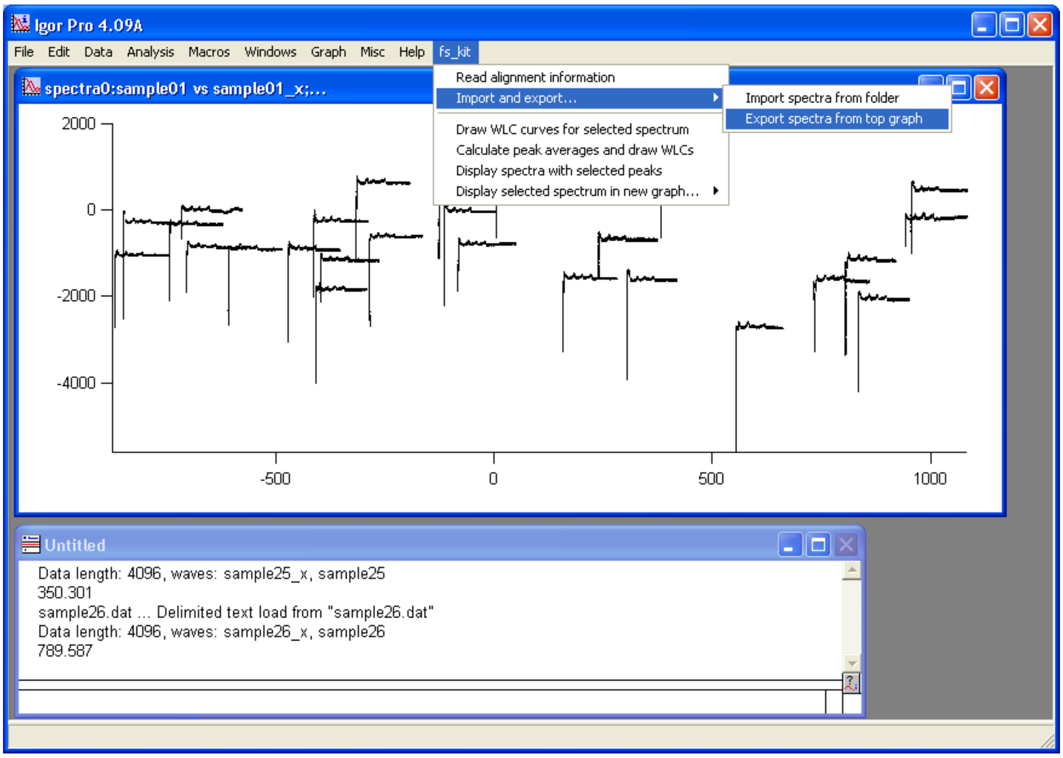
Tutorial for fs_kit

Michael Kuhn, 2003-2005

Import and export of spectra

1. Download the sample spectra, which are some partly modified real spectra. Create a new folder, which will be called main folder from now on. Extract the file into this folder. Now, the main folder should contain the spectra folder.
2. Start Igor Pro. Open the procedure file fs_igor.ipf, which resides in the igor folder of the fs_kit installation. Minimize the procedure window, or close it and then select "Hide." A new menu has been added to the menu bar: fs_kit. (To automatically load the procedure every time you start Igor or create a new experiment, create a shortcut to the igor folder in the Igor Procedures folder, which can be found in the Igor installation directory. If you don't want to do that, you'll at least be able to find)
3. Open the new menu. The upper part displays option for loading spectra and import and export of spectra. The lower part is used once spectra have been loaded. (It is not feasible to program dynamic menus with Igor, hence all menu points are active at all times.)
4. In the fs_kit menu, click on "Import spectra from folder" under "Import and export..." Navigate to the spectra folder. Click one any of the files. All files in this folder with the same extension are automatically loaded. After the files have been loaded, a new graph window with the loaded spectra is created. As you can see, the traces are distributed over a wide range, and some have visible baseline slopes.

