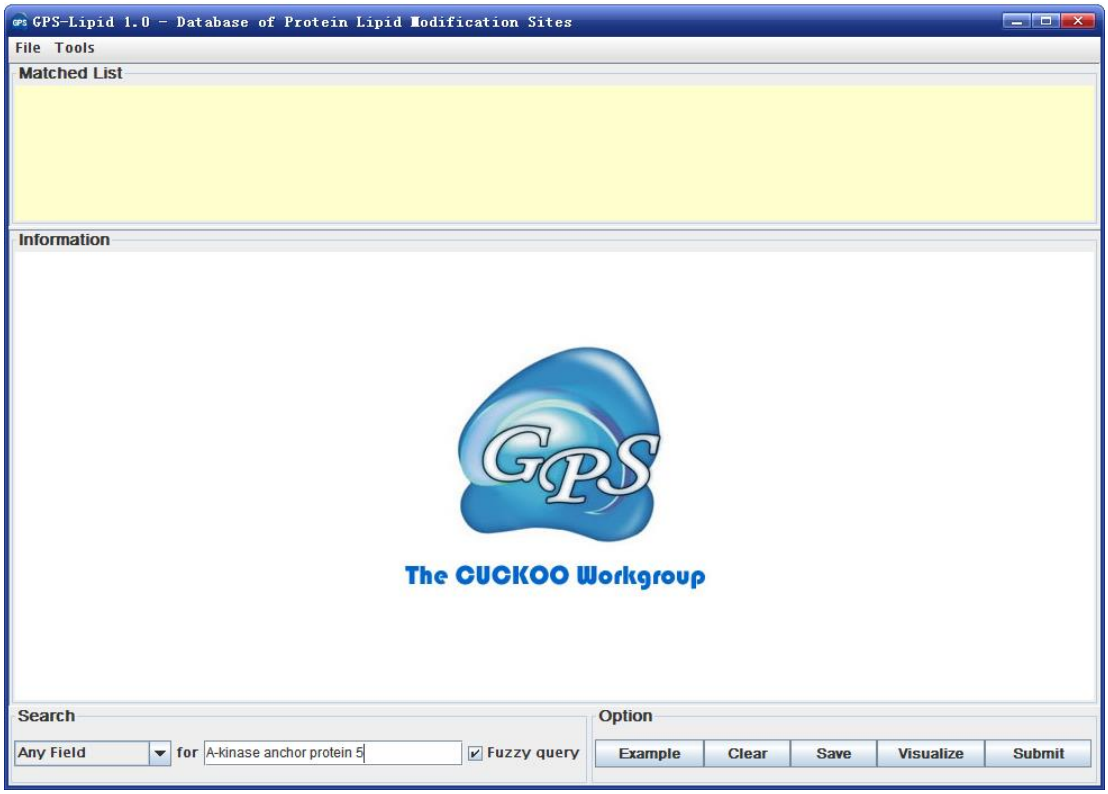
The GPS-Lipid was also provide a database for lipid modifications. Click on the **Tools -> Lipid Database** menu to open the Lipid database.



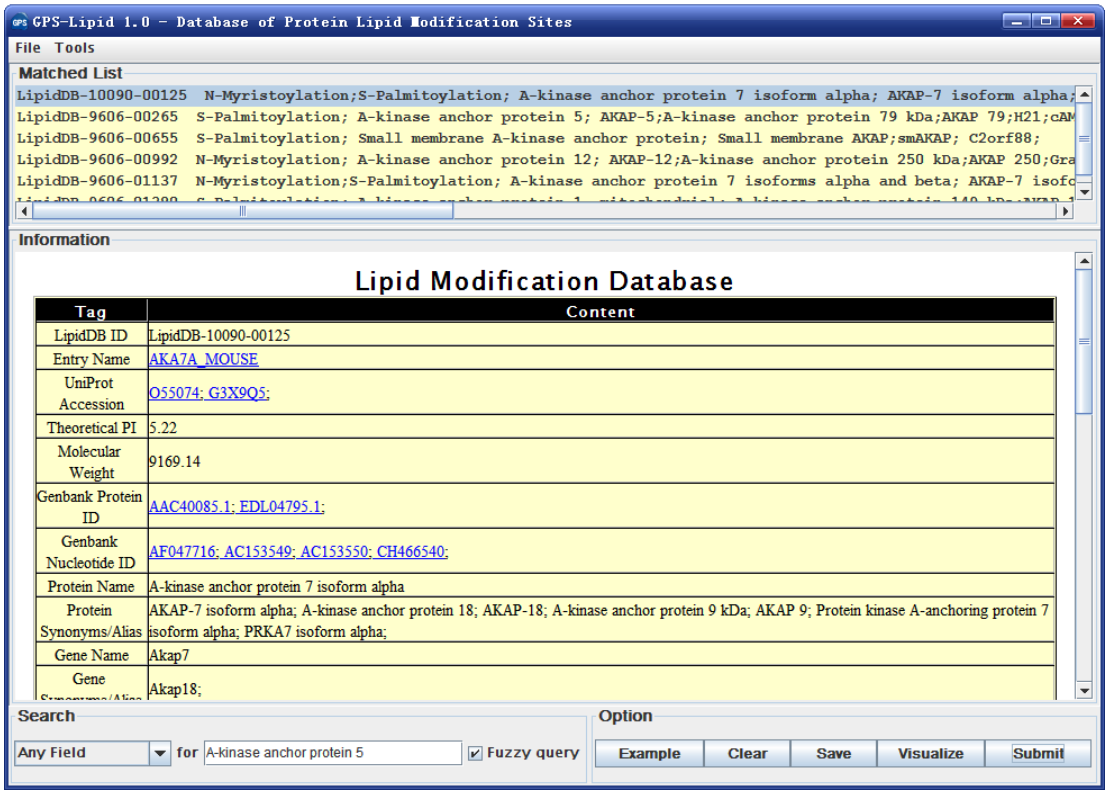
For simple search, users could input a LipidDB ID with LipidDB-XXXX-XXXXXX, a UniProt ID (Q07065), protein/gene names/aliases (eg., cytoskeleton-associated protein 4) or functions. Users could click the “Example” button one or several times to view the instances.



For example, users could input a protein/gene name/aliase, e.g., A-kinase anchor protein 5, specify the "**Any Field**", and then click on the "**Submit**" button to search the related information for this protein.

