
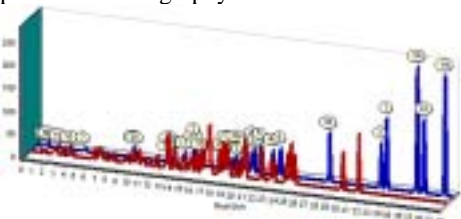


# GPMAW version 6.00

**3D graphs.** The graphs for a number of functions in GPMAW have been updated in looks and have at the same time received a number of new functions. The most obvious of these is the ability to show the graph in 3D.

When you click on the 'toolbox' button  a small tool window opens where you can set the 3D options, turn the graph around the horizontal and vertical axis, change the colors etc. The picture shows the peptide distribution of two different digests of BSA by reversed phase chromatography:



**Mass search.** The mass search results window has changed as the top frame showing hits, peptides etc. in stick bar format has been removed, as I did not judge it very important. Instead you can now open a window, which will show a mass/precision graph for all checked peptides. The advantage of this graph is that you can quickly see if your peptides are likely to be hits or not. Additionally you can perform a linear fit of your data to the protein adding extra security to your determination. More information on this feature on page 2.

**Fragment analyzer.** On page 4 is described a new tool (under the 'Utilities' section of the main menu) that can assist you in *de novo* sequencing using ms/ms analysis of peptides.

**Sequence tag.** When in the department of *de novo* sequencing, the ability of GPMAW to extract sequence tags from an ms/ms or PSD spectrum has been greatly increased ('Utilities|MS peak analysis -> Sequence tag). A check box has been included ('N-term sulf. + K as hR) that, when checked, will modify the N-terminus of the output peptide to be derivatized with sulfonic acid and the lysine residue to be treated as homoarginine (i.e. the modifications that take place when using the Amersham Biotech CAF© chemistry (chemically assisted fragmentation).

**Peptide list.** In some cases it can be an advantage to work with a peptide list instead of a protein sequence, particularly in the case where you cannot specify an enzyme that will cleave your protein into the wanted peptides. This could be the case if you worked with a number of synthetic peptides or a number of peptides from different proteins. GPMAW now

gives you the opportunity to import a list of peptides. This has to be present as a text file with the peptide sequences listed one per line and in upper case 1-letter code. You activate the command Utilities|Search peptide list. You are then asked to specify the file, which is imported as a sequence with cleavage points specified 'manually'. The peptide list opens directly and you have now access to all the usual peptide analysis functions.

**Minor changes department.** The select directory dialog box has changed to a tree view component. The same has happened with the directory selection in the Protein Explorer and the sequence information panel. This panel can also now dynamically show the coverage of highlights and underlines. The protein info panel can also show multiply charges of the current selection. Many of the 'Print' dialogs now have the option of selecting the printer and set printer orientation prior to printing (previously you had to use the Printer setup option). Some small changes have been made to the Make digest database wizard to make the handling of databases easier. The modification file can now work with a 'mass only' modification so you do not need to know the exact composition of a given modification to search for.

**Download.** If you download the upgrade from the web site, you can now also obtain a small text file called 'version changes' that will list most if not all changes and modifications made to the program (note: most of the information is rather tedious).



**Hamburger Leuchtturm.** The lighthouse to accompany the current version of GPMAW is the 23 m high lighthouse situated at the mouth of the river Elbe where it runs into the North Sea. Originally built in 1803 it has been modernized many times since. Originally kerosene powered, it became gas powered in 1905 and electrified in 1913. Its light is visible 22 kilometers at sea.

## Page 1

GPMAW version 6.00

## Page 2

Letter from the editor  
Mass search – fitting peptides to a protein

## Page 3

In the works  
Generating database indices

## Page 4

Upgrading  
Fragment analyzer

## Mass search – fitting peptides to a protein

### Letter from the editor

This issue of *From the Lighthouse* introduces version 6.00 of GPMAW. It has been half a year since the last upgrade of the program (v. 5.11), and it was time to get all the improvements out of the door.

