HX-Express2 User Guide

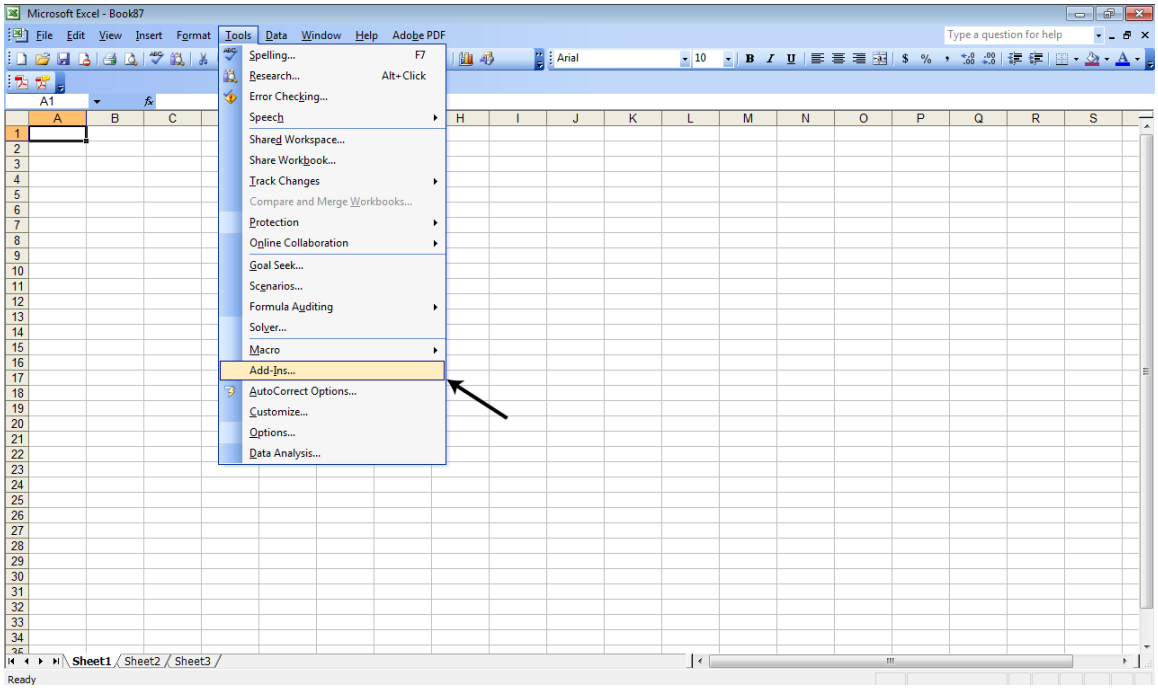
For questions or to report bugs contact

Miklos Guttman
University of Washington, Seattle
1959 NE Pacific St. Box 357610
Seattle WA, 98195
mguttman@uw.edu

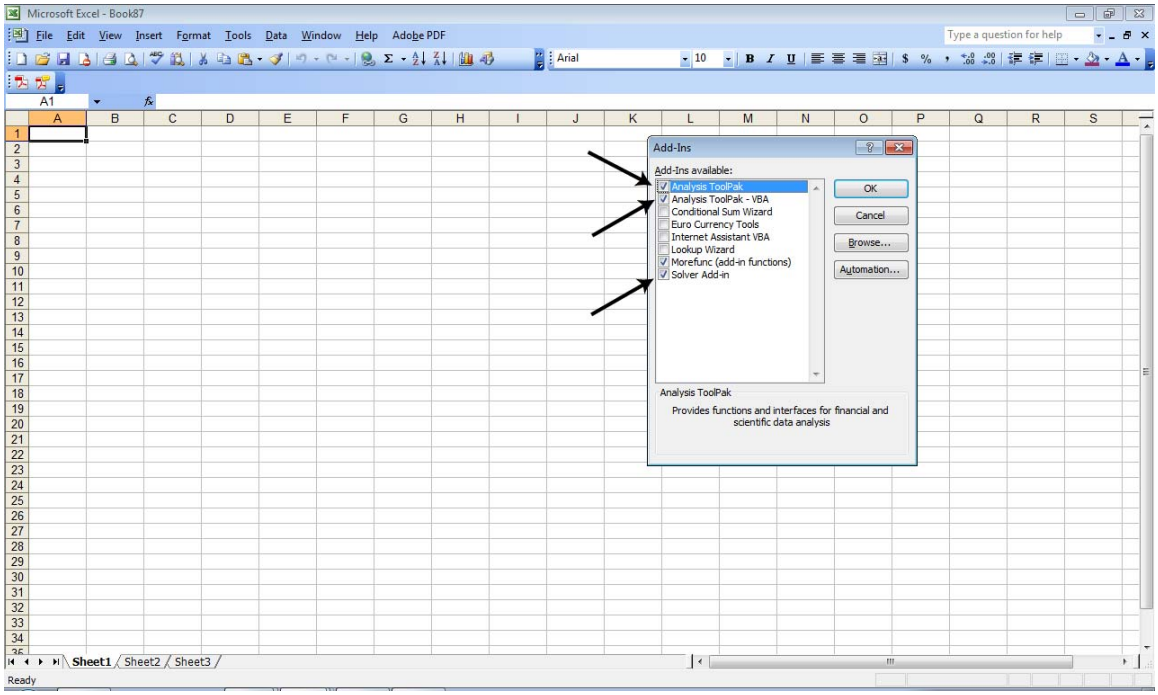
Contents

Installing the necessary add-ins; Excel 2003:

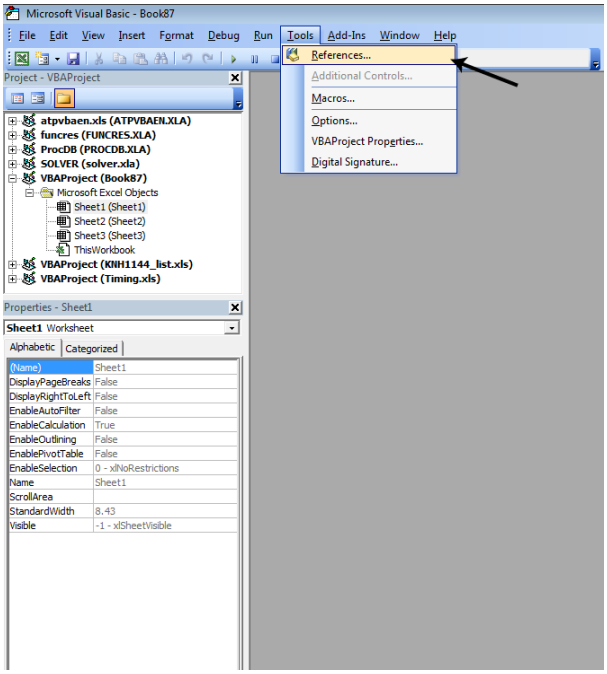
Within Excel, select the “add-ins…” option under the Tools menu bar.



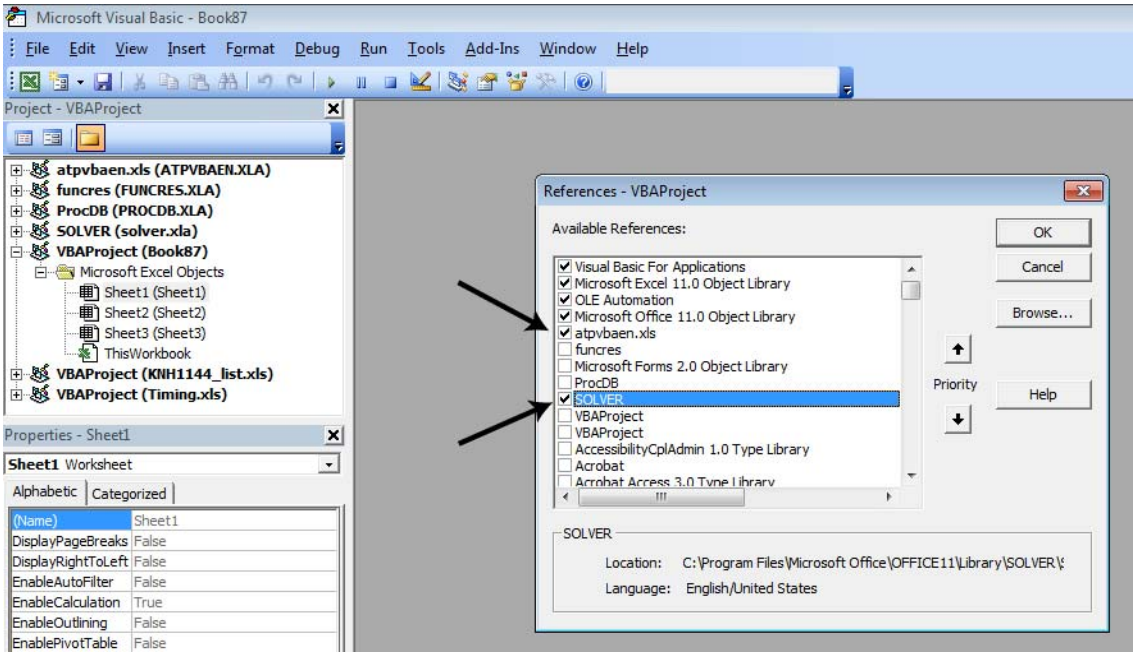
Check the boxes for “Analysis ToolPak”, “Analysis ToolPak – VBA”, and “Solver Add-in”, and click OK. In some cases the it may ask for the Microsoft Office CD to install the add-in.



Next add the references for these add-ins in visual basic. Go to the Visual basic editor (press Alt+F11). Select “References…” from the Tools menu bar.