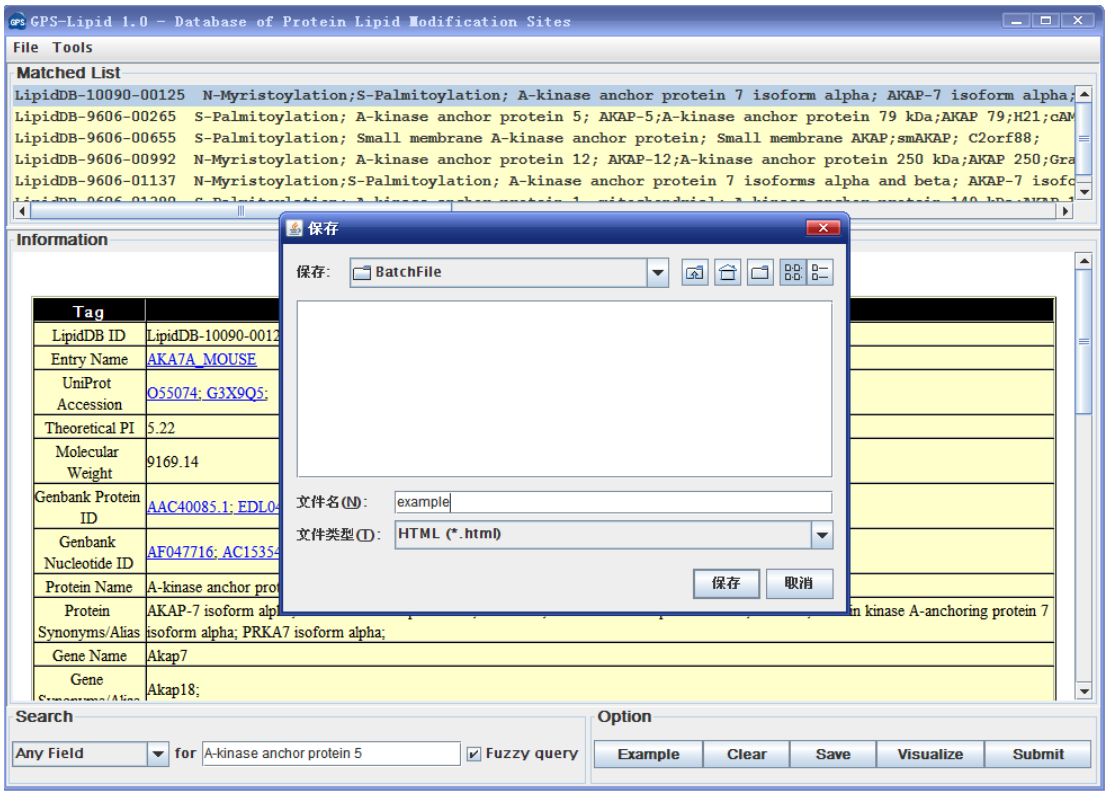


Then the information for human kinase anchor protein 5 will be shown in the "Information" form.

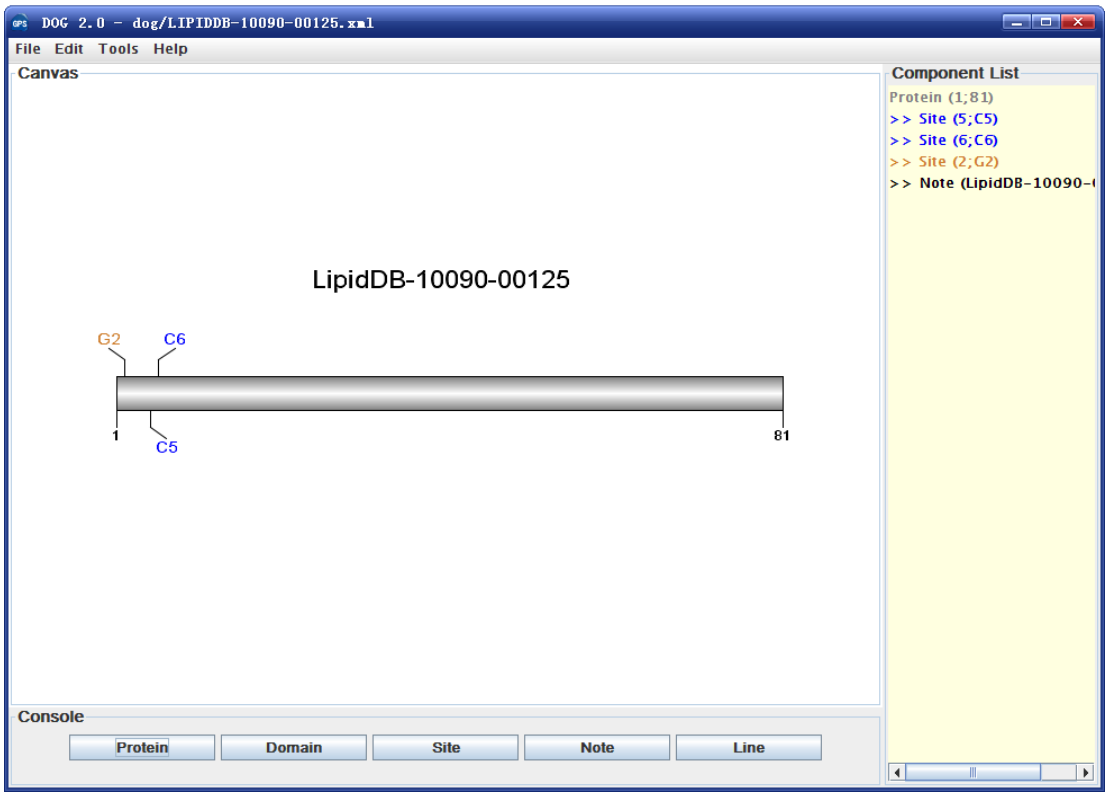


The protein and nucleotide sequences for this protein could be shown by clicking on the related NCBI or Uniprot identifiers:

The searched results could be saved in an HTML format.

