TCA Cycle Turnover And Serum Glucose Sources By Automated Bayesian Analysis Of NMR Spectra

Matthew E. Merritt[1], Shawn Burgess[1], F. Mark Jeffrey[1], A. Dean Sherry[1,3], Craig Malloy[1,4], and G. Larry Bretthorst[2]

1. University of Texas Southwestern Medical Center, Mary Nell and Ralph B. Rogers Magnetic Resonance Center, Dallas, TX

2. Department of Chemistry and Radiology, Washington University, St. Louis, MO

3. Department of Chemistry, University of Texas at Dallas, Dallas, TX

4. VA North Texas Health Care System, Dallas, TX

Abstract. Changes in sources of serum glucose are indicative of a variety of pathological metabolic states. It is possible to measure the sources of serum glucose by the administration of deuterated water to a subject followed by analysis of the ^2H enrichment levels in glucose extracted from plasma from a single blood draw by ^2H NMR. Markov Chain Monte Carlo simulations of the posterior probability densities may then be used to evaluate the contribution of glycogenolysis, glycerol, and the Kreb's cycle to serum glucose. Experiments with simulated NMR spectra show that in spectra with a S/N of 20 to 1, the resulting metabolic information may be evaluated with an accuracy of about 4 percent.

SOURCES OF SERUM GLUCOSE

With the current epidemic proportions of diabetes mellitus, the study of *in vivo* metabolism is being driven to a greater sophistication. It is becoming increasingly apparent that simple oral glucose tolerance tests are a valuable but such tests do not provide detailed insights into total body metabolism. Further information about the sources of serum glucose would greatly aid in describing the metabolic basis of diabetes. One method for assessing the sources of serum glucose first proposed by Landau, et. al.[1] uses orally administered ^2H_2O as an isotopic tracer of glucose synthesis within a living organism. Following administration of a bolus of ^2H_2O orally, blood is drawn from the subject. Typically enough deuterated water is given to achieve 0.5% total body water enrichment. Following extraction, the glucose is subjected to a carbon-by-carbon degradation and the site specific enrichment by deuterium is measured by mass spectroscopy. This technique has already been used in a variety of different subjects including low birth weight infants, children, and lactating women. While the mass spectroscopy analysis has been useful, it is complicated and time consuming due to the number of steps in degrading the carbon skeleton of glucose. Recently, ^2H nuclear magnetic resonance (NMR) measurements of a derivative of glucose have yielded comparable results, *albeit* with decreased sensitivity relative to the mass spectroscopy.[2] However, the ^2H NMR spectrum allows the measurement of the isotopic enrichment at each position of the glucose simultaneously, which may be an overwhelming advantage.

Isotopomer analysis by NMR typically differs from more standard chemical applications of NMR in that the spectra are recorded for the same compound under

CP707, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 23rd International Workshop*, edited by G. Erickson and Y. Zhai

many different conditions of isotopic labeling. The spectra from run to run will have resonances at the same chemical shift, but differences of amplitudes of certain peaks within the spectrum will encode the relevant metabolic information. The repetitive nature of the spectra lends their analysis to a Bayesian formulation (prior information is abundant). Not only do accurate models exist of the NMR spectra, but the NMR signal amplitudes are related to the relevant metabolic parameters by simple ratios of peak amplitudes. By allying the Bayesian formulation of the probability densities with Markov Chain Monte Carlo analysis, it is possible to accurately estimate the metabolic parameters and the associated uncertainties.

In this paper, the practical advantages of the Bayesian/MCMC methodology will be outlined for estimating metabolic parameters by NMR isotopomer analysis.

