An publication from Lighthouse data Volume 1, Number 1 October 2001

From the Lighthouse

Page 1
GPMAW version 5.0
Page 2
Letter from the editor
Working with the new toolbar
Page 3
In the works
Post-translational modifications
Page 4
Upgrading
Peak erazor

GPMAW version 5.0

Welcome to the first issue of *From the Lighthouse*.

Version 5.0 of GPMAW introduces some of the first major changes in the user interface to take place in a long time. The first change you will see is that the toolbar of the main window has changed to a collection of smaller toolbars that can be turned on and off by the user, or even 'torn off' and used as free-floating toolbar windows. Additionally, the number of tool buttons has increased dramatically, making most of the common functions available. The toolbars have also changed style, as the new 'flat' style introduced by Microsoft has been adopted. The flat buttons 'pop up' when the mouse cursor passes over them, giving a nice positive feedback that something happens. The toolbar is further discussed on page 2.

The modifications have also received a small but welcome change: the simple modification. Both the 'insert modification' dialog box and the main window pop-up menu has acquired a drop-down box listing a number of the most common modifications (e.g. phosphorylation, methylation, oxidation etc.). The modifications are context-sensitive (e.g. phosphorylations can only be accepted for Thr, Ser and Tyr), a feature which can be turned off, so any modification can be accepted by all residues. In the process, the items in the sequence window pop-up menu have been rearranged with a number of items moved into sub-menus in order to increase visibility. The way GPMAW handles modifications is discussed further on page 3.

The peptide list (the result of a protein cleavage) has received a button for hiding the low molecular mass peptides. This is most efficient for MALDI mass spectrometry where the low molecular weight peptides often are hidden in matrix peaks. The 'cut-off' value for the peptide list can be set in System setup, on the peptide page (range 100-1000 Da.). The peptide list can now also be sorted by sequence position by clicking on the 'From-to' header button. The level of overlaps (missed cleavages) is now shown as a blue superscript after the peptide number.

In connection with the peptide list, the peptide info box (can be called by double-clicking on a peptide in the list, or by right-clicking on a highlighted peptide in the sequence window and selecting 'Peptide info') has been rearranged with the addition of an isotope distribution, both in graph and list form. Please note that the mass value given in the list is not exact – this is likely to be changed with the next release of GPMAW.

In the search for masses of cross-linked peptides (Search | Protein MS X-links) you can now specify an overlap value larger than 2. An overlap value of 2 is necessary, as you very often use amines for cross-linking, and lysine is a target for trypsin, the most common cleavage reagent. This means that you have to have at least one missed cleavage before you can have two cross-linked peptides. You can now also search the resulting list of cross-linked peptides with a mass list. The results are presented in a two-panel window.

The protein mass search window has an extra entry in the pop-up menu: Color overlaps. Peptide hits are colored according to this scheme: Gray: no overlaps; green: one overlap; yellow: two overlaps; red: three or more overlaps. This color scheme will be expanded to other functions, and will be given a button in the toolbar when fully developed.

The simulated 2D gel has received an overhaul. The N- and C-terminal truncations (tails) can now be specified by the user, both by length and frequency. Furthermore, you can simulate the addition of phosphyrylations. If you rest the mouse above one of the dots, a fly-by help will show mass and pI.

In the mass vs. charge window (also called the titration window) you can now choose to have a sidebar with a list of mass and charge values. This list can of course be copied and saved.

Figure 1 The lighthouse in this issue of GPMAW is the Old Lighthouse at Point Loma. The lighthouse is situated at the entry to the San Diego harbour in southern California. The installation CD and the About box in GPMAW shows the whole lighthouse building.

Letter from the editor

The introduction of version 5.0 of GPMAW also marks the introduction of *From the Lighthouse*. This small publication is intended as a means of communicating new features of GPMAW and related programs for the analysis of biostructural data and bioinformatics in general.
The current issue introduces the new features of version 5.0 on the front page with elaboration of the menu system on this page. However, as many of the features in GPMAW sort of creeps in sideways due to the incremental upgrades, articles covering broader items will be included in each issue. This time an overview of the many ways of including post-translational modifications in sequences is shown on page 3. Finally a 'work in progress', the Peak Erazor, is introduced on page 4. You are all invited to join in on developing this 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).

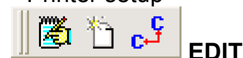
Working with the new toolbar

In recent incarnations of GPMAW, the menu has gradually increased with the addition of new commands, while the toolbar has remained static. In version 5.0 the toolbar has changed and is now a collection of smaller toolbars, each of which controls a section of the menu.