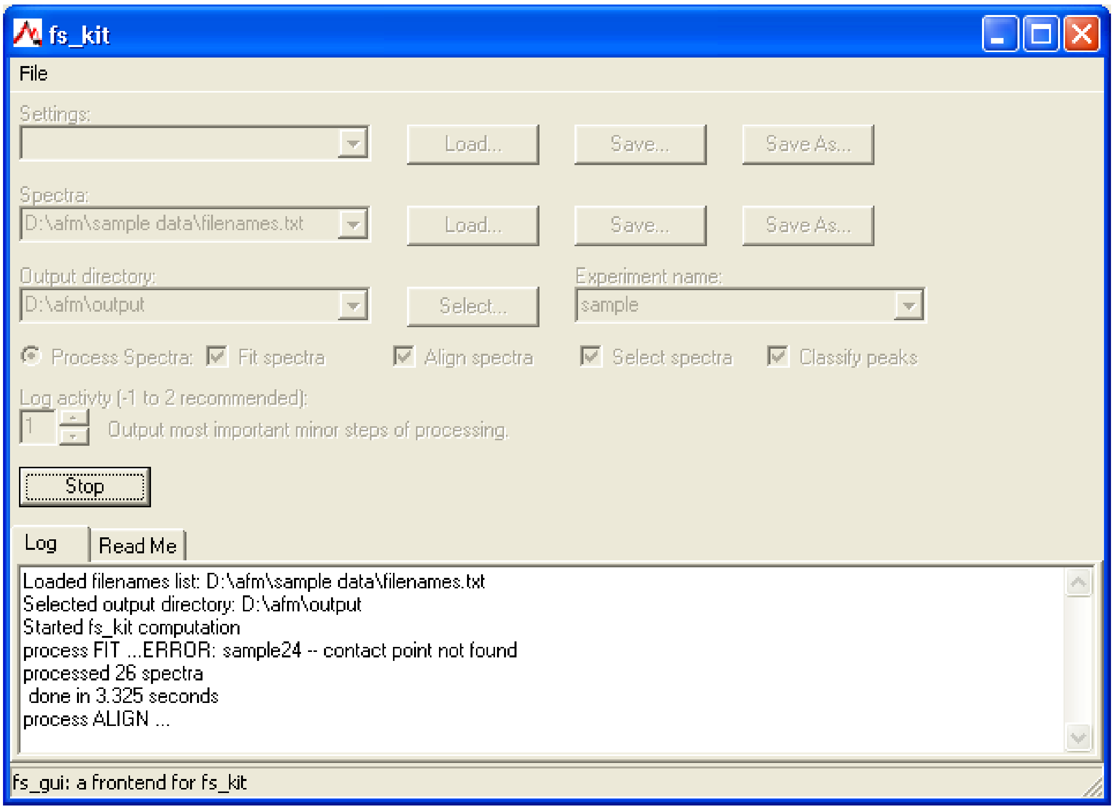

Igor Pro 4.09A
File
Edit
Data
Analysis
Macros
Windows
Graph
Misc
Help
fs_kit
spectra0:sample01 vs sample01_x;...
Read alignment information
Import and export...
Import spectra from folder
Export spectra from top graph
Draw WLC curves for selected spectrum
Calculate peak averages and draw WLCs
Display spectra with selected peaks
Display selected spectrum in new graph...
2000
0
-2000
-4000
-500
0
500
1000
Untitled

```
Data length: 4096, waves: sample25_x, sample25
350.301
sample26.dat ... Delimited text load from "sample26.dat"
Data length: 4096, waves: sample26_x, sample26
789.587
```

5. Export the spectra with the option “Export spectra from top graph”. Create a new folder in the main folder by entering a new folder name at the end of the path.

Running fs_kit: Fitting and aligning spectra

1. Start the fs_kit GUI application.
2. Load the spectra filenames by clicking on “Load” in the row labeled “Spectra”. Navigate to the folder you just created and open “filenames.txt”.
3. Using your favorite way of doing it, create a new folder in the main folder and call it “output”. This will be our output folder. Click on “Select...” in the “Output directory” row. Then, navigate to the output folder and select it.
4. Enter an experiment name, for example “sample”. The output will end up in the output folder, and all filenames will carry the same prefix – namely, the experiment name.
5. Mark all four check boxes in the column “Process Spectra”.
6. Click “Start”! You are able to stop the program by clicking the same button while it runs. A word of warning: The program might become unresponsive or crash altogether. It has been tested under a number of conditions and should work fine, but there are many things that might go wrong. If the program crashes, download the command line version and let it run in a terminal. You can specify an option file which you previously created with “fs_kit –o settings.ini”. That way, you will be able to analyze the log output even if it crashes.



