Exploring Breast Cancer Risk Using NMR Metabonomics: Interstitial Breast Fluid and Breast Cancer Biomarkers

by

Rachel LeAnne Reese

Department of Chemistry
Duke University

Date:_____

Approved:

_____
Warren S. Warren, Supervisor

_____
Rosa T. Branca

_____
David N. Beratan

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science of the Department of Chemistry in the Graduate School
of Duke University

2011

ABSTRACT

Exploring Breast Cancer Risk Using NMR Metabonomics: Interstitial Breast Fluid and Breast Cancer Biomarkers

by

Rachel LeAnne Reese

Department of Chemistry
Duke University

Date:_____

Approved:

_____
Warren S. Warren, Supervisor

_____
Rosa T. Branca

_____
David N. Beratan

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Chemistry in the Graduate School of Duke University

2011

Copyright by
Rachel LeAnne Reese
2011

Abstract

For women with a high risk of developing breast cancer, the treatment options can be drastic, unnecessary, and emotionally and physically scarring. While histological evaluation of tissue samples is reliable for invasive breast cancers, the malignant potential from ductal carcinoma in situ (DCIS) cannot be reliably predicted from histopathology. Biomarkers in the breast have appeal as being directly reflective of underlying preneoplastic processes and could be used to accurately predict short-term breast cancer risk even in DCIS. By using 1H NMR metabonomics on interstitial breast fluid provided from random periareaolar fine-needle aspirations, as well as wash fluid from breast extractions, some potential biomarkers have been uncovered that could aid physicians in the prognosis of high-risk women.

Dedication

I dedicate this thesis to Gee. Thank you for your unwavering support and encouragement. Without you, I would not be nearly as excited for what the future holds and the happiness that I feel every morning when I wake up.

Contents

Abstract.....iv

List of Tables ix

List of Figures x

Acknowledgements xiii

1. Introduction 1

1.1. Introduction to Nuclear Magnetic Resonance 1

1.1.1. The Vector Model 4

1.1.2. Fourier Transform in NMR 7

1.2. Spectroscopy 8

1.2.1. 2D Spectroscopy 8

1.3. Water Suppression 15

1.3.1. Water Suppression Using Presaturation 17

1.3.2. Water Suppression Using WATERGATE 19

1.4. Multivariate Statistical Analysis 22

1.4.1. Principal Component Analysis 24

2. Metabonomics 27

2.1. What is Metabonomics? 27

2.1.1. Using NMR as a Tool in Metabonomics 28

2.1.2. Metabolites as Biomarkers for Breast Cancer 30

2.2. Breast Cancer 31

2.2.1. High Risk Patients 34

2.2.2. Treatment Options for High Risk Patients 35

2.2.3. Ductal Carcinoma in Situ 36

2.2.4. Male Breast Cancer 37

2.2.5. Random Periareolar Fine Needle Aspiration 39

3. Using Interstitial Breast Fluid and NMR Spectroscopy to Assess Risk for Breast Cancer 43

3.1. Materials and Methods 43

3.1.1. Random Periareolar Fine-Needle Aspiration Samples 43

3.1.2. Sample Preparation 44

3.1.3. High Resolution NMR Spectroscopy 44

3.1.4. Data Processing 45

3.2. Results 48

3.3. Discussion 65

4. Hyperpolarizing Using Dynamic Nuclear Polarization and the Hypersense 69

4.1. Hyperpolarization 69

4.2. Dynamic Nuclear Polarization 71

4.2.1. Solid Effect 71

4.2.2. Thermal Mixing 72

4.3. The Hypersense 73

4.3.1. Calibration and Monitoring Polarization 76

4.3.2. Troubleshooting 78

Appendix A **86**

Matlab Code for PCA **86**

Works Cited **91**

List of Tables

Table 1 Explanation of Massood cytology index scoring system.....41

Table 2 Samples of women at high risk of developing breast cancer. Twenty-six samples were included for the PCA. The patient number denotes which patient the samples were taken from and whether it was taken from the left or right breast (L or R). The patient's menopausal status was also recorded for some patients. The samples are color coded based on their Masood score risk, red being high risk, purple mid-range risk and blue little to no risk. For more details on the meaning of the Masood cytology scores, see Table 1.....47

Table 3 Amino acids present in the interstitial breast fluid samples and an example of a typical amount in μM concentration. Also present were cysteine and lysine, lactic acid, fatty acids, cholesterol, beta-hydroxybutyric acid and 2-(4-hydroxyphenylacetic) acid.....59

List of Figures

Figure 1 The net magnetization, or bulk magnetization M, is aligned along the z axis. M is stationary and aligned parallel to the magnetic field B_0. By applying a radiofrequency (RF) pulse, the magnetization can be tipped into the transverse plane, as illustrated on the right. During the RF pulse, M precesses around the B_1 field, which is orthogonal to B_0. In the illustration above, the arrow represents the bulk magnetization. 5

Figure 2 A 90 degree pulse would rotate M completely to the transverse plane. The sense and rate of the precession of M in the rotating frame depends on the resonance offset frequency, Ω. If Ω is equal to zero, then M is on resonance. If Ω is greater or less than zero, the frequency of the precession of M is greater or less than the resonance frequency, as illustrated to the right. 6

Figure 4 Traditional COSY sequence. 10

Figure 5 The DQF-COSY pulse sequence. A third 90 degree pulse is added directly after the second pulse in the conventional COSY sequence. 12

Figure 6 Possible water suppression schemes (can use one or a combination of techniques) for NMR experiments. 16

Figure 7 An example of a pulse sequence using a presaturation pulse for water suppression. The presaturation pulse is added prior to the excitation pulse, though in some sequences, it has been added in the main body of the sequence, between excitation pulses. 18

Figure 8 WATERGATE pulse sequence with nonselective 90, symmetric short field-gradients and a selective 180 degree sandwich. 20

Figure 9 WATERGATE sequence with 3-9-19 selective pulse train included for optimization of water suppression. 21

Figure 10 Breast anatomy and typical progression of breast cancer from hyperplasia (enlarged cells caused by an increase in production rate) to fully invasive ductal cancer. 33

Figure 11 Schematic for random periareolar fine-needle aspiration as designed by Carol Fabian. RPFNA is a breast biopsy technique that is minimally invasive... 40

Figure 13 Example of the spectral binning used to correct for shifts in the spectra between samples. This is applied prior to PCA calculations. 50

Figure 15 Metabolic pathway of tyrosine, provided by the KEGG database. 52

Figure 16 The top figure illustrates how much variance is accounted for in each PC, becoming less significant as you move from left to right. By PC six, less than five percent of the variance is left to consider. The plot on the bottom shows the grouping of the data with the Masood cytology scores included for context. These groups were divided using cluster analysis in PCA space. 55

Figure 17 ROC curve and plot of the sensitivity and specificity of PC 1 using various horizontal divides of the plot. 58

Figure 18 Two-dimensional DQF-COSY with water suppression that is labeled with the compounds present in the mass spectroscopy results. 61

Figure 19 The principal components and loadings plot for the tyrosine/2-(4-hydroxyphenylacetic) acid region of the spectrum. 63

Figure 20 Scores plot of the tyrosine/2-(4-hydroxyphenylacetic) acid region of the spectrum. 64

Figure 21 Distribution of the Masood cytology scores of the patient samples included in the PCA. The majority of the Masood scores fell in the mid-range of 13-14. 66

Figure 22 Saturation of the forbidden transitions, either zero- or double-quantum, leads to a positive or negative enhancement through the solid effect. 71

Figure 23 Illustration of thermal mixing. 72

Figure 24 Hypersense diagram taken from the Hypersense User's Manual. The Hypersense is based on a standard 3.35 T vertical bore magnet. The Hypersense has been altered to contain a polarizing chamber, which receives helium from the helium can of the magnet, and a sample holder that sits in the center of the magnet where the sample is placed and irradiated by microwaves at 1.4 K. Once polarized, the dissolution stick drops into the sample holder and removes the sample from the helium bath and rapidly dissolves the sample and shoots it from a capillary tube out of the Hypersense..... 74

Figure 25 Microwave sweep of pyruvic acid. 78

Acknowledgements

I would like to acknowledge the members of my group. In particular, I extend special thanks to Ashley, Yesu and Tom for all of their support and friendship through my years at Duke. I would say you all have made my frustrating days bearable and my good days great. I consider you invaluable friends. I would also like to acknowledge Morris for his help whenever I had spectroscopy questions or the 360 was doing something weird. Thank you for looking out for me. Last, I would like to thank Warren for all of the opportunities he gave me to better myself academically and challenge myself personally. Thank you, all of you.

1. Introduction

1.1. Introduction to Nuclear Magnetic Resonance

Nuclear spin is the driving force behind all magnetic resonance experiments. However, the concept of spin is challenging and abstract; as a result, the concept of spin is used as a convention developed to describe experimental evidence. Spin is an intrinsic property of a particle and is considered a form of angular momentum that is not produced by a rotation. The reason it can be considered a form of angular momentum is that spin behaves as expected from a classical description of a rotating object and fits the quantum mechanical descriptions of angular momentum. The total angular momentum of a particle with spin takes values of the form $\hbar[I(I+1)]^{1/2}$, with $2I+1$ possible sublevels. The spin quantum number, I, depends on the inherit nature of the particle being considered.

Nuclear magnetic resonance is based on the interaction of a nuclear spin with an external magnetic field, referred to as B_0. This interaction causes the spin to feel a torque, forcing it to precess around the external magnetic field (B_0). The size of the torque experienced by the spin depends both on the characteristics of

the spin as well as the size of the magnetic field. Therefore, the rate the spin precesses depends on the magnetic field and the characteristics of the nucleus. Taking into account all these factors, the frequency of the precession (ω_0) is given by the Larmor equation:

$$\omega_0 = \gamma(1-\sigma)B_0 \tag{1}$$

where γ is the gyromagnetic ratio (2.68×10^8 rad/s/Tesla for ^{1}H) and σ is the shielding constant created by electrons orbiting the nucleus.

For spin ½ nuclei, there are two stationary states; one is aligned with the magnetic field and one anti-aligned with the magnetic field, which is referred to as spin up or spin down, respectively. The magnetic moment can assume any orientation, but precession occurs in all orientations other than the spin up or spin down orientation. The difference in energy between the two states is the Larmor frequency ($\hbar\omega_0$). At room temperature, the energy difference between the states is small compared to kT, where k is the Boltzmann constant and T is the temperature in Kelvin. There is only a slight excess of spins aligned with the magnetic field (in the lower energy level) compared to those aligned against the field (in the higher energy level) for nuclei with a positive gyromagnetic ratio. Some nuclei, like nitrogen, have a negative gyromagnetic ratio, meaning the

spin-up can be found in the lower energy state. The slight excess of spins yields a net magnetic moment, which is the signal detected in an NMR experiment. The excess of spins in the lower energy level is small, often measuring approximately 1 ppm. However, taken in context, there are 10^{23} spins in a mole of water, meaning the net magnetization is not negligible. Rather than consider a single spin, Bloch [1] showed that it was possible to consider the ensemble of spins and bulk magnetization. The net magnetization for an ensemble of spin ½ nuclei is written as:

$$M_0 = \frac{\rho_0 \gamma^2 \hbar^2 B_0}{4kT} \tag{2}$$

where ρ_0 is the spin density, γ is the gyromagnetic ratio, B_0 is the external main magnetic field, k is the Boltzmann constant, 1.380 x 10^{-23} J·K^{-1}, ħ is Planck's constant divided by 2π, 1.054 x 10^{-34} J·s, and T is the temperature in Kelvin. By taking the vector sum over all spins, the behavior of the spins in a magnetic field can be described accurately by the Bloch equations, written in the rotating frame [1, 2].

$$M_{x'}(t) = e^{-t/T_2}(M_{x'}(0)\cos\Delta\omega t + M_{y'}(0)\sin\Delta\omega t) \tag{3}$$

$$M_{y'}(t) = e^{-t/T_2}(M_{y'}(0)\cos\Delta\omega t - M_{x'}(0)\sin\Delta\omega t) \quad (4)$$

$$M_z(t) = M_z(0)e^{-t/T_1} + M_0(1 - e^{-t/T_1}) \quad (5)$$

M_x, M_y and M_z are the x, y and z components, respectively, of the macroscopic magnetization of a sample placed in a magnetic field B_0. Created by an RF pulse is the B_1 applied magnetic field that is orthogonal to the main magnetic field. T_1 and T_2 are the longitudinal and transverse relaxation times, respectively. The longitudinal relaxation time is the time it takes for M_z to be restored to its equilibrium value of M_0. T_1 relaxation is the result of non-radiative interactions between the spin system and the lattice. T_2 relaxation comes from the spin-spin interactions, which cause the Larmor frequency of the spins to change and dephase the magnetization. In the absence of broadening effects (such as those from magnetic field imperfections), the spin-spin interactions determine the width of a peak. [3, 4]

1.1.1. The Vector Model

In the vector model, the bulk magnetization can be viewed as a magnetization vector that points along the direction of the applied magnetic field, by convention the z-direction. When you tip the magnetization vector

using an RF pulse by some angle β to the z-axis (Figure 1), the magnetization vector rotates about the direction of the magnetic field, sweeping out a cone shape. This precession is at the Larmor frequency, shown in equation (1). It is the precession of this magnetization vector that is detected in an NMR experiment.

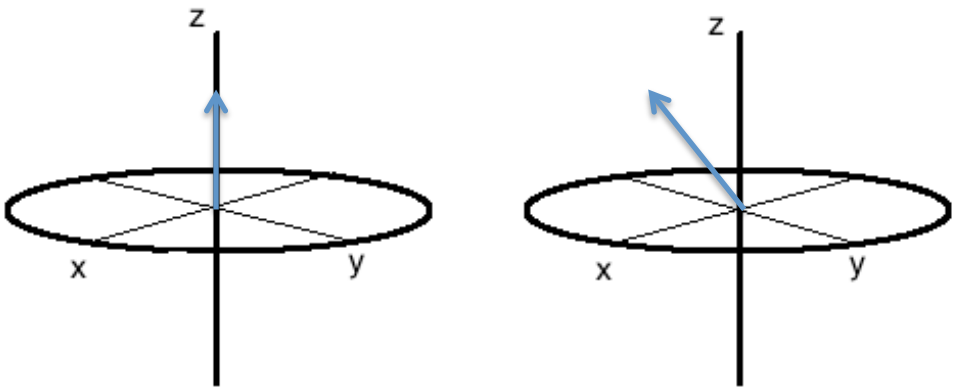

Figure 1 The net magnetization, or bulk magnetization M, is aligned along the z-axis. M is stationary and aligned parallel to the magnetic field B_0. By applying a radiofrequency (RF) pulse, the magnetization can be tipped into the transverse plane, as illustrated on the right. During the RF pulse, M precesses around the B_1 field, which is orthogonal to B_0. In the illustration above, the arrow represents the bulk magnetization.

There is a small coil of wire around the sample that the magnetization vector **M** "cuts" as it precesses in the transverse plane. Current is induced,

which is amplified and recorded. The detectable signal can be broken down into its x and y components, which are simple oscillations at the Larmor frequency.

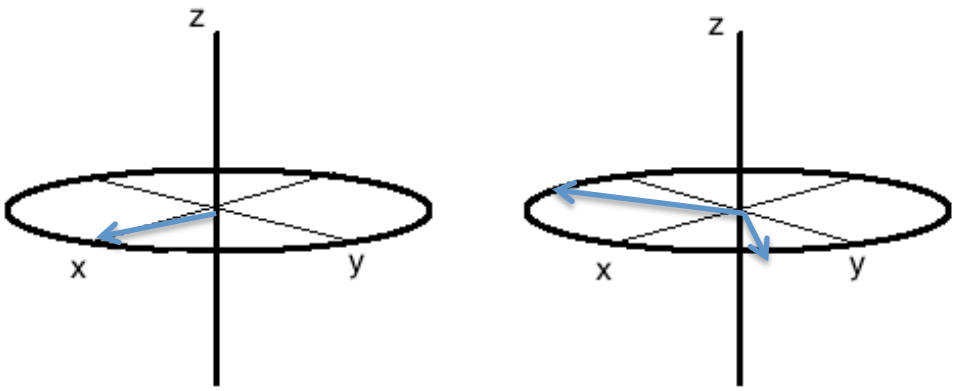

Figure 2 A 90 degree pulse would rotate M completely to the transverse plane. The sense and rate of the precession of M in the rotating frame depends on the resonance offset frequency, Ω. If Ω is equal to zero, then M is on resonance. If Ω is greater or less than zero, the frequency of the precession of M is greater or less than the resonance frequency, as illustrated to the right.

In the rotating frame, a 90 degree RF pulse will rotate M into the transverse plane and a 180 degree pulse will rotate M from the positive to negative z direction. Pulses are created by a radiofrequency that produce a static B_1 magnetic field perpendicular the main magnetic field. The flip angle of the RF pulse it determined by the nutation frequency and the pulse length. However, the magnetization produced in the transverse plane from a pulse, such as a 90

degree pulse, will not precess in the xy plane indefinitely. As discussed earlier, the magnetization will return to thermal equilibrium, or rather, experience T_1 relaxation. In addition to this relaxation, the vector model also illustrates T_2 relaxation well. This spin-spin relaxation is a loss of phase coherence between individual spins. As the magnetization precesses in the xy plane, some of the spins may have frequencies greater or less than the resonant frequency and begin to fan outward with respect to the resonant frequency.

1.1.2. Fourier Transform in NMR

The Fourier transform is a mathematical technique that converts data from the frequency domain to the time domain, and vice versa. This is an invaluable tool in data processing in magnetic resonance. As discussed briefly, the magnetization is detected by the coil in the NMR probe. This magnetization is recorded as a free induction decay (FID), which is a collection of discrete amplitude and time values gathered at a constant time interval. This FID is detected using two orthogonal detection channels along the x and y axes; this method is called quadrature detection and is used in all modern scanners. For each resonance in the spectrum, the two signals from these x and y channels are cosine and sine functions that have the offset frequency Ω and decay at the rate

$1/T_2$. These two functions are regarded as the real and imaginary components of the complex signal and can be converted into the frequency domain through a Fourier transform.

