Identification and Quantitation of Cannabinoid Solute-Column Interactions Using the Hydrophobic-Subtraction Model to Predict Separation in High-Performance Liquid Chromatography

by

Heather Tunnell

July 2024

Director of Thesis: Jack E. Pender

Major Department: Chemistry

The Hydrophobic-Subtraction Model (HSM) has been widely used for identification of similar and orthogonal chromatography columns since its inception in 2002 but only had limited uses as a predictive tool. Recent refinement of the model has improved the prediction accuracy to a level that offers promise for use in high-performance liquid chromatography (HPLC) separation development. The HSM quantitatively describes the parameters affecting solute retention on HPLC columns including hydrophobic, steric, hydrogen-bond, cation exchange, and dipole interactions. To demonstrate the process, 16 cannabinoid compounds were chosen as the test solutes due to increasing interest and applications.

The HSM was used to identify and quantify the various cannabinoid solute retention parameters based on actual retention of each solute on a set of calibration columns with previously quantified column retention parameters. The resulting solute parameters were used to predict relative retention of each cannabinoid on over 550 columns in the HSM column database. A sorting system was then developed to rank the columns. Overall, this Thesis demonstrates the application of the HSM to cannabinoid compounds starting with identification of the best columns for the separation of these solutes.

Identification and Quantitation of Cannabinoid Solute-Column Interactions Using the Hydrophobic-Subtraction Model to Predict Separation in High-Performance Liquid Chromatography

A Thesis

Presented to the Faculty of the Department of Chemistry

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Chemistry

By

Heather Tunnell

July, 2024

Director of Thesis: Jack Pender, Ph.D.

Thesis Committee Members:

Jack Pender, Ph.D.

Eli Hvastkovs, Ph.D.

Eric Anderson, Ph.D.

Acknowledgements

I would first like to thank my mentor, Dr. Jack Pender, for giving me the best graduate student experience and teaching me everything I know about analytical chemistry. Thank you for always pushing your piles of important papers and never-ending to-do lists aside to help me with any research problem that I came across. Thank you for pushing me, cheering me on, giving me constructive criticism, letting me make mistakes, being my social outlet, and encouraging me to go forward regardless of the circumstances.

I would also like to thank my fiancé, Ryan, for always supporting me through the good and the bad days of undergraduate and graduate school. I could not have come this far without his love, advice, and care. Much love to my parents, all my siblings, and friends who kept me sane throughout these brief years of my life. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

CHAPTER 1: INTRODUCTION

1.1 Cannabinoid Overview

Cannabinoids are naturally occurring compounds that constitute the Cannabis sativa plant. These cannabinoids are biosynthesized in an acidic form within the plant tissues. However, when the herbal product is dried, heated, and stored, the acids decarboxylize which results in the creation of neutral cannabinoids.[1] The most commonly known cannabinoids include cannabidiol (CBD) and tetrahydrocannabinol (THC), with CBD being widely known for its "calming" properties, and THC being known for its "psychoactive" properties. Additional cannabinoids include, but are not limited to, cannabidiolic acid (CBDA), tetrahydrocannabinolic acid (THCA), delta-8-tetrahydrocannabinol (detla-8-THC), cannabigerolic acid (CBGA), cannabigerol (CBG), cannabichromeric acid (CBCA), and cannabinol (CBN).

These structures are featured in Figure 1.1, where an example degradation pathway of the aromatic oxygenated hydrocarbons is demonstrated.[2] The structures of these cannabinoids are generally carbocyclic, with approximately 20 carbon atoms that are usually formed by three rings: cyclohexane, tetrahydropirane, and benzene.[3] The similarity of these cannabinoid structures contributes to the challenge in the separation of these compounds in herbal products and poses a variety of analytical challenges requiring reliable, reproducible, and sensitive analytical methods and techniques.

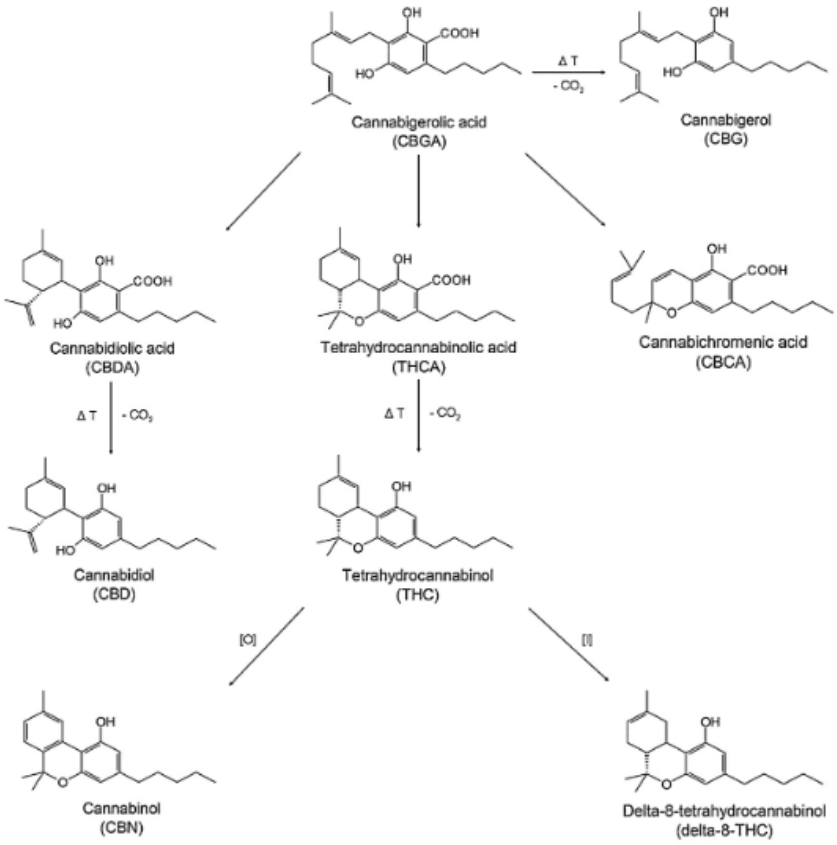


Figure 1.1 Example cannabinoid degradation pathway (ΔT = heating, [O] = oxidation, [I] = isomerization).[2] Other cannabinoids exist, but are not shown in the above figure.

With the cannabinoid market thriving, the demand for reliable analytical methods has increased dramatically to identify and quantitate these compounds.[4] Common ways to find a suitable separation range from adopting a published method, to starting from scratch by choosing a readily-available column and adjusting conditions such as mobile phase strength, column temperature, and pH.[5] Other modifications include increasing column length or using smaller particle sizes.[6] If the column is found to be unacceptable, the process continues on another column, then another, until a suitable separation is achieved. One drawback of this approach is the actual chemical differences of the columns are not often considered or utilized. This is a

missed opportunity since the stationary phase is a critical factor in the resulting retention and selectivity of various solutes.

Published methods are available and have been successful in separating various combinations of cannabinoid compounds. Many have been provided by column manufacturers to promote their products. Table 1.1 below shows some examples of alkyl-silica columns (C8, C18, etc.) that were recommended by various column manufacturers and independent authors for the separation of cannabinoids. Cannabinoids were chosen as a set of test compounds as these solutes would be beneficial to a wider audience and can be used for the identification and quantification of solute-column interactions to identify suitable columns.

Table 1.1. Manufacturers and authors that recommend specific columns and modes used for cannabinoid separation.

Column Name	**Manufacturer/Author**	**Mode**
Cortecs C18[7]	Waters	Gradient
Cortecs Shield RP18[8]	Waters	Isocratic
Accucore C8[9]	Thermo	Gradient
Kinetex C18[10]	Phenomenex	Isocratic
Ascentis Express C18[11]	Eurofins	Gradient
Ace-3 C18-Amide[12]	McRae, Melanson	Gradient
Ascentis Express C8[13]	Sigma Aldrich	Isocratic
Ascentis Express C18[13]	Sigma Aldrich	Gradient
Shimadzu Shim-pack XR-ODS II[14]	UNODC	Gradient
Waters UHPLC HSS[14]	UNODC	Gradient
InfinityLab Poroshell 120 EC-C18[15]	Agilent	Gradient

1.2 UPLC-PDA Analysis

In the published methods above, a variety of alkyl silica columns were used for cannabinoid separation, but each of these columns have different selectivities and produce

different chromatograms. To obtain information about column selectivity, liquid chromatography (LC) can be used. LC can be very effective in separating the components of a sample mixture. High-performance liquid chromatography (HPLC) and ultra-performance liquid chromatography (UPLC) both serve the same purpose of separating analytes in a mixture, as they can both be used for identification and quantitation. However, UPLC reduces the analysis run time, decreases the consumption of solvent, can handle higher pressures (15,000 psi) and has higher sensitivity.[16]

Regardless of which instrument is used, the theoretical principles remain the same. Reversed-phased HPLC utilizes a relatively polar mobile phase and nonpolar stationary phase. A schematic of an HPLC is shown in Figure 1.2. The process starts by mobile phase being drawn up from the solvent reservoir and pushed throughout the system by high pressure pumps. The sample is injected into the flowing mobile phase and is transported to the column containing the stationary phase. The components of the sample are separated in time based on the partitioning of the analytes between the mobile and stationary phases. The analyte's time spent inside the column depends on how strong its interactions are with that stationary phase. As analytes elute from the column, they are carried by the mobile phase to a detector which responds to a property of the analyte. The detector then converts the response into a digital signal consistently with time which is displayed as a resulting chromatogram. The response caused by each eluting compound is called a peak. The retention time of each peak on the chromatogram corresponds to the identity of each analyte, and the size of the peak is related to the amount of analyte present.

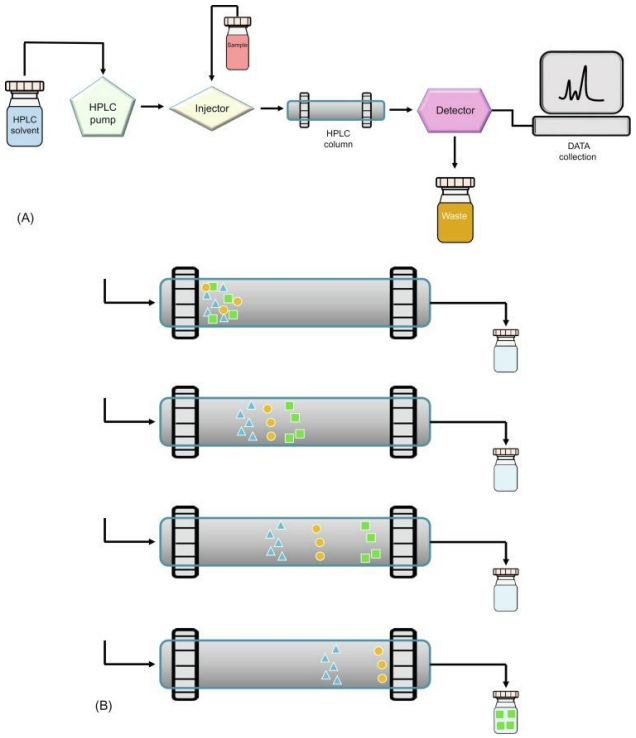


Figure 1.2. Schematic of the analyte's flow through an HPLC to produce chromatographic peaks.[17]

Many different detectors can be coupled to the HPLC or UPLC instrument to detect the analytes eluting from the column. A photodiode array (PDA) detector was chosen for the cannabinoid analyses because of its ease of use and ability to measure a range of wavelengths in real time (200-400 nm). Figure 1.3 below shows a diagram of a PDA optical system. A tungsten (W) and deuterated discharge (D_2) lamp serves as the light source, and that combined light is shown directly onto the flow cell. The light that passes through the flow cell is dispersed by the diffraction grating, and the intensity of the light is converted into the electrical signal for each wavelength via an array of photodiodes. The PDA spectrum of individual cannabinoids can be used to obtain a spectral absorbance profile which can be utilized as a form of analyte confirmation when they elute at similar times.[18] An ultra-violet (UV) absorbance detector was

chosen for this work due to sensitivity of the cannabinoid double bonds as well as cannabis labs often preferring the inexpensive and readily available UV absorbance detectors.

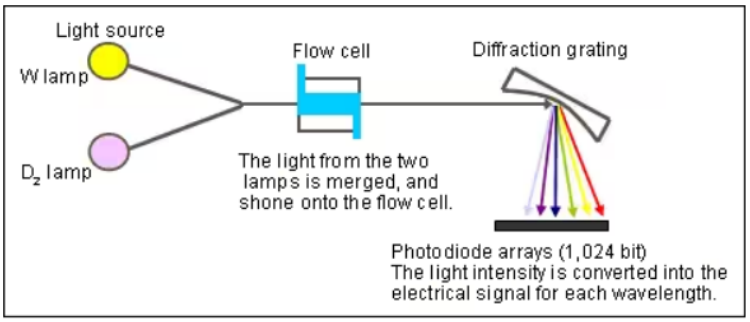


Figure 1.3. Diagram of a PDA optical system.[19]

1.3 Stationary Phases

As was shown in Figure 1.2, the analytes in the sample flow though the instrument via the polar mobile phase. They are then separated by and eluted from the column, and their signals are converted into chromatographic peaks by the detector and computer software. If the column was absent, then all the analytes would elute at the same time and no separation of the compounds would occur. Therefore, the HPLC column is one of the most important factors in obtaining sufficient separation, which then allows for correct identification and accurate quantitative calculations. Therefore, it is imperative to choose a column that would provide a suitable separation for the compounds of interest.

Reversed-phase HPLC columns are packed with porous silica particles that have a primary bonded phase, such as a nonpolar C_{18} ligand, attached. Figure 1.4 (left) below shows a schematic of the bare silica surface on a reversed-phase HPLC column. The bare silica surface is composed of silica atoms attached to alcohol groups and is therefore referred to as silica “silanols”. Here, the column has no reversed-phase selectivity. Figure 1.4 (right) displays how various ligands can be attached to the silanols. In this case, the column does have selectivity as the nonpolar alkyl C8

ligand will attract nonpolar solutes, resulting in a later elution of those solutes. In addition to alkyl-silica columns, there are also phenyl, cyano, amino and other phases. These can provide different selectivity to a variety of solutes, depending on the properties of the solute.

