nature protocols

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41596-018-0091-9

In the format provided by the authors and unedited.

Application of anisotropic NMR parameters to the confirmation of molecular structure

Yizhou Liu[1,5*], Armando Navarro-Vázquez[2*], Roberto R. Gil[3], Christian Griesinger[4], Gary E. Martin[1,6] and R. Thomas Williamson[1,7]

[1]Structure Elucidation Group, Process and Analytical Research and Development, Merck Research Laboratories, Rahway, NJ, USA. [2]Departamento de Química Fundamental, CCEN, Universidade Federal de Pernambuco, Cidade Universitária, Recife, Brazil. [3]Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA. [4]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. [5]Present address: Analytical Research and Development, Pfizer Worldwide Research and Development, Groton, CT, USA. [6]Present address: Department of Chemistry and Biochemistry, Seton Hall University, South Orange, NJ, USA. [7]Present address: Department of Chemistry and Biochemistry, University of North Carolina Wilmington, Wilmington, NC, USA. *e-mail: yizhou.liu@pfizer.com; armando.navarro@ufpe.br

Supporting Information

Application of Anisotropic NMR Parameters to the Confirmation of Molecular Structure

Yizhou Liu*[1†], Armando Navarro-Vázquez*[2], Roberto R. Gil[3], Christian Griesinger[4], Gary E. Martin[1], and R. Thomas Williamson[1]

[1] Structure Elucidation Group, Merck Research Laboratories, Process and Analytical Research and Development, 126 E. Lincoln Ave., Rahway, NJ 07065, United States

[2] Departamento de Química Fundamental, CCEN, Universidade Federal de Pernambuco, Cidade Universitária, Recife, PE 50740-560, Brazil

[3] Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, United States

[4] Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

* Authors to whom inquiries should be addressed
yizhou.liu@pfizer.com
armando.navarro@ufpe.br

† Current address: Analytical Research and Development, Pfizer Worldwide Research and Development, 445 Eastern Point Road, Groton, CT, 06340, USA.

Step-by-step RCSA analysis with MNova/MSpin program.

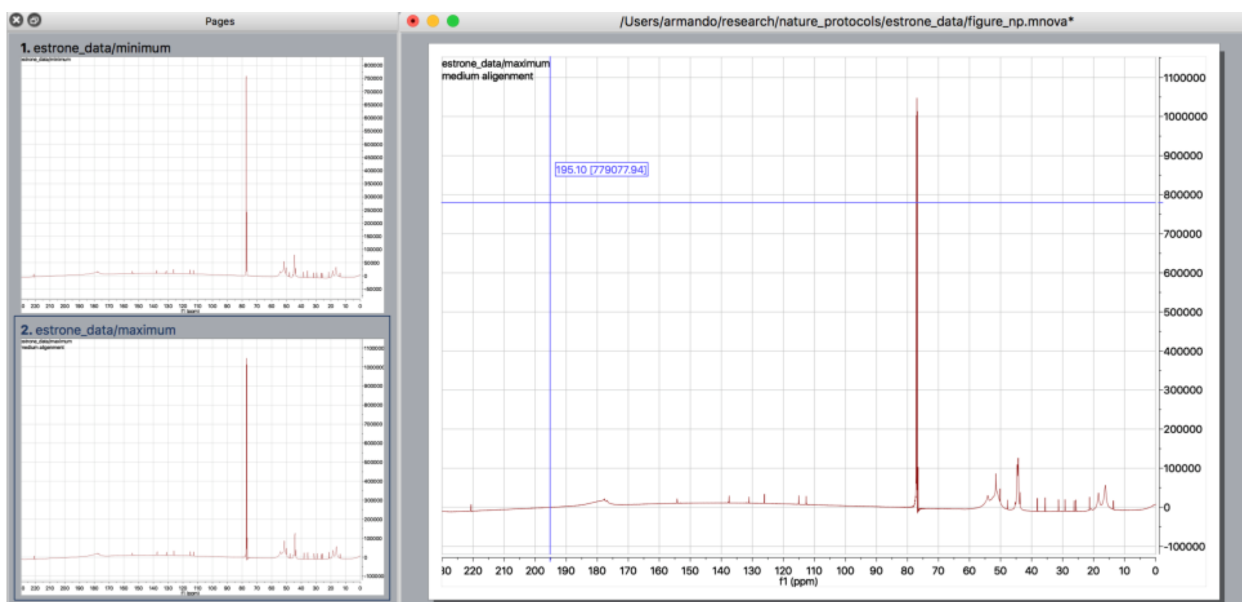
The following step-by-step protocol for estrone is intended to help investigators interested in utilizing the compression device to collect RCSA data for structure confirmation to work through the process using the MSpin program component of the MNova NMR data processing software. The equation used for the isotropic chemical shift change correction, as well as data that can be used for practicing analysis in MSpin, are described and available in the supporting information in the following reference: Nath, N. et al., Determination of relative configuration from residual chemical shift anisotropy. *J. Am. Chem. Soc*., **138**, 9548-9556 (2016).

A subset of these steps also comprises the protocol for analyzing RCSA data collected in a stretched gel, as described later.