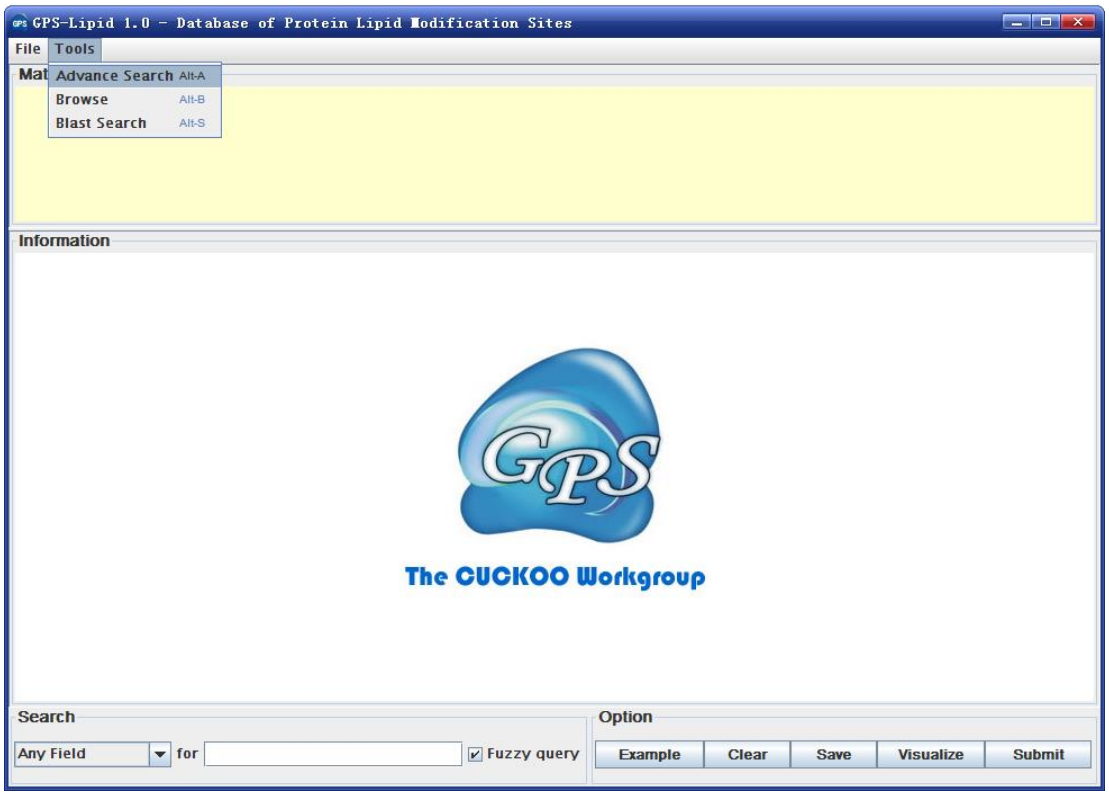


Also, when clicking on the **Visualize** button, you can view the schematic diagram for your search result.