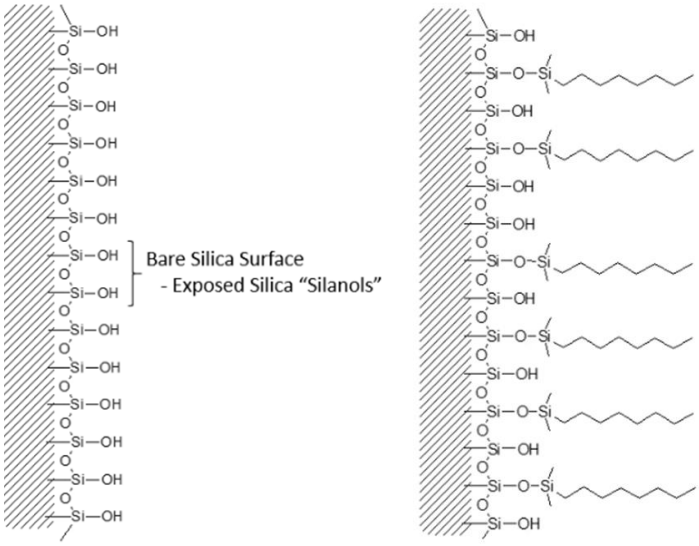


Figure 1.4. The bare silica surface of a stationary phase (left) and corresponding nonpolar carbon ligands that can be chemically bonded to the bare silanols (right).[20]

1.4 Retention and Selectivity

The overall goal of analytical HPLC is to separate the compounds of interest so that they can be identified and quantified. The peaks must be sufficiently separated to accomplish this. As shown in Figure 1.5 below, each peak in a chromatogram is described by a corresponding retention time. The retention time is the time that it takes for the analyte to flow through the instrument, interact with the column, and reach the detector. Here, t_r is the retention time of a solute and t_0 is the retention time of an unretained solute, such as thiourea. If the retention times of various solutes are compared across various HPLC columns and instruments, then a retention factor k can be calculated as shown in Eq. 1. The retention time of the unretained solute needs to

be included in the retention factor calculation because different columns and instruments have different dead times, or the time it takes to flow through the system from the injection to the detector when no column interaction is present. The retention factor is reported as multiples of the void time t_0 and accounts for the degree of retention of a solute on that column under the analysis conditions.

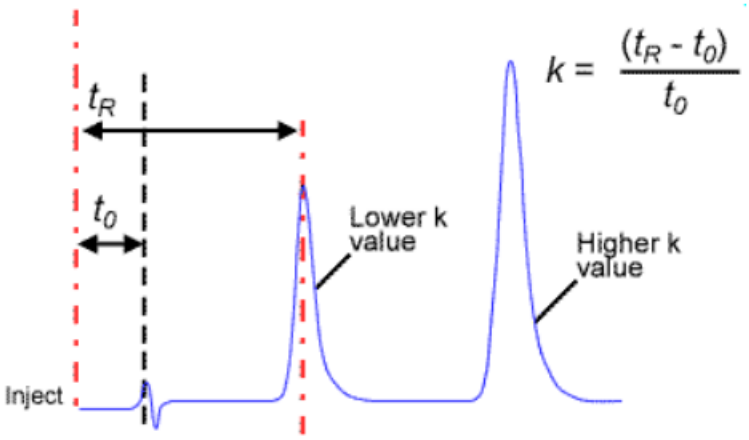


Figure 1.5. The calculation of the retention factor k for a solute using t_r and t_0.[21]

$$k = \frac{(t_r - t_0)}{t_0} \qquad \text{Eq. 1}$$

To compare the retention of two peaks, the retention factor ratio of the compounds, called the selectivity factor α, can be determined. Equation 2 displays the relationship between the retention factors of the solute and reference solute, and the selectivity factor. Equation 2 also sets these values on a logarithmic scale since these numbers are often quite small and cover a wide range. The selectivity factor can be described as the ability of the chromatographic system to 'chemically' distinguish between sample components and can be visualized as the difference between the apices of the two peaks, as shown in Figure 1.6. In a chromatogram, the least

resolved peaks are referred to as the critical pair. If the selectivity factors of the two solutes in the critical pair are similar, then it is likely that the peaks will not be separated. If the selectivity values of the two solutes in the critical pair are sufficiently different, then it is likely that the peaks will be sufficiently separated. It should be noted that resolution is the definitive measure of peak separation, considering both the retention time differences as well as peak width. The selectivity factor addresses the retention time differences only.

$$\log(\frac{k}{k_{ref}}) = \log \alpha \qquad \text{Eq. 2}$$

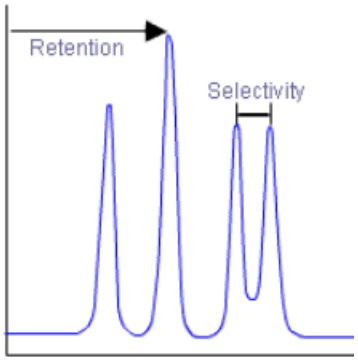


Figure 1.6. Retention and selectivity differences between two peaks.[21]

1.5 The Hydrophobic-Subtraction Model

The Hydrophobic-Subtraction Model (HSM) was created by Snyder and Dolan in 2002 to gain a better understanding of the factors that affect column selectivity.[22] In contrast to determining the selectivity factor for neighboring peaks, the HSM determines the selectivity for each compound relative to a constant reference compound, ethylbenzene. Ethylbenzene is used as the reference solute as its elution only reflects hydrophobic interactions. If the hydrophobic interactions are subtracted from the retention factor of the solute, then the remaining value would be the result of the interactions that are not primarily hydrophobic in nature. This would thus

help to explain the other types of solute-column interactions that exist, other than the primary hydrophobic interaction. Hence, the name Hydrophobic-Subtraction Model was deduced.

The relationship between the selectivity factor and these interactions is described by Equation 3, where log α can be used to quantify solute (η', σ', β', α', and κ') and column (H, S, A, B, and C) parameters. These parameters are representative of the solute's and column's chemical characteristics and how they will affect the selectivity. The solute parameters are unique to the solute of interest, whereas the column parameters are unique to a particular brand and phase of column. A database has been created which lists the H, S, A, B, and C column parameters for more than 600 different types of HPLC columns from more than 30 manufacturers.[23] The original purpose of this model was to develop a simple procedure for comparing the relative selectivity of any two columns, and to determine they were either equivalent or orthogonal by use of a single measure. This helped in the selection of equivalent replacement columns for routine assays or orthogonal columns for different selectivity in method development.

$$\log \alpha = \eta\text{'H} - \sigma\text{'S} + \beta\text{'A} + \alpha\text{'B} + \kappa\text{'C} \qquad \text{Eq. 3}$$

The products of these solute and column parameters, η'H, σ'S, β'A, α'B, and κ'C, describe various solute-column interactions which affect retention and selectivity. These terms are represented by hydrophobic interactions (η'H), steric selectivity (σ'S), hydrogen bonding between acceptor solutes and non-ionized silanols in the stationary phase (β'A), or hydrogen bonding between donor solutes and an unidentified proton acceptor in the stationary phase (α'B), and the attraction of protonated bases by ionized silanols (κ'C). Figure 1.7 shows a simplified representation of these contributions to HPLC retention and selectivity.

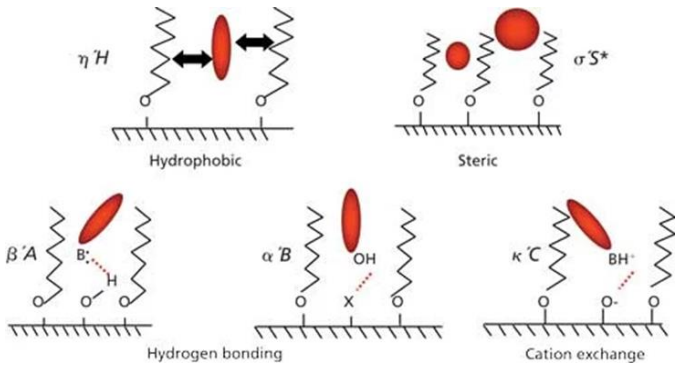


Figure 1.7. Solute-column interactions represented by the HSM.[22]

The first term of Equation 3 represents hydrophobic interactions. Hydrophobicity accounts for the general interaction of hydrophobic solutes with the hydrophobic stationary phase, and it is not surprising that it is the dominant contributor to HPLC retention. In fact, H is such a dominant factor that its contribution to retention overshadows the minor (but often more important from a selectivity standpoint) S, A, B, and C terms. Steric selectivity corresponds to a solute's ability to penetrate the particular bonded stationary phase. A more "bulky" solute will have a more difficult time penetrating and interacting with the phase than a less bulky solute.

For the β'A interaction, neutral proton acceptors (B:) interact with the silanols (-OH) on the stationary phase. Therefore, solutes that are better proton acceptors will be more strongly retained. Although the silica stationary surface is acidic by nature, it can still express basic characteristics with acidic solutes. This interaction is expressed by the α'B term. One of the key sources of column basicity is the interaction with vicinal silanol groups (X) on the silica surface. These groups comprise two adjacent silanol groups that are internally hydrogen bonded, which enables them to easily hydrogen-bond with acidic solutes as shown in Figure 1.8.[24] Figure 1.8(a) shows how vicinal silanols doubly hydrogen-bond with carboxylic acids to form a stable ring structure, whereas Figure 1.8(b) shows that phenolic solutes are less retained by vicinal silanols because of the single hydrogen-bond interaction. Finally, the more well-known cation-exchange

term represents the interaction between a protonated base and an ionized silanol on the stationary phase. In the κ'C term, κ' is a measure of the ionization state of the solute and C represents the column cation-exchange activity, which varies with mobile phase pH. The HSM Equation provides an explanation of how the solute parameters and the column parameters can be used to calculate and predict the selectivity, log α, of each solute on each column.

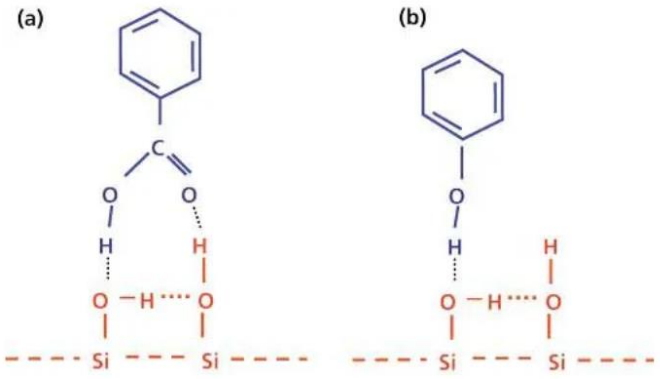


Figure 1.8. Schematic of the α'B interaction with a) carboxylic acids and b) phenolic solutes.[24]

While the HSM has been used widely, it has been recognized that it is not a global model. HSM, as it was originally formulated, did not have sufficient predictive accuracy to predict the selectivity of a solute set on a wide variety of columns. The original model was created using a limited type of solutes, as well as stationary phases of limited scope. Stoll et al. revised the original HSM model (HSM1) based on six parameters that were determined using principal component analysis (PCA) to create the HSM2 model as seen in Equation 4.[25] During the development of HSM2, it was found that the refined model improved prediction of retention factors, with the standard error being reduced from 1.0 to 0.35.[25]

$$\log \alpha = \eta'\mathrm{H} + \beta'\mathrm{A} + \alpha'\mathrm{B} + \kappa'\mathrm{C} + v\mathrm{V} + d\mathrm{D} \qquad \text{Eq. 4}$$

Stoll's revised model includes two different solute-column interaction terms, vV and dD. The vV term replaced the σ'S term and more accurately reflects the molecular volume of the solute. It was obtained from the Connolly solvent-excluded volume calculated using Chem3D for the HSM2 test compounds. The Connolly solvent-excluded volume is the volume that is described precisely by the surface defined by rolling a spherical solvent across the surface of the molecule. The dD term was also added to address the selectivity of dipolar and/or polarizable phases, such as cyano and phenyl phases. Additionally, a version of the linear solvation energy relationship (LSER) scale of hydrogen bond basicity and acidity replaced the HSM1 β' and α'parameters. Stoll's HSM2 column database converted all the column parameters from Snyder and Dolan's HSM1 database into the appropriate HSM2 column parameters after removing outliers and known errors. In total, the original database was reduced from 600+ columns to around 550 columns in the HSM2 database.[25]

Although the HSM1 was originally designed to identify similarities and differences of two columns,[26, 27] attempts have been made to use it for predicting columns for complementary separations by two-dimensional LC (2DLC). Because one of the major challenges in implementing 2DLC methods is the selection of suitable column pairs, Lindsey et al. proposed a computational screening method whereby virtual 2D chromatograms were calculated using HSM1.[28] Compared to other screening approaches, this method was highly predictive for column pairs that were able to resolve the largest number of analytes and showed a strong sensitivity to the choice of the second-dimension column. The success of this column selection method resulted in the development of 2DLC methods for pharmaceutical compounds using the HSM1 by Liu et al.[29] HSM solute parameters were experimentally determined for a small molecule pharmaceutical and its impurities. The predicted critical resolution values were compared for

both single-phase columns and tandem column pairs, and it was shown that tandem column pairs outperformed the single-phase columns.

There are only a few publications that utilize the HSM as a tool in method development. Zhang et al. selected suitable columns for the separation of beta-lactam antibiotics using the HSM1.[30] The goal was to specify a recommended range of these key parameters within which the critical pairs would be separated well. It was found that column parameter A was the most important factor affecting the separation of the critical pairs. However, this approach only considered critical pairs that were shown in their calibration columns and only a narrow range of C_{18} columns were used for predictions. Overall, this approach was mathematically complicated and would not be easily understood by method developers wishing to calculate suitable columns for other test compounds.

The most successful application of HSM for retention prediction and column choice was developed by Zhang and Hu to select optimal columns for clarithromycin impurity analysis using the HSM1.[31] This approach focused on the separation of three sets of critical pairs. This approach established an acceptable minimum log α value for each critical pair and used a calibration column selection process that normalized each column parameter by its H term. However, the authors forced the solute elution order to be the same throughout the column prediction process, which could result in the elimination of suitable columns that had a different solute elution order. Additionally, insufficient detail was given regarding the determination of the accepted minimum log α for each critical pair. This method also predicted elution on only C_{18} phases due to limitations of the original HSM. The approach presented in this Thesis builds upon the work of Zhang and Hu's demonstration of the potential for developing a separation focusing

on systematic selection of the column based on relative retention prediction using the HSM. This work also benefits from the improved prediction accuracy of the HSM2.

1.6 Scope of Thesis

This Thesis aims to show the application of the Hydrophobic-Subtraction Model to the separation of a cannabinoid solute set. These cannabinoid compounds were chosen here as test solutes due to the increasing public interest and potential future applicability. It is known that most analytical methods modify common HPLC variables such as temperature, mobile phase composition, pH, and gradient time to obtain a sufficient separation. However, our approach considers the column itself as a quantifiable variable and demonstrates a methodical approach to choosing a column. Therefore, the main goals of this Thesis were to develop a process for applying the HSM2 principles to:

1) Quantify and report the HSM2 solute parameters for 16 common cannabinoids.
2) Predict cannabinoid elution on more than 550 columns using the HSM2 Equation.
3) Identify columns with the best chance of cannabinoid separation.
4) Explain separation trends utilizing solute and column parameter differences.

In Chapter 2, I discuss the experimental procedure that was employed for the preparation and analysis of two sets of solutes: the HSM reference solutes and the cannabinoid solutes. Additionally, the process in which the calibration columns were selected is explained and justified. In Chapter 3, I present cannabinoid chromatographic data showcasing column separation differences. Moreover, I discuss how the verification of the HSM was performed and present direct numerical data of how the cannabinoid solute parameters were calculated through multiple linear

regression analysis. In Chapter 4, I support the validation of the model's predictability by comparing the predicted retention of the cannabinoid solutes to the experimental retention. In Chapter 5, I apply a sorting approach to identify possible columns with the best chances of cannabinoid separation, justify this selection using separation trends, and provide possible common HPLC variables that could be used to improve the separation of the solute set. I also acknowledge the limitations of the HSM and the approaches shown within. In Chapter 6, I discuss the conclusions of this Thesis and future areas of study to address the above limitations.

