Raven Tutorial

Raven was created by Tom Aharoni & Prof. Amir Goldbourt

School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 6997801, Tel Aviv, Israel

For technical support please contact us at tomaharoni88@gmail.com or amirgo@tauex.tau.ac.il.

Our website: http://kuwari.tau.ac.il/index.html

Recent version: 2.0, 28-August-2018.

Table of Contents:

1. The structure and operation of Raven

Raven was programmed by utilizing the ANSI C integrated development environment of LabWindows\CVI. It contains two main modules; (i) the cross-peak assignment tool and (ii) the cross-peak analysis tool.

(i) <u>Cross peaks assignment tool:</u>

The goal of this tool is to assign a given list of cross-peaks, generated by peak picking a 2D or a 3D spectrum, based on an assignment table. A protein sequence, and if exists a structural coordinates file, are used as additional inputs. The peak matching process can utilize isotope labeling (pre-selected or custom-made), spectral resolution, and the maximal observable distance between atoms in an existing structure. The various matches are scored according to the difference between the input value and a functional distribution around the assignment table, and an indication to bias is presented through calibration histograms.

The resulting peak matches are sorted by the type of correlation (e.g. intra-residue, sequential, helical, long-range, or inter-subunit for biomolecular complexes) and are either ambiguous or non-ambiguous. It is therefore possible to limit the output to peaks having a single match, or a pre-defined number of matches (ambiguity level). The matched cross-peaks can be output according to correlation-type, or in the form of input for structure calculation programs such as Xplor-NIH. Currently, Raven allows to analyze cross peaks of the 2D ^{13}C-^{13}C and ^{15}N-^{13}C experiments and of the 3D ^{15}N-^{13}CO-^{13}CX and ^{15}N-^{13}CA-^{13}CX experiments. Given the advance in proton detection techniques, ^{1}H-based experiments are to be included, and handling sets of peaks will be generalized.

(ii) <u>Cross peaks analysis tool:</u>

This tool produces intensity curves, which provide information on dynamics or other modalities of choice. As input, the user provides a set of 2D planes, and a list of cross-peaks to be analyzed. These cross-peaks may be assigned by the cross peak assignment tool, or remain unassigned. It is also possible to visualize each curve individually in the same interface. Since the location of the cross peaks may change slightly between consecutive spectra, the software allows to search for the highest intensity for each cross peak, within a chosen error range.
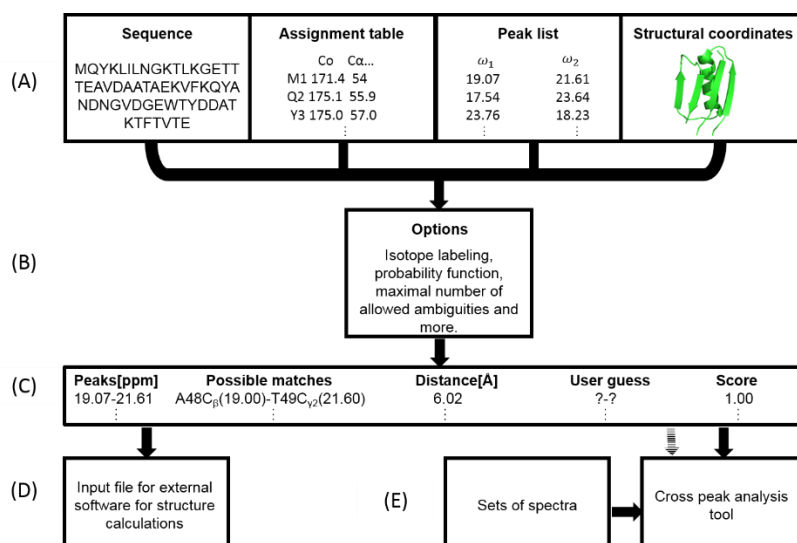
Flowchart of the Raven algorithm. The four input files are indicated in (a) and include the sequence, the assignment table, the peak list to be assigned, and potentially a pdb file. (b) This input is filtered by different options. (c) Raven outputs a list of matched cross peaks and optional distances (if a structure is provided). (d) Output data as structure restraints. (e) Output to the Cross peaks Analysis tool. The protein image is for GB1 (pdb file: 2JSV).

2. Setup

Requirements: Windows 7/8/10.

Run the setup file "Setup.exe" and follow the instructions.

Note: Installation in the \program files\ folder is not recommended. This folder is somewhat protected by the system, hence, the installation might failed and output files that were saved in this folder will not work properly.

In case Raven was installed in this folder or other protected folder, one of the following approaches could be used:

1. Reinstall Raven in another folder.
2. Run Raven as administrator (by right-clicking the shortcut icon).
3. Make sure to save your results in another folder, which is not protected by the system, for instance: "C:\Raven\Protein Results\".

In case the installation failed, or do not progress for a very long time:

1. Run the setup file as administrator and reinstall Raven in another folder which is not protected by the system, for instance D:\Raven\.
2. Delete the temp files on Windows directory:
2.1 Press Windows Key + R (or start and then search for %temp% directory) and enter %temp%.
2.2 Delete all files from the %temp% directory.
2.3 Reinstall Raven.

In case you are not able to install or update Raven, try to uninstall "National instruments Software" and reinstall Raven.

In case you still experience problems, please contact us at tomaharoni88@gmail.com or amirgo@tauex.tau.ac.il.

You can run Raven by using the shortcut created on your desktop.

3. Panels overview:

Raven launches with the following cross peaks assignment tool, which consist of three tabs. The tab "Load Files" is designed for loading 4 input files and to configure several parameters during the loading step. The tab "Output configuration & Run" is designed for configuring the output files and to run the assignment process. The third tab is "Summary Data" which contains a graphical summary of the output data.
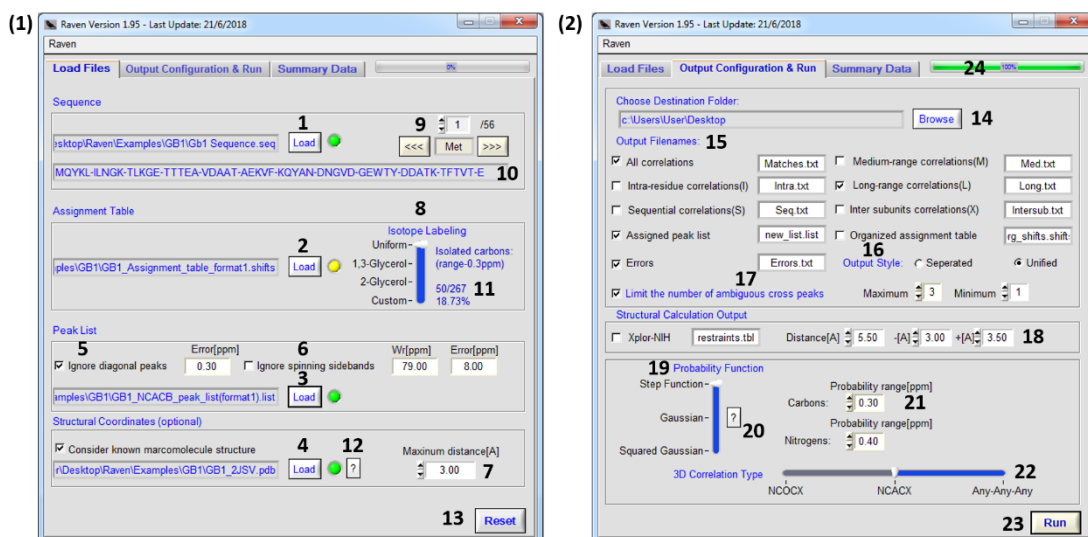
3.1 The "Load Files" tab:

Items #1, #2, #3 and #4 are buttons that are used for loading the sequence (4.1.1), the assignment table (4.1.2), the peak list (4.1.3) and the atomic coordinates (4.1.4) (optional) files respectively and the green "lights" indicate that the loading processes were successful. The yellow light is displayed when there one or more outliers in the loaded assignment table. Outliers have assigned chemical shift that are larger than 3 standard deviations of the average chemical shifts reported in the BMRB database.