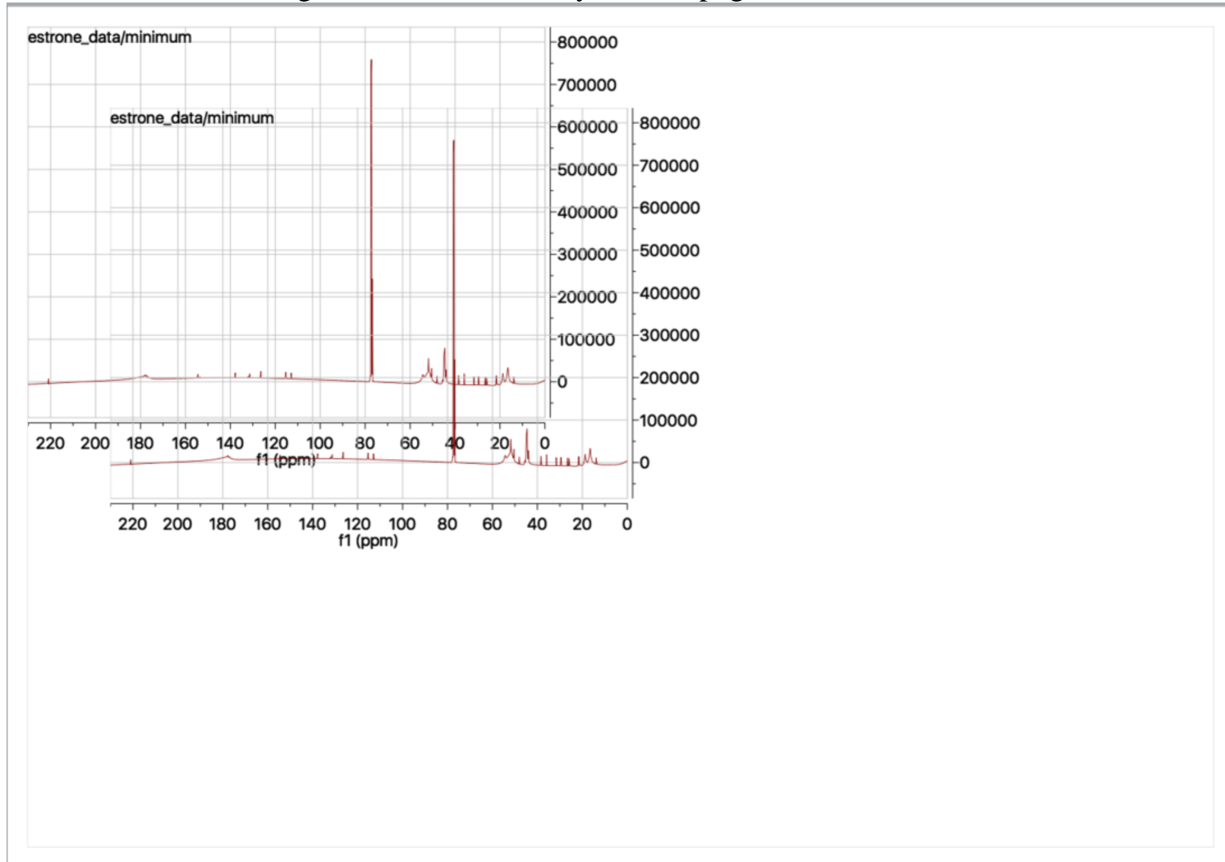
Protocol for RCSA input data preparation for the compression method.

1. Data should be acquired at two different degrees of alignment. The first data set corresponds to the gel in the relaxed state (or minimally compressed state) whereas the second set of data corresponds to the gel fully compressed using the NewEra® compression device.

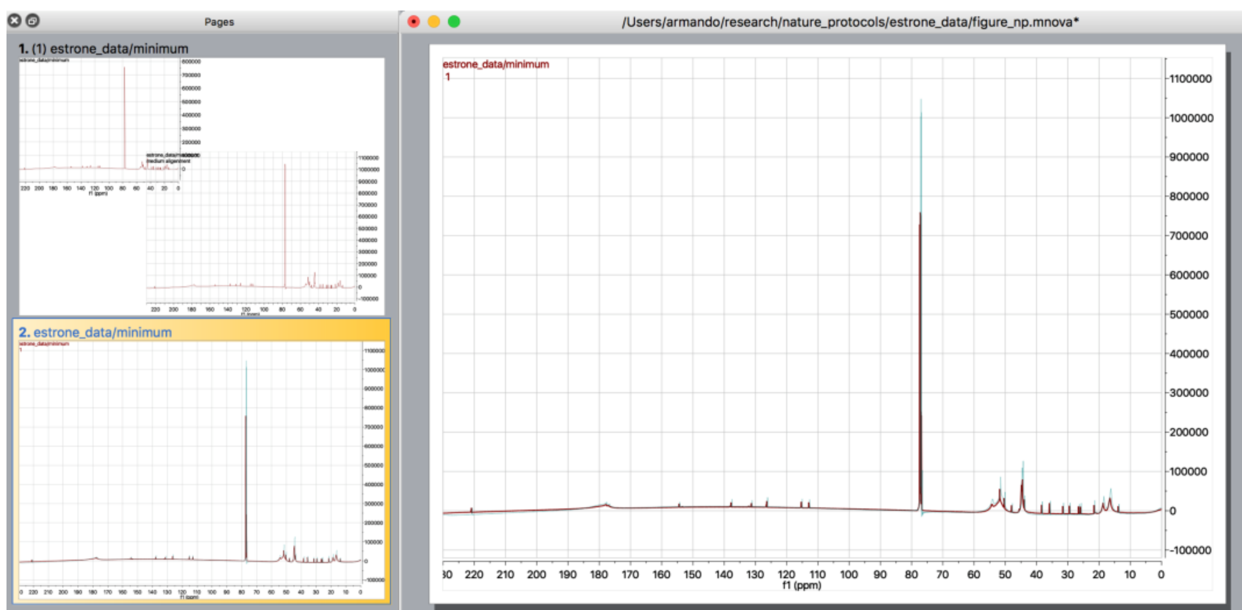
2. Open the two sets of data in the MNova program. At this stage both spectra will be processed with default parameters. The following screen shot shows what the interface will be showing at this point.



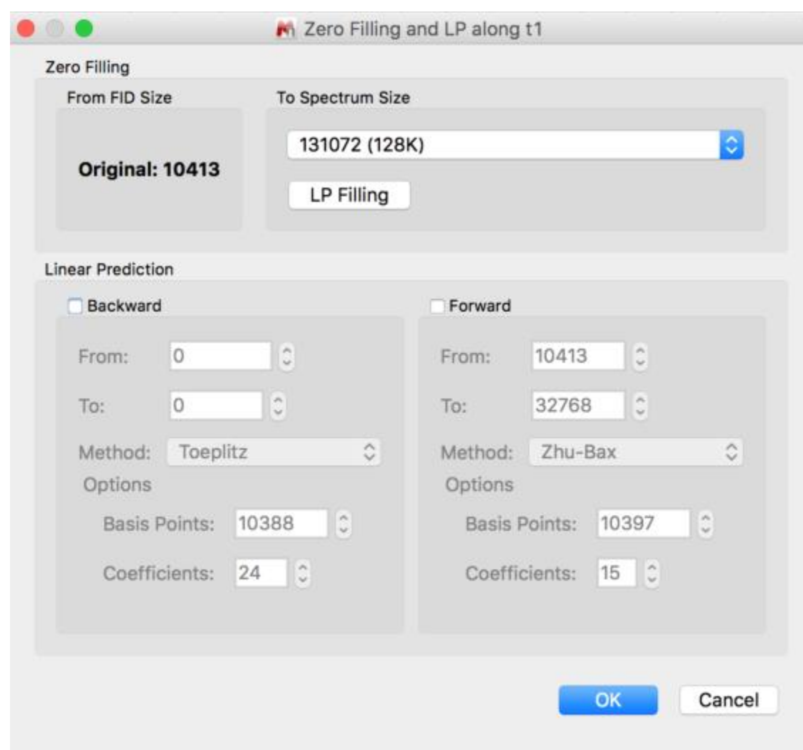
3. Create a new page (**Edit→Create New Page**) and add two copies of the spectrum at minimal alignment onto the newly created page as shown below.