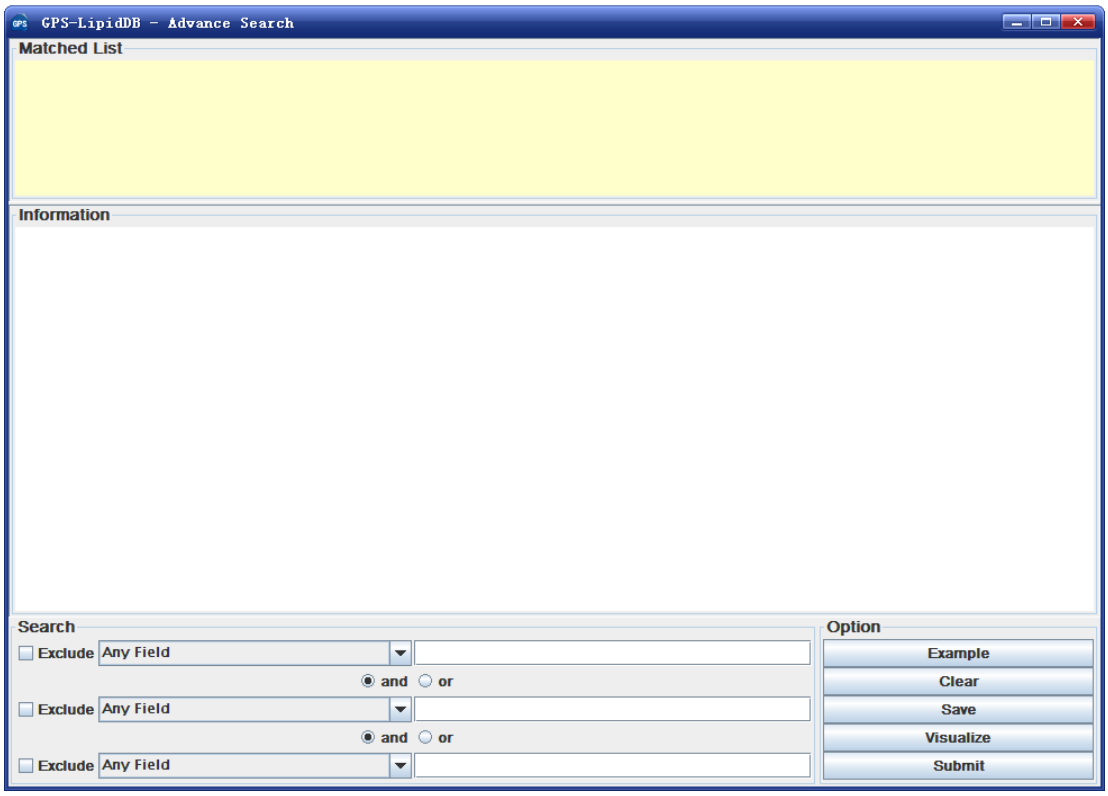


Advance Search

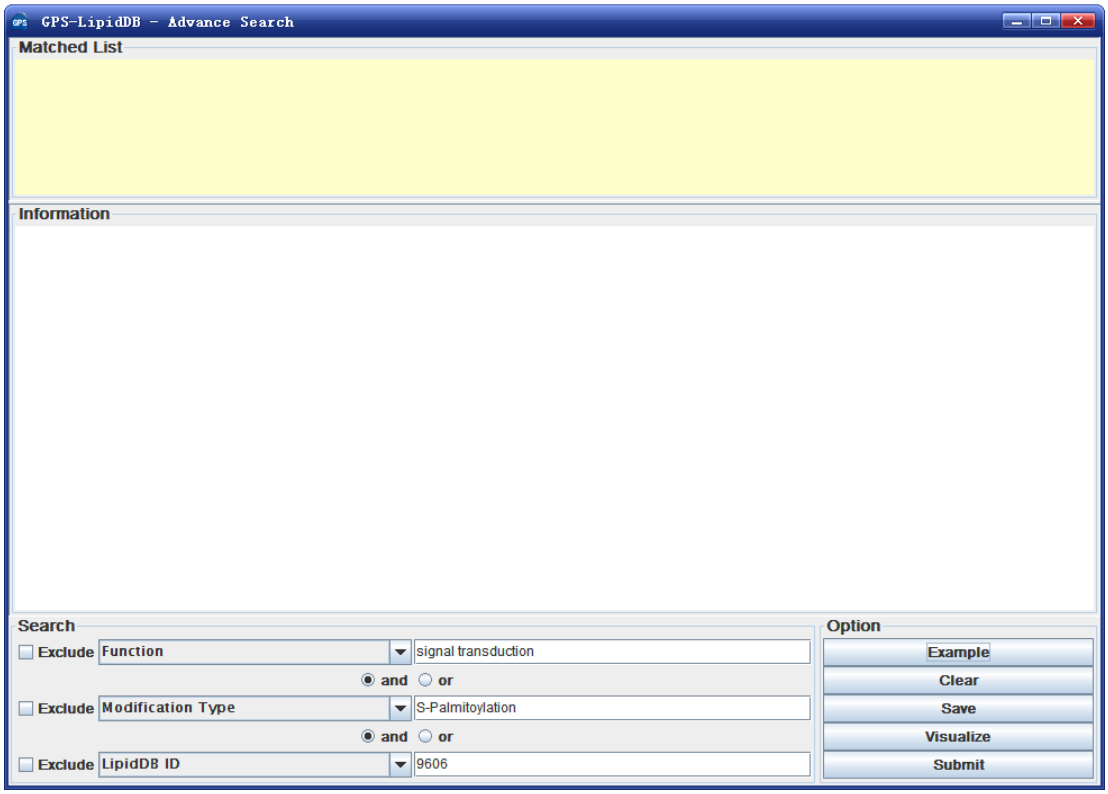
The GPS-Lipid 1.0 database supports three advance options, including advance search, browse, BLAST search. The Advance search option allows you to input up to three terms to find the information more specifically. The querying fields can be empty if fewer terms are needed.



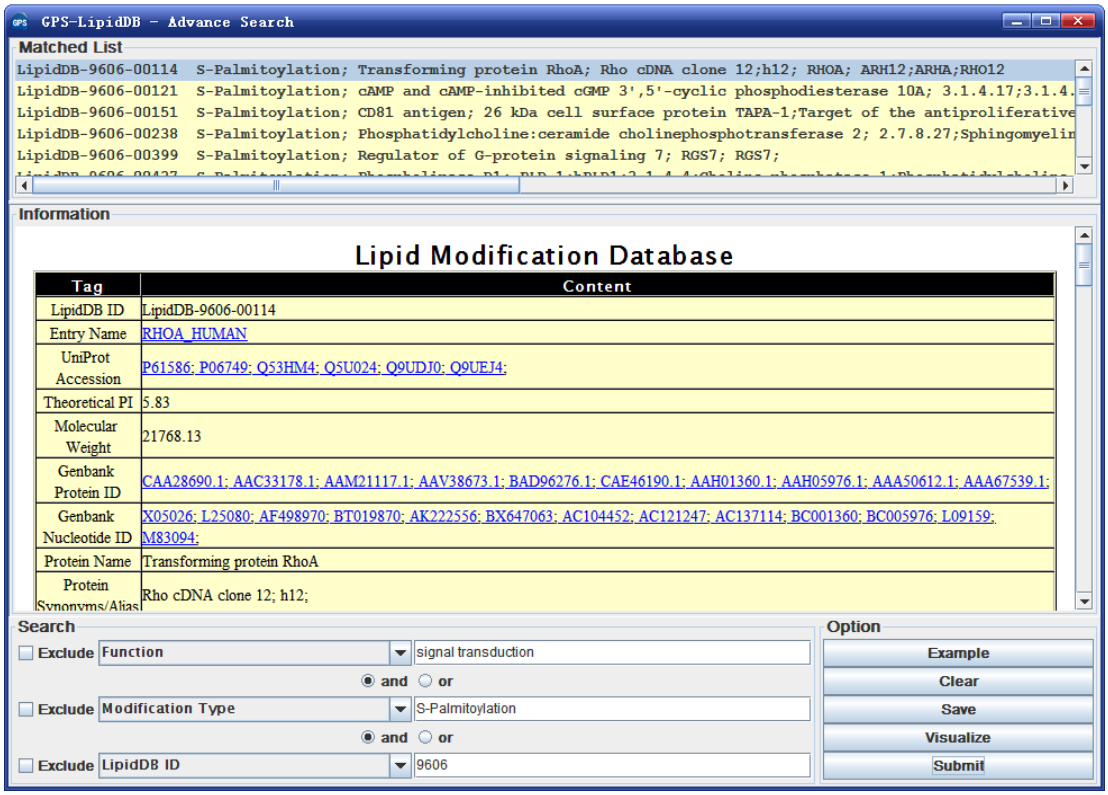
First, users could click on the "**Tools**" button then click on the "**Advance Search**" button to open this option.



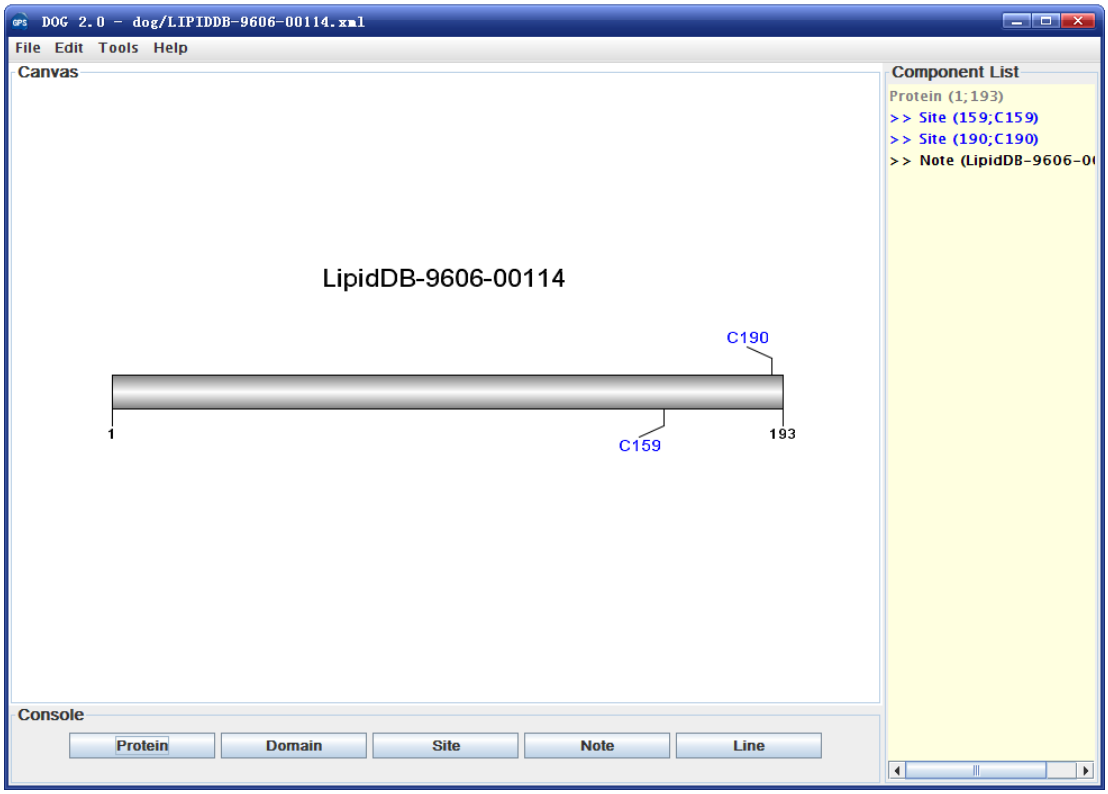
By clicking the "**Example**" button, you can try an instance for usage. You can input Phosphatidylcholine (Protein Name), S-Palmitoylation (Modification type), and 9606 (LipidDB ID) for querying.



