# RSpin: A Dipolar Couplings Tool for the Analysis of Flexible Molecules

Suraj Mirpuri, William R. Carroll, John D. Roberts*

Department of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena CA, 91126
USA

## Abstract

- This work presents a tutorial on how to use software to determine the dihedral angle in substituted ethane systems using a combination of scalar and dipolar couplings. The tutorial is based on the software package RSpin, released free of charge, subject to use and modification under the GNU-AGPLv3 software license. Beginning with generation of appropriate structure files for the software through conformer searching or single bond rotation, it proceeds to construction of text files for coupling constant input. The tutorial also includes instruction on how to obtain useful output from the program in the form of plots and spreadsheets. The goal of this software is to provide a convenient way to determine the dihedral angle of flexible ethanes in solution.

## Table of Contents/Graphic Abstract

-

## Keywords

- NMR Spectroscopy, Computer-Based Learning

In the field of physical organic chemistry, there has been much discussion on the population dynamics of disubstituted ethane fragments. Some of these molecules have attractive interactions between substituents leading to the gauche conformer being favored, while some have repulsive interactions leading to the trans conformer being favored. In either case, the dihedral angle of the molecule is key to understanding quantitatively what fraction of the population is gauche and what fraction is trans. Previously, various schemes have been used to approximate this angle in solution, including using crystal structure angles and computationally provided angles. However, there seems to be no good way to measure what the angle would be in solution nor does there seem to be an easy way to quantitatively predict the conformer population from this angle and compare it to solid phase results and calculations.

To solve this issue, a new method has been developed that uses the often neglected concept of dipolar couplings. These dipolar couplings are spin-spin couplings extracted from the NMR spectra of molecules in ordered media that give information on the absolute orientation and arrangement of the molecule's atoms. These couplings are combined with the familiar scalar couplings from isotropic solution, which provide information on the relative arrangement in space of groups, to give a dihedral angle for a molecule that can be used for further conformational analysis. The combined approach of using both types of coupling can give much more insight into the structure of a molecule than previously thought possible. The downside of

this technique is that it involves advanced knowledge of the theory of NMR couplings, and finesse in preparing ill behaved liquid crystal and lyotropic solutions.

Despite this experimental difficulty, the real issue is how to apply the combined coupling approach in a convenient manner. A commercial program, MSpin, has been made to estimate dipolar couplings for given input structures but it is not user friendly and also cannot use this data to measure dihedral angles. It is closed source and therefore is not transparent to anyone who wishes to learn how the mathematics behind these dipolar coupling calculations work. It is also very costly, being approximately eighty-six dollars for a single annual license at the time of writing. The question is then one of whether or not a free alternative can be made to provide exactly the same function that the commercial software provides, and to also provide a method by which one can measure dihedral angles. This software exists in the form of an open source, free of charge, free for use and modification package called RSpin.

For rigid systems RSpin can take structural files as inputs, as well as dipolar coupling tables, and it can not only back calculate the dipolar couplings for each one, obtaining the same results as commercial software, but it can also output a goodness-of-fit parameter by which to score each conformer that is also the same as commercial software. For flexible systems that can interact with solvent, RSpin can use the given NMR data to select the best fitting input structure and thusly can be given a list of conformers to output the actual dihedral angle of the molecule in solution. Secondarily it acts as an interface to the openbabel cheminformatics project and can read in and save structures as nearly any chemical file format. It can also internally rotate molecules around a single bond to generate conformers, performing minimization at each rotation, or it can conformer search using three different algorithms to generate as many structures as are needed to fit.

RSpin can be a powerful and user friendly tool to do these calculations. What follows is a brief tutorial on how to use it to both calculate dipolar couplings and fit a dihedral angle.