1.7 References

1) Atakan Z. Cannabis, A Complex Plant: Different Compounds and Different Effects on Individuals. *Ther Adv Psychopharmacol*., **2012**, *2*(6), 241-54.
2) De Backer et al., Innovative Development and Validation of an HPLC/DAD Method for the Qualitative and Quantitative Determination of Major Cannabinoids in Cannabis Plant Material. *J Chrom. B.* **2009**, *877*(32), 4115-4124.
3) Barrales-Cureño et al., Chemical Characteristics, Therapeutic Uses, and Legal Aspects of the Cannabinoids of Cannabis sativa. *Braz. Arch. Biol. Technol*. **2020**.
4) Brown et al., Leaner and Greener Analysis of Cannabinoids. *Anal. And Bioanal. Chem*. **2017**, 409, 3153-3163.
5) Sigma Aldrich. Factors Affecting Resolution in HPLC. https://www.sigmaaldrich.com/US/en/technical-documents/technical-article/analytical-chemistry/small-molecule-hplc/factors-affecting-resolution-in-hplc (accessed 7 Sep 2023).
6) Chromatography Today. How Do You Improve Separation in Chromatography? https://www.chromatographytoday.com/news/columns-gc/52/breaking-news/how-do-you-

improve-separation-in-column-chromatography/58691#:~:text=When%20you're%20using%20columns,column%20length%2C%20as%20discussed%20above (accessed Sep 7, 2023).

7) Van Tran, K., Twohig, M., Hudalla, C. Analysis of Cannabinoids in Cannabis Plant Materials and Edible Products Using Ultra-Performance Liquid Chromatography (UPLC) with PDA and Mass Detection. Waters Application Note. **2021**.
8) Aubin, A., Layton, C., Helmueller, S. Separation of 16 Cannabinoids in Cannabis Flower and Extracts Using a Reversed-Phase Isocratic HPLC Method. Waters Application Note. **2018**.
9) McRae, G., Melanson, J. Quantitation of Seventeen Cannabinoids in Dried Cannabis, Hemp and Vape Oils by LC-MS/MS. Thermo Scientific Application Note. **2021**.
10) Krepich, S., Orlowicz, S., Aqeel, Z., Layne, J. 18 Cannabinoids for Potency Testing By LC-UV. Phenomenex Application Note. **2020**.
11) Vaclavik, L., et. al. Determination of Cannabinoids in Plant Materials, Oils and Concentrates using UHPLC–DAD: Single Laboratory Validation for AOAC First Action Official Method Consideration. Eurofins Food Integrity and Innovation. **2020**.
12) McRae, G., Melanson, J. Quantitative Determination and Validation of 17 Cannabinoids in Cannabis and Hemp Using Liquid Chromatography-Tandem Mass Spectrometry. *Anal. Bioanal. Chem.* **2020**, 412:7381–7393.
13) Millipore Sigma. Analysis of 17 Cannabinoids in Hemp and Cannabis. Millipore Sigma Application Note. https://www.sigmaaldrich.com/US/en/technical-documents/protocol/analytical-chemistry/small-molecule-hplc/analysis-of-cannabinoids (accessed 25 May 2024).

14) Recommended Methods for the Identification and Analysis of Cannabis and Cannabis Products. United Nations Office on Drugs and Crime. **2022**.

15) Storm, C., Zumwalt, M., Macherone, A. Dedicated Cannabinoid Potency Testing in Cannabis or Hemp Products Using the Agilent 1220 Infinity II LC System. Agilent Application Note. **2020**.

16) Chromatography Today. HPLC vs. UPLC- What's the Difference? International Labmate Limited. **2021**.

17) Shen, C. Quantification and Analysis of Proteins. Diagnostic Molecular Biology **2019**. 187-214.

18) Shimadzu. UV vs. Diode-Array (PDA) Detectors for (U)HPLC. https://www.ssi.shimadzu.com/service-support/faq/liquid-chromatography/knowledge-base/uv-vs-pda-detectors/index.html (accessed 7 Sep 2023).

19) Hitachi. Principle and Feature of Various Detection Methods. www.hitachi-hightech.com/global/en/knowledge/analytical-systems/hplc/hplc-basics/course7.html (accessed 8 Sep 2024).

20) Shimadzu. High Performance Liquid Chromatography (HPLC) Basics. www.ssi.shimadzu.com/service-support/faq/liquid-chromatography/knowledge-base/hplc-basics/index.html (accessed 12 Feb 2024).

21) Crawford Scientific. The Theory of HPLC Chromatographic Parameters. https://www.inacom.nl/wp-content/uploads/2020/04/theory_of_hplc_chromatographic_parameters__1_.pdf (accessed 12 Feb 2024).

22) Snyder et al., Column Selectivity in Reversed-Phase Liquid Chromatography I. A General Quantitative Relationship. *J Chrom A*. **2002**, 961(2), 171-193.

23) HPLC Columns. Column Selectivity Database. https://hplccolumns.org/about/index.php (accessed 7 Sep 2023).

24) Dolan, J., Snyder, L. The Hydrophobic-Subtraction Model for Reversed-Phase Liquid Chromatography: A Reprise. LCGC North America. **2016**, 34(9), 730-741.

25) Stoll, D. et al. Improvements in the Predictive Accuracy of the Hydrophobic Subtraction Model of Reversed-Phase Selectivity. *J Chrom A*. **2021**, 1636, 461682.

26) Zhang, Y., Carr, P. A Visual Approach to Stationary Phase Selectivity Classification Based on the Snyder-Dolan Hydrophobic-Subtraction Model. *J Chrom A*. **2009**, 1216, 6685-6694.

27) Johnson, A., Johnson, C., Stoll, D., Vitha, M. Identifying Orthogonal and Similar Reversed-Phase Liquid Chromatography Stationary Phases Using the System Selectivity Cube and the Hydrophobic-Subtraction Model. *J Chrom A*. **2012**, 1249, 62-82.

28) Lindsey, R., Eggimann, B., Stoll, D., Carr, P., Schure, M., Siepmann, J. Column Selection for Comprehensive Two-Dimensional Liquid Chromatography Using the Hydrophobic-Subtraction Model. *J Chrom A*. **2019**, 1589, 47-55.

29) Liu, Z., Zhou, Y., Wang, Q., Foley, J., Stoll, D., Shackman, J. Development of Tandem-Column Liquid Chromatographic Methods for Pharmaceutical Compounds Using Simulations Based on Hydrophobic-Subtraction Model Parameters. *J Chrom A*. **2023**, 1695, 463925.

30) Zhang, W., Hu, Q., Zhang, X., Li, Y., Wang, M., Hu, C. The Selection of Suitable Columns for Reversed-Phase Liquid Chromatographic Separation of Beta-Lactam Antibiotics and

Related Substances via Chromatographic Column Parameters. *J Chrom A*. **2014**, 1323, 87-96.

31) Zhang, X., Hu, C. Selecting Optimal Columns for Clarithromycin Impurity Analysis According to the Quantitative Relationship of Hydrophobic Subtraction Model. *J Pharm Biomed Anal*. **2017**, 136, 162-169.

CHAPTER 2: EXPERIMENTAL

2.1 Materials

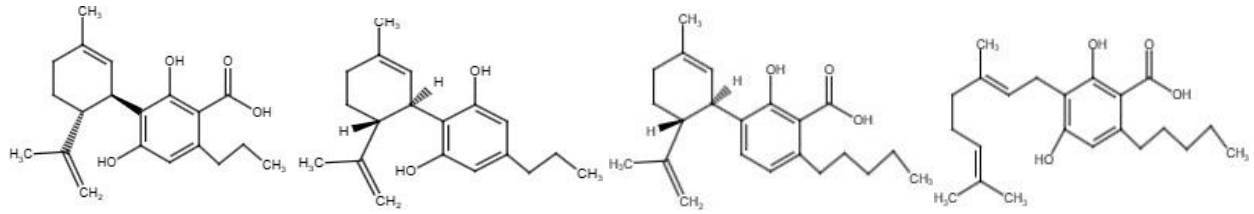
HPLC grade acetonitrile was from Aepex Chemicals. Potassium phosphate was from Acros Organics, and phosphoric acid was from Fisher. The cannabinoid solute standards were purchased from Supelco, Restek. The HSM reference solutes were purchased and/or provided from Tokyo Chemical Industry (TCI) and East Carolina University's organic chemistry chemical inventory. HPLC columns used for model verification were newly purchased from a variety of manufacturers including Agilent, Phenomenex, and Waters. The best-chance columns were acquired by either donation from Waters Corporation or other institutions, loaned out from local businesses, or found within East Carolina University's analytical HPLC column inventory.

2.2 Cannabinoid Marker Solutions

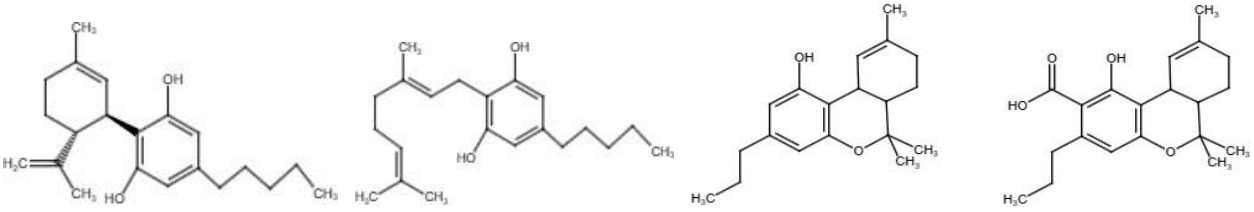
The cannabinoid solutes were purchased as 1000 μg/mL solutions in acetonitrile. To decrease run time and mobile phase usage, solute "mixes" were prepared based on a Phenomenex analysis of a cannabinoid mixture.[1] In the published Phenomenex analysis of the cannabinoid compounds, individual cannabinoid solutes were injected onto a Phenomenex Kinetex C18 column, and a separation was obtained. Two solutes that had large retention time differences in the published chromatogram were included in a solute mix here. For the 17 cannabinoid solutes listed below in Table 2.1, nine mixes were prepared by adding 7.5 μL of the two 1000 μg/mL solutions to a vial and diluting with acetonitrile to make 50 μg/mL marker solutions. The structures of these cannabinoid solutes are displayed in Figure 2.1.[2]

Table 2.1. Cannabinoid solutes and corresponding abbreviations.

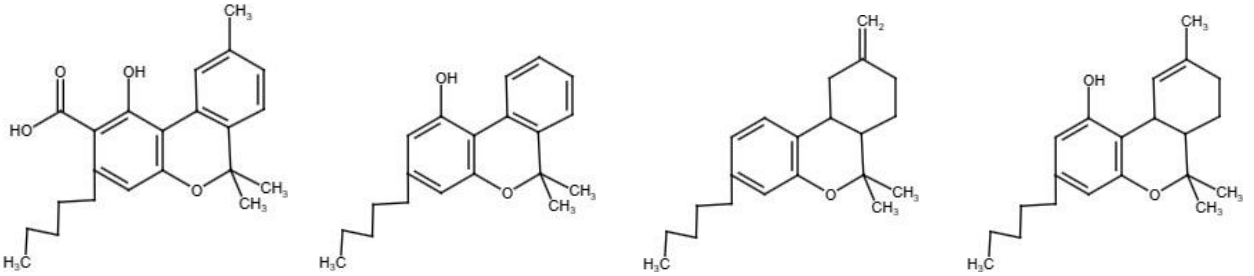
Cannabinoid Solute	**Abbreviation**
Cannabidivarinic Acid	CBDVA
Cannabidivarin	CBDV
Cannabidiol Acid	CBDA
Cannabigerolic Acid	CBGA
Cannabigerol	CBG
Cannabidiol	CBD
Tetrahydrocannabivarin	THCV
Cannabinolic Acid	CBNA
Cannabinol	CBN
exo-Tetrahydrocannabinol	exo-THC
Delta-9-Tetrahydrocannabinol	Delta-9-THC
Delta-8-Tetrahydrocannabinol	Delta-8-THC
Delta-9-Tetrahydrocannabinolic Acid A	THCA-A
Cannabichromenic acid	CBCA
Cannabichromene	CBC
Cannabicyclol	CBL
Cannabicyclolic acid	CBLA



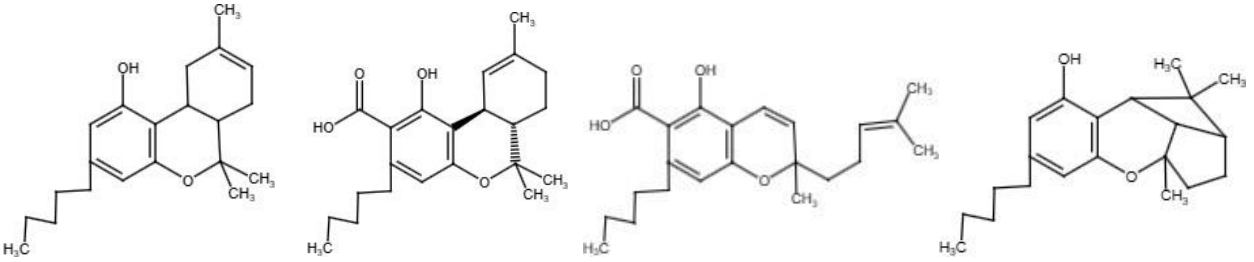
CBDVA CBDV CBDA CBGA



CBD CBG THCV THCVA



CBNA CBN Exo-THC D9-THC



D8-THC THCA-A CBCA CBL

CBC CBLA

Figure 2.1. Cannabinoid chemical structures.[2]

2.3 HSM Reference Solute Marker Solutions

The verification of the published column parameters was performed before attempting to calculate the cannabinoid solute parameters. This was done using the same approach used to create the column parameters in the database. To do this, 16 reference solutes shown in Table 2.2 were used. These 16 reference solutes were chosen by the developers of the HSM1 model, Snyder and Dolan.[3] Each solute possesses different characteristics which allows for the various interactions shown in the HSM Equation to be showcased.

Table 2.2. The 16 reference solutes and corresponding solute parameters η', σ', β', α', and κ'. Lighter highlighted cells indicate smallest solute parameters, and darker highlighted cells indicate largest solute parameters.

