SAM SAGE V1.2 MANUAL

1. Purpose

The Sample Analysis at Mars (SAM) uses a quadrupole mass spectrometer (QMS) to detect molecules during evolved gas analysis (EGA) and gas chromatography mass spectrometry (GCMS) experiments. The EGA and GCMS data from the QMS is often undersampled (that is, there are not many points in each peak), so the area under the peak is inaccurate, and the data is sampled at irregular time intervals, so it cannot be analyzed with commercial GCMS software (i.e., ACD, MassHunter, Excalibur) directly. **S**oftware for the **A**nalysis of **G**CMS and **E**GA, or SAGE, was developed to solve this problem.

The SAGE Igor tool will allow the user to fit curves to the peaks in the data, fit the "quiet"(that is, non-peak) segments of the data realistically, and then resample the data at a shorter time interval so that peaks are defined by many points and so that the data is aligned such that there is a full mass spectra available at every time step.

2. Setup and Data Preparation

If you don't have Igor Pro, install it from Wavemetrics. Versions 6.2 and 6.3 should both work. SAGE was primarily developed in versions 6.22A and 6.30B of Igor Pro.

2.1. **Required software for preparing input data for SAGE.** If you plan on processing tm.sam files you need to install the Python tools (`699util`), XINA, and SAM Data View, all of which can be found on the SAM Software page on the 699/SAM wiki:

`http://23.21.243.36/wiki/index.php/SAM_Software`

The "Individual Software Pages"section links to download/install instructions for each of these software tools, and also links to the page for the SAGE tool.

The `gcms.py` script requires files installed with SAM Data View, so that should be installed as well even though you wont need to use it directly.

If you cant install any of this software for some reason, all you need in order to open a TID in SAGE is either the IGORdata.zip file for that TID or five text files (stored in the same folder) that were generated by the `699util` Python scripts (details in the Data Preparation section):

`gcms.txt`
`gcms_b.txt`
`tmmarker.txt`

```
gcsci.txt
tmmsg.txt
```

The tmmsg.txt file is optional but useful.

2.2. **Software compatible with SAGE output.** SAGE can generate mass spectra files (.msp) and chromatogram files (.jdx). The .msp files are NIST-format text files that can be opened in the MS Search program of the NIST 11 software. Earlier versions of NIST are not guaranteed to be able to open our output files.

The .jdx file is a JCAMP-format text files that can be opened with the MS Manager program of ACD software. This file will consist of the selected user-generated peak fits ("selected"means checked in the peak list in the Main Graph), and automatically generated filler data in the "quiet"regions between the fitted peaks. Currently the filler data is a line connecting the last point of one fitted region to the first point of the next fitted region. The entire dataset is zeroed to the start of the GCMS run (specifically, to the time of the `GC_IT_FLASH` TM marker), goes until the end of the GCMS run (the time of the `GC_RUN END` TM marker), and is sampled at a rate of 200 milliseconds. This sampling time step will be reduced to 20ms in a future version of the SAGE tool, because that time step is more SAM-relevant.

2.3. **Data preparation.** First, download the desired TID from XINA (checkout all).

If there is an IGORdata.zip file in XINA, all the user has to do is unzip this file into the TID folder.

If there is no IGORdata.zip file, the user has to extract data from the tm.sam using the `699util` Python scripts. To generate the SAM data for use with Igor, navigate in the terminal (Mac) or command line (PC) into the folder where the tm.sam is located and then run the following commands:

```
gcms.py > gcms.txt
gcms.py -b > gcms_b.txt
tmmarker.py > tmmarker.txt
gcsci.py > gcsci.txt
tmmsg.py > tmmsg.txt
```

The tmmsg.txt file is optional but useful.

3. Opening SAGE

An experiment in Igor is an environment which contains data files, procedure files (code), and any graphs or tables that user creates. SAGE is an Igor experiment file.

Double-click the icon for the latest version of SAGE to open the experiment file, or open the experiment from within Igor Pro (File > Open Experiment...).

4. Loading SAM Data into SAGE

Under the Macro menu select "Load TID Folder" (see Fig. 1). Navigate into the folder of the TID you want to view, and press the "Choose" button.

Because of the way SAGE reads in information, the TID folder name needs to have the following format: YYYY-MM-DD-hh.mm.ss-tid#-description

This is what the folder will be named if it is downloaded directly from XINA.

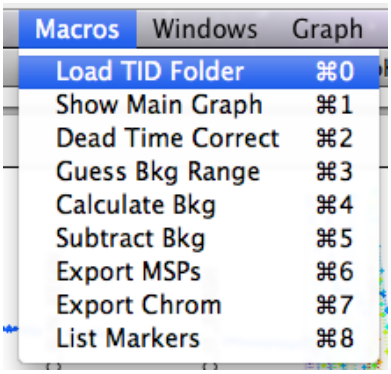


Figure 1. Select Load TID from the Macros menu to import SAM data into Igor.

5. Main Graph

The Main Graph (or Main View), as shown in Fig. 2 is the window where the corrections (deadtime correction and background subtraction) will be applied, and where the peak regions will be selected.

The Main Graph opens automatically when a TID is loaded. If the Main Graph window gets closed, it can be reopened from the Macro menu (Show Main Graph).

NOTE: If you want to close the Main Graph or the Peak Graph, you should choose the No Save option. Any other graph in Igor you can choose the Save option if you want to be able to regenerate the plot, but the Main Graph and the Peak Graph don't need to be saved in order to be regenerated after you close them (and they won't regenerate in the expected way if you try to use Igor's built-in graph macros to reload them).

WAY MORE IMPORTANT NOTE: If you want to close something (i.e., procedure files, notebooks) and Igor gives you options to Kill or Hide the file, ALWAYS choose Hide. Kill will irreversibly delete the file, whereas Hide will just close it.

5.1. Displaying TM markers.

To show or hide TM markers in the Main Graph, check or uncheck the Show Markers checkbox (Fig. 2d) in the top left corner of the Main Graph. Clicking this checkbox will open a window containing a list of the TM markers (Fig. 3). Only the TM markers which are checked in the TM marker window will be displayed when Show Markers is selected (Fig. 4).

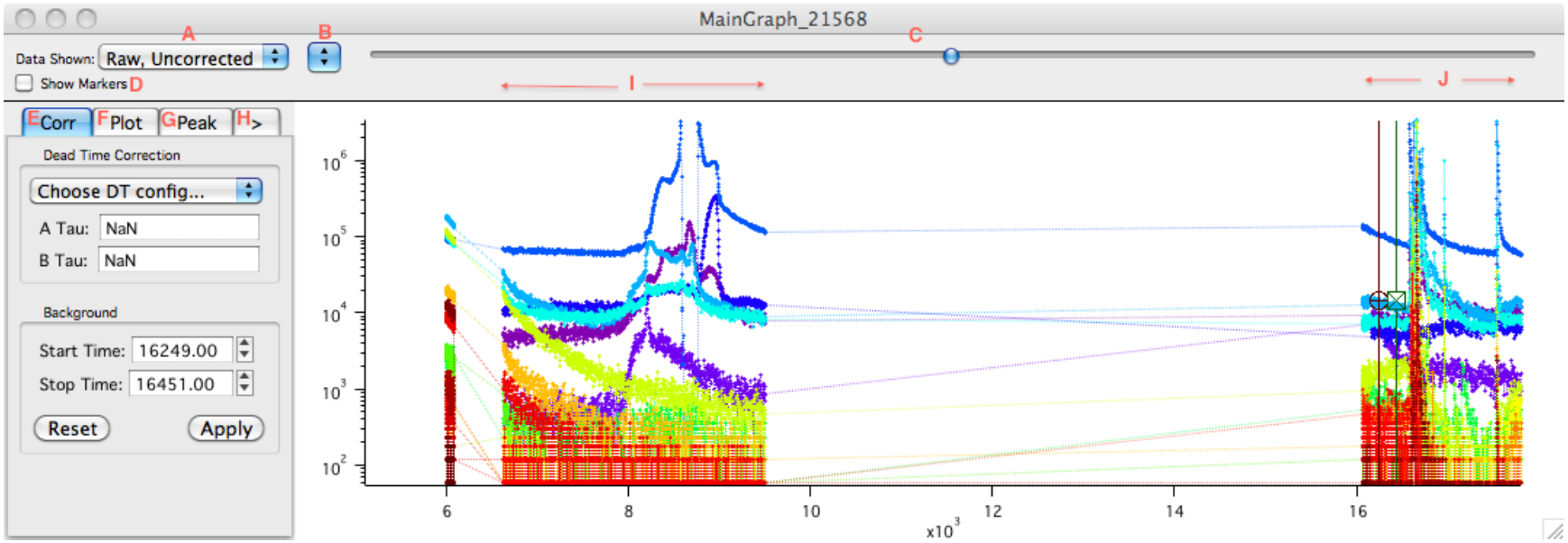


Figure 2. Example of the initial appearance of the Main Graph. **(a)** Data selection dropdown menu. **(b)** Axis slider options. **(c)** Axis slider. **(d)** Show Markers checkbox. **(e)** Correction tab. **(f)** Plot tab. **(g)** Peak tab. **(h)** Export tab. **(i)** EGA data (known from TM markers). **(j)** GCMS data (known from TM markers).