Items #5-#8 allow the user to filter out peak matches that correspond to diagonal signals (#5), to spinning sidebands (#6), to contacts larger than a certain cutoff defined by the structure, if loaded (#7), and to contacts that are forbidden by the labeling scheme (#8, see also 3.4).

Items #9-#10 display the protein sequence.

Item #11 displays how disperse the ^{13}C spectrum is by indicating the number of ^{13}C shifts that are isolated with 0.3 ppm. This feature is not used by Raven for any type of calculation but provides general information to the user that gives an estimation of data complexity.

Item #12 displays (by clicking the '?' mark) a python script that can translate a pdb file to a special atomic coordinates file (*.rav) by executing it with the software Pymol. The format of the '*.rav' files is described in 4.1.4.2. Loading a pdb file directly is possible as well.

Item #13 is a 'Reset' button, which allows to remove all the data from the allocated memory in order to reload the files. Note that the sequence and assignment table can be uploaded once and not replaced unless the 'reset' is used.

3.2 The "Output Configuration & Run" tab:

Item #14 defines the output directory.

Items #15-#16 control the output content (Detailed in section 4.3):

#15 is a feature that allows the user to control which files would be created according to the type of the matched correlation. #16 represents an option to choose between two different output styles.

Item #17 is an option to control the number of ambiguous cross peaks, up to which an output result is given. This option makes the output files simpler and more informative. It also allows to simplify the data given to other analysis software.

Item #18 configures and outputs the results to files that are suitable as input for external software for structure calculations such as Xplor-NIH. This file consists of both non-ambiguous and ambiguous constraints, depending on the choice in #17.

Item #19 is a vertical level slide, which allows the user to choose between different types of probability functions, as demonstrated and explained by clicking item #20 (see also Section 4.2). The items in #21 allow to define the width of the distribution for nitrogens and carbons.

Item #22 appears when 3D correlations are loaded and allows to choose a selective filter for NCOCX and NCACX ssNMR experiments. This filter is selective towards single bond N-Cα or N-CO transfer, hence, each cross peak match that does not fit this pattern will be eliminated from the match. The option Any-Any-Any switches off the filter.

Item #23, the "RUN" button, starts the cross-peak assignment process, inclusive of all the above options. The progress bar #24 visualizes the progression of Raven operation.

3.3 The "Summary Data" tab:
The tab "Summary Data" represents portions of the output information graphically.

3.3.1 Cross Peaks Summary:
Text 1: Indicates how many cross peaks were loaded successfully. Diagonal peaks and spinning side bands are not included if filtered as described in 3.1 items #5,#6.
Text 2: Indicates how many cross peaks have at least one possible match. It does not include cases for which all the possible matches were filtered by distance.
Text 3/red dots in the graph: Unmatched cross peaks, and cross peaks in which all the possible matches were filtered by distance.
Text 4/magenta dots in the graph: Cross peaks that were filtered according to the chosen ambiguity level. This could be controlled by the maximum and minimum option in the ambiguity filter (Item #17, "Output Configuration and Run" tab).
Text 5/blue dots in the graph: Indicates the number of non-ambiguous cross peaks. The minimum value of the ambiguity filter must be set to 1 otherwise the result is zero.
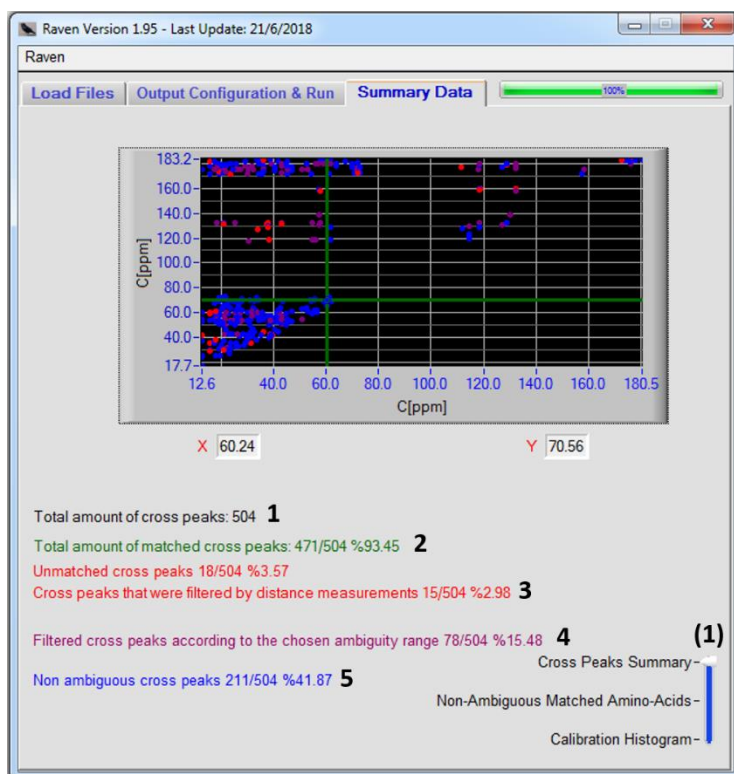
3.3.2 Non-Ambiguous Matched Amino-Acids:

This feature presents a graphical summary of the amino acids that participate in non-ambiguous matches.
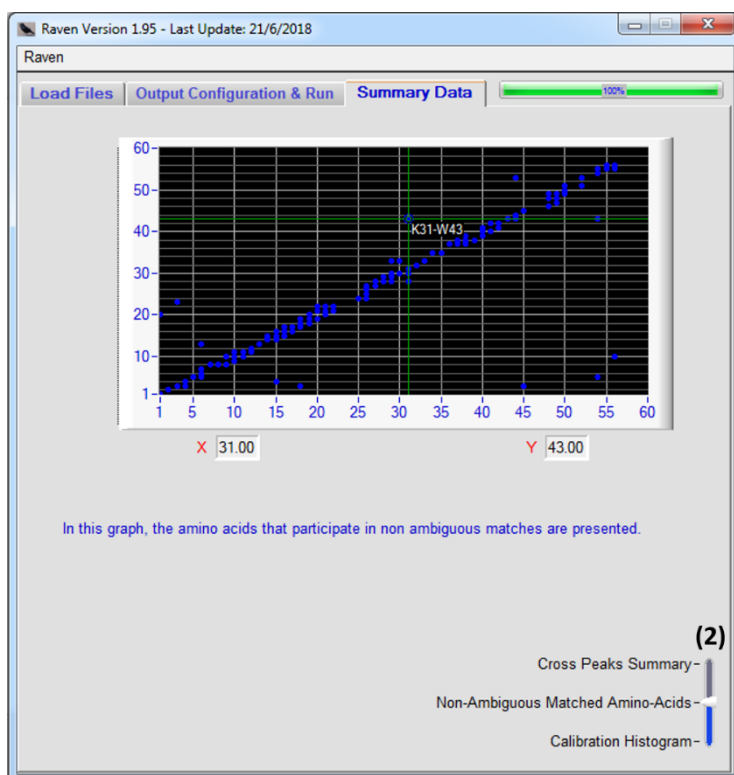
3.3.3 Calibration Histogram:

The two histograms indicate the deviation between the average value of the absolute difference between the chemical shift in the experimental data and the matched chemical shift in the assignment table for carbons (#6) and nitrogens (#7). It works

only when the Step function is chosen (Item #19, "Output Configuration and Run" tab) as the probability function. It is therefore an indicator of wrong referencing and may be used to correct the data. For example, and average value of +0.1 ppm for nitrogens may indicate that the experimental values of nitrogens should be recalibrated by -0.1 ppm.