Solute	**η' (hydro-phobic)**	**σ' (steric)**	**β' (basic)**	**α' (acidic)**	**κ' (cation-exchange)**
Amitriptyline	-1.094	0.163	-0.041	0.300	0.817
n-butylbenzoic acid	-0.266	-0.223	0.013	0.838	0.045
N,N-diethylacetamide	-1.390	0.214	0.369	-0.215	0.047
5-phenylpentanol	-0.495	0.136	0.030	0.610	0.013
Ethylbenzene	0	0	0	0	0
N,N-dimethylacetamide	-1.903	0.001	0.994	-0.012	0.001
5,5-diphenylhydantoin	-0.940	0.026	0.003	0.568	0.007
Toluene	-0.205	-0.095	0.011	-0.214	0.005
Nortriptyline	-1.163	-0.018	-0.024	0.289	0.845
Acetophenone	-0.744	0.133	0.059	-0.152	-0.009
Mefenamic Acid	0.049	0.333	-0.049	1.123	-0.008
p-nitrophenol	-0.968	0.040	0.009	0.098	-0.021
Anisole	-0.467	0.062	0.006	-0.156	-0.009
Benzonitrile	-0.703	0.317	0.003	0.080	-0.030
Cis-chalcone	-0.048	0.821	-0.030	0.466	-0.045
Trans-chalcone	0.029	0.918	-0.021	-0.292	-0.017

All HSM reference solutes were diluted to 1000 ug/mL before being diluted to 50 μg/mL marker solutions. In addition to the 16 reference solute names, Table 2.2 also shows the published solute parameters for each solute. The lighter highlighted cells represent the lowest solute parameter values, and the darker highlighted cells represent the highest solute parameter values. For example, trans-chalcone was selected because of its bulky structure, and therefore it has the highest σ' solute parameter out of all the solutes. It should be noted that the solute parameters for the ethylbenzene reference solute are zero because its retention factor is subtracted from all the other solutes' retention factors, as seen in Eq. 2. Additionally, there are no solute parameters for thiourea as its only purpose is to account for the column and HPLC dead time. Even though these solute parameters shown here are from HSM1, they should suffice when verifying the model. For the purpose of using the solute parameters for predicting elution order in the future, the revised HSM2 parameters will be used.

2.4 Calibration Columns

The HPLC columns that were chosen to build the cannabinoid prediction model were selected based on their relative column selectivity parameters to provide a wide range of H, S, A, B, and C for the solute parameter calculations. The ratios of S/H, A/H, B/H, and C/H were used to compare the selectivity of the various stationary phases rather than the absolute values of H, S, A, B, and C. This is because differences in stationary phase selectivity only exist when the ratios of the phase coefficients differ, not when their absolute values differ.[4] Therefore, the columns were selected based on their relative column parameters (X_i). The relative column parameters of more than 100 columns from Agilent, Phenomenex, and Waters were calculated using Equation 5, where *I* represents S/H, A/H, B/H, or C/H ratios and Φ_i represents the weighting factors of

each column parameter.[5] These manufacturers were chosen because of their column popularity and wide availability. The names and parameters for these columns are shown in Table 2.3, with the lighter shade representing the minimum column parameter across all columns, and the darker shade representing the maximum column parameter. The dimensions for all columns were 4.6 x 150 mm, ranging from 2.6 to 5-micron particles which are similar to those used to create the HSM database. The columns were allowed to statically equilibrate in the mobile phase overnight.

$$X_i = (I - I_{min}) * \Phi_i \qquad \text{Eq. 5}$$

Table 2.3. Calibration column names and corresponding database column parameters H (hydrophobic), S (steric), A (acidic), B (basic), and C (cation exchange, pH 2.8). Lighter highlighted cells indicate smallest column parameters, and darker highlighted cells indicate largest column parameters.

Column Name	**Type**	**H**	**S**	**A**	**B**	**C**
Agilent InfinityLab Poroshell 120 SB-AQ	EP	0.581	-0.120	-0.133	0.051	-0.014
Agilent InfinityLab Poroshell 120 EC-CN	CN	0.421	-0.057	-0.476	0.002	0.045
Agilent Zorbax Eclipse PAH	B	1.030	-0.010	0.680	-0.050	0.070
Agilent Zorbax Eclipse Plus C8	B	0.880	0.010	-0.170	0.000	-0.04
Agilent Zorbax SB-C8	B	0.790	-0.070	0.130	0.010	0.010
Agilent Zorbax SB-C18	B	0.990	-0.030	0.260	0.000	0.130
Agilent Zorbax SB-Phenyl	Phenyl	0.620	-0.160	0.060	0.030	0.030
Agilent Zorbax Eclipse XDB-CN	CN	0.450	-0.060	-0.310	0.000	0.070
Agilent Zorbax StableBond 300A C18	B	0.900	-0.050	0.040	0.040	0.250
Agilent Zorbax Eclipse XDB-Phenyl	Phenyl	0.660	-0.120	-0.240	0.010	0.060
Phenomenex Kinetex C18 100A	B	0.960	0.000	-0.130	-0.010	0.000
Phenomenex HyperClone MOS C8 120A	EP	1.050	0.060	0.010	-0.020	-0.300
Waters SymmetryShield C8	EP	0.840	-0.050	0.040	0.130	1.140

2.5 Chromatography Conditions

A Waters Acquity UPLC equipped with a PDA detector was utilized for chromatographic analysis. For the cannabinoid solutes, the following experimental conditions were modified from a published analysis of a cannabinoid mixture.[1] Isocratic elution was performed using a 30/70 ratio of 10 mM potassium phosphate buffer at pH 2.9/acetonitrile mobile phase with a flow rate of 1.5 mL/min and an injection volume of 1.0 μL. The columns were used at ambient temperature (~20˚C). For the reference solute verification analysis, the original HSM experimental conditions were used.[3] This included a 50/50 ratio of 30 mM potassium phosphate buffer at pH 2.8/acetonitrile with a flow rate of 2.0 mL/min and an injection volume of 10.0 μL. The column temperature was set to 35 ˚C. This mobile phase was prepared using gravimetric addition for maximum mobile phase consistency between batches.

2.6 Data Analysis

Empower 3 software was used for data acquisition and qualitative chromatogram processing. Statistical and multiple linear regression analysis were performed using the Data Analysis ToolPak in Excel for Microsoft 365 MSO, version 2401.

2.7 References

1) Krepich, S., Orlowicz, S., Aqeel, Z., Layne, J. 18 Cannabinoids for Potency Testing By LC-UV. Phenomenex Application Note. **2020**.
2) Radwan, M., Chandra, S., Gul, S., ElSohly, M. Cannabinoids, Phenolics, Terpenes and Alkaloids of Cannabis. *Molecules* **2021**, 26, 2774.

3) Snyder et al., Column Selectivity in Reversed-Phase Liquid Chromatography I. A General Quantitative Relationship. *J Chrom A*. **2002**, 961(2), 171-193.
4) Zhang, Y., Carr, P. A Visual Approach to Stationary Phase Selectivity Classification Based on the Snyder–Dolan Hydrophobic-Subtraction Model. *J Chrom A*. **2009**, 1216(39), 6685-6694.
5) Zhang, X., Hu, C. Selecting Optimal Columns for Clarithromycin Impurity Analysis According to the Quantitative Relationship of the Hydrophobic-Subtraction Model. *J Pharm Biomed Anal.* **2017**, 136, 162-169.

CHAPTER 3: CANNABINOID SOLUTE PARAMETER DETERMINATION

3.1 Model Verification Using HSM Reference Solutes

Before the cannabinoid solutes were evaluated, the HSM column parameters were verified in-house to confirm the ability to replicate the published HSM2 column parameter database. To do this, the 16 HSM reference solutes were injected onto the UPLC, and chromatograms were produced for each solute on each column. An example chromatogram is shown below in Figure 3.1, which displays the retention of these HSM reference solutes on the Zorbax Eclipse Plus C8 column.

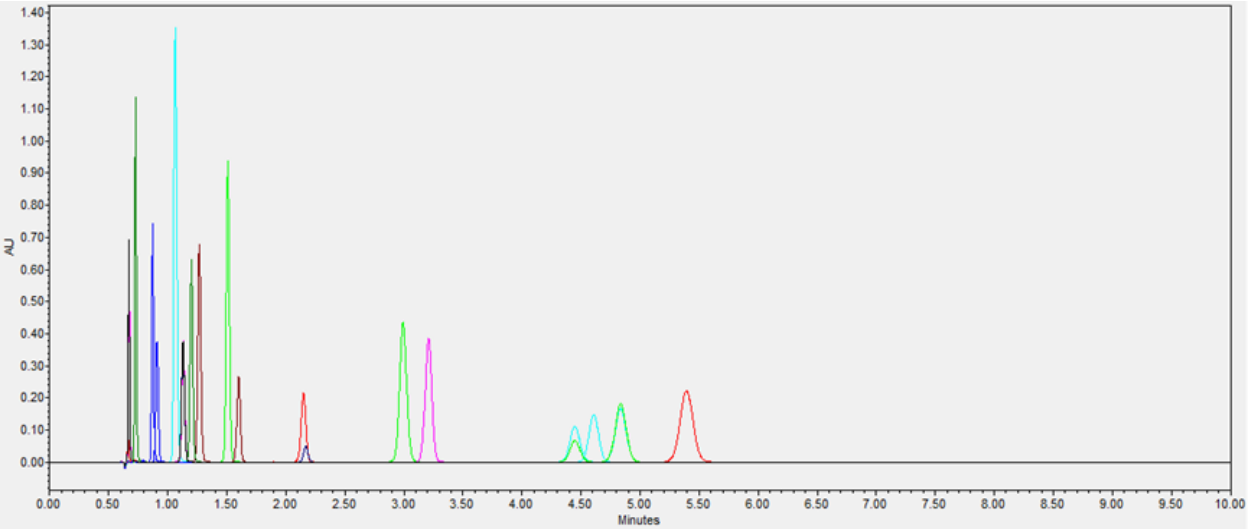


Figure 3.1. Chromatogram of 16 HSM reference solutes on the Zorbax Eclipse Plus C8 column.

Referring to the HSM Equation (Eq. 3), it is clear that if the solute parameters (η', σ', β', α', and κ') and their log α values on those columns are known, then the column parameters (H, S, A, B, and C) can be calculated. To utilize the relationship displayed by the HSM Equation, the retention times of each solute were converted into retention factors (Eq. 1), then into log α values (Eq. 2). The solute parameters of the 16 HSM reference solutes have previously been calculated, as seen in Table 2.2, and published by the original HSM authors.[1]

The column parameters were calculated by performing a multiple linear regression of all the solutes' log α values and their solute parameters. This was done using log α as the dependent variable and the five solute parameters as the independent variables, with the y-intercept being set to zero.[2] The resulting coefficients from the regression represented the experimental column parameters that were produced in-house. These were then compared to the published database column parameters. To determine if the experimental and published column parameter values were "similar", the 3% change limit was used and is shown in Table 3.1.[1] This limit was established by the original authors as a measure of similarity. Here, Δ(allowed) is the allowed change in H, S, etc. for a maximum change in log α by 0.004 (equal to 1% in α).

Table 3.1. Effect of a change in column parameters on separation.[1]

Δ(allowed)	H	S	A	B	C
Standard Deviation	0.0005	0.0041	0.0012	0.0061	0.00326
For 1% change in α	0.080	0.010	0.033	0.007	0.012
For 3% change in α	0.240	0.029	0.100	0.020	0.037

Most of the columns were determined to be similar to the database values, which confirmed the ability to execute the HSM procedure. However, a few columns produced column parameters determined to be "different" from the database values. These included the Agilent Poroshell 120 Bonus-RP, Agilent Zorbax Bonus-RP, Agilent Zorbax Extend-C18, Phenomenex Luna CN, and Phenomenex Kinetex PFP columns. This discrepancy is unsurprising, since it is known that the HSM1 model does not account well for all solute-column interactions that exist outside of the interactions that occur in alkyl-silica columns. It is known that the model's standard error for non-alkyl-silica columns is generally larger. Some of the differences can be

attributed to manufacturer reproducibility issues as well.[2] The above columns were subsequently eliminated from the column set and are not listed in Table 2.3. It can be seen in Table 2.3 that some non-alkyl-silica columns are still included in the model's column list. This is because, even though those columns had larger column parameters differences, they were still in the range of being acceptable. The eliminated columns were then replaced with additional alkyl-silica columns: the Phenomenex HyperClone MOS C8 and the Waters SymmetryShield C8. This was done to ensure a sufficient number of columns were included to build a reliable multiple linear regression model.

After injecting the HSM solutes and performing multiple linear regression on these new columns, their experimental column parameters were compared to the database, and it was determined that they were similar. These new columns were therefore included in the calibration column set. Overall, it was found that the HSM model and the multiple linear regressions could be performed accurately and interpreted correctly, which gave confidence that the cannabinoid solute parameters could be reliably calculated in the next step.

3.2 Calculation of Cannabinoid Solute Parameters

The cannabinoid markers were injected onto all 13 calibration columns. All the resulting chromatograms were then analyzed, and each solute's identity was confirmed using UV spectral analysis for the PDA. Example chromatographic overlays of the cannabinoid peaks on the Zorbax Eclipse Plus C8 and Zorbax SB-Phenyl columns are shown in Figures 3.2 (a) and (b) to display the differences in column retention and selectivity. Calculations were then performed to determine the cannabinoid solute parameters. The retention times were converted into retention factors and the log α values were determined. Multiple linear regressions were performed for

each solute across all columns, with the dependent variable being log α, the independent variables being the six HSM2 column parameters, and the y-intercept not forced to zero.[3]

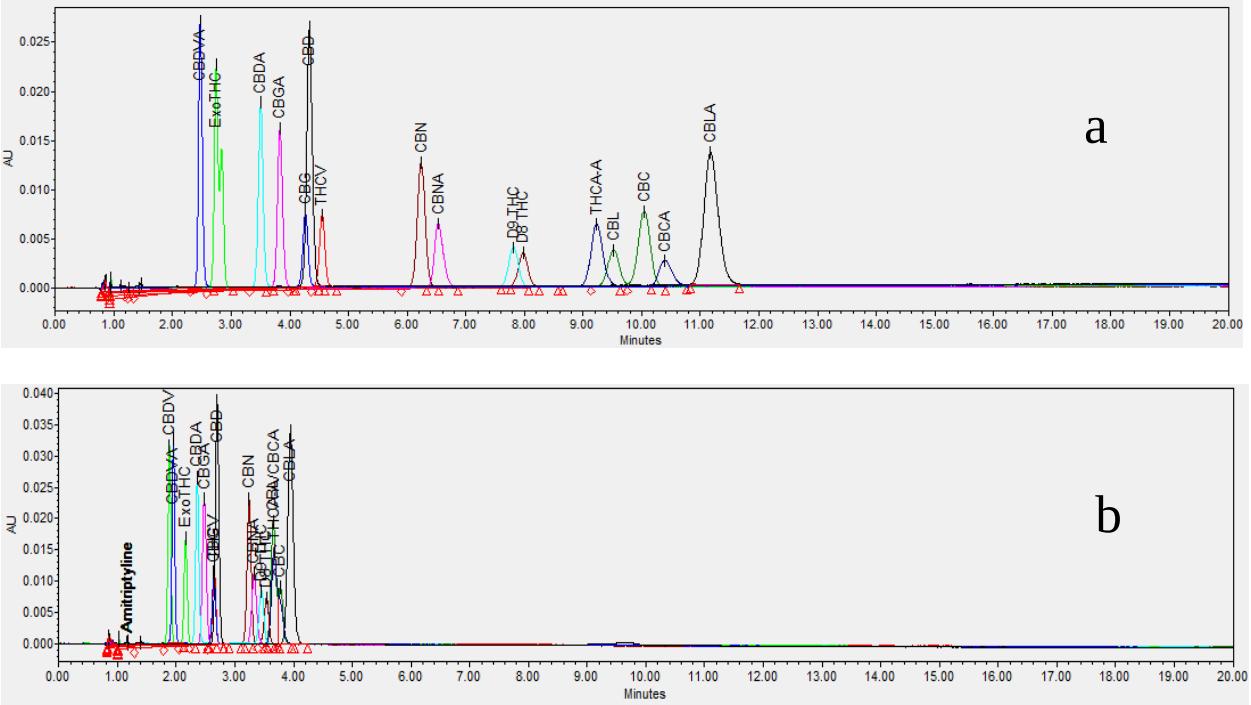


Figure 3.2. Example chromatograms of the cannabinoids on the a) Zorbax Eclipse Plus C8 column and b) Zorbax SB-Phenyl column.