Deuterium NMR for Assessment of Serum Glucose Origins

Landau first proposed the use of a deuterated water tracer to study glucose metabolism.[1] The method is based upon a detailed understanding of the pathways related to glucose production *in vivo*. Figure 1a displays a simplified representation of the pathways with labels indicating where a deuterium tracer is incorporated into the glucose molecule. If the glucose was derived from the breakdown of glycogen, then it will pass through the glucose-6-phosphate metabolite, which is in fast equilibrium with fructose-6-phosphate. During the isomerization reaction the H2 will attain the ^2H enrichment level of the body water. If the glucose was synthesized from glycerol obtained from fat stores in the body, it will pick up a deuterium at H5 in addition to H2. Finally, if the precursors to the glucose molecule were derived from the TCA cycle, the H6(s) position will also be labeled. Figure 1b shows a typical deuterium NMR spectrum that can be obtained from a sample of blood glucose that has been derivatized to 1,2-O-isopropylidene-D-glucofuranose, referred to hereafter as monoacetone glucose (MAG). This derivative of glucose has superior NMR properties due to the elimination of the boat and chair conformations of the six-membered glucose ring.

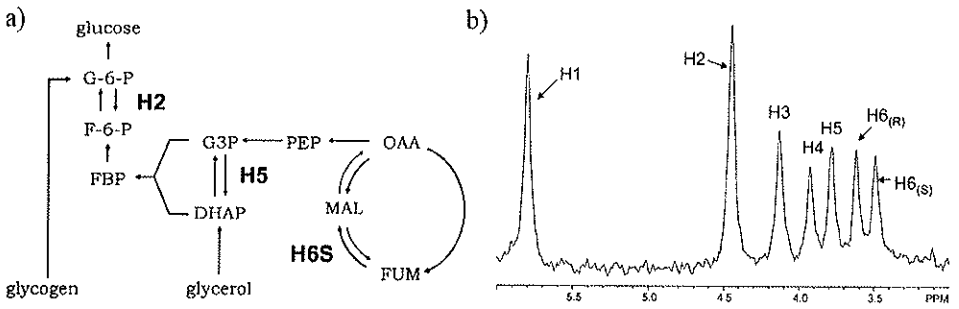

FIGURE 1. a) Schematic of deuterium tracer entering a glucose molecule in the bloodstream. b) ^2H NMR spectrum of a typical sample following derivatization. Note that ^2H label enters other sites within the molecule, but does not interfere with the metabolic estimates.

There are several key issues involved in the ^2H spectroscopy associated with the deuterium isotopomer analysis. The central issue associated with these experiments is quantifying the uncertainty associated with an estimate of a metabolic parameter. Since many of the experiments carried out are on a single sample and the low sensitivity precludes making multiple measurements, the Bayesian approach to estimating uncertainty is the appropriate formulation. The question that must be answered is, "At what signal to noise in the NMR spectrum do we achieve a satisfactory uncertainty in the metabolic measurements?" Being able to answer this question accurately has tremendous implications for throughput of the NMR samples. A corollary to this question is "Is the metabolism of the patient represented by an NMR sample really different from that of a healthy person?" Since multiple measurements are not feasible, how do we assign an uncertainty to the metabolic inference? Obviously, the Bayesian formulation utilizing all the prior information about the NMR spectrum will be optimal for making estimates of the uncertainties. Another issue is that ideally any program used to make the metabolic estimates should be transparent to the end user to avoid a lengthy training period in the use of the software. Also, with the marginal signal to noise ratios sometimes achieved in these experiments due to low concentrations of ^2H labeled tracers, any analysis should be carried out by the computer to avoid user bias in the spectral fitting stage.

Experimental

The sample spectrum analyzed here was generated by extracting blood (20mL) from a single volunteer following a 42 hour fast according to a protocol approved by the Institutional Review Board. The sample was centrifuged at 2500 rpm to separate plasma from erythrocytes and extracted with perchloric acid to remove proteins. Following extraction the blood glucose was converted to the MAG derivative.[4] The sample was dissolved in natural abundance acetonitrile and NMR was performed on a 600MHz ^1H (92MHz ^2H) Unity Inova console (Varian Inc., Palo Alto, CA) with a probe optimized for deuterium detection (Nalorac, Inc., Martinez, CA). The spectrum was acquired with ^1H waltz decoupling for deuterium line narrowing at a temperature of 50 degrees celsius. A ^2H 90 degree pulse was used with an acquisition time of 1 second with no delay between scans. The short T_1s of the deuterium nuclei preclude the need for post acquisition weighting of peak intensities. The spectral width was 10ppm.