New toolbar layout:

FILE

Open sequence (file). The down-arrow opens the most recently used file list (identical to the bottom of the 'File' menu).
Close sequence window.
Save sequence to file.
Print
Printer setup



EDIT

Edit current sequence.
Edit new sequence.
Edit cross-links of current sequence.

IMPORT / EXPORT

Import from file.
Import from clipboard.
Search FastA formatted database.
Search Web Entrez.
Copy to clipboard.

BASIC

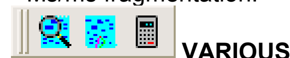
System setup.
On-line help. The down-arrow open a list of toolbars. Each toolbar can be turned on and off.
SS-button – enable/disable disulfide bridges (oxidized/reduced Cys)..
Mass file selection list (drop-down list). The list is gray when the standard mass file (AA_MASS) is selected, white for all other mass files.
Exit – close GPMAW.

GRAPH

Hydrophobicity.
Secondary structure prediction (GOR).
Charge vs. pI (titration).
Dot-plot graph.

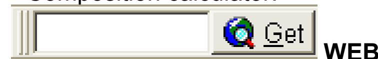
SEQUENCE

Highlight residues (motif). The down-arrow opens the 'Quickcolor' menu.
Search for mass.
Automatic cleavage (digest). The down-arrow opens the 'Quick-cleave' menu.
Ms/ms fragmentation.



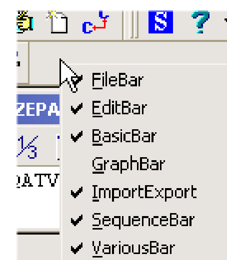
VARIOUS

Digest (peptide) mass search (PMS).
Simulated 2D gel.
Composition calculator.



WEB

Quicksearch of the Entrez Web server for an accession number or other unique ID. Retrieval of the sequence takes place through the Import ASCII dialog. This functions only works when connected to the Internet. May not work through a firewall.



Each toolbar item can be turned on and off, either by clicking on the down-arrow next to the 'Help' button or right-click in an empty part of the toolbar container. Each of the toolbars shown is checked in the pop-up menu.
The individual toolbar items can be moved around in the toolbar container by grabbing the bar handle (the vertical stripes at the left end) with the mouse and drag to the new position.

The individual toolbars can be dragged completely off the toolbar container to be either free-floating windows or they can be docked at the bottom of the window, which contains a pop-up toolbar container, like the top of the main window.
When all visible bars are in the top toolbar container, their positions are saved upon exit of GPMAW and will appear in the same positions when you restart the program – this allows you to customize the toolbar for your specific purposes. If the bars are either free-floating or positioned in the bottom container, they will re-appear in the top container upon restart (but still only those that are checked in the toolbar menu).
Please note that two buttons have moved from the sequence window to the main toolbar. The 'mass search' and the 'automatic cleavage' buttons have been moved. This result in a little more space in the toolbar of the sequence window – I'm sure it can be used for other purposes.

Future developments

The next step in toolbar development will be to make the main menu part of the toolbar (this will make better use of the top area of the main window). Another feature will be to make the user able to configure the toolbar.

Post-translational modifications

Post-translational modifications (understood as chemical modification of residues) are in GPMAW stored as a name combined with a chemical composition and 'allowed' residues (i.e. Thr, Ser and Tyr for phosphorylation, Met for oxidation etc). Other relevant information (i.e. mass, position) is calculated on the fly.

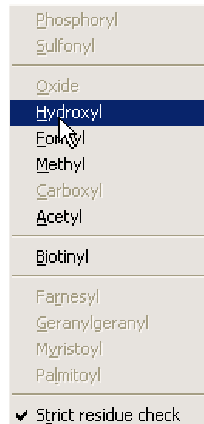
A few modifications, called 'simple modifications', are coded directly into GPMAW. When these modifications are insufficient, you can create 'modification files', which are small databases containing the relevant information (name, elemental composition and allowed residues) that can be called upon when needed.

Internally in a protein sequence, the modifications are stored in a separate list (along with the sequence) that contains the above-mentioned information together with the mass (calculated based on the current mass file) and sequence position. On the display, modified residues are shown in red characters. **Note**: you have to save the sequence after entering a modification, it is not saved automatically.

Chemical modifications may be entered either by right-clicking or double-clicking on the relevant residue (the modification box can also be called from the 'Edit sequence' dialog).