4. Superimpose the two identical spectra at minimal compression created in the new page in Step 3. Go to the edit menu pulldown and perform the following operations. First, **Edit→Select All**. Then choose **Stack→Superimpose** Items. Both spectra will be overlaid as shown in the panel below.

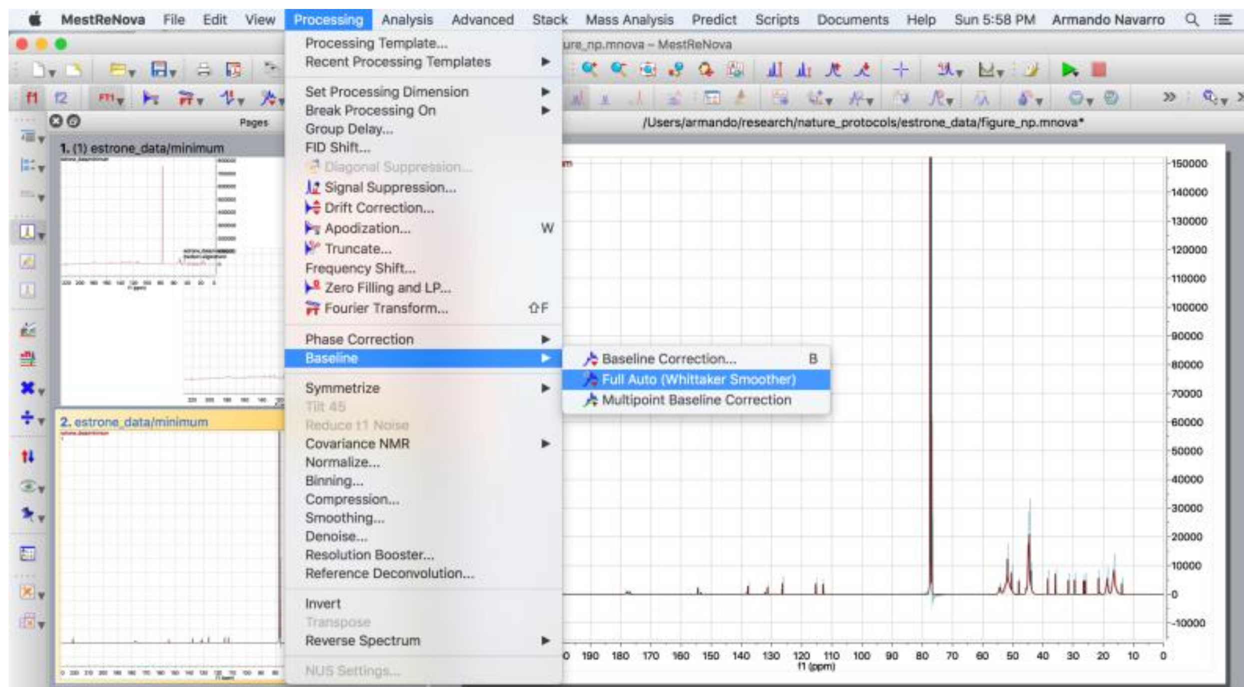


5. In order to accurately measure what can be relatively small chemical shift differences, it is important that peak shapes are well digitized. To satisfy this requirement, go to **Processing→Zero Filling and LP** and select at least 131072 points. The screen shot below shows what you will see during this operation.

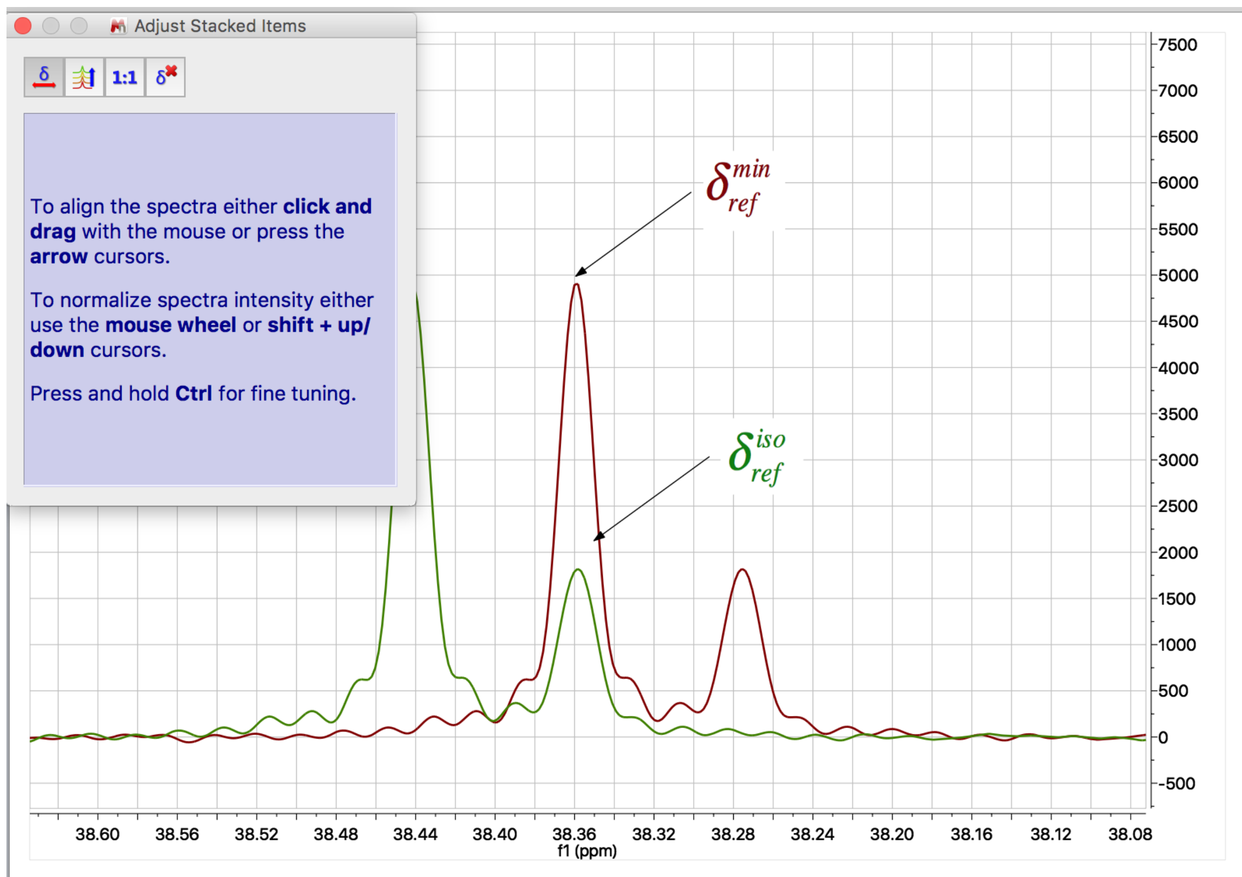


6. Go to **Processing→Phase Correction** and manually correct the zero and first order phase terms.

7. After phasing the spectra correctly, it is necessary to correct the baseline. Go to **Processing→Baseline→Whittaker smoother (or “Polynomial fit”)** to complete this operation.