3.4 Custom isotope labeling tool:
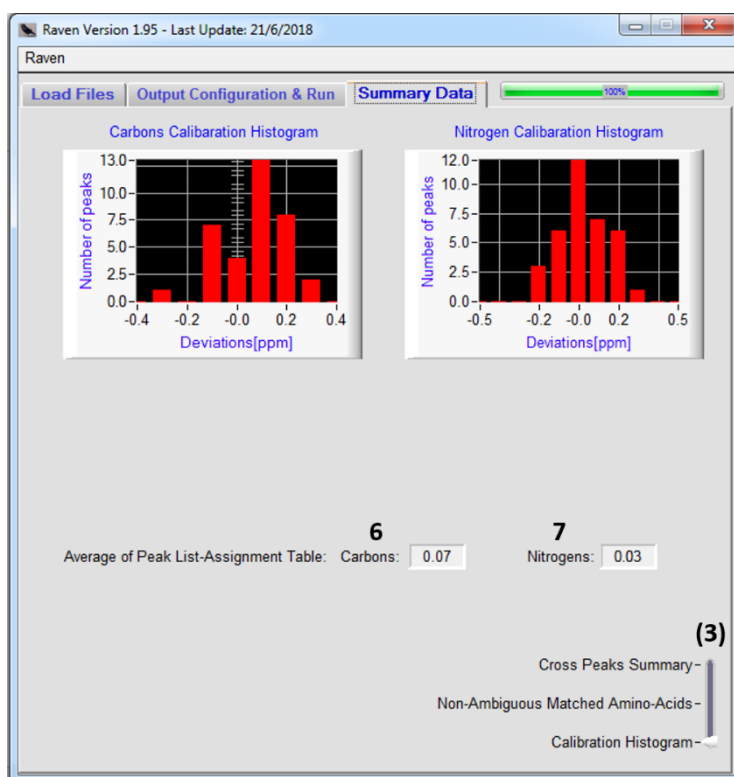
The tool is started from the dropdown menu by selecting 'Raven'→Isotope labeling file generator'. The tool is demonstrated below for the choice of 1,3-^{13}C-glycerol. This feature allows to define custom isotope labeling or view the present choice of labeling. The atoms are represented by the different columns and the 20 amino acids are represented by the rows. The color could be changed in one click, between red indicating unlabeled atoms and blue indicating labeled atoms. Entire column and row could be changed by clicking an atom or an amino acid. Unlabeled atoms will not be considered during the cross peaks assignment process.

By right clicking on a square the user can determine a probability for an atom to be labeled, which will be marked in green. In the current version of Raven (2.0, August 2018), this option is cosmetic and probabilities that are higher than 0 will be considered at the cross peaks assignment process. In the "Options" menu bar the user may inverse the isotope labeling (red squares become blue and vice versa), save a file and edit a saved file. The saved ".ilf" file could be used in the custom isotope labeling option, which is located in the vertical level slide of isotope labeling (configure before the assignment table file is loaded, item #8, "Load Files" tab).

Raven Version 1.95 – Custom Isotope Labeling Tool (*.ilf) — Options

	C	Ca	Cb	Cg	Cg1	Cg2	Cd	Cd1	Cd2	Ce	Ce1	Ce2	Ce3	Cz	Cz2	Cz3	Ch2	N	Nd1	Nd2	Ne	Ne1	Ne2	Nz	Nh1	Nh2
ALA																										
ARG	0.40	0.80	0.60																							
ASP	0.50	0.67	0.83	0.33																						
ASN	0.50	0.67	0.83	0.33																						
CYS																										
GLU	0.40	0.80	0.60																							
GLN	0.40	0.80	0.60																							
GLY																										
HIS				0.50						0.50																
ILE	0.50	0.67			0.83			0.33																		
LEU																										
LYS	0.75	0.42	0.92	0.33			0.83			0.42																
MET	0.50	0.67	0.83	0.33																						
PHE																										
PRO	0.40	0.80	0.60																							
SER																										
THR	0.50	0.67	0.83			0.33																				
TRP			0.50																							
TYR																										
VAL																										

Unlabeled   Scrambled(Right Click)   Labeled
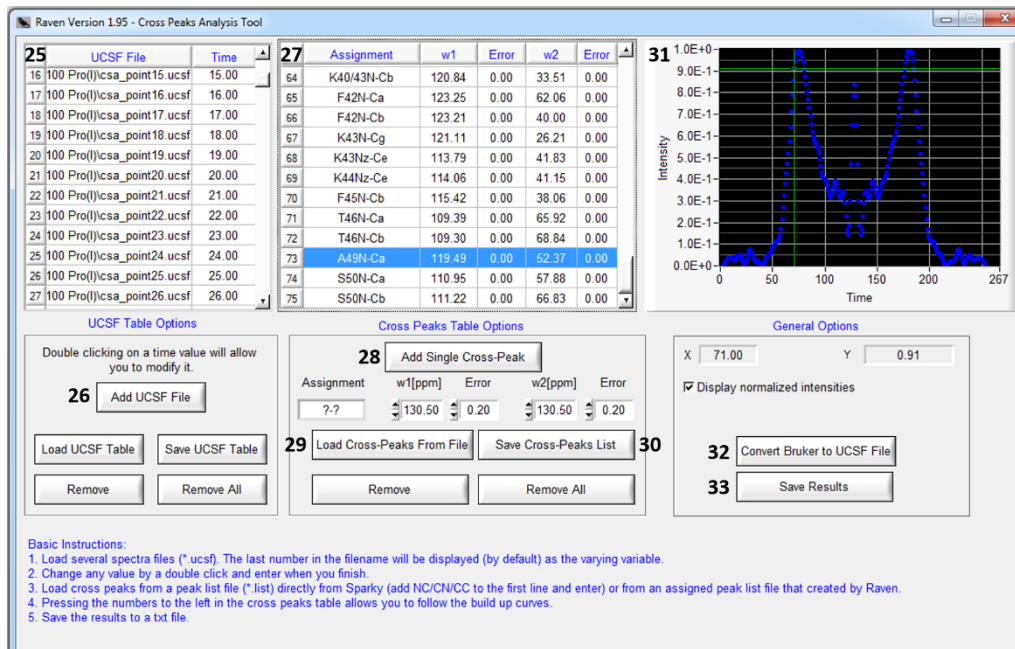
3.5 Cross peaks analysis tool:

The tool is started from dropdown menu by selecting 'Raven'→'Cross peak analysis tool'. The operating window is shown in the figure below.

After loading several UCSF files (output of spectra generated by the NMRPipe processing software), each corresponding to an experiment acquired at a different condition, the Cross Peaks Analysis tool allows to analyze automatically and simultaneously the intensities of all cross-peaks across the various spectra.

