

E – Getting data out of GPMAW

In a field as varied as protein chemistry, it is not possible just to use a single rather specialized program like GPMAW for all your protein analyses. Particularly when using the Internet with the ready availability of (free) programs, it is of interest to be able to quickly and efficiently transport protein sequences around. Another aspect is the handling of larger projects, where you typically use a word processor or spreadsheet to keep track of your data. For these programs, GPMAW also have some functions that enable you to do as little handling as possible in the target program.

1 – Protein sequence.

The most obvious way of getting a sequence from GPMAW to a report is to copy to the clipboard.

Select the sequence window you want to copy, press Ctrl-C or select **Edit | Copy to clipboard**. GPMAW shows a timed dialog box (the box is on screen for a few seconds and terminates automatically) telling you that the sequence has been copied. This operation copies the sequence to the clipboard in the format on screen (1- or 3-letter code).

Note: If you have one or more peptides (regions) highlighted, only those portions of the sequence will be copied to the clipboard! Different highlighted regions will be copied as separate lines.

When you paste a sequence into a report, you should in most cases select a monospaced font for display, as the sequences will line up correctly eg.

```
Courier      AGSYLLEELFEGHLEKECWEEICVYEEAREVFEDDETTDE  40
             FWRTYMGGSPCASQPCLNNGSCQDSIRGYACTCAPGYEGP  80
             NCAFAESECHPLRLDGCQHFCYPGPESYTCSCARGHKLGG  120
```

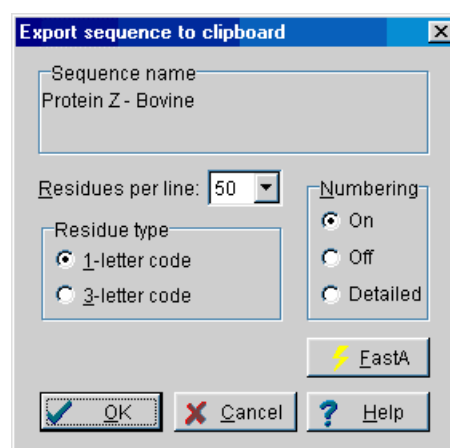
```
Arial (Swiss) AGSYLLEELFEGHLEKECWEEICVYEEAREVFEDDETTDE  40
              FWRTYMGGSPCASQPCLNNGSCQDSIRGYACTCAPGYEGP  80
              NCAFAESECHPLRLDGCQHFCYPGPESYTCSCARGHKLGG  120
```

Although the Arial font is nicer to look at, it is much more difficult to find your way around a sequence. If you export in the detailed mode, things get really screwed up unless you choose a monospaced font.

The method above has one disadvantage: you only copy the sequence and not the name. In order to include the name you have to 'export' the sequence.

Select **File | Export sequence | To clipboard**. This opens the dialog shown on the right. Here you have a number of options to format the output to fit with your report. The **Residues per line** is a dropdown box that lets you select from 10 to 100 residues per line. The **Residue type** is set as your screen display, but can be changed. **Numbering** can be selected **On** (see example above), **Off** (no numbers) or **Detailed**:

```
Protein Z - Bovine (396 res.)
      10      20      30      40
AGSYLLEELFEGHLEKECWEEICVYEEAREVFEDDETTDE
      50      60      70      80
FWRTYMGGSPCASQPCLNNGSCQDSIRGYACTCAPGYEGP
      90     100     110     120
NCAFAESECHPLRLDGCQHFCYPGPESYTCSCARGHKLGG
```



The **FastA** button puts a '>' in front of the name (e.g. '> Protein Z – Bovine'), selects 60 residues per line, 1-letter code and numbering off. This makes it easy to copy a FastA formatted sequence to another program (e.g. on the Internet).

When you click '**OK**' the sequence is copied to the clipboard.

The **annotation page** has a small trick when copying to the clipboard. If you select the **Edit | Copy to clipboard** or **Ctrl-V** command you will always get the whole annotation copied. However, you can copy part of the annotation by highlighting the relevant portion, right-click in the window and select the **'Copy'** command from the pop-up menu.

2 – The peptide list.