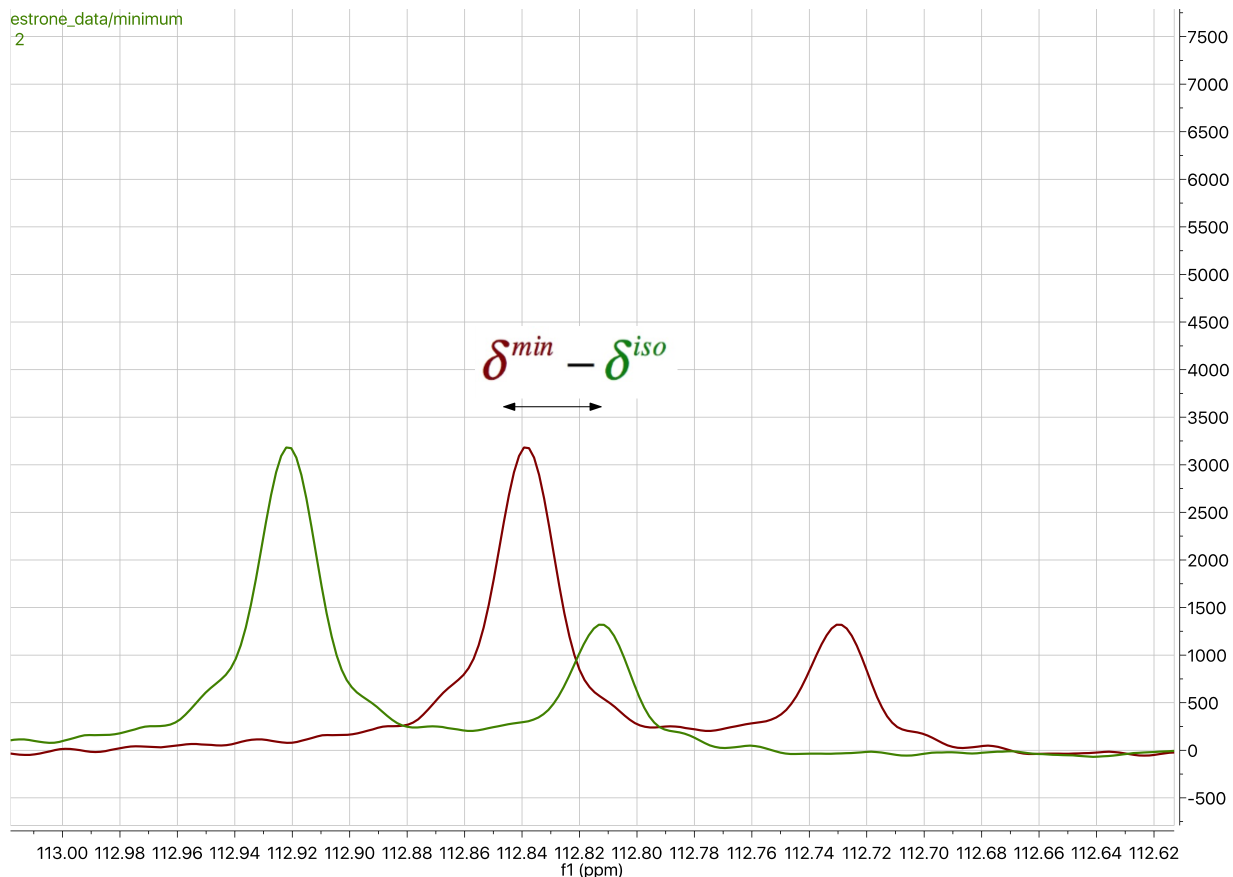


8. This step requires the two spectra to be offset horizontally. To do this, go to **Stack→Adjust** stacked items and in the opened window select the **Shift Horizontally** tool. Zoom on your reference peak, specifically the C8 carbon at 38 ppm in this example, as shown below.



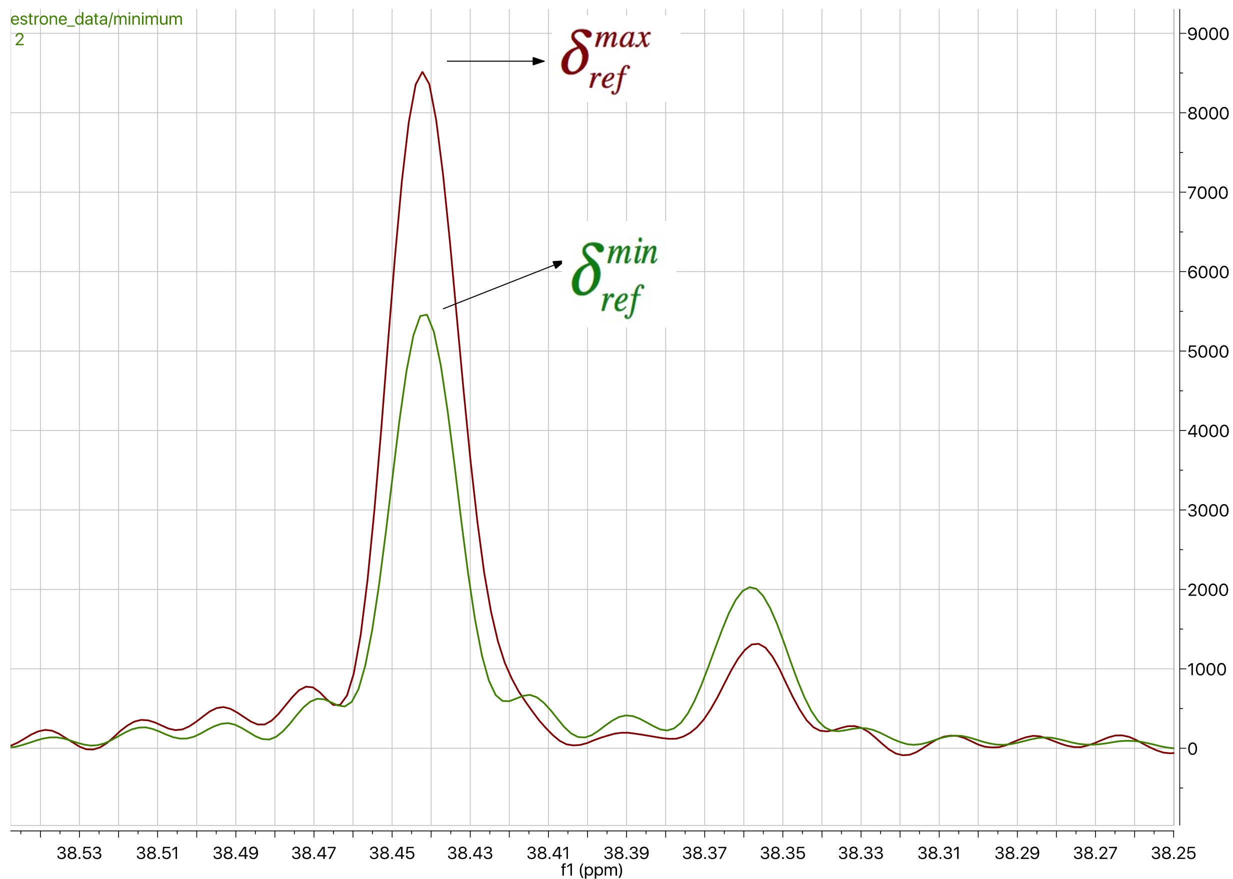
9. Move one of the spectra (in this case the trace colored in red) until you overlap the C8 peak corresponding to the molecule inside the gel in one of the spectrum (the strong red peak) over the peak corresponding to the molecule outside the gel in the other spectrum (the weak green peak). See figure above.