Raven Version 1.95 – Cross Peaks Analysis Tool

#25 UCSF File / Time
	UCSF File	Time
16	100 Pro(I)\csa_point15.ucsf	15.00
17	100 Pro(I)\csa_point16.ucsf	16.00
18	100 Pro(I)\csa_point17.ucsf	17.00
19	100 Pro(I)\csa_point18.ucsf	18.00
20	100 Pro(I)\csa_point19.ucsf	19.00
21	100 Pro(I)\csa_point20.ucsf	20.00
22	100 Pro(I)\csa_point21.ucsf	21.00
23	100 Pro(I)\csa_point22.ucsf	22.00
24	100 Pro(I)\csa_point23.ucsf	23.00
25	100 Pro(I)\csa_point24.ucsf	24.00
26	100 Pro(I)\csa_point25.ucsf	25.00
27	100 Pro(I)\csa_point26.ucsf	26.00

#27
	Assignment	w1	Error	w2	Error
64	K40/43N-Cb	120.84	0.00	33.51	0.00
65	F42N-Ca	123.25	0.00	62.06	0.00
66	F42N-Cb	123.21	0.00	40.00	0.00
67	K43N-Cg	121.11	0.00	26.21	0.00
68	K43Nz-Ce	113.79	0.00	41.83	0.00
69	K44Nz-Ce	114.06	0.00	41.15	0.00
70	F45N-Cb	115.42	0.00	38.06	0.00
71	T46N-Ca	109.39	0.00	65.92	0.00
72	T46N-Cb	109.30	0.00	68.84	0.00
73	A49N-Ca	119.49	0.00	52.37	0.00
74	S50N-Ca	110.95	0.00	57.88	0.00
75	S50N-Cb	111.22	0.00	66.83	0.00

UCSF Table Options

Double clicking on a time value will allow you to modify it.

#26 Add UCSF File

Load UCSF Table   Save UCSF Table
Remove   Remove All

Cross Peaks Table Options

#28 Add Single Cross-Peak

Assignment   w1[ppm]   Error   w2[ppm]   Error
?-?   130.50   0.20   130.50   0.20

#29 Load Cross-Peaks From File   Save Cross-Peaks List #30
Remove   Remove All

General Options

X 71.00   Y 0.91
☑ Display normalized intensities

#32 Convert Bruker to UCSF File
#33 Save Results

Basic Instructions:
1. Load several spectra files (*.ucsf). The last number in the filename will be displayed (by default) as the varying variable.
2. Change any value by a double click and enter when you finish.
3. Load cross peaks from a peak list file (*.list) directly from Sparky (add NC/CN/CC to the first line and enter) or from an assigned peak list file that created by Raven.
4. Pressing the numbers to the left in the cross peaks table allows you to follow the build up curves.
5. Save the results to a txt file.

#25: A list of binary data files in 'ucsf' format that contain several 2D spectra differing in intensity. The mixing time, or any other varying variable of choice, should be manually inserted by double clicking a value, or incorporated automatically by the software from the last number in the file name (e.g. DARR100.ucsf, csa_point9.ucsf).
#26: Select and upload the list shown in #25 (multi-selection is available and recommended). The selection list can be saved and uploaded for future use by using "Save UCSF Table" and "Load UCSF Table".

#27: A list of cross peaks to analyze across the ucsf spectra that appear in #25. Raven searches for the maximum intensity around the cross peak in a range that is defined for each dimension individually. For instance, errors of 0.1 ppm and 0.2 ppm will form a rectangle of 0.2 ppm x 0.4 ppm around an (x,y) crosspeak (e.g. from x-0.1 to x+0.1 and from y-0.2 to y+0.2) and the maximum intensity within this rectangle will be obtained. Any value in this list could be changed by double clicking on it.

#28: Manually load a single cross peak according to its two shifts and the corresponding error.

#29: Load a list of cross peaks generated by the cross peaks assignment tool (by checking "assigned peak list", #15), or by sparky. The format is as described in Section 4.1.3.

#30: Save the list of cross peaks.

#31: Display the intensity curve of a chosen cross peak (marked in a blue rectangle in #26) across the ucsf files appearing in #25. The red dots are intensities of the cross peak A49N-Cα in 256 ^{15}N-^{13}C planes (the symmetry-recoupled CSA lineshape of A49N). The graph is displayed by clicking a row number in #27.

#32: A tool to convert Bruker files to ucsf format (uses bruk2ucsf execution file from the Sparky software)

#33: Save all the cross peaks and their corresponding intensity curves to a file. A demonstration to a partial output appears below:

Assigned cross peak	Chemical shifts		Time (or any other variable)					
			0	1	2	3	4	5
D5N-Ca	121.36	52.06	1931322	2126112	2601209	3089985	3317215	3150656
D5N-Cb	121.55	41.51	1174550	1159043	1106046	1001396	835995.8	618513.4
P6N-Ca	138.27	64.16	1308432	1280195	1193797	1047938	850366.5	625153.5

(Intensities)

4. Cross Peaks Assignment

4.1 The input files:

Raven requires 4 input files from the user:

4.1.1 A sequence file (*.seq):

The sequence file contains the amino acid sequence of the probed macromolecule. The purpose is to verify whether the amino acids are written and placed correctly at the other inputs. This file should contain one long line of the sequence of amino acids in one letter code. The sequence files of the GB1 protein and the Y21M fd bacteriophage coat protein are given in the example files here:

```
…\Raven\Examples\GB1\GB1.seq
…\Raven\Examples\Y21M fd Bacteriophage\Y21M.seq
```

Y21M.seq: `AEGDDPAKAAFDSLQASATEMIGYAWAMVVVIVGATIGIKLFKKFTSKAS`

4.1.2 An assignment table file (*.shifts):

This file contains the chemical shift of each atom (or all the available atoms) in the corresponding amino acid. This input file can be obtained directly from the software Sparky (or exported in Sparky format by the file convertor of CCPN). An incomplete assignment table can be loaded, where amino acids are missing entirely, or particular atoms are missing, or chemical shifts are missing.

Raven supports 2 assignment table formats as shown here:

```
…\Raven\Examples\GB1\GB1_Assignment_table_format1.shifts
…\Raven\Examples\GB1\GB1_Assignment_table_format2.shifts
```

Format1: Sparky Assignment Table (can be obtained by the command 'tb' on Sparky):

	C	Ca	Cb	Cg	Cg1	Cg2
M1	171.4	54.24	32.28	30.18	-	-
Q2	175	55.75	30.36	35.29	-	-
Y3	174.9	57.2	43.77	128.7	-	-

The columns represent the atoms, the rows represent the amino acids and their location on the sequence and the matrix is filled with the corresponding chemical shifts. Atoms that are not assigned or do not exist **must be marked with the character "-"**. It is possible to use the character "*" for unassigned but existing atoms (Raven can also indicate the unassigned atoms for you).

Format2: Resonance List – The original version of this format, which contains at least 5 columns can be obtained by the command 'rl' on Sparky, hence a removal of unnecessary columns by Microsoft Excel or any other software should be done for proper reading by Raven.

Group	Atom	Shift
M1	C	171.4
M1	CA	54.24
M1	CB	33.28
Q2	C	175.0
Q2	CA	55.71

The first line must be: "Group Atom Shift".
The first column represent the amino acids and their location on the sequence; the second column represent the atoms; the third column contain the corresponding chemical shifts. Additional columns and non-assigned atoms must not be present in this file.

Technical note: In case that an assignment table was loaded in format2, Raven would convert it to format 1 and save it as "\My Documents\Raven_assignmentcopy.txt". The converted file in format1 will be loaded instead.

Amino Acids:

The amino acids are represented by the rows in one letter-code:

A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V.

Alternatively a three-letter-code can also be used (or combination of the two formats):

Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val.