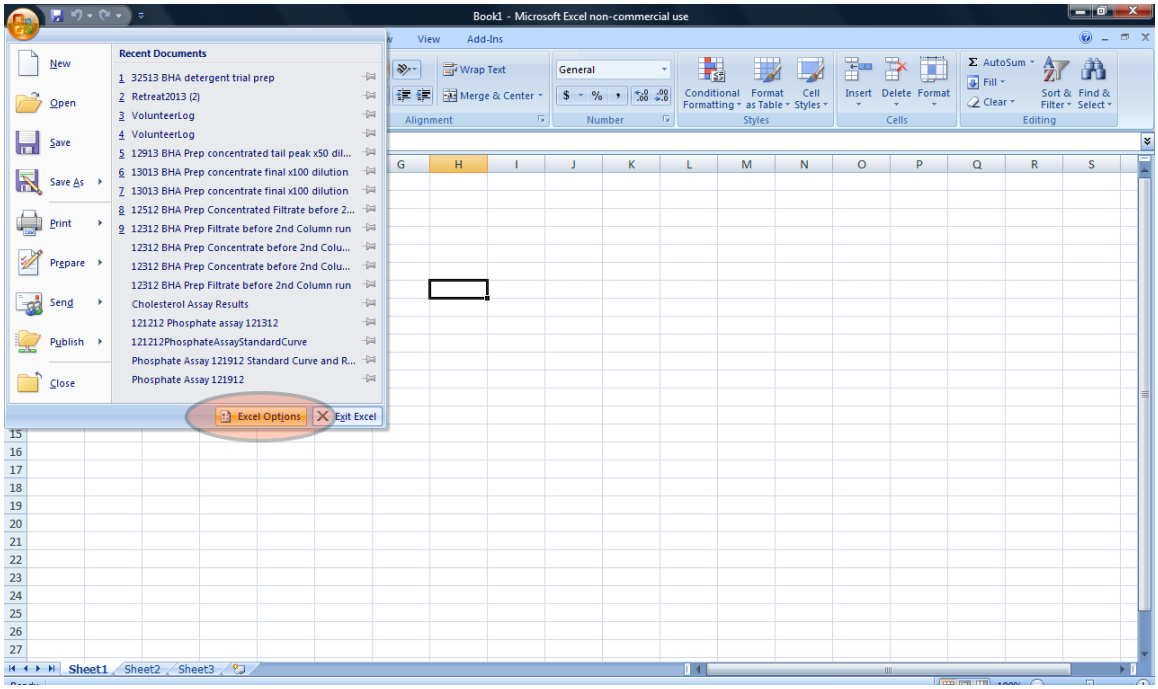


Check the boxes for “atpvbaen.xls” and “SOLVER”, and click OK. If they are not visible it may be necessary to activate them in Excel. This can be done by exiting Visual basic editor, clicking “Data Analysis…” and “Solver…” under the “Tools” menu bar, and clicking cancel when the respective windows pop up.

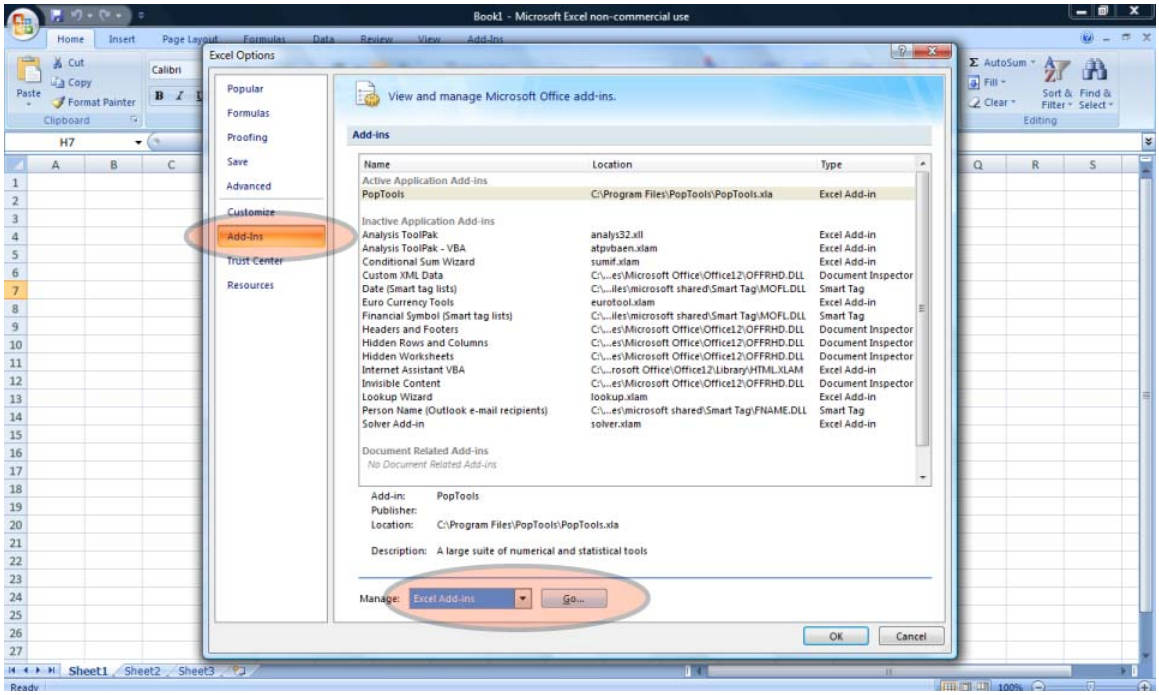


Installing the necessary add-ins; Excel 2007:

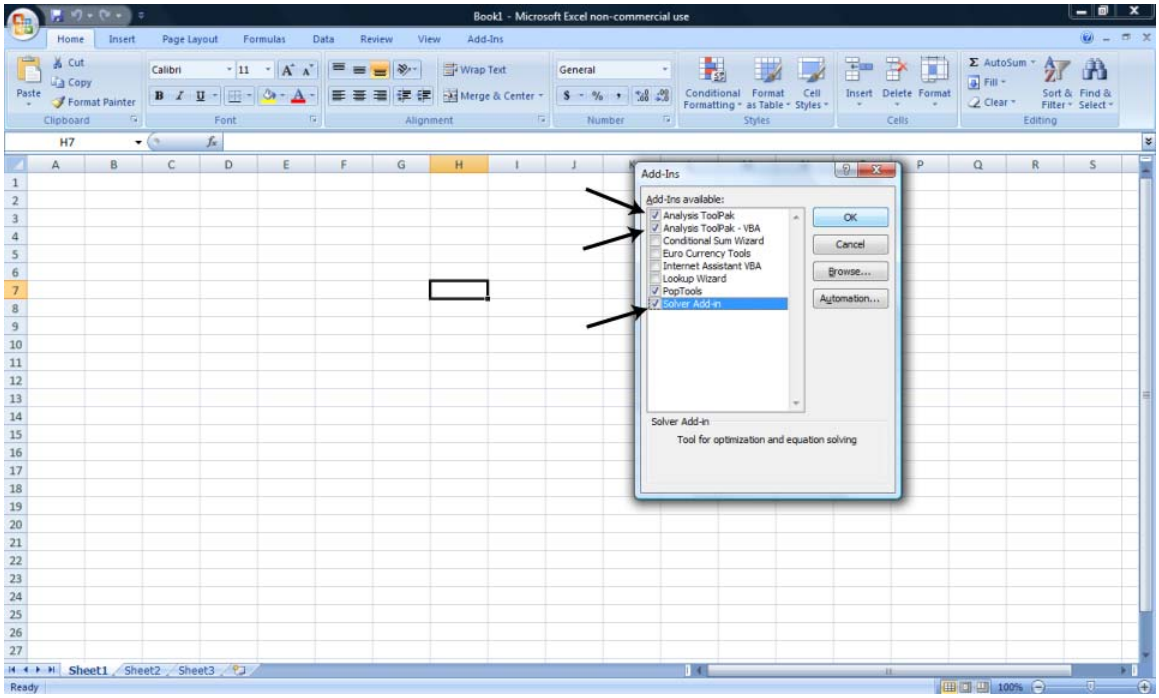
Under the main windows button select “Excel Options”



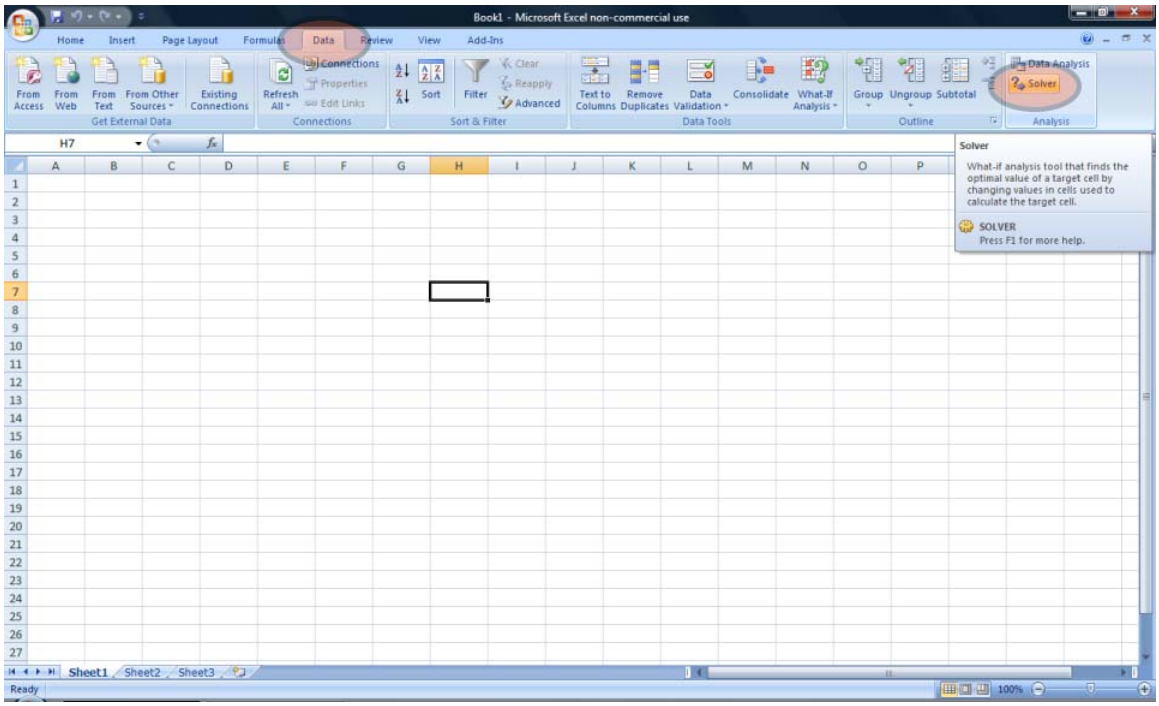
Then under the “Add-ins” options, select “Excel Add-ins” at the bottom and click Go…



Check the boxes for “Analysis ToolPak”, “Analysis ToolPak –VBA” and “Solver Add-in”, and click OK. The program might ask for the Microsoft Office installation CD.