10. Zoom on each of the peaks and annotate the resonance difference between the gel-residing peak in one spectrum (red) and the weak isotropic peak on the second spectrum(green). Tabulate these differences, which correspond to the term $\left(d^{min} - d^{min}_{ref}\right) - \left(d^{iso} - d^{iso}_{ref}\right)$, for peaks as shown below.

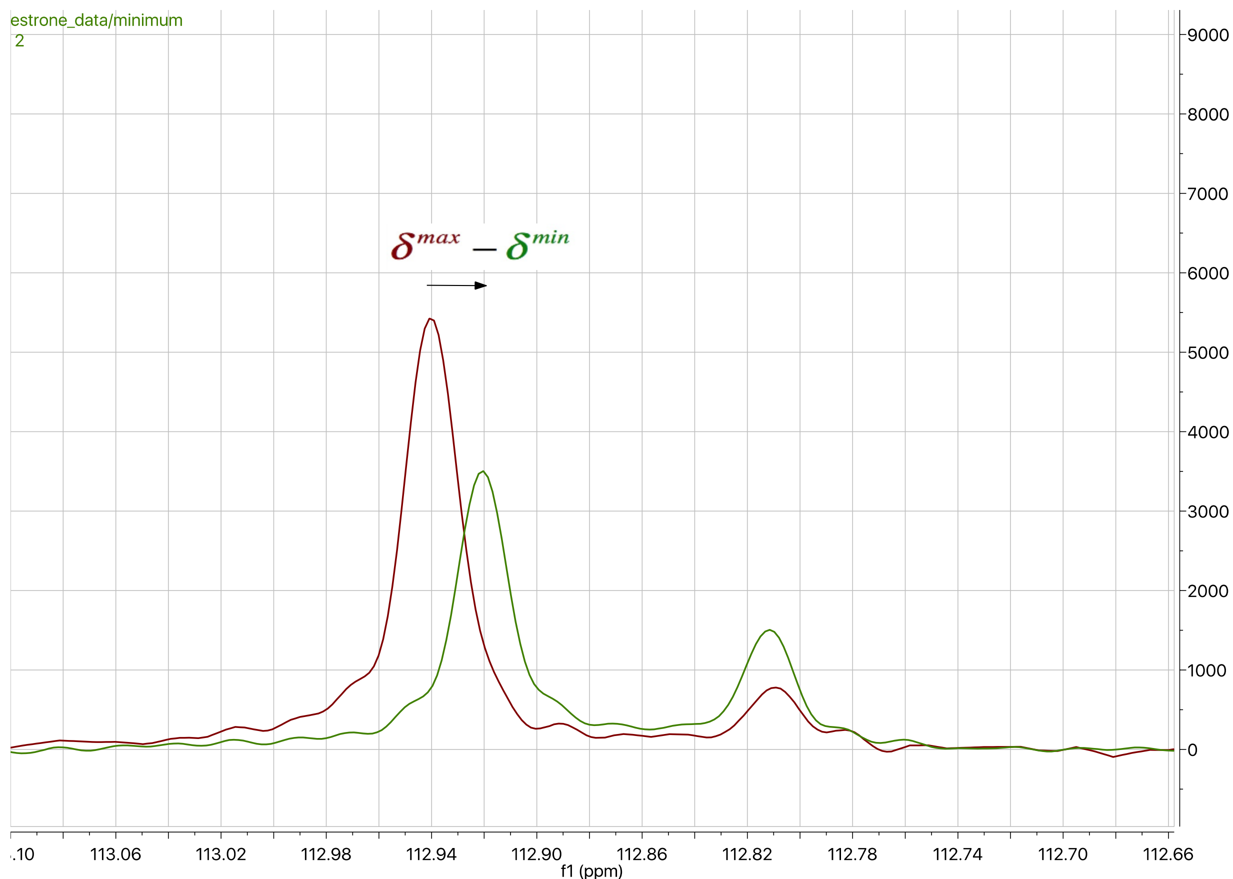


11. Open another page (**Edit →Create New Page**) and copy the spectra at minimum and maximum alignment onto the new page. Repeat steps #4 to #7. Note that the earlier operations in Steps 4-7 were performed on two identical copies of the minimally aligned spectra, whereas here the same operations were repeated but for the two different spectra at minimal and maximal alignments.

12. Go to **Stack→Adjust stacked items**. Zoom on the reference carbon (C8 at 38 ppm) and overlap the peak corresponding to the molecule on the compressed gel (the larger red peak) with the peak corresponding to the molecule inside the relaxed (or minimally aligned) gel (the larger green peak, green), which will give you the presentation shown below.



13. Repeat this step for each of the resonances and record the difference between the peaks for compressed gel (largest peaks, red) with those corresponding to the relaxed gel. These differences will correspond to the term $\left(d^{max} - d^{max}_{ref}\right) - \left(d^{min} - d^{min}_{ref}\right)$. The differences being referred to here are illustrated in the screenshot below.



