INSIGHTS INTO COLUMN CHROMATOGRAPHIC PROCESSES WITH APPLICATION TO BIOMARKER DETECTION

by

Nicole M. Devitt

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Honors Bachelor of Chemical and Biomolecular Engineering with Distinction

Spring 2018

INSIGHTS INTO COLUMN CHROMATOGRAPHIC PROCESSES

WITH APPLICATION TO BIOMARKER DETECTION

by

Nicole M. Devitt

Approved: _____
Mark R. Schure, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Christopher J. Roberts, Ph.D.
Committee member from the Department of Chemical and Biomolecular Engineering

Approved: _____
Susan Groh, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved: _____
Paul Laux, Ph.D.
Director, University Honors Program

ACKNOWLEDGMENTS

I would like to extend my deepest thanks and gratitude to Dr. Mark Schure for his guidance, encouragement, contributions, and invaluable advice throughout this work. He has taught me so much, and inspired me to work my hardest, appreciate the power of science and engineering, and dream big. Without his ideas, mentorship and ingenuity this thesis would not have been possible, and I am grateful to call him a lifelong mentor and friend.

I would like to thank my committee members, Dr. Christopher J. Roberts and Dr. Susan Groh, for their guidance throughout this process, and for giving me a chance to use the knowledge and skills they taught me in class. I would also like to thank the Undergraduate Research Program and the Honors Program for their direction, and for providing me the opportunity to write an undergraduate thesis. I am deeply grateful to the Department of Chemical and Biomolecular Engineering for the rigor and depth of my education, and for experiences I will take throughout my life. I would also like to thank the people of Advanced Materials Technology, Inc. for introducing me to the wonders of chromatography, and for providing me so many opportunities to develop both professionally and academically.

Finally, I would like to thank my family and friends for their support of me and my education. I would especially like to thank my brother, Tyler, and my sister, Shannon, who have always been my cheering squad, and my parents, Alan and Peggy, who have inspired me to follow my dreams and been there for me every step of the way. I would not have found success without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

High Performance Liquid Chromatography (HPLC) is a high-resolution separation technique with applications in medical research and natural mixture analysis. This study sought to increase the capacity of current HPLC methods to identify and quantify the presence of biomarkers for disease in two ways. First, the development of a mass-based method of determining column porosity was proposed to better quantify the performance of a column. Second, statistical overlap theory was employed in the simulation of synthetic chromatograms to illustrate the impacts of noise and peak overlap in chromatographic signals. Both of these aspects sought to assess the ability of current methods to detect and identify biomarkers present at low concentrations, providing insight into the potential of chromatographic methods to analyze biological and natural mixtures.

Chapter 1

INTRODUCTION

1.1 Background and Motivation for Study

Separation science is one of the branches of the chemical engineering field. The continual advances being made in the separation sciences are pushing the frontier of modern knowledge and increasing the application of chemical engineering principles and processes to a multitude of areas, including analysis of materials to aid the study and knowledge of diseases and how they affect the body, a pertinent cause being investigated today.

High performance liquid chromatography, or HPLC, is a widely employed analytical method characterized by the fact that moderate to high pressure is used to flow solvent through the chromatographic column. The column itself is packed with a stationary phase chemically bound on the surface of supporting particles, while the mobile phase is comprised of a mixture of solvents flowing through the column under high pressure. The sample of interest, known as the analyte, is then separated based on the interactions of its components with the stationary phase; constituents of the mixture interacting the most strongly with the stationary phase take longest to elute from the column, and are therefore detected at longer times. This technique can be applied to almost any sample which is soluble in an HPLC-compatible solvent and can be detected by a chosen detector. HPLC is the most common analytical technique in use[1], with applications in pharmaceuticals, food, cosmetics, industry and basic medical research.

1.1.1 Goals of the Work

The search for more selective, specific, and high-resolution separations is being aided by the development of porous materials and their integration into chromatography systems. Knowledge of porous materials and how they serve to aid separations is essential to both the development and analysis of chromatographic separations. The porosity of a column, or portion of a packed column that is "empty space," is made up of the spaces between the support particles, referred to as the interstitial space, and the surface-accessible space within the porous particles, referred to as the interparticle space. One of the goals of this thesis is to define a completely mass-based method for determining the interstitial and interparticle porosities of a packed column. This is especially relevant for providing information on chromatographic columns in an effort to design very high performance liquid phase separation systems.

One aspect of fundamental separation science is the development of a means to estimate the loss in resolvable chromatographic information due to noise (background signals caused by the equipment used) and finite detection limits (the lowest, or highest, signals discernable by the method). The understanding of these aspects could clarify the use of separation methods for the identification of biomarkers. Biomarkers are substances present in blood and other fluids that are sought as indicators of disease often long before symptoms start. A number of researchers believe, however, that some of these biomarkers may be present at much lower concentrations than traditionally distinguished by analyses which almost always employ chromatographic methods. The second goal of this thesis is to employ statistical overlap theory (SOT), using simulated chromatograms in MATLAB, to quantify the presence of biomarkers below traditional detection limits.

1.1.2 Significance to the Field

The ability to understand the packing of a chromatographic column and obtain information about the inner contents of the column without relying on elution techniques has great implications for high performance separations and column manufacture. This work also contributes to resolution of the resulting chromatographic signals, and highlights a need to define and estimate the loss in a chromatogram due to peak overlap and signals present below the detection limit. With the capacity to quantify this information, separation science will be better enabled to aid the search for biomarkers, and the principles of the chemical engineering field may be better applied to biomedical research.

1.2 High Performance Separations Using Liquid Chromatography

HPLC enables the separation of complex mixtures by their affinity for the stationary phase, which in packed columns is comprised of supportive particles chemically coated with molecules that define the separation mode of the column. Most commonly for HPLC, the stationary phase used is non-polar (for example, C18 bonded silica). The separation performance of the method, commonly called chromatographic resolution, depends on both the mechanical separation power, or efficiency, and the chemical separation power, or selectivity. The efficiency is dependent on the column length, particle size, and packed-bed uniformity of the column, which depend on transport properties, while the selectivity is often determined from the nature of the interactions between stationary phase and analyte in solution and most often under thermodynamic control from local phase equilibria. To this end, HPLC contributes to both faster analytical times and greater resolving power

with higher pressures, shorter columns, smaller particles, and core-shell particle technology.

1.3 Particle Technology for High Performance Liquid Chromatography and Ultra-High Performance Liquid Chromatography

Many advances in HPLC performance and separation quality have come from the utilization of particle technology to achieve separations that are faster and possess higher resolution. Monolithic columns, where porous material inside the column creates channels for the mobile phase to move through, have encountered difficulties with reproducibility, mechanical stability, and column performance. Packed columns have found great success in these areas[2].

The size of the particles used to pack the column has a large impact on the column performance; in general, porous silica particles can achieve higher performance separations with smaller diameters. As noted by Hayes and Zhang[2], although halving the particle diameter may double the number of theoretical plates (and therefore roughly double separation performance), this can result in a four-times increase in back pressure required as a result of the proportionality between pressure drop and the inverse of particle diameter squared[3]. At higher pressures, especially seen in ultra-high performance liquid chromatography (UPLC), the need for stricter (and more expensive) equipment standards arises. The solution to this issue is found in core-shell (superficially porous) particles.

1.3.1 Fully Porous vs. Superficially Porous Particles

Superficially porous particles (SPPs) have been shown to provide fast, high resolution separations without the need to employ extreme pressures. These silica particles, known as superficially porous, core-shell or fused-core particles, enable

better separations at lower pressures and larger particle sizes than both nonporous and fully porous particles (FPPs)[2]. SPPs offer an advantage over nonporous materials via a larger surface area of analyte interaction with the solid phase via pores; in nonporous particles, separation can occur only on the particle surface. SPPs also demonstrate superior mass transfer over FPPs; 2.7μm-diameter fused-core particles were shown to achieve the efficiency of sub-2μm FPPs at the pressure typically required for 3μm particles[2,4]. This eliminates the need to employ smaller diameter particles to achieve higher resolution, thus enabling more efficient separations without the need for new, higher-pressure equipment. Larger diameter SPPs have also opened the door for exploration of a variety of particle pore sizes, to tailor separations based on application.

1.4 Small to Large Pore Materials

The pore size of SPPs employed in chromatography is extremely influential in select separations, and often defines the power of the packing material to successfully separate the components of a complex mixture. Whereas small pore packings (typically on the order of 90 Å) may be useful for separating samples containing low-molecular weight constituents, current work in HPLC is attempting to tackle the significant challenge of separating proteins with wider pore materials. Human cells contain at least 20,000 proteins; the implications of an ability to separate mixtures of proteins are huge for medical, pharmaceutical and diagnostic applications[4,5]. As the size of proteins being separated increases, larger pores (e.g. 400 Å and larger) elute sharper peaks; smaller pores become increasingly limited by restricted diffusion of material[6,7].

1.4.1 Measuring Porosity

To understand column mechanics, it is important to be able to measure the porosity of a column. As mentioned previously, porosity depends on the total pore volume, also called the "void" or "dead" volume of a column, which can be broken down into two main parts: the interstitial volume, or space between the particles in a packed column; and the intraparticle volume, the space in the column due to the pores of the particles[8]. Although several methods exist to determine the interstitial volume experimentally, an example of which is inverse size exclusion chromatography (ISEC), a significant advantage would be associated with a completely mass-based method of determination[9–13]. One of the contributions of this work is a means of determining the porosity of a column quickly and accurately via measurements of mass, thus eliminating the need to rely on void times or tracer transport through a column. Total pore blocking (TPB) is employed in this study to cover the intraparticle pores using an immiscible solvent technique introduced previously in the literature.

1.5 The Search for Biomarkers by Liquid Chromatography and Mass Spectrometry

In analysis of biological mixtures of proteins, it is important to recognize the influence of biomarkers on current medical research and separation analyses. Biomarkers are typically lower concentration biomolecules present in bodily fluids such as blood and urine that can play a vital role in a number of diseases[14]. Although these components can take the form of a wide variety of natural components, the form being investigated most extensively today is protein biomarkers, i.e. those derived from the proteome. Smaller molecules, especially those of environmental origin, are still of interest in researching disease states. However biomarkers, if properly separated and detected, could serve as indicators of neurodegeneration in Alzheimer's,

development of cancer, and predictors of cardiovascular disease, among many others[14–16]. The difficulties posed in detecting these components, however, come from the fact that they are often present in concentrations currently undetectable by modern methods.

1.5.1 Statistical Overlap Theory and Limit of Detection

Low-concentration peaks may be lost in a chromatographic signal as a result of peak overlap, the limit of detection of the equipment, or a combination of both. Because of this, it is possible that a majority of biomarkers cannot currently be distinguished via HPLC with either a single channel ultraviolet detector or some form of mass spectrometry detector. The capacity to estimate this, however, lies in the use of statistical overlap theory (SOT) to estimate the resolution of a chromatogram[17,18]. By determining a likely probability density function (PDF) for the resulting peak amplitudes, it is possible to model the signals present below the method's limit of detection, quantify the signals being lost due to overlap by neighboring signals, and measure the missing information in an existing chromatogram. This work will employ SOT, using simulated chromatograms in MATLAB, to contribute to the identification of low-concentration and overlapped peaks, thus aiding the search for detecting and identifying biomarkers that may not be currently accessible.

Chapter 2

MEASURING POROSITIES OF CHROMATOGRAPHIC COLUMNS UTILIZING MASS-BASED MEASUREMENTS AND THE TOTAL PORE BLOCKING TECHNIQUE

2.1 Introduction

The experimental determination of porosity, or proportion of pore volume within a column, provides a means of predicting and understanding column performance. While methodology exists currently to estimate the different volumes present within a chromatographic column, a certain advantage can be gained by relying on mass measurements rather than elution measurements. A better understanding of porous materials, and how they can aid more specific and higher-resolution separations, may then be utilized in both development and real-world analysis. In pursuit of the first goal of this thesis, a mass-based method for determining the interstitial and intraparticle porosities of a packed column, it is first necessary to define the inner contents of a column conceptually.

The measurement of the total pore volume, V_0, has been discussed in numerous publications and reviews[9–11]. The total pore volume, often called the "void volume," is the volume that is accessible in a liquid chromatographic column to low molecular weight solvents and solutes. This excludes solids and "inaccessible" pores which may be in the packing material. Obtaining values of the total pore volume is extremely useful in liquid chromatography when new column materials development, packing studies and column performance monitoring are to be performed

quantitatively. In this regard, the total porosity, ε_T, is calculated by dividing the pore volume V_0 by the volume of an empty column, V_T.

In addition to the total volume and total porosity, the interstitial volume, V_i, and interstitial porosity are of great importance in all forms of chromatography. The interstitial volume and porosity have been determined traditionally by a number of methods, for example by using excluded polymers which are assumed to not enter the pores of the particle like blue dextran. In addition, a more complete pore size distribution may be obtained by inverse size exclusion chromatography or ISEC[12,19–21], and by subtractive methods such as taking the total pore volume, obtained by measuring an unretained chromatographic peak elution time, and subtracting out the intraparticle pore volume, obtained by various methods such as nitrogen adsorption; this requires knowing the amount of packing material in the column.

2.1.1 Current Methods

Two types of measurements are usually associated with the determination of both V_0 and V_i. These are dynamic methods, usually based on some form of elution measurement, and static methods, usually based on pycnometric (density) and/or mass measurements. The static measurement has a long history[8,13,22–24] and was one of the first methods used to probe the porosity of packed columns. In early studies it was determined that the dynamic methods gave results that differed from the static measurements and a good deal of effort was made to explain these differences[9–11,13]. Nonetheless, both types of measurements continue to be utilized.

One of the most interesting methods for excluding solutes from pores is to use the Donnan exclusion principle[11,25–29], whereby electrostatic repulsion of charged solutes from pores which possess some charge can be utilized to measure the

interstitial pore volume. This exclusion process requires that the pore chemistry be suitable for this type of analysis and that the solvent conditions be suitable for promoting charge-charge interactions between test solutes and pores. This dynamic method works well when the conditions support the electrostatic mechanism but appears to fail under non-suitable conditions, apparently with wide-pore materials[29].

A recent addition to the dynamic methods are the so-called "Total Pore Blocking" (TPB) methods. These techniques utilize a fluid which is held in the pores to block a solute from entering and thus enable the measurement of the interstitial volume[30–33]. This technology has been applied to both hydrophobic materials (i.e. reversed-phase materials)[30,31] and to hydrophilic materials (i.e. "normal phase" materials)[32]. Additional applications using the TPB method include determining interstitial porosity when studying the pressure drop characteristics of reversed-phase (RPLC) materials[33] and studying the differences between hydrophilic interaction chromatography (HILIC) and RPLC with the same packing[34]. The TPB methods can also be used as aids in studying mass transport effects[35,36] when it is desired to run experiments which probe both the interstitial transport and pore transport mechanisms of zone broadening by shutting off the particle porosity *in situ*. Note that all of the pore blocking studies mentioned here use the static pycnometric method to determine the total column porosity. A review of previous work using the TPB method is shown here in Table 1.

Table 1: Summary of TPB methods in the literature. The tracers and pore blockers listed are those tested and do not necessarily correspond to pairs.

Author	Column	Particle	Solvent	Tracers	Pore Blockers
Cabooter et. al.[30]	Thermo Hypersil	C_{18}, 175 Å	Hydrophilic buffer	Uracil	Cyclohexane
				Thiourea	Octane
				KI	Decane
				$NaNO_3$	
Liekens et. al.[31]	Hypersil Gold	C_{18}, 175 Å, 5µm	Hydrophilic buffer	KI	C_7
					C_8
					C_9
					C_{10}
					C_{12}
	Hypersil Gold	C_{18},175 Å, 19µm	Hydrophilic buffer	KI	C_{10}
	Zorbax	C_{18}, 80 Å, 5µm	Hydrophilic buffer	KI	C_{10}
	X Bridge	C_{18}, 300 Å, 5µm	Hydrophilic buffer	KI	C_{10}
Liekens et. al.[32]	Zorbax Rx-SIL	Silica, 80 Å, 5µm	C_{10}	Addtl [30]	Hydrophilic buffer
	Sunfire Prep	Silica, 94 Å, 5µm	C_{12}		Water
Cabooter et al.[33]	Hypersil Gold	C_{18}, 175Å, 1.9µm	Hydrophilic buffer	KI	C_{10}
	Acquity BEH	C_{18},135Å, 1.7 µm	Hydrophilic buffer	KI	C_{10}
	Zorbax	C_{18}, 80 Å, 1.8 µm	Hydrophilic buffer	KI	C_{10}
Song et al.[34]	Zorbax Eclipse Plus	C_{18}, 95 Å, 5µm	Hydrophilic buffer	KI	C_{10}
		(stripped particle)	C_{10}	Addtl [30]	Hydrophilic buffer
Gritti et al.[35]	Luna (Phenomenex)	C_{18}, 100 Å, 3 µm	Water	$NaNO_3$	C_9
	Jupiter (Phenomenex)	C_{18}, 320 Å, 5 µm	Water	NaNO3	C_9
Gritti et al.[36]	Gemini (Phenomenex)	C_{18}, 110 Å, 5 µm	Water	$NaNO_3$	C_9
	Sunfire (Waters)	C_{18}, 90 Å, 5 µm	Water	$NaNO_3$	C_9

2.1.2 Focus of the Work

What is needed for the TPB method is to find liquids that can be held in place in the particle pores because they are soluble with the pore surface chemistry and be cleared out in excess because they are insoluble in the solvent. Typical pairs of fluids used in these studies include a long-chain hydrocarbon and water, which are insoluble and can be used as solvent and pore blocker depending on the particle surface chemistry. One of the difficulties with pore blocking implemented with a dynamic method is that the solutes which are used as void time markers will interact with the pore blocking fluid. This pore blocking fluid will act as a stationary phase even if the solute does not enter the particle pore. This will affect the accuracy of the void measurement for V_i. Other problems exist for the dynamic method and will be discussed in detail below.

In this paper we focus on utilizing a static measurement based on simple but precise mass measurements that rely on the TPB method but do not require a transport determination of the void time through the channel. There will still be an interaction between the interstitial solvent and pore blocking fluid, but static measurements should minimize transport artefacts. In addition, we sample a wide range of materials from small pore reversed-phase materials to wide-pore materials based on a host of stationary phase chemistries. These will be compared to previous literature measurements and an assessment made with respect to ease of use, accuracy and precision. Comparison of these results with HPLC elution times are made and differences noted. A critical assessment of the generality of the TPB technology is made with respect to the phase chemistry and pore sizes used in this study and

contrasted with a theory of pore blocking stability based on the Young and Laplace equation.

2.2 Experimental

For measuring the total pore volume in the C_4 and C_{18} column experiments, methanol was from Honeywell (Morris Plains, NJ), methylene chloride was from MilliporeSigma, and tetrahydrofuran with added stabilizer was from J.T. Baker, a part of Avantor (Center Valley, PA). Isopropyl alcohol (2-propanol) was used as the flushing solvent and octane served as the pore blocker, both purchased from MilliporeSigma. Laboratory deionized water was used as an interstitial solvent along with methanol from Honeywell (Morris Plains, NJ) and propanol (1-propanol) from MilliporeSigma.