Settings:

Spectra: D:\afm\sample data\filenames.txt

Output directory: D:\afm\output

Experiment name: sample

● Process Spectra: ☑ Fit spectra ☑ Align spectra ☑ Select spectra ☑ Classify peaks

Log activty (-1 to 2 recommended): 1 Output most important minor steps of processing.

```
Loaded filenames list: D:\afm\sample data\filenames.txt
Selected output directory: D:\afm\output
Started fs_kit computation
process FIT ...ERROR: sample24 -- contact point not found
processed 26 spectra
 done in 3.325 seconds
process ALIGN ...
```

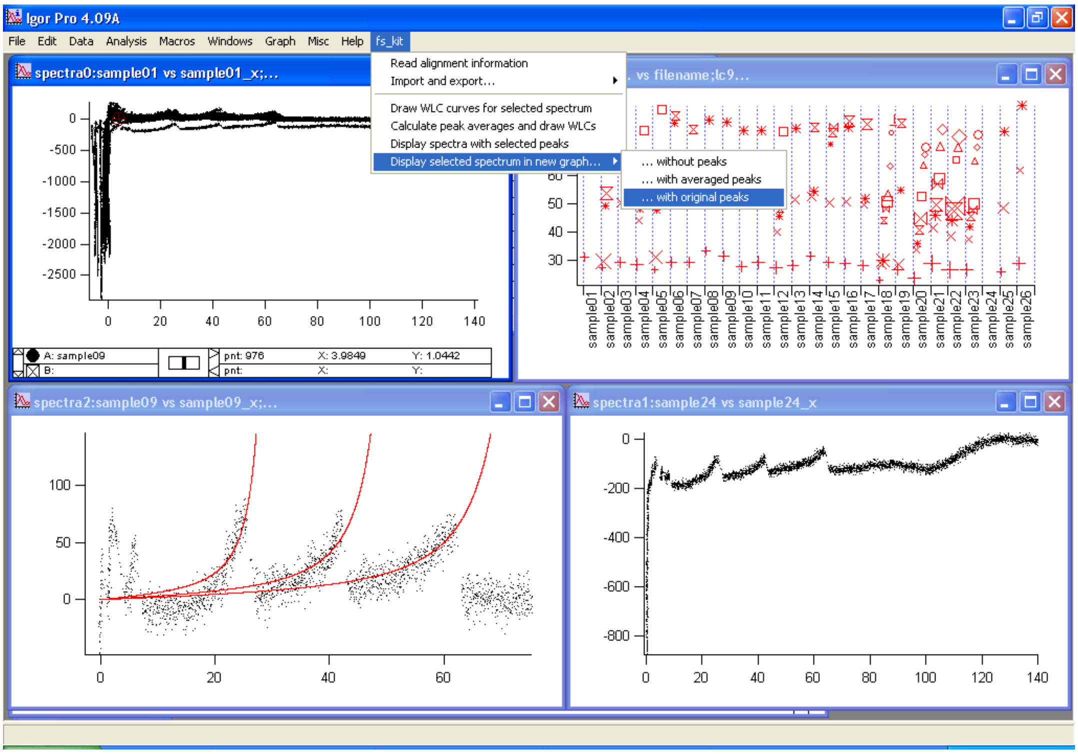
7. Wait until the computation finishes. With just 26 spectra, this will be fairly fast. Be aware: The alignment procedure is in O(n^2): If you double the number of spectra,

the computation time increases by factor four. If you don't set the log level lower than 1, fs_kit will give you an estimate for the duration of the distance calculation, which is the most time-consuming step.

8. Now, you can save your settings. (It is currently not possible to edit either the settings or the spectra to be analyzed in the program. You'll need a text editor for that.) Click on “Save...” or “Save as...” in the top row. fs_kit will save all settings into the new file. If you need to adjust specific settings, refer to the documentation.
9. Take a look at the results in the output directory. If you want to, explore the table format with a text or spreadsheet editor. A later part of this tutorial and the documentation give you more information about the format of the files.

