DISCLAIMER

This workbook was written using Excel X on Mac OS X. Whilst it will probably work on recent versions
of Excel for windows, you may experience some problems.

If you benefit from the use of this spreadsheet, please be kind enough to provide an acknowledgment
in any publications resulting from its use.

Some General Comments on this Workbook

This workbook has been designed with the primary goal of facilitating mass calculations on synthetic peptides. As such, it is restricted to
calculations on molecules containing the following elements only: C, H, N, O, F, Cl, Br, I, S, P, Na, K, Mg, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn,
 Ag, Cd, Pt, Au, Hg.

The calculations are based around amino acid residual masses - basically the mass of the amino acid minus water. This means that the
N- and C-terminal residues need to be declared explicitly for all except backbone-cyclic peptides.

All calculations for average mass are based upon the full isotope distribution of the element. Calculations of monoisotopic masses
however, use only the most abundant (lower mass) isotopes. Isotope masses are the IUPAC values from 2003 (see:
De Laeter, J. R. *et al, Pure Appl. Chem.* **75**, 683-800 (2003)).

Standard abbreviations are used throughout, with the exception of TFA (used here to represent 2,2,2-trifluoroacyl).

As this workbook focuses on peptides, it has been restricted to calculations related to positive ions.

How to Use This Workbook

This workbook comprises two calculation sheets and two reference sheets:
> "Mass-Calc" is used to calculate the mass of peptides and proteins
> "Search" is used to calculate the composition of a peptide using MS data
> "library" contains reference data for elements and amino acids.
>> Note 1:   the area marked in turquoise contains the elemental data, and can be adjusted, if appropriate.
>>   If adjusted, all of the other values in the workbook will automatically be recalculated.
> "tmp" is hidden and contains data relating to calculations. ***This should only be modified with caution***

A. Performing Mass Calculations

1. Enter the sequence of your peptide or protein on the sheet "Mass-Calc", starting row 5, with one residue per row.

2. Select the groups at the N- and C-termini of the peptide

3. Select the appropriate options in the other boxes (phosphates, disulphides, backbone cyclic)

4. If you have any groups that are unusual, you may include these in the "additional group" boxes.

> Note 1:   These should be entered according to their formula, in the order C, H, N, O, F, Cl, Br, I, S, P, Na, K
> *e.g.* C2H3O for acetyl

> Note 2:   If the group is at the N-terminus, select the appropriate option on the drop-down menu in the first box
> This simply adds the mass of the group to the mass of the peptide sequence with the necessary
> corrections and adjusts the amino acid formula.

> Note 3:   One aspect to pay attention to is the "number of H-atoms replaced" drop-down box. The appropriate
> value needs to be selected here. For example, substitution of I for H on an aromatic ring increases
> the mass by 1 x I, minus the H-atom replaced - *i.e.* the number of H-atoms replaced = 1
> In the case of acetonide formation (*e.g.* pseudoprolines), the number of atoms replaced = 2

5. If your protein has a bound metal, select the appropriate option from the drop-down box.

> Note 1:   For a neutral protein, the number of H replaced will be the same as the charge on the metal
> *e.g.* for Zn^{2+}, the number of H replaced = 2

6. Other elements may be included using the box at the top right of the sheet.

> Note 1:   You will need to enter the element symbol and the average and monoisotopic masses.
> Note 2:   This will also work for any group, provided that an average and monoisotopic mass is available.
> For example, this could be used to calcualte the mass of a protein containing selenomethionine
> by entering a mass that corresponds to the difference between the elements S and Se.

7. The monoisotopic and average masses are calculated automatically, as are the formulae for the elemental and amino acid compositions.

8. The expected ions for ESMS of the target structure are displayed at the bottom.

> Note 1:   The 1+ ion is also the species expected to be observed by MALDI MS.

EXAMPLE 1:   Below is an example calculation for Gramicidin A

HCO-Val-Gly-Ala-D-Leu-Ala-D-Val-Val-D-Val-Trp-D-Leu-Trp-D-Leu-Trp-D-Leu-Trp-NHCH2CH2OH

Sequence Enter 1 Letter Code		
	3-letter	**1-letter**
residue 1	Val	V
residue 2	Gly	G
residue 3	Ala	A
residue 4	Leu	L
...	Ala	A
...	Val	V
...	Val	V
	Val	V
	Trp	W
	Leu	L
	Trp	W
	Leu	L
	Trp	W
	Leu	L
	Trp	W

N-terminus For

C-terminus NH2

Gramicidin A

Backbone Cyclic? No

Number of Disulphides 0

Number of Phosphates 0

Formula of Additional Group	C2H5O	
Mass of Additional Group	44.0532824	44.02621475