Total pore volume for the normal phase columns was obtained using the same solvents as were used for the reversed phase columns. Isopropyl alcohol from MilliporeSigma was again used as the flushing solvent before pore blocking in these columns. Laboratory deionized water served as the pore blocker for the normal phase columns. Octane, methylene chloride, and cyclohexane were used as interstitial solvents, and came from MilliporeSigma. Supplemental methylene chloride was obtained from J.T. Baker. Ethyl acetate was also used as an interstitial solvent and was purchased from MilliporeSigma.

Uracil, naphthalene and acetonitrile (HPLC grade) were obtained from MilliporeSigma. Carboxylate-modified polystyrene latex microspheres, 0.1 µm, were obtained from Bangs Laboratory (Fishers, IN) and used as exclusion markers. This particle size is clearly a compromise intended to minimize hydrodynamic effects with small size, yet large enough to be excluded by most of the pores. Potassium Chloride

was obtained from MilliporeSigma. The columns used in this study were all from Advanced Materials Technology and are listed in Table 2. These columns are all packed with superficially porous particles. Particles were packed into standard column hardware using a proprietary column packing process.

Table 2: Column specifications where the abbreviations ID, HILIC are inner diameter and hydrophilic interaction liquid chromatography, respectively.

Column number	ID (mm)	Length (mm)	Bonded phase	Pore size (Å)
1	2.1	50	C_{18}	90
2	2.1	50	C_{18}	160
3	2.1	50	C_{18}	400
4	2.1	50	HILIC	90
5	4.6	50	HILIC	160
6	4.6	50	HILIC	1000
7	2.1	50	C_4	1000
8	2.1	50	C_4	1000

Chromatographic void times were determined using a Shimadzu Nexera HPLC instrument (Columbia, MD). The mobile phase was 50/50 water/acetonitrile with 50 mM KCl and contained a 180 µL incorporated mixer. Total extra column volume included connections between the injector and column (6.3 µL), column and detector (6.7 µL), and detector cell (1 µL) for a total of 14 µL. This additional dwell volume was subtracted from the chromatographically determined column volumes and porosities. Uracil and naphthalene were used as dead time markers for hydrophobic and hydrophilic phases respectively and detected at 254 nm. Modified microspheres

were used to determine interstitial porosity and were detected at 210 nm. Data was analyzed using integrated LabSolutions software (Shimadzu).

All mass-based measurements were performed using a 5 -digit balance (Sartorius model CP225D, Bohemia, NY). An HPLC pump (Shimadzu Scientific Instruments, model LC-30AD, Columbia, Maryland) was used for all filling and blocking experiments.

2.2.1 Total Pore Volume Experiments

To establish the procedure of obtaining the total pore volume within a chromatographic column, we must first establish an analysis of the contents of a column, and then distinguish which variables will be determined experimentally. The ultimate goal in these analyses is to obtain numerical values for total pore volume, V_0, the particle pore volume, V_p, and the interstitial volume, V_i, described below. These will then be converted to porosities.

Let V_T denote the total internal volume of the column:

$$V_T = \pi r^2 l \qquad 2.1$$

where l is the length of column and r is the internal radius. We can break down V_T into two components:

$$V_T = V_S + V_0 \qquad 2.2$$

where V_S is the volume of solid material in the column and V_0 represents the total pore volume of the column which is the portion of the column that is not silica, bonded phase, or inaccessible pores. In this way, the total pore volume can be viewed as the empty space in the column that can be occupied by solvent.

The total pore volume within a chromatographic column can be further broken down into two constituents: the interstitial space, or space between the packed

particles, and the intraparticle space due to the pores in the particle. There is no clear delineation surface between the interstitial and intraparticle space, and this will be discussed below. In these terms,

$$V_0 = V_i + V_p \qquad 2.3$$

where V_i is the interstitial volume of the column and V_p is the intraparticle or pore volume.

Once V_0 is known and V_T is calculated from the column specifications, V_S (volume of solid material in the column) can be found by subtraction in Equation 2.2. The pore blocking experiments, detailed in the next section, will seek to determine V_i, and subsequently V_p by subtraction in Equation 2.3.

To determine V_0 experimentally, the chromatographic column was wetted with various solvents to provide a basis of comparison for their total masses. Because of the possibility of pore exclusion, the solvents used were chosen based on low molecular weights and their ability to fully wet all of the pores of the column. A variety of solvents were used in these experiments with the goal of obtaining the highest accuracy of an average value for V_0.

The following equation, adapted from Alhedai, Martire, and Scott[8], can be used:

$$V_0 = \frac{m_1 - m_2}{\rho_1 - \rho_2} \qquad 2.4$$

where m_1 is the mass of the column wetted with solvent 1, m_2 is the mass of the column wetted with solvent 2, ρ_1 is the density of solvent 1 and ρ_2 is the density of solvent 2.

From Equation 2.3 to find V_0, we note two areas of potential uncertainty: mass and density measurements/values. Since the densities were obtained from the solvent

manufacturers, we may neglect the contributions of the density values to the uncertainty of the calculations.

It is also important to note the choice of solvents 1 and 2 used in calculating V_0. Due to the large value of error associated with small differences in solvent mass, solvent pairs for calculation of V_0 from Equation 2.4 were chosen to maximize the difference in density between the two. In fact, theoretical calculations made using Equation 2.4 determined that for two solvents with a density difference of about 0.01 g/cm^3, as in the case of methanol and acetonitrile, the uncertainty associated with the calculation of V_0 is greater than V_0 itself, and therefore not a significant result.

2.2.2 Pore Blocking Experiments

The next step is to find what proportion of the total pore volume in the column, V_0, is due to the interstitial space, V_i, and the pore volume of the particles, V_p. Adapting the procedure for TPB used by Cabooter et al.[30] with the condition that this applies for an RPLC column, steps are as follows:

1. Rinse the column with a solvent to dissolve both the hydrophilic and hydrophobic liquids it contains (Cabooter et. al.[30] used isopropanol at 0.2 mL/min for 60 min.).

2. Fill the column pores with a hydrophobic solvent to replace the isopropanol (or other initial solvent) – the hydrophobicity of this solvent will allow it to pass through and be retained in the intraparticle pores because of its affinity for the C_4 or C_{18} layer covering their walls.

3. Flush the column with a hydrophilic buffer to fill the interstitial space without displacing the hydrophobic solvent contained in the pores (the two substances are immiscible).

4. Record the mass of the column and repeat the procedure for various hydrophilic buffers.

Similarly to the method for calculating V_0, V_i, the volume of the interstitial space in the column, can be obtained by:

$$V_{\mathrm{i}} = \frac{m_1 - m_2}{\rho_1 - \rho_2} \qquad 2.5$$

Data were collected for several hydrophilic buffers to compute an average value for V_i of each column. The same procedure is applied in reverse for a hydrophilic stationary phase and pore blocking solvent and a hydrophobic fluid to flush excess pore blocking fluid. The particle pore volume was determined by manipulating Equation 2.3 to yield an explicit expression for V_p:

$$V_P = V_0 - V_i \qquad 2.6$$

The total porosity, ε_T, interstitial porosity, ε_i, and the intraparticle porosity, ε_p, are calculated from the total pore volume, interstitial pore volume and intraparticle pore volume divided by the empty column volume respectively. Note that there are two interparticle porosity systems used in practice[37]. The system most often used in chromatography has the total porosity, ε_T, as the sum of the interstitial porosity, ε_i, and the intraparticle or particle porosity, ε_p. In some of the papers referenced here, the interparticle porosity is expressed as those use by chemical engineers, ε_p^{CE}, and is computed using the relationship $\varepsilon_p^{CE}=(\varepsilon_T-\varepsilon_i)/(1-\varepsilon_i)$. These particle porosities can be converted back to the additive intraparticle porosity used in chromatography through the relationship $\varepsilon_p=(1-\varepsilon_i)\,\varepsilon_p^{CE}$.

2.3 Results

The raw data for the total pore volume mass measurements are given in Table A1 in Appendix A. Table A2 gives the mass measurements found during the pore blocking experiments. In addition, Table A3, also in Appendix A, gives the total pore

volume and the interstitial volumes. These data are then combined to give the porosity results shown in Table 3.

Table 3: Total, interstitial and particle porosity measurements from pore blocking experiments. The interstitial porosities obtained by elution of the 0.1 μm particle are given in red, except for column 6 which was not performed.

Column	Type	Total Porosity (ε_T)	Interstitial Porosity (ε_i)		Particle Porosity (ε_p)
1	2.1x50 90A C18	0.522 ± 0.022	0.220 ± 0.127	0.39	0.301 ± 0.129
2	2.1x50 160A C18	0.598 ± 0.005	0.387 ± 0.014	0.35	0.210 ± 0.014
3	2.1x50 400A C18	0.542 ± 0.004	0.400 ± 0.009	0.43	0.142 ± 0.010
4	2.1x50 90A HILIC	0.628 ± 0.003	0.597 ± 0.007	0.42	0.0317 ± 0.0074
5	4.6x50 160A HILIC	0.652 ± 0.006	0.604 ± 0.007	0.37	0.0481 ± 0.0072
6	4.6x50 1000A HILIC	0.594 ± 0.004	0.589 ± 0.013	—	0.0048 ± 0.0140
7	2.1x50 1000A C4	0.577 ± 0.056	0.257 ± 0.016	0.40	0.319 ± 0.058
8	2.1x50 1000A C4	0.598 ± 0.027	0.337 ± 0.027	0.40	0.261 ± 0.038

Note: standard deviations for the total and interstitial porosities in this table were calculated from duplicate or triplicate measurements.

These results indicate a number of important points. First, the total porosity numbers are very reproducible in each set of experiments; these are shown in Appendix B in Table B1 and are performed in duplicate or triplicate. The interstitial porosities, which are performed as TPB measurements, show larger variations in results per each set of experiments.

The numbers in Table 3 for columns 1 and 4, which are 90 Å SPPs with C_{18} and HILIC surfaces, can be contrasted with previous studies[38–43] where columns of these particles were also measured. In the case of the C_{18} material for this particle, the total porosity averaged is in the range 0.498[40] to 0.540[38] with the exception of 0.45[43]. This is in contrast to the average of 0.522 obtained via the mass-based measurement, albeit they are all not the same columns. The average interstitial porosity of the C_{18} material in the literature[38–44] is in the range of 0.38[43] to 0.432 as compared with the 0.220 average determined in this work. However, considering the variability in the interstitial measurement, this would be comparable if one standard deviation is added to the interstitial value – i.e. 0.39, which is within the range of values reported by others[37–42,44]. The interstitial porosity obtained by the elution measurement of the 0.1 μm particle is exactly this value and provides a consistency check between the elution technique and the mass-based TPB method.

For column 4, which contains the same basic particle as column 1 without the bonded phase, the average total porosity from mass-based measurements is 0.628 as compared to 0.614 from other literature[39]. These results are in quite good agreement. In addition, the interstitial porosity of the HILIC (unbonded silica) material is 0.597 from mass-based measurements, in agreement with 0.614[39]. It is surprising that the presence of a stationary phase makes a difference in both the total and the interstitial

porosities between the bonded and unbonded particle columns. However, these columns are packed with a different solvent and the packing protocol is different, so this may explain the difference between these values. For this column the exclusion marker porosity is low and this may possibly be due to charge-charge repulsion of the exclusion marker and porous particle. If the exclusion marker cannot sample all of the available interstitial space, the exclusion volume and interstitial porosity would be underestimated.

Two other columns listed in Tables 2 and 3 have C_{18} bonded phases; these are columns 2 and 3 with average pore diameters of 160 Å and 400 Å respectively. Again, the total porosities from two separate measurements give reasonably good total porosity self-agreement, as judged by their standard deviations, and appear to be reasonable values. The literature value for the total porosity of the 160 Å particle column is 0.563[39] compared with the value of 0.598 from Table 3. The interstitial porosity is 0.402[39] compared to 0.387, the TPB method result from Table 3 – a difference of ≈3.87%. The exclusion marker for this column gives an interstitial porosity of 0.35, a value deviating from the mass-based TPB method by ≈10%.

The 160 Å HILIC column (column 5) result has a higher total porosity than the bonded result, consistent with the 90 Å bonded and unbonded column values. However, the interstitial porosity for the unbonded column appears to be unreasonably high, noting there are no other literature values for this column. The exclusion marker for this column is 0.37, probably closer to the actual interstitial porosity, but this is unfortunately hard to ascertain.

In the case of the 400 Å C_{18} column, both the total porosity and interstitial porosity (0.542 and 0.400) appear reasonable because this particle has a very thin shell

and one would expect that the intraparticle porosity would be particularly small. Unfortunately there are no literature values to compare these to, however, the exclusion marker gives a value of 0.43, a difference of ≈ 2.2% with the TPB mass-based measurement showing consistency for these results.

In the case of column 6, the 1000 Å silica (HILIC) particle, the total porosity appears reasonable but the intraparticle porosity is very low. Although this particle type can be compared with the two C_4 1000 Å particle columns which have very similar total porosities, the 1000 Å HILIC column has almost no pore volume. This is surprising, but may be explained by the sensitivity of pressure and flow velocity when clearing out the excess pore blocking solvent, as explained in section 2.3.1 below.

The columns listed as 7 and 8 in Table 3 are two columns from the same lot. As can be seen, the total porosities are close, ≈3.6% different. However the interstitial values and intraparticle porosities are clearly different, and this difference may also be explained by the sensitivity of large pore materials to the clearing of excess pore blocking solvent as discussed below. Interestingly, the exclusion marker consistently gives an interstitial porosity of 0.40 between the two columns which are packed with particles from the same lot.

In contrast to using mass-based measurements, we compare the results of using void measurements with HPLC to the porosities determined in Table 3. For two columns, the total porosities were determined by injection of 1 μL of uracil and naphthalene with 50 mM KCl added to the mobile phase to avoid any Donnan exclusion effects with the charged uracil solute. The flow rate was varied in these cases from 0.05 mL/min to 0.5 mL/min. Two cases were examined in some detail, column 1, the 90 Å C_{18} column, and column 7, the 1000 Å C_4 column. The results of

these experiments are shown in Figures 1 and 2 respectively, plotted as total porosity obtained by measuring the void time and converting it to total porosity by multiplying by the flow rate and dividing by the column volume.

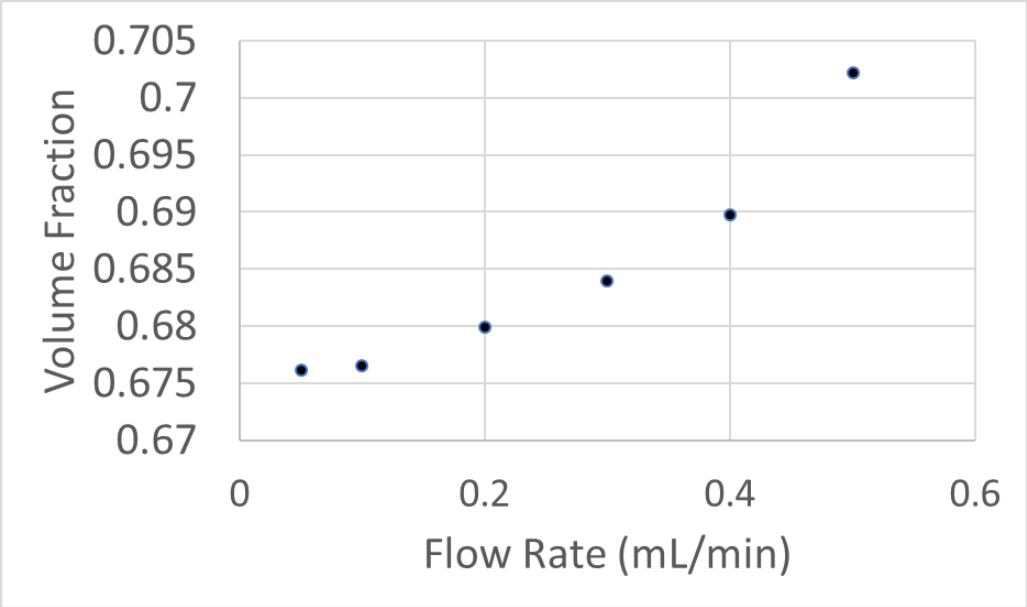

Figure 1: The total porosity of unretained zones of uracil and naphthalene for column 1. Conditions as given in the text.

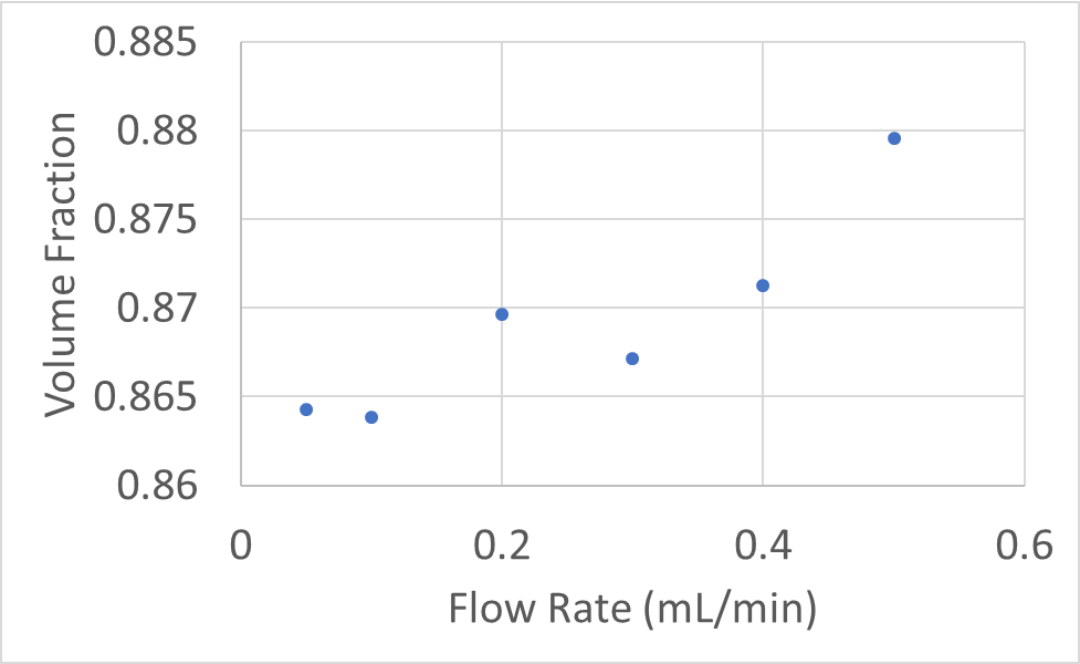

Figure 2: The total porosity of unretained zones of uracil and naphthalene for column 7. Conditions as given in the text.

The dependence on flow rate is interesting as it suggests that there are diffusion limitations and/or residual charge effects for these two columns. Although these are small effects, they are present nonetheless and account for approximately 3% variation in the 90 Å pore diameter particle and the 1000 Å pore diameter particle. Nonetheless, the total porosity for the 90 Å pore diameter particle was determined in Table 3 by mass as 0.522 and by elution as 0.675, a surprisingly large difference of ≈ 20%. The precision of the elution measurements are typically a few percent, so this is not the source of the discrepancy. For the 1000 Å pore diameter particle, the total porosity was measured by mass as 0.577 and the elution experiment gave ≈ 0.864, approximately 33% in disagreement. The physical mechanism to explain this discrepancy is discussed further in the discussion section.

