

GPMAW version 6.01

Ms/ms analysis. Finally, after a lot of threats the number of fragmentation patterns has been expanded. The new fragments include loss of side-chains, loss of water and in-chain cleavages. As the number of possible fragments is so huge, particularly with in-chain fragmentation, only the most common fragments are included. The function is not complete yet, as the fragments is in three different tables, but hopefully by the next version all will be collected into a single table for searching and printing. For more information, please see the article on page 3.

Sequence retrieval. GPMAW has for a long time been able to retrieve sequences through the Internet. For this purpose a server at the NCBI (<http://www.ncbi.nlm.nih.gov>) and the server at ExPasy (<http://www.expasy.org>) has been used for retrieving records from Entrez and Swiss-Prot respectively. However, over the summer, both the NCBI and ExPasy servers has changed format, at it was no longer possible to retrieve sequences in text format. This has resulting in breaking the GPMAW code with the result that you will have experienced error messages like "Unable in getting web info" or "HTTP 1.1 301 Moved permanently". A solution for the Entrez server has been implemented, but for now, the Swiss-Prot retrieval is not possible. Hopefully it can be implemented in version 6.02. More information on Web retrieval on page 4.

Proxy server: If your computer is located behind a firewall and a proxy server GPMAW is likely unable to retrieve information from the web. In the 'Setup proxy' dialog (accessed through the main menu item 'Setup') you can now enter information on the proxy server. For security reasons, the password is not saved between settings and has to be entered whenever GPMAW is restarted. This function may work, but is not guaranteed and you will have to get help from your IT department to fill out all the fields. If anyone gets experience with this, I would like to hear about it on php@bmb.sdu.dk.

EMBL and ATLAS CD-ROM. As the release of the EMBL and ATLAS CD-ROM has been discontinued for a long time now, support for the two CD-ROM formats have been removed from the program.

Modification and mass file: The number of entries in the atomic mass file has been increased and a large number of additional atoms has been added. The modification file now has an entry where you can specify mass values without knowing the composition. Furthermore, you can now in the peptide mass

list specify a fixed mass addition and display mass values based on an alternate mass file. More on this theme on page 2.

Limits: For most items (e.g. size of protein, number of items in a list etc.) GPMAW have limit to the number of items or the size of them. When these turn out to be limiting in praxis, they can usually be modified without problems. In the present release, the number of peptides in a mass search list has been increased to 300. The number of residues in ms/ms fragmentation is now limited to 400. This limit was also in effect previously, the user was just not notified that the bottom of the list was cut off – sorry for that.

Specialist department. In many lc-ms/ms applications you often want to look for particular mass values, or you want to exclude certain mass values. The masses are usually based on peptides and potential modifications. You can now have GPMAW make these include/exclude files, for more details see the sidebar on page 3.

Minor changes department. The 'Quick cleave' function now has the option of specifying 'one missed cleavage' so the resulting mass list shows both completely cleaved peptides and peptides with one missed cleavage.

The number and type of protein databases installed in GPMAW can be quite confusing. A preliminary attempt to bring order to this has been implemented in the 'Setup' dialog on the 'Directories' page.

Although direct import of Swiss-Prot records no longer is possible through the Internet, you can still install the Swiss-Prot database and copy/paste complete records from elsewhere. Import of modifications from the records has been improved with support for deamidation, methylation, citrulline and sulfation.



Galley Head Lighthouse. Galley Head Lighthouse is beautifully situated on the tip of a peninsula on the rugged southwestern coast of Ireland. The lighthouse was first established in 1878 and is 21 meters high. The light is 53 meters above mean high water and has a range of 23 nautical miles. The lighthouse became unwatched on 31st January 1997.

Page 1

GPMAW version 6.01

Page 2

From the editor
Modified residues

Page 3

Ms/ms fragmentation
of peptides
In-/exclusion list

Page 4

Upgrading
Sequence retrieval

From the editor

This issue of *From the Lighthouse* accompanies version 6.01 of GPMW. It has been some time since the last update, but a number of events have conspired to delay the current version.

As indicated by the small increment in numbering, the current release does not contain major changes. The focus for this release has been to improve the functionality of the existing function instead of trying to improve the feature race (although a couple of new functions has been included).