Include how many of this group? 1

Is this group at the N-terminus? No

Number of H atoms Replaced 1

Formula of Additional Group		
Mass of Additional Group	0	0

Include how many of this group? 0

Number of H atoms Replaced 0

Formula	C99H140N20O17
Amino Acid Composition	For-Ala2GlyLeu4Trp4Val4-NH2 (C2H5O)

Monoisotopic Mass	1881.07053
Average Mass	1882.33193

Electrospray		M<n+>	m/z mono	m/z average
		1	1882.078358	1883.339909
Charge Agent	H+	2	941.543091	942.173942
		3	628.031336	628.451954
Charge Range	1-5	4	471.275458	471.590959
		5	377.221932	377.474363

EXAMPLE 2:   Gramicidin S

[cyclo-(D-Phe-Pro-Val-Orn-Leu)$_2$]

Sequence Enter 1 Letter Code		
	3-letter	1-letter
residue 1	Phe	F
residue 2	Pro	P
residue 3	Val	V
residue 4	Orn	O
...	Leu	L
...	Phe	F
...	Pro	P
	Val	V
	Orn	O
	Leu	L

N-terminus [H]

C-terminus [OH]

Gramicidin S

Backbone Cyclic? [Yes]

Number of Disulphides [0]

Number of Phosphates [0]

Formula of Additional Group		
Mass of Additional Group	0	0

Include how many of this group? [0]

Is this group at the N-terminus? [No]

Number of H atoms Replaced [0]

Formula of Additional Group		
Mass of Additional Group	0	0

Include how many of this group? [0]

Number of H atoms Replaced [0]

Formula	C60H92N12O10
Amino Acid Composition	Leu2Orn2Phe2Pro2Val2

Monoisotopic Mass	1140.70594
Average Mass	1141.46973

Electrospray		M<n+>	m/z mono	m/z average
		1	1141.713762	1142.477705
Charge Agent	H+	2	571.360794	571.742840
		3	381.243137	381.497885
Charge Range	1-5	4	286.184309	286.375408
		5	229.149012	229.301922

B. Performing Mass Searches

1. Select the "Search" sheet and enter your MS data into any of the cells in the rose-coloured section of the sheet

Note 1:	The search will use all of the data that you enter here (including combinations of MALDI and ES data).
Note 2:	Use 'mono' for monoisotopic masses, and 'average' for average ion masses.
Note 3:	For searches on ES data, the calculation will determine which ionic species best fits the raw data, and will report this in the answer along with the associated charges of the calculated ions. Calculations may be performed on between 1 and 5 ions, with the highest observed mass entered in the top-most row. Gaps are tolerated, as long as the masses of ions entered are in the correct relative locations.

2. For searches on MALDI data, select the ionic form of the (singly-charged) species using the drop down menu.
3. In the amino acids section of the sheet, set the parameters for all of the amino acids

Note 1:	The initial values on the first occasion that the sheet is opened are: Include all amino acids set to 'no', the minimum number of amino acids in the peptide = 0 and the maximum = 10. The default of 'no' is the better option, as searches are considerably faster if only those amino acids that can potentially be in the peptide are used.
Note 2:	Searches execute considerably faster (like, hours faster) if the minimum and maximum values are within ± 2 of the expected value. For a synthetic peptide, it is typically best to start with the maximum value for each amino acid used set to the anticipated number of that amino acid in the peptide, and the minimum set to the maximum value -2 or -1.
Note 3:	Searches are restricted by the usual limitations: Ile and Leu cannot be distinguished, and some dipeptide combinations have identical masses, *e.g.* G+G=N, I+G=V+A *etc.* Nevertheless, for most synthetic peptides, where the ideal combination of amino acids is known, this is less of a restriction, allowing informed conclusions to be made.

4. Select suitable options for N- and C-terminal residues, number of disulphides *etc.*

Note 1:	It is possible to do an explicit search for a backbone cyclic product. If however, you are unsure that a cyclisation reaction has occurred, the 'backbone cyclic' option should be set to no, and the 'possible dehydration' option set to yes. The latter enables *both* products with no-dehydration and products minus one water molecule to be returned as solutions. Dehydrated products are indicated by "-H2O" in the formula.

5. In the "Other Groups" section, tick all of the boxes that correspond to groups that you would like to include in the search.
6. In the "Additional Groups" section, tick all of the boxes and enter mass data for extra groups that you would like to include in the search.

