Development of a Quantitative Structure–Activity Relationship (QSAR) Model relating Solvent Structure to Ibuprofen Crystal Morphology using 2D and 3D Molecular Descriptors

by

John Colin Haser

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 3, 2013

Keywords: crystallization, ibuprofen, descriptors, CAMD, QSAR, PCA

Approved by:

Mario R. Eden, Department Chair, Joe T. and Billie Carole McMillan Professor
Allan E. David, John W. Brown Assistant Professor
Marko Hakovirta, Professor & AC-PABE Director

Abstract

The objective of this thesis is to develop a quantitative structure-activity relationship (QSAR) that relates solvent structure to the morphology of ibuprofen crystals grown within that solvent. Morphology can be quantified by aspect ratio, and ibuprofen aspect ratio data was obtained for crystals grown in 16 different organic solvents. Developing this QSAR requires accurate geometry optimization using empirical force fields to estimate the three-dimensional structure of the solvent molecules. Three different force fields are implemented and their effect on the developed models is analyzed. Next, a combination of 2D and 3D molecular descriptors are calculated using those structures to provide a quantitative representation of the geometry optimized solvent molecules. The descriptor data matrix is then reduced in size for regression into linear models. This stage is executed using Bayesian Information Criterion (BIC) methods and also Principal Component Analysis (PCA) with Principal Component Regression. The final step in the development is to evaluate the predictive capabilities of the resulting models. The QSAR models developed with either technique were all able to fit the training set data and PCA models generally had better predictive capabilities than the models developed using BIC. However, it was also shown that the applicability domain for the models is very small and the predictive capabilities were less than expected. The principal conclusion from this work is that both methods produce models that can fit the training set data, but that additional experimental

data should be obtained to produce better predictive models that could be used for crystallization solvent design for pharmaceutical or other industrial applications.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Mario Richard Eden, for all his help and support in completing my classes and research at Auburn University. I would also like to thank the entire faculty and staff within the Samuel Ginn College of Engineering and the Department of Chemical Engineering, especially Dr. Allan David and Dr. Marko Hakovirta for devoting their time to serve on my thesis committee. In addition, I would like to thank Subin Hada and Robert Herring for their support as I began my research work and helping me complete my thesis. I would like thank Dr. Charles Acquah of the University of Connecticut and Dr. Arunprakash Karunanithi of the University of Colorado Denver for their willingness to supply me with the data used within this work. I would like to thank the Auburn Department of Chemical Engineering and the Department of Energy National Energy Technology Laboratory (DOE-NETL), the Department of Agriculture (USDA-NIFA-AFRI) and the Walt Woltosz Fellowship Program for the financial support of my degree work. I also want to thank my parents, Ed and Peggy Haser, for their help and support throughout my entire academic career. Finally, I would like to thank my fiancée Cortney Crouse for her love and support as I completed my degree.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
AMBER	Assisted Model Building with Energy Refinement
AR	Aspect Ratio
BIC	Bayesian Information Criterion
CAMD	Computer-Aided Molecular Design
FDA	Food and Drug Administration
GAFF	Generalized AMBER Force Field
GA	Genetic Algorithm
GC	Group Contribution
MMFF94	Merck Molecular Force Field
MMFF94s	Merck Molecular Force Field for static molecules
NSAID	Non-Steroidal Anti-Inflammatory Drug
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PRESS	Sum of squares of the prediction errors
R^2	Coefficient of determination
RMSE	Root Mean-Squared Error
RSS	Residual sum of squares

Q^2	Predictive squared correlation coefficient
QM	Quantum Mechanics
QSAR	Quantitative Structure-Activity Relationship
SMILES	Simplified Molecular-Input Line-Entry System
TSS	Total Sum of Squares

Chapter 1: Introduction

The purpose of this thesis work is to develop a quantitative structure-activity relationship (QSAR) to relate ibuprofen crystal aspect ratio (AR) to the structure of the solvents that it is crystallized within. A valid QSAR with strong predictive capabilities can be used to design or select crystallization solvents while minimizing time, cost and environmental impacts associated with experimental work.

Crystallization is a much less-studied separation unit operation compared to vapor-liquid distillation and liquid-liquid extraction [1]. In addition to the lack of study of crystallization processes, there are also more variables that can affect the quality and usefulness of the end-product of this unit operation. The important output variables from distillation columns and liquid-liquid extraction columns are mainly product flow rate and product composition. In crystallization, crystal clarity and morphology are also crucial to the final product quality in addition to flow rate and composition. Along with the increase in product variables, the driving forces behind how crystals grow in solution are not nearly as well-known. In distillation, the difference in boiling points between two compounds is what drives the separation for most mixtures that need to be separated. For crystallization, the interactions between solute and solvent are more difficult to quantify and can vary with every solute-solvent combination.

Crystallization is often used within the pharmaceutical industry; therefore the quality of the end-product is very important because any defects within the crystalline drug product could lead to injury or death of patients. Crystal morphology can have a significant impact on downstream processing of pharmaceutical products and also how the drug is metabolized within the human body [2]. For example, needle-like ibuprofen crystals (high AR) tend to stick to tablet presses and dies much more than plate-like crystals (low AR) [3].

Because crystallization processes tend to be used on much smaller processes than the more traditional separation unit operations, the products are typically value-added specialty chemicals and pharmaceutical products. Therefore, any deviation from the desired product specifications, even for a small quantity of product, can result in a significant loss of revenue due to off-specification product. Issues such as crystal clarity and morphology that are not considered in other separation processes become crucial to meeting specifications for crystallization products.

1.1 Building a CAMD Framework for Crystallization Solvent Design

An application of the QSAR developed within this work would be to use it within a computer-aided molecular design (CAMD) framework to design solvents that can be used to grow crystals of a specific morphology. A pharmaceutical company could utilize a framework of this manner to develop solvents to crystallize a novel drug with a desired morphology. Using a CAMD framework to design an ideal solvent (or solvents) can reduce the need for money- and time-consuming experimental work. As computers continue to gain processor speed and more data becomes available, higher quality models can be developed and implemented. If a CAMD framework can be utilized to develop crystallization solvents that will yield the desired product, thousands of dollars of experimental work can be saved. If a CAMD framework produces five candidate solvents for a given crystallization process, then experiments can be performed using just those solvents. Otherwise, dozens or hundreds of solvents would need to be evaluated experimentally to find an optimal solvent. Saving time is crucial as well, especially within the pharmaceutical industry. In the United States, approval of a drug from the Food and Drug Administration (FDA) can take up to 15 years [4]. If a CAMD framework can reduce the amount of time needed to develop the drug, then pharmaceutical companies can begin to produce the drug and gain FDA approval much quicker. There is currently a push within industry and the

government for corporations to become more sustainable and produce less waste. Companies can reduce their environmental footprint and improve sustainability by replacing experimental design work with CAMD frameworks. Many of the solvents utilized within the bulk and specialty chemicals industry can have negative environmental impacts, so any simulation work that can reduce the amount of solvent waste produced is beneficial environmentally. Finally, it has become easier and cheaper to acquire vast amounts of chemical data, and CAMD frameworks can utilize that data to save time and money and reduce environmental impact.

1.2 Descriptors

In order to relate the structure of solvent molecules to the aspect ratio of the ibuprofen grown within it, solvent structure must be quantified somehow. This can be achieved through the calculation of molecular descriptors. These descriptor values can be arranged into a data matrix and then regressed to fit aspect ratio data. There are several dimensions of descriptors from 0D through 4D. As the dimensionality of the descriptors increase, the calculations behind them become more complex. Many 0D and 1D descriptors, such as atom counts and structural fragment lists, can be determined by looking at a text string of a molecule, but 2D, 3D and 4D descriptors typically require significant computational software. This work utilizes 2D and 3D descriptors to quantify solvent molecules.

1.2.1 2D Descriptors

2D molecular descriptors are calculated simply based on the bonds and identities of atoms within the molecules. Descriptor calculation software can calculate 2D descriptors from simplified molecular-input line-entry system (SMILES) notation which for propylene glycol is CC(O)CO. The calculations can also be made from simple 2D molecular graphs of molecules, and an example is shown in Figure 1.1:

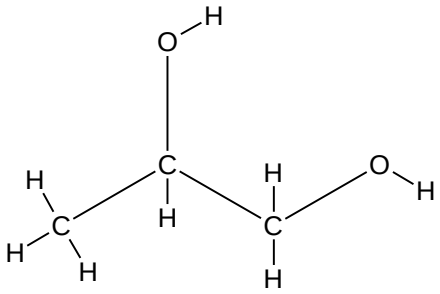


Figure 1.1: Molecular graph of propylene glycol

2D descriptors have a lower computational expense but do not provide as much information as higher dimensionality descriptors.

1.2.2 3D Descriptors

As the name suggests, 3D molecular descriptors account for the position of atoms within a molecule in the x, y, and z directions. Bond lengths, angles and rotations, as well as non-bonded interactions, are used in the calculation of 3D molecular descriptors. The calculation of 3D descriptors requires geometry optimization utilizing molecular force fields, which are further described in Chapter 2, to estimate the location of the atoms within three-dimensional space. Geometrically optimized propylene glycol is shown below in Figure 1.2:

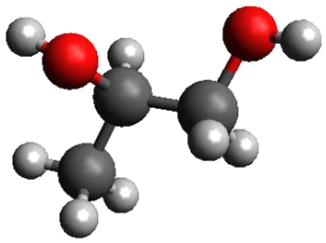


Figure 1.2: 3D representation of propylene glycol

3D descriptors carry a higher computational expense but can provide more information than lower dimensional descriptors.

1.3 Motivation for Using 2D and 3D Descriptors

In this work, a combination of 2D and 3D molecular descriptors are calculated for each solvent molecule. Each class of descriptors has its advantages and disadvantages according to the information provided within that descriptor value. When referring to molecular descriptor values, degeneracy is synonymous with uniqueness. A representation of these advantages and disadvantages is shown in Figure 1.3:

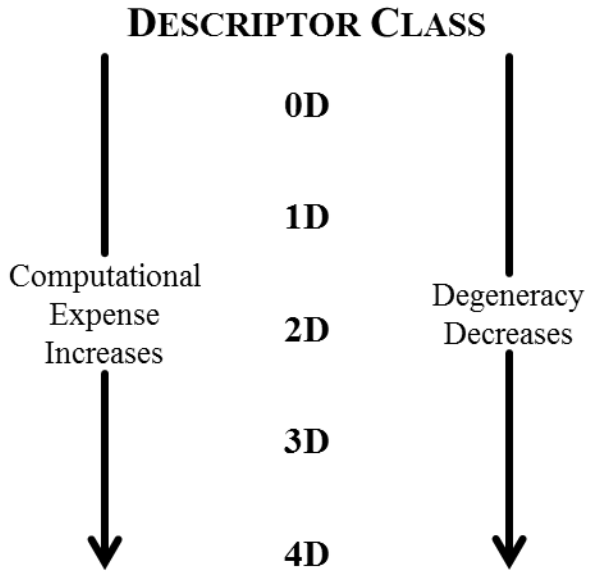


Figure 1.3: Computational expense vs. degeneracy using 0-4D descriptors

0D descriptors, such as atom counts, are very inexpensive to calculate but the resulting descriptor value can be shared with thousands of other molecules. 1D descriptors, such as lists of structural fragments, are more expensive to calculate but the degeneracy of that descriptor value is reduced. The pattern continues for 2D descriptors, like the Wiener Index, with the descriptor being more expensive to calculate but having lower degeneracy. For 3D descriptors such as the 3D-Balaban Index, the calculations become more complex but the degeneracy is further decreased. Finally for 4D descriptors which include conformers with 3D coordinates, the computational expense is great but degeneracy of the resulting descriptor value is very low or zero.

The rationale for using a combination of 2D and 3D descriptors is that they provide a solid middle ground between the 0D and 1D descriptors which do not provide much useful information for this analysis and 4D descriptors that are very difficult to calculate, require specialized software, and could be subject to statistical noise.

1.4 Thesis Outline

Chapter 2 of this thesis details the background work completed in the fields of molecular descriptors, crystallization solvent research and QSAR development. Chapter 3 explains the methodology utilized to develop QSAR models. Chapter 4 presents the results of this methodology in relating solvent structure to crystal aspect ratio. Chapter 5 contains the conclusions that can be drawn from the presented results. Finally, Chapter 6 explores potential future work that could be performed based on the results presented in this thesis.

Chapter 2: Theoretical Background

In this chapter, the previous work in the fields of molecular descriptors, crystallization solvent design and QSAR developments will be presented. With the increase in computational work in many industrial fields, the development of accurate predictive models has become more and more important. Greater and greater amounts of data are becoming available and this data can be harvested and transformed into useful predictive models. The work presented in this chapter will detail the developments in each of these areas and how they relate to the work within this thesis.

2.1 Molecular Mechanics

Utilizing 3D descriptors requires geometry optimization of molecular structures. This can be done using software packages utilizing molecular force fields. These force field methods, also known as molecular mechanics, calculate the energy of a system using nuclear positions and can be effectively used on systems with high numbers of particles. This varies from quantum mechanics (QM) in that QM uses the positions of electrons and is more applicable for smaller systems with fewer particles. There are four key elements to a molecular mechanics force field that determine the 3D structure of a molecule: bond stretching, angle bending, bond torsion, and non-bonded interactions [5]. A functional form for an energy minimization force field is shown in Equation 2.1. The optimized structure will have the lowest potential energy according to this equation.

$$\Upsilon(\boldsymbol{r}^N) = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \quad (2.1)\ [5]$$

$\Upsilon(\boldsymbol{r}^N)$ is the potential energy as a function of the positions of N atoms $\boldsymbol{r}$. l is the bond length ,θ is the bond angle, ω is the torsion angle, V_N is referred to as the barrier height, n is the

multiplicity and γ is the phase factor. The final term of the equation represents the non-bonded electrostatic interactions between point charges (atoms within the molecule) [5].

These four contributions are shown in Figure 2.1:

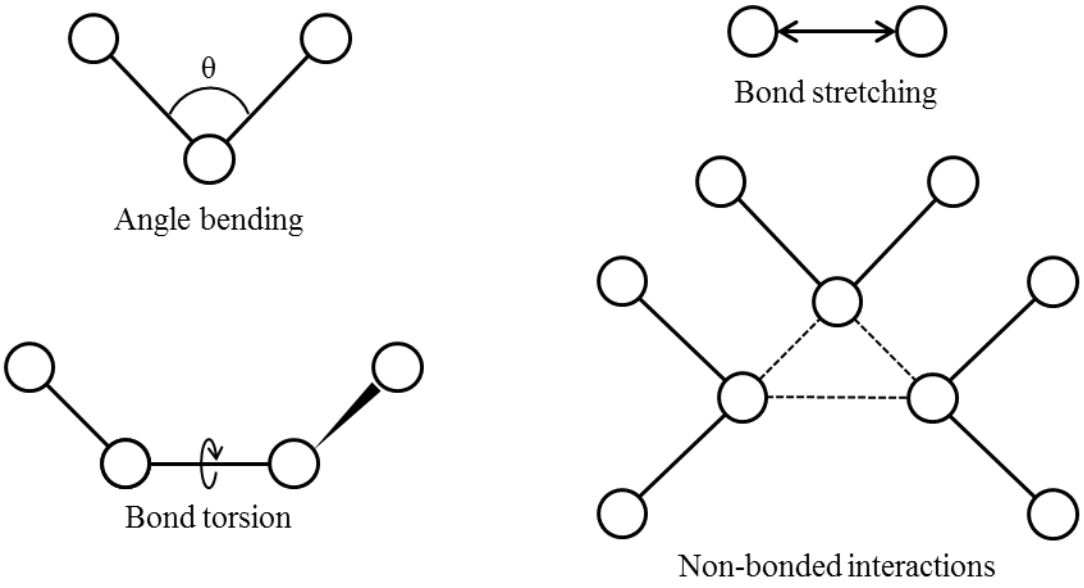


Figure 2.1: Key contributions to a molecular mechanics force field (Adapted from [5])

In order to calculate values for 3D descriptors, geometry optimized structures are needed. Three different force fields were chosen to estimate the three-dimensional structures of solvents. As further described in Chapters 3 and 4, QSAR models were developed using descriptors calculated from structures estimated with each of these force fields and then compared.

2.1.1 GAFF

The Generalized AMBER Force Field (GAFF) was developed by Wang *et al.* in 2003 to extend the Assisted Model Building with Energy Refinement (AMBER) force field to most organic molecules that include the atoms of H, C, N, O, S, P and halogens [6]. This force field was chosen because it was developed with the aim of being applicable to drug-like molecules and the molecules with which they interact.

2.1.2 Ghemical

The Ghemical force field was developed in the computational chemistry software package of the same name. The Ghemical force field was chosen because it is applicable to simple organic molecules, under which all of the solvents used in the experimental work fall.

2.1.3 MMFF94s

The Merck Molecular Force Field (MMFF94) force field was originally introduced in 1996 by Halgren to reproduce a number of molecular properties including: molecular geometries, intermolecular-interaction energies and vibrational frequencies [7]. It was developed to handle the large quantity of molecules within the *Merck Index.* However, the MMFF94 force field was not optimal for energy-minimization studies. Therefore, a static option was developed by Halgren in 1999 and labeled as MMFF94s. The two force fields share many parameters so most molecules that can be handled by MMFF94 can be handled with the static option, called the Merck Molecular Force Field for static molecules (MMFF94s) [8]. This force field was chosen because the compound classes used in its core parameterization are expansive and each of the solvents used in this work fall within that expansive space.

2.2 2D and 3D Descriptors

Utilizing a combination of 2D and 3D molecular descriptors has shown to be promising in predicting the biological targets of ligand probes [9]. In this study by Nettles *et al.*, 2D and 3D molecular descriptors are used to bridge the gap between chemical and biological space by identifying the molecular target for a single chemical entity by evaluating a new compound's activity against structures whose activities are already well known. The application of this work is in the pharmaceutical industry, where any viable method needs to work quickly and able to handle millions of calculations. The speed with which 2D descriptors can be calculated can be

combined with the low degeneracy of 3D descriptors to meet these requirements. The results of these studies show that the slower 3D descriptors can be successfully utilized in conjunction with 2D descriptors. The goals of the work in this thesis are obviously different, but the requirements for useful application are the same which is the reason both 2D and 3D molecular descriptors were utilized to construct the QSAR models.

Other studies have been performed using a combination of 2D and 3D methods to develop QSAR models [10]. This work by Gavernet *et al.* was intended to select new anticonvulsant candidate molecules from a natural product library. The approach used was to first select candidates using solely the quick computing 2D descriptors to save on computational expense. Next, the candidates that made it through the first filters were subjected to the more complex 3D pharmacore superposition process. While this methodology differs significantly from the work presented in this thesis, it does show that combination of both 2D and 3D methods can yield better results in QSAR development than using single dimensionality methods.

Regardless of the dimensionality of the descriptors used, using combinations of descriptor dimensionalities has shown to be consistently more effective than using just one. Helguera *et al.* used a combination of 0D, 1D and 2D descriptors to construct QSAR models for predicting carcinogenic potency of nitroso-compounds [11]. QSAR models using 0D descriptors unsurprisingly had low coefficient of determination (R^2) and predictive squared correlation coefficient (Q^2) values and models with 2D descriptors had significantly higher R^2 and Q^2 values. However, the highest R^2 and Q^2 values were obtained by models using a combination of 0D, 1D and 2D descriptors.

These previous works have shown that employing combinations of 2D and 3D descriptors is more effective than using just one dimensionality of descriptors for the development of QSAR

models. The quick calculation of 2D descriptors combined with the low degeneracy and vast amount of information provided by 3D descriptors are ideal for the construction of QSAR models.

2.3 QSAR Development

The development of QSAR models consists of four basic steps [12]:

1. Calculation of molecular descriptors
2. Descriptor selection for model building
3. Finding an optimal relationship between selected descriptors and target activity
4. Validation of that relationship's predictive capabilities and applicable domain

There are many different methods of analyzing large amounts of molecular descriptor data and constructing QSARs from that data. Two of these methods are explored within this thesis, Bayesian Information Criterion (BIC) and Principal Component Analysis (PCA). The main difference between them is that BIC selects the descriptors that best model the desired activity and eliminates the remainder. PCA calculates principal components (PC) that are linear combinations of the all of the original descriptor values.

2.3.1 BIC

Bayesian information criterion (BIC) is a method of selecting an optimal model from a finite set of models. It was developed by Schwarz in 1978 and is similar to the Akaike Information Criterion (AIC) which was developed in 1974 [13] [14].

BIC recommends selecting a model that maximizes the equation [15]:

$$S = \log L(Y|\theta) - {}^{1}/_{2}\, k \log n \qquad (2.2)$$

where $L(Y|\theta)$ is the likelihood of the data, θ is the vector of model parameters, k is the number of parameters and n is the sample size. When used in linear regression, maximizing S is essentially the same as minimizing BIC in this equation [15]:

$$BIC = n\log\left({}^{RSS}/_{n}\right) + k\log n \qquad (2.3)$$

where *RSS* is the residual sum of squares, calculated from regression.

The second term in the equation is a penalty term that increases each time a parameter is added to the mode. This reduces overfitting by encouraging models to be constructed using fewer parameters.

2.3.2 PCA

Principal component analysis (PCA) is a method that is used to find systematic patterns in data and visualizes multivariate data using as few variables as possible. It maps a large multi-dimensional matrix of data onto lower dimensions with a minimal loss of information. PCA converts a correlated data matrix into a new set of uncorrelated factors that are linear combinations of the original variables. PCA extracts the most important data from the correlated large matrix and creates a smaller matrix of uncorrelated principal component factors that represent a percentage of the variance in the original data. The first factor is a linear combination of the original variables that have the greatest possible variance and each following factor is another linear combination of the original variables that have the greatest possible variance and has zero correlation with the previous factors.

In the past, PCA has been used for the prediction of the mechanism of action of anti-cancer drugs by Lauria *et al.* [16]. PCA was used to reduce a large data matrix containing over 600 descriptor values for 60 compounds within the training set down to five PC factors. These five PC factors are able to cover over 84% of the total variance within the original data matrix. All of the QSAR

models developed are able to attain high R^2 and Q^2 values. In this study, a large matrix of descriptor data was reduced to far fewer uncorrelated factors that contain most of the information from the original matrix and relate it to the desired target activities. The algorithm developed by Lauria *et al.* correlates very strongly with the work presented in this thesis except with a very different application.

2.3.3 Validation Methods

The methods that can be used to validate the predictive ability of a model are numerous. The two that are used within this thesis are commonly used to evaluate the predictive capabilities of QSAR models. A very commonly used method is to use the predictive squared correlation coefficient (Q^2) for leave-one-out cross validations. Q^2 is defined as [17]:

$$Q^2 = 1 - \frac{PRESS}{TSS} \tag{2.4}$$

where *PRESS* is the sum of squares of the prediction errors for the test set, and *TSS* is the total sum of squares which is the sum of squared deviations from the training set mean. Obviously, the highest possible value for Q^2 is one.

In order to evaluate a model's ability to match training set data, the coefficient of determination (R^2) is used. R^2 is defined as [17]:

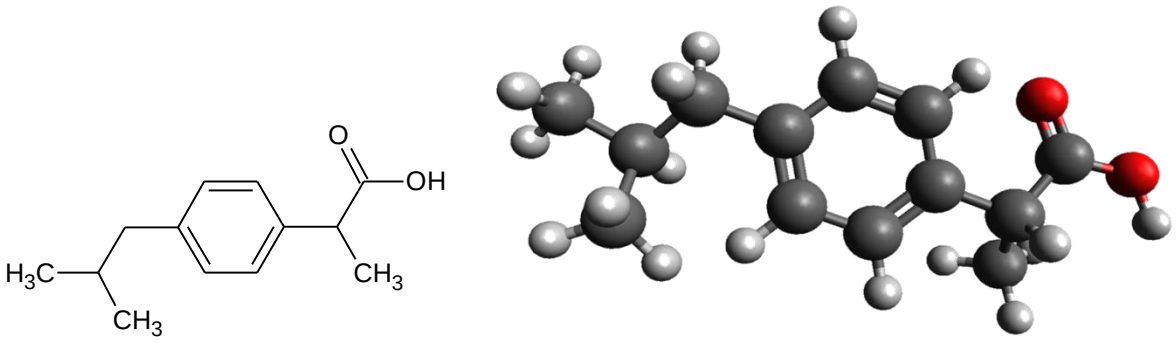
$$R^2 = 1 - \frac{RSS}{TSS} \tag{2.5}$$

where *RSS* is the residual sum of squares from the training set data and *TSS* is the total sum of squares for the training set data. Similar to Q^2, the highest possible value of R^2 is one.

2.4 Crystal Morphology Prediction

Ibuprofen is a common non-steroidal anti-inflammatory drug (NSAID) that is typically used for relief of muscular and skeletal pain. In addition to its anti-inflammatory effects, it has analgesic (pain relieving) and antipyretic (fever reducing) effects. Ibuprofen is sold over-the-counter under

trade names such as Advil and Motrin. Its risks are assumed to be less severe than aspirin and therefore it is a very common and widely applicable drug [18]. Ibuprofen's structure consists of a isobutyl chain and a propionic acid group attached on opposite (carbons 1,4) sides of a phenyl ring. Its full chemical name is iso-butyl-propionic-phenolic acid which is where the common name of ibuprofen is derived. A molecular graph of ibuprofen and also a geometry optimized structure of ibuprofen are shown in Figure 2.2:



(a) **(b)**

Figure 2.2: 2D structure (a) and 3D structure (b) of ibuprofen

The Boots Company, who discovered ibuprofen in the 1960s, developed a method for synthesizing ibuprofen that was used for many years [19]. This process, known as the brown synthesis method, produced the millions of pounds of desired ibuprofen product, but also millions of pounds of undesired byproducts that had to be disposed of. The percentage atom economy, which is a ratio of the molecular weight of ibuprofen divided by the molecular weight of all reactants, for the brown method was about 40 percent, meaning that more waste was produced from the process than ibuprofen product [19].

More recently, the Hoechst Celanese Corporation and the Boots Company agreed to a joint venture known as the BHC Company in order to develop a greener ibuprofen synthesis process [19]. This process has a percentage atom economy of 77 percent, meaning that the amount of unwanted byproducts is significantly reduced with this new method.

This new greener synthesis method consists of three different steps [19]:

1. Acylation of isobutylbenzene using hydrogen fluoride as a catalyst
2. Hydrogenation using Raney nickel catalyst
3. Carbonylation using palladium as a catalyst

Each of the catalysts used within the process are recycled and reused to reduce the amount of waste produced during production. Figure 2.3 shows the three step process commonly used to produce ibuprofen that was described earlier.

