8-2016

# Enhancing Silymarin Fractionation via Molecular Modeling using the Conductor-like Screening Model for Real Solvents

Emma C. Brace
*Purdue University*

*Graduate School Form*
*30 Updated 12/26/2015*

**PURDUE UNIVERSITY**
**GRADUATE SCHOOL**
**Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Emma C. Brace

Entitled

ENHANCING SILYMARIN FRACTIONATION VIA MOLECULAR MODELING USING THE CONDUCTOR-LIKE SCREENING MODEL FOR REAL SOLVENTS

For the degree of Master of Science in Agricultural and Biological Engineering

Is approved by the final examining committee:

Abigail S. Engelberth
Chair

Mario G. Ferruzzi

Nathan S. Mosier

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Abigail S. Engelberth

Approved by: Bernard A. Engel 7/14/2016

Head of the Departmental Graduate Program Date

# ENHANCING SILYMARIN FRACTIONATION VIA MOLECULAR MODELING USING THE CONDUCTOR-LIKE SCREENING MODEL FOR REAL SOLVENTS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Emma C. Brace

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Agricultural and Biological Engineering

August 2016

Purdue University

West Lafayette, Indiana

*To all the strong women in my life – thank you for the support and inspiration.*

# ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Abby Engelberth, for her mentorship and support through graduate school. I would also like to thank my committee members, Dr. Nate Mosier and Dr. Mario Ferruzzi, for their support on this project.

Second, I would like to thank my friends and colleagues in the Laboratory of Renewable Resources Engineering (LORRE) and in Agricultural & Biological Engineering (ABE). In particular, my labmates Samira, Mahdieh, and Studie provided much needed and much appreciated moral and technical support throughout the past two years! I also really appreciated Linda Liu's assistance in the laboratory and in use of the HPLC.

Additionally, I would like to thank the ABE Department and Graduate Committee for their genuine commitment to helping me succeed at Purdue.

Last but not least, I want to thank my friends and family near and far for their support. My mom and dad have always believed I can do anything and I appreciate their love and support as I pursue my goals. Many thanks to Krystal for all of the hours on the phone and to Grant for never letting me take myself too seriously. To all my wildcat friends, friends at Purdue and in Women in Engineering – thank you for the never ending encouragement! Charlotte Lee – thanks for being the best and most encouraging thesis accountability partner!

TABLE OF CONTENTS

Page

Page

Page

# LIST OF TABLES

# LIST OF FIGURES

Figure Page

# LIST OF ABBREVIATIONS

| | |
|---|---|
| CCC | Counter-Current Chromatography |
| COSMO-RS | COnductor-like Screening MOdel for Real Solvents |
| CPC | Centrifugal Partition Chromatography |
| EtOAc | Ethyl Acetate |
| FID | Flame Ionization Detector |
| GC | Gas Chromatography |
| HPLC | High Performance Liquid Chromatography |
| IA | Isosilybin A |
| IB | Isosilybin B |
| IL | Ionic Liquid |
| MeOH | Methanol |
| RMSE | Root Mean Square Error |
| SA | Silybin A |
| SB | Silybin B |
| SC | Silychristin |
| SD | Silydianin |

# ABSTRACT

Brace, Emma C. M.S.A.B.E., Purdue University, August 2016. Enhancing Silymarin Fractionation via Molecular Modeling using the Conductor-like Screening Model for Real Solvents. Major Professor: Abigail S. Engelberth.

The market for bio-based products from plant sources is on the rise. There is a global challenge to implement environmentally clean practices for the production of fuels and pharmaceuticals from sustainable resources. A significant hurdle for discovery of comparable plant-derived products is the extensive volume of trial-and-error experimentation required. To alleviate the experimental burden, a quantum mechanics-based molecular modeling approach known as the **CO**nductor-like **S**creening **M**odel for **R**eal **S**olvents (COSMO-RS) was used to predict the best biphasic solvent system to purify silymarins from an aqueous mixture. Silymarins are a class of flavonolignans present in milk thistle (*Silybum marianum* L.), which has been used in traditional eastern medicine to treat liver disease. More recently, silymarins have been studied as a cancer treatment therapy due to their antioxidant properties, but effective large-scale separation methods need to be developed. Previous research has shown that these compounds can be fractionated using centrifugal partition chromatography (CPC), but not an acceptable level of purity. Due to previous incomplete fractionation, the silymarins are ideal compounds to assess the use of a molecular modeling approach to

prediction partitioning in a CPC separation. The COSMO-RS method was implemented using the software programs HyperChem, TmoleX, and COSMOthermX in order to calculate partition coefficients for the six silymarin compounds in various solvent systems. The partition coefficient for each silymarin in each solvent system was verified by experimentation using the shake flask method and compared to the results of the model.

# CHAPTER 1. INTRODUCTION

In the past two decades, efforts to make the biorefinery and production of biofuels more economically viable has resulted in increased research into the potential of biomass as a source for production of chemicals and other products. The U.S. Department of Energy (DOE) in conjunction with the Department of Agriculture (USDA) have set goals to simultaneously stimulate rural economies by supporting growth of agriculture and forestry production and decrease the need for oil imports, and aim to do this by fostering the new biorefinery industry, which could produce fuels, chemicals, and other products [1,2]. As the biofuels industry has shifted from food crops to increased research into lignocellulosic materials, the research and industrial opportunities for producion of bio-based chemicals has also increased. In 2015, the worth of the bio-based chemical market was estimated at $3.6 billion, and is expected to triple in value by 2021 [3].

In addition to fuels and chemicals, nutraceuticals are a growing market and are an opportunity to develop high value/low volume bioproducts. Nutraceuticals are compounds that occur naturally in plants and are used as additives in food or medicine. Biomass-derived nutraceuticals can be found in many plants, and remain an under-explored area of research. Major inhibitors to research into plant-derived

nutraceuticals are: 1) complex molecules existing in complex mixtures when removed from the plant,and 2) high cost of separations techniques. Liquid-liquid extraction methods have been explored as techniques for extraction and purification of a variety of molecules from plant sources, including, but not limited to, solanesol from tobacco [4], ginsenosides from American ginseng [5], silymarins from milk thistle [6,7], and xylose oligomers from switchgrass hemicellulose [8,9]. Developing a cost-effective extraction method typically involves significant experimental efforts, including time and the use of costly solvents and purified standards. Use of the Conductor-like Screening Model for Real Solvents (COSMO-RS) as a molecular modeling tool to predict partition coefficients of well-characterized biomolecules in two-phase solvent may alleviate the experimental burden and improve the development of such processes.

COSMO-RS was first developed by Andreas Klamt in 1995 as a tool for studying behavior of molecules in solution, building upon previous studies that were able to predict how molecules behave in the gas phase [10]. It has since expanded into a fully developed model that uses principles of quantum chemistry and statistical thermodynamics to calculate the chemical potential of a molecule in solution. Once the chemical potential is known, various other values can be determined, such as solubility, activity, and partition coefficients [11–13]. The diverse applications of COSMO-RS and the ability to derive properties using only molecular structure and composition of the solvent system makes it uniquely valuable. COSMO-RS has been shown to accurately predict partition coefficents of small biomolecules [14–17], and there has been some use of COSMO-RS in studying agriculturally significant products, including the extraction of lignin from

lignocellulosic biomass [18] and the solubility of cellulose [19] in ionic liquids. The present thesis assesses the use of COSMO-RS for predicting partition coefficients of silymarins in solvent systems meant for use in centrifugal partition chromatography. Silymarins are a class of flavonolignans found in milk thistile (*Silybum marianum* L.). Milk thistle has been used in traditional Eastern medicine to treat maladies of the liver for the past 2000 years, and has been studied for use in liver disease therapy [20]. Today, it can be found in a variety of dietary supplements as an antioxidant and herbal remedy [21]. There are six primary compounds in the silymarin: silychristin, silydianin, silybin (which has two stereoisomers), and isosilybin (which has two stereoisomers). Because fractionation and purification of these six silymarins using CPC has been studied previously [6,7], they are an ideal starting point for using modeling software to develop a way to predict partition coefficients, are an excellent molecule to use to evaluate the accuracy of the COSMO-RS model in predicting partition coefficients of plant-derived biomolecules.

Centrifugal partition chromatography is a type of countercurrent chromatography, in which two immiscible liquids are placed in contact under a centrifugal field to form a bi-phasic solvent system. Fractionation is a function of the partition coefficient of the solutes in the biphasic system. Fractionation of silymarins has been previously demonstrated using countercurrent chromatography [7] and centrifugal partition chromatography [6], and use of COSMO-RS would allow for prediction of the partition coefficient in a variety of solvent systems in order to hone in on the best solvent system

and improve previous fractionation methods in order to achieve high purity and high yield.

With the intent to study the use of COSMO-RS to identify the best solvent system for fractionation of naturally occurring biomolecules in solvent systems used in countercurrent chromatography, the objectives for this research are:

1. Use a theoretical method based on molecular modeling to predict partition coefficients of six silymarins in various solvent systems.
2. Hone in on best solvent system and use experimental methods to choose the best solvent system for fractionation.
3. Compare the experimental results and evaluate the model accuracy.

# CHAPTER 2. LITERATURE REVIEW

## 2.1 Conductor-like Screening Model for Real Solvents (COSMO-RS)

### 2.1.1 Development

The Conductor-like Screening Model for Real Solvents (COSMO-RS) was first published by Andreas Klamt in 1995 and was designed as an approach to address the challenges of studying solvation phenomena using computational techniques [10]. However, the origins go back further to Klamt's COSMO: an algorithm developed to accurately calculate dielectric screening effects in order to create a realistic dielectric continuum model in which to compute geometry optimization of solutes [22]. Prior to the development of COSMO-RS, the computational study of molecules in the gas phase or in vacuum had been well developed, but there was a real need to address how to properly complete theoretical calculations of molecular behavior in solution and to study fluid thermodynamics.

COSMO was one of many continuum solvation models (CSMs) to arise between the 1970s and early 2000s as a way to study the effects of solvation. Electrostatic interactions are an important and challenging component of modeling solvation, and so the goal of most continuum solvation models was to create a continuous dielectric medium for the solvents and model a surface around the solute in order to calculate the

interactions between the solute surface and the continuous solvent medium [23]. The primary difference between COSMO and other apparent surface charge dielectric continuum solvation models (DCSM) is that COSMO uses a scaled conductor boundary condition, rather than the dielectric boundary condition [11]. One issue that many DCSMs fail to address is the 'outlying charge', solute electrons which lie outside the cavity/boundary. The scaled boundary conditions created by COSMO make the electrostatic field operator more sensitive to any outlying charger, whereas the electrostatic potential operator in other DCSMs is less sensitive [24]. This results in COSMO being a more robust model in terms of being able to handle different solutes, including larger molecules.

After COSMO was developed and applied to industrial applications, limitations of the dielectric continuum solvation model arose. For example, the DCSMs generally cannot distinguish between solvents which have the same dielectric constant, even if their other properties are very different. Examples of this would be cyclohexane and benzene, or methoxyphenol and heptanone [11]. There were also issues in predictions of thermodynamic properties of polar systems and non-neutral systems [11]. This limitation led Klamt to develop COSMO-RS, which combines the DCM of COSMO with a statistical thermodynamics approach to interacting surfaces [11]. By considering the solvent and solute to be equal in terms of quantum chemical and statistical thermodynamics considerations, COSMO-RS is able to predict more consistent thermodynamic mixtures and vary the temperature, which greatly expands the computational power of the model. Because of this, COSMO-RS allows for theoretical

calculation of a variety of thermodynamic equilibrium properties of liquids, such as the free energy of hydration, activity coefficients, vapor pressure, capacity, and solubility, to name a few [12]. In addition to COSMO-RS, other methods such as UNIFAC and the CLOGP methods allow for calculation of the partition coefficient.

### 2.1.2 Theory

The major limitation of dielectric continuum solvation models like COSMO is that they are only able to describe the behavior of polarizable solvents on a macroscopic scale. No solvent system is uniformly polarizable at a molecular level, and so at best these types of models can qualitatively evaluate how the solvent will screen the electric fields of solutes, which are much weaker than the van der Waals surfaces of ions or polar molecules. As it is unknown if solvents really behave like a dielectric medium at the molecular scale, it is interesting that water cannot be characterized as a dielectric medium which is uniformly polarizable, nor as a conductor with free charges. Because neither model gives an accurate description of the behavior of water, a virtual experiment was developed and the results would lead to COSMO-RS [10].

The following virtual experiment is included to better understand the COSMO-RS model.

*If there were a set of cubic solvent molecules S, and a cubic solute X, that were are placed in a conductor, were perfectly screened by surface charges, and exhibited no interaction between the molecules, the molecules could essentially be arranged in any configuration with no resulting change in energy. Next, assume that one face of the solvent cube has a surface charge density of –σ, and the solute cube has a surface charge density of +σ, the solvent cubes can be arranged to surround the solute so that*

*the net charge is 0. In this configuration, the presence or lack of a conductor (substance capable of transferring electric charge) between the molecules becomes irrelevant, and the solute is being screened by the solvent as if it were a conductor. The conclusion of this virtual experiment is that solvents which offer the opposite of the surface charge densities for all faces of the solute will be able to screen the solute equally as well as conductor could and is an explanation for why water behaves as a conductor-like screen for many solutes.* [10]

The behavior of water in this manner was a fundamental shift in how to consider solutions. Instead of looking at solutions as an array of molecules interacting through electric fields and van der Waals interactions, they can be considered as having pairwise van der Waals interactions with adjacent surface charge densities. In order to consider these interactions, COSMO is used to calculate the screening charge density of surface segments on the molecule. For a solute molecule X, COSMO calculates the difference between a molecule's energy in vacuum and in a conducting continuum; this is the ideal screening energy, ΔX. COSMO uses the concept of screening energy to calculate the screening surface charge densities of surface segments around the molecule.

The next challenge is how to model molecules. Real molecules are not cubic or simply shaped, and so a method to identify surface segments was needed. The primary criterion for surface segments was that the screening charge density should be reasonably uniform across a segment, but a segment should have a significantly different screening charge density from its neighboring segments. Using this criterion, a segment size of 3 $Å^2$ was generally acceptable but not necessarily a strict standard [10].

The COSMO-RS model creates a *σ*-profile which represents the probability distribution of the screening charge densities of the surface segments. Only molecular structure and a few basic constants are needed in order to calculate the *σ*-profile, which is a major advantage of using COSMO-RS.

#### 2.1.2.1 Calculation of Chemical Potential

The chemical potential, $\mu_s(\sigma)$, of a surface piece with polarity $\sigma$ is calculated from the probability $p_s(\sigma')$ of finding a certain polarity inside the solvent and uses an interaction term [$E_{int}(\sigma,\sigma')$] that takes into account coulombic, hydrogen bond, van der Waals interactions, and shape [10].

$$\mu_s(\sigma) = -kTln \int p_s(\sigma') exp\left\{-\frac{E_{int(\sigma,\sigma')-\mu_s(\sigma')}}{kT}\right\} d\sigma' \qquad (2.1)$$

This iterative solution using statistical thermodynamics makes COSMO-RS unique from group contribution methods. Some of the assumptions of this approach are that all segments can interact with all other segments, the energy of all segment-segment interactions must be evaluated, and segments with similar σ values will have similar energies [10,11]. When $\mu_s(\sigma)$ has been calculated for all surface segments, the chemical potential of a whole component, solute X, in a system S can be calculated using **Equation 2.2** [10,11]**.**

$$\mu_S^X = \sum_\sigma p^X(\sigma)\mu_S(\sigma) + \mu_{C,S}^X \qquad (2.2)$$

Where $\mu_S^X$ is the chemical potential of solute X in system S, $p^X(\sigma)$ is the probability of finding X of a given polarity σ and is multiplied by the chemical potential of surface

segments, $\mu_S(\sigma)$, and $\mu_{C,S}^X$ is a combinatorial term which accounts for the size and shape of the molecules.

#### 2.1.2.2 Calculation of the Partition Coefficient

COSMO-RS can be used to calculate the chemical potential of solute X in an infinite dilution of a mixture S, and the activity coefficient $\gamma_S^X$ is derived from the chemical potential [11].

$$\gamma_S^X = \exp\left(\frac{\mu_S^X - \mu_1^X}{RT}\right) \tag{2.3}$$

The partition coefficient at thermodynamic equilibrium can then be calculated using the activity coefficients. S is the stationary phase, M is the mobile phase, and X is the solute. The composition of the mixture is needed for the activity coefficient calculation, and so for a biphasic solvent system the phase equilibrium data is needed and can be taken from literature if available, or determined experimentally using gas chromatography or other methods [14].

$$K_X^{SM} = \frac{\gamma_X^M}{\gamma_X^S} \tag{2.4}$$

### 2.1.3 Applications

COSMO-RS was developed as a novel method for predicting thermodynamic properties of liquids, and in particular, various types of partition coefficients [11,25]. It has become widely used in chemistry and chemical engineering for theoretical phase equilibrium calculations, and has been used as a tool for efficient solvent screening to optimize the solvent screening process [11,14,15,26]. COSMO-RS was originally parameterized only

using neutral compounds, but since 2002 it has been able to effectively model ionic liquids by simulating them as mixtures containing cations and anions [11,27,28], with root-mean-square-error values less than 0.5 in comparing to experimental results [29]. COSMO*logic* GmbH &Co. KG, Germany, has developed the Turbomole and COSMOtherm software programs in order to make use of the COSMO-RS model. As COSMO-RS and the COSMOlogic programs continue to expand into new areas, the variety of applications has expanded to include solvent screening [14,15,30,31], ionic liquids and solvent design [29,32], prediction of partition coefficients [13–15,33], studying vapor-liquid equilibrium [34,35], vapor pressure and enthalpy of vaporization [36–39], and prediction of flash points [40].

