

NMR BASED STUDIES OF METABOLIC PROCESS

by

JANANI VARATHARAJAN

(Under the Direction of James H. Prestegard)

ABSTRACT

Metabonomics, the study of small molecule metabolic profiles, together with proteomic and transcriptomic information gives a rich picture of the function of living organisms. In order to understand the function of the multiple enzymes involved in these systems we need massive quantities of real-time data. This often requires the development of new analysis procedures that can model interactions of multiple components. The methodology presented here is based on Nuclear Magnetic Resonance (NMR) observation of real time variation of metabolites, on new spectral analysis methods that can deconvolute the signatures of each molecular species, and network modeling methods that can identify interactions in complex systems. As a model system, the enzyme acetyl-CoA synthetase is selected. The role of this enzyme is to activate acetate to acetyl-coenzyme A (Ac-CoA). The importance of this enzyme is that it provides the cell a two-carbon metabolite which can be used in energy generation processes and functional modification of both proteins and carbohydrates.

Index words: Acetyl-CoA Synthetase, NMR (Nuclear Magnetic Resonance), STOCSY (Statistical Total Correlation Spectroscopy), Monte Carlo simulations, Network modeling, Metabolism.

NMR BASED STUDIES OF METABOLIC PROCESS

by

JANANI VARATHARAJAN

B.Tech., Mepco Schlenk Engineering College, Anna University, India, 2007

A Thesis Submitted to the Graduate Faculty of The University of Georgia

in Partial Fulfillment of the Requirements for the Degree

MASTERS OF SCIENCE

ATHENS, GEORGIA

2009

© 2009

JANANI VARATHARAJAN

All Rights Reserved

NMR BASED STUDIES OF METABOLIC PROCESS

by

JANANI VARATHARAJAN

Major Professor: James H. Prestegard

Committee: H.B.Schuttler
Andrew T. Sornborger
Maor Bar-Peled

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2009

DEDICATION

I wish to thank my Parents, Varatharajan and Ranganayaki Varatharajan for motivating me to go to graduate school and for believing in me and my fiancé Sriram who always supported and encouraged me to do the best in all matters of life. To them I dedicate this thesis.

ACKNOWLEDGEMENTS

Foremost I would like to thank my advisor Prof. James Prestegard who shared with me a lot of his expertise and research insight. He became for me the role model of a successful researcher. He was always accessible and willing to help his students with their research. As a result, research life became smooth and rewarding for me. He has made a great difference in my life and I am grateful to him. Prof. H.B.Schuttler, Prof. Andrew T. Sornborger, Prof. Maor Bar-Peled deserves special thanks as my thesis committee members. In particular, I would like to thank Prof. H.B.Schuttler for patiently teaching me lots about physics and simulations. I would also like to thank Prof. Andrew T. Sornborger, Prof. Maor Bar-Peled for considering me as a part of their metabolomics project and for constantly giving many research ideas. I especially want to thank John Glushka with whom I began to learn NMR. He is very friendly and always willing to help. I would like to extend my thanks to Eric, who helped me by providing data and for all the useful discussions. All my lab friends at CCRC made it a convivial place to work. In particular I would like to thank Wendy for her friendship and Victor for patiently teaching me all the basic concepts of NMR. In addition, I would like to thank all the folks in lab who inspired me in research and life through our interactions during the long hours in lab. I wish to thank my roomies Harshita, Spandana and Krithika for providing me such a homely atmosphere and also I am tempted to thank all my friends especially Shirlee, Saranya, Bharathy and Swetha. I cannot finish without saying how grateful I am with my family, my Parents, my fiancé, my in-laws, my grand parents, uncle, aunt, cousins, nephews, all have given a loving environment in which to develop.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	v
LIST OF FIGURES	viii
LIST OF TABLES.....	ix
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW	1
Statement of objective.....	1
Metabonomics.....	2
Kinetic analysis of the system.....	3
Options for the approach.....	3
Network modeling of metabolite fluxes.....	7
2 NMR ANALYSIS OF ACETYL Co-A SYNTHETASE REACTIONS.....	9
Data collection for acetyl-CoA Synthetase: ¹ H NMR for the enzymatic reaction.....	9
Conclusion.....	12
3 DATA ANALYSIS USING ALTERNATIVE PROCESSING STRATEGIES.....	13
STOCSY-Statistical Total Correlation Spectroscopy.....	13
Conclusion.....	19
4 NETWORK MODELING OF THE ACS REACTION.....	20

	Network modeling of ACS system using Monte Carlo methods.....	20
	Conclusion.....	29
5	FUTURE WORK.....	30
6	REFERENCE.....	31
	APPENDIX.....	35

LIST OF FIGURES

	Page
Figure 2.1: Production of acetyl CoA.....	9
Figure 2.2: ¹ H NMR for reaction ATP +acetate+CoA+buffer.....	10
Figure 2.3: ¹ H NMR for reaction ATP+acetate+buffer.....	11
Figure 3.1: ¹ H NMR spectra-Lactose.....	15
Figure 3.2: ¹ H NMR spectra-Sucrose.....	16
Figure 3.3: Contour plot for random mixtures of sucrose and lactose.....	17
Figure 3.4: Covariance matrix for acetyl-CoA.....	18
Figure 4.1: Network modeling for enzyme system ACS: Model 1&2	21
Figure 4.2: Network modeling for enzyme system ACS: Model 3&4	22
Figure 4.3: Network modeling for enzyme system ACS: Model 5&6.....	23
Figure 4.4: Monte Carlo simulation run for Model 1.....	24
Figure 4.5: Monte Carlo simulation run for Model 2.....	26
Figure 4.6: Comparison of Model 1& Model 2.....	28

LIST OF TABLES

	Page
Table 4.1: Rate constants for the forward reactions of Model 1.....	25
Table 4.2: Rate constants for the backward reactions of Model 1.....	25
Table 4.3: Rate constants for the forward reactions of Model 2.....	27
Table 4.4: Rate constants for the backward reactions of Model 2.....	27

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Statement of objective

The synthesis of biomolecules is a very complicated process. Multiple enzymes can share substrates, be inhibited by any number of products, and be controlled by transporters that regulate the availability of both products and substrates. Real-time data is needed to analyze and model the interactions of multiple components. A new methodology needs to be designed that is able to model complex synthetic systems and can be applied to both a well-defined system and more complex cellular systems. The methodology can be based on Nuclear Magnetic Resonance (NMR) observation of real time variation of metabolites, on new spectral analysis methods that can deconvolute the signatures of each molecular species, and also on network modeling methods that can identify interactions and control points in complex systems. (*Lindon et.al, 2004*)

In order to develop methodology, a well-controlled system of moderate complexity must be selected. Here we chose the enzyme acetyl-CoA synthetase. The product produced by the action of this enzyme, Acetyl-CoA, is important in the acetylation of a number of more complex carbohydrate products that will be future targets of research. Real-time NMR spectroscopy methods can be used to monitor this enzyme system and obtain data that reflect the conversion of substrates to intermediates, to final products. Data analysis and computational methods can be developed for extracting spectra of individual enzymatic substrates and product from the complex, time-evolving, spectra, and using these for identification of molecular species. A

general enzymatic network model can be developed and tested to simulate metabolic response of the enzyme systems to varying concentrations of polysaccharide-synthesis precursors, and their derivatives. Once developed and tested, a real-time NMR protocol can be developed for monitoring the response of more complex biological systems (*Nicholson et.al, 2004*). In eukaryotic systems, many of the enzymes involved in glycosylation and polysaccharide synthesis are localized to the Golgi, an intracellular organelle that executes an important subset of cellular biochemistry. Studies of this organelle can serve as a future step in the analysis of a more complex biological system.

Metabonomics

Metabonomics is defined as “The quantitative measurement of the dynamic multi parametric metabolic response of living systems to physiological stimuli or genetic modification.”(*J.K.Nicholson et.al, 1999*). It is distinct from the related concept of metabolomics that includes specific studies of reaction pathways, similar to what is presented here.

Metabonomics, on the other hand, has shown particular promise in the area of toxicology and drug development. In both preclinical screening and metabolic profiling, it offers rapid toxicological information that is extremely useful. Metabonomics can use a combination of nuclear magnetic resonance spectroscopy and mass spectrometry together with chemometrics for profiling metabolism and interpreting metabolic fingerprints in extremely complex biological systems. The methods are used in pharmaceutical research and development, including drug safety assessment, characterization of genetically modified animal models of disease, diagnosis of human disease, understanding physiological variation and drug therapy monitoring.

Kinetic analysis of the system:

At its simplest level metabolomics requires the consideration of how a collection of enzymes, membrane barriers, and transporters must act together to accomplish a defined task. We need to analyze the limitations of the current methods and also understand how real-time NMR methodology will help to understand this particular complex biological problem. We can approach the problem in three basic steps, data collection, data analysis and network modeling.

- Systems development
- Data collection
- Data analysis
- Network modeling.

The strong interactions among these areas with continual feed-back of results allows the design of new biology-based experiments, the implementation of new NMR methodology, the improved extraction of molecular signals. To understand the synthesis of biomolecules we should develop new approaches that can gather massive quantities of data on complex systems evolving in time, and new analysis procedures that can model the interactions of multiple components.