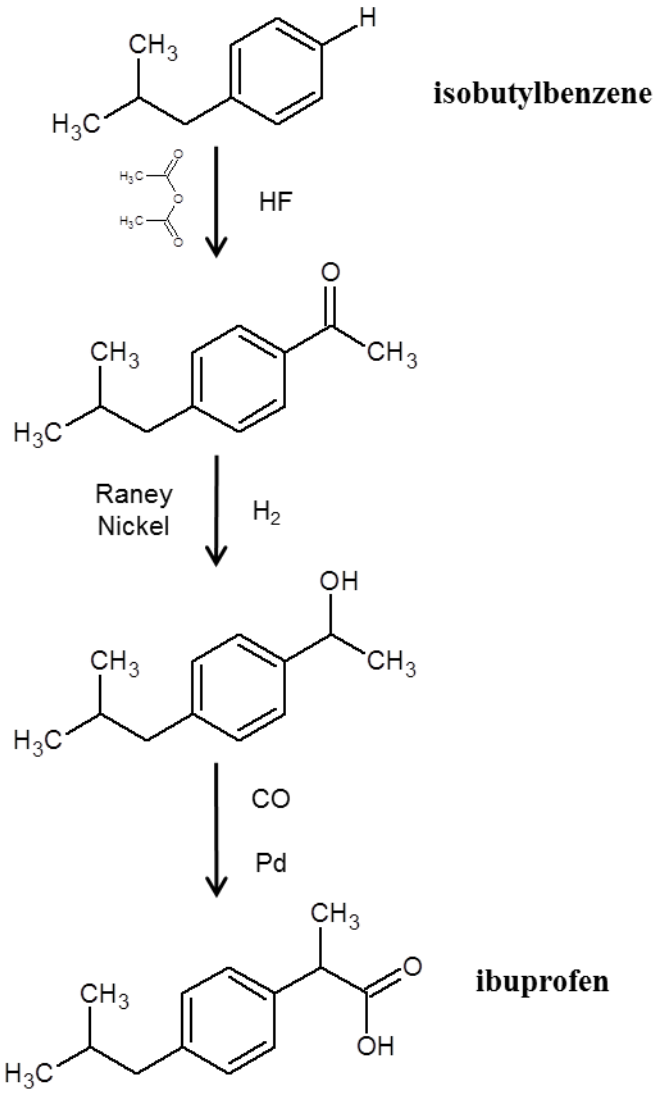
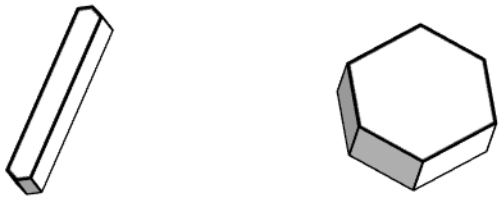


Figure 2.3: Three step ibuprofen production mechanism (Adapted from [19])

Winn emphasized the importance of crystal morphology prediction in relation to industrial and pharmaceutical processes [20]. The morphology of crystals can affect the efficiency of downstream processes for instance filtering, washing and drying. It can also influence material properties such as bulk density and mechanical strength [20]. It can also have an impact on particle flowability, agglomeration and mixing characteristics and also redissolution properties [20]. This is especially important for pharmaceutical products because the crystallization step is one of the last in production and the characteristics of the crystal can have an effect on a drug's usefulness. Because of the impact that crystal morphology can have on downstream processing and the utility of industrial and pharmaceutical products, it is important to be able to predict and control crystal morphology. Aspect ratio is commonly used to quantify crystal morphology, and is simply the ratio of the longest to shortest crystal dimension [21]. Lower aspect ratio ibuprofen crystals are preferred for the easier downstream processing and higher product quality reasons mentioned earlier.

It has been established that for many organic solvents, the solvent utilized in crystallization processes has a strong influence on the resulting crystal morphology [22]. There has been significant study of how crystals of carboxylic acids, and specifically of ibuprofen, are grown within solvents and how those solvents affect the resulting morphology. Before packing into a crystal, ibuprofen molecules form hydrogen-bonded dimers with other ibuprofen molecules via dispersion forces. It is believed that the growth unit of ibuprofen crystals is the non-polar entity of the dimer (isobutyl) [21]. Due to this, ibuprofen will crystallize much differently within polar solvents versus non-polar solvents. Winn and Doherty predict needle-like ibuprofen crystals using non-polar *n*-hexane as the solvent and plate-like ibuprofen crystals using polar methanol as

the solvent [21]. Their predictions accurately match previous experimental data [23] and a representation of those predictions is shown in Figure 2.4:



(a) **(b)**

Figure 2.4: Ibuprofen crystal shape when grown in *n*-hexane (a) and methanol (b)

The experimental data used in the work presented in this thesis was previously presented in 2009 by Acquah *et al.* [24]. In this work, linear models were constructed relating crystal morphology to hydrogen bonding propensities of 16 different solvents. These 16 solvents and their structure were used in the analysis within this thesis and are presented in Chapter 3. Aspect ratio was used to quantify crystal morphology. The solvents selected were not necessarily industrially important or common pharmaceutical solvents. Cooling crystallization was utilized to grow the ibuprofen crystals and the aspect ratio was measured using optical microscopy images [24]. The aspect ratio data was then regressed with 9 different solvent properties. The coefficient of determination for each regression with the property as the independent variable and aspect ratio as the dependent variable is shown in Table 2.1 [24]:

Table 2.1: Coefficients of determination for relating AR to hydrogen bonding properties

Property	**Symbol**	**R^2**
Hansen's dispersion parameter ($MPa^{1/2}$)	δ_D	0.000
Dielectric constant (dimensionless)	ε	0.510
Kamlet-Taft hydrogen bond acceptor (dimensionless)	β	0.599
Hansen's polar parameter ($MPa^{1/2}$)	δ_P	0.671
Hildebrand's total solubility parameter ($MPa^{1/2}$)	δ	0.739
Kamlet-Taft hydrogen bond donor (dimensionless)	α	0.751
Hansen's hydrogen bonding solubility parameter ($MPa^{1/2}$)	δ_H	0.815
Kosower's parameter (kcal/mol)	Z	0.833
Acceptance number (dimensionless)	AN	0.925

The highest R^2 value was achieved with acceptance number, which is defined as the ability of the solvent to form a hydrogen bond by accepting an electron pair of a donor atom from a solute molecule, as the independent variable [25]. Because acceptance number depicts the underlying solute-solvent interaction, it is not surprising that it has the best correlation with the aspect ratio of the ibuprofen crystals.

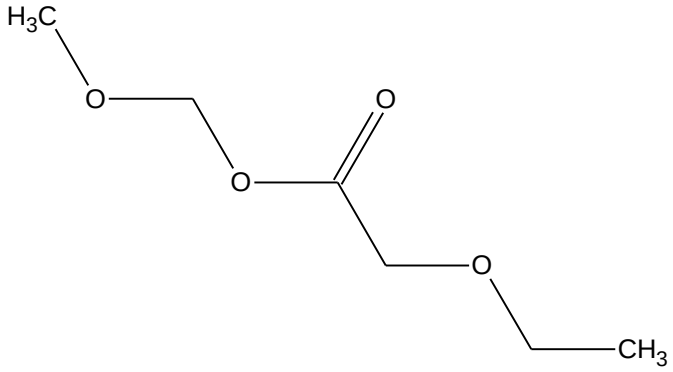
The predictive power of the acceptance number model was evaluated by comparing model predictions to previous experimental work. Lower root mean-squared error (RMSE) values were observed which indicates that the models predict very well. The hydrogen bonding solubility parameter model was used to predict the aspect ratio of ibuprofen in 2-ethoxyethyl acetate, which was not in the original training set. The predicted value (3.4) falls within the interval of the experimental value (3.1±0.5) [24]. The data from this particular study was acquired for the analysis performed in this thesis. The previous analysis related solvent hydrogen bonding properties to ibuprofen aspect ratio, but the work in this thesis will relate the solvent structure to ibuprofen aspect ratio.

2.5 Crystallization Solvent Design Framework

A CAMD framework has been developed for ibuprofen crystallization solvent design by Karunanithi *et al.* [1]. In this work, a mixed-integer nonlinear programming problem was solved to design an ideal solvent for the crystallization of ibuprofen. Their framework incorporated seven solvent properties in order to design a solvent that is safe for pharmaceutical use and also will provide the desired crystal morphology. The seven properties studied within the work are solubility, potential recovery, crystal morphology (estimated using hydrogen bonding solubility parameter), flammability limit, toxicity, viscosity and liquid state of the solvent [1]. Potential recovery was the property that was maximized within the framework. Within the CAMD, group

contribution (GC) was used to design optimal solvents for ibuprofen crystallization. The result from this CAMD framework is an overall optimal solvent, methoxymethyl ethoxyacetate, and an optimal solvent among readily available compounds, 2-ethoxyethyl acetate [1]. Both of these molecules are shown in Figure 2.5:

Methoxymethyl ethoxyacetate



2-Ethoxyethyl acetate

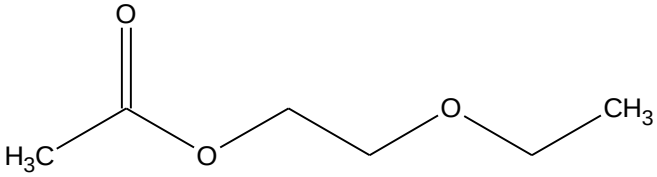


Figure 2.5: Results from CAMD framework for crystallization solvent design

The results of this framework were then experimentally verified [3]. Ibuprofen was crystallized using the cooling crystallization method in 2-ethoxyethyl acetate and also *n*-hexane for comparison purposes. It has been predicted that ibuprofen crystals grown in *n*-hexane will have a high aspect ratio [20]. The experimental verification validated the CAMD results in that crystals from 2-ethoxyethyl acetate were considerably larger in size and lower in aspect ratio than those from *n*-hexane [3]. 2-ethoxyethyl acetate also maximizes potential recovery of ibuprofen among known solvents.

2-ethoxyethyl acetate was selected for use in the test set within this analysis because it was shown to be the optimal solvent for ibuprofen crystallization. The other two solvents chosen for use in the test set were chloroform and decanol. These were chosen because data was available

and they were not included in the training set data used by Acquah *et al.* in their study of linear models using hydrogen bonding properties as independent variables [24]. The experimental data for each of the test set solvents was developed using cooling crystallization by the same research group [26].

Chapter 3: Methodology

As described in Chapter 1, the intent of this work is to develop a QSAR that relates solvent structure to the aspect ratio of ibuprofen crystals grown within that solvent. The original training set included 16 solvents for which experimental aspect ratio data was obtained [24]. The training set was later expanded with 35 additional solvents for a total of 51 solvents to expand the application domain for the developed QSAR models. The test set contained three solvents for which experimental aspect ratio data was also obtained. The first step in the method developed in this thesis was to estimate the 3D structure of each molecule within both the training and test sets. Multiple force fields were used to geometry optimize each solvent structure to determine the ultimate effects of different geometry estimations on the QSAR developed. Once the structure geometries had been estimated, their descriptor values were calculated. Finally, two different data regression methods were applied to the descriptor matrices to determine a linear relationship between the descriptor values and the aspect ratio of the ibuprofen crystals. Linear regression was utilized in developing the QSAR models because it has been shown that ibuprofen crystal aspect ratio has a linear relationship with solvent properties [24]. Once models were developed, internal and external validation was performed on each to check the fit and predictive prowess of each QSAR in prediction of ibuprofen crystal aspect ratio.

3.1 Estimating Solvent Geometry

The first step in developing this relationship is to estimate the 3D geometry of the solvent molecules. Estimations were made using the Avogadro molecular modeling software applying various force fields that produced different estimations for the structure of each solvent [27] [28]. The three force fields used for structure estimation were the GAFF, Ghemical and MMFF94s described in Chapter 2. Three force fields were chosen to examine the effect of using different

geometries for solvent structure estimation and see how the determination of the QSAR changed. In Figure 3.1, *n*-hexane is optimized using different force fields. In general, the structures estimated using the Ghemical force field and MMFF94s were similar in geometry while the structures estimated with GAFF tended to have very distinct geometries. The criterion for choosing force fields was that they be applicable to the solvent molecules in the experimental data matrix as well as the ibuprofen molecules. While not important within this study, future work includes expanding the models to handle multiple solutes which would require accurate 3D geometries for those molecules as well.

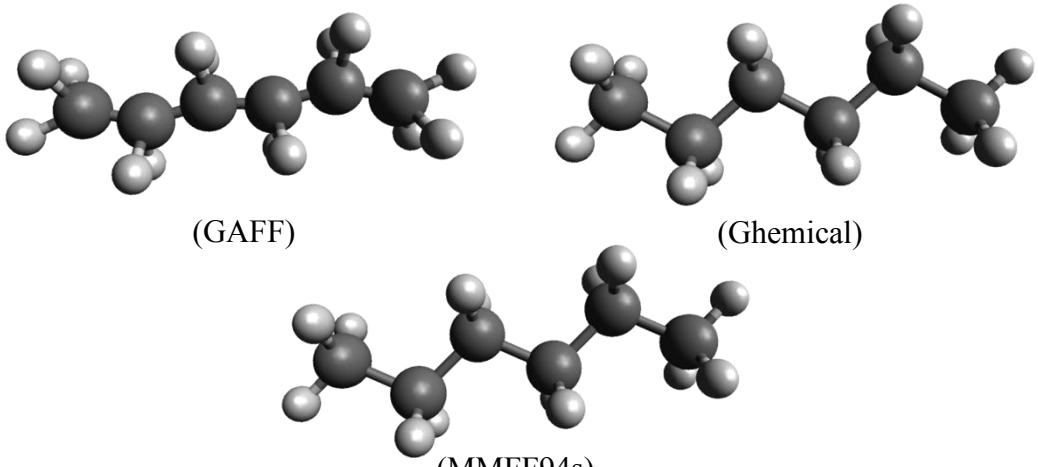


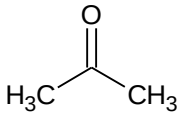
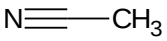
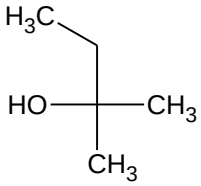
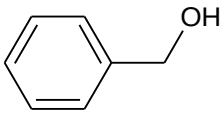
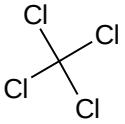
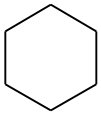
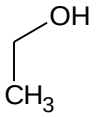
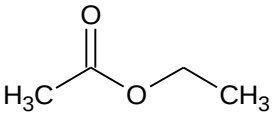
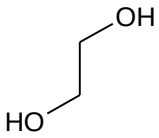
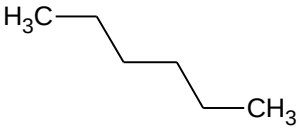
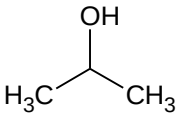
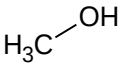
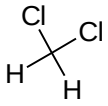
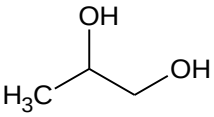
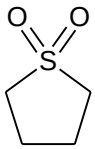
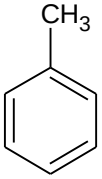
Figure 3.1: *n*-Hexane geometry optimized with Avogadro

3.1.1 Optimizing Structures with Avogadro

Each of the sixteen solvent structures were drawn and then geometrically optimized using the three different force fields. 10000 steps were included in each optimization using the steepest descent algorithm option with a convergence criterion of 1×10^{-6}. The energetically minimized molecules were saved as *.mol files for input into the E-DRAGON web applet [29].

The 16 solvents and their two-dimensional structures are shown below in Table 3.1:

Table 3.1: Original 16 solvents and their two-dimensional structures

 **Acetone**	 **Acetonitrile**	 ***t*-Amyl alcohol**	 **Benzyl alcohol**
 **Carbon tetrachloride**	 **Cyclohexane**	 **Ethanol**	 **Ethyl acetate**
 **Ethylene glycol**	 ***n*-Hexane**	 **Isopropanol**	 **Methanol**
 **Methylene dichloride**	 **Propylene glycol**	 **Sulfolane**	 **Toluene**

3.2 Calculating Descriptors

The next step in relating solvent structure to crystal aspect ratio was to calculate the descriptor values for each solvent within the data set. E-DRAGON was chosen to calculate these values because it is provided as a free java web applet and also can easily handle the files produced within Avogadro. The applet calculated over 1600 descriptors for each molecule in less than 10 seconds. This was convenient with respect to calculating descriptors for data regression, but will also be advantageous in any CAMD algorithm because candidate molecules' descriptor values can be calculated with a relatively small computational expense compared to the expense involved with geometry optimization.

3.2.1 E-DRAGON

E-DRAGON was chosen to calculate descriptors [29]. Within E-DRAGON, all available descriptors were calculated for each solvent. 1666 descriptors were calculated and then classified according to dimensionality from 0D to 3D and then by descriptor class. Only 2D and 3D descriptors were kept for this analysis and the classes analyzed are shown in Table 3.2:

Table 3.2: 2D and 3D descriptor classes

2D Descriptors	**3D Descriptors**
Topographical descriptors	Randic molecular profiles
Walk and path counts	Geometrical descriptors
Connectivity indices	RDF descriptors
Information indices	3D-MoRSE descriptors
2D autocorrelations	WHIM descriptors
Edge adjacency indices	GETAWAY descriptors
Burgen eigenvalue indices	
Eigenvalue-based indices	

3.2.2 Eliminating Zero Descriptors

This left 578 2D descriptors and 721 3D descriptors for each solvent. Seven different descriptors lists were then composed. One list contained only 2D descriptors. Three lists (one for each optimization force field implemented) contained only 3D descriptors. The final three lists added the 2D descriptors to each list of 3D descriptors. The next step was to remove descriptors from each matrix that had values of zero for multiple solvents within the data set and also the descriptors that were constant across all solvents. If a descriptor had values of zero for more than eight solvents, then that descriptor was removed from the data matrix. Table 3.3 shows the size of each of the data matrices used in the regression stage of the QSAR development.

Table 3.3: Data matrix sizes

Descriptor Set	Matrix Size (solvents × descriptors)
2D	16 × 341
3D GAFF	16 × 481
2D & 3D GAFF	16 × 822
3D Ghemical	16 × 490
2D & 3D Ghemical	16 × 831
3D MMFF94s	16 × 490
2D & 3D MMFF94s	16 × 831

3.2.3 Normalizing Descriptors

The final step in preparing the data was normalizing the data so that no descriptor's importance was artificially inflated/deflated by having a large/small value. The equation for normalization is shown below as Equation 3.1:

$$Normalized\ Value = \frac{Original\ Value - Mean\ Value}{Standard\ Deviation} \quad (3.1)$$

With this normalization procedure, descriptor values that were the same as the mean have a normalized value of zero. Values that were one standard deviation above the mean have a normalized value of positive one and conversely, values one standard deviation below the mean are normalized to negative one. The ibuprofen aspect ratio for each solvent was normalized in the same manner. All regression operations and predictions were carried out with the normalized vales. Predicted aspect ratios from the QSAR models were therefore also normalized, so then Equation 3.1 was applied in reverse to determine the actual predicted value. Validation methods produced the same results if applied on the normalized value or the actual value.

3.3 Regression and Analysis Methods

Due to the large number of descriptors calculated for each solvent and the small size of the training set, traditional linear regression was not an option for building the QSARs. The data

matrix needed to be reduced in size while still capturing the important information within it. The first method used was the Bayesian Information Criterion (BIC) method which chooses a certain number of descriptors to be used and discards the others such that the BIC value is minimized. Then linear regression was performed using the remaining descriptors to determine the model equation. The second method used was Principal Component Analysis (PCA) which calculates a number of factors which cover a certain percentage of the variance within the original data set. Linear regression was performed using these factors to determine the model equation. Both of the regression methods were applied on the normalized descriptor values.

3.3.1 BIC in JMP®

The BIC algorithm selects variables that increase the likelihood of model fit without overfitting the model. BIC introduces a penalty for addition of variables into the model. In this way, BIC selects the optimal number of variables that results in a better model fit [30]. The models produced using BIC are in the form of Equation 3.2:

$$AR = a_0 + a_1 D_1 + a_2 D_2 + a_3 D_3 + \cdots + a_n D_n \quad (3.2)$$

where a is a regression coefficient and D is a descriptor selected by JMP®.

3.3.1.1 Variable Selection

One of the benefits of the BIC method relative to PCA is that it selects certain descriptors to be used in linear regression while PCA calculates factors that are linear combinations of all the descriptors within the training set matrix. Therefore a BIC model will require a lot less computational expense than a PCA model when used in CAMD frameworks. PCA models will require each of several hundred descriptors to be calculated while BIC models will only need the descriptor values needed in the model (14 and fewer in this analysis).

3.3.2 PCA

PCA was also used to create a QSAR relating solvent structure to crystal aspect ratio. PCA is very different from BIC methods because it does not eliminate descriptors from the analysis, but instead transforms the large, highly correlated data matrix into a smaller matrix of uncorrelated factors that are linear combinations of the original data set. Each factor covers a certain percentage of the variance within the original data matrix. The total number of factors is equal to one less than the number of solvents included in each analysis. XLSTAT®, an add-in within Microsoft Excel®, was utilized to carry out all the PCA and linear regression calculations [31].

When PCA is applied to one of the descriptor matrices, it calculates eigenvalues for each factor that shows how much variance of the data that factor covers. These eigenvalues can be graphed on a Scree plot, shown below in Figure 3.2:

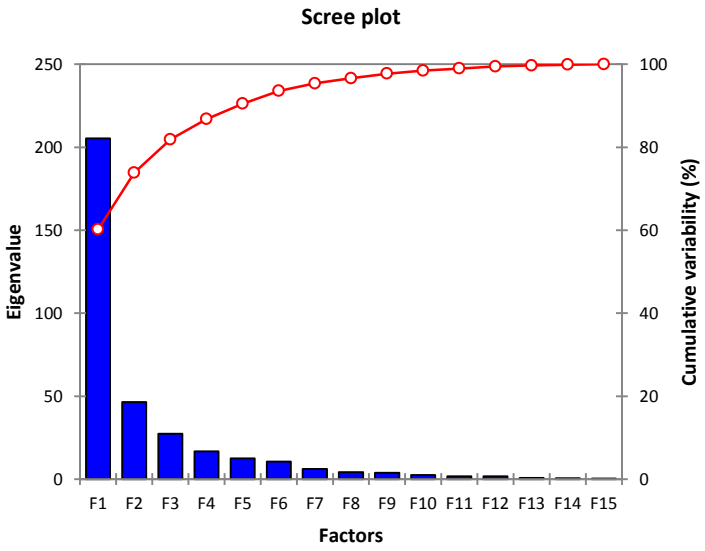


Figure 3.2: Scree plot

Typically, there is an elbow in the cumulative variability covered at which adding additional factors do not cover much more of the variance in the data set. In the above plot, that point

appears to be at eight or nine factors. One method of selecting how many factors to use in linear regression is to choose a minimum variability coverage percentage (i.e. 85%).

In order to calculate the factor scores for each solvent within the data set, PCA calculates an eigenvector (loadings) matrix and a scores matrix to decompose the original data matrix. These two matrices are calculated so that the error between the decomposition matrices and the original matrix is minimized. Figure 3.3 below shows how the original data matrix $\hat{X}$ is decomposed into a factors matrix T and a loadings matrix P. The number of factors calculated within the scores matrix is A. PCA decomposes the large matrix with O variables to a smaller matrix containing A factors where $A \ll O$. The scores matrix contains the principal component factors that represent the original data.

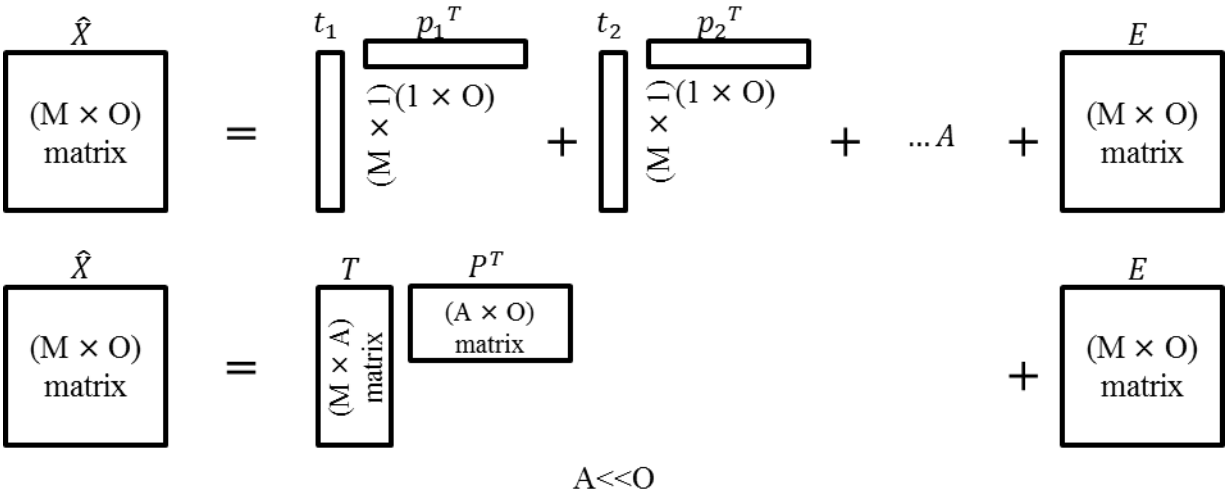


Figure 3.3: PCA matrix decomposition.

PCA was applied to each of the seven descriptor matrices and seven different factor matrices were produced. The descriptor matrices were different sizes depending on the types of descriptors they contain, but through PCA each scores matrix had the same dimension. The scores matrices contain the independent variables for linear regression. Linear regression was chosen to build the models over nonlinear, power or exponential regression because a strong

linear relationship between ibuprofen aspect ratio and many solvent properties has been shown [24].

Principal component regression (PCR) was performed to determine the multivariable linear equations that make up the QSAR models to predict aspect ratio. For each of the seven factor matrices, models were regressed repeatedly, starting with just one factor and then adding factors until the maximum number was reached. XLSTAT's® data input system allowed this to be done quickly and easily. This was necessary for the internal and external validation steps. Each of the models produced were in the form of Equation 3.3:

$$AR = a_0 + a_1F_1 + a_2F_2 + a_3F_3 + \cdots + a_nF_n \quad (3.3)$$

where a is a regression coefficient and F is a PC from PCA.

Using the eigenvector matrix P from Figure 3.3, the factors for the test set solvents can be calculated in order to insert them into the model equation to determine the predicted aspect ratio for ibuprofen grown in that particular solvent.

3.3.3 Validation

The validation techniques used to determine the predictive ability of the QSAR models include percent error analysis, the coefficient of determination (R^2) and predictive squared correlation coefficient (Q^2) methods described in Chapter 2.

External validation for the models produced using BIC was performed by analyzing R^2 values and the percent error between the experimental values and predicted values within the test set. The experimental aspect ratios and predicted values were also plotted to analyze any trends within the data. In addition the percent error analysis, external validation was also performed R^2 and Q^2 values for each model developed using the BIC method. The test set for this analysis included 2-ethoxyethyl acetate, chloroform and decanol. For the PCA models, external and

internal validation was performed using the same R^2 and Q^2 analysis as with BIC. The external test set used one solvent, 2-ethoxyethyl acetate. 2-ethoxyethyl acetate was chosen as the sole solvent for the test set because it was the same solvent used by Acquah for his external validation analysis. The internal test set removed isopropanol from the training set to create the test set. Isopropanol was chosen because its value was close to the mean value of the training set and through inspection it was structurally similar to many solvents within the training set. Cross validation was performed using Q^2 and R^2 values. As shown in Chapter 2, Q^2 is defined as [17]:

$$Q^2 = 1 - \frac{PRESS}{TSS} \tag{2.4}$$

where *PRESS* is the sum of squares of the prediction errors for the test set, and *TSS* is the total sum of squares which is the sum of squared deviations from the training set mean. As shown in Chapter 2, R^2 is defined as [17]:

$$\boldsymbol{R^2 = 1 - \frac{RSS}{TSS}} \tag{2.5}$$

where *RSS* is the residual sum of squares from the training set data and *TSS* is the total sum of squares for the training set data.

Similar to how R^2 values near one represent a strong correlation between the predicted and experimental values within the training set; a Q^2 value closer to one represents a stronger predictive model. The Q^2 equation takes into account the error between the experimental value and the predicted value but also how far the experimental value is from the mean of the training set. If the values in the test set are closer to the values within the training set, then the error between predicted and experimental needs to be smaller in order to achieve a high Q^2 value. Conversely, if the experimental values within the test set are further from the values in the training set, the error between predicted and experimental values can be larger and still achieve a high Q^2 value.

It is important for a model to have strong correlation with both the training set and the test set data. A model with a high R^2 value and a low Q^2 value fits the training set data well, but is not useful for predictive purposes. A model with a low R^2 value and a high Q^2 value does not fit the training set data well but does have some predictive capabilities. The strongest models will have R^2 values and Q^2 values very close to each other, and obviously the higher those values are the more powerful the model is overall.

3.4 Data Set Expansion

The original set of experimental data, with only 16 solvents, lacks the diversity and variability of molecular structures to build a QSAR with strong predictive capabilities. In order to improve the predictive capabilities of the models developed, more solvents were added to the data set. Solvent were not randomly chosen to fill in the data set. The criterion for solvents to be added to the data set were that it be a liquid at the temperatures used in the experimental work and also similar in structure to the solvents in the original data set. Solvents were also chosen to expand the chemical space covered by the model so that the solvents used for external validation would be within or very close to that chemical space. This is why several long chain alkanes and alcohols were included within the expansion set. A “bridge” was built between the chemical space covered by the original data set to those solvents utilized for external validation. Figure 3.4 shows the original solvents along with the expansion solvents. The lines show the connections between the original data solvents and the expansion solvents.