2.3.1 Mechanism

The fluid mechanics of the pore blocking method closely resembles the displacement and replacement of a wetting fluid by a non-wetting fluid in porous media[45–49]. Applications related to fluid replacement in a porous medium include soil remediation[50] and oil recovery with pumped fluids[51,52]. The petrochemical applications of this displacement process are obviously of great importance. A closely related problem of chromatographic significance is the loss of retention when pure water solvent is utilized with a reversed-phase column after depressurization[53].

A nonwetting fluid can invade a pore filled with a wetting fluid when the driving pressure exceeds the capillary pressure p_c, also called the "threshold pressure"[45–49,54,55]:

$$p_c = p_{nw} - p_w \qquad 2.7$$

where p_{nw} and p_w are the pressures needed to drive the non-wetting and wetting fluids into a pore, respectively. In the petroleum literature, it is customary to define an oil-wet pore with a negative capillary pressure[48]. The capillary pressure is a measure of the porous media's ability to fill with the wetting fluid or to expel the non-wetting phase[45]. These two fluids are considered to be immiscible in this treatment.

The capillary pressure may be expressed using the Young and Laplace equation[45,47,48,52,56,57]:

$$p_c = \gamma\,(1/r_1 + 1/r_2) \qquad 2.8$$

where γ is the interfacial tension between wetting and non-wetting fluids and r_1 and r_2 are principle radii of curvature of the interface. Replacing the radii of curvature in Equation 2.8 with a mean radius of curvature, r_c,[48] and assuming a spherical interface between two fluids with a finite contact angle gives the usual form of the capillary pressure:

$$p_c = 2\,\gamma \cos\theta\, r_c^{-1} \qquad 2.9$$

where θ is the contact angle between immiscible fluids. Different forms of Equation 2.9 are available for different geometries, although these forms are more complicated. A more accurate geometry for this problem is that of intersecting and neighboring spheres[58], however this would greatly complicate the simple form of Equation 2.9.

The contact angle for a wetting fluid is < 90°, and is 0° for a fully wetting fluid[59]. Spontaneous pore filling occurs with wetting fluids unless a non-wetting fluid is held in the pores above the capillary pressure for the non-wetting fluid. Notice that cos θ is positive for $\theta < 90°$. The pressure with which the pore blocking fluid will be displaced by the (immiscible) solvent is the capillary pressure. For a non-wetting fluid, the contact angle at the pore surface is typically > 90°, with extreme values being $\theta \approx 120°$ for water contacting a fluorinated surface[60]. In this case cos θ is negative. Note that p_c can be positive or negative depending on the curvature of the interface.

The two pertinent cases of interest here include the forced filling of a pure silica pore by hydrophobic alkane solvents for HILIC phases where water is used as the pore blocker and the wetting of alkyl chain bonded-phases with pure water when a hydrophobic species is used as the pore blocking fluid. The interfacial tension of an octane/water interface is reported as 51.16 mN m^{-1} [61] at 25.0 °C and 52.5 mN m^{-1} [62] at 22.0 °C. The values for the decane/water interfacial tension are similar within a few percent[61,62]. The contact angles relevant for understanding water adsorbing on a hydrophobic surface include water on a paraffin surface of θ = 110.6° [63] and water on an octadecyl trichlorosilane-modified Si wafer giving θ = 109° [64] although a host of surfaces used in water adsorption studies[64] show a variety of contact angles with water that are highly dependent on the surface material.

For a hydrophilic pore wall like bare silica, the wetting of the pore by hydrophobic fluids like octane is a somewhat different story. This is because for a freshly prepared silica, which has been cleaned by elevated temperature and outgassing, a hydrocarbon will adsorb on the surface[65] to some small degree. However, under chromatographic conditions, silica is well known to maintain a tightly held water layer at the surface[66–69] and this will cause a non-wetting surface to develop for medium to large alkanes under normal aqueous solvent conditions.

Some examples using Equation 2.9 are shown in Figure 3 to illustrate the effect of pore diameter on the capillary pressure needed to replace the pore-wetting blocking fluid with the solvent for both hydrophobic and hydrophilic cases discussed above. The interfacial tension mentioned above is held constant and the contact angle varied with θ = 110° in the middle. In addition, contact angles of lower (θ = 100) and higher (θ = 120°) hydrophobicity are shown. Two other curves are also shown with higher and lower interfacial tension.

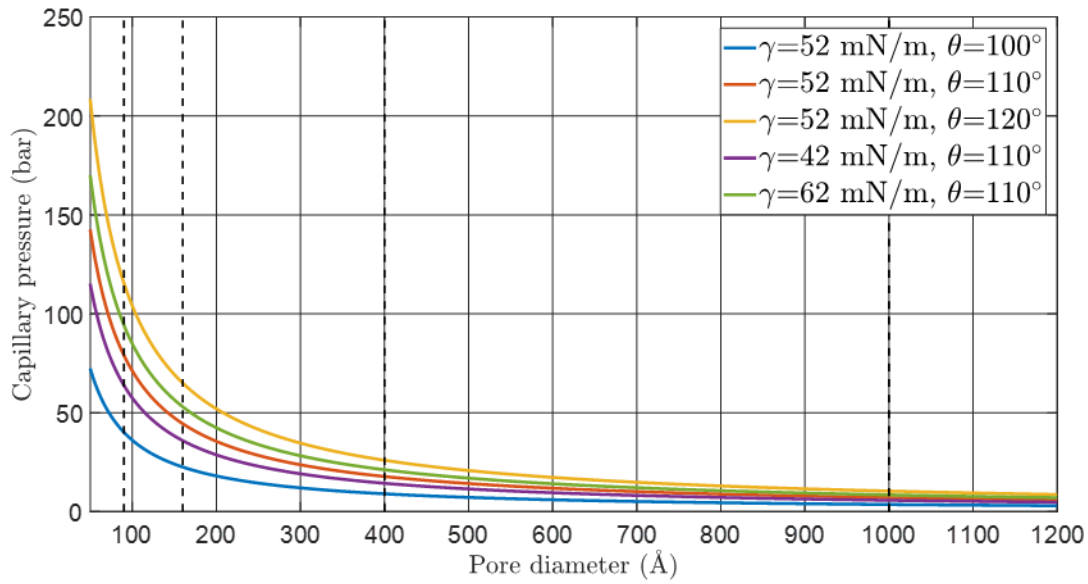

Figure 3: The effect of pore diameter on the capillary pressure needed to force the non-wetting fluid into the pore. The conditions for each curve are shown in the legend. The pore diameters of 90 Å, 160 Å, 400 Å and 1000 Å are delineated.

This plot shows that substantially higher pressure needs to be applied to displace the pore blocking fluid with 90 Å particles than with larger pore size particles. This suggests that the TPB method is more reliable for smaller pore materials; if larger pressures are utilized, the pore-blocking fluid may be displaced.

An important and practical aspect of this plot is that the velocity of the clearing process, where a non-wetting fluid removes the excess pore blocking solvent, is critical because it determines the capillary pressure in the column given in Equation 2.9. Too high a velocity and the pore blocker will be displaced, thereby washing it out. The pressure is related to flow velocity (and flow rate) using Darcy's law[45,48] for single-phase fluid transport:

$$\bar{v} = -(k/\mu)\,\nabla p \qquad 2.10$$

where k is the column permeability, μ is the viscosity, $\bar{v}$ is the average velocity in the packed bed and ∇p is the pressure gradient driving force for fluid flow in the column. This equation shows that the fluid velocity and pressure gradient are linearly related so that higher velocity will require higher pressure. Conversely, Equation 2.10 can be rearranged to express the pressure as a function of the velocity. Hence, the pressure at the column head, which is measured by the HPLC pump and associated transducer, is adjustable through the flow rate and the average velocity. Thus, it should be possible with these equations to optimize the minimum velocity necessary to ensure the capillary pressure is not exceeded, although pressure is easily monitored and may be directly viewed so that it is not exceeded as flow rates are adjusted. Figure 3 also shows the critical nature of using larger pore materials with the TPB method. It doesn't take much pressure with large pore materials to displace the pore blocking fluid. Hence, these must be run with the lowest pressures possible and consequently very small flow rates (and velocities). This will dictate the time necessary to clear the excess pore blocking solvent from a large pore particle; this can be very long as the pressure must be exceedingly small or else the method will fail due to disturbing the pore blocking fluid in the pore.

Further insight into the TPB method can be obtained through calculation of the nondimensional capillary number, C, which is the ratio of viscous to capillary force[48,54,56,57]:

$$C = \mu\, v_p\, \gamma^{-1} \qquad 2.11$$

where μ is the viscosity and v_p is the velocity of a fluid in a capillary pore. For C values approximately less than 1, the fluid movement is dominated by capillary forces, and for C approximately greater than 1, fluid movement is due to viscous forces. In a

recent study of the computational fluid mechanics of SPPs with 1000 Å pores[58], the mean pore velocity v_p is ≈ 0.01 $\bar{\boldsymbol{v}}$. Assuming a viscosity of water of 8.90 x 10^{-4} Pa. s, the surface tension of the octane /water interface is 52 mN m^{-1} and $\bar{\boldsymbol{v}}$ is 10^{-3} meters per second (v is then 10^{-5} m s^{-1}), such that C = 1.71 x 10^{-6}. This shows that for the largest pore material, capillary force dominates the flow of fluid. The capillary number will be even smaller for smaller pore materials, showing for the TPB method, capillary forces will dominate over viscous forces for most all chromatographic pore diameters of interest.

These equations are simplistic for a number of reasons now discussed but overall show the magnitudes of numbers used in the TPB process. First, when a less viscous fluid displaces a more viscous fluid, such as when water displaces octane (the viscosity of octane is 5.195 x 10^{-5} Pa.s[70]) the liquid-liquid interfacial region becomes unstable and produces viscous fingering[71,72]. In addition, "Haines jumps"[48,49] occur due to kinetic effects where droplets are ejected in almost random collections as a result of interfacial instability and nonuniform pore effects. Furthermore, "snap off" effects[48] occur when a nonwetting fluid loses contact with the wall. These effects occur due to flow; using a mass-based approach to interstitial pore volume measurement can minimize these extra effects, but not eliminate them, since the pore filling occurs under flow. Hysteresis effects[45–49] are also known to occur where the repeated cycling between pore wetting and dewetting gives changing results. Hysteresis is thought to occur due to surface contamination, surface roughness and surface liquid immobility[48,56].

2.4 Discussion

The use of TPB techniques for porosity measurements is wrought with difficulties but can return potentially valuable information if a number of variables are recognized and controlled. The results of the mechanism study show that wide pore materials will be the most difficult because even a small pressure will purge the pore blocker out of the particle with the displacement fluid. The displacement process of the excess pore blocking fluid is the critical step in all of the samples that were run here. For pores in excess of 400 Å, the fluid pressure will probably be too high to leave the pore blocking fluid intact when clearing out the residual amount of pore blocking fluid. One way to access this is to utilize an extremely small flow rate (and accompanying pressure) for samples of this nature. Another way to do this is to monitor the effluent for any signs of the pore blocking fluid. In other TPB studies utilizing HPLC-based measurements, the detector signal indicated when the flushing step was complete.

The HPLC data show systematic variation, and it is well-known that this method is not a reliable method for determining total porosities. However, it is interesting to see that a huge overestimation of the wide pore material studied here is probably due to flow through the particle[58] which has been calculated from models and would explain how the elution measurement would so overestimate the total pore volume and total porosity. Other problems with determining total, interstitial and intraparticle porosity for large pore particles are unique: the dividing surface between what is a pore and what is interstitial space is not defined for any porous particle system. This is discussed in the modeling paper referred to previously[58]. The physical manifestation of a dividing surface in the TPB method is the meniscus at the interface between immiscible fluids, and the geometry of this is highly dependent on contact

angle and the pore shape. This is an interesting area of research as characterization of very small pore and very wide pore materials is a challenging area that still offers room for innovation. There is room for innovation with the TPB method, but its limitations become clear with the physical insight discussed above on the mechanism of pore filling with respect to pressure and pore size.

The exclusion marker results agree in most cases with the interstitial porosities determined by the mass-based TPB method. This appears to be a good consistency check with the TPB method in general. However, the results for the widest pore particle columns may be questionable because the pore size distribution is wide enough for these particles that some penetration of the exclusion particle is possible.

The ISEC method, although time consuming and not free of problems, seems to be the most readily adopted to making accurate measurements of porosity. However, its use, although amenable to automation with an autosampler, may still have problems with wide pore materials because finding excluded materials may be difficult if not impossible; many high molecular weight solutes have non-ideal behavior in the vicinity of particles like hydrodynamic and slalom chromatographic modes, and these can complicate the ISEC data interpretation.

Chapter 3

A STATISTICAL OVERLAP MODEL OF CHROMATOGRAPHY WITH FINITE DETECTION LIMIT: REQUIREMENTS FOR BIOMARKER DETECTION

3.1 Introduction

A clearer understanding of the capacities of the proposed mass-based method of determining porosity provides insight into the physical potential of a chromatographic column; the search for biomarkers, however, also necessitates an investigation of the resulting chromatographic data. Biomarkers, especially those substances present at extremely low concentrations in bodily fluids, have the potential to be used universally as a medical diagnostic tool to indicate trauma and disease in the body before the first onset of symptoms. The potentials of biomarker detection require a thorough analysis of the resolution of chromatographic signals, and an understanding of how, and to what extent, individual peaks may be lost in a chromatogram due to noise and finite detection limits. The quantification of these losses using theoretical methods has the potential to frame the practical considerations of missing analytes and provide further insight into the nature of experimental chromatography. The application of statistical overlap theory to simulated chromatograms defines the second goal of this thesis, with the target of quantifying peak loss in the context of biomarkers present below traditional detection limits.

The search for biomarkers indicative of diseases is one of the most intensively investigated aspects of modern biomedical research. For example, cardiovascular

disease biomarkers[73,74] are sought that can distinguish healthy individuals and those with early developing disease. Cancer biomarkers derived from the plasma proteome[75,76] can be very diverse and may be found in exceedingly smaller concentration, specifically in the picogram per milliliter basis[76]. However, as is well-known, the concentration range of proteins often exceeds 11 orders of magnitude[76] and this poses great problems in analytical detection methods when the analysis scheme is examined in detail. Many of these molecules may be present in less than a picogram per milliliter, i.e. parts per trillion (ppt), and these can offer exceeding difficulty in analysis. Furthermore, dynamic range limitations inherent in the detection process affect the useful concentration range that can be studied[77] although high concentration proteins can be removed by affinity chromatographic methods.

This large dynamic range poses a great problem in identifying cancer biomarkers in the proteome, especially when analyzing intact proteins using the so-called "top-down" methods[78,79]. These methods employ very high resolution detection by mass spectrometry (MS), often with tandem mass spectrometry stages (i.e. MS/MS) of detection.

Although these methods can be very powerful, the selectivity of liquid chromatography (LC) is insufficient to separate these species due to limited peak capacity; the possibility of tens of thousands of compounds pose a nearly-impossible separation task for both 1D and 2D chromatography. Even with a four-dimensional separation[79], the human proteome has far too many components to be utilized directly for biomarker detection, let alone at the concentration levels that are thought to be present. Many options exist to supplement the separation selectivity, such as using

affinity chromatography for retaining very specific compounds for further analysis or as a rejection of specific compounds or compound types.

In the “bottom-up” approach to proteomics, proteins are digested so that specific peptide fragments can be selectively detected. This also eases the difficulties of limited peak capacity in LC column methodology, although this problem is still extremely complex. One of the most selective approaches appears to use the analysis of the carbohydrate component of the glycoproteins, which appear to offer a number of possibilities for cancer detection[80,81]. This is due to specific cellular processes which are often associated with the carbohydrate part of the protein. Hence, glycomic analysis of proteins is becoming a very active part of the biomarker discovery process[16,80–82] and mass spectrometry detection methods have been driven by the requirements needed for selective and sensitive detection of these glycans. Towards this end a great deal of study has gone into understanding the noise characteristics of high performance mass spectrometers[83–88] used in biomarker detection.

As is the case for top-down (intact) protein studies, bottom-up proteomics also requires unraveling of the complex mixture that defines the sample. This is most often accomplished using liquid chromatography (LC) with MS or MS/MS detection[89–92]. Both bottom-up and top-down analysis place severe demands on the separation stage. In many cases, capillary techniques[93] are utilized due to the common case of limited sample volume, less dilution for higher sensitivity and very high resolution in spite of long run times.

Resolution is extremely important as these samples have characteristically large overlapping of peaks and this can mask very low-level signals. For example, if a low concentration peak is partially overlapped with a high concentration peak, then the

low concentration peak will essentially be indistinguishable when mass spectrometry cannot resolve these on separate mass channels. This overlap also causes problems in quantitation due to ion suppression[94]. 2D chromatography has also been tried on numerous occasions[95] but will not be discussed further; this study is focused on 1D LC.

3.1.1 Statistical Overlap Theory and Natural Datasets

A number of years ago, a statistical model of chromatography was developed by Davis and Giddings[17] which examined chromatographic elution profiles and expressed the fraction of peaks resolved as a function of the ratio of the component number m to peak capacity n_c; this was deemed the saturation, α. A simple Poisson model of random retention times was utilized. Their results showed for a typical case, a random chromatogram will never resolve more than 37% of its potential peaks and only 18% of its potential peaks are single component peaks (SCPs). This theory also predicts that in order for a 90% probability of detecting an isolated SCP, the chromatogram must be 95% vacant. In theory, this would give decent detection, however in applications where thousands of components are present, such as in proteomics research, this is improbable, impractical, time consuming, and costly.

This research was continued by Davis and coworkers; these models were collectively known as "Statistical Overlap Theory" (SOT). A review of SOT has been given[96] and explored for a number of different peak spacing models using different probability density functions (pdfs), different peak amplitude pdfs, etc. In all of these previous SOT models, there is an assumption present: a single detector is utilized that can detect all components and no noise is present on this single idealized channel.

Allowance for a threshold below which detection is impossible was introduced into SOT[97] and subsequently refined[18].

Nagels and coworkers[98] have looked at sample concentrations of plant extracts and concluded that the relative peak areas are exponentially distributed. Enke and Nagels[99] have examined model sample datasets of natural materials and concluded these systems adhere to a log -normal distribution in signal intensities. Further work on this problem[100] examined the response factors of the signals in an attempt to statistically correlate concentration and signal intensity factors. The exponential distribution suggests that most sample components are found at very low concentrations whereas the log-normal distribution suggests that at low concentrations, the number of components decreases after a maximum number of components which exist near the center of the distribution. These observations may be very sample dependent, but it is impossible to describe the statistical distribution of components below the limit of detection (LOD) when one can't measure the concentration of these components. This suggests a more general approach to understanding this problem is needed.

3.1.2 Focus of the Work

In the work reported here, we impose a finite signal-to-noise ratio (SNR) into the SOT to see what signal requirements are necessary to avoid missing low concentration signals that might represent useful biomarkers. A finite LOD is imposed on the signal to be measured. Although the detector is still assumed to be a single channel detector, some statistical statements can be made about multiple detection channels common to MS and MS/MS. A number of amplitude (concentration) pdfs are investigated here including the log-normal, exponential and Weibull functions. The

latter, in particular, can be varied with two parameters so that the basis pdf changes between an exponential signal distribution and one looking almost like that of a log - normal distribution. This is necessary to see what effect the loss in detectable signals appears to have, depending on which model is chosen.