Then the human's sphingomyelin synthase 2 will be shown.

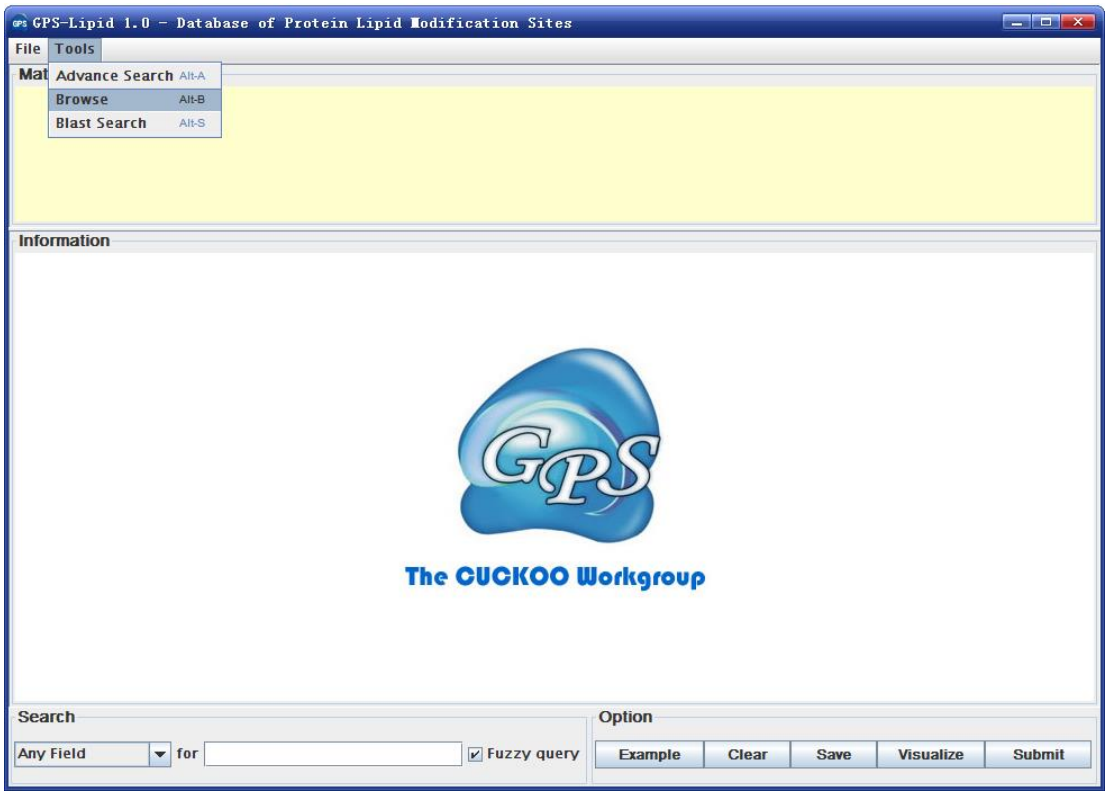


Notably, by clicking the "**Visualize**" button, you can open a schematic diagram of the human's sphingomyelin synthase 2 with a palmitoylation site marked in blue.

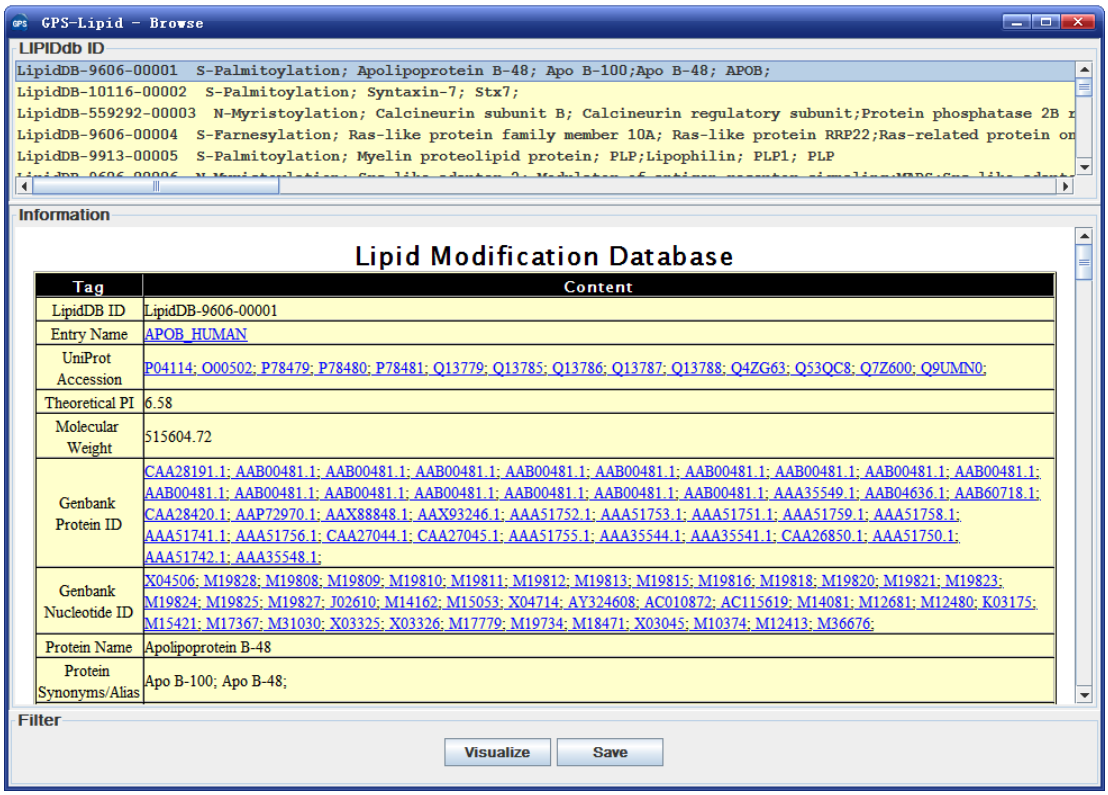


Browse

The GPS-Lipid 1.0 database supports the browse function. The Browse search allows users to view all entries in GPS-Lipid 1.0 database.