Analyzing the results with Igor: Fitting

1. Start Igor Pro. If you chose not to automatically load fs_igor.ipf (see first part of tutorial), you have to open the procedure file, and minimize the procedure window.
2. In the fs_kit menu, click on “Read alignment information”. Navigate to the output folder and select “sample.fit.txt”. This file contains an optimal fit of peaks for each spectrum. Igor will also ask you to select a file in the spectra folder.
3. Notice the two new windows. One of the contains all spectra that were just loaded, the other contains an overview about the distribution of peaks.
4. If you're not too familiar with Igor, notice that double-clicking in the spectra graph will take you to a window “Modify Trace Appearance”. There, you can select individual traces (spectra) and use the “Offset” command to shift them away from the other spectra.
5. At first, let's look at the spectrum that doesn’t seem quite right: “sample24”. Here, the baseline is curved and no peaks have been found.
6. Now, let's look at happier examples. In the fs_kit menu, Do “Graph”, “Show Info” and drag the A cursor on any spectrum in the spectra window. (Except “sample24”.)
7. In the fs_kit menu, select “Display selected spectrum in new graph ...”, “... with original peaks”. You will learn about averaged peaks in the next part. A new window should come up, with the spectrum you selected and its peaks. Zoom in on the relevant region.


Igor Pro 4.09A
File Edit Data Analysis Macros Windows Graph Misc Help fs_kit
Read alignment information
Import and export…
Draw WLC curves for selected spectrum
Calculate peak averages and draw WLCs
Display spectra with selected peaks
Display selected spectrum in new graph…
… without peaks
… with averaged peaks
… with original peaks
spectra0:sample01 vs sample01_x;…
0
-500
-1000
-1500
-2000
-2500
0 20 40 60 80 100 120 140
A: sample09
B:
pnt: 976
X: 3.9849
Y: 1.0442
pnt:
X:
Y:
. vs filename;lc9…
50
40
30
sample01 sample02 sample03 sample04 sample05 sample06 sample07 sample08 sample09 sample10 sample11 sample12 sample13 sample14 sample15 sample16 sample17 sample18 sample19 sample20 sample21 sample22 sample23 sample24 sample25 sample26
spectra2:sample09 vs sample09_x;…
100
50
0
0 20 40 60
spectra1:sample24 vs sample24_x
0
-200
-400
-600
-800
0 20 40 60 80 100 120 140

8. Notice that the red curves (the WLC models) start at (0,0). For most spectra, the graph of the spectrum passes that point also. There are a few exceptions, where that part of the spectrum is more horizontal. fs_kit finds the contact point by averaging the data points in the vicinity of the first intersection between the baseline and the spectrum.

9. It might be the case that you have many peaks, where you only see three main peaks with the naked eye. Well, try to smooth the curve of the spectrum. (Go to “Analysis”, “Smooth...”; select “From target” and smooth “sampleNN” and “sampleNN_x”. Now you also know how the waves are called internally.) What you should see is that there are little spikes everywhere where a red line crosses the spectrum. Why don't all spikes have a WLC model? After fitting, fs_kit combines those WLC models that have almost equal contour length.

