

GPMAW version 6.10

The local BLAST search (which is an implementation of the NCBI BLAST) now has an improved interface with a **redo** function that remembers the input and a **clear text** button.

The **cleavage analysis** window has been reworked. The layout has changed and a new graph comparing the cleavages of the different enzymes has been included. The hardcopy has also changed, with the inclusion of the graph and minor differences in the layout.

Since ExPasy last summer changed the access to their web search engine, it has only been possible to retrieve Swiss-Prot entries over the web through Entrez. As GPMAW cannot parse Entrez records to retrieve modifications, the only ways of getting access to 'easy' incorporation of posttranslational modifications were either to cut and paste from your browser or download and index the entire Swiss-Prot database (easily done, but it will eat up 700 MB of your harddisk). Now web access has been established to the SRS search engine at EBI, UK, and you can retrieve sequences over the web either through the main toolbar



, after a BLAST search or the 'retrieve sequence list'. The complete annotation is copied to the annotation page of the sequence. This can be accessed through the 'a' button in the sequence toolbar. The **feature table** of the annotation window now enables quick retrieval of modifications (just check the appropriate boxes and press the 'retrieve' button).

If you need to retrieve a number of sequences, you can use the **File|Retrieve|Retrieve accession number list** command which enables you to enter a list of accession numbers and retrieve all sequences either to sequence windows in GPMAW or to a file on disk (in various formats).

The retrieval function works in the way that if the accession number starts with O, P or Q (or IPI) the search is initiated on the SRS server, otherwise the search is done on the Entrez server.

The **Dot-plot window** has been changed. The major layout change has been that the command bars at the top and bottom of the window have replaced the small floating command window. Some additional changes have been made to the implementation of the algorithm. In particular, the window size has been increased to allow for easier visualization of weak homologies. The default window size is now thus 13 with a maximum of 29. The setting of the cutoff value has been changed to a slider bar and finally a gray value slider has been introduced. The function of this is to make

a gray gradient of values on the limit of the cutoff value thus displaying differences in 'similarity'. The alignment displayed at the bottom of the window can be copied to the clipboard by right-clicking and select 'copy'.

Peculiarities. A new function **DigestAlyzer** will show you two parameters for a given number of peptide windows as a scatter plot. You can choose between mass, HPLC index, pI and hydrophobicity. The function can appropriately be combined with the new **digest all proteins** function. For more details on the DigestAlyzer see the feature later.

Minor changes department. Setup system | Display - Default window size: **Closing size:** if checked, the main window will open with the same size as it closes. The coloring of residues under Windows 2000 sometimes behaves erratically. Changes have been made in the control, but the problem is partly system specific, as it is not observed to the same degree in Window XP. Peptide compositions reported in the title of the ms/ms window was not reported correctly if the peptides were modified – this is now corrected.

The peptide window can now show mass values with **6 decimals** (check the setup dialog box).

If you want to work with the (relatively) new **IPI databases** (International Protein Index) it is now possible as both the most recent version of DBindex as the present version of GPMAW supports them. As the databases are built upon the Swiss/EMBL format, you have to convert to FastA before indexing and creating digest database, but if all the database files are in the same directory, GPMAW will read the original IPI record when reading the sequence.

Searching peptide lists have now been extended to 'peptide' FastA databases and includes searching for fragments mass values – more details are available in the sidebar on page 3.

Version 6.01.1: After the release of version 6.01 a few corrections were made and posted on the web site. These changes involve: sequences in peptide mass search reports can now span multiple pages; negatively charged ions in the sequence report now prints correctly; reading of sequences from FastA formatted files has been corrected; color schemes for alpha helical wheel are now implemented.

Page 1

GPMAW version 6.05

Page 2

From the editor
Searching for masses in
GPMAW

Page 3

Peptide lists
DigestAlyzer

Page 4

Predicting HPLC retention
indices in GPMAW

From the editor

Version 6.10 of GPMW is now available for upgrades and download.