Since the goal was to eventually use these cannabinoid solute parameters in the elution prediction calculations of the cannabinoids, the more reliable HSM2 column parameters were used. This was done only after verifying that the HSM2 column parameter values could be obtained using Stoll's HSM2 solute parameters for the HSM reference solutes. Column parameters for five out of the 13 columns chosen for the calibration column set were not available in the HSM2 column database. Therefore, experimental HSM2 column parameters were determined in-house and used in subsequent calculations. The values are presented in Appendix Table A.1. Once the multiple linear regressions of the HSM2 column parameters and their respective log α values were complete, coefficients were displayed from the model output, which represented the cannabinoid solute parameters. These cannabinoid solute parameters,

along with the y-intercepts associated with each solute, are shown in Table 3.2. Note that the solute parameters for CBCA are not included in Table 3.3 because its retention time could not be determined on the HyperClone MOS C8 column. This ultimately excluded CBCA from the solute list.

Table 3.2. Cannabinoid solute parameters.

Solute	**y-int**	**h**	**b**	**a**	**k**	**v**	**d**
CBD	-0.114	0.500	0.205	0.560	-0.058	0.564	0.263
CBLA	-0.374	1.354	0.075	1.541	0.006	0.866	0.477
CBDVA	-0.054	0.071	0.293	1.571	-0.012	0.703	1.485
CBG	-0.092	0.463	0.239	0.998	-0.066	0.739	0.658
CBDV	-0.025	0.122	0.296	0.563	-0.050	0.571	0.488
EXOTHC	-0.358	0.451	0.556	0.888	0.007	1.459	0.378
CBDA	-0.107	0.382	0.240	1.587	-0.047	0.714	1.284
D9THC	-0.329	1.075	0.231	0.554	-0.048	0.930	-0.208
CBGA	-0.075	0.410	0.258	2.241	-0.022	0.894	1.856
CBNA	-0.267	0.968	0.217	1.632	-0.023	0.790	0.949
CBN	-0.248	0.871	0.232	0.586	-0.062	0.764	0.031
D8THC	-0.342	1.102	0.299	0.635	-0.054	1.054	-0.191
THCA-A	-0.425	1.302	0.231	1.322	0.002	1.099	0.366
CBL	-0.303	1.144	0.308	0.861	-0.042	1.213	-0.169
CBC	-0.400	1.286	0.216	0.851	-0.035	1.166	-0.202
THCV	-0.230	0.658	0.214	0.449	-0.045	0.668	0.040

3.3 References

1) Gilroy, J., Dolan, J., Snyder, L. Column selectivity in reversed-phase liquid chromatography: IV. Type-B alkyl-silica columns. *J Chrom A.* **2003**, 1000(1-2), 757-778.
2) Liu, Z., Zhou, Y., Wang, Q., Foley, J., Stoll, D., Shackman, J. Development of Tandem-Column Liquid Chromatographic Methods for Pharmaceutical Compounds Using Simulations Based on Hydrophobic-Subtraction Model Parameters. *J. Chrom A.* **2023**, 1695, 463925.

3) Personal communication with John Dolan.

CHAPTER 4: PREDICTABILITY OF THE MODEL

4.1 Predicted Log α on Calibration Columns

The cannabinoid solute parameters and the database column parameters were used to predict the log α of each solute on all 13 calibration columns. This was done to evaluate the accuracy of the predicted log α values compared to the actual log α values as a best-case scenario. Comparing the predicted and actual log α values on the calibration columns verifies that the model would predict on future columns with high accuracy.

To do this, the HSM2 equation (Eq. 4) was utilized. The solute parameters of each cannabinoid solute were multiplied by the respective database column parameters. These products were then added according to Eq. 4 to yield a predicted log α. The y-intercepts that resulted from the multiple linear regression for each solute's solute parameters were then added to this sum for each column. The predicted and actual log α values for all the cannabinoid solutes were then plotted for each column. An example of the correlation is shown in Figure 4.1 for the Waters SymmetryShield C8 column. Here, a regression equation and determination coefficient (R^2) for each column was obtained. An R^2 and regression coefficient that both approximated one, as well as a constant term that approached zero, indicated an accurate MLR model.[1] The model was determined to accurately predict the values of log α on all calibration columns.

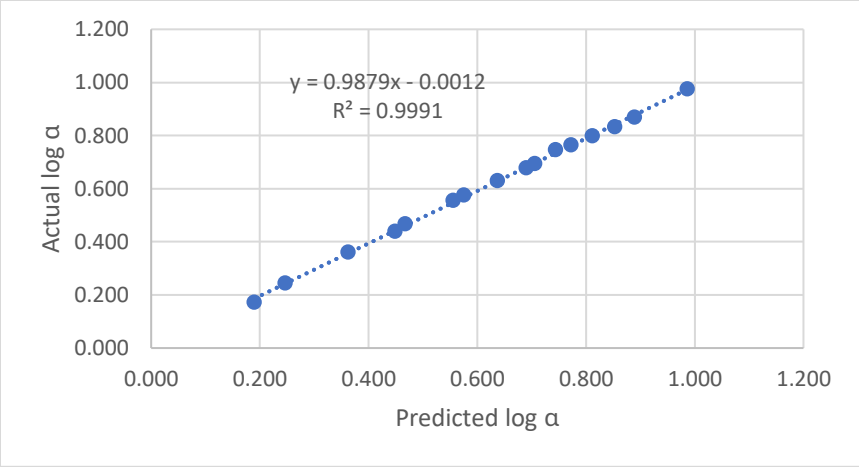


Figure 4.1. Predicted versus actual log α correlation on the Waters SymmetryShield C8 column.

4.2 Standard Error of the Regression

In order to make decisions about the relative separations of the cannabinoids on the database columns, the error in the predictions needed to be determined. To estimate the error associated with the predictions, the standard error of the estimate was calculated using Equation 6 on the actual errors observed in the calibration column data set. Here, s_{est} is the standard error of the estimate, $\hat{y}$ is the predicted value, y is the actual value, and n is the number of observations.[2] For each solute on each column, the difference between the predicted and actual log α was calculated then squared. Once these differences were summed across all the columns in the column set and the number of observations was determined, the standard error of the estimate was calculated. Figure 4.2 shows a histogram of the error associated with the predictions for all solutes across the 13 calibration columns. The shape reasonably estimates a normal distribution.

$$s_{est} = \sqrt{\frac{\Sigma(\hat{y}-y)^2}{n-2}} \qquad \text{Eq. 6}$$

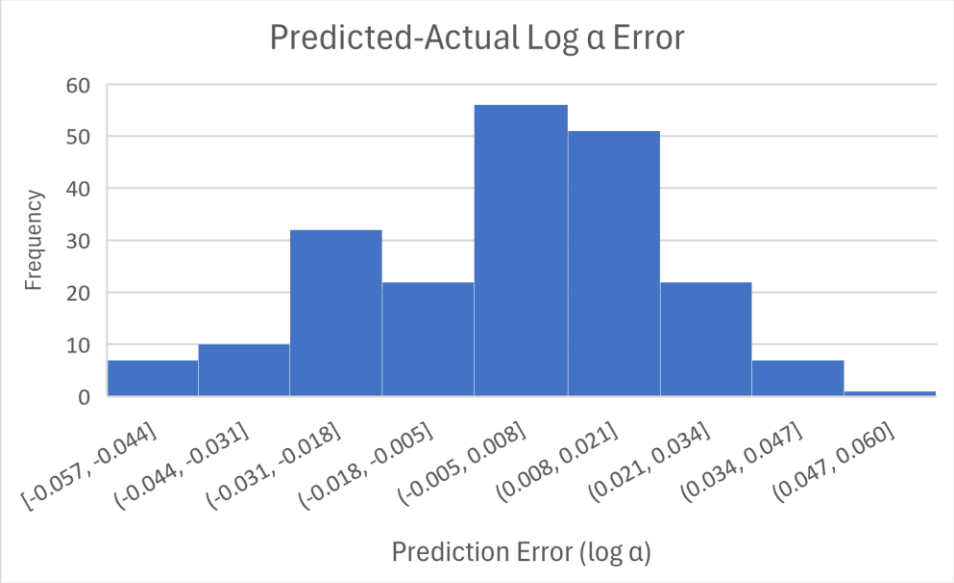


Figure 4.2. Histogram of the prediction error.

The result for the standard error of the regression was 0.0215 log α units. This value only accounts for ±1 standard deviation (SD), or approximately 68% of the variation expected in prediction accuracy. Multiplying this result by two yields 0.0431 log α units, which accounts for ±2 SDs, or approximately 95% of the variation expected in prediction accuracy. Therefore, there is a 95% chance that the actual log α of a single solute will be within this predicted range. It should be noted that some cannabinoid solutes are more or less reliable than the average error of the model. This 0.0431 log α value was also used as a minimum log α limit between neighboring solutes. In other words, two adjacent solutes must have a log α difference greater than 0.0431 to be considered "separate". To provide more leeway for the separation of the neighboring solutes, the 0.0215 log α standard error was multiplied by 1.5 to give 0.0323 log α, which accounts for ±1.5 SDs or approximately 87% of the variation expected in prediction accuracy. This approach is discussed in more detail in Chapter 5.

4.3 References

1) Zhang, X., Hu, C. Selecting Optimal Columns for Clarithromycin Impurity Analysis According to the Quantitative Relationship of Hydrophobic Subtraction Model. *J Pharm Biomed Anal.* **2017**, 136, 162-169.
2) Gemperline, P., et al. Interpreting Regression ANOVA Tables. In *Practical Guide to Chemometrics*. 2nd Edition; Taylor & Francis: Boca Raton, **2006**; 127-130.

CHAPTER 5: PREDICTING COLUMNS WITH BEST CHANCE OF SEPARATION

5.1 Selection of the Best-Chance Column

Confident that the model can accurately predict the log α values for each of the cannabinoids within a reasonable amount of error, the elution of the solutes was predicted on a wide variety of columns. Since the revised HSM2 improved the predictive accuracy of the model and rationalized the selectivity of new stationary phases, the new HSM2 HPLC column database developed by Stoll was utilized. This database includes the updated column parameters of HSM2 (V and D) for approximately 550 HPLC columns.

All the column parameters included in the HSM2 column database were multiplied by the respective cannabinoid solute parameters. These products, in addition to the y-intercept associated with each solute, were summed to yield a predicted log α for each solute on each column. The log α values were then sorted lowest to highest predicted log α to give the predicted elution order, then the predicted difference of log α between each neighboring solute was calculated. Finally, it was determined whether the differences between each neighboring solutes' log α was greater than the 0.0431 and/or the 0.0323 log α limit. Ideally, the best column would be one in which all the neighboring solutes have a predicted difference greater than 0.0431 log α units. Based on the columns that had the highest predicted number of solute pairs sufficiently separated from one another, the column(s) with the best chance of separation was selected.

The column list was primarily sorted such that the columns whose predicted log α difference between each neighboring solute was greater than 0.0431 were the top priority.

A secondary sorting was implemented to determine the number of solute pairs that had predicted log α differences greater than 0.0323. This way, if some of the solute pairs had a log α difference slightly under 0.0431, these would be accounted for in the 0.0323 sorting. Table 5.1 lists column names, manufacturers, silica types, and how many neighboring solute pairs had predicted differences greater than 0.0431 and 0.0323 log α units for each column.

Table 5.1. Example ranking information.

Ranking	**Column Name**	**Manufacturer**	**Type**	**log α Δ > 0.0431**	**log α Δ > 0.0323**
1	Adsorbosphere C18	Grace/Alltech	A	13	13
6	Resolve C18	Waters	A	11	13
8	Cortecs UPLC Shield RP18	Waters	EP	11	12
16	Accucore XL C18	Thermo/Hypersil	B	10	12
165	Cortecs C18	Waters	B	8	11

As seen in the table, the Adsorbosphere C18 Type A column was predicted to be the column with the best chance of separating the cannabinoid solutes, with 13 solute pairs exceeding the 0.0431 and 0.0323 log α thresholds. Although there were other columns that also had high rankings, the Resolve C18, Cortecs UPLC Shield RP18, and Accucore XL C18 columns were chosen as potential suitable columns because of their availability in the ECU chemistry laboratory or recommendations from manufacturers. The Cortecs UPLC Shield RP 18 is recommended by Waters in an application note in which 16 cannabinoids were separated in a reversed-phase isocratic HPLC method.[1] The Accucore C18 column, which has almost identical column parameters and therefore the same underlying chemistry as the Accucore XL C18 column, is recommended by Thermo Fisher in an application note in which 17 cannabinoids were separated in a reversed-phase gradient HPLC method.[2] The fact that this model's high-ranking columns have been used by manufacturers for cannabinoid separation in the past gives

confidence that the HSM does predict columns with suitable separation. Next, the Waters Cortecs C18 was chosen because it has also been recommended by the manufacturer for cannabinoid separation, even though its ranking is lower.[3] The predicted and actual separation of a lower-ranking column was compared to the higher-ranking columns to better understand the model's ranking performance.

It should be noted that the two top-ranking columns for cannabinoid separation were Type A columns. Type A silica is prepared from metal silicate and has high metal content, acidic surfaces, and low silica purity. The high metal content affects the peak shape of basic compounds, and peak tailing typically results.[4] On the contrary, Type B silica is prepared by a metal-free process, with high purity silica and a small amount of metal. Type B silica can be used to separate ionic, ionizable, and basic compounds, and is the most used silica type on the market today because it does not have the acidic and tailing properties of Type A columns. Since Type A columns are not as reliable as Type B columns, the goal here was to recommend a column that was not Type A. However, to show the model's prediction accuracy on Type A columns, the cannabinoids were experimentally analyzed on them.