NMR and Metabolic Models

Landau[1] first expressed the equations for serum glucose sources:

$$\text{glycogenolysis contribution} = 1-(H5/H2) \qquad (1)$$

$$\text{glycerol contribution} = (H5 - H6(s))/H2 \qquad (2)$$

$$\text{Krebs cycle contribution (PEP)} = H6(s)/H2 \qquad (3)$$

H5, H2, and H6(s) are the enrichments at the specific carbons of monoacetone glucose (figure 1). There are assumed to be 3 sources of serum glucose including breakdown from stored glycogen (1), synthesis from the glycerol head groups obtained from stored fatty acids (2), or synthesis from phosphoenol pyruvate through the Kreb's cycle (3). This defines a fourth equation:

$$H2 + H5 + H6(s) = 1 \qquad (4)$$

These four equations comprise the metabolic model. The correct model for the NMR data was written down in its full form in the papers by Bretthorst,[5] but can be summarized by equation 5.

$$M(t) = \sum_{i=1}^{3} H_i e^{\{2\pi f_i (t+t_0) - a_i t\}} \qquad (5)$$

where H_i is the amplitude of one of the three relevant resonances. The exponential term is a complex sinusoid with frequency f_i and decay rate a_i. The t_0 term represents a time offset that often appears in experiments due to the effects of filters in the acquisition hardware. The NMR model also included elements for a DC offset in the data as well as bad first points of the free induction decay. Each deuterium spectrum contained ten resonances total, 7 from monoacetone glucose, 1 for residual water, and two to model the large natural abundance acetonitrile resonance from the solvent. Each resonance not in the metabolic model, termed nuisance resonances, was also simulated. The Markov Chain Monte Carlo simulation computes the joint posterior probability for the three relevant amplitudes independent of the other quantities in this equation and then outputs the posterior probability for the glycogenolysis, glycerol and Krebs cycle contributions by an appropriate transformation of variables at the end of the calculations.

Simulations

The probability densities were formulated according to Bayes Theorem and simplified using the sum rule. Equation 6 is simply Bayes Theorem without the normalization term in the denominator and with the addition of the β, which is a simulated annealing parameter.

$$P(H|D,I) = P(H|I) * P(D|H,I)^{\beta} \qquad (6)$$

In order to estimate the metabolic parameters, Markov Chain Monte Carlo Simulations of the posterior probability densities were carried out. For these simulations, 50 different Markov Chains were used with Gaussian priors for the NMR parameters. Each chain was allowed to propagate for 50 steps as the annealing

parameter, β, increased linearly from 0 to 1 in steps of .02, for a total of 50 increments of β. After each step of β, 20 to 25 of the NMR parameters were varied. The slow increase allows the samples to migrate from the prior probabilities to the posterior probability without becoming trapped in local minima. For this type of experiment, the chains usually converged before β reached .2. Once the annealing was complete, each chain was allowed to propagate for 50 more steps, so that a total of 2500 samples of the posterior probability density were made. Estimates of the probability densities were made by calculating the mean and standard deviation of the 2500 samples. Histograms were generated by binning the samples. Hereafter the program is referred to as BayesMetabolite.

NMR Data Simulations And Analysis Via The BayesMetabolite Program

In order to fully test the program, a set of free induction decays (FID's) were generated by summing frequency shifted exponentially decaying sinusoids with known amplitudes. Gaussian white noise was scaled to three different amplitudes and added to the FID's to simulate decreasing signal to noise in the NMR spectrum. These FID's were analyzed with BayesMetabolite. Figure 2 shows three spectra with a decreasing signal to noise ratio from top to bottom. The top trace has a signal to noise of approximately 20 to 1. The same set of noise has been scaled and added to the FID prior to fourier transformation for each of the three spectra. Figure 3 displays the histograms generated by BayesMetabolite that correspond to the different possible contributions to the serum glucose. The set of histograms shown here were derived from the data set in the top panel of Figure 2. As would be expected, the posterior probability distributions are fairly symmetric for all three sources. It is also of note that each of the three parameters have approximately equivalent uncertainties (Table 1). When all three of the sets of data are analyzed and the histograms superimposed, the systematic error caused by scaling the same set of noise and adding it to the FID's is pronounced (Figure 3, inset). Further simulations were carried out at equivalent noise levels but with different noise sets to observe how the variation in the noise would manifest in the BayesMetabolite simulations.