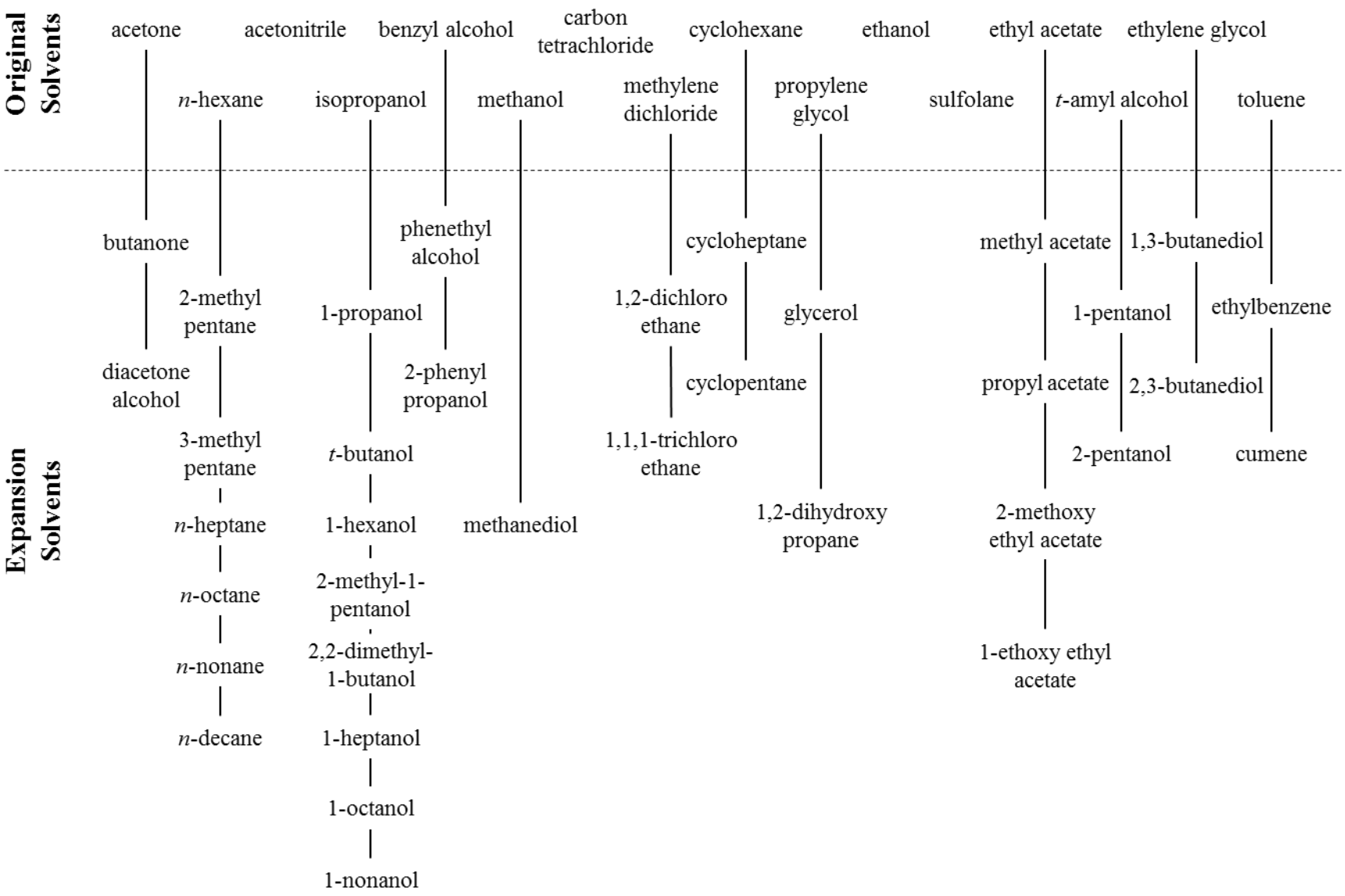


Figure 3.4: Original solvents and expansion solvents.

As shown in Figure 3.4, 35 new solvents were added to the data set. Since the solvents chosen were structurally very similar to at least one original solvent, an assumption was made that the predicted aspect ratio would be a good approximation of the actual experimental aspect ratio. The predicted aspect ratio for the expansion solvents was calculated using a model constructed using 2D descriptors with the PCA regression method. Also, the maximum number of factors available was used for each linear regression. The model built with only 2D descriptors was chosen because it had a strong correlation to the training set data ($R^2 > 0.95$). Instead of arbitrarily choosing one of the force fields for the expansion solvents, 2D descriptors were chosen because their values do not change with geometry optimization. Adding the additional solvents was an iterative process. One solvent would be added and its aspect ratio value was estimated. Then that solvent was added to the training set and a prediction was made for the next solvent. This was done repeatedly until all the expansion solvents had been added. Figure 3.5 shows the iterative process repeated for each solvent added.

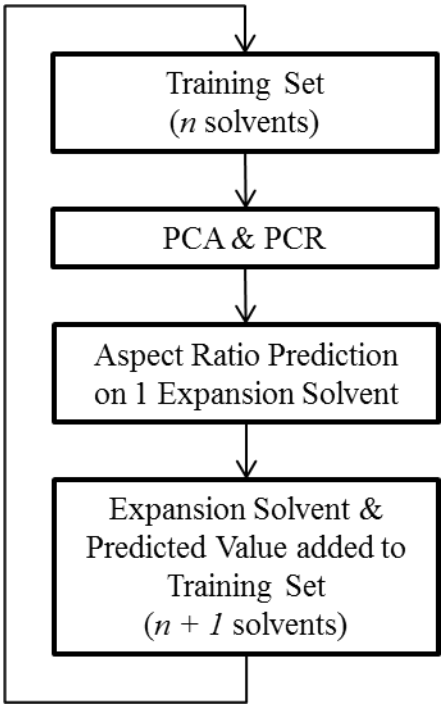


Figure 3.5: Iterative process to add expansion solvents to training set

Once the expanded data set of 51 solvents had been developed, PCA with linear regression was performed on the data set in a similar manner to that done on the original set of 16 solvents. Models were built using the same seven sets of descriptor values described earlier and then internal and external validation was performed also in the same manner.

3.5 Summary

In this chapter, a method has been presented for constructing QSAR models that relate solvent structure to the aspect ratio of ibuprofen crystals grown within them. Three empirical force fields were used to estimate the three-dimensional structure of each solvent and then a combination of 2D and 3D molecular descriptors were calculated to quantify the solvent structures. These descriptors were then arranged into matrices for regression with aspect ratio data. Two techniques, BIC and PCA, were used to linearly relate the information in the descriptors to the aspect ratio data. External and internal validation was performed to assess the model's ability to fit the training set and test set data. The original solvent data set included 16 solvents, but was expanded to 51 solvents in order to increase the chemical space covered by the models.

Chapter 4: Results and Discussion

In this section, the results from each analysis are presented and analyzed. The BIC method was first used to construct a QSAR model that contained a minimal number of descriptors. While the models built using this method were able to achieve a high correlation to the training set aspect ratio values, they were unable to make satisfactory predictions of aspect ratio for solvents in the test set. In an attempt to develop a model with stronger predictive capabilities, PCA was used using the same training set data. Most of the models built with this method had strong internal validation with isopropanol as the solvent transferred from the training set to the test set. However, most of the QSARs developed left much to be desired when they were externally validated with 2-ethoxyethyl acetate in the test set. It was hypothesized that expanding the data set to include many more solvents would increase the predictive capabilities of the models developed using PCA. With 51 solvents in the training set, the QSARs developed maintained their strong internal validation with isopropanol as the solvent shifted from the training set to the test set. Similarly to the models constructed with the smaller training set, the external validation with 2-ethoxyethyl acetate in the test set did not show powerful predictive capabilities, although they were improved over the smaller training set.

4.1 Training Set Aspect Ratio Data

The training set solvents displayed in Table 3.1 were utilized for the BIC analysis and the first PCA analysis. The aspect ratios for these 16 solvents were acquired from Acquah *et al.* and are shown in Table 4.1 [24]:

Table 4.1: 16 solvent training set aspect ratio data

Solvent	Aspect Ratio
Acetone	4.27
Acetonitrile	3.01
Benzyl alcohol	2.63
Carbon tetrachloride	4.81
Cyclohexane	5.64
Ethanol	2.85
Ethyl acetate	4.65
Ethylene glycol	2.20
n-Hexane	7.23
Isopropanol	3.10
Methanol	1.85
Methylene dichloride	3.20
Propylene glycol	3.02
Sulfolane	4.05
t-Amyl alcohol	3.21
Toluene	4.94

4.2 BIC in JMP®

The BIC models constructed using JMP® software were constructed only using the original 16 solvent data points. Each model had a very strong correlation with the training set data, with an R^2 value of 1.00 on each one. However, the predictive capabilities of these models were much poorer than expected. As shown in this section, this regression method did not produce reliable and consistent results to merit using it on the expanded training set with additional solvent molecules added.

7 different BIC models were constructed using the different sets of descriptors explained in Chapter 3. They all were in the form of Equation 3.2 and each contained a unique set of 14 different descriptors. 14 descriptors were chosen by the program which covered all the degrees of freedom in the regression and allowed for the high R^2 values mentioned earlier. The descriptors selected by JMP® are shown in Table 4.2 below, with the meaning on each descriptor abbreviation shown in Table A.1.

Table 4.2: Descriptors selected for BIC models

2D	3D GAFF	2D & 3D GAFF	3D Ghemical	2D & 3D Ghemical	3D MMFF94s	2D & 3D MMFF94s
Ram	SP01	S2K	SP01	BAC	SP13	SPI
TI2	L/Bw	AAC	SP03	X0v	PJI3	MAXDP
Rww	DISPe	SIC1	MEcc	SIC2	RDF040e	ECC
Jhetv	RDF015m	MATS1v	RDF040e	MATS3p	Mor11m	DECC
Jhete	RDF025m	EPS0	RDF045p	GATS1v	Mor12e	IC2
S1K	Mor32u	ESpm04d	Mor27u	BEHe8	Mor05p	EEig01d
Lop	Mor13p	DISPv	Mor18m	MEcc	E1s	BEHm1
X0v	E1e	Mor03u	Mor01v	RDF035v	H3u	BEHv4
SIC1	L2s	Mor03e	Mor13e	Mor16v	HATS3u	BEHp8
CIC2	G2s	E1e	Mor28p	Mor13e	HATS3m	Mor06p
MATS2p	Ks	G3s	Dm	Gs	R3v+	E2v
ESpm15u	R3m+	Kv	HATS5v	ISH	R1e+	E1e
ESpm04d	R5e	Kp	R5u	HATS3u	RTe+	R5u+
BEHm4	R3p	R2u	R5u+	R1u	R3p+	RTe+

There is very little overlap in descriptors selected over the seven different descriptor matrices. This outcome was not expected in that the important interactions between the ibuprofen and the solvents were thought to be same; therefore the effects of certain descriptors would be magnified over all sets of descriptors. Therefore, it was expected that similar 2D descriptors would be selected by the program in the model built solely with those descriptors as well as the models built with 2D descriptors and a set of 3D descriptors. The QSAR model equations constructed via BIC method in JMP® are included in Table A.2.

4.2.1 External Validation

In order to determine the predictive capabilities of each QSAR model, they were used to predict the aspect ratio of ibuprofen crystal grown in test set solvents. The test set for these models included 2-ethoxyethyl acetate, chloroform and decanol. These solvents were chosen for the test set because experimental data was readily available and they were not used in the linear models constructed by Acquah [24]. Figure 4.1 through Figure 4.3 compare the predicted values for each solvent compared with their experimental values. The solid black lines represent the

experimental value and the markers represent the predicted aspect ratio values from the models developed using the BIC method for each set of descriptors.

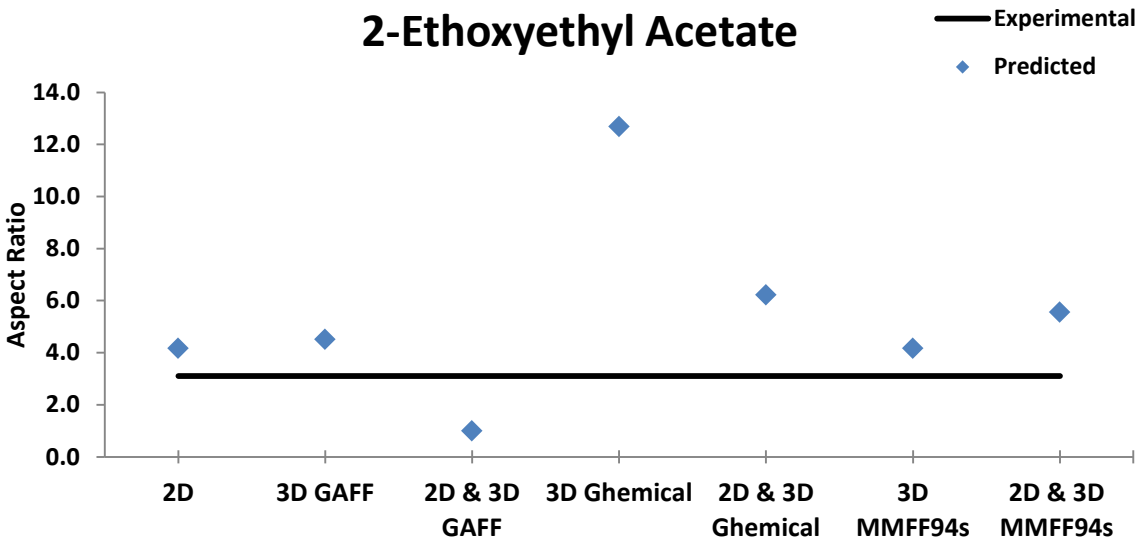


Figure 4.1: BIC experimental comparison with 2-ethoxyethyl acetate

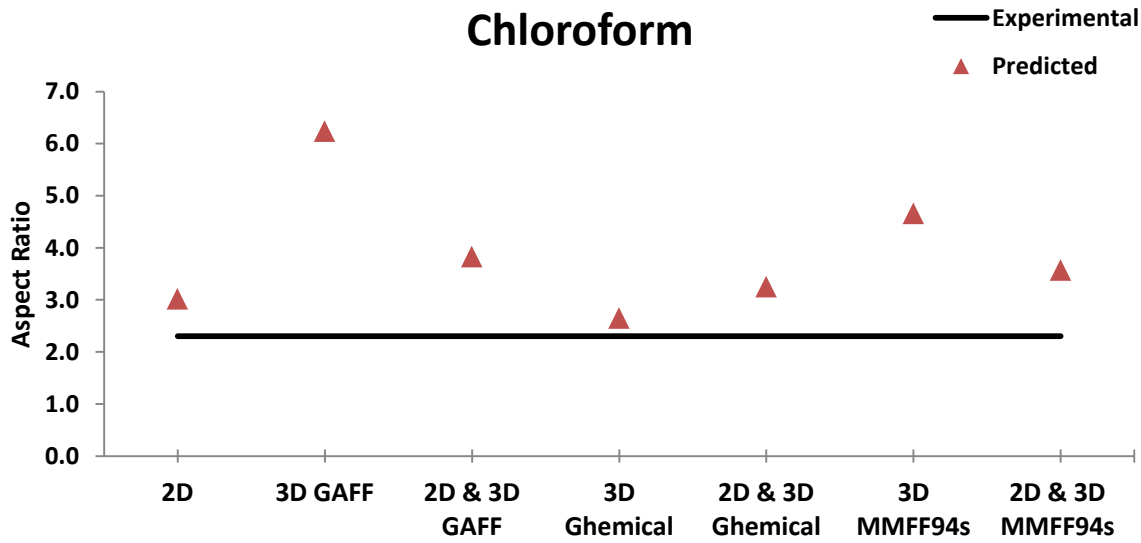


Figure 4.2: BIC experimental comparison with chloroform

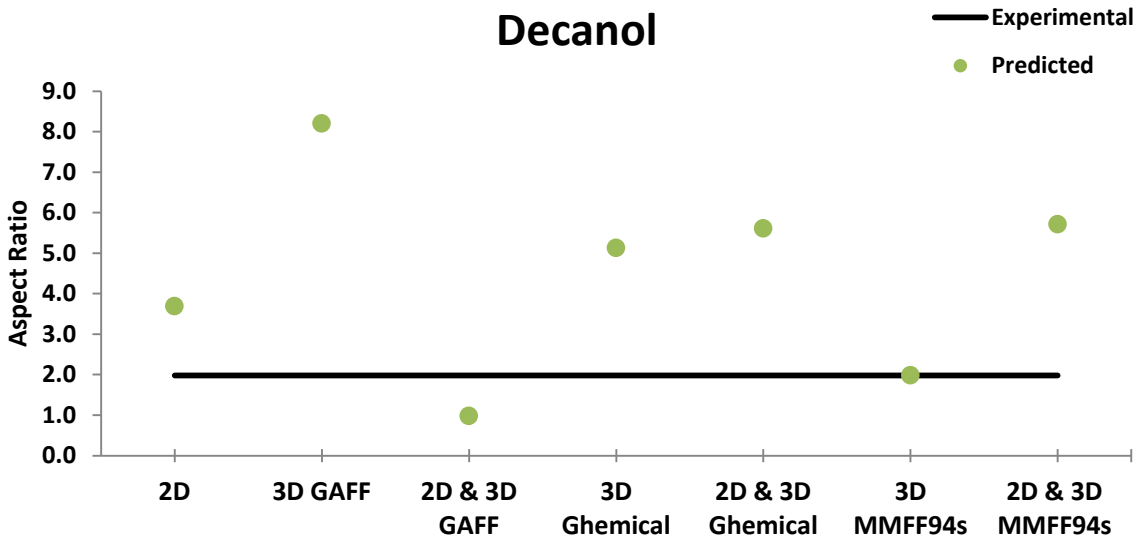


Figure 4.3: BIC experimental comparison with decanol

There is very little consistency between the experimental and predicted AR values. The models constructed with 2D descriptors provided the best predictions for the test set solvents, but it consistently predicted aspect ratios above what was observed in experimental work. The model built using the 3D Ghemical descriptor set makes an accurate prediction for the aspect ratio of ibuprofen grown in chloroform but inaccurate predictions with the other two test set solvents. The same is true with 3D MMFF94s in predicting an accurate aspect ratio of ibuprofen grown in decanol but poor predictions in the other solvents. Table 4.3 below shows the percent errors between the predicted and experimental values for each QSAR model and each solvent within the test set.

Table 4.3: External validation of BIC models

	2-Ethoxyethyl acetate		Chloroform		Decanol	
Experimental AR	3.1		2.3		1.98	
Model	**Predicted AR**	**Percent Error**	**Predicted AR**	**Percent Error**	**Predicted AR**	**Percent Error**
2D	4.17	34%	3.02	31%	3.69	86%
3D GAFF	4.51	45%	6.24	171%	8.20	314%
2D & 3D GAFF	1.00	68%	3.83	67%	0.98	51%
3D Ghemical	12.68	309%	2.65	15%	5.13	159%
2D & 3D Ghemical	6.22	101%	3.25	41%	5.61	183%
3D MMFF94s	4.17	34%	4.66	103%	1.98	0%
2D & 3D MMFF94s	5.56	79%	3.57	55%	5.71	188%

As shown in Figure 4.1 through Figure 4.3 and Table 4.3, there are no easily seen patterns between model predictions across the different solvents or between the models across the same solvent.

The models described above each used the maximum number of descriptors chosen by the BIC algorithm. The results above are consistent with a model that has been overfitted to the training set data, sacrificing its predictive capabilities. In order to improve the predictive power of models developed with method, some of the descriptors were removed from the model equations to eliminate overfitting and improve the predictive capabilities of the QSARs. The descriptors were arbitrarily removed from the end of the BIC equation. These models were analyzed using R^2 and Q^2 values. The test set for this analysis contained the same three solvents as in the previous analysis. The results of this analysis are shown below in Figure 4.4 through Figure 4.10. The vertical axis scale is fixed from zero to one, therefore any Q^2 value less than zero is not shown.

While the Q^2 value was less than zero for most QSAR models constructed using the BIC method, significant improvements were seen when PCA methods were used.

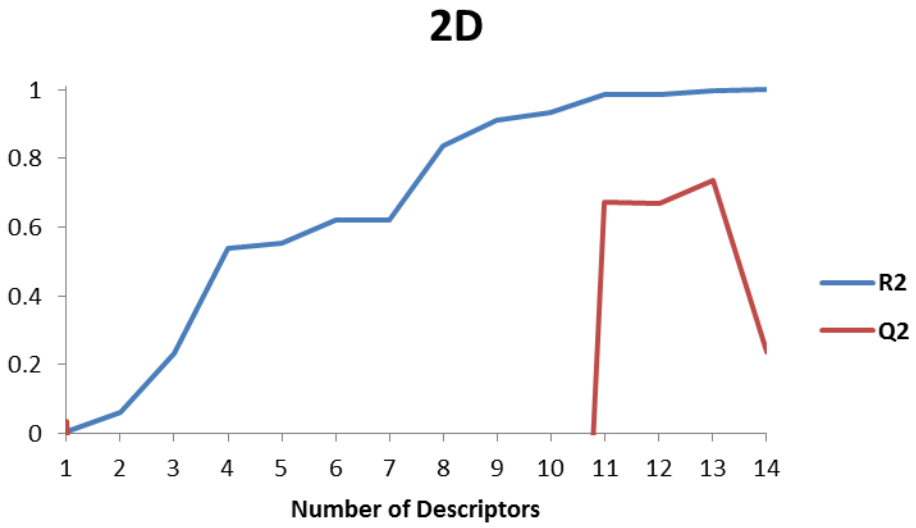


Figure 4.4: 2D BIC Q^2 external validation

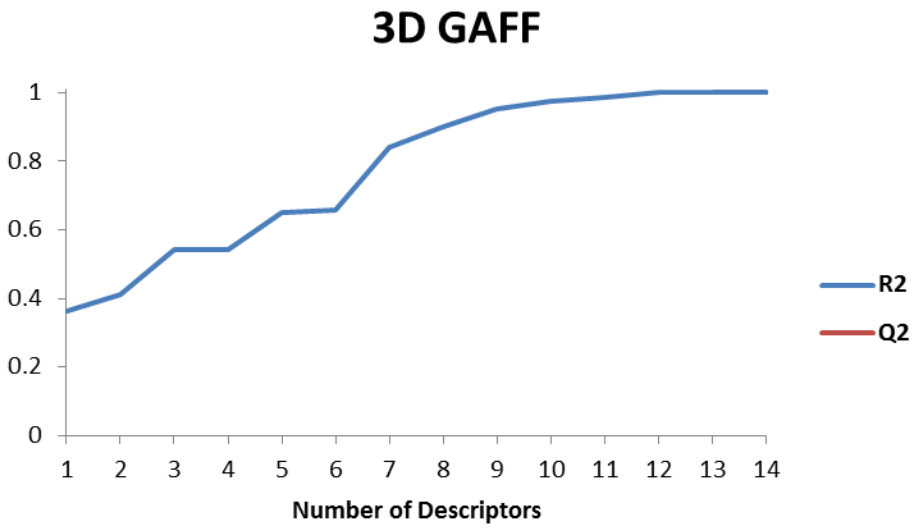


Figure 4.5: 3D GAFF BIC Q^2 external validation

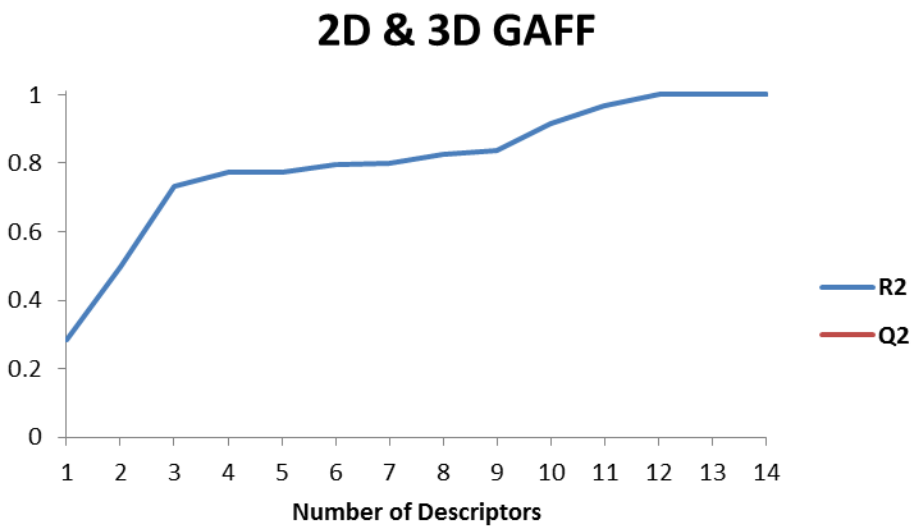


Figure 4.6: 2D & 3D GAFF BIC Q^2 external validation

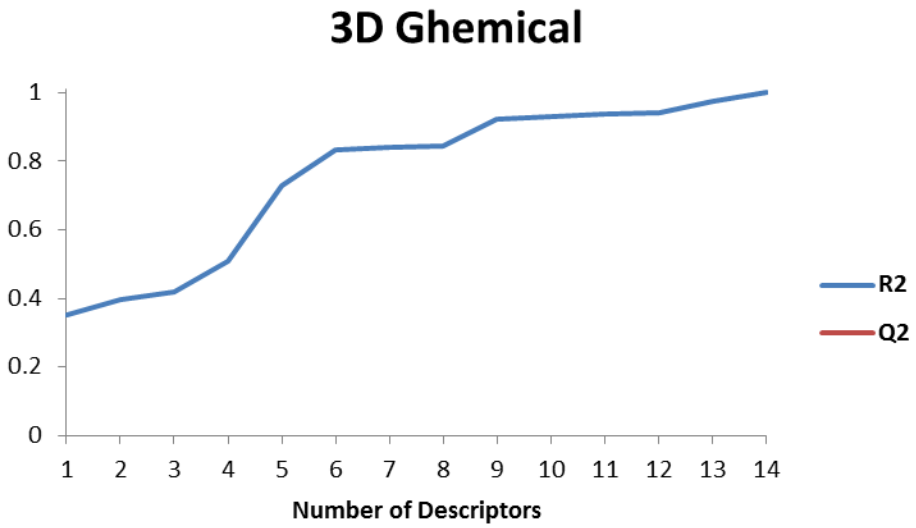


Figure 4.7: 3D Ghemical BIC Q^2 external validation

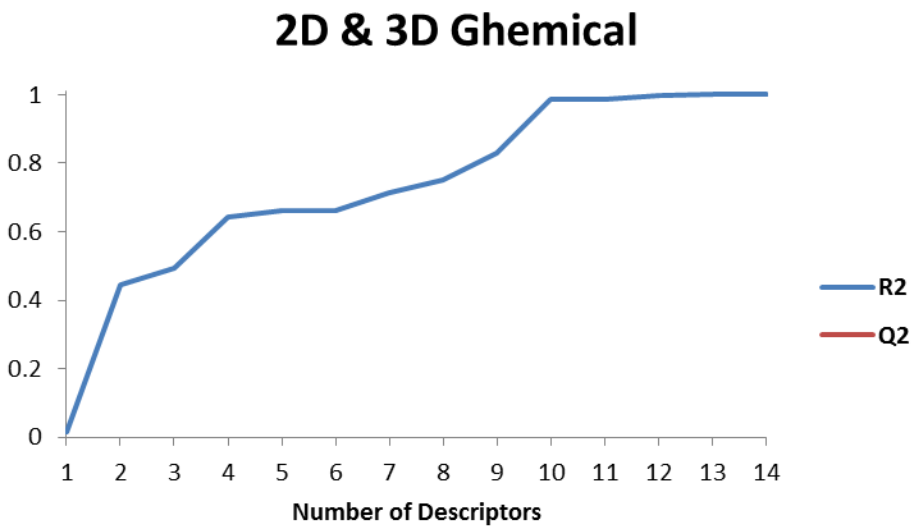


Figure 4.8: 2D & 3D Ghemical BIC Q^2 external validation

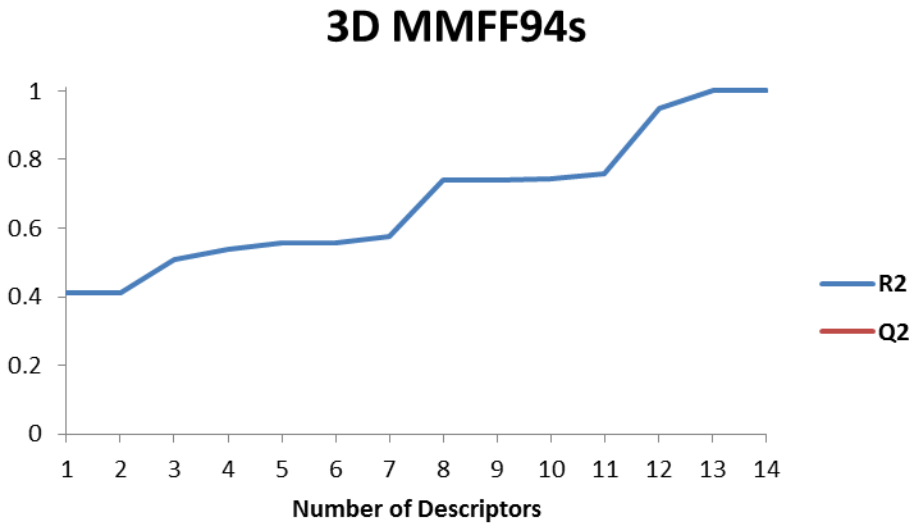


Figure 4.9: 3D MMFF94s BIC Q^2 external validation

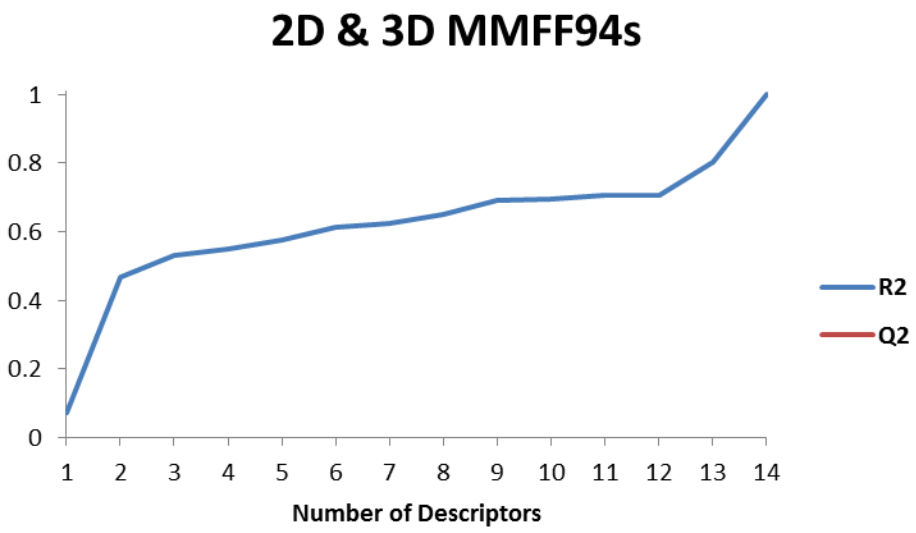


Figure 4.10: 2D & 3D MMFF94s BIC Q^2 external validation

All of the models built using the BIC method were able to achieve high correlation with the training set data. For all the models except those with 3D MMFF94s descriptors, an R^2 value above 0.9 could be achieved using 10 descriptors in the training set. For the 3D MMFF94s and 2D & 3D MMFF94s descriptor sets, an R^2 value of 1.00 was achieved with 14 descriptors. The only QSAR that had any positive Q^2 values was the one built with only 2D descriptors. In that model, the highest Q^2 value attained was around 0.7 when 13 descriptors were used. In the remainder of the models, the Q^2 values were negative which represents very poor predictive capabilities. It had been hypothesized that eliminating some of the descriptors would remove the overfitting within the models and improve predictive capabilities. Through this analysis, it can be determined that the BIC method of regressing QSARs does not provide adequate predictive capabilities for ibuprofen crystal aspect ratio. Therefore, a different method of analysis will be needed to create the desired model.