1.2. Spectroscopy

1.2.1. 2D Spectroscopy

A traditional two-dimensional experiment can be broken down to four basic time intervals. These time intervals include the preparation (τ_P) period, evolution (t_1), mixing (τ_m), and the detection time (t_2). The preparation period allows the nuclear spin system to come to an equilibrium or a nonequilibrium state with unique and desirable properties. This can be accomplished simply through a time delay or through a more elaborate pulse sequence scheme to "prepare" the system for the remainder of the two-dimensional experiment. The final step in the preparation period is to rotate the magnetization M into the xy plane, often accomplished by a simple 90° pulse.

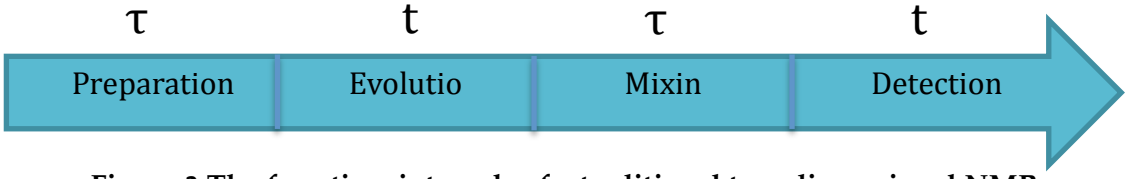

Figure 3 The four time intervals of a traditional two-dimensional NMR experiment. The time intervals include the preparation, evolution, mixing and detection time. By systematically adjusting the evolution time t_1 in a series of experiments, the connectivity and/or exchange dynamics of a compound can be determined.

The evolution time (t_1) is where the magnetization evolves freely so that M_{xu} precesses at the Larmor frequency. Each nuclear magnetization is "tagged" according to its individual characteristic Larmor frequency, which depends on their particular Hamiltonian. Mixing follows the evolution period, which gives the magnetization in the xy plane time to interact or "mix" their wave functions. It is in this time that the processes to be investigated predominately take place. The interval τ_m may be short but can also be an involved sequence of pulses that rotate M out of the xy plane and allow time for the spin system to relax. Following this mixing time is the last time interval, detection (t_2). This is the normal data acquisition period to acquire the free induction decay. Techniques in two-dimensional spectroscopy differ mainly in the structure of their mixing period. The information to determine the connectivity and exchange dynamics in a traditional 2D-sequence is accomplished by a systematic variation of t_1 from one experiment to the next. [5, 6]

1.2.2.1. The COSY Sequence

Two-dimensional correlated spectroscopy (COSY), shown in Figure 4, is often used for the delineation of the spin-spin coupling networks of a system for the purpose of spectral assignment. The spectra themselves can be either

absolute value or phase sensitive. It is through the resolved fine structure of the cross peaks that resonance assignments can be obtained to yield the quantitative measurements of spin-spin coupling constants. The cross peaks themselves are generated by the transfer of anti-phase magnetization of directly coupled spins, thus revealing connectivity in a compound.

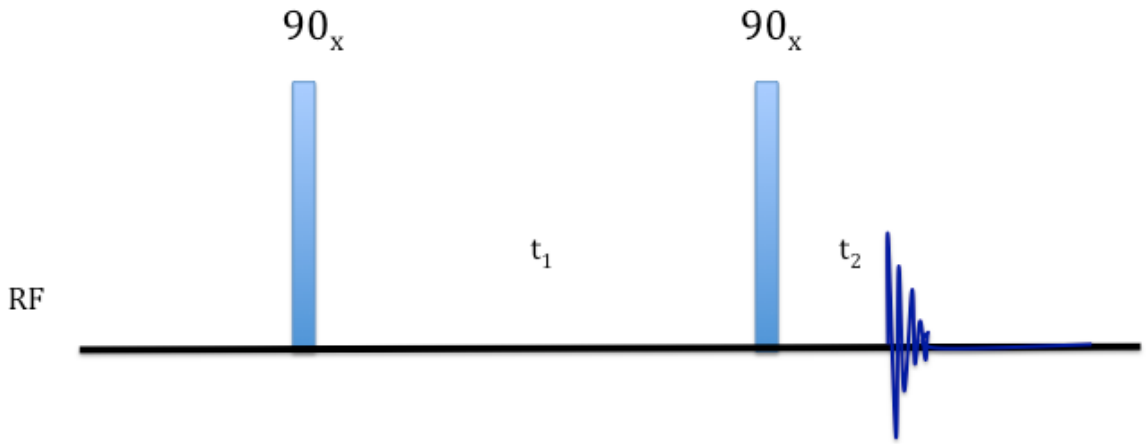

Figure 4 Traditional COSY sequence.

For a two-spin system IS, calculating the spin for I after the first pulse and evolution time t_1, the product operator can be written as follows.

$$\begin{aligned} & I_z \xrightarrow{(90_x)_{IS}} \xrightarrow{\Omega_I t_1} \xrightarrow{\pi J_{IS} t_1} \\ & -I_y \cos\Omega_I t_1 \cos\pi J_{IS} t_1 + 2I_x S_z \cos\Omega_I t_1 \sin\pi J_{IS} t_1 \\ & +I_x \sin\Omega_I t_1 \cos\pi J_{IS} t_1 + 2I_y S_z \sin\Omega_I t_1 \sin\pi J_{IS} t_1 \end{aligned} \tag{6}$$

Immediately following the second pulse, we have the product operator shown in Eq. 7. The first two terms are unobservable. The third term will evolve at the

same frequency in t_2 as it did in t_2 and thus give rise to the diagonal peaks in the spectrum. These diagonal peaks do not reveal any information beyond what can be seen in a one-dimensional experiment. The fourth and final spin evolves at Ω_2 opposed to Ω_1 in t_2 giving rise to the cross peaks that show the correlations in the spectrum.

$$\begin{aligned} & I_z \xrightarrow{(90_x)_{IS}} \\ & -I_z \cos\Omega_I t_1 \cos\pi J_{IS} t_1 - 2I_x S_y \cos\Omega_I t_1 \sin\pi J_{IS} t_1 \\ & +I_x \sin\Omega_I t_1 \cos\pi J_{IS} t_1 - 2I_z S_y \sin\Omega_I t_1 \sin\pi J_{IS} t_1 \end{aligned} \tag{7}$$

The diagonal peaks will have an in-phase multiplet structure whereas the cross peaks will reveal an anti-phase structure with respect to the J_{IS} coupling. One of the drawbacks to the traditional COSY sequence is that the diagonal peaks are intense because they cover a large area of the 2D spectrum because of their dispersive properties. In a phase sensitive COSY spectrum, the cross peaks and diagonal peaks have different lineshapes and intensities, making it difficult to analyze the information rich cross peaks that are in close proximity to the more intense diagonal peaks. One way to get around this is to suppress the signal on the diagonal of a two-dimensional experiment. To accomplish that, a sequence

that uses a double quantum filter can be applied. This is shown in Figure 5 and is referred to as a DQF-COSY. [7, 8]

1.2.2.2. The DQF-COSY Sequence

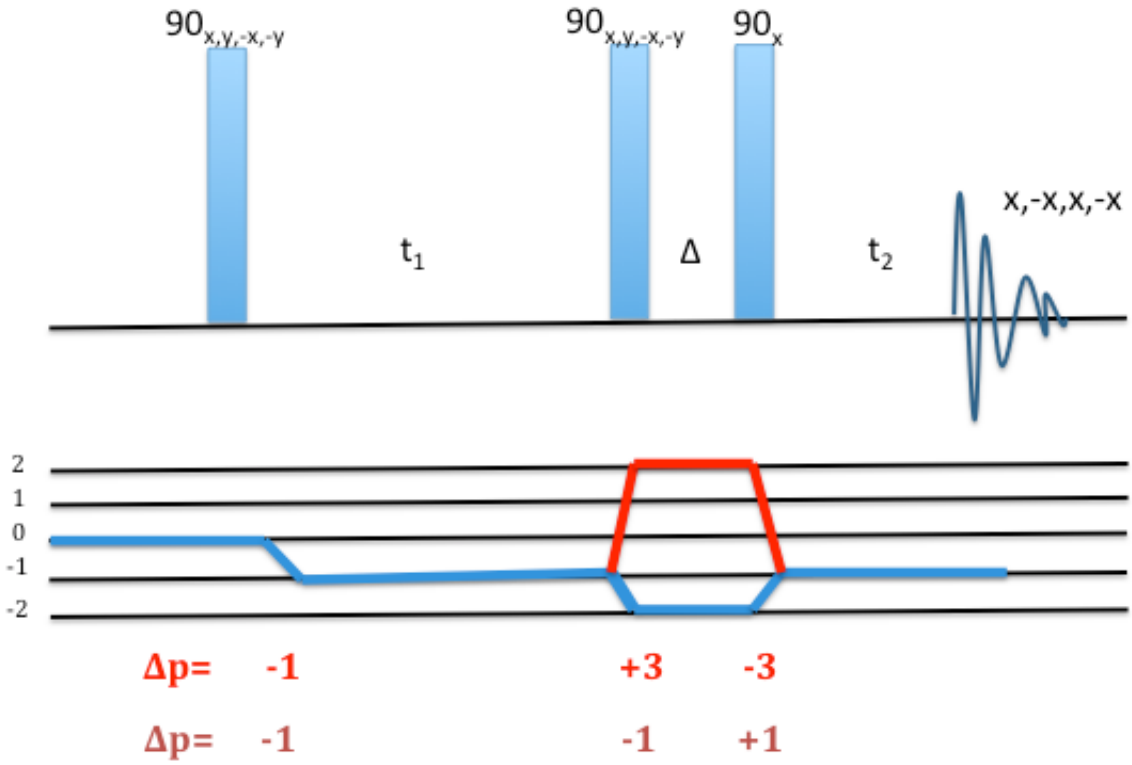

Figure 5 The DQF-COSY pulse sequence. A third 90 degree pulse is added directly after the second pulse in the conventional COSY sequence.

The DQF-COSY is much like the COSY experiment except for the addition of a third 90° RF pulse added immediately after the second pulse in the COSY sequence. This particular sequence attempts to solve the intense diagonal

problem by causing the diagonal peaks to be antiphase and nearly pure absorption. In additional to suppressing the diagonal signal, singlet resonances from isolated spins are also suppressed. The phase of the first two pulses are cycled to select only pure double quantum coherences during the negligible time interval, Δ, between the second and third RF pulses. The only term retained is the one that contains the desired double quantum coherence, as shown in equation (8).

$$I_z \xrightarrow{(90_x)_{IS}} \xrightarrow{\Omega_I t_1} \xrightarrow{\pi J_{IS} t_1} \xrightarrow{(90_x)_{IS}} -2I_x S_y \cos\Omega_I t_1 \sin\pi J_{IS} t_1 \tag{8}$$

The term that contains the desired DQ coherence to be selected by phase cycle is retained. This product operator I_xS_y is equivalent to DQ_y-ZQ_y; the zero-quantum component is removed by the phase cycling meaning the term $2I_xS_y$ can be replaced by the operator $DQ_y=(2I_xS_y + 2I_yS_x)/2$ to obtain the following terms.

$$\xrightarrow{selectDQ} -\frac{1}{2}(2I_xS_y + 2I_yS_x)\cos\Omega_I t_1 \sin\pi J_{IS} t_1$$
$$\xrightarrow{(90_x)_{IS}} -\frac{1}{2}(2I_xS_z + 2I_zS_x)\cos\Omega_I t_1 \sin\pi J_{IS} t_1 \tag{9}$$

Both of the resulting terms are observable. The first term represents the diagonal peak, which, as discussed before, evolves at the Ω_1 frequency in both the mixing

and detection time intervals. This term will be anti-phase with respect to the J_{IS} coupling in F_2 and F_1. The second term represents the cross peak, which evolves at frequency Ω_I during the mixing time interval and Ω_S in the detection time interval; this term is also antiphase with respect to the J_{IS} coupling in both F_1 and F_2. Both diagonal and cross peaks have the same modulation in the mixing time interval t_1 and phase in the detection interval t_2. This means the spectrum can be phased to be pure absorption in both dimensions.

The coherence pathway is selected by phase cycling. Phase cycling suppresses contributions from changes in coherence and selects pathways with changes in coherence order Δp. For observable signal, the overall Δp must be equal to $|1|$. For the DQF-COSY, the pathways $\Delta p=\pm 1$ are inherently selected by the first pulse in the sequence. One possible phase cycle would by **0 1 2 3**, where 0 is 90 degrees, 1 is 180 degrees, 2 is 270 degrees and 3 is 360 degrees, applied to the first two pulses in the sequence. The third pulse would remain fixed and the receiver, with the same number to degree conversions, would need to be cycled **0 2 0 2.** Such a cycle suppresses the pathway $\Delta p=\pm 1$ and so the axial peaks are suppressed. The final Δp= -3 and +1, as seen in Figure 5, and the only detectable signal has p= -1. The selection of these two pathways guarantees that the only

contributors to the observed signal will be from the coherences with orders p=±2 present just before the third pulse. [3]

The disadvantage to using a DQF-COSY protocol is that the sensitivity is a factor of ½ the sensitivity of a basic COSY experiment. [9]

1.3. Water Suppression

Water is a generic solvent used in NMR studies, but using this solvent can create an overwhelming proton signal (~100M) that can overlap or saturate weaker signals. A large water signal can also create other problems with spectra, such as baseline distortions and a phase roll. The easiest and perhaps most obvious solution to eliminate any problems with water saturation is to use a deuterated solvent. However, even deuterated solvents can contain small amounts of undeuterated/partially-deuterated solvent and deuterated solvents can be unpractical in protein studies and impossible in some biological samples. Exchange with labile protons can cause OH and NH to disappear from a spectrum, as well. To cope with a larger water signal that may otherwise be unavoidable, many methods have been developed to suppress water using

various combinations of RF pulses and gradients or post-processing techniques. [10]

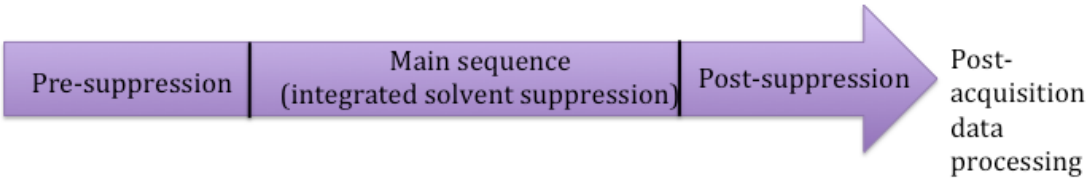

Figure 6 Possible water suppression schemes (can use one or a combination of techniques) for NMR experiments.

The "best" approach for water suppression can depend on a variety of factors, including but not limited to sample, type of information desired, spectrometer, skill of operator, time available to setup the experiment, degree of suppression desired, and the NMR experiment to be performed. Clearly, with this many considerations, there is not one answer satisfying how to suppress a solvent, in particular, a water signal. [8] The most conventional approach to water suppression is on-resonance presaturation of the water resonance. This is discussed in greater detail in the next section. However, this method can attenuate signals near the water frequency. To relieve some of the issues with unwanted attenuation, methods have been developed that use pulse schemes that leave the spin population of the water unperturbed during the relaxation delay. These techniques can be grouped in three categories.

- Jump and return/selective excitation schemes used as an acquisition read pulse
- Schemes that diphase the transverse component of water magnetization during a spin-lock purge pulse
- Schemes that apply pulsed gradients of static magnetic field that allow for the efficient selection of the coherence transfer pathway.

However, many of the methods that fall in these categories can suffer from low signal to noise in a single acquisition. When time is a crucial factor in an NMR experiment, a method would be required that could quickly produce a reliable spectrum with high signal to noise in a short amount of time. WATERGATE (water suppression by gradient-tailored excitation), discussed in more detail in a later section, produces efficient water suppression in a single acquisition. Unlike the methods described earlier, WATERGATE uses gradient pulses to remove unwanted coherences rather than select appropriate coherence-transfer pathways. [11]

1.3.1. Water Suppression Using Presaturation

Presaturation (Figure 7) is a technique where a continuous, weak (and therefore frequency-selective) radiofrequency irradiation is used prior to excitation and

acquisition. This eliminates the water magnetization in the simplest and most effective way for a pre-suppression method. In recent years, however, this method has been integrated into the main sequence, such as in the mixing time of a NOESY sequence. [10]

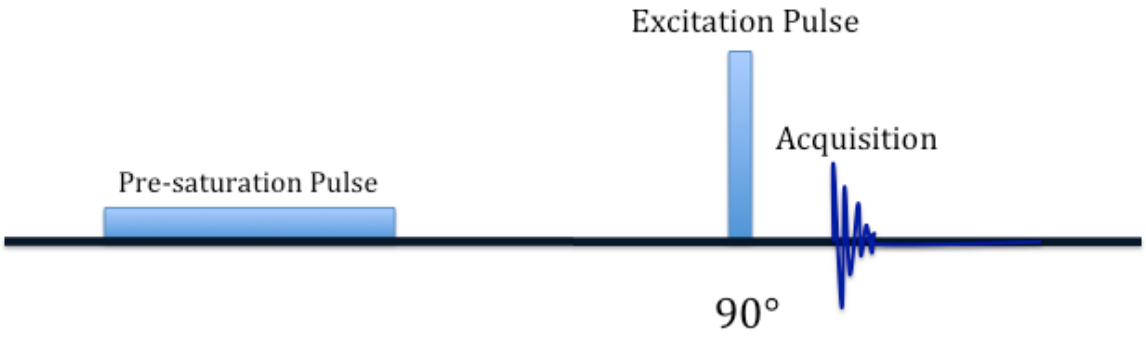

Figure 7 An example of a pulse sequence using a presaturation pulse for water suppression. The presaturation pulse is added prior to the excitation pulse, though in some sequences, it has been added in the main body of the sequence, between excitation pulses.