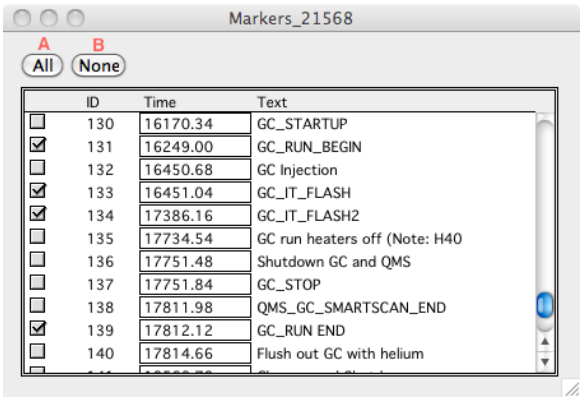


Figure 3. The TM markers window lists the ID number, time, and name of each TM marker for the current TID. The checkboxes hide or display the markers in the Main Graph. **(a)** "All" button plots all TM markers. **(b)** "None" button hides all TM markers.

There are two buttons in the TM markers window: one which allows you to quickly display all TM markers (All button, Fig. 3a), and one which unchecks all TM markers (None button, Fig. 3b).

5.2. **Displaying EGA temperature.** In order to display EGA temperature, you need to have a text file containing the temperature vs time, which you can then load into Igor by using Overlay Pyro Temp in the Macros menu. You may need to change the file type to All Files in the file opening dialog that pops up.

5.3. **Deadtime correction – Correction (Corr) tab.** Before doing anything else, the user will have to specify the deadtime correction parameters. The deadtime correction parameter dropdown menu (Fig. 5a) allows the user to select which detector electron multiplier (EM) was used for this TID so that the deadtime correction can be properly applied.

Getting these coefficients exactly right is important. The deadtime correction coefficients are different between the SAM flight models EM1 and EM2, and the SAM flight models coefficients are different from the SAM test beds. In addition, the deadtime correction coefficients change over time for any given detector, so at some point that will have to be dealt with.

Currently the default options provided by SAGE are the coefficients for the flight model's EM1 and EM2, and the testbed's EM2. All of these values were calculated by Heather Franz.

If the user knows that the coefficients are different from the choices provided by SAGE, they must enter these manually into the text fields for the deadtime correction coefficients (Fig. 5b). If the user wants to specify different values, they must choose the "Custom" option in the deadtime correction parameters dropdown menu, and enter values into the text fields.

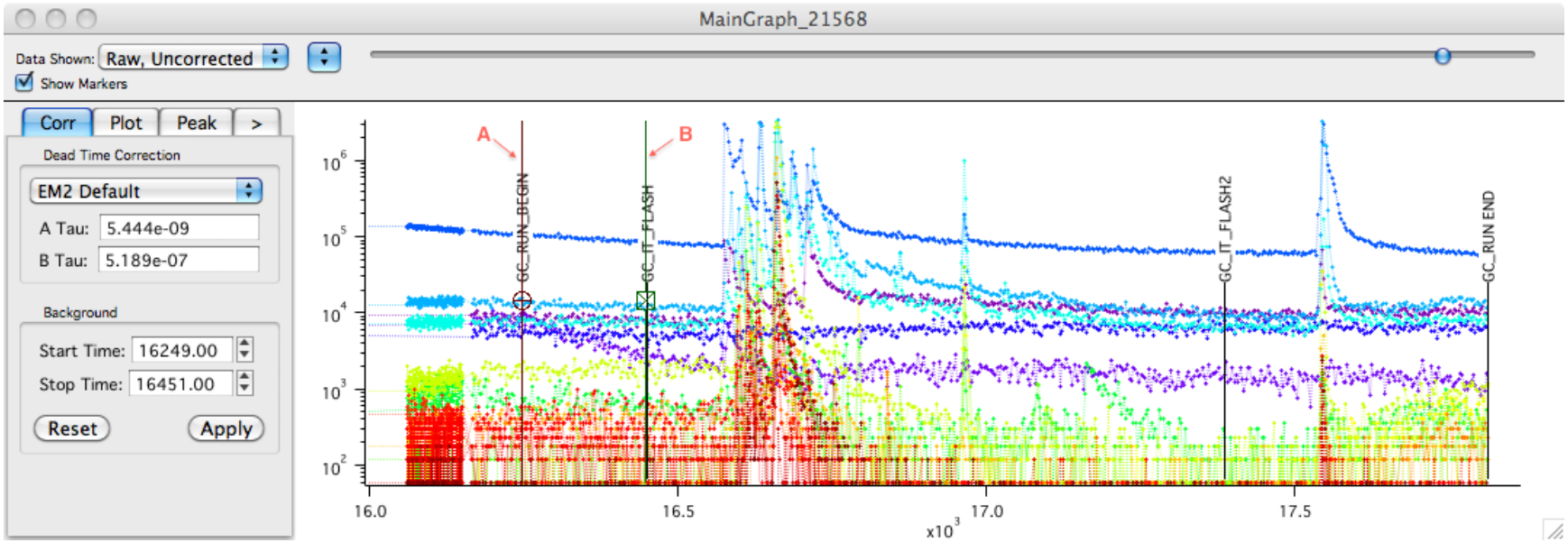


Figure 4. Examples of TM markers displayed in the Main Graph (zoomed in on GCMS part of the experiment). Note that the default positions of the background cursors (**(a)** and **(b)**) line up with TM markers GC_RUN_BEGIN and GC_IT_FLASH.

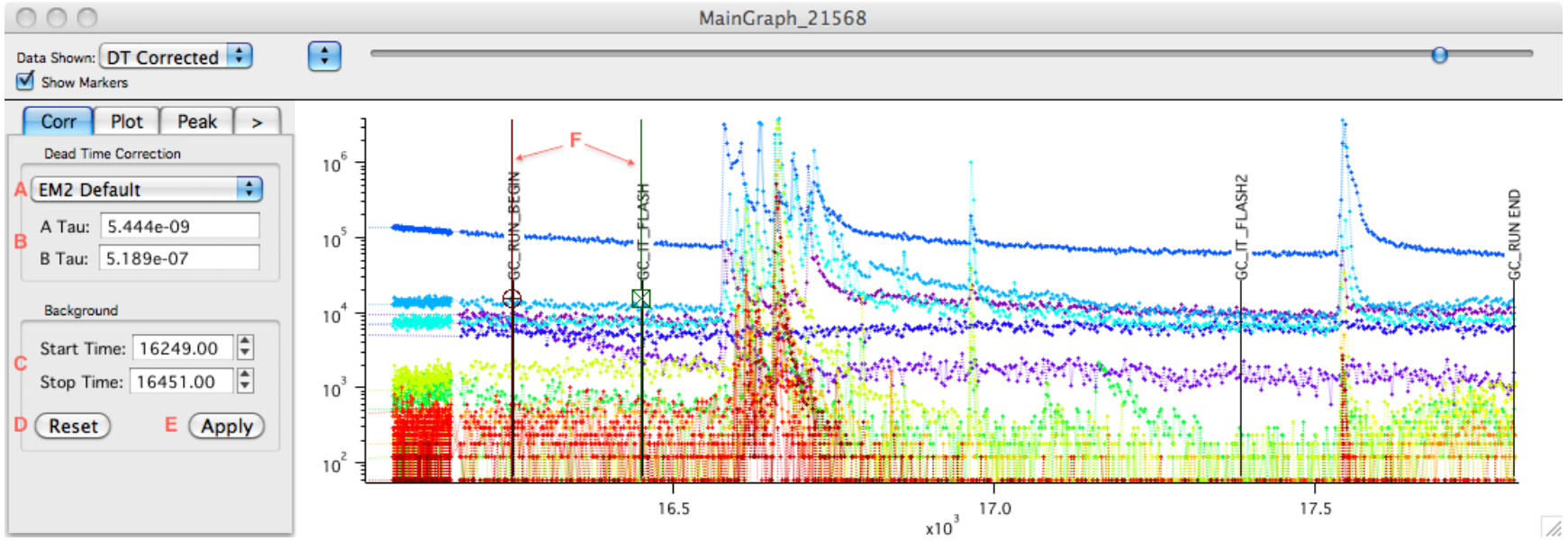


Figure 5. Correction tab in the Main Graph, zoomed in on the GCMS part of the run. **(a)** Electron multiplier selection dropdown menu. **(b)** Deadtime correction coefficients. **(c)** Start and stop time of background subtraction region. **(d)** Reset button (resets background region to default). **(e)** Apply button (applies background subtraction). **(f)** Background subtraction cursors.

The multiplier information for the current TID is printed out to the Igor command line during the TID load. This printout is the first line in the TM messages which contains the word "Multiplier."However, some of the older TIDs may not have the information in this format in the TM messages, or may not even have a TM message log. In this case, you will have to dig up the information yourself from the housekeeping data. The plots of the currents of EM1 and EM2 can be found in either the hs.hk viewer or the cdh_e.hk viewer within SAM Data View, by right clicking on either CDH:EM1_Imon or CDH:EM2_Imon and selecting "Plot Scientific."