Options for Data Collection**Mass Spectrometry**

Mass spectrometry has recently emerged as the foremost technology in endogenous metabolite research. The advantages of mass spectrometry include a wide dynamic range, high sensitivity and the ability to observe a diverse number of molecular species. These attributes are important in addressing the issue of metabolite profiling, as the dynamic range easily exceeds nine orders of magnitude in biofluids, and the diversity of species ranges from simple amino acids to complex carbohydrates. The goals of the application of mass spectrometry range from basic

biochemistry to data analysis, and structurally characterizing physiologically important metabolites. Mass spectrometry has been successful in drug-metabolite analysis and pharmacokinetic studies. Mass spectrometry also facilitates the identification of previously uncharacterized metabolites. Mass spectrometry is already well established as a quantitative tool for small molecule analysis. (*Elizabeth J. Want et.al, 2005*).

The combination of high-resolution capillary gas chromatography with mass spectrometry (GC/MS) is the most widely used form of MS in metabolite research and disease diagnosis. GC/MS enables the identification of key small molecules—such as fatty acids, amino acids, and organic acids—in biofluids, particularly in urine and blood. GC/MS has provided diagnostic information for many inherited diseases, including numerous metabolic disorders. These include disorders of the metabolism of amino acids, thyroid hormones, bile acids, steroids, organic acids, and fatty acids, with 20–30 disorders of the latter alone having been characterized. (*Want et.al, 2005*). An important example of more quantitative mass spectrometry metabolite analysis is the measurement of phenylalanine and tyrosine in the diagnosis of phenylketonuria (PKU). Many clinical laboratories around the world still use GC/MS for the diagnosis of some of these metabolic diseases, particularly those disorders of organic acid synthesis, in which, at present, LC/MS techniques have offered little advancement. GC/MS has also been applied successfully to metabolite profiling for plant functional genomics, which, combined with principal-component analysis (PCA), has been used to significantly improve upon existing approaches and understanding.

However, there are limitations. (*Elizabeth J. Want et.al, 2005*). In particular, metabolomics applications of MS are limited by difficulties with quantitation, the need to do pre-analysis separations, and the lack of sufficient molecular structure information. Quantitation of various

biological products in mass spectrometry has been primarily accomplished through the use of various isotope tagging strategies (*Goshe and Smith, 2003*), but these usually require advanced knowledge of the types of molecules of interest.

NMR

NMR spectroscopy has emerged as a key tool for understanding metabolic processes in living systems. The enormous value and utility of the approach to studies of metabolism in tissues is undisputed. The new approaches, which fall in the realm of metabonomics analysis, are based on a minimum of preconceptions of the biochemical processes that occur in a living system and use advanced spectroscopic acquisition and analysis techniques. (*Muireann Coen and Philip W. Kuchel, 2004*).

Traditional biochemical analysis techniques have not been sufficient to fully elucidate the chemical mechanisms associated with the highly integrated processes of organelles and whole cells and there is a need for new investigative methods. Novel nuclear magnetic resonance (NMR) techniques are one possible investigative route that is particularly appealing given that NMR can be used to determine the structure of molecules as well as the biochemical course of their action. Furthermore, NMR is very efficient in real-time metabonomics analysis; the simultaneous systematic determination of metabolite levels in even whole organisms and their changes over time as a consequence of stimuli can be accomplished. NMR provides an effective tool to identify the various compounds formed and also helps to follow these compounds through time to observe their synthesis and consumption. NMR has been applied to metabolite-profiling studies in diverse areas that includes plant metabolism. The primary advantage of NMR is its ability to measure analytes in biofluids quickly and accurately, without the need for initial processing or separation. It can also be made more selective using isotope labeling. In addition to

the ability to observe magnetic isotopes of higher abundance (^1H and ^{31}P) substrates can be enriched in ^{13}C or ^{15}N to selectively monitor absolute pathways. The improvements in NMR have included higher resolution and lower instrument cost. However, a major weakness of NMR is that it has relatively low sensitivity and a poor dynamic range (100–1000) that results in only the most abundant components being observed.

The detection of a broad spectrum of metabolites in ^1H NMR has advantages, but one must deal with enormous quantities of superimposed and noisy data. Resonances belonging to single compounds must be deconvoluted and assigned. A new analysis method called statistical total correlation spectroscopy (STOCSY) is a useful tool and can be applied to NMR spectra in metabonomics applications. This method is usually applied to small molecules in complex mixture analysis and in metabolic profiling studies. The name stochastic TOCSY comes from the similarity in appearance of processed data to the normal spin-physics based 2D TOCSY spectrum. A TOCSY spectrum yields through bond correlations via spin-spin coupling usually by 3 bond ^1H - ^1H couplings. Correlations are seen throughout the coupling network and intensity is not related to the number of bonds connecting the protons. TOCSY is usually used in large molecules with many separated coupling networks such as peptides, proteins, oligosaccharides and polysaccharides. STOCSY delivers intra molecular information and in metabolism studies can generate information on pathway-related correlations (*Couto Alves A, Holmes E, 2009*). In STOCSY, spectra from a set of samples with randomly varying concentrations of compounds, offset by an average spectrum, are simply collected in a 2D matrix and a covariance matrix is calculated. When they are plotted, cross-peaks appear that connect resonances belonging to each distinct compound. Unlike normal TOCSY spectra, connections of resonances belonging to different parts of molecules are established, despite the lack of 3-bond proton-proton scalar

couplings. For example, resonances from different sugars in a disaccharide are connected despite the lack of coupling across a glycosidic bond. The stochastic TOCSY procedure thus allows extraction of individual spectra from columns or rows containing cross-peaks. The process of STOCSY mainly relies on the random variation of concentrations of various components in the spectra collected. Random variations can also be simulated by randomly ordering spectra obtained from more systematically varying processes. (*Couto Alves A, Holmes E, 2009*).

Network modeling of metabolite fluxes

A goal in network modeling is to connect identified components in a reaction network and assign activities to the enzymes and transporters that govern the connectivity's. The peaks in data sets are assigned to specific molecules, those belonging to substrates, intermediates and products are integrated and converted to concentrations as a function of time. Complex reactions, in which a number of substrates are converted to a number of products by the action of a set of enzymes, have been successfully modeled. The simulation starts from an initial value that is chosen completely randomly. The simulation then proceeds as a random walk in parameter space. From the current location in space, a new location is chosen and if the proposed new set of parameters improves the fit to the experimental data then it is automatically accepted and becomes the next point from which a random walk is initiated. If that point is rejected, then the next point is identical to the current point and a new random step is explored. A large sample of points visited by the random walk represents the model ensemble. (*Landau, D. P. & Binder, K, 2000*)

If the possible substrates and most of the products and intermediates can be identified, then a network can be constructed in which all participants are connected by links assigned variable weights. Analysis of correlated changes in concentrations of each of the products and

intermediates in response to a supply of particular concentrations for substrates can be used to set the weights in the network. The weights can be interpreted in terms of activities of enzymes interconverting various compounds and also in terms of the inhibitory effects of particular compounds on particular enzymes. Rate limiting steps in complex pathways can be identified and these can often become targets for drugs that may optimally inhibit a pathway.

For example a chemical reaction network for the regulation of the quinic acid (qa) gene cluster of *Neurospora crassa* was proposed. An efficient Monte Carlo method for walking through the parameter space of possible chemical reaction networks was developed to identify an ensemble of deterministic kinetics models with rate constants consistent with RNA and protein profiling data. This method was successful in identifying a model ensemble fitting available RNA profiling data on the qa gene cluster .(*Battogtokh, D., D. K. Asch, et al., 2002*)

CHAPTER 2

NMR ANALYSIS OF ACETYL Co-A SYNTHETASE REACTIONS

Data collection for Acetyl-CoA Synthetase: ^1H NMR for the enzymatic reaction

NMR-methodology is non-destructive and capable of measuring concentration of analytes. We will investigate the nature of the enzymatic products formed by following the reaction assay in real time by ^1H NMR spectroscopy. In principal, this technique can detect all the enzymatic products, reactants and stable intermediates as long as they are present at a sufficient concentration.

Acetyl-CoA Synthetase (ACS) catalyzed reaction:

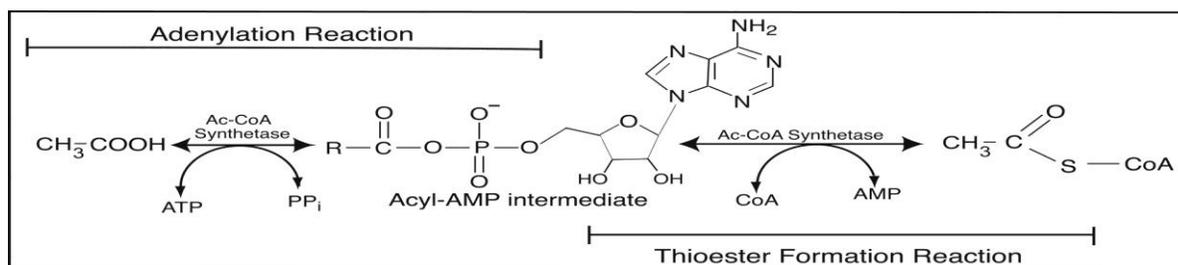


Figure 2.1. Production of Acetyl CoA

Figure 2.1 shows the reaction catalyzed by ACS. It is hypothesized that acetic acid then reacts with ATP in the presence of Ac-CoA Synthetase to form an Acyl-AMP intermediate. The Acyl-AMP reacts with CoA to form Acetyl CoA. (Starai *VJ*, *Escalante-Semerena JC*, 2004).

