

GPMAW version 7.00

Updates and new features.

The main new feature of version 7 is a new interface library. This is most clearly displayed in the dialog boxes, where most have received a background consisting of a smooth gradient from white to gray. In addition, several of the secondary buttons now change color when the mouse passes over it, thus making it easier to identify which button will be activated. Input lists (e.g. mass lists) have also changed, so the movements of the focus for the active cell are slightly different, but should be more intuitive. To enter edit mode, you can click twice, press the F2 button or just start typing. For many edit fields, the associated button has now moved inside the edit field and will only show up when the edit field is active.

Annotation: When reading a record in Swiss-Prot or GenPept format, the whole record is kept with the GPMAW sequence as an annotation. The Swiss-Prot format is particularly attractive, as GPMAW will parse the data and extract information on post-translational modifications and put this information on a separate page in the annotation window where it can easily be turned on and off. Unfortunately the way the modifications were annotated was changed, so this function became broken. It is now fixed (and expanded) in version 7, and in addition you can now easily add your own modifications to the annotation.

An added bonus is that the annotated IPI databases are in the same format and can be downloaded from www.ebi.ac.uk/IPI/. After download you can convert them to FastA and index them just like Swiss-Prot using the DBIndex utility. The species compiled into IPI databases now comprise human, rat, mouse, zebrafish, Arabidopsis and chicken.

Modifications: The unimod database (www.unimod.org) can now be downloaded and accessed in the 'Edit modifications' dialog of GPMAW. A copy is included in the upgrade.

Peptide window: In the peptide window it has always been a bit longwinded to change column layout (except for changing between normal and alternate display). This has now changed, as you can right-click a column header to call up a list of all peptide list options. Click on any option to change the layout to that particular option. In addition the peptide window has received a few more options, e.g. more multiply charged species and mono/ave peptide mass.

Antigenic site prediction: A new graph window has been included with the purpose of prediction where in a protein sequence the most antigenic sites are located. The function enables you to combine various prediction methods.

Coverage analysis: The coverage analysis window introduced in the last version of GPMAW has been expanded and improved, both with regard to how the coverage is displayed, how the peptides are controlled, and how the lists are imported and exported.

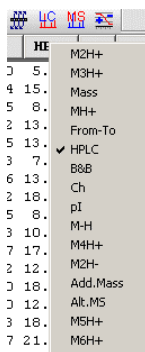
Ms/ms analysis: Minor errors have been corrected and you can now show neutral ion loss.

Fragment analysis: The fragment analysis dialog for interpreting ms/ms mass spectra of peptides has been made into a separate program. This helps to bring it on top of the ms/ms spectrum program. In addition it has been expanded with a b/y ion equivalent mass list. Fragment analysis can still be called from GPMAW.

Minor changes:

Sequence files can now be up to 256.000 bytes. Immonium ions can be displayed with 4 decimals.

Copy graphical information to the clipboard: this has in general been improved. However, you may still experience problems when copying to PowerPoint, Excel etc. Remembering to start the target application (e.g. PowerPoint) before copying from GPMAW to clipboard helps in most cases. If it still does not paste correctly, try the 'Paste special' command, and select 'Enhanced metafile'



Hook Head Lighthouse is believed to be one of the oldest operational lighthouses in the world. It has been in almost continuous operation since it was built in the 13th century.

The lighthouse is situated on the southern tip of Ireland, 50 km from Wexford. The tower is just under 36.6 m high and 13 m in diameter, 6.3 m at the top. It became automated in 1996 and is now open to the public.