#### 2.1.3.1 Ionic Liquids and Solvent Design

Ionic liquids (ILs) have become popular in the study of environmentally friendly solvents due to their low vapor pressure, high stability, and high solvent capacity [41–43]. ILs are sometimes referred to as designer solvents because their cations and anions can be carefully selected to create a unique liquid for a designated application [32]. However, there is limited experimental data regarding the properties of the specialized cation/anion mixtures. COSMO-RS was used to investigate how well it could predict thermodynamic properties of ILs and was found to be well-suited for such an application due to the reliance on quantum chemical calculations based on structure rather than experimental data [32]. COSMO-RS was used to calculate molar volume and specific density of 18 1-alkyl-2-methylimidazolium ILs at 298 K and a highly linear relationship

was found between experimental data and the COSMO-RS predicted data, with R = 0.999 and a root mean square error (RMSE) less than 1.7% [32]. COSMO-RS has also been used in a variety of other ionic liquid studies, including screening for green solvents for denitrification [44], predicting cellulose solubility in ILs [19], predicting hydrocarbon solubility in ILs [45], studying physical absorption of $CO_2$ in ILs [46], and predicting enthalpies of vaporization [35].

#### 2.1.3.2 Prediction of Partition Coefficients

COSMO-RS has been used as a tool to predict partition coefficients of a variety of compounds in different solvent systems and applications, including prediction of micelle/water partition coefficients [47], nonpolar organics in water-surfactant systems [33], solutes in polymers [13], and G.U.E.S.S.-mix compounds [16] in solvents used in countercurrent chromatography [14,15].

Solubilization in solutions of micelles have applications in separation of biosynthesis products, and the partition coefficient can be used to measure the solubilization of a solute between the micellar and aqueous phases [47]. By treating micelles as a macroscopic phase in equilibrium with and surrounded by an aqueous phase, then thermodynamic equilibrium properties can be used to evaluate the partitioning of a solute between the two phases. This pseudo-phase approach allows statistical thermodynamic models like COSMO-RS to estimate the chemical potential between the aqueous phase and micelle phase, from which the partition coefficient can be derived. A comparison of the ability of UNIFAC (a group contribution method) and COSMO-RS (an

*a priori* model) to model the partitioning of nonpolar organic solutes in water-surfactant systems concluded that COSMO-RS had an advantage in modeling ions [33]. The study concluded that COSMO-RS can make quantitative predictions that are in reasonable agreement with experimental data and is able to deal with both polar and non-polar organic solutes in ionic and non-ionic surfactant solutions [47].

In 2014, Klamt and COSMOlogic showed that polymers can be treated as solutions of monomers or oligomers in order to predict thermodynamic properties and partition coefficients [13]. COSMOtherm can only represent a polymer as a complete molecule if it has a low degree of polymerization, but there are other ways to model polymers in COSMOtherm, as noted in COSMOtherm version C3.0-Revision 14.01. A few repeat units can be used and capped with appropriate end groups, and studies conducted at COSMOlogic have shown there is little difference when including more repeat units, as shown in an example using polyethylene glycol (PEG) [13]. More importantly, accurate representation of the free volume of the polymer, as polymers typically have less free volume than smaller molecules, is possible. The Bondi van der Waals volumes can be implemented in COSMO to overcome this challenge [13,48]. The study concluded that prediction of partition coefficients can be made using COSMO-RS and COSMOtherm as long as the density, molecular weight, and crystallinity of the polymer is known, and a sufficient free volume estimation can be made [13].

In an effort to reduce the experimental burden for choosing solvent systems for use in counter-current chromatography, COSMO-RS was used to predict the partition coefficient of five different case studies in HEMWAT and ARIZONA solvent systems

[14,49,50]. First, COSMO-RS was used to distinguish between homologues of n-alkylbenzenes and steroids. Next, the partition coefficient of one solute (benzyl alcohol) in many solvent systems was calculated. The fourth case study used phenols in ARIZONA systems of heptane/ethyl acetate/methanol/water at different ratios, and to predict partitioning of G.U.E.S.S. mix compounds. Proposed by Friesen and Pauli, G.U.E.S.S. mix is representative of common compounds found in extracts of natural products [14,16]. Comparing all of the COSMO-RS studies to experimental determination of the partition coefficients via the shake flask method, they found that best solvent system as determined by COSMO-RS was within one of the best solvent system experimentally, and so COSMO-RS can be used as a screening tool for predicting what solvent system will yield the best partitioning and separation of compounds [14,15].

#### 2.1.3.3 Vapor-Liquid Equilibrium

COSMO-RS has been investigated as a method for predicting vapor-liquid equilibria (VLE) in a variety of studies [51,29,52]. COSMO-RS was used to calculate the activity coefficients of 38 compounds at infinite dilution in various ionic liquids [29]. The compounds included alkanes, alkenes, alkylbenzenes, alcohols, polar organics, and chloromethanes. It was found that although COSMO-RS was developed for neutral solvents, it is capable of predicting activity coefficients in ionic liquids with the same accuracy with no adjustment [29], meaning vapor-liquid equilibria can be reliably predicted for solvent systems where experimental data is unavailable. The applicability

of COSMO-RS to binary VLE was evaluated by studying interactions of 136 binary systems using mixtures of alkanes, alkenes, cycloalkanes, alcohols, ethers, ketones, aldehydes, and alkyl benzenes [51]. Most systems had deviations between experimental and predicted phase behavior of less than 5%, and so COSMO-RS is useful in predicting phase behavior of mixtures [51]. In 2012, Guzel & Xu used COSMO-RS to study the liquid-liquid equilibria (LLE) and VLE of solvents containing fatty acid ethyl ester components for biodiesel fuel refining [53]. They reported 2.10% mean deviation between experimental data and COSMO-RS predicted phase equilibria and concluded the usefulness of COSMO-RS in predicting phase equilibria where no experimental data is available [53].

#### 2.1.3.4 Vapor Pressure and Enthalpy of Vaporization

Prediction of vapor pressures and enthalpies of vaporization can be done by breaking solvation free energy into three components: dispersion, cavity formation, and electrostatic contributions, the last of which can be predicted by the COSMO solvation model [37]. Model parameters were determined using 371 pure substances including alkanes, alcohols, ketones, esters, amines, aromatics, and multifunctional compounds [37]. The average accuracy of the comprehensive model was found to be 76%. An improvement on this method uses the property-relationship COSMO statistical associating equation of state (PR+COSMOSAC) [39]. The predicted vapor pressure for 1140 substances found that the average deviation was 1/3 of the original model [39]. The PR+COSMOSAC equation of state has been used for study of vapor-liquid equilibria

[54] and demonstrates the usefulness of COSMO in calculating vapor pressure and enthalpy of vaporization when experimental data and other methods are not readily available. Additionally, COSMO-RS has been used to predict vapor pressures and enthalpies of vaporization for ionic liquids [38] and COSMO has been used in predicting temperature dependent vapor pressures of explosives [36].

#### 2.1.3.5 Flash Points

Knowledge of the flash point of a chemical is important for safety and hazard prevention, and minimum flash point behavior (MFPB) is a phenomenon in which the flash point of a mixture is lower than the flash points of the individual components [40,55]. Although the flash point and MFPB can be determined experimentally, a prediction and estimate of the flash point of a mixture is needed prior to experimental determination. The COSMO-RS and UNIFAC methods were used to calculate the MFPB of methanol-water, ethanol-water, octane-ethanol, and octane-1-butanol mixtures [40]. Although Liaw's mathematical method is commonly used for theoretical determination of the flash point of binary mixtures [56], it is not always able to calculate the MFPB. Comparison of COSMO-RS and UNIFAC in theoretically calculating the MFPB concluded that mixture flash point is a function of flash point, vapor pressure, and activity coefficients, which can be accurately predicted using COSMO-RS. The equation was fit for over 1200 compounds with less than 1% error when compared with experimental data [40]. In conclusion, when COSMO-RS is available, in can greatly improve the prediction of the flash point of mixtures, especially those exhibiting minimum flash point behavior.

### 2.1.4 Comparison to Other Partition Coefficient Prediction Methods

A variety of other methods exist for theoretical prediction of partition coefficients, and these methods continue to evolve as the needs of the chemical industry shift and as new compounds require computational study. Toxicity of compounds and solvents as well as a desire to speed up the solvent selection process drives the development of theoretical methods for prediction partition coefficients. The types of models used to predict partition coefficients can be divided into molar methods and group contribution methods, and these can be compared to COSMO-RS.

#### 2.1.4.1 Molar Methods

The primary distinction of molar methods is that they require molecular characteristics such as molar volume or solubility in order to calculate the partition coefficient. They were common prior to the development of more sophisticated models in the 1990s.

1) Molar volume method: This method was very common prior to the development of more advanced theoretical models. It is based upon the assumption that the chemical potential of compounds is equal in pure liquid solution and in a saturated aqueous solution, and utilizes the following relationship between solubility and molar volume based on Hansen's solubility parameters [57].
2) Solubility and CLOGP: The CLOGP method was first proposed by Hansch as part of his work in developing parameters for quantitative structure-activity relationship (QSAR) equations. Using 1-octanol as the hydrophobic solvent, the 1-octanol/water partition coefficient (log P) could be determined and used to

evaluate molecular hydrophobicity [58–60]. CLOGP, now commercialized and available as a tool in several software programs is particularly popular among biochemists and medical chemists since it has been developed primarily for studying the partitioning of lipids in water.

#### 2.1.4.2 Group Contribution Methods

The primary basis of group contribution methods for calculating the partition coefficient is to consider a compound as a sum of groups and then predict properties of the compounds (such as the activity coefficient or partition coefficient) by analysis and parameterization of the groups making up the compound. The most prominent group contribution method is UNIFAC, but there are many other group and fragment contribution methods. **Table 2.1** compares the difference in the bases and results of these different continuum solvation models.

1) UNIFAC: The UNIFAC group contribution method was first designed as a method to estimate activity coefficients in non-ideal liquid mixtures [61]. UNIFAC united the ideas of functional group solvation and the quasi chemical theory of liquid mixtures (UNIQUAC). UNIQUAC was derived from Guggenheim's quasi-chemical theory and was extended to include solution of different functional groups. In order to do this, parameters were chosen and evaluated for different functional groups. Primarily, size and area of different functional groups was calculated from pure-component molecular structure, and phase equilibrium data for mixtures of functional groups was used to evaluate interaction parameters. This

is the fundamental idea of UNIFAC, Derr and Deal's ASOG model, and other fragment contribution methods: that activity coefficients can be derived from the interactions between structural groups [61,62]. The group interaction parameters are obtained from vapor-liquid equilibria data at the Dortmund Data Bank at Universität Oldenburg, Germany [63,64].

2) Other fragmentation and group contribution methods: like UNIFAC, these methods are based on the idea that activity and partition coefficients can be derived by adding up the interactions between structural or functional groups (fragments of the whole molecule or compound). The idea of group contribution methods originated in 1971 when Leo *et al*. proposed that classical statistical thermodynamics could be used to calculate the chemical potential of structural groups, and that a whole molecule could be represented as the sum of the groups [65]. The atom/fragment contribution method instead considered atom contributions in order to estimate the octanol-water partition coefficient of 130 simple chemical structures [66]. This method is highly accurate but has a very high degree of parameterization. The contributions of second and third-order groups has also been calculated using simple linear regression analysis of 9,560 octanol-water partition coefficient values [67]. Quantum mechanics can similarly be used by creating two parameters: one based on shape, and one to account for size, which in conjunction with an energy parameter were able to determine the octanol-water partition coefficient [68]. The quantum mechanics approach

is still a group contribution method as the size and shape parameters are both based on functional groups.

**Table 2.1.** Comparison of solvation models to calculate the partition coefficient.

| Model | Authors | Basis | Results/Examples |
|---|---|---|---|
| Molar Volume | Hansen *et al.* 2007 | Assumes chemical potential of pure liquids is equal to saturated solutions; uses the relationship between solubility and molar volume based on Hansen's solubility parameters | Simple method for calculation of chemical potential and partition coefficient when some experimental data is available |
| CLOGP | Hansch *et al.* 1968; Ghose *et al.* 1997; Hansch *et al.* 1998. | Uses the 1-octanol as the hydrophobic solvent to determine the 1-octanol/water partition coefficient (log P) and evaluate hydrophobicity | Popular in biochemistry and medicine for studying partitioning of lipids in water. |
| UNIFAC and other group contribution methods | Derr *et al.* 1969; Fredenslund *et al.* 1975; Leo *et al.* 1971; Meylan *et al.* 1995; Merrano *et al.* 2002 | A molecule can be considered the sum of its functional groups, and functional group interactions can be used to model the behavior of the molecule in solution. | Requires some experimental data on functional groups; lacks sensitivity. |
| COSMO-RS | Klamt 1995 | Small surface segments can be used to create the surface of the molecule with little to no experimental data required. | Higher sensitivity, more capable of dealing with a variety of polar (or ionic) and nonpolar systems. |

#### 2.1.4.3 Comparison of COSMO-RS to Other Theoretical Methods

Presently, UNIFAC and COSMO-RS are competing models for calculation of partition coefficients. Competing researchers and business interests lead to both sides claiming superiority [69–72], but numerous independent studies have often found the two produce similar results [47,33,40], with some citing advantage of COSMO-RS in dealing with ionic or polar solutions. COSMO-RS is not unlike UNIFAC and other fragmentation methods, in that it does, in a sense, create a network of groups in order to calculate the activity coefficient. However, COSMO-RS is unique in that it considers the surface of the molecule to be made up of very small surface segments, rather than large functional groups. This increases the sensitivity of the charge density of the molecule, and can therefore improve the accuracy of the calculation of the activity coefficient [69,71]. Localization and scaling of the screening charge density as opposed to considering the dielectric boundary constant of a functional group leads to an advantage in sensitivity of the COSMO-RS model and an improved analysis of non-neutral thermodynamic behavior. When using COSMOtherm, the user also has the choice to use the most dominant conformer of a molecule or use a weighted Boltzmann mixture of all the conformers, which can lead to more accurate results.

## 2.2 Silymarins

Silymarins are a group of six compounds most commonly found in milk thistle plants, and are of interest due to their antioxidant properties and long history of being used to treat liver disease [20]. Their antioxidant properties have led to their availability in dietary supplements and recent research has investigated the use of silymarins in canter treatment therapies [73]. Silymarins are an excellent starting point for evaluating the ability of modeling software to predict partition coefficients of biomolecules, due to previous studies on fractionation of silymarins using countercurrent chromatography methods [6,7]. There is interest in using COSMO-RS to predict the partition coefficients of the six silymarin compounds in CCC solvent systems, and try to improve upon and fine-tune the previously studied solvent systems.

### 2.2.1 Plant Sources

Silymarins are most commonly found in the seeds of milk thistle (*Silybum marianum* L.), which is an herbaceous weed native to the Mediterranean [74] and found in many regions around the world [75]. Milk thistle seeds are 28-34% fat, and 0.1% flavonoids, vitamins, trace elements, and other components, and flavonolignans make up 4% of the flavonoids [75]. Milk thistle has been used to treat liver disorders for over 2000 years [20] and silymarins and milk thistle extracts are used for similar purposes and as antioxidants in herbal remedies and dietary supplements [76,77].

### 2.2.2 Structures

There are six primary compounds in the silymarin: silychristin, silydianin, silybin (which has two stereoisomers), and isosilybin (which has two stereoisomers). Their structures are shown in **Figure 2.1**.