The choice of pulse power and duration usually require some amount of compromise, as increased power allows for efficient elimination of solvent but can saturate nearby chemical shifts and increased duration yields great saturation but reduces the efficiency of signal averaging. When considering a presaturation pulse in an NMR experiment, it is crucial that the magnet be well shimmed, but the pulse can easily be used in conjunction with most existing NMR experiments. There are some drawbacks to the presaturation technique, including saturation transfer and cross-relaxation that can reduce the intensity of

exchanging protons and that some protons lie directly beneath the water or sufficiently close that they cannot escape the saturation themselves. Some of these issues can be remedied by minimizing rates of exchange, but this is not always possible. [8]

1.3.2. Water Suppression Using WATERGATE

Water suppression by gradient-tailored excitation (WATERGATE) is a highly selective water suppression technique developed to ensure greater suppression in a single acquisition. This combination of tailored excitation pulses with magnetic field gradients is conducive to most multiple dimension NMR experiments. [11, 12] The traditional WATERGATE scheme is shown in Figure 8.

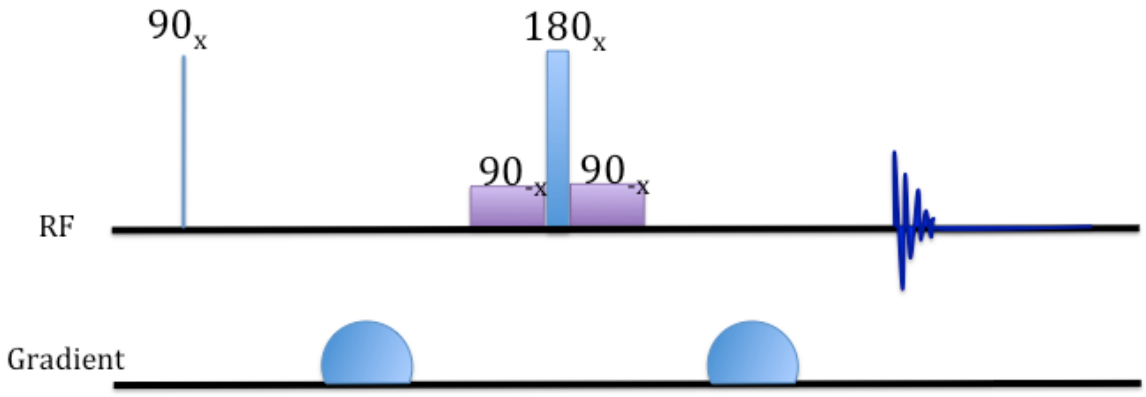

Figure 8 WATERGATE pulse sequence with nonselective 90, symmetric short field-gradients and a selective 180 degree sandwich.

A non-selective 90° RF pulse excites all resonances uniformly. The subsequent echo segment is formed by two short field-gradient pulses that have the same amplitude and sign with a central selective 180° RF pulse sandwich. The 180° RF pulse is sandwiched such that the net rotation of the water, which is set on resonance, approaches zero while the rest of the spectrum is flipped by the 180° pulse. [12] This means there is no water signal rephrased as the acquisition window opens. All coherences excited by the first nonselective pulse are dephased by the first field gradient and then rephrased by the second gradient provided they experience the 180° rotation by the selective sandwich. [11, 13]

To minimize the effects of J-modulation and spin-spin relaxation, the echo length is kept as short as possible. The length of the non-excited region in the

sandwich is determined by the length of the selective 90° pulses, given time τ. The excitation bandwidth is limited by the RF field intensity of the non-selective pulses. The spectral range is reduced to frequencies ±1/τ from the carrier frequency, which is the water frequency. [12]

Optimal refocusing can be achieved by using specialized pulse trains in place of the 180° sandwich. One of the most frequently used optimized pulse trains is the binomial 3-9-19, which denotes a symmetric sequence 3α-τ-9α-τ-19α-τ-19α-τ-9α-τ-3α (64α=180), as shown in Figure 9. The train delivers a net excitation on resonance and at frequencies at $\Delta\nu_{zero}$= ±(k)1/τ, where k= 0,1,2,3,…. [13]

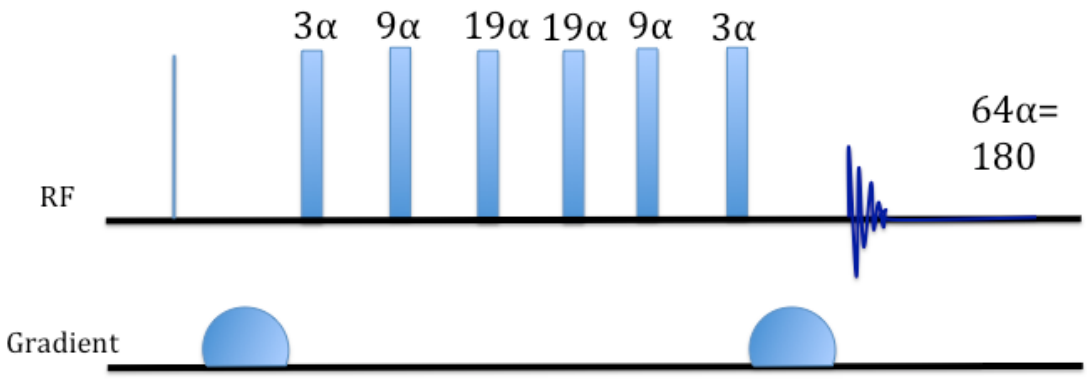

Figure 9 WATERGATE sequence with 3-9-19 selective pulse train included for optimization of water suppression.

1.4. Multivariate Statistical Analysis

The interpretation of large datasets relies on computational approaches for analysis and visualization. Methods to untangle the complicated networks of biological systems include principal component analysis (PCA), multidimensional scaling, clustering techniques, and discriminant function analysis among others. [14-17] Common to most of these methods is that they build upon the interdependencies between classes as expressed by a covariance or correlation matrix. When applying multiparametric techniques for pattern recognition in biological systems, it is important to consider where variance may be derived. Variation comes from three sources in a biological system. [14]

- Specific perturbations- These perturbations in a biological system are a specific and localized intervention.
- Global perturbations- The changes induced at multiple sites within a network of a biological system or interventions brought about by external factors that have a significant, simultaneous influence at multiple sites can be classified as global in nature.

- Intrinsic variability- Variance derived from changes that do not represent a deliberate perturbation or changes of the physiological state but rather are derived by intrinsic variability are considered intrinsic to the system.

The statistical methods mentioned previously fall into two categories, unsupervised and supervised. Techniques such as PCA are classified as unsupervised because they require no *a priori* knowledge of the class of samples and are based on the calculation of latent variables. Techniques in this category are useful when the data is rich and plentiful but the hypothesis or information regarding the data is poor. [15, 16] When there is a stronger hypothesis and more information regarding the sample, supervised techniques can be useful. These techniques include discriminant analysis and partial least squares analysis. Supervised techniques have some patterns that have a desired response that are known to the investigator. The goal of supervised methods is to model the data so there is an accurate map for the inputs with the targets. The idea is develop a "training" set that can be applied to the analysis of all the samples in a dataset. [15]

1.4.1. Principal Component Analysis

Principal component analysis is a simple, non-parametric method used to extract relevant information from large or confusing datasets. The aim of PCA is to identify the most meaningful basis in which to re-express the data while filtering out noise and revealing hidden structure. Through PCA, it is hoped that the information that is important, redundant or noise can all be discerned.

Mathematically, we must make three assumptions. First, we assume linearity, so the problem can be framed as a change of basis. Second, the assumption is made that large variances in the data have important structures that reveal useful information. Third, the principal components (PCs) are orthogonal, which is an intuitive simplification that makes PCA soluble with linear algebra decomposition techniques. PCs are linear combinations of the original descriptors that are calculated to be uncorrelated with one another and describe a decreasing amount of data variance as you sequentially move through the PCs. [16-18] Consider a data set **X**, where each column is a single sample (or snapshot in time), and **Y**, which is another $m \; x \; n$ matrix related by linear transformation to matrix **P.** **X** is the original dataset and **Y** is the new representation of the data set.

$$\mathbf{PX} = \mathbf{Y} \tag{10}$$

P is the matrix that transforms **X** into **Y** through a rotation and a stretch. The rows of **P** are a set of the new basis vectors used to express the columns in **X**. The row vectors in this transformation become the PCs of **X**.

$$\mathbf{PX} = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \tag{11}$$

$$\mathbf{Y} = \begin{bmatrix} p_1 x_1 & \cdots & p_1 x_m \\ \vdots & \ddots & \vdots \\ p_m x_1 & \cdots & p_m x_m \end{bmatrix} \tag{12}$$

An orthonormal matrix **P** must be found in **Y=PX** such that $\mathbf{C_y = (1/n)\ YY^T}$ is a diagonal matrix. The makes the rows of **P** the PCs of **X**. Using eigenvector decomposition techniques, matrix manipulation can yield the following results.

$$\begin{aligned} \mathbf{C_y} &= \frac{1}{n}\mathbf{YY^T} \\ &= \frac{1}{n}(\mathbf{PX})(\mathbf{XP})^\mathbf{T} \\ &= \frac{1}{n}\mathbf{PXX^TP^T} \\ &= \mathbf{P}(\frac{1}{n}\mathbf{XX^T})\mathbf{P^T} \\ \mathbf{C_y} &= \mathbf{PC}_x\mathbf{P^T} \end{aligned} \tag{13}$$

The matrix P is selected to be a matrix where each row, p_i, is an eigenvector of $\mathbf{(1/n)\ XX^T}$. By this definition, **P** is equivalent to $\mathbf{E^T}$.

$$
\begin{aligned}
\mathbf{C_y} &= \mathbf{P}\mathbf{C}_x\mathbf{P}^{\mathbf{T}} \\
&= \mathbf{P}(\mathbf{E}^T\mathbf{D}\mathbf{E})\mathbf{P}^T \\
&= \mathbf{P}(\mathbf{P}^T\mathbf{D}\mathbf{P})\mathbf{P}^T \\
&= (\mathbf{P}\mathbf{P}^T)\mathbf{D}(\mathbf{P}\mathbf{P}^T) \\
&= (\mathbf{P}\mathbf{P}^{-1})\mathbf{D}(\mathbf{P}\mathbf{P}^{-1}) \\
\mathbf{C_y} &= \mathbf{D}
\end{aligned}
\tag{14}
$$

In summary, PCA require that the data is organized in a *m x n* matrix where *m* is the measurement types and *n* is the number of samples, the mean is subtracted from each measurement type, and finally the eigenvectors of the covariance is calculated with the method outlined above. [18] The fundamental properties of PCs is that they are orthonormal and thus uncorrelated with one another and that the first PC contains the largest part of the variance of the dataset, with subsequent PCs containing correspondingly smaller amounts of variance. This means that a plot of the first two or three PCs gives the best representation the meaningful structure of the dataset. [16, 17]

2. Metabonomics

2.1. What is Metabonomics?

Genomics and proteomics have been crucial techniques in unraveling the mysteries of biological systems, unlocking secrets of disease and human health. To apply the genomic and proteomic knowledge effectively between variation and disease, a process had to be established that used a systems biology approach that examined metabolic pathway activity. [19] The metabolome offers information beyond that of genes and proteins alone, as it reflects more closely the activities of cells at a functional level. Changes in the metabolome are amplified with respect to changes in the transcriptome and proteome and can yield further information about changes in the genetic expression through protein production and integrated cellular biochemical regulation and control. This characterization of protein function through the phenotyping has become an important objective in functional genomics and is accomplished through metabonomics. [15, 16, 19-21] Metabonomics allows for the quantitative measurement of the dynamic metabolic response of living systems to physiological stimuli, biological stressors, environmental conditions, or genetic

modification. [16, 17, 20, 22] The ultimate goal of metabonomics is to quantitatively and qualitatively analyze all of the metabolites in a given biological system, providing an unbiased spatial and temporal snapshot that provides a global picture of actual biological events in the system. [15-17, 21, 23] Uncovering end-point biomarkers that can be used for prognosis or monitoring treatment plans has great importance in the field of personalized medicine, as will be discussed in this chapter.

2.1.1. Using NMR as a Tool in Metabonomics

For metabonomic profiling, five major modalities are currently employed and include high performance liquid chromatography (HPLC), gas chromatography-mass spectrometry (GC-MS), mass spectrometry (MS) and NMR spectrometry. Traditionally NMR and MS are considered to be the primary, most universal approaches in the field. [20] Other modalities offer strengths in metabonomics, however, NMR is unique and advantageous in that it is inherently non-destructive, non-invasive, quantitative in nature and provides a wealth of information about nuclei that are NMR visible (i.e. 1H, 13C, 15N).[16, 23, 24] While MS is more sensitive than NMR, differential ionization suppression can make pattern quantification quite difficult.[16] In addition, NMR is more

robust than MS when it comes to sample preparation, reproducibility, and absolute and relative quantitation.[20] Because NMR provides information about a chemicals structure, an NMR spectrum is a physical characteristic of a compound is this is how NMR spectroscopy is highly reproducible. [20]

Nuclear magnetic resonance spectroscopy can produce fast and detailed analysis of the biomolecular composition of crude extracts such as biofluids.[20] Offering a uniform detection system with equal sensitivity for all proton-containing molecules, 1H-NMR is able to detect changes in metabolism that occur before any gross morphological changes. In addition, numerous novel biomarkers of organ-specific toxicity have shown that NMR can excel when used as an exploratory tool.[15, 17, 25] With these qualifications in its favor, the field of metabonomics uses NMR to measure biochemical perturbations caused by disease. This has been successfully demonstrated as a viable technique with disease in the prostate, cardiac disease and Parkinson's disease through the analysis of prostatic fluid, urine, and plasma. [26, 27] The current study is aimed at uncovering a multiparametric metabolic response of interstitial breast fluid to cancer, a pathophysiological stimulus. [17] Implementing an NMR-based

metabonomic approach allows for relatively crude samples, such as interstitial breast fluid, to be analyzed with little sample preparation. [20]

2.1.2. Metabolites as Biomarkers for Breast Cancer

Metabolites can be classified as either endogenous, which means they are coded in the genome and are intrinsic to cellular function, or xenobiotic, which means they are derived from a foreign compound. [16, 19]

The need for biomarkers to refine short-term breast cancer risk estimates and measure the response to prevention interventions is evident. Biomarkers in breast tissues and fluid have appeal as being directly reflective of underlying preneoplastic processes and could be used to accurately predict short-term breast cancer risk. [28, 29] Biomarkers that vary with risk and response to prevention interventions are often termed "surrogate end point biomarkers" and must fulfill the following requirements [30]:

- Biologically and statistically significant in cancer development
- Present in a reasonable proportion of at-risk individuals
- Be obtained in minimally invasive procedure/s
- Be reversible with prevention interventions that have been validated in decreasing cancer incidence.

However, distinct surrogate end point biomarkers that predict the clinical outcome of fully developed invasive cancers with high degree of certainty and/or the response of breast cancers to anticancer therapies, as well as those that identify benign lesions that will later progress into invasive breast cancers, have yet to be identified. [31] Metabolites offer a close reflection of cell activity at a functional level. [15] Because metabolites are synthesized from other metabolites, identifying these small molecules and interpreting the correlations in the networks in which they participate offers rich potential in uncovering biomarkers that could fulfill the needs of physicians diagnosing breast cancer. However, surrogate end point biomarkers need to present in a reasonable proportion of a sample populate of at-risk individuals, should be obtainable in minimally invasive procedures, and should show reversible trends when interventions are applied. [14, 23, 28]

2.2. Breast Cancer

According to the National Cancer Institute, 1 in 8 women will be diagnosed with breast cancer in her lifetime and breast cancer results in over 40,000 deaths in the United States each year. This number is staggering, but greater still are the number of patients diagnosed with breast cancer in the United States each year,

equally more than 178,480 cases yearly. Breast cancer, in fact, is the most common malignancy among women worldwide. [32]

Numerical tools have the amazing potential of humanizing the practice of medicine, allowing the patients to become informed partners in the decisions about therapeutic options. [33]

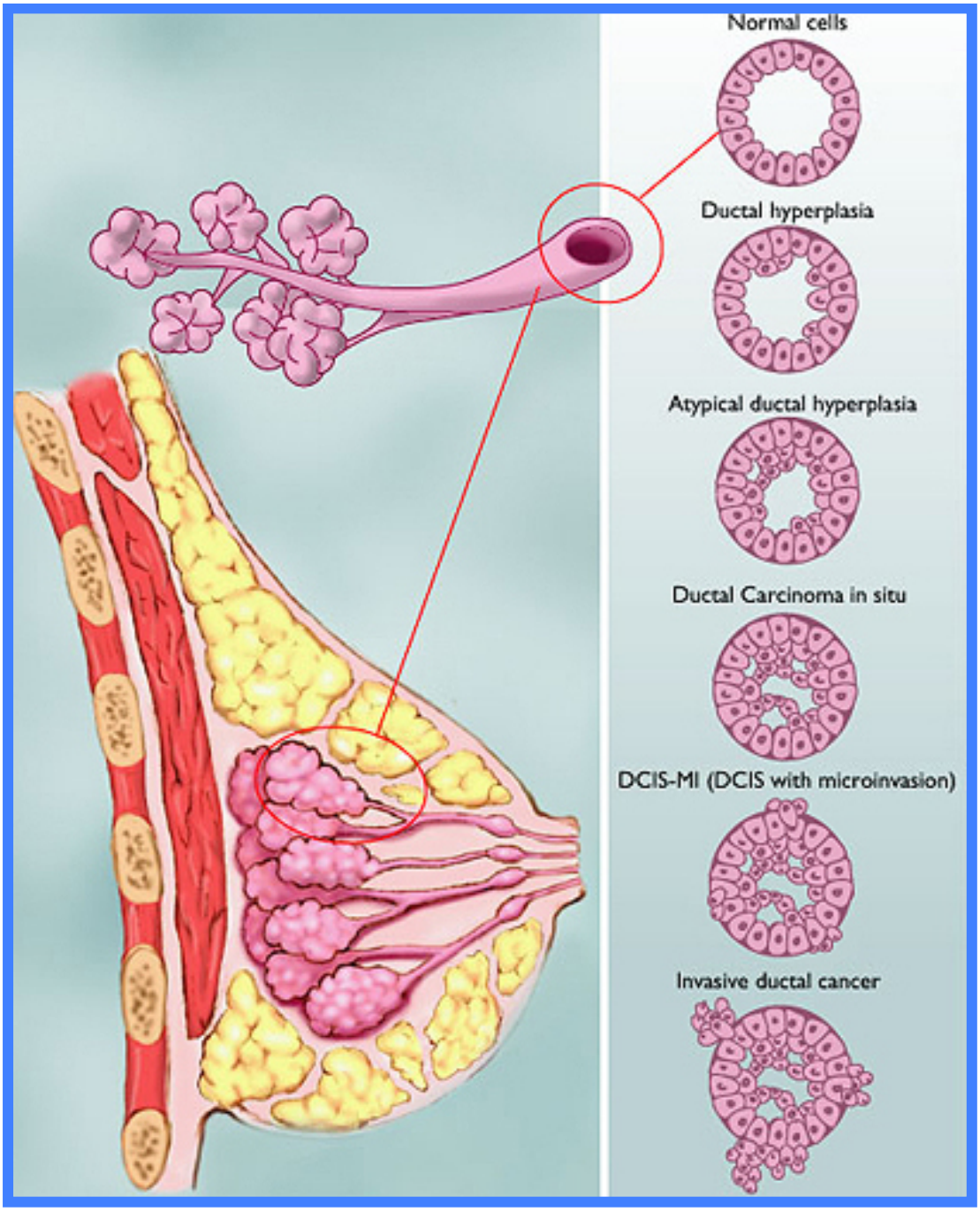

Figure 10 Breast anatomy and typical progression of breast cancer from hyperplasia (enlarged cells caused by an increase in production rate) to fully invasive ductal cancer.