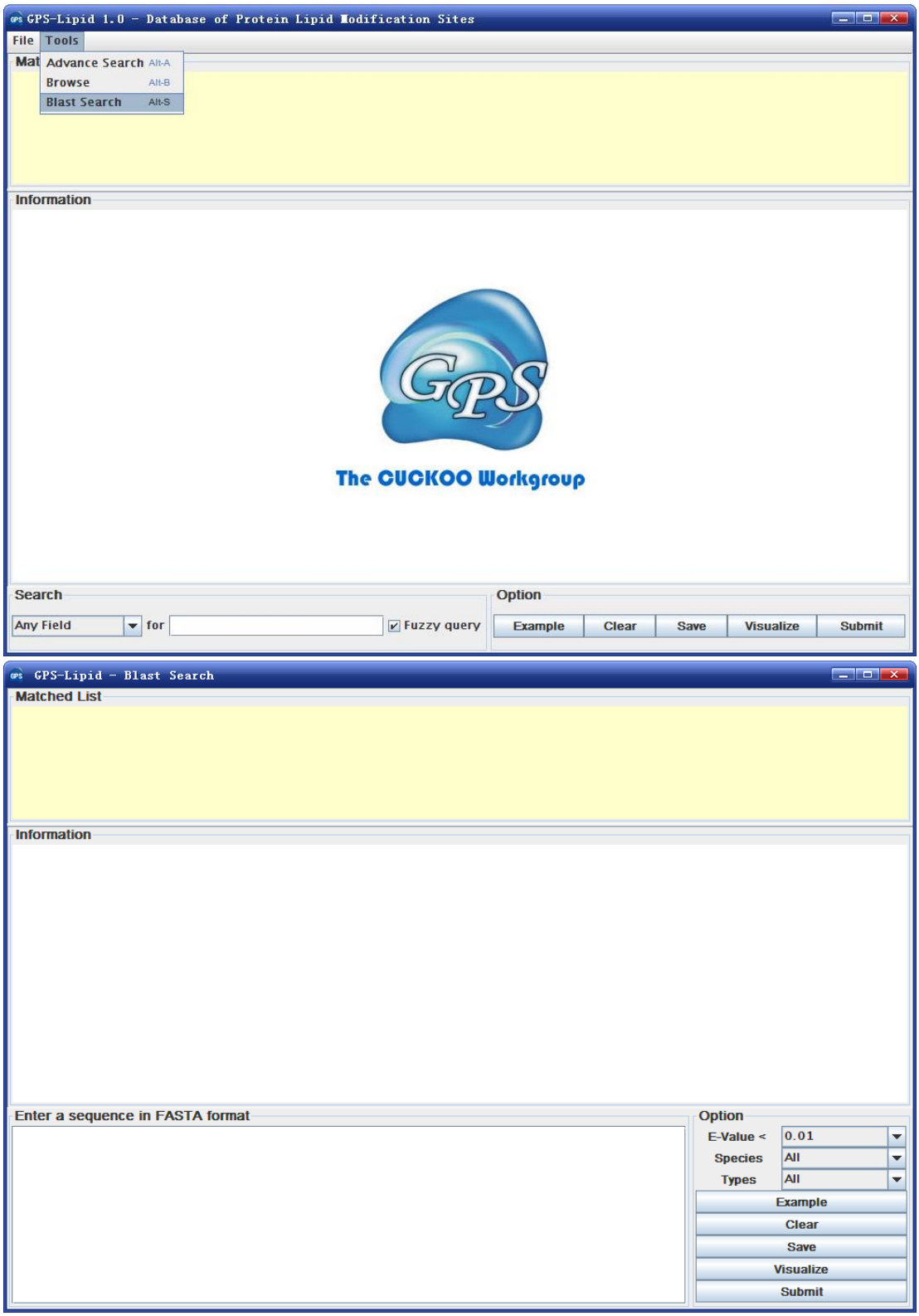
First, users could click on the "**Tools**" button then click on the "**Browse**" button to visualize all GPS-Lipid proteins. Users could visualize any protein by click on the entries listed in the "LipidDB ID" form.



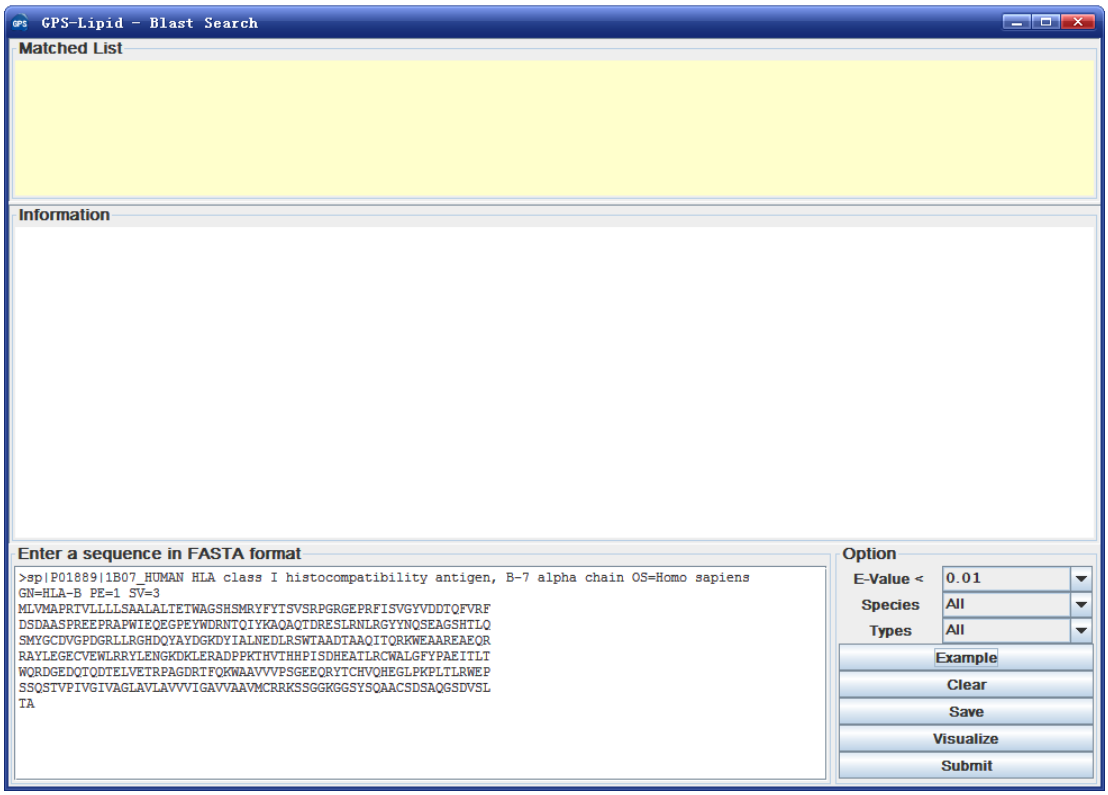
The result can be saved in HTML format or viewed in a schematic form.