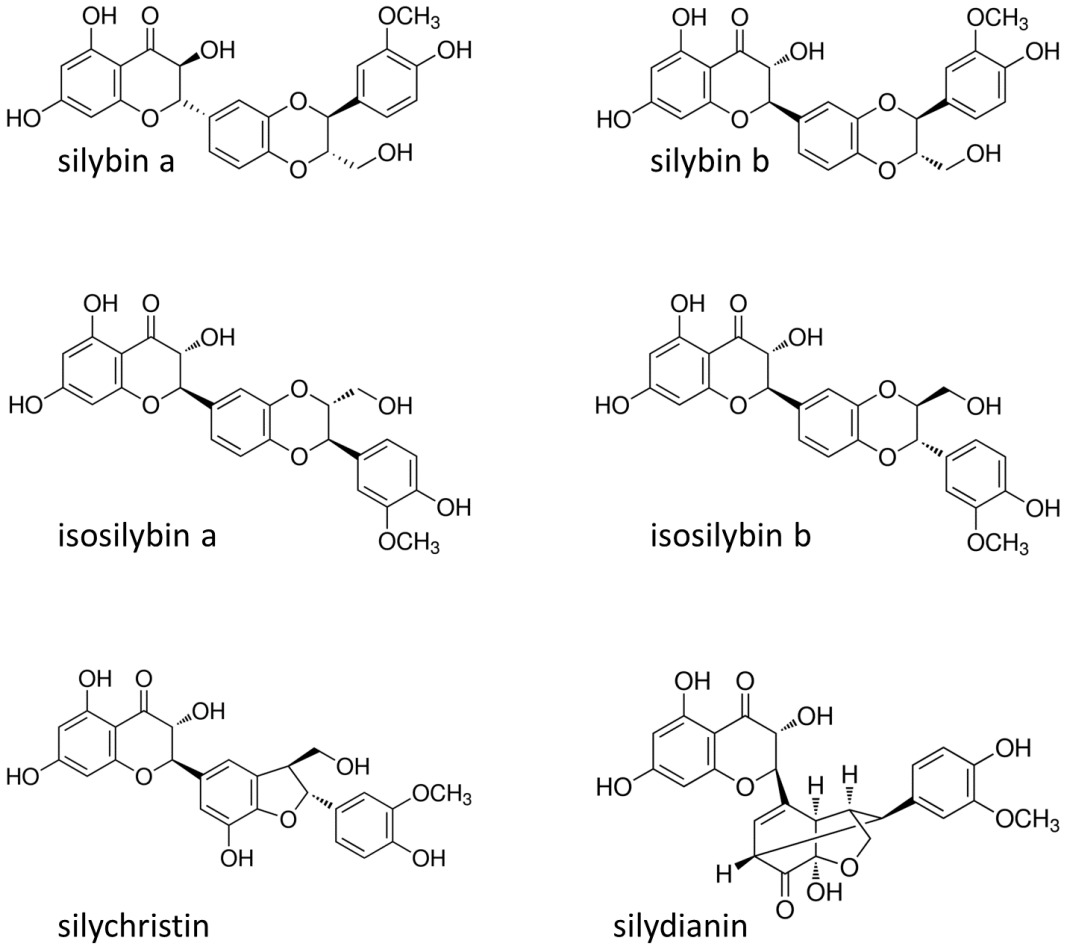


**Figure 2.1.** Structure of six silymarin compounds. Silybin and isosilybin are regioisomers each of which have two stereoisomers.

### 2.2.3 Applications

For over 2000 years, silymarins have been used for treatment of liver disorders [20], and stimulate liver regeneration while preventing hepatotoxins from entering by stabilizing

the cell membrane [78]. Milk thistle has also been studied as a cancer preventative [79,80] or anti-cancer drug [81,82], and studies have been done on isolation of silymarins from cultured cells of *S. marianum* [74]. Milk thistle is commonly marketed as an antioxidant and dietary supplement [77]. The solid wastes of the milk thistle fruits from silymarins production has also been studied for use in feed quality improvement, and it has been shown that it can be used with *Aspergillus niger* and *Candida tropicalis* in fermentation and will result in better blood serum biochemistry, including higher high density lipoprotein cholesterol, blood urea nitrogen, and improved immune responses [83].

### 2.2.4 Separation and Purification

Silymarins can be extracted from milk thistle using hot water extraction [6,84,85] or organic solvents [86]. Pressurized hot water extraction of silymarins has been well studied by Carrier *et al*. [6,84,85]. Hot water extraction and pretreatment methods are becoming more popular in biomass processing due to an increased desire to use green solvents and avoid harsh organic chemicals. Hot water extraction at 140 °C for 55 minutes was sufficient for maximum yield of the compounds, but first order degradation kinetics were observed at these conditions [84].

Countercurrent chromatography (CCC) has been a method employed to separate the silymarins into pure product fractions [7]. Using pure standards in a biphasic solvent system of water/methanol/ethyl acetate/n-hexane, HPLC analysis of the fractions showed the following: silychristin at a purity of 93.1%, silybin at a purity of 95.7%, and

isosilybin at a purity of 89.7% [7]. Silydianin was unable to be isolated, as were the stereoisomers of silybin and isosilybin.

Building upon the hot water extraction technique and fractionation via CCC, Engelberth *et al*. [6] used fast centrifugal partition chromatography (FCPC) – a type of CCC – with a heptane/ethyl acetate/methanol/water system at varying solvent ratios to improve upon the yield and purity of fractions. This method proved very useful in separating silydianin from the crude extract. The fractions from FCPC fractionation of the hot water extract resulted in silychristin at 70.2% purity, silydianin at 93.7% purity, isosilybin B at a purity of 96.1%, and a mixture of isosilybin A, silybin A, and silybin B [6]. FCPC separation of pure silymarins standards yielded silychristin at 85.7% purity, silydianin at 62.9% purity, silybin B at 78.6% purity (with the remainder being silybin A), and isosilybin B at 96.1% purity (the remainder being silybin A, B, and isosilybin A) [6]. The use of CPC and various solvent systems by Engelberth *et al*. [6] showed clear improvement on the work by Du *et al*. [7].

Other studies have demonstrated gram scale purification of silymarins, including a study using a combination of flash chromatography and preparative HPLC, achieving greater than 97% purity of silybin A, silybin B, isosilybin A, and isosilybin B from powdered extract [87]. Similarly, a binary-column preparative HPLC was able to achieve greater than 98% purity separation of the four diastereomers from silymarin powder [88].

**Table 2.2** outlines the different methods, solvent systems, and results for chromatographic separation of silymarins.

**Table 2.2.** Chromatography methods for purification of silymarins.

| Authors | Methods | Solvent System | Results |
|---|---|---|---|
| Du *et al.* 2002 | Countercurrent chromatography | Hexane-ethyl acetate-methanol-water | From standards: silychristin at 93.1% purity, silybin at 95.7% purity, isosilybin at 89.7% purity |
| Engelberth *et al.* 2008 | Centrifugal partition chromatography | Heptane-ethyl acetate-methanol-water | From standards: silychristin at 85.7% purity, silydianin at 62.9% purity, silybin B at 78.6% purity, isosilybin B at 96.1% purity |
| Graf *et al*. 2007 | Flash chromatography and preparative HPLC | Methanol-water | >97% purity silybin A, silybin B, isosilybin A, and isosilybin B from pure silymarins |
| Zhao *et al*. 2014 | Binary column preparative HPLC | Methanol-water | >98% purity silybin A, silybin B, isosilybin A, and isosilybin B from pure silymarins |

## 2.3 Centrifugal Partition Chromatography

### 2.3.1 Introduction

Countercurrent chromatography (CCC) was developed in the 1960s by Ito *et al*. [89]in an effort to utilize centrifugal force and biphasic solvent systems for separation of solutes. The technique of countercurrent chromatography is based on fluid dynamics and the theory that the target compound (solute) will be carried by a mobile phase as it passes through a second phase that is held stationary by centrifugal force. The solute will separate based on the difference between the partition coefficients of the solute in

each phase. Centrifugal partition chromatography (CPC) was pioneered by Nunogaki *et al.* [90], and is one of many technologies that grew out of CCC [91]. The original CPC apparatus was intended to simplify the set-up of a CCC apparatus and allow solvents to be continuously pumped into the rotating separation columns of the centrifuge. CPC is based on the idea of the hydrostatic equilibrium system, in which gravity and pressure are balanced to keep a fluid stationary [92]. The centrifugal force generated by a spinning rotor keeps the stationary phase in the cells of the apparatus, while allowing the mobile phase to quickly pass through, which can allow for very fast separation times [91]. The column of a CPC instrument consists of ducts or cells connected by channels. The column rotates about the shaft which creates the centrifugal force field. Meanwhile, the mobile phase is pumped in and passes through the stationary phase and the channels. The CPC can be operated in two modes: ascending and descending. Ascending mode should be used when the less dense upper phase is used as mobile phase; descending mode should be used when the denser lower phase is used as the mobile phase. The ability to choose whether the upper or lower phase is mobile is a major advantage of CPC. Choice of solvent systems and operating modes make CPC a highly flexible technology.

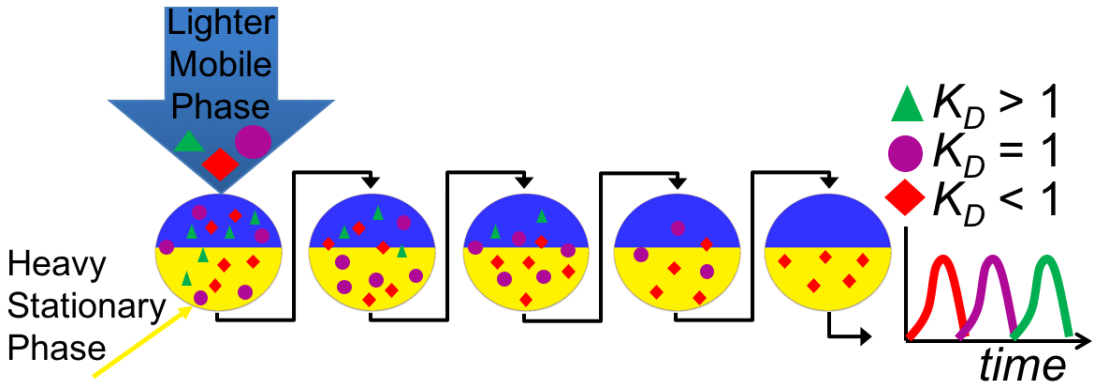


**Figure 2.2.** Centrifugal partition chromatography works by pumping the less dense mobile phase through a series of cells under a centrifugal field; the components will separate based on their affinity for the mobile or stationary phase.

### 2.3.2 Solvent System Selection

CPC separations are dependent on using a bi-phasic solvent system constructed of two immiscible phases. Achieving the best separation is dependent on identifying the best solvent system, and so this must be the first step in any CPC separation. The best solvent system is measured by two parameters: the partition coefficient and the separation factor if there is more than one target compound (multiple solutes). The partition coefficient, $K_D$, is the ratio of the solute's concentration in the upper phase to its concentration in the lower phase at equilibrium [91], as shown in **Equation 2.5**.

$$K_D = \frac{c_U}{c_L} = \frac{concentration\ in\ the\ upper\ phase}{concentration\ in\ the\ lower\ phase} \tag{2.5}$$

The separation factor (α) is defined as the ratio of the partition coefficient ($K_D$) of two components (A and B) in the mixture, as shown in **Equation 2.6**. The partition coefficient indicates which phase the solute prefers, while the separation factor indicates how well multiple solutes can be separated from each other.

$$\alpha_{A/B} = \frac{K_A}{K_B}, where\ K_A > K_B \quad (2.6)$$

The selected solvent system can be binary, ternary, or quaternary [91]. The best solvent should be the one where the solute is most soluble, and the solvents should not decompose any of the solutes [93]. All components of a sample should have acceptable partition coefficients and separation factors if possible; $K_D$ should be between 0.4 and 2.5 and as close to 1 as possible. The separation factor should be greater than 1.5 between all target compounds [31].

Solvent retention time for CPC can be determined by measuring the time needed for the system to form two distinct phases after thorough mixing. Generally, a phase separation time less than 30 seconds indicates the stationary phase will be easily retained, and a phase separation time greater than 30 seconds indicates the stationary phase will be less easily retained , although a CPC settling time greater than 30 seconds is usually fine because the high rotational speed of CPC will help retain the stationary phase [94].

High purity separations rely directly on choosing the solvent system that yields the best partition coefficient and separation factor for the target compound(s), and so the solvent selection step is the most crucial part of the process. A ternary bi-phasic system can be constructed by choosing the solvent the solute is most soluble in as the “best solvent” and then adding a more polar solvent and a less polar solvent to create the two phases [91]. A similar theory based on polarity was used to build the ARIZONA system of quaternary solvent systems [92] containing heptane, ethyl acetate, methanol, and

water. Quaternary solvent systems has also been constructed using the Generally Useful Estimate of Solvent Systems (G.U.E.S.S. mix) and a range of polarities can be chosen using the HEMWat systems containing various ratios of hexane, ethyl acetate, methanol, and water [93].

### 2.3.3 Applications in Natural Products

Almost since its invention, CCC and CPC have been used as techniques for separating pure fractions of natural products. In particular, the heptane-ethyl acetate-methanol-water solvent system has been used for a variety of natural product separations [92].

#### 2.3.3.1 Hexane/Heptane-Ethyl Acetate-Methanol-Water Separations

HEMWat systems (hexane-ethyl acetate-methanol-water) and ARIZONA (heptane-ethyl acetate-methanol-water) systems have been used in centrifugal partition chromatography purification of a variety of natural products, including *Annonaceous acetogenins* from tropical plants of the *Annonacea* family [92], 10-deacetyl-baccatin III from *Taxus baccata* [92], and silymarins from *Silybum marianum* L. [6].

#### 2.3.3.2 Purification of Alkaloids

Alkaloids are nitrogen-containing organic compounds present throughout nature. They have complex ring structures and significant biological activity.

1) pH-zone refining countercurrent chromatography using a binary methyl tertiary butyl ether (MtBE)-water solvent system has been used to isolate alkaloids from *Crinum moorei* [95].

2) pH-zone refining CPC using a MtBE-acetonitrile-water solvent system has been used to isolate vindoline, vindolinine, catharanthin and vincaleukoblastine from *Caranthus roseus* [96].
3) Ascending mode CPC using a MtBE-acetonitrile-water solvent system has been used to isolate lotusine from *Zizyphus lotus* [97].
4) Ascending mode CPC using a chloroform-methanol-acetic acid solvent system has been used to isolate 14-membered cyclopeptides paliurines G, H, and F from *Paliurus ramossisimus* [98].
5) Ascending mode CPC using a MtBE-acetonitrile-water solvent system has been used to isolate chanoclavine and lysergol from *Ipomoea muricata* [99].

#### 2.3.3.3 Purification of Polyphenols

Plant-derived polyphenols are useful for their antioxidant activity and ability to remove free radicals when found with vitamins C, E, and carotene [100]. CPC has also been used to extract phenols from grapes and polyphenols from grape seeds and vines using ethyl acetate [101]. Subsequently, a hexane-ethyl acetate-ethanol-water system was used in CPC to obtain fractions of various polyphenols and flavonols. CPC has also been used for separation of dammarane saponin from *Zizyphus lotus* [102] and glycosides from *Holmskioldia sanguinea* [103].

#### 2.3.3.4 Other Natural Product Separations

1) Slow rotary CCC has been shown to effectively separate solanesol from tobacco leaves extract in a non-aqueous solvent system [4]. Solanesol is a low-polarity,

non-cyclic alcohol and can be used as a food additive if separated from the tobacco tar, pigments, and other impurities in the tobacco leaves. Use of CCC and a non-aqueous solvent system of sunflower oil and ethanol led to 89% recovery of solanesol at 27% purity.

2) Fast CPC has been used to recover six ginsenosides from ginseng saponins in a heptane-n-Butanol-water solvent system, after extraction using pressurized hot water [5].
3) The use of CPC ternary solvent systems to recover three xylooligosaccharides – xylose, xylobiose, and xylotriose – from hemicellulose has been investigated using COSMO-RS and the shake flask experimental method [104].

# CHAPTER 3. MATERIALS AND METHODS

A major objective of this study is to evaluate the ability of molecular modeling to function as a solvent screening tool for fractionating biomolecules using countercurrent chromatography. In order to do this, the partition coefficients of six silymarin molecules were predicted using the Conductor-like Screening Model for Real Solvents. The shake flask method was used to experimentally determine the partition coefficients for comparison to the model. The theoretical and experimental methods are outlined in this chapter.

## 3.1 Theoretical Method: the Conductor-like Screening Model for Real Solvents (COSMO-RS)

The COSMO-RS method for calculating the partition coefficient is based on the molecular structure of the solute molecules, and on the composition of the phases of the solvent system. **Figure 3.1** outlines the method for theoretical determination of the partition coefficient using COSMO-RS.