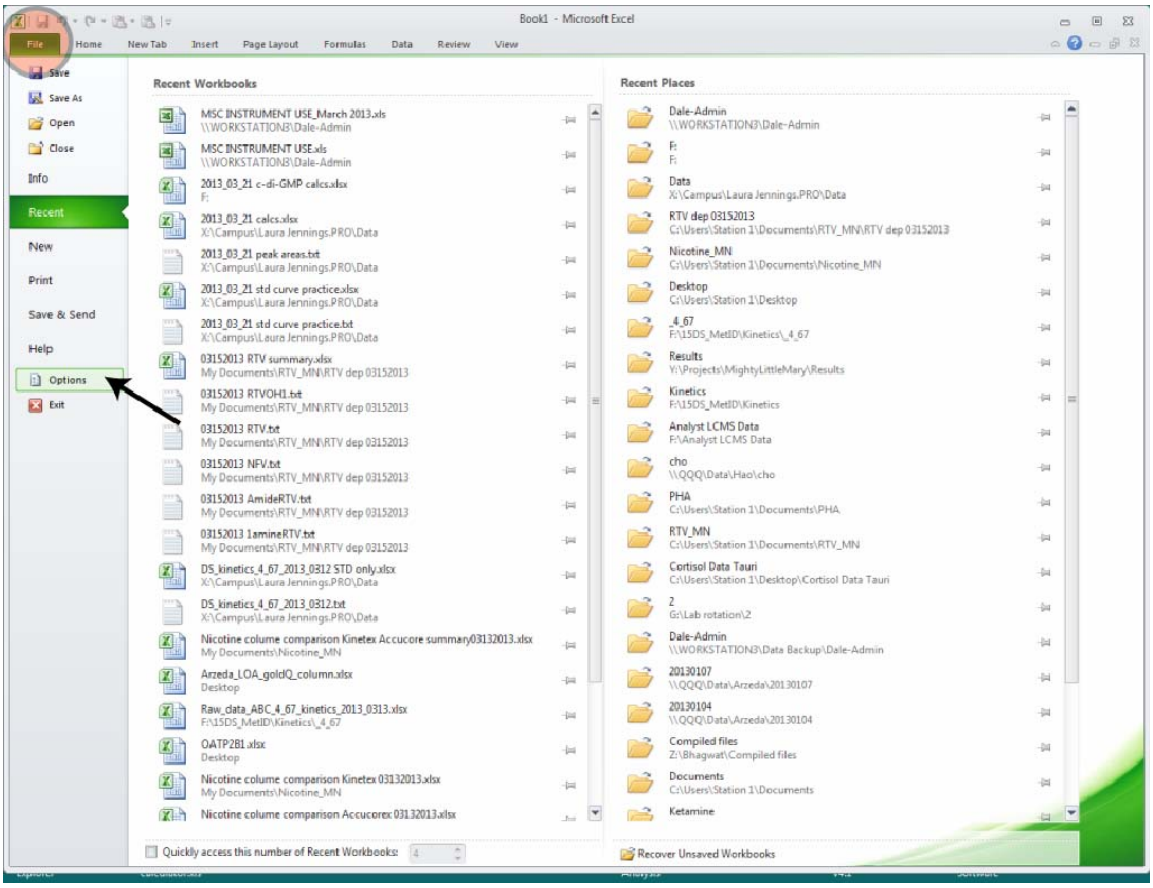


Then go to Visual Basic editor (Alt + F11) and add the references for the add-ins as shown for Excel 2003 above. It may be necessary to first activate each add-in within excel. The “Solver” and “Data Analysis” functions can be found under the “Data” tab.



Installing the necessary add-ins; Excel 2010:

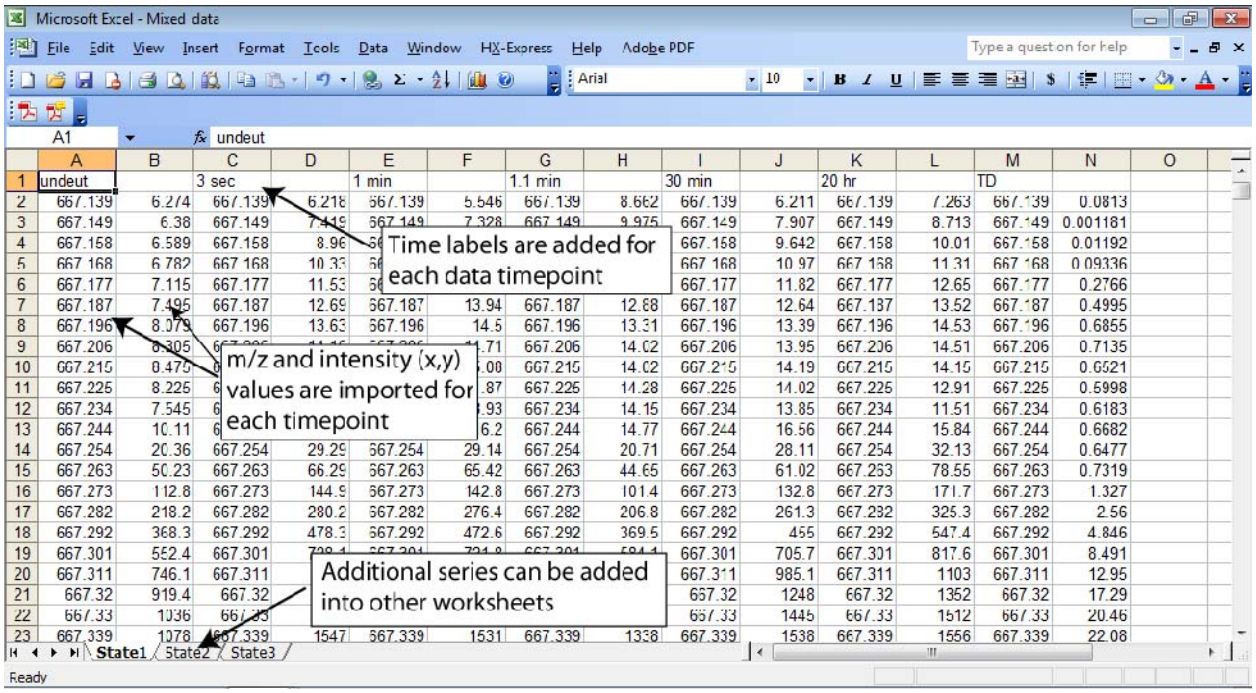
For Excel 2010 the installation is nearly identical to Excel 2007, the only difference is the first step. Under the “File” tab select “Options”, and follow the instructions for installation with Excel 2007.



Importing Data:

The first step is to import the spectral data for each deuteration time point into an excel worksheet. The x (m/z) and y (intensity) values are arranged in pairs of columns and the deuteration time is specified within the first row. Data for other states (ligand bound, mutants, etc) can be imported into a separate worksheet within the same workbook for parallel analysis. Data should be at least minimally smoothed (typically 4 x 2 Savitsky-Golay). If smoothing is not possible, there is a function to perform minimal smoothing in HX-Express (described on page 9).

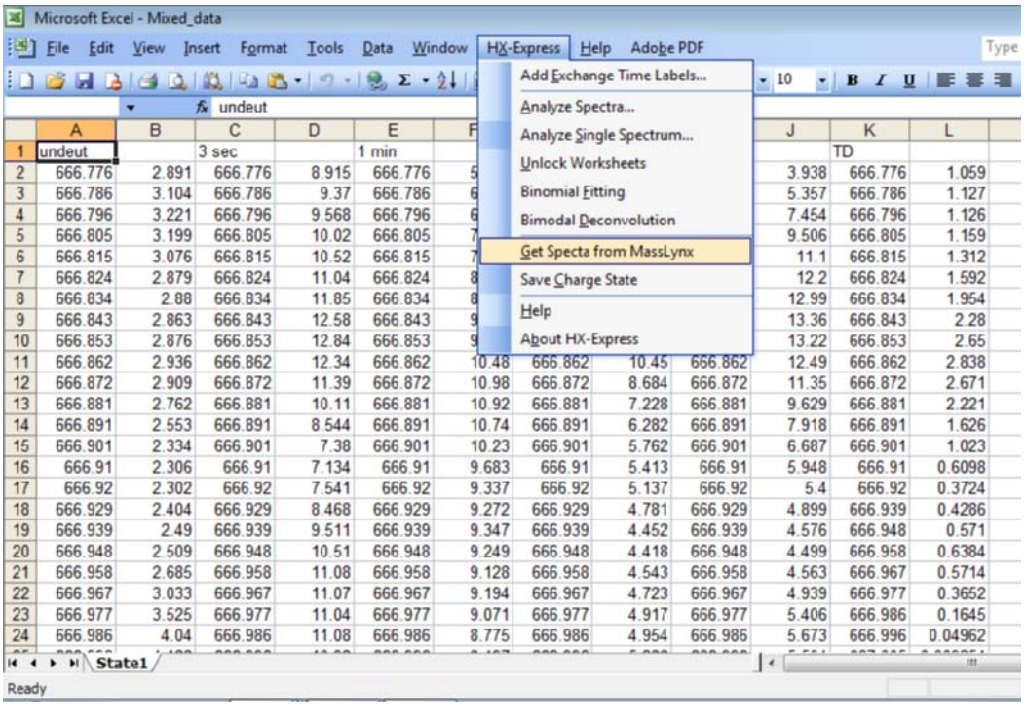
*Be sure that the data contains no redundant or repeating x values, this will cause errors during fitting. In MassLynx, setting the peak labels to 3 decimal places solves this issue.



