

A Novel Method for Computing Conformer Populations

RAJ MIRPURI

smirpuri@caltech.edu

California Institute of Technology

Department of Chemistry and Chemical Engineering

1200 E California Blvd

Pasadena CA, 91126

Abstract

A new method for computing conformer populations from NMR scalar coupling data is shown. This method is significantly more quantitatively accurate than any previously used method, it is bounded correctly, and it backs out additional useful information such as the free energy landscape that cannot be found by any other method. This model is built on a generalization of the previous model to correctly count accessible equilibrium states, and it is rooted in statistical mechanics, giving this model more theoretical grounding than the previous model. Finally, using this model explains several experimental details that could not be explained by prior methods.

INTRODUCTION

In computational chemistry, particular attention is often given to molecular structure during a simulation. There are several theories in place to track molecular trajectories at various levels of rigor, including molecular dynamics and quantum mechanical methods such as Hartree-Fock theory^[1-2]. However, all of these methods are confronted by the same issue: it is difficult to demonstrate how well they work as there is still no good way to examine molecular structure in solution without significantly perturbing the structure. One of the few methods that exists for doing this in general is nuclear magnetic resonance spectroscopy (NMR). NMR spectroscopy relies on the fact that molecules are basically spin systems, and when they encounter a large magnetic field, there is a splitting in energy between spin up and spin down states. By exciting transitions

in these spin states and watching how it affects nearby spins, the environment of every spin can be examined and assembled into a concrete picture of the molecule in solution. Most of this information about the connections between spins is given by the scalar coupling between those spins, also known as the J-coupling^[3-4]. To make this less abstract, consider the ethane system shown in **Figure 1**.

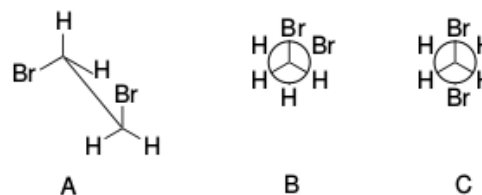


Figure 1: Sawhorse and Newman projections for a simple disubstituted ethane fragment. Part A shows that the substituted groups are free to rotate around the carbon carbon sigma bond. Part B shows what is known as a

“gauche” configuration and part C shows what is known as a “trans” configuration.

As molecules are free to rotate around sigma bonds, such as the bond between the two carbons in part A of the figure, there exist many rotational conformers or rotamers as shown in parts B and C. Part B shows the “gauche” rotamer, part C shows the “anti-” or “trans” rotamer. These two broad categories of rotamers are differentiated solely by the dihedral angle: the angle in between the substituent groups if the molecule is flattened as shown in the Newman projection. In the gauche configuration the dihedral angle range is between 30 to 90 degrees or 330 to 270 degrees as one group can either be to the right or left of the other, and in the trans configuration the dihedral angle is between 150 and 210 degrees^[5]. In any case, every molecule with an ethane fragment will have an associated fraction of the population in the gauche configuration and an associated population in trans based on which configuration minimizes the free energy. Furthermore each fraction will have a specific angle between the groups that represents an energy minimum, and this angle is known as the gauche angle or the trans angle depending on which fraction is being discussed. To relate this to J-couplings, the J-coupling is basically a function of the dihedral angle, so if one is known the other can be obtained from some empirical correlations known as the Karplus curves^[6]. This provides an easy method for computing the equilibrium gauche and trans angles. Even with this, there is still one issue: the fraction gauche cannot be calculated from these J-couplings unless some assumptions are made. The basic idea is to calculate this fraction by assuming the observed value for the J-coupling is a sum of two contributors, a trans configuration J-coupling and a gauche configuration J-coupling, each weighted by the fraction trans and fraction gauche. When values for the J-coupling of each configuration are obtained from Karplus curves and plugged in, the fraction gauche can be easily solved. This approach is lacking in several respects however. Firstly, the Karplus curves are sinu-

soidal, so by selecting only two conformers to represent the continuum of states available a two point approximation to a sine wave has been made, which is very inaccurate. In fact this method often reports fractions above 1, because there is no hard constraint on what the fraction should be when only the two most likely states are counted. If lower probability states are included in the calculation this will bring down the fraction in those higher probability states and count correctly. Clearly, a more subtle approach is needed. A new method has been developed that stems from statistical mechanical treatment of the system that is correctly bounded on the fraction gauche, similarly computationally inexpensive, and overall more accurate.