The effect of chromatographic efficiency on the detection of these low-level components is included in this SOT model via the saturation parameter. We will show that the detectability of these trace components places severe demands on chromatographic efficiency which affects peak heights and peak overlaps. When neighboring peaks have much different peak heights the smallest signal is often undetectable due to swamping out by the neighboring signal(s). Estimates of the fraction of components lost due to finite SNR and LOD as a function of column efficiency are given in detail to provide insight into the potential loss of components relevant to biomarker research.

3.2 Theory

The peak is the fundamental unit of measure here and all peaks in this treatment are Gaussian within any chromatogram. Chromatograms are formed by summing Gaussian single component peaks (SCPs). The SCP of the jth peak is defined as being composed of discrete data points of length n with index i for the peak amplitude, g_i, which is a function of the time t_i:

$$g_i(h_j, t_i, t_{R,j}, \sigma_j) = h_j \exp[-\frac{(t_i - t_{R,j})^2}{2\sigma_j^{\,2}}] \qquad 3.1$$

noting the peak maxima h_j occurs at retention time $t_{R,j}$ with standard deviation σ_j. The vector index of these quantities $j \in [1,m]$ where m is the number of SCPs with p observable peaks in a chromatogram noting that $p \leq m$ due to peak overlap. The peak

area A_j can be equated with the peak amplitude h_j for a Gaussian peak with standard deviation σ_j so that:

$$A_j = h_j \, \sigma_j \sqrt{2\pi} \qquad 3.2$$

Noiseless, pure signal chromatograms $H_S\,(t_i)$ are composed of the superposition of m peaks:

$$H_S(t_i) = \sum_{j=1}^{m} g_j(h_j, t_i, t_{R,j}, \sigma_j) \qquad 3.3$$

3.2.1 Amplitude Distributions and Concentration Model

The amplitude distributions used here are idealized systems guided by previous models of the exponential pdf[98] and a log-normal model[99,100] of signal strengths derived from experimental data of complex mixtures. To capture a more general model, the Weibull pdf[101] is utilized. A comparison of these pdfs is shown in Figure 4 and the parameters used to derive these were curve fit from the data discussed below with the exception of the Weibull 2 model. The functional form of these models is given in Table 4.

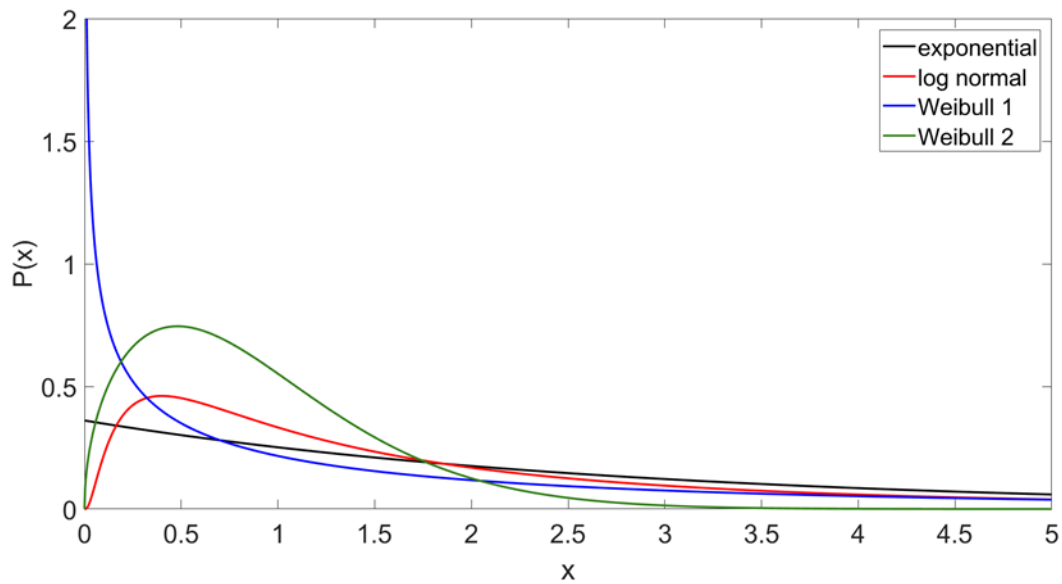

Figure 4: The 4 probability density functions (pdfs) used in this study. The independent variable is signal amplitude x and P(x) is the probability of finding that signal amplitude. The mean and variance used in the exponential, log-normal and Weibull 1 cases are from the curve fits of the natural product data described in the text with parameters mean 2.77 and standard deviation of 4.5. The parameters for the Weibull 2 curve are mean 0.9027 and standard deviation of 0.6129.

Table 4: The amplitude pdfs used in this study. ξ_u is a uniform random variable between 0 and 1.

Name	pdf form	Mean	Variance	Random deviate
Exponential	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$	$\xi_e = -\frac{1}{\lambda}\ln \xi_u$
Log-normal	$\frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	Matlab lognrnd
Weibull	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp-(x/\lambda)^k$	$\lambda\Gamma(1+1/k)$	$\lambda^2\left[\Gamma(1+\frac{2}{k})-(\Gamma(1+\frac{1}{k}))^2\right]$	Matlab wblrnd

Additional notes:

1. For the log-normal density: $\mu = \ln\left(\frac{m^2}{\sqrt{\upsilon+m^2}}\right)$ and $\sigma = \sqrt{\ln(1+\frac{\upsilon}{m^2})}$ where the mean is m and the variance is υ.
2. For the Weibull distributions k is the shape factor and λ is the scale parameter.
3. The random number generators listed as Matlab are available in the Matlab Statistics and Machine Learning toolbox.

The Weibull pdf varies in shape from exponential (k=1, equal to an exponential pdf) to an almost Gaussian-like density at higher k values. Both of these shapes are apparent in Figure 4 for the Weibull distribution. The relationships between the Weibull and log-normal distributions have been discussed[102].

In these examples, highlighted in Figure 4, two primary cases of interest emerge. Case 1 is where there is a larger probability of trace components relative to the mean. Case 2 is where a smaller probability of trace components relative to the mean exists. The exponential pdf is an example of case 1, while the log-normal distribution clearly represents a case 2 amplitude distribution. The Weibull pdf, depending on the parameters chosen, can mimic both case 1 and case 2; and this is shown for the Weibull 1 and Weibull 2 pdfs respectively in Figure 4.

The signal amplitude dictated by the vector of random peak values, h_j, can be equated to the peak maximum in concentration units. To accomplish this, it must be realized that the signal amplitude is dependent on chromatographic efficiency[103] and the instrument response factor[100]. The duality between the signal amplitude and the concentration profile of chromatographic peaks is illustrated in Figure 5, along with the noise amplitude discussed below.

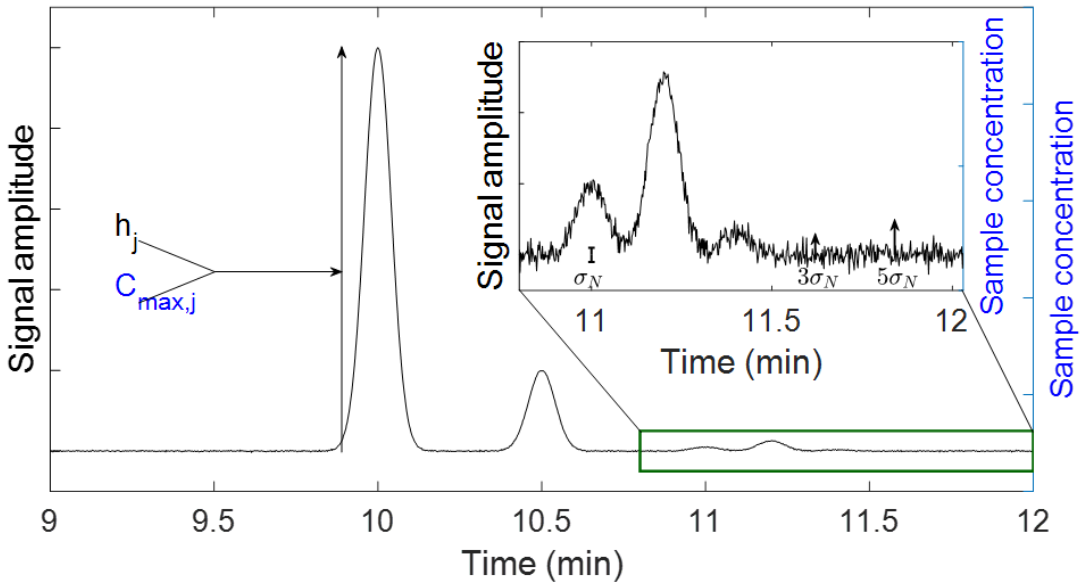

Figure 5: The relationship between the concentration and signal amplitudes illustrating the limit of detection LOD_S for signals and $C^{\neq}_{max}$ for concentration, the concentration of the jth peak at the peak maximum, $C_{max,j}$, and the signal amplitude h_j. The blue labels apply to concentration and the black labels apply to signal strength. The signal to noise ratio (SNR) is 500. Inset: The range illustrated above the noise standard deviation, σ_N, is 1 standard deviation wide. The limit of detection (LOD) is shown with both $3\sigma_N$ and $5\sigma_N$ height.

The concentration present in the detector at the peak maximum for component j, $C_{max,j}$ is[103]:

$$C_{\max,j} = \frac{m_{inj,j}}{V_{dil,j}} \qquad 3.4$$

where $m_{inj,j}$ is the mass of injected component j and $V_{dil,j}$ is the dilution volume for component j[103]. The dilution volume has been derived for single dimension[103] and multidimensional[104] separation systems. For single dimension separation systems, it takes the simple form of:

$$V_{dil,j} = \frac{V_{R,j}\sqrt{2\pi}}{\sqrt{N_j}} \qquad 3.5$$

where the retention volume is given as $V_{R,j}$ and N_j is the number of plates for component j. The number of plates is equal to $(t_j / \sigma_j)^2$, however we will use a constant σ_j model as this is much closer to the experimental results seen in gradient elution liquid chromatography. Hence, $\sigma_j = t_1 / \sqrt{N}$, where t_1 is the retention time of the first peak.

The response factor of each component j, R_j, equates the component concentration in the detector with the instrument amplitude response:

$$h_j = C_{\max,j} R_j \qquad 3.6$$

noting that for this study all R_j are assumed unknown and with units of signal amplitude per unit concentration. In one study[100] it was reported that these response factors may follow the same distribution (log-normal) as the amplitude function that was being studied. The response factor R_j includes contributions from detectors, electronics and other parts of the signal chain. Although the response factor remains unknown, in this treatment we only need to specify the distribution of peak heights h_j.

3.2.2 Signal-to-Noise Ratio and Limit of Detection

The SNR can be defined in a number of ways and these are discussed in the detector and signal processing literature[83–88,105,106]. Most differences exist in defining the signal. The first of the two most common approaches is to take an amplitude, usually a voltage, and square this voltage to make a power by assuming a certain load resistance (1 Ω) and dividing by the noise variance[104]. The other way, and the approach used in mass spectrometry-based investigations of noise[83–88] is to record a peak amplitude from a known injected compound, usually one with a large response

factor. Then the background signal is sampled and the noise standard deviation determined. The SNR is then defined as the ratio of peak signal amplitude to noise standard deviation. Deviations from this scheme are known.

In this work we use the average peak amplitude, $\bar{h}$, which is *a priori* determined when setting the amplitude distribution. The corresponding noise variance is then scaled to this peak level. Hence, the SNR is defined as the ratio of the mean peak signal amplitude to noise signal standard deviation:

$$SNR = \frac{\bar{h}}{\sigma_N} \qquad 3.7$$

Rearranging Eq 7 gives the noise standard deviation as:

$$\sigma_N = \frac{\bar{h}}{SNR} \qquad 3.8$$

As shown in Eq 8, given $\bar{h}$ and the SNR one can obtain σ_N. In the development of the threshold cutoff[18] for SOT, the symbol h_n was used as the normalized (to the mean) threshold value so that:

$$\gamma = h_n / \bar{h} \qquad 3.9$$

This threshold represents a signal buried within the noise; relative signal heights greater than γ are assumed to be detectable, but may not be resolvable if a SCP is next to a large peak that makes the smaller signal difficult or impossible to detect. This last case defines signals lost to peak overlap.

Noise can be explicitly added to the signal by producing a vector of random Poisson, Gaussian or mixed noise deviates ξ_i with mean zero and standard deviation σ_N. By adding the signal vector $H_s(t_i)$ to this noise vector the total signal is calculated:

$$H(t_i) = H_S(t_i) + \xi_i \qquad 3.10$$

The addition of uncorrelated, random Gaussian deviates is often referred to as additive white Gaussian noise (AWGN) and its power spectrum is flat. This form of noise is often used as a model for noise in electronic circuits. However, Poisson noise is more characteristic of the random arrival of ions[88] and this contribution may also be important. An illustration of how noise appears for low level signals is shown in the inset to Figure 5.

The addition of Gaussian noise to the signal impacts the resulting chromatograms; an example of chromatograms generated at varying SNR can be seen visually in Figure 6. This effect is shown for a log-normal amplitude distribution with 100 SCPs, and an expansion of the baseline highlighted to further illustrate the effect of noise on the signal.

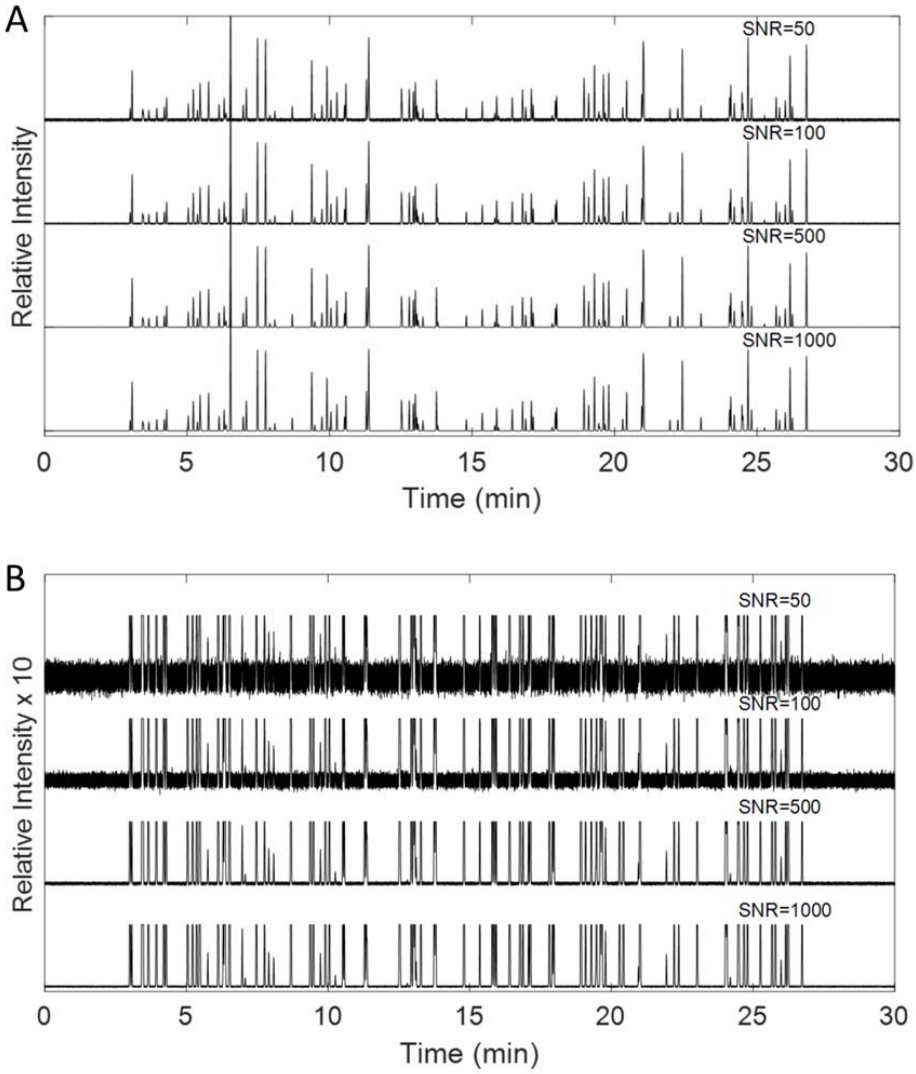

Figure 6: Four synthetic chromatograms showing the various signal to noise ratios (SNRs) when Gaussian noise is added at various levels. The conditions for these chromatograms are α=0.16, m=100 SCPs, the amplitude distribution is log-normal with mean and standard deviation equal to 0.5. The standard deviation of the Gaussian peaks is constant and equal to $\sigma_j = t_1/\sqrt{N}$ where t_1 is the retention time of the first peak noting that the efficiency of the first peak is 100,000 theoretical plates. A: Total chromatogram B: Amplitudes multiplied by time and clipped at a signal level to further show the noise amplitude in detail.

The SNR can also be specified in units of decibels (dBs) as:

$$SNR_{dB} = 10\log_{10}(SNR) \qquad 3.11$$

In this paper we don't explicitly add noise to the signal, as in Equation 3.10, but rather designate a level above which signals are considered detectable, and below which signals are assumed lost in the noise.

The LOD in signal terms[107,108] is given as:

$$LOD_S = 3\sigma_N \qquad 3.12$$

when the baseline or offset signal is essentially zero or nulled to zero. This states that the minimum detectable signal LOD_S that can be detected with a certain reliability is three times the noise standard deviation, assuming the noise has zero mean. Other treatments have used a constant of 5 instead of 3[107,108]. We will contrast results for SCPs at the LOD with the $3\sigma_N$ criteria below, realizing that given the SNR and $\bar{h}$, the LOD can be calculated.

3.2.3 Statistical Overlap Theory

Two important results from SOT are the definition of the effective saturation, α_e, and the simplest result of SOT, which is establishing the fraction of observable peaks given the saturation. The peak capacity, n_c, is commonly defined as[109,110]:

$$n_c = \frac{t_m - t_1}{4\sigma R_s} \qquad 3.13$$

where t_m is the last SCP and t_1 is the first SCP in a chromatogram. In Equation 3.13 σ is the temporal Gaussian standard deviation of all SCPs and R_s is the resolution

between two neighboring peaks; we will use $R_s = 1$ for this study. In addition, the effective saturation is given as[111–113]:

$$\alpha_e = \frac{m}{n_c R_s} \qquad 3.14$$

where m is the number of SCPs and n_c is the peak capacity. $R_s = 1$ is utilized in Equation 3.14. Of additional interest is the primary result from SOT studies which equates the saturation to the fraction of observable peaks in a chromatogram[17,113]:

$$\frac{p}{m} = e^{-\alpha_e} \qquad 3.15$$

3.2.4 Simulations

The synthesis and analysis of synthetic chromatograms is written in MATLAB (Mathworks, Natick, Massachusetts). The program makes extensive use of spreadsheets for parameter input and calculated output and provides most of the graphs calculated in this chapter. The amplitude distributions are generated with functions inherent in MATLAB, as shown in Table 4, except for the exponential distribution which is generated explicitly with the inverse transform sampling method[114,115]. In addition, the Minitab statistical software system (Minitab, Inc. State College, Pennsylvania) was used for some of the analysis of the data reported by Enke and Nagels[99].

