Decomposition of Two Dimensional $^{13}$C-$^{13}$C Solid State NMR Correlation Spectra for Structural Determination of Proteins

C. Ramos$^1$, Z. Gan$^2$, I. Hung$^3$

1. Florida State University, Tallahassee, FL 32310
2. National High Magnetic Field Laboratory, Tallahassee FL 32310

Introduction

Proteins are sequences of amino acids that fold into different 3D structural conformations. The function and mechanism of proteins are closely related to their structure. Therefore, the distance between amino acids is crucial information for refining protein structures.

By performing 2D $^{13}$C-$^{13}$C NMR correlation experiments it is possible to obtain qualitative distance information between amino acids. This information is obtained through analysis of the rate of build up of the cross-peaks intensities. However, the large amount of sites causes spectral overlap and obscures the cross-peaks due to long range interactions. In order to observe the entire spectrum, singular value decomposition (SVD) is used to categorize and separate curves according to rate of build up. Thus, isolating the cross-peaks due to long range interactions from the intense close range signals.

Singular Value Decomposition (SVD)

\[
R_{ij} \cdot A = C
\]

Cross-peak variations showing different build up rates for NMR spectrum

Exponential Growth Rate Constants

Exponential growth rates composing the different build up curves

Amplitudes From SVD Curve Fit

Amplitudes describing each component of the sum of exponentials that fit the experimental data

Conclusion

We found that SVD analysis creates a robust fit using a sum of exponentials model with six terms. After preliminary testing, it shows stable results when decomposing and separating out build up curves for specified ranges of build up rates.

By developing a successful preliminary mathematical algorithm, further research should be done to enhance the specificity of the model. Further developments of the model include increasing the density of the coefficient matrix to better describe the data to eliminate values that cause instability. Additionally, it should be applied to the entire 2D spectrum to observe how well the data can be simulated. This will give information about its ability to describe the long range build up curves which are difficult to observe in the spectrum.

Acknowledgements

Thank you Dr. Zhehong Gan and Dr. Ivan Hung for their mentorship during the 2015 REU program. Also a thank you to the National High Magnetic Field Laboratory Center for Integrated Research Learning. This project was sponsored by the National Science Foundation DMR-1157490.

References


X-ray Diffraction Structure of GB1

Figure 1. Crystal structure of GB1 protein

Figure 2. Two Dimensional overlay NMR plot of $\tau=15$ ms (black) and $\tau=1000$ ms (red). Specified area indicates A20CA-CB binding site (blue arrow). Other binding site analyzed not present due to the use of a fast mixing time during full assignment of the protein structure.

NMR Contour Plot Overlay

Figure 3. Normalized experimental cross-peak data for binding site A20CA-CB (blue) and 3x scaled long range signal (black) with SVD curve fit (red) for 1750 ms.

Figure 4. Range of exponential rate constants from 1.17 ms to 250 ms for a mixing time of 1750 ms.

Figure 5. Normalized amplitude changes vs. composition of exponential growth rates approximated with SVD curve fit.