

GPMAW version 6.20

pI: Several questions have been raised regarding the calculation of pI of modified residues. The feature was not part of the original plan of GPMAW, as I was unable to locate the pKa of modified residues (there is enough confusion about the pKa values of standard residues, which is why GPMAW includes three different tables). However, the questions have been so persistent, that version 6.20 enables you to specify a pKa of any modification. Go into the 'Edit modification file' option, and you can now specify charge (-1 or +1) and pKa. Remember to save the modification file after change. You can also now specify that a given modification is only enabled on either the N- or the C-terminal. This particularly useful for a modification like pyroglutamate from an N-terminal Gln. **Note:** The changing of a terminal does not change the calculations of the pI, thus giving rise to an erroneous pI value of the terminal peptide (and the protein of course). Work is also in progress for user-defined pKa values of the mass files. The function may make it into 6.20, if not it will be part of the first beta release. Check it in the 'Edit mass file' dialog.

Mass search report: The report of the mass search now allows for copying of the graph and results in color. The export is done in Windows metafile format with the result that it can be scaled without loss of detail.

If you have defined Cys-Cys links in your protein, GPMAW will now also search for disulfide linked peptides. However, there is a limit in that only singly linked peptides will be considered. You find the linked peptides in a third page on the bottom of the result list.

Printout: The print dialog box from Sequence and Peptide windows now has a line for entering a comment in the printout. The sequence window also has the option to print the annotation. If the annotation contains more than 10 lines, the printout switches to 8 point to conserve space, otherwise it is printed in the defined font size.

Coverage analysis: A new window has been made under utilities "Coverage analysis". This window is able to load files saved from 'Digest analysis' and from 'Mass search reports'. The files can be edited, copied and printed. The function is still in its infancy, but will be expanded over the next versions – feel free to come with suggestions. See also page 4.

Analyze **Report** **SS linked**

Sequence coverage: The sequence coverage window has been reworked. More details on page 4.

Peptide list: When you have defined modifications in your protein and perform a digest, you can now see a list of partial modifications. Click on the partial mods button



and then select the modified peptide. A window will open below the peptide list, showing all combinations of modifications.

```
[Ser35] Phospho 79.97 Da - [Ser36] Phospho 79.97 Da -
Unmodified peptide:
5 32- 38 737.4192 778.4229 FVLSGGK
Partially modified peptide:
5 32- 38 817.3855 858.3893 Phospho
5 32- 38 897.3519 938.3556 Phospho Phospho
```

Minor changes department. Once again the format of the NCBI web search engine changed slightly, but sufficiently to make the GPMAW web retrieval give up its ghost. This should now be fixed and I will try to keep a vigilant eye on the web retrieval and keep it working in the beta-versions.

Deviation type (e.g. ppm/Da) in the Mass search report now corresponds to the Analysis page.

The sequence information window now has a page showing the isotopic distribution of the intact protein. The function works up to 70 kDa (approx.). The absolute and relative distribution of the isotopes is also reported.

A small error in the calculations of cross-links (Search | MS X-links) resulted in cross-linked peptides being off by 1 Da. The error apparently crept in with one of the beta-versions – shows that it pays to keep your eyes open. The error was fixed from day to day.



The lighthouse for version 6.20 is kind of peculiar, as the northeastern tower of Kronborg Castle, Elsinore, Denmark is a lighthouse. The Castle is well known from Shakespeare as the castle of Hamlet.

The tower was converted to a lighthouse in 1772 in order to lead ships through the narrow strait of Oresund between Denmark and Sweden. However, the seamen were unwilling to pay extra fees, so the light was turned off again the same year, not to be re-lighted before 1800.

In 1842, the lighthouse was, as the first in Denmark, equipped with lenses, and in 1925 the burner was exchanged for a bulb. The light is still in use today.

Page 1

GPMAW version 6.20

Page 2

From the editor
Searching sequences
Modifications and mass files

Page 3

Beta versions
GPMAW and networks

Page 4

Upgrading
Cleavage analysis

Searching multiple (similar) sequences

From the editor

Merry Christmas. This issue of *From the Lighthouse* relates to the release of version 6.20 of GPMW, and it seems to be released just before Christmas. It has been some time since the last upgrade, but a number of events have conspired to delay the current version.