## Using RSpin

RSpin is a Python-2.7 script consisting of a backend compiled down to C code by Cython and a top level GUI written in Python-2.7 using Qt4 libraries. What this means is that the code is both fast and easy to use due to it's compiled backend and simple interface. **Figure 1** is a screenshot of the main RSpin screen showing the command window, plotting window, and output window.
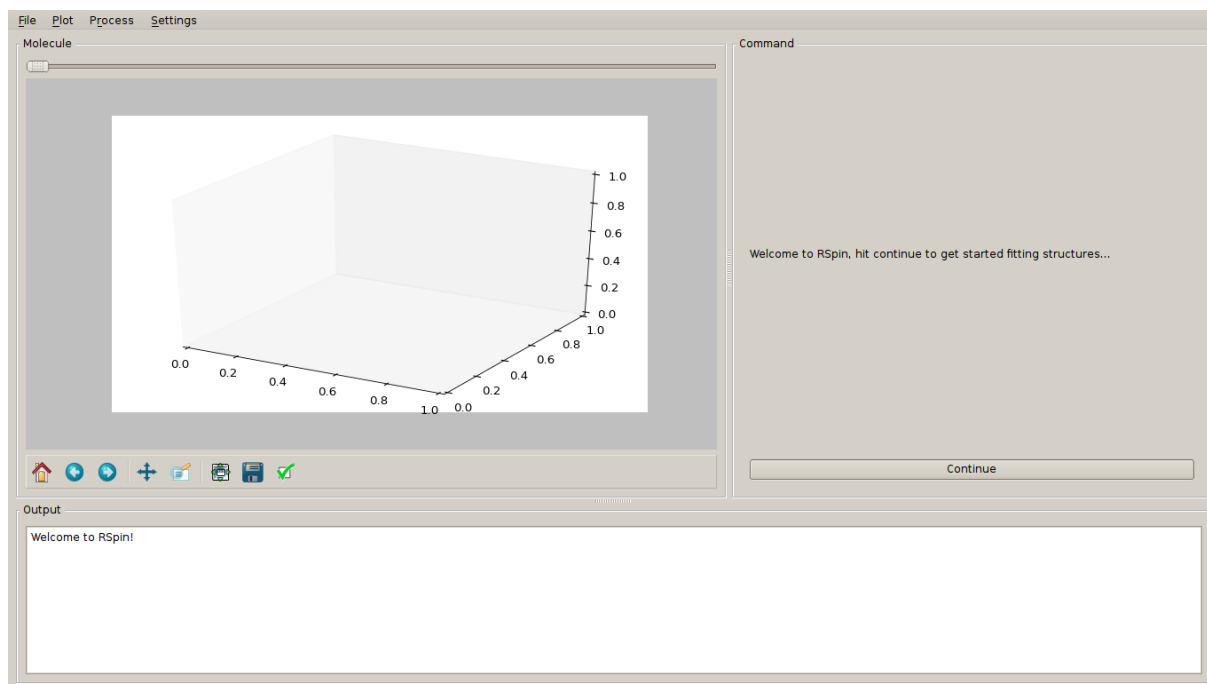
Figure 1: the main window of RSpin. Molecules can be viewed in the plotting window in the top left, calculations can be performed in the top right window, and output and error messages are given in the bottom window.

The plotting window is the top left window displayed in **Figure 1**. To open a molecule through the file menu and view it, click on plot in the plot menu. Click on the plotting window in the main window and the molecule will appear as a ball and stick model. For opening directories, the slider on top of the plot controls which molecule is to be displayed. If the slider is moved and the plotting window is clicked, it will update with the new molecule. The output window is the bottom window displayed **Figure 1**. The output logged in this window should be used only for diagnostic purposes. The actual, more organized results of the calculations must be saved to a file after the calculation is submitted. After a calculation is done, simply use the File → Save menu to save the output. The structure of the file saved will vary depending on exactly the type of calculation performed. Finally, the command window in the top right of **Figure 1** features a set of buttons in a loop that one might use to calculate dipolar couplings and/or measure a dihedral angle. All of this may be achieved through the Process menu as well, and more.

In any case, the user will need a set of chemically reasonable structures, a dipolar coupling file, and a scalar coupling file. The next sections explain how these files are generated.

## Preparing Structural Inputs for Use in RSpin

Structural input to RSpin can be created with any popular molecular modeling program. RSpin can open any file type that is openbabel compatible, including Protein Data Bank files, MDL .mol files, and SDF files. In preparing structures there are many options but there is only one limitation if the user plans to measure a dihedral angle: there must be at least one structure file given for every angle within a set of predefined ranges. To demonstrate this point, consider

the following example options menu shown in **Figure 2**.



Figure 2: the RSpin options menu. From this menu a force field can be selected, as well as optimization parameters and angle bounds for the dihedral angle calculation.