4.3 PCA Using 16 Solvents

PCA was selected to build the QSARs after the BIC method was unable to produce models that adequately fit the training set and also had strong predictive capabilities. PCA was applied to the

training set containing the same 16 solvents as used in the BIC method. The models were in the form of Equation 3.3. Unlike the BIC method, PCA used all the descriptor values within the training and test sets. Also, in these analyses presented from this point on, chloroform and decanol were removed from the external validation test set. These two solvents were removed from the test set because they were possibly far outside the applicable domain for the model. Also, the study that provided this data only included 2-ethoxyethyl acetate in the test set therefore that same test was used to analyze the PCA models [24]. A strong predictive model will have high R^2 and Q^2 values for both internal and external validation at the same number of factors.

4.3.1 Internal Validation

Internal validation was performed on the models constructed with PCA using isopropanol as the solvent to be left out and then used within the test set. As isopropanol was left out of the training set, the models were built using the remaining 15 solvents. The results of this analysis are shown in Figure 4.11 through Figure 4.17. In each of the seven descriptor sets, the R^2 value increases as more factors are added to the model.

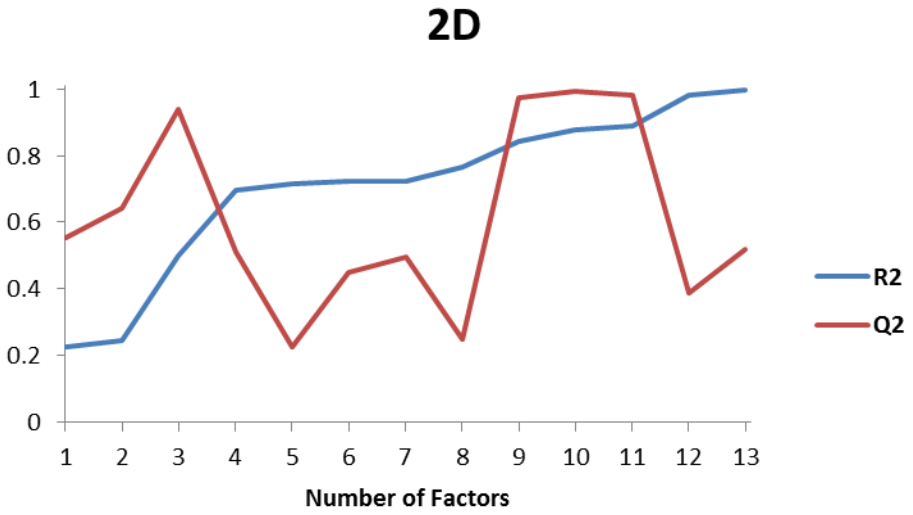


Figure 4.11: 2D PCA Q^2 internal validation

In the 2D model, the Q^2 value fluctuates as more factors are added. At the highest R^2 values, the Q^2 value is very low. The optimal model using only 2D descriptors contains 11 PC factors.

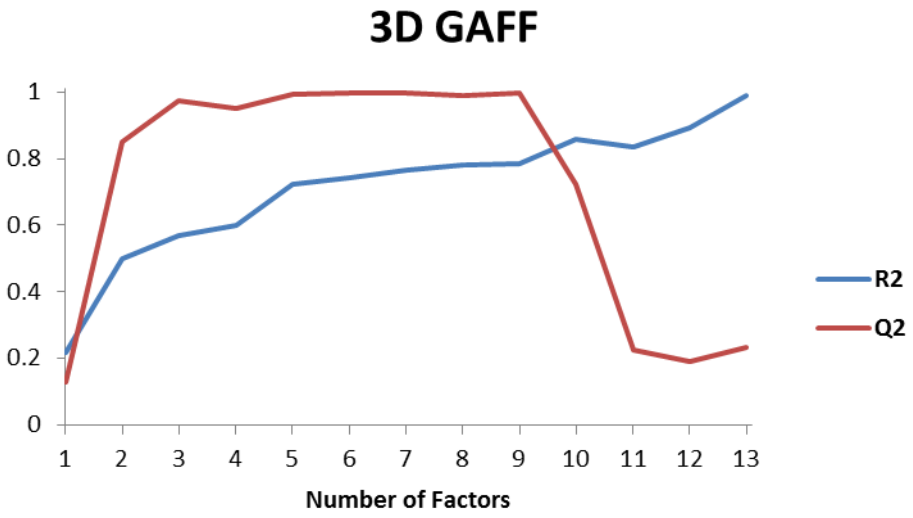


Figure 4.12: 3D GAFF PCA Q^2 internal validation

In the 3D GAFF model, the Q^2 reaches a plateau at 3 and then decreases after 9 factors. At the highest R^2 values, the Q^2 value is very low. The optimal model for 3D GAFF descriptors utilizes 9 factors.

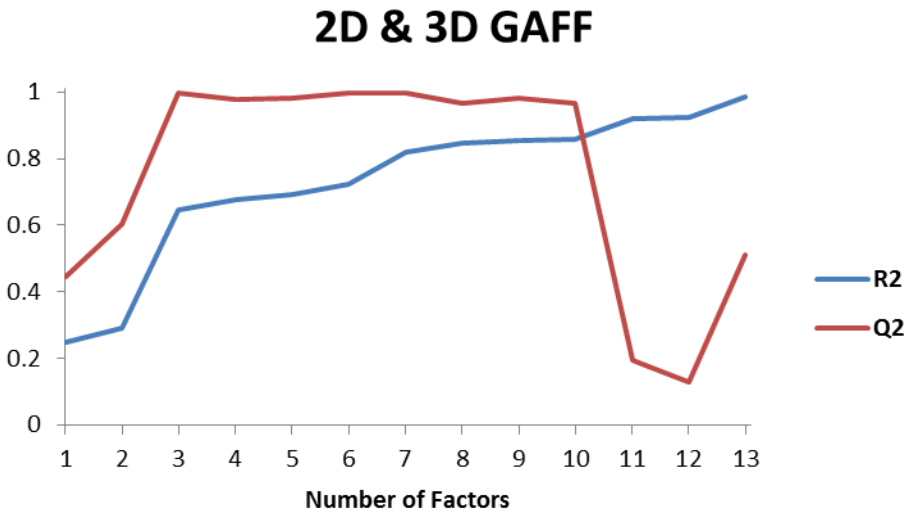


Figure 4.13: 2D & 3D GAFF PCA Q^2 internal validation

In the 2D & 3D GAFF model, the Q^2 model again plateaus with 3 factors but then decreases as more than 10 factors are added. As would be expected in a model combining the descriptors from the previous two, at the highest R^2 values the Q^2 value is very low. The optimal model for 2D & 3D GAFF descriptors is one containing 10 factors.

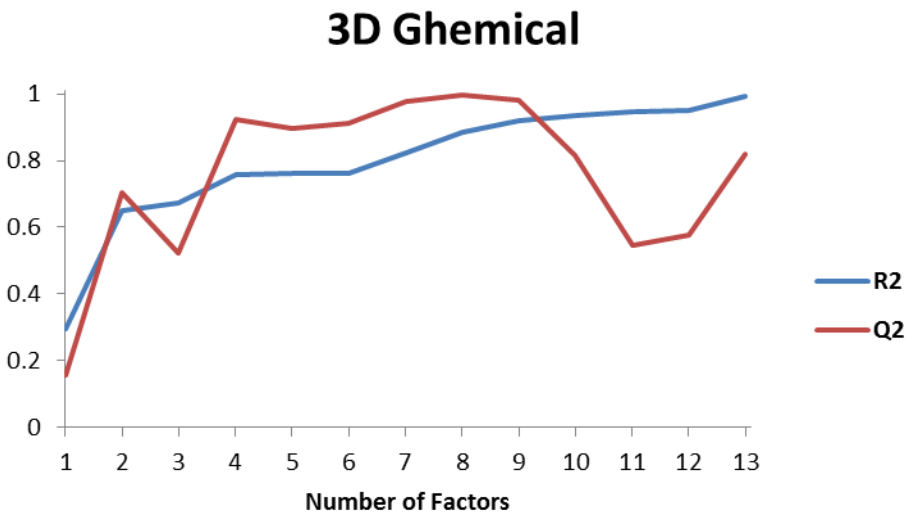


Figure 4.14: 3D Ghemical PCA Q^2 internal validation

In the 3D Ghemical model, the Q^2 value plateaus with at 3 factors until 9 factors and then decreases before rising again at 13 factors. The optimal model for 3D Ghemical descriptors is one using 9 factors.

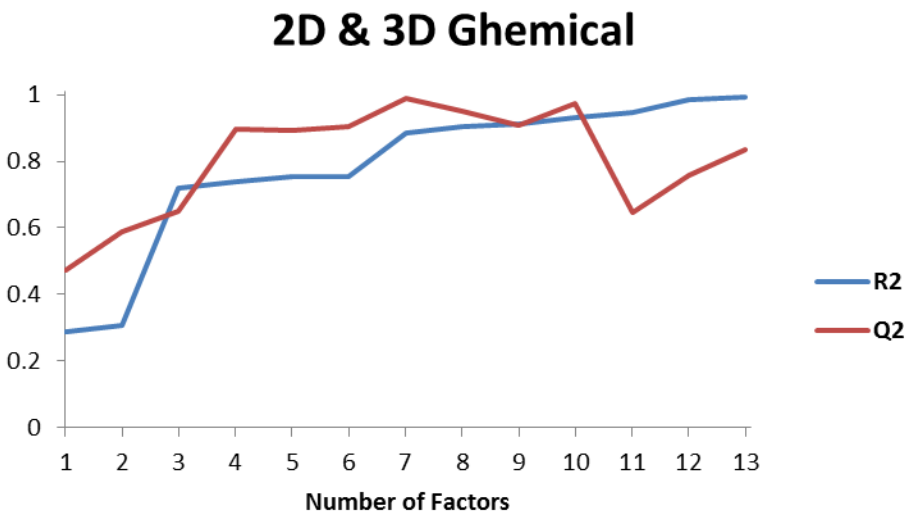


Figure 4.15: 2D & 3D Ghemical PCA Q^2 internal validation

In the model with 2D & 3D Ghemical descriptors, the Q^2 value has a similar profile to the model built with only 3D Ghemical descriptors, but the decrease after 10 factors is much smaller. The optimal model for 2D & 3D Ghemical descriptors utilizes 10 factors.

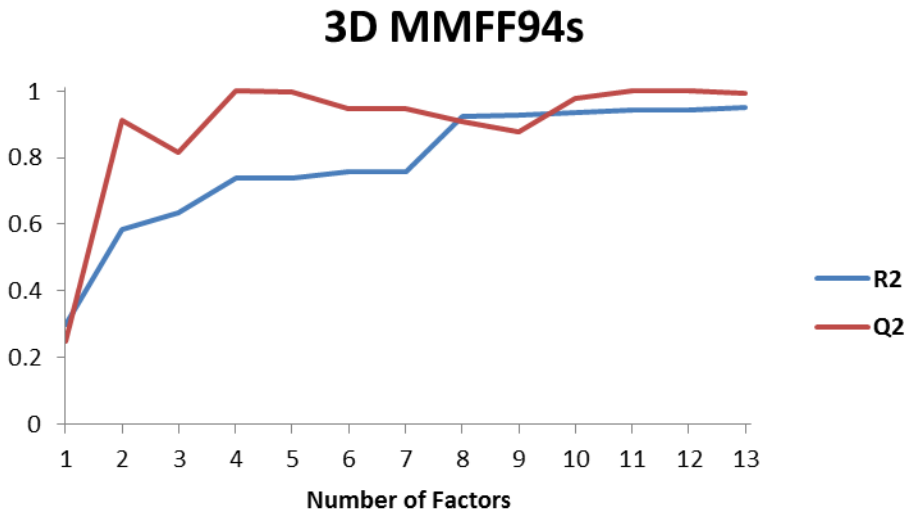


Figure 4.16: 3D MMFF94s PCA Q^2 internal validation

The model with 3D MMFF94s descriptors has high Q^2 and R^2 values as more factors are added. There are no significant decreases with more factors. Optimal models are observed using between 8 and 13 factors.

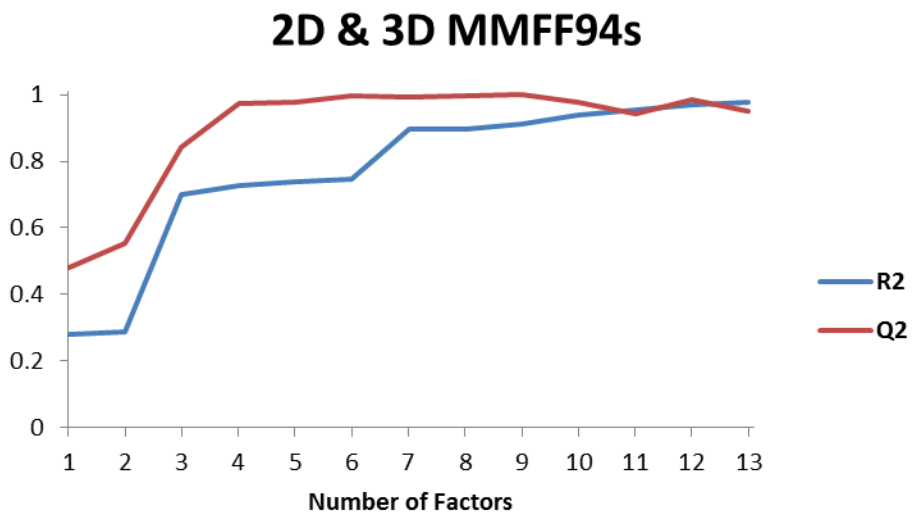


Figure 4.17: 2D & 3D MMFF94s PCA Q^2 internal validation

With 2D descriptors added to the 3D MMFF94s model, the results are very similar with high R^2 and Q^2 values as more factors are added with no significant decreases. Similar to the models using only 3D MMFF94s descriptors, optimal models for 2D & 3D MMFF94s are observed with 7 to 13 factors.

Overall, the 2D, 3D GAFF and 2D & 3D GAFF models validated adequately internally. The 3D Ghemical, 2D & 3D Ghemical, 3D MMFF94s and 2D & 3D MMFF94s models internally validated very strongly with the MMFF94s models validating very strongly with isopropanol. Since these models are able to internally validate strongly, the next step in the analysis is to insert isopropanol back into the training set and using 2-ethoxyethyl acetate for external validation.

4.3.2 External Validation

The results of the external validation are not as strong as those from the internal validation. They are presented as Figure 4.18 through Figure 4.24. The vertical axis scale is fixed from zero to one, therefore any Q^2 value less than zero is not shown.

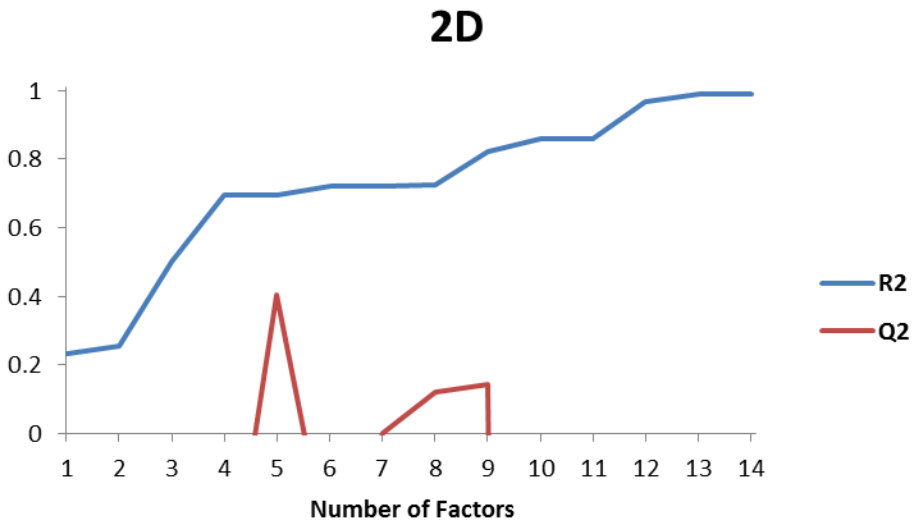


Figure 4.18: 2D PCA Q^2 external validation

The R^2 value of the 2D model increases as the number of factors increases, but the Q^2 value is only positive for models including 5, 8, and 9 factors.

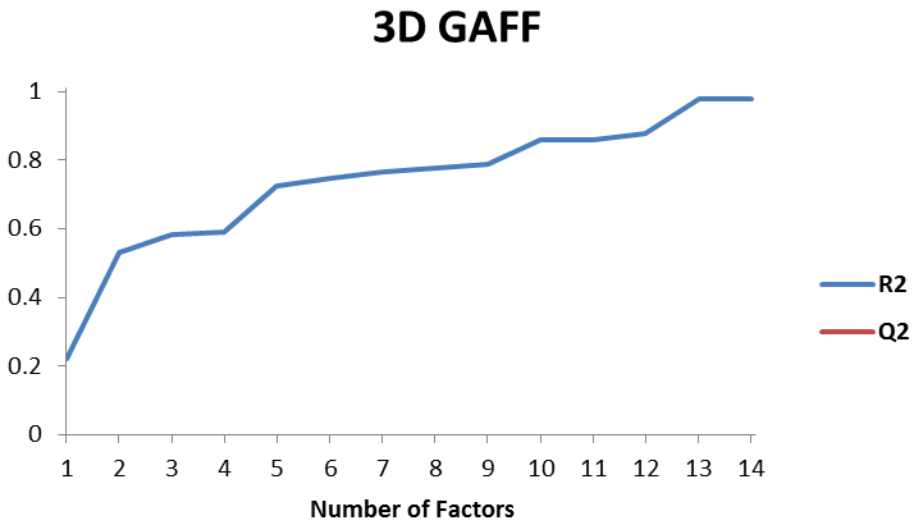


Figure 4.19: 3D GAFF PCA Q^2 external validation

The 3D GAFF model has a high R^2 value with 13 and 14 factors, but the Q^2 value is very low regardless of the number of factors.

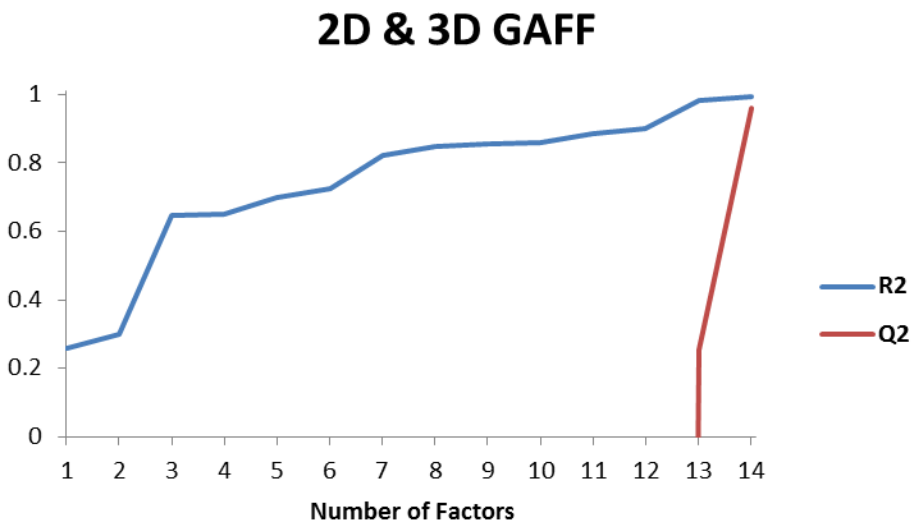


Figure 4.20: 2D & 3D GAFF PCA Q^2 external validation

The 2D & 3D GAFF model has high R^2 values at 13 and 14 factors, and also has a high Q^2 value with 14 factors included. Overall however, the models that include the GAFF descriptor sets do not internally validate strongly.

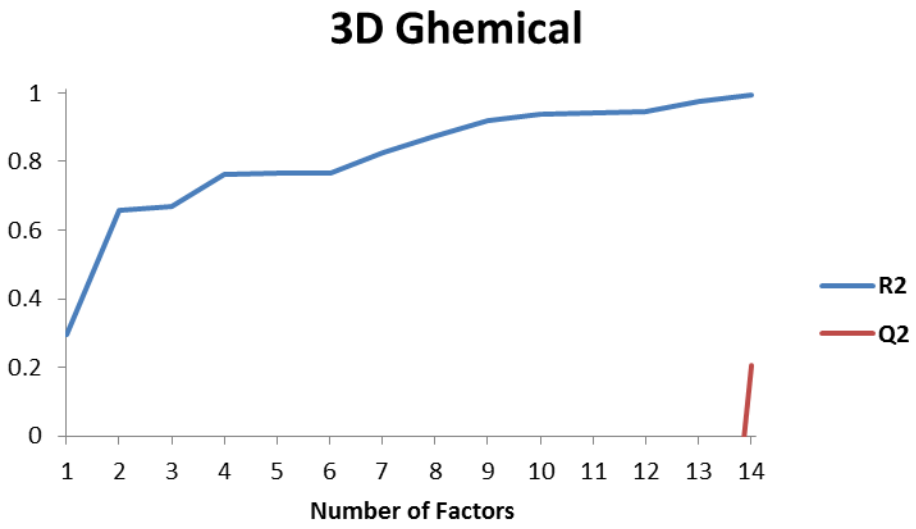


Figure 4.21: 3D Ghemical PCA Q^2 external validation

The 3D Ghemical model once again has high R^2 values as more factors are added, but only has a positive Q^2 value with 14 factors, and that value is relatively low.

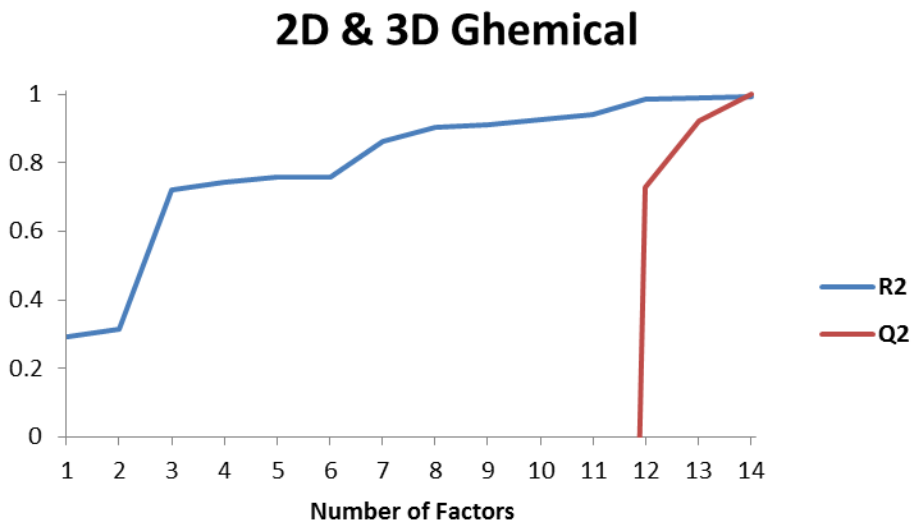


Figure 4.22: 2D & 3D Ghemical PCA Q^2 external validation

With the 2D descriptors added to the 3D Ghemical model, the same high R^2 values are seen. In addition, a high Q^2 value is seen with 13 and 14 factors. This model also internally validates strongly and is the best QSAR model built using the 16 solvent training set.

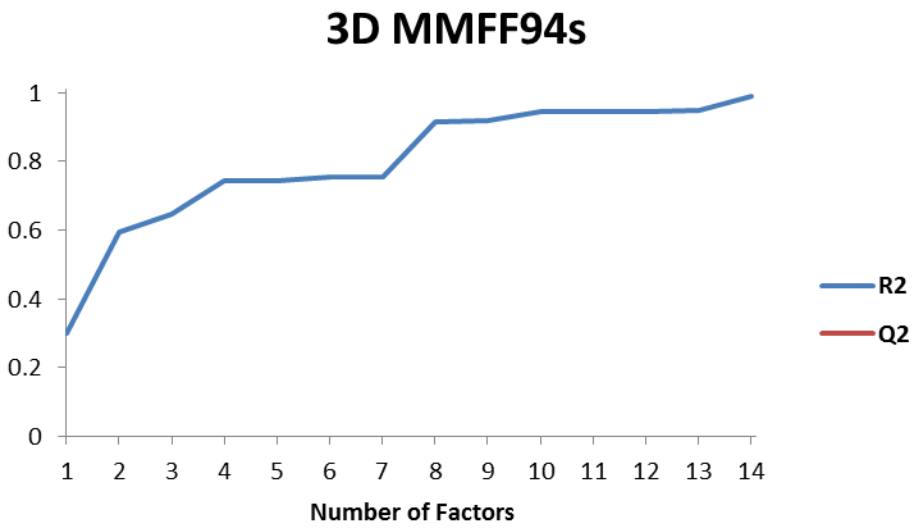


Figure 4.23: 3D MMFF94s PCA Q^2 external validation

The 3D MMFF94s model has a high R^2 value with 8 through 14 factors, but the Q^2 value is very low regardless of the number of factors.