TABLE 1. Mean values and standard deviations as measured by BayesMetabolite

Source	Noise	Noise 2	Noise 3
Glycogen	.07 +/- .04	.11 +/- .04	.11 +/- .04
Glycerol	.56 +/- .05	.52 +/- .05	.50 +/- .04
Kreb's Cycles	.37 +/- .04	.37 +/- .03	.39 +/- .03

Different sets of noise do not significantly change the estimates made by the program, as can be seen by following the rows across table 1. Each estimate in table 1 was made at the same rms noise level.

Another way to study the effects of noise upon the metabolic measurements is illustrated in Figure 4. The bottom panel illustrates data that was acquired for a study with a test subject that probed the depletion of glycogen due to a 42 hour fast. The signal to noise in this spectrum would be described as fair. Previously the spectrum

would have been analyzed by fitting the peaks with Lorentzian lines to approximate the areas.

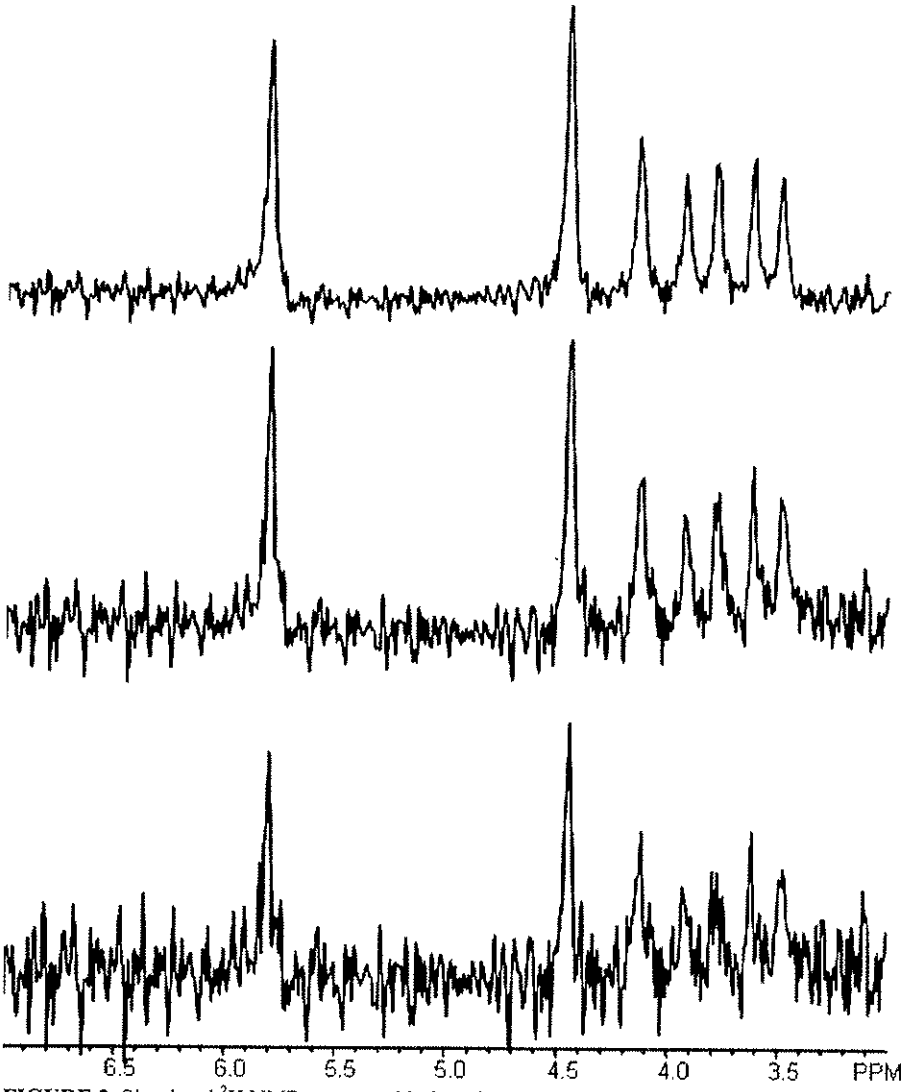

FIGURE 2. Simulated ^{2}H NMR spectra with three increasing levels of Gaussian white noise added.