Given the input parameters contained in the spreadsheet: peak width σ, the number of components m, the start and stop time of the chromatogram, t_1 and t_m, and the mean and standard deviation of the peak height function, the chromatograms are calculated and then analyzed. The pdf parameters of the peak height function are

calculated from the inputs of function mean and standard deviation. The chromatograms in this work are generated from t_1 = 60s to t_m = 7200s, at m = 3000 with 25 points per peak. The analysis consists of detecting peaks using the zero-crossing of the derivative signal, which is calculated by the finite difference of the signal $H_s(t_i)$. In addition, the two points used for zero-crossing determination must exceed a certain noise threshold of 10^{-6} to prevent roundoff error from triggering a false peak.

The detected peak retention times are then compared to the SCPs originally calculated from a uniformly random density of components between t_1 and t_m and sorted prior to generation. Because summation of the peaks using Equation 3.3 can slightly alter the peak maxima retention times, a range window of 1.00% is used to identify if the peak is matched with the component. All SCPs that can't be identified with a unique peak are considered to be undetectable due to summation, i.e. SCP overlap. Identifiable peaks are coded green in the map of SCPs which accompany the chromatograms and unidentifiable SCPs are coded red in this map; we refer to these diagrams as "loss diagrams." The peak capacity and saturation are stored in a separate row of the output worksheet contained within the spreadsheet. The SCPs that did not have a peak within the window were stored as time-amplitude pairs, and the pdf of these SCPs was determined using histogram analysis. This facilitates determining what quantities of signals get obliterated by peak overlap and what peaks get taken out because they are below a threshold in amplitude. The synthesized chromatograms, loss diagrams and lost component densities produced via these procedures are shown below.

3.3 Results

3.3.1 Analysis of the Natural Mixtures Datasets

Enke and Nagels[99] have proposed that the components present in a complex natural mixture follow a "natural law" in terms of their response. If this response could be modeled and fit to a known function, more information could be realized regarding the presence of materials in a mixture below traditional detection limits, predicting characteristics of that portion of the mixture that may be hidden. These authors suggest that the natural concentration law of biological and natural mixtures is a log-normal pdf. Their paper attempts to show that analytical responses of compounds in three natural mixtures: 1) extracellular metabolites, 2) light crude oil and 3) plant extracts obey the log-normal pdf. By seeking to model the responses and thus determine the log-normal parameters of the distribution, it is theoretically possible to predict the "degree of analytical selectivity and dynamic range that would be required to detect any additional fraction of the components present"[99]. Here, we will examine the data reported for two of the natural mixtures discussed: light crude oil and plant extracts, as catalogued by Nagels et al.[98]. Data for the extracellular metabolites mixture was not readily available from the publication.

Although there has been a significant focus on the log-normal distribution to describe the response of natural mixtures[18,97,99,100], a significant limitation to this approach lies in its lack of specificity. Despite significant evidence to support the log-normal fit, it remains possible that this distribution is only one of several that can be considered. An analysis of two of the natural mixtures considered by Enke and Nagels[99] in the context of the Weibull pdf, in addition to the log-normal pdf, follows.

The first data set considered was a sample of light crude oil analyzed by high resolution mass spectrometry; the compounds identified were arranged in equally-spaced groupings of response intensity, as reported by these authors in Table 2[99]. One of the quantifying assumptions made by Enke and Nagels, following analysis of a log-log parabolic plot they generated, was that the number of undetected components would be the same as the number of detected components; from the values reported, this would mean roughly 16,000 components of the mixture were not detected by the method.

In Minitab (State College, PA, USA), which was used by Enke and Nagels[99], least squares (LSXY) estimates with arbitrary censoring were employed to fit the proposed distributions (Stat → Reliability/Survival → Distribution Analysis (Arbitrary Censoring), and 16,135 missing data points were specified as per the assumption made in the work[99].

It is important to note that this is a result of the postulation that the data follow a log-normal distribution. In removing this assumption, the data were also fit without explicitly specifying the number of missing points. The results of this analysis are below.

The second data set considered was a sample of plant extract data obtained by HPLC and UV detection, initially from work by Nagels et al.[98]. The relative abundance of peak areas observed in the raw data was smoothed by the author to estimate the frequency distribution of all component peaks; here, we consider only the data on observed peaks to fit to potential pdfs, again with Arbitrary Censoring in Minitab. These results are included in the table below.

Table 5: Fits of crude oil and plant extract data to the pdfs. Data 1 includes the assumption that 50% of the components are undetected. In Data 2 no assumption is specified for undetected components.

	Mean		**Std Dev**		**Correlation Coeff.**	
Crude Oil - MS	**Data 1**	**Data 2**	**Data 1**	**Data 2**	**Data 1**	**Data 2**
Log-normal	0.1629	0.0861	0.0677	0.1551	0.983	0.984
Weibull	0.1547	0.0481	0.0717	0.0313	0.991	0.963
Normal					0.992	0.904
Plant Extracts - HPLC/UV Vis						
Log-normal	2.7276		4.127		0.997	
Weibull	2.9411		3.4837		0.966	
Normal					0.851	

Although the work by Enke and Nagels provides evidence to consider the log-normal pdf to describe the mixtures considered, examination of the Weibull pdf suggests that this distribution may also fit the data well for both the crude oil and plant extracts datasets, as shown in Table 5. Without specifying a number of undetected components, the log-normal and Weibull distributions both provide fits notably better than a simple normal distribution, as judged by the correlation coefficients. However, this also identifies the influence of assumptions made regarding undetected and unresolved signals, and highlights a difficulty in ensuring the accuracy of the fitting procedure. It is well recognized that fitting data over a limited range of a pdf may result in false extrapolation. In the fitting process used for the Enke and Nagels datasets, no data below the detection limit was or can be fit because it cannot be detected, and this tends to make the argument of a unique pdf less plausible.

Data from both natural mixtures were also analyzed via MATLAB parameter estimates for the exponential, Weibull and log-normal pdfs using the MATLAB functions expfit, wblfit, and lognfit respectively. A comparison of these parameter values with those obtained by Minitab for the exponential, Weibull and log-normal

pdfs are shown below in Table 6. Although the parameters estimated from the plant extract data are comparable between the two methods, these results suggest there is a justification for using other pdfs than log-normal as models for the natural mixture data, including the Weibull pdf.

Table 6: Comparison of curve fitting for the Enke and Nagels crude oil and plant extract datasets with Minitab and MATLAB.

Crude Oil	**Mean**			
	Minitab	**MATLAB**		
Exponential	0.062	0.004		
	Shape Parameter		**Scale Parameter**	
	Minitab	**MATLAB**	**Minitab**	**MATLAB**
Weibull	1.569	0.910	0.054	0.004
Log-normal	-3.176	-6.031	1.203	0.807

Plant Extracts	**Mean**			
	Minitab	**MATLAB**		
Exponential	5.247	2.765		
	Shape Parameter		**Scale Parameter**	
	Minitab	**MATLAB**	**Minitab**	**MATLAB**
Weibull	0.848	0.926	2.699	2.640
Log-normal	0.408	0.469	1.091	0.956

One important area of note is the difference observed between the parameters determined from Minitab software and those determined using MATLAB. We may offer some insight into the differences between the algorithms behind these programs, while the exact estimation processes are property of Minitab, Inc. and MathWorks, Inc. respectively. The Arbitrary Censoring distribution analysis in Minitab employed least squares estimates to the data provided, fitting a regression curve to the points via a minimum of the average squared difference between the observed and predicted

values, or the least square error. The parameter estimates in MATLAB, however, use the principle of maximum likelihood to search over all possible parameter values for the most likely model fit. While this subtle difference in regression tactic is likely responsible for the differences between the parameter estimates, we again note that underlying assumptions and models are not readily available for these programs. A significant conclusion from this work remains a justification for probability distribution functions other than log-normal to be examined as potential models for natural mixture data.

We note the potential that several pdfs may be considered to describe signals associated with natural mixtures. A thorough analysis, however, necessitates more readily-available data and a means of estimating undetected and unresolved signals. Since this problem itself arises from an inability to find these signals in real data, we instead turn to simulated results to better quantify signal loss. For the Weibull pdf, it can be adjusted so that all scenarios can be covered from an exponential-like behavior (a Weibull pdf can be exactly equivalent to an exponential pdf with the proper parameters) to log-normal-like behavior. Hence, modeling can mimic the case from an excess of the lowest concentration components, as embodied in the exponential pdf, to a distribution like log-normal where the lowest concentration components appear with almost zero probability.

3.3.2 Theoretical Chromatograms and Fraction of Peaks Lost

Figure 7 shows a comparison of synthetic chromatograms from the log-normal, exponential and Weibull amplitude pdfs shown in Figure 4. The mean and standard deviation used for the log-normal, exponential and Weibull pdfs are from the raw data on plant extract signals reported by Enke and Nagels. The mean and standard

deviation for the Weibull 2 pdf were chosen to encompass a case with a smaller probability of trace components relative to the mean. The saturation, α, is 0.1 in all four cases. Also shown in these figures is the location where SCPs are missing due to loss by peak overlap. It is easy to see that these peaks and the corresponding amplitude distributions look different by inspection. For example, the log-normal distribution appears to have a number of high amplitude peaks among the mean amplitude heights; this is due to the power law behavior of both log-normal and Weibull distributions[116] which are both called "heavy-tailed" distributions due to the persistence at higher values of the independent variable. In the case of the Weibull 1 distribution, there is a high probability of low amplitude peaks, and these are hard to see in the chromatogram. In the case of the Weibull 2 distribution, the parameters were chosen (referring to Figure 4) so that the tail is suppressed and more low amplitude peaks is favored. In the case of the exponential distribution, the raw data parameters favor almost a uniform distribution of SCP heights (shown in Figure 4) and this uniformity reveals itself as shown in Figure 7. The loss in specific SCPs, shown as red bars below these graphs, is discussed below for a number of α values.

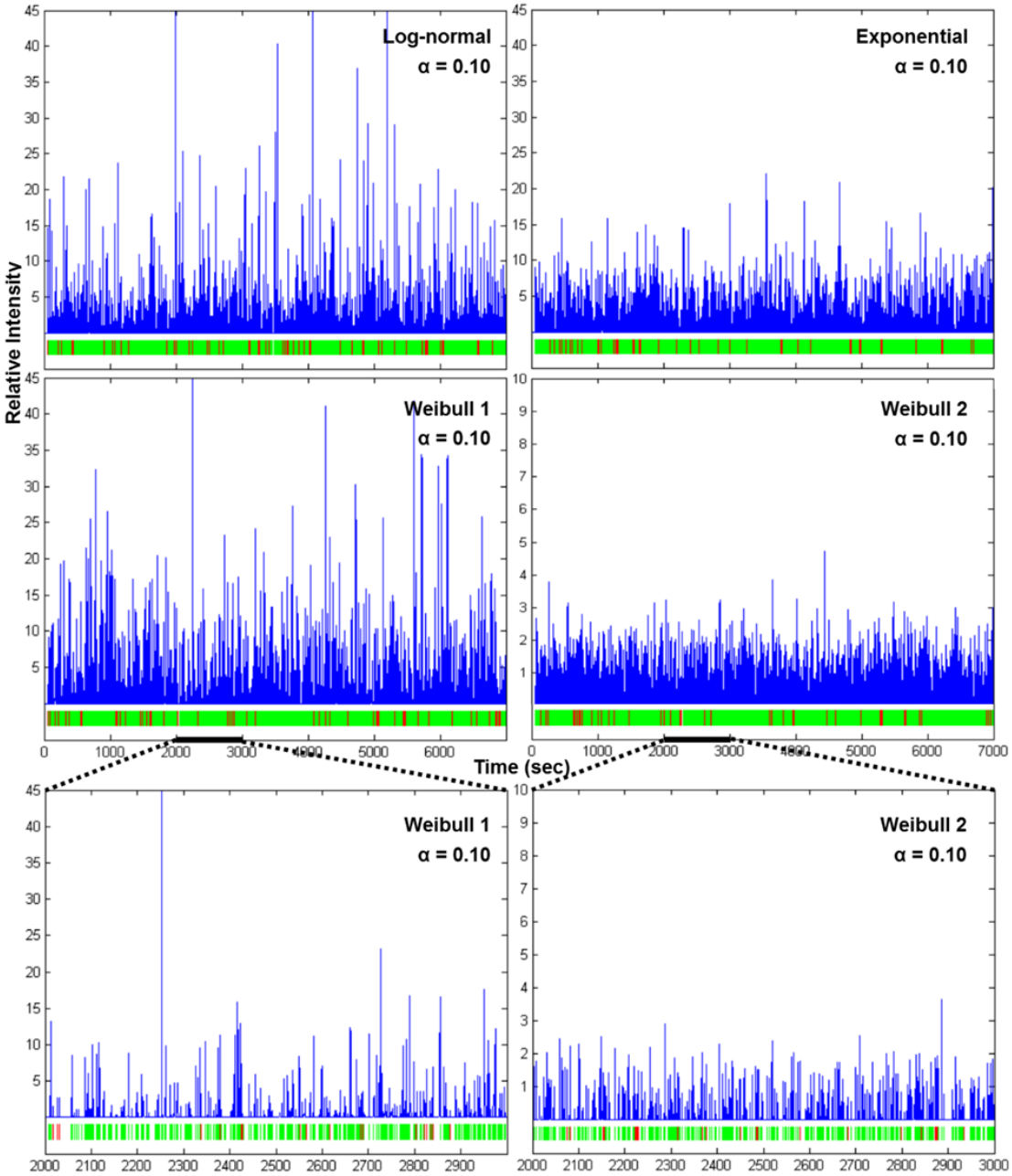

Figure 7: A comparison of the chromatograms generated with the different peak amplitude probability density functions (pdfs) and the loss of resolved peaks shown as red lines. The green lines are single component peaks (SCPs) that have recognizable maxima. The amplitude scale for all but the Weibull 2 result is the same. The Weibull 2 amplitude scale is expanded to show comparison. The bottom two are expanded scale.

A comparison of these case 1 and case 2 results utilizing the two Weibull pdfs with different saturation values is shown in Figure 8. As can be seen from the loss diagrams given here at different saturation values α, as saturation increases, the loss of SCPs increases as expected. The loss diagram, although visually useful, does not give a quantitative estimate of the loss of SCPs as a function of SCP amplitude. That loss of peaks is given in Table 7 and will be discussed shortly. However, Table 7 does show the fraction of visible peaks p/m from both the simulation results and from Equation 3.15 given α. These numbers indicate a slight loss of SCPs for all amplitudes at α=0.10 and shows approximately 10% and 8% loss of SCPs for the Weibull 1 and 2 cases, respectively. Higher saturations show more severe loss of SCPs with about 50% loss of SCPs at α=1.00. Note that the simple theory from SOT given in Equation 3.15 works well at low α, but underestimates p/m values for higher saturations.

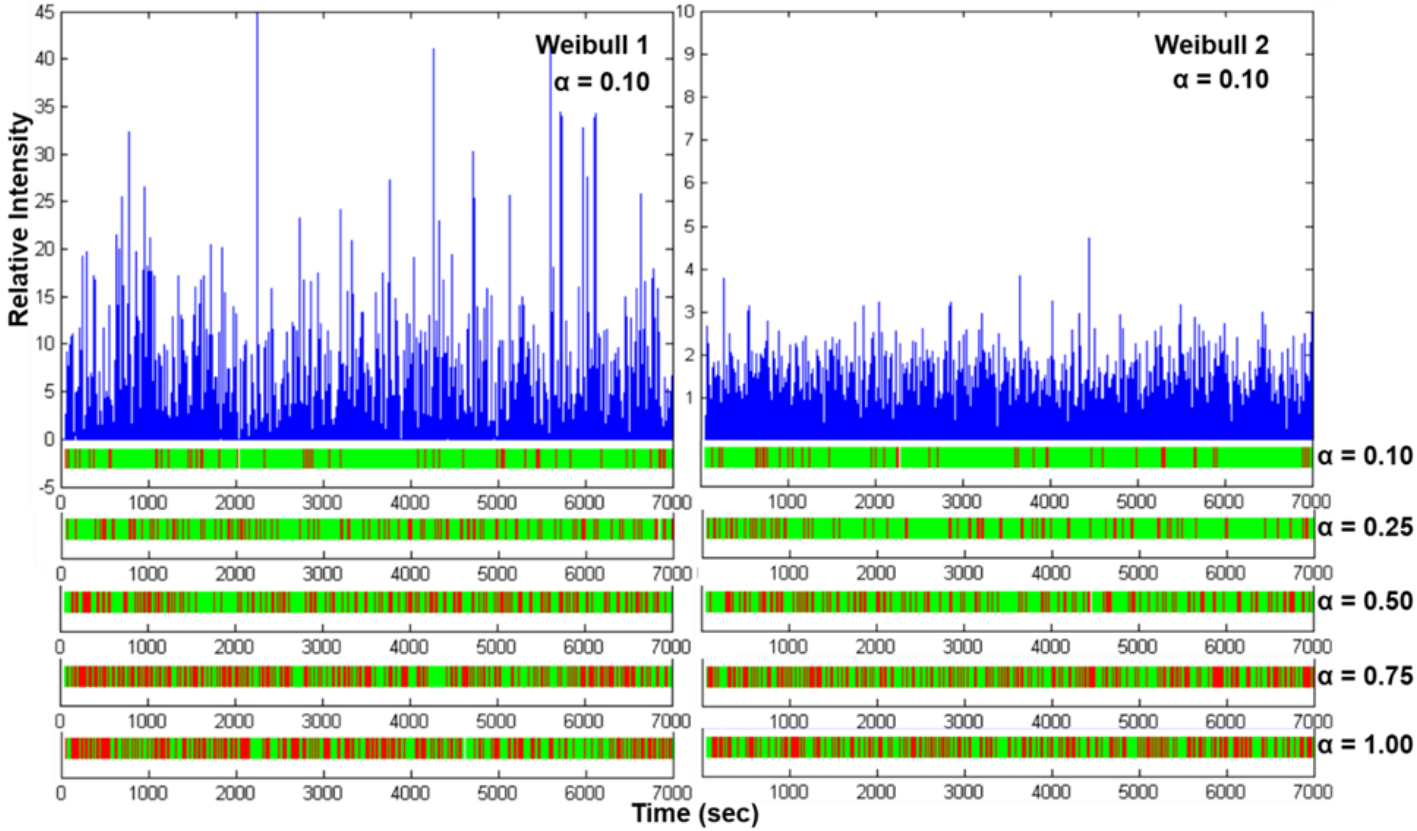

Figure 8: The loss of peaks for the Weibull 1 and Weibull 2 models. The chromatograms are calculated at α=0.10. The loss diagrams are shown for α=0.10, 0.25, 0.50, 0.75 and 1.00 for both models. The loss in SCPs due to peak overlap are tabulated in the figure caption to Figure 9A and 9B. The percent loss of SCPs is given in Table 7 at the different signal-to-noise ratios at the limit of detection (LOD).

Table 7: The fraction of peaks lost as a function of the signal-to-noise ratio (SNR). The α=0 results are obtained from the cumulative distributions. The fraction of SCPs found, p/m, are from the simulation and from Equation 3.15.