The third top-ranking column was a Type EP, or embedded polar group, column. As seen in Figure 5.1, a Type EP column includes an alkyl chain much like its C18 counterparts but contains a polar group intrinsic to the chain.[5] The specific functional groups that are incorporated into the alkyl chain can include urea, carbamate, ether, and amine, but many manufacturers classify this group as proprietary.[6] Embedded polar groups offer many advantages such as increased retention factors for polar and basic analytes, with nonpolar analytes being less affected. Silanol activity is also suppressed, which leads to better peak shape and decreased tailing of basic compounds, particularly at intermediate pH values. Additionally, hydrogen bond

donors can interact through hydrogen bonding with the polar-embedded groups. This interaction might explain the increased retention for this class of compounds compared with conventional alkyl phases.[6]

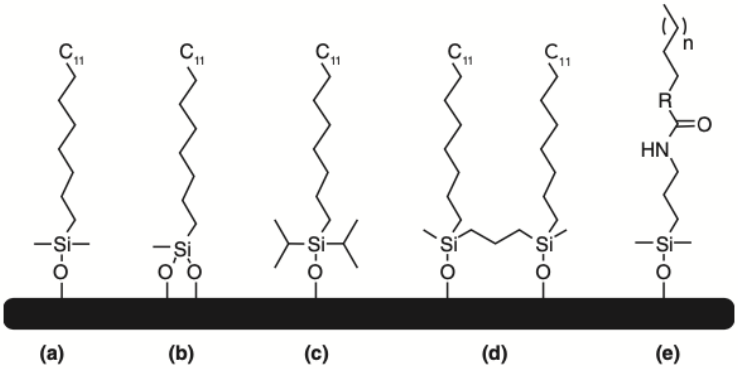


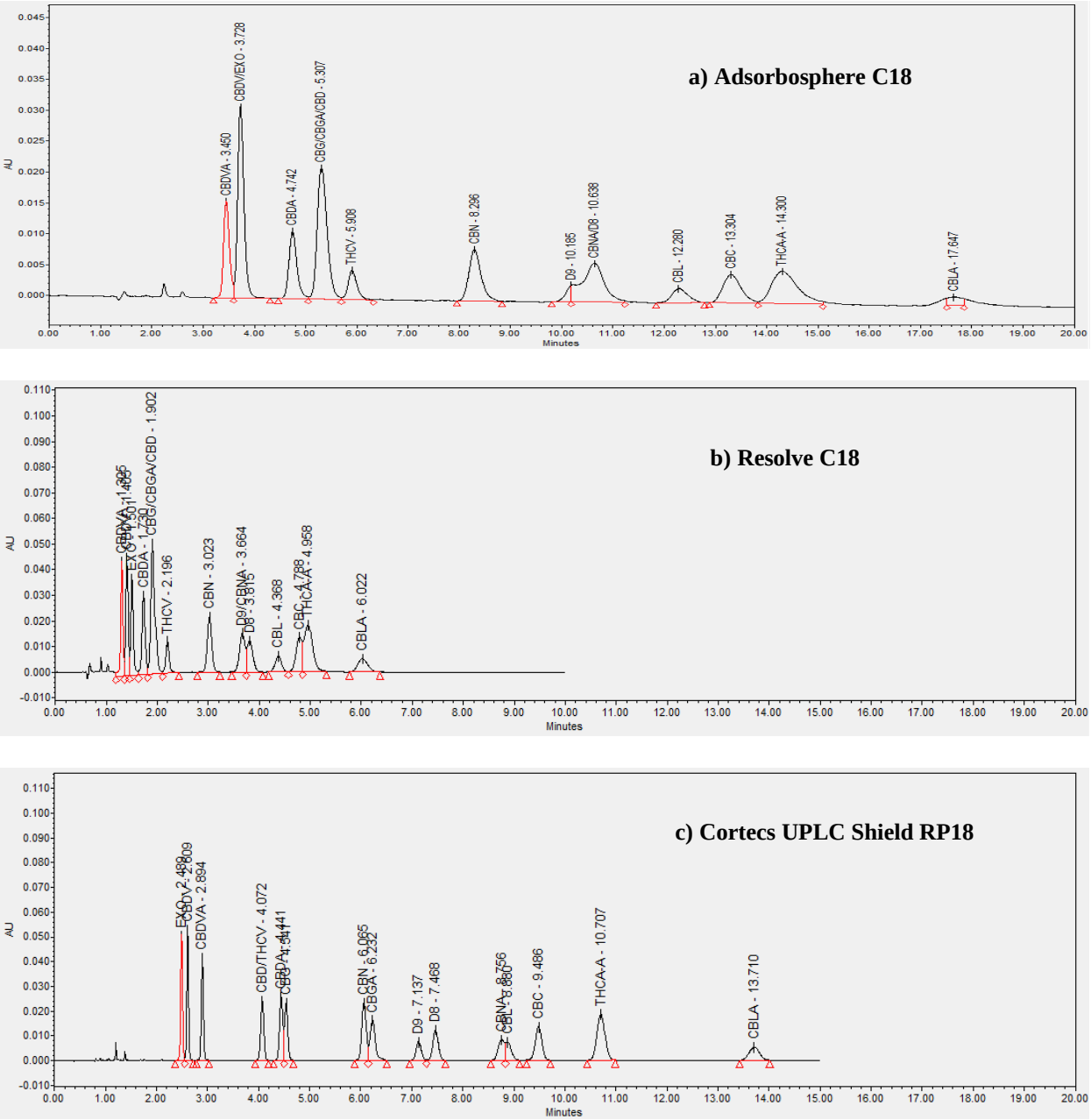
Figure 5.1. Comparison of a traditional alkyl chain (a)-(d) to an embedded polar group (e).[7]

5.2 Experimental Verification of Best-Chance Column Prediction

Once the above columns had been selected and obtained, they were analyzed to determine the accuracy of the HSM predictions. Instead of individual mixes of two solutes being prepared as seen in Section 2.2, a fresh cannabinoid solute mix that included all the cannabinoids was prepared and used for the injections onto the columns. This was done to reduce mobile phase, run time, and to most importantly view the resulting chromatogram with all analytes being visible. Fresh mobile phase for the cannabinoid analyses was prepared as seen in Section 2.5.

The Adsorbosphere C18 and Resolve C18 columns were already in the ECU chemistry department's laboratory, whereas the Cortecs UPLC Shield RP18, Cortecs C18, and Accucore XL C18 columns were donated by manufacturers and colleagues. All these columns were flushed with mobile phase and allowed to statically equilibrate overnight. It is important to note that the Adsorbosphere C18 column had 4.6-mm inner diameter by 250-mm length dimensions because no other dimensions were available for purchase. This will result in an increase of solute

retention and peak broadening. Finally, the cannabinoid solutes were injected onto the various columns to produce the following chromatograms in Figure 5.2 (a)-(e). The predicted versus actual log α correlations on these columns are shown in Figure 5.3 (a)-(e).



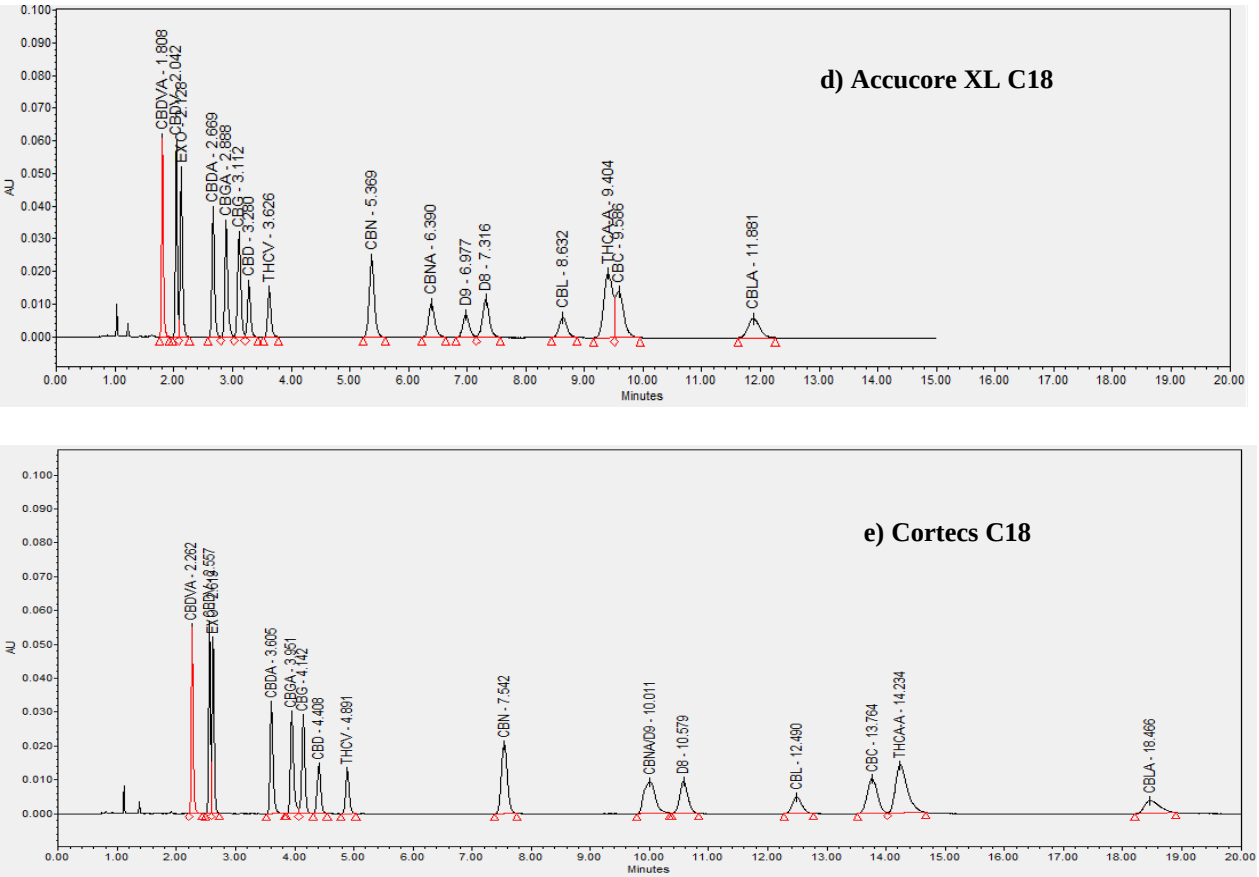


Figure 5.2. Cannabinoid separation on the a) Adsorbosphere C18, b) Resolve C18, c) Cortecs UPLC Shield RP18, d) Accucore XL C18, and e) Cortecs C18 columns.

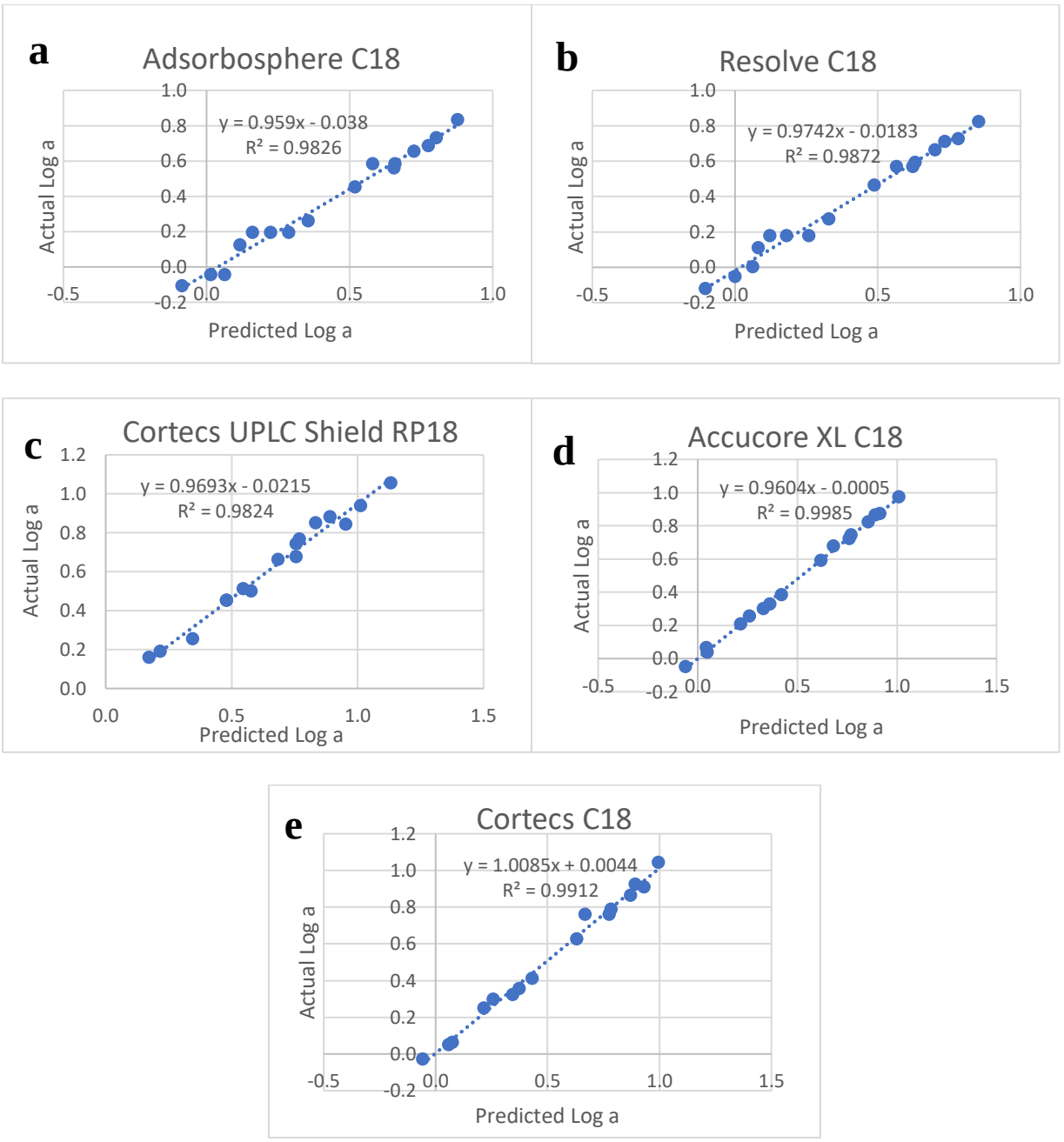


Figure 5.3. Predicted versus actual log α correlations on the a) Adsorbosphere C18, b) Resolve C18, c) Cortecs UPLC Shield RP18, d) Accucore XL C18, and e) Cortecs C18 columns.

Table 5.2 shows the ranking for each column in the HSM2 column database, the column name, silica type, the number of solute pairs predicted to be separate (with 0.0431, 0.0323 log α limits) and the number of peaks observed in the actual separation. The Adsorbosphere C18

column in Figure 5.2(a) does not have an ideal separation even though it was ranked as #1 in the log α predictions. In total, there were four coelutions of the cannabinoids on Adsorbosphere C18 column, the peaks were broad, and the retention time of the last eluting solute was approximately 18 minutes. The peak broadness and long retention time are a result of the column's 250-mm length dimensions. However, with the longer column, the separation of the cannabinoids would theoretically be better, but it was not in this case.

Table 5.2. Each columns' predicted column ranking and predicted number of separated solute pairs compared to actual number of distinguishable peaks.

Ranking	Column Name	Type	Predicted Solute Pairs Separated (0.0431, 0.0323)	Actual Peaks Distinguishable in Chromatograms
1	Adsorbosphere C18	A	13, 13	12
6	Resolve C18	A	11, 13	13
8	Cortecs UPLC Shield RP18	EP	11, 12	15
16	Accucore XL C18	B	10, 12	16
160	Cortecs C18	B	8, 11	14

Additionally, when correlating the predicted and actual log α in Figure 5.3(a), the predictions are notably worse than the Type B columns in Figure 5.3 (d) and (e). This makes sense as there was a total of 11 out of the 16 solutes that had differences in the predicted and actual log α greater than the 0.0431 limit on this column. The cause of this could be from the Adsorbosphere C18 being a Type A column and having all the Type A silica disadvantages (see Section 5.1), as well as the calibration columns not including any Type A columns. Therefore, although the model predicted this column to be the column with the best chance of cannabinoid separation, it was not. For these same reasons associated with Type A silica, the Resolve C18 column was not a successful contender either as seen in Figure 5.3(b). Although the Resolve C18 ranked #6 in Table 5.2, the retention time of the last eluting solute was extremely short at

approximately 6 minutes, which resulted in multiple coelutions and critical pairs. Overall, the model predicted two Type A columns to be the best-chance column. This was true mathematically but was in reality false. For these reasons, these two Type A columns were dismissed from the best-chance column list, and the solid black line in Table 5.2 represents the distinction between the Type A columns from the other ranked columns.