2.2.1. High Risk Patients

Beyond women who are diagnosed with this disease, there has also been a category of women who are at a higher risk for developing breast cancer. Risk factors include factors such as prior biopsy exhibiting suspicious lesions, known BRCA1/2 or suspected mutation carrier, or prior contralateral breast cancer. The BRCA1/2 are breast cancer susceptibility genes that are responsible for a group of heritable breast cancer. In women, mutations in these genes confer with a 40-70% lifetime increase in breast cancer. Other factors, such as age, menopausal status, hormone or oral contraceptive use, age of menarche and menopause, location history, family cancer history (including breast, colon, ovarian and prostate cancers), radiation exposure and race also have been considered when determining whether a woman was at higher risk for developing breast cancer. [34-37] A popular model to assess the risk of in situ and invasive breast cancer, originally in white women but then expanded to include many races, is the 5-year Gail model. This model, as well as its subsequent derivatives, was used in the Breast Cancer Prevention Trials, but all versions of the model under-predict or over-predict the risk for breast cancer in some subsets of women. [28] The implication is that work still needs to be done to accurately diagnose high-risk

women to develop a personalized treatment plan based on patients' needs that may avoid the more radical treatment options. Studies show [28, 29, 34] that cancer of the breast can be greatly reduced in these high-risk women by treatment with tamoxifen, prophylactic oophorectomy (surgical removal of the ovaries) and/or prophylactic mastectomy (surgical removal of the breast). However, not all women need or benefit from drug therapy or drastic, life-altering surgery. [28, 29, 38]

2.2.2. Treatment Options for High Risk Patients

There are several treatment options available for women at a high risk of developing breast cancer or those with early onset breast cancer. However, the degree of benefits in terms of disease free survival and overall survival can be uncertain enough to make the decision for treatment or the preference for which treatment option extremely variable among patients. [33] The most widely accepted treatment options for women at high-risk of developing breast cancer are chemoprevention with tamoxifen, prophylactic oophorectomy and/or prophylactic mastectomy. [28, 29, 38] However, the risk reduction strategies carry with them serious side effects and the costs incurred by these procedures is expensive. [30] While tamoxifen has been shown to reduce breast cancer risk by

38%, side effects can include endometrial cancer, pulmonary embolism, stroke and deep-vein thrombosis. [34] In addition, chemotherapy has led to premature menopause that can lead to a chemically induced deep depression in patients. The depression rate for breast cancer has been shown to be higher than most other cancers, with the exception of pancreatic and oropharyngeal cancers. [32] In addition, known BRCA1/2 carriers have a significant risk of a cancer event in either breast, even after more conservative surgery, to the point that further interventions, such as oophorectomy, are often deemed prudent. The excess risk can take at least six to eight years to manifest itself after an initial diagnosis of high risk, so the need for long-term follow-up is crucial. [39]

2.2.3. Ductal Carcinoma in Situ

The gold standard for diagnosis of all types of breast disease is histological evaluation of tissue samples. This method has been accurate and reproducible for fully invasive breast cancer, but the criteria for identifying benign lesions at high risk for later developing into breast cancer has not been optimal. [[31, 38]] Also, there has been great uncertainty in developing an effective treatment plan for ductal carcinoma in situ (DCIS), which, in the past twenty years, has seen an increase from <1-5% of detected mammographic

malignancies to >50% of all detected mammographic malignancies. DCIS is a category of breast lesions with varied malignant potential and morphological heterogeneity making reproducible histopathological diagnosis almost impossible. The current curative treatment of DCIS has been mastectomy; however, this has often been seen as a radical overtreatment for many women who could be cured with breast-conserving treatment. [38]

2.2.4. Male Breast Cancer

While male breast cancer is considered a rare disease, it accounts for almost one percent of all breast cancer diagnoses and has seen an increase in reported cases in the past 25 years. Male breast cancer, in general, is an understudied disease, but risk factors have been identified. These risk factors include hormonal levels, testicular abnormalities, congenital inguinal hernia, orchlectomy, orchitis, infertility as well as men diagnosed with Klinefelter's syndrome, a syndrome in which patients carry XXY chromosomes. The presence of Klinefelter's syndrome alone results in a 50-fold increase in male breast cancer risk. Men with germline mutations of the BRCA2 oncogene have an estimated six percent lifetime risk of breast cancer that represents a 100-fold increase over the general male population. [36, 37]

DCIS comprises approximately ten percent of breast cancers in men, with the most common growth patterns being papillary and cribriform. However, a much larger 93.7% of male breast cancers are ductal or unclassified carcinomas. Lobular carcinomas are rare among male breast cancer patients, unlike in female patients where there is a higher incidence of lobular carcinomas in situ; apart from this departure, most histologic subtypes seen in women are also seen in men. [28, 36]

Assessing male breast cancer risk, prognosis, and treatment plans important, even if the incidence is much less than its female counterpart. Due to the similarities in the disease between the sexes, finding biomarkers for either gender could lead to great advances in risk assessment across the board. The diagnosis of male breast cancer usually does not come until there is a palpable lump with a distinct shape found in the breast, which often is very late in disease progression. The current treatment plan for men is modified radical mastectomy with auxiliary lymph node dissection or sentinel node biopsy. There is little data regarding the benefits of adjuvant chemotherapy in men, though tamoxifen is a chemotherapy drug used in male breast cancer.

2.2.5. Random Periareolar Fine Needle Aspiration

Random periareolar fine-needle aspiration (RPFNA) is a biopsy technique developed by Carol Fabian that causes little discomfort to the patient and extracts epithelial cells, interstitial breast fluid and fat from the breast. [29, 38] RPFNA samples may be used to define a cohort at very high short-term risk for DCIS and invasive breast cancer in a group that have indicators of being at high risk for the disease. [28] A 1.5 inch 21-guage needle attached to a 10-ml syringe is used for the aspiration. Eight to ten samples typically are typically pulled from each breast [28, 29] from the 10:00 and 2:00 positions around the nipple and placed in duplicate tubes, as illustrated in Figure 11. Women are instructed to wear a spandex sports bra for four to five days following the procedure to minimize breast movement. [28, 30] It is from these tubes the cytology is prepared.

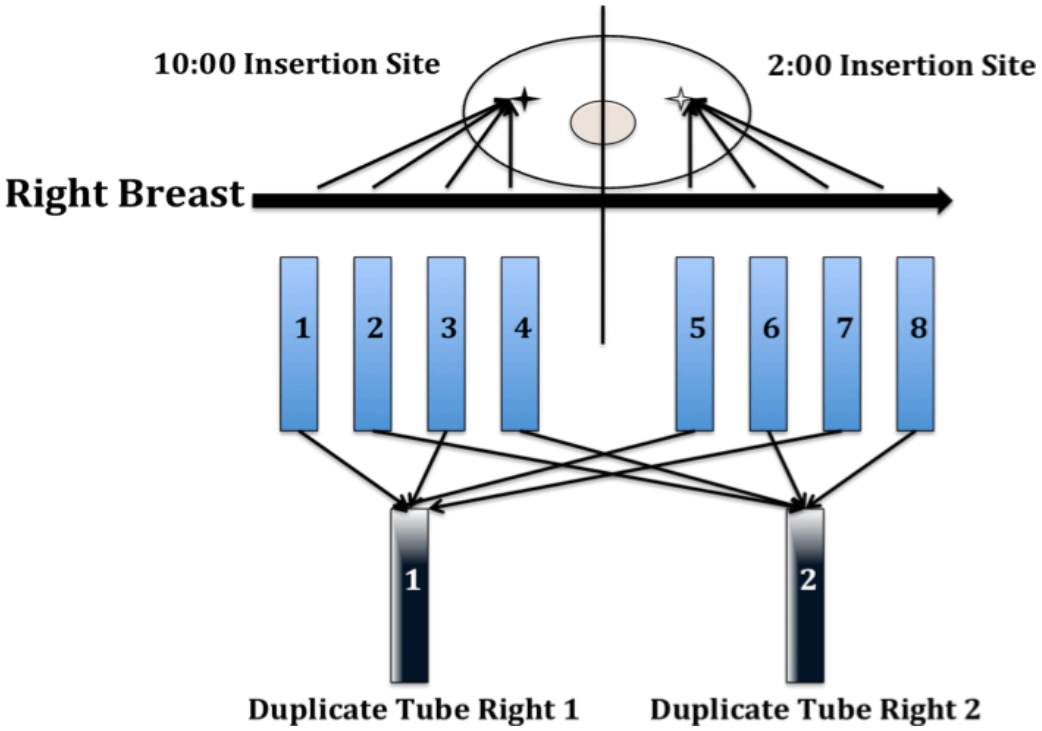

Figure 11 Schematic for random periareolar fine-needle aspiration as designed by Carol Fabian. RPFNA is a breast biopsy technique that is minimally invasive.

Table 1 Explanation of Masood cytology index scoring system.

Masood cytology index scoring system						
Cellular morphology	**Cellular pleomorphism**	**Myoepithelial cells**	**Aniso-nucleosis**	**Nucleoli**	**Chromatin clumping**	**Score**
Monolayer	Absent	Many	Absent	Absent	Absent	1
Nuclear overlap	Mild	Moderate	Mild	Occasional micronucleoli	Rare	2
Clustering	Moderate	Few	Moderate	Micronucleoli	Occasional	3
Loss of cohesion	Conspicuous	Absent	Frequent	Macronucleoli	Frequent	4

A minimum of one epithelial cell cluster with at least ten epithelial cells were required to determine the pathology and the Masood cytology index score (Table 1) was assigned for six morphologic characteristics without a knowledge of the subjects clinical history. [29, 40]

The presence of hyperplasia and hyperplasia with atypia in the FNA (fine-needle aspirates) were more prevalent in high-risk women than low-risk women and the presence of atypia in the initial RPFNA was associated with an increase in the decision to take tamoxifen when compared with women who had borderline atypia or nonproliferative cells. [28, 34] However, to date, the greatest limitations of RPFNA used to predict short-term breast cancer risk and monitor response to chemoprevention agents has been its focus on morphologic analysis. Molecular analysis has the potential to enhance the reproducibility and prognostic value. [29] With this sight in mind, metabonomics can open a new opportunities using RPFNA that have prognostic value.

3. Using Interstitial Breast Fluid and NMR Spectroscopy to Assess Risk for Breast Cancer

3.1. Materials and Methods

3.1.1. Random Periareolar Fine-Needle Aspiration Samples

Random periareolar fine-needle aspiration (RPFNA) was performed as the biopsy technique using the standards developed by Carol Fabian. This procedure extracted epithelial cells, interstitial breast fluid and fat from the breast of the patients participating in the study. [29, 38] Our collaborators collected the breast samples at the Women's Wellness Center on Duke University's Center for Living Campus as part of an ongoing clinical trial under the supervision of Victoria Seewaldt, MD. A 1.5 inch 21-guage needle attached to a syringe was used for aspiration. Eight to ten samples were pulled from each breast and collected in a test tube that was immediately centrifuged. From this tube, the interstitial breast fluid was then removed by pipette to a separate test tube containing heparin, an anti-coagulant. If the spectroscopy was performed within 24 hours of sample

acquisition, the samples were kept on ice and refrigerated; otherwise the samples were frozen until they could be analyzed at a later time.

3.1.2. Sample Preparation

Frozen samples were thawed. For each sample, 550 µl of interstitial breast fluid and 50µl of a 5mM 3-trimethylsilyl-1-propanesulfonic acid sodium salt (DSS) solution in deuterium oxide were pipetted into a 5mm glass NMR tube. If the solution volume from the patient was not 550 µl, the volume difference was accommodated by adding more of the DSS solution to reach 600 µl for all samples. The values were recorded and the internal standard concentration was calculated according to the specific values listed for each sample

3.1.3. High Resolution NMR Spectroscopy

The interstitial breast fluid samples were run on a Varian 800Mhz ^{1}H NMR. The software used was VNMRJ Version 2.1 Revision B with CHEMPACK Version 4.1. The probe is a Varian 800 MHz Cold Triple (HCN) PFG probe for work on NMR samples between 10 and 50 degrees Celsius (283 - 323 Kelvin). The NMR coil is cooled to 28 Kelvin with a continuous flow of cold helium gas supplied from a helium compressor, a helium refrigerator and two pumps. A standard presaturation (PRESAT) pulse was used to suppress the water signal.

For one set of samples that were frozen and stored in DMSO, a unique pulse design was created to suppress the signal at the chemical shifts of both water and DMSO. This was not part of the standard PRESAT sequence. The duration of the saturation pulse was two seconds. A spectral width of 9kHz was used with a 5.333 second acquisition time. The relaxation time was two seconds. Two dummy scans were included. The transmitter offset was -298.2 Hz.

3.1.4. Data Processing

The data was processed using Chenomx NMR Suite Professional Version 6.01 and MestReNova Version 6.0.3-5604. To accurately model the metabolites in interstitial breast fluid, it was necessary to determine the reproducibility in the sample collection and resulting NMR analysis. There was variation associated with the dynamic nature of the individual biological systems that had to be accounted for before significant conclusions could be made. To do this, a computer-based pattern recognition tool was used to classify metabolites by identifying inherent patterns in a set of NMR spectra. This information was used in conjunction with data obtained from mass spectroscopy analysis done at the Sarah W. Stedman Nutrition and Metabolism Center under the supervision of Christopher Newgard, MD.

The samples were catalogued according to the patient number. The PCA was an unsupervised method and the Masood cytology was not taken into account when conducting the statistical analysis, making this a blind study. Below is a table that includes the patient numbers, whether the sample was from the right or left breast, the Masood cytology index score and the menopausal status of the patient, if the information was available. Following the completion of the scores plot, the Masood scores were included in the graph to add context to the grouping and clustering that were seen; this was done so that the sensitivity and specificity could be calculated.

Table 2 Samples of women at high risk of developing breast cancer. Twenty-six samples were included for the PCA. The patient number denotes which patient the samples were taken from and whether it was taken from the left or right breast (L or R). The patient's menopausal status was also recorded for some patients. The samples are color coded based on their Masood score risk, red being high risk, purple mid-range risk and blue little to no risk. For more details on the meaning of the Masood cytology scores, see Table 1.

Sample	Score	Menopausal Status	Sample	Score	Menopausal Status
130.6L	14	post	296L	12	
130.6R	15	post	188.3L	14	post
76.6L	16	pre	188.3R	13	post
76.6R	16	pre	297L	13	
291L	14		297R	14	
291R	10		298L	9	
292R	15	pre	299L	16	
293L	14		299R	13	
293R	18		300R	16	
294L	10	post	95.5L	13	pre
294R	13	post	95.5R	13	pre
295L	14	pre	301L	14	
295R	16	pre	301R	17	

3.2. Results

After collecting and preparing the samples according to the above outlined method, a one-dimensional proton spectrum was taken of all the interstitial breast fluid. The resulting spectra had a large water signal that saturated the less intense signals of the metabolites present in the sample. Different combinations of water suppression techniques were tested, as shown in Figure 12; the best suppression was achieved using a presaturation pulse with a gradient prior to the excitation pulse and a 3-9-19 WATERGATE sequence before data acquisition. The bottom spectrum in Figure 12 illustrates the superior reduction in the water peak compared to the top and middle spectrum. This allowed for many of the metabolites that were present in much smaller concentrations to be seen above the baseline of the spectrum, particularly peaks near 7.2 ppm. (These peaks were examined in greater detail at a later time.)

After collecting the water suppressed, high-resolution spectra, they were normalized to eliminate variance due to overall sample concentration in the NMR tube (especially because these biofluids came directly from a patient and sometimes the sample size was limited or slightly diluted). To normalize the data, a peak was selected that was the same relative concentration in every

spectrum and normalized to one. This peak was at 3.45 ppm. There are other methods of normalization, but none were explored here.