5.4. **Selecting a background region – Correction (Corr) tab.** The two vertical lines (background subtraction cursors, Fig. 5f) that appear when you first go into this tab denote the start and stop time of the background region. Either drag the cursors or change the numbers in the Start Time and Stop Time fields (Fig. 5c) to set the background region and then lock them in with the Apply button (Fig. 5e). The initial values for these cursors are grabbed from the TM markers (it looks for the markers `GC_RUN_BEGIN` or `QMS_UBGSCAN` and then uses the time corresponding to that marker for the first cursor, and the time of the following marker for the second cursor).

If you wish to alter your background region, hit the Reset button (Fig. 5d) to go back to the initial TM marker guesses, adjust the start and stop time as desired, then hit Apply again to recalculate.

5.5. **Details of background subtraction.** The background is calculated separately for each mass and each band, but is a constant value within a mass/band. Within the background region, the signal (cps) for a given mass/band is averaged, which becomes the background value for that mass/band, and a standard deviation of the background is calculated. The background value for each mass/band is then subtracted from every deadtime-corrected value in that mass/band.

5.5.1. *Optional: Bypassing background subtraction in Igor.* If you want to do all background subtraction in ACD or other commercial GCMS software (because SAGE assumes a constant background for each mass/band, although the background will not necessarily be constant throughout the run), there are two options for bypassing background subtraction in SAGE. Either you can add the background back when you export the chromatogram (see Section 5.9), or you can set the background waves to zero and then apply the background.

The first option is the simplest, and allows you to just add the background to each of the waves after the end of the peak-fitting process. Once all peaks have been fitted with whatever background you originally chose, check the checkbox labeled "Add Bkg Back to Output"(Fig. 8g) in the Export tab of the Main Graph. If this option is checked, the background will be added back to the output when the Export button (Fig. 8h) is clicked.

The second option lets you set the background to zero before any analysis or fitting is done. To do this, navigate to the *tid######* folder in the data browser, then type the following commands into the command line and hit Enter:

```
bkg_mass = 0
bkg_band = 0
```

These commands set the background of all masses and bands to 0. After zeroing these waves you have to use the Subtract Bkg option in the Macros menu in order to subtract the background without automatically recalculating the background levels from the selected background range.

5.6. **Displaying different data corrections.** The dropdown menu in the top left corner of the Main Graph (Fig. 2a) allows the user to switch between three different corrections of the data: the raw uncorrected data ("Raw, Uncorrected"), the deadtime-corrected data ("DT Corrected"), and the background-subtracted data (background subtraction is always applied to deadtime-corrected data, so this option is abbreviated to just "BKG Subtracted"even though it is also deadtime-corrected).

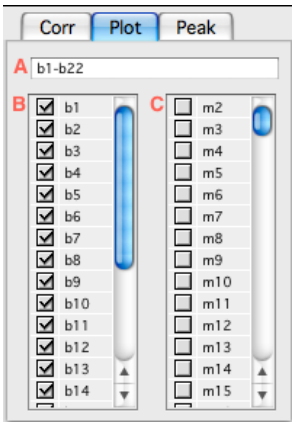


Figure 6. Main Graph: Plot tab. **(a)** Editable text field containing a list of the selected masses and bands. **(b)** Band checklist. **(c)** Mass checklist.

5.7. **Viewing the mass/band data - Plot tab.** The default set of data which is displayed in the Main Graph is the full set of bands, but individual masses can be added as well. Which data sets are displayed can be changed in the Plot tab (Fig. 2f, Fig. 6).

The text field at the top of the Plot tab area (Fig. 6a) tells you which bands/masses are currently displayed. Any changes to this string will be reflected in the display and the checkbox once you hit enter. Bands are denoted as *b#* and masses as *m#* where # is the band or mass number. Masses or bands can be listed as a range

(e.g., $b1-b22$ to plot all the bands), or as individual comma-separated entries (e.g., $m78, m91$), or as some combination of these (e.g., $b1, b4-b7, m73-79, m91$).

The band and mass checklists (Fig. 6b - 6c) provide a fast way to scan through the masses and bands and see which ones have peaks of interest. Clicking a band in the band checklist (Fig. 6b) will highlight the included masses in the mass checklist (Fig. 6c), and will scroll the mass checklist to the start of that band. Double clicking on a band will check all of the masses in that band, and if all of the masses in a band are checked, double clicking on the band will uncheck all of those masses.

Once you have selected a mass or band, you can navigate through the checklist with the up and down keys (move between adjacent entries in the checklist). Another useful way to navigate the checklist is to use the J and K keys to go backwards and forwards (respectively) between only checked boxes. $Shift + J$ or $Shift + K$ can be used to scroll through only unchecked masses.

When a mass or band is selected, pressing the space bar will check or uncheck it (which will display or hide it, respectively). When a mass or band is selected it will appear as a solid black line on the graph so that it is easier to see amongst all the colored lines.

If you add masses or bands, their traces will default to black. Use the Make Traces Different window to quickly change the color or appearance of all the plotted data at once (to find this window if it has been closed, use the menu bar to go to Windows > Other Windows > Make Traces Different). For instance if you wiggle any of the three sliders in the Color Wheel section, all the data will be assigned a different color, which is useful if you are trying to distinguish between a few different masses you just added.

5.8. **Manual peak selection - Peak tab.** The Peak tab (Fig. 7) is where the user keeps track of all the peaks (or peak regions in the case of overlapping peaks) they wish to fit.

To add a peak, click the "+", or Add Peak, button (Fig. 7b). This adds a new entry to the peak list (Fig. 7e) and creates two cursors which will define the start and stop of a peak or peak region (Fig. 7f). You can either drag these to the peak boundaries or you can specify the start and stop time in the text fields (Fig. 7a) at the top of the Peak tab.

If a peak overlaps another peak, you can attempt to deconvolve them during your fitting if you include the overlapping peaks within the same peak region (as in Peak 2 in Fig. 7).

You can rename the peak by typing in the "Name"field of the peak list.

When a peak is selected in the peak list, you can delete it by clicking the "−", or Delete Peak, button (Fig. 7c).

If the checkbox to the left of the peak name is checked the regions of the peaks are greyed out, so that it is easy to see what areas still need to be looked at.

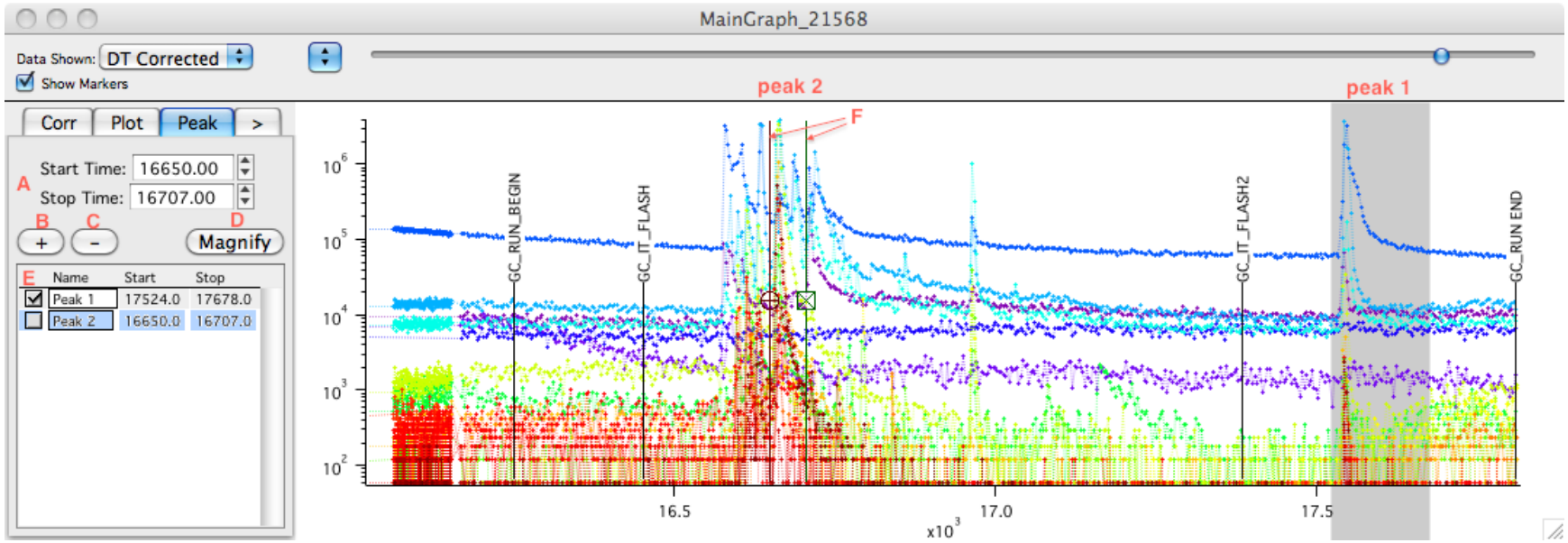


Figure 7. Peak tab with example peaks. **(a)** Start Time and Stop Time text fields (linked to peak cursors). **(b)** Peak add button. **(c)** Peak delete button. **(d)** Magnify button. **(e)** Peak list (with example peaks). Peak 1 is checked, which both greys out that region in the graph and means it will be included in the output. Peak 2 is not checked, but since it is currently selected in the peak list, the cursors which define its start and stop time are available. Peak 2 is not just one peak, it is a region containing two overlapping peaks. **(f)** Peak region start and stop cursors for Peak 2.