As the numbering should imply, more than the usual number of improvements have been made to the program this time. The most obvious is the change of several of the graphs which now have a much more modern look and which can be changed to a 3D mode. This will in some cases make complicated graphs easier to evaluate.

A few additional functions have found their way into the program (e.g. working with peptide lists) and one option, the utility 'Fragment analyzer' got sneaked in so late that it didn't even make it into the manual.

Otherwise, most of the changes have been in response to input from users who were in need of one or other function or extension of an existing function.

Quite a lot of errors have been corrected, thanks to a collaborating hardware company who had their professional testers go through the corners of the program whirling up quite a lot of dust I thought buried a long time ago. However, the result is a much more stable program.

If anyone would like to contribute or have suggestions for themes to cover in the next issue of *From the Lighthouse* please contact me by e-mail (php@bmb.sdu.dk).

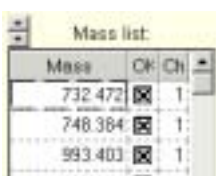
Peter Højrup

Fitting peptide masses from a mass spectrometric peptide map is one of the most common functions when working with GPMAW.

If you have identified the protein in a peptide mass search, you will have to import the sequence into GPMAW. If the peptide mass search was done in GPMAW, the retrieval is straightforward, as you can import directly from the database with a click on a button. If the search was done using a different search engine (e.g. Mascot or ProFound) you will have to cut and paste from the search program into GPMAW (File | Import text | From clipboard). If the pasted sequence is in either Entrez, Swiss-Prot or FastA format, it will be recognized and parsed directly, otherwise you will have to parse it manually, please see the manual.

Once you have the sequence loaded in GPMAW you can select Search | Mass search (or F6) to open the Search for mass dialog box.

Here you will have to paste in your mass list (if you made a peptide mass search it will be the same list) and enter a number of parameters: Ion type, Mass type, Modification list (if you want to search for modified residues), precision and enzyme to match against resulting peptides. Fortunately GPMAW remembers all settings so you will only have to modify the ones that have changed since the last search.



Mass	OK	Ch
732.472	<input checked="" type="checkbox"/>	1
748.384	<input checked="" type="checkbox"/>	1
993.403	<input checked="" type="checkbox"/>	1

When you enter or paste mass values into the mass list they will automatically be selected (X in the OK column) and the charge will be set to 1.


You can deselect individual values and change the charge state of the input m/z values in order to search e.g. for double charged ions. Note here that if you mark the 'Multicharged' check-box, GPMAW will search each input value as if was charged +1, +2, +3 and +4 – in this case you will have to check the 'Search' and 'Found' columns in the results to see whether a multiply charged ion was assumed.

The results of a mass search is presented in a tabular fashion:



Search	Found	Ser.	Modif.	From	To	Sequence
<input type="checkbox"/>	no match					
<input type="checkbox"/>	no match					
<input type="checkbox"/>	no match					
<input type="checkbox"/>	no match					
<input checked="" type="checkbox"/>	1271.63/ 1270.66	23	-	32-	42	13-LPTWHPETLEE-PL
<input type="checkbox"/>	no match					
<input checked="" type="checkbox"/>	1378.60/ 1377.63	32	-	64-	77	13-ETVLTALGQILK

Depending on the state of the 'List type'


button , the table will be either in an extended format (multiple lines per hit/search

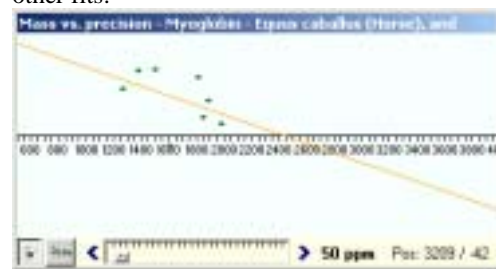
value) or compact format – a single line for each hit/search value.