After excluding the Type A columns, the next column that was predicted to have the best chance of separating the cannabinoids was the Cortecs UPLC Shield RP18. This column was ranked #8, with 11 ($\Delta > 0.0431$) to 12 ($\Delta > 0.0323$) solute pairs predicted to be separated out of the 16 total solutes. In actuality, 15 out of the 16 total peaks were distinguishable in the chromatogram. This prediction seems to be good at first glance. However, when comparing the differences between the predicted and actual log α values for the cannabinoids on this particular column, there were six solutes that had differences greater than the 0.0431 log α maximum limit. The resulting poor correlation is displayed in Figure 5.3(c). When this occurred on the Type A columns, it made sense because the model itself did not include any Type A silica columns, so those types of interactions were not accounted for in the original model. It should be noted that although the correlation coefficients for the Type A columns and the Cortecs UPLC Shield RP18 column are very similar, the Cortecs UPLC Shield RP18 column had only six solutes that surpassed the 0.0431 log α limit, which is much more accurate compared to the Adsorbosphere C18's 11.

Ranking at #16, the Accucore XL C18 predicted that 10 ($\Delta > 0.0431$) to 12 ($\Delta > 0.0323$) solute pairs would be separated out of the 16 total solutes. In actuality, 16 out of the 16 total peaks were distinguishable in the chromatogram. Because all the solutes' predicted log α values were relatively accurate and all 16 peaks were distinguishable, the Accucore XL C18 column

was selected as the column with the best chance of separating the cannabinoid solutes under the original conditions used to build the model. This column is also the highest-ranking Type B column that is currently available by a widely known manufacturer. From an efficiency standpoint, the Accucore XL C18 is also the best choice of column because the run time is the shortest at approximately 12 minutes, which will in turn reduce analysis time and mobile phase waste. Even though this column is classified as the column with the best chance of separation, it can be seen in Figure 5.2(d) that the cannabinoid solute pairs CBDV/EXO-THC and THCA-A/CBC would be classified as critical pairs on the Accucore XL C18 column. These critical pairs can be separated using common HPLC conditions to obtain a more suitable separation as seen in Section 5.5.

It is apparent that the model underestimates the number of solutes that will be separated in the actual chromatogram for these columns. An explanation for this is that the limits that were set for the chances of a possible coelution between neighboring solutes were broad. So, even if a column had a high number of solute pairs that were predicted to have a chance of coeluting, that does not necessarily mean that they in fact would coelute under the given conditions.

This was the case for the Accucore XL C18 column. The model predicted that there were five solute pairs that had the chance to coelute. However, none of them coeluted because of each solute uncertainty in their predictions. For example, Table 5.3 shows that D8-THC and D9-THC were a solute pair that had the chance to coelute since their differences in predicted log α values were 0.009, which is much less than the 0.0431 limit. However, the actual log α for D9-THC was earlier than the predicted log α by 0.040 log α units. The actual log α for D8-THC was earlier than the predicted log α by 0.026 log α units. D9-THC moved much farther left on the chromatogram than D8-THC, but still within the expected 0.0431 log α range. This is how two

solutes that were predicted to coelute with one another can be fully separated. In fact, the Accucore XL C18 column had no solutes that eluted outside of the predicted 0.0431 log α limit, which supports the accuracy of the predictions. This is shown by the significant correlation in Figure 5.3(d).

Table 5.3. Example of predicted coelutions being separated on a chromatogram.

Solute	Predicted log α	Δ Predicted log α	Actual log α	Δ Actual log α
D9-THC	0.761		0.721	
D8-THC	0.770	0.009	0.744	0.023

The Cortecs C18 column was also considered for confirmation because although it ranked #160 in the database predictions, it has been shown by the manufacturer and other users to have the capacity to separate the cannabinoid solutes. The model predicted that the Cortecs C18 could have 8 ($\Delta > 0.0431$) to 11 ($\Delta > 0.0323$) solute pairs separated out of the 16 total solutes. In actuality, 14 peaks were distinguishable in Figure 5.2(e) with only the solute pairs CBDV/EXO-THC and D9/CBNA coeluting. The model predicted CBDV and EXO-THC to have the chance to coelute with each other (Δ 0.016 log α), but it did not predict the same for the D9/CBNA solute pair (Δ 0.107 log α). This is because the prediction of CBNA was inaccurate by 0.092 log α units, which greatly exceeds the expected limit of 0.0431 log α units. Overall, the correlation seen in Figure 5.3(e) for the Cortecs C18 was good with the CBNA prediction clearly standing out.

This may prompt the question of at what point the model would predict that the separations of the cannabinoid solutes to not be sufficient. It cannot be stated with certainty where in the prediction list the separation is sufficient or “good enough” on a column. There is a

large gap between the #6 ranking of the Accucore XL C18 and the #160 ranking of the Cortecs C18 column. Even then, from #160 to #556, it is not clear where the cutoff should be to recommend a column for separation. Regardless, the main aim of this Thesis is to identify a column with the *best* chance of separation of the cannabinoid solutes. The answer to this question will require future analyses and advanced data treatment of the lower-ranking columns.

5.3 Model Confirmation with Other Columns

In Section 5.2 above, it appeared that the model predicted the worst on Type A columns, slightly better on Type EP columns, and best on Type B columns. The Accucore XL C18 and Cortecs C18 had the best predicted versus actual log α correlation, suggesting that the model predicted best on Type B silica columns. However, there was still a need to analyze additional columns to conclude how the model predicted on other Type B and EP columns. Therefore, the Waters XBridge C18 and Waters XBridge Shield RP18 columns were chosen. These columns were also chosen since Waters is a well-known column manufacturer and readily available. Figure 5.4 (a) and (b) below show the cannabinoid separations on these columns. Figure 5.5 (a) and (b) show the predicted versus actual log α correlations. Table 5.4 shows the ranking of each column, the number of solute pairs that the model predicted to be separate, and the number of peaks distinguishable in the actual chromatograms. It should be noted that the dimensions of the XBridge Shield RP18 column were 4.6 by 250-mm, which will result in longer retention and broader peaks, much like the Adsorbosphere C18.

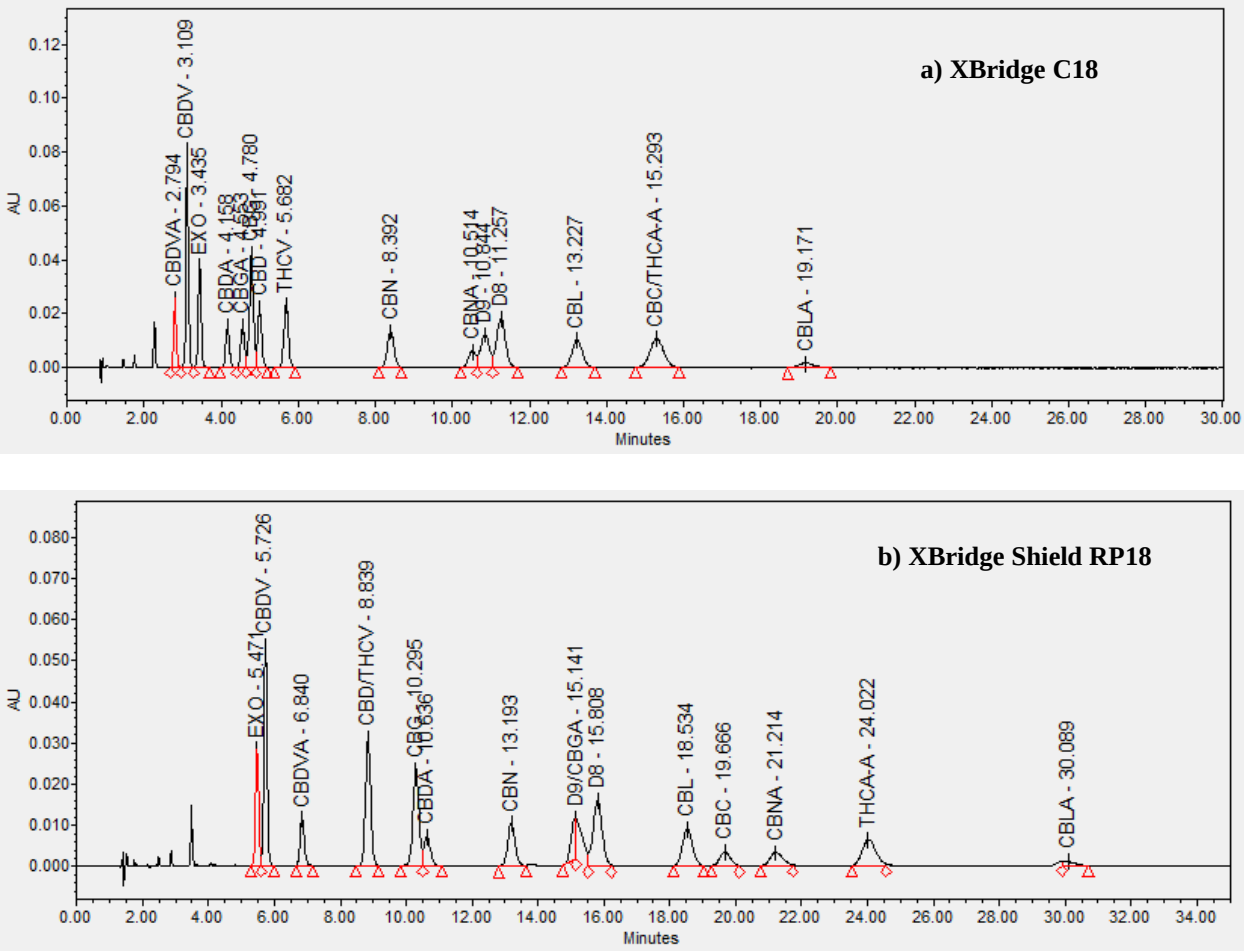


Figure 5.4. Cannabinoid separation on the a) XBridge C18 and b) XBridge Shield RP18 columns.

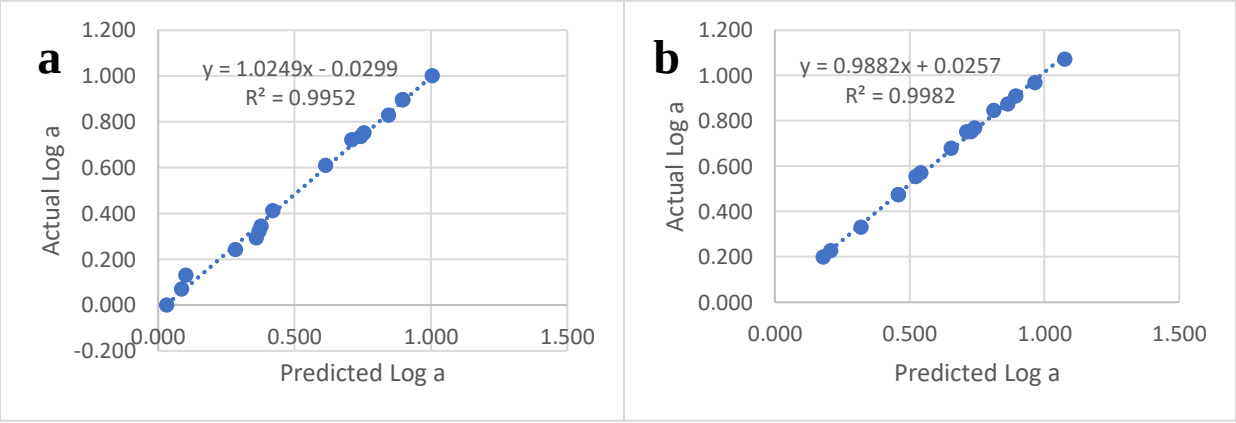


Figure 5.5. Predicted versus actual log α correlations on the a) XBridge C18 and b) XBridge Shield RP18 columns.

Table 5.4. Additional columns chosen for prediction accuracy confirmation.

Ranking	Column Name	Type	Predicted Solute Pairs Separated (0.0431, 0.0323)	Actual Peaks Distinguishable in Chromatograms
137	XBridge C18	B	9, 10	15
147	XBridge Shield RP18	EP	9, 9	14

The XBridge C18 column ranked #137. The model predicted that this column could have 9 ($\Delta > 0.0431$) to 10 ($\Delta > 0.0323$) solute pairs separated out of the 16 total solutes. In actuality, it can be seen in Figure 5.4(a) that 15 peaks were distinguishable. This is because the CBC/THCA-A solute pair coeluted with one another, and the model did in fact predict this solute pair to have a chance to coelute. In Figure 5.5(a), the correlation between the predicted and actual log α is strong, but there are a few deviations. This is because there were two solutes on the XBridge C18 column that had differences between their actual and predicted log α values greater than the 0.0431 log α limit (Δ 0.068 and Δ 0.051). Regardless, the model predicted relatively well on this Type B column, similarly to how it predicted on the Cortecs C18 column.

The XBridge Shield RP18 column ranked #147. The model predicted that this column could have 9 ($\Delta > 0.0431$ and 0.0323) solute pairs separated out of the 16 total solutes. In actuality, it can be seen in Figure 5.4(b) that 14 peaks were distinguishable. This is because of the coelution of each solute in the CBD/THC and D9/CBGA critical pairs. However, the model predicted these solute pairs to have a chance to coelute. In Figure 5.5(b), the correlation between predicted and actual log α is very strong. This is because there were no solutes that had differences between their actual and predicted log α values that were greater than the 0.0431 log α limit. Thus, the correlations showed the accuracy of the model's predictions on this Type EP column. The reasoning for the model's strong accuracy on the XBridge Shield RP18 column and low accuracy on the Cortecs UPLC Shield RP18 column remains unknown. At a minimum, with

the confirmation of these XBridge columns, it can be inferred that the relatively poor prediction on the Cortecs UPLC Shield RP18 column was not a Type EP column-specific issue.

5.4 Solute and Column Parameter Contributions to Separation Trends

Although the HSM is a model that can be used to predict solute elution, it also shows the individual contributions of each solute-column interaction that affects the retention of a solute on a column. The mathematical predictability of the model has been investigated here, but chemical understanding can also be gained. Thus, the solute and column parameters can be used to explain the separation trends on various columns, and those differences can be utilized to maximize the separation ability of various columns.

To explore this further, the critical pair THCA-A/CBC will be used as an example. The differences between the solute parameters for THCA-A and CBC are shown below in Table 5.5. It is apparent that the largest differences between these two solutes are attributed to the *a* and *d* solute parameters. Because these two solutes have such large differences (0.471 for the *a* parameter and 0.568 for the *d* parameter), then a column with relatively large B and D terms would maximize these differences. The summation of the solute and column parameter products would therefore theoretically provide better separation.

Table 5.5. Solute parameter differences between THCA-A and CBC.