	Weibull 1					
SNR	**100**	**500**	**1000**	**5000**	**p/m**	**p/m**
LOD	8.31E-02	1.66E-02	8.31E-03	1.66E-03	**simulation**	**exp(-α_e)**
α = 0.00	12.30%	4.56%	2.95%	1.05%	1	1
α = 0.10	1.17% ± 0.16%	0.483% ± 0.227%	0.342% ± 0.166%	0.0667% ± 0.0272%	0.908 ± 0.005	0.905
α = 0.50	3.61% ± 0.16%	1.33% ± 0.21%	0.817% ± 0.164%	0.250% ± 0.155%	0.696 ± 0.007	0.607
α = 1.00	5.91% ± 0.59%	2.22% ± 0.24%	1.46% ± 0.20%	0.633% ± 0.136%	0.515 ± 0.002	0.368
α = 1.50	7.18% ± 0.62%	2.67% ± 0.38%	1.78% ± 0.30%	0.700% ± 0.181%	0.400 ± 0.006	0.223
	Weibull 2					
SNR	**100**	**500**	**1000**	**5000**	**p/m**	**p/m**
LOD	2.71E-02	5.42E-03	2.71E-03	5.42E-04	**simulation**	**exp(-α_e)**
α = 0.00	0.444%	0.0397%	0.0140%	0.00120%	1	1
α = 0.10	0.0250% ± 0.0319%	0.00% ± 0.00%	0.00% ± 0.00%	0.00% ± 0.00%	0.925 ± 0.006	0.905
α = 0.50	0.142% ± 0.032%	0.00833% ± 0.01667%	0.00% ± 0.00%	0.00% ± 0.00%	0.728 ± 0.001	0.607
α = 1.00	0.175% ± 0.050%	0.00833% ± 0.01667%	0.00% ± 0.00%	0.00% ± 0.00%	0.549 ± 0.008	0.368
α = 1.50	0.233% ± 0.067%	0.00833% ± 0.01667%	0.00% ± 0.00%	0.00% ± 0.00%	0.436 ± 0.006	0.223

These previous results lump together the total number of SCPs lost due to peak overlap regardless of SCP amplitude. The distribution of lost amplitudes is shown in Figures 9A and 9B for four different α values used in the study. As shown in both Weibull 1 and 2 cases, the largest number of lost peaks due to overlap is from the smallest amplitudes. These figures also show the cumulative percent loss. For both Weibull 1 and 2 cases, it appears that the shapes of the frequency of peaks lost stays relatively constant as α increases. Obviously, more total peaks are lost as α is increased, but the shapes of the distributions, as a function of peak heights, are similar.

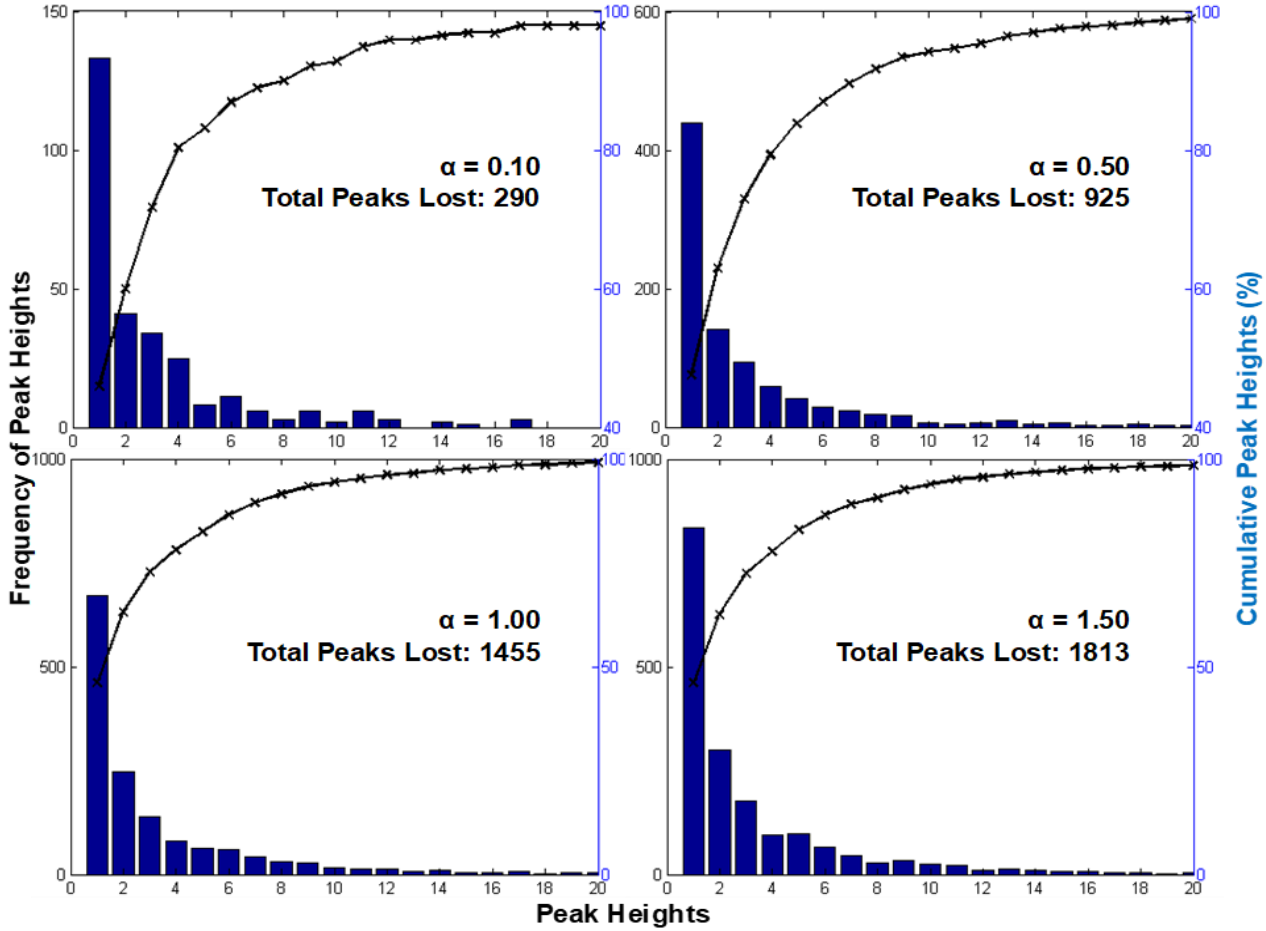

Figure 9A: Lost peaks as a function of peak heights for the Weibull 1 distribution at 4 different α values.

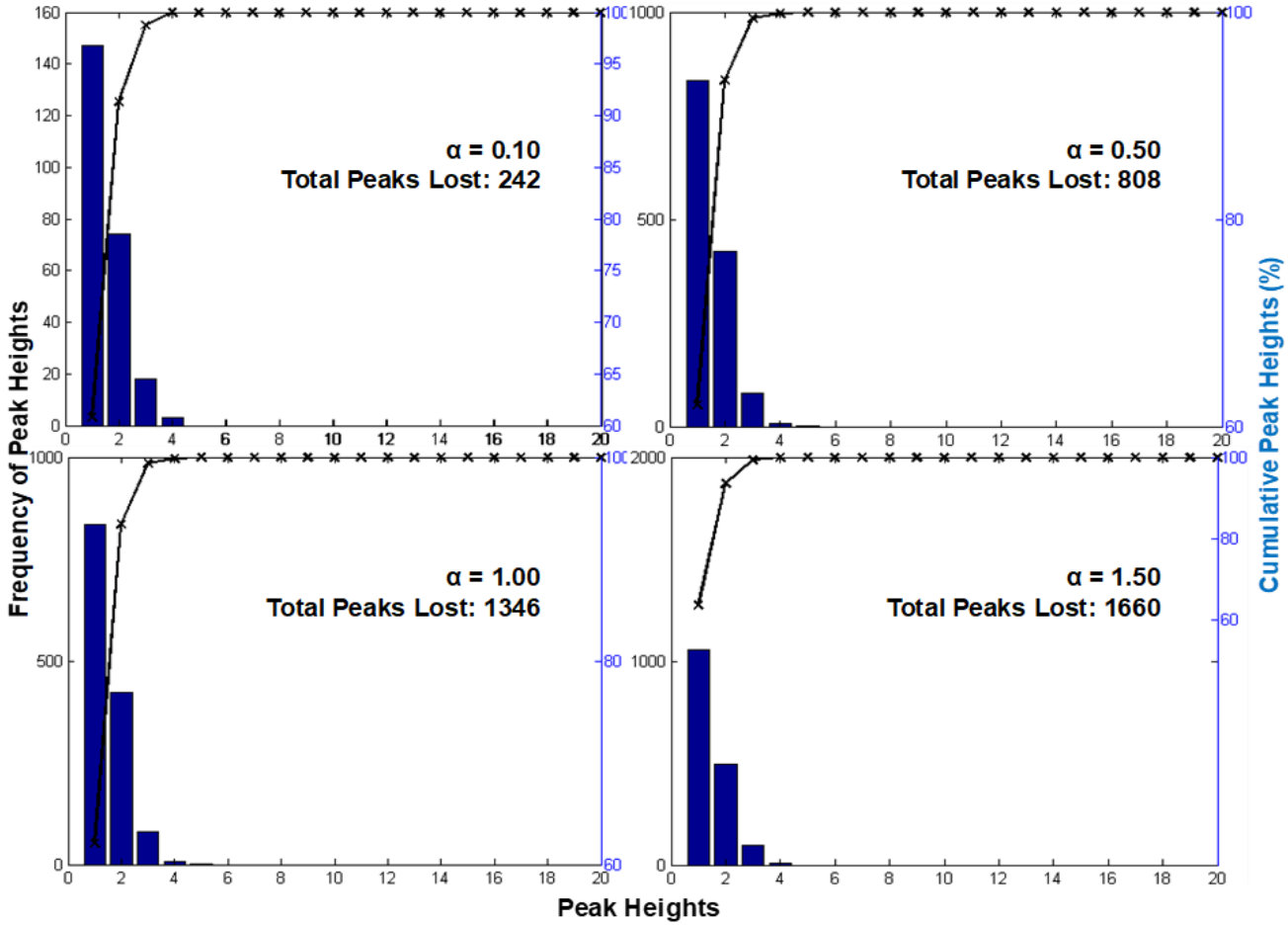

Figure 9B: Lost peaks as a function of peak heights for the Weibull 2 distribution at 4 different α values.

Closer inspection between the Weibull case 1 and case 2 plots in Figure 9A and 9B show that more total peaks are lost with the Weibull 1 case. This is most likely due to having more low amplitude peak heights for the Weibull 1 case, shown clearly in the nature of the pdfs given in Figure 4. For the Weibull 2 peak profiles in Figure 9B, few higher amplitude peak heights are eliminated and this is tied to the basic shape of the Weibull 2 pdf shown in Figure 4. Although both Weibull amplitude models appear different, the plots of number of SCPs lost versus peak heights all appear exponential in nature. This may be due to the loss mechanism as a superposition effect where smaller amplitudes are much more susceptible to loss of uniqueness from the overlap of neighboring SCPs. This underlying mechanism is the heart of the problem; both case 1 and case 2 pdfs lose small amplitude SCPs due to superposition overlap and lose the peak characteristic when surrounded by larger neighbors.

The peak heights lost due to overlap can be further expanded to a finer scale and the LOD superimposed on this scale, as shown in Figure 10. These results show that for the Weibull 1 model, different amounts of SCP loss are experienced over the SNR range from SNR=100 to SNR=5000 and it appears that approximately 5-fold more SCPs are lost for the highest SNR considered here to the lowest SNR. This is surprising but not unexpected due to the exponential-like behavior inherent in the data shown in Figure 10. A much smaller number of peaks are lost at the same SNRs and LODs for the Weibull 2 model. This is again due to the shape of the Weibull 2 amplitude pdf, and demonstrates the sensitivity of this loss to the shape of the amplitude pdf.

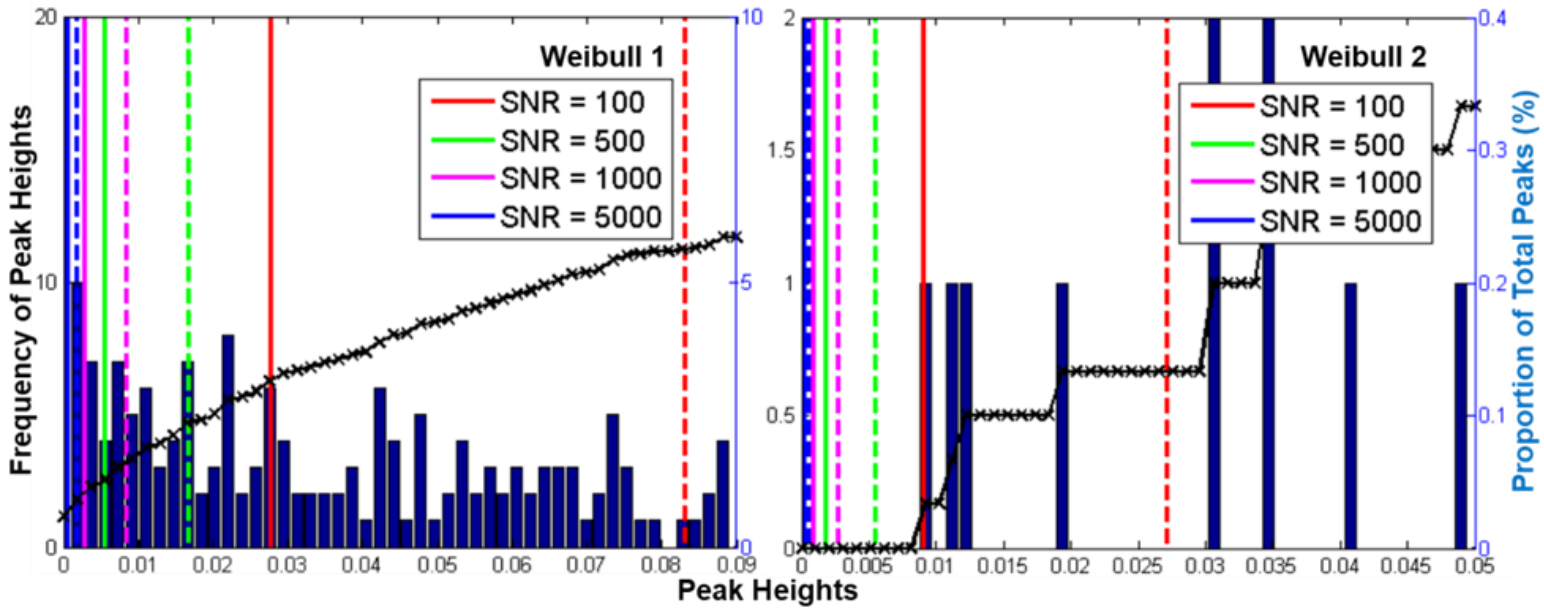

Figure 10: The distribution of lost signal amplitudes for the Weibull 1 and Weibull 2 amplitude distributions for α=1.00.

These results highlight the very nature of this study. In the case 1 scenario for the Weibull 1 amplitude pdf, SCPs are lost at the rate of a few percent, at the stated SNRs. For the case 2 distribution, much less loss is shown at the stated SNRs. In both cases, the overlap superposition mechanism is causing peak loss.

An infinite resolution model (α=0.00) can be derived from the cumulative density function (cdf) of the Weibull model[101]. This model, in principle, can be used to probe what fraction of SCPs would be masked by noise below the LOD when peak overlap is not present. This model assumes that all peaks are resolvable, i.e. the peaks are non-overlapping, and in the extreme would be represented as Dirac delta functions[117] at each retention time. The results of these calculations are given in Table 7 for the Weibull 1 and 2 models at zero saturation and the cumulative densities are shown in Figure 11 for these models.

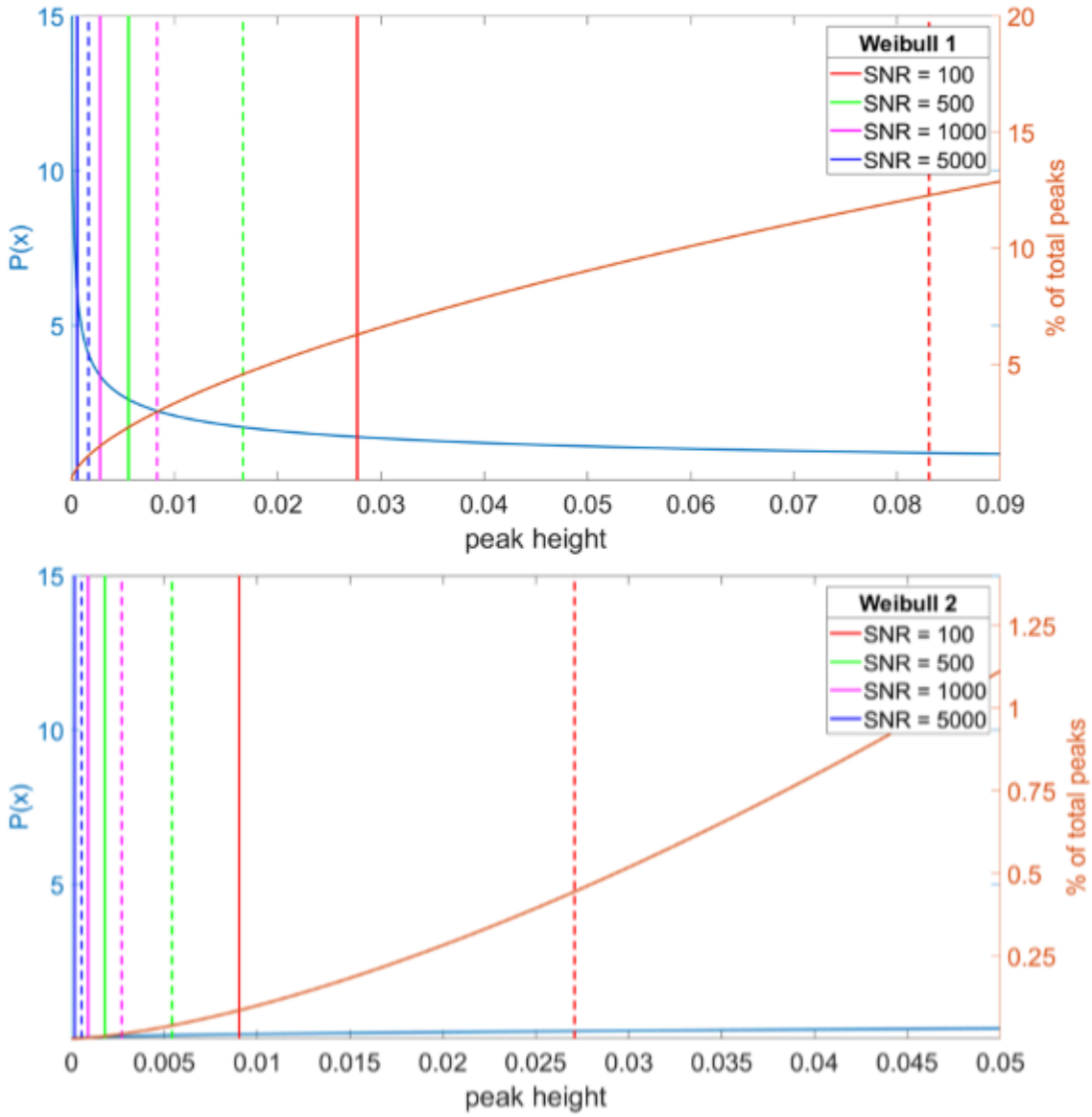

Figure 11: The loss of total SCPs due to gating the amplitude of potential peaks in the $\alpha = 0$ limit of infinitely thin peaks. The two plots are for the Weibull 1 (top) and Weibull 2 (bottom) amplitude pdfs. The blue curves are the probability density function and the tan curves are the cumulative density functions. The dashed lines are the corresponding limit of detection to the SNR noise amplitude shown in vertical solid colors.