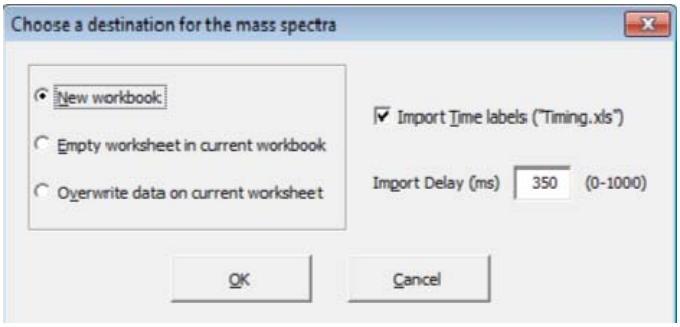
Time labels can be added using the “Add Exchange Time Labels…” option form the HX-express menu bar.

Each time point must be unique. Therefore the easiest way to analyze replicate experiments together is to place all the data into the worksheet and slightly offset the timing labels for the replicates (ie 1.01 min for the duplicate 1 min time point).

For Masslynx (Waters) users there is a function within HX-express for directly importing multiple spectra from Masslynx.



The target location for the data is specified within the next window. The Masslynx window with the spectra must be visible (not minimized) for HX-express to successfully import the data.

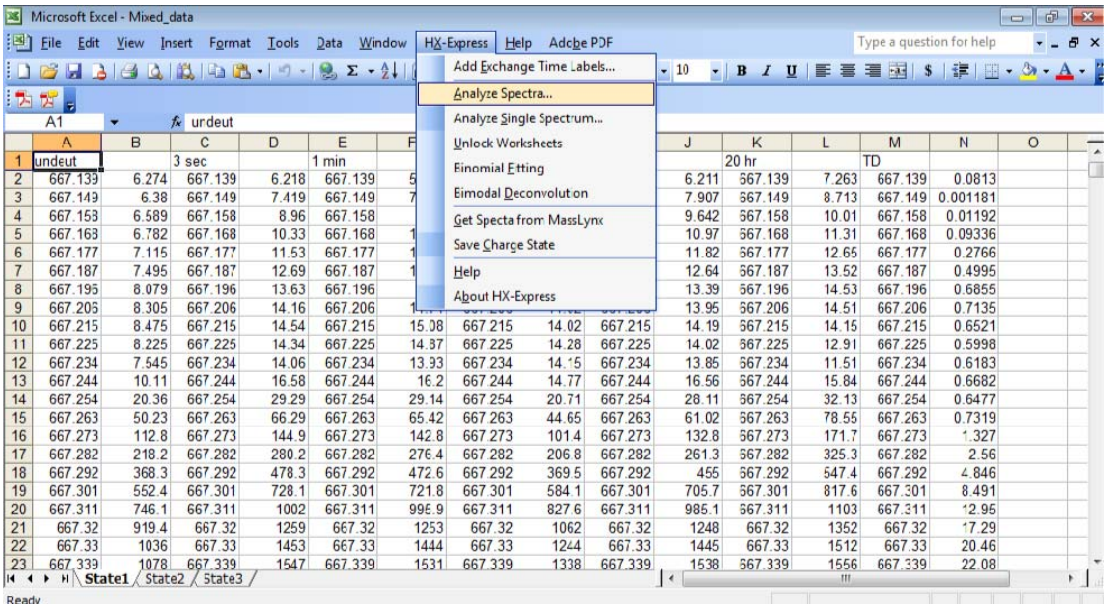


There is an optional checkbox for importing time labels from an existing workbook named Timing.xls (or Timing.xlsx). If this box is selected HX-express will look for the time labels from Timing.xls and import them to the current project. If the import time labels box is checked then the Timing.xls window must be open otherwise HX-express will generate an error.

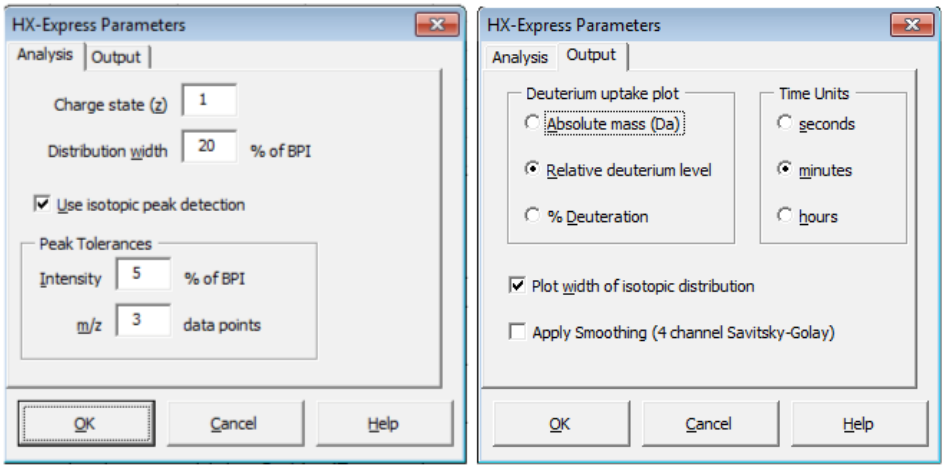
When importing data from Masslynx within Windows 7 it is necessary to use an import delay time of 350msec or more. If errors are generated during importing this value should be increased. In Windows xp the timing delay can be left at “0”.

Initial Analysis:

Once the data has been imported it ready for the first step of analysis. Click “Analyze Spectra” in the HX-express menu bar.



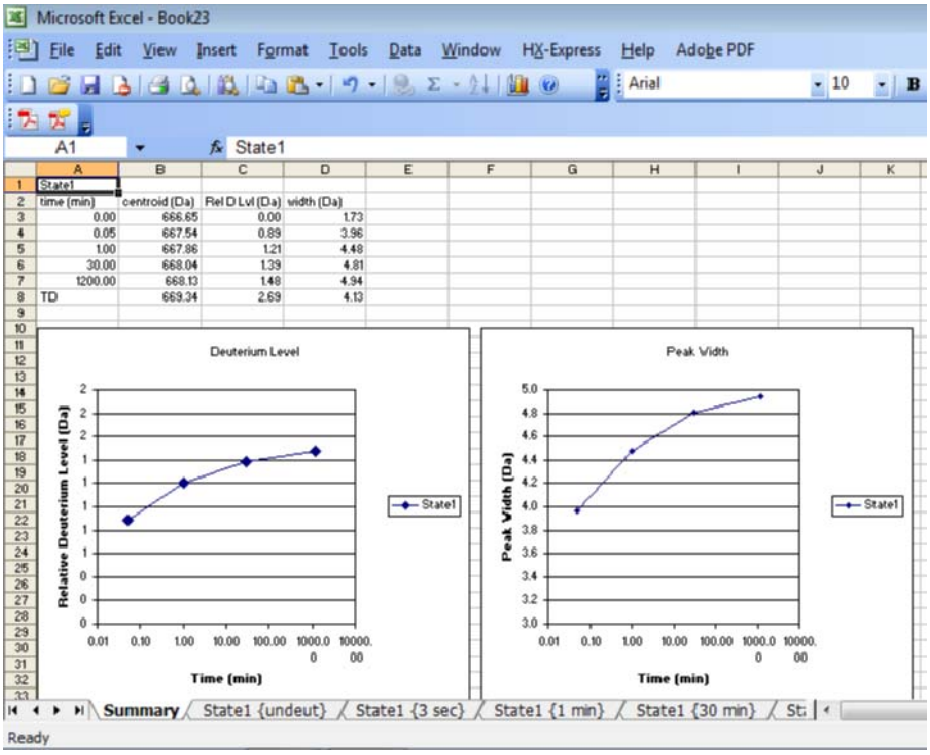
The charge state of the fragment of interest and distribution width must be specified by the user.



For isotopic peak detection (recommended in most cases) check the “use isotopic peak detection” checkbox. Values for intensity threshold and m/z tolerance will also need to be set for using this option. The Output tab contains options for the how to plot the final deuterium uptake plots, which units to use for the x axis, and whether to plot the width of the isotopic distribution (recommended). For downstream binomial fitting a low value for intensity threshold (5%) is recommended.

Minimal smoothing (4 channel Savitsky-Golay) will be applied to the data if the “Apply Smoothing” box is checked. This only needs to be used if the imported data was not previously smoothed.

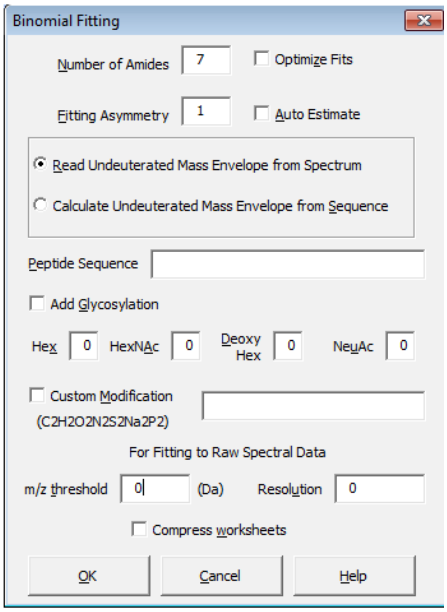
All spectra are then processed and the summary worksheet is generated.



The data is summarized in columns A-D and the two plots correspond to deuterium uptake (left) and the width of the isotopic distribution (right). These correspond to data obtained from centroid analysis.

Binomial Fitting:

The data can then be further processed using automated binomial fitting by clicking "Binomial Fitting" in the HX-express menu bar.