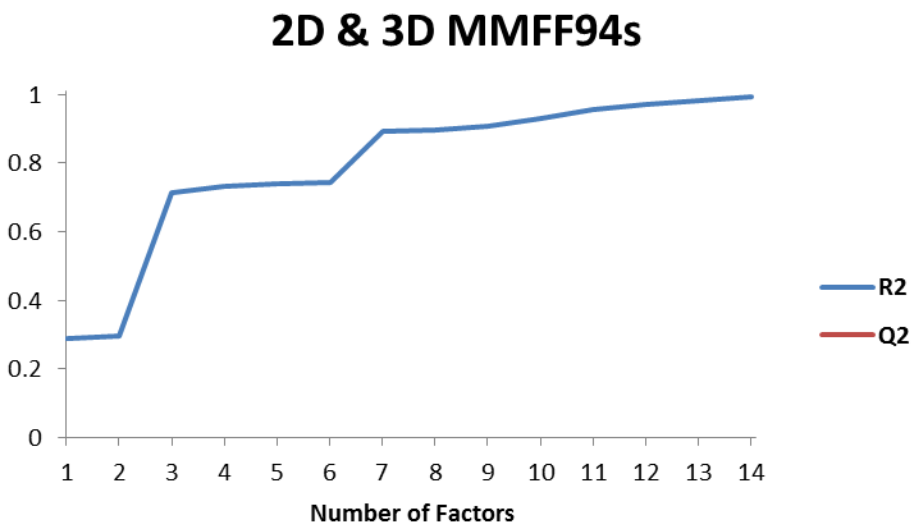


Figure 4.24: 2D & 3D MMFF94s PCA Q^2 external validation

The 2D & 3D MMFF94s model has a high R^2 value with 7 through 14 factors, but the Q^2 value is very low regardless of the number of factors.

Only one model is able to validate strongly both internally and externally and that model is constructed using 2D & 3D Ghemical descriptors with a high number of factors included (Figure 4.15 and Figure 4.22). This model is able to fit the training set data and also make accurate an accurate prediction for the aspect ratio of ibuprofen grown in 2-ethoxyethyl acetate.

With only one model able to meet both criteria for a strong predictive model, it was hypothesized that increasing the size of the training set would improve both the internal and external validation of models.

4.4 Expanded Training Set Aspect Ratio Data

The expansion solvents displayed in Figure 3.4 along with the 16 solvents shown in Table 3.1 were utilized in the second PCA analysis. The expansion solvents and their corresponding ibuprofen crystal aspect ratios are shown in Table 4.4. The aspect ratios for these solvents were estimated using the method outlined in Section 3.4. The order of the solvents within the table corresponds to the order in which they were estimated. For example, the aspect ratio of ibuprofen

grown in butanone was estimated using the original 16 solvents in the training set. This step was then iterated for each expansion solvent as shown in Figure 3.5. In the final iteration, the aspect ratio of ibuprofen grown in 1-nonanol was estimated using the original 16 solvents and all of the expansion solvents within the training set.

Table 4.4: Expansion solvent training set estimated aspect ratio data

Solvent	**Aspect Ratio**
Butanone	3.97
Diacetone alcohol	3.70
Phenethyl alcohol	1.97
2-Phenylpropanol	1.36
Cycloheptane	4.85
Cyclopentane	5.13
Propyl acetate	4.64
Methyl acetate	4.21
1,3-Butanediol	3.75
2,3-Butanediol	2.88
3-Methylpentane	6.30
2-Methylpentane	6.16
Propanol	2.81
t-Butanol	2.97
Methanediol	3.00
1,2-Dichloroethane	3.27
1,1,1-Trichloroethane	3.85
Glycerol	3.45
1,3-Dihydroxypropane	3.61
1-Pentanol	4.85
2-Pentanol	4.09
Ethylbenzene	4.32
Cumene	3.50
2,2-Dimethyl-1-butanol	3.82
1-Hexanol	4.84
2-Methyl-1-pentanol	4.25
n-Heptane	6.89
n-Octane	5.98
n-Nonane	5.68
n-Decane	5.02
1-Ethoxyethyl acetate	4.31
2-Methoxyethyl acetate	4.44
1-Heptanol	4.03
1-Octanol	3.77
1-Nonanol	3.14

4.5 PCA Using 51 Solvents

Using the method outlined in Chapter 3, the training set was expanded with 35 additional solvents to bring the total to 51 solvents. The same analysis described in the previous section was performed again with the expanded training set. Isopropanol was removed from the training set and moved to the test set for internal validation. It was then reinserted into the training set for external validation with 2-ethoxyethyl acetate in the test set. Just as with the smaller training set, the ideal model will have high R^2 and Q^2 values with the same amount of factors within the model equation.

4.5.1 Internal Validation

The results of internal validation are shown in Figure 4.25 through Figure 4.31. The vertical axis scale is fixed from zero to one, therefore any Q^2 value less than zero is not shown.

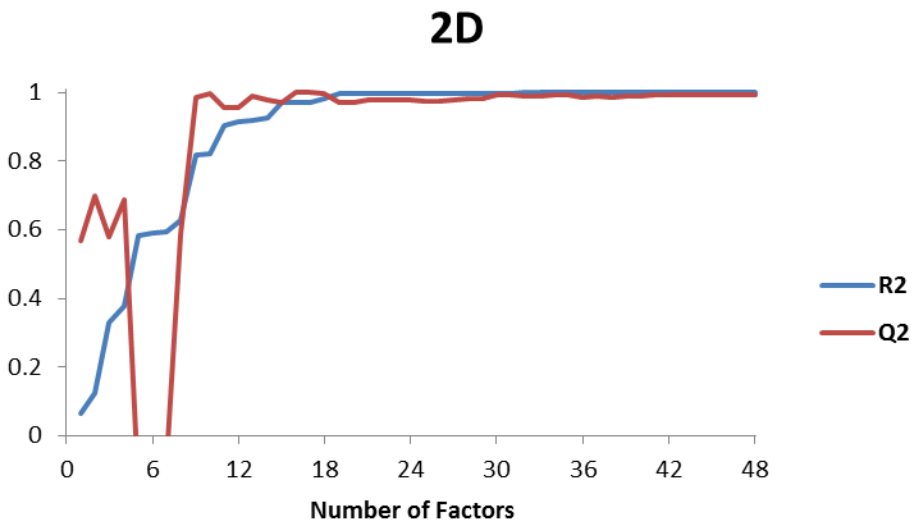


Figure 4.25: 2D PCA Q^2 internal validation with 51 solvents

The 2D QSAR has consistently high R^2 and Q^2 values from about 12 factors and higher. However, there is a sharp decline in Q^2 around 6 factors. Optimal models are seen with 12+ factors.

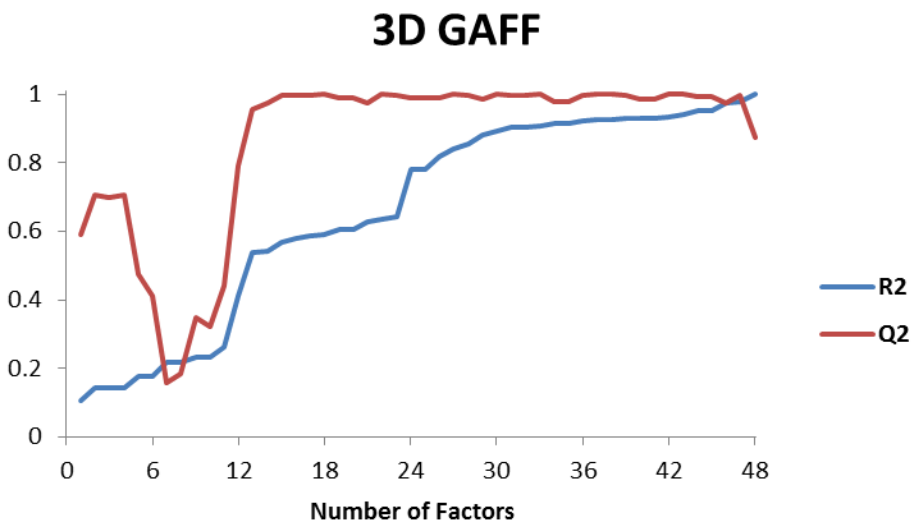


Figure 4.26: 3D GAFF PCA Q^2 internal validation with 51 solvents

The 3D GAFF model requires a lot of factors to reach a high R^2 value and about 12 factors to reach a high Q^2 value. Similar to the 2D model, there is a decline in Q^2 around 6 factors. The optimal model using 3D GAFF descriptors is seen with 46 factors, and adequate models are observed with 30+ factors.

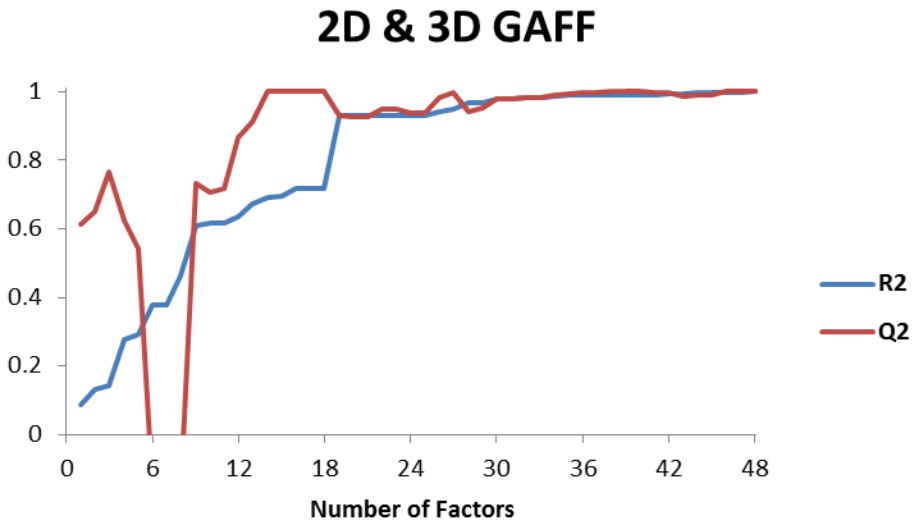


Figure 4.27: 2D & 3D GAFF PCA Q^2 internal validation with 51 solvents

The 2D & 3D GAFF model achieves high R^2 and Q^2 values from about 20 factors and higher, making these the optimal models for this descriptor set. There is the same sharp decline around 6 factors as seen in the 2 previous models.

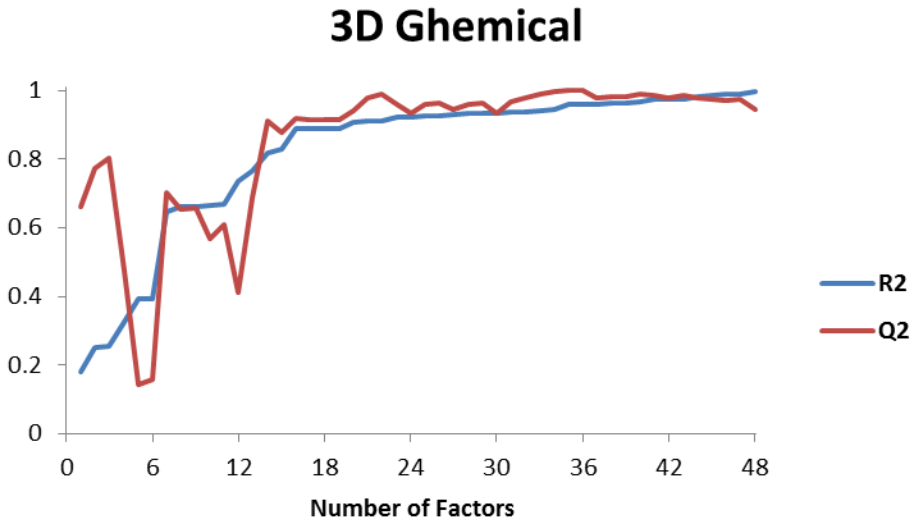


Figure 4.28: 3D Ghemical PCA Q^2 internal validation with 51 solvents

The 3D Ghemical model achieves high R^2 and Q^2 values from about 16 factors and higher. Following the trend of the previous models, there is decline in Q^2 value around 6 factors. Optimal models are seen with 16+ PC factors used for PCR.

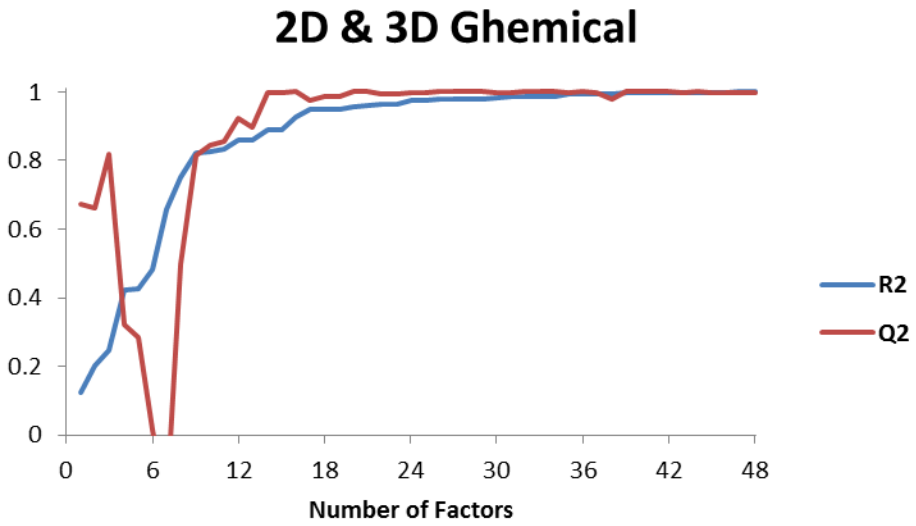


Figure 4.29: 2D & 3D Ghemical PCA Q^2 internal validation with 51 solvents

The 2D & 3D Ghemical model reaches and maintains high R^2 and Q^2 values about 18 factors. Once again, the same sharp decline in Q^2 value is seen around 6 factors. Similar to the 3D Ghemical models, optimal models include 18+ descriptors.

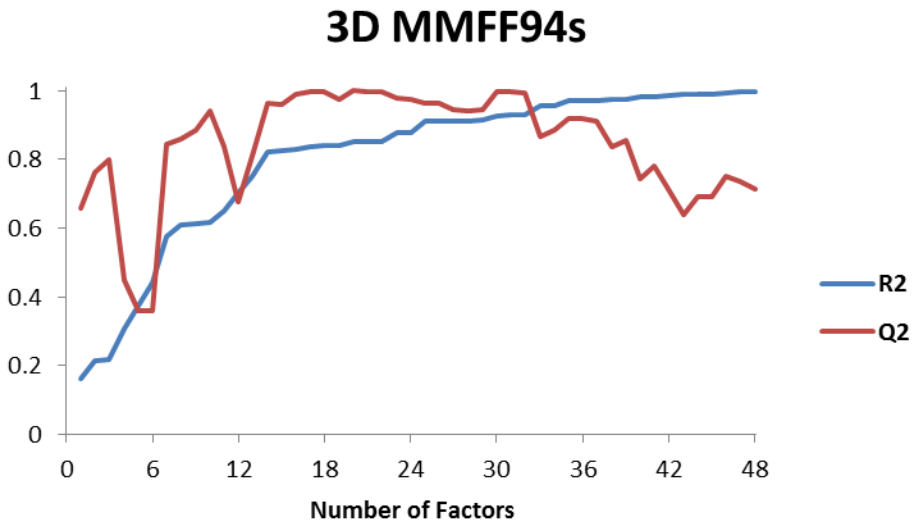


Figure 4.30: 3D MMFF94s PCA Q^2 internal validation with 51 solvents

The 3D MMFF94s model Q^2 value fluctuates up to a plateau from about 15 to 30 factors and then decreases. The R^2 value steadily increases as more factors are added. The top models are seen with about 32 factors included.

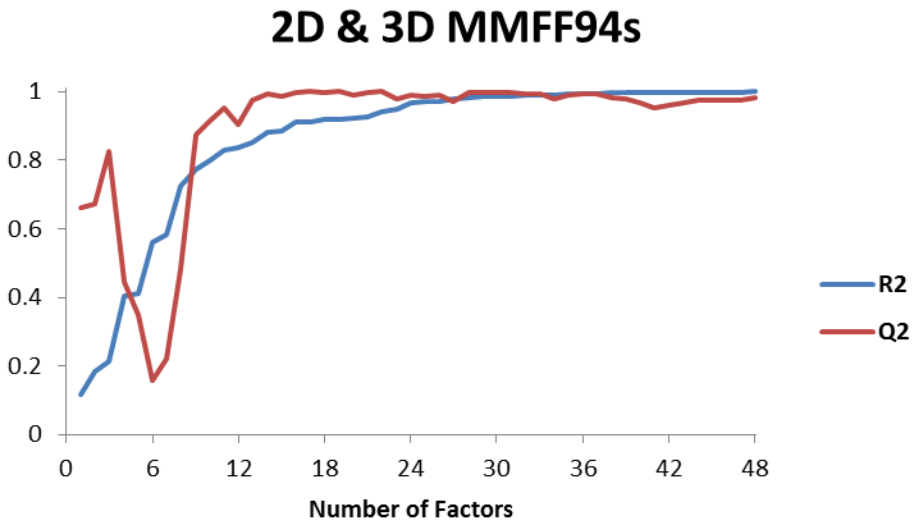


Figure 4.31: 2D & 3D MMFF94s PCA Q^2 internal validation with 51 solvents

The 2D & 3D MMFF94s model achieves high R^2 and Q^2 values around 20 factors and higher. Similar to most of the previous models, there is a sharp decline in Q^2 around 6 factors. Optimal models are seen when 24+ factors are used.

Overall, most of the models internally validated very strongly. There was a significant improvement in internal validation when the expansion solvents were added to the training set. The models were able to consistently reach high R^2 and Q^2 values as more factors were added. This was not seen in the model built using the original 16 solvents. The models using 3D descriptors only were improved with the addition of 2D descriptors to their data matrix. The next step was to move isopropanol back into the training set and then use 2-ethoxyethyl acetate in the training set for external validation.

4.5.2 External Validation

The results of the external validation analysis with 2-ethoxyethyl acetate in the test set are shown in Figure 4.32 through Figure 4.38. The vertical axis scale is fixed from zero to one, therefore any Q^2 value less than zero is not shown.

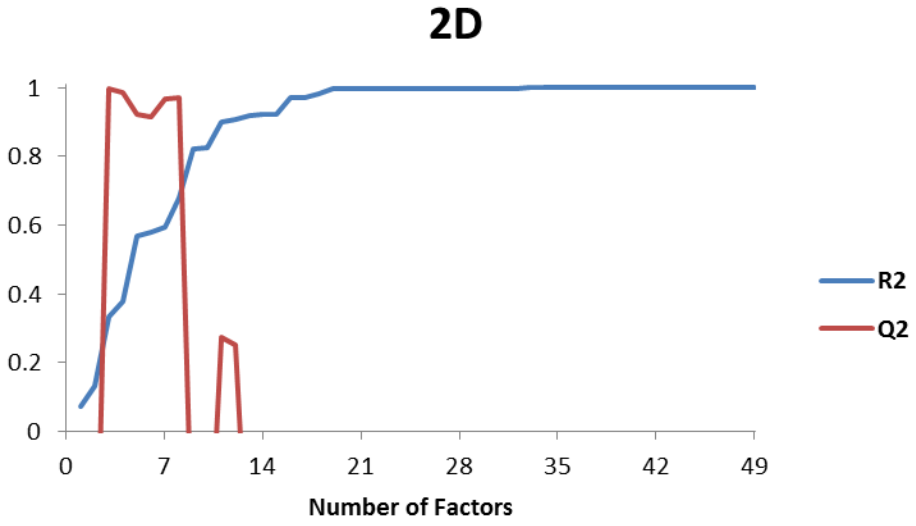


Figure 4.32: 2D PCA Q^2 external validation with 51 solvents

Similar to the internal validation, the R^2 value reaches and maintains a very high level with about 14 factors. The Q^2 value is very high from 3 to 9 factors and then as more factors are added it decreases significantly.

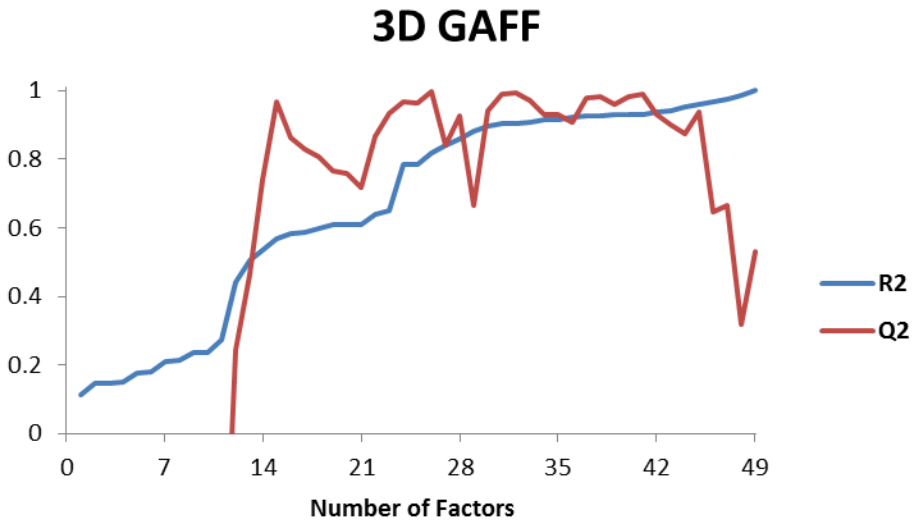


Figure 4.33: 3D GAFF PCA Q^2 external validation with 51 solvents

In the 3D GAFF model, many factors are needed to reach a high R^2 value. The Q^2 value fluctuates at a fairly high level from 14 to 44 factors. Optimal models are observed when 33 to 45 PC factors are used.

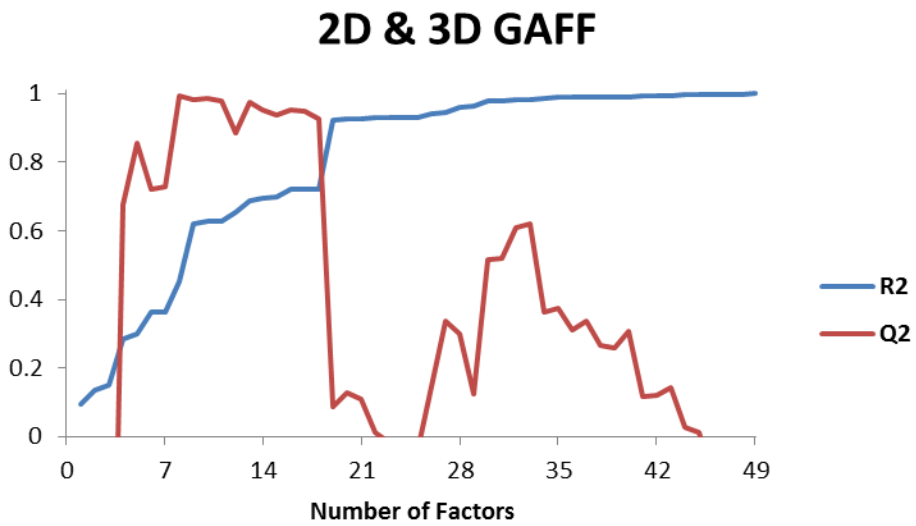


Figure 4.34: 2D & 3D GAFF PCA Q^2 external validation with 51 solvents

In the 2D & 3D GAFF model, the Q^2 value is initially very high, but has a sharp decline as soon as the R^2 value reaches very high levels.

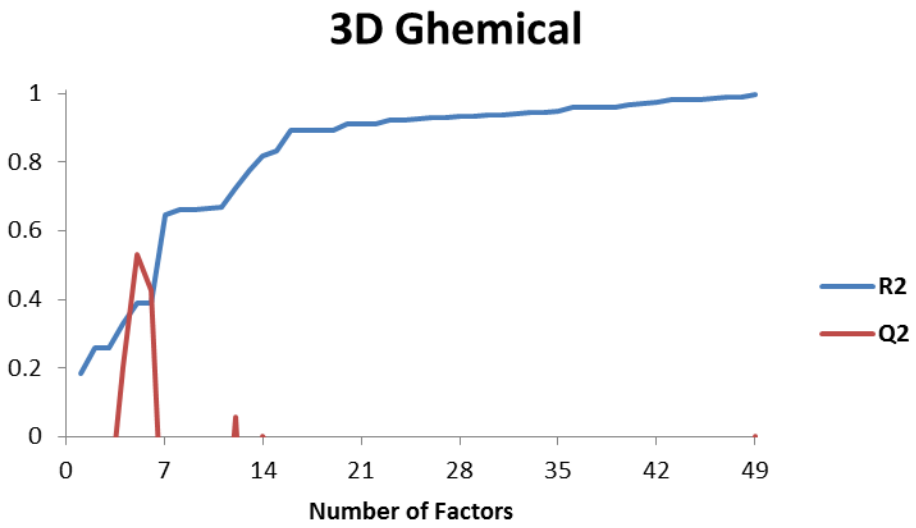


Figure 4.35: 3D Ghemical PCA Q^2 external validation with 51 solvents

In the 3D Ghemical model, the R^2 value reaches a high level about 15 factors but the Q^2 value never reaches a high level.

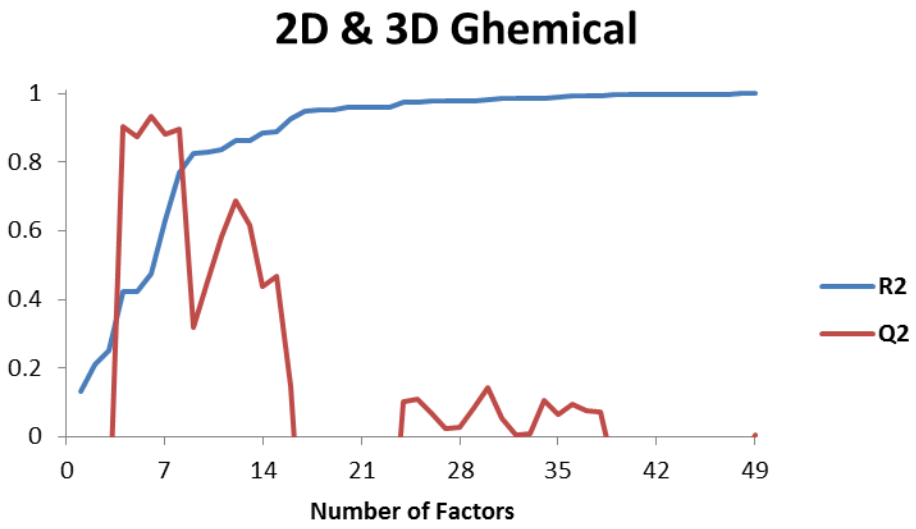


Figure 4.36: 2D & 3D Ghemical PCA Q^2 external validation with 51 solvents

In the 2D & 3D Ghemical mode, the Q^2 value is high from 5 to 8 factors, but decreases sharply with more factors. The R^2 value reaches a high level with about 14 factors and higher.