Peptide sequences of all hits are listed to the right. If a peptide fits the cleavage specification of the selected enzyme at either end, it will be marked with a green '<', and if it fits at both ends it will be underlined. The settings of enzyme and whether only exact matches should be listed are quickly set in the toolbar, enzyme in a drop-down selection box, 'check fit' by a button:




A new feature in version 6.0 is the change of the graph frame that showed the match between expected and found m/z values to a different graph displaying the difference

between mass and precision (button ). This little graph shows all checked hits in the result list as dots, green for perfect fits, red for other fits:



If the checkmark button  is unselected, all hits will be displayed. In the 'checked' mode, dots can be turned on and off by marking the relevant lines in the check boxes. A yellow line shows the best linear fit calculated on all the displayed values, and if the 'Linear fit'

button  is pressed; the best linear fit is calculated and transferred to the mass search results which are recalculated and redisplayed.

In the presented example (ABRF ms sample from some years ago), the match of the search data is very good with an average precision of 19 ppm, but after a linear fit recalibration the precision increases to 9 ppm and two more hits fall inside the 50 ppm limit set.

Moving the slider below the graph will change the displayed precision. Setting a low precision (high value) will usually catch even badly calibrated mass spectra. Two gray curves will be displayed at low precision; they represent +1 and -1 mass unit. If any dots fall on these lines, it could represent wrong assignment of the monoisotopic peak, or deamidation of Gln or Asn.

When the fitting of the search data is finished, you can switch to the 'Report' page of the search result window by clicking on the 'Report' tab at the bottom of the window. The report gives you a summarized view of the results, including sequence coverage, data fit, non-fitting data etc.

## Generating database indices

### In the works

One of the major operations of the current version was to implement the latest version of a new compiler and all the libraries that follow. A new compiler will often break existing code, and often in not very obvious ways. I hope I have succeeded in swapping most of the bugs out of the system, so what to expect next?

I have postponed the ability of improving the ms/ms analysis for some time now. It should be upgraded with the options of searching for side chain loss, searching a small collection of sequences with mass lists and recovery report, better mass search report.

Many of the printouts are not informative enough, and I will try to improve both the content and the layout, hopefully without detracting from the important contents.

Several times I am contacted by users who would like a specific feature "here and now". In many cases the feature can be coded in a matter of days, and I will then like to distribute the added function, however, there may still be month before the next release of the program. For these people there is now the possibility to download the modified version of the program from the web site. Go to the 'Download' page, select the 'beta version' and you are taken to a page where you can download the 'gpmaw3.exe' file. Installation is manual, i.e. you have to replace the existing .exe file and there is no updated help file. Your only 'help' is the accompanying file 'version changes' that lists all changes made since the last release.

The program is developed in close collaboration with the Protein-Research Group at the University of Southern Denmark, so a lot of the input to the program comes from the research carried out here. However, if you have any suggestions, please contact Lighthouse data.

Identifying a protein (e.g. from a 1- or 2-dimensional gel) is typically carried out by mass spectrometry after digestion of the protein with an enzyme, typically trypsin. Following the acquisition of the mass spectrum, the collection of m/z values are then used to search a protein database, either a specific species database or a non-redundant database representing the sum of all known proteins. A number of specialized programs exist for this kind of peptide mass search (PMS) analysis (e.g. Mascot, ProFound, SpectrumMill etc.). A common feature of these programs is that a license for a local implementation is quite expensive and although some of them can be accessed through the web (see f. ex. [www.matrixscience.com](http://www.matrixscience.com)) they suffer from a limited number of databases to access and a varying processing time due to different loading of the web server.

GPMaw also contains an implementation of a PMS search engine. Although not quite as fast as the specialized search programs and without a probability based scoring system it does have the advantage that you can implement almost any local FastA formatted database. When searching the proteins from a single genome, search times are usually measured in 1-3 seconds, and the results can be evaluated manually with only a little experience.