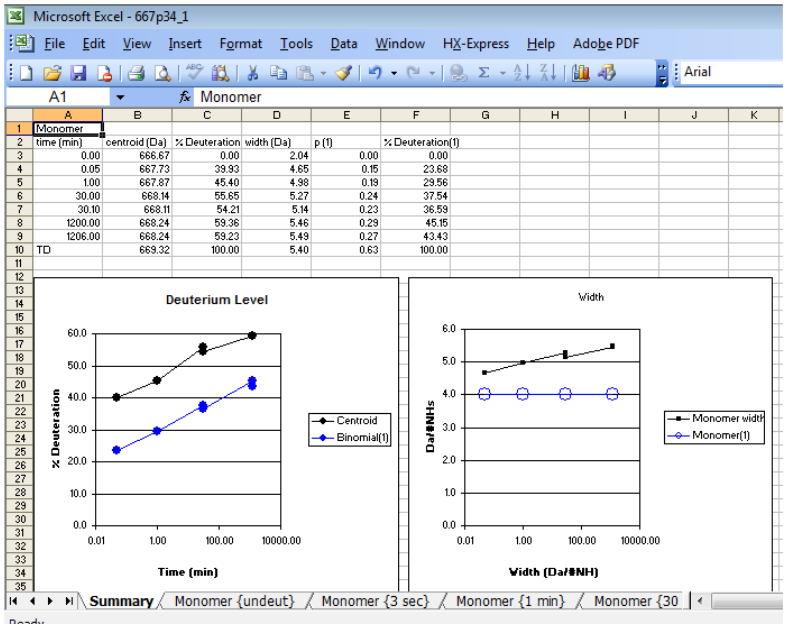


The binomial fitting algorithm requires the natural abundance isotopic distribution. If "Read Undeuterated Mass Envelope from Spectrum" is checked then the undeuterated data is used for this (faster option). This requires an value for the number of slow exchanging amides (most often this is the number of residues in the peptide minus the number of prolines and minus one for the N-terminal residue). Alternatively the user can provide the peptide sequence and check the "Calculate Undeuterated Mass Envelope from Sequence" for calculation of the undeuterated profile within Excel. For glycopeptides the "Add glycosylation" box should be checked and the number of hexose (Hex), N-acetyl hexose (HexNAc), DeoxyHexose (ie Fucose) (DeoxyHex), and Neuraminic acid (NeuAc) can be specified as integer values in the following boxes. If the peptide sequence is provided then the Number of Amides box can be set to "0" in which case the peptide sequence (and glycans) will be used to calculate this value.
= (Number of residues) - 1 – (Number of Prolines) + (Number of HexNAc).

Custom modifications can be added when providing the peptide sequence. The custom modification box should be checked, and the chemical formula of the modification entered into the field (C3H6O4N15). Having a minus sign as the first character ("-") will indicate a subtractive modification. Only C,H,O,N,S,P, and Na atoms are supported.

The fitting asymmetry term can be used to obtain better fits with noisy/overlapped data. Its use will be discussed in the section on "Improving fits with overlapped data". For clean (non-overlapped) data it should be left at "1". The fitting to raw spectral data parameters can be left blank for now, this is discussed in a later section.

The binomial fitting function will go through and fit the spectra in each worksheet with the specified parameters. The summary worksheet will now include the results from the binomial fits.

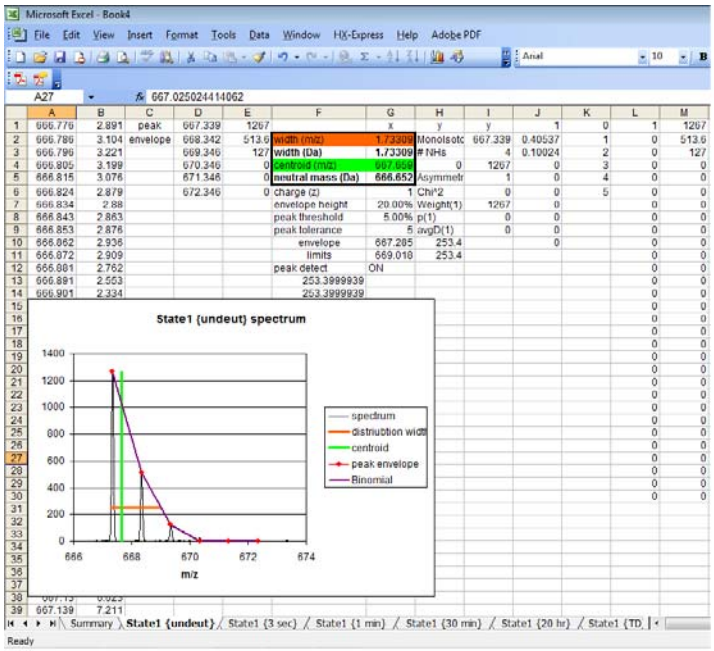


	A	B	C	D	E	F
1	Monomer					
2	time (min)	centroid (Da)	% Deuteration	width (Da)	p (1)	% Deuteration(1)
3	0.00	666.67	0.00	2.04	0.00	0.00
4	0.05	667.73	39.93	4.65	0.15	23.68
5	1.00	667.87	45.40	4.98	0.19	29.56
6	30.00	668.14	55.65	5.27	0.24	37.54
7	30.10	668.11	54.21	5.14	0.23	36.59
8	1200.00	668.24	59.36	5.46	0.29	45.15
9	1206.00	668.24	59.23	5.49	0.27	43.43
10	TD	669.32	100.00	5.40	0.63	100.00

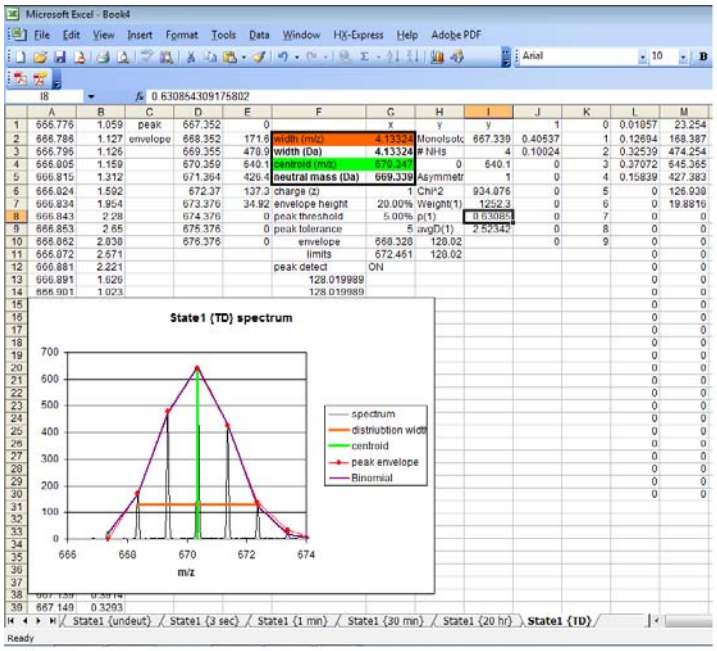
Columns A-D contain the same centroid data as before, and columns E and F now contain the labeling probability (*p*) and the average deuterium uptake (or percent deuteration, depending on what was selected for the deuterium uptake plots in the initial analysis). The uptake plot now shows the results from both centroid (black) and binomial analysis (blue).

The width plot on the right now also shows the #NHs at each time point in the binomial fit (blue circles). These values are also summarized in column N.

The fits should be inspected manually be scrolling through the worksheets.



Red dots indicate the detected isotopic peaks and the purple line depicts the resulting binomial fit. The labeling probability (*p*) is stored in cell “I8” and the intensity scale (weight) is stored in cell “I7”. The average deuterium incorporation (*p* times the number of amides) is stored in cell “I9”. The fitting routine automatically adds three extra points of zero (“zero padding”). This is to achieve a smoother binomial fit, as missing data points may distort the binomial distribution calculation.



If a fully deuterated sample is available it is a good check to make sure that the binomial fitting parameters are accurate. For this example the fit shows good agreement with the data (shown above).

In some cases the distribution width of the binomial function may be wider than the fully deuterated spectra. This is most commonly attributed to fast back-exchanging amides within certain peptides. In this case the binomial fitting should be re-run using a smaller value for the number of amides to achieve more accurate fits.

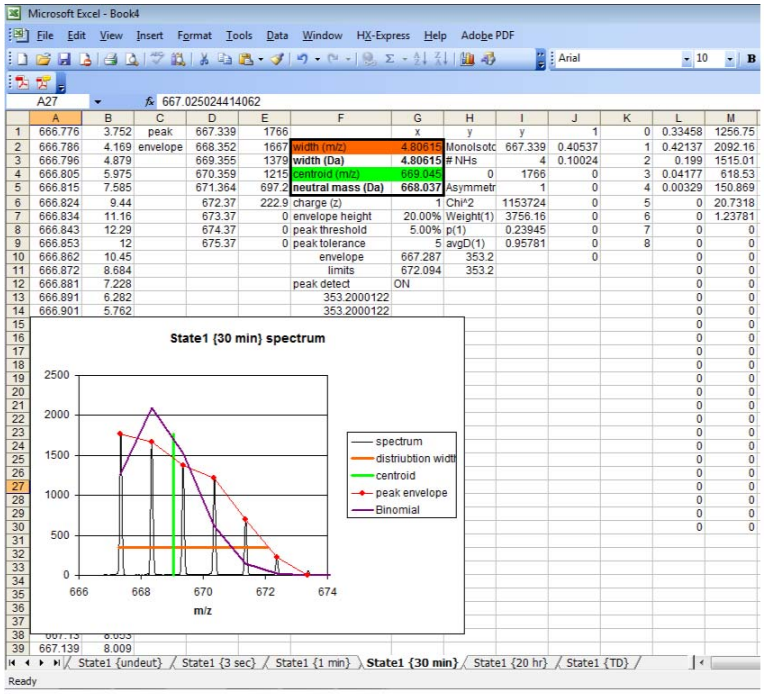
It is also possible that only a few time points for a given peptide have distributions that are narrower than that predicted from the binomial fit. This is most likely due to the presence of highly protected amides that have yet to exchange, thereby not contributing to the binomial distribution. For better fits the number of amides can be manually adjusted by changing the value in cell “I3” and re-running solver (“Tools” menu bar, “Solver” in Excel 2003 or “solver” in the “Data” ribbon in Excel 2007/2010).

If the “Optimize Fits” option within the original binomial fitting window is checked then HX-express will also vary the # NHs parameter to for obtaining the best fit.

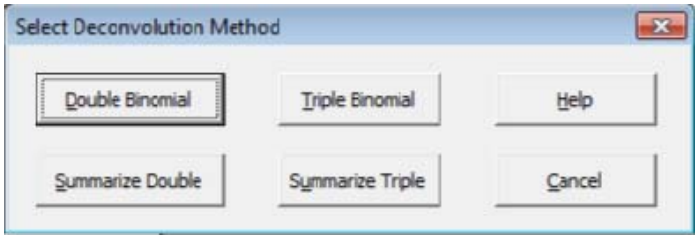
*Note: Different values for the number of amides will generate different probability (*p*) values. Therefore, if different number of amides are used for different time points, then the relative deuterium content (*p* times the number of amides) should be used for subsequent analysis, and not the % deuterium.