Although the changes this time mostly is in the minor details (or specialist) department, we are not sitting still. It seems that there forever is a demand for new ways of interpreting mass spectrometric data and bioinformatics in general. One of the changes is the ability to set all the parameters for the simulated HPLC chromatogram. This was done on the request of Joseph Justice, Emory University, who has also been to kind to write a little about his results (p. 4).

Another addition is the DigestAlyzer (p. 3) which arose out of a specific problem I encountered in my research – how to determine the best enzyme for a group of almost identical proteins. I tried to make the function a bit more general, and I hope others may find use for it.

For the next release of GPMW, expected around the end of summer, we expect to have some ability to search databases with ms/ms data. Don't expect a high-throughput competitor for Mascot, but more likely a manual search of specialized databases

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



The lighthouse this time is Sletterhage fyr (lighthouse). The tower is not very big, but is beautifully situated on the tip of the peninsula Helgenæs where it commands a view of the entrance to the harbor of Aarhus, the second-largest city in Denmark

Searching for masses in GPMW

Finishing touches to a peptide mass search.


When you perform a peptide mass search (PMS) you often end up with a hit that has a score, which places it in the twilight zone where the score becomes significant.

Alternatively you want to check for total coverage (i.e. including all cleavage variants, not just a single missed cleavage).

In any case you can verify your data by 'repeating' the search using the GPMW 'Search for mass' function.


Getting the sequence.


If you performed the PMS in GPMW, you can retrieve the sequence through the

'Retrieve' button  in the results window.

This has the further advantage that peptide hits are underlined to immediately give you the coverage.


If you do the search using another search engine (e.g. Mascot) the fastest way of retrieving the sequence is to enter (or copy) the accession number of the protein into the web

retrieval box  in the main toolbar. If you are directly coupled to the web, GPMW queries either Entrez or EBI SRS server to get the sequence directly into the program. If this doesn't work you need to navigate to the sequence display of the search program, copy the sequence to the clipboard and paste it into GPMW using the 'Paste

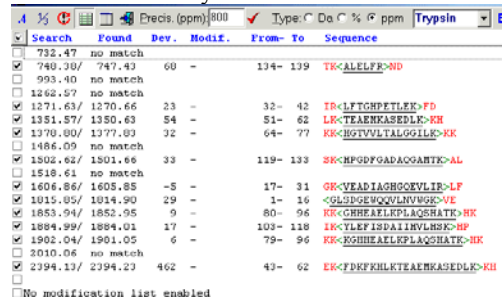
sequence' command . If the search program displays a complete sequence record (i.e. FastA or Swiss-Prot format) be sure to copy the complete record, as GPMW recognizes several formats and will try to parse the record directly. If the sequence is not recognized and parsed correctly, you can quickly do it manually (see the on-line help).

Mass search.

With the sequence safely on board you

perform a mass search  using the same mass list that you used for the Peptide Mass Search. Instead of using narrow search specificity (e.g. 10-30 ppm) that is common in PMS in order to narrow down protein hits, you should enter a high value (e.g. 800 ppm).

Once you have the peptide match window you will be able to see your 'exact' hits.