The number of errors fixed this time has not been particularly high (if you find one check out the beta versions on the web site). One of the more serious errors that have occurred is the breaking of the web links to both Entrez and Swiss-Prot. In both cases it was the server that changed its protocol (e.g. *I am innocent*) and it took a little time to create a substitute. At present only the NCBI (Entrez) retrieval is working, but work is in progress to include Swiss-Prot in the next release.

In the new features department, the focus has been on ms/ms analysis with a long due overhaul of the system. A request from a user who found the creation of exclusion lists for his Micromass Q-TOF to be too laborious a feature to put to regular use, led to the inclusion of automatic creation of inclusion lists in GPMW (see page 3). If you need additional formats in relation to the ones included, please contact me.

Due to the large size of the non-redundant database, it is no longer included in the compilation on the installation CD, I am trying to find an alternative. The Swiss-Prot database is still included, and the nr database can be downloaded from <ftp://ftp.ncbi.nlm.nih.gov/blast/nr/>, select the nr.tar.gz file (you will need to decompress the file before use).


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 (nhn@bmb.sdu.dk).

Modified residues – specifying and using them

Posttranslationally modified amino acid residues plays an essential part in the working of proteins (i.e. signal transduction, solubility, structure, activity, sorting etc.). Most of these modifications result in a change of mass and can thus be detected by mass spectrometry. When using GPMW for analysis, you need to inform the program of these (potential) modifications. As the way you work with these modifications can change from a 'simple' 'what if' to the chemical modification of all residues of a given kind (e.g. carboxymethylation of Cys) to a global search for large number of potential modifications, the way that GPMW handles modifications varies.

1) Modifying a single residue: This is the simple intuitive approach where you double-click on the residue to be modified and are greeted by a dialog box where you at the top see the residue and below in the field 'Modification' you can enter name and chemical composition of the modification. The chemical composition is then used for calculating the mass change from the default composition. If the modification results in the loss of atoms, you enter them as negative numbers. To the right of the 'Elemental composition' field there is a button that will open a composition calculator. If you need atoms that are not present in the composition calculator, you will have to enter them in the mass file (Edit | Edit mass file on the 'Atomic masses page'). Instead of entering a name and composition manually, you can choose to use one of the built-in 'simple' modifications from the drop-down selection labeled 'Insert simple modification'. Finally, you can select a modification from a modification file in the bottom part of the dialog box. These files are created and edited from the main menu Edit | Edit modification file. The modification files are also used in mass searches so it can be useful to take a little time to construct the files that are most useful for your present purposes.

Note: The simple modification can also be quickly inserted by right-click on the residue and select the appropriate option from the pop-up menu.

Note: A new feature in the modification file is that you can specify a modification by mass alone  Mass only (i.e. no composition). This enables you to display and search for 'unknown' modifications.


Note: Individually modified residues are colored red. The modifications are carried on to daughter windows (peptide window, mass search etc.) but are not taken into account when calculating physical/chemical characteristics like pI, hydrophobicity, HPLC retention etc.

2) Modifying all residues: This is typically

done when chemically modifying a protein, i.e. carboxymethylation of all cysteine residues. To cater for this possibility, GPMW enables you to have different mass files (for details see vol.2,1 of *From the Lighthouse*). The different mass files are easily accessed in the drop-down box in the main toolbar. This is a global selection, i.e. is effective for all sequences opened in the program. As this option is particularly used for Cys residues, an SS/SH button is placed next to the drop-down box. When a mass file where Cys is defined to 102/103 Da, this button is enabled and makes it fast to change between reduced Cys (SH) and oxidized Cys (SS). Again this selection is global.


Different mass files are of course not restricted to Cys residues, but can be implemented for any residue. As the mass file can contain 30 different residues you can also specify completely new residues. This is particularly useful if you have a large number of modified residues of a specific kind as you may exceed the number of individually changed residues (20) allowed by GPMW.

Peptide list: When generating fragments of a protein you end up with the 'Peptide list' window. The number of parameters you can display for each peptide is quite large. For a complete list you have to open the 'Peptide setup' dialog by pressing the white-on-blue 's'

button  in the peptide list window and then press the 'Setup' button for the 'Column layout'. In addition to the various charged mass values you can now also specify an 'Add mass' value. This can be used for 'unknown' adducts where you know the mass but not the composition or residue involved. When you know the residue involved, you can choose an alternative mass file with charge state and N- and C-terminus.