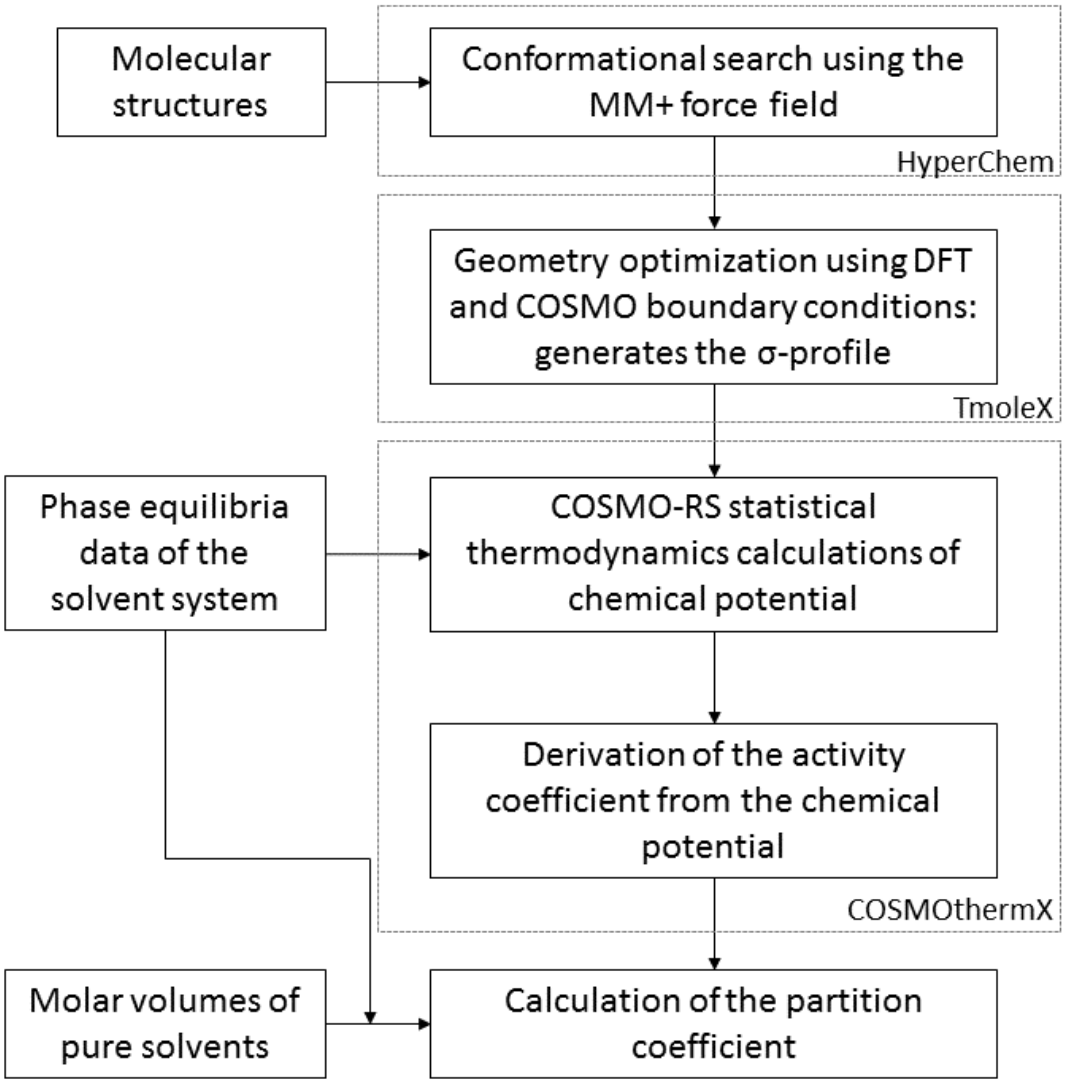


**Figure 3.1.** Hyperchem, TmoleX, and COSMOthermX were used to perform the theoretical calculations leading to the partition coefficient. Molecular structures and molar volumes of pure solvents were obtained from reference standards [105], and the phase composition was determined using gas chromatography. This closely follows a method outlined by the Minceva group [14,15].

### 3.1.1 Molecular Structures and Conformational Search

Because the COSMO-RS calculations are so dependent on the molecular structure, it is important that conformations are considered. Conformers are stereoisomers corresponding to potential energy minima, and are known to have an impact on the

COSMO-RS calculations [33]. HyperChem (release 8.0, Hypercube, Inc.) was used to draw the two-dimensional molecular structure, and then generate the three-dimensional using the Molecular Mechanic + (MM+) force field. The MM+ force field uses the most recent MM parameter sets and includes molecular dynamics calculations. Conformations of each silymarin were generated using the conformational search feature. The number of conformations generated is dependent on the parameters chosen, including limits on the energy range and root mean square error (RMSE). The energy range was set for 0.05-1 kcal/mol and the RMSE range was set to 0.5-2 Å. The number of conformations found for each silymarin was: 24 Silybin A, 19 Silybin B, 14 Isosilybin A, 35 Isosilybin B, 40 Silychristin, 17 Silydianin. To give an idea of how the conformers appear in the program, two conformations of silychristin are depicted in **Figure 3.2**.

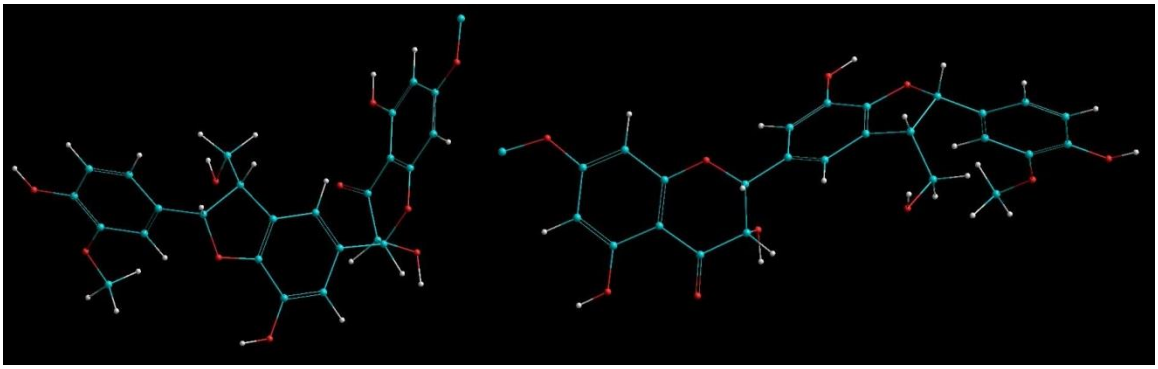

**Figure 3.2**. Example of some conformations of silychristin generated by HyperChem.

3.1.2 Geometry Optimization and Calculation of the Screening Charge Density

The σ-profile representation of screening charge density must be calculated for each conformation. This was completed using the program TmoleX (Version 3.4, COSMO*logic* GmbH &Co. KG, Germany). Within TmoleX, density functional theory using the Becke-

Perdew (B-P) functional and triple zeta valence polarized (TZVP) basis set [25] was applied. B-P is a density functional theory model proposed by Becke [106] and TZVP is a basis for molecular calculations [107]. **Equation 3.1** is used by TmoleX in calculating the average screening charge density $\sigma_m$ of a standard surface segment (*m*) with original screening charge density $\sigma_n^*$ and the average radius of a standard surface segment, $r_{eff}$ [12]. $r_n$ is the radius of segment *n*, and $d_{mn}$ is the distance between segments m and n.

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{eff}^2}{r_n^2 + r_{eff}^2} \exp(-\frac{d_{mn}^2}{r_n^2 + r_{eff}^2})}{\frac{r_n^2 r_{eff}^2}{r_n^2 + r_{eff}^2} \exp(-\frac{d_{mn}^2}{r_n^2 + r_{eff}^2})} \tag{3.1}$$

The average screening charge density, $\sigma_m$, is then used to calculate the $\sigma$-profile, which represents the probability of a surface segment having a screening charge density of $\sigma_m$. The average screening charge density and $\sigma$-profile of a compound, *i,* is calculated using **Equation 3.2** [30], where *$n_i(\sigma)$ is the number of segments with a screening charge density σ, and $n_i$ is the total number of segments.* **Figure 3.3** displays the $\sigma$-profile of one conformation of Silybin A.

$$p^i(\sigma) = \frac{n_i(\sigma)}{n_i} \tag{3.2}$$

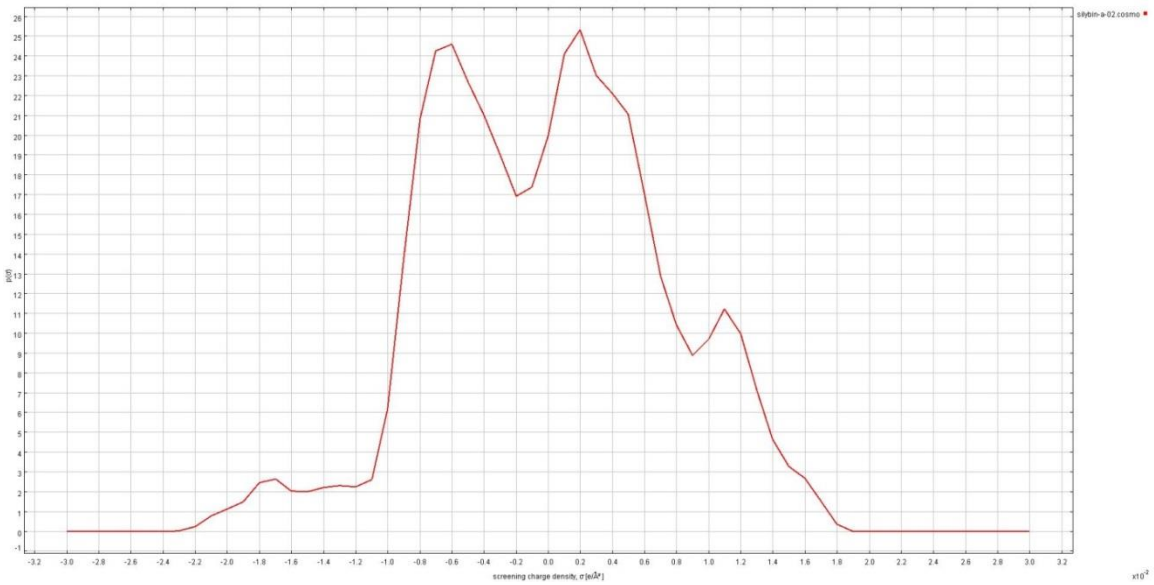


**Figure 3.3.** The σ-profile of one conformation of Silybin A.

A more qualitative way of considering the *σ*-profile is to look at the corresponding COSMO surface, in which blue areas represent likely hydrogen donors and red areas are more likely to be acceptors. The COSMO surface of the same conformation of Silybin A from **Figure 3.3** is shown in **Figure 3.4**.

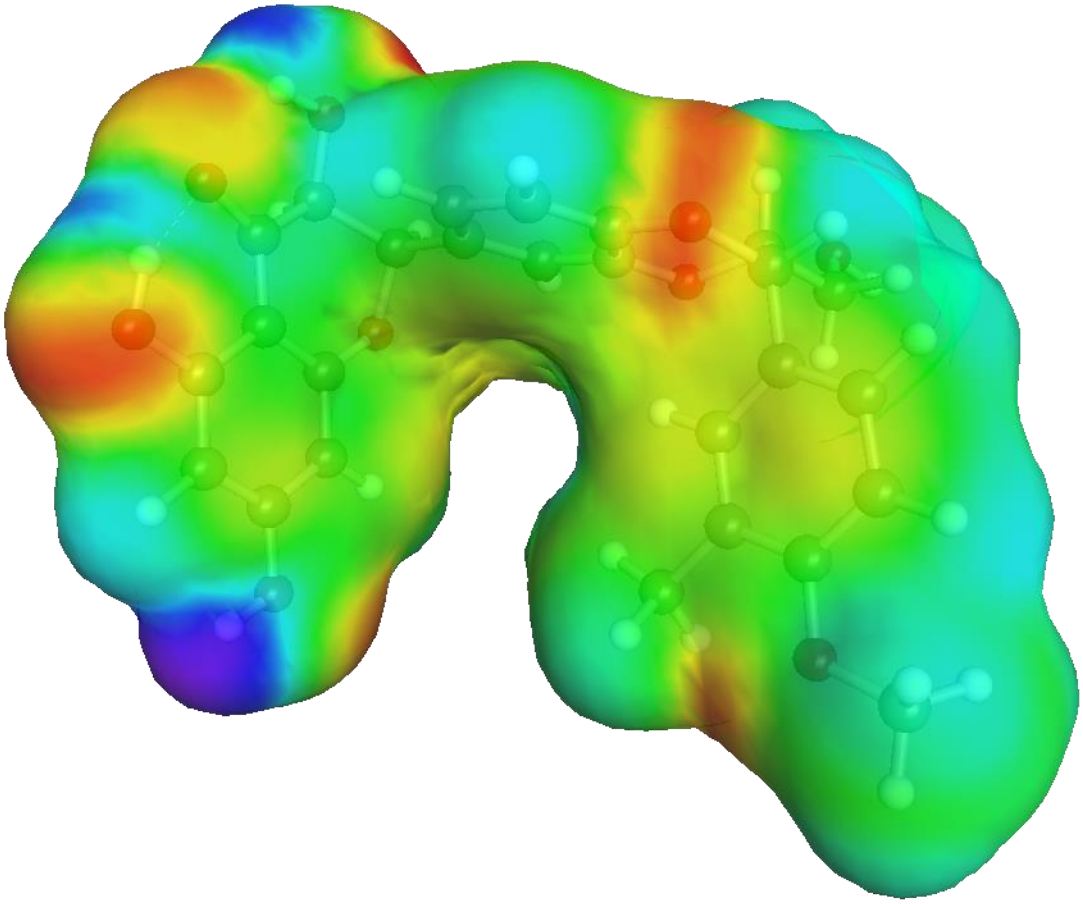

**Figure 3.4.** COSMO surface of a conformation of Silybin A. Blue areas are where the molecule is likely to be a hydrogen donor and red areas are areas where the molecule is likely to be a hydrogen acceptor.

Although the COSMO surface in **Figure 3.4** appears continuous, it is an illustration of all of the tiny surface segments which have their own screening charge density and different potential for interactions, and these surface segments can be quantitatively considered in the $\sigma$-profile or qualitatively examined in the COSMO surface. The $\sigma$-profile is stored and is used in the next step to calculate chemical potential.

### 3.1.3 Calculation of Chemical Potential and Derivation of the Partition Coefficient

COSMOthermX (Version 13, COSMO*logic* GmbH & Co. KG, Germany) was used for calculation of chemical potential and the activity coefficient. The *σ*-profile of a mixture, $p_s(\sigma)$, which contains several solutes (*i*), can be calculated by summing the *σ*-profiles of all components, weighted by their mole fraction, $x_i$, as shown by **Equation 3.3** [10].

$$p_S(\sigma) = \frac{\sum_i x_i p^i(\sigma)}{\sum_i x_i} \tag{3.3}$$

COSMOthermX calculates the chemical potential of compound *i* in mixture *S* according to **Equation 3.4** [12].

$$\mu_S^i = -\frac{RT}{a_{eff}} ln\left[\int p_S(\sigma') exp\left(\frac{a_{eff} \times (\mu_S^i(\sigma') - E_{misfit}(\sigma,\sigma') - E_{HB}(\sigma,\sigma'))}{RT}\right) d\sigma'\right] \tag{3.4}$$

$E_{misfit}$ represents the electrostatic interactions; $E_{HB}$ is hydrogen bonding; $a_{eff}$ is the effective contact area between two surface segments, σ and σ' are the screening charge density of two interacting surface segments. The electrostatic energy in the misfit term is calculated using **Equation 3.5**, and includes the effective contact area between segments, the screening charge densities, and the interaction parameter α'. The hydrogen bonding term is calculated using **Equation 3.6** and takes into account the hydrogen bond strength ($c_{HB}$), the threshold for hydrogen bonding ($\sigma_{HB}$), and the screening charge density of two segments in contact with each other: $\sigma_{donor}$ and $\sigma_{acceptor}$.

$$E_{misfit}(\sigma,\sigma') = a_{eff}\frac{\alpha'}{2}(\sigma + \sigma')^2 \tag{3.5}$$

$$E_{HB}(\sigma,\sigma') = a_{eff}c_{HB}\min\left(0;\min(0;\sigma_{donor+}\sigma_{HB})\max(0;\sigma_{acceptor}-\sigma_{HB})\right) \quad (3.6)$$

COSMOthermX then uses **Equation 3.7** to derive the activity coefficient of a solute, *i*, infinitely diluted in solution, *m*, ($\gamma_m^{\infty,i}$) using the chemical potential of solute *i* in solvent *m* ($\mu_m^i$) and the chemical potential of the pure solute ($\mu_i^i$).

$$\gamma_m^{\infty,i} = exp\left(\frac{\mu_m^i - \mu_i^i}{RT}\right) \quad (3.7)$$

Once these calculations were completed in COSMOthermX for a solute in each phase of the binary system, the partition coefficient of solute *i* in the stationary (*S*) and mobile (*M*) phase ($K_i^{SM}$) was calculated using **Equation 3.8** [14].

$$K_i^{SM} = \frac{\gamma_i^M}{\gamma_i^S} \quad (3.8)$$

The partition coefficient $K_i^{SM}$ can be converted to the partition coefficient based on molar fraction ($P_i^{SM}$), which is useful for comparison to experimental data. **Equation 3.9** uses a weighted sum of the molar fraction of solute *i* in each phase ($x_i$) multiplied by the molar volumes of pure compounds ($v_{0j}$), which can be found in the Design Institute for Physical Properties database [105].

$$P_i^{SM} = K_i^{SM} \times \frac{\sum x_i^M v_{0j}}{\sum x_i^S v_{oj}} \quad (3.9)$$

## 3.2 Experimental Methods

### 3.2.1 Chemicals

Pure standards of the six silymarins (purity ≥ 93%) were purchased from Cerrilliant (Round Rock, TX). Powdered silymarin extract was purchased from Sigma-Aldrich (St. Louis, MO). N-Heptane was purchased from Alfa Aesar (Ward Hill, MA). Methanol was purchased from OmniSolv (Gibbstown, NJ). Ethyl acetate and dichloromethane were purchased from J.T. Baker (Center Valley, PA). All solvents were of HPLC grade and water was deionized in house.

### 3.2.2 Equipment

HPLC analyses were performed using a Waters e2695 model equipped with a Waters 2414 Refractive Index Detector and a Waters Symmetry C18 column (Waters, Milford, MA).

Gas chromatography analyses were performed using an Agilent Technologies 7820A system equipped with an Agilent Technologies 7693A auto sampler (Agilent Technologies, Santa Clara, CA).