A few things stand out here. First, without overlap the fraction of lost SCPs would appear much higher than that for finite resolution (finite α) models given also in Table 7. However, this model has no overlap, so it has no influence on adjacent, neighboring SCPs. Hence, this is strictly valid for only the height distribution and says nothing about the superposition loss. In the limit this suggests a much larger effect, based on pure amplitudes, rather than due to overlap superposition. However, the comparison is interesting because partial additivity of neighboring zones can boost an SCP over the LOD and artificially lower the number of lost SCPs. Hence, the comparison of the infinite resolution model is not a good comparison but does emphasize some important points. The superposition models at finite α most likely do not sample the complete amplitude density, and that would suggest numbers in Table 7 for finite α are lower estimates than for an exhaustively sampled density function. Nonetheless, having low-side estimates still shows that a significant number of SCPs can still be lost under these conditions.

3.4 Discussion

This work clearly suggests that the loss of detectable components in single channel detection systems is highly dependent on the shape of the amplitude density function and that for the two cases examined here the results are different, with one model being sensitive to the low signal level components and the other not sensitive to these components. The problem here is that models can be ambiguous because one can’t get at the underlying distribution; its shape and form is hidden below the LOD.

Another set of problems are also present. We have used the assumption that the response factor is not explicitly determined here because we are using a signal-based assumption about the distribution and not asking questions about the concentration

distribution. To relate the concentration model to the signal model, additional information is needed. This includes the sensitivity distribution inherent in the response factor, given in Equation 3.6. However, this factor is extremely difficult to ascertain and is dependent on a multitude of instrument factors, perhaps too many to accurately predict. These may be dependent on the amplitude distribution itself so that the response is a function of concentration.

Many detection systems are well-filtered to mask baseline noise, an effect known in chromatography[118,119] to add an additional zone broadening-like contribution which increases peak width. In this regard, the noise is masked and visually makes the signal look better, but doesn't change the SNR because both signal and noise are typically reduced. This affects the signal chain, and thus careful use of filters must be applied when working near the LOD.

Finally, mass spectrometry should be effective in reducing the saturation α when used as a multichannel detector. This is because detection on separate mass channels should drastically reduce the crowding of signals seen on a single channel detector. This strategy should effectively distribute the signal across lower occupancy detection channels. However, in many scenarios pertinent to biomarkers, such as top-down proteomics, the chromatography is run under such high saturation conditions, even with long columns and very long run times[120], that the mass spectra at specific masses is too crowded and strategies like MS/MS are needed. That increases the complexity and is not immune to other SNR limitations such as chemical noise when electrospray interfaces are used. Additionally, two-dimensional LC (2DLC) can help sort out the components in these highly complex samples and help reduce the crowding in the MS or MS/MS detection system. However, this places even more

burden on the detector because dilution will lower the signals proportionally[104], as is well-known in 2DLC. Nonetheless, these techniques will help in the future in the ability to sort out the separation and create a demand for more sensitive and lower noise detection systems to aid in the search for low-abundance biomarkers.

Chapter 4

CONCLUSION

HPLC and separation science hold endless potential in the investigations of the modern medical community. The study of bodily samples and other biological constituents for the identification, prevention and treatment of diseases relies on an ability to accurately analyze these materials. Chromatography provides a method of mixture analysis that is highly specific and tailorable, and HPLC in particular offers the advantage of faster, higher resolution separations of complex mixtures. This work sought to increase the capacity of current methods in HPLC to identify and quantify the presence of biomarkers for disease in two ways. First, a method for the determination of column porosity using mass-based measurements was proposed to better realize the physical capability of a chromatographic column for separations. Second, synthetic chromatograms were generated in order to estimate the loss in the resultant signals due to noise and peak overlap. In all, the work of this thesis sought to improve upon current chromatographic methods to enable better biological separations and diagnostics.

This research contributed valuable findings to the overall knowledge of HPLC and chromatographic signals. The mass-based method for determination of column porosity has the potential to provide information on the performance of a column based solely on its physical contents. Despite the issues encountered for larger pore materials using this method, a great deal of success was observed in replicating measurements for the total pore volume of smaller pore materials. Static column mass

measurements made using total pore blocking remove the dependence on void time determination seen in dynamic methods, and thus offer an advantage over current techniques. This method highlights the possibility of improving upon current separation techniques with a better estimation of column physical performance.

The work of this thesis in employing statistical overlap theory to generate synthetic chromatograms contributes to a means of estimating the resolution of resulting signals. A better understanding of the impacts of noise and peak overlap in a chromatogram provides another possibility for improving separations, and seeks to define where low-level biomarkers could be found. This investigation into the underlying distribution of natural mixtures identified several possibilities, each with their own implications for unidentified components. While some cases suggested higher quantities of low-level signals than others, many of the simulations run demonstrated a commonality in the relative spread of peaks lost due to overlap. By using simulated chromatograms to identify low-level signals, this work was able to draw conclusions about the portion of a natural mixture that cannot be seen with current methods.

Through this work, further insight was provided into the ability of current methods to analyze natural mixtures. The development of a method to easily quantify the inner contents of a column using mass-based measurements demonstrated promise for small-pore materials, and highlighted phenomena in larger-pore materials, increasing knowledge of the physical nature of chromatographic separations. A series of computer simulations shed light on the underlying distribution of a resulting chromatographic signal and its lost components, providing the possibility of quantifying part of a mixture that cannot currently be observed. Despite these

successes, this work has only begun to explore the opportunities in the area of chromatography and the search for biomarkers. The findings presented here will hopefully inspire future work to improve upon current methods in chromatography for biological separations.

REFERENCES

1. S. Moldoveanu, V. David, Selection of the HPLC Method in Chemical Analysis, First, Elsevier, 2016.

2. R. Hayes, A. Ahmed, T. Edge, H. Zhang, Core-shell particles: Preparation, fundamentals and applications in high performance liquid chromatography, J. Chromatogr. A. 1357 (2014) 36–52.

3. K.K. Unger, R. Skudas, M.M. Schulte, Particle packed columns and monolithic columns in high-performance liquid chromatography-comparison and critical appraisal, J. Chromatogr. A. 1184 (2008) 393–415.

4. D.R.N. and J.J.K. B. M. Wagner, S. A. Schuster, B. E. Boyes, W. L. Miles, Tools to Imporve Protein Separations, LCGC North Am. 33 (2015) 856–865.

5. H.J. Issaq, The role of separation science in proteomics research, Electrophoresis. 22 (2001) 3629–3638.

6. S.A. Schuster, B.M. Wagner, B.E. Boyes, J.J. Kirkland, Optimized superficially porous particles for protein separations, J. Chromatogr. A. 1315 (2013) 118–126.

7. B.M. Wagner, S.A. Schuster, B.E. Boyes, J.J. Kirkland, Superficially porous silica particles with wide pores for biomacromolecular separations, J. Chromatogr. A. 1264 (2012) 22–30.

8. A. Alhedai, D.E. Martire, R.P.W. Scott, Column “dead volume” in liquid chromatography, Analyst. 114 (1989) 869.

9. A.M. Krstulovlć, H. Colin, G. Guiochon, Comparison of Methods Used for the Determination of Void Volume in Reversed-Phase Liquid Chromatography, Anal. Chem. 54 (1982) 2438–2443.

10. J.H. Knox, R. Kaliszan, Theory of solvent disturbance peaks and experimental determination of thermodynamic dead-volume in column liquid chromatography, J. Chromatogr. A. 349 (1985) 211–234.

11. C.A. Rimmer, C.R. Simmons, J.G. Dorsey, The measurement and meaning of void volumes in reversed-phase liquid chromatography, J. Chromatogr. A. 965 (2002) 219–232.

12. Y. Yao, A.M. Lenhoff, Determination of pore size distributions of porous chromatographic adsorbents by inverse size-exclusion chromatography, J. Chromatogr. A. 1037 (2004) 273–282.

13. F. Gritti, Y. Kazakevich, G. Guiochon, Measurement of hold-up volumes in reverse-phase liquid chromatography. Definition and comparison between static and dynamic methods, J. Chromatogr. A. 1161 (2007) 157–169.

14. S.B. Nimse, M.D. Sonawane, K.-S. Song, T. Kim, Biomarker detection technologies and future directions, Analyst. 141 (2016) 740–755.

15. G.D. Rabinovici, Amyloid biomarkers: pushing the limits of early detection, Brain. 139 (2016) 1002–1013.

16. P. Drake, B. Schilling, B. Gibson, S. Fisher, Elucidation of N-Glycosites Within Human Plasma Glycoproteins for Cancer Biomarker Discovery, Methods Mol. Biol. 951 (2013) 307–322.

17. J.M. Davis, J.C. Giddings, Statistical Theory of Component Overlap in Multicomponent Chromatograms, Anal. Chem. 55 (1983) 418–424.

18. J.M. Davis, New theory for distribution of minimum resolution in multi-component separations with noise/detection limits, J. Chromatogr. A. 1251 (2012) 1–9.

19. I. Halász, K. Martin, Pore Sizes of Solids, Angew. Chemie Int. Ed. English. 17 (1978) 901–908.

20. J.H. Knox, H.P. Scott, Theoretical models for size-exclusion chromatography and calculation of pore size distribution from size-exclusion chromatography data, J. Chromatogr. A. 316 (1984) 311–332.

21. H. Guan, G. Guiochon, Study of physico-chemical properties of some packing materials, J. Chromatogr. A. 731 (1996) 27–40.

22. E.H. Slaats, J.C. Kraak, W.J.T. Brugman, H. Poppe, N. Achtergracht, Study of the Influence of Competition and Solvent Interaction on Retention in Liquid-Solid Chromatography by Measurement of Activity Coefficients in the Mobile Phase, J. Chromatogr. A. 149 (1978) 255–270.

23. R. McCormick, B. Karger, Distribution phenomena of mobile-phase components and determination of dead volume in reversed-phase liquid chromatography, Anal. Chem. 52 (1980) 2249–2257.

24. F. Gritti, G. Guiochon, Unusual behavior of the height equivalent to a theoretical plate of a new poroshell stationary phase at high temperatures, J. Chromatogr. A. 1169 (2007) 125–138.

25. M.J.M. Wells, C.R. Clark, Liquid Chromatographic Elution Characteristics of Some Bonded Solutes Used To Measure Column Void Volume on Phases, Water. (1981) 1341–1345.

26. D.P. Rillema, L.G. Daignault, D.C. Jackman, Behavior of Simple Salts on Silica and C18 Columns: Retention Dynamics of Cations, Anions and Ion Pairs, J. Chromatogr. 462 (1989) 71–84.

27. I. Nischang, G. Chen, U. Tallarek, Electrohydrodynamics in hierarchically structured monolithic and particulate fixed beds, J. Chromatogr. A. 1109 (2006) 32–50.

28. F. Gritti, G. Guiochon, Experimental evidence of the influence of the surface chemistry of the packing material on the column pressure drop in reverse-phase liquid chromatography, J. Chromatogr. A. 1136 (2006) 192–201.

29. S. Jung, S. Ehlert, M. Pattky, U. Tallarek, Determination of the interparticle void volume in packed beds via intraparticle Donnan exclusion, J. Chromatogr. A. 1217 (2010) 696–704.

30. D. Cabooter, F. Lynen, P. Sandra, G. Desmet, Total pore blocking as an alternative method for the on-column determination of the external porosity of packed and monolithic reversed-phase columns, J. Chromatogr. A. 1157 (2007) 131–141.

31. A. Liekens, D. Cabooter, J. Denayer, G. Desmet, A study of the parameters affecting the accuracy of the total pore blocking method, J. Chromatogr. A. 1217 (2010) 6754–6761.

32. A. Liekens, J. Denayer, G. Desmet, Extending the Total Pore Blocking method to normal phase high performance liquid chromatography, J. Chromatogr. A. 1218 (2011) 7781–7787.

33. D. Cabooter, J. Billen, H. Terryn, F. Lynen, P. Sandra, G. Desmet, Detailed characterisation of the flow resistance of commercial sub-2 μm reversed-phase columns, J. Chromatogr. A. 1178 (2008) 108–117.

34. H. Song, G. Desmet, D. Cabooter, Evaluation of the Kinetic Performance Differences between Hydrophilic-Interaction Liquid Chromatography and Reversed-Phase Liquid Chromatography under Conditions of Identical Packing Structure, Anal. Chem. 87 (2015) 12331–12339.

35. F. Gritti, G. Guiochon, Impact of Retention on Trans-Column Velocity Biases in Packed Columns, AIChE J. 56 (2010) 1495–1509.

36. G. Guiochon, F. Gritti, New insights on mass transfer kinetics in chromatography, AIChE J. 57 (2011) 333–345.

37. G. Guiochon, A. Felinger, Fundamentals of Preparative and Nonlinear Chromatography, Academic Press, New York, 2006.

38. F. Gritti, G. Guiochon, Comparison between the loading capacities of columns packed with partially and totally porous fine particles. What is the effective surface area available for adsorption?, J. Chromatogr. A. 1176 (2007) 107–122.

39. F. Gritti, I. Leonardis, J. Abia, G. Guiochon, Physical properties and structure of fine core-shell particles used as packing materials for chromatography. Relationships between particle characteristics and column performance, J. Chromatogr. A. 1217 (2010) 3819–3843.

40. F. Gritti, G. Guiochon, The mass transfer kinetics in columns packed with Halo-ES shell particles, J. Chromatogr. A. 1218 (2011) 907–921.

41. F. Gritti, G. Guiochon, Comparative study of the performance of columns packed with several new fine silica particles. Would the external roughness of the particles affect column properties?, J. Chromatogr. A. 1166 (2007) 30–46.

42. F. Gritti, G. Guiochon, Performance of columns packed with the new shell Kinetex-C18 particles in gradient elution chromatography, J. Chromatogr. A. 1217 (2010) 1604–1615.

43. D. Cabooter, A. Fanigliulo, G. Bellazzi, B. Allieri, A. Rottigni, G. Desmet, Relationship between the particle size distribution of commercial fully porous and superficially porous high-performance liquid chromatography column packings and their chromatographic performance, J. Chromatogr. A. 1217 (2010) 7074–7081.

44. F. Gritti, A. Cavazzini, N. Marchetti, G. Guiochon, Comparison between the efficiencies of columns packed with fully and partially porous C18-bonded silica materials, J. Chromatogr. A. 1157 (2007) 289–303. doi:10.1016/j.chroma.2007.05.030.

45. J. Bear, Dynamics of fluid in porous media, Dover Publications, Inc., New York, 1972.

46. R. Lenormand, C. Zarcone, A. Sarr, Mechanisms of the displacement of one fluid by another in a network of capillary ducts, J. Fluid Mech. 135 (1983) 337–353.

47. R. Lenormand, E. Toboul, C. Zarcone, Numerical models and eperiments on immiscible displacements in porous media., J. Fluid Mech. 189 (1988) 165–187.

48. F.A.L. Dullien, Porous Media: Fluid transport and pore structure, 2nd ed., San Diego, CA, 1992.

49. Y. Edery, D. Weitz, S. Berg, Surfactant Variations in Porous Media Localize Capillary Instabilities during Haines Jumps, Phys. Rev. Lett. 120 (2018) 28005.

50. Y.B. Acar, A.N. Alshawabkeh, Principles of Electrokinetic Remidiation, Environ. Sci. Technol. 27 (1993) 2638–2647.

51. V. Alvarado, E. Manrique, Enhanced oil recovery: An update review, Energies. 3 (2010) 1529–1575.

52. P. Raffa, A.A. Broekhuis, F. Picchioni, Polymeric surfactants for enhanced oil recovery: A review, J. Pet. Sci. Eng. 145 (2016) 723–733.

53. T.H. Walter, P. Iraneta, M. Capparella, Mechanism of retention loss when C8 and C18 HPLC columns are used with highly aqueous mobile phases, J. Chromatogr. A. 1075 (2005) 177–183.

54. P.G. DeGennes, F. Brouchard-Wyatt, D. Quere, Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves, Springer, New York, 2004.

55. D.G. Kiriakidis, G.H. Neale, E. Mitsoulis, Slow immiscible displacement flow in disordered porous media, Can. J. Chem. Eng. 70 (1992) 1220–1223.

56. A.W. Adamson, A.P. Gast, Physical Chemistry of Surfaces, 6th ed., Wiley-Interscience, New York, 1997.

57. R.F. Probstein, Physicochemical hydrodynamics: an introduction, 2nd ed., Wiley-Interscience, New York, 2003.

58. M.R. Schure, R.S. Maier, T.J. Shields, C.M. Wunder, B.M. Wagner, Intraparticle and interstitial flow in wide-pore superficially porous and fully porous particles, Chem. Eng. Sci. 174 (2017) 445–458.

59. K.Y. Law, Definitions for Hydrophilicity, Hydrophobicity, and Superhydrophobicity: Getting the Basics Right, J. Phys. Chem. Lett. 5 (2014) 686–688.

60. W.A. Zisman, Relation of the Equilibrium Contact Angle to Liquid and Solid Constitution, Adv. Chem. 43 (1964) 1–51.

61. S. Zeppieri, J. Rodriguez, a L.L. de Ramos, Interfacial tension of alkane + water systems, J. Chem. Eng. Data. 46 (2001) 1086–1088.

62. A. Goebel, K. Lunkenheimer, Interfacial Tension of the Water/ *n* -Alkane Interface, Langmuir. 13 (1997) 369–372.

63. B. Janczuk, T. Bialopiotrowicz, W. Wojcik, The Surface Tension Components of Aqueous Alcohol Solutions, Colloids and Surfaces. 36 (1989) 391–403.

64. B. Samuel, H. Zhao, K.Y. Law, Study of wetting and adhesion interactions between water and various polymer and superhydrophobic surfaces, J. Phys. Chem. C. 115 (2011) 14852–14861.

65. L.J.M. Schlangen, L.K. Koopal, M.A.C. Stuart, J. Lyklema, M. Robin, H. Toulhoat, Thin Hydrocarbon and Water Films on Bare and Methylated Silica: Vapor Adsorption, Wettability, Adhesion, and Surface Forces, Langmuir. 11 (1995) 1701–1710.

66. R.K. Iler, The Chemistry of Silica, New York, 1979.

67. A.P. Legrand, The Surface Properties of Silicas, New York, 1998.

68. J. Yang, S. Meng, L. Xu, E.G. Wang, Water adsorption on hydroxylated silica surfaces studied using the density functional theory, Phys. Rev. B - Condens. Matter Mater. Phys. 71 (2005) 1–12.

69. R.K. Lindsey, J.L. Rafferty, B.L. Eggimann, J.I. Siepmann, M.R. Schure, Molecular simulation studies of reversed-phase liquid chromatography, J. Chromatogr. A. 1287 (2013) 60–82.