In addition to entering the mass values/file name, you also have to select a column in which to display the value – either the default or the alternate column.

Note: If you have multiple modified residues in the same peptide, you can highlight (select) the peptide and press

the 'Partial modif' button  in the toolbar to display a list of all the different combinations of modifications at the bottom of the list.

Mass search: When performing a mass search, you can specify a modification mass list to include in the search. All mass values are then considered +/- all enabled modifications in the list, but only for the residues specified.



Inclusion / exclusion lists

When working with complex peptide digests in lc-ms/ms analysis it is often desirable to exclude certain mass values from analysis or to look for specific mass values (e.g. wholly or partially modified peptides). These mass lists can be very tedious to create, but GPMW is now well suited to create them. You start out with a peptide window (e.g. a digest of a protein). Right-click in the window and select 'Export | Create in-/exclusion list'. This action will open a small 'Wizard' in four pages. The first page enables you to select 'All peptides' or just 'Modified peptides'. You also select whether you only want 'Fully modified peptides' or 'All partial modifications'. Partial modifications means that if a peptide contains three phosphorylations, then the list will include none, peptide + 1, peptide + 2 and peptide + 3 phosphorylations. A check box enables you to exclude the none modified peptide mass from the list.

NOTE: The modifications that are calculated for the list have to be individually defined beforehand in the sequence window (e.g. double-click on a residue and select modification or right-click and select modify).

Pressing the 'Next' button leads you to the second page where you can select variable modifications (e.g. oxidized Met). These modifications are calculated for all peptides containing the relevant residue, but only none and one modification is included for each peptide. You also select which charge states to include in the list (from 1 to 4 charges).

The third page of the wizard enables you to select output format and mass mode (average or monoisotopic mass). **NOTE:** The resulting list is always saved as a text file on disk, it is not possible to save to the clipboard.

The last page lets you review the resulting mass list before saving to a file on disk. If you want to change a parameter, you just press the 'Previous' button to go back and change parameters. The number after the 'Review' legend tells you how many mass values are in the list.

The 'Finish' button opens a 'Save file dialog' that enables you to save the file in the locations of your choice.

Ms/ms fragmentation of peptides

When fragmenting polypeptides by CID (collision induced fragmentation) or PSD (post-source dissociation) in the mass spectrometer, the resulting fragments will primarily be the result of fragmentation in the peptide bond where the charge can be retained either on the C-terminal part (typically generating so-called 'y' ions) or on the N-terminal part (typically generating so-called 'b' ions). These spectra are generally called ms/ms spectra and are used either for identifying peptides (proteins) with a higher degree of certainty than is possible with peptide mapping, or they can be used for *de novo* sequencing.

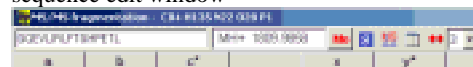
These fragments were described and labeled by Roepstorff and Fohlman (Biomed. Mass Spectrom., 11, 601, 1984) and the prediction of the basic set of fragments (a,b,c,x,y,z) has been part of GPMW for a long time.



The Roepstorff notation for peptide fragmentation.

However, a number of additional fragments often occur in the ms/ms mass spectra, particularly loss of side-chains or part of a side chain, or multiple cleavages may occur resulting in in-chain cleavages. Predicting the masses of these fragments have now been implemented in GPMW and will be described below.

The starting sequence: The simplest way of using the ms/ms window is to open it through the main menu Cleavage | Ms/ms fragmentation or using the fragmentation button . In the resulting ms/ms window you can now enter a sequence in the top sequence edit window



As you enter the sequence, the mass and the fragment ion (below) will be updated. The button to the left of the mass value enables you to toggle between monoisotopic and average mass.

However, in most cases you will already have a sequence in GPMW. If you have a sequence window open, you will be asked whether you want to see an ms/ms fragmentation on the intact protein. If you answer 'yes', please note that there is a limit of 400 residues for the ms/ms window. If you have highlighted part of the sequence in the sequence window, this selection will be taken as for the fragmentation – modified residues will be transferred and colored red in the ms/ms list.