Atoms:

Only the following atoms could be written in the columns:

C,Ca,Cb,Cg,Cg1,Cg2,Cd,Cd1,Cd2,Ce,Ce1,Ce2,Ce3,Cz,Cz2,Cz3,Ch,Ch2,N,Nd1,Nd2,Ne,Ne1,Ne2,Nz,Nh1,Nh2.

For placing the atoms correctly, one could be assisted by the bmrb website:
http://www.bmrb.wisc.edu/ref_info/statful.htm

This file contains a list of cross peaks (with a possible user-proposed identity) from experimental spectra. The file is identical to the peak list output of the software Sparky (or could be exported in Sparky format by the file convertor of CCPN) with the exception of the first line, which determines the number of dimensions and the atom type for each column.

2D example

```
CN

Assignment w1 w2

G3Ca-N    45.3    111.2
D4Ca-N    54.4    120.3
D5Ca-N    51.9    121.5
?-?   64.2    138.2
?-?   54.9    119.5
?-?   59.8    119.3
```

3D example

```
CNC
Assignment      w1        w2        w3

?-?-?          178.069   117.998   58.820
?-?-?          178.966   122.995   54.967
?-?-?          175.848   111.526   38.077
?-?-?          178.074   110.065   177.944
?-?-?          177.573   124.388   60.821
?-?-?          177.589   117.283   61.669
Ala1C'-Glu2N-Ala1C'   174.165   120.135   174.058
Glu2C'-Gly3N-Glu2C'   177.177   111.285   177.025
Asp4C'-Asp5N-Asp4C'   175.958   122.033   176.052
```

The first line should be "CC" for homonuclear ^{13}C-^{13}C correlations, "NC"/"CN" (according to the column) for heteronuclear ^{15}N-^{13}C/^{13}C-^{15}N correlations. For 3D NCOCX/NCACX experiments the first line should be "CNC" so that the first chemical shift column contains the chemical shifts of the correlated Co (or Ca) atom, the second column contains the chemical shifts of the correlated nitrogens, and the third column is the acquisition dimension. The file can contain the user's guess (or some symbol such as '?-?', which must be a character or a word without spaces, and the correlated chemical shifts. Note that the user's guess will be displayed in the output files, but it does not have any impact on the process of assigning cross peaks.

Examples to peak list files can be found here:

```
…\Raven\Examples\GB1\GB1_CC_peaklist_DARR_100ms_13KHz.list
…\Raven\Examples\GB1\GB1_NCACB_peak_list(format1).list
…\Raven\Examples\Y21M\Y21M_Gly_CC_peak_list.list
…\Raven\Examples\Y21M\Y21M_NCO_peak_list.list
…\Raven\Examples\Y21M\Y21M_CAN_peak_list.list
```

4.1.4 An atomic coordinates file (Optional, *.rav or *.pdb file):

This file contains the 3D structural coordinates of each atom, if available. In case of complex macromolecules, the data may also include coordinates of atoms that are placed in multiple subunits.

Loading a pdb file:

The standard pdb file must be written in a proper manner. Raven will skip any lines until reaching the first line starting with the word "ATOM". Proper lines should be written as follows:

The word "ATOM" | An integer | The atom name | The amino acid | The subunit name, must be in a one letter code | The amino acid location | Structural coordinates(x,y,z) | Any other text

Lines that begin with the word "TER" or contains an atom which is not a carbon or a nitrogen will be ignored. In order to get proper results, each assigned atom must have coordinates and the subunits must be identical, so that each subunit must contain the same atoms.

In case that there are several subunits, the distances will be measured only from the subunit that was written first (at the top of the file) and within this subunit. It is therefore important to put the subunit of interest written at the top of the file. Moreover, in case that several models exist, only the first model will be considered, hence, the model of interest must also be written at the top of the file.

Loading a rav file:

Instead of a pdb file, one may also create and load a rav file by copying a python script to the software Pymol. Any non-vital data is eliminated from the file and it has the advantage of flexibility (by editing the python code) to choose any model of interest or even to output the average location of the atoms in all the existing models. In order to create a *.rav file press the question mark at the "Structural Coordinates" section in the "Load Files" tab. A text box that contains the script will be opened. Change the filename "my_atoms_coords.rav" and copy the script as is to Pymol. The rav file will be created and ready to be loaded. Examples to rav files can be found here:

```
…\Raven\Examples\GB1\GB1_atomic_coordinates.rav
…\Raven\Examples\GB1\Y21M fd Bacteriophage\y21m atomic coordinates.rav
```

The format of the rav file is as follows:

Number of subunits | The number of carbons in each subunit | The number of nitrogens in each subunit

Begin with the carbons within the subunit of interest. The subunit name must be written in one letter

```
35 237 57
Q 1 CA -18.6979999542 22.4839992523 147.604003906
Q 1 C -17.9230003357 23.7080001831 148.136001587
Q 1 CB -19.9650001526 22.2520008087 148.417007446
Q 2 CA -15.9300003052 24.5100002289 149.389007568
Q 2 C -14.4600000381 24.0699996948 149.319000244
Q 2 CB -16.343000412 24.8269996643 150.839996338
Q 2 CG -16.2409992218 23.6439990997 151.817001343
Q 2 CD -17.0970001221 23.8150005341 153.06300354
Q 3 CA -12.123000145 24.7150001526 149.005004883
Q 3 C -11.732000351 24.3050003052 147.57699585
```

The number of the amino acid on the sequence | The name of the atom | Structural coordinates (x,y,z)

The file is ordered as follows: Carbons of all subunits are positioned first followed by the nitrogens in the same order of subunits. In the case above (fd-y21m), 237

carbons are written for each of the 35 subunits (A, B, C…), and then 57 nitrogens are written for each of the 35 subunits (A, B, C…).

4.2 Peak match probability:

<u>Note:</u> This feature is graphically demonstrated in Raven. (The question mark, item #20, "Output Configuration & Run" tab).

The default probability function is the step function, which is defined by $|\delta_{AT} - \delta_{PL}| \leq$ probability range. δ_{AT} is a shift in the assignment table. δ_{PL} is a shift in the peak list. Raven calculates whether the absolute of the difference between each chemical shift in the peak list file and each chemical shift with the same atom type from the assignment table file is smaller than the probability range, which was chosen by the user. For a step function, the calculated probability value for each match can only be 1 or 0.

Raven offers additional probability functions such as a Gaussian Function and a squared Gaussian function. These functions allow to score the peak matching corresponding to a chosen half width at half maximum and a minimum intensity cutoff.

The Gaussian function is defined according to $e^{-\frac{\ln(2)}{Half\,FWHM^2}(\delta_{AT} - \delta_{PL})^2} \geq$ cutoff .

The squared Gaussian function is given by $e^{-\frac{\ln(2)}{Half\,FWHM^4}(\delta_{AT} - \delta_{PL})^4} \geq$ cutoff. It offers a probability decay with a sharper slope and closer to the edges. The output of these two functions is presented only if the match for each chemical shift has a probability that is larger than the chosen cutoff (marker in the figure as a straight blue line). The output to the file is the multiplication of these probabilities.
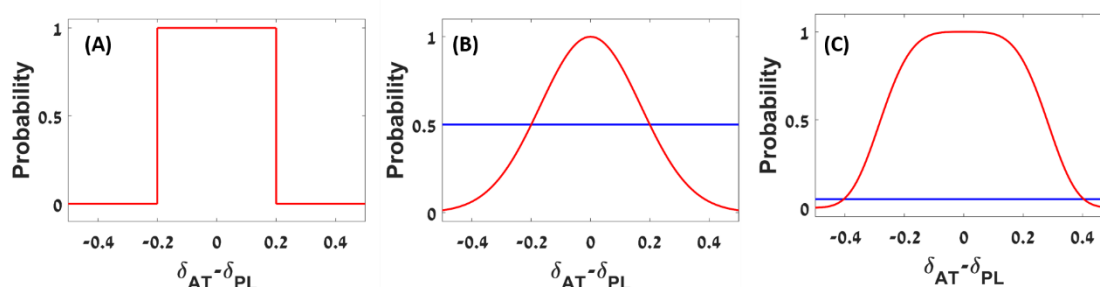
Illustration of (A) Step function; (B) Gaussian function; (C) Squared Gaussian function.