Page 1

GPMAW version 7.00

Page 2

From the editor
Antigenic site prediction

Page 3

Beta versions
Fragment analysis

Page 4

Upgrading
Modifications

Antigenic site prediction

The theoretical prediction of antigenic sites in a protein can be a very useful starting point for the generation of antibodies for analytical purposes. As the synthesis of peptides and generation of antibodies is both costly and time consuming, tools to predict the most antigenic sites in a protein based on the primary structure

can be extremely helpful. No method currently exists that precisely predicts antigenic sites, but the method of Kolaskar and Tongaonkar claims a 75% accuracy. This method has now been implemented as a graphical window in GPMW (Graph | Antigenic site prediction):

From the editor

Once again it took slightly longer than expected to bring out the new version. One of the obvious differences this time is that most of the dialog boxes now look different, as they have obtained a background with a gradient of gray.

Although the colored gradient was put in to 'smarten up' the interface, it only adds a negligible overhead to the running of the program. In addition, a large number of other changes have been made like new buttons with colored 'mouse-over' to make it easier to discern which button will be activated, colored tabs on the notebooks, and new input tables that hopefully have a more logical feel to them than the old ones.

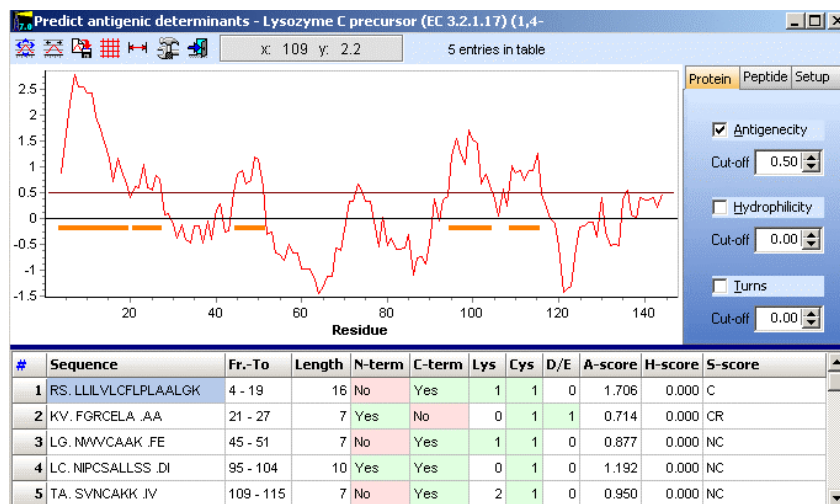
An additional bonus is that as I had to go through all dialog boxes in the program quite a number of small annoyances (definitely not bugs!) were corrected and small improvements were carried out.

For the next version I will concentrate on improving the analysis of individual programs, i.e. when making a peptide map of a known protein, how to identify each peptide and locate modifications. Hopefully I will also get the VEMS search-engine implemented (R. Matthiesen, J. Proteome Res., 2005) to replace the old search engine of GPMW. Furthermore, I will try and interface GPMW to some of the good but scattered programs available out there on the Internet.

If you have any specific requests, I will try and squeeze them in. However, no chance of running out of work.

Merry Christmas.

Peter Højrup



The window is divided into three parts:

1) The central window, which shows a sliding-window graph of the currently selected method(s). In addition you will see the selected cut-off value as a colored horizontal line and the areas that fulfill the criteria as outlined on the 'Peptide' and 'Setup' tabs will be highlighted by orange bars.

2) The right-hand panel shows a three tab notebook where the first tab enables you to select any of three methods: Antigenicity as predicted by Kolaskar and Tongaonkar, hydrophilicity (Hopps and Wood), turns (Garnier). The first method builds on experimentally determined antigenic sites, where hydrophobic residues Cys, Leu, Val are observed to be antigenic if present on the surface, the second method on the hypothesis that hydrophilic areas are surface exposed (and may therefore be antigenic), and the last method on the observation that antigenic sites are often observed in beta-turn regions.

GPMW allows you to mix and match each method and select a separate cut-off value for each. **Note:** As all three methods are based on the sliding window principle, you should always aim for the high decisive peaks.