Checking the peaks also means that they will be included when the output to ACD is generated.

Once you have a peak in your peak list, you are ready to start fitting! Select the peak, and click the Magnify button (Fig. 7d) or double click on it in the peak list to enter the Peak Graph (see Section 6).

5.9. **Exporting chromatograms to ACD – Export tab (> tab).** After all the peak fitting has been done, there is enough information to reconstruct a chromatogram incorporating all of the fitted peaks and all of the non-fitted areas. The data for each mass will be resampled at time steps of 20 milliseconds, using either the calculated fit equations (in peak areas) or the filler.

The settings for chromatogram output are located in the Export tab (Fig. 8).

Only the data in between the user-defined Time Zero (Fig. 8b, 8i) and Maximum Time (Fig. 8c, 8l) will be exported to the chromatogram. If a TM marker is displayed (see Section 5.1), you can easily set either the Time Zero or Maximum Time to exactly the value of that TM marker by dragging the Zero or Maximum cursor directly on top of the plotted TM marker.

The offset axis (Fig. 8), which is displayed at the top of the Main Graph when the "Show Offset" option (Fig. 8a) is checked, indicates the time axis that will correspond to the data in the output file, based on the specified Time Zero. This feature is necessary when it comes to comparing two different TIDs to each other, because corresponding experiment segments will not start at the same time in two different TIDs. In other words, the GCMS retention times or EGA oven temperatures should not be calculated relative to the starting time of the TID, but rather relative to the starting time of one particular segment of the overall experiment.

There are currently two choices for how to fill the non-fitted areas in each mass, which you can switch between using the "Fill-in Data "dropdown menu. The Line Fit option generates a linear fit to each gap between fitted peak areas. The linear fit in each gap is sampled at 20ms intervals when filling in the output wave.

The Average option does not perform a fit, but rather uses the average value for a given set of points. This method makes more sense if the baseline seems nonlinear. In each of the gaps between fitted peaks, the average of each five point bins is used as the value in the output across the time range of that five point bin.

If you wish to bypass the background subtraction (for instance, if you want to have the ability to perform localized peak-by-peak background subtractions in commercial GCMS software), check the Add Bkg Back checkbox (Fig. 8g). See Section 5.5.1 for more information on bypassing background subtraction.

5.10. **Other features in Main Graph.** The slider at the top of the Main Graph (Fig. 2c) provides convenient ways to move around the graph. The default action is for dragging the cursor to cause the graph to scroll horizontally. To see how else

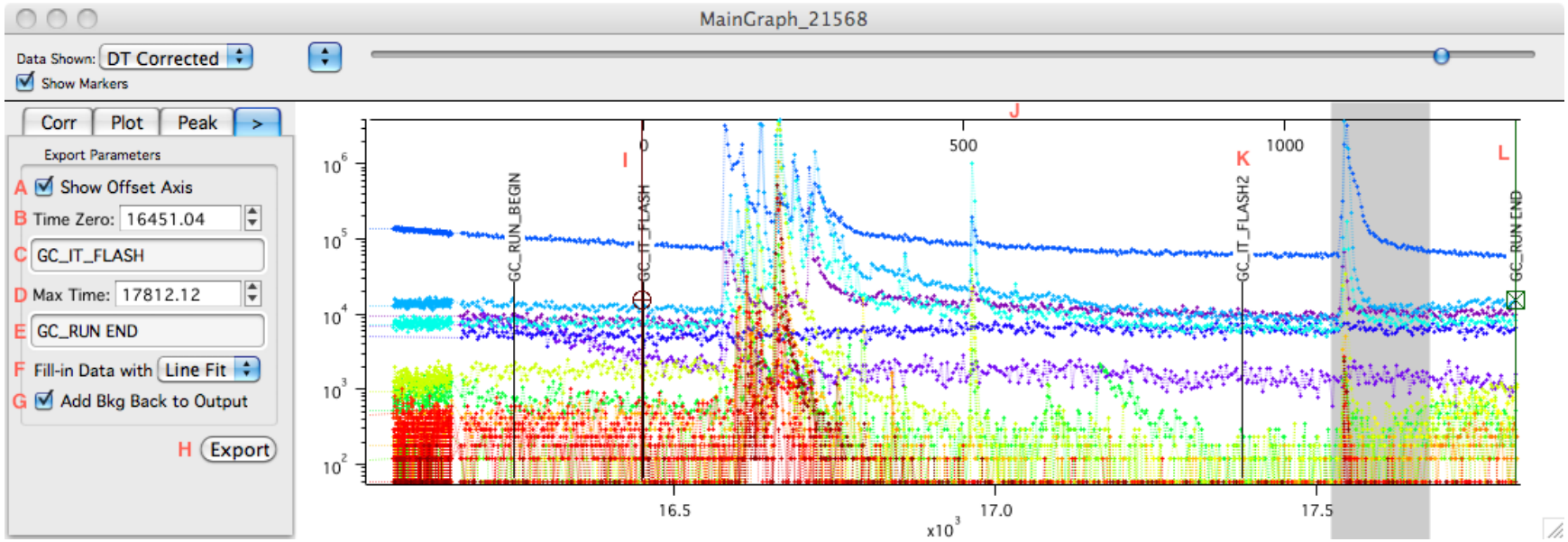


Figure 8. Export tab (> tab), with a smattering of TM markers plotted. Note that the default positions of the Zero Time and Maximum Time cursors correspond to the GC_IT_FLASH and GC_RUN END markers respectively. **(a)** Option to show or hide the offset axis. **(b)** Time Zero text field. **(c)** TM marker for Time Zero (if the time in the Time Zero field does not correspond to a TM marker, this box is blank). **(d)** Maximum Time text field. **(e)** TM marker for Maximum Time (if the time in the Maximum Time field does not correspond to a TM marker, this box is blank). **(f)** Choose the method for filling in the gaps between fitted peaks. **(g)** Option to add subtracted background back to output **(h)** Export button. **(i)** Time Zero cursor. **(j)** Offset Axis. **(k)** An example of a displayed TM marker. **(l)** Maximum Time Cursor.

this slider can be used, click the drop-down list to the left of the slider (Fig. 2b), and click "Instructions".

You can modify the Main Graph in all the usual Igor ways. For example, you can double-click on axis to modify axis properties such as range, label, linear/log.

6. Peak Graph

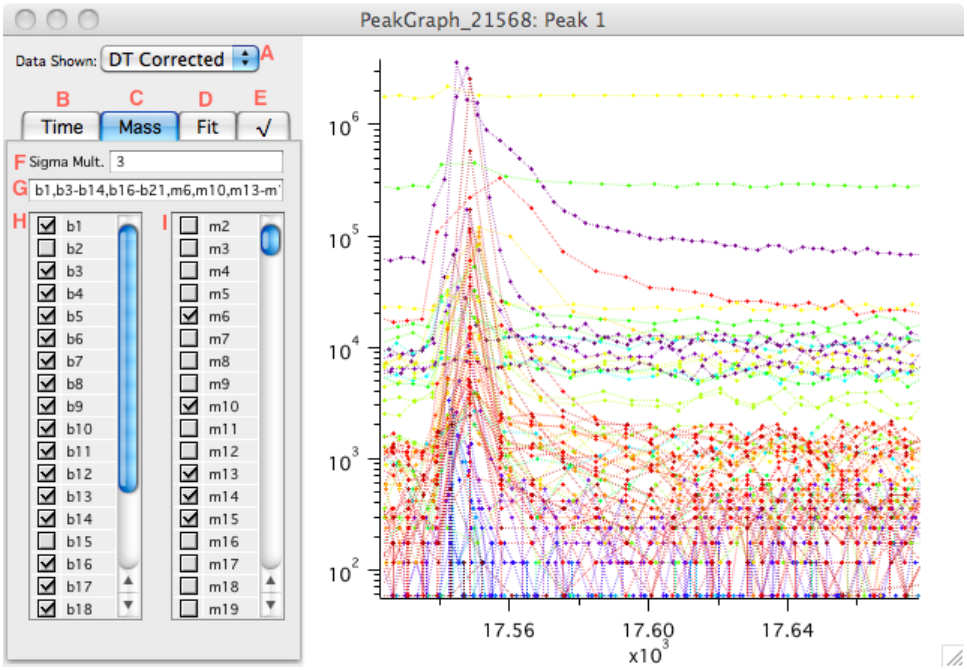


Figure 9. Peak Graph showing example Peak 1 (from Fig. 7). **(a)** Data selection dropdown menu. **(b)** Time tab. **(c)** Mass tab. **(d)** Fit tab. **(e)** Review tab (✓tab). **(f)** Sigma multiplier (noise threshold) field. **(g)** Editable band and mass text list. **(h)** Band checklist. **(i)** Mass checklist.

To get to the Peak Graph (or Peak View) window (Fig. 9), define a peak region in the Main Graph window (by clicking the "+"button in the Peak tab of the Main Graph), then either "Magnify"it or double-click on it in the peak list.

All peak fitting will be done in this window. When you first open the Peak Graph it will be in the Mass tab, and the graph will display a zoomed-in region of the Main Graph, based on the start and stop time of the peak.

