# A Novel Method for Determining Dihedral Angles from Dipolar Couplings in Rotationally Dynamic Systems

Suraj Mirpuri †,
Mentor: John D. Roberts
Co-mentors: Bill Carroll, Bright Emenike

Gates and Crellin Laboratories of Chemistry
California Institute of Technology
1200 E California Blvd
Pasadena, CA 91125

**ABSTRACT:** In computational chemistry, the problem of identifying the dihedral angle around a "$CH_2CH_2$" fragment of ethane systems is a difficult one due to solvent effects, making solutions unreliable for flexible systems. This project is to write a program that will, given dipolar and vicinal coupling constants from NMR, be able to fit a dihedral angle for the gauche conformer, trans conformer, and their relative proliferation in solution holistically. The program obtains structural data from the user, narrows down a list of dipolar couplings using singular value decomposition, and then superimposes structures to find a fraction gauche with best fit. The program then uses this value for the gauche fraction in the Haasnoot-Altona relation to back calculate a dihedral angle for the gauche conformer. This method is both chemically intuitive and mathematically sound, and it reliably provides dihedral angles for a variety of rotationally dynamic systems.

## BACKGROUND

In recent times, nuclear magnetic resonance spectroscopy (NMR) has become the tool of choice for chemists who need to obtain structural information for organic molecules. It delivers a wealth of information very quickly, much of which is hard to analyze without the help of specialized software and a trained eye. However, even with modern computational facilities, there is still no good way to determine accurate structural information for molecules that rapidly interconvert between different conformations. NMR simply does not have the time resolution to image each conformation that something can take so any structural data gleaned from the NMR is then a time average of the real states. This presents a serious problem: the "averages" we derive from NMR experiments cannot directly confirm our structural guesses. For example, succinic acid is a rotationally flexible molecule that strongly prefers a gauche conformation in water. This is due to the gauche effect, which changes the molecule's dihedral angle from the classically held 60°. Current software solutions still output or even assume these stereotypical dihedral angles despite the fact that the angles can be computed directly from NMR data. Also, current software does not solve for conformer populations and also has difficulties in accounting for solvents. The goal of this research therefore, was to contrive of a software solution that can, given structural input and NMR data, output the dihedral angles for a given molecule's gauche and trans conformers and the relative abundance of each population. This was found to work on a large variety of molecules that can be treated as disubstituted ethane fragments. The software has not yet been adapted to work on trisubstituted or fully substituted ethane systems; however, the disubstituted case acts as proof of concept of the idea that dihedral angles can be calculated for rotationally dynamic systems.

## METHODS

The program, RSpin, was written in the popular scripting language Python 2.7, in four separate modules made to handle different facets of the problem. The first module written was "inputparser.py," a module built to read common structure files used widely in computational chemistry. It contains several parsers for a variety of formats, manual input subroutines, and error checking for input files. The second module was "coord.py" which contains plotting functions, functions to retrieve geometries from structure files, and a molecule object class that wraps a coordinate system, atoms, bonds, and geometry for a given input molecule into a common class. The third module is the main function "gui.py" which contains a graphical user interface class that presents a simple mouse button driven interface for the entire program. Finally, the last module is “moleculemath.py” which as the name implies, contains all functions relating to calculation of secular constants, conformer populations, and dihedral angles from input dipolar couplings.

In order for the program to work, it must first be given structural data in the form of Cartesian Coordinate files. This file format was selected as it can be used by the popular software ChemBioDraw and by the OpenBabel project, and is simple to parse unlike more well-known formats such as MOL 3000 and Protein Database formats. The structures of the Cartesian Coordinate files are given in **Figure 1**. In order to create conformations to test, these files may be rotated by OpenBabel’s “obrotate” program or individual files may be generated from ChemBioDraw’s dihedral driver module. The module “inputparser.py” accepts this input format and decomposes it into a single list, the entries of which are atoms, labels, a tuple of coordinates, and a tuple of linked atoms. The main function of the program then accepts this list and passes it to the molecule class to store each parsed input file as an individual molecule. After the program stores structural information for all given conformations to be tested, it must be given a single set of NMR data from the test molecule. In order to calculate dihedral angles the program must be given: dipolar couplings and assignments, vicinal couplings, λ values for the 1994 reparametrization Haasnoot-De Leeuw-Altona (HLA) relation, a theoretical gauche angle, and a theoretical trans angle upon which to iterate.