Based on the assumption that higher values of $|\delta_{AT} - \delta_{PL}|$ are less probable for assignments than low values, it is hypothesized that (after correcting for any referencing bias) scoring the peak matching leads to better cross peaks assignments by reducing ambiguities.

4.3 The output files:

Correlations types in the following cartoon of two subunits are defined as follows:
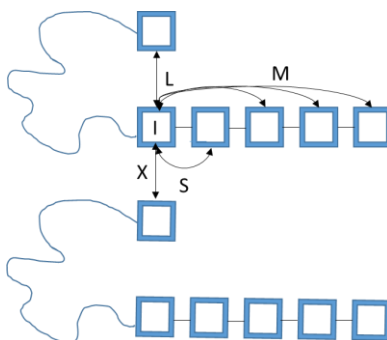
Intra-residue correlations (marked as: "I"): correlations between atoms within the same amino acid - $i \leftrightarrow i$.
Sequential correlations (marked as "S"): correlations between atoms that are located in a sequential amino acids - $i \leftrightarrow i \pm 1$.
Medium range correlations (M): $i \leftrightarrow i \pm x; x = 2 - 4$.
Long range correlations (marked as "L"): $i \leftrightarrow i \pm x; x > 4$.
Inter-reside correlations (marked as "X"): Any correlation in which two different subunits are involved. The distinction of long-range and inter-residue correlations is straightforward if the structure is given, however, in its absence, it is not always possible to differentiate between the two possibilities.

The output panel is shown below:

#1: <u>All correlations:</u> (default filename: Matches.txt) an output file that contains all the matched cross-peaks.
#2-#6: <u>Intra-residue(I)/Sequential(S)/Medium-range(M)/Long-range (L)/Inter-subunit(X):</u> Individual output files are generated to matches of the specific chosen correlation. Since peaks can match two or more correlation options, the "Overall" entry would declare if any other correlations exist for a particular match (see below).

<u>Output file format - Separated Style:</u>

Here each peak is reported individually.
Example:
```
 The peaks match for: 73.10-57.56 (User Guess: ?-?)
         Peaks              Type  Distance  Subunits  Probability
T44Cb(73.09)-W43Ca(57.49)    S      4.65      X-X         1
T44Cb(73.09)-Y45Ca(57.70)    S      4.51      X-X         1
Overall: 0I2S0M0L0X
```

<u>Output file format - Unified Style:</u>

Here all peaks are listed uniformly.
Example:

```
   Peaks           Match          User Guess Type Distance Subunits  Overall   Probability
38.70-33.36 E56Cg(38.64)-E56Cb(33.38)   ?-?    I     1.53    X-X    1I0S0M0L0X      1
69.79-43.22 T49Cb(69.96)-K50Ce(43.30)   ?-?    S     7.85    X-X    0I1S0M0L0X     0.82
```

The columns of the output file (separated by tab)**:**

<u>Peaks</u> – The cross peaks that was taken from the peak-list file.
<u>Match</u> – The possible matches generated by Raven.
<u>User Guess</u> – The user's guess that was taken from the peak-list file.
<u>Type</u> – The type of correlation as obtained by matching to the sequence and subunit letter (I/S/M/L/X).
<u>Distance:</u> The distance between the atoms taken from the structural coordinates file, if loaded. Otherwise, no distance column appears. In NCOCX/NCACX, the distance is measured between the carbons.
<u>Subunits:</u> The subunit name for each matched atom. For instance, Q-X means that the first atom is from subunit Q, and the second from subunit X.

<u>Overall:</u> A counter of matches for each type of correlation. For instance, "7I1S0M3L0X" means that there are 7 intra-residue matches, 1 sequential match and 3 long range matches. When such an entry appears in the 'seq.txt' files it informs the user that it is not unique since there are also intra-residue matches.

#7: This checkbox allows to create a new peak list in a sparky format, in which the non-ambiguous cross peaks are assigned. This list can be loaded to the Cross peaks Analysis tool (See 3.5).
#8: This check box allows to create an organized assignment table file, which considers the order of the atoms (Co, Ca, Cb, Cg…), the chosen isotope labeling (non-labeled atoms will be marked as "$") and indicates the atoms that do not have an assigned chemical shift (marked as "*"). This feature could assist in assigning chemical shifts by pointing on the remaining unassigned atoms.
#9: The errors output file contains information about unmatched cross peaks, matches that were filtered by distance measurements or cross peaks that were filtered according to their ambiguity level.

<u>Output file for structure calculation (standard Xplor-NIH format):</u>

Controls the standard output file for structural calculation – filename, the distance and the errors in [Å]. This output could also contain ambiguous cross peaks.
Example:

```
assign (( resid 55 and name Cb )) (( resid 55 and name Ca )) 5.50 3.00 3.50
assign (( resid 51 and name Cb )) (( resid 49 and name Ca )) 5.50 3.00 3.50
or (( resid 51 and name Cb )) (( resid 53 and name Ca ))
or (( resid 53 and name Cb )) (( resid 53 and name Ca ))
assign (( resid 51 and name Cb )) (( resid 51 and name Ca )) 5.50 3.00 3.50
or (( resid 53 and name Cb )) (( resid 51 and name Ca ))
```

Additional information about the software Xplor-NIH can be found here.

5. Troubleshooting:

Error Number	Corrective action
001/002	Correct the chemical shift in case of an error.
003/700/704	Re-install Raven.
111	Correct the sequence file accordingly
112	Raven is limited to a sequence of 500 amino acids, please contact us for help.
402	Correct the sequence file and/or the assignment table file accordingly.
100/101/102/500	Correct the assignment table file accordingly.
200/201	Correct the first line in the peak list file to be CC/CN/NC for 2D experiments, or CNC for a 3D experiment.
202	Edit the peak list in excel or any other software so that the first chemical shift column contain the Ca (or Co) chemical shifts, the second column contain the N chemical shifts and the last column contain the Cx chemical shifts. Write CNC and enter at the first line of the file.
212/213	The peak list is limited to 180,000 cross peaks. Divide the peak list file into several files and reload each of them individually.

599/600/601	One of the paths of the ucsf files is incorrect. Remove the ucsf files from the list and re-load them in the right path.
604/605	Make sure that all the cross peaks are within the correct chemical shift ranges of the ucsf files.
606	Load a proper ucsf file.
607	In this version of Raven only 2D ucsf can be loaded to the Cross Peaks Analysis tool. You may load 2D planes of a 3D experiment instead.
650	Only 2D CC/CN/NC files are allowed. You may load 2D planes of a 3D experiment instead.
701/702/703	Correct the atomic coordinates file (pdb/rav) accordingly.

6. Exercises:

Exercise 1: Load and display the GB1 protein sequence.

In the load files tab, press the Load button which is located in the <u>sequence section</u>. Select the following assignment table file using the loading panel:

`…\Raven\Examples\GB1\GB1.seq`

As shown below, in case of successful loading:
(1) A green light will appear to the right of the Load button.
(2) The location of the file will appear to the left of the Load button.
(3) The sequence of the T2Q GB1 protein will appear below the Load panel.