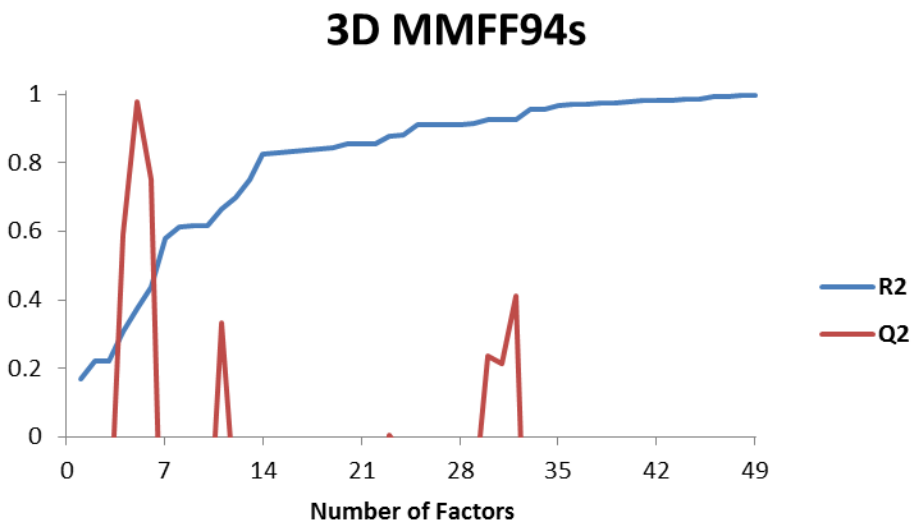


Figure 4.37: 3D MMFF94s PCA Q^2 external validation with 51 solvents

In the 3D MMFF94s model, there is a sharp spike in Q^2 with 5 factors and smaller spikes as more factors are added. However, the R^2 value only reaches high levels after many factors are added.

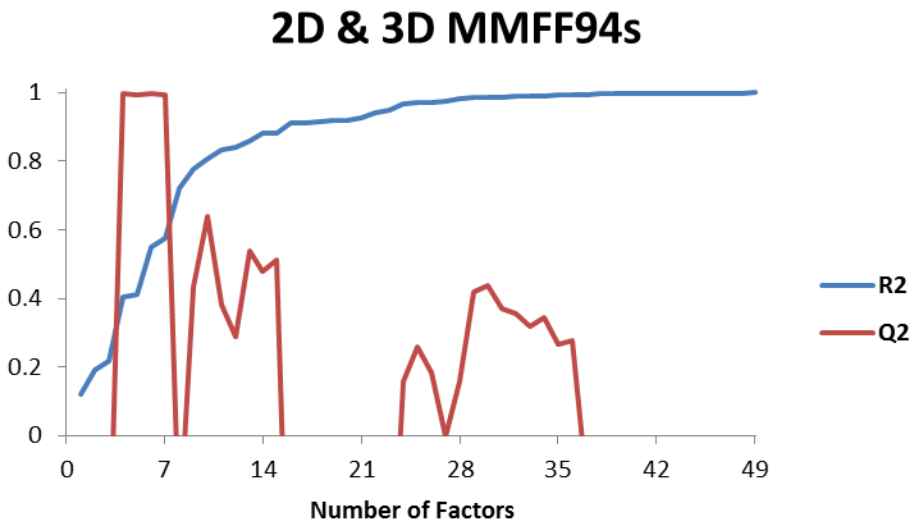


Figure 4.38: 2D & 3D MMFF94s PCA Q^2 external validation with 51 solvents

In the 2D & 3D MMFF94s model, a high Q^2 level is reached with 5 to 7 factors, however, the R^2 value only reaches high values after many factors are added.

With internal validation having improved greatly with the addition of the 35 expansion solvents, it was expected that the external validation would show a similar improvement. And while the models made better predictions using the expanded training set than with the original training set, they did not produce strong predictive models.

While there are certain factor numbers from certain models that do have high R^2 and Q^2 values in both internal and external validation, it is desired that the models consistently have those high values. In comparing many models, the Q^2 value in internal validation is relatively low with fewer factors and then reaches a high plateau as more factors are added. However, the Q^2 value in external validation is high with fewer factors and then tends to decrease as more factors are added.

4.6 Summary

Ultimately, there is no model type that consistently produces strong internal and external validation results. The QSARs produced using the BIC method appear to be overfitted in that they fit the training set data very well but provide poor, inconsistent predictions for the test set solvents. The results shown using PCA show improvement over BIC, especially with the internal validation analysis. Almost all of the models were able to produce the desired results with isopropanol with both the original and expanded training set. Therefore, it can be concluded that isopropanol is well within the descriptor space of the both the original and expanded training sets. 2-Ethoxyethyl acetate should produce similar results if it is within the chemical space of the training set. However, it has been shown that 2-ethoxyethyl acetate is not within the chemical space of the original or expanded training sets based upon ibuprofen crystal aspect ratio.

Chapter 5: Conclusions

The final results of the QSAR building technique developed in this project were not as powerful as expected but do show significant promise. The models constructed using BIC methods did not have very strong predictive capabilities with 16 solvents included in the training set. The QSARs constructed with PCA using the 16 solvent training set showed an improvement over the BIC method models, but did not consistently produce strong predictive models. When the expansion solvents were added to the training set to bring the total to 51 solvents, the QSARs showed significant increases in predictive capabilities but still left much to be desired. However, the improvement seen with a larger training set, especially with internal validation, shows that acquiring more experimental data could provide substantial improvements in predictive capabilities.

5.1 BIC in JMP®

As shown in Section 4.2, the models built with BIC methods were able to fit the training set data very well but did not fit the test set data. Across all of the descriptor sets, the coefficient of determination was 1.00. However, the models were unable to accurately predict the aspect ratio of ibuprofen crystals grown in the solvents within the test set. There did not appear to be any systematic error in the predictions. When descriptors were left out of the QSAR regression in an attempt to increase the predictive power of the model by sacrificing the fit to the training set data, no significant improvements were observed. No particular model produced consistent results for each test set solvent, nor was there any consistency in error for each particular solvent across all descriptor sets.

The model selects a small number of descriptors from the data matrix, and it is possible that the descriptors selected do not provide enough information to make accurate predictions of the

aspect ratio of ibuprofen grown in the test set solvents. The information in the descriptors excluded from the regression could be necessary in order to produce QSARs that can accurately predict crystal aspect ratios. It would be preferred that as few descriptors as possible be included in the model because then any CAMD frameworks would be less computationally expensive because they only need to calculate for those few descriptors chosen and not the entire set. So while using all of the descriptors in the data matrix adds to the computational expense of utilizing any predictive model, it will be necessary to capture enough information to create an accurate predictive model. This was the reason that PCA was then used to build models because it includes all of the information included in the data matrix as every descriptor is used to calculate the factors.

5.2 PCA

Overall, the models built with PCA performed better than the models built using BIC. Because PCA methods use all of the descriptors, they capture all of the relevant information within the data set. They also capture all of the data that may be extraneous and not relevant to the resulting crystal aspect ratio which can make any resulting CAMD framework more computationally expensive.

5.2.1 16 Solvents

The results of PCA with 16 solvents show much more consistency than previous models. Internal validation using isopropanol in the test set showed very good results, especially for the models using Ghemical and MMFF94s descriptors. The similarities between these two are not unexpected because the geometry optimized molecules from these two force fields were very similar in shape. Internal validation using GAFF descriptors did not produce the same results, which can likely be attributed to the very different shape of the molecules that were geometry

optimized using the GAFF force field. Internal validation of the models that only used 2D descriptors were also not very good, but the addition of 2D descriptors to the models that used 3D descriptors produced better QSARs.

When isopropanol was returned to the training set and 2-ethoxyethyl acetate was used as the test set for external validation, the results were significantly worse. The models were unable to consistently and correctly predict the aspect ratio of ibuprofen crystals grown in 2-ethoxyethyl acetate. Only the 2D & 3D GAFF and the 2D & 3D Ghemical models with a high number of factors made accurate predictions.

When looking at the results of internal and external validation together, only one model was able to accurately predict the crystal aspect ratio for internal and external validation. That model was constructed using 2D & 3D Ghemical optimized descriptors.

It can be concluded from the internal validation analysis that isopropanol is firmly within the chemical space outlined by the remainder of the training set solvents. The external validation analysis shows that 2-ethoxyethyl acetate is not within that same chemical space. Therefore, the data set was further populated in an effort to increase the size of the chemical space.

5.2.2 51 Solvents

The expansion solvents were added to the 16 original solvents, making the training set contain 51 solvents. It was hypothesized that populating the data set in the manner would increase the chemical space to better include 2-ethoxyethyl acetate and produce better predictive models.

The results seen for internal validation with isopropanol in the test set were as expected. Isopropanol is certainly within the chemical space where these models are applicable. As seen with the original solvents, the best results were seen with the 2D & 3D Ghemical and 2D & 3D

MMFF94s models. Marked improvements were seen with the 2D and 2D & 3D GAFF models as well.

When isopropanol was returned to the training set and 2-ethoxyethyl acetate was added to the test set, an improvement along the same lines as seen with internal validation was expected as the chemical space would be expanded to contain the new external test set solvent. The results for external validation were once again disappointing in that while improvements were made, they did not match the results seen during internal validation. In an unexpected result, the 3D GAFF model was the only one to simultaneously achieve both high R^2 and Q^2 values for both internal and external validation.

From the analysis of the expanded data set, it can be concluded that isopropanol is still well within the chemical space of the rest of the training set. However, while it can be concluded that 2-ethoxyethyl acetate is closer to being within the chemical space of the training set with the expansion solvents added, it still falls outside that chemical space.

5.3 Impact of Geometry Optimization Force Fields

From observing the shapes of the molecules produced through geometry optimization, there was a sharp contrast between the predictions for the GAFF force field versus the Ghemical and MMFF94s force fields. This can be seen very clearly in Figure 3.1 in the different shapes of the *n*-hexane molecule. Because of this, any difference that appeared in the GAFF models versus the Ghemical and MMFF94s models was not unexpected. With the three-dimensional geometries being so different, the descriptor values would obviously be different and that difference would propagate through the QSAR development process. What was unexpected was the difference in descriptors chosen using BIC methods between the Ghemical and MMFF94s models. There was no overlap between the descriptors selected by JMP® to minimize the BIC value.

In the PCA models, there was still a difference between the GAFF models and the Ghemical and the MMFF94s models, as seen in Figure 4.12 through Figure 4.38. There is, however, a similarity between the shapes of the internal and external validation curves with the Ghemical and MMFF94s models. None of the three force fields are able to distinguish themselves from the others in leading to more accurate QSAR models for ibuprofen crystallization. However, the variation in results observed with the three different force fields highlights the importance of selecting a force field that can provide accurate models for the molecules of interest.

5.4 Summary

Overall, the methodology presented in this work shows potential for constructing QSAR models that can relate solvent structure to crystal aspect ratio. The QSAR model developed show a good ability to match training set data but that same ability is not seen when the predictive capabilities are measured using the test set solvents. When the methodology is applied to this particular set of experimental data, the chemical space of the resulting models are relatively small and do not appear to include any of the solvents within the test set. In order for this methodology to produce strong predictive models, the applicable domain will need to be expanded and this need is expanded upon further in Chapter 6. While not able to indicate an optimal molecular mechanics force field, the results do underline the significance of choosing a force field that can accurately or consistently estimate the three-dimensional shapes of the molecules studied.

Chapter 6: Future Work

In this chapter, future work that could be executed using this thesis as a basis will be presented. There are many different directions that this research could take. The first possibility would be to expand the training set used to build the QSARs by including more experimental work. Improvements were observed with the addition of the expansion solvents that were only estimates. Expanding the chemical space within the training set would lead to a stronger predictive model. The second possibility for future work would be to implement a new strategy for regressing the current data. There are other data regression techniques such as the genetic algorithm procedure that could be used to create a better predictive model from the existing ibuprofen crystal aspect ratio data. In order to increase the model's usefulness in crystallization solvent design with novel drug molecules, the QSAR model would need to be expanded to include multiple solute molecules.

6.1 Acquiring Additional Experimental Data

The work presented in this thesis was inspired by a study of the relationship between ibuprofen crystal aspect ratio and hydrogen bonding properties of the solvents in which ibuprofen is crystallized [24]. In this work, ibuprofen was crystallized in 16 different organic solvents, resulting in a relatively small training set in this analysis. The results of the analysis using that small training set were not as good as desired and an improvement was seen when the 35 expansion solvents were added to it. The aspect ratios of the expansion solvents were estimated using the 16 solvent 2D model. With 51 solvents now in the training set, the predictive capabilities of the developed QSARs increased, but the results were once again not as good as desired. However, the progress seen with a larger training set shows that models with better predictive powers can be developed if more experimental data can be acquired.

The cooling crystallization experimental methods used by Acquah *et al.* are relatively simple and could be replicated in most chemical labs [24]. It would be recommended that the solvents used to populate the data set be included in the additional experimental work. First, these solvents were selected with the intent of increasing the applicable chemical space of the developed QSARs and that would obviously be better achieved with actual laboratory experiments over the assumptions made in this work. Secondly, the method of populating the original data set could be validated if these solvents are utilized in ibuprofen crystallization experiments.

Having adequate experimental data is crucial to constructing reliable predictive QSAR models. The results presented in this thesis would be significantly improved if more data was available for constructing the models.

6.2 Genetic Algorithm QSAR Models

There are many different techniques available to construct QSAR models from molecular descriptor data. One such method that may be able to create better models is to use the Genetic Algorithm (GA) procedure for variable selection. Like BIC methods, GA can select the most relevant variables that describe a data set and then those variables can be regressed into a QSAR model. This method was successfully implemented by Gramatica and Papa to create QSAR models of bioconcentration factors using theoretical molecular descriptors [32]. The GA procedure can take thousands of molecular descriptors and select a very small number of them that best model training set data. Gramatica and Papa used the GA procedure to select the five most important descriptors from an input of 1150. The QSAR models they developed achieved R^2 and Q^2 values over 0.80 through external and internal validation [32].

The disadvantage of using PCA for QSAR construction is that each descriptor remains in the model therefore the calculations are much more complex than when variable selection methods

are employed. The GA procedure could provide clues in determining the solute-solvent interactions between ibuprofen and the solvents it is crystallized within.

6.3 Expanding from Single Solute to Multiple Solute

Once a QSAR has been developed that can accurately predict the aspect ratio of ibuprofen crystals, the next step in making the model useful in designing solvents for novel drug crystallization is expanding the model to handle additional solvent molecules. This would require additional experimental work using the same solvents with various solute molecules. The work presented earlier in Chapters 3-5 is better for a multiple solute model because it takes into account more than one solvent property. It has been shown that models built using only the hydrogen bonding solubility parameter as the independent variable do not make accurate predictions for the aspect ratios of carboxylic acid crystals other than ibuprofen [22]. Because the method presented in this thesis incorporates the entire solvent structure into the model, it may be better suited for predicting aspect ratio for many different solute molecules. Because crystal aspect ratio is determined by the nature of the solute-solvent interaction, it is hypothesized that other NSAID molecules within the same propionic acid derivatives family could interact with organic solvents in the same manner as ibuprofen. Fenoprofen, flurbiprofen, ketoprofen, and naproxen are all within this class of drug molecules. The structures of these molecules, and also ibuprofen for comparison purposes, are shown below in Figure 6.1:

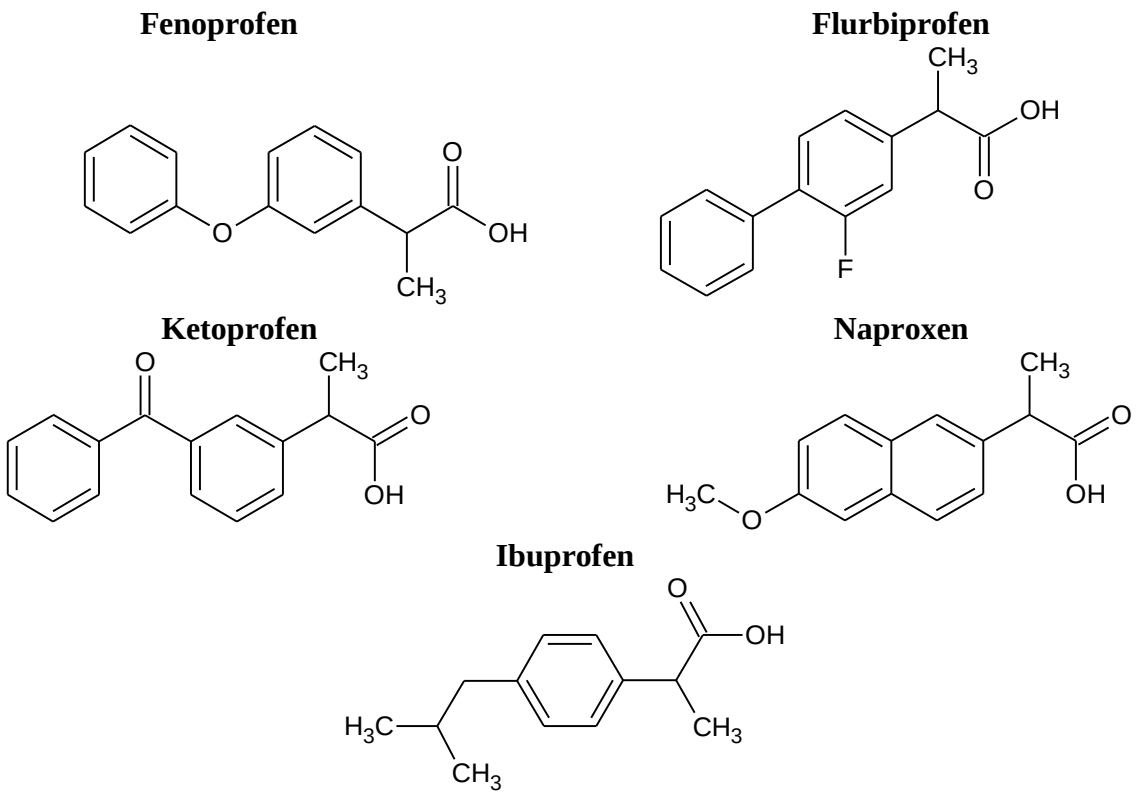


Figure 6.1: NSAIDs within the propionic acid derivative family

Each of these NSAID molecules has the same propionic acid group attached to phenolic rings. This structural similarity could mean that their interactions with crystallization solvents could be quite similar to that of ibuprofen and therefore the aspect ratio of their crystals could accurately be predicted with the same QSAR. This theory would need to be evaluated using the same crystallization process used for ibuprofen.

Having a QSAR model that can accommodate multiple solutes is crucial for using it within a CAMD framework, especially for pharmaceuticals. Expanding the chemical space for solutes to accommodate any phenolic organic compound with a propionic acid group attached would allow the model to be applied to any drug molecule that has that structure within it.

6.4 Summary

The results of this QSAR development algorithm show that while the method is promising, additional work will be needed to construct a model that can accurately model ibuprofen crystallization characteristics and also have strong predictive capabilities. Expanding the chemical applicability domain through additional experimental data would be one step to improve future models. Different regression techniques could also be employed in order to produce more accurate and predictive models. In any case, once an effective QSAR model is developed, it has been shown that it can be used within a CAMD framework to design optimal solvents for many crystallization processes.

References

[1] Karunanithi, A, Achenie, L., and Gani, R. (2006). A computer-aided molecular design framework for crystallization solvent design. Chemical Engineering Science, 61(4), 1247–1260. doi:10.1016/j.ces.2005.08.031

[2] Rasenack, N., and Müller, B. (2002). Ibuprofen crystals with optimized properties. International journal of pharmaceutics, 245, 9–24. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378517302002946

[3] Karunanithi, A. T., Acquah, C., Achenie, L. E. K., Sithambaram, S., Suib, S. L., and Gani, R. (2007). An experimental verification of morphology of ibuprofen crystals from CAMD designed solvent. Chemical Engineering Science, 62(12), 3276–3281. doi:10.1016/j.ces.2007.02.017

[4] Shuler, M. L., and Kargi, F. (2002). Bioprocess engineering, basic concepts. (2nd ed., pp. 8-9). New Jersey: Pearson College Div.

[5] Leach, A. (2001). Molecular modelling: Principles and applications. (2nd ed., pp. 165-167). Harlow, England: Pearson Education Limited

[6] Wang, J., Wolf, R., Caldwell, J., Kollman, P., and Case, D. (2004). Development and testing of a general amber force field. Journal of Computational Chemistry, 25(9), 1157–1174. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/jcc.20035/full

[7] Halgren, T. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of computational chemistry*, *17*(5-6), 490–519. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1096-987X(199604)17:5/6%3C490::AID-JCC1%3E3.0.CO;2-P/full

[8] Halgren, T. (1999). MMFF VI. MMFF94s option for energy minimization studies. Journal of Computational Chemistry, 20(7), 720–729. Retrieved from http://doi.wiley.com/10.1002/%28SICI%291096-987X%28199905%2920%3A7%3C720%3A%3AAID-JCC7%3E3.0.CO%3B2-X

[9] Nettles, J., Jenkins, J., and Bender, A. (2006). Bridging chemical and biological space:"target fishing" using 2D and 3D molecular descriptors. Journal of Medicinal Chemistry, 49(23), 6802–6810. Retrieved from http://pubs.acs.org/doi/abs/10.1021/jm060902w

[10] Gavernet, L., Talevi, A., Castro, E. A., and Bruno-Blanch, L. E. (2008). A Combined Virtual Screening 2D and 3D QSAR Methodology for the Selection of New Anticonvulsant Candidates from a Natural Product Library. QSAR and Combinatorial Science, 27(9), 1120–1129. doi:10.1002/qsar.200730055

[11] Helguera, A., Cordeiro, M., Gonzalez, M., Perez, M., Ruiz, R., and Castillo, Y. (2007). QSAR modeling for predicting carcinogenic potency of nitroso-compounds using 0D-2D

molecular descriptors. 11th International Electronic Conference on Synthetic Organic Chemistry. Retrieved from https://usc.es/congresos/ecsoc/11/hall_gCC/g003/g003.pdf

[12] Burden, F. R., and Winkler, D. A. (2009). Optimal Sparse Descriptor Selection for QSAR Using Bayesian Methods. QSAR and Combinatorial Science, 28(6-7), 645–653. doi:10.1002/qsar.200810173

[13] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461–464. Retrieved from http://projecteuclid.org/euclid.aos/1176344136

[14] Akaike, H. (1974). A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 19(6), 716–723. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1100705

[15] Bogdan, M., Ghosh, J. K., and Doerge, R. W. (2004). Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci. Genetics, 167(2), 989–99. doi:10.1534/genetics.103.021683

[16] Lauria, A., Ippolito, M., and Almerico, A. M. (2009). Combined Use of PCA and QSAR/QSPR to Predict the Drugs Mechanism of Action. An Application to the NCI ACAM Database. QSAR and Combinatorial Science, 28(4), 387–395. doi:10.1002/qsar.200810062

[17] Consonni, V., Ballabio, D., and Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. *Journal of Chemometrics*, *24*(3-4), 194–201. doi:10.1002/cem.1290

[18] University of Oxford Department of Chemistry. (2002, December). Ibuprofen. Retrieved May 23, 2013, from http://www.chem.ox.ac.uk/mom/ibuprofen/ibuprofen.html

[19] Cann, M.C.; and Connelly, M.E. *Real World Cases in Green Chemistry*, American Chemical Society: Washington, DC, 2000

[20] Winn, D., and Doherty, M. (1998). A new technique for predicting the shape of solution-grown organic crystals. AIChE journal, 44(11), 2501–2514. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/aic.690441117/abstract

[21] Winn, D., and Doherty, M. (2000). Modeling crystal shapes of organic materials grown from solution. AIChE journal, 46(7), 1348–1367. Retrieved from http://dx.doi.org/10.1002/aic.690460709

[22] Karunanithi, A. T., Acquah, C., Achenie, L. E. K., Sithambaram, S., and Suib, S. L. (2009). Solvent design for crystallization of carboxylic acids. *Computers and Chemical Engineering*, *33*(5), 1014–1021. doi:10.1016/j.compchemeng.2008.11.003

[23] Storey, R. A. (1997). The Nucleation, Growth and Solid-State Properties of Particulate Pharmaceuticals. Ph.D. Thesis, University of Bradford.

[24] Acquah, C., Karunanithi, A., Cagnetta, M., Achenie, L., & Suib, S. (2009). Linear models for prediction of ibuprofen crystal morphology based on hydrogen bonding propensities. *Fluid Phase Equilibria*, 277(1), 73–80. doi:10.1016/j.fluid.2008.11

[25] Marcus, Y. (1998). The Properties of Solvents. Chichester, New York: Wiley

[26] Acquah, C. (2008). Quantitative indices for tuning the crystal morphology of carboxlic acids. Ph.D. Thesis, University of Connecticut.

[27] Avogadro: an open-source molecular builder and visualization tool. Version 1.1.0. http://avogadro.openmolecules.net/

[28] Hanwell, M., Curtis, D., Lonie, D., Vandermeersch, T., Zurek, E., and Hutchinson, G. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of cheminformatics, 4(1), 17. doi:10.1186/1758-2946-4-17

[29] Tetko, I. V, Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V. a, *et al.* (2005). Virtual computational chemistry laboratory--design and description. Journal of computer-aided molecular design, 19(6), 453–63. doi:10.1007/s10822-005-8694-y

[30] JMP, Version 9.0. SAS Institute Inc., Cary, NC, 1989-2012.

[31] XLSTAT, Version 2013.2.04. Addinsoft, New York, 1995-2013.

[32] Gramatica, P., and Papa, E. (2003). QSAR modeling of bioconcentration factor by theoretical molecular descriptors. QSAR and Combinatorial Science, 22(3), 374–385. doi:10.1002/qsar.200390027

Appendices

A. BIC Method

Table A.1 shows the descriptors chosen by JMP® that result in models with minimum BIC values [30]. The class, symbol, and name of the descriptor are shown below for the seven descriptor sets employed in the construction of the QSAR models.