| | | | | | | |
|---|---|---|---|---|---|---|
| Line 1 | Total Number of Atoms | | (Coordinates in Angstroms) | | | |
| Line 2 | **Atom Symbol** | **Atom Label** | **X** | **Y** | **Z** | **Bonded Atoms** |
| Line 3 | **Atom Symbol** | **Atom Label** | **X** | **Y** | **Z** | **Bonded Atoms** |
| Line 4 | . | . | . | . | . | . |

**Figure 1:** An example of the coordinate file format

The secular couplings are gathered in a variety of ways. They may be derived from liquid crystal NMR studies, lyotropic NMR studies, or other well-ordered media studies of molecules in a variety of solvents. For this program, spectra of various gauche preferring and trans preferring molecules in lyotropic solutions were used as test cases, generated by another 2012 SURF student, Ben Arbeiter. The lyotropic solution was prepared with myristyltrimethylammonium bromide (MTAB) as the ordering media and D2O as the solvent. The final solutions consisted of a total of 2.15% of the test molecule, approximately 31% MTAB, 61% D2O, and decanol to balance. Spectra of these solutions are iteratively fit in gNMR, a program that can simulate spectra and calculate couplings from those simulated spectra. When a table of dipolar constants is generated, these couplings are then assigned to an interspin vector representing a bond in the molecule or a vector through space connecting nuclei. These assignments are manually correlated with the labels that RSpin reads from coordinate files, and then the assignment / coupling pair is put into a comma separated value (CSV) file. This CSV file is a common type of spreadsheet file that can be opened by any spreadsheet program, and the format was selected for RSpin as it is notably easy to parse. The format of one of these files is given in **Figure 2**.

```
Header 1, Header 2
Coupling in Hz, Label Pair in Quotes
1.5, "[1,12]"

., .

., .
```

**Figure 2:** Format of the comma separated values file with example data in the third line. The format is space separated as is the Cartesian Coordinate file.

The last four inputs for RSpin may also be gathered in a number of ways. The vicinal coupling data used to test this program was taken from gNMR when the secular couplings for the test molecules were calculated. The electronegativities used in the HLA relation were calculated in house (Bendelsmith, 2011) for a variety of substituents in myriad solvents. Finally, the theoretical input angles are calculated using SPARTAN ’10. Note that these angles are starting points upon which RSpin may operate and they are in no way assumed to give an end product. These values are used to generate a chemically valid fraction gauche in later calculations. This calculation of fraction gauche is what allows RSpin to solve for rotationally dynamic systems and when combined with information from the dipolar couplings, RSpin can select a best fitting dihedral angle.

## THEORY

The first type of NMR data on which the program operates is the secular coupling, also known as the dipolar or D coupling. The dipolar coupling between two spins A and B is given by:

$$D = \frac{\kappa}{R^3}\langle 3\ \cos^2\theta - 1\rangle \qquad (1)$$

$$\kappa = \frac{-3\mu_0\hbar}{8\pi^2}\ \gamma_A\gamma_B \qquad (2)$$

R is the distance between both spins, θ is the angle of the interspin vector with the external magnetic field, κ is a constant term defined in (2), and γ represents the gyromagnetic ratio of the spin A or B. It is important to note that the cosine term is in brackets, as in isotropic solution, the molecule is tumbling in all directions such that there is no net alignment with the magnetic field. Therefore the alignment that exists, if any, is an average and in isotropic solution averages to zero, which is why no dipolar couplings are seen without ordered media. These dipolar couplings give information about the distances between spins, which in rigid cases is enough to completely determine geometry. A more relevant equation that has been shown to be equivalent to (**1**) by Kramer and coworkers in 2003 is:

$$D = \frac{\kappa}{R^3}\langle\overrightarrow{r^T}\boldsymbol{A}\vec{r}\rangle \qquad (3)$$