Bimodal Deconvolution:

The binomial function serves as a good indicator of spectra exhibiting deviations from simple exchange kinetics (data showing evidence of bimodal exchange patterns).

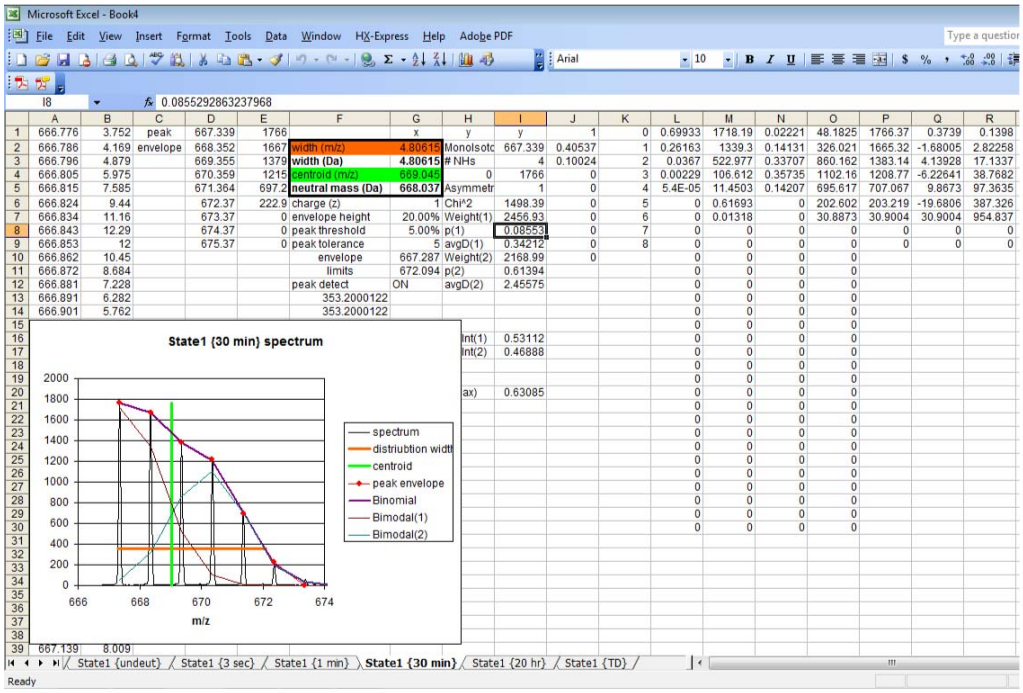


In this case the data clearly show deviation from the expected binomial distribution. The spectra can be deconvoluted by applying a double binomial fit. Click on the “Bimodal Deconvolution” option from the HX-express menu bar.



The Double Binomial button will fit two binomial functions to the data in the current worksheet. The same parameters used in initial binomial fitting will be used for the double binomial fitting. [Ctrl + Shift + D] is a keyboard shortcut that also triggers double binomial fitting.

The plot will now show the results from the bimodal deconvolution.



The individual binomial distributions are shown in brown and blue and their sum is shown in purple. The parameters for the second binomial function are stored in cells “I10” thru “I12”.

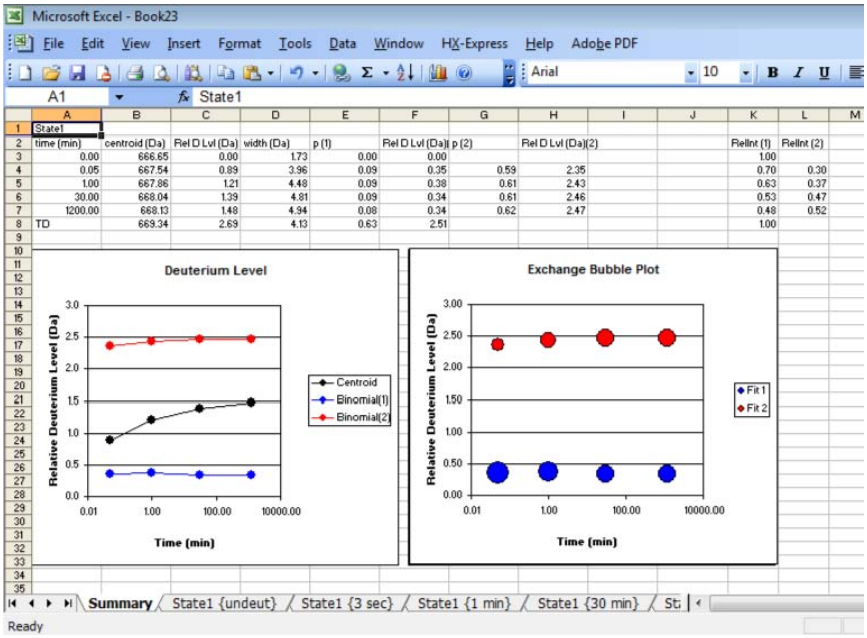
If a fully deuterated data set is included in the analysis then the labeling probability for each binomial component will be restricted so it that it does not exceed the deuteration measured for the fully deuterated sample.

In some cases the initial parameters for fitting need to be adjusted (cells in column “I” that specify each labeling probability (*p*) and intensity (weight). The worksheets can be unlocked using the “Unlock worksheets” from the HX-Express menu bar, after which the values can be adjusted and the fitting re-ran by initializing “solver” within the tools menu bar in Excel 2003 or the Data ribbon in Excel 2007/2010.

Sometimes better fits can be achieved by varying the number of amides used for one of the binomial components, usually if a highly protected species is present. The number of amides can be adjusted for each binomial component by changing the value in cells “I3” and “I22”, followed by re-running solver.

The bimodal deconvolution should be applied to all spectra showing significant deviation from the single binomial fit. Once all of those have been processed the deconvoluted data can be summarized by clicking “Bimodal Deconvolution” in the HX-express menu bar and selecting “Summarize Double”.

The summary worksheet now shows combined results from all of the analyses.



The results from the double binomial fitting are listed in columns E-H along with the relative intensities of the two binomials in columns K-L. The uptake plot now shows the original centroid fit along with both binomial fits.

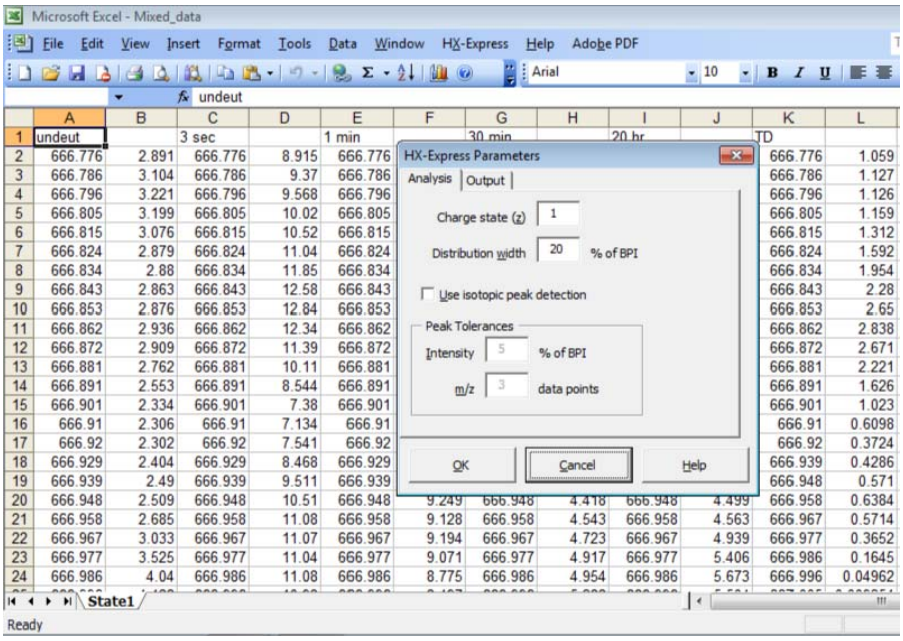
The additional bubble plot (right) is useful for interpreting the bimodal exchange profile. This plot shows the uptake profile for both binomial fits and the size of each point is scaled to its relative intensity at that time point. This is useful for differentiating between EX1 type exchange kinetics (where the relative bubble sizes shifts over time) and two distinct populations (where the relative intensities remain constant over time).

We stress that bimodal deconvolution should be interpreted with caution. Close visual inspection of the fits is required to ensure that the data is displaying a true bimodal character. This is especially true when the separation between the two binomial components is marginal. It should be kept in mind that other effects, such as carry over during LC-MS, can also lead to bimodal patterns in mass envelopes.

Bubble plots can only show data from a single sample series. For datasets with multiple conditions in the same workbook, individual bubble plots are generated for each series. Plot names will indicate which dataset each plot corresponds to.

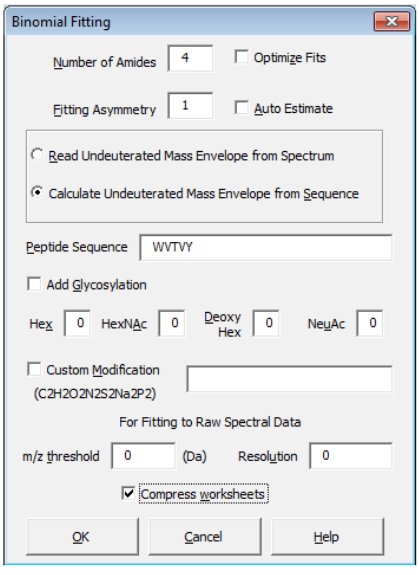
Fitting to raw spectra data:

After the data is imported to Excel, the initial analysis is done using no isotopic peak detection ("Use isotopic peak detection" is unchecked).