Prior to using the PMS function in GPMaw you have to generate indices. Doing this is quite straightforward, as there is a 'wizard' available (Setup|Make digest database) that guides you through the various options. However, several problems present themselves when you start working with the databases:

Downloading from the net: Most of the major non-redundant databases and a number of genomes are available on the web for download by FTP. The first problem is that the servers delivering the databases are usually UNIX based servers and the line format is different from Windows (DOS) based computers, the second problem is that the name line, particularly in the non-redundant databases, can be several hundred characters long while GPMaw can only handle names with a length of 250 characters.

Both problems can be handled by the auxiliary program Dbindex included with GPMaw and called from the 'Utilities' menu. Go to the 'Convert file' page, set the 'Maximum name length' to an appropriate size (lower values like 100-150 can significantly reduce the size of nr databases and will usually carry enough information to identify any given protein), select

database type and press the 'Convert' button. You now have to select the appropriate database, enter a name for the converted database when asked and the conversion will be done quickly. After you have performed the conversion, the original database can be deleted in order to conserve space (the full NCBI nr database takes up approximately 1 GB of hard disk space). Once converted, the databases can usually be indexed by GPMaw without problems.

However, when working with proprietary databases, typically translated genomes or in-house generated databases, problems can be located the construction of the name line. FastA formatted databases are known in a 2-line format (most common – first line is accession number and name, second line is the sequence) and a 3-line format (first line is accession number, second line is name and third line is the sequence). In order to differentiate between these formats, GPMaw looks at the length of the first sequence line. If the length of this line on average is less than 14 characters for the first 50 or so sequences, it is deemed to be a 3-line format. If the file is really in 2-line format, the result may initially look fine, but when you start searching, you will get weird looking results. A drop-down list has now been included in the *Make digest database* wizard (second page) that shows the result of the database record analysis by GPMaw and you can use the drop-down list to override the further processing (e.g. if the list says '3 line FastA, DOS format' you can change it to '2 line FastA, DOS format' if you know that your database has short name lines. If the list says 'VMS format', you will have to convert the database to DOS format using DBindexer as outlined in the beginning.

In most FastA databases, the first line starts with the accession number before the actual sequence name. GPMaw will extract this first word on the name line (i.e. all characters up to the first space) and assumes this is the accession number. In this word it will look for the characters '|', ';', and ':' and if found, the line will be parsed and the most likely accession number extracted. So if you construct your own databases make sure of the following: If making a 2-line FastA formatted database make sure 1) name lines are longer than 14 characters 2) the first word contains the accession number (or other kind of identifier) 3) the second and following words can be used as a name of the protein.

### Exporting list data to a spreadsheet.

A common question is that when copying list data (typically from the peptide window) and pasting the data into Excel a whole line of data gets copied into a single spreadsheet cell. The reason for this is that GPMaw has copied each column item together, separated by space characters. This is fine if you want to copy the data into a report, but when copying into Excel you want each item to go into a separate column. You can easily correct this by opening the Setup|Setup system dialog box, switch to the 'Peptide' tab and in the top right-hand side you will find the 'Copy table to clipboard' option. Switching the option to 'Copy tab delimited' will make certain that GPMaw use the tabulator character (#9), and transfer of data from tables to spreadsheets should proceed without problems.

## Upgrading

Included in a license of GPMaw is the right to upgrade your program to the latest version within one year of purchase. Current releases of the program are coded to accept licenses that are up to 18 month old. The reason for this is that OEM versions of the program may be several month underway before reaching the end-user.

You can check whether your copy of GPMaw can be upgraded by opening the 'About' box (Help | About). In the middle of the window you can read 'License date:' followed by the month and year of your license. If the current release is within 18 month of this date, you can upgrade.

The upgrade is easily performed if you have access to the Internet. Point your web browser at <http://welcome.to/gpmaw>, go for the 'Update' button and locate the update to most recent version of the program. Click on the name of the download, and when asked whether to download answer 'Yes' and specify the download location.

The upgrade is an executable file that you just double-click from 'Explorer'. The install program searches your disk drive for the present location of GPMaw, and if found you can just accept the default for upgrading.