14. Record and tabulate the differences in the chemical shift for each of the resonances in the spectrum. The C8 reference resonance at 38 ppm will have a chemical shift difference of 0.00 ppm as shown below. Make sure the carbon numbering is correctly mapped between MNova and *Gaussian*. Put values obtained in Step 13 in the "dmax-dmin" column, and values obtained in Step 10 in the "dmin-diso" column.

	carbon number	gaussian file	dmax-dmin	dmin-diso
	17	18	-0.0078	-0.1774
	3	4	0.1150	0.8513
	5	6	0.0056	-0.3756
	10	11	0.0202	-0.9017
	1	2	-0.0134	-0.2345
	4	5	-0.0070	-0.0356
	2	3	0.0191	0.0234
	14	15	-0.0182	-0.0855
	13	14	-0.0188	-0.0855
	9	10	0.0004	-0.0518
Reference	8	9	0.0000	0.0000
	16	17	0.0012	-0.0855
	12	13	0.0044	-0.0513
	6	7	-0.0089	-0.0425
	7	8	0.0060	0.0028
	11	12	-0.0022	-0.0508
	15	16	-0.0107	-0.0855
	18	19	0.0089	-0.0855

15. Build the MSpin file. Carbon indexes should correspond to those in the Gaussian (or other *ab initio* programs) file. In the "rcsa_data" block, enclose data in the format:

    $carbon_id $dmax-dmin gel_shift=$dmin-diso";
    for example,

    "18 -0.0078 gel_shift=-0.1774" ,
    which is simply adopted from the first row of the table.
    Indicate the index of the carbon to be used as reference in the "rcsa_reference" block. The "gel_shift" value corresponds to "dmin-diso"

```
rcsa_data {
#
  18    -0.0078 gel_shift=-0.1774
   4     0.1150 gel_shift=0.8513
   6     0.0056 gel_shift=-0.3756
  11     0.0202 gel_shift=-0.9017
   2    -0.0134 gel_shift=-0.2345
   5    -0.0070 gel_shift=-0.0356
   3     0.0191 gel_shift=0.0234
  15    -0.0182 gel_shift=-0.0855
  14    -0.0188 gel_shift=-0.0855
  10     0.0004 gel_shift=-0.0518
#C8 reference
   #9    0.0000
  17     0.0012 gel_shift=-0.0855
  13     0.0044 gel_shift=-0.0513
   7    -0.0089 gel_shift=-0.0425
   8     0.0060 gel_shift=0.0028
  12    -0.0022 gel_shift=-0.0508
  16    -0.0107 gel_shift=-0.0855
  19     0.0089 gel_shift=-0.0855
}

rcsa_reference {
9
}
```

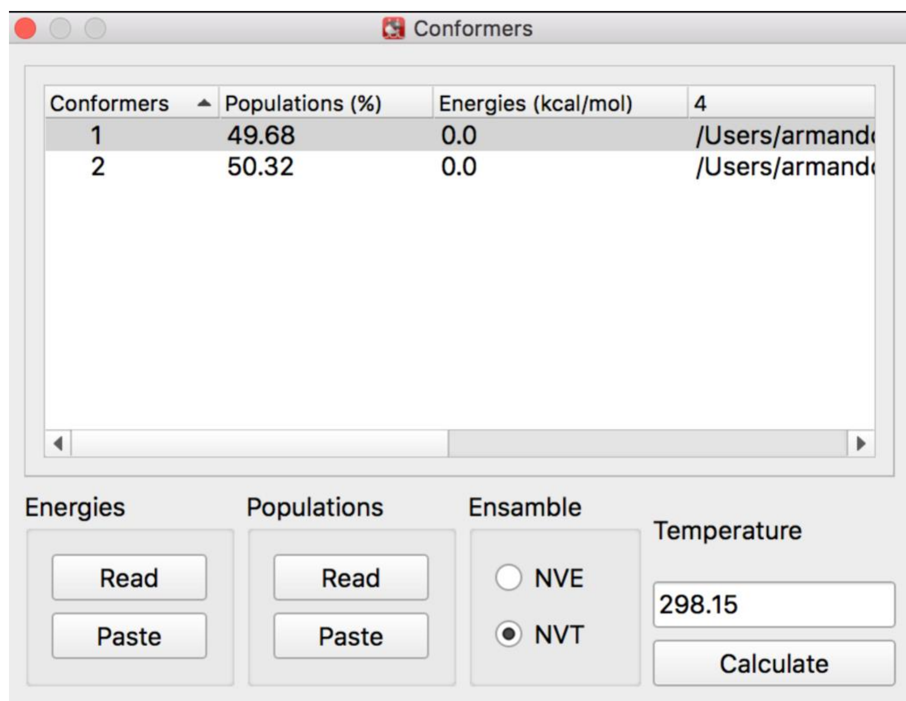
Protocol for RCSA input data preparation for the stretching method

RCSA data collected with the stretching method can be used directly in MSpin after referencing without the need for correction, i.e., the term $\left(d^{max} - d^{max}_{ref}\right) - \left(d^{min} - d^{min}_{ref}\right)$ described in Step 13 can be used directly in MSpin. Therefore, one only needs to go through steps 11-13 to annotate the value associated with $\left(d^{max} - d^{max}_{ref}\right) - \left(d^{min} - d^{min}_{ref}\right)$, which also corresponds to "dmax-dmin" in Step 14, for each carbon atom. In the final RCSA input file, just write the "dmax-dmin" after the carbon ID.