In this equation, $r^T$ and r are unit vectors corresponding to the interspin vector. **A** is the commonly used alignment tensor which is rooted in the probability ellipsoid that governs the alignment of the interspin vector with the magnetic field. In this equation it is still time averaged as the alignment tensor changes with molecular motion, however in anisotropic media there is a definite alignment tensor due to restricted movement and **A** is nonzero. The alignment tensor is 3x3, real, symmetric, and traceless and therefore consists of only five independent parameters $A_{xx}$, $A_{xy}$, $A_{xz}$, $A_{yy}$, $A_{yz}$:

$$\begin{bmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{xy} & A_{yy} & A_{yz} \\ A_{xz} & A_{yz} & (-A_{xx} - A_{yy}) \end{bmatrix} \qquad (4)$$

If the unit vector is set equal to some vector $\langle x_1\, y_1\, z_1\rangle$, where $x_1, y_1, z_1 \in \mathbb{R}$, and if D is a vector of five dipolar couplings, the alignment tensor can be multiplied by the five corresponding unit vectors to give a simple equation of the form $Ax = b$, given in equation **5**.

$$\begin{bmatrix} x_1{}^2 - z_1{}^2 & y_1{}^2 - z_1{}^2 & 2x_1y_1 & 2x_1z_1 & 2y_1z_1 \\ x_2{}^2 - z_2{}^2 & y_2{}^2 - z_2{}^2 & 2x_2y_2 & 2x_2z_2 & 2y_2z_2 \\ x_3{}^2 - z_3{}^2 & y_3{}^2 - z_3{}^2 & 2x_3y_3 & 2x_3z_3 & 2y_3z_3 \\ x_4{}^2 - z_4{}^2 & y_4{}^2 - z_4{}^2 & 2x_4y_4 & 2x_4z_4 & 2y_4z_4 \\ x_5{}^2 - z_5{}^2 & y_5{}^2 - z_5{}^2 & 2x_5y_5 & 2x_5z_5 & 2y_5z_5 \end{bmatrix} \begin{bmatrix} A_{xx} \\ A_{yy} \\ A_{xy} \\ A_{xz} \\ A_{yz} \end{bmatrix} = \frac{\vec{R}^3}{\kappa} \begin{bmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \end{bmatrix} \qquad (5)$$

If five dipolar couplings are known, and corresponding unit vectors are supplied, this equation can be solved with a least squares method to give the five independent parameters of the unit vector. If more than five dipolar couplings are available, the list may undergo singular value decomposition (SVD) to generate the five best dipolar couplings to use. SVD is a form of eigendecomposition that will calculate a singular value – the square root of the eigenvalue - for each dipolar coupling. If these values are placed in ascending order, the corresponding dipolar couplings will be placed in the order in which they should be used. The top five couplings are those values with the lowest eigenvalues, and thusly the least covariance. The values of least covariance are desired as they represent the values that will produce the best line of fit for the data set and if they are used to calculate an alignment tensor, the best fitting alignment tensor for the given dipolar couplings. This alignment tensor can then be used to back calculate other dipolar couplings filtered by the SVD algorithm. Finally then, the root mean square deviation (RMSD) between the calculated and experimental data set is calculated and used as a single measure of how well the given dipolar couplings fit a given structure's alignment tensor. The equation used to calculate RMSD is given in (**6**).

$$RMSD = \sqrt{\langle(D_{obs} - D_{calc})^2\rangle} \qquad (6)$$

The structure that gives the minimal RMSD is the best fitting structure, and if several structures are loaded into the program that represent one degree rotations about a single carbon-carbon sigma bond, then the best fitting structure represents the true dihedral angle of the given molecule. This is only true for rotationally inflexible molecules however. If there is anything other than a strong preference for one conformer, for example if the preference is less than 95% for one conformer, then the D couplings / RMSD method will not work alone.