Solute	**y-int**	**h**	**b**	**a**	**k**	**v**	**d**
THCA-A	-0.425	1.302	0.231	1.322	0.002	1.099	0.366
CBC	-0.400	1.286	0.216	0.851	-0.035	1.166	-0.202
Δ	-0.025	0.016	0.015	0.471	0.036	-0.067	0.568

The XBridge C18, Accucore XL C18, and Cortecs C18 columns were chosen for comparison because each one separated the two solutes with varying degrees of success. Additionally, the model predicted relatively well on these columns so the values could be trusted. Table 5.6(a) shows the individual solute-column interactions associated with each column and solute, and how those contributions lead to the predicted log α value. The XBridge C18 column was not able to separate the two solutes. This was indicated and confirmed by the predicted log α difference between the solutes on this column being miniscule at 0.002 log α units. The Accucore XL C18's chromatogram showed that the THCA-A peak merged into the CBC peak, but there was not an absolute coelution like the XBridge C18 column. This better separation is indicated by the predicted log α difference of 0.025 between the two solutes. The Cortecs C18 column was the column predicted to best separate the THCA-A and CBC solutes, with the predicted log α difference being 0.039. The differences between the THCA-A/CBC predicted log α values across the columns show that the Cortecs C18 was best able to separate the critical pair.

Table 5.6 a) Solute-column interactions and predicted (pred) log α differences between THCA-A and CBC on the XBridge C18, Accucore XL C18, and Cortecs C18 columns and b) Accucore XL C18 and Cortecs C18 solute-column differences compared to XBridge C18.

Table 5.6 (a)

Column	Solute	y-int	hH	bA	aB	kC	vV	dD	Pred log a
XBridge C18	THCA-A	-0.425	1.288	-0.027	-0.001	0.000	0.057	0.004	0.896
	CBC	-0.400	1.272	-0.025	-0.001	-0.006	0.061	-0.002	0.899
	Δ	0.025	-0.016	0.002	0.000	-0.006	0.003	-0.006	**0.002**
Accucore XL C18	THCA-A	-0.425	1.327	-0.016	0.050	0.000	-0.021	-0.026	0.890
	CBC	-0.400	1.311	-0.015	0.032	-0.005	-0.022	0.014	0.915
	Δ	0.025	-0.016	0.001	-0.018	-0.006	-0.001	0.040	**0.025**
Cortecs C18	THCA-A	-0.425	1.323	-0.039	-0.037	0.000	0.078	-0.008	0.892
	CBC	-0.400	1.307	-0.037	-0.024	-0.002	0.083	0.004	0.932
	Δ	0.025	-0.016	0.003	0.013	-0.002	0.005	0.012	**0.039**

Table 5.6 (b)

Column	Solute	y-int	hH	bA	aB	kC	vV	dD	Pred log a
XBridge C18	Δ	0.025	-0.016	0.002	0.000	-0.006	0.003	-0.006	0.002
Accucore XL C18	Δ	0.025	-0.016	0.001	-0.018	-0.006	-0.001	0.040	0.025
	Δ XBridge C18	0.000	0.000	-0.001	-0.018	0.001	-0.005	0.046	0.022
Cortecs C18	Δ	0.025	-0.016	0.003	0.013	-0.002	0.005	0.012	0.039
	Δ XBridge C18	0.000	0.000	0.001	0.013	0.004	0.001	0.018	0.037

However, a deeper understanding of column selectivity can be acquired when focusing on the size of the specific solute-column interactions that contributed to the resulting log α predictions. The different contributions of the solute-column interactions across the three columns are shown in Table 5.6(b). The Δ represents the differences in the interaction values between the two solutes on each column. The Δ XBridge C18 represents the comparison in differences between the Accucore XL C18 and Cortecs C18 with the XBridge C18 column. Put simply, if the differences between the two columns for each solute-column interaction are similar to the XBridge C18 column, then that interaction had no contribution to the difference in column

selectivity for those two solutes. In Table 5.6(b), the difference in the dD term between the Accucore XL C18 and XBridge C18 is the largest at 0.046. This contribution seems like it would yield the biggest difference in the log α predictions, but it does not. Some of the separation gained by the dD term is lost by the aB interaction, resulting in a predicted log α that is different from the XBridge C18 column, but not different enough to fully separate the critical pair. For the Cortecs C18 column, the differences in the aB and dD contributions are both positive and larger compared to the other contributions. This results in the highest predicted log α difference between the two solutes out of these three columns, yielding a sufficient separation of the THCA-A and CBC solutes. Knowing the most important solute-column interactions can then aid in the decision of which common HPLC conditions to modify to acquire ideal separation. It also allows for the selection of other columns that magnify these specific differences.

5.5 Modification of Common HPLC Conditions to Optimize Best Separation

A main goal of this approach was to use the HSM to identify the column with the best chance of separating a solute mixture, but additional method development will likely be necessary to optimize the separation. Modification of common HPLC conditions could then be performed to acquire a more suitable separation. Common HPLC conditions that are changed throughout a regular method development process include temperature, mobile phase pH, gradient ratios, organic percentages, and buffer concentration. The Accucore XL C18 column was determined to provide the best chance of separation of the cannabinoid solutes. The critical pairs on this column included CBDV/EXO-THC and THCA-A/CBC. As can be seen in Figure 5.2(d), CBDV and EXO-THC are the second and third earliest eluting peaks, whereas THCA-A and CBC are the second and third latest eluting peaks.

Although this study did not include final optimization of the chromatographic conditions, Thermo's Accucore C18 cannabinoid analysis application note can be used as a guide to understand which chromatographic conditions could be modified to obtain a better separation. The application note utilized a UPLC gradient in the separation of the solutes, whereas the approach shown here applied an isocratic elution method with similar pH and organic percentage. The application note began its gradient with 60% acetonitrile which slowly increased to 68% over the course of approximately 14 minutes, then washed the column with 95% acetonitrile.[8] This shallow gradient below 70% should affect both sets of critical pairs. The earlier eluting CBDV and EXO-THC solutes would have a better chance of separating due to the lower organic composition at the beginning of the chromatographic run. The lower organic percentage could also aid in the separation of the later eluting peaks. The later eluting THCA-A and CBC solutes will also have a chance to become sharper and more resolved when the organic percentage increases as this was shown in Thermo's application note. This demonstrates that only slight adjustments would likely need to be made once a best-chance column has been chosen using this approach.

5.6 Limitations

There are a variety of limitations to this method of identifying the column with the best chance of separation. First, the error from the multiple linear regression of the solute parameters themselves were not included in the prediction of the best-chance column. The overall standard errors of the regressions for the cannabinoid solute parameters were noted, as that could account for more error associated with the predictions of those solutes. Some solutes had higher standard errors than others. For example, CBNA had one of the highest standard errors (0.045) from its

solute parameter multiple linear regression output, which resulted in more error in the predicted log α calculations than was accounted for in the best-chance column rankings. An alternative to acquire a better prediction for each solute would be to compare the actual and predicted log α of each solute on all 13 calibration columns to estimate each solute's individual error. This would allow for a more accurate prediction of the individual solutes compared to taking the average of all 16 solutes' errors. However, it would be a daunting task to apply an individual error for all 16 solutes across more than 550 columns in the HSM2 database using Excel.

As mentioned above, there were instances when a solute's log α prediction exceeded the 0.0431 limit. Because this limit was also used as a threshold for determining possible coelutions of neighboring solutes, the error associated with each solute's prediction accuracy affected the accuracy of the predicted coelutions. Therefore, if the predicted log α values were inaccurate, then the model might have determined that a solute pair would or would not have a chance of coeluting based on that assumption. This could lead to possible errors in the ranking system used in this approach.

The 0.0431 log α limit was used as the minimum difference between each neighboring solute in order to be considered separate. In hindsight, this limit might not have been correct, as it does not consider the fact that *both* solutes' predicted log α has an error range of ±0.0431. Although this value is correct for determining the error associated with the predictions, more advanced statistical analysis is needed to determine the limit in which two adjacent peaks can be considered separate.

Another contributor to inaccuracies in the log α predictions are from the published column parameters themselves. The lot-to-lot variability is a major concern because there are no reproducibility measurements associated with either the HSM1 or HSM2 databases. Therefore,

differences between the same brand of columns across different lots can cause inaccurate predictions using the approach shown here.

Another disadvantage of this approach is the upfront effort required for the identification of the best-chance column. Although there is a lot of work required on the front end, the process is relatively smooth and provides logical evidence for the pathway of choosing a column for a separation. Additionally, there is no consideration of peak width in this approach, which is a major contribution to resolution. Lastly, the approach shown here can only consider the predicted separation of the solute set at the conditions which were used to build the model. This means that a column that ranked low in the prediction table could have been able to separate the solutes at other LC conditions; by changing either the mobile phase pH or buffer, temperature, or organic composition. The organic percentage can be easily extrapolated but pH and temperature are more difficult to predict.

5.7 References

1) Aubin, A., Layton, C., Helmueller, S. Separation of 16 Cannabinoids in Cannabis Flower and Extracts Using a Reversed-Phase Isocratic HPLC Method. Waters Application Note. **2018**.
2) McRae, G., Melanson, J. Quantitation of Seventeen Cannabinoids in Dried Cannabis, Hemp and Vape Oils by LC-MS/MS. Thermo Scientific Application Note. **2021**.
3) Van Tran, K., Twohig, M., Hudalla, C. Analysis of Cannabinoids in Cannabis Plant Materials and Edible Products Using Ultra-Performance Liquid Chromatography (UPLC) with PDA and Mass Detection. Waters Application Note. **2021**.

4) Hawach Scientific. Introduction to HPLC Column and Reversed-Phase HPLC Columns. www.hawachhplccolumn.com/introduction-to-hplc-column-and-reversed-phase-hplc-columns/ (accessed 12 June 2024).
5) Majors, R., Przbyciel, M. Columns for Reversed-Phase LC Separations in Highly Aqueous Mobile Phases. LCGC International Column Watch. **2002**.
6) Waters. Polar Compound Retention. In Understanding What Works and What Doesn't Work in Reversed-Phase Polar Compound Retention. Waters Webinar. **2016**.
7) Bell, D., Green, A. Columns- What is on Your HPLC Particle? A Look at Stationary Phase Chemistry Synthesis. LCGC Asia Pacific. **2021**, 24(1), 28-33.
8) McRae, G., Melanson, J. Quantitation of Seventeen Cannabinoids in Dried Cannabis, Hemp and Vape Oils by LC-MS/MS. Thermo Scientific Application Note. **2021**.

CHAPTER 6: CONCLUSIONS AND FUTURE OUTLOOK

6.1 Conclusions

The HSM has widely been utilized in the determination of similar and orthogonal columns but has had limited use as a predictive tool. The improved accuracy of the refined HSM2 model has made predicting relative elution plausible. This improvement has made it possible to reasonably predict the elution of a widely relevant set of cannabinoid solutes, given the increasing interest from consumers and manufacturers alike. This approach is also all-inclusive, with no bias associated with promoting a certain brand or phase of column. The HSM2 model was therefore utilized to calculate the cannabinoid solute parameters and predict their elution on more than 550 columns in a public HPLC column database.

A novel sorting approach that utilized the prediction error and a ranking system was then applied to identify the column with the best chance of cannabinoid separation. After dismissing columns on which the model did not accurately predict (Type A columns), the Accucore XL C18 column was selected as the column with the best chance of separation. The Accucore XL C18 had no significant errors in its log α predictions, a low analysis time, and only two critical pairs in the chromatogram left to be separated. Solute and column parameters were then focused on to explain separation trends on a few columns.

6.2 Future Outlook

With this being the first attempt at using the HSM as a prediction tool for this laboratory, there were inevitably limitations to the approach. Future studies are planned to explore these limitations to improve the methodology described in this Thesis. The most obvious limitation is

the 0.0431 log α limit used for determining the number of predicted critical pairs on a column. Although this value is correct for determining the accuracy of the predictions, it cannot be stated with certainty that it should also be the limit to determine if adjacent peaks will be separate. Therefore, advanced data treatment is needed to statistically defend the determination of critical pairs. Once this dilemma is corrected, a ranking cutoff could be determined. This would be useful since there was a large gap between the #16 ranking Accucore XL C18 column, which included all 16 distinguishable peaks, and the #160 ranking Cortecs C18 column, which included 14 distinguishable peaks.

It was also shown that a key contributor to inaccuracies in the log α predictions could be from the column parameters. A major concern is the lot-to-lot variability of the columns in the HSM1 and HSM2 databases, since there are no reproducibility measurements reported. Therefore, a large lot-to-lot study is being considered to estimate the overall variability of the columns, including the differences between lots.

Another avenue for future studies includes estimating the accuracy of the cannabinoid solute parameters across different HPLC conditions such as organic composition, temperature, and pH. It is known that solute parameters are dependent on the conditions used to calculate them. However, conditions could be systematically varied to determine if there are any prediction models that could be applied to predict the solute parameters at different conditions.

Lastly, the development of the HSM2 model was an improvement from the HSM1 model which reduced the standard error in predictions.[1] This is because the updated model accounted for more accurate solute-column interactions that exist in reversed-phase chromatography. Since then, Stoll has developed a new revision of the HSM, HSM3, which is based on a much larger dataset (43,329 total measurements) containing selectivities for compounds covering a broader

range of physicochemical properties compared to HSM1.[2] This HSM3 model is based on retention measurements for 75 compounds using 13 reversed-phase stationary phases. The updated model also uses an ammonium formate mobile phase which is compatible with mass-spectrometry. This is highly sought after for high throughput since the identity of each solute could be easily determined using the mass-to charge ratios, and coelutions do not matter. However, in its current form, the HSM3 model is impractical due to using such a high number of solutes to characterize solute-column interactions. The HSM2 model is therefore still applicable while the predictability of HSM3 is being refined. Regardless, the cannabinoid solute parameter values can still be used by others to predict solute retention and compare potential separations on columns without laboratory activities.

6.3 References

1) Stoll, D. et al. Improvements in the Predictive Accuracy of the Hydrophobic Subtraction Model of Reversed-phase Selectivity. *J Chrom A*. **2021**, 1636, 461682.
2) Stoll, D., Rutan, S., Kempen, T., Dahlseid, T., Kruger, Z., Pirok, B., Shackman, J., et al. Improved Hydrophobic Subtraction Model of Reversed-phase Liquid Chromatography Selectivity Based on a Large Dataset with a Focus on Isomer Selectivity. *J Chrom A*. (In Press, Journal Pre-proof). **2024**, 465127.

APPENDIX

Table A1. Experimentally calculated HSM2 column parameters due to database unavailability.

Name	**Manufacturer**	**Type**	**Phase**	**k_{EB}**	**H**	**A**	**B**	**C**	**V**	**D**
Kinetex C18	Phenomenex	B	C18	5.92	0.996	0.000	-0.035	0.015	0.025	0.031
Poroshell 120 EC-CN	Agilent	CN	CN	0.84	0.542	-0.149	-0.296	-0.113	0.181	0.268
Eclipse Plus C8	Agilent	B	C8	5.88	0.882	-0.088	-0.032	-0.050	0.087	0.022
SymmetryShield C8	Waters	EP	EP	5.81	0.972	-0.246	-0.173	-0.755	0.181	0.373
HyperClone MOS C8	Phenomenex	EP	EP	4.57	1.047	0.177	-0.109	1.533	-0.040	0.196