Analysis by BayesMetabolite gave estimates of contributions from glycogen, glycerol, and Kreb's cycle as .09, .11, and .80, respectively. As would be expected, the glycogen contribution is severely decreased with most of the need being taken up by Kreb's cycle activity. To illustrate the uncertainty in the metabolic estimates due to noise in the NMR spectrum, another spectrum was simulated, one drawn from the distribution inferred by BayesMetabolite at the level of one standard deviation. This corresponded to metabolic estimates of .03, .06, and .91 for glycogenolysis, glycerol,

and Kreb's cycle contributions. Difference spectra were taken from the simulations and data to produce the two traces that show only noise.

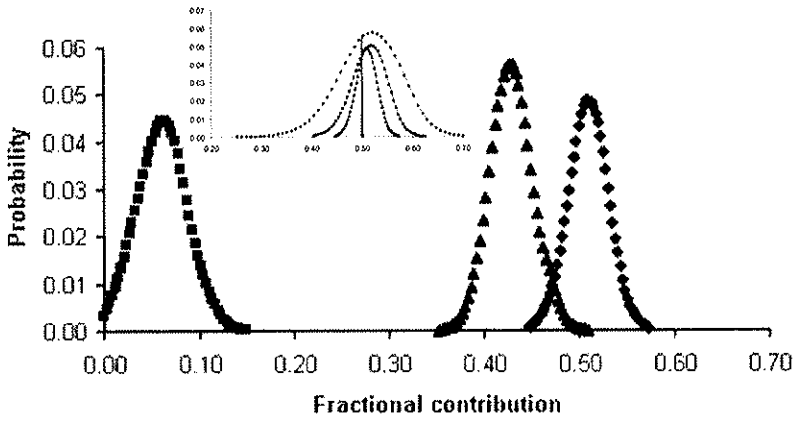

FIGURE 3. Histograms of the probability densities for the fractional contribution of glycogenolysis (squares), glycerol (triangles), and Kreb's cycle (diamonds) contributions to total glucose from the spectrum with the highest signal to noise in Figure 2. The actual values were glycogenolysis=.1, glycerol=.4, and Kreb's cycle =.5. The inset shows the change in the probability density associated with the glycerol contribution as the S/N decreased.

To the eye, the two residuals do not appear to be significantly different. Without an unbiased method for assessing the quality of the fits, there is no way to distinguish the quality of the two residuals.

Discussion

The results summarized in Table 1 affirm the accuracy of the BayesMetabolite program. The MCMC simulation consistently made estimates of the metabolic parameters that were the same as the beginning parameters used to synthesize the FID's to within the estimated uncertainty, as illustrated by the histograms of Figure 3. Surprisingly, even at the lowest noise level chosen, (Figure 2, top) the mean of the MCMC simulations was one standard deviation away from the actual value. Since the simulation exactly reproduced the starting metabolic values with no noise added to the spectrum, the observation must be made that even at signal to noise levels of approximately 20 to 1, the metabolic values are subject to error at a level that might be construed as physiologically relevant.

This point is emphasized in dramatic fashion in Figure 4. The residuals calculated from the data and the MCMC simulations from the mean and at one standard deviation are impossible to distinguish from one another by eye. From a physiological stand point, the two answers would have dramatic consequences if there was no estimate of the uncertainty. If the glycogen contribution was really at a three percent level serious doubts about the health of the subject's liver would be raised.