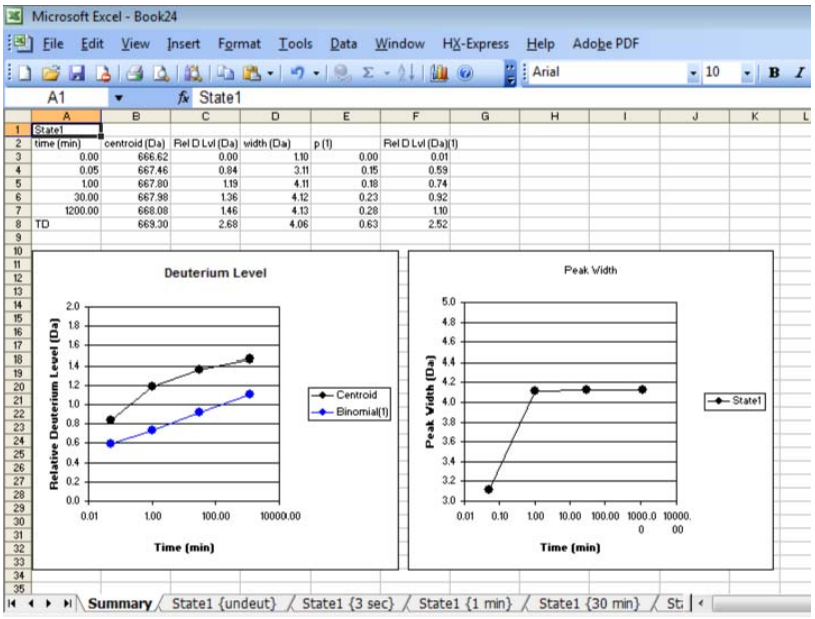
	A	B	C	D	E	F	G	H	I	J	K	L
1	undeut		3 sec		1 min		30 min		20 hr		TD	
2	666.776	2.891	666.776	8.915	666.776						666.776	1.059
3	666.786	3.104	666.786	9.37	666.786						666.786	1.127
4	666.796	3.221	666.796	9.568	666.796						666.796	1.126
5	666.805	3.199	666.805	10.02	666.805						666.805	1.159
6	666.815	3.076	666.815	10.52	666.815						666.815	1.312
7	666.824	2.879	666.824	11.04	666.824						666.824	1.592
8	666.834	2.88	666.834	11.85	666.834						666.834	1.954
9	666.843	2.863	666.843	12.58	666.843						666.843	2.28
10	666.853	2.876	666.853	12.84	666.853						666.853	2.65
11	666.862	2.936	666.862	12.34	666.862						666.862	2.838
12	666.872	2.909	666.872	11.39	666.872						666.872	2.671
13	666.881	2.762	666.881	10.11	666.881						666.881	2.221
14	666.891	2.553	666.891	8.544	666.891						666.891	1.626
15	666.901	2.334	666.901	7.38	666.901						666.901	1.023
16	666.91	2.306	666.91	7.134	666.91						666.91	0.6098
17	666.92	2.302	666.92	7.541	666.92						666.92	0.3724
18	666.929	2.404	666.929	8.468	666.929						666.939	0.4286
19	666.939	2.49	666.939	9.511	666.939						666.948	0.571
20	666.948	2.509	666.948	10.51	666.948	9.249	666.948	4.418	666.948	4.499	666.958	0.6384
21	666.958	2.685	666.958	11.08	666.958	9.128	666.958	4.543	666.958	4.563	666.967	0.5714
22	666.967	3.033	666.967	11.07	666.967	9.194	666.967	4.723	666.967	4.939	666.977	0.3652
23	666.977	3.525	666.977	11.04	666.977	9.071	666.977	4.917	666.977	5.406	666.986	0.1645
24	666.986	4.04	666.986	11.08	666.986	8.775	666.986	4.954	666.986	5.673	666.996	0.04962

Then, for binomial fitting, the undeuterated mass envelope can not be read from the spectrum. Therefore the "Calculate Undeuterated Mass Envelope from Sequence" box should be checked and the peptide sequence is entered in the peptide sequence box.

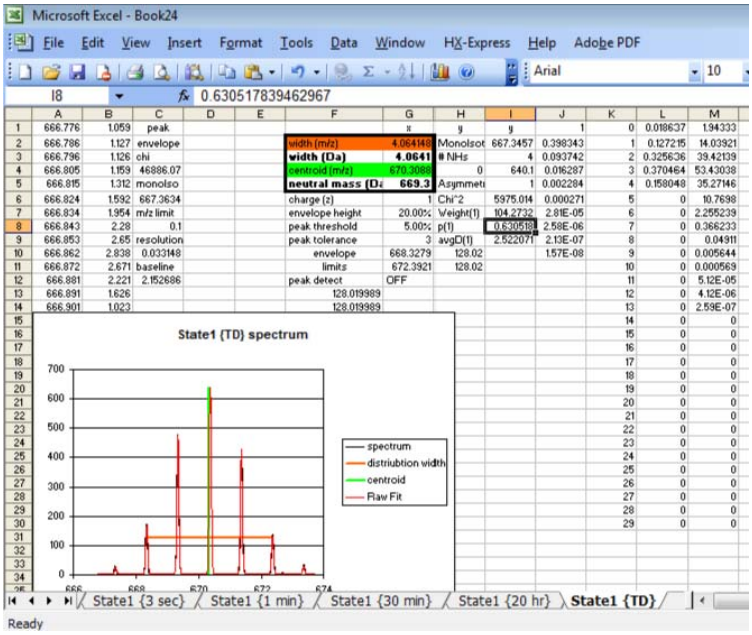


An m/z threshold and resolution is entered (based on instrument accuracy and resolution). As a rough guide, on a modern Q-TOF instrument, values of 0.1 and 10000 are appropriate. HX-Express reconstructs the theoretical spectra composed of Gaussian peaks for each isotope based on these values.

The summary worksheet will show the combined results from the centroid and binomial fitting.

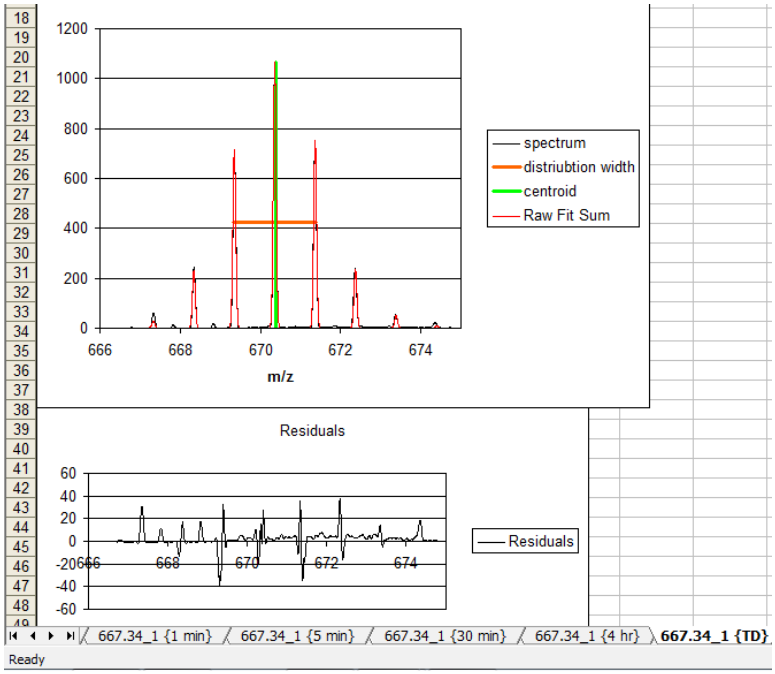


The fits can be manually checked by scrolling through the worksheets.



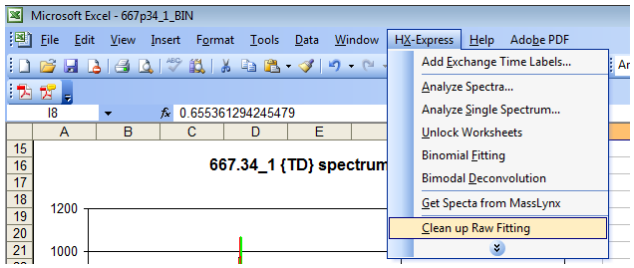
The resulting binomial fit (red) is overlaid with the data (black).

A residual plot for the raw spectral fit is shown below the main spectral plot, which can be useful for assessing the accuracy of the fit.



The fitting algorithm uses an initial guess at the labeling probability from the centroid value (in cell G4). For some noisier data this value may be inaccurate and the fitting may be misled. Better fits may be achieved by unlocking the sheets ("Unlock Worksheets" within the HX-express menu bar), and re-analyzing that particular spectrum. For this it will be necessary to adjust the *p* and weight values (cells I7 and I8) and re-run Solver. Sometimes the fitting gets caught in a local minimum during the initial trial, which can be overcome by offsetting the initial parameters slightly.

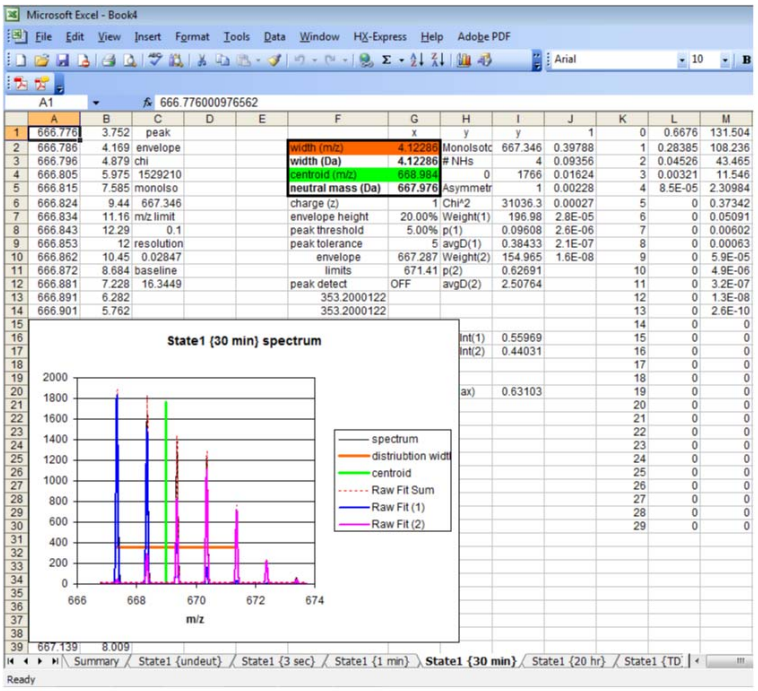
After all spectra have been verified or corrected, you can run the "Clean up Raw Fitting" option under the HX-Express menu bar. This will dramatically reduce the file size by clearing all the cells temporarily used for the fitting routine. It will also re-summarize the results and update any manual corrections made to any of the spectra.