Note: If a residue in the sequence is modified, it is no longer possible to edit the sequence.

From the peptide window list you can select any peptide, right-click and select MS/MS.

The fragment list: The list of fragment ions is currently a 3-page 'notebook':

Fragment	Mass	Label	Fragment	Mass	Label
1029.000	1029.000	Y	1029.000	1029.000	Y
218.000	218.000	B	218.000	218.000	B
336.000	336.000	B	336.000	336.000	B
454.000	454.000	B	454.000	454.000	B
572.000	572.000	B	572.000	572.000	B
690.000	690.000	B	690.000	690.000	B
808.000	808.000	B	808.000	808.000	B
926.000	926.000	B	926.000	926.000	B
1044.000	1044.000	B	1044.000	1044.000	B

The first page 'Backbone fragments' enables you to show a,b,c,x,y,z and the side chain fragments d, v, w. Each type of fragment has its own column with the columns to the left of the residue label representing N-terminal fragments and the ones to the right are C-terminal fragments. If a residue is modified, it will be colored red, and the modification listed in at the bottom of the window. The setting up of the columns and determining which type of fragments to show is done in the

'Edit ms parameters' box ():

Name	Compex	Proton	Di	FragmentType
1	-C(O)OH	1	<input checked="" type="checkbox"/>	Backbone cleavage
2	-OH	1	<input checked="" type="checkbox"/>	Backbone cleavage
3	-O-NH2	2	<input type="checkbox"/>	Backbone cleavage
4	-O-CH2NH2	1	<input type="checkbox"/>	Side chain loss
5	O	1	<input type="checkbox"/>	Backbone cleavage

The first column edits the fragment label, the second the composition change from the corresponding peptide (only the change in the backbone), the third column enables you to add protons (0, 1 or 2), the fourth column enables/disables the display of the column, and the fifth column sets the side-chain loss type. As the side-chain varies between residues, it is hard-coded into the program and cannot be changed by the user. Pressing the 'Default' button puts default values into the table.

'd' and 'w' fragments are cleavage in the backbone next to the alpha carbon (x and z type) and side-chain cleavage after the beta-carbon. 'v' type cleavage is a 'y' ion with loss of the complete side-chain (cleavage next to the alpha-carbon). As 'c' ions are rarely observed in low energy CID, they are by default not shown in the list. If the fragment contains Arg or Ser/Thr you may observe loss of NH₃ and H₂O respectively. These fragments are shown on the 'Fragment losses' page.

Internal fragments are shown on the last page of the notebook. As the numerical number of internal fragments can be quite enormous even for a relatively small peptide, only the most commonly observed fragments are shown. These are directed by the following residues: Pro, His, Lys, Asp and Glu.

In the right-hand side of the window you can open a frame containing a sorted list of all the backbone fragments by pressing the 'Frame' button (). The current version of GPMW does not support a combined list of all the fragments, this is expected in the following version.

When printing the list, you are given a choice of installed printers, page orientation and which list(s) to print.



Sequence retrieval

Upgrading

Included in a license of GPMW 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 GPMW 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 GPMW, and if found you can just accept the default for upgrading.

If the program does not find your copy of GPMW 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 GPMW 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 GPMW license number.

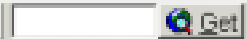
If you want to upgrade and your license is too old, you can upgrade to the latest version for US\$ 130.-. This represents 50% off the price of a full version of the program. If you need additional copies you may buy them for just \$195.- each. This represents a discount of 25%. These prices includes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

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


GPMW has for a long time been able to retrieve protein sequences from the web. In order to do this, it has been using the NCBI web server which offered a facility for searching the Entrez database engine and retrieving in text format – ideal for easy integration into a program. Expaty has offered a similar service for the Swiss-Prot database. However, both services has terminated over the summer, leaving GPMW high and dry.

The services were used in several places:

Web Entrez search is accessed through the File menu item and gives you a dialog box with three search fields and a number of display and database options.

The **web toolbar**  in the main toolbar can only search on a single item, typically the accession number. If the item started with P, Q or O, the search was performed first on the Expaty server (looking for a Swiss-Prot entry) otherwise the search was performed on the Entrez server. If the search retrieves more than one entry, only the first is displayed.