ATP + acetate + CoA + buffer (pH7.0) (37°C, 50 min and NMR 600)

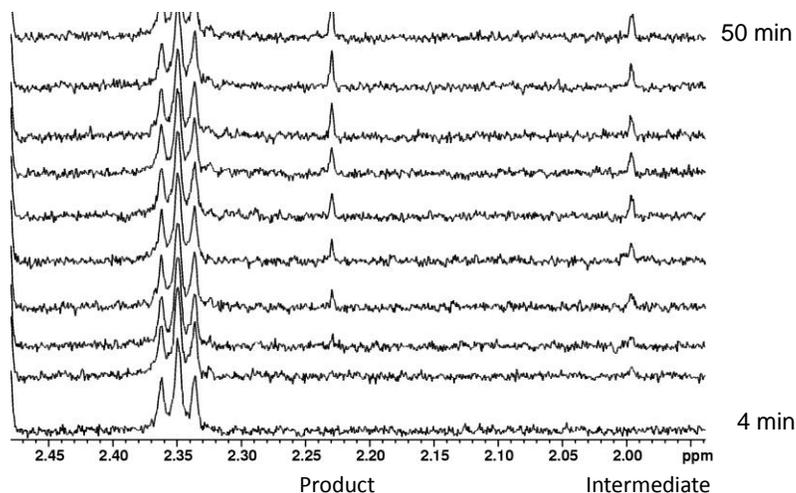


Figure 2.2. ^1H NMR for the reaction of ATP+acetate+CoA+buffer

Figure 2.2 show ^1H NMR spectra as a function of time for a sample containing ATP+acetate+CoA. ACS reactions were performed in 50 mM sodium phosphate, pH 7.5, in a mixture of $\text{D}_2\text{O}:\text{H}_2\text{O}$ (8:1 v/v, 180 μl) containing 5 mM MgCl_2 , 1 mM ATP, 0.46 mM CoA, 1 mM acetate and 1.5 μg recombinant ACS supplied in H_2O -buffer. Immediately upon addition of enzyme, the reaction mixture was transferred to a 3 mm NMR tube. Real-time ^1H NMR spectra were obtained at 37°C on a Varian (Palo Alto, CA) Inova spectrometer with a cryogenic probe operating at 600 MHz. Data acquisition was not started until 2 minutes after the addition of enzyme to the reaction mixture due to spectrometer set-up requirements (shimming). Sequential 1D proton spectra were acquired over the course of the enzymatic reaction. All spectra were referenced to the water resonance at 4.765 ppm. Reactions were initiated by addition of acetate, ATP, CoA and buffer along with the enzyme Acetyl-CoA synthetase and 1D NMR spectra were collected as above. The enzyme Acetyl-CoA synthetase reacts with acetate and ATP to form an

intermediate with a peak around 2.0 ppm as shown in figure. This is a postulated intermediate and is not yet identified. The postulated intermediate then reacts with CoA and the enzyme Acetyl-CoA synthetase to form the product, acetyl CoA around 2.2 ppm. The acquisitions are carried out every 4 minutes and the formation of product can clearly be observed after 50 minutes.

ATP + acetate + buffer (pH7.0) (37°C, 50 min and NMR 600)

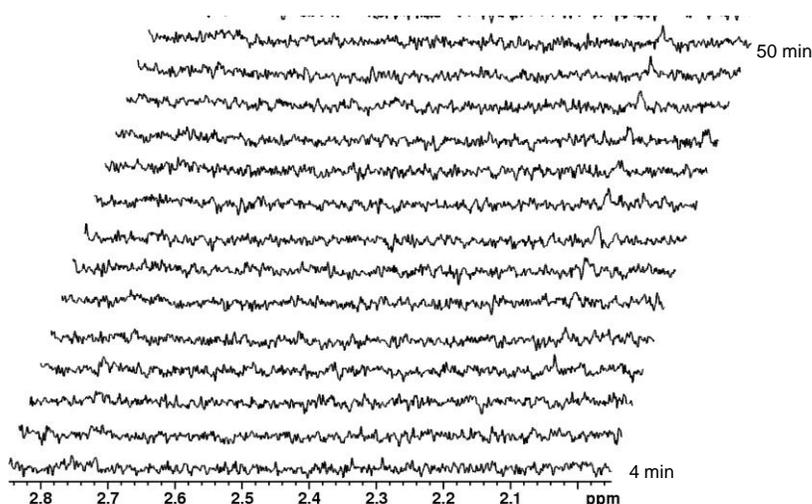


Figure 2.3. ¹H NMR for the reaction using just ATP + acetate

To clarify the origin of the peaks at 2.0 ppm, peaks at the reaction was run without the CoA present. Figure 2.3 shows the ¹H NMR reactions for ATP+acetate. ACS reactions were performed in 50 mM sodium phosphate, pH 7.5, in a mixture of D₂O:H₂O (8:1 v/v D₂O, 180 μl) containing 5 mM MgCl₂, 1 mM ATP, 1 mM acetate and 1.5 μg recombinant ACS supplied in H₂O-buffer. Immediately upon addition of enzyme, the reaction mixture was transferred to a 3 mm NMR tube. Real-time ¹H NMR spectra were obtained at 37°C on a Varian (Palo Alto, CA)

Inova spectrometer with a cryogenic probe operating at 600 MHz. Data acquisition was not started until 2 minutes after the addition of enzyme to the reaction mixture due to spectrometer set-up requirements (shimming). Sequential 1D proton spectra were acquired over the course of the enzymatic reaction. All spectra were referenced to the water resonance at 4.765 ppm. Reactions were carried out by addition of acetate, ATP and buffer are added with the enzyme Acetyl-CoA synthetase and the 1D NMR spectra was collected as above. The enzyme Acetyl-CoA synthetase reacts with acetate and ATP to form a methyl containing intermediate around 2.0 ppm. This is the postulated intermediate and is not yet detected. The acquisitions are carried out every 4 minutes and we could see the formation of the postulated intermediate after 50 minutes.

CONCLUSION

The enzyme acetyl-CoA synthetase when it reacts with acetate and ATP forms a methyl containing released intermediate or side-product with a resonance around 2.0 ppm as shown in figure 2.2. If it were an intermediate, it would return to react with CoA under the influence of the enzyme acetyl-CoA synthetase to form the product, acetyl CoA with a resonance around 2.2 ppm. In figure 2.3, it is shown that the enzyme acetyl-CoA synthetase reacts with acetate and ATP to form the same intermediate, the postulated intermediate around 2.0 ppm. Thus, we could observe that without the addition of CoA the reaction stops with the formation of an intermediate or side-product. It has been suggested that the intermediate is acetyl AMP. However, comparing to chemical shift data of standards indicates that acetyl phosphate is equally consistent with observation. This could be formed by decomposition of a reactive intermediate and may or may not be on the reaction pathway. The intensity of the resonance is relatively large inconsistent with an enzyme-bound intermediate. It has to be an exchange with a species on the enzyme.

CHAPTER 3

DATA ANALYSIS USING ALTERNATIVE PROCESSING STRATEGIES

STOCSY-Statistical total correlation spectroscopy

The detection of metabolites in ^1H NMR has some advantages, but there is also the issue of having to cope with enormous quantities of superimposed data. Structural assignment of resonances is another important problem in NMR spectroscopy. A new analysis method called statistical total correlation spectroscopy (STOCSY) is a useful tool and can be applied to NMR spectra in metabonomics applications. This method is usually applied to small molecules in complex mixture analysis and metabolic profiling studies. The name stochastic TOCSY comes from the similarity in appearance of processed data to the normal spin-physics based 2D TOCSY spectrum. A TOCSY spectrum yields through bond correlations via spin-spin coupling. Correlations are seen throughout the coupling network and intensity is not related to the number of bonds connecting the protons. TOCSY is usually used in large molecules with many separated coupling networks such as peptides, proteins, oligosaccharides and polysaccharides. STOCSY delivers intra molecular information and in metabolism studies can generate information on pathway-related correlations (Analytic Properties of Statistical Total Correlation Spectroscopy Based Information Recovery in $(1) \text{H}$ NMR Metabolic Data Sets, *Couto Alves A, Holmes E*, 2009). In STOCSY, spectra from a set of samples with randomly varying concentrations of compounds, offset by an average spectrum, are simply collected in a 2D matrix and a correlation matrix is calculated. When they are plotted, cross-peaks appear that connect resonances belonging to each distinct compound. Unlike normal TOCSY spectra, connections of resonances

belonging to different sugar residues in a polysaccharide can be established, despite the lack of 3-bond proton-proton scalar couplings across glycosidic bonds. The stochastic TOCSY procedure allows extraction of individual spectra from columns or rows containing cross-peaks. The process of STOCSY mainly relies on the random variation of concentrations of various components in the spectra collected. Random variations can also be made by randomly ordering spectra. **(Analytic Properties of Statistical Total Correlation Spectroscopy Based Information Recovery in (1)H NMR Metabolic Data Sets, Couto Alves A, Holmes E, 2009).**