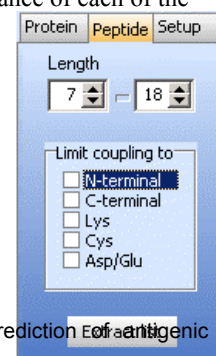
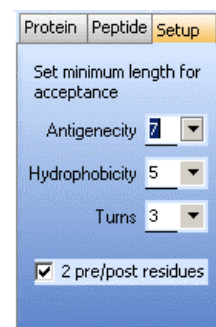
3) The bottom panel lists the peptides that are detected by the above methods and are shown as orange bars in the graph. From left to right in the table is listed: The sequence with preceding and following residue, the location

in the sequence, length of peptide, is the N-terminus the only amino group available for coupling (Yes/No), is the C-terminal carboxyl group the only available (Yes/No), number of Lys, Cys, Asp+Glu residues present, average antigenic propensity score, average hydrophobicity score, and finally (S-score) residues that may be problematic during synthesis.

The 'Peptide' tab enables you to filter the peptide list based on peptide length (peptides of various length that fit within the limits will be generated) and you can limit coupling to

peptides that only couple through the N- or C-terminus or have at least one residue of Lys / Cys or Asp+Glu. Press the 'Extract list' button to create the filtered list.

The last tab, 'Setup', lets you set the minimum peptide length for acceptance of each of the three methods. You can also select whether you want to display two residue before and after each peptide in the result list.



References:

- A.S. Kolaskar and Prasad C. Tongaonkar: 'A semi-empirical method for prediction of antigenic determinants on protein antigens'. FEBS Lett. **276**, 172-172, 1990.
- T.P. Hopp and K.R. Woods: 'Prediction of protein antigenic determinants from amino acid sequences'. Proc. Natl. Acad. Sci. USA. **78**, 3824-3828, 1981.
- J. Garnier. D.J. Osguthorpe and B. Robson. 'Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins.'. J. Mol. Biol. **120**, 97-120, 1978

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).

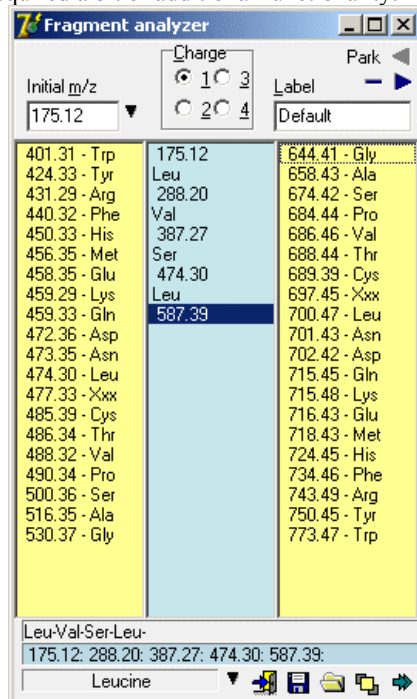
You are as always welcome to suggest improvements and changes to the program. If you find errors in the program, try first to download the current beta version (or at least check the web site if the error has been corrected) before contacting Lighthouse data. However, if you find an error, we always 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".

Fragment analysis

The interpretation of ms/ms fragmentation spectra of peptides can be a daunting task. Many programs have been developed to this automatically, but even the best of programs sometimes make it wrong and have difficulties rivaling an experienced interpreter. But even the experienced user have a lot of work to do calculating mass differences and keeping track of related ions (e.g. b/y ions).

To help both the novice and the experienced ms/ms interpreter, GPMaw has for some time had the 'Fragment analysis' dialog box

(called from the toolbar  or through Utilities | Fragment analyzer). In version 7.0 this dialog box has changed to a separate program, changed view slightly and has acquired a bit of additional functionality.




It still runs as rather small window where you in the top left input box type an initial mass value. The drop-down arrow lets you select an initial mass of 175/147 for the C-terminal y-ion of Arg/Lys terminated peptides, but you can enter any value you like. A good starting point is often a high peak in the middle or at slightly higher mass.