6.1. **Data selection.** Once again, the data used and plotted in the graph can be either the raw uncorrected data, the deadtime-corrected data, or the deadtime-corrected and background-subtracted data. The data set can be switched by changing the selection in the data selection dropdown menu (Fig. 9a).

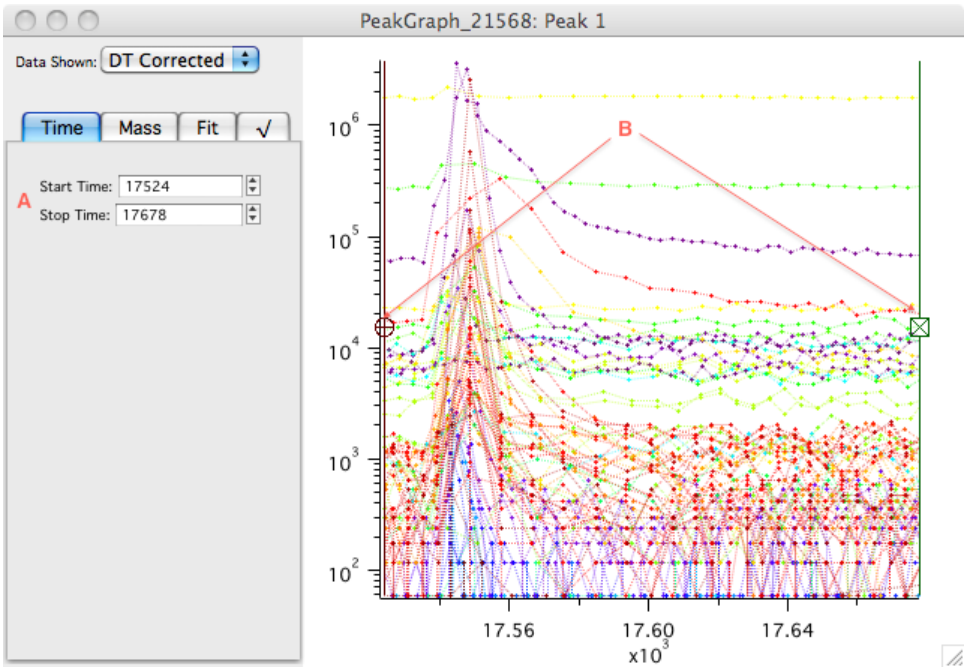


Figure 10. Peak Graph: Time tab. **(a)** Start and stop time text fields. **(b)** Peak time cursors.

6.2. **Time tab.** The Time tab (Fig. 9b) is used to adjust which points are included in the fit. Changing the values in the text fields (Fig. 10a) or the cursor positions (Fig. 10b) will update the region will be used in the fit.

6.3. **Mass tab.** The Mass tab (Fig. 9c) is where the user selects the set of bands/masses to be used in the fit for this particular peak (or set of overlapping peaks).

The number in the Sigma Multiplier field (Fig. 9f) is multiplied by the standard deviation of the background for each band/mass, and this is used as the noise threshold for each band/mass. If the band/mass data goes above the noise level within the start and stop time of the peak, it is plotted, checked in the checklist (Fig. 9h - 9i), and included in the band/mass text list (Fig. 9g). Changing this number and pressing the *Enter* or *Return* key will update the masses plotted.

Double-clicking on a band in the band checklist will plot all of the masses in that band. If all of the masses in that band are plotted, double-clicking on the band will uncheck them all.

The checklist and text list functionality that is present in the Plot tab of the Main Graph also applies in the Mass tab of the Peak Graph.

Once the list of bands/masses includes everything you think is relevant, go to the Fit tab.

6.3.1. *Notes on mass selection.* On one hand, if you don't include all of a molecules major masses into the fit, you may have a hard time identifying the molecule later, and it won't necessarily be obvious what molecules might be present at this early stage. However, if you add too many masses, you can remove many unwanted masses quickly with the "Remove Bad Fits"button after the fit calculation has been done. (The "Remove Bad Fits"button will deselect any masses whose fit has a negative amplitude.)

Both of the above points argue for adding as many masses as possible. However, the more masses you add to the fit, the longer the fit calculation will take. (Additionally, the more overlapping peaks you are trying to deconvolve, the longer the fit calculation will take, which is a tangential point to consider.) The more junk data you include in the fit, the worse the resulting fit will be, because the algorithm will try to fit to the junk data as well as the data where there is a real peak.

A suggested technique for catching all of the relevant masses and not including too much junk data is to add all the masses from each of the bands which have a peak above some noise threshold. This can also be done iteratively: add one set of masses, perform a fit, remove bad fits, add next set of masses, and so on. If any of the masses you added obviously do not have a peak in this time range, you can manually deselect it to help feed better information to the fitting algorithm.

6.4. **Fit tab.** The Fit tab (Fig. 11) is where it all happens.

First, select a fit function from the dropdown menu (Fig. 11a). The EMGFit (exponentially modified Gaussian) is recommended. If there is more than one peak, select one of the functions starting with "Multi-". Ignore the "External Functions", as these were not functions added by the SAGE developers.

6.4.1. *Fit Functions.* EMGFit/MultiEMGFit – Exponentially Modified Gaussian (Fig. 12). Accurately describes most chromatographic peaks, because most real peaks are not symmetric. EMG equation (Fig. 13) and method of calculating peak parameters were drawn from Yau and Kirkland (2001).

GaussianFit/MultiGaussFit – A Gaussian curve is symmetric and emphasizes the maximum of the peak. See Figures 12 and 14 for examples.

LorFit/MultiLorFit – A Lorentzian curve is symmetric and emphasizes the tails of the peak. See Fig. 14 for an example.

6.4.2. *Peak parameters.* The built-in GlobalFit function requires a reasonably good initial guess of all coefficients in the function you are fitting to, otherwise the fit will not converge. Depending on which function is selected, the user will have to manually set a few of the peak parameters using the fit input cursor(s) (Fig. 11f or the fit input text field(s) (Fig. 11b). 1 summarizes which peak coefficient initial guesses are determined directly from user input (which the user has the most control over), and which are calculated automatically by the code.

To return the cursor positions to default, press the Reset button (Fig. 11c).

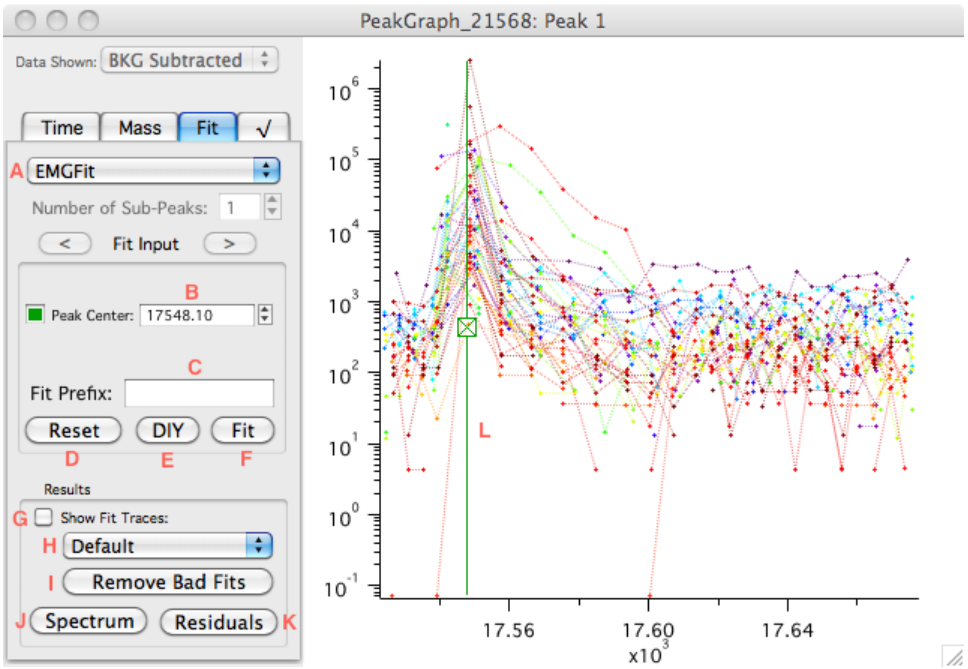


Figure 11. Peak Graph: Fit tab. **(a)** Fit function selection dropdown menu. **(b)** Fit parameters text field(s). **(c)** Fit Prefix field. **(d)** Reset button. **(e)** DIY button. **(f)** Fit button. **(g)** Fit traces display checkbox. **(h)** Currently displayed fit prefix dropdown menu. **(i)** Remove Bad Fits button. **(j)** Spectrum button. **(k)** Residuals button. **(l)** Fit parameter cursor.

When there is more than one peak, as in the case of example Peak 2, you must enter the number of sub-peaks into the number field (Fig. 15a), and then scroll back and forth between the sub-peaks using the sub-peak arrows (Fig. 15b), setting the fit input cursors for each sub-peak.

6.4.3. *Fit Prefix.* Sometimes it might not be clear what fit equation or what parameter values will provide the best fit for a given data set. If you want to try out a few different ways of fitting the peak(s), prefixes can be used to identify each of them. They are called prefixes because the wave names of the fitted data will all start with that string.