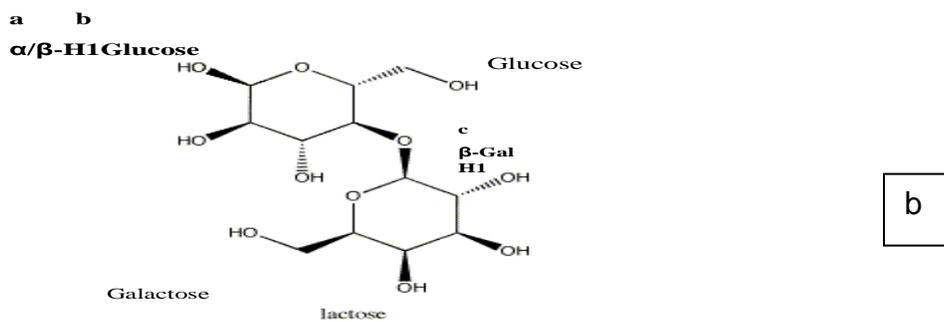
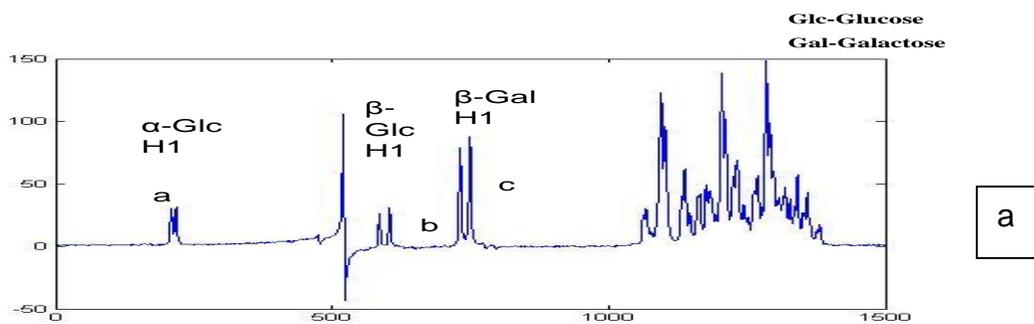
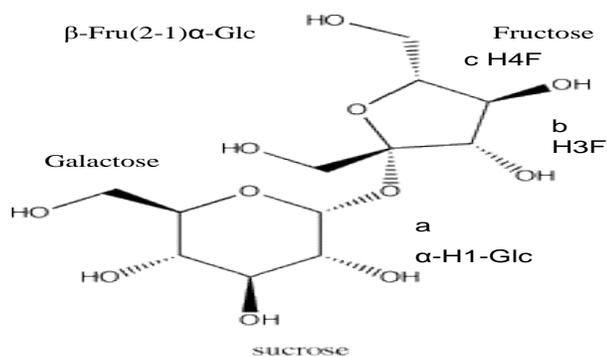


Figure 3.1 1H NMR spectra -Lactose

The figure 3.1-a is the 1H NMR spectra for pure lactose and 2.1-b is the structure of lactose. For the spectrum, the ASCII file was obtained and the first 1500 points were extracted by the MATLAB. The systematic name for lactose is β -D-galactopyranosyl-(1 \leftrightarrow 4) β -D-glucopyranose. It is a disaccharide that consists of β -D-galactose and α/β -D-glucose fragments bonded through a β 1-4-glycosidic linkage. The assignments for the anomeric resonances are made in the figure 2.1-a comparing it with the structure in figure 2.1-b.

a



b

Figure 3.2 ^1H NMR spectra -Sucrose

The figure 3.2 is the ^1H NMR spectra for pure sucrose. The ASCII file for the spectrum was obtained and the first 1500 points were extracted by the MATLAB. The figure 3.2-a is the ^1H NMR spectra for pure sucrose and 3.2-b is the structure of sucrose. The ASCII file was obtained for the spectra and the first 1500 points were extracted by the MATLAB. The systematic name for sucrose is β -D-fructofuranosyl-(2 \rightarrow 1)- α -D-glucopyranoside. It is a disaccharide that consists of glucose and fructose with an alpha 1, 2 glycosidic linkages. The assignments are made in the figure 2.2-a comparing it with the structure in figure 2.2-b.

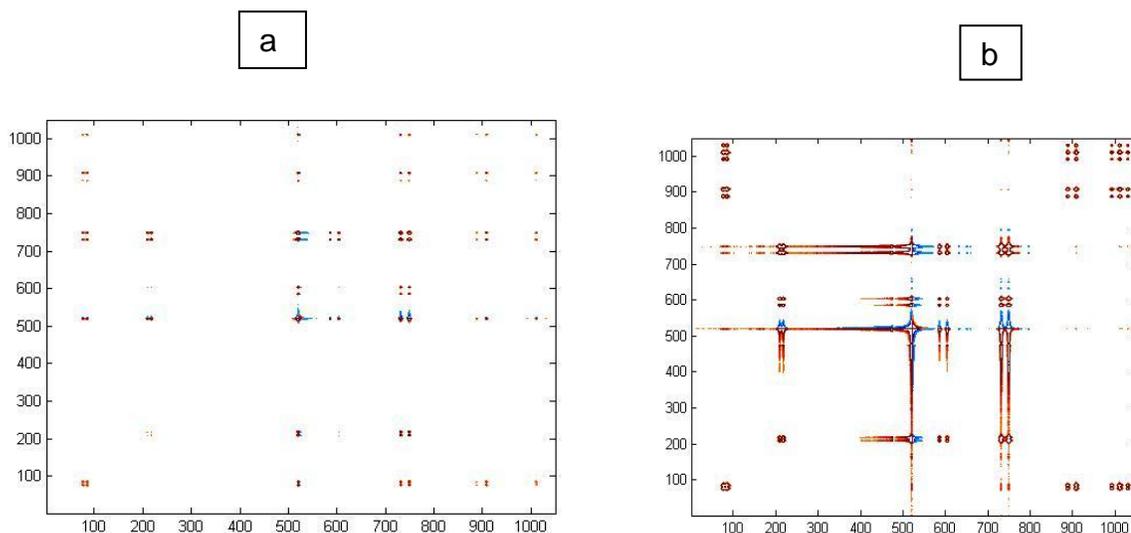


Figure 3.3 Contour plot for STOCSY spectrum produced from 1D spectra of random mixtures of sucrose and lactose

The figure 3.3 (a) and (b) shows the statistical total correlation spectrum for random mixtures of lactose and sucrose. The process of STOCSY mainly relies on the random variation of concentrations of various components in the spectra collected. Spectra from a set of samples with randomly varying concentrations of compounds are simply collected in a 2D matrix and a correlation matrix is calculated. (*Couto Alves A, Holmes E, 2009*). The correlation matrix is formed using MATLAB (see appendix). The diagonal peaks correspond to the superposition of the downfield regions of 1D spectra in figure 3.1 a and 3.2 a. The figure 3.3(a) corresponds to the random mixtures with 50 columns whereas figure 3.3(b) corresponds to random mixtures with 150 columns and random variations. The column at data point 750 in figure 3.3(a) shows the peak which belongs to lactose which should not appear as it is the spectra for completely random mixtures and the correlation should be observed only for the sucrose peaks at 80, 900 and 1000. The figure 3.3(b) shows a complete correlation between sucrose peaks at 80, 900 and 1000. Complete positive correlation is observed when we increase the number of peaks to 150. Thus we can say that at least 150 random variations are required to observe the correlation.