In addition, the navigate arrows can be used to display the amino acids one-by-one in a three letter code according to the sequence.
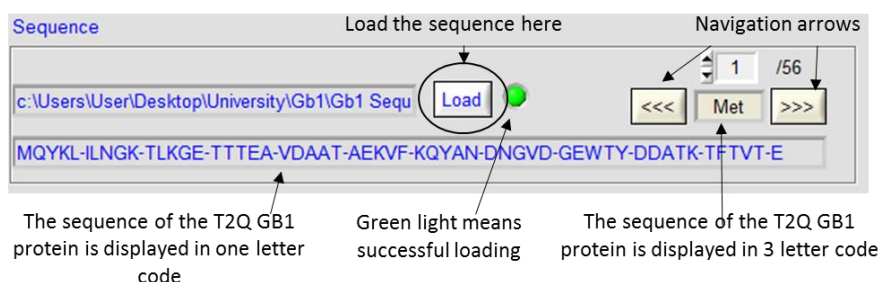
Exercise 2: Reveal an outlier in the GB1 assignment table.

In the GB1 protein the nitrogen in the N-terminus has a chemical shift of ~40 ppm since it is not connected to a carbonyl ($-NH_3^+$). Since currently we treat it as a regular amide nitrogen, a notification will appear. Outliers are important since they may indicate wrong assignments.

1. Press 'Reset'.
2. Load the sequence file: `…\Raven\Examples\GB1\GB1.seq`.
3. In the <u>assignment table</u> section (which is located in the load files tab) select the uniform isotope labeling option in the vertical bar.
4. In the assignment table section press the 'Load' button and select the following assignment table file using the loading panel:

`…\Raven\Examples\GB1\GB1_Assignment_table_format1.shifts`

You will receive the following alert:
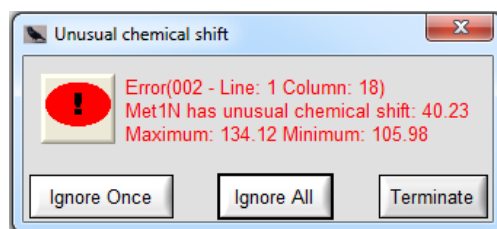
The message contains the following information:
Error number: 002 (for troubleshooting).
The location of the error in the assignment table: Line 1 and column 18.
The atom and the amino acid and their location on the sequence: Met1N
The unusual chemical shift of the related atom: 40.23 ppm
3 standard deviations differences from the average chemical shift:
Maximum 134.12 Minimum: 105.98
5. Since this outlier is a valid chemical shift, press 'Ignore Once' or 'Ignore All' (if you know that all your assignments are valid). As a result, a yellow light will be displayed. The yellow light means that the loaded assignment table contained outliers, but the user chose to ignore them. The program can proceed regularly.
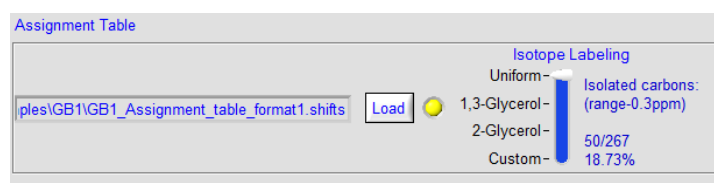
Exercise 3: Reveal common errors in an assignment table.

For technical programming reasons, the user could not change the selection of a sequence file or an assignment table file and the choice of isotope labeling scheme. The right way to do that is to press the Reset button, which is located in the right-bottom corner of the "Load Files" tab. This button allows to remove all the data from the allocated memory in order to reload the files.

1. In the right bottom of the Load Files tab press the button 'Reset', consequently the panel will return to its original state.
2. Create a copy file of the assignment table file:

`…\Raven\Examples\GB1\GB1_Assignment_table_format1.shifts`

You may use the following name GB1_Assignment_table_format1_copy.shifts (do not change the ".shifts" ending).
3. Open the copy and change one of the following:
(a) Change an assigned chemical shift to a very high number, for instance change the chemical shift of T17Ca from 60.32 ppm to 600.32 ppm.
(b) Change one of the amino acids so it will not match the sequence. For instance, change L5 (amino acid leucine located number #5 on the sequence) to K5 (Lys, 5).
(c) Change an atom name to a wrong one. For instance, change the atom "ce" to "ce4".
(d) Delete a "-" mark, and instead add a chemical shift to an atom that does not exist in the corresponding amino acid. For instance, add a chemical shift to the Cb atom of Glycine #9.

Save the changes (make sure you will be able to find the file) and exit the editor.
4. Load the sequence file GB1.seq.
5. In the "Isotope Labeling" vertical bar select the uniform isotope option.
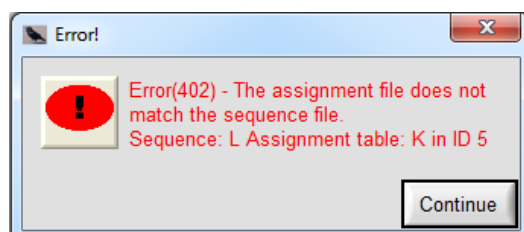6. Load the new assignment table file with the change you made.
At first, you will receive the same outlier notification as displayed in Exercise 2, press 'Ignore once', afterwards one of the following error notifications will appear according to the modification that you made:
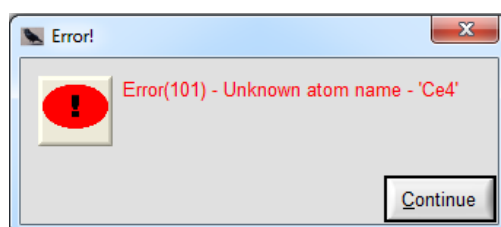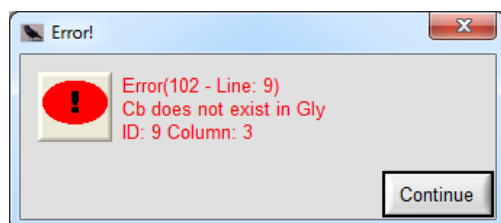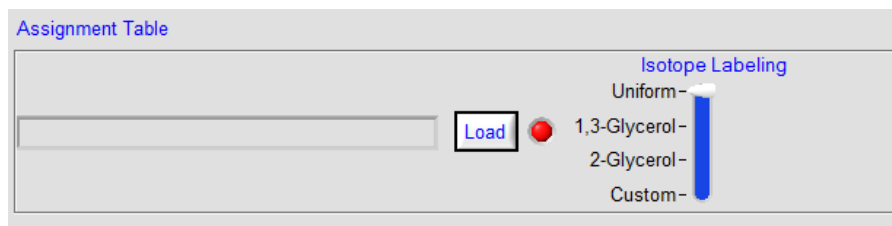
Case (a):
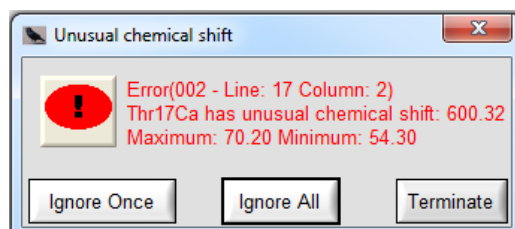
Press "Terminate".

Case (b):

Case (c):

Case (d):

A red light will appear in the assignment table section, which means that an error occurred and you will have to correct the assignment table accordingly. Since the loading was not complete, you can make the correction accordingly and reload the file (no need to press 'Reset').