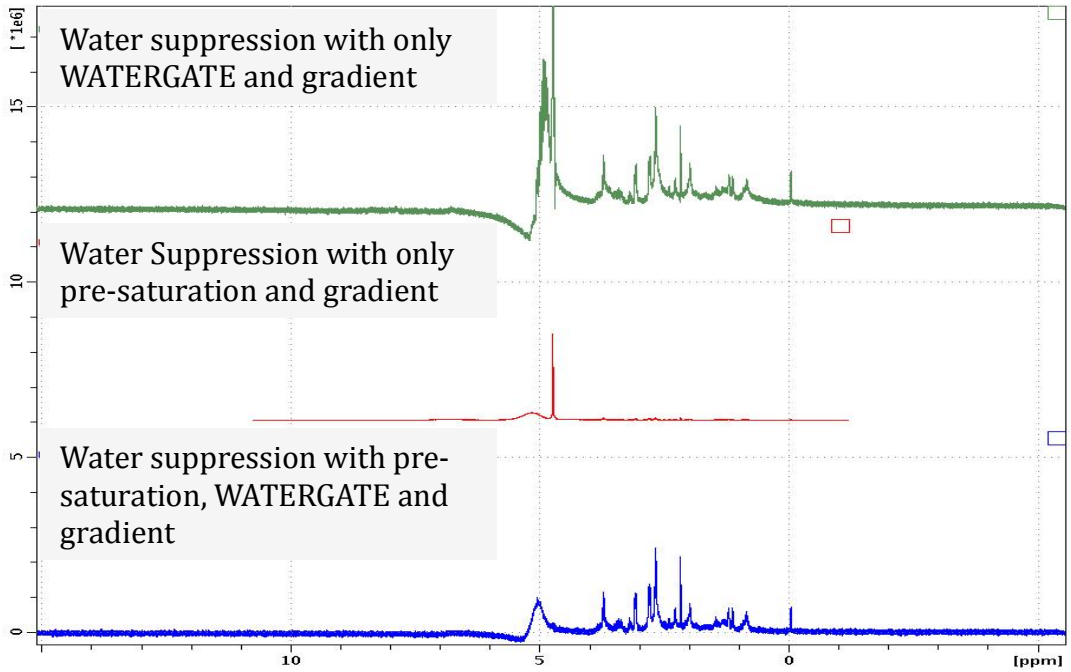

Figure 12 Example of spectra taken of the same sample with various combinations of water suppression. The best water suppression was achieved using a combination of presaturation, WATERGATE and a gradient pulse. This is represented in the bottom spectrum.

The resulting spectra were binned in increments of 0.04 ppm to eliminate the spectral shift and differences in linewidth. An example of this binning is illustrated in Figure 13. As is readily evident, the resolution of the spectrum was greatly diminished by this binning technique, subsequent work could focus on a supervised binning technique that uses smaller bin sizes in targeted regions of the spectra where metabolites of interest are present. However, for more

accurate PCA analysis, the binning was necessary to eliminate meaningless variance in the data caused by the before-mentioned spectral shifts and linewidth variation. After the normalization and binning, the PCA was completed using a Matlab code (see Appendix A).

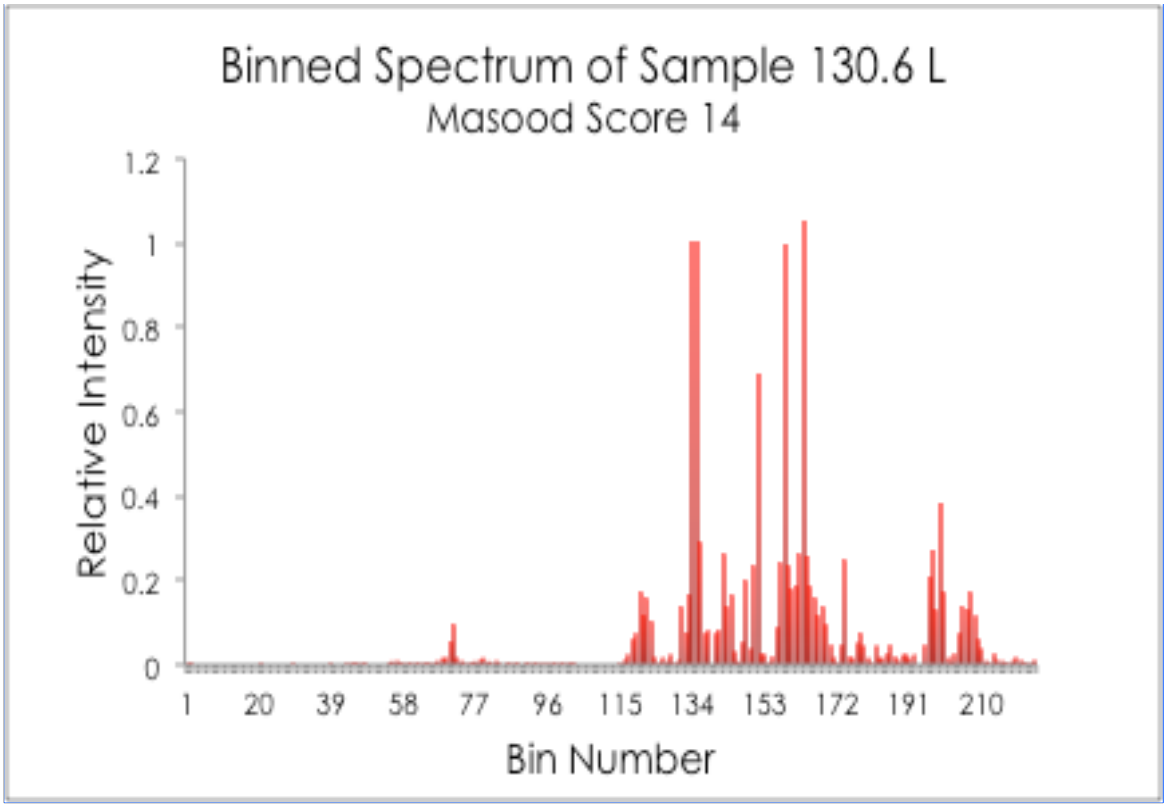

Figure 13 Example of the spectral binning used to correct for shifts in the spectra between samples. This is applied prior to PCA calculations.

Shown in Figure 14 is the loadings plot from the PCA applied to the 26 interstitial breast fluid samples. The significance of this plot is that it illustrates the areas of the spectra that contributed the most to the variance in the dataset. For the interstitial breast fluid samples, it is evident that there was a significant source of variance around 1 ppm, a significant amount of variance in the 2.5-3.7

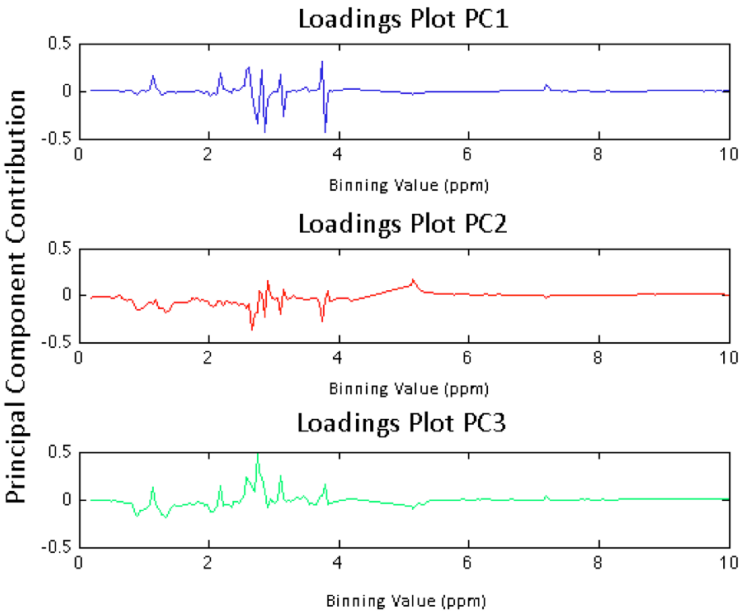

Figure 14 PCA loadings plot of the first three PCs of the interstitial breast fluid sample set. The loadings plot shows the source of variance in the dataset. There is a large variance contribution from the peak around 1 ppm, many peaks contributing in the 2-4 ppm range, and a small contribution further up field near 7 ppm.

ppm range and a small contribution around 7 ppm. Further inspection led to the conclusion that the 2.5-3.7 ppm region could be excluded from making meaningful conclusions because this is the sugar region; it can be assumed that

the glucose levels depended most heavily on what the patient ate prior to the biopsy more than indicating a meaningful underlying indicator of breast cancer risk. However, the region around 7 ppm showed great promise, which is the area associated with the tyrosine peak; this area was looked at in greater detail in two-dimensional spectroscopy to tease out finer details in this region.

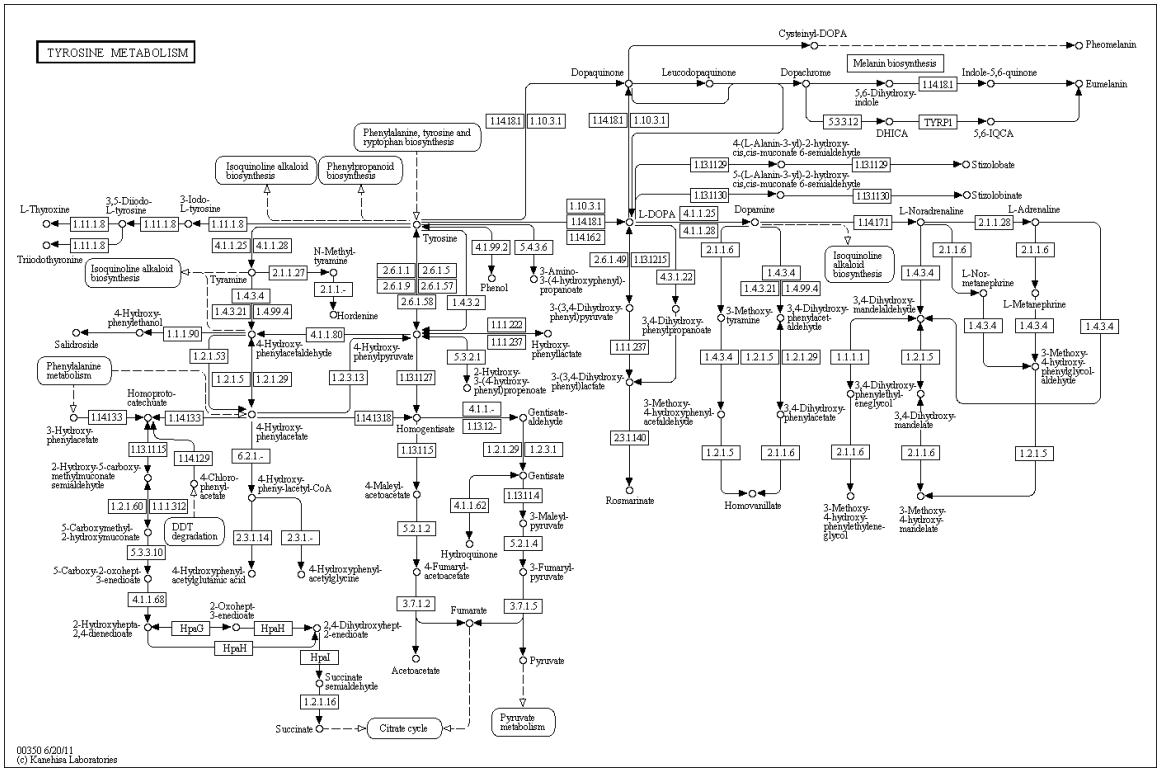

Figure 15 Metabolic pathway of tyrosine, provided by the KEGG database.

Tyrosine metabolism includes hormone production and skin pigment production as well as signal transduction. Because hormone production has been linked to breast cancer in both men and women, this was a worthwhile

metabolite to pursue. [30] The intricate metabolic pathway of tyrosine is mapped out in Figure 15. [41]

The total variance from each PC, plotted on the top of Figure 16, shows that most of the variance is accounted for within the first few PCs. There is the same number of PCs as there are bins from in a spectrum; in this particular case, there were 225 bins and therefore 225 PCs. However, the significance of each PC decreases as you move from the first PC to subsequent PCs; as illustrated in Figure 16; by PC ten, you are looking at less than one percent of the total variance in the data. By minimizing the PCs to the first few, it creates a more manageable dataset; patterns and clustering can be sought analyzing these PCs and referring back to the loadings plot for context. With the binning techniques and PCA conducted for the breast biofluid, the best clustering results are shown in the bottom scores plot of Figure 16.

The PCA was done blind; there was no organization of samples prior to using the statistical analysis technique. The samples were not sorted to higher or lower risk, they were run without knowing the Masood cytology index score. The Masood cytology scores were matched to the plot as a final step to add information to the data and meaning to the clusters that were calculated. This

means that the sensitivity and specificity shown are a result of the data separating itself through the PCA based on the variance of different metabolites in the system. Because the aim is to find women in the high-risk category that possess indicators that they will develop breast cancer, this type of separation is important and can lead to isolating key metabolites in breast cancer progression. More work needs to be done looking at samples from cohorts of healthy women, as well as those diagnosed with breast cancer, so the clusters seen in Figure 16 can be compared to both healthy and cancerous samples. However, the initial grouping of this data from high-risk women is an important fist step to uncovering meaningful biomarkers from a less invasive procedure that can lead to more informed decisions from both physicians and patients in disease prognosis.

Using a cluster analysis in the Matlab Multivariate Statistics Toolbox, we were able to separate the data into two main groups using the Masood cytology scores, as seen on the bottom of Figure 16. While not a perfect split between high and low Masood scores, there was a meaningful separation between the data that divided the lower Masood scores from the higher (meaning the lower risk scores

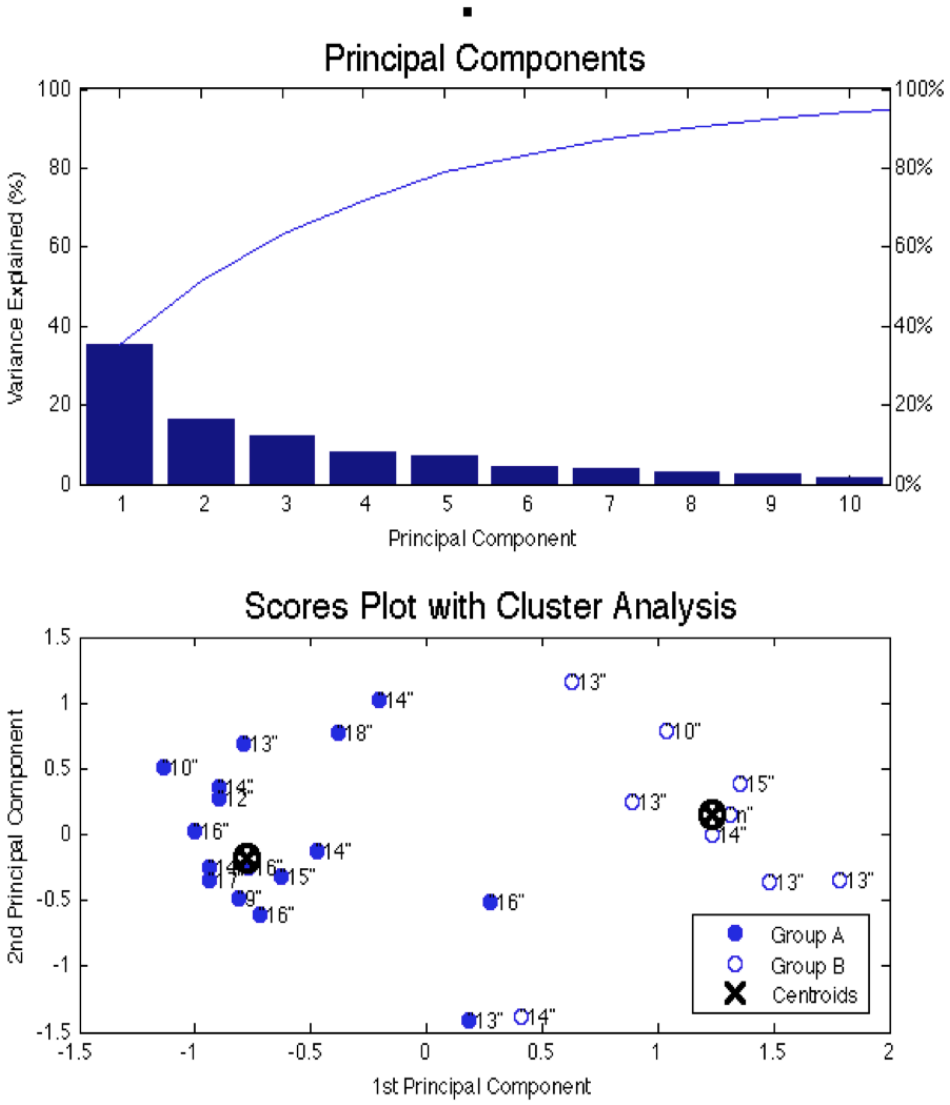

Figure 16 The top figure illustrates how much variance is accounted for in each PC, becoming less significant as you move from left to right. By PC six, less than five percent of the variance is left to consider. The plot on the bottom shows the grouping of the data with the Masood cytology scores included for context. These groups were divided using cluster analysis in PCA space.

from the higher risk scores). It is from this kind of grouping that we can hone in on specific metabolites using the loadings plot and identify where the

contributions of variance were derived. A numerical value to the separation in the low and high scores can be assigned using the sensitivity and specificity calculations.

By looking at only the first PC, and thus reducing the amount of variance we were taking into consideration, we could separate the data into groups with a sliding vertical line to determine the best placement for optimum sensitivity, the statistical measurement of true positives, and specificity, the statistical measurement of true negatives. In the case of breast cancer, the most desirable outcome is to minimize the number of false negatives. False negatives would result in a failure to treat patients to prevent the progression of breast cancer. In disease prognosis, physicians tend to err on the side of caution, which leads to the overtreatment or more radical treatment options for women who are labeled as high risk for developing breast cancer.