From this screen the first gauche angle range is set from 40° to 90°, the trans range is 165° to 195°, and the second gacche angle range is set from 270° to 320°. This means that one molecule with a dihedral angle in between 40° and 90° must be provided, one with a dihedral angle in between 165° and 180°, one in between 180° and 195°, and finally one in between 270° and 320°. Note that while the trans angle range is given as one range it is internally split into two ranges and therefore must be given one file in the lower range and one in the upper range. The reason for these angle ranges is that the program adopts a brute force approach and calculates the fraction gauche, and dipolar coupling fit for every set of angles and simply plots and selects the best fitting combination. Therefore it can only operate if there is at least one structure per angle range. This is the only limitation to generating structures for testing. However, if the user is interested only in the dipolar couplings for each molecule this limitation does not apply and an arbitrary amount of structures may be given.

   A particularly easy way to obtain structures is to draw the molecule in any molecule editor, and then use Rspin's built in Generate function located under the Process menu. This will rotate the given single input molecule around a rotatable bond of the user's choice, in one degree increments, and it will minimize the structure at a locked dihedral angle. The output files for all 360 structures will then be saved to a directory of the user's choice that can be opened later to select a dihedral angle. This may also be done directly from the Process → Calculate Dihedral Angle dialog. The user should just check the generate box, provide a structure in the top input box, and a folder in which to save output in the bottom input box. Separately from single bond rotation, the user can use Rspin's built in molecular mechanics conformer search to generate all of the energy minima conformers, or totally random conformers in a similar fashion. This can be done by going to the Process menu and selecting Conformer Search, a dialog will appear with

an option to select the search algorithm as well as input boxes for starting file, save directory, and number of conformers. Note that if the Systematic algorithm is selected, the number of conformer options is irrelevant and will not be taken into consideration no matter what is given as input.

## Preparing Scalar and Dipolar Coupling Input Files

Coupling input for RSpin can be generated by text editing a file or using a spreadsheet editor. For dipolar couplings the file must be in a specific format, it must be a three column spreadsheet, with the first row being headers. The first column must be the dipolar couplings in Hz, the second column must contain the first nuclei of the coupled pairs, and the in the third column there must be the corresponding terminal nucleus for each nucleus in the second column. An example is shown in **Figure 3**.
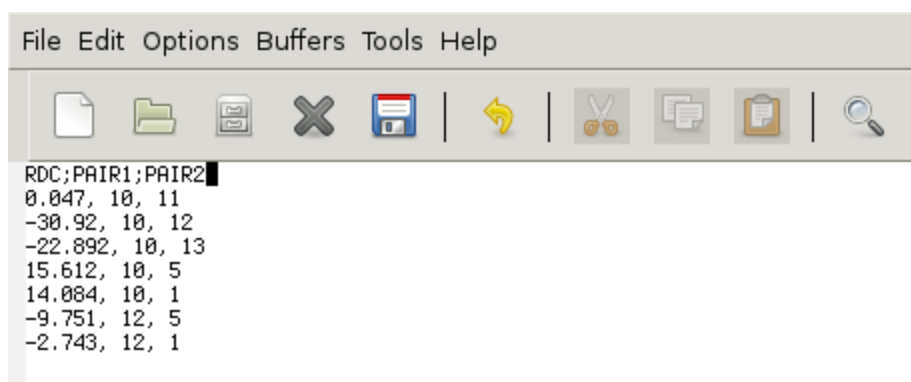


Figure 3: a GNU Emacs window showing the dipolar coupling file's structure. The first row is a header, it can be anything as it is discarded, the rows after are organized in columns. The first column represents the dipolar coupling in Hz, the second column is the label of the first nucleus in the coupling, and the third column is the label of the second nucleus.

This file must be saved as a text csv, or a comma separated values file. The program will not accept the couplings if they are not in the correct format and an error will appear stating that the user must check the dipolar coupling input.