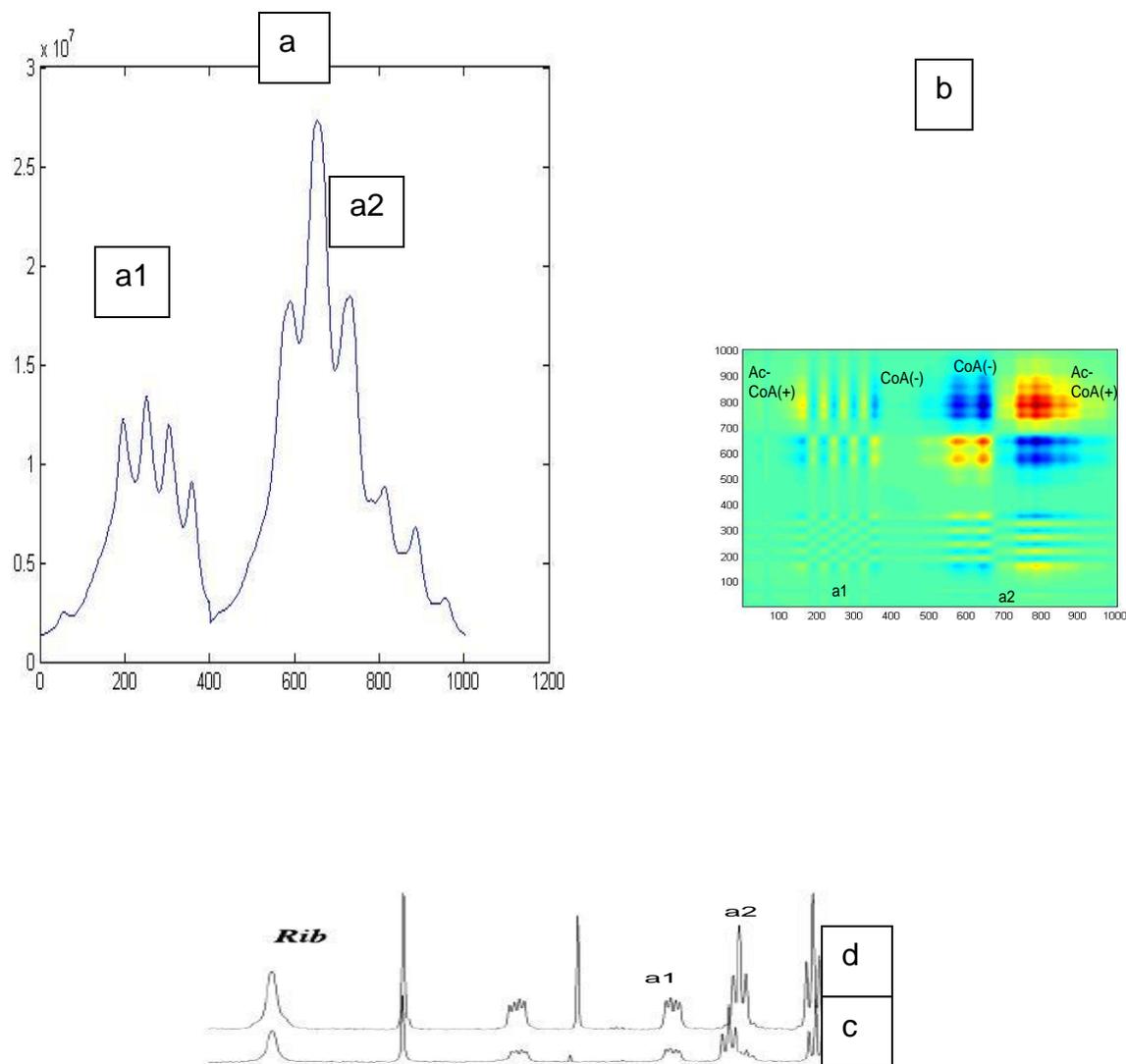


Figure 3.4 Covariance matrix for Acetyl-CoA Synthetase catalysed reaction

The figure 3.4 shows the covariance matrix for the Acetyl-CoA catalysed reaction (Figure 2.1). From the full ACS spectra few peaks were sliced using MATLAB to observe correlation between them. Figure 3.4-(a1)(a2) shows the two CoA peaks observed. The Figure 3.4-(b) shows the covariance matrix constructed using MATLAB. We observe the yellow to be positively correlated and the peaks in blue to be negatively correlated. The covariance matrix (figure 3.4-(b)) has two regions a1 and a2 corresponding to the ones in Figure 3.4-(a). In Figure 3.4-(c), the a1 and a2 CoA regions seem like they overlapped and later convert into acetyl-CoA in (d)

according to the reaction. The conversion of CoA to acetyl-CoA can be observed in figure 3.4(c) and 3.4(d) respectively. We can observe positive and negative correlations between the Ac-CoA and the CoA peaks respectively which is clear from figure 3.4-(b).

CONCLUSION

STOCSY mainly relies on the random variation of concentrations of various components in the spectra collected. (**Analytic Properties of Statistical Total Correlation Spectroscopy Based Information Recovery in (1)H NMR Metabolic Data Sets, Couto Alves A, Holmes E, 2009**). We have tested its ability to separate spectra as described in figure 2.3 for the samples that vary randomly in composition as well as ones related by kinetic conversion of one to another. The total number of random samples needed for adequate separation should be determined and later it could be applied to systems that don't vary randomly. The code which was used to generate fields and process spectra is included in appendix A. Separation of peaks belonging to different molecules and separation of overlapping peaks allows not only identification of the molecules but integration of additional peaks for kinetic analysis.

CHAPTER 4

NETWORK MODELING OF THE ACS REACTION

Network modeling of the ACS system using Monte Carlo methods

Figure 2.1 can be used to carry out network modeling to connect components in a reaction network. The peaks in data sets obtained in Figure 2.2 and 2.3 could be assigned to specific substrates, intermediates and products and were integrated and converted to concentrations as a function of time. The network constructed will represent the function of the enzyme system and reproduce the real-time variations of concentrations represented in the NMR data. The data obtained is tabulated and simulated using Monte Carlo simulations. Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling of parameters followed by computation of results and comparison with observables. It is useful for modeling phenomena with significant mechanistic uncertainty. This method is often used when the model is complex or involves more than just a couple of uncertain parameters. A simulation can typically involve over 10,000 evaluations of the model. The inputs are chosen that most closely match data already known. Several network forms are usually tested. The data obtained should fit well with a proper network. In these cases the form of the networks will suggest connections not directly observed, as well as unobservable intermediates. The ability to eliminate unnecessary connections and also to identify previously unknown pathways or intermediates is a useful outcome.

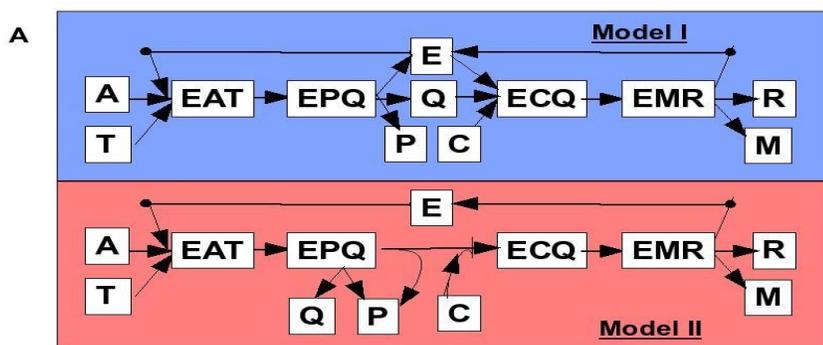


Figure 4.1. Network forms for modeling for the enzyme system ACS: Model 1 and 2

A: Acetate, **T:** ATP, **EAT:** Enzyme+Acetate+ATP,

EPQ: Enzyme+PPi+AMP-Acet **Q:** AMP-Acet, **P:** PPi, **C:** CoA, **E:** Enzyme ACS

ECQ: Enzyme+CoA+AMP-Acet, **EMR:** Enzyme+AMP+Acet CoA, **R:** Acet CoA, **M:** AMP

Figure 4.1 represents two models for the function of the enzyme system ACS and the proper form will reproduce the real-time variations of concentrations represented in the NMR data. For the ACS system, the inputs are simply the substrates. All other nodes are either intermediates or products. The concentrations are used as a function of time for various compounds are used to refine weights of various connections in the model. In Model 1 the AMP-Acet, (postulated intermediate) is assumed to be an intermediate on the path to Acet CoA. In Model 2, the AMP-Acet (postulated intermediate) is assumed to be bound to the enzyme. It is found as a complex EPQ and then converts to another intermediate ECQ. It is difficult to observe the ECQ intermediate because it is always bound to the enzyme.

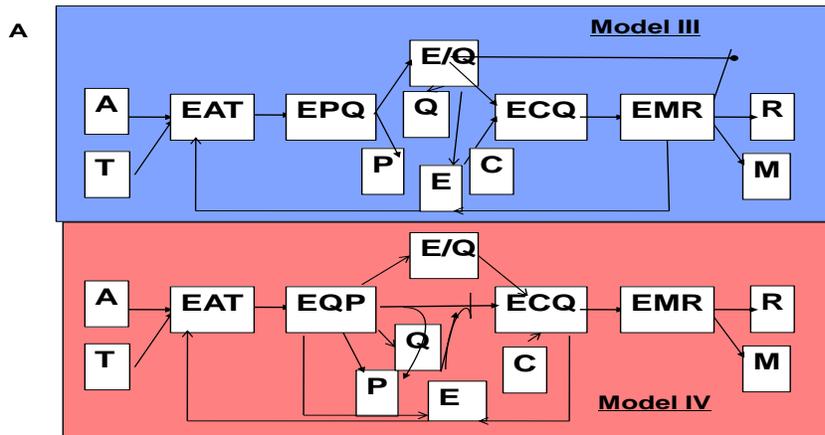


Figure 4.2. Network modeling for enzyme system ACS: Model 3 and 4

A: Acetate, **T:** ATP, **EAT:** Enzyme+Acetate+ATP,

EPQ: Enzyme+PPi+AMP-Acet **Q:** AMP-Acet, **P:** PPi, **C:** CoA, **E:** Enzyme ACS

ECQ: Enzyme+CoA+AMP-Acet, **EMR:** Enzyme+AMP+Acet CoA, **R:** Acet CoA, **M:** AMP

Figure 4.2 also represents the function of the enzyme system ACS. Here the ACS system the inputs are again simply the substrates. All other nodes are either intermediates or products. The concentrations as a function of time for various compounds are used to set weights in the model. In Model 3 and Model 4 the AMP-Acet, (postulated intermediate) is bound with the enzyme ACS on the path to Acet CoA. Both models are almost the same with a very few modifications. The difference is the intermediate is not involved in the pathway to product in model 3 whereas it is an intermediate in model 4.