As always the improvements made to the program represents a gradual change. We try to make all improvements compatible with previous versions in the way that you can upgrade without performing any conversions of your existing files. However, in many cases it is not possible (at least not convenient) to “downgrade” as changes in file formats may not be downward compatible. This is mainly the case for ‘system specific’ files like modification lists, .ini files etc. as files that are exchanged (e.g. sequence files) we always try also to make downwards compatible.

For the next version of GPMW I hope to be able to include a new search engine to perform peptide mass searches, and searches using ms/ms data. It is likely to be the JASS search program (Just Another Search System), and while it will not compete with programs like Mascot, I hope that for the small-time user it will prove a worthy competitor.

Other changes that are likely to happen is the general inclusion of user-defined pKa values and related calculations for all of GPMW. Furthermore, small accessory program PeakEraser is being further developed, and a new version is likely to be shown at the ABRF conference and released immediately after.

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 Høirup

Several requests have centered around getting better tools for analyzing multiple sequences, in particular multiple closely similar sequences. In the last version the ‘Digestalyzer’ was introduced, a function that enables you to predict the best conditions for differentiating the sequences based on a couple of parameters, any two of mass, HPLC index, pI or hydrophobicity – see article in previous issue of ‘From the lighthouse’. Version 6.20 brings couple of improvements:

First: The control panel has been improved, so you better can see the color change and the ‘Dot change’ command has changed to a slider, enabling you to choose any value for the change between diamond and square shaped dots.

Second: A third page has been introduced in the control panel, ‘Table’.

This page shows you a list of peptides, either a list of the unique peptides or a list of the identical peptides (i.e. peptides of a mass that occurs at least twice in the list).

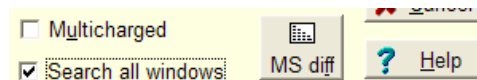
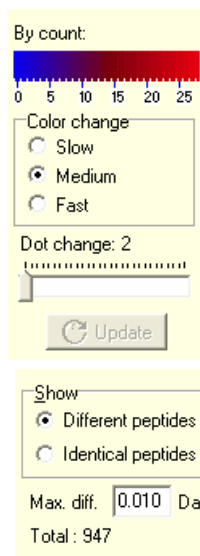
Below the selection of list type, you can enter the maximum mass difference that will include peptides in the ‘Identical’ list.

Another initiative is the ability to perform a mass search of all sequence windows opened on the desktop with a single command.

Proceed as follows:

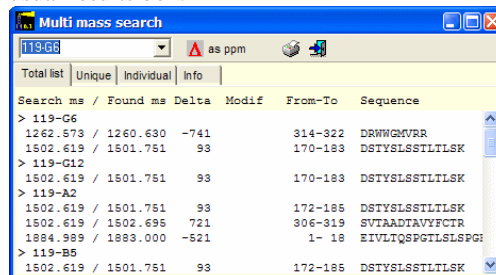
Open all relevant sequences on the desktop, close all irrelevant windows.

Select the ‘Search for mass’ command as usual.



In the bottom of the dialog box you check the ‘Search all windows’ option.

When you now select ‘OK’, GPMW will search all sequences (the screen will flicker a bit) and a new window opens instead of the usual results box.



This window is so-called ‘terminal’ window, in that you have to close it to continue your work.

It consists of a four page control. The **first page** shows a list of all results: sequence name followed by a list of peptides with theoretical and found mass values, deviation in either Dalton or ppm (selected in the command bar at the top), potential modification (if any), location in sequence and finally the sequence.

The **second page** of the control shows all the unique peptides, i.e. the ‘hit’ peptides that only occur once in the list. The **third page** shows all ‘hits’ on an individual sequence by sequence basis, the sequence being selected by the dropdown list in the toolbar. Finally, the **fourth page** shows information regarding the search, input mass values, proteins searched etc.