The sensitivity is calculated by taking the number of true positives divided by the number of true positives added with false negatives. With this binary system, we would call a true positive a patient that was placed in some cluster or group that was labeled as high risk and the patient actually had a high Masood score to support its placement. A false negative would be a patient that

had a high Masood score that was placed in the group or cluster that was labeled as low risk. Specificity is calculated in the same fashion as sensitivity but uses the rate of true negatives. (This would be the number of true negatives divided by the number of true negatives plus false positives.)

With additional tools that provide information about the rate of false negatives, treatment options for patients may be improved, such as a tendency toward more breast conserving treatment opposed to mastectomy. The sensitivity of the plot on the bottom of Figure 16 is 79%, which is very encouraging. However, the specificity is 50%, meaning there would still be a tendency to over-treat patients if using this plot as a standard. According to the Receiver Operating Characteristic (ROC) curve shown in Figure 17 for the first PC, the optimum separation of the single PC yielded 70% sensitivity. From these results, it is clear that there are metabolites that contribute to a significant separation of the data and that it is possible to, with some accuracy, divide higher and lower Masood scores using NMR spectroscopy. From the loadings plot of our PCA, the spectral regions that contribute the most, and therefore the metabolites that contribute the most, to the separation of the samples can be isolated.

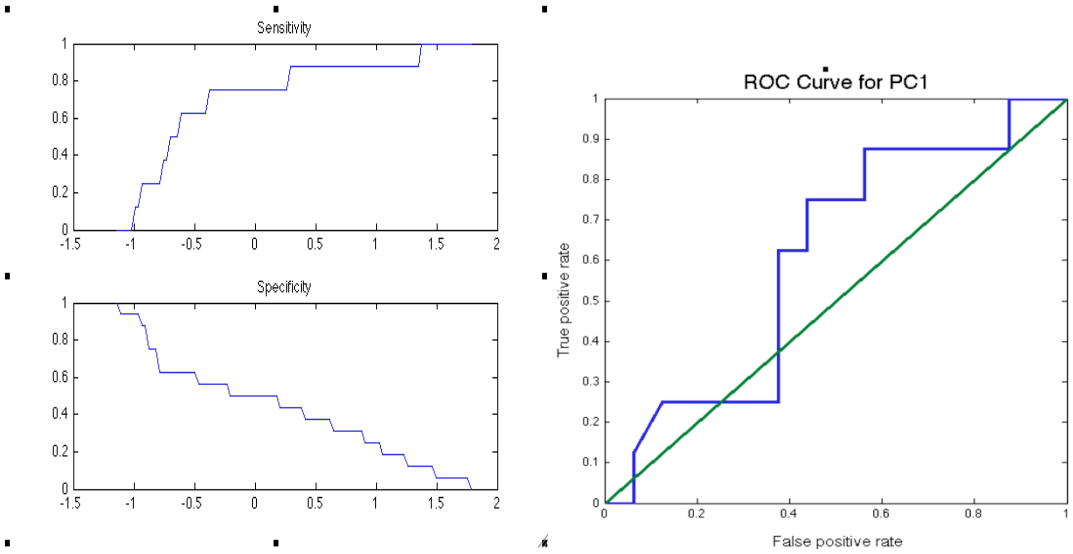

Figure 17 ROC curve and plot of the sensitivity and specificity of PC 1 using various horizontal divides of the plot.

Because the samples produced spectra with a significant amount of overlap, mass spectroscopy was also used to help identify key metabolites in the interstitial breast fluid. The most prominent metabolites are listed in Table 3. These metabolites were cross-referenced with the peaks that were seen in the two-dimensional spectroscopy done on the samples, as shown in Figure 18.

Table 3 Amino acids present in the interstitial breast fluid samples and an example of a typical amount in µM concentration. Also present were cysteine and lysine, lactic acid, fatty acids, cholesterol, beta-hydroxybutyric acid and 2-(4-hydroxyphenylacetic) acid.

Gly	Ala	Ser	Pro	Val	Leu/ Ile	Met	His	Phe	Tyr	Asx	Glx	Orn	Cit	Arg
142.4	3.7	136.1	139.1	92.7	400.1	68.5	30.6	68.5	86.1	137.2	246.2	23.2	5.5	1112.0

Using Table 3, we were able to identify many of the metabolites seen in the two-dimensional spectroscopy. The two-dimensional spectroscopy also allowed for a more detailed look at the region around the tyrosine peak. There were finer structures present that were not seen in the one-dimensional spectroscopy, and with the help of the mass spectrometry, we were able to identify the smaller peaks as 2-(4-hydroxyphenylscetic) acid (HPA). HPA is known to increase the risk of disease in humans and was of specific interest in this experiment as an indicator of breast cancer risk. [42] HPA is also part of the tyrosine metabolic pathway. [41] Some of the peaks seen in the two-dimensional spectroscopy remain to be identified. Hopefully, this can be accomplished in the near future using the metabolic databases available.

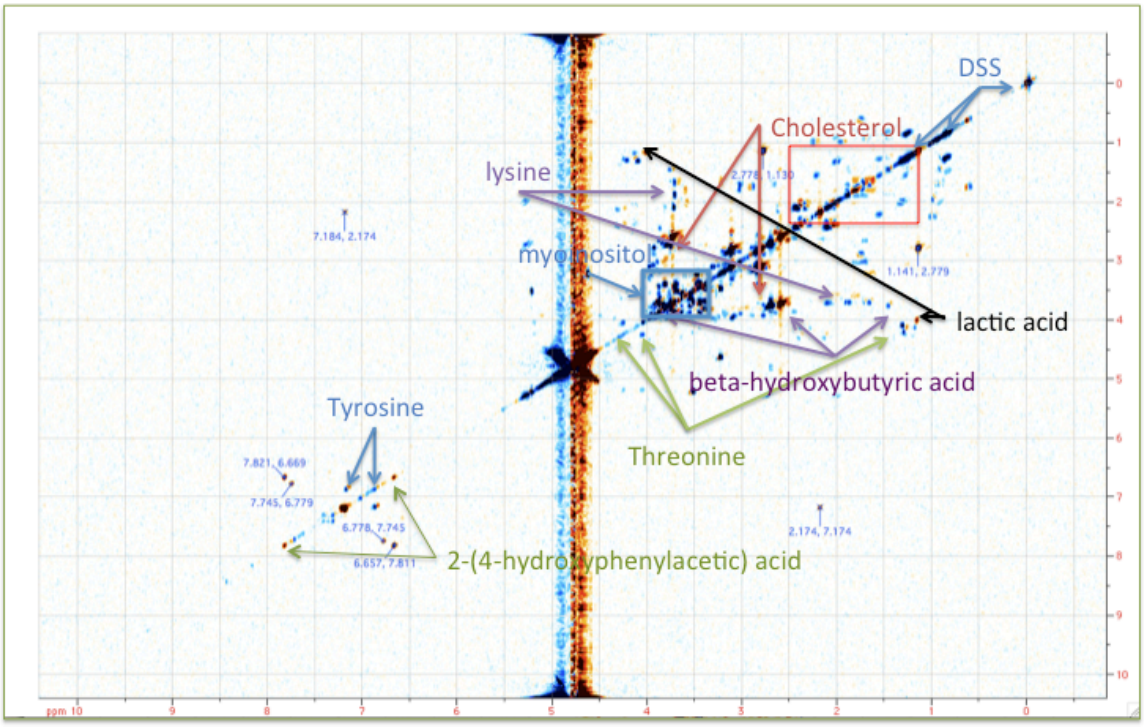

Figure 18 Two-dimensional DQF-COSY with water suppression that is labeled with the compounds present in the mass spectroscopy results.

By isolating specific regions of the spectrum, PCA could be done on targeted regions, such as in the tyrosine/2-(4-hydroxyphenylacetic) acid region. (This is the region of interest near 7 ppm.) By looking at PC 2 of this targeted region, which contained about 20% of the variance in the data and could be attributed to tyrosine and 2-(4-hydroxyphenylacetic) acid by looking at the peaks in the loadings plot, a higher sensitivity (79%) could be achieved with a single PC, though the specificity (60%) was not much changed. The loadings plot shown in Figure 19 enabled us to see that the second PC had the strongest

contribution from the region of the tyrosine peak. It also had small contributions from the 2-(4-hydroxyphenylacetic) acid peak. It was from this information that the second PC was chosen for the scores plot instead of the first PC, which clearly accounts for more of the overall variance in the data, as seen in the top of Figure 18.

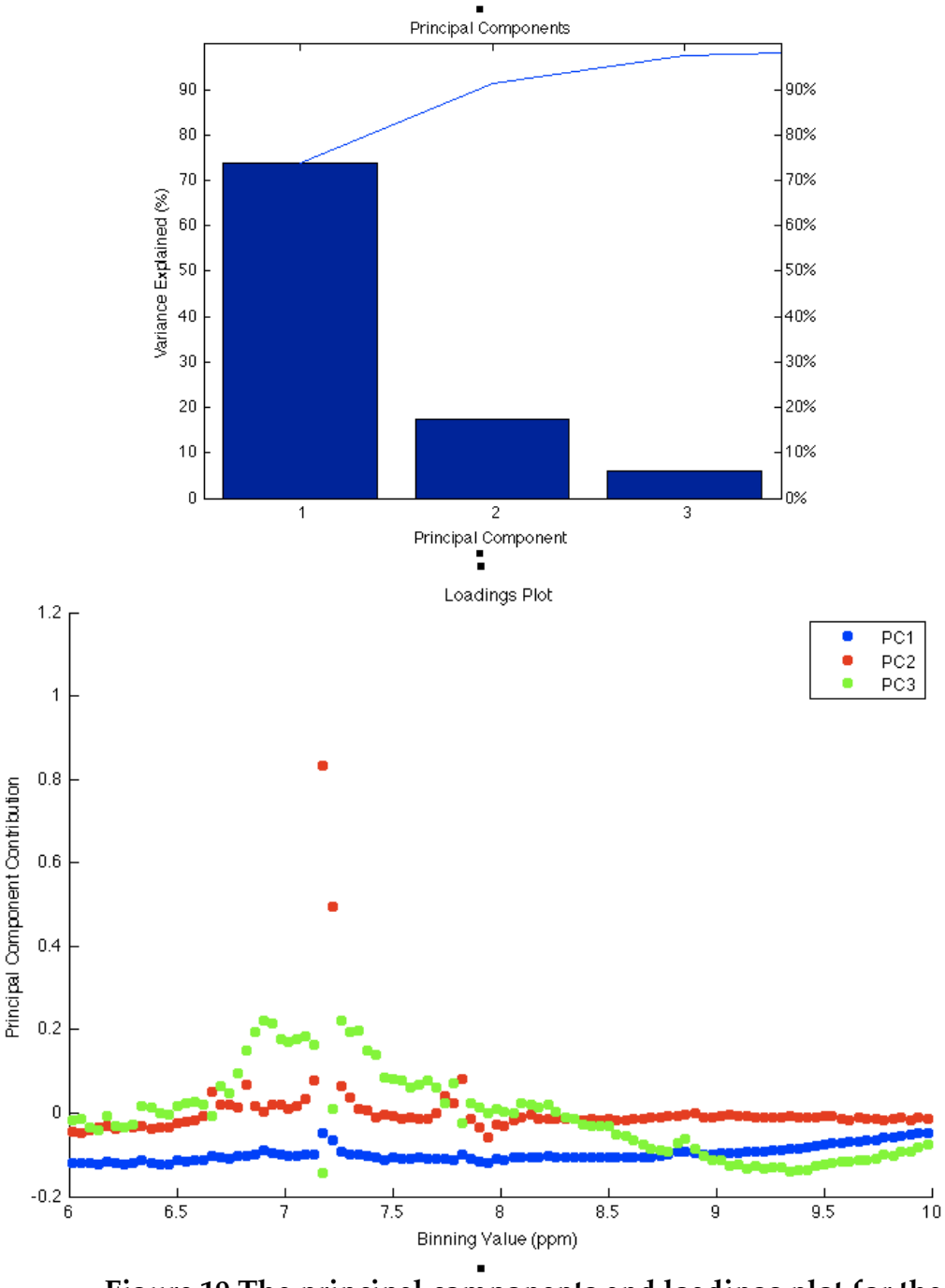

Figure 19 The principal components and loadings plot for the tyrosine/2-(4-hydroxyphenylacetic) acid region of the spectrum.

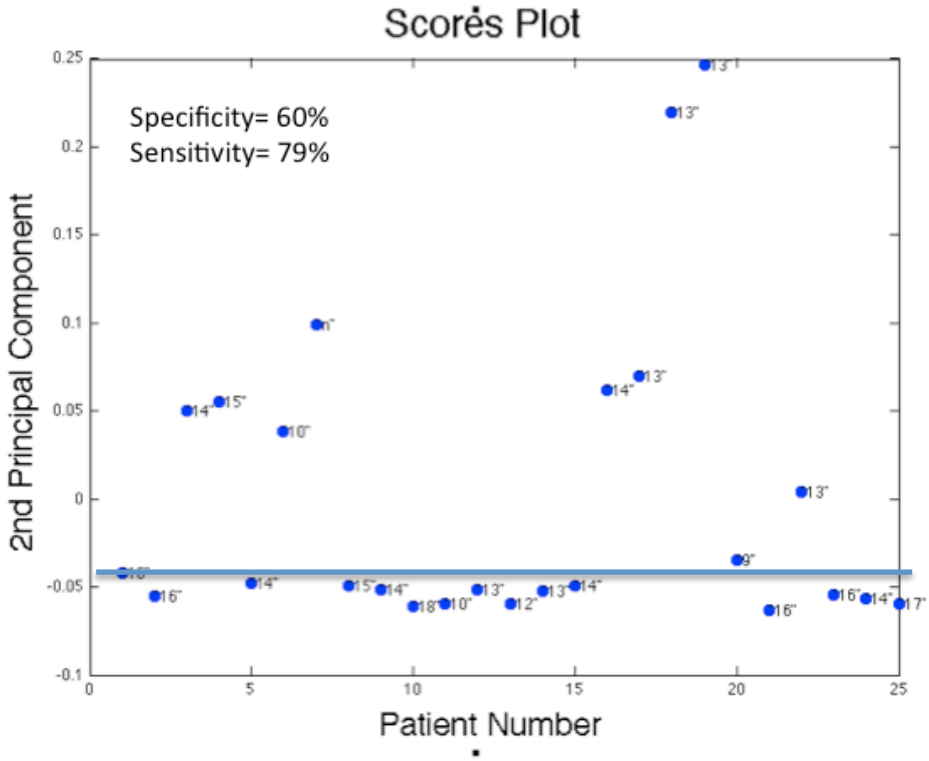

Figure 20 Scores plot of the tyrosine/2-(4-hydroxyphenylacetic) acid region of the spectrum.

Unlike the scores plot shown in Figure 16, the plot above is of only one PC. As discussed earlier, this is a plot of the second PC in the targeted region. Looking at just one PC allows a different perspective in dividing the data, and if enough targeted variance is derived from one PC, it can contain all the information needed to accurately measure sensitivity. In this case, the sensitivity was 79%, which, in conjunction with other diagnostic tools, could aid physicians in treatment plan assessment. While this dataset did not yield 100% certainty in diagnosis, it can be applied as a tool, giving more information to the

physicians that must make tough decisions concerning the personalized medicine of women considered at high risk for breast cancer. It is the combination of all three graphical representations of the PCA that we are able to optimize the way in which we examined the data. The variance bar graph allowed us to limit our PCs, the loadings plot allowed us to target the appropriate PCs and finally, the scores plot illustrated the specific variance of interest. This is an example of targeting specific metabolites for further analysis that could reveal more about the role of specific metabolites in breast cancer risk. HPA has already been identified as a key player in liver disease; its presence in the PCA of interstitial breast fluid is therefore thought be meaningful and warrants further investigation.

3.3. Discussion

Metabonomics has proven itself an important tool in the analysis of biosystems. NMR-based metabonomics has found use in a variety of fields, including the study of external stresses on a biological system, such as temperature change, profiling gender to help simplify disease detection through biomarkers, and the study of oxidation susceptibility of serum lipids using chocolate. [43-45] By looking at interstitial breast fluid with a metabonomics

perspective, we were able to gain a small insight into the elaborate metabolic networks that come together and had moderate success in using this information to uncover distinct subgroups in a cohort of high-risk women.

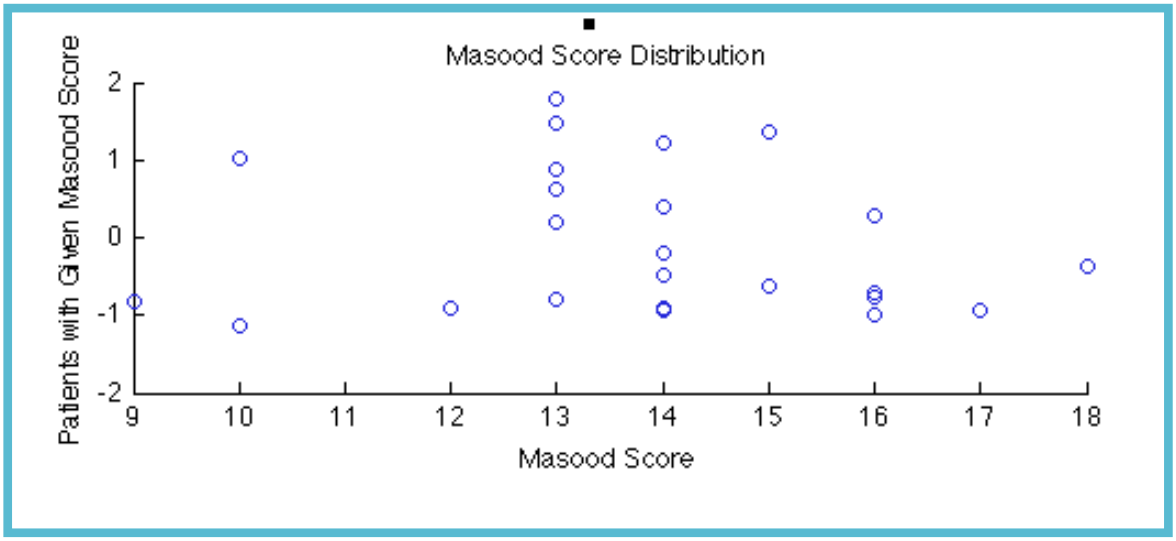

Figure 21 Distribution of the Masood cytology scores of the patient samples included in the PCA. The majority of the Masood scores fell in the mid-range of 13-14.