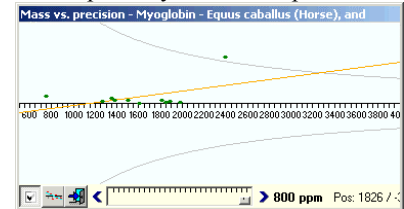
| Search | Found | Dev. | Modif. | From | To | Sequence |
|-------------------------------------|----------|----------|--------|------|----------|---------------------------------------|
| <input type="checkbox"/> | 732.47 | no match | | | | |
| <input checked="" type="checkbox"/> | 740.30/ | 747.43 | 60 | - | 134- 139 | TK<u>CALELFR</u><u>ND</u> |
| <input type="checkbox"/> | 993.40 | no match | | | | |
| <input type="checkbox"/> | 1262.57 | no match | | | | |
| <input checked="" type="checkbox"/> | 1271.63/ | 1270.66 | 23 | - | 32- 42 | TR<u>LFTGHPDTLEK</u><u>FD</u> |
| <input checked="" type="checkbox"/> | 1351.57/ | 1350.43 | 54 | - | 51- 62 | LR<u>TEARKASELE</u><u>KH</u> |
| <input checked="" type="checkbox"/> | 1370.80/ | 1377.83 | 32 | - | 64- 77 | HK<u>GDTVLTALGGILK</u><u>KK</u> |
| <input type="checkbox"/> | 1486.09 | no match | | | | |
| <input checked="" type="checkbox"/> | 1502.62/ | 1501.66 | 33 | - | 119- 133 | SK<u>CHFGDFGADACGARTK</u><u>AL</u> |
| <input type="checkbox"/> | 1518.61 | no match | | | | |
| <input checked="" type="checkbox"/> | 1606.86/ | 1605.85 | -5 | - | 17- 31 | GR<u>VEADTAGBCEVLLIE</u><u>LF</u> |
| <input checked="" type="checkbox"/> | 1615.05/ | 1614.90 | 29 | - | 1- 16 | <u>GLDPEKQQLLVGSE</u><u>VE</u> |
| <input checked="" type="checkbox"/> | 1853.94/ | 1852.95 | 9 | - | 80- 96 | HK<u>GHEAELEHPLAQSHATE</u><u>HK</u> |
| <input checked="" type="checkbox"/> | 1884.99/ | 1884.01 | 17 | - | 103- 118 | IK<u>CYLEFISDAIILVLSNKK</u><u>RP</u> |
| <input checked="" type="checkbox"/> | 1902.04/ | 1901.05 | 6 | - | 79- 96 | HK<u>GQHEAELEHPLAQSHATE</u><u>HK</u> |
| <input type="checkbox"/> | 2010.06 | no match | | | | |
| <input checked="" type="checkbox"/> | 2394.13/ | 2394.23 | 462 | - | 43- 62 | EK<u>DFKFKHLKTEARAKSEDLK</u><u>KH</u> |

If the 'Exact mass fit only' options were selected, the list will show only 'hits' that fit

with the enzyme specificity- without limit on missed cleavages, you can thus find more hits than in a normal standard PMS search.

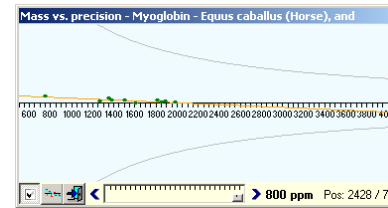
However, you will notice from the deviation ('Dev.') column that your precision may not be optimal (i.e. your calibration is slightly off).

Click on the 'Show graph' button  to get a scatter plot of your hits vs. precision:




This graph shows the mass along the x-axis and the precision along the y-axis. The gray curved lines indicate +/- 1 Da. Dots falling on these lines typically indicate cases where the monoisotopic peak has been wrongly assigned, deamidation of an amide (Gln, Asn), or perhaps reduced cysteines.


The yellow line shows the best multipoint linear calibration based on all the displayed dots. If the last item in the mass list is unchecked, the calibration line will move and look much more 'correct'.




An alternative approach is to increase the precision in the *Mass search results* window – as the 'hits' move out of the precision window, they are removed from the graph.

Press the *Calibration* button  and the mass list is recalibrated and the data transferred to the *Mass search results* window. In this case the 'hits' changed from -5 to 68 ppm to -29 to 19 ppm. In borderline cases this could change a 'potential hit' to a 'confirmed hit'.

By moving the slider below the scatter plot, you can move to a detailed view of the precision. In the results window you can

change from 'exact enzyme fit'  to 'check fit' – this will show you all potential non-specific cleavage sites. These can be shown in the scatter plot as red dots by clicking on the

checkmark button .