If the "Compress worksheets" option was checked in the binomial fitting window, then the worksheets will automatically be cleaned up immediately after the fitting routine. However, with this option you will not be able to edit and re-run solver within each worksheet to correct the fits.

Bimodal deconvolution with raw spectral data:

The bimodal deconvolution can also be invoked the same way after binomial fitting to the raw spectral data.

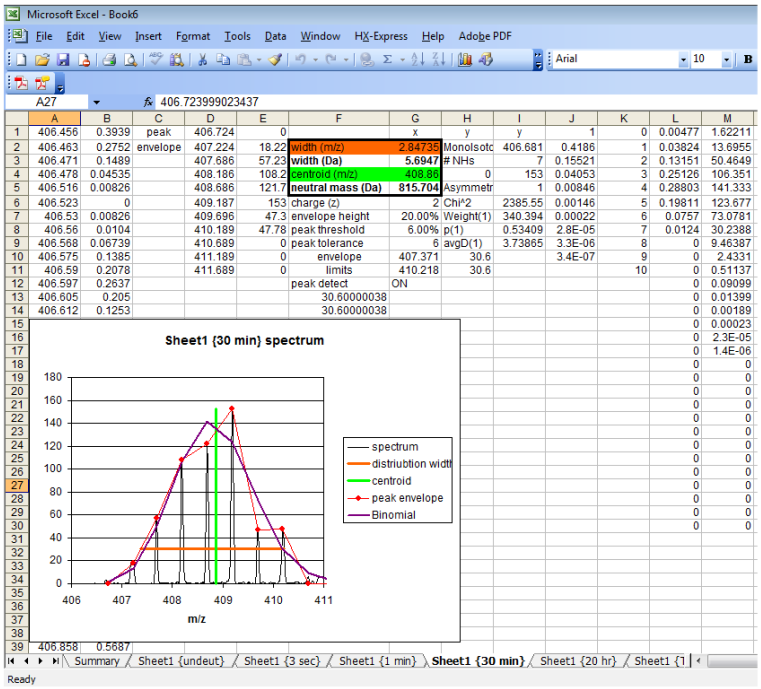


The two species in the mass envelope are shown along with their sum (dotted red line). In some cases the routine may get trapped in a local minima, and not find a reasonable answer. Better fits might be achieved by making manual adjustments to the weight terms (cells I7 and I10) along with the probability terms (cells I8 and I11), and re-running solver.

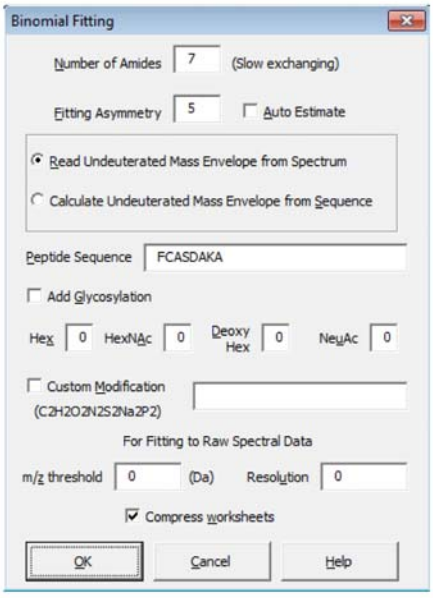
*Warning: This fitting routine is more computationally intensive and is roughly 10 times slower compared to bimodal deconvolution using isotopically picked peaks.

Improving fits with overlapped data: Asymmetric linear squares regression

In some cases data will contain overlapping peaks that contaminate the mass envelope.

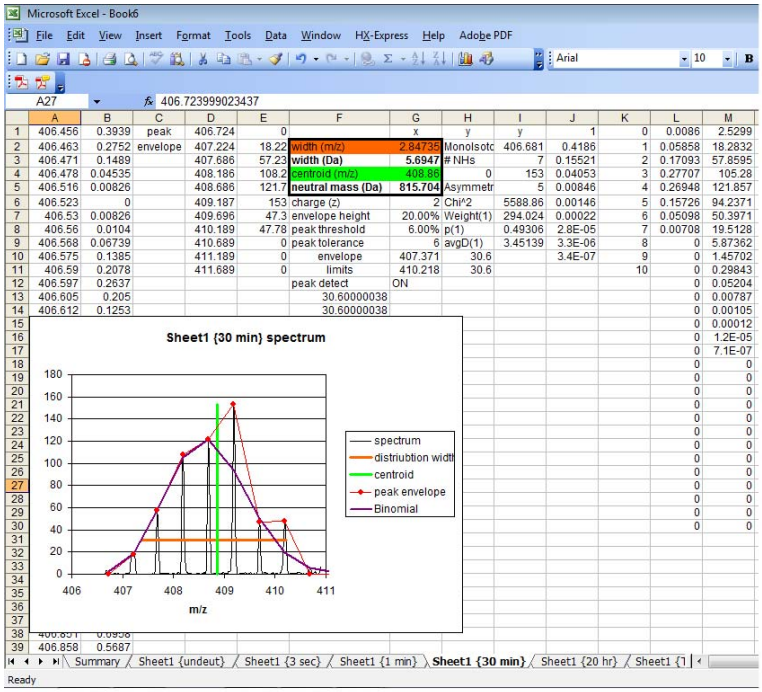


In this case there is a 1+ ion overlapped with the 2+ ion of interest. These contaminated peaks will offset the labeling probability obtained from binomial fitting. For these cases the “Fitting asymmetry” term can be increased (typically around 2-10) depending on the severity of the overlap.



This causes data with intensity values above the theoretical fit to contribute less to the total fit than data with intensity values below the theoretical fit. In this case a value of 5 is applied for achieving better fits.

The resulting fit is now more consistent with the true mass envelope of the 2+ ion.



Generally for this approach to work the number of contaminated peaks should not exceed the number of “clean” peaks. It is therefore effective at fitting a higher charged state mass envelope with lower charge state contaminants, but not the opposite. The asymmetric fitting can also be used in conjunction with raw spectra fitting, but typically, larger values (5-50) are needed to have significant effects for alleviating overlap.

The “Auto Estimate” box can be used to test different values of the asymmetry term. This runs iterations of the fitting increasing the asymmetry term and comparing the resulting *p* values until the *p* value is within 0.01 of the last iteration. The Fitting Asymmetry box can be left empty when this is checked as the calculations will always start with asymmetry = 1. This is relatively fast for isotopically picked peaks, but can take significantly longer when using fitting to raw spectral data.

* Selecting both the “Auto Estimate” and “Optimize Fits” options is not recommended as it can sometimes result in strange fits.

Release Notes for HX-Express2

This software has been tested in Excel 2013, 2010, 2007, and 2003 within Windows 7 and Windows XP. It has not yet been tested in Windows 8.

New Features in version 2:

- Import delay function to wait for the clipboard to update when pulling in data directly from Masslynx. The delay will depend on the system performance and architecture, but 350 msec seems to work well within Windows 7. This can be left zero when working in WinXP.

- Added the option of automatically importing time labels from an open workbook named "Timing" to save time when importing many sets of spectra.

- Added binomial fitting as an alternative method to centroiding for measuring the mass shift. This algorithm requires the "solver" add-in within Excel. It also requires the mass envelope of the undeuterated peptide which can either be read from the undeuterated spectrum, or calculated theoretically from the peptide sequence. The later requires the Analysis TookPak add-in within Excel. The calculation has been written to handle common glycan modifications and custom modifications.

- An asymmetric weighting factor can be used for binomial fitting to obtain more accurate fits for overlapped spectra. This results in larger errors for data points below the theoretical fit, thereby minimizing the effect of contaminant peaks within a mass envelope.

- Binomial distributions can also be fit using the raw spectral data. It is much slower, but can be useful for weaker fragments, where isotope detection is ineffective.

- Double and triple binomial fitting can be applied to any single spectra to deconvolute bimodal/trimodal spectra.

- A summary macro is also available for visualizing the results with the aid of a bubble plot.

Known issues:

- Automatic calculation must be enabled within Excel for the routines to function.
  In Excel 2003 this is under Tools - Options - Calculations Tab.
  In Excel 2007+ this is under the main excel window - Excel Options

- Solver needs to be open to activate before HX-Express2 modules will read it in for binomial fitting. For some systems it will need to be activated every time excel is started before HX-express will recognize it. For this simply call the solver function (Tools - Solver) and cancel out of it.

- Some functions including the Masslynx import tool will not work in 64 bit versions of Microsoft Excel.

- For importing spectra from Masslynx the import delay needs to be used in Windows 7. If this delay is too short the data won't be read and an error will be generated. In rare cases data may still be imported, but it will contain duplicates of a single spectra. A slight increase in the import delay time appears to resolve this issue.

- The import spectra from Masslynx function will also reset the CapsLock and NumLock keys.

- Changing the number of amides used in binomial fitting alters the calculated probability, therefore the plotting by “percent deuteration” should only be used if the same number of amides were used for fitting all spectra within a series. If not, then the “relative deuterium level” option should be used for visualizing the data.

- For very large peptides (>4000Da) the distribution may become too wide for effective binomial fitting. The easiest solution is to make the peak threshold slightly higher to cut down the number of picked peaks below 30.

- In some versions of Excel 2007 the prediction of the natural isotopic distribution from the sequence might fail. This is usually because an earlier version of the Analysis ToolPak add-in is installed. You can check within visual basic editor whether the ATPVBAEN has a ".XLA" or ".XLAM" suffix. The ".XLA" is the previous Excel 2003 version.

- If many excel worksheets are open the binomial calculations may become sluggish. For best performance, use the “Clean up Raw Fitting” before re-processing, or have the “Compress worksheets” option enabled.

- Using the “Optimize Fits” function for binomial fitting to raw spectral data will add considerable computation time and is not recommended.