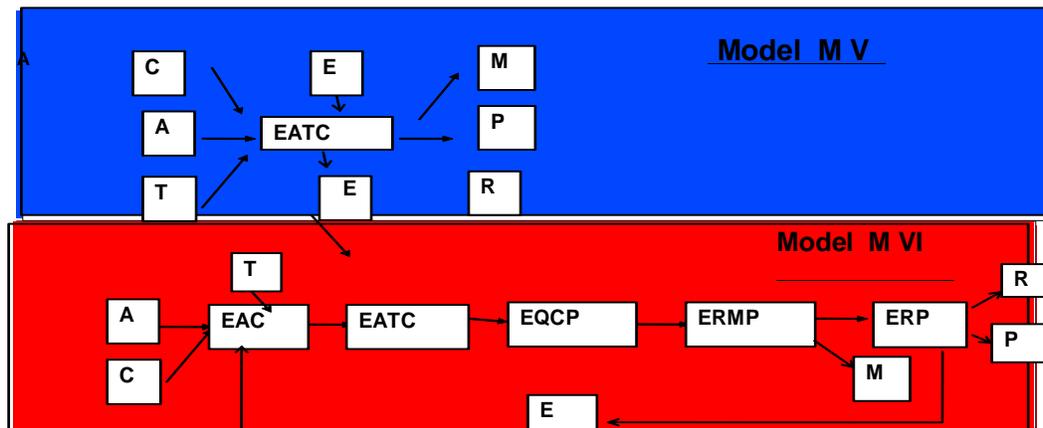


Figure 4.3. Network modeling for enzyme system ACS: Model 5 and 6

A: Acetate, **T:** ATP, **EAT:** Enzyme+Acetate+ATP,

EPQ: Enzyme+PPi+AMP-Acet **Q:** AMP-Acet, **P:** PPi, **C:** CoA, **E:** Enzyme ACS

ECQ: Enzyme+CoA+AMP-Acet, **EMR:** Enzyme+AMP+Acet CoA, **R:** Acet CoA, **M:** AMP

Figure 4.3 also represents the function of the enzyme system ACS and will reproduce the real-time variations of concentrations represented in the NMR data. For the ACS system, the inputs are simply the substrates. All other nodes are either intermediates or products. The concentrations are used as a function of time for various compounds to set the weights. Model 5 represents a very simple two step reaction to the formation of Acet-CoA. In model 6, there is formation of five substrate complexes EAC, EATC, EQCP, ERMP and ERP. Then there is formation of Acet-CoA from the ERP complex with the release of free enzyme.

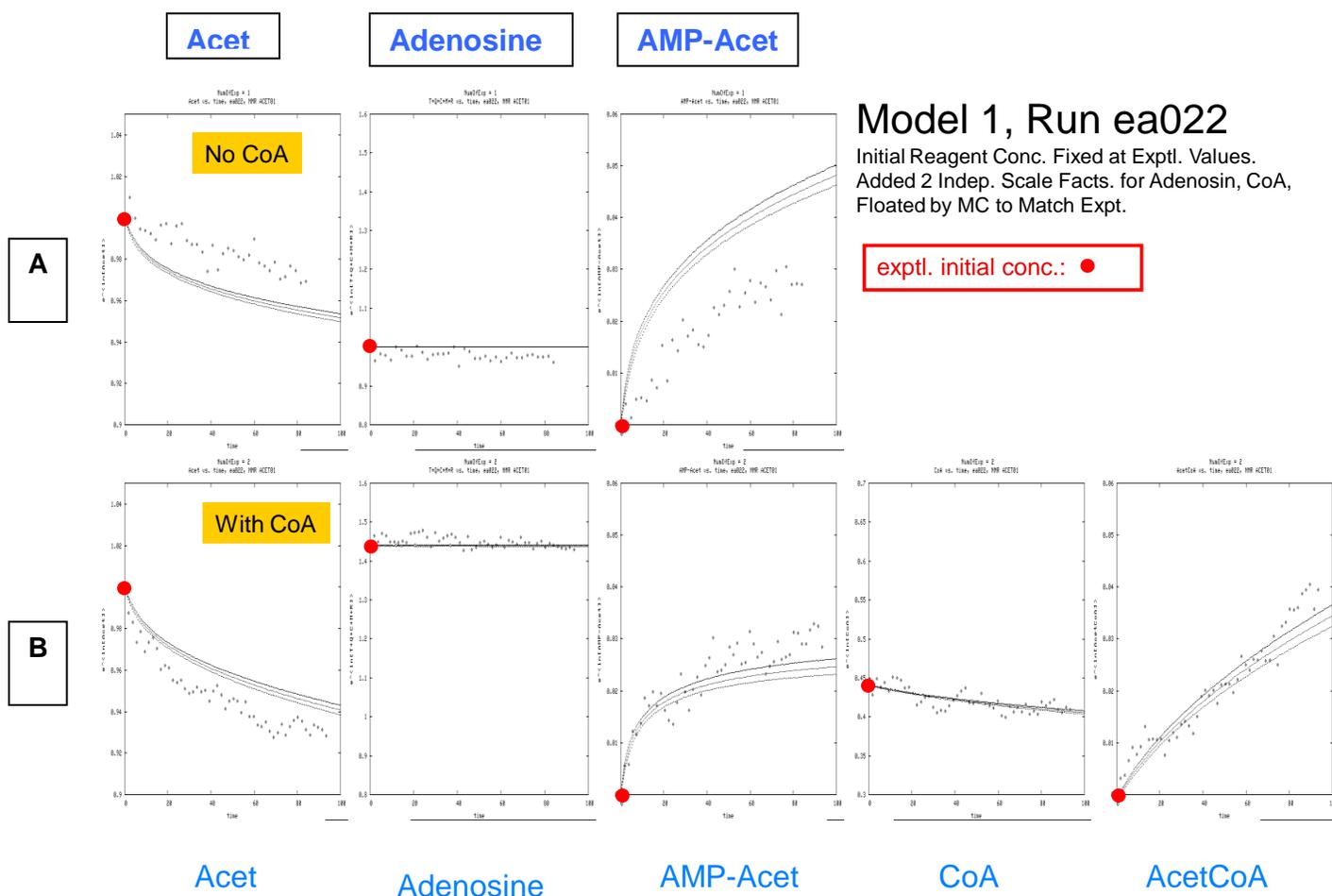


Figure 4.4 Monte Carlo simulations run for Model 1

Figure 4.4. shows simulation of concentrations of various compounds as a function of time using MC, determines parameters for model 1 as lines and dots are experimental data. We are going to compare model 1 to model 2 in more detail. The 'A' in the figure represents the experimental run for No CoA and the 'B' in the figure represents the experimental run for With CoA. The kinetic parameters are extracted from Figure 2.2 and Figure 2.3. The experimental data is the input for modeling. The initial reagent concentrations are fixed at experimental values, and then it is floated by the Monte Carlo methods to best simulate the experimental values. From figure 2.1, we know that the substrate Acetate should be decreasing with time and the product AcetCoA should be increasing. This is evident from both 'A' and 'B'. Model 1 insists that AMP-Acet is an intermediate on the path to Acet-CoA. In 'A', we can see that the AMP-Acetate would be formed and shows an increasing pattern. The reactions stops with it as there is no addition of CoA, whereas in 'B' we can observe that AMP-Acet again reacts with CoA and forms the product Acet-CoA. When we compare the fit of AMP-Acet (the postulated intermediate) in 'A' and 'B' it is seen that it fits well to the data in 'B'. Thus, we can say that AMP-Acet fits better when we give it a pathway to CoA. In 'A' the fit of AMP-Acet is not good which means that there is something missing in 'A'. Overall, we could see that the Model 1 does not give a very good fit to the data.

Table 4.1 Rate constants for the forward reactions of Model 1

No:	Reaction	Rate constants	Standard deviation
1	ATP+Acet+ACS→E/A/T	0.11218441D+05(1/min*mM ²)	3.0035D+05
2	E/A/T→E/Q/P	0.10311299D+04(1/min)	1.886D+03
3	E/Q/P→AMP-Acet+PPi+ACS	0.67328855D+03 (1/min)	1.449D+03
4	AMP-acet+CoA+ACS→E/Q/C	0.31223684D+05 (1/min*mM)	8.2071D+04
5	E/Q/C→E/R/M	0.25620575D+04 (1/min)	2.617D+03
6	E/R/M→Acet- CoA+AMP+ACS	0.86249880D+03 (1/min)	1.463D+03

Table 4.2 Rate constants for the backward reactions of Model 1

No:	Reaction	Rate constants	Standard deviation
1	E/A/T→ ATP+Acet+ACS	0.79541847D+03 (1/min)	1.731D+03
2	E/Q/P →E/A/T	0.76332343D+02(1/min)	1.4374D+02
3	AMP-Acet+PPi+ACS →E/Q/P	0.67705345D+05 (1/min*mM ²)	9.9661D+04
4	E/Q/C→AMP-acet+CoA+ACS	0.47784184D+03 (1/min*mM)	8.5923D+02
5	E/R/M → E/Q/C	0.22217598D+03 (1/min)	6.806D+02
6	Acet-CoA+AMP+ACS→ E/R/M	0.10033296D+06 (1/min*mM ²)	1.99672D+05