When you press enter, the mass is copied to the central blue column. You can then expand the sequence tag by double clicking on the right-hand column to move to a higher mass or the left-hand column to move to a lower mass. The new 'residue' is always added to the point that has the focus in the central column. This means that if you make an erroneous assignment (or want to try another possibility) you just select the point of change and make another choice.

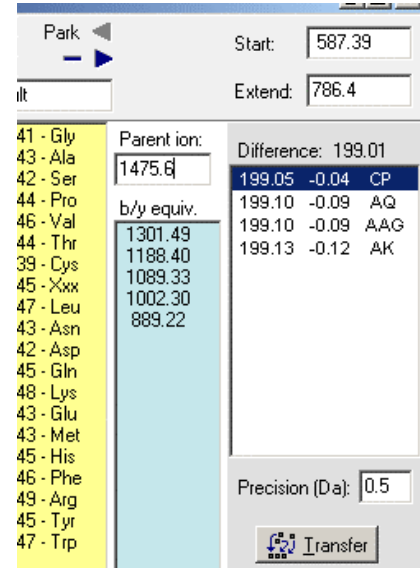
The **Label** is used if you want to save your sequence tag or if you want to use the **Parking** area. In the top right corner you can click on the blue arrow to 'save' your status.

You can then retrieve it by clicking on the left turned blue arrow. Up to five peptides can be saved.

The bottom right-hand corner has a button

 to bring the window to float over other

windows and a button  for expanding the window to include two additional columns:



If you enter the parent ion for the ms/ms fragmentation you will get a list of all b/y ions corresponding to the main fragment list. This means that if your parent ion has an m/z of 1475.6, and you have determined a fragment ion to be 474.3, then if this ion is a b or y ion, the opposite fragment (y or b respectively) will be 1002.3. Identifying both the y- and the b-ion is a great confirmation of the fragment assignment. The math is not exiting (b-ion + y-ion = parent ion + 1), but having the numbers in a table beats the shit out of doing it on a calculator.

The last column is an 'extension column'. Often when working with an ms/ms spectrum, you will find that you are missing a fragment or two in order to 'get to' the next ion. Here the extension column comes in. Click on the ion from where you want to start, and it will be copied into the 'Start' box. Now enter the m/z of the ion to extend to and the difference and all possible amino acid residue combinations that fit within the given precision will be listed. Often you can use this list to 'hunt' for peaks in the background; alternatively you can just click on the 'Transfer' button to copy the list to your main fragment list. **Note**, the order of residues transferred is random.

When called from GPMaw, the mass list for the residues will be the one used in GPMaw. An alternative mass list can be loaded through the pop-up menu (duplicated in the down-arrow in the bottom panel).

Upgrading

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

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

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

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

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

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


If you want to upgrade and your license is too old, you can upgrade to the latest version for US\$ 150.-. This represents 50% off the price of a full version of the program. If you need additional copies you may buy them for just \$210.- each. This represents a reduction of 25%. These prices include postage and handling.

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

Modifications

Analysis of secondary modifications seems to be an ever-returning question. Two problems are common: One is to identify the modification (i.e. you find a mass addition of 71 Da what can that be?), the other problem is: Once the modification has been determined, how get it into GPMaw? Several modification databases are available on the Internet. One of the oldest and most used is DeltaMass (www.abrf.org/index.cfm/dm.home), however, this reference contains integer mass values and most of the modifications are recorded without a chemical formula, which makes it unsuitable for immediate use in GPMaw.

Another resource, although considerably smaller, is the Unimod database. This database in general has chemical formulas for each modification and is available for download in XML or Mascot format. In the Edit | Edit modification file, GPMaw now has the option to read the downloaded XML version of Unimod and add selected modifications to your own local modification file.

In the 'Edit modification file' dialog, you click on the 'Unimod' button  to open the Unimod database (see below).