Blast Search

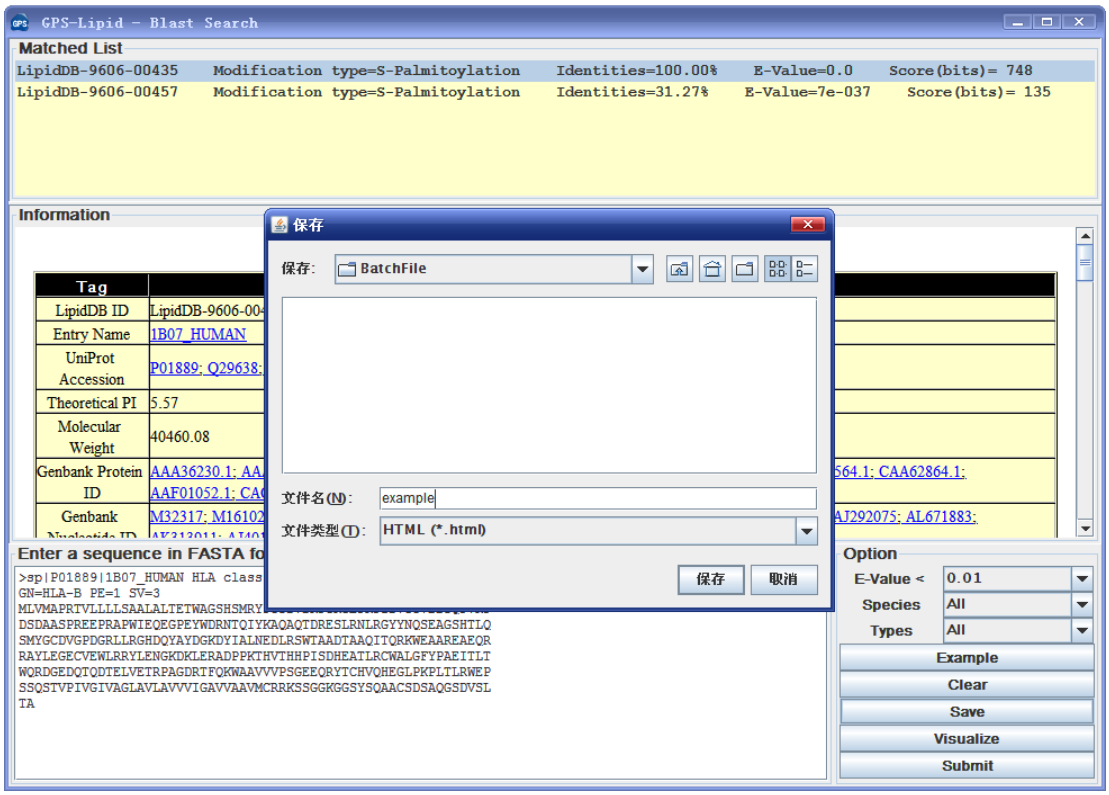
The GPS-Lipid 1.0 database also supports the searching function by sequence alignment. The blastp program from NCBI BLAST+ packages was included in GPS-Lipid 1.0 database. Users could input one protein (not mRNA sequence) in FASTA or RAW format a time to search identical or homologous entries. First, users could click on the "**Tools**" button then click on the "**Blast Search**" button to open the Blast search window.



Then users could either click on the "**Example**" button in the Option form or directly input a protein sequence in FASTA or RAW format. Please note that only one protein is permitted a time. Then please click on the "Submit" button to search identical or homologous entries. The E-value cut-off could be user-defined in the Option form.



Again, users could visualize any GPS-LipidDB proteins by clicking on the entries listed in the "Matched list" form. And the results could be saved by clicking on the "**Save**" button in the Option form.


GPS-Lipid - Blast Search
Matched List
LipidDB-9606-00435 Modification type=S-Palmitoylation Identities=100.00% E-Value=0.0 Score(bits)= 748
LipidDB-9606-00457 Modification type=S-Palmitoylation Identities=31.27% E-Value=7e-037 Score(bits)= 135
Information
保存
保存: BatchFile
文件名(N): example
文件类型(T): HTML (*.html)
保存
取消
Tag
LipidDB ID
Entry Name
1B07_HUMAN
UniProt Accession
P01889; Q29638;
Theoretical PI
5.57
Molecular Weight
40460.08
Genbank Protein ID
Genbank
Enter a sequence in FASTA fo
>sp|P01889|1B07_HUMAN HLA class
GN=HLA-B PE=1 SV=3
MLVMAPRTVLLLLSAALALTETWAGSHSMRY
DSDAASPREEPRAPWIEQEGPEYWDRNTQIYKAQAQTDRESLRNLRGYYNQSEAGSHTLQ
SMYGCDVGPDGRLLRGHDQYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQR
RAYLEGECVEWLRRYLENGKDKLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLT
WQRDGEDQTQDTELVETRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEP
SSQSTVPIVGIVAGLAVLAVVVIGAVVAAVMCRRKSSGGKGGSYSQAACSDSAQGSDVSL
TA
Option
E-Value < 0.01
Species All
Types All
Example
Clear
Save
Visualize
Submit