THEORY

This new method is built on essentially the same principle as the old method: the observed J-coupling is the sum the J-coupling at every angle weighted by the fraction of the population in that angle. This definition follows directly from the definition of expectation value, and it really is just a generalization of the old method from two points to many points. This can also be easily made continuous by replacing the sum with an integral over the dihedral angle, however integrals are harder to work with in the optimization mechanics. For this reason, a discrete approximation will be used here without loss of generality, but with one point in the spectrum per degree of dihedral angle. To make this more concrete, for one rotation around a sigma bond there are 360 states the system could be partitioned into conveniently, by degrees of rotation. The observed J-coupling is some average of all of these 360 states. With this established the following difference is minimized for a set of J-couplings observed from experiment:

$$J_{obs} - \sum_{\phi=1}^{360} P(\phi)J(\phi)$$

$J(\phi)$ is the Karplus curve for the system

in question, as it relates the J-coupling to the dihedral angle at every angle, and $P(\phi)$ is a probability distribution that is varied to minimize this difference. This probability distribution is key to the problem as there will be one single probability distribution for every observed J-coupling, and this probability distri-

bution is related to the free energy landscape of the system via Boltzmann factors. If the probability distribution is to be supplied and varied, the general form of this distribution must be obtained. This can be easily backed out by looking at the example free energy landscape shown in **Figure 2**.