The peptide list is the result of the cleavage of a protein. It is one of the main features of GPMW. This is mainly due to the fact that although the primary structure of a large number of proteins is known (mostly based on nucleotide data), the analysis of intact proteins is still very difficult. Thus you have to cleave the protein into specific smaller fragments, peptides, using specific enzymes or chemicals. Although the calculation of chemical/physical parameters like mass, pI and HPLC retention times is possible, either by hand or programs freely available on the web (e.g. www.expasy.ch), these programs are often cumbersome to work with and are often not designed for the mass spectrometrist.

The generation of the peptide list is fairly straightforward, and will not be covered here (see the manual and online help for more details).

Num	From-To	Mass	HPLC	Ch	pI	Sequence
13	87- 88	234.09	2.00	1.0	3.85	Ser-Glu-
4	16- 17	275.15	1.51	2.0	6.11	Lys-Glu-
7	28- 30	374.19	2.81	2.0	6.36	Ala-Arg-Glu-
9	34- 36	377.11	2.03	0.9	3.07	Asp-Asp-Glu-
8	31- 33	393.19	10.47	1.0	3.85	Val-Phe-Glu-
2	9- 11	407.21	14.60	1.0	3.85	Leu-Phe-Glu-
3	12- 15	454.22	9.39	2.0	5.36	Gly-His-Leu-Glu-

Once generated you may sort the list based on any of the displayed parameters by clicking on the header. The first click will sort in descending order, while clicking a second time will sort in ascending order.

Pressing **Ctrl-C** or selecting **Edit | Copy to clipboard** (or from the pop-up menu) will copy the entire content to the clipboard.

The copy will be just like the list displayed, so you should make sure the sorting, monoisotopic / average mass, 1-/3- letter code etc. is correct before copying.

You may copy only part of the list by selecting only some lines. You can select a continuous range of entries by clicking on the first one and holding down 'Shift' while clicking on the last one. If you hold down 'Ctrl' you can add and remove single entries from the selection. If you start by selecting a continuous stretch, you can add and remove single entries afterwards.

30	345-347	370.23	4.81	2.0	10.35	Val-Pro-Arg
35	369-372	416.21	4.53	2.0	10.35	Gly-Gln-Gly
33	362-365	429.23	4.48	2.0	10.35	Ala-Ser-Pro
28	318-320	440.21	3.70	3.0	7.21	Glu-His-Arg
16	202-205	523.32	9.82	3.0	10.55	Leu-His-Val
15	198-201	545.27	10.43	3.0	9.92	Ser-His-Phe
8	118-122	587.30	9.49	1.9	5.94	Leu-Gly-Gln
25	296-300	593.24	9.94	2.0	6.36	Thr-Ser-Cys
26	301-308	643.34	5.70	2.0	10.35	Gly-Ala-Ala
21	268-273	743.40	13.60	2.0	6.36	Glu-Met-Val
12	160-166	747.38	7.14	2.0	6.36	Leu-Thr-Asn
17	206-212	804.44	8.23	4.0	10.36	Gly-Val-His
31	348-353	820.46	21.76	2.0	9.00	Tyr-Ala-Leu
10	124-130	825.37	15.37	3.0	6.75	Ser-Cys-Leu
27	309-317	913.51	15.51	2.0	10.35	Trp-Val-Ala
26	378-381	940.43	9.68	1.9	3.07	Leu-Glu-Glu

When you want copy only part of the list, you have to select at least two peptides, as GPMW will otherwise just go ahead and copy the whole list.

Two entries in the **Peptide | Setup** are important when copying: copy as text vs. tab delimited and copy full sequence vs. limited sequence.

When you copy as text, each column is separated from the next by a space character. This makes it easy to align columns if you use a monospaced font (e.g. Courier). If you select 'tab delimited', each column is separated from the next by a 'tab' character. This means that you have to set the tabs properly in the report. E.g.