Simple modification:

When you right-click on a residue in the sequence window, the second item in the pop-up menu list the residue type and number, along with a secondary menu listing all the allowed residue for the selected residue. If the residue is already modified, the second menu option will be 'Remove modification'.



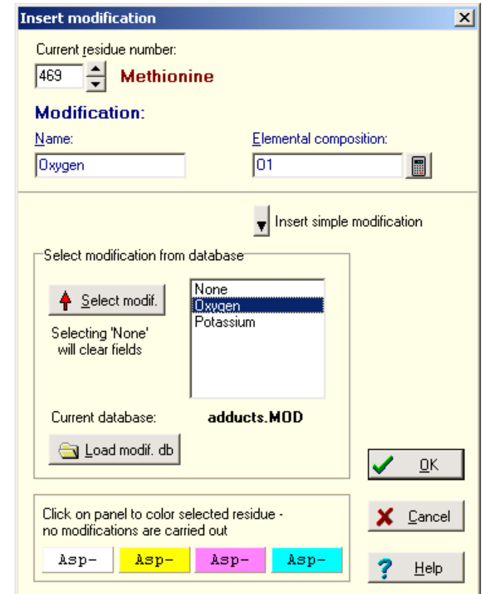
Selecting any modification from the list will immediately modify the residue.

By default any modification not allowed for the current residue will be grayed and not selectable. However, by selecting the last item on the secondary menu 'Strick residue check' you can allow for any modification to be entered on any residue.

'Standard' modifications:

Double-clicking on a residue will activate the 'standard' modification box.

The top left corner shows the residue clicked upon. You may enter a different number in the box or use the up/down arrows for selecting neighboring residues. In the two boxes below you can enter the name and elemental composition of the modification. The elemental composition is most easily entered by clicking



on the calculator button. Remember to subtract elements removed as a result of the modification – the composition shows the difference in composition (and may thus be negative). You may insert 'simple modifications' by clicking on the down-arrow button below the line. This menu is identical to the 'simple modification' above.

User-edited modification files can be selected through the 'Load modif. db' button and selecting from the presented list. The user modified modification lists can be edited through the menu 'Edit | Edit modification file'. Up to 30 modifications can be entered into a single file, and the number of files is only limited by the size of your harddisk.

Note: The bottom, colored panels on the dialog box are only for temporary coloring of individual residues. Clicking on a colored panel will immediately close the modification box without any modifications carried out.

Other modifications:

If you import a sequence from the Swiss-Prot database you can save the complete record on the annotation page (this may be done automatically). If you have a Swiss-Prot annotation, the '**a**' in the toolbar will be colored green. The 'feature table' of the record contains information about post-translational modifications and several of these may be imported into GPMAW by checking them and clicking on the 'Import' button. The import also covers 'signal' and 'pro' peptides that are removed upon import.

Tip: As the annotation page is a free-text page, you can enter your own modifications in a Swiss-Prot entry by following the standard rules (two character identification, three spaces, from–to and modification name). Please see the on-line help for modifications that can be imported. Also remember to save the sequence after making changes to the annotation page.

In the works

What can we expect of the next versions of GPMAW?

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.

One feature you can expect to see improved is the 'peptide mass search'. A research project is looking into the possibilities of extracting more information from MALDI mass data. We all know about the oxidation of Met that shows up as a +16 peak, but there is much more information waiting to be included in 'intelligent' search programs. This information is also likely to be included in the general mass search function (a function that is in need of a complete re-design).

The compiler used for creating GPMAW has recently been released in a new version. However, some of the accompanying libries have been slow in updating, so it will not be until the next release that new program features can be implemented.

These features will mainly include user interface improvements, but you can also expect an improved stability, and perhaps a slight improvement in program speed.

This realease has brought a new toolbar with a number of small bars that can be turned on and off. This design will be expanded to a design where the toolbars can be completely user-designed. Microsofts recent adventure into 'intelligent' menus that hide seldomly used features will not be implemented (unless a general consensus calls for it).

A missing feature is the retrieval of BLAST sequences, and we will also try to implement a multiple alignment program.