Analyzing the results with Igor: Alignment

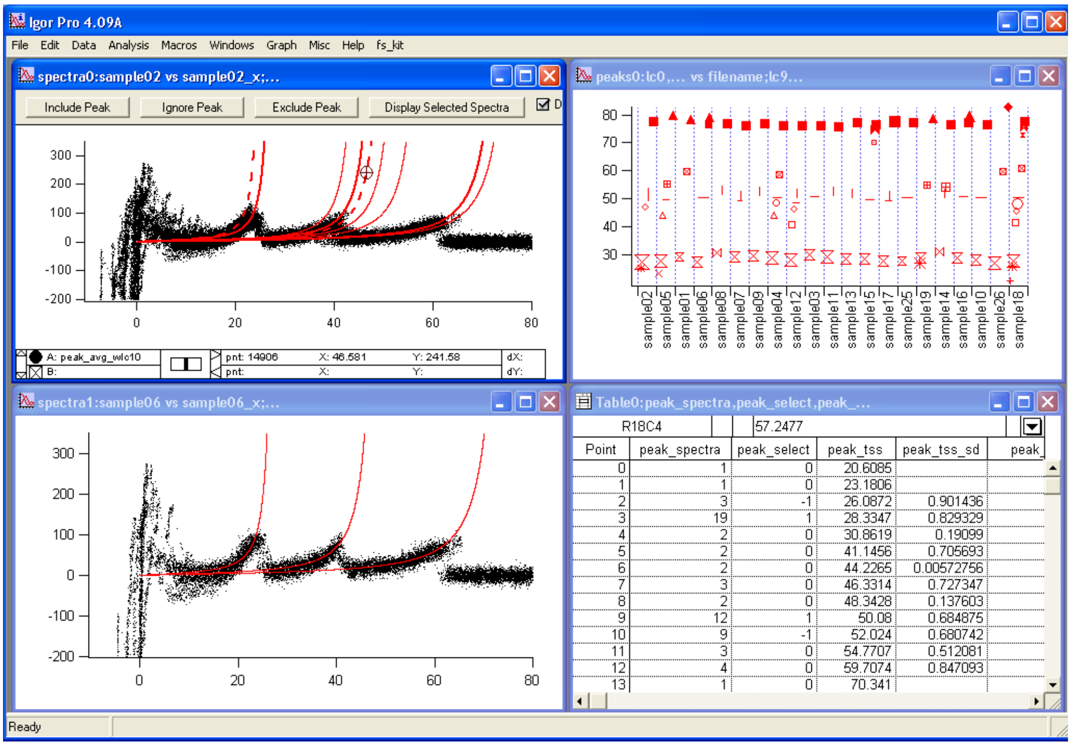
1. Start a new experiment. Load fs_igor.ipf if you have to. (You should always start with a new experiment if you load the same set of spectra with a different alignment. The frontend loads the spectra and subtracts the baseline. However, it does not overwrite existing data.)

2. Now, load the alignment “sample.align.txt”. If you zoom in a little, you will notive that the peaks seem to be much more uniform. On the other hand, some spectra have been shifted quite a bit. Their contact point does not coincide with (0,0) anymore.

3. Actually, you will rarely analyze the raw alignment files. fs_kit offers two refinements: At first, it looks at the similarity between spectra and divides them up into clusters. That is why you also saw files called “sample.1.align.txt”. Each cluster is assigned a number increasing from one. All spectra that don't fit into any cluster are put in the file “sample.0.align.txt”. Also, clusters only contain spectra

with peaks, that is to say, faulty spectra like the one with the curved baseline are removed. In this case, two clusters were created. One of them contains just two spectra, the other one 21 spectra. In a later part of the tutorial you will learn how this comes about. The other refinement is the classification of peaks. In the peak plot, every peak with the same number has the same symbol. However, most peaks with the same symbol are different actual peaks. The peak classification routine looks for peaks with similar contour length and groups them together.

4. Start a new experiment and load the alignment file “sample.1.pc.txt”. “pc” stands for “peak classification”.
5. You will notice two things: First, the alignment looks slightly better. Second, in the peak window, each symbol is almost on a horizontal line. This means that fs_kit grouped peaks with different numbers from each spectrum together. It did not shift the spectra themselves!
6. Click “fs_kit”, “Calculate peak averages and draw WLCs”. You will be asked how many spectra should share a peak for it to be displayed. If you select 0, all peaks will be shown. If you select a high number, like 10, only major peaks shared by half of the spectra (in this case) are shown. Select something in between, like 3. Now, the frontend searches through the table of peaks and calculates the “average peaks”. A new table will appear with statistical data about the peaks. The first column contains the number of spectra featuring that peak. The second column contains a flag to be used in the next few steps.
7. Go to the spectra window. Drag a cursor on the second WLC. (It should be the trace “peak_avg_wlc3”. Click on “Include peak”. Now, click “Display selected spectra”. A new window with 19 spectra will show.
8. Play around with it. Drag the cursor to new WLC models and choose one of the three options, then display your selection. Sometimes, you will just get one or even no spectrum. You should always use only the original spectra window to include/exclude peaks. Otherwise, the display (bold/dashed lines) will get out of sync with the actual data. Notice that the “peak_select” column in the table changes accordingly.