### 3.2.3 Methods

#### 3.2.3.1 Determination of Phase Composition by Gas Chromatography

The composition of each phase in a biphasic solvent system is needed for the calculations of the activity coefficient in the theoretical COSMO-RS method. In order to determine the composition of each phase at equilibrium, the solvent systems of interest were made up to a volume of 10 mL, mixed for 3 minutes using a vortex genie, and

allowed to settle for a minimum of 2 hours in order to reach equilibrium. 1.5 mL of each phase was removed and added to a GC vial and analyzed using an Agilent HP-5 column (30 m x 0.32 mm x 0.25 µm, 7 inch cage). **Table 3.1** displays the GC parameters.

**Table 3.1.** Parameters for gas chromatography analysis of phase composition.

| **Condition** | **Value** |
|---|---|
| Oven initial temperature | 45 °C |
| Oven initial time | 2 min |
| Oven ramp rate | 20 °C/min |
| Oven final temperature | 70 °C |
| Total run time | 7.0 min |
| Inlet mode | Split |
| Inlet temperature | 200 °C |
| Inlet pressure | 8.50 psi |
| Split ratio | 20:1 |
| Split flow | 35.6 mL/min |
| Average velocity | 30.125 cm/s |
| Carrier gas | Helium |
| Detector | Flame ionization detection (FID) |
| FID Temperature | 250 °C |
| FID gas flow | Hydrogen 30 mL/min; Air 400 mL/min |
| FID makeup gas | Helium, 25 mL/min |

The GC was calibrated for heptane, ethyl acetate, and methanol by mixing each solvent individually with dichloromethane at the following volume percentages: 1, 2, 5, 10, 15, 25, 45, 65, 85, and 100%. A calibration curve was created with two injections at each level. The split mode with a split ratio of 20:1 was used to avoid maximum loading on the flame ionization detector (FID). The phases of the different solvent systems (all varying ratios of heptane, ethyl acetate, methanol, and water) were then run under these same conditions. The heptane, ethyl acetate, and methanol calibration curves were used to determine the volume percent of each solvent in each phase and water

was assumed to make up the remaining volume fraction. Water is difficult to assess using gas chromatography, but this method was found to be suitable, as it was able to accurately reproduce the volume fractions of heptane-ethyl acetate-methanol-water systems in the literature [14]. The phase composition of the different solvent systems can be found in **Table 4.1**.

#### 3.2.3.2 Shake-flask Method

The conventional shake flask method was used to experimentally determine the partition coefficients of the six silymarin compounds. Each solvent system was prepared to a total volume of 10 mL in a 15 mL centrifuge tube. The solvents were mixed using a vortex genie for 2 minutes, and then 50 mg of powdered silymarin extract was added to each tube. The mixtures were vortexed for 3 minutes, and then allowed to settle for 2 hours. For each mixture, 1 mL was removed from the lower phase and added to a glass tube, and then 1 mL was removed from the upper phase and added to a glass tube. Each sample was dried under nitrogen at 70°C. Once fully dried, each sample was reconstituted in 1 mL of methanol and analyzed using HPLC.

#### 3.2.3.3 HPLC Analysis

HPLC analyses were carried out based on a method used by the Carrier group [6,85], using a Waters e2695 model equipped with a Waters 2414 Refractive Index Detector and a Waters Symmetry C18 column (150 mm, 4.6 mm, 5 μm) set at 40°C. Solvent A was 80% water, 20% methanol; Solvent B was 20% water and 80% methanol. A UV detector was set at 290 nm. The mobile phase flow rate was 0.75 mL/min and the

injection volume was 10 µL. The mobile phase started with a ratio of 85:15 solvent A:B over 5 minutes. The ratio changed to 45:55 (solvent A:B) over 15 minutes and held for 20 minutes, before being linearly ramped down to 85:15 (solvent A:B) over 10 minutes (50 minute total run time). The HPLC was calibrated by adding 1 mg of each silymarin compound to 1 mL of methanol. Serial dilutions were carried out to create 6-level calibration curves including the following concentrations: 0.03125, 0.0625, 0.125, 0.25, 0.50, and 1 mg/mL. A 1 mg/mL standard of each silymarin in methanol were individually run to confirm the retention times, and silychristin and silydianin were individually calibrated due to their close and sometimes overlapping retention times. The samples from the shake flask method were analyzed on HPLC using this methodology to determine the concentration of each silymarin compound in each phase. The partition coefficient, $K_D$, is the ratio of the concentration in the upper phase divided by the concentration in the lower phase.

# CHAPTER 4. ENHANCING SILYMARIN FRACTIONATION USING THE CONDUCTOR-LIKE SCREENING MODEL FOR REAL SOLVENTS[1]

## 4.1 Abstract

Silymarins are a class of flavonolignans found in milk thistle (Silybum marianum L.), and have potential applications in medicine due to their antioxidant properties. A significant hurdle for discovery of plant-derived products is the extensive volume of trial-and-error experimentation required to develop an effective purification strategy. To overcome this, a quantum mechanics-based molecular modeling approach known as the COnductor-like Screening MOdel for Real Solvents (COSMO-RS) was used to predict the best two-phase solvent system to purify six silymarins from an aqueous mixture. Previous research has shown these compounds can be fractionated using centrifugal partition chromatography (CPC), but not to an acceptable level of purity. Due to previous incomplete fractionation, the silymarins are ideal for assessing the use of a molecular modeling approach to predict partitioning in a CPC separation. The COSMO-RS model results predicted the partition coefficients in nine solvent systems of various ratios of heptane, ethyl acetate, methanol, and water. Predicted results displayed a

[1] Chapter 4 is intended for submission to the Journal of Chromatography A with the title "Enhancing Silymarin Fractionation using the Conductor-like Screening Model for Real Solvents" by Emma C. Brace and Abigail S. Engelberth.

similar trend to experimental results determined by the shake flask method, with a root-mean-square-error less than 3.6.

## 4.2 Introduction

An increased interest in improving the economics of the biorefinery and developing useful products from lignocellulosic biomass is the driving force behind the need for innovative bioseparations techniques for purification of biomolecules as precursors for commodity chemicals and other products. The U.S. Departments of Energy and Agriculture have each recognized a need to foster the emerging biorefinery industry, which could produce fuels, chemicals, nutraceuticals, and other bioproducts [1,2]. Nutraceuticals are naturally occurring compounds in plants that have properties that make them useful for additives in food or medicine. Biomass-derived nutraceuticals can be found in a variety of plants, but the complexity of the biomolecules and of the mixtures from which they must be removed have stalled this type of research, along with the high cost of separation techniques and long time it takes to research and develop methods for extraction and purification. Some examples of liquid-liquid extraction of biomolecules include solanesol from tobacco [4], ginsenosides from American ginseng [5], silymarins from milk thistle [6,7], and xylose oligomers from switchgrass hemicellulose [8,9,104]. Even more specifically, various forms of countercurrent chromatography have been used to for fractionation of molecules from plant extracts, including alkaloids from *Crinum moorei* [95], indole alkaloids from *Caranthus roseus* [96], lotusine from *Zizyphus lotus* [97].

The first, and arguably most critical step, to successful fractionation of biomolecules in counter-current chromatography is choosing solvents that will yield relatively pure fractions to minimize the need for further downstream processing. Traditionally, biphasic solvent systems can be selected through solubility experiments to identify the solvent the compounds are most soluble in, and then choosing a solvent that is more polar and less polar to create a biphasic ternary solvent system. Ternary phase diagrams can be developed or found in literature in order to select appropriate volume ratios of each solvent, and then the best system can be identified through experimentation. This critical solvent screening and selection step is often very costly due to the use of high volumes of expensive solvents and standards, and requires a significant amount of time. In order to more efficiently and effectively select solvents, the Conductor-like Screening Model for Real Solvents (COSMO-RS) has been investigated as a molecular modeling tool for predicting partition coefficients of molecules in a two-phase solvent system.

## 4.3 Theoretical Method: COSMO-RS

The COSMO-RS method for calculating the partition coefficient requires only the molecular structure of the solute molecules, and on the composition of the phases of the solvent system. While COSMO-RS has many applications in solvent screening and deriving properties of molecules in solution based on calculations of chemical potential, its usefulness in predicting partition coefficients and the methodology for using COSMO-RS to complete those calculations has been demonstrated by Hopmann et al. [14]. COSMO-RS calculations are dependent on the molecular structure and known to be

impacted by use of different energy conformations of molecules [33], so these must be considered. HyperChem (release 8.0, Hypercube, Inc.) was used to draw and generate the three-dimensional structure of silymarins using the Molecular Mechanic + (MM+) force field. Conformations of each silymarin were generated using the conformational search feature. The energy range was set for 0.05-1 kcal/mol and the RMSE range was set to 0.5-2 Å. The number of conformations found for each silymarin was: 24 Silybin A, 19 Silybin B, 14 Isosilybin A, 35 Isosilybin B, 40 Silychristin, 17 Silydianin. The next step was to generate the σ-profile representations of screening charge density for each conformation using the program TmoleX (Version 3.4, COSMO*logic* GmbH &Co. KG, Germany). Within TmoleX, density functional theory using the Becke-Perdew (B-P) functional and triple zeta valence polarized (TZVP) basis set [25] was applied. The σ-profiles generated in TmoleX were then imported into COSMOthermX (Version 13, COSMO*logic* GmbH & Co. KG, Germany) and used for calculation of chemical potential and the activity coefficient. COSMOthermX calculates the chemical potential of compound *i* in mixture *S* according to **Equation 4.1** [12].

$$\mu_S^i = -\frac{RT}{a_{eff}} ln\left[\int p_S(\sigma') exp\left(\frac{a_{eff} \times (\mu_S^i(\sigma') - E_{misfit}(\sigma,\sigma') - E_{HB}(\sigma,\sigma'))}{RT}\right) d\sigma'\right] \quad (4.1)$$

$E_{misfit}$ represents the electrostatic interactions; $E_{HB}$ is hydrogen bonding; $a_{eff}$ is the effective contact area between two surface segments, σ and σ' are the screening charge density of two interacting surface segments. The electrostatic energy in the misfit term is calculated using **Equation 4.2**, and includes the effective contact area between

segments, the screening charge densities, and the interaction parameter α'. The hydrogen bonding term is calculated using **Equation 4.3** and takes into account the hydrogen bond strength ($c_{HB}$), the threshold for hydrogen bonding ($\sigma_{HB}$), and the screening charge density of two segments in contact with each other: $\sigma_{donor}$ and $\sigma_{acceptor}$.

$$E_{misfit}(\sigma, \sigma') = a_{eff} \frac{\alpha'}{2} (\sigma + \sigma')^2 \tag{4.2}$$

$$E_{HB}(\sigma, \sigma') = a_{eff} c_{HB} \min\left(0; \min(0; \sigma_{donor+}\sigma_{HB})\max(0; \sigma_{acceptor} - \sigma_{HB})\right) \tag{4.3}$$

COSMOthermX then uses **Equation 4.4** to derive the activity coefficient of a solute, *i*, infinitely diluted in solution, *m*, ($\gamma_m^{\infty,i}$) using the chemical potential of solute *i* in solvent *m* ($\mu_m^i$) and the chemical potential of the pure solute ($\mu_i^i$).

$$\gamma_m^{\infty,i} = exp\left(\frac{\mu_m^i - \mu_i^i}{RT}\right) \tag{4.4}$$

Once these calculations were completed in COSMOthermX for a solute in each phase of the binary system, the partition coefficient of solute *i* in the stationary (*S*) and mobile (*M*) phase ($K_D$ ) was calculated using **Equation 4.5** [14].

$$K_D = \frac{\gamma_i^M}{\gamma_i^S} \times \frac{\sum x_i^M v_{0j}}{\sum x_i^S v_{oj}} \tag{4.5}$$

The equation uses a weighted sum of the molar fraction of solute *i* in each phase ($x_i$) multiplied by the molar volumes of pure compounds ($v_{0j}$), which can be found in the Design Institute for Physical Properties database [105].

## 4.4 Experimental Methods

### 4.4.1 Reagents and Equipment

Pure standards (purity > 93%) of the six silymarin compounds were purchased from Cerrilliant (Round Rock, TX). Silymarin extract was purchased from Sigma-Aldrich (St. Louis, MO). All solvents were of HPLC grade and water was deionized in house. N-Heptane was purchased from Alfa Aesar (Ward Hill, MA). Methanol was purchased from OmniSolv (Gibbstown, NJ). Ethyl acetate and dichloromethane were purchased from J.T. Baker (Center Valley, PA). All solvents were of HPLC grade and water was deionized in house.

HPLC analyses used a Waters e2695 model and a Waters 2414 refractive index detector and a Waters Symmetry C18 column.

Gas chromatography analyses were performed using an Agilent Technologies 7820A system combined with an Agilent Technologies 7693A auto-sampler (Agilent Technologies, Santa Clara, CA).

### 4.4.2 Experimental Determination of Partition Coefficients

The conventional shake flask method was used to experimentally determine the partition coefficients of the six silymarin compounds in heptane/ethyl acetate/methanol/water solvent systems. Each solvent system was prepared to a volume of 10 mL and mixed using a vortex genie for 2 minutes. 50 mg of powdered silymarin extract was added to each tube and then mixed using a vortex genie for 3 minute and allowed to settle for 2 hours. From each mixture, 1 mL was removed from the lower phase followed by 1 mL from the upper phase, and added to separate glass

tubes. Each sample was dried under nitrogen at 70°C and then reconstituted in methanol for HPLC analysis.

### 4.4.3 HPLC Analysis

HPLC analyses were carried out based on a method used by the Carrier group [6,85], using a Waters e2695 model equipped with a Waters 2414 Refractive Index Detector and a Waters Symmetry C18 column (150 mm, 4.6 mm, 5 µm) set at 40°C. Solvent A was 80% water, 20% methanol; Solvent B was 20% water, 80% methanol. A UV detector was set at 290 nm. The mobile phase flow rate was 0.75 mL/min and the injection volume was 10 µL. After six-level calibration curves were constructed with 2 injections per level, the samples from the shake flask method were analyzed on HPLC using this methodology to determine the concentration of each silymarin compound in each phase. The partition coefficient, $K_D$, is the ratio of the concentration in the upper phase divided by the concentration in the lower phase.

### 4.4.4 Gas Chromatography Analysis of Phase Composition

As a precursor to molecular modeling, the composition of each phase of bi-phasic quaternary solvent systems was determined using gas chromatography. This data is an input for calculation of the activity coefficient of molecules in each phase and derivation of the partition coefficient. Solvent systems were made up to a volume of 10 mL, mixed for 3 minutes using a vortex genie, and were allowed to equilibrate for a minimum of 2 hours. 1.5 mL of each phase was removed and added to a GC vial and analyzed using an Agilent HP-5 column (30 m x 0.32 mm x 0.25 µm, 7 inch cage). The oven initial temperature was 45°C and increased at a rate of 20°C/min to a final temperature of

70°C. The total run time was 7 minutes and the split mode was used with a split ratio of 20:1 to avoid overloading the flame ionization detector (FID), which was held at 250°C. Helium was used as the carrier gas. The GC was calibrated for heptane, ethyl acetate, and methanol by mixing each solvent with dichloromethane at the following volume percentages: 1, 2, 5, 10, 15, 25, 45, 65, 85, and 100%. Calibration curves were created with two injections at each level, and the phases of the different solvent systems were run under these same conditions. The heptane, ethyl acetate, and methanol calibration curves were used to determine the volume percent of each solvent in each phase and water was assumed to make up the remaining volume fraction.

## 4.5 Results and Discussion

### 4.5.1 Solvent System Selection

The general procedure for selecting a solvent system for fractionation of target compounds using centrifugal partition chromatography is to use the “best solvent” approach outlined by Foucault & Chevolot [91]and used specifically by the Minceva group when applying COSMO-RS as a solvent screening method prior to experimentation [14]. This approach identifies the solvent the target compounds are most soluble in and then chooses a less polar solvent and a more polar solvent to create a biphasic system. The development of ARIZONA solvent systems containing varying ratios of heptane, ethyl acetate, methanol, and water used this approach to develop a quaternary biphasic solvent system [92]. Previous studies on fractionation of silymarins using countercurrent chromatography identified methanol as the solvent silymarins are most soluble in, and so ratios of this solvent system based on those studies were chosen

for identifying a system capable of higher purification of the silymarins [6,7]. The previously identified best solvent system was the 1:4:3:4 heptane/ethyl acetate/methanol/water system [6], and so in addition to testing impacts of increasing heptane, the impact of varying the amount of water in the system was of interest due to the impact of water on polarity. The solvent systems chosen are shown in **Table 4.1**, along with the analysis of the phase composition as determined by gas chromatography.