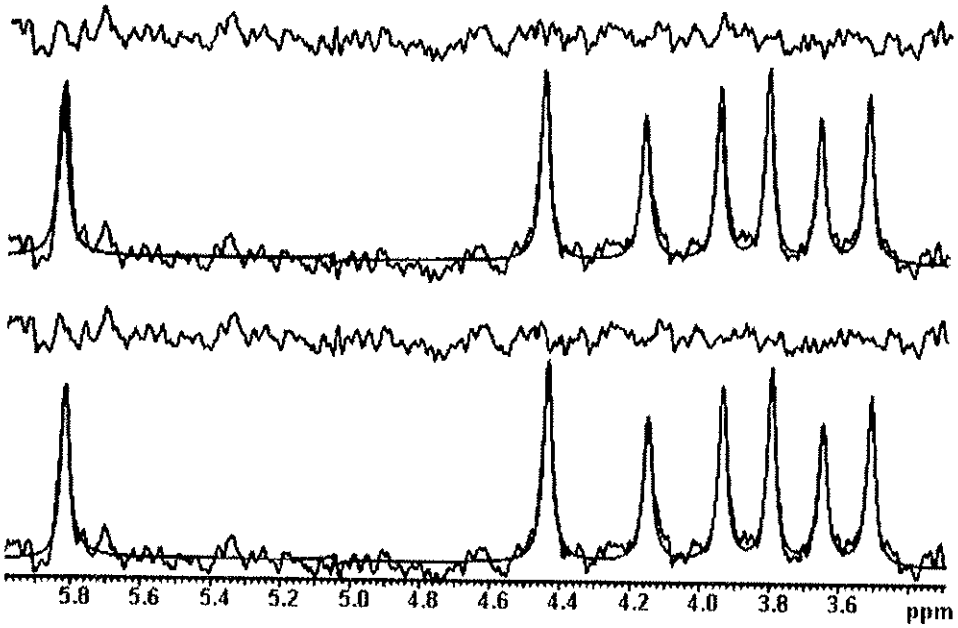

FIGURE 4. (bottom) ^{2}H NMR spectrum and fit made with metabolic parameters drawn from the mean of the probability distribution with the resulting residual. (top) Spectrum and residual drawn from the probability density at one standard deviation.

Conclusions

It has been shown that BayesMetabolite can accurately estimate the sources of serum glucose through ^{2}H NMR of a derivative of glucose. Simulated NMR data was used to test the accuracy of the program. During the tests, the deleterious effects of noise in the NMR FID upon the metabolic estimates were well characterized. Even at signal to noise levels as high as 20 to1, errors in the metabolic estimates of one standard deviation are possible. From a clinical standpoint, the need for improved signal to noise is emphasized if metabolic estimates within four percent of the true value are needed.

With BayesMetabolite, it is possible to rigorously assign an uncertainty to a single metabolic measurement. Clearly, this has enormous benefit in diagnosing pathological states correctly and allows the practitioner to craft the experiment to the level of accuracy that is required.

ACKNOWLEDGMENTS

Dr. Merritt acknowledges the support of the NSF and the NIH-NCRR, grant 02584. Dr. Bretthorst acknowledges the support of NCI-NCRR, grant 02584.

REFERENCES

1. Landau, B. R., Wahren, J., Chandramouli, V., Schumann, W., C., Ekberg, K., Kalhan, S. C., *J. Clin. Invest.* **95**, 172-178, (1995).
2. Jones, J. G., Carvalho, R. A., Sherry, A. D., Malloy, C. R., *Anal. Biochem*, **277**, 121-126, (2000).
3. Burgess, S.C., Nuss, M., Chandramouli, V., Hardin, D. S., Rice, M., Landau, B.R., Malloy, C.R., Sherry, A.D., Anal. Biochem. 318, 321-324, (2003).
4. Burgess, S. C., Weis, B., Jones, J. G., Smith, E., Merritt, M. E., Margolis, D., Sherry, A. D., Malloy, C. R., *Anal. Biochem*, **312**, 228-234, (2003).
5. Kotyk, John J.; Hoffman, Norman G.; Hutton, William C.; Bretthorst, G. Larry; Ackerman, Joseph, J. H., *J. Mag. Res.*, **98(3)**, 483-500, (1992).