Note 1:	Monoisotopic and average Masses may be calculated using the "Mass Search" worksheet.
Note 2:	In the '#H Replaced' section, indicate the number of H-atoms substituted by the group (see A: Note 3 above)

7. Click on the search button in the blue box, which will start the search

Note 1:	The solver performs an integer search using the groups/values that you entered. This may take a long time to complete if you have allowed the 'min' and 'max' variables to cover too wide a range. In most cases, the search will complete within 10 min.
Note 2:	After 15 min, a dialogue box will appear informing you that the maximum time limit has expired. This is here as a failsafe - you may either terminate the calculation at this point, or allow it to continue.

8. Once the search is complete the result is displayed in the red-coloured area. You may keep this result by clicking on the 'Store' button.

Note 1:   When the store button used for the first time, you will be prompted for the name of a new worksheet on which to store the values. If you delete this sheet ("notes"), you will need to edit the 'keepData' macro or create more than one results sheet.

Note 2:   If a results sheet already exits, the current results will be added to it. **N.B. it is important that the results sheet is the third sheet in the workbook** - only move it once you have finished calculations.

Note 3:   It is sometimes useful to store a result and then restart the search, in order to find other results that are close to the target mass (particularly when searching against low resolution spectra). All of the 'good' results may added to the results sheet as you proceed.

Note 4:   **IMPORTANT**. IF the search ends almost immediately, or if an answer is returned that contains fractions of amino acids, then the search has not converged. Upon convergence, the rms difference should be of the order of ≤ 1.

9. If the search does not converge, or you wish to search for similar solutions, you may restart the search.

Note 1:   There are two options - either restart immediately, or after adjusting the search parameters. The latter approach is particularly useful if the search finished after a few seconds without converging.

Note 2:   **IMPORTANT**. Remember to hit the 'reset' button if you want to restart a search from scratch. This does not clear any of the search criteria that were entered on the "Search" sheet (amino acids included, min/max values *etc.*), but resets values on the "tmp" worksheet that may otherwise carry over between calculations.

EXAMPLE: The following peptide was the target of a synthesis that involved α-allyl protection of an aspartate group that was linked to an amine solid-support *via* the side chain carboxyl (to produce Asn after cleavage). The allyl deprotection was performed on the solid-support, and the peptide cleaved by normal procedures following cyclisation. HPLC and low resolution MS (MALDI) analysis indicated that a minor product had a mass of 3343.4, ~176 mass units less than the target.
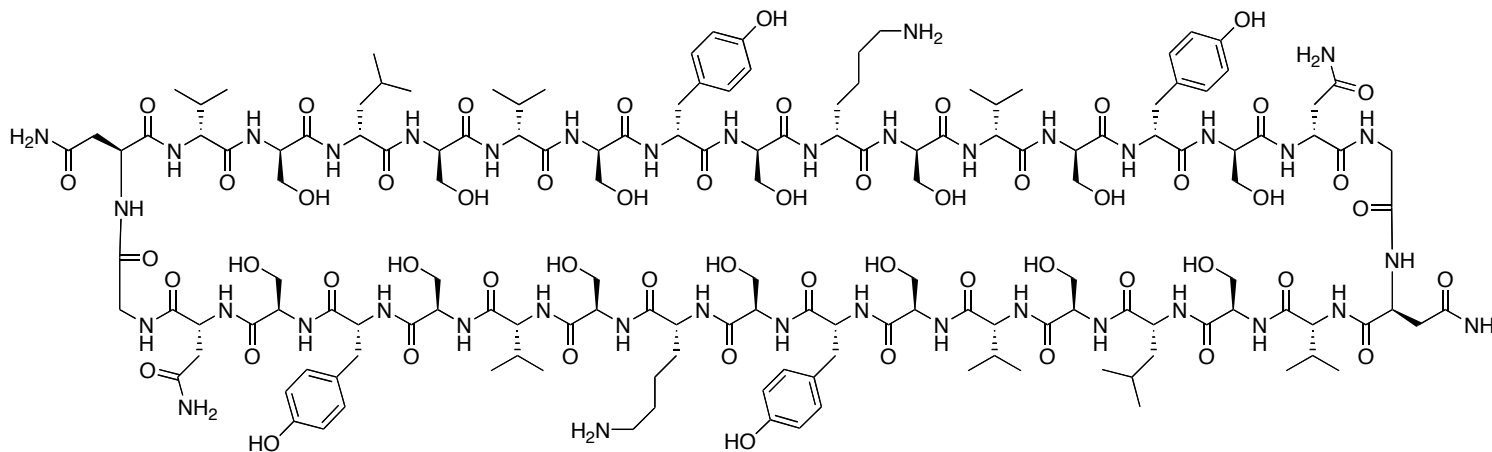
mass calculation:

	Sequence		
Enter 1 Letter Code			
	3-letter	**1-letter**	
residue 1	Asn	N	
residue 2	Gly	G	
residue 3	Asn	N	
residue 4	Val	V	
...	Ser	S	
...	Leu	L	
...	Ser	S	
	Val	V	
	Ser	S	
	Tyr	Y	
	Ser	S	
	Lys	K	
	Ser	S	
	Val	V	
	Ser	S	
	Tyr	Y	
	Ser	S	
	Asn	N	
	Gly	G	
	Asn	N	
	Val	V	
	Ser	S	
	Leu	L	
	Ser	S	
	Val	V	
	Ser	S	
	Tyr	Y	
	Ser	S	
	Lys	K	
	Ser	S	
	Val	V	
	Ser	S	
	Tyr	Y	
	Ser	S	

N-terminus H

C-terminus OH

Test 2K Peptide

minor product

Backbone Cyclic? Yes

Number of Disulphides 0

Number of Phosphates 0

Formula of Additional Group		
Mass of Additional Group	0	0
Include how many of this group?	0	
Is this group at the N-terminus?	No	
Number of H atoms Replaced	0	
Formula of Additional Group		
Mass of Additional Group	0	0
Include how many of this group?	0	
Number of H atoms Replaced	0	

Formula	C152H236N40O56
Amino Acid Composition	Asn4Gly2Leu2Lys2Ser14Tyr4Val6
Monoisotopic Mass	3517.68489
Average Mass	3519.78993

Electrospray		M<n+>	m/z mono	m/z average
		1	3518.692711	3520.797905
Charge Agent	H+	2	1759.850268	1760.902940
		3	1173.569454	1174.271286
Charge Range	1-5	4	880.429047	880.955458
		5	704.544802	704.965962

sample search result:

Amino Acids

Code	Include?	Min	Max
Ala, A	No	0	10
Arg, R	No	0	10
Asn, N	Yes	3	4
Asp, D	No	0	10
Cys, C	No	0	10
Gln, Q	No	0	10
Glu, E	No	0	10
Gly, G	Yes	1	2
His, H	No	0	10
Ile, I	No	0	10
Leu, L	Yes	0	2
Lys, K	Yes	1	2
Met, M	No	0	10
Orn, O	No	0	10
Phe, F	No	0	10
Pro, P	No	0	10
Ser, S	Yes	11	14
Thr, T	No	0	10
Trp, W	No	0	10
Tyr, Y	Yes	3	4
Val, V	Yes	4	6

Search Data

MALDI		ES	mono	average
mono	average	Ion 1		
	3343.4	Ion 2		
		Ion 3		
MALDI Ion	H+	Ion 4		
		Ion 5		

Results

MALDI		ES	mono	average
mono	average			
	3343.60104			

rms diff. 0.2010422
formula H-Asn3Gly2Leu2Lys2Ser12Tyr4Val5BuTFAAllyl-OH
M.W. mono 3340.5963 average 3342.5931

N-terminus H   C-terminus OH

Backbone Cyclic? No   Number of Disulphides 0

Possible Dehydration? Yes

Other Groups

ID	Include?	Min	Max
OH		0	1
Me		0	1
Et		0	1
Pr		0	1
Bu	X	0	2
For		0	1
Ac		0	1
TFA	X	0	2
Trt		0	1
Pbf		0	1
Pmc		0	1
Acm		0	1
Allyl	X	0	1
Benzyl		0	1
NH2		0	1
Acetonide		0	1
Boc		0	1
Fmoc		0	1
Alloc		0	1
Phosphate		0	1
Cl		0	1
Br		0	1
I		0	1

Additional Groups

ID	Mono Mass	Av. Mass	#H replaced	Include	Min	Max
Other1			1		0	2
Other2			1		0	2

This result <u>suggests</u> one possible solution for the identity of the product - an analogue of the parent peptide that is shorter by four residues (2 x Ser, 1 x Asn and 1 x Val) and has retained the allyl group (and is therefore linear). The peptide has also been modified by the addition of Bu and a trifluoroacyl group, possibly due to incomplete trapping of these species during the deprotection reaction.
Given that G+G=N, it is also possible that this product has arisen in part through inefficient coupling of Gly

during the synthesis.

There are, of course, other products with similar mass that are returned from repeated searches, but this example indiciates how suggestions for the identity of this byproduct may be generated *before* high-resolution spectra are obtained.