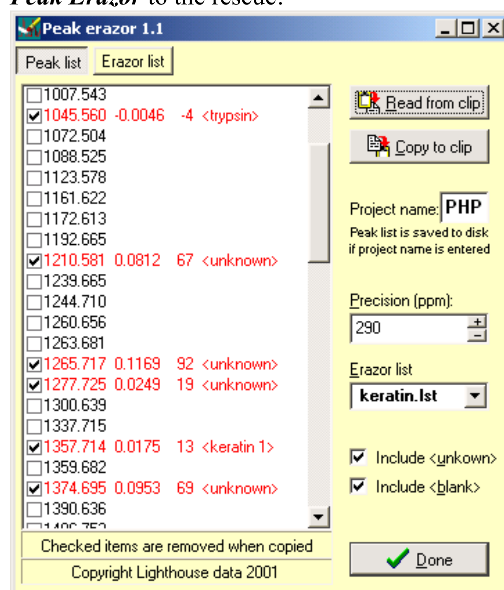
Other features that are under consideration is a more extended use of colors for showing different features of sequences and numbers, an improved web interface that will connect to the most 'stable' programs on the web, and more details in printouts.

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 resent 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 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$ 120.-. This represents 50% off the price of a full version of the program. If you need additional copies you may buy them for just $180.- each. This represents a discount of 25%. These prices includes postage and handling.
!!NEW!! We can now accept credit cards from MasterCard, EuroCard and VISA. Please contact Lighthouse data for information.

Peak Erazor

A common characteristic when analysing protein peptide maps created from 2D gels is the presence of contaminating peaks. Particularly when you are working with very small amounts (i.e. radioactive, silverstained, or other high-sensitivity stains) the occurrence of keration (in particular), tryptic autodigest and other unidentified peaks very often leads to ambiguous identifications.
You can of course manually edit your peak list to remove peaks, but the number and character of contaminating peaks may vary from sample to sample. So what should you do other than throw all those identifications of keratin 1, keratin 9 and keratin 10 in the wastebasket?
Peak Erazor to the rescue!



The operation of the program is very simple:

1) Copy your peak list to the clipboard (works with most programs that copy the masses as the first part of the peak list).
2) Change to 'Peak Erazor' and push the 'Read from clip' button. The peak list is copied from the clipboard to the peak list in the program. A comparison is done with the currently loaded peak list, and all masses that fall within the precision are checked and colored red.
3) Press the 'Copy to clip' button and all non-checked masses are copied back onto the clipboard.
4) Change to your favorite peptide mass search program (ProFound, GPMAW, Mascot etc) and perform the search.

Note: you may also use the keyboard shortcuts Ctrl+C and Ctrl+V to paste and copy the mass list respectively.
If you want a particular mass value to be copied to the clipboard, you just make certain that it is unchecked, in the same way if a mass is unwanted you just check it (click the check box with the mouse).

The masses that match (to the erazor list) and are colored red are displayed with the absolute difference in dalton, the deviation in ppm (parts per million) and the peak ID.
By comparing the deviations, it is fairly obvious which peaks are true 'hits' and which peaks are accidental hits. Note, by looking at this list you also get an idea of the precision of your mass calibration.
IMPORTANT: if you put characters into the 'Project name' box, all mass lists that are copied to the clipboard are also copied to a file on disk called 'allmass.mas'. An analysis module will be built later for analyzing all the saved mass lists (e.g. for unknown contaminats, better precision of unknowns etc.). **Note**, this module has not been built yet and the specifications are still uncertain. If the 'Project name' field is left blank, the mass lists will not be saved.
Other control options:
You may view and modify the Erazor mass list by clicking on the 'Erazor list' button in the top. On the Erazor list you can modify, delete and add mass values, and you can save a modified list under a new name. The list that is included in 'Peak Erazor' contains a large number of common contaminants found in our lab.
If you right-click in the Peak list, you also have the option of saving the current mass list (minus checked items) to the current Erazor list with the ID of <blank>. This can be useful if you are analyzing a series of samples where you have a blank mass spectrum to start with.
The check boxes 'Include <unknown>' and 'Include <blank>' adds the two peak erazor Ids to the peaks removed.
The 'Precision' box determines the precision (in ppm) with which your mass list is compared to the erazor list.
The 'Erazor list' drop-down box lets you choose between all the lists you have saved.
All the options you choose are saved between sessions.
The 'Peak Erazor' is very much a work in progress, which you are welcome to participate in. If you want a copy of the program, please contact Lighthouse data for a copy.
Note: The program is the copyright of Lighthouse data, but you are welcome to use it with charge.

Conclusion: A handy small program with the main advantage being the ease of use, a small footprint, and the added option of a check for mass precision.

Future developments: As mentioned an analysis module will be developed for the analysis of saved mass lists. Furthermore, options like graphical view of the mass fits, linear fit to selected 'contaminants', calibration on keratin (when internal clibrants are missing) etc.