To set a fit prefix, type into the Fit Prefix field (Fig. 11c). This string will be added to the start of all the GlobalFit-generated waves for the next fit you do. If you already used this prefix, the old fit will be overwritten.

Important Notes on Prefix Conventions:

- Currently the MAXIMUM PREFIX LENGTH is 8 characters. Igor caps wave names at 32 characters total, so any more than 8 characters in the

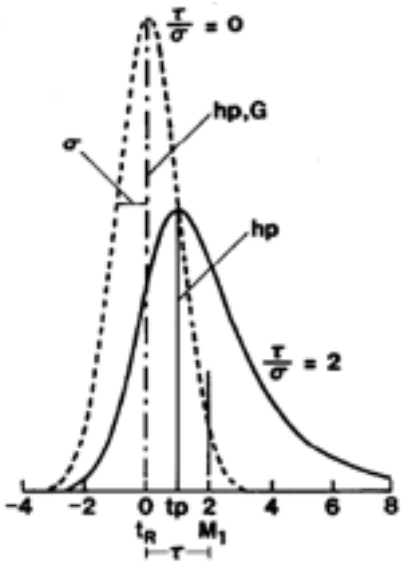


Figure 12. Model of an EMG curve (solid). The Gaussian curve which the EMG was based on is also plotted (dashed). (Yau and Kirkland, 2001)

The following EMG peak-shape model is used:

$$h(t) = \left(\frac{A}{2\tau}\right) \exp\left(\frac{\sigma^2}{2\tau^2} - \frac{t - t_R}{\tau}\right)[1 + \mathrm{erf}(Z/\sqrt{2})]$$

where

$$Z = \frac{t - t_R}{\sigma} - \frac{\sigma}{\tau}$$

and

$$\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy$$

Figure 13. EMG equation (Yau and Kirkland, 2001).

prefix may cause some wave names to be truncated, which will create a lot of problems.

- The prefix name should only use letters, numbers, and underscores – all other characters will be turned into underscores
- If you start the prefix with a number, an "X"will be added to the front of the prefix

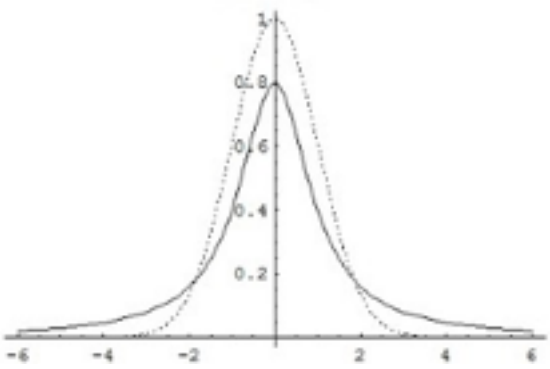


Figure 14. Comparison of Lorentzian (or Cauchy) curve (solid) and a Gaussian curve (dashed).

Fit Function	User-selected		Calculated			
EMGFit (exponentially modified Gaussian)	peak center		y-offset	peak area	exponential decay constant	width
MultiEMGFit (multi-peak EMG)	peak center	right divide	y-offset	peak area	exponential decay constant	width
GaussianFit	peak center	FWHM bounds	y-offset	amplitude		
MultiGaussFit	peak center	FWHM bounds	y-offset	amplitude		
LorFit (Lorentzian)	peak center	FWHM bounds	y-offset	amplitude		
MultiLorFit	peak center	FWHM bounds	y-offset	amplitude		

Table 1. Fit Function Coefficients

- DO NOT start your prefix with "Coef_ ", "sig_ ", "GRes_ ", or "GFit_ "because these are reserved for GlobalFit's use.
- Keep in mind that Igor is case-insensitive

6.4.4. *DIY button.* Clicking the DIY button (Fig. 11d) opens the GlobalFit user interface. For instructions on how to use this interface, see appendix A. For a far less complicated way of calling the same fitting algorithm, use the Fit button (Fig. 11e).

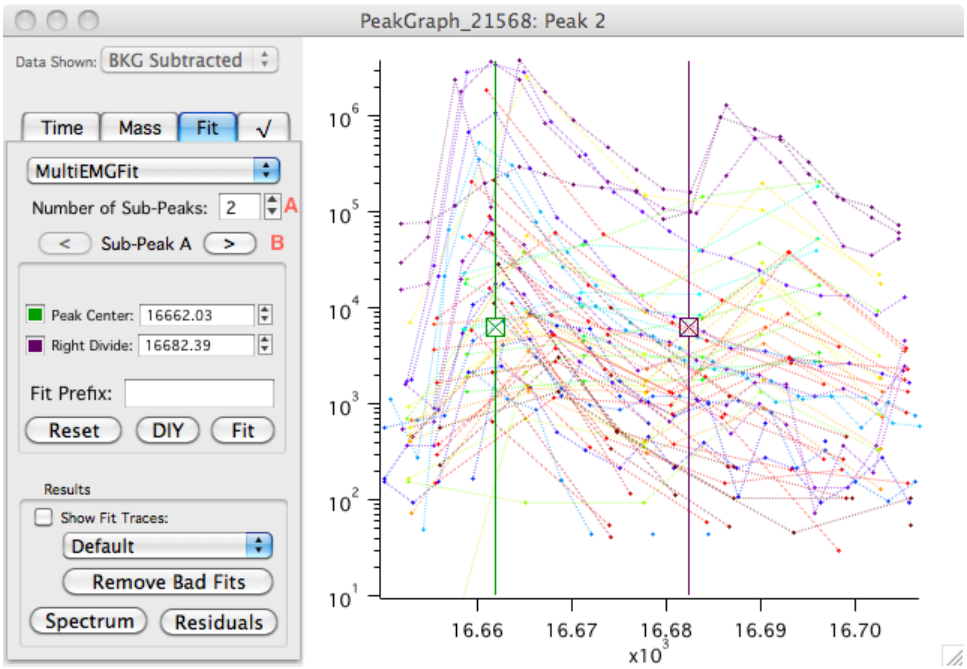


Figure 15. Overlapping peak example (Peak 2a). MultiEMGFit with 2 sub-peaks. The Right Divide fit input cursor defines the boundary where the second subpeak starts to dominate over the first subpeak. This is necessary because the EMG peak coefficients are estimated based on the area of the peak.

6.4.5. *Fit button.* Clicking the Fit button (Fig. 11e) will call GlobalFit programmatically, and the fit traces will automatically be plotted in the Peak Graph window.

The fit prefix (Fig. 11h) is set to "auto"when the Fit button is used.

This method is highly recommended over the DIY method, but if you need to tweak the automatically calculated (not the user-selected) values of the initial guesses, the DIY might be a better option.

6.4.6. *Displaying Fit Traces.* Fit traces from the most recent fit will automatically be plotted (Fig. 17a). If you want to hide the fit traces, uncheck the "Show Fit Traces "checkbox (Fig. 11g).

If you renamed your fit outputs in the DIY GlobalFit interface, multiple fit prefixes will be listed in the fit traces dropdown menu (Fig. 11h), and switching between the items in this menu will plot the different fits in the Peak Graph.

6.4.7. *Remove Bad Fits button.* The Remove Bad Fits button (Fig. 11i) removes the fits of all the masses/bands whose fitted peaks have a negative amplitude. If a Multi- fit function was used, this button removes only masses/bands which have negative amplitudes for all sub-peaks (Fig. 17b).

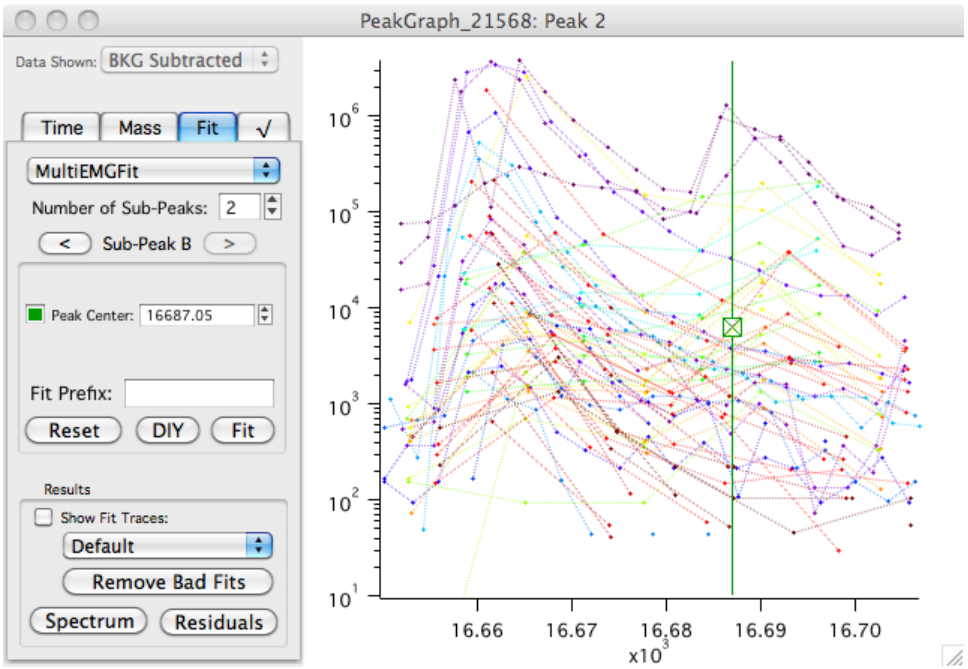
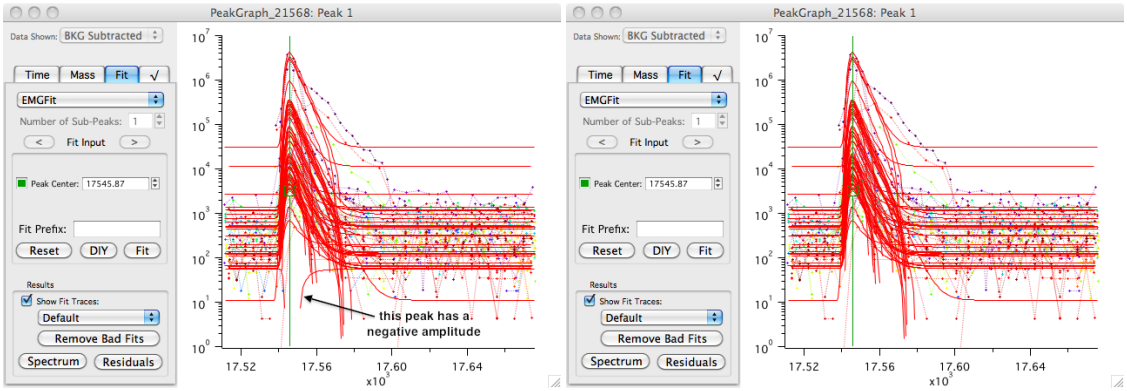