Table as text delimited (**Courier** font):

Num	From-To	Mass	HPLC	Ch	pI	Sequence
33	362-365	429.23	4.48	2.0	10.35	Ala-Ser-Pro-Arg-
28	318-320	440.21	3.70	3.0	7.21	Glu-His-Arg-
16	202-205	523.32	9.82	3.0	10.55	Leu-His-Val-Arg-
15	198-201	545.27	10.43	3.0	9.92	Ser-His-Phe-Arg-

Copy table to clipboard

Copy table as text

Copy tab delimited

Table sequence

Limited sequence

Full sequence

Arial font:

Num	From-To	Mass	HPLC	Ch	pI	Sequence
33	362-365	429.23	4.48	2.0	10.35	Ala-Ser-Pro-Arg-
28	318-320	440.21	3.70	3.0	7.21	Glu-His-Arg-
16	202-205	523.32	9.82	3.0	10.55	Leu-His-Val-Arg-
15	198-201	545.27	10.43	3.0	9.92	Ser-His-Phe-Arg-

Table as tab delimited (**Arial** font):

Num	From-To	Mass	HPLC	Ch	pI	Sequence
33	362-365	429.23	4.48	2.0	10.35	Ala-Ser-Pro-Arg-
28	318-320	440.21	3.70	3.0	7.21	Glu-His-Arg-
16	202-205	523.32	9.82	3.0	10.55	Leu-His-Val-Arg-
15	198-201	545.27	10.43	3.0	9.92	Ser-His-Phe-Arg-

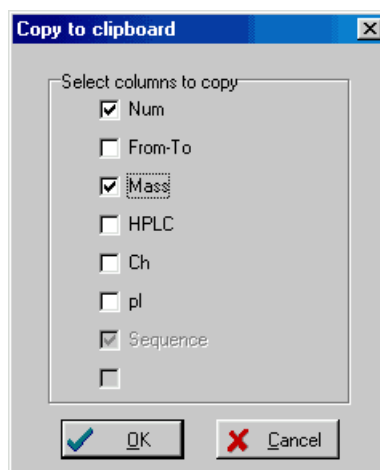
Courier font:

Num	From-To	Mass	HPLC	Ch	pI	Sequence
33	362-365	429.23	4.48	2.0	10.35	Ala-Ser-Pro-Arg-
28	318-320	440.21	3.70	3.0	7.21	Glu-His-Arg-
16	202-205	523.32	9.82	3.0	10.55	Leu-His-Val-Arg-
15	198-201	545.27	10.43	3.0	9.92	Ser-His-Phe-Arg-

When you copy columns to a **spreadsheet** (e.g. Excel) you **delimited** as this will transfer columns to individual columns. You can set your spreadsheet up to accept space-delimited columns, but this is fraught with errors.

Instead of copying the complete table, you may be interested in only copying a few of the columns. You could go into **Peptide | Setup** and change the layout of the peptide table, but it is much easier to right-click in the table and select **Copy/Export | Copy columns to clipboard** from the pop-up menu. In the copy to clipboard dialog box, you can then select the columns to copy. The title for each tick-box is taken from the actual header of the peptide table. The 'Sequence' column is always selected.

Like when copying the complete table you can select a range of peptides before you start the copy operation.



3 – Mass search results

The results of the mass search works very much like the peptide list described above. The main difference lies in the selection of lines to report. Where the peptide list is a standard multiple selection list, the results of the mass search is a check box selection list.

This works in the way that if no lines have been selected (checked) the whole list is copied. If one or more lines have been checked you are asked whether you want only the selected lines copied (Yes – selected lines only; No – whole list; Cancel – cancel copy operation).

<input checked="" type="checkbox"/>	Search	Fr
<input type="checkbox"/>	735.48/	73
<input type="checkbox"/>	- /	73
<input checked="" type="checkbox"/>	- /	72
<input type="checkbox"/>	- /	73
<input type="checkbox"/>	- /	73
<input type="checkbox"/>	748.44/	74

The **'Check'** button atop the check boxes, works to check/uncheck all lines in a single operation.

The individual check boxes can be checked/unchecked by clicking on them with the left mouse button. Alternatively you can use the arrow keys to move up and down the list and use the space bar to check/uncheck lines (usually faster than using the mouse).

A shortcut exists in the pop-up menu to check all peptides that fits with the cleavage pattern of the enzyme used in the search (**Selected peptides | Check perfect fits**). Another option inverses all selections (e.g. unchecks all checked items and visa versa - **Selected peptides | Toggle selections**).

Unlike the peptide list, you cannot select individual columns for transfer.

Depending on how you make out your report and whether you copy to a spreadsheet, you have to set the **Peptide | Setup** correctly (see 'Peptide list' above).