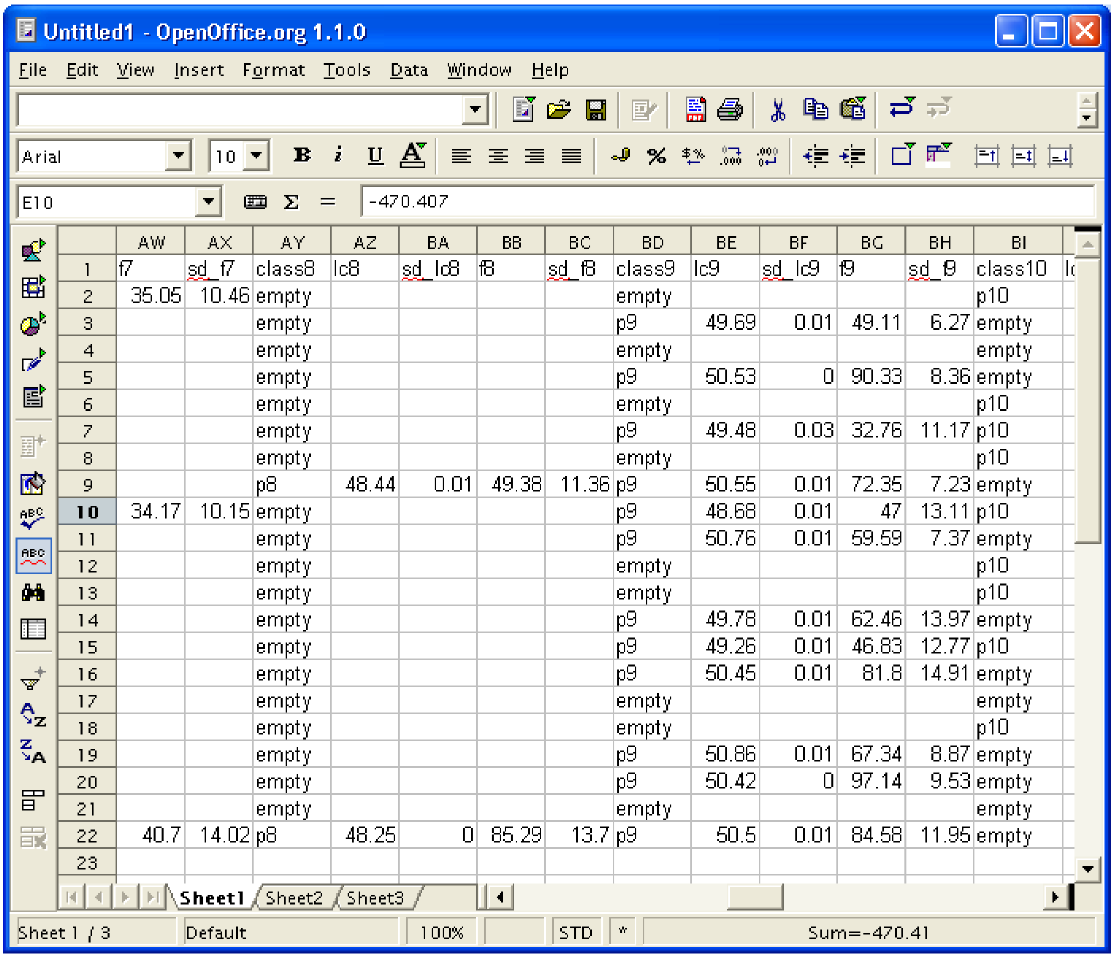
A: peak_avg_wlc10 pnt: 14906 X: 46.581 Y: 241.58 dX:
B: pnt: X: Y: dY:

R18C4 57.2477

Point	peak_spectra	peak_select	peak_tss	peak_tss_sd	peak_
0	1	0	20.6085		
1	1	0	23.1806		
2	3	-1	26.0872	0.901436	
3	19	1	28.3347	0.829329	
4	2	0	30.8619	0.19099	
5	2	0	41.1456	0.705693	
6	2	0	44.2265	0.00572756	
7	3	0	46.3314	0.727347	
8	2	0	48.3428	0.137603	
9	12	1	50.08	0.684875	
10	9	-1	52.024	0.680742	
11	3	0	54.7707	0.512081	
12	4	0	59.7074	0.847093	
13	1	0	70.341		

Behind the scenes: Data files and clustering trees

1. Look at the different files in a spreadsheet program or text editor.
2. Files with the suffix “fit”, “align” and “pc” always have the same tabular structure: The names of the spectrum and its x-trace name in Igor. The status with which fitting ended. Baseline information. Information about the peaks. However, there is a big difference between “pc” and the other two: The peak classification contains “empty” peaks. These are peaks that the spectrum under consideration does not have. Thus, if you tell your spreadsheet program to calculate statistical data for one column of the peak classification file, it will actually give you the data for just this one “real” peak. If you were to do the same for a column in the fit data, you'd get a huge standard deviation – a sign that you actually average across many different peaks. A complete description of the data formats can be found in the documentation under Data Formats.