At the bottom of the results list, you can change the view to a report page that gives you a graphical overview of sequence coverage along with a list of hits which is colored according to tryptic PMS cleavage rules (if trypsin was selected as the cleavage enzyme). For details, see the manual or the on-line help under *Mass search result - Report*.

Peptide lists

You may end up with a list of peptides that you want to analyze. GPMW has a couple of ways in which to do this:

If you want to get the parameters you can generate a peptide window, just like a digest of a protein by 1) save the peptides to a text file, remember to use 1-letter code and have only a single sequence per line.

2) Select the main menu command *File | Import text | Peptide list from file* and select the file containing the peptides.

The list of peptides will be loaded into a sequence window in the order they are listed in the file. The size of each peptide is stored, as when all the peptides are read, a peptide window will open displaying the peptides in the file along with their physical/chemical properties. The peptides that are collected into a sequence can of course be saved to disk, but the peptide length information is lost as soon as the peptide window is closed.

You can also perform a mass search of a peptide list. GPMW lets you either perform a search for intact mass values or search for any fragment mass.

Start by selecting *Utilities | Search peptide list*. In the dialog box, you open a list of peptides – this can be either just a text file like described above, or it can be a simple FastA formatted file – first a name line starting with '>' with the sequence on the following line in upper case 1-letter code (this box will only read single line sequences of up to 250 residues and the maximum is 250 peptides).

When the list of peptides has been read, the buttons for reading or pasting a mass list becomes enabled. You can read the standard GPMW peak list formats.

You now set the parameters defining the mass list (charge and mass type) and the relevant search precision (in parts per million).

If you click on the 'Search peptides' button you will see a search of intact masses only, listed by search mass values. Selecting the 'Search fragments' button results in a search of all possible fragments of the peptides. The results are listed in order of search mass values and reported for each peptide sequence.

The result list can be copied to clipboard, saved to disk and printed using the buttons below the search buttons. In the edit box labeled 'Terminals' you can enter a composition that will be added to each peptide – works only for searching intact peptide masses.

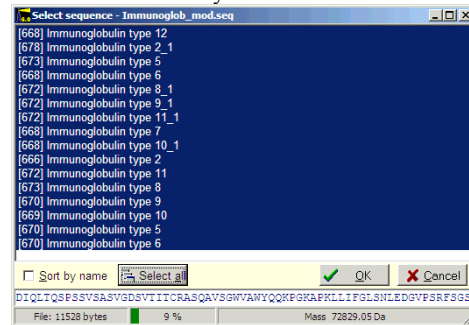
You may have occasions where you want to compare the physical/chemical of the peptides from different digests, either of the same protein or of different proteins. The immediate reaction is to copy the tables to the clipboard and transfer the data to a spreadsheet (Excel) or perhaps do it manually by pen and paper.

GPMW now offers you a different option. Create your digests on the GPMW desktop, right-click in any of the peptide windows and select the **DigestAlyzer** option. A new window opens, and all the peptide parameters from the various digests are copied. A number of parameters can now be compared in a 2D dot-plot.

Example:

A number of immunoglobulins had to be identified in a mixture. Their sequences were known, but we needed a positive identification of their presence. Just digesting with trypsin and identifying all we could would not guarantee a success, as the constant regions could easily 'drown' the variable parts of the molecules – we needed to establish conditions for separating the variable regions from the constant regions.

Initially all the sequences were saved in a single file so they easily could be opened with a single command – Open sequence, select the file, click on **Select all** followed by **OK**.