Table A.1: Descriptors used in BIC method regression

2D		
Class	**Symbol**	**Name**
Topographical	Ram	ramification index
Topographical	TI2	second Mohar index TI2
Topographical	Rww	reciprocal hyper-detour index
Topographical	Jhetv	Balaban-type index from van der Waals weighted distance matrix
Topographical	Jhete	Balaban-type index from electronegativity weighted distance matrix
Topographical	S1K	1-path Kier alpha-modified shape index
Topographical	Lop	Lopping centric index
Connectivity	X0v	valence connectivity index chi-0
Information	SIC1	structural information content (neighborhood symmetry of 1-order)
Information	CIC2	complementary information content (neighborhood symmetry of 2-order)
2D Autocorrelations	MATS2p	Moran autocorrelation - lag 2 / weighted by atomic polarizabilities
Edge Adjacency	ESpm15u	Spectral moment 15 from edge adj. matrix
Edge Adjacency	ESpm04d	Spectral moment 04 from edge adj. matrix weighted by dipole moments
Burgen Eigenvalue	BEHm4	highest eigenvalue n. 4 of Burden matrix / weighted by atomic masses
3D GAFF		
Class	**Symbol**	**Name**
Randic Molecular Profiles	SP01	shape profile no. 01
Geometrical	L/Bw	length-to-breadth ratio by WHIM
Geometrical	DISPe	d COMMA2 value / weighted by atomic Sanderson electronegativities
RDF	RDF015m	Radial Distribution Function - 1.5 / weighted by atomic masses
RDF	RDF025m	Radial Distribution Function - 2.5 / weighted by atomic masses
3D-MoRSE	Mor32u	3D-MoRSE - signal 32 / unweighted
3D-MoRSE	Mor13p	3D-MoRSE - signal 13 / weighted by atomic polarizabilities
WHIM	E1e	1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
WHIM	L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states
WHIM	G2s	2st component symmetry directional WHIM index / weighted by atomic electrotopological states
WHIM	Ks	K global shape index / weighted by atomic electrotopological states
GETAWAY	R3m+	R maximal autocorrelation of lag 3 / weighted by atomic masses
GETAWAY	R5e	R autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
GETAWAY	R3p	R autocorrelation of lag 3 / weighted by atomic polarizabilities

2D & 3D GAFF		
Class	**Symbol**	**Name**
Topographical	S2K	2-path Kier alpha-modified shape index
Information	AAC	mean information index on atomic composition
Information	SIC1	structural information content (neighborhood symmetry of 1-order)
2D Autocorrelations	MATS1v	Moran autocorrelation - lag 1 / weighted by atomic van der Waals volumes
Edge Adjacency	EPS0	edge connectivity index of order 0
Edge Adjacency	ESpm04d	Spectral moment 04 from edge adj. matrix weighted by dipole moments
Geometrical	DISPv	d COMMA2 value / weighted by atomic van der Waals volumes
3D-MoRSE	Mor03u	3D-MoRSE - signal 03 / unweighted
3D-MoRSE	Mor03e	3D-MoRSE - signal 03 / weighted by atomic Sanderson electronegativities
WHIM	E1e	1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
WHIM	G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states
WHIM	Kv	K global shape index / weighted by atomic van der Waals volumes
WHIM	Kp	K global shape index / weighted by atomic polarizabilities
GETAWAY	R2u	R autocorrelation of lag 2 / unweighted
3D Ghemical		
Class	**Symbol**	**Name**
Randic Molecular Profiles	SP01	shape profile no. 01
Randic Molecular Profiles	SP03	shape profile no. 03
Geometrical	MEcc	molecular eccentricity
RDF	RDF040e	Radial Distribution Function - 4.0 / weighted by atomic Sanderson electronegativities
RDF	RDF045p	Radial Distribution Function - 4.5 / weighted by atomic polarizabilities
3D-MoRSE	Mor27u	3D-MoRSE - signal 27 / unweighted
3D-MoRSE	Mor18m	3D-MoRSE - signal 18 / weighted by atomic masses
3D-MoRSE	Mor01v	3D-MoRSE - signal 01 / weighted by atomic van der Waals volumes
3D-MoRSE	Mor13e	3D-MoRSE - signal 13 / weighted by atomic Sanderson electronegativities
3D-MoRSE	Mor28p	3D-MoRSE - signal 28 / weighted by atomic polarizabilities
WHIM	Dm	D total accessibility index / weighted by atomic masses
GETAWAY	HATS5v	leverage-weighted autocorrelation of lag 5 / weighted by atomic van der Waals volumes
GETAWAY	R5u	R autocorrelation of lag 5 / unweighted
GETAWAY	R5u+	R maximal autocorrelation of lag 5 / unweighted
2D & 3D Ghemical		
Class	**Symbol**	**Name**
Topographical	BAC	Balaban centric index
Connectivity	X0v	valence connectivity index chi-0
Information	SIC2	structural information content (neighborhood symmetry of 2-order)
2D Autocorrelations	MATS3p	Moran autocorrelation - lag 3 / weighted by atomic polarizabilities
2D Autocorrelations	GATS1v	Geary autocorrelation - lag 1 / weighted by atomic van der Waals volumes
Burgen Eigenvalue	BEHe8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic Sanderson electronegativities
Geometrical	MEcc	molecular eccentricity

RDF	RDF035v	Radial Distribution Function - 3.5 / weighted by atomic van der Waals volumes
3D-MoRSE	Mor16v	3D-MoRSE - signal 16 / weighted by atomic van der Waals volumes
3D-MoRSE	Mor13e	3D-MoRSE - signal 13 / weighted by atomic Sanderson electronegativities
WHIM	Gs	G total symmetry index / weighted by atomic electrotopological states
GETAWAY	ISH	standardized information content on the leverage equality
GETAWAY	HATS3u	leverage-weighted autocorrelation of lag 3 / unweighted
GETAWAY	R1u	R autocorrelation of lag 1 / unweighted
3D MMFF94s		
Class	**Symbol**	**Name**
Randic Molecular Profiles	SP13	shape profile no. 13
Geometrical	PJI3	3D Petijean shape index
RDF	RDF040e	Radial Distribution Function - 4.0 / weighted by atomic Sanderson electronegativities
3D-MoRSE	Mor11m	3D-MoRSE - signal 11 / weighted by atomic masses
3D-MoRSE	Mor12e	3D-MoRSE - signal 12 / weighted by atomic Sanderson electronegativities
3D-MoRSE	Mor05p	3D-MoRSE - signal 05 / weighted by atomic polarizabilities
WHIM	E1s	1st component accessibility directional WHIM index / weighted by atomic electrotopological states
GETAWAY	H3u	H autocorrelation of lag 3 / unweighted
GETAWAY	HATS3u	leverage-weighted autocorrelation of lag 3 / unweighted
GETAWAY	HATS3m	leverage-weighted autocorrelation of lag 3 / weighted by atomic masses
GETAWAY	R3v+	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes
GETAWAY	R1e+	R maximal autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
GETAWAY	RTe+	R maximal index / weighted by atomic Sanderson electronegativities
GETAWAY	R3p+	R maximal autocorrelation of lag 3 / weighted by atomic polarizabilities
2D & 3D MMFF94s		
Class	**Symbol**	**Name**
Topographical	SPI	superpendentic index
Topographical	MAXDP	maximal electrotopological positive variation
Topographical	ECC	eccentricity
Topographical	DECC	eccentric
Information	IC2	information content index (neighborhood symmetry of 2-order)
Edge Adjacency	EEig01d	Eigenvalue 01 from edge adj. matrix weighted by dipole moments
Burgen Eigenvalue	BEHm1	highest eigenvalue n. 1 of Burden matrix / weighted by atomic masses
Burgen Eigenvalue	BEHv4	highest eigenvalue n. 4 of Burden matrix / weighted by atomic van der Waals volumes
Burgen Eigenvalue	BEHp8	highest eigenvalue n. 8 of Burden matrix / weighted by atomic polarizabilities
3D-MoRSE	Mor06p	3D-MoRSE - signal 06 / weighted by atomic polarizabilities
WHIM	E2v	2nd component accessibility directional WHIM index / weighted by atomic van der Waals volumes
WHIM	E1e	1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
GETAWAY	R5u+	R maximal autocorrelation of lag 5 / unweighted
GETAWAY	RTe+	R maximal index / weighted by atomic Sanderson electronegativities

Table A.2 shows the equations for each of the seven descriptor sets. They are linear equations in the form of Equation 3.2. These models minimize the BIC value calculated using Equation 2.2.

Table A.2: Equations developed with BIC method regression

Descriptor Class	Equation to calculate normalized AR
2D	$-0.022 \times (Ram) + 0.5049 \times (TI2) + -0.4619 \times (Rww) + 0.1205 \times (Jhetv) + 0.486 \times (Jhete) + -0.2023 \times (S1K) + 0.0534 \times (Lop) + -0.2367 \times (X0v) + -1.5073 \times (SIC1) + -0.7135 \times (CIC2) + 0.6381 \times (MATS2p) + 0.0276 \times (ESpm15u) + 0.1546 \times (ESpm04d) + -0.2598 \times (BEHm4)$
3D GAFF	$0.022 \times (SP01) + 0.4887 \times (L/Bw) + -0.7619 \times (DISPe) + 0.1441 \times (RDF015m) + 0.0948 \times (RDF025m) + -0.1023 \times (Mor32u) + 0.3981 \times (Mor13p) + 0.8508 \times (E1e) + 0.2341 \times (L2s) + 0.1378 \times (G2s) + -0.5789 \times (Ks) + -0.3081 \times (R3m+) + -0.0498 \times (R5e) + -0.0007 \times (R3p)$
2D & 3D GAFF	$-0.385 \times (S2K) + 0.2663 \times (AAC) + -0.2883 \times (SIC1) + 0.7871 \times (MATS1v) + -0.0153 \times (EPS0) + 0.0041 \times (ESpm04d) + -0.389 \times (DISPv) + -0.1906 \times (Mor03u) + -0.071 \times (Mor03e) + 0.2651 \times (E1e) + 0.0354 \times (G3s) + 0.2342 \times (Kv) + 0.5072 \times (Kp) + 0.0355 \times (R2u)$
3D Ghemical	$-0.401 \times (SP01) + -0.0137 \times (SP03) + 0.8226 \times (MEcc) + 0.3312 \times (RDF040e) + 0.7006 \times (RDF045p) + 1.0263 \times (Mor27u) + -0.0271 \times (Mor18m) + -0.405 \times (Mor01v) + 1.2038 \times (Mor13e) + 0.538 \times (Mor28p) + -0.0007 \times (Dm) + -0.6611 \times (HATS5v) + 0.4669 \times (R5u) + 0.7206 \times (R5u+)$
2D & 3D Ghemical	$0.32 \times (BAC) + 0.0202 \times (X0v) + 0.4154 \times (SIC2) + 0.249 \times (MATS3p) + -0.2381 \times (GATS1v) + -0.1528 \times (BEHe8) + 0.4436 \times (MEcc) + -0.0027 \times (RDF035v) + -0.0313 \times (Mor16v) + 0.6902 \times (Mor13e) + 0 \times (Gs) + -0.41 \times (ISH) + 0.0395 \times (HATS3u) + 0.1177 \times (R1u) +$
3D MMFF94s	$0 \times (SP13) + -0.4863 \times (PJI3) + -0.8888 \times (RDF040e) + 0.0078 \times (Mor11m) + 0.1359 \times (Mor12e) + 0.0225 \times (Mor05p) + -0.7705 \times (E1s) + -0.2998 \times (H3u) + 0.0024 \times (HATS3u) + 0.0329 \times (HATS3m) + 0.3499 \times (R3v+) + -0.0731 \times (R1e+) + -1.3313 \times (RTe+) + -0.5439 \times (R3p+)$
2D & 3D MMFF94s	$0.457 \times (SPI) + -0.7142 \times (MAXDP) + -0.0001 \times (ECC) + -0.1901 \times (DECC) + 0.4275 \times (IC2) + 0.0012 \times (EEig01d) + 0.0004 \times (BEHm1) + -0.0643 \times (BEHv4) + 0.0179 \times (BEHp8) + 0.0566 \times (Mor06p) + 0.0918 \times (E2v) + 0.0964 \times (E1e) + -0.3047 \times (R5u+) + -0.8518 \times (RTe+)$

B. PCA Method

In this section, the matrices used in PCA are shown. The matrices shown in Table B.1 through Table B.3 are from the PCA analysis using 16 solvents and the 2D descriptor set. The same procedure is used for the other six descriptor sets and also when the training set is expanded to 51 solvents. Table B.1 contains the normalized (Equation 3.1) descriptor values for each solvent. Table B.2 contains the eigenvectors that are used in the calculation of the principal component factors shown in Table B.3. These factors are then linearly regressed to produce the QSAR model.

Table B.1: 2D descriptor matrix

	Acetone	Aceto-nitrile	Benzyl Alcohol	Carbon Tetra-chloride	Cyclo-hexane	Ethanol	Ethyl Acetate	Ethylene Glycol	Hexane	Iso-propanol	Methanol	Methylene Dichloride	Propylene Glycol	Sulfolane	*t*-Amyl Alcohol	Toluene
ZM1	-0.504	-1.094	1.659	0.283	0.676	-1.094	0.283	-0.701	0.086	-0.504	-1.487	-1.094	-0.111	1.659	0.676	1.266
ZM1V	0.242	-0.196	1.555	-1.057	-0.853	-0.634	1.701	0.387	-1.072	-0.415	-0.780	-1.539	0.606	1.498	0.023	0.533
ZM2	-0.592	-1.030	1.776	0.022	0.723	-1.030	0.197	-0.680	0.022	-0.592	-1.293	-1.030	-0.153	1.776	0.548	1.337
ZM2V	0.102	-0.262	2.193	-0.787	-0.262	-0.808	1.647	-0.262	-0.626	-0.398	-1.126	-1.212	0.193	-0.323	0.374	1.556
Qindex	-0.364	-0.894	1.225	0.695	0.695	-0.894	-0.364	-0.894	-0.894	-0.364	-0.894	-0.894	-0.364	2.285	0.695	1.225
SNar	-0.652	-0.906	1.941	-0.473	1.256	-0.906	0.212	-0.473	0.392	-0.652	-1.338	-0.906	-0.220	1.256	-0.041	1.509
HNar	-0.666	-0.666	1.569	-0.749	2.102	-0.666	-0.026	-0.206	0.372	-0.666	-1.358	-0.666	-0.306	0.880	-0.448	1.500
GNar	-0.583	-0.776	1.539	-0.569	1.784	-0.776	0.099	-0.244	0.355	-0.583	-1.676	-0.776	-0.185	1.130	-0.244	1.504
Xt	0.664	1.415	-0.910	0.345	-0.746	1.415	-0.308	0.345	-0.409	0.664	-2.235	1.415	0.035	-0.746	-0.129	-0.813
Dz	-0.541	-1.251	1.731	0.121	0.311	-1.109	0.879	-0.257	0.311	-0.541	-1.677	-1.204	0.311	1.447	0.595	0.879
Ram	0.323	-0.968	0.323	1.614	-0.968	-0.968	0.323	-0.968	-0.968	0.323	-0.968	-0.968	0.323	1.614	1.614	0.323
Pol	-0.891	-0.891	2.328	-0.891	0.489	-0.891	0.489	-0.431	0.489	-0.891	-0.891	-0.891	0.029	0.948	0.489	1.408
LPRS	-0.614	-1.094	1.932	-0.103	0.531	-1.094	0.671	-0.547	0.755	-0.614	-1.550	-1.094	-0.010	1.099	0.544	1.187
VDA	-0.647	-1.075	2.034	-0.204	0.402	-1.075	0.791	-0.531	1.024	-0.647	-1.463	-1.075	-0.018	0.902	0.480	1.102
MSD	-0.006	1.073	-1.049	-0.737	-0.897	1.073	-0.220	0.627	0.235	-0.006	2.232	1.073	-0.184	-1.334	-0.817	-1.066
SMTI	-0.680	-0.935	2.426	-0.324	0.542	-0.935	0.364	-0.655	0.491	-0.680	-1.088	-0.935	-0.273	1.153	0.211	1.318
SMTIV	-0.416	-0.697	2.607	-0.785	-0.223	-0.816	0.962	-0.208	-0.253	-0.608	-0.964	-1.098	0.162	1.301	0.058	0.977
GMTI	-0.688	-0.822	2.569	-0.495	0.695	-0.822	0.175	-0.628	0.353	-0.688	-0.896	-0.822	-0.376	1.082	-0.063	1.424
GMTIV	-0.526	-0.657	2.957	-0.735	-0.254	-0.722	0.641	-0.085	-0.379	-0.608	-0.815	-0.819	0.183	1.142	-0.204	0.880
Xu	-0.538	-1.103	1.702	-0.046	0.597	-1.103	0.742	-0.416	0.865	-0.538	-1.857	-1.103	0.104	1.008	0.563	1.120
SPI	-0.039	-0.804	0.545	0.711	-1.848	-0.804	1.265	-0.168	0.898	-0.039	-1.848	-0.804	0.640	0.748	1.415	0.130
W	-0.697	-0.979	2.402	-0.303	0.317	-0.979	0.599	-0.641	0.768	-0.697	-1.148	-0.979	-0.190	0.993	0.373	1.162
WA	-0.561	-1.010	1.551	-0.293	0.245	-1.010	1.140	-0.113	1.677	-0.561	-1.905	-1.010	0.245	0.398	0.425	0.782
Har	-0.580	-1.067	1.860	-0.011	0.748	-1.067	0.382	-0.625	0.314	-0.580	-1.474	-1.067	-0.101	1.434	0.504	1.328
Har2	-0.607	-1.062	1.957	-0.039	0.643	-1.062	0.416	-0.645	0.348	-0.607	-1.402	-1.062	-0.115	1.400	0.529	1.306
QW	-0.682	-1.043	2.149	-0.176	-0.068	-1.043	0.980	-0.610	1.197	-0.682	-1.260	-1.043	-0.032	0.836	0.691	0.787
TI1	-0.086	0.372	1.139	-0.550	1.139	0.372	-1.621	-0.222	-1.880	-0.086	0.811	0.372	-0.761	1.139	-1.276	1.139
TI2	-0.660	0.040	0.351	-1.080	-1.360	0.040	1.551	0.826	2.468	-0.660	-0.660	0.040	0.479	-0.788	0.122	-0.712
HyDp	-0.708	-0.917	2.564	-0.411	0.184	-0.917	0.660	-0.619	1.017	-0.708	-1.036	-0.917	-0.232	0.749	0.244	1.047
RHyDp	-0.586	-1.067	1.880	-0.008	0.715	-1.067	0.387	-0.634	0.310	-0.586	-1.453	-1.067	-0.104	1.437	0.522	1.321
w	-0.648	-0.785	2.505	-0.456	0.833	-0.785	-0.017	-0.620	0.065	-0.648	-0.867	-0.785	-0.401	1.162	-0.127	1.573
ww	-0.611	-0.676	2.681	-0.518	0.832	-0.676	-0.186	-0.583	-0.075	-0.611	-0.713	-0.676	-0.463	0.989	-0.315	1.599
Rww	-0.084	-0.872	0.102	0.861	-1.266	-0.872	1.507	-0.163	1.381	-0.084	-1.502	-0.872	0.703	0.073	1.727	-0.638

D/D	-0.544	-1.142	1.499	0.253	-0.305	-1.142	1.248	-0.544	1.248	-0.544	-1.540	-1.142	0.253	0.687	1.248	0.465
Wap	-0.612	-0.702	2.555	-0.485	0.854	-0.702	-0.196	-0.594	-0.141	-0.612	-0.757	-0.702	-0.449	1.216	-0.268	1.596
WhetZ	-0.671	-0.996	2.195	-0.777	0.749	-0.930	0.579	-0.573	1.320	-0.591	-1.126	-1.079	-0.037	0.244	0.731	0.963
Whetm	-0.669	-0.994	2.194	-0.791	0.749	-0.928	0.580	-0.571	1.320	-0.589	-1.124	-1.081	-0.036	0.244	0.731	0.963
Whetv	-0.813	-1.173	2.011	-0.386	0.279	-0.995	1.553	-0.402	0.762	-0.636	-1.234	-1.111	0.196	0.861	0.627	0.460
Whete	-0.756	-1.081	2.111	-0.363	0.663	-1.015	0.498	-0.657	1.234	-0.676	-1.212	-1.040	-0.121	0.892	0.646	0.877
Whetp	-0.772	-1.136	2.038	-0.561	0.262	-0.942	1.775	-0.311	0.729	-0.580	-1.187	-1.127	0.297	0.408	0.671	0.437
J	0.264	-0.997	-0.099	1.542	-0.327	-0.997	0.817	-0.373	0.292	0.264	-2.152	-0.997	0.658	0.403	1.804	-0.103
JhetZ	-0.147	-0.378	-0.179	3.172	-0.718	-0.796	0.139	-0.545	-0.517	-0.399	-1.113	0.838	-0.225	0.920	0.060	-0.113
Jhetm	-0.160	-0.378	-0.190	3.198	-0.701	-0.776	0.112	-0.538	-0.510	-0.399	-1.077	0.886	-0.234	0.857	0.038	-0.126
Jhetv	0.430	0.254	0.769	1.405	-0.059	-1.331	-0.503	-1.066	0.426	-0.381	-2.187	-0.584	-0.297	0.729	0.995	1.401
Jhete	0.491	-0.112	0.410	1.831	-0.994	-1.204	1.225	-0.551	-0.469	-0.168	-2.034	-0.890	0.285	0.562	1.033	0.585
Jhetp	0.186	0.055	0.509	1.961	-0.109	-1.263	-0.664	-1.069	0.293	-0.494	-1.942	-0.082	-0.449	1.307	0.658	1.102
MAXDN	0.098	-0.731	-0.057	1.601	-1.559	-0.178	0.514	0.236	-1.399	0.098	-0.455	-0.117	0.533	2.267	0.328	-1.178
MAXDP	0.797	-0.252	0.885	-0.815	-1.481	-0.019	1.149	0.034	-1.265	0.435	-0.549	-0.873	0.487	1.717	1.153	-1.404
DELS	0.093	-0.564	0.147	0.336	-1.334	-0.418	0.840	0.563	-1.064	-0.134	-0.750	-0.572	1.003	2.671	0.316	-1.133
TIE	-0.472	-1.095	0.753	-0.233	0.013	-0.990	0.369	-0.471	0.435	-0.417	-1.222	-1.122	0.526	2.359	1.631	-0.063
S0K	-0.377	-0.594	2.064	-0.793	-1.423	-0.594	1.281	-0.725	0.235	-0.377	-1.074	-0.942	0.601	0.957	0.933	0.826
S1K	-0.556	-1.480	0.731	1.374	-0.172	-1.115	0.977	-0.363	1.262	-0.333	-1.897	-0.638	0.419	0.562	1.232	-0.003
S2K	-1.080	-0.623	0.388	-0.381	0.124	-0.144	0.770	0.842	2.974	-0.824	-1.130	0.482	0.077	-0.639	-0.509	-0.326
S3K	-1.197	-0.130	-0.486	-1.197	-0.486	-0.130	1.479	0.941	1.649	-1.197	-0.423	-0.130	0.896	-0.576	1.630	-0.643
PHI	-0.939	-0.676	-0.289	0.162	-0.390	-0.017	0.688	0.864	2.890	-0.634	-0.923	1.050	0.179	-0.831	-0.340	-0.794
BLI	-0.555	-0.874	-1.189	0.782	0.241	0.335	-0.721	-0.748	0.911	0.007	-0.184	2.675	-0.646	1.086	-0.106	-1.014
PW2	0.321	-0.794	0.488	0.989	0.321	-0.794	0.341	-0.233	-0.053	0.321	-3.017	-0.794	0.435	0.956	0.936	0.575
PW3	-1.057	-1.057	1.380	-1.057	0.967	-1.057	0.578	0.295	0.675	-1.057	-1.057	-1.057	0.562	1.218	0.562	1.161
PJI2	0.939	0.939	0.001	0.939	-1.879	0.939	0.939	-0.470	0.001	0.939	-1.879	0.939	-0.470	-0.470	-0.470	-0.940
CSI	-0.761	-0.938	2.240	-0.585	0.827	-0.938	0.533	-0.467	0.945	-0.761	-1.173	-0.938	-0.173	0.710	0.121	1.357
ECC	-0.750	-0.988	2.103	-0.513	0.557	-0.988	0.795	-0.394	1.270	-0.750	-1.345	-0.988	-0.037	0.557	0.319	1.152
AECC	-0.790	-0.884	1.622	-0.733	0.629	-0.884	1.007	0.062	1.764	-0.790	-1.641	-0.884	0.175	0.142	0.251	0.954
DECC	-0.231	0.126	1.221	-0.515	-2.168	0.126	1.278	0.415	1.278	-0.231	-2.168	0.126	0.312	0.364	0.126	-0.060
MDDD	-0.415	-0.854	1.379	-0.113	-1.491	-0.854	1.379	-0.056	1.697	-0.415	-1.491	-0.854	0.461	0.735	0.742	0.149
UNIP	-0.734	-1.028	1.908	-0.440	1.028	-1.028	0.734	-0.440	1.028	-0.734	-1.321	-1.028	-0.147	0.734	0.147	1.321
CENT	-0.467	-0.900	2.348	0.183	-1.117	-0.900	0.616	-0.683	0.616	-0.467	-1.117	-0.900	0.074	1.265	1.049	0.399
VAR	-0.495	-0.855	2.385	-0.135	-1.215	-0.855	0.945	-0.495	0.945	-0.495	-1.215	-0.855	0.225	0.585	0.945	0.585
BAC	0.297	-0.582	-0.934	1.527	-1.461	-0.582	1.000	-0.055	0.648	0.297	-0.758	-0.582	0.824	-0.582	2.054	-1.110
Lop	-0.137	0.124	0.630	-0.355	-2.120	0.124	1.447	0.325	1.755	-0.137	-2.120	0.124	0.254	0.122	0.124	-0.159
ICR	-0.208	0.035	1.495	-0.410	-2.049	0.035	1.263	0.221	1.549	-0.208	-2.049	0.035	0.155	0.187	0.035	-0.090
MWC01	-0.573	-1.055	1.839	-0.090	0.874	-1.055	0.392	-0.573	0.392	-0.573	-1.538	-1.055	-0.090	1.357	0.392	1.357

MWC02	-0.204	-1.095	1.220	0.486	0.736	-1.095	0.486	-0.445	0.341	-0.204	-2.313	-1.095	0.181	1.220	0.736	1.046
MWC03	-0.257	-1.086	1.237	0.357	0.796	-1.086	0.484	-0.380	0.357	-0.257	-2.305	-1.086	0.213	1.237	0.701	1.074
MWC04	-0.145	-1.081	1.135	0.555	0.717	-1.081	0.462	-0.427	0.272	-0.145	-2.392	-1.081	0.209	1.233	0.770	0.998
MWC05	-0.192	-1.077	1.157	0.452	0.755	-1.077	0.463	-0.377	0.296	-0.192	-2.384	-1.077	0.237	1.244	0.747	1.022
MWC06	-0.115	-1.068	1.092	0.584	0.702	-1.068	0.451	-0.411	0.242	-0.115	-2.441	-1.068	0.227	1.232	0.787	0.971
MWC07	-0.154	-1.066	1.112	0.501	0.732	-1.066	0.452	-0.373	0.264	-0.154	-2.435	-1.066	0.247	1.243	0.768	0.991
MWC08	-0.097	-1.058	1.066	0.600	0.693	-1.058	0.443	-0.400	0.224	-0.097	-2.476	-1.058	0.239	1.230	0.796	0.953
MWC09	-0.130	-1.056	1.084	0.531	0.717	-1.056	0.445	-0.370	0.244	-0.130	-2.470	-1.056	0.254	1.241	0.780	0.972
MWC10	-0.085	-1.049	1.048	0.609	0.686	-1.049	0.438	-0.392	0.212	-0.085	-2.502	-1.049	0.246	1.227	0.801	0.942
TWC	-0.232	-1.084	1.275	0.420	0.736	-1.084	0.468	-0.434	0.312	-0.232	-2.289	-1.084	0.180	1.262	0.723	1.062
SRW01	-0.542	-1.119	1.769	0.036	0.614	-1.119	0.614	-0.542	0.614	-0.542	-1.697	-1.119	0.036	1.192	0.614	1.192
SRW02	-0.573	-1.055	1.839	-0.090	0.874	-1.055	0.392	-0.573	0.392	-0.573	-1.538	-1.055	-0.090	1.357	0.392	1.357
SRW04	-0.483	-1.097	1.604	0.376	0.622	-1.097	0.253	-0.729	0.008	-0.483	-1.466	-1.097	-0.115	1.727	0.744	1.236
SRW06	-0.535	-1.050	1.552	0.467	0.522	-1.050	0.088	-0.779	-0.237	-0.535	-1.240	-1.050	-0.183	1.931	0.901	1.199
SRW08	-0.593	-0.971	1.558	0.427	0.439	-0.971	-0.045	-0.791	-0.383	-0.593	-1.059	-0.971	-0.272	2.100	0.934	1.191
SRW10	-0.627	-0.889	1.568	0.345	0.347	-0.889	-0.152	-0.776	-0.469	-0.627	-0.928	-0.889	-0.352	2.264	0.906	1.169
MPC01	-0.573	-1.055	1.839	-0.090	0.874	-1.055	0.392	-0.573	0.392	-0.573	-1.538	-1.055	-0.090	1.357	0.392	1.357
MPC02	-0.447	-1.098	1.505	0.529	0.529	-1.098	0.203	-0.773	-0.122	-0.447	-1.423	-1.098	-0.122	1.830	0.854	1.179
piPC01	-0.130	-0.130	1.668	-0.130	0.561	-1.177	0.561	-0.588	0.245	-0.588	-2.011	-1.177	-0.130	1.294	0.245	1.489
piPC02	0.128	-0.365	1.496	0.315	0.315	-1.208	0.477	-0.714	-0.094	-0.365	-2.050	-1.208	-0.094	1.462	0.477	1.428
piPC03	-0.899	-0.899	1.884	-0.899	0.713	-0.899	0.434	-0.325	0.249	-0.899	-0.899	-0.899	0.011	1.287	0.249	1.792
TPC	-0.626	-0.969	2.074	-0.318	0.921	-0.969	0.289	-0.556	0.377	-0.626	-1.331	-0.969	-0.187	1.261	0.149	1.481
piID	-0.538	-0.716	2.414	-0.392	0.516	-0.869	0.173	-0.567	0.117	-0.618	-1.134	-0.869	-0.296	1.015	-0.050	1.813
PCR	0.101	1.295	2.202	-0.602	-0.602	-0.602	-0.005	-0.602	-0.602	-0.602	-0.602	-0.602	-0.602	0.244	-0.602	2.186
CID	-0.579	-1.074	1.857	-0.099	0.802	-1.074	0.494	-0.542	0.537	-0.579	-1.594	-1.074	-0.045	1.225	0.436	1.309
BID	-0.487	-1.107	1.767	0.136	0.624	-1.107	0.523	-0.614	0.473	-0.487	-1.722	-1.107	-0.002	1.279	0.630	1.202
X0	-0.426	-1.190	1.542	0.384	0.158	-1.190	0.815	-0.569	0.671	-0.426	-1.810	-1.190	0.194	1.110	1.004	0.921
X1	-0.661	-1.036	1.936	-0.344	0.836	-1.036	0.564	-0.446	0.734	-0.661	-1.525	-1.036	-0.026	1.080	0.318	1.301
X2	-0.102	-1.127	1.079	1.167	0.287	-1.127	0.350	-0.834	-0.127	-0.102	-1.834	-1.127	-0.032	1.538	1.081	0.910
X3	-0.912	-0.912	1.922	-0.912	0.935	-0.912	0.154	-0.297	0.266	-0.912	-0.912	-0.912	0.093	1.498	0.394	1.420
X0A	0.597	0.696	-1.476	0.671	-1.725	0.696	-0.173	0.100	-0.508	0.597	1.912	0.696	0.137	-1.054	0.274	-1.439
X1A	-0.154	0.810	-0.791	-0.724	-0.724	0.810	-0.324	0.298	-0.109	-0.154	2.981	0.810	-0.220	-1.035	-0.635	-0.835
X2A	0.648	1.369	-0.756	0.221	-0.589	1.369	-0.129	0.221	-0.184	0.648	-2.553	1.369	-0.051	-0.684	-0.245	-0.650
X0v	-0.530	-1.346	0.856	1.276	0.603	-1.170	0.417	-1.039	1.099	-0.431	-1.770	-0.474	-0.300	1.135	0.951	0.725
X1v	-0.728	-1.217	0.674	0.356	1.101	-0.912	-0.015	-0.801	1.014	-0.515	-1.499	-0.321	-0.365	2.356	0.372	0.501
X2v	-0.453	-1.027	0.164	2.020	0.564	-0.949	-0.439	-0.840	0.217	-0.297	-1.214	-0.452	-0.349	2.277	0.602	0.174
X0Av	-0.002	-0.626	-1.273	2.237	-0.162	-0.074	-0.450	-1.201	0.622	0.230	-0.026	2.117	-0.730	-0.250	0.390	-0.801
X1Av	-0.557	-0.872	-1.186	0.781	0.241	0.338	-0.718	-0.751	0.910	0.007	-0.186	2.675	-0.646	1.087	-0.106	-1.017