In the case of flexible molecules, there is yet more information to be gleaned from the NMR, in the form of J, or scalar, couplings. These couplings arise from the spin Hamiltonian, which in isotropic solutions consists of only the contribution from magnetic interactions between nuclear and electron spin. They can be used to determine the relative arrangement of groups on a backbone and also to determine fraction gauche (*vide infra*). In any case, these J couplings may be determined directly from the NMR spectra, or if the molecule's geometry is known, estimated with any of the equations in the family of Karplus curves. One can also, given J couplings from experiment, estimate dihedral angles using these curves, however in flexible cases this method is of dubious accuracy at best. To solve for flexible systems, the system's conformer population must first be deduced using the following simultaneous equation **7**:

$$\begin{cases} J_{obs} = f_G J_G + f_T J_T \\ 1 = f_G + f_T \\ J_{G,T} = 14.63\, cos^2(\theta) - \\ .78\cos(\theta) + .60 + \sum_i \lambda_i \begin{Bmatrix} .34 - 2.31\, cos^2 \\ [s_i(\theta) + 18.4\,|\lambda_i|] \end{Bmatrix} \end{cases} \qquad (7)$$

The first equation states that observed couplings are a weighted sum of the fraction gauche and trans multiplied by the coupling in an all gauche population and the coupling in an all trans population. The second equation states that the fraction gauche and trans sum to one. The final equation is the 1994 reparametrization of the HLA relation which accounts for group electronegativities in its estimation of J couplings. The general idea is to use the third equation to estimate $J_G$ and $J_T$, which are then used in the first equation with observed couplings to find the fraction gauche.

After a fraction gauche has been found, then the flexible case may be solved through a combination of these J and D couplings. This combined approach is the key to breaking down the conformational averaging seen in the NMR. It works by assuming that the dipolar couplings gathered from the spectrum are a weighted average of a single gauche conformer's dipolar couplings and a single trans conformer's dipolar couplings. This assumption can be made safely in disubstituted systems where the gauche conformers are symmetric. However in ethane fragments with a greater number of substituents this symmetry assumption fails to hold and more terms must be added to the weighted average to account for the lack of symmetry, which is an issue in current research. Equation **8** shows the main equation derived from this assumption:

$$\begin{bmatrix} D^{Calc} \\ \vdots \end{bmatrix} = \begin{bmatrix} D_{Gauche} \\ \vdots \end{bmatrix} f_G + \begin{bmatrix} D_{Trans} \\ \vdots \end{bmatrix} f_T \qquad (8)$$

As this function is a bijection, there exists only one pair of gauche and trans molecules for a given fraction gauche such that the difference between the calculated dipolar couplings and the observed dipolar couplings is minimized. As the fraction gauche can always be determined reliably, the error in this method is only limited by the error in the original set of dipolar couplings. In summary then, the method to solve for rotationally flexible systems is to determine a fraction gauche, and then fit all possible gauche and trans pairs using equations **8** and **6** to effectively treat each conformer as its own rigid molecule and use dipolar couplings to assign geometry in this rigid case.

## RESULTS / DISCUSSION

The first experiment tested with RSpin was to fit observed dihedral angles for five molecules: tetramethylethylenediamine oxide (TMEDO), β-alanine, a trans standard - trimethyldimethylbutylamine (TMDMBA), the diacid and dianion of succinic acid, and ethanolamine. These fits are then compared to angles derived from molecular simulation in order to provide a basis of comparison. Dihedral angles are given in **Table 1**. For each molecule the average RMSD of fit was less than .4. This indicates a good quality fit for all molecules, as if this average RMSD is above .4, it suggests that either the input structures are poor and do not fit the molecule, or that the dipolar data input is high in error. Because the dipolar data fits are difficult to iterate further and improve, the easiest way to get better quality fits is to check minimization schemes and use better structures. Both MMF94 minimization (RMS Gradient of .01) and MM2 minimization schemes were used to generate structures. MMF94 gave lower average RMSDs however MM2 was very close and much faster, so it was chosen for all of the molecules.