Exercise 4: *Assign the non-ambiguous 2D ^{13}C-^{13}C cross peaks.*

1. Load the sequence file GB1.seq.
2. In the vertical bar of "Isotope Labeling" select the uniform isotope option.
3. Load the assignment table file:

`…\Raven\Examples\GB1\GB1_Assignment_table_format1.shifts`

4. You will get notification on an outlier, press 'Ignore Once' or 'Ignore All'.
5. In the peak list section select the option "Ignore diagonal peaks" with an error of 0.3 ppm. Deselect the option "Ignore spinning side bands". Press 'Load', and select the following assignment table file using the loading panel:

`…\Raven\Examples\GB1\GB1_CC_peaklist_DARR_100ms_13KHz.list`

A green light will appear to the right of the Load button.
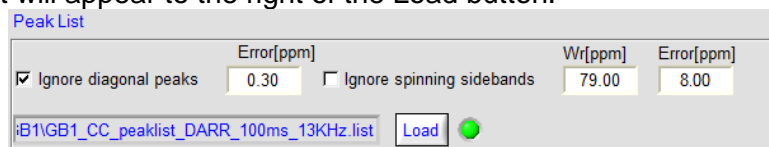
6. In the structural coordinates press the 'Load' button and use the loading panel to load one of the following files:

`…\Raven\Examples\GB1\GB1_2JSV.pdb`
`…\Raven\Examples\GB1\GB1_atomic_coordinates.rav`

7. A green light will appear to the right of the 'Load' button.
8. Set the maximum distance to be 8 Å and the checkbox of "Consider known macromolecule structure" on.
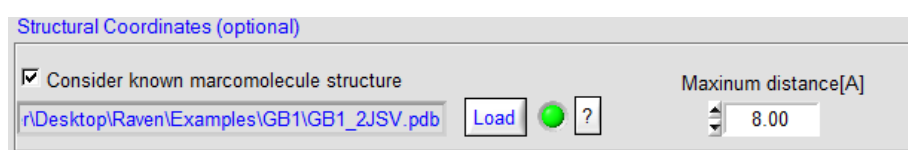
9. Click on the "Output Configuration & Run" tab
10. Press the 'Browse' button, go to a chosen output directory and press 'Done'. The location of the directory will be displayed in the box next to the 'Browse' button.
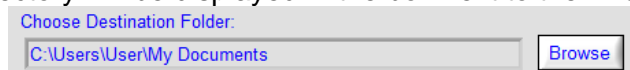
11. Select the output files you desire by turning the checkboxes on/off. You may also select their names.
12. Select the Output style to be 'Unified'.
13. Select the level of ambiguity to be maximum '1' and minimum '1'. It will therefore output only the non-ambiguous cross peaks.

14. Select the probability function to be "Step Function" and choose a probability range of 0.3 ppm. Press the question mark for graphic demonstration of how this function works.

15. In the bottom right corner press 'Run'.
16. Open the files in the chosen destination folder. In case you chose to output the long range correlations file, this output file contains only the matched non-ambiguous long range correlations.
17. Change the ambiguity level and press 'Run' again to observe the ambiguous cross peaks.

Exercise 5: *Assign 2D ^{15}N-^{13}CO cross peaks.*

1. Press 'Reset' if an assignment table or sequence file is already loaded.
2. In the vertical bar of "Isotope Labeling" select the uniform isotope option.
3. Load the sequence file located at …\Raven\Examples\Y21M fd Bacteriophage\Y21M.seq.
4. In the same folder, load the assignment table file y21m-assignment.shifts.
5. Load the peak list file Y21M_NCO_peak_list.list which contains 2D correlations of ^{15}N-^{13}CO.
6. Load the structural coordinates file "y21m atomic coordinates.rav".
7. Set the maximum distance to 3 Å.
8. In the "Output Configuration & Run" tab make sure that the checkbox responsible for the output of sequential correlations is on.
9. Press the 'Browse' button, go to a chosen output directory and press 'Done'.
10. Do not limit the number of ambiguous peaks.
11. Choose a step function in a probability range of 0.3 ppm for carbons and 0.4 ppm for nitrogens.
12. Press 'Run'.
13. Open the output file of the sequential correlations; the file contains all the matches.

Exercise 6: *Cross peak analysis of ^{13}C-^{13}C buildup curves*

1. Press 'Reset' if an assignment table or sequence file is already loaded.
2. In the vertical bar of "Isotope Labeling" select the uniform isotope option.
3. Load the sequence file located at …\Raven\Examples\Y21M fd Bacteriophage\Y21M.seq.

4. In the same folder, load the assignment table file y21m-assignment.shifts.
5. In the same folder, load the following peak list file y21m_ILE_example.list. This file contains ^{13}C-^{13}C cross peaks of the ILE region of fd-y21m that were taken from a DARR experiment.
6. Load the structural coordinates file "y21m atomic coordinates.rav".
7. Set the maximum distance to 8Å.
8. In the "Output Configuration & Run" tab make sure that the "Assigned peak list" checkbox is on. This output will contain the cross peaks that was assigned by Raven.
9. Press the 'Browse' button, go to a chosen output directory and press 'Done'.
10. Do not limit the number of ambiguous peaks.
11. Choose a step function in a probability range of 0.3 ppm.
12. Press 'Run'.
13. Open the output files and observe the results.
14. In the menu bar select 'Raven'→'Cross peak analysis tool'.
15. Press 'Add UCSF File' (#26) and load the following files (one may use ctrl+A in order to select all the files):

```
…\Raven\Examples\Y21M fd Bacteriophage\UCSF\DARR_ILE_10ms.ucsf
…\Raven\Examples\Y21M fd Bacteriophage\UCSF\DARR_ILE_100ms.ucsf
…\Raven\Examples\Y21M fd Bacteriophage\UCSF\DARR_ILE_250ms.ucsf
```

15. Press 'Load Cross Peaks From File' (#29) and load the assigned peak list that was gained from step 12.
16. Select an error of 0 ppm in both dimensions and Press 'OK'.
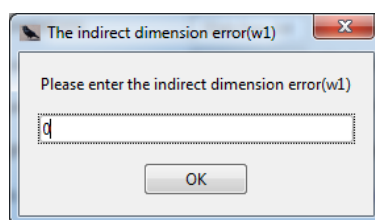
17. In the cross peaks table (#27), press the numbers to the left to visualize the intensities in the graph. In this stage, the panel should look like this:
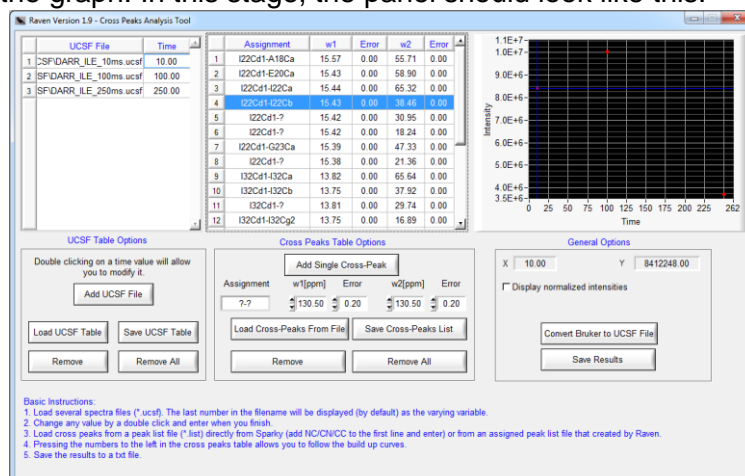
18. Press the 'Save Results' (#33) button in order to save the results to a text file.
19. Open the text file and observe the results.