70. M.A. Barrufet, K.R. Hall, A. Estrada-Baltazar, G.A. Iglesias-Silva, Liquid Viscosity of Octane and Pentane + Octane Mixtures from 298.15K to 373.15K up to 25 MPa, J. Chem. Eng. Data. 44 (1999) 1310–1314.

71. P.G. Saffman, G.I. Taylor, The Penetration of a Fluid into a Medium or Hele-Shaw Cell Containing a More Viscous Liquid, Proc. R. Soc. A. (1958) 312–329.

72. G.M. Homsy, Viscous Fingering in Porous Media, Annu. Rev. Fluid Mech. 19 (1987) 271–311.

73. R.E. Gerszten, T.J. Wang, The search for new cardiovascular biomarkers, Nature. 451 (2008) 949–952.

74. L. Anderon, Candidate-based proteomics in the search for biomarkers of cardiovascular disease, J. Physiol. 563 (2005) 23–60.

75. M. Polanski, N.L. Anderson, A list of candidate cancer biomarkers for targeted proteomics., Biomark. Insights. 1 (2007) 1–48. http://www.ncbi.nlm.nih.gov/pubmed/19690635%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2716785.

76. T. Liu, W.-J. Qian, M.A. Gritsenko, W. Xiao, L.L. Moldawer, A. Kaushal, M.E. Monroe, S.M. Varnum, R.J. Moore, S.O. Purvine, R. V Maier, R.W. Davis, R.G. Tompkins, D.G. Camp, R.D. Smith, High dynamic range characterization of the trauma patient plasma proteome, Mol. Cell. Proteomics. 5 (2006) 1899–913.

77. J.M. Jacobs, J.N. Adkins, W.J. Qian, T. Liu, Y. Shen, D.G. Camp, R.D. Smith, Utilizing human blood plasma for proteomic biomarker discovery, J. Proteome Res. 4 (2005) 1073–1085.

78. B. Baddanov, R.D. Smith, Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up, Mass Spectrom. Rev. 24 (2005) 168–200.

79. J.C. Tran, L. Zamdborg, D.R. Ahlf, J.E. Lee, A.D. Catherman, K.R. Durbin, J.D. Tipton, A. Vellaichamy, J.F. Kellie, M. Li, C. Wu, S.M.M. Sweet, B.P. Early, N. Siuti, R.D. Leduc, P.D. Compton, P.M. Thomas, N.L. Kelleher, Mapping intact protein isoforms in discovery mode using top-down proteomics, Nature. 480 (2011) 254–258.

80. J. Stahl-Zeng, V. Lange, R. Ossola, K. Eckhardt, W. Krek, R. Aebersold, B. Domon, High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N -Glycosites, Mol. Cell. Proteomics. 6 (2007) 1809–1817.

81. B. Adamczyk, T. Tharmalingam, P.M. Rudd, Glycans as cancer biomarkers, Biochim. Biophys. Acta - Gen. Subj. 1820 (2012) 1347–1353.

82. D.D. Taylor Allen, W.S. Hancock, M. Hincapie, N. Taniguchi, S.M. Hanash, Towards an integrated proteomic and glycomic approach to finding cancer biomarkers, Genome Med. 1 (2009) 1–10.

83. C.C.L. Wong, D. Cociorva, J.D. Venable, T. Xu, J.R. Yates, Comparison of Different Signal Thresholds on Data Dependent Sampling in Orbitrap and LTQ Mass Spectrometry for the Identification of Peptides and Proteins in Complex Mixtures, J. Am. Soc. Mass Spectrom. 20 (2009) 1405–1414.

84. M. Petković, J. Schiller, J. Müller, M. Müller, K. Arnold, J. Arnhold, The signal-to-noise ratio as the measure for the quantification of lysophospholipids by matrix-assisted laser desorption/ ionisation time-of-flight mass spectrometry, Analyst. 126 (2001) 1042–1050.

85. L. Chen, C.E. Cottrell, A.G. Marshall, Effect of signal-to-noise ratio and number of data points upon precision in measurement of peak amplitude, position and width in fourier transform spectrometry, Chemom. Intell. Lab. Syst. 1 (1986) 51–58.

86. A. Makarov, E. Denisov, O. Lange, S. Horning, Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer, J. Am. Soc. Mass Spectrom. 17 (2006) 977–982.

87. K.O. Zhurov, A.N. Kozhinov, L. Fornelli, Y.O. Tsybin, Distinguishing analyte from noise components in mass spectra of complex samples: Where to cut the noise?, Anal. Chem. 86 (2014) 3308–3316.

88. P. Du, G. Stolovitzky, P. Horvatovich, R. Bischoff, J. Lim, F. Suits, A noise model for mass spectrometry based proteomics, Bioinformatics. 24 (2008) 1070–1077.

89. A.L. McCormack, D.M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, J.R. Yates, Direct Analysis and Identification of Proteins in Mixtures by LC/MS/MS and Database Searching at the Low-Femtomole Level, Anal. Chem. 69 (1997) 767–776.

90. W.-J. Qian, J.M. Jacobs, T. Liu, D.G. Camp, R.D. Smith, Advances and Challenges in Liquid Chromatography-Mass Spectrometry-based Proteomics Profiling for Clinical Applications, Mol. Cell. Proteomics. 5 (2006) 1727–1744.

91. Y. V. Karpievitch, A.D. Polpitiya, G.A. Anderson, R.D. Smith, A.R. Dabney, Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects, Ann. Appl. Stat. 4 (2010) 1797–1823.

92. S. Hua, H.N. Jeong, L.M. Dimapasoc, I. Kang, C. Han, J.S. Choi, C.B. Lebrilla, H.J. An, Isomer-specific LC/MS and LC/MS/MS profiling of the mouse serum N-glycome revealing a number of novel sialylated N-glycans, Anal. Chem. 85 (2013) 4636–4643.

93. J.P. Chervet, M. Ursem, J.P. Salzmann, Instrumental requirements for nanoscale liquid chromatography, Anal. Chem. 68 (1996) 1507–1512.

94. C.R. Mallet, Z. Lu, J.R. Mazzeo, A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts, Rapid Commun. Mass Spectrom. 18 (2004) 49–58.

95. P. Horvatovich, B. Hoekman, N. Govorukhina, R. Bischoff, Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples, J. Sep. Sci. 33 (2010) 1421–1437.

96. J.M. Davis, Statistical Theories of Peak Overlap in Chromatography, in: P.R. Brown, E. Grushka (Eds.), Adv. Chromatogr. Vol 34, Marcel Dekker, Inc., New York, 1994: pp. 109–176.

97. J.M. Davis, Computation of distribution of minimum resolution for log-normal distribution of chromatographic peak heights, J. Chromatogr. A. 1218 (2011) 7841–7849.

98. L.J. Nagels, W.L. Creten, P.M. Vanpeperstraete, Determination Limits and Distribution Function of Ultraviolet Absorbing Substances in Liquid Chromatographic Analysis of Plant Extracts, Anal. Chem. 55 (1983) 216–220.

99. C.G. Enke, L.J. Nagels, Undetected components in natural mixtures: How many? What concentrations? Do they account for chemical noise? What is needed to detect them?, Anal. Chem. 83 (2011) 2539–2546.

100. A. Gundlach-Graham, C.G. Enke, Effect of response factor variations on the response distribution of complex mixtures, Eur. J. Mass Spectrom. 21 (2015) 471–479.

101. H. Rinne, The Weibull Distribution: A Handbook, CRC Press, Boca Raton, FL, 2009.

102. W.K. Brown, K.H. Wohletz, Derivation of the Weibull distribution based on physical principles and its connection to the Rosin-Rammler and lognormal distributions, J. Appl. Phys. 78 (1995) 2758–2763.

103. B.L. Karger, M. Martin, G. Guiochon, Role of Column Parameters and Injection Volume on Detection Limits in Liquid Chromatography, Anal. Chem. 46 (1974) 1640–1647.

104. M.R. Schure, Limit of detection, dilution factors, and technique compatibility in multidimensional separations utilizing chromatography, capillary electrophoresis, and field-flow fractionation, Anal. Chem. 71 (1999) 1645–1657.

105. T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, New York, 1991.

106. Signal-to-noise ratio, Wikipedia. (2018). https://en.wikipedia.org/wiki/Signal-to-noise_ratio (accessed April 9, 2018).

107. A.L. Currie, Limits for Qualitative Detection and Quantitative Determination, Anal. Chem. 40 (1968) 586–593.

108. G.L. Long, J.D. Winefordner, Limit of Detection: A Closer Look at the IUPAC Definition, Anal. Chem. 55 (1983) 712–724.

109. J. Calvin Giddings, Maximum Number of Components Resolvable by Gel Filtration and Other Elution Chromatographic Methods, Anal. Chem. 39 (1967) 1027–1028.

110. E. Grushka, Chromatographic Peak Capacity and the Factors Influencing It, 42 (1970).

111. J.M. Davis, P.W. Carr, Effective saturation: A more informative metric for comparing peak separation in one- and two-dimensional separations, Anal. Chem. 81 (2009) 1198–1207.

112. J.M. Davis, Dependence on effective saturation of numbers of singlet peaks in one- and two-dimensional separations, Talanta. 83 (2011) 1068–1073.

113. M.R. Schure, J.M. Davis, The Simple Use of Statistical Overlap Theory in Chromatography, LCGC North Am. 33 (2015) 14–18.

114. R.W. Hamming, Numerical Methods for Scientists and Engineers, 2nd ed., Dover Publications, Inc., New York, 1987.

115. L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, New York, 1986.

116. M.R. Schroeder, Fractals, Chaos, Power Laws, Dover Publications, Inc., New York, 1991.

117. G.B. Arfken, H.J. Weber, F.E. Harris, Mathematical Methods for Physicists, 7th ed., Academic Press, Boston, Massachusetts, 2013.

118. N. Dyson, Chromatographic Integration Methods, 2nd ed., The Royal Society of Chemistry, London, 1998.

119. M.F. Wahab, P.K. Dasgupta, A.F. Kadjo, D.W. Armstrong, Sampling Frequency Response Times and Embedded Signal Filtration in Fast, High Efficiency Liquid Chromatography: a Tutorial, Anal. Chim. Acta. 907 (2016) 31–44.

120. Y. Shen, N. Tolic, P.D. Piehowski, A.K. Shukla, S. Kim, R. Zhao, Y. Qu, E. Robinson, R.D. Smith, L. Pasa-Tolic, High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics, J. Chromatogr. A. 1498 (2017) 99–110.

121. P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York, 1969.

Appendix A

POROSITY MEASUREMENTS

This Appendix contains three tables: the total pore volume obtained through mass measurements, the interstitial volume measurements made with pore blocking experiments and the volume determination results leading to the porosity values used in Table 3. Numbers shown here were read off of the scale itself and do not necessarily account for significant digits.

Table A1: Total pore volume mass measurements.

Column		Mass CH_3OH	Mass CH_2Cl_2	Mass $(CH_2)_4O$	Mass CH_3CN	Vo (cm3)
Reversed Phase						
1	A	28.3286	28.3741	28.3347	28.3240	0.08951
	B	28.3254	28.3738	28.3347	28.3245	0.09038
	C	28.3181	28.3664	28.3272	28.3168	0.09139
2	A	28.2739	28.3291	28.2837	28.2728	0.10396
	B	28.2761	28.3309	28.2859	28.2750	0.10306
3	A	28.3235	28.3732	28.3327	28.3228	0.09314
	B	28.3263	28.3766	28.3353	28.3254	0.09456
7	A	28.1965	28.2475	28.2046	28.1902	0.09987
8	A	28.4925	28.5477	28.5013	28.4934	0.10353
Normal Phase						
4	A	28.3144	28.3730	28.3257	28.3132	0.11028
	B	28.3076	28.3651	28.3208	28.3051	0.10738
5	A	29.9731	30.2588	30.0255	29.9670	0.54016
	B	29.9697	30.2569	30.0221	29.9638	0.54291
6	A	30.1440	30.4049	30.1918	30.1386	0.49324
	B	29.9013	30.1627	29.9480	29.8979	0.49374

Table A2: Interstitial volume measurements (pore blocking experiment).

Column		Mass H_2O	Mass CH_3OH	Mass C_3H_8O	Vi (cm^3)
Reversed Phase					
1	A	28.3359	28.3253	28.3298	0.04132
	B	28.3379	28.3266	28.3268	0.05574
	C	28.3221	28.3181	28.3191	0.01739
2	A	28.2925	28.2754	28.2756	0.08414
	B	28.2870	28.2766	28.2772	0.04998
3	A	28.3378	28.3248	28.3259	0.06143
	B	28.3424	28.3264	28.3272	0.07717
7	A	28.2007	28.1916	28.1919	0.04457
8	A	28.5043	28.4924	28.4925	0.05835

		Mass C_8H_{18}	Mass CH_2Cl_2	Mass $C_4H_8O_2$	Mass C_6H_{12}	Mass C_4H_8O	Vi (cm^3)
Normal Phase							
4	A	28.3094	28.3701	28.3273	28.3132		0.10100
	B	28.3070	28.3647	28.3176		28.3070	0.10568
5	A	29.9521	30.2488	30.0341			0.49230
	B	29.9833	30.2588	30.0310		29.9769	0.51087
6	A	30.1110	30.1988	30.4081	30.1384		0.48891
	B	29.8608	30.1628	29.9577		29.9074	0.49016

Table A3: Volume determination results

Col	Trial	V_0 (cm^3, avg)	Std Dev V_0	Rel Std Dev	V_i (cm^3, avg.)	Std Dev V_i	Rel Std Dev
1	a	0.08951	0.00383	4.27%	0.04132	0.00976	23.62%
	b	0.09038	0.00102	1.13%	0.05574	0.00128	2.30%
	c	0.09139	0.00114	1.25%	0.01739	0.00214	12.28%
2	a	0.10396	0.00047	0.45%	0.08414	0.00211	2.51%
	b	0.10306	0.00048	0.46%	0.04998	0.00011	0.22%
3	a	0.09314	0.00040	0.42%	0.06143	0.00067	1.09%
	b	0.09456	0.00035	0.37%	0.07717	0.00051	0.66%
4	a	0.11028	0.00133	1.21%	0.10100	0.00333	3.30%
	b	0.10738	0.00500	4.65%	0.10568	0.01078	10.20%
5	a	0.54016	0.00369	0.68%	0.49230	0.01526	3.10%
	b	0.54291	0.00326	0.60%	0.51087	0.05707	11.17%
6	a	0.49324	0.00322	0.65%	0.48891	0.00968	1.98%
	b	0.49374	0.00114	0.23%	0.49016	0.00333	0.68%
7	a	0.09987	0.00558	5.59%	0.04457	0.00071	1.59%
8	a	0.10353	0.00277	2.68%	0.05835	0.00155	2.65%

Note: Standard deviations in this table were calculated from duplicate measurements.

Appendix B

PROPAGATION OF ERROR FOR THE POROSITY CALCULATIONS

For the interstitial volume formula, assuming that the masses of columns filled with solvents 1 and 2 were uncorrelated, we may use propagation of error methods as specified in Bevington, et al.[121] to determine:

$$V_0 = \frac{m_1 - m_2}{\rho_1 - \rho_2}$$

$$\frac{\sigma_{V_0}^2}{{V_0}^2} = \frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2} + \frac{\sigma_{\rho 1}^2 + \sigma_{\rho 2}^2}{(\rho_1 - \rho_2)^2}$$

Which can also be expressed to yield σ_{V_0} directly, given values of V_0 and errors associated with mass and density measurements:

$$\sigma_{V_0} = V_0 \sqrt{\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2} + \frac{\sigma_{\rho 1}^2 + \sigma_{\rho 2}^2}{(\rho_1 - \rho_2)^2}}$$

However, since values for density were obtained from the solvent manufacturers, we may neglect the contributions of the density values to the systematic uncertainty, and eliminate the second term under the radical to yield the final formula for σ_{V_0}:

$$\frac{\sigma_{V_0}^2}{{V_0}^2} = \frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}$$

$$\sigma_{V_0} = V_0 \sqrt{\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}}$$

By the same manipulations, we can also obtain expressions for V_i, the interstitial pore volume of the column, noting that the values of m_1, m_2, ρ_1, and ρ_2 in the below expression correspond to measurements taken to determine V_i rather than V_0:

$$V_i = \frac{m_1 - m_2}{\rho_1 - \rho_2}$$

$$\frac{\sigma_{V_i}^2}{{V_i}^2} = \frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}$$

And explicitly for σ_{V_i},

$$\sigma_{V_i} = V_i \sqrt{\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}}$$

In estimating the error associated with solving for pore volume, identified as V_p, we recall the relation established in Equation 3, which can be manipulated explicitly for V_p:

$$V_p = V_0 - V_i$$

Recalling the expression for error propagation through terms being added or subtracted (in a case with no multiplicative constants), we see that

$$\sigma_{V_p}^2 = \sigma_{V_0}^2 + \sigma_{V_i}^2 + 2\sigma_{V_0 V_i}^2$$

And again assuming that fluctuations in each of the volume calculations are uncorrelated, the error equation becomes:

$$\sigma_{V_p}^2 = \sigma_{V_0}^2 + \sigma_{V_i}^2$$

Solving explicitly for σ_{V_p} and plugging in the values obtained for errors of V_0 and V_i:

$$\sigma_{V_p}^2 = V_0^2 \left(\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}\right) + V_i^2 \left(\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}\right)$$

$$\sigma_{V_p} = \sqrt{V_0^2 \left(\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}\right) + V_i^2 \left(\frac{\sigma_{m1}^2 + \sigma_{m2}^2}{(m_1 - m_2)^2}\right)}$$

Where the references to m_1 and m_2 are for the measurements associated for V_0 and V_i and will not be the same in both terms of error associated with V_p.

Table B1: Volume determination results with propagation of error calculations.

Col	**Trial**	**V_0 (cm^3, avg)**	**Error (σ_{V0}, avg)**	**Relative SD (%)**	**V_i (cm^3, avg.)**	**Error (σ_{Vi})**	**Relative SD (%)**
1	a	0.08951	0.000502222	0.56%	0.04132	0.0003435	0.83%
	b	0.09038	0.000034630	0.04%	0.05574	0.0001105	0.20%
	c	0.09139	0.000281555	0.31%	0.01739	0.0003705	2.13%
2	a	0.10396	0.000092260	0.09%	0.08414	0.0000749	0.09%
	b	0.10306	0.000205732	0.20%	0.04998	0.0002007	0.40%
3	a	0.09314	0.000088160	0.09%	0.06143	0.0000699	0.11%
	b	0.09456	0.000047084	0.05%	0.07717	0.0001596	0.21%
4	a	0.11028	0.000034630	0.03%	0.10100	0.0000464	0.05%
	b	0.10738	0.000208949	0.19%	0.10568	0.0001223	0.12%
5	a	0.54016	0.000049024	0.01%	0.49230	0.0000444	0.01%
	b	0.54291	0.000316029	0.06%	0.51087	0.0001532	0.03%
6	a	0.49324	0.000044302	0.01%	0.48891	0.0000260	0.01%
	b	0.49374	0.000141487	0.03%	0.49016	0.0003361	0.07%
7	a	0.09987	0.006624896	6.63%	0.04457	0.0056116	12.59%
8	a	0.10353	0.003621220	3.50%	0.05835	0.0053513	9.17%