| Molecule | Observed Gauche Angle | Simulated Gauche Angle | Observed Trans Angle | Simulated Trans Angle |
|---|---|---|---|---|
| TMEDO | 78 | 65 | 180 | 180 |
| B Alanine | 73 | 56 | 180 | 180 |
| TMDMBA | 36 | 110 | 180 | 180 |
| Succinic Acid | 79 | 74 | 175 | 175 |
| Succinate Dianion | 52 | 60 | 180 | 180 |
| Ethanolamine | 78 | 56 | 180 | 180 |

**Table 1:** A table of observed dihedral angles fit using RSpin with lowest energy conformers displayed for comparison.

It is important to note that the simulated angles do not account for solvent effects and have very difficult and time consuming methods for doing so. RSpin can only fit on couplings gathered from solution and as such is forced to account for solvent effects. To demonstrate this difference, instead of fitting to 360 one degree rotation files, RSpin was only given structures generated from a conformer search and made to select which lowest energy conformer most accurately reflects the NMR data. The results of this experiment are shown in **Table 2**.

| Molecule | Lowest Energy Gauche Angle | Selected Angle |
|---|---|---|
| TMEDO | 65 | 82 |
| B Alanine | 56 | 56 |
| TMDMBA | 110 | 85 |
| Succinic Acid | 74 | 91 |
| Succinic Dianion | 60 | 76 |
| Ethanolamine | 56 | 56 |

**Table 2**: A table of the lowest energy conformer and the conformer selected by RSpin to best represent the NMR data

It is important to note that tests will not be consistent with each other for a number of reasons. Firstly, in the latter experiment, the number of conformers for each molecule was in the range of 2 to 500 depending on how many energy minima the conformer generator can find. For succinic acid, there were ~300 conformers in the 70-80 degree range and as such the lowest energy minima were found in this range. For ethanolamine, RSpin only had to choose between two structures and therefore has high accuracy with fewer conformers. Secondly the structures are completely different. In the former experiment all atoms are essentially static except for any displacements made necessary after rotations about the central sigma bond. If two groups are to collide after rotation, this is not accounted for due to the nature of the rotation program, thus producing some chemically invalid structures and poorer fits. In the latter experiment, all conformers are chemically valid and have not only different dihedral angles but the groups are minimized, and thusly rotated at different angles to avoid collisions depending on the dihedral angle.

## CONCLUSIONS / FUTURE WORK

Combining data from each type of coupling allows for fitting of dihedral angles in rotationally flexible systems. The current embodiment only functions for simple disubstituted ethane fragments however. The current embodiment also assumes good NMR fits, and good structural inputs, which may not always be given. Finally, this method assumes tetrahedral geometries on the central carbons, as the HLA/Karplus relations will not function accurately without this geometry assumption.

After this investigation however, one thing is still unclear, and this is the overall effect of solvent on backbone angles. Further investigation must be carried out using a variety of solvents on a test molecule to examine trends in dihedral angle given solvent characteristics like dielectric constant, proticity, and pKa. Another SURF student from 2012, Gregory Moore, found trends in fraction gauche in a large list of solvents, so if dipolar couplings for his molecule were found and fraction gauche values substituted in, one could find the exact trend between solvent and dihedral angle. In any case, RSpin and more importantly, the theory behind it, have proven to be useful tools in conformational studies.

## AUTHOR INFORMATION

### Corresponding Author

Email address to which correspondence may be sent: smirpuri@caltech.edu. Alternatively, correspondence may be sent to mirpurisuraj1@gmail.com.

### Present Address

[†] 1200 E California Blvd, MSC 711, Pasadena, CA 91126

## ACKNOWLEDGMENT

We would like to thank Dr. Roberts for his guidance in this project, Dr. Bill Carroll and Dr. Bright Emenike for their help in the mathematical foundations of this project, Albert Smith, Andrew Bendelsmith, and Yichen Liu for their previous work on this subject, and all of the members of the Roberts group, in particular Ben Arbeiter, for their support and data. We would also like to acknowledge the Caltech SURF Office and Robert T Herzog for financial support.

## REFERENCES

[1]. Chidichimo, G., P. Formoso, A. Golemme, and D. Imbardelli. "Conformation of Succinic Acid: Its PH Dependence by Licry-NMR Analysis." *Molecular Physics* 79.1 (1993): 25-38. Print.