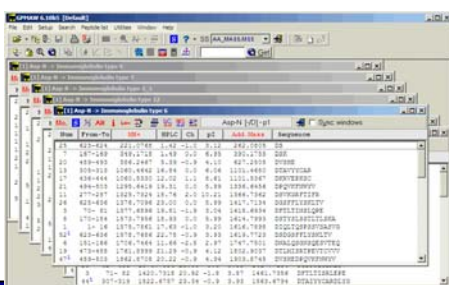
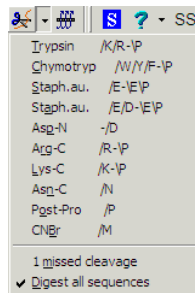


This opens all the sequences on the GPMW desktop. Now, to speed up digestion, in the

Quickdigest menu in the toolbar (see right) you check the **Digest all sequences**. We

could also check the **1 missed cleavage** to account for missed cleavages of the selected enzyme. Now

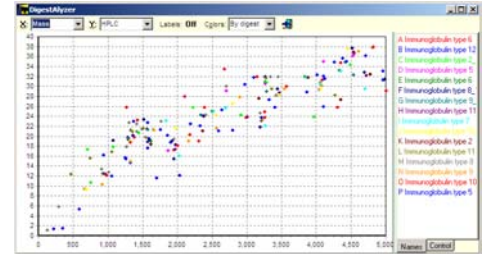
it is just to select the appropriate enzyme, in this case Asp-N, and all sequence windows will generate a digest peptide window. **Note**, the 'Digest all sequences' option is turned off as soon as you perform a digest – if you need to perform multiple 'all sequences' digests you



DigestAlyzer

have to turn the option on each time. The '1 missed cleavage' is a persistent selection and have to be switched off manually.

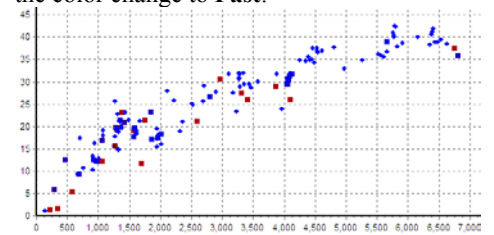
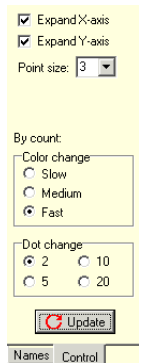
Right-click in any of the peptide windows and select **DigestAlyzer** from the pop-up menu. The resulting window shows a scatter plot of all the peptides with each digest in a different color.



The graph shows the peptide mass along the x-axis and the HPLC retention time along the y-axis. These values can be changed around, or you can choose to view pI or hydrophobicity index. In the mass/HPLC retention time index there is a clear correlation, which is not present when choosing any of the other indices.

As this is a rather complex mixture with many overlapping sequences (e.g. the constant regions) we switch the **Colors** options (top toolbar) from **By digest** (e.g. each digest has its own color) to **By color**.

To clarify things even further, we switch the right-hand panel to **Control**, expand the x- and y-axis, and set the color change to **Fast**.



The graph now shows co-eluting spots as squares and multiple co-eluting spots as progressively more reddish colored squares.

In order to get rid of the constant regions (the red squares) we can see that we have by selecting the late-eluting peaks (>31 min) we get a mixture of variable region peptides with only a single constant region peptide. By switching back to 'By digest' color, we can check that all the proteins are represented in this region.

The described procedure can then be rerun for other enzymes and cleavage specificities, but it turned out that Asp-N apparently was the best enzyme. However, when the 'real-life' experiment was run, the result was not encouraging, most likely because the peptides were too large to be efficiently separated on the reversed phase HPLC system.

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. Remember to supply your GPMAW license number.

If your license is too old for an upgrade (>1½ year), you can upgrade to the latest version for US\$ 140.-. This represents 50% off the price of a full version of the program. If you need additional copies you may buy them for just \$280.- first copy, \$210.- additional copies. This represents a discount of 25%. These prices includes new manual, postage and handling.

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

Predicting HPLC retention indices in GPMAW

Current research in my lab uses photoaffinity labeling with ¹²⁵I radioligands to identify the binding site of cocaine at the human dopamine transporter, a membrane bound protein that removes dopamine from dopaminergic synapses following release. The transporter, which is thought to contain twelve membrane-spanning domains, is expressed in a cell line (HEK 293) from which membranes are prepared for photoaffinity labeling.