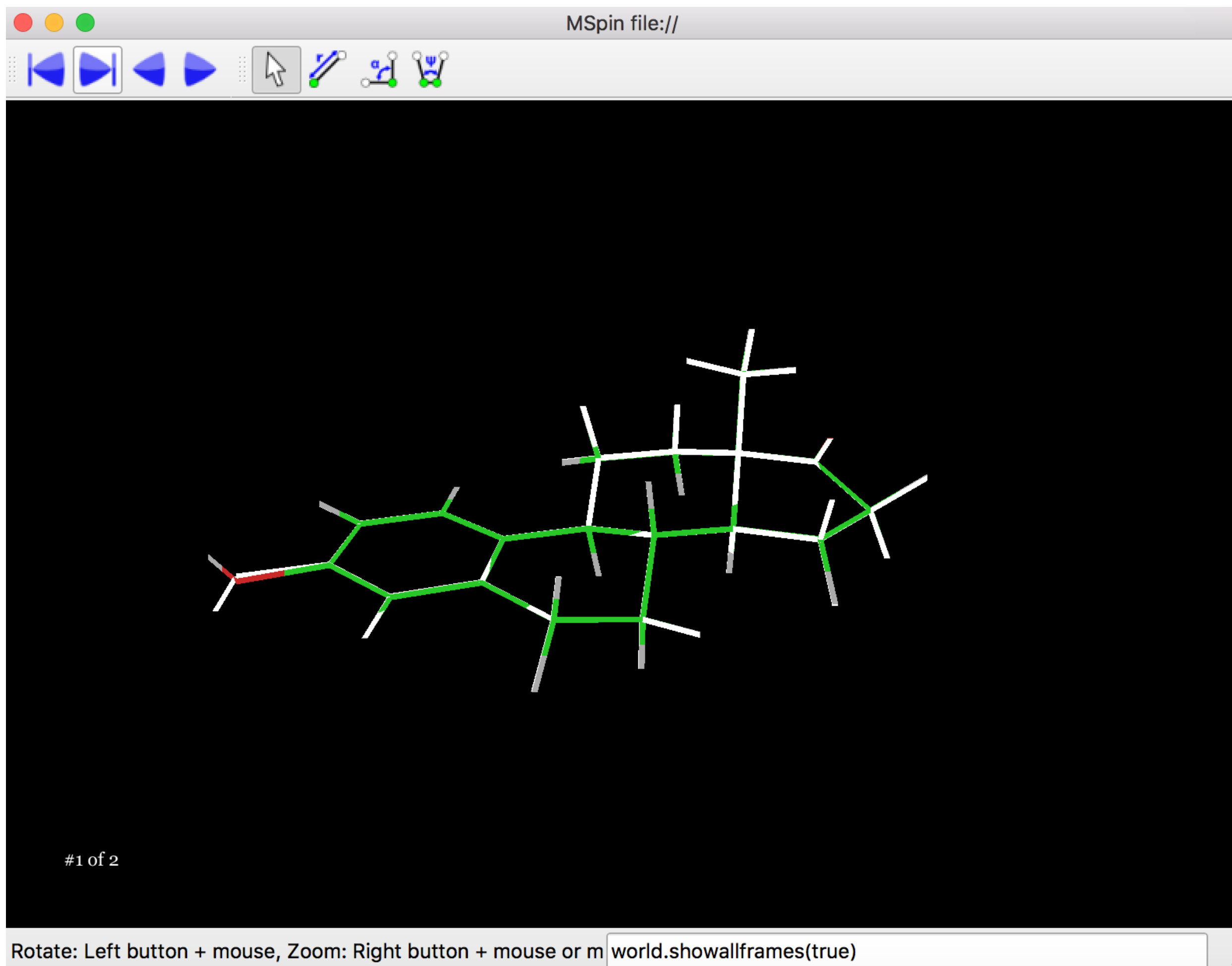
The following steps are for RCSA data analysis in MSpin.

16. Open the MSpin (version 2.3 or higher) program. Select **File→Open Directory** and select the folder containing the Gaussian output files.

17. Open the **Conformers** window. (**Window→Conformers**). Select the NVT ensemble and set the temperature in the Kevin unit (298 in this case). Boltzmann populations will be computed from *Gaussian* energies. The results of this operation are shown below.

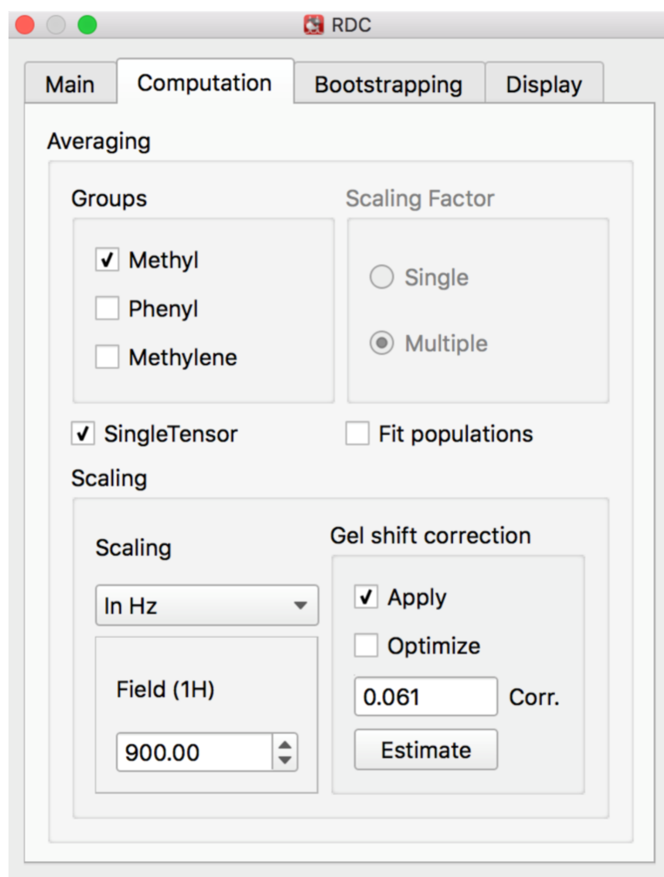


18. Next, go to **Window→Command Line** and type **world.eckarttransform()** in order to align atomic coordinates for single tensor approximation. Type world.showallframes(true) to verify that conformers have been properly aligned. The screenshot below displays the results of this operation.