**Table 4.1.1.** Solvent system phase composition determined by gas chromatography.

| Volume Ratio | Upper phase composition (volume %) | | | | Lower phase composition (volume %) | | | |
|---|---|---|---|---|---|---|---|---|
| | **Heptane** | **EtOAc** | **MeOH** | **Water** | **Heptane** | **EtOAc** | **MeOH** | **Water** |
| 1:4:3:2 | 29.09% | 48.46% | 10.94% | 11.52% | 1.38% | 29.55% | 39.15% | 29.93% |
| 1:4:3:3 | 23.30% | 55.23% | 9.96% | 11.52% | 0.41% | 21.55% | 38.60% | 39.43% |
| 1:4:3:4 | 20.48% | 58.95% | 8.62% | 11.95% | 0.12% | 16.09% | 36.20% | 47.58% |
| 1:4:3:5 | 20.36% | 61.48% | 7.15% | 11.01% | 0.06% | 13.01% | 33.72% | 53.22% |
| 1:4:3:6 | 20.24% | 64.66% | 6.12% | 8.99% | 0.03% | 10.85% | 30.72% | 58.41% |
| 1:4:3:7 | 19.81% | 65.17% | 5.20% | 9.79% | 0.02% | 9.65% | 28.59% | 61.74% |
| 1.2:4:3:4 | 23.61% | 59.01% | 8.59% | 8.79% | 0.12% | 15.55% | 36.70% | 47.63% |
| 1.5:4:3:4 | 27.64% | 55.95% | 6.77% | 9.64% | 0.10% | 14.26% | 36.02% | 49.61% |
| 2:4:3:4 | 33.02% | 52.23% | 5.50% | 9.25% | 0.10% | 13.60% | 37.34% | 48.96% |

### 4.5.2 Determination of the Partition Coefficient

After selecting the solvent systems and determining the phase composition via gas chromatography, the partition coefficient ($K_D$) of each silymarin in each solvent system was calculated using the COSMO-RS method. Initially, all conformations of the six silymarin compounds were used, and solvents were selected from the TZVP database available in COSMOthermX. The phase composition was converted from volume fraction to mole fraction – using molecular weights and densities of the solvents – and

input into COSMOthermX to determine the activity of each silymarin in each phase of each solvent system. These initial results were two orders of magnitude away from the expected range for $K_D$ values. After confirming the gas chromatography method for determination of phase composition through repeated experiments and accurately reproducing the volume fractions of other heptane/ethyl acetate/methanol/water systems published by Hopmann et al. 2011 [14], as well as confirming the calculations leading from the activity coefficient to the partition coefficient by replicating Case Study 4 from this same paper, the calculations being performed in COSMOthermX were investigated. In a comparison of the UNIFAC group contribution method and the COSMO-RS model, COSMO-RS was found to have significant error in predicting properties of molecules in aqueous systems, because COSMO-RS takes a simplistic approach to modeling hydrogen bonding and does not account for the fact that only one hydrogen bond can be formed between two donor-acceptor sites [69]. This inaccuracy in hydrogen bond modeling becomes a concern when the system being modeled is aqueous, and in particular when other components of the system – like alcohols – are also capable of forming hydrogen bonds. Based on this analysis, the decision was made to turn off hydrogen bonding in the COSMOthermX model. Additionally, COSMOtherm has an unofficial upper limit of 800 Da molecular weight for solute molecules, and has been found to have difficulties in dealing with larger molecules. Silymarins, with a molecular weight of 482 Da, are significantly larger than other molecules which have been studied for comparison between COSMO-RS prediction of the partition coefficient and experimental results. Studies by Hopmann et

al. primarily included molecules of 100-200 Da, with a few steroids of molecular weight 300-400 studied as well [14]. In a publication on prediction of partition coefficients in polymers by members of COSMOlogic, it was noted that low energy conformations best represented the solubilized form of solutes and polymers, and so only the lowest-energy conformations were used for polymers. Additionally, it was noted that the combinatorial contribution to chemical potential was developed for 'molecules of small and moderate size', and so for this analyses it was switched off. Finally, by using only the lowest energy conformer for each silymarin, and removing the combinatorial contribution and hydrogen bonding, the COSMO-RS model was re-ran and values for $K_D$ were in a more realistic range, as shown in **Figure 4.1**. Ideally, values for $K_D$ are between 0.4 and 2.5, and values for log$K_D$ are between -0.4 and 0.4 for complete fractionation using CPC.

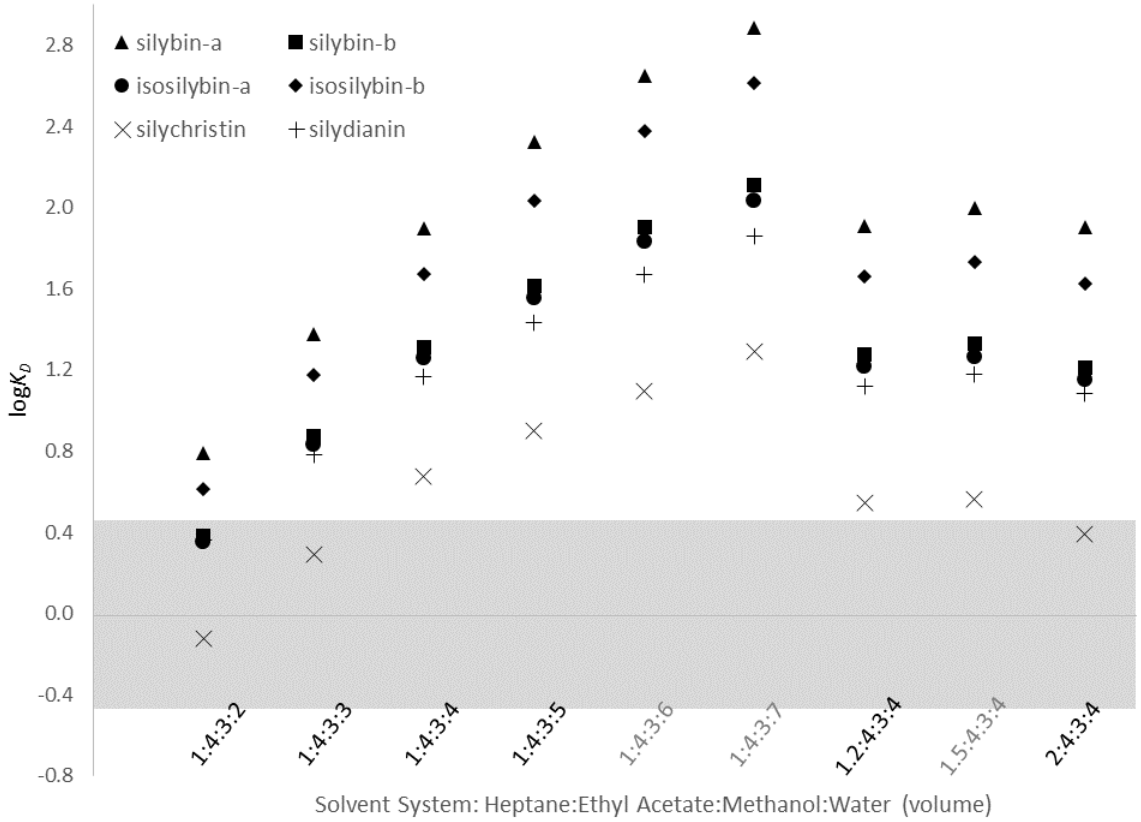


**Figure 4.1.** log$K_D$ values of the six silymarin compounds in solvent systems of varying volume ratios of heptane:ethyl acetate:methanol:water. Log$K_D$ values should be between -0.4 and 0.4 for high purity fractionation using CPC.

The COSMO-RS estimation of the partition coefficient shows an increase in log$K_D$ for all six silymarins as the volume fraction of water is increased, and a slight decrease in log$K_D$ when the volume fraction of heptane is increased.

In order to evaluate the accuracy of the COSMO-RS predictions, the partition coefficients were experimentally determined using the shake flask method and analyzing the concentration of each silymarin in each phase of each solvent system using HPLC. The results of the shake flask method are shown in **Figure 4.2**.

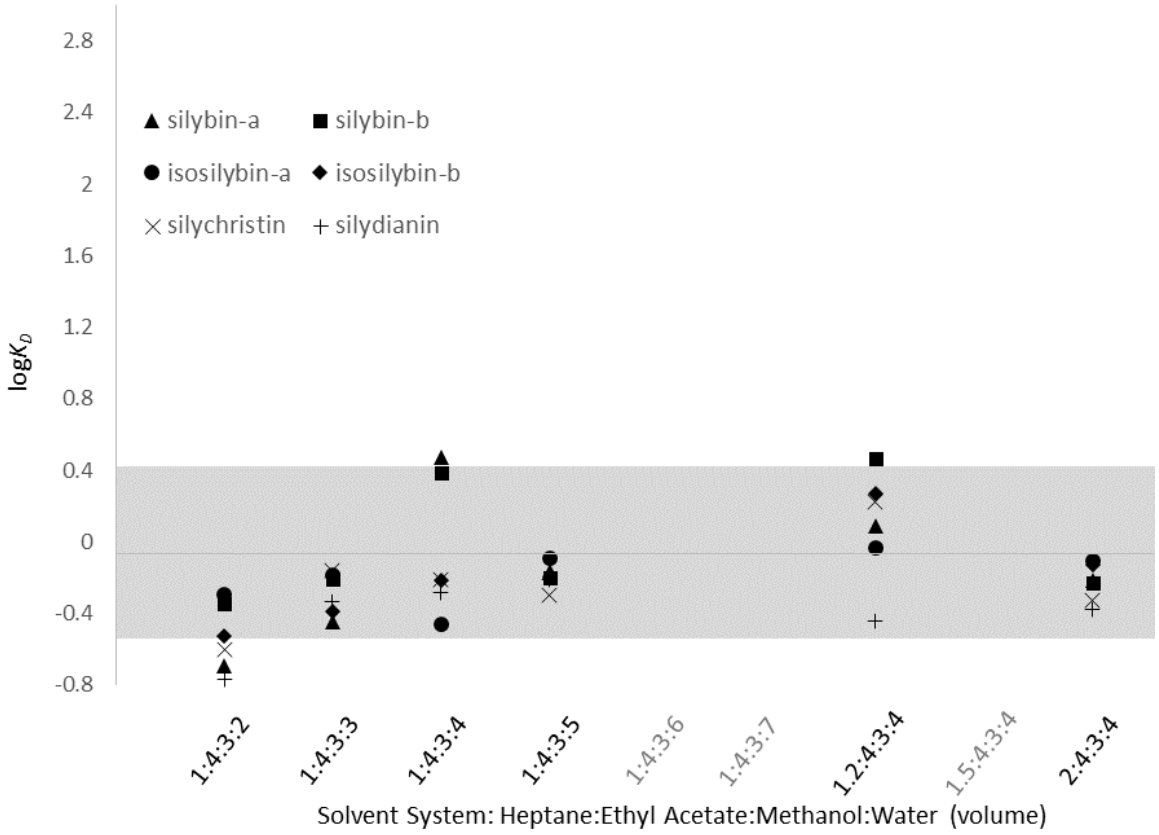


**Figure 4.2.** Partition coefficients of silymarin compounds as experimentally determined via the shake flask method analyzed on HPLC.

As seen in **Figure 4.2**, the partition coefficients determined experimentally via the shake flask method have a similar trend as those determined by COSMO-RS; the relationship between the experimental logK and theoretical logK is further examined in **Figure 4.3**. In the experimental determination of the partition coefficients, there is much less variation in the partition coefficients, as they all fall between -0.8 and 0.4. Experimental partition coefficients for the 1:4:3:5, 1:4:3:6, and 1.5:4:3:4 systems are not included because the HPLC results had very large peaks with no distinction between silychristin and silydianin (which have very close retention times), and in some cases no separation between the isosilybins. Future work should include replicating the shake flask

experiments for all nine solvent systems and using a lower concentration of the silymarins mixture.

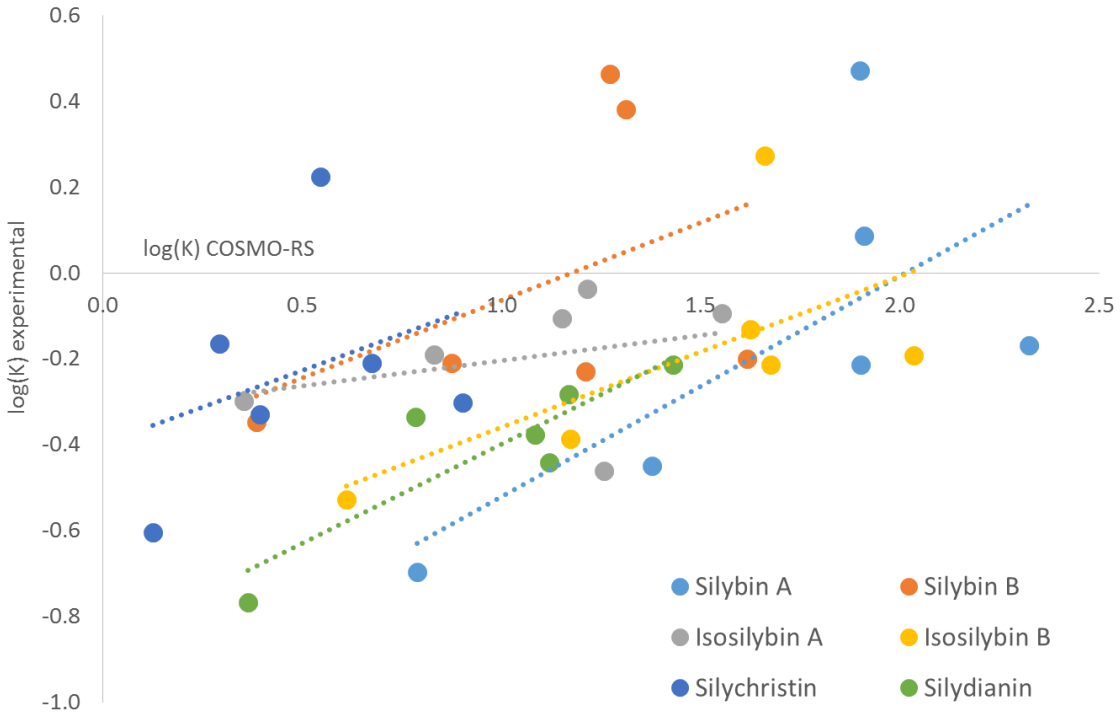


**Figure 4.3.** Qualitative overview of the relationship between experimental and predicted logK for each silymarin. COSMO-RS predicted values are on the x-axis; experimental values are on the x-axis.

**Figure 4.3** shows the relationship between the COSMO-RS predicted log*K* (x-axis) and the shake flask experimental log*K* (y-axis). If the results were equal, a trendline would have a slope of 1. These trendlines have an average slope of 0.36 and an average $R^2$ value of 0.34, unsurprising giving the scatter in the data. The positive relationship between the predicted and experimental log*K* values and confirms that the predictions for each silymarin compounds follow similar trends, which is useful information for beginning to assess the model accuracy. The root-mean-square-error (RMSE) is calculated using **Equation 4.1** to quantitatively evaluate the difference between the

predicted partition coefficients and experimental partition coefficients, and the results are shown in **Table 4.2**.

$$RMSE = \left[\frac{1}{n}\sum_{i}^{n}\left(log_{10}\left(K_D^{experimental}\right) - log_{10}\left(K_D^{predicted}\right)\right)^2\right]^{1/2} \quad (4.1)$$

**Table 4.2.** COSMO-RS predicted log$K_D$ values; experimental log$K_D$ values; the difference between the predicted and experimental log$K_D$ values, and the RMSE for each solvent system.

| Solvent | Silybin A | | | Silybin B | | | Isosilybin A | | | Isosilybin B | | | Silychristin | | | Silydianin | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| System | COSMO-RS | Experimental | ΔlogK | COSMO-RS | Experimental | ΔlogK | COSMO-RS | Experimental | ΔlogK | COSMO-RS | Experimental | ΔlogK | COSMO-RS | Experimental | ΔlogK | COSMO-RS | Experimental | ΔlogK | RMSE |
| 1/4/3/2 | 0.79 | -0.70 | -1.49 | 0.39 | -0.35 | -0.73 | 0.35 | -0.30 | -0.65 | 0.61 | -0.53 | -1.14 | 0.12 | -0.60 | -0.73 | 0.36 | -0.77 | -1.13 | 2.40 |
| 1/4/3/3 | 1.38 | -0.45 | -1.83 | 0.87 | -0.21 | -1.09 | 0.83 | -0.19 | -1.02 | 1.17 | -0.39 | -1.56 | 0.29 | -0.16 | -0.46 | 0.78 | -0.34 | -1.12 | 2.89 |
| 1/4/3/4 | 1.90 | 0.47 | -1.43 | 1.31 | 0.38 | -0.93 | 1.26 | -0.46 | -1.72 | 1.67 | -0.21 | -1.89 | 0.68 | -0.21 | -0.89 | 1.17 | -0.28 | -1.45 | 3.39 |
| 1/4/3/5 | 2.32 | -0.17 | -2.49 | 1.62 | -0.20 | -1.82 | 1.55 | -0.09 | -1.65 | 2.03 | -0.19 | -2.23 | 0.90 | -0.30 | -1.21 | 1.43 | -0.21 | -1.65 | 4.50 |
| 1.2/4/3/4 | 1.91 | 0.09 | -1.82 | 1.27 | 0.46 | -0.81 | 1.22 | -0.04 | -1.25 | 1.66 | 0.27 | -1.39 | 0.54 | 0.22 | -0.32 | 1.12 | -0.44 | -1.56 | 2.92 |
| 2/4/3/4 | 1.90 | -0.21 | -2.12 | 1.21 | -0.23 | -1.44 | 1.15 | -0.11 | -1.26 | 1.62 | -0.13 | -1.76 | 0.39 | -0.33 | -0.72 | 1.08 | -0.38 | -1.46 | 3.57 |

**Table 4.2** offers a quantitative look at the difference between predicted and experimental log$K_D$ values. In a study of 580 compounds in heptane-water, a comparison between COSMO-RS predicted values and experimental values had an RMSE = 1.45 [108]. A value less than this can be considered an acceptable RMSE value. That there is a higher RMSE for all of these solvent systems is unsurprising considering the much wider spread of the COSMO-RS predicted values versus the little variation in the experimentally determined values. The RMSE increase seemingly corresponds to increased volume percent of water in the solvent system, which supports the assertion from the literature that COSMO-RS can fall short when modeling behavior of aqueous systems [69].