If the program does not find your copy of GPMaw you will have to specify a location where the program will be located. From here you have to move the two files "gpmaw3.exe" and "gpmaw3.hlp" to replace the files with the same name. The default location of GPMaw is C:\gpmaw\bin\.

If you do not have access to the Internet, you will have to contact Lighthouse data to obtain an upgrade on CD-ROM. Remember to specify your GPMaw license number.

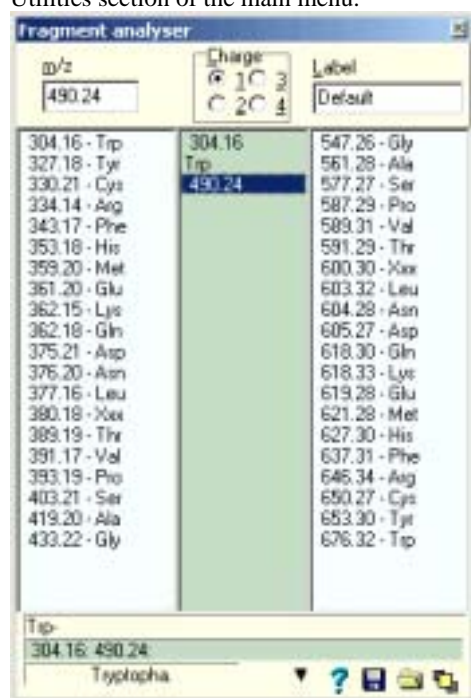
If you want to upgrade and your license is too old, you can upgrade to the latest version for US\$ 130.-. This represents 50% off the price of a full version of the program. If you need additional copies you may buy them for just \$195.- each. This represents a discount of 25%. These prices includes postage and handling.

The MasterCard, EuroCard and VISA credit cards are accepted (not available in Denmark).

## Fragment analyzer

Trying to interpret an ms/ms chromatogram, particularly a *de novo* interpretation, can quickly take quite a long time as there are many possibilities to extend a given sequence tag. If you can extract a peak list, GPMaw can help to extract the longest sequence tag using the Utilities | MS peak analysis - SeqTag function, but as this function does not take peak intensities into consideration, it will in many cases just be a starting point for a manual interpretation.

Here the Fragment analyzer comes to the rescue. The dialog box can also be found in the Utilities section of the main menu.



You start by entering an m/z value from the ms/ms spectrum in the m/z field. Select the appropriate charge and when pressing enter (or click elsewhere in the dialog) the m/z value is transferred to the central list and the left and right hand list-boxes are filled with mass values corresponding to subtraction (left) or addition of

amino acid residues to the selected mass. In the case below, m/z 490.2 was the largest fragment ion, when displayed in the central box, it was obvious that the next residue to the low mass side was 304.1 (Trp). This was double-clicked, which added the value to the list and updated all list boxes. The focus then moved back to 490.2. Looking in the right-hand box it can be seen that 603.3 fits with Leu (~isobaric with Ile which is not shown in the list). Double-click on this and you can extend the sequence tag.

The bottom lines show the sequence tag and mass values in a linear fashion. Through the right-click pop-up menu you can select different display options. In the pop-up menu you can also copy lists to the clipboard or print the sequence tag. The pop-up menu can also be activated as a drop-down menu through the triangle in the bottom toolbar.



These buttons activate from the left:

### Local menu

### On-line help

**Save:** The sequence tag is saved in the current user directory under the name written in the **Label** box (top right) with the extension .fan. 'Default' will always be the initial selection. If a previous file of this name exists, it will be overwritten without warning. If you want to save several sequence tags, you should enter a specific name for each in the 'Label' box.

**Open:** Shows an 'Open file' dialog box where you can select a .fan file. The initial choice will be 'default.fan'.

**Stay-on-top:** When this button is depressed, the 'Fragment analyzer' will stay on top of the GPMaw main window and will not be hidden.

**NOTE:** The up/down list boxes shows all amino acid residues entered in the currently selected mass file. If you want to include post-translationally modified residues in your calculations you have to add them to the mass file and they will show up in the +/--residue lists.