Before you use the ‘Search all’ option, it is a good idea to optimize your input values first, using one of the closely similar sequences first to make sure you have the optimal calibration and know the precision of your search data.

This command will hopefully save a lot of manual comparison of output data.

Modifications, mass files and GPMW.

Two lists are essential for the correct working of GPMW. The first list is the list of atoms. This comprises up to 32 atoms and forms the basis for the second important list, the mass file. The list of atoms is saved in the .ini file (the initialization file which is user-specific) while the mass file is stored as a separate file. The rationale behind this is that you should never (or rarely at least) change the mass values of the atoms, but you can, by the click of a button, select an alternative mass file, typically if you modify all residues of a given kind in you sequence (i.e. carboxymethylate all cysteine residues). As the mass file can contain up to 30 residues, you can enter new modified residues in the list. The down-side to this is that if you use multiple mass lists, you have to enter these modified residues in all the lists whenever you make a change.

An alternative way of specifying modifications are through the Modification lists. These are small files of up to 30 modifications that specify changes to specific amino acid residues. The information that is saved for each entry in the modification file is: name of modification, formula for the **change** from the standard amino acid residue, valid residues (i.e. which residues can be affected by this modification). From version 6.20 you can further specify charge (-1 or +1), pKa of the modified residue and whether the modification is located to either polypeptide terminal (works in combination with ‘valid residues’).

A modification can be assigned to a residue in a sequence by double clicking on the residue in the sequence window. If modifications are specified for the particular residue in the currently loaded modification file, you can select it and click ‘OK’. As an alternative you can enter the relevant information directly into the edit boxes provided, but this is discouraged, as it is difficult to reuse the modification and make a consistent nomenclature. A last alternative is the insertion of ‘simple modifications’, which is a list of often encountered modifications that can be found by right-clicking on a residue and select ‘Modify -Xxx111-’.

Continues on next page.

Beta versions

As the development of GPMaw is a continuous exercise with many inputs from the end-users, several interim versions are developed between each release of the program. These versions are called beta-versions, and will usually have a number like 6.12b2 (the second beta after version 6.11 was released).

Usually the beta version consists just of the executable file (i.e. the program) without any of the accessory files like help file, modification files etc. It is thus also without any installation routines, which means that the user have to copy the downloaded file into the \gpmaw\bin\ directory and replace the existing copy of the file.

All users with a valid upgrade path are welcome to download these beta versions. The current upgrade rules are such that after each purchase, you have 18 month of free upgrades. You can check the current status of your license by selecting Help | About... from the main menu of GPMaw. In the About box you can read a License date telling you the starting month of your license (e.g. 5 – 2004 means May 2004; you will then have free upgrades through October 2005). After expiration of the license you will have to buy an upgrade, this is usually half price of a new license.

The release of each beta is usually the result of a request for a specific function in the program from a user. Instead of just sending the modified program to the user, it is posted on the web site for all to use.

No updated help file is usually included with the beta version, but information of the main changes can usually be found on the download site (otherwise you may consider to send a mail to php@bmb.sdu.dk).

You are as always welcome to suggest improvements and changes to the program. If you find errors in the program, try first to download the current beta version (or at least check the web site if the error has been corrected) before contacting Lighthouse data. However, if you find an error, we always try to rectify it as soon as possible, often from day to day, and post the correction on the web site. This also means that the beta version may have an unfinished or dysfunctional command or dialog box that was "in the works" during the "fixing".

Running GPMaw over a network

There are cases where it is convenient to run programs across a network, e.g. when having 20 or more students each running a separate computer in a teaching situation – or even when a department runs a common set of programs for all computers. The main advantage lies in that upgrades only have to be carried out on a single, central computer instead of for a multitude of users.

In principle GPMaw is able to run across a network, except for one thing: each user has his/her own .ini file storing all the local settings, and GPMaw will report an error if it is not able to store the .ini file due to write-protection of the server.