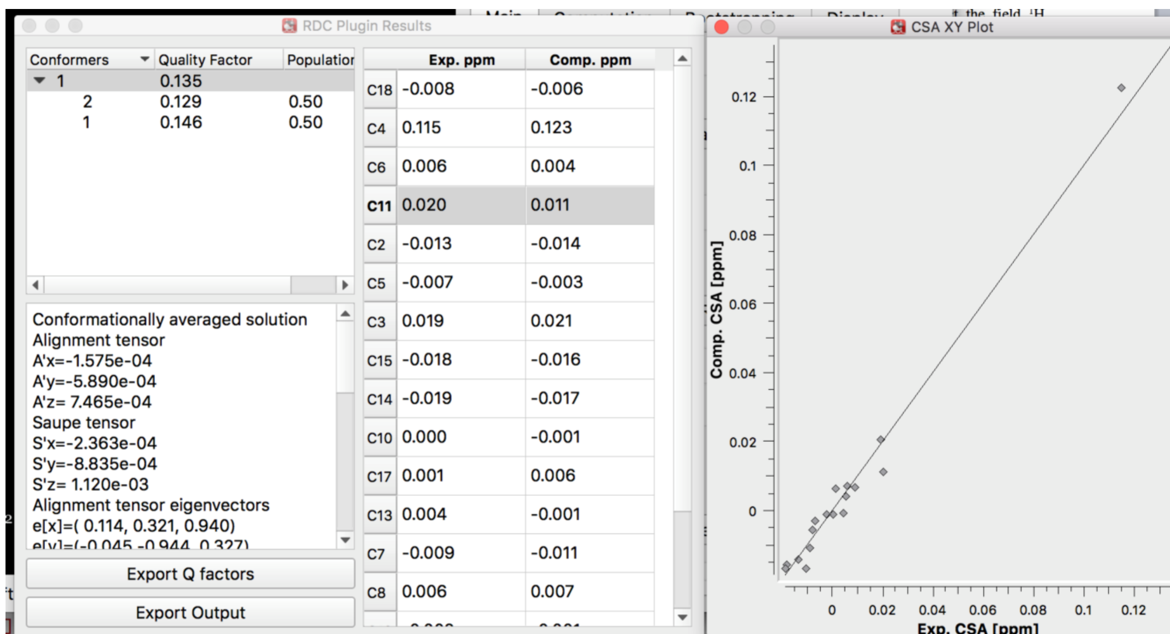


19. Select the MSpin RDC plugin in **Analysis→RDC**. In the **Main** tab of the RDC interface click the **Load experimental data** button. Check the RCSAs box (see below) to verify that data have been well assigned.

20. Select the computation tab. Check the Single Tensor box (see below). Select the field ^{1}H Larmor frequency (Note, however, that this only has consequences if RDCs are also included in the computation). In the **Gel Shift Correction** box select **Apply** and click the **Estimate** button (a value of 0.061 was estimated for the present data set). Return to Main tab and click the **Calculate b**utton. Note that if the RCSA data was collected with the stretched gel, leave the “Gel Shift Correction” box unchecked.

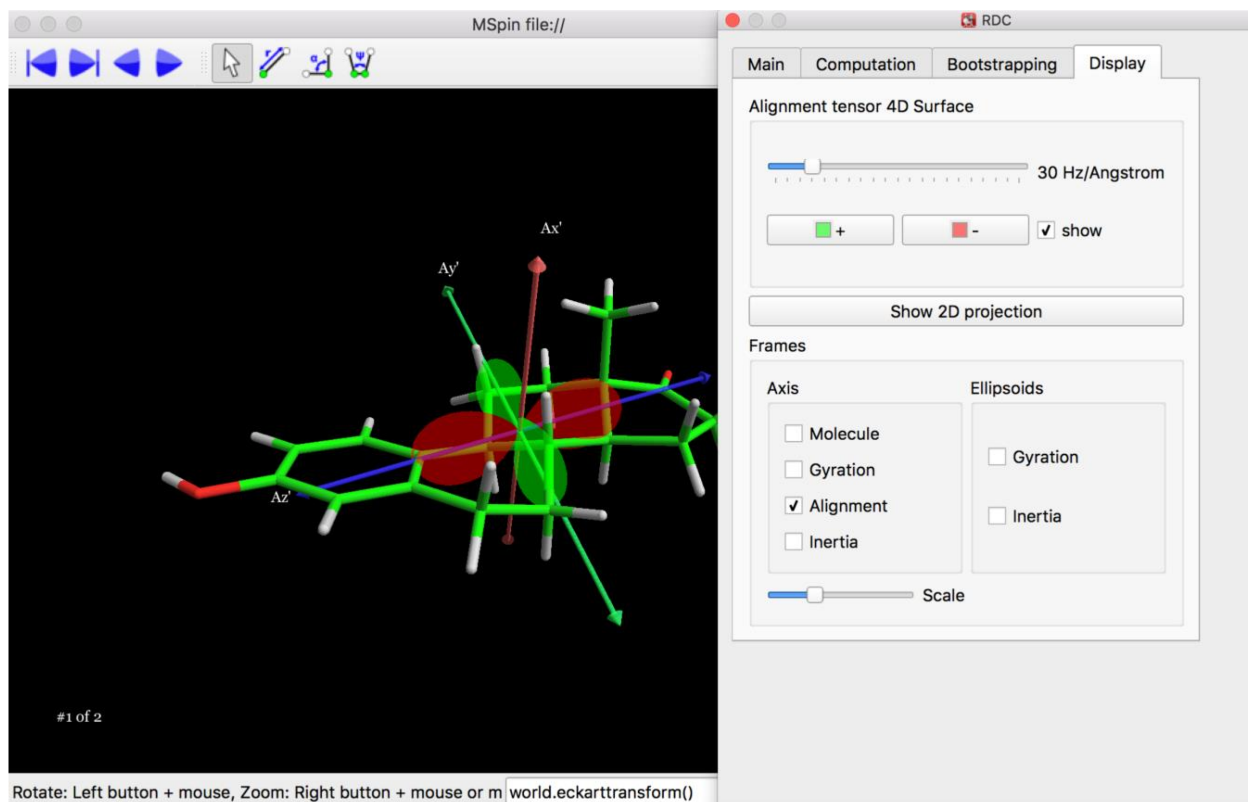


21. Click the **Show Results** button. The quality factor, *Q*, will be displayed for the conformational ensemble along with information about the computation (condition number, alignment tensor principal frame etc.) along with a table of computed *vs*. experimental values. Right click on the table to display a gravimetric plot as shown below.



22. Click the **Export Output** button (see figure above) to open a new window with a complete output text file that can be copied to a text editor and saved to disk.

23. Finally, the principal frame and valued surface of the alignment tensor can be displayed by going to the **Display** tab and selecting the respective options.



24. From step #16 repeat the procedure for the epi-estrone conformational ensemble a quality factor *Q* of 0.354 should be obtained.

Handling equivalent RDC sites in MSpin-2.4/StereoFitter.

The upcoming MSpin-2.4 and StereoFitter programs from MestReLab Research implement a more general way to handle equivalent RDC sites. Both programs can handle any number of equivalent (or averaged) RDC vectors as a list of nuclei pairs and an experimental value in the format:

(I1,J1) (I2,J2) (I3,J3) ……experimental value

For instance, the *N*-methyl groups in cryptospirolepine are input in MSpin version 2.3 or earlier as

```
rdc_data {
…
#NMe 31
31 53 -6.3
31 54 -6.3
31 55 -6.3
#NMe 32
18 49 -6.3
18 50 -6.3
18 51 -6.3
}
```

whereas the input for MSpin-2.4/stereofitter should read as

```
rdc_data {
…
#NMe 31
(31,53) (31,54) (31,55) -6.3
#NMe 32
(18,49) (18,50) (18,51)
}
```