The very low labeling efficiency of the aromatic azido group on the ligands, combined with the limited amount membrane preparation available, results in low picomole amounts of labeled protein. To identify the site of labeling, the ¹²⁵I photolabeled hDAT is digested enzymatically or chemically after an initial SDS-PAGE separation, leading to HPLC separation and scintillation counting of collected HPLC fractions. However, due to the very hydrophobic nature of the protein, it is not possible to predict a priori, which potential cleavage sites are actually cleaved. It is necessary to calculate the HPLC retention times of a range of peptides containing missed cleavage sites, which can be compared to the experimental results. As the photoaffinity label itself shifts the retention time to longer times, this must also be taken into account. We have labeled synthetic peptides of different sizes to estimate the shift produced by the label.

GPMAW has been very useful in the analysis of our data. Theoretical digests can be defined and generated for any set of amino

acid cleavages and the experimental HPLC conditions stored in the program to produce HPLC retention times closely matching our experimental results, as shown in figure 1. The programs ability to reorder the peptide list by different criteria has been helpful in comparing different digests.

To interpret the data from the different digests, the sets of peptides near the experimental retention times from different parallel and serial digests (trypsin, chymotrypsin, cyanogen bromide, thermolysin, and others) are combined to suggest the set of overlapping peptides containing the labeled site. These are synthesized and tested for a consistent interpretation of the site of labeling across digests.

Jay Justice, Emory University

Figure 2. Parameter settings for the peptide retention time prediction. For details on the parameters, please see the on-line help and the reference therein

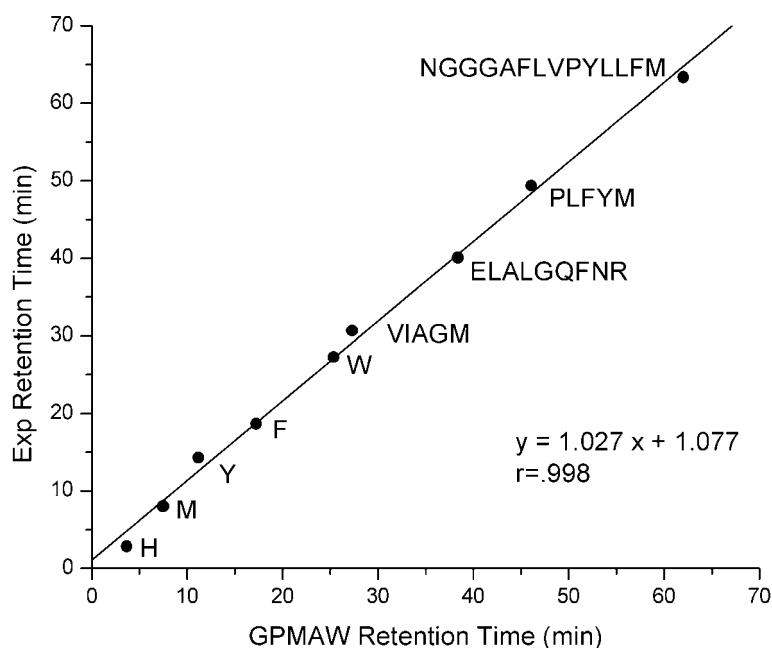


Figure 1. Correlation of experimental and calculated HPLC retention times for amino acids and peptides. The peptides used are synthetic peptides of sequences in the human dopamine transporter. The mobile phase was A = H₂O, 0.1%TFA, B = 100% acetonitrile, 0.1%TFA, at a flow rate of 1.0 ml/min. The gradient used was 0.67% B per minute for 90 minutes, after an initial 3 minute delay. The HPLC column was a Dupont Zorbax C8 column, 4.6 x 250 mm with a Supelco 4 x 2 mm C8 guard column. The GPMAW parameters used for the gradient HPLC were: T0 = 3.1 min, unretained sample; DT = 3.0 min, gradient delay time; B = 0.0067, gradient.