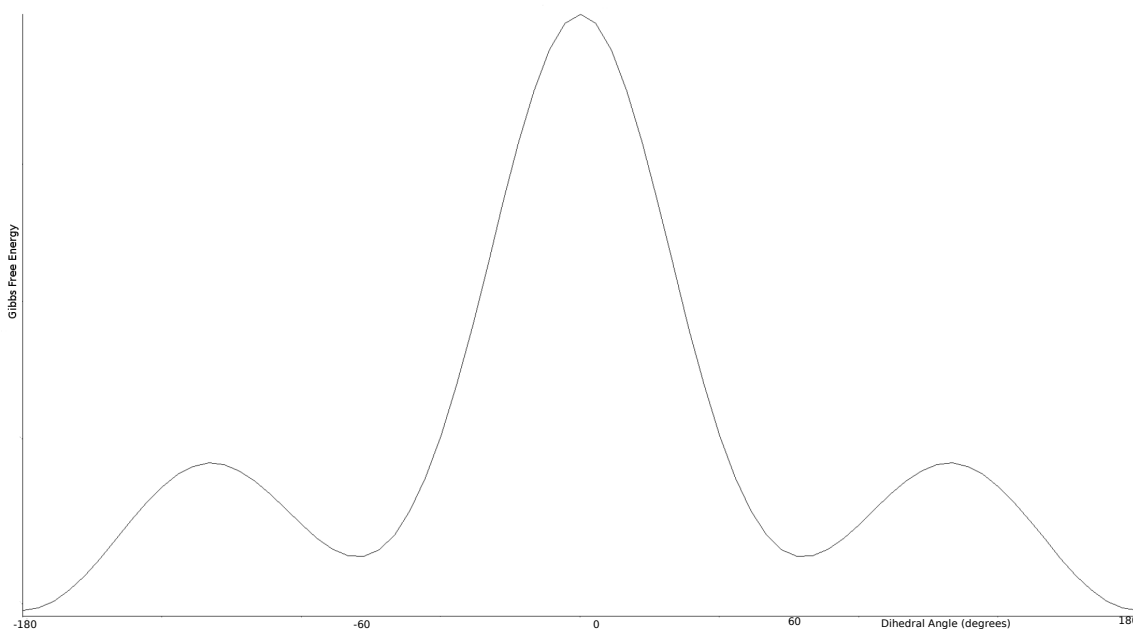


Figure 2: A free energy landscape as a function of dihedral angle. There are three minima, associated with the two gauche configurations at positive and negative sixty degrees of rotation and one associated with the trans configuration at one hundred eighty degrees of rotation. The lower the energy, the higher the probability of finding the molecule at that angle in equilibrium.

There are three minima in energy corresponding to the two gauche configurations and one trans configuration. This means that the probability distribution should have 3 peaks, one for each gauche configuration and one for the trans configuration. Furthermore, the two peaks for each gauche configuration should be symmetric, and since the fractions in each configuration may vary the area of these peaks should not be fixed. Therefore the peak height and standard deviation should be varied. Also, consider that each configuration has one single angle that minimizes the free energy, therefore each peak in the distribution should have a mean centered around these angles. This is really all that can be stated about the nature

of the peaks in the distribution other than that they must be Gaussian peaks, as the central limit theorem requires that for a large sample size the distribution of conformers in these states will be normal. This means there are 6 parameters to vary in total: gauche and trans peak heights, standard deviations, and means. We can then minimize the difference between the observed couplings and the calculated couplings to find what probability distribution, hence, what free energy landscape, fits the data the best. Numerical methods such as constrained optimization by linear approximation and sequential least squares programming can be used to do the minimization quickly and a single probability distribution can be fit to a

set of six J-couplings in less than one second.

RESULTS

The new method for calculating fraction gauche not only gives the complete free energy landscape, but also gives an error measure to know how good the fit is. In practice the root mean square deviation between the observed and calculated values is on the order of 10^{-5} , meaning that fantastic fits are usually achieved. Rarely, the error rises to being on the order of a J-coupling, about 7 Hz, which implies more often that the J-couplings inputted are incorrect than there is no probability distribution that fits well. Also, this method gives more intuitive information about the system, as the standard deviations of the probability distributions are a rough measure of how sharply divided the different configurations are. For example a large standard deviation to the gauche peak means that there is not in actuality a single gauche angle but a broad range of possible states. This clarifies dynamics in many real situations, such as those for the molecule TMEDO, a common polymer initiator. In experiment it was found that TMEDO has a higher substituent electronegativity than TMDMBA, a similar molecule, and this means it should have a lower gauche fraction from the Karplus curves. However, the old method for calculating gauche fractions showed a contra-

dition, because the fractions for TMEDO and TMDMBA were both above 1, and TMEDO had a higher fraction gauche than TMDMBA. When the new method was applied, it was instantly seen why this discrepancy arises: the distribution for both TMEDO and TMDMBA is very broad in the gauche configuration, so there were many contributing states ignored and in reality TMEDO does have a lower gauche fraction than TMDMBA. The older method failed to work as the TMEDO peak is higher but more broad. This new model for calculating gauche fractions is far more accurate then, and much more intuitive.

REFERENCES

1. BJ Alder, TE Wainwright - "Studies in Molecular Dynamics. I. General Method". J Chem Phys 31, 459 (1959)
2. A Szabo, NS Ostlund - "Modern Quantum Chemistry". Dover Publishing. (1996)
3. J Roberts - "The ABCs of FT-NMR". University Science Books. (2000)
4. M Levitt - "Spin Dynamics". Wiley Publishing. (2008)
5. W Klyne, V Prelog - "Description of Steric Relationships Across Single Bonds". Experientia 16:521-23. (1960)
6. M Karplus - "Vicinal Proton Coupling in NMR". J Am Chem Soc 85:2870-2871. (1963)