References

1. Draper, J.M., Xia, Z. and Smith, C.D. (2007) Cellular palmitoylation and trafficking of lipidated peptides. *Journal of lipid research*, **48**, 1873-1884.
2. Linder, M.E. and Deschenes, R.J. (2007) Palmitoylation: policing protein stability and traffic. *Nature reviews. Molecular cell biology*, **8**, 74-84.
3. Smotrys, J.E. and Linder, M.E. (2004) Palmitoylation of intracellular signaling proteins: regulation and function. *Annual review of biochemistry*, **73**, 559-587.
4. Resh, M.D. (2004) Membrane targeting of lipid modified signal transduction proteins. *Sub-cellular biochemistry*, **37**, 217-232.
5. Levental, I., Grzybek, M. and Simons, K. (2010) Greasing their way: lipid modifications determine protein association with membrane rafts. *Biochemistry*, **49**, 6305-6316.
6. Nadolski, M.J. and Linder, M.E. (2007) Protein lipidation. *The FEBS journal*, **274**, 5202-5210.
7. Greaves, J. and Chamberlain, L.H. (2007) Palmitoylation-dependent protein sorting. *The Journal of cell biology*, **176**, 249-254.
8. Casey, P.J. (1995) Protein lipidation in cell signaling. *Science*, **268**, 221-225.
9. Kurayoshi, M., Yamamoto, H., Izumi, S. and Kikuchi, A. (2007) Post-translational palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *The Biochemical journal*, **402**, 515-523.
10. Maurer-Stroh, S. and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome biology*, **6**, R55.
11. Hougland, J.L., Hicks, K.A., Hartman, H.L., Kelly, R.A., Watt, T.J. and Fierke, C.A. (2010) Identification of novel peptide substrates for protein farnesyltransferase reveals two substrate classes with distinct sequence selectivities. *Journal of molecular biology*, **395**, 176-190.
12. Gangopadhyay, S.A., Losito, E.L. and Hougland, J.L. (2014) Targeted reengineering of protein geranylgeranyltransferase type I selectivity functionally implicates active-site residues in protein-substrate recognition. *Biochemistry*, **53**, 434-446.
13. McGuire, T.F., Qian, Y., Vogt, A., Hamilton, A.D. and Sebti, S.M. (1996) Platelet-derived growth factor receptor tyrosine phosphorylation requires protein geranylgeranylation but not farnesylation. *The Journal of biological chemistry*, **271**, 27402-27407.
14. Pereira-Leal, J.B., Hume, A.N. and Seabra, M.C. (2001) Prenylation of Rab GTPases: molecular mechanisms and involvement in genetic disease. *FEBS letters*, **498**, 197-200.
15. Fivaz, M. and Meyer, T. (2005) Reversible intracellular translocation of KRas but not HRas in hippocampal neurons regulated by Ca2+/calmodulin. *The Journal of cell biology*, **170**, 429-441.
16. Bivona, T.G., Quatela, S.E., Bodemann, B.O., Ahearn, I.M., Soskis, M.J., Mor,

A., Miura, J., Wiener, H.H., Wright, L., Saba, S.G. *et al.* (2006) PKC regulates a farnesyl-electrostatic switch on K-Ras that promotes its association with Bcl-XL on mitochondria and induces apoptosis. *Molecular cell*, **21**, 481-493.

17. Hoffman, G.R., Nassar, N. and Cerione, R.A. (2000) Structure of the Rho family GTP-binding protein Cdc42 in complex with the multifunctional regulator RhoGDI. *Cell*, **100**, 345-356.
18. An, Y., Shao, Y., Alory, C., Matteson, J., Sakisaka, T., Chen, W., Gibbs, R.A., Wilson, I.A. and Balch, W.E. (2003) Geranylgeranyl switching regulates GDI-Rab GTPase recycling. *Structure*, **11**, 347-357.
19. Seabra, M.C., Mules, E.H. and Hume, A.N. (2002) Rab GTPases, intracellular traffic and disease. *Trends in molecular medicine*, **8**, 23-30.
20. Maurer-Stroh, S., Eisenhaber, B. and Eisenhaber, F. (2002) N-terminal N-myristoylation of proteins: prediction of substrate proteins from amino acid sequence. *Journal of molecular biology*, **317**, 541-557.
21. Towler, D.A., Gordon, J.I., Adams, S.P. and Glaser, L. (1988) The biology and enzymology of eukaryotic protein acylation. *Annual review of biochemistry*, **57**, 69-99.
22. Zha, J., Weiler, S., Oh, K.J., Wei, M.C. and Korsmeyer, S.J. (2000) Posttranslational N-myristoylation of BID as a molecular switch for targeting mitochondria and apoptosis. *Science*, **290**, 1761-1765.
23. Resh, M.D. (1994) Myristylation and palmitylation of Src family members: the fats of the matter. *Cell*, **76**, 411-413.
24. Selvakumar, P., Lakshmikuttyamma, A., Shrivastav, A., Das, S.B., Dimmock, J.R. and Sharma, R.K. (2007) Potential role of N-myristoyltransferase in cancer. *Progress in lipid research*, **46**, 1-36.
25. Sebti, S.M. (2005) Protein farnesylation: implications for normal physiology, malignant transformation, and cancer therapy. *Cancer cell*, **7**, 297-300.
26. Mukai, J., Liu, H., Burt, R.A., Swor, D.E., Lai, W.S., Karayiorgou, M. and Gogos, J.A. (2004) Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia. *Nature genetics*, **36**, 725-731.
27. Yanai, A., Huang, K., Kang, R., Singaraja, R.R., Arstikaitis, P., Gan, L., Orban, P.C., Mullard, A., Cowan, C.M., Raymond, L.A. *et al.* (2006) Palmitoylation of huntingtin by HIP14 is essential for its trafficking and function. *Nature neuroscience*, **9**, 824-831.
28. Maurer-Stroh, S. and Eisenhaber, F. (2004) Myristoylation of viral and bacterial proteins. *Trends in microbiology*, **12**, 178-185.
29. Chen, W.N., Zhang, J., Lin, Y., Chen, N., Zhan, Z.H., Chung, H.S.H., Li, Y. and Shi, Y.H. (2013) Particle Swarm Optimization With an Aging Leader and Challengers. *Evolutionary Computation, IEEE Transactions on*, **17**, 241-258.

Release Note

1. June 22th, 2014, the GPS-Lipid 1.0 was released. The online service was developed with PHP + MySQL + JavaScript, while the local packages were developed in JAVA 1.6 (J2SE 6.0). Several redundant entries were removed, with 738 lipid-modified proteins in GPS-Lipid 1.0.