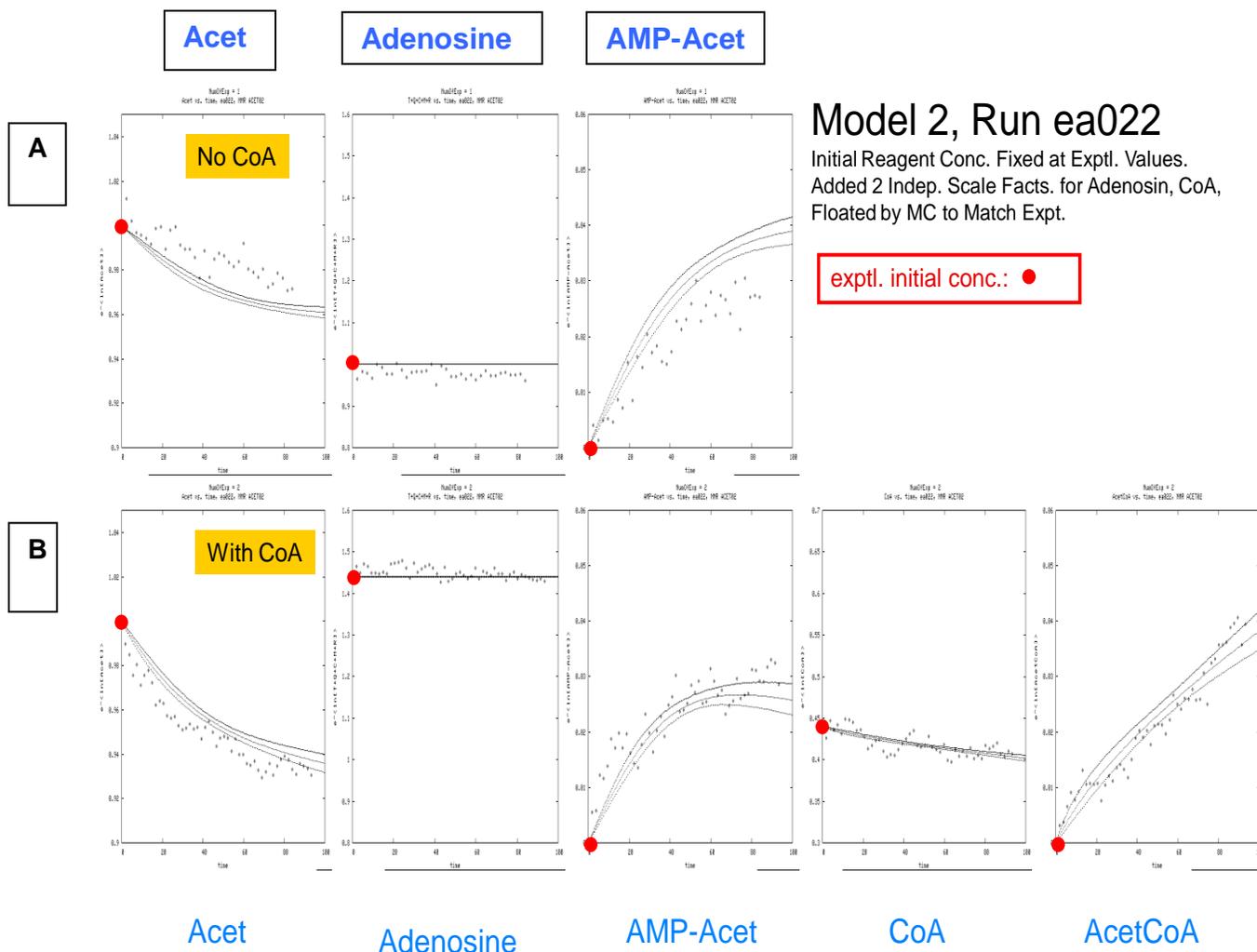


Figure 4.5 Monte Carlo simulations run for model 2

Figure 4.5 is the result of the run for model 2 (Figure 4.4). The 'A' in the figure represents the experimental run for No CoA and the 'B' in the figure represents the experimental runs for With CoA. The kinetic parameters are extracted from Figure 2.2 and Figure 2.3. The experimental data is the input for modeling. The initial reagent concentrations are started at experimental values, and then it is floated by the Monte Carlo methods to match the experimental values. From figure 1.1 we know that the substrate Acetate should be decreasing with time and the product acet-CoA should be increasing. This is evident from both 'A' and 'B'. Model 2 insists that AMP-acet is always bound to the enzyme. In 'A', we can see that the AMP-acetate is formed and shows an increasing pattern. The reaction stops with it, as there is no addition of CoA, whereas in 'B' we can observe that AMP-acet again reacts with CoA and forms the product Acet-CoA. In addition, we can observe there is a decreasing pattern of CoA in 'B' because it is converted into Acet-CoA. Overall by comparing model 1 (Figure 4.4) and model 2 (Figure 4.5) we can say that the Model 2 gives a better fit.

No:	Reaction	Rate constants	Standard deviation
1	ATP+Acet+ACS→E/A/T	0.20304647D+04 (1/min*mM ²)	4.577D+03
2	E/A/T→E/Q/P	0.10732D+03(1/min)	1.192D+02
3	E/Q/P→AMP- Acet+PPi+ACS	0.44395D+04(1/min)	3.687D+03
4	E/Q/P+CoA→E/Q/C+PPi	0.22103D+03(1/min*mM)	1.736D+02
5	E/Q/C→E/R/M	0.16988D+04(1/min)	2.287D+03
6	E/R/M→Acet- CoA+AMP+ACS	0.9873651D+03	1.963D+03

Table 4.3 Rate constants for the forward reactions of Model 2

Table 4.4 Rate constants for the backward reactions of Model 2

No:	Reaction	Rate constants	Standard deviation
1	E/A/T→ ATP+Acet+ACS	0.154104D+04 (1/min)	2.834D+03
2	E/Q/P → E/A/T	0.148997D+00(1/min)	1.5598
3	AMP-Acet+PPi+ACS → E/Q/P	0.3075D+06(1/min*mM ²)	3.72090D+05
4	E/Q/C+PPi → E/Q/P+CoA	0.7133D+00(1/min*mM)	8.8687
5	E/R/M → E/Q/C	0.112686D+03(1/min)	3.3697D+02
6	Acet-CoA+AMP+ACS → E/R/M	0.29362D+05(1/min*mM ²)	4.8222D+04

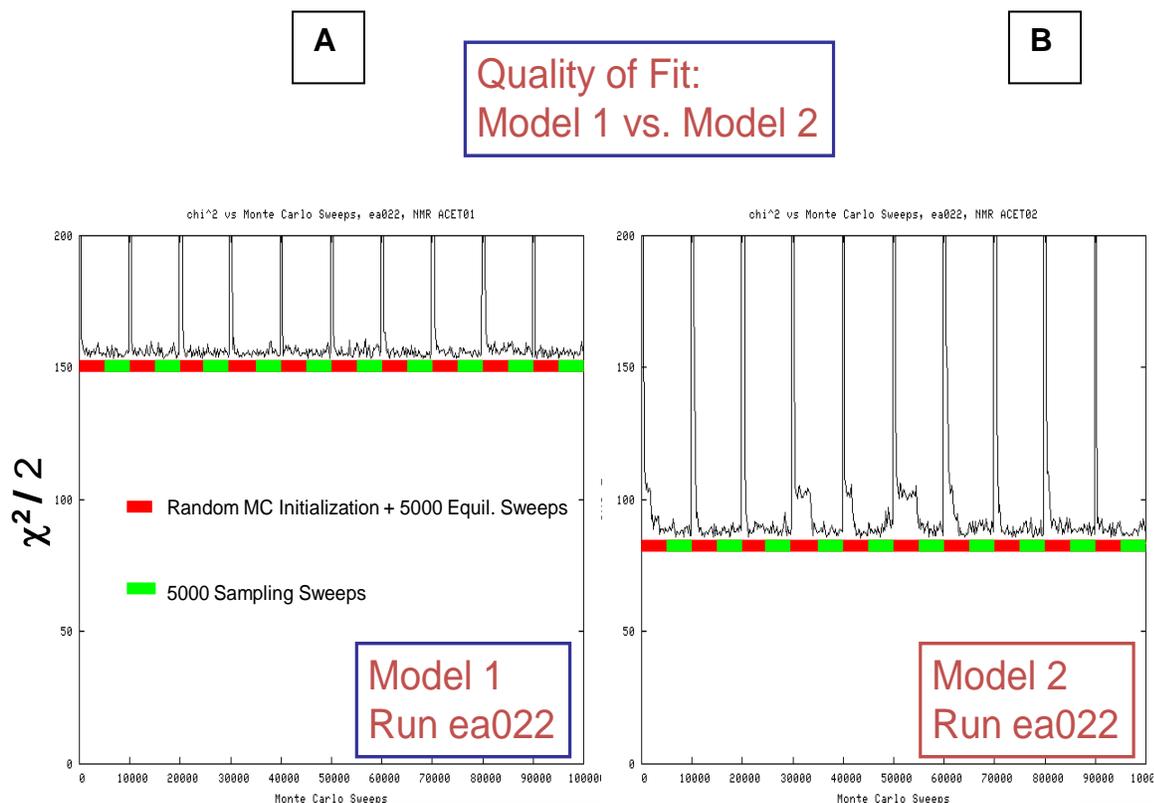


Figure 4.6 Comparisons of Model 1 and Model 2:

Figure 4.6 compares the quality of fit of model 1 and model 2. We totally show 10,000 Monte Carlo sweeps, of which 5000 are equilibration and 5000 are sampling sweeps. The run starts with a random Monte Carlo initialization then 5000 equilibration sweeps (indicated in red) are carried out followed by 5000 sampling sweeps (indicated in green) and then completely start again with a random Monte Carlo initialization. The chi square test indicates the overall quality of fit of the Model. In Monte Carlo simulations, minimization of chi square indicates a better quality in fit of the data by the Model. We can observe in 'A' (Model 1) that the χ^2 is very high when compared to 'B' (Model 2) which has lower χ^2 , thus we can conclude saying that Model 2 gives a better quality of fit compared to Model 1 which means AMP-Acetate is considered to be bound to the enzyme and gets converted from one complex to another and it is difficult to observe it as an intermediate. The intermediate we observed is off pathway. It may or may not be acetyl AMP.