Hint: You can use the web toolbar for retrieving entries by their gi number (Entrez gene index) by entering a vertical bar between gi and the number (e.g. `gi|10639239`).

After a **local BLAST** search, if you highlight a hit and press the 'Retrieve' button , depending on the accession number and settings, the sequence would be retrieved either from the Expaty or the NCBI server. Only as a last choice would the sequence be retrieved from the local database due to the fact that a local FastA formatted database only contains the name and sequence, not any annotation.

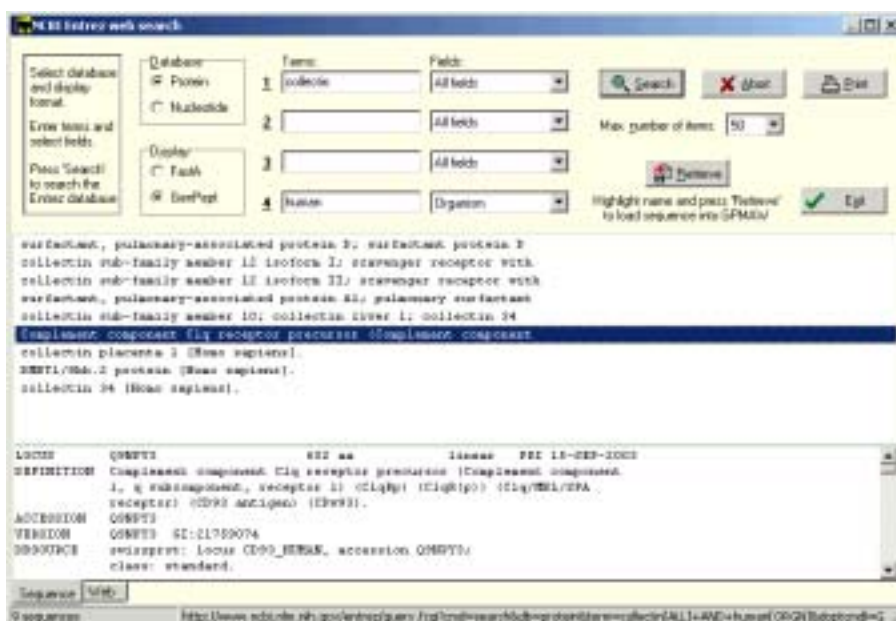
In version 6.01 this has been partly remedied, as the sequences are now retrieved from the NCBI Entrez server in html format. This makes parsing of the sequence a bit more tricky, as any changes in the html or display format may break the code, resulting in no sequence retrieval. In part to compensate for this, the result page of the Web

Entrez Search is now a two-page notebook with tabs at the bottom (see below) where the first page is the traditional sequence name on top and selected sequence at bottom while the second page is a web page. This web page does not have a web address edit bar, and can thus not be used as a traditional browser, but you can follow any of the links in the returned page and you can even use the Entrez search facilities. In order to retrieve a sequence from this page into GPMW, you have to highlight it, right-click and copy the record to the clipboard (or Ctrl+C) and paste it into a new sequence window (File | Import text (ASCII) | From clipboard) or Ctrl+Alt+V.

The Entrez web search dialog has been changed slightly and now has four input fields, slightly different search parameters and a selectable number for maximum number of records retrieved.

Alternative sequence retrievals: If you paste a sequence record in Entrez or Swiss-Prot format (e.g. from a web browser) using the File | Import text (ASCII) | From clipboard command, the format will be recognized and the 'Import ASCII' dialog box will open and the record will already be parsed. In most cases you can just push the 'OK' button and the sequence will be ready. The annotation will be present in the 'Annotation' window which can be accessed through the 'a' button (green for Swiss-Prot, blue for Entrez record, red for other). Please note that the Annotation window is a free field editor and any text can be entered here (remember to save the sequence afterwards).

You can also download any database in FastA format, index it with the DBIndexer utility (see the manual for details) and search it with the File | Open FastA database command. If you download the Swiss-Prot database you have to convert to FastA format first, but will retrieve the complete sequence record when searching the database.



The new NCBI Entrez web search dialog box – sequence names on top, selected sequence record shown at the bottom. Select a sequence in the top part and press 'Retrieve' to import.