Entries in the Unimod can be copied to your local modification file (first available slot) by either double-click on the entry or select an entry and press the 'Copy to list' button. However, a few comments on how GPMaw treats the Unimod:

The database is shown in numerical mass order. The long and short name is taken from the Unimod record. When copied to GPMaw the short name is used, and it may be truncated by GPMaw. The name can be edited in the GPMaw part of the dialog box after being copied.

The mass value is taken from Unimod and may be calculated slightly differently in GPMaw. The 'Formula' is calculated by GPMaw based on the Unimod formula (next column) – please

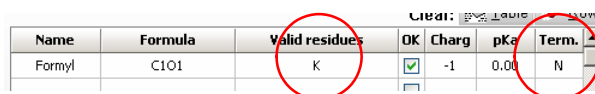
check that it has been translated correctly – if not you can correct it after it has been copied.

The ID is the record ID in Unimod and is only of interest if you need to refer to Unimod.

Residues are the ones that can accept the modification. A slightly different convention than the one normally used in GPMaw is found here: '>' means N-terminal and '<' means C-terminal modification.

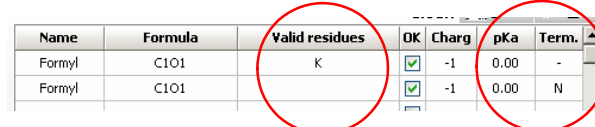
If an entry in Unimod have both **residues and a terminus** in the 'Residues' column (e.g. modifications of amine that affects both N-terminal and lysine side chain) both functions will be transferred to GPMaw, but in separate entries that will have the same name and formula.

This means that instead of this:



Name	Formula	Valid residues	OK	Charg	pKa	Term.
Formyl	C1O1	K	<input checked="" type="checkbox"/>	-1	0.00	N

You will see a result like this:



Name	Formula	Valid residues	OK	Charg	pKa	Term.
Formyl	C1O1	K	<input checked="" type="checkbox"/>	-1	0.00	>
Formyl	C1O1		<input checked="" type="checkbox"/>	-1	0.00	N

The Unimod database is now part of the standard installation of GPMaw, but you can download updated versions directly from the Unimod web site and install it in GPMaw: On the Unimod home page click on 'Help' in the top line. On the help page you will find the 'XML' about one page down. Right-click on the 'XML' and select 'Save target as...' from the pop-up menu. In the 'Save as' box you save the file as 'unimod.xml' (default) in the \gpmaw\system\ directory, overwriting the existing file. Upon restarting, GPMaw will recognize the new file and parse it for use.

NOTE: If your pet modification is missing from the Unimod database, do not report it to me, but add it directly to the database yourself. If you log on to the web site, it is possible for the end-user to add records – the help of all other users.

Double-click or select and press "Copy to list"		Copy to list					
#	Long name	Short name	Mass	Formula	Unimod	ID	Residues
70	dimethyl labeling (D)	Dimethyl-D	34.07		H(-2) H2(6) C(2)	189	>K
71	dimethylated arginine	Dimethyl-Heavy	36.08		H(-2) H2(6)	330	R
72	Acrolein addition +38	Acrolein38	38.02	H2C3	H(2) C(3)	207	K
73	Propionaldehyde +40	Propionald+40	40.03	H4C3	H(4) C(3)	256	KH
74	amidination of lysines or N-terminal	amidine	41.03	H3C2N1	H(3) C(2) N	141	>K
75	Acetate labeling reagent light form	Acetyl_light	42.01	H2C2O1	H(2) C(2) O	57	>K
76	N-Acetylation	N-Acetyl	42.01	H2C2O1	H(2) C(2) O	15	>
77	Acetylation	Acetyl	42.01	H2C2O1	H(2) C(2) O	1	>K
78	Guanidination	Guanidination	42.02	H2C1N2	H(2) C N(2)	52	K
79	tri-Methylation	tri-Methylation	42.05	H6C3	H(6) C(3)	37	KR