CONCLUSION

The Network modeling analysis was done for six models as explained in the figures. The Monte Carlo simulation results are explained in figure 4.6. The chi square test indicates the overall quality of fit of the model. We know that in Monte Carlo simulations minimization of chi square indicates a better quality in fit of the data by the model. Model 2 has the lowest Chi square which indicates that model 2 is best when compared to all the other models, which says in figure 4.4 that the postulated intermediate is always bound to the enzyme in a complex making it difficult to observe. We actually observe a side product, which is possibly acetyl phosphate. Model 1 can be considered as the second best fit. Monte Carlo simulation run results are observed in figure 4.4 and figure 4.5 for model 1 and model 2 respectively. We can see that model 2 fits the data better when compared to the model 1. Overall, from the simulations we conclude that model 2 is the best. In addition, when the experiments were repeated with a new purified enzyme it was found that chemical shift data of standards of acetyl phosphate is equally consistent with the above observation. This could be formed by decomposition of a reactive intermediate and may or may not be on the reaction pathway. The intensity of the resonance is relatively large inconsistent with an enzyme-bound intermediate which suggests that it has to be an exchange with a species on the enzyme.

CHAPTER 5

FUTURE WORK

The new methodology discussed above provides models for the mechanisms of complex synthetic systems and it can be applied to both well-defined and very complex systems. A multi enzyme system can be designed and application to this system can provide a next step towards modeling more complex biological process. Analyzing and developing real-time NMR data for a complex enzyme system, doing network modeling to establish connections between unknowns and known's to identify unidentified intermediates and products and finally applying statistical methods to understand the actual synthesis might lead to the understanding of more complex process.

CHAPTER 6

REFERENCE

1. Nicholas V. Reo (2002) NMR based Metabolomics. *Drug and chemical toxicology* Vol. 25, No. 4, Pages 375-382.
2. Lindon JC, Holmes E, Nicholson JK (2004) Metabonomics: Systems biology in pharmaceutical research and development.
3. Keun HC (2006) Metabonomic modeling of drug toxicity. *Pharmacology and therapeutics*, 92-106.
4. Want EJ, Cravatt BF, Siuzdak G (2005) The expanding role of mass spectrometry in metabolite profiling and characterization. *ChemBiochem*, 1941-51.
5. Couto Alves A, Rantalainen M, Holmes E, Nicholson JK, Ebeels TM (2009). Analytic Properties of Statistical Total Correlation Spectroscopy Based Information Recovery in (1)H NMR Metabolic Data Sets. *Analytic chemistry*.
6. Want EJ, Cravatt BF, Siuzdak G (2005) The Expanding Role of Mass Spectrometry in Metabolite Profiling and Characterization. *ChemBiochem*.
7. Lu X, Zhao X, Bai C, Zhao C, Lu G, Xu G (2008) LC-MS –based Metabonomics analysis. *Journal of Chromatography.B, Analytic technologies in biomedical and life sciences*, 866(1-2):64-76.
8. Li H, Jiang Y, He FC (2008). Recent developments of Metabonomics and its applications in clinical research. *Yi chuan=Hereditas*:388-99.
9. Lindon JC, Holmes E, Nicholson JK. (2007). Metabonomics in Pharmaceutical R&D. *FEBS Journal*, 1140-57.

10. Lindon JC, Holmes E, Nicholson JK.(2006).Metabonomics techniques and applications to pharmaceutical research & development.*Pharmaceutical Research*,1075-88.
11. Robertson DG (2005).Metabonomics in toxicology:a review.*Toxicological Sciences*,809-22.
12. Keun HC.(2007). Biomarker discovery for drug development and translational medicine using metabonomics.Ernst Schering found symposium proceedings,79-98.
13. Lindon JC, Holmes E, Bollard ME, Stanley EG, Nicholson JK.(2004). Metabonomics technologies and their applications in physiological monitoring ,drug safety assessment & disease diagnosis.*Biomarkers* 1-31.
14. Lindon JC, Holmes E, Nicholson JK. (2004).Metabonomics and its role in drug development & disease diagnosis.
15. Schlotterbeck G, Ross A, Dietele F,Senn H.(2006). Metabolic profiling technologies for biomarker discovery in biomedicine & drug development.*Pharmacogenomics*,1055-75.
16. Coen M, Holmes E, Nicholson JK.(2008).NMR-based metabolic profiling and metabonomic approach to problems in molecular toxicology.*Chemical research in toxicology*,9-27.
17. Shockcoe JP, Holmes E.(2002).Metabonomic applications in toxicity screening and disease diagnosis.*Current topics in medicinal chemistry*,35-51.
18. Clarke CJ, Haselden JN.(2008).Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicology Pathology*,140-7.
19. Schlotterbeck G, Ross A, Dieterie F, Senn H.(2006).Metabolic profile technologies for biomarker discovery in biomedicine & drug development. *Pharmacogenomics*,1055-75.
20. Monton MR, Soga T.(2007).Metabolome analysis by capillary electrophoresis-Mass spectrometry.*Journal of Chromatography A*.237-46.

21. Chen C, Gonzalez FJ, Idle JR(2007). LC-MS based metabonomics in drug metabolism. *Drug metabolism reviews*. 581-97.
22. Starai VJ, Escalante-Semerena JC(2004). Acetyl-Coenzyme A synthetase (AMP-forming). *Cellular and Molecular Life sciences* .2020-30.
23. Hegg EL. (2004). Unraveling the structure & mechanism of acetyl-coenzyme A synthase. *Accounts of chemical research*, 775-83.
24. Servalli J, Gu W, Tam A, Strauss E, Begley TP, Cramer SP, Ragsdale SW. (2003). Functional copper at the acetyl-CoA synthase active site. *Proceedings of National Academy of Sciences of United States of America*. 3689-94.
25. Brunold TC. (2004) Spectroscopic & computational insights into the geometric & electronic properties of the A-cluster of acetyl-coenzyme A synthase. *Journal of biological inorganic chemistry*. 533-41.
26. Nicholson, J. K. & Wilson, I. D (2003). Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. *Nature Rev. Drug Discov.* **2**, 668–676.
27. Lenz, E. M. & Wilson, I. D (2007). Analytical strategies in metabonomics. *J. Proteome Res.* **6**, 443–458.
28. Landau, D. P. & Binder, K. (2000). *A Guide to Monte Carlo Simulations in Statistical Physics* Cambridge Univ. Press, Cambridge, U.K.
29. Binder, K., ed. (1987) *Topics in Current Physics* (Springer, Berlin), Vol. 36, 2nd Ed.
30. Kalos, M.H. & Whitlock, P.A. (1986) *Monte Carlo Methods* (Wiley, New York).
31. Hammersley, J. M. & Handscomb, D. C. (1964) *Monte Carlo Methods* (Methuen, London).
32. Heinz-Bernd Schuttler et.al (2006), A genetic network for the clock of *Neurospora crassa*.

33. Battogtokh, D., D. K. Asch, et al (2002), An ensemble method for identifying regulatory circuits with special reference to the qa gene of *Neurospora crassa*. *Proc Natl Acad Sci U S A* . **99**(26): 16904-9.

APPENDIX

Matlab code

```
%function [] =trial1 ()
load ('C:\Users\janav862003\Desktop\Final nmrdata\samp5.txt');
load ('C:\Users\janav862003\Desktop\Final nmrdata\sample1.txt');
```

Loading the data from the location.

%mixing sucrose (sample5) and lactose (sample1)

We need the mixture of sucrose and lactose spectra

```
l=sample1 (1:4:4200, 1);
```

%In Lactose spectra, the first point to 4200 is taken in increments of 4.

```
l=l';
```

%Take transpose of matrix.

```
[m, n]=size(l);
```

```
s=samp5(1:4:4200,1);
```

```
s=s';
```

```
X=zeros(50,length(l));
```

```
r=zeros(m,1);
```

```
for i=1:50
```

```
    r=rand; %generate random number
```

```
    q=1-r;
```

```
    t=r*s+q*l;
```

Random number multiplied with spectra from sucrose and lactose.

```

    X(i,:)=t;

end

flag=1;

while(flag)

in=input('are you interested in a particular spectrum from the random matrix? (y/n)','s')

flag=strmatch('y',in);

if(flag)

i=input('what is the spectrum required number?')

end

%This command for getting the required spectra.

figure(i);plot(X(i,:))

end;

A=X';(For finding co-variance).

[m,n]=size(A);

B=mean(A,1);    % get average of columns

C= repmat(B,m,1); % create matrix of average values

D=A-C;        %subtract averages

F=(1/(m-1))*D*D';

% multiply with transpose

[fm,fn]=size(F);

% plot

G=max(F);%
```

```
figure(52);  
plot(G);  
axis tight;  
th=input('threshold ');  
n=input('number of contour lines ');  
v=[(-2)*th:th/n:(-1)*th th:th/n:2*th];  
figure(53);  
contour(F,v); % to get the final contour plot or the correlation matrix
```