Another significant difference between this study and others is the use of six nearly-identical compounds and how they may behavior differently in the right solvent system. Previous studies which compare the partition coefficients predicted by COSMO-RS to those determined by the shake flask method or countercurrent chromatography used molecules of low molecular weight, such as phenols and n-alkyl benzenes [14]. This demonstrated the ability of COSMO-RS to distinguish between stereoisomers such as hydroquinone and pyrocatechol, which is useful, but does not necessarily translate to larger stereoisomers such as the silybins. Additionally, analysis of a GUESS mix containing caffeine, estradiol, coumarin, vanillin, and other molecules showed successful use of COSMO-RS, but this ideal mixture contains molecules of various polarities, molecular weights, and other properties. Successful modeling of the silymarins is challenging not only because they are large molecules in an aqueous

solution but also because they are naturally found together and have nearly identical properties. Without significant differences between the molecules, it can be challenging to parameterize the model to differentiate between them.

## 4.6 Conclusion

The primary objectives of this study were to evaluate the use of COSMO-RS in predicting partition coefficients of the six silymarin compounds in order to quickly eliminate potential solvent systems and hone in on the best one. Although the COSMO-RS predicted partition coefficients appear to be considerably higher than they are in actuality, the trends appear to be similar, which makes the program still useful in identifying a small group of solvent systems that can then be confirmed experimentally. The next steps are to confirm the best solvent system by carrying out fractionation using centrifugal partition chromatography. Additional investigation into the parameterization of the COSMO-RS models and options available in COSMOthermX could lead to better modeling of large, naturally occurring molecules including stereoisomers, like the silymarins.

# CHAPTER 5. CONCLUSIONS AND FUTURE WORK

## 5.1 Conclusions

The first objective of this thesis was to use a theoretical method based on molecular modeling to predict partition coefficients of six silymarins in various solvent systems. COSMO-RS was used to predict the partition coefficients of the six silymarins in nine solvent systems containing varying ratios of heptane, ethyl acetate, methanol, and water. Some of these solvent systems were chosen based on literature, and from there the amount of water in the system was altered in order to manipulate polarity and examine the effect of increased water in the systems. After altering the COSMO-RS methodology to use only the lowest energy conformer of each silymarin, turning off the combinatorial term since it is based on small molecules, and eliminating hydrogen bonding due to COSMO-RS failure to recognize only one hydrogen bond can form between any donor/acceptor site, the predicted partition coefficients fell in a reasonable range.

The second objective was to use COSMO-RS to speed up the solvent screening process and alleviate the experimental burden by honing in on the best solvent systems and testing only those experimentally. Due to the unexpected changes in the COSMO-RS model (mentioned above), all nine solvent systems were tested using the shake flask

method and HPLC analysis to determine the partition coefficient. The experimental results follow a similar trend to the COSMO-RS results, and so although the COSMO-RS results may not be accurate, they can still be useful in identifying a few solvent systems which are most likely to provide the best fractionation.

The third objective was to compare the experimental and predicted results and evaluate the accuracy of the COSMO-RS model. This was done quantitatively by calculating the root-mean-square-error of each solvent system. Although the root-mean-square-error is higher than desirable, this is to be expected given the difficulty in predicting partition coefficients of such large and similar molecules in an aqueous system. Further examination of the COSMO-RS parameters and options within COSMOthermX could lead to a better method for using the COSMO-RS model to predict partition coefficients of complex systems.

## 5.2 Future Work and Centrifugal Partition Chromatography

As was noted in the literature review, centrifugal partition chromatography has many applications in natural products purification, and heptane:ethyl acetate:methanol:water systems are commonly used in CPC. The next step of this study is to fractionate silymarins using these solvent systems in CPC, and validate the model and experimental results. The gap between COSMO-RS and CPC lies in the current inability of COSMO-RS to accurately model large, naturally occurring molecules in aqueous systems, and the fact that these types of molecules and solvent systems are what CPC is best used for. By removing the hydrogen bonding effect in COSMOthermX, and using only the lowest energy conformers of the molecules, progress has been made in improving the accuracy

and ability of the model to predict partition coefficients of these complex molecules and systems. Specifically, future work will include replicating the shake flask experiments in order to determine experimental partition coefficients for all six silymarins in all nine solvent systems. The concentration of silymarins used in the shake flask method (50 mg/10 mL) will be lowered to ensure lower concentrations in the HPLC analysis to avoid overlapping peaks between Silychristin and Silydianin, and Isosilybin A and B. Although the HPLC calibration curves which included known concentrations of all six silymarins in one standard sample were for sufficient for this study, HPLC calibration curves will be constructed for each silymarin individually to improve accuracy and reduce error in the experimental results. What makes this study novel is the use of COSMO-RS to quickly narrow down the infinite possibilities for solvent systems and identify a small range of solvents likely to work. Once confirmed by experimental results using the shake flask method, the final step is to use CPC to fractionate the silymarins and confirm if the solvent system is able to produce high purity fractions of each silymarin, including high purity and high yield separation of the stereoisomers. Based on the data at this point, the CPC would be run using the 1:4:3:2, 1:4:3:3, or 1:4:3:4 solvent systems. The 1:4:3:4 is the best solvent system in the literature [6,7], and the goal is to identify a solvent system that can achieve higher purity fractions and separation of Isosilybin A and B. The CPC will be operated in descending mode, using the organic upper phase (primarily heptane and ethyl acetate) as the stationary phase and the aqueous lower phase as the mobile phase. The solvent system will be prepared in a large separation funnel, mixed, and then the phases separated and promptly used. The stationary phase will be

introduced at 16 mL/min with the rotor set at 200 rpm. When the rotor is completely filled with stationary phase, the mobile phase will be introduced at 4 mL/min with the rotor spinning at 1300 rpm until the column has attained equilibrium. A sample of the powdered silymarin extract constituted in methanol will then be placed in the sample loop for injection onto the column. A UV detector set to 290 nm will monitor the flow out of the column before fractions are collected in a fraction collector; fractions can be collected based on time or volume. The fractions will then be dried and reconstituted in methanol for analysis using HPLC to determine the concentration of different silymarin compounds in each fraction.

REFERENCES

# REFERENCES

[1] R.D. Perlack, L.L. Wright, A.F. Turhollow, R.L. Graham, B.J. Stokes, D.C. Erbach, Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply, 2005. http://www.eere.energy.gov/biomass/publications.html. (accessed May 16, 2016).

[2] R.D. Perlack, L.M. Eaton, A.F. Turhollow Jr, M.H. Langholtz, A.F. Jr, D.J. Jr, J. Richard, D. La Torre Ugarte, J.M. F, E.P. Jr, D. Andrew, R.S. Jr, P. Daniel, U.S. Billion-ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry, (2011). http://lib.dr.iastate.edu/abe_eng_reports (accessed May 16, 2016).

[3] M. Aylott, A. Higson, How will the changing biofuels market affect the development of bio-based chemicals?, Biofuels, Bioprod. Biorefining. 7 (2013) 9–11. doi:10.1002/bbb.1384.

[4] Y. Zhao, Q. Du, Separation of solanesol in tobacco leaves extract by slow rotary counter-current chromatography using a novel non-aqueous two-phase solvent system, J. Chromatogr. A. (2007) 1151–193. doi:10.1016/j.chroma.2007.01.009.

[5] A.S. Engelberth, E.C. Clausen, D.J. Carrier, Comparing extraction methods to recover ginseng saponins from American ginseng (Panax quinquefolium), followed by purification using fast centrifugal partition chromatography with HPLC verification, Sep. Purif. Technol. (2010). doi:10.1016/j.seppur.2009.12.002.

[6] A.S. Engelberth, D.J. Carrier, E.C. Clausen, Separation of Silymarins from Milk Thistle ( Silybum Marianum L.) Extracted with Pressurized Hot Water using Fast Centrifugal Partition Chromatography, J. Liq. Chromatogr. Relat. Technol. (2008). doi:10.1080/10826070802424907.

[7] Q. Du, W. Cai, Y. Ito, PREPARATIVE SEPARATION OF FRUIT EXTRACT OF ***SILYBUM MARIANUM*** USING A HIGH-SPEED COUNTERCURRENT CHROMATOGRAPH WITH SCALE-UP COLUMNS, J. Liq. Chromatogr. Relat. Technol. (2002). doi:10.1081/JLC-120014271.

[8] K. Bunnell, C.-S. Lau, J.O. Lay, J. Gidden, D.J. Carrier, Production and Fractionation of Xylose Oligomers from Switchgrass Hemicelluloses Using Centrifugal Partition Chromatography, J. Liq. Chromatogr. Relat. Technol. 38 (2015) 801–809. doi:10.1080/10826076.2014.973505.

[9] C.-S. Lau, E.C. Clausen, J.O. Lay, J. Gidden, D.J. Carrier, Separation of xylose oligomers using centrifugal partition chromatography with a butanol–methanol–water system, J. Ind. Microbiol. Biotechnol. 40 (2013) 51–62. doi:10.1007/s10295-012-1209-7.

[10] A. Klamt, Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena, J. Phys. Chem. 99 (1995) 2224–2235.

[11] A. Klamt, The COSMO and COSMO-RS solvation models, Wiley Interdiscip. Rev. Comput. Mol. Sci. 1 (2011) 699–709. doi:10.1002/wcms.56.

[12] A. Klamt, V. Jonas, T. Bürger, J.C.W. Lohrenz, T. Bu, J.C.W. Lohrenz, Refinement and Parametrization of COSMO-RS, J. Phys. Chem. A. 102 (1998) 5074–5085. doi:10.1021/jp980017s.

[13] C. Loschen, A. Klamt, Prediction of solubilities and partition coefficients in polymers using COSMO-RS, Ind. Eng. Chem. Res. (2014). doi:10.1021/ie501669z.

[14] E. Hopmann, W. Arlt, M. Minceva, Solvent system selection in counter-current chromatography using conductor-like screening model for real solvents, J. Chromatogr. A. 1218 (2011) 242–250. doi:10.1016/j.chroma.2010.11.018.

[15] E. Hopmann, A. Frey, M. Minceva, A priori selection of the mobile and stationary phase in centrifugal partition chromatography and counter-current chromatography, J. Chromatogr. A. (2012). doi:10.1016/j.chroma.2012.03.035.

[16] J.B. Friesen, G.F. Pauli, GUESS - A generally useful estimate of solvent systems in CCC, J. Liq. Chromatogr. Relat. Technol. 28 (2005) 2777–2806. doi:10.1080/10826070500225234.

[17] J.B. Friesen, G.F. Pauli, Rational development of solvent system families in counter-current chromatography, J. Chromatogr. A. 1151 (2007) 51–59. doi:10.1016/j.chroma.2007.01.126.

[18] C. Balaji, T. Banerjee, V. V. Goud, COSMO-RS Based Predictions for the Extraction of Lignin from Lignocellulosic Biomass Using Ionic Liquids: Effect of Cation and Anion Combination, J. Solution Chem. 41 (2012) 1610–1630. doi:10.1007/s10953-012-9887-3.

[19] J. Kahlen, K. Masuch, K. Leonhard, Modelling cellulose solubilities in ionic liquids using COSMO-RS, Green Chem. 12 (2010) 2172. doi:10.1039/c0gc00200c.

[20] K. Flora, M. Hahn, H. Rosen, K. Benner, Milk Thistle (Silybum marianum) for the Therapy of Liver Disease, Am. J. Gastroenterol. 93 (1998) 139–143. doi:10.1111/j.1572-0241.1998.00139.x.

[21] S.M. Henning, Y. Zhang, V.G. Rontoyanni, J. Huang, R.-P. Lee, A. Trang, G. Nuernberger, D. Heber, Variability in the Antioxidant Activity of Dietary Supplements from Pomegranate, Milk Thistle, Green Tea, Grape Seed, Goji, and Acai: Effects of in Vitro Digestion, (n.d.). doi:10.1021/jf500106r.

[22] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, J. Chem. Soc., Perkin Trans. 2. (1993) 799–805. doi:10.1039/P29930000799.

[23] J. Tomasi ’, M. Persico, Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent Con ten ts, Chem. Rev. 94 (1994) 2027–2094.

[24] A. Klamt, V. Jonas, Treatment of the outlying charge in continuum solvation models, J. Chem. Phys. 105 (1996) 9972. doi:10.1063/1.472829.

[25] A. Klamt, F. Eckert, COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids, Fluid Phase Equilib. 172 (2000) 43–72. www.elsevier.nlrlocaterfluid (accessed May 22, 2016).

[26] A. Frey, E. Hopmann, M. Minceva, Selection of biphasic liquid systems in liquid-liquid chromatography using predictive thermodynamic models, Chem. Eng. Technol. (2014). doi:10.1002/ceat.201400234.

[27] J.E. Bara, J.D. Moon, K.R. Reclusado, J.W. Whitley, COSMOTherm as a Tool for Estimating the Thermophysical Properties of Alkylimidazoles as Solvents for CO 2 Separations, (n.d.). doi:10.1021/ie400094h.

[28] A. Casas, J. Palomar, M.V. Alonso, M. Oliet, S. Omar, F. Rodriguez, Comparison of lignin and cellulose solubilities in ionic liquids by COSMO-RS analysis and experimental validation, Ind. Crop. Prod. 37 (2011) 155–163. doi:10.1016/j.indcrop.2011.11.032.

[29] M. Diedenhofen, F. Eckert, A. Klamt, Prediction of Infinite Dilution Activity Coefficients of Organic Compounds in Ionic Liquids Using COSMO-RS, J. Chem. Eng. Data2. (2003). doi:10.1021/je025626e.

[30] F. Eckert, A. Klamt, Fast solvent screening via quantum chemistry: COSMO-RS approach, AIChE J. 48 (2002) 369–385. doi:10.1002/aic.690480220.

[31] A. Marston, K. Hostettmann, Developments in the application of counter-current chromatography to plant analysis, J. Chromatogr. A. 1112 (2006) 181–194. doi:10.1016/j.chroma.2005.10.018.

[32] J. Palomar, V.R. Ferro, J.S. Torrecilla, F. Rodríguez, Density and Molar Volume Predictions Using COSMO-RS for Ionic Liquids. An Approach to Solvent Design, Ind. Eng. Chem. Res. 46 (2007) 6041–6048. doi:10.1021/ie070445x.

[33] M. Buggert, L. Mokrushina, I. Smirnova, R. Schomäcker, W. Arlt, Prediction of Equilibrium Partitioning of Nonpolar Organic Solutes in Water- Surfactant Systems by UNIFAC and COSMO-RS Models, Chem. Eng. Technol. 29 (2006) 567–573. doi:10.1002/ceat.200600015.

[34] C. Loschen, A. Klamt, Using COSMO-RS for the Prediction of Vapor- Liquid Equilibria, Gas Solubilities and Partition Coefficients in Polymers, (n.d.).

[35] B. Schröder, J.A.P. Coutinho, Predicting enthalpies of vaporization of aprotic ionic liquids with COSMO-RS, Fluid Phase Equilib. 370 (2014) 24–33. doi:10.1016/j.fluid.2014.02.026.

[36] S. Alnemrat, J.P. Hooper, Predicting Temperature-Dependent Solid Vapor Pressures of Explosives and Related Compounds Using a Quantum Mechanical Continuum Solvation Model, J. Phys. Chem. A. 117 (2013) 2035–2043. doi:10.1021/jp400164j.

[37] S.-T. Lin, J. Chang, S. Wang, W.A. Goddard, S.I. Sandler, Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model, J. Phys. Chem. A. 108 (2004) 7429–7439. doi:10.1021/jp048813n.

[38] M. Diedenhofen, A. Klamt, K. Marsh, A. Schäfer, Prediction of the vapor pressure and vaporization enthalpy of 1-n-alkyl-3-methylimidazolium-bis-(trifluoromethanesulfonyl) amide ionic liquids, Phys. Chem. Chem. Phys. 9 (2007) 4653. doi:10.1039/b706728c.

[39] L.-H. Wang, C.-M. Hsieh, S.-T. Lin, Improved Prediction of Vapor Pressure for Pure Liquids and Solids from the PR+COSMOSAC Equation of State, Ind. Eng. Chem. Res. 54 (2015) 10115–10125. doi:10.1021/acs.iecr.5b01750.

[40] M. Vidal, W.J. Rogers, M.S. Mannan, Prediction of Minimum Flash Point Behaviour for Binary Mixtures, Process Saf. Environ. Prot. 84 (2006) 1–9. doi:10.1205/psep.05041.

[41] J.P. Hallett, T. Welton, Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2, Chem. Rev. 111 (2011) 3508–3576. doi:10.1021/cr1003248.