This project would benefit from a larger variety of samples. In the current study, all of the women are labeled as high risk and subsequently assigned a score after biopsy. The distribution of Masood scores is shown in Figure 21; most of the Masood scores fell in the 13-14 score range. For a control, women who are not at high risk, or even women who are known to have breast cancer, could shed more light on the groupings we saw in the scores plot of the PCA and show validation for the Masood scores. However, a group of women that are labeled

healthy are not guaranteed to be risk free or even cancer free. In addition, women who do not believe themselves to be at risk of developing breast cancer may be reluctant to undergo the RPFNA biopsy procedure and procuring samples from women sick from breast cancer may prove equally difficult. This study was at the mercy of an ongoing clinical trial being conducted by Victoria Seewaldt, M.D., and altering the study group may lie beyond the scope of what is feasible for our group.

However, even with the limited sample number and type, we were able to see promising results in the groupings shown in the scores plot of the PCA. With the information we gathered, we were able to group the women into higher and lower risk groups with over a 50% selectivity using metabonomic techniques alone. Because RPFNA is a much less invasive biopsy technique than taking a core biopsy sample and results in no scare tissue, it would be helpful to get as much information from these RPFNA samples as possible. The scores plots and subsequent groupings of our data could help physicians develop an individualized treatment plan for high risk by using this analytical technique. While our data does not yield 100% accuracy, it does provide another outlet of information that could be very useful when deciding on a difficult prognosis.

Furthermore, focusing on targeted groups of metabolites yields additional information. Using a global snapshot of the entire spectrum of the sample in conjunction with targeted PCA reveals very distinct groups within the high-risk women that were part of this study. Overall, our study provides initial evidence for the association of specific metabolites with breast cancer risk according to the separation of the Masood scores in the PCA. Although the exact mechanisms remain unclear and further work is necessary, the novel finding that the interstitial breast fluid metabolome is in some way connected with risk supports the idea of applying this method on a larger scale to study breast cancer risk. Further studies with a larger sample population and risk spectrum may provide new opportunities and insights to identify additional meaningful biomarkers.

4. Hyperpolarizing Using Dynamic Nuclear Polarization and the Hypersense

4.1. Hyperpolarization

In general terms, hyperpolarization refers to the significant enhancement of signal intensity of NMR visible nuclei that have a low natural abundance and are otherwise difficult to see. It is an important tool in structural and mechanistic studies of, specifically, biologically relevant molecules. [46]

In a collection of spin ½ nuclei, the polarization is determined by the following equation:

$$P = \frac{N\uparrow - N\downarrow}{N\uparrow + N\downarrow} \tag{15}$$

This equation illustrates the difference in population between energy levels where N represents the number of spins in the -½ or ½ state, respectively. The population in each energy level is determined by the energy difference between the energy levels and are calculated using the Boltzmann distribution (Eq. X)

$$\frac{N\uparrow}{N\downarrow} = e^{-\Delta E/kT} \tag{16}$$

where $-\Delta E$ is the change in energy, k is the Boltzmann constant and T is the temperature in Kelvin. The magnitude of the observable magnetization is proportional to the polarization, as shown by Eq. X

$$M_0 = \frac{1}{2} N_s \gamma \hbar P \tag{17}$$

where N_s is the number of spins in the sample. In the case of thermal equilibrium in the presence of an external magnetic field B_0, the polarization is written as

$$P = \tanh\left(\frac{\gamma \hbar B_0}{2kT}\right) \tag{18}$$

where k is the Boltzmann constant and T is the temperature in Kelvin. [47]

There are a number of different methods to achieve enhanced polarization. These include optical pumping for noble gases [48], para-hydrogen induced polarization [49, 50], and dynamic nuclear polarization (DNP) [46, 51]. The method that will be discussed in some amount of detail here is DNP, as it is the method used in the Hypersense, an instrument that sits in Bryan Research Center at Duke University.

4.2. Dynamic Nuclear Polarization

4.2.1. Solid Effect

The solid effect is a two-spin process that relies on the mixing of states. This mixing is caused by the nonsecular component of the hyperfine coupling. Irradiation at the frequency $\omega_{0S}+\omega_{0I}$ or $\omega_{0S}-\omega_{0I}$ leads to a zero-quantum or double-quantum transition. The solid effect requires a polarizing agent with a narrow electron spin resonance (ESR) spectrum so that only one of the forbidden transitions is excited at a time. This is achieved by an ESR linewidth that is much smaller than the width of ω_{0I}. If both transitions are irradiated, a partial or complete cancellation of the polarization effect occurs. [46]

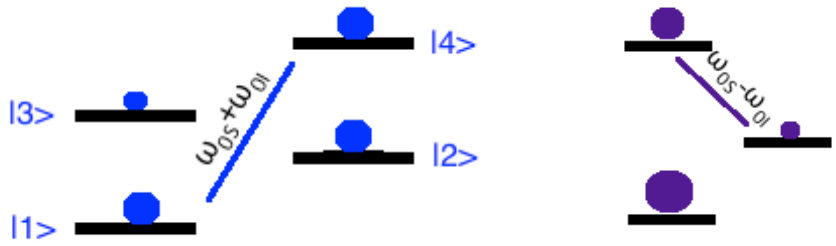

Figure 22 Saturation of the forbidden transitions, either zero- or double-quantum, leads to a positive or negative enhancement through the solid effect.

4.2.2. Thermal Mixing

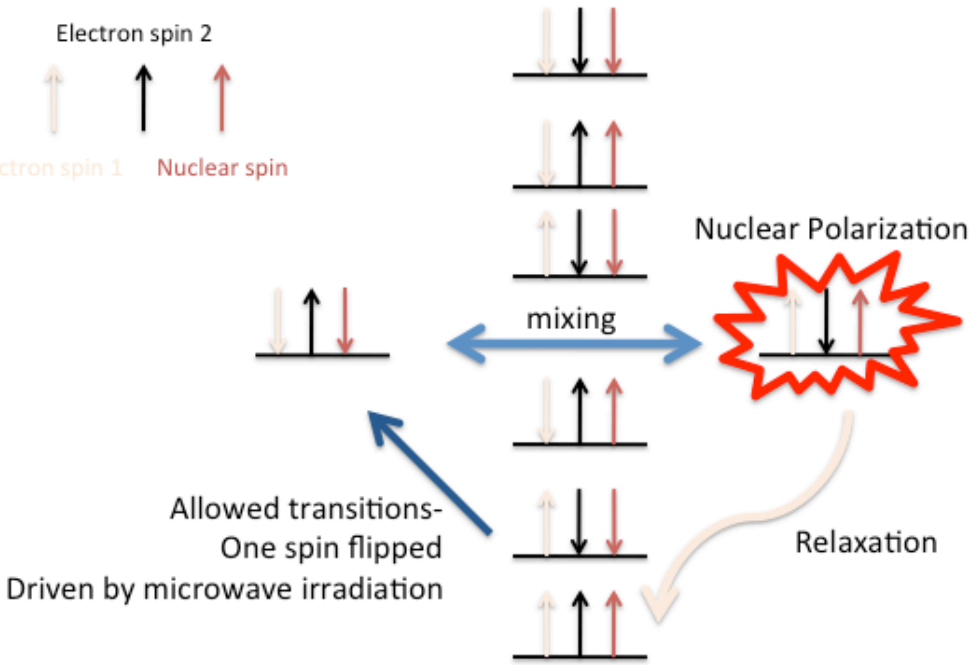

Figure 23 Illustration of thermal mixing.

Our Hypersense uses dynamic nuclear polarization where thermal mixing, shown in Figure 22, is the dominant mechanism. The thermal mixing phenomenon is often described as a three-spin electron-electron-nucleus process. DNP takes advantage of the energy difference of electrons. The method uses ESR, which is a sophisticated spectroscopic technique that detects free radicals or inorganic complexes in chemical and biological systems. In our case, the ESR linewidth is greater than the NMR frequency. For this technique, you start with a 50/50 population distribution between the two lowest energy levels. Because we are irradiating a radical, we use microwave irradiation because electrons

have a gigahertz energy separation. The microwave irradiation is set to the frequency of *e1*, which excites the *e1* spin flip. From this energy level there can be relaxation back to the original state or mixing to the equivalent energy level where the polarization is transferred to the nuclei. If mixing occurs, relaxation of *e2* can occur down to the ground state. This process slowly depletes the second energy level and builds up the population in the ground state. However, through the mixing process, there is a build-up of the polarization of the nuclei.

4.3. The Hypersense

To polarize the sample using DNP, the sample must be cooled to 1.4 K to form a glassy solid. This is accomplished by placing the sample in a liquid helium bath under vacuum. The sample is then irradiated with microwaves of a particular frequency; determining the appropriate frequency is discussed in the next section. Outside of the Hypersense, the sample is placed in a sample cup, which is attached to the end of the sample insertion stick and is inserted directly into a sample holder that is located inside the VTI. The sample holder is lowered into a helium bath.

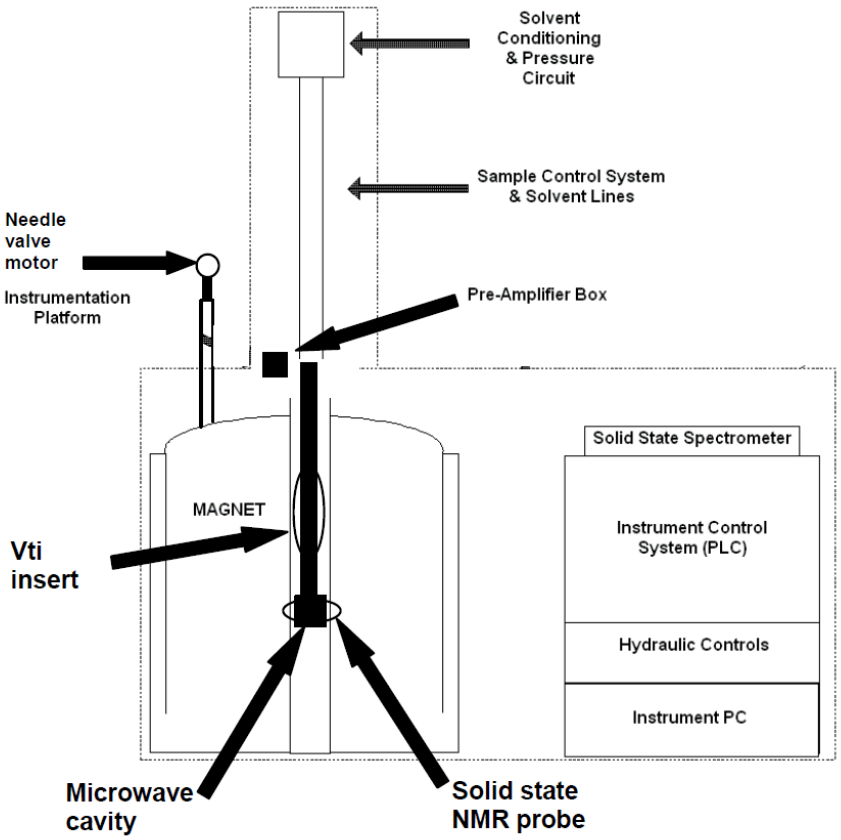

Figure 24 Hypersense diagram taken from the Hypersense User's Manual. The Hypersense is based on a standard 3.35 T vertical bore magnet. The Hypersense has been altered to contain a polarizing chamber, which receives helium from the helium can of the magnet, and a sample holder that sits in the center of the magnet where the sample is placed and irradiated by microwaves at 1.4 K. Once polarized, the dissolution stick drops into the sample holder and removes the sample from the helium bath and rapidly dissolves the sample and shoots it from a capillary tube out of the Hypersense.

When inside the Hypersense and ready for polarization, the sample is in the center of the magnet where it is irradiated by the microwave source and is held at a temperature of 1.4 K. While the sample is polarizing, the polarization

build-up is monitored using a series of low flip angle pulses. The maintenance of the helium bath under vacuum is a challenge, and the Hypersense constantly balances the flow of helium into the sample chamber and the vacuum. The helium in the sample chamber comes from the helium can of the magnet.

The probe installed on the Hypersense is designed only to detect the signal from carbon-13, although the system is capable of polarizing a wide range of nuclear spin types. However, without the appropriate probe, other nuclei are hyperpolarized blindly, it is impossible to know the polarization level. If the solid-state probe for nitrogen-15 is used, the probe is used much like the insertion stick. The sample is placed on the end of the probe (where the coil for build-up monitoring is located) and inserted directly into the Hypersense exactly like the insertion stick. However, unlike the insertion stick, the nitrogen probe should not be removed prior to the polarization procedure nor should the cup be ejected from the end of the stick. Also, the BNC cable in the back of the Hypersense must be changed from the carbon-13 cable to the longer nitrogen-15 cable. In addition, 6 dB of attenuation must be added to the system. There is a box that can be added in series to the attenuation already located on the Hypersense and is labeled with "6 dB Attenuation." The microwave frequency

must also be changed to match that of nitrogen. The probe must also be tuned and matched, which can be done using a Morris instrument.

Because the nitrogen probe remains in the system during the polarization buildup, a dissolution cannot be run. The nitrogen probe should be used to calibrate the system for the nitrogen compound and determine the optimum polarization time. Then, the sample and probe can be removed together by following the steps on the software interface. The sample should be reinserted with the insertion stick and a blind polarization should be run without the probe. This can be done exactly how a dissolution would be run on a carbon compound. [52]

4.3.1. Calibration and Monitoring Polarization

During polarization, it is possible to monitor the polarization buildup using the NMR coil located in the VTI. This coil, however, is only tuned for carbon-13 and cannot monitor the buildup of other nuclei. We have the capability of monitoring the polarization buildup of nitrogen-15 through the use of our solid-state probe. This probe has to be inserted into the Hypersense and left inserted through the duration of the polarization. This will be described in a later section. Polarization rates and levels depend on the sample formulation. Some samples,

such as neat pyruvic acid, form a glass on their own without any additional solvents. The buildup times for pyruvic acid usually range from around 45 minutes to 1.5 hours. Some compounds, however, can take upward of five hours to reach their maximum polarization. A polarization is complete when the graphic display shows a plateau in the buildup plot and there is no longer any polarization growth.

Efficient polarization requires that the frequency of the microwave source be tuned to the specific compound being polarized. This calibration is done by slowly sweeping the microwave frequency to find the frequency that corresponds to the maximum buildup. An example of a microwave sweep curve is provided in Figure 24. The microwave sweep should be completed for every compound you wish to look at, as the frequencies could vary and will affect the overall success of your polarization.

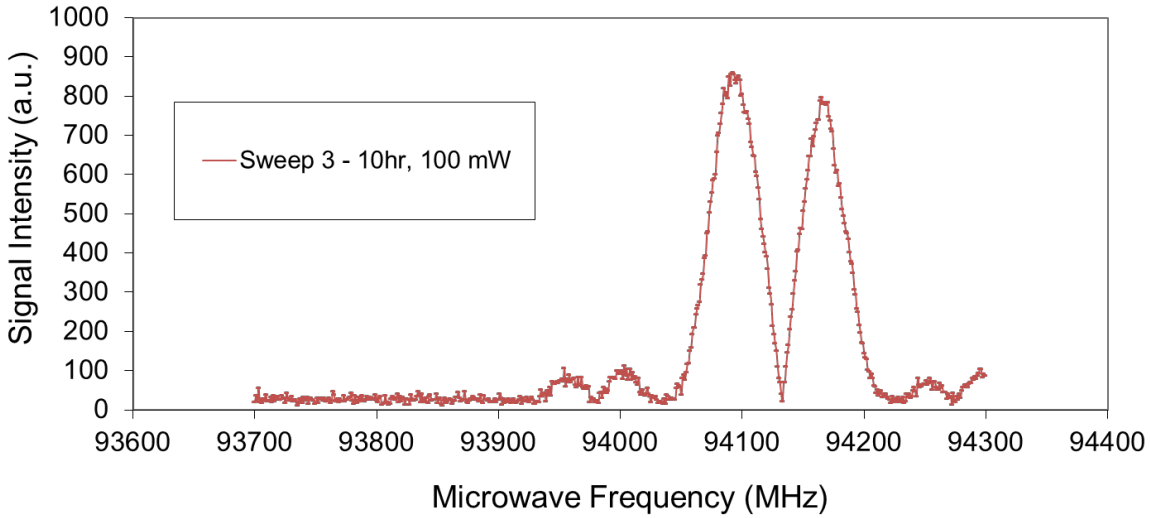

Figure 25 Microwave sweep of pyruvic acid.

4.3.2. Troubleshooting

4.3.3.1. Ice

The sample chamber is filled from the helium reservoir through a capillary tube with a needle valve. The vacuum that maintains the temperature inside the sample chamber also causes a vacuum in the helium reservoir. As a result, it is vitally important that all of the fittings to the outside environment are completely airtight. As part of routine Hypersense care, the O-rings need to be checked regularly for cracks or wear. Any cracks or irregularities mean that the O-ring is no longer creating a good seal, which allows air to penetrate the system create ice. To clear ice from the helium reservoir requires the system be warmed to room temperature (which requires de-energizing the magnet). The system

then needs to be pumped and flushed with pure helium gas. This is a costly procedure that requires the assistance of an Oxford Instrument service engineer.