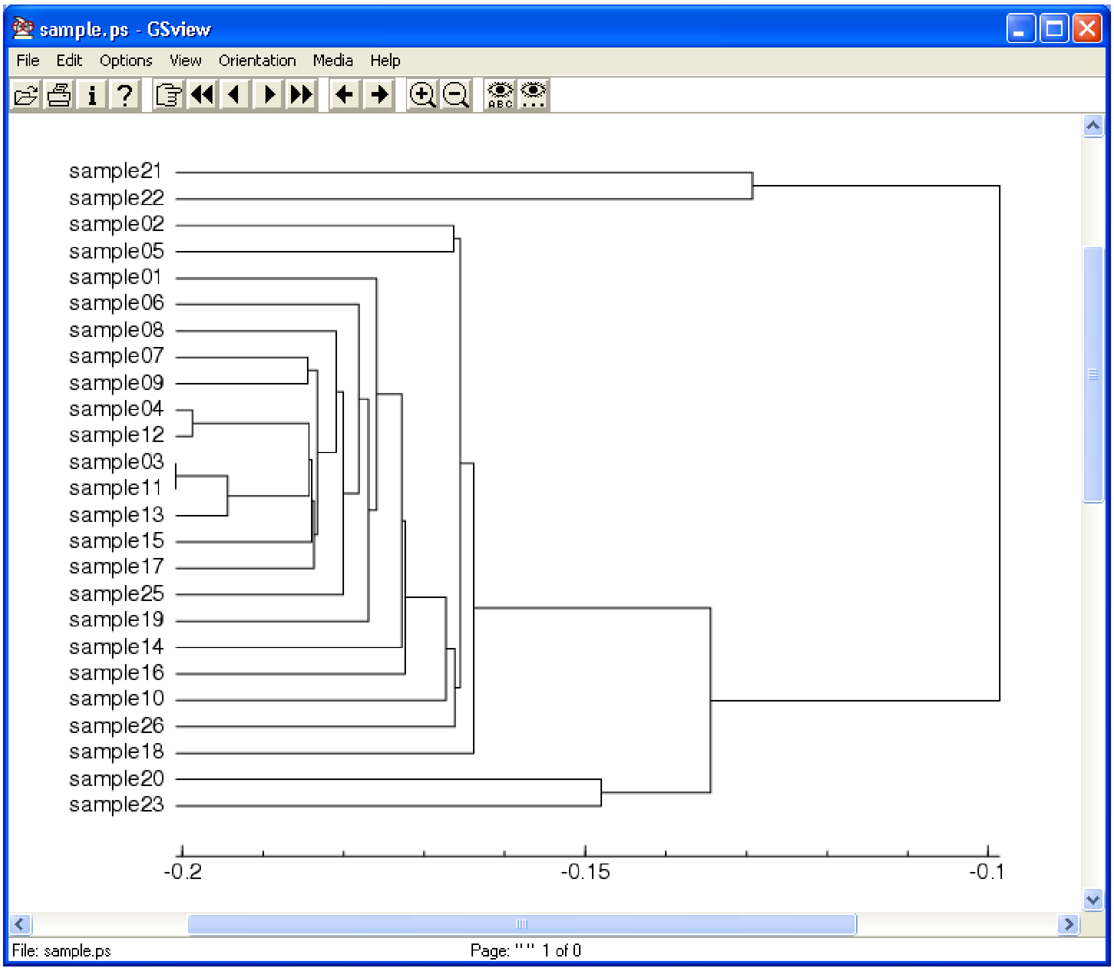
-470.407

	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	
1	f7	sd_f7	class8	lc8	sd_lc8	f8	sd_f8	class9	lc9	sd_lc9	f9	sd_f9	class10	lc
2	35.05	10.46	empty					empty					p10	
3			empty					p9	49.69	0.01	49.11	6.27	empty	
4			empty					empty					empty	
5			empty					p9	50.53	0	90.33	8.36	empty	
6			empty					empty					p10	
7			empty					p9	49.48	0.03	32.76	11.17	p10	
8			empty					empty					p10	
9			p8	48.44	0.01	49.38	11.36	p9	50.55	0.01	72.35	7.23	empty	
10	34.17	10.15	empty					p9	48.68	0.01	47	13.11	p10	
11			empty					p9	50.76	0.01	59.59	7.37	empty	
12			empty					empty					p10	
13			empty					empty					p10	
14			empty					p9	49.78	0.01	62.46	13.97	empty	
15			empty					p9	49.26	0.01	46.83	12.77	p10	
16			empty					p9	50.45	0.01	81.8	14.91	empty	
17			empty					empty					empty	
18			empty					empty					p10	
19			empty					p9	50.86	0.01	67.34	8.87	empty	
20			empty					p9	50.42	0	97.14	9.53	empty	
21			empty					empty					empty	
22	40.7	14.02	p8	48.25	0	85.29	13.7	p9	50.5	0.01	84.58	11.95	empty	
23														

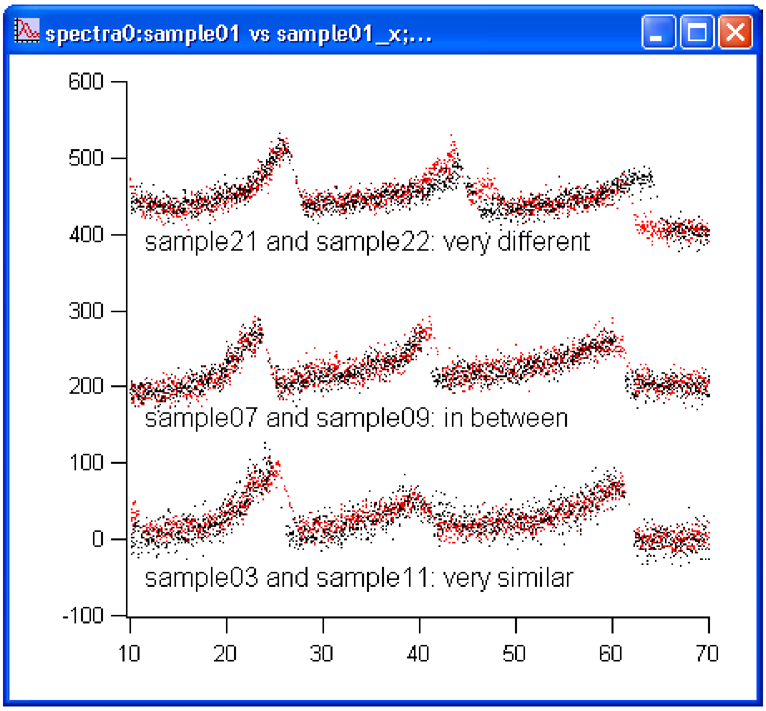
Sum=-470.41

3. The “peakinfo” file is generated during fitting and used during alignment. It contains information about where peaks start and end in the spectrum. You probably won't need to analyze it.
4. More interesting, yet more cryptic is the “tree” file. You can convert it into a postscript file using den.exe (http://odur.let.rug.nl/~kleiweg/clustering/clustering.html): Get a command prompt or terminal, and type “den.exe sample.tree.txt > sample.tree.ps”. You might need

to adjust paths. Now, look at it in a postscript viewer like GhostView.



5. The tree contains the following information: which two spectra or cluster of spectra were merged at which distance. The way distance is measured in this plot, 0 would be the largest distance (no similarity). In this plot, sample03 and sample11 are the most similar spectra, having a merge distance of about -0.2.



6. Now you can understand how fs_kit distributed the spectra into cluster. By default, it cuts the tree at the average plus one standard deviation of the merge distances. Here, this ends up being -0.147. When the tree is cut, 4 new trees get formed: Two tiny ones with just sample21 and sample22, a small one containing the cluster of sample20 and sample23, and the remaining big cluster. Those trees with just one spectrum are put in the “0” file, the other clusters are put in the files with higher numbers.