For scalar couplings however, the format is slightly different. It must also be in a text csv file, however there is no header line. The first column rather is a header column, and all of the data must be entered in this exact order in the second column: J13, J14, Lambda 1, Lambda 2, Gauche Angle, Trans Angle. The J13 and J14 couplings are the scalar couplings in Hz, lambda one and two are the group electronegativity coefficients for the HLA relation, and the gauche and trans angles do not necessarily need to be in the file unless one wishes to do a gauche fraction calculation for a given set of angles. An example of the file is given in **Figure 4**.

```
"J13", 2.01
"J14", 9.73
"Lambda 1", .2
"Lambda 2",.41
"Gauche Angle", 65
"Trans Angle", 180
```
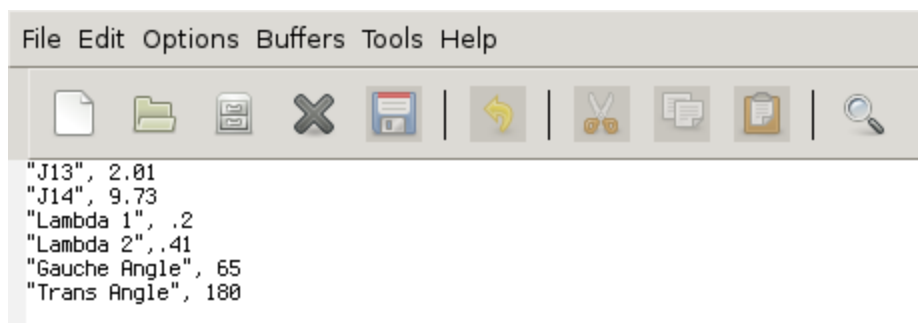
Figure 4: a GNU Emacs window showing the structure of the scalar couplings file. It has no header row, but the first column is composed of labels that are discarded much like the dipolar coupling file's header row. The labels can be any string, but they must be in the order shown and contained in double quotes. Also, the gauche and trans angle will not be used and can be omitted, unless a gauche fraction calculation is performed.

## Using RSpin

After inputs have been generated and saved, either navigate to the process menu and select "Calculate Rigid System Couplings" to back calculate dipolar couplings or "Calculate Dihedral Angle" to find the best fitting structure for a flexible system. This can also be done by clicking continue in the main command window and then clicking the appropriate button. A dialog will appear with various input boxes. Click on the browse buttons to select the appropriate locations for the files that were prepared earlier. The top of the dialog contains file input and should only have the middle text box open for input unless the generate structures button is checked. The path of the directory containing the structure files must be entered into this text box. The middle section of the dialog contains the inputs for the coupling files. In the first box, the location of the dipolar coupling csv should be inputted while the scalar coupling file location is inserted into the second box. In a calculation of dipolar couplings without dihedral fits, the scalar coupling box will be disabled as it is not necessary. Finally the last box contains an input for a SMARTS search string, four characters long, that represents the atoms for the dihedral angle in question. For example, the backbone angle of succinic acid is represented by the string "CCCC". This must be given for both flexible and rigid fits as when saving the data, it is indexed by dihedral angle and in order to do this it must be read from each input file. If the molecule has more than one possible dihedral angle for the given SMARTS query, a dialog will appear asking for the user to select the appropriate backbone atoms listed by atom label. Then simply select "Run" and the calculation will begin. All error messages and updates will be shown in the output window. A progress bar should also appear showing how long the calculation should take. If the calculation was a dihedral angle fit, a plot of goodness of fit against gauche fraction should appear. Use the save icon in the bottom left to save a Portable Network Graphics (.png) image for the plot. To zoom in or out use the zoom tool in the same toolbar.

When the calculation is finished, simply close the plot and view the results in the output window, or save them to a text CSV by using the File → Save option. If a dipolar coupling calculation was done, the file will be saved as a four column spreadsheet containing the dihedral angle of each file, the dipolar couplings, the force field total energies, and the rigid system

goodness of fit parameters. If the calculation was a dihedral angle fit, then the file will contain a column of permutations of angles, the fraction gauche for each permutation, the flexible system goodness of fit, and the dipolar couplings of the combined system. These text CSV files can be opened by any spreadsheet editor for further plotting and data analysis.

Finally, a word about the options menu: it can be used if the user wishes to change the force field type used in generation of molecules and conformer searching. MMFF94 and UFF are both available as is the force field from Ghemical. Convergence criteria can be tweaked in this menu as well. In addition to all of this, one can change the range of the angles examined by RSpin here. This may provide extraneous or useless information however, and it will certainly cause the calculation to take longer. It is not recommended to change the angle ranges.

## Associated content

## Supporting Information

Supporting information includes a folder full of structures to test with the program for the succinic acid monoanion test case, a dipolar coupling file, and a scalar coupling file. This material is available via the Internet at *http://pubs.acs.org.*

## AUTHOR INFORMATION

## Corresponding Author

- *E-mail: <Dr. Roberts' Email>

## Notes

The author declares no competing final interest.

## REFERENCES

1.