[2]. Freudenberger, J. Christoph, Sebastian Knor, Kyryl Kobzar, Dominik Heckmann, Thomas Paululat, Horst Kessler, and Burkhard Luy. "Stretched Poly(vinyl Acetate) Gels as NMR Alignment Media for the Measurement of Residual Dipolar Couplings in Polar Organic Solvents." *Angewandte Chemie* 117.3 (2005): 427-30. Print.

[3]. Gil, Roberto R., Chakicherla Gayathri, Nicolay V. Tsarevsky, and Krzysztof Matyjaszewski. "Stretched Poly(methyl Methacrylate) Gel Aligns Small Organic Molecules in Chloroform. Stereochemical Analysis and Diastereotopic Proton NMR Assignment in Ludartin Using Residual Dipolar Couplings And Coupling Constant Analysis." *The Journal of Organic Chemistry* 73.3 (2008): 840-48. Print.

[4]. Kramer, F., M.V. Deshmukh, H. Kessler, and S.J. Glaser. "Residual Dipolar Coupling Constants: An Elementary Derivation of Key Equations." *Concepts in Magnetic Resonance* 21A.1 (2004): 10-21. Print.

[5]. Thiele, Christina M., and Stefan Berger. "Probing the Diastereotopicity of Methylene Protons in Strychnine Using Residual Dipolar Couplings." *Organic Letters* 5.5 (2003): 705-08. Print.

[6]. Thiele, Christina M., Volker Schmidts, Benjamin Böttcher, Iria Louzao, Robert Berger, Arnold Maliniak, and Baltzar Stevensson. "On the Treatment of Conformational Flexibility When Using Residual Dipolar Couplings for Structure Determination." *Angewandte Chemie International Edition* 48.36 (2009): 6708-712. Print.

[7]. Yan, Jiangli, Frank Delaglio, Andreas Kaerner, Allen D. Kline, Huaping Mo, Michael J. Shapiro, Tim A. Smitka, Gregory A. Stephenson, and Edward R. Zartler. "Complete Relative Stereochemistry of Multiple Stereocenters Using Only Residual Dipolar Couplings." *Journal of the American Chemical Society* 126.15 (2004): 5008-017. Print.

[8]. Altona, Cornelis, Robert Francke, Rudy De Haan, Johannes H. Ippel, Godefridus J. Daalmans, Aldert J. A. Westra Hoekzema, and John Van Wijk. "Empirical Group Electronegativities for Vicinal NMR Proton-proton Couplings along a C-C Bond: Solvent Effects and Reparameterization of the Haasnoot Equation." *Magnetic Resonance in Chemistry* 32.11 (1994): 670-78. Print.

[9]. Losonczi, J. "Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition." *Journal of Magnetic Resonance* 138.2 (1999): 334-42. Print.

[10]. Sánchez-Pedregal, Víctor M., Raquel Santamaría-Fernández, and Armando Navarro-Vázquez. "Residual Dipolar Couplings of Freely Rotating Groups in Small Molecules. Stereochemical Assignment and Side-Chain Conformation of 8-Phenylmenthol." *Organic Letters* 11.7 (2009): 1471-474. Print.

[11]. Thiele, Christina M. "Use of RDCs in Rigid Organic Compounds and Some Practical Considerations concerning Alignment Media." *Concepts in Magnetic Resonance Part A* 30A.2 (2007): 65-80. Print.

[12]. Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek and Geoffrey R Hutchison; "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform" *Journal of Cheminformatics* **2012**, 4:17.

[13]. Smith, Albert A. “Determination of the Dihedral Angle of the Monoanion of Succinic Acid”. *Summer Undergraduate Research Program*. SURF, (2009)

[14]. Bendelsmith, Andrew J. “The Determination of the Dihedral Angles of 1,2-Disubstituted Ethanes in Aqueous Lyotropic Solutions”. *Summer Undergraduate Research Program*, SURF, (2011)

[15]. Yichen Liu. “Conformation Analysis of Succinate Dianion of Liquid Crystal Media”. *Summer Undergraduate Research Program*. SURF, (2011)