Figure 16. Overlapping peak example (Peak 2a). MultiEMGFit with 2 sub-peaks. Note that there is no Right Divide cursor for the final sub-peak (the peak stop time is considered the right divide).

Figure 17. Fit Traces for Example Peak 1



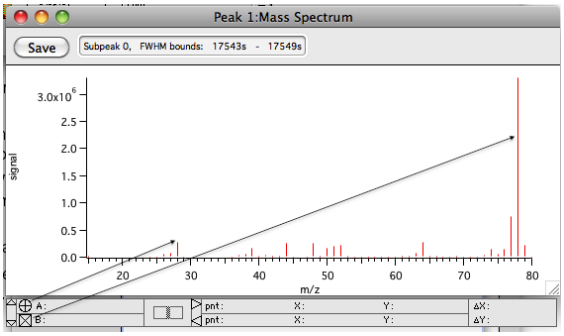
(a) The fit traces are plotted in red over the original data. One mass has a peak with a negative amplitude, which means that mass is not accurately represented by the shape of the fit.

(b) After clicking the Remove Bad Fits button, the mass whose peak has a negative amplitude has been removed. After removing bad fits, you should recalculate the fit to make sure the fit has stabilized.

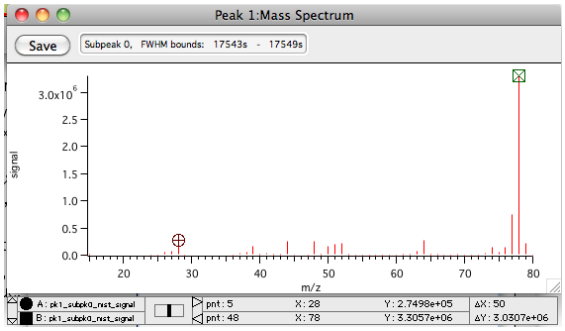
6.4.8. *Spectrum.* The Spectrum button (Fig. 11j) plots a mass spectrum by calculating the average peak height within the peaks full-width half-maximum range for each of the masses/bands included currently in Peak Graph. The fitted data is used in this calculation, not the original data, so this button does not work if no fit has been done yet.

In the mass spectrum graph, there is a save button (Fig. 18a) which can be used to export this spectrum to a NIST-format text file (as opposed to the Export MSPs option in the Macros menu, which exports a mass spectrum for each fitted peak or sub-peak).

Figure 18. Mass Spectra for Example Peak 1



(a) Spectrum from the fit of example Peak 1. To find out which mass a line in the mass spectrum represents, press Command + I to display the Show Info bar, and then drag the cursors onto the graph as shown.



(b) Once the cursors are on the graph, the Show Info bar displays the cursors'current x- and y-coordinate values, as well as point number, Δx and Δy, and the name of the wave. The cursors can be dragged to different data points, or you can use the left and right arrow keys to scroll from point to point.

If you want to know which masses correspond to which lines, use the Show Info bar to get the values of selected data (see Fig. 18a - 18b).

6.4.9. *Residuals.* When the Residuals button (Fig. 11k) is clicked, a plot of the automatically calculated GlobalFit residuals is displayed for all the masses and for all the different fits in the fit dropdown menu. If you turn off residual calculation in the DIY GlobalFit interface, this button will not work.

This was developed for comparing different fits of the same peak (where each fit was named with a different prefix).

6.5. **Reviewing the fit – Review tab (✓tab).** In the Review tab (or ✓tab), the real data (graphed in black) for individual masses or bands can be compared to the corresponding fitted data (graphed in red). If a fitted peak seems to not be

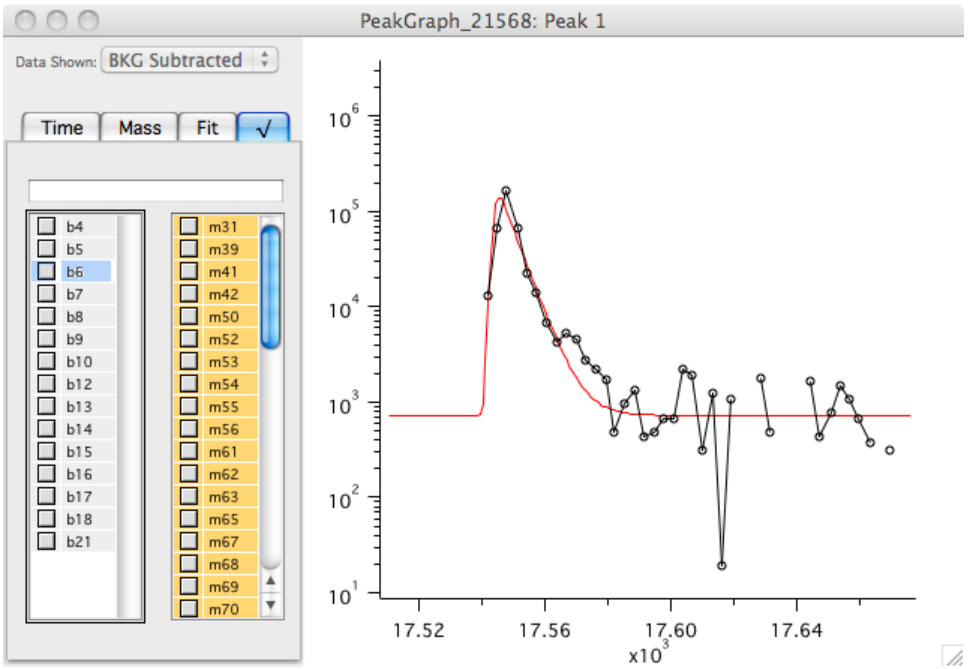


Figure 19. Peak Graph: Review (✓) tab. Real data (black) vs fitted data (red) for band 6 in Example Peak 1.

accurately representing the original data, you know that you need to either adjust the refine the set of masses you are fitting or initial guesses of the peak parameters.

In this tab, you can zoom out if you want to see the full time range of the real data set, because the peaks context might need to be taken into consideration.

7. Output from Igor

7.1. **Chromatograms.** After you have fit all of the peaks you can find, the chromatogram should be exported to ACD for analysis (Export Chrom in the Macros menu).

Only masses (not bands) are used in constructing the chromatogram. For each mass, the regions where that mass was used in a fit are resampled by plugging the calculated fit coefficients into the fit equation at evenly spaced timesteps. Then each quiet region is approximated by drawing a straight line between the two adjacent peak regions, and the new data in the quiet regions is generated by resampling at evenly spaced timesteps along that line.

The result is that for every timestep there is a mass spectrum, and for every mass there is a signal vs time curve. This is then exported to a JCAMP-format text file (.jdx), which ACD can open.

The output is not necessarily scientifically sound yet for a few reasons: overlapping peak regions overwrite each other, a simple line fit may not be the best

way to estimate the values in the quiet regions, and there is not yet a good way to compare the full time range of fitted data to the original data.

7.2. **Mass spectra.** Individual mass spectra can be exported to NIST-format text files (.msp) using the Export MSPs command in the Macros menu.

8. Macros Menu

Load TID Folder – Load a new TID into your Igor experiment.

Show Main Graph – Opens the Main Graph.

Dead Time Correct – Runs the deadtime correction calculation. (Done automatically when TID is loaded.)

Guess Bkg Range – Resets the background range cursors to the background range guesses extracted from the TM markers. (Done automatically when Apply button in Corr tab of Main Graph is clicked.)

Calculate Bkg – Calculates background and standard deviation of the data. (Done automatically when Apply button in Corr tab of Main Graph is clicked.)