4.3.3.2. Contamination

A sample is inserted into the Hypersense using a specially designed insertion stick. This insertion must be done quickly and carefully for a number of reasons. First, the VTI is open to the outside environment during the insertion process and is vulnerable to contaminants. The faster the stick is inserted and removed (or the solid-state probe is inserted), the sooner the VTI is closed to outside contamination. Second, the insertion stick must remain straight during the insertion process. This not only prevents the stick from being damaged during the procedure, but it also minimizes the chances the sample will spill. If the sample splashes up the side of the cup into the insertion stick, it can freeze to the stick and you will not be able to deposit the cup into the VTI. A worse case scenario is that you can spill sample directly into the VTI, which would introduce more contamination and affect the polarization buildup of future polarization experiments by giving false signal. As the contamination builds in the VTI, the vacuum pump becomes less effective at reducing the pressure in the

VTI and the temperature during polarization increases. At a certain point, the polarization efficiency diminishes significantly, and the system must be cleaned.

There is an automated routine called a bake-out that is implemented to clean the system. The automated procedure heats the VTI to 300 K under vacuum and then flushes the system with UHP (ultra high purity) helium gas to wash out contaminants. The combination of the high temperature (300 K) and low pressure (~0 mbar) causes any standard contamination to boil off. This pump-flush procedure is repeated many times over the course of 3 h. It is possible that one bake-out is not sufficient to clean the system, especially if something has been spilled. In more severe cases, repeating the bake-out several times (i.e., re-starting the bake-out every 3 hours for 24 hours) cleans the system effectively. However, for larger spills inside the VTI, Oxford Instruments should be called and a service engineer should be notified. Because the bake-out procedure involves heating the VTI (which is close to the magnet), helium will boil off at faster rate. It is important that there be an adequate amount of helium prior to the start of a bake-out, otherwise you risk a possible quench. In the Hypersense in the Warren group, there is usually a loss of 6% helium over the course of a 3-hour bake-out.

In addition to contamination in the VTI, there can also be contamination in the capillary tube. This is the tube that feeds the liquid helium from the helium reservoir to the VTI. If this happens, the rate at which the helium fills the VTI is diminished or sometimes blocked completely. An early indicator that the capillary tube suffers from contamination is reduced pressure during the filling period. When the Hypersense begins filling the VTI, the needle valve opens to 100%, as does the butterfly valve (to provide a vacuum on the system). For the first few minutes, only gas comes through the capillary because the VTI is too warm for liquid helium. If the capillary is unblocked, the pressure should be > 50 mbar. As the VTI cools, liquid begins to flow and the pressure will drop to 30-50 mbar. At this point, the liquid helium gauge on the VTI will begin to register the liquid. If there is a block in the capillary tube, these numbers will be significantly lower than those detailed above.

Heating the VTI and the needle valve can clear the contamination in the capillary tube. Both have small heaters that can be controlled by the software while under vacuum. It is important to monitor the temperature on both heaters; they should be turned off before the temperature gets above 300 K. The heating combined with the vacuum can clear the block after several hours.

4.3.3.3. VTI Level Calibration

The level sensor in the VTI measures the amount of liquid helium present in the VTI. The sample must be covered by liquid helium to cool properly, and an overfull VTI can result in large pressure spikes when the sample is inserted or removed. This will cause a system alarm to trip. The alarm must be acknowledged and the system reset. The reason for a poor calibration is the contamination that gets into the VTI level sensor, causing it to misread the volume of liquid or even short out. To determine if the sensor is misreading the level, watch the level of helium as the VTI begins to fill. The level meter only switches on when the temperature gets below 4 K. Once it switches on, it may read an unlikely number if the sensor is misreading. For example, if the system has been idle for a week, meaning the VTI should be empty, and is then put into cool-down that begins to fill the VTI with helium, The level sensor should not immediately read 20% or have a number like -15%. Should these types of numbers appear, the system is clearly not calibrated.

The system must be baked out to remove any contamination that might be causing a poor calibration. The level-check script in ScopeView should then be started. The system should be put into cool-down to start the VTI filling, which

can be viewed in the software as the curve collected in ScopeView. At first, there will be an adjustment in the level of the line as the system cools. As the VTI begins to fill, the line will gently slope up. When the plot has a zero slope prior to the upward slope seen in the ScopeView plot, you have found the new zero point. The cool-down can be terminated and the settings adjusted for the level meter on the HMI Hypersense software. The 0% and 100% marks are linearly related; therefore, a change in the zero point by +0.02 V will result in a change by +0.02 V for the 100% mark. The yearly service that is part of the warrantee on the Hypersense should include a calibration of the cryogen levels.

If the calibration sensor is shorting out, when the system goes into cool-down and switches on the level meter, the level meter will send a "very full" message to the computer, causing it to trip the alarm. The system can be baked out several times to remove the contamination. If that does not resolve the problem, then one of the wires connected to the level meter must be faulty or broken and the whole VTI insert should be removed by one of the Oxford service engineers.

4.3.3.4. Sample Cups

Cups can appear lost in the instrument for several reasons. Determining if a cup has been lost in the instrument is straightforward. If you start the polarization and there is no signal, then either the sample is bad or the cup is not in the polarization chamber. In this case, do not run a dissolution because the dissolution will spray solvent into the VTI. Instead, the cup can be retrieved without a polarization. You should click the button that reads "Abort" to terminate the polarization. Then, you can remove the sample. However, if after you walk through the steps to remove the sample and the cup does not appear on the end of the dissolution stick, contact Oxford Instruments.

Also, when inserting a sample, if the cup freezes to the insertion stick, there is a procedure to follow to return the system back to its ready state. When the cup remains on the insertion stick, the Hypersense "thinks" the cup has been deposited in the polarization chamber; thus the software menu options appear ready to start a polarization. To reset system so it does not "think" there is a cup in the polarization chamber, you will follow the procedure to remove a sample like there was, in fact, one in the VTI. The dissolution stick should be moved into the operating position and allowed to drop down into the VTI to pick up the cup.

Once this is completed, you may start the process over again. It is important the insertion stick be allowed time to dry after removing it from the VTI. The frost that develops on the stick causes a lot of moisture that can freeze the stick and/or the cup inside the VTI.

Often times, when a dissolution is completed and the dissolution stick is moved away from the VTI, the cup can seem stuck on the end of the dissolution stick. Allow a few minute before an attempt to remove the cup, it can be frozen to the end of the dissolution stick. Also, a gentle wiggling motion has proven effective in getting lose the toughest of cups.

Appendix A

Matlab Code for PCA

```
close all
%The following loads all of the data files for the binned,
normalized data
load tyrosine.mat
load tyrosinebins.mat

load acid.mat
load acidbins.mat

load bins.mat
load data1.mat
load masood.mat

load pre_post.mat
load pre_post_labels.mat

%The is the command for PCA; it outputs data for a loadings
%plot, scores plot, biplot, and distribution of variance
%along the PCs

[coefs,scores,variances,t2] = princomp(data);
```

```
%This will give a figure or a heirachical dendogram of the
%data, it uses the scores data from the PCA

Z = linkage(scores,'ward','euclidean');
c = cluster(Z,'maxclust',2);
figure
[H,T] = dendrogram(Z,'colorthreshold','default');
set(H,'LineWidth',2)
xlabel('Patient Number')
ylabel('Distance Between Spectra')
title('Hierarchical Tree for Groupings of High Risk Women')
labels=cellstr(masood);
 %This will plot the scores information of PCA and include
 %the Masood score
 %labels
figure
plot((1:25),scores(:,1),'b.')
text((1:25),scores(:,1), labels);
ylabel('Patient Number')
xlabel('1st Principal Component')
title('Scores Plot')
mapcaplot(scores,data);

%This will plot the loadings plot for any of the PCs that
```

```
%may be useful.
%Simply change the index to change the PC that is being
%plotted.
figure
subplot(3,1,1);
plot(bins(:,1:224), coefs(:,1),'b')
title('Loadings Plot PC1')
xlabel('Binning Value (ppm)')
subplot(3,1,2);
plot(bins(:,1:224), coefs(:,2),'r')
title('Loadings Plot PC2')
xlabel('Binning Value (ppm)')
ylabel('Principal Component Contribution')
subplot(3,1,3);
plot(bins(:,1:224), coefs(:,3),'g')
title('Loadings Plot PC3')
xlabel('Binning Value (ppm)')

%This will plot the percent of variance from each othe PCs
%in bar graph form along with the total sum of percent
%variance as a %line.

percent_explained = 100*variances/sum(variances);
figure
```

```
subplot(2,1,1);
pareto(percent_explained)
xlabel('Principal Component')
ylabel('Variance Explained (%)')
title('Principal Components')

%This will create the centroids and groups in the cluster
%analysis of the PCA scores results.

[idx,ctrs] = kmeans(scores,2,...
                    'Distance','cityblock', 'start',
                    'uniform');
subplot(2,1,2)
plot(scores(idx==1,1),scores(idx==1,2),'.','MarkerSize',20)
hold on
plot(scores(idx==2,1),scores(idx==2,2),'o','MarkerSize',7)

plot(ctrs(:,1),ctrs(:,2),'kx',...
     'MarkerSize',12,'LineWidth',2)
plot(ctrs(:,1),ctrs(:,2),'ko',...
     'MarkerSize',12,'LineWidth',2)
legend('Group A','Group B','Centroids',...
       'Location','NW')
text(scores(:,1), scores(:,2), labels);
```

```
xlabel('1st Principal Component')

ylabel('2nd Principal Component')

title('Scores Plot with Cluster Analysis')
```

Works Cited

1. Bloch, F., *Nuclear Induction.* Physical Review, 1946. **70**(7-8): p. 460-474.

2. E. Mark Haacke, et al., *Magnetic Resonance Imaging: Physical Principals and Sequence Design*1999.

3. Keeler, J., *Understanding NMR Spectroscopy*, 2004.

4. Jenista, E.R., *Development of Novel Physical Methods to Enhance Contrast and Sensitivity in Magnetic Resonance Imaging*, in *Department of Chemistry* 2010, Duke University: Durham. p. 238.

5. Becker, E.D., *High resolution NMR: theory and chemical applications*2000, San Diego: Academic Press.

6. Jeener, J., et al., *Investigation of exchange processes by two-dimensional NMR spectroscopy.* Journal of Chemical Physics, 1979. **21**(11): p. 4546-4553.

7. Rance, M., et al., *Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering.* Biochemical and Biophysical Research Communications, 1983. **117**(2): p. 479-485.

8. Hore, P.J., *Solvent Suppression.* Methods in Enzymology, 1989. **176**: p. 64-77.

9. Hore, P.J., J.A. Jones, and S. Wimperis, *NMR: The toolkit*. Oxford Chemistry Primers, ed. R.G. Compton2000, Oxford: Oxford University Press.

10. Zheng, G. and W. Price, *Solvent signal suppression in NMR.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2010. **56**: p. 267-288.

11. Sklenar, V., et al., *Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity* Journal of Magnetic Resonance 1993. **102**: p. 241-245.

12. Piotto, M., V. Saudek, and V. Sklenar, *Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions.* Journal of Biomolecular NMR, 1992. **2**: p. 661-665.

13. Lui, M., et al., *Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy.* Journal of Magnetic Resonance, 1998. **132**: p. 125-129.

14. Steuer, R., *On the analysis and interpretation of correlations in metabolomic data.* Briefings in Bioinformatics, 2006. **7**(2): p. 151-158.

15. Goodacre, R., et al., *Metabolomics by numbers: acquiring and understanding global metabolite data.* Trends in Biotechnology, 2004. **22**(5): p. 245-252.

16. Nicholson, J.K., et al., *Metabonomics: a platform for studying drug toxicity and gene function.* Nature Reviews Drug Discovery, 2002. **1**: p. 153-161.

17. Nicholson, J.K., J.C. Lindon, and E. Holmes, *'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data.* Xenobiotica, 1999. **29**(11): p. 1181-1189.

18. Shlens, J., *A tutorial on principal component analysis.* 2009.

19. Nicholson, J.K. and I.D. Wilson, *Understanding 'global' systems biology: metabonomics and the continuum of metabolism.* Nature Reviews, 2003. **2**: p. 668-676.

20. Verpoorte, R., et al., *Metabolomics: back to the basics.* Phytochem Rev, 2008. **7**: p. 525-537.

21. Griffin, J.L., *Metabonomics: NMR spectroscopy and pattern recognition analysis od body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis* Current Opinion in Chemical Biology, 2003. **7**: p. 648-654.

22. Weckwerth, W. and O. Fiehn, *Can we discover novel pathways using metabolomic analysis?* Current Opinion in Biotechnology, 2002. **13**(2): p. 156-160.

23. Kell, D.B., *Metabolomics and systems biology: making sense of the soup.* Current Opinion in Microbiology, 2004. **7**: p. 296-307.

24. Weljie, A.M., et al., *Targeted profiling: Quantitative analysis of 1H NMR metabolomics data.* Analytical Chemistry, 2006. **78**(13): p. 4430-4442.

25. Bolan, P., et al., *Imaging in breast cancer: Magnetic resonance spectroscopy.* Breast Cancer Research, 2005. **7**: p. 149-152.

26. Lynch, M.L. and J.K. Nicholson, *Proton MRS of human prostatic fluid: correlations between citrate, spermine, and myo-inositol levels and changes with disease.* The Prostate, 1997. **30**: p. 248-255.

27. Ahmed, S.S., et al., *Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection.* Journal of Biomedical Science, 2009. **16**(63).

28. Fabian, C.J., et al., *Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the gail risk model.* Journal of the National Cancer Institute, 2000. **92**(15): p. 1217-1227.

29. Bean, G.R., et al., *Retinoic Acid Receptor-B2 Promoter Methylation in Random Periareolar Fine Needle Aspiration.* Cancer Epidemiology, Biomarkers & Prevention, 2005. **14**(4): p. 790-708.

30. Ibarra-Drendall, C., et al., *Reproducibility of Random Periareolar Fine Needle Aspiration in a Multi-Institutional Cancer and Leukemia Group B (CALGB) Cross-Sectional Study.* Cancer Epidemiology, Biomarkers & Prevention, 2009. **18**(5): p. 1379-1385.

31. Davidson, N. and H. Moses, *Charting the course: Priorities for breast cancer research*, in *Report of the Breast Cancer Progress Review Group*1998, National Cancer Institute.

32. Fann, J.R., et al., *Major depression after breast cancer: a review of epidemiology and treatment.* General Hospital Psychiatry, 2008. **30**: p. 112-126.

33. Ravdin, P.M., et al., *Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer.* Journal of Clinical Oncology, 2001. **19**(4): p. 980-991.

34. Goldenberg, V.K., et al., *Atypia in random periareolar fine-needle aspiration affects the decision of women at high risk to take tamoxifen for breast cancer chemoprevention.* Cancer Epidemiology, Biomarkers & Prevention, 2007. **15**(5): p. 1032-1034.

35. Ihemelandu, C.U., et al., *Molecular breast cancer subtypes in premenopausal and postmenopausal African-American women: age-specific prevalence and survival.* Journal of Surgical Research, 2007. **143**: p. 109-118.

36. Giordano, S.H., *A review of the diagnosis and management of male breast cancer.* The Oncologist, 2005. **10**: p. 471-479.

37. Nathanson, K.L., R. Wooster, and B. Weber, *Breast cancer genetics: Whate we know and what we need.* Nature Medicine, 2001. **7**(5): p. 552-556.

38. Lindon, J.C., J.K. Nicholson, and E. Holmes, eds. *The Handbook of Metabonomics and Metabolomics*. First ed. 2007, Elsevier: The Netherlands. 355-364.

39. Haffty, B., et al., *Outcome of conservatively managed early-onset breast cancer by BRCA1/2 status.* The Lancet, 2002. **359**: p. 1471-1477.

40. Bean, G.R., et al., *Morphologically normal-appearing mammary epithelial cells obtained from high-risk women exhibit methylation silencing of INK4a/ARF.* Cancer Susceptibility and Prevention, 2007. **13**(22).

41. Laboratories, K., *Tyrosine metabolism: Reference pathway*, in http://www.genome.jp/kegg/pathway/map/map00350.html2011, KEGG Database. p. 00350.

42. Manna, S., et al., *Identification of noninvasive biomarkers for alcohol-induced liver disease using urinary metabolomics and the Ppara-null mouse.* Journal of Proteome Research, 2010. **9**(8): p. 4176-4188.

43. Gandhi, S., et al., *Metabolic regulatory variations in rats due to acute cold stress & <i>Tinospora Cordifolia</i> intervention: high resolution ¹H NMR approach.* Metabolomics, 2011: p. 1-10.

44. Zhang, S., et al., *Metabolic profiling of gender: Headspace-SPME/GC–MS and ¹H NMR analysis of urine.* Metabolomics, 2011: p. 1-12.

45. Tynkkynen, T., et al., *NMR protocol for determination of oxidation susceptibility of serum lipids and application of the protocol to a chocolate study.* Metabolomics, 2011: p. 1-13.

46. Maly, T., et al., *Dynamic nuclear polarization at high agnetic fields.* The Journal of Chemical Physics, 2008. **128**.

47. Goodson, B.M., *Nuclear Magnetic Resonance of Laser-Polarized Noble Gases in Molecules, Materials, and Organisms.* Journal of Magnetic Resonance, 2002. **155**(2): p. 157-216.

48. Goldman, M., *Overview of spin temperature, thermal mixing and dynamic nuclear polarization.* Applied Magnetic Resonance, 2008. **34**: p. 219-226.

49. Canet, D., et al., *Para-hydrogen enrichment and hyperpolarization.* Concepts in Magnetic Resonance, 2006. **28A**(5): p. 321-330.

50. Bhattacharya, P., et al., *Towards hyperpolarized 13C-succinate imaging of brain cancer.* Journal of Magnetic Resonance, 2007. **186**(1): p. 150-155.

51. Prisner, T. and W. Kockenberger, *Dynamic nuclear polarization: New experimental and methodology approaches and applications in physics, chemistry, biology and medicine.* Applied Magnetic Resonance, 2005: p. 213-218.

52. Woods, T., http://www.oxford-instruments.com/products/dnp/hyperSense/Pages/hyperSense.aspx, 2011, Oxford Instruments.