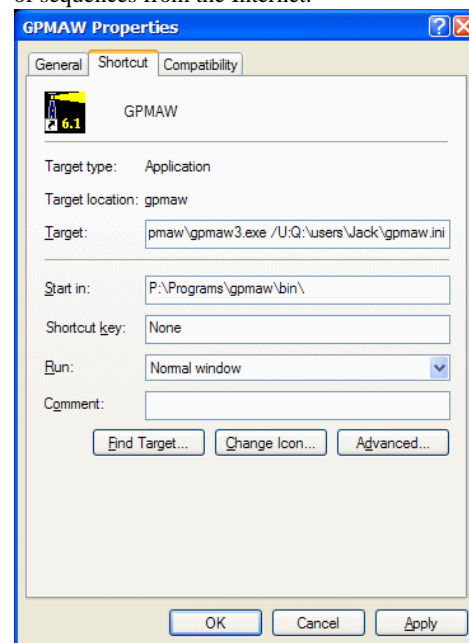
The ideal thing is to store the .ini file among the users own data or another place that is not write-protected. The trick is to use the '/U:' inline command in the properties dialog of the shortcut. Proceed as follows:

- 1) Install GPMaw on the central computer.
- 2) For each user decide on a location for the .ini file. The user needs to have write-access to this location.
- 3) On each user computer, you place a shortcut to the gpmaw3.exe file.
- 4) Right-click on the shortcut and select 'Properties' from the pop-up menu.
- 5) Select the 'Shortcut' tab. In the top edit line (called 'Target') you make a space after the program name, and then you enter '/U:' + directory path to the ini file.

The .ini file may be called anything. If you prefer to locate the .ini files in the same directory, you can name them after the user, e.g. jack.ini, lucy.ini etc.

The main disadvantage to the above method is that the setup has to be done for each computer separately, but once done you can upgrade the central program to your hearts content.

Each .ini file only uses a few kilobytes, so with a central repository there is no need for a large space. However, GPMaw also uses this directory for temporary files, e.g. when downloading a series of sequences from the Internet.



Modifications, mass files and GPMaw....continued from previous page.

Once a modification has been entered/selected for a given residue and the sequence saved, it will follow the sequence and is not dependent upon the corresponding modification file being loaded. Modified residues can be seen in the sequence as being red, and when the mouse cursor 'hovers' over the residue, the information panels in the sequence window toolbar will change color to yellow and show the name and molecular composition of the modification.

When GPMaw starts it will always opens with the default mass file 'aa_mass' no matter which was last selected, but the program will load the modification file last used.

Editing the modification table:

Select 'Edit | Edit modification file' from the main menu. The edit dialog opens with the content of the currently loaded modification file. In the right-hand part you can select a new file and save any changes you make to the current one.

When entering a modification, you have to fill out the first two fields in the line (name and formula) – the rest are optional. When in the formula field, you can call on the formula editor (right-hand side of the dialog) to enter the formula correctly. If 'Valid residues' is not filled out, all residues are considered valid. The 'OK' column enables the current modification for selection. This enables you to have a full modification table (30 entries) and only use a limited number for a given command (e.g. mass search) where the full complement would give information overload. To enter a pKa value, you have to select both a charge and a pKa value (only values between 0.1 and 12 are accepted. Selecting a terminal, 'Term', works in conjunction with 'Valid residues' (both fields have to be correct to be selectable).

Modification file: **adducts.MOD**

Name	Formula	Valid residues	OK	Charg	pKa	Term
Oxygen	O1	M	<input checked="" type="checkbox"/>	0	0	-
Methylatio	C1H2	DE	<input checked="" type="checkbox"/>	0	0	-
Dehydroxy	-H2O1	All residues	<input type="checkbox"/>	0	0	-
PyroGln	-H3N1	Q	<input checked="" type="checkbox"/>	0	0	N-
Carboxy	C1O2	E	<input checked="" type="checkbox"/>	-1	4.5	-

After editing, the list have to be saved to disk, either using the 'Save' command, or the 'Save as' to save it as a new file with a new name. You may have to reload the file to make the changes active in the current session – this is a shortcoming that is being looked into.