Subtract Bkg – Subtracts background from all the data. (Done automatically when Apply button in Corr tab of Main Graph is clicked.)

Export MSPs – Exports one mass spectrum for each peak region (that is, once per item in the peak list in the Main Graph) to a NIST-format text file (.msp) which can be opened in NIST 11.

Export Chrom – Export a full chromatogram to a JCAMP-format text file (.jdx) that can be opened in ACD.

9. Advanced

9.1. **Data Structure.** An Igor packed experiment file (.pxp) organizes its data in a system of data folders. The data folders are internal to the experiment, and cannot be viewed externally in things like Finder (Mac) or Explorer (Windows). The data structure can be viewed only in the Data Browser.

The Data Browser is the behind-the-scenes way to view or alter the data. All the basic tasks in SAGE can be performed without needing to worry about what's in the Data Browser, because the navigation and creation of data folders is handled in the code, which is even further behind-the-scenes for the average user. However, knowing how to use the Data Browser can be quite helpful.

9.2. **Using the Data Browser.** Display or hide the contents of a folder by clicking the triangle to the left of the folder name.

9.3. **Importing data from one experiment to another.** It is recommended to have a separate experiment file for each TID so that within an experiment you dont need to keep track of what the current TID is.

If you want to look at two TIDs at the same time, or perhaps if you want to compare/merge the same TID processed by two different users, data can be easily imported into an Igor experiment from one or more other experiments.

First, open the experiment you want to load the data into. This can be a new experiment or an existing one. You may want to use Save As to make a copy of the experiment first if it is an existing experiment that you dont want to accidentally make unwanted changes to.

Then open the Data Browser (from the Data menu). Click the "Browse Expt"button on the left. Navigate to the experiment you wish to open using the file browser dialogue, and open it.

The experiment you just loaded will appear in a subwindow to the right of the original experiment in the data browser. Folders or files can now be dragged from the opened experiment to the original experiment in order to copy them from one to the other. When dragging and dropping, the file or folder can be dropped only when hovering over a folder in the original experiment, otherwise Igor wont know where to put it. Be sure to put all TID folders directly into the *root* folder, and in general always maintain the data structure that was set in place by SAGE, or you will run into problems.

9.4. **Importing a single peak into an experiment.** If multiple scientists are working on the same TID, they can merge and compare their work by adding each of their peaks to the same TID. Currently this is a very manual process, so detailed instructions are provided here.

First, make sure none of the names of the peak folders you want to import are the same as any peak folder names in the original experiment. Then open the two experiments you want to merge using the process described above in the "Importing data from one experiment to another"section.

Drag the desired peak folder(s) from the imported experiment into the original experiments TID folder. You may close the imported experiment once you have imported all of the peaks you wanted.

The data has been imported, but since it wasnt generated in this experiment, there are now some discrepancies in the data structure you will have to fix. Go to $tid\#\#\#\#\#\# > ui > peakList$. There are three waves here you will have to edit: $peakLabels$, $peakIDs$, and $peakChecks$. You may or may not have to update the variable $curPeakIDValue$.

$peakIDs$: In $peakIDs$, you have to add the same number of rows as peaks you imported (to add rows, type in the greyed out box at the bottom of the wave). This wave must contain the exact peak folder name of every peak folder you want SAGE to know about.

$peakLabels$: In $peakLabels$, you have to add the same number of rows as peaks you imported (to add rows, type in the greyed out boxes at the bottom of the wave). The peak labels should be typed in the second column. The order of entries in $peakLabels$ and $peakIDs$ should match. Entries in the $peakLabels$ second column

can be anything (i.e., do not need to match peak folder names exactly), because they are only used for display purposes. The third and fourth column can be left blank – they will automatically be filled in with the start and stop times of the peak(s) when you next go to the Peak tab of the Main Graph.

peakChecks: This wave controls whether the checkbox to the left of the peak is checked or unchecked. You must add the same number of rows as peaks you imported (to add rows, type in the greyed out boxes at the bottom of the wave), and you can initialize each of them as either checked or unchecked. Checked: 48,2,0,0. Unchecked: 32,2,0,0. You can either type these numbers into the correct boxes or copy and paste entire rows. (When pasting, the dimensions of the paste area must match the dimensions of the copy area.)

The variable $curPeakIDValue$ determines the next peak ID (its value is incremented when the "+ "button is clicked to add a new peak and then the new peak ID is $pk\#$ where # is the value of $curPeakIDValue$). This means if you added peaks whose peak IDs will conflict with these automatically generated peak IDs, you should change this variable so that a conflict does not occur. If you do not do this, your data may be overwritten if you add a new peak from the Main Graph.

9.5. **Peak area.** To find the area for a single mass (e.g. $m78$) with a single peak (that is, no overlapping peaks),

You have the option of using the `area()` function or the `Integrate\T` function, but `Integrate\T` involves the extra step of having to sum everything in the "destination wave" together.

You can put these functions into a new procedure file (Windows > New... > Procedure...):

```
Function FindIntArea()
        Integrate/T  autoGFit_cps_m78 /D=int_m78    //in this case, the destinat
        Variable i, int_area_m78
        for (i=0;i<numpnts(int_m78);i+=1)
                int_area_m78 = int_m78[i]
        endfor
        Print int_area_m78
End

Function FindAreaArea()
        Variable area_m78 = area(autoGFit_cps_m78)
        Print area_m78
End
```

Then you can call them by copying and pasting either of these lines into the command line:

```
FindAreaArea()
FindIntArea()
```

The area will be printed out after the function call. These two functions should output the same number, because both `area()` and `Integrate\T` utilize a trapezoidal integration.

10. Contact Information

Send feedback or questions to:
Anna Brunner (anna.e.brunner@nasa.gov)
Micah Johnson (micah.s.johnson@nasa.gov)

11. Appendix A – DIY Global Fit instructions

1. Use the cursors to select the peak center (x coordinate of the peak maximum). If the fit function youis Gauss the beginning and end points which define the peak's full-width at half maximum (FWHM).

2. Press the DIY button to automatically generate a wave named $CoefWave$ containing guesses for the fit coefficients which is compatible with the Global Analysis window.

3. Open the Global Analysis window by selecting "Global Fit"from the Analysis menu, if it is not already open.

4. Go to the Data Sets and Functions tab.

5. Load data:

5a. If there is already data loaded in the Data Sets area, use the Remove All option in the Remove dropdown menu. 5b. Make sure that the relevant peak plot is the top graph by clicking on it, then clicking on the Global Analysis window. 5c. Choose Add Data Sets > All From Top Graph.

6. Click the top of the Function column in the Data Sets area to highlight the entire column, then select "GaussianFit"from the Choose Fit Function dropdown menu.

7. Link coefficients:

7a. Highlight column $K2$ in the Coefficients area, then click the Link Selection button. 7b. Repeat step 7a for column $K3$.

8. Go to the Coefficent Control tab.

9. To set the initial guesses, select $CoefWave$ from the Set From Wave dropdown menu.

10. After optionally adjusting the Options at the bottom of the Global Analysis window, click Fit!

11. If the fit did not converge, repeat steps 1-2 (adjusting the cursors to be more accurate), then proceed from step 8 onward. If the fit still does not converge, any initial guesses that seem wildly off can be manually adjusted in the Coefficient Control tab.

12. Optional: After the fit is done, you can save the calculated fit coefficients to a wave using the Save to Wave dropdown menu.

12. Appendix B – Igor Tips

Trick for having two Igor experiments open at once (Mac): Go into Applications > Igor 6.3 Folder and then copy and paste the "Igor Pro.app"file. An "Igor Pro copy.app"should appear, and these two instances of Igor can be open simultaneously. Having multiple desktops is useful for keeping the Igor instances separate, since Igor (for Macs) puts its windows everywhere.

Graph Interface

- Click and drag marquee on plot, click inside marquee for Expand/Shrink options
- Hover cursor over axis, scroll wheel to zoom in and out (side scroll slides data)
- Press Option + click to slide data with inertia, click to stop if it's still moving
- To find out the values of specific data points, use Igor's Show Info capability (Command + I on Macs, or the Show Info option in the Graph menu) to show a bar below the graph which has information about two cursors which you can click and drag to any point on the graph, and then you can scroll through the data points with the left and right arrow keys.
- Holding Command + Option will hide cursors, if you want to get an unobstructed view of the data for a screenshot, for instance.

Global Fit

- If the fit does not converge, adjust the cursors if they seem off, or vary the value of the EMG modification factor. The following Igor-generated error messages are common and do not necessarily mean there is anything wrong with the data set or the program - they usually just mean the initial guesses weren't quite right. If you see these errors, do not panic; try tweaking the parameters, use the Remove Bad Fits button, or manually remove masses or bands that don't look like they have peaks. If these errors continue to occur, however, there may be something wrong besides the parameters.

```
- While executing FuncFit, the following error occurred:
  Singular matrix or other numeric error.
- While executing FuncFit, the following error occurred:
  The fitting function returned NaN for at least one X value.
- While executing a wave read, the following error occurred:
  Index out of range for wave "W_sigma".
```

- Global Fit does not do well if the peak you are trying to fit is not lined up precisely amongst the masses/bands - if there are slight time shifts, the fits do not look as good