[42] R. Sheldon, Catalytic reactions in ionic liquids, Chem. Commun. 3 (2001) 2399–2407. doi:10.1039/b107270f.

[43] R.D. Rogers, K.R. Seddon, eds., Ionic Liquids as Green Solvents, American Chemical Society, Washington, DC, 2003. doi:10.1021/bk-2003-0856.

[44] R. Anantharaj, T. Banerjee, COSMO-RS-Based Screening of Ionic Liquids as Green Solvents in Denitrification Studies, Ind. Eng. Chem. Res. 49 (2010) 8705–8725. doi:10.1021/ie901341k.

[45] S.S. AlSaleem, W.M. Zahid, I.M. AlNashef, M.K. Hadj-Kali, Solubility of Halogenated Hydrocarbons in Hydrophobic Ionic Liquids: Experimental Study and COSMO-RS Prediction, J. Chem. Eng. Data. 60 (2015) 2926–2936. doi:10.1021/acs.jced.5b00310.

[46] J. Palomar, M. Gonzalez-Miquel, A. Polo, F. Rodriguez, Understanding the Physical Absorption of $CO_2$ in Ionic Liquids Using the COSMO-RS Method, Ind. Eng. Chem. Res. 50 (2011) 3452–3463. doi:10.1021/ie101572m.

[47] L. Mokrushina, M. Buggert, I. Smirnova, W. Arlt, R. Schomäcker, COSMO-RS and UNIFAC in Prediction of Micelle/Water Partition Coefficients, Ind. Eng. Chem. Res. 46 (2007) 6501–6509. doi:10.1021/ie0704849.

[48] A.A. Bondi, Physical Properties of Molecular Crystals, Liquids, and Glasses, 1968. doi:10.1126/science.163.3868.666.

[49] F. Oka, H. Oka, Y. Ito, Systematic search for suitable two-phase solvent systems for high-speed counter-current chromatography, J. Chromatogr. A. 538 (1991) 99–108. doi:10.1016/S0021-9673(01)91626-7.

[50] J.B. Friesen, G.F. Pauli, Rational development of solvent system families in counter-current chromatography, J. Chromatogr. A. 1151 (2007) 51–59. doi:10.1016/j.chroma.2007.01.126.

[51] O. Spuhl, W. Arlt, COSMO−RS Predictions in Chemical EngineeringA Study of the Applicability to Binary VLE, Ind. Eng. Chem. Res. 43 (2004) 852–861. doi:10.1021/ie034009w.

[52] T. Banerjee, M.K. Singh, A. Khanna, Prediction of Binary VLE for Imidazolium Based Ionic Liquid Systems Using COSMO-RS, Ind. Eng. Chem. Res. 45 (2006) 3207–3219. doi:10.1021/ie051116c.

[53] G. Güzel, X. Xu, Phase equilibria (LLE and VLE) of refining operations for enzymatic biodiesel production via quantum mechanical COSMO-RS method, AIChE J. 58 (2012) 3504–3516. doi:10.1002/aic.13731.

[54] C.-M. Hsieh, S.-T. Lin, First-Principles Prediction of Vapor−Liquid−Liquid Equilibrium from the PR+COSMOSAC Equation of State, Ind. Eng. Chem. Res. 50 (2011) 1496–1503. doi:10.1021/ie100781a.

[55] D.A. Crowl, J.F. Louvar, Chemical Process Safety: Fundamentals with Applications, Third, Prentice Hall, 2011.

[56] H.-J. Liaw, Y.-H. Lee, C.-L. Tang, H.-H. Hsu, J.-H. Liu, A mathematical model for predicting the flash point of binary solutions, J. Loss Prev. Process Ind. 15 (2002) 429–438. doi:10.1016/S0950-4230(02)00068-2.

[57] C. Hansen, Hansen Solubility Parameters, CRC Press, 2007. doi:10.1201/9781420006834.

[58] C. Hansch, Quantitative structure-activity relationships and the unnamed science, Acc. Chem. Res. 26 (1993) 147–153. doi:10.1021/ar00028a003.

[59] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods, 17 (1997).

[60] C. Hansch, J.E. Quinlan, G.L. Lawrence, Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids, J. Org. Chem. 33 (1968) 347–350. doi:10.1021/jo01265a071.

[61] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, AIChE J. 21 (1975) 1086–1099. doi:10.1002/aic.690210607.

[62] E.L. Derr, C.H. Deal, Analytical solutions of groups: Correlation of activity coefficients through structural group parameters, Inst. Chem. Eng. Symp. Ser. (1969).

[63] H.K. Hansen, P. Rasmussen, A. Fredenslund, M. Schiller, J. Gmehling, Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension, Ind. Eng. Chem. Res. 30 (1991) 2352–2355. doi:10.1021/ie00058a017.

[64] J. Gmehling, J. Li, M. Schiller, A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties, Ind. Eng. Chem. Res. 32 (1993) 178–193. doi:10.1021/ie00013a024.

[65] A. Leo, C. Hansch, D. Elkins, Partition coefficients and their uses, Chem. Rev. 71 (1971) 525–616. doi:10.1021/cr60274a001.

[66] W.M. Meylan, P.H. Howard, Atom/Fragment Contribution Method for Estimating Octanol–Water Partition Coefficients, J. Pharm. Sci. 84 (1995) 83–92. doi:10.1002/jps.2600840120.

[67] J. Marrero, R. Gani, Group-Contribution-Based Estimation of Octanol/Water Partition Coefficient and Aqueous Solubility, Ind. Eng. Chem. Res. 41 (2002) 6623–6633. doi:10.1021/ie0205290.

[68] S.-T. Lin, S.I. Sandler, Prediction of Octanol−Water Partition Coefficients Using a Group Contribution Solvation Model, Ind. Eng. Chem. Res. 38 (1999) 4081–4091. doi:10.1021/ie990391u.

[69] Z. Xue, T. Mu, J. Gmehling, Comparison of the a Priori COSMO-RS Models and Group Contribution Methods: Original UNIFAC, Modified UNIFAC(Do), and Modified UNIFAC(Do) Consortium, Ind. Eng. Chem. Res. 51 (2012) 11809–11817. doi:10.1021/ie301611w.

[70] Z. Xue, T. Mu, J. Gmehling, Correction to “Comparison of the a Priori COSMO-RS Models and Group Contribution Methods: Original UNIFAC, Modified UNIFAC(Do), and Modified UNIFAC(Do) Consortium,” Ind. Eng. Chem. Res. 51 (2012) 16163–16163. doi:10.1021/ie303175r.

[71] A. Klamt, Comment on “Comparison of the a Priori COSMO-RS Models and Group Contribution Methods: Original UNIFAC, Modified UNIFAC(Do), and Modified UNIFAC(Do) Consortium,” Ind. Eng. Chem. Res. 51 (2012) 13538–13540. doi:10.1021/ie302247q.

[72] J. Gmehling, Z. Xue, T. Mu, Reply to “Comments on ‘Comparison of the a Priori COSMO-RS Models and Group Contribution Methods: Original UNIFAC, Modified UNIFAC(Do), and Modified UNIFAC(Do) Consortium,’” Ind. Eng. Chem. Res. 51 (2012) 13541–13543. doi:10.1021/ie3024142.

[73] J. Bosch-Barrera, J.A. Menendez, Silibinin and STAT3: A natural way of targeting transcription factors for cancer therapy, Cancer Treat. Rev. 41 (2015) 540–546. doi:10.1016/j.ctrv.2015.04.008.

[74] A. Elwekeel, S. AbouZid, N. Sokkar, A. Elfishway, Studies on flavanolignans from cultured cells of Silybum marianum, Acta Physiol. Plant. 34 (2012) 1445–1449. doi:10.1007/s11738-012-0942-x.

[75] N.U. Mirzaeva, T. V Chernenko, S.D. Gusakova, S. Sh Sagdullaev, S. Yu Yunusov, Lipids and Flavolignans from Silybum marianum, Chem. Nat. Compd. 47 (2011) 318–320.

[76] R. Gazak, D. Walterova, V. Kren, Silybin and Silymarin - New and Emerging Applications in Medicine, Curr. Med. Chem. 14 (2007) 315–338. doi:10.2174/092986707779941159.

[77] K. Berginc, Pharmacokinetic interactions between drugs and dietary supplements: herbal supplements, in: K. Berginc, S. Kreft (Eds.), Diet. Suppl. Safety, Effic. Qual., Woodhead Publishing Series in Food Science, Technology and Nutrition, 2014.

[78] F. Fraschini, G. Demartini, D. Esposti, Pharmacology of Silymarin, Clin. Drug Investig. 22 (2002) 51–65. doi:10.2165/00044011-200222010-00007.

[79] N. Bhatia, J. Zhao, D.M. Wolf, R. Agarwal, Inhibition of human carcinoma cell growth and DNA synthesis by silybinin, an active constituent of milk thistle: comparison with silymarin, Cancer Lett. 147 (1999) 77–84.

[80] G. Deep, N.H. Oberlies, D.J. Kroll, R. Agarwal, Identifying the differential effects of silymarin constituents on cell growth and cell cycle regulatory molecules in human prostate cancer cells, Int. J. Cancer. 123 (2008) 41–50. doi:10.1002/ijc.23485.

[81] H.O. El Mesallamy, N.S. Metwally, M.S. Soliman, K.A. Ahmed, M.M. Abdel Moaty, The chemopreventive effect of Ginkgo biloba and Silybum marianum extracts on hepatocarcinogenesis in rats, Cancer Cell Int. 11 (2011) 38. doi:10.1186/1475-2867-11-38.

[82] Y. Dong, J. Tu, Y. Zhou, X. Zhou, B. Xu, S. Zhu, Silybum marianum oil attenuates oxidative stress and ameliorates mitochondrial dysfunction in mice treated with D-galactose, Pharmacogn. Mag. 10 (2014) 92. doi:10.4103/0973-1296.127353.

[83] F. Li, F. Li, T. Zhao, G. Mao, Y. Zou, D. Zheng, M. Takase, W. Feng, X. Wu, L. Yang, Solid-state fermentation of industrial solid wastes from the fruits of milk thistle Silybum marianum for feed quality improvement, Appl. Microbiol. Biotechnol. 97 (2013) 6725–6737. doi:10.1007/s00253-013-5002-y.

[84] L. Duan, D.J. Carrier, E.C. Clausen, Silymarin Extraction from Milk Thistle Using Hot Water, Appl. Biochem. Biotechnol. 114 (2004) 559–568. doi:10.1385/ABAB:114:1-3:559.

[85] J.F.A. Barreto, S.N. Wallace, D.J. Carrier, E.C. Clausen, Extraction of Nutraceuticals from Milk Thistle: I. Hot Water Extraction, Appl. Biochem. Biotechnol. 108 (2003) 881–890. doi:10.1385/ABAB:108:1-3:881.

[86] S.N. Wallace, D.J. Carrier, E.C. Clausen, Extraction of Nutraceuticals from Milk Thistle: Part II. Extraction with Organic Solvents, Appl. Biochem. Biotechnol. 108 (2003) 891–904. doi:10.1385/ABAB:108:1-3:891.

[87] T. Graf, M. Wani, R. Agarwal, D. Kroll, N. Oberlies, Gram-Scale Purification of Flavonolignan Diastereoisomers from *Silybum marianum* (Milk Thistle) Extract in Support of Preclinical *in vivo* Studies for Prostate Cancer Chemoprevention, Planta Med. 73 (2007) 1495–1501. doi:10.1055/s-2007-990239.

[88] W. Zhao, G. Yang, F. Zhong, N. Yang, X. Zhao, Y. Qi, G. Fan, Isolation and purification of diastereoisomeric flavonolignans from silymarin by binary-column recycling preparative high-performance liquid chromatography, J. Sep. Sci. 37 (2014) 2300–2306. doi:10.1002/jssc.201400270.

[89] Y. ITO, M. WEINSTEIN, I. AOKI, R. HARADA, E. KIMURA, K. NUNOGAKI, The Coil Planet Centrifuge, Nature. 212 (1966) 985–987. doi:10.1038/212985a0.

[90] W. Murayama, T. Kobayashi, Y. Kosuge, H. Yano, Y. Nunogaki, K. Nunogaki, A new centrifugal counter-current chromatograph and its application, J. Chromatogr. A. 239 (1982) 643–649. doi:10.1016/S0021-9673(00)82022-1.

[91] A.. Foucault, L. Chevolot, Counter-current chromatography: instrumentation, solvent selection and some recent applications to natural product purification, J. Chromatogr. A. 808 (1998) 3–22. doi:10.1016/S0021-9673(98)00121-6.

[92] P. Duret, A.-I. Waechter, R. Margraff, A. Foucault, R. Hocquemiller, A. Cavé, High-Speed Countercurrent Chromatography : A Promising Method for the Separation of the Annonaceous Acetogenins [1], J. Liq. Chromatogr. Relat. Technol. 20 (1997) 627–635. doi:10.1080/10826079708010949.

[93] J.B. Friesen, G.F. Pauli, G.U.E.S.S.—A Generally Useful Estimate of Solvent Systems for CCC, J. Liq. Chromatogr. Relat. Technol. 28 (2005) 2777–2806. doi:10.1080/10826070500225234.

[94] K. Hostettmann, M. Hostettmann, A. Marston, Preparative Chromatography Techniques, Springer Berlin Heidelberg, Berlin, Heidelberg, 1986. doi:10.1007/978-3-662-02492-8.

[95] Y. Ma, Y. Ito, E. Sokolosky, H.M. Fales, Separation of alkaloids by pH-one refining counter-current chromatography, J. Chromatogr. A. 685 (1994) 259–262.

[96] J.-H. Renault, J.-M. Nuzillard, G. Le Crouérour, P. Thépenier, M. Zèches-Hanrot, L. Le Men-Olivier, Isolation of indole alkaloids from Catharanthus roseus by centrifugal partition chromatography in the pH-zone refining mode, J. Chromatogr. A. 849 (1999) 421–431. doi:10.1016/S0021-9673(99)00495-1.

[97] G. Le Crouéour, P. Thépenier, B. Richard, C. Petermann, K. Ghédira, M. Zèches-Hanrot, Lotusine G: a new cyclopeptide alkaloid from Zizyphus lotus, Fitoterapia. 73 (2002) 63–68. doi:10.1016/S0367-326X(01)00363-X.

[98] H. Lin, C.-H. Chen, K.C.S. Chen Liu, S.-S. Lee, 14-Membered Cyclopeptides fromPaliurus ramosissimus andP. hemsleyanus, Helv. Chim. Acta. 86 (2003) 127–138. doi:10.1002/hlca.200390002.

[99] A. Maurya, R.K. Verma, S.K. Srivastava, Quantitative determination of bioactive alkaloids lysergol and chanoclavine in Ipomoea muricata by reversed-phase high-performance liquid chromatography, Biomed. Chromatogr. 26 (2012) 1096–1100. doi:10.1002/bmc.1753.

[100] H. El Gharras, Polyphenols: food sources, properties and applications - a review, Int. J. Food Sci. Technol. 44 (2009) 2512–2518. doi:10.1111/j.1365-2621.2009.02077.x.

[101] J.-C. Delaunay, C. Castagnino, C. Chèze, J. Vercauteren, Preparative isolation of polyphenolic compounds from Vitis vinifera by centrifugal partition chromatography, J. Chromatogr. A. 964 (2002) 123–128. doi:10.1016/S0021-9673(02)00355-2.

[102] J.-H. Renault, K. Ghedira, P. Thepenier, C. Lavaud, M. Zeches-Hanrot, L. Le Men-Olivier, Dammarane saponins from Zizyphus lotus, Phytochemistry. 44 (1997) 1321–1327. doi:10.1016/S0031-9422(96)00721-2.

[103] E. Helfrich, H. Rimpler, Iridoid glycosides and phenolic glycosides from Holmskioldia sanguinea, Phytochemistry. 50 (1999) 619–627. doi:10.1016/S0031-9422(98)00559-7.

[104] H. Zhang, E.C. Brace, A.S. Engelberth, Selection of a Non-Aqueous Two-Phase Solvent System for Fractionation of XOS Prebiotics using the Conductor-Like Screening Model for Real Solvents, J. Liq. Chromatogr. Relat. Technol. (2016) *Under Review.*

[105] T.E.D.R.P. Daubert, Physical and thermodynamic properties of pure chemicals : data compilation, Hemisphere Pub. Corp., New York, 1989.

[106] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A. 38 (1988) 3098–3100. doi:10.1103/PhysRevA.38.3098.

[107] M.F. Peintinger, D.V. Oliveira, T. Bredow, Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations, J. Comput. Chem. 34 (2013) 451–459. doi:10.1002/jcc.23153.

[108] C. Wittekindt, A. Klamt, COSMO-RS as a Predictive Tool for Lipophilicity, QSAR Comb. Sci. 28 (2009) 874–877. doi:10.1002/qsar.200810175.

VITA

# VITA

Emma C. Brace is originally from Topeka, KS. She received her bachelor's degree in Biological Systems Engineering with a minor in Chemistry from Kansas State University in 2014. During her time at Purdue, Emma has been an active participant in the ABE Graduate Student Association and leadership team member for the Women in Engineering Graduate Mentoring Program. Emma plans to remain at Purdue, working in the Engelberth research group and pursuing a doctoral degree in Agricultural & Biological Engineering.