X2Av	-0.143	-0.523	-0.721	1.494	0.103	-0.080	-0.711	-0.523	0.454	0.156	-1.601	2.774	-0.360	0.401	-0.114	-0.605
X0sol	-0.550	-1.244	1.240	1.783	-0.018	-1.244	0.579	-0.680	0.449	-0.550	-1.808	-0.446	0.014	1.047	0.751	0.676
X1sol	-0.794	-1.145	1.633	0.605	0.605	-1.145	0.351	-0.593	0.510	-0.794	-1.602	-0.365	-0.201	1.775	0.120	1.039
X2sol	-0.290	-0.916	0.431	2.774	-0.052	-0.916	-0.014	-0.737	-0.305	-0.290	-1.347	-0.376	-0.247	1.524	0.432	0.328
XMOD	-0.840	-1.191	1.360	1.145	0.320	-1.133	0.338	-0.521	0.235	-0.840	-1.495	0.034	-0.190	2.099	-0.032	0.710
RDCHI	-0.707	-0.941	1.843	-0.574	1.212	-0.941	0.468	-0.344	0.805	-0.707	-1.458	-0.941	-0.120	1.028	-0.023	1.402
RDSQ	-0.617	-1.020	1.986	-0.068	0.694	-1.020	0.280	-0.676	0.181	-0.617	-1.276	-1.020	-0.186	1.547	0.449	1.363
ISIZ	-0.463	-1.192	0.804	-1.352	1.260	-0.656	0.364	-0.463	1.728	-0.059	-1.192	-1.352	0.150	0.582	1.260	0.582
IAC	-0.257	-1.000	1.050	-1.910	0.375	-0.600	0.865	-0.124	0.569	-0.026	-1.220	-1.203	0.520	1.802	1.064	0.096
AAC	0.287	0.954	0.193	-2.043	-1.246	-0.002	0.629	0.596	-1.397	-0.144	0.112	1.210	0.450	1.690	-0.364	-0.925
IDE	-0.583	-0.733	1.353	-0.636	0.370	-0.733	1.077	0.255	1.516	-0.583	-2.410	-0.733	0.370	0.319	0.341	0.807
IDM	-0.319	-1.100	1.346	0.262	0.677	-1.100	0.645	-0.364	0.616	-0.319	-2.268	-1.100	0.221	1.068	0.689	1.045
IDDE	-0.580	-0.432	1.289	-0.704	-1.708	-0.432	0.958	-0.318	0.496	-0.580	-1.708	-0.432	0.964	1.003	0.783	1.400
IDDM	-0.387	-1.122	1.418	0.192	0.714	-1.122	0.658	-0.387	0.656	-0.387	-2.128	-1.122	0.187	1.079	0.658	1.092
IDET	-0.733	-0.933	2.453	-0.505	0.305	-0.933	0.662	-0.563	0.885	-0.733	-1.103	-0.933	-0.165	0.831	0.290	1.177
IDMT	-0.705	-0.904	2.600	-0.351	0.253	-0.904	0.465	-0.682	0.584	-0.705	-0.977	-0.904	-0.284	1.035	0.304	1.175
IVDE	-0.254	0.000	0.340	-0.466	-2.184	0.000	1.287	0.195	0.000	-0.254	-2.184	0.000	1.078	1.097	0.795	0.550
IVDM	-0.535	-1.064	1.566	-0.159	0.900	-1.064	0.649	-0.307	0.786	-0.535	-1.969	-1.064	0.124	1.043	0.424	1.204
HVcpx	-0.834	-0.737	1.484	-1.072	0.787	-0.737	0.976	0.325	1.704	-0.834	-1.734	-0.737	0.276	0.119	0.042	0.973
HDcpx	-0.089	-0.891	1.023	0.377	0.611	-0.891	0.602	-0.158	0.569	-0.089	-2.881	-0.891	0.326	0.879	0.656	0.849
Uindex	-0.512	-0.985	1.300	0.161	0.188	-0.985	1.134	-0.436	1.182	-0.512	-2.203	-0.985	0.238	0.663	1.044	0.708
Vindex	0.716	1.462	-0.973	0.652	-0.789	1.462	-0.461	-0.118	-0.702	0.716	-1.924	1.462	-0.090	-0.629	0.071	-0.857
Xindex	0.619	1.085	-0.949	0.697	-0.694	1.085	-0.042	0.243	-0.304	0.619	-2.733	1.085	0.240	-0.580	0.426	-0.797
Yindex	-0.959	-0.959	0.103	-0.959	0.266	-0.959	0.612	0.908	0.208	-0.959	-0.959	-0.959	1.375	0.864	2.131	0.246
IC0	0.287	0.954	0.193	-2.043	-1.246	-0.002	0.629	0.596	-1.397	-0.144	0.112	1.210	0.450	1.690	-0.364	-0.925
TIC0	-0.257	-1.000	1.050	-1.910	0.375	-0.600	0.865	-0.124	0.569	-0.026	-1.220	-1.203	0.520	1.802	1.064	0.096
SIC0	0.152	1.597	-0.446	-0.504	-1.260	0.119	-0.080	0.343	-1.393	-0.330	0.933	2.353	-0.089	0.401	-0.828	-0.969
CIC0	-0.263	-1.450	0.663	-0.825	1.354	-0.369	0.269	-0.362	1.603	0.225	-1.177	-1.880	0.185	0.055	1.068	0.903
BIC0	0.249	1.053	-0.313	0.070	-0.801	0.361	0.076	0.509	-0.881	-0.035	-2.160	2.542	0.138	0.317	-0.443	-0.684
IC1	-0.327	0.159	1.548	-2.192	-1.761	0.352	0.820	0.445	-1.236	0.135	0.159	-0.434	1.055	0.412	0.159	0.708
TIC1	-0.518	-1.016	1.802	-1.735	-0.436	-0.396	0.849	-0.165	0.229	0.051	-1.016	-1.332	0.778	0.778	1.145	0.982
SIC1	-0.264	1.271	0.664	-1.394	-2.028	0.573	0.273	0.476	-1.693	-0.096	1.271	1.006	0.580	-0.159	-0.564	0.085
CIC1	0.029	-1.291	-0.212	-0.179	2.096	-0.606	-0.022	-0.454	1.976	0.102	-1.291	-1.280	-0.315	0.372	0.892	0.186
BIC1	-0.012	0.865	0.509	-0.607	-1.356	0.807	0.398	0.695	-1.080	0.212	-2.526	1.519	0.738	-0.044	-0.198	0.079
IC2	-0.496	-0.220	0.946	-2.456	-1.310	0.562	1.268	-0.058	0.097	0.197	-0.220	-1.208	1.321	0.588	0.492	0.496
TIC2	-0.653	-0.988	1.238	-1.716	-0.597	-0.242	1.113	-0.415	1.055	0.011	-0.988	-1.378	0.948	0.762	1.163	0.686
SIC2	-0.421	0.614	0.563	-2.648	-1.613	0.944	1.043	0.078	-0.417	0.144	0.614	-0.619	1.203	0.294	0.021	0.200
CIC2	0.387	-0.988	-0.298	1.205	2.538	-1.047	-0.945	-0.116	1.138	-0.032	-0.988	-0.228	-1.158	-0.020	0.467	0.086

BIC2	-0.195	0.462	0.478	-2.076	-1.201	1.130	1.043	0.334	-0.159	0.366	-2.076	-0.087	1.281	0.299	0.223	0.179
ATS1m	-0.625	-1.242	0.886	1.510	0.320	-1.141	0.318	-0.479	0.017	-0.625	-1.845	0.293	-0.097	2.005	0.123	0.582
ATS2m	-0.313	-1.162	0.566	2.380	0.136	-1.080	0.251	-0.580	-0.237	-0.313	-2.017	0.498	-0.036	1.081	0.413	0.413
ATS3m	-1.050	-1.050	1.553	-1.050	0.621	-1.050	0.717	0.181	0.621	-1.050	-1.050	-1.050	0.654	1.174	0.717	1.110
ATS1v	-0.402	-0.950	1.428	0.328	1.025	-1.095	0.127	-0.712	0.707	-0.402	-2.146	-0.727	-0.121	1.110	0.531	1.300
ATS2v	-0.301	-1.116	1.316	0.883	0.883	-1.278	0.020	-0.867	0.408	-0.301	-1.861	-0.884	-0.080	1.154	0.786	1.237
ATS3v	-0.928	-0.928	1.717	-0.928	0.953	-0.928	0.776	-0.612	0.953	-0.928	-0.928	-0.928	-0.151	0.584	0.776	1.504
ATS1e	-0.476	-1.259	1.438	0.364	0.723	-1.132	0.715	-0.294	0.339	-0.476	-2.022	-0.982	0.189	1.351	0.471	1.054
ATS2e	-0.124	-1.212	1.005	1.044	0.454	-1.105	0.597	-0.467	-0.023	-0.124	-2.302	-0.948	0.232	1.355	0.807	0.810
ATS3e	-1.050	-1.050	1.554	-1.050	0.623	-1.050	0.718	0.176	0.623	-1.050	-1.050	-1.050	0.649	1.173	0.718	1.113
ATS1p	-0.445	-0.985	1.310	0.622	0.939	-1.116	0.012	-0.783	0.637	-0.445	-2.142	-0.434	-0.204	1.387	0.449	1.200
ATS2p	-0.368	-1.151	1.217	1.320	0.813	-1.300	-0.092	-0.934	0.360	-0.368	-1.804	-0.553	-0.173	1.198	0.685	1.151
ATS3p	-0.913	-0.913	1.724	-0.913	0.978	-0.913	0.778	-0.656	0.978	-0.913	-0.913	-0.913	-0.221	0.500	0.778	1.531
MATS1m	0.013	-0.201	0.185	-1.360	1.730	-0.201	-0.433	-0.330	1.730	0.013	-1.360	-1.360	-0.073	-0.205	0.123	1.730
MATS2m	-0.663	-1.301	0.159	0.495	1.573	-1.301	-0.007	-1.301	1.573	-0.663	0.136	0.855	-0.701	-0.114	-0.315	1.573
MATS3m	-0.523	-0.523	-0.921	-0.523	1.641	-0.523	-0.523	1.641	1.641	-0.523	-0.523	-0.523	0.018	-0.521	-0.956	1.641
MATS1v	-0.387	-0.604	-0.214	1.345	1.345	-0.604	-0.838	-0.733	1.345	-0.387	-1.774	1.345	-0.475	-0.434	-0.277	1.345
MATS2v	-0.715	-1.286	0.020	1.286	1.286	-1.286	-0.129	-1.286	1.286	-0.715	0.000	1.286	-0.750	0.121	-0.404	1.286
MATS3v	-0.350	-0.350	-0.693	-0.350	1.514	-0.350	-0.350	1.514	1.514	-0.350	-0.350	-0.350	0.116	-1.958	-0.723	1.514
MATS1e	-0.028	-0.242	0.142	-1.395	1.680	-0.242	-0.473	-0.370	1.680	-0.028	-1.395	-1.395	-0.114	0.421	0.081	1.680
MATS2e	-0.676	-1.314	0.146	0.482	1.560	-1.314	-0.021	-1.314	1.560	-0.676	0.123	0.841	-0.715	0.088	-0.328	1.560
MATS3e	-0.297	-0.297	-0.617	-0.297	1.443	-0.297	-0.297	1.443	1.443	-0.297	-0.297	-0.297	0.138	-2.269	-0.645	1.443
MATS1p	0.045	-0.166	0.214	-1.307	1.735	-0.166	-0.394	-0.292	1.735	0.045	-1.307	-1.307	-0.040	-0.686	0.153	1.735
MATS2p	-0.690	-1.325	0.130	0.465	1.539	-1.325	-0.037	-1.325	1.539	-0.690	0.107	0.823	-0.728	0.323	-0.343	1.539
MATS3p	-0.482	-0.482	-0.872	-0.482	1.633	-0.482	-0.482	1.633	1.633	-0.482	-0.482	-0.482	0.046	-0.942	-0.906	1.633
GATS1m	-0.282	-0.156	-0.535	2.496	-1.293	-0.156	0.412	0.222	-1.293	-0.282	0.222	0.980	-0.031	1.373	-0.384	-1.293
GATS2m	0.877	1.130	-0.472	-1.145	-1.145	1.130	0.561	1.130	-1.145	0.877	-1.145	-1.145	0.751	0.082	0.806	-1.145
GATS1v	0.306	0.509	-0.104	-1.330	-1.330	0.509	1.429	1.123	-1.330	0.306	1.123	-1.330	0.713	0.593	0.142	-1.330
GATS2v	0.880	1.133	-0.469	-1.143	-1.143	1.133	0.564	1.133	-1.143	0.880	-1.143	-1.143	0.754	0.044	0.808	-1.143
GATS1e	-0.164	-0.030	-0.433	2.785	-1.237	-0.030	0.573	0.372	-1.237	-0.164	0.372	1.176	0.103	-0.577	-0.272	-1.237
GATS2e	0.865	1.118	-0.481	-1.154	-1.154	1.118	0.550	1.118	-1.154	0.865	-1.154	-1.154	0.739	0.238	0.794	-1.154
GATS1p	-0.273	-0.144	-0.531	2.563	-1.305	-0.144	0.436	0.242	-1.305	-0.273	0.242	1.016	-0.016	1.172	-0.376	-1.305
GATS2p	0.894	1.147	-0.453	-1.126	-1.126	1.147	0.579	1.147	-1.126	0.894	-1.126	-1.126	0.768	-0.191	0.823	-1.126
EPS0	-0.641	-0.729	1.854	-0.505	0.899	-0.729	0.504	-0.216	0.810	-0.641	-2.180	-0.729	-0.009	0.923	0.165	1.223
EPS1	-0.486	-0.954	1.752	-0.019	0.916	-0.954	0.385	-0.567	0.368	-0.486	-1.889	-0.954	-0.118	1.332	0.323	1.351
EEig01x	0.295	0.295	0.797	0.917	-0.145	-1.207	0.433	-0.767	-0.430	-0.145	-2.269	-1.207	0.036	1.648	1.008	0.743
EEig02x	-0.573	-0.573	1.585	-1.115	0.736	-1.115	0.838	-0.190	0.736	-1.115	-1.115	-1.115	0.098	1.178	0.207	1.534
EEig01d	0.694	1.149	0.013	0.996	-0.675	-1.114	0.810	-0.759	-0.920	-0.359	-1.716	-0.492	-0.194	2.184	0.533	-0.152

EEig02d	-0.221	0.371	1.045	-1.106	0.273	-1.683	1.233	0.108	0.273	-1.479	-0.906	-0.670	0.337	1.890	-0.322	0.859
EEig03d	-0.556	-0.113	1.271	-0.348	1.271	-0.113	0.594	-1.564	-0.113	-1.497	-0.113	-0.113	-0.916	1.793	-0.754	1.271
EEig04d	0.429	0.429	2.093	0.124	-1.365	0.429	0.140	0.429	-1.365	0.429	0.429	0.429	-1.871	0.413	-1.365	0.192
EEig01r	0.297	-0.933	0.760	0.659	0.220	-0.983	0.416	-0.530	-0.075	0.149	-2.197	-1.538	0.332	1.383	1.364	0.675
EEig02r	-0.730	-0.889	1.343	-1.372	0.916	-0.937	0.863	-0.140	0.916	-0.844	-0.844	-1.372	0.172	1.238	0.403	1.280
EEig03r	-0.518	-0.518	1.664	-1.173	1.664	-0.518	-0.072	-1.083	0.573	-0.667	-0.518	-0.518	-0.627	1.166	-0.518	1.664
ESpm02u	-0.083	-1.088	1.102	0.652	0.652	-1.088	0.454	-0.483	0.215	-0.083	-2.393	-1.088	0.215	1.222	0.822	0.970
ESpm03u	0.471	-1.025	0.471	1.450	-1.025	-1.025	0.471	-1.025	-1.025	0.471	-1.025	-1.025	0.471	1.450	1.450	0.471
ESpm04u	-0.005	-1.265	0.895	1.018	0.450	-1.265	0.412	-0.515	0.063	-0.005	-2.015	-1.265	0.283	1.302	1.122	0.791
ESpm05u	0.422	-1.046	0.711	1.298	-1.046	-1.046	0.541	-1.046	-1.046	0.422	-1.046	-1.046	0.541	1.407	1.348	0.634
ESpm06u	0.061	-1.331	0.793	1.133	0.368	-1.331	0.401	-0.554	-0.011	0.061	-1.823	-1.331	0.311	1.332	1.215	0.703
ESpm07u	0.400	-1.051	0.788	1.252	-1.051	-1.051	0.571	-1.051	-1.051	0.400	-1.051	-1.051	0.552	1.383	1.310	0.705
ESpm08u	0.103	-1.364	0.751	1.166	0.330	-1.364	0.402	-0.575	-0.046	0.103	-1.725	-1.364	0.330	1.343	1.241	0.668
ESpm09u	0.390	-1.053	0.817	1.235	-1.053	-1.053	0.584	-1.053	-1.053	0.390	-1.053	-1.053	0.553	1.370	1.294	0.736
ESpm10u	0.127	-1.383	0.735	1.176	0.306	-1.383	0.406	-0.587	-0.065	0.127	-1.668	-1.383	0.343	1.347	1.249	0.653
ESpm11u	0.386	-1.054	0.830	1.228	-1.054	-1.054	0.589	-1.054	-1.054	0.386	-1.054	-1.054	0.553	1.364	1.287	0.752
ESpm12u	0.142	-1.396	0.728	1.179	0.290	-1.396	0.410	-0.595	-0.077	0.142	-1.630	-1.396	0.353	1.347	1.252	0.646
ESpm13u	0.384	-1.054	0.836	1.225	-1.054	-1.054	0.591	-1.054	-1.054	0.384	-1.054	-1.054	0.552	1.361	1.284	0.760
ESpm14u	0.153	-1.405	0.726	1.180	0.278	-1.405	0.414	-0.599	-0.086	0.153	-1.604	-1.405	0.360	1.346	1.252	0.642
ESpm15u	0.383	-1.054	0.839	1.223	-1.054	-1.054	0.592	-1.054	-1.054	0.383	-1.054	-1.054	0.552	1.359	1.282	0.765
ESpm01x	-0.130	-0.130	1.668	-0.130	0.561	-1.177	0.561	-0.588	0.245	-0.588	-2.011	-1.177	-0.130	1.294	0.245	1.489
ESpm02x	0.015	0.015	1.286	0.364	0.508	-1.230	0.508	-0.618	0.111	-0.327	-2.423	-1.230	0.015	1.266	0.576	1.167
ESpm03x	0.207	0.207	1.070	0.633	0.251	-1.196	0.478	-0.663	-0.082	-0.146	-2.587	-1.196	0.060	1.230	0.728	1.005
ESpm04x	0.285	0.285	0.933	0.723	0.163	-1.170	0.471	-0.646	-0.145	-0.073	-2.662	-1.170	0.099	1.228	0.793	0.886
ESpm05x	0.326	0.326	0.856	0.765	0.095	-1.145	0.469	-0.626	-0.184	-0.033	-2.704	-1.145	0.127	1.231	0.827	0.815
ESpm06x	0.346	0.346	0.809	0.784	0.064	-1.127	0.469	-0.609	-0.203	-0.012	-2.732	-1.127	0.143	1.234	0.844	0.771
ESpm07x	0.357	0.357	0.780	0.793	0.044	-1.113	0.470	-0.596	-0.214	0.001	-2.753	-1.113	0.154	1.235	0.852	0.743
ESpm08x	0.364	0.364	0.762	0.798	0.035	-1.102	0.472	-0.587	-0.221	0.009	-2.768	-1.102	0.160	1.236	0.857	0.724
ESpm09x	0.368	0.368	0.750	0.800	0.029	-1.094	0.472	-0.580	-0.224	0.014	-2.779	-1.094	0.165	1.235	0.858	0.713
ESpm10x	0.370	0.370	0.742	0.801	0.026	-1.088	0.473	-0.575	-0.226	0.017	-2.788	-1.088	0.168	1.234	0.859	0.705
ESpm11x	0.371	0.371	0.736	0.802	0.025	-1.083	0.473	-0.571	-0.227	0.020	-2.796	-1.083	0.169	1.233	0.860	0.699
ESpm12x	0.372	0.372	0.732	0.802	0.025	-1.078	0.474	-0.568	-0.227	0.021	-2.802	-1.078	0.171	1.232	0.859	0.696
ESpm13x	0.372	0.372	0.729	0.801	0.024	-1.075	0.474	-0.566	-0.227	0.022	-2.807	-1.075	0.172	1.231	0.859	0.693
ESpm14x	0.373	0.373	0.727	0.801	0.025	-1.072	0.474	-0.564	-0.227	0.023	-2.811	-1.072	0.172	1.229	0.858	0.691
ESpm15x	0.373	0.373	0.726	0.800	0.025	-1.070	0.474	-0.562	-0.227	0.024	-2.815	-1.070	0.173	1.228	0.858	0.690
ESpm01d	0.960	1.366	-0.079	0.800	-1.622	-0.594	1.288	0.035	-1.622	-0.594	-0.594	0.404	0.035	1.576	-0.594	-0.763
ESpm02d	0.310	0.688	0.831	0.498	0.242	-1.422	0.728	-0.688	-0.250	-0.451	-2.444	-0.817	-0.047	1.688	0.497	0.636
ESpm03d	0.789	1.059	0.146	0.962	-1.714	-0.762	0.922	-0.387	-1.714	-0.028	-1.389	-0.090	0.120	1.587	0.606	-0.105

ESpm04d	0.679	1.027	0.328	0.913	-0.237	-1.369	0.814	-0.783	-0.606	-0.325	-2.305	-0.507	-0.098	1.710	0.576	0.182
ESpm05d	0.774	1.035	0.244	0.951	-1.679	-0.820	0.855	-0.410	-1.679	-0.009	-1.528	-0.127	0.132	1.530	0.665	0.066
ESpm06d	0.736	1.070	0.215	0.963	-0.363	-1.325	0.832	-0.764	-0.717	-0.263	-2.267	-0.426	-0.070	1.701	0.610	0.068
ESpm07d	0.768	1.026	0.278	0.944	-1.663	-0.839	0.839	-0.409	-1.663	-0.008	-1.580	-0.132	0.135	1.512	0.667	0.125
ESpm08d	0.749	1.078	0.186	0.972	-0.415	-1.301	0.838	-0.743	-0.767	-0.237	-2.262	-0.396	-0.052	1.694	0.622	0.033
ESpm09d	0.765	1.022	0.291	0.940	-1.656	-0.843	0.834	-0.407	-1.656	-0.008	-1.607	-0.132	0.136	1.505	0.665	0.149
ESpm10d	0.754	1.079	0.178	0.975	-0.445	-1.285	0.841	-0.729	-0.796	-0.225	-2.262	-0.382	-0.042	1.690	0.627	0.022
ESpm11d	0.764	1.020	0.297	0.938	-1.651	-0.844	0.832	-0.406	-1.651	-0.007	-1.621	-0.131	0.136	1.501	0.664	0.160
ESpm12d	0.756	1.080	0.177	0.976	-0.465	-1.275	0.842	-0.720	-0.815	-0.218	-2.262	-0.374	-0.036	1.687	0.630	0.017
ESpm13d	0.763	1.019	0.299	0.937	-1.649	-0.844	0.831	-0.405	-1.649	-0.007	-1.629	-0.131	0.136	1.499	0.664	0.165
ESpm14d	0.757	1.080	0.177	0.976	-0.480	-1.267	0.843	-0.714	-0.829	-0.213	-2.261	-0.369	-0.032	1.685	0.632	0.016
ESpm15d	0.763	1.019	0.300	0.936	-1.647	-0.843	0.831	-0.404	-1.647	-0.007	-1.635	-0.130	0.136	1.498	0.663	0.168
ESpm01r	-0.043	-0.771	1.412	-0.985	0.961	-0.840	0.592	-0.346	0.659	-0.240	-1.709	-1.709	0.136	1.070	0.592	1.223
ESpm02r	-0.004	-0.967	1.159	0.328	0.739	-1.012	0.504	-0.453	0.360	-0.104	-2.305	-1.428	0.197	1.175	0.781	1.030
ESpm03r	0.229	-0.878	0.945	0.475	0.546	-0.934	0.482	-0.446	0.231	0.117	-2.380	-1.639	0.294	1.126	0.968	0.865
ESpm04r	0.297	-0.862	0.851	0.565	0.460	-0.922	0.476	-0.420	0.169	0.181	-2.444	-1.612	0.337	1.115	1.027	0.784
ESpm05r	0.338	-0.836	0.792	0.605	0.398	-0.897	0.471	-0.392	0.135	0.221	-2.462	-1.643	0.369	1.108	1.061	0.731
ESpm06r	0.357	-0.823	0.762	0.624	0.368	-0.884	0.471	-0.376	0.115	0.240	-2.480	-1.643	0.385	1.106	1.077	0.702
ESpm07r	0.367	-0.812	0.744	0.634	0.349	-0.874	0.470	-0.365	0.104	0.251	-2.489	-1.649	0.395	1.105	1.085	0.684
ESpm08r	0.373	-0.805	0.733	0.639	0.338	-0.867	0.471	-0.358	0.096	0.257	-2.497	-1.649	0.401	1.105	1.089	0.674
ESpm09r	0.377	-0.800	0.727	0.642	0.331	-0.862	0.471	-0.353	0.091	0.261	-2.502	-1.650	0.404	1.105	1.091	0.668
ESpm10r	0.379	-0.797	0.723	0.644	0.328	-0.859	0.472	-0.351	0.087	0.263	-2.505	-1.650	0.406	1.104	1.092	0.664
ESpm11r	0.380	-0.795	0.720	0.645	0.326	-0.857	0.472	-0.348	0.085	0.264	-2.508	-1.649	0.407	1.104	1.093	0.661
ESpm12r	0.381	-0.793	0.718	0.645	0.324	-0.855	0.472	-0.347	0.084	0.266	-2.510	-1.649	0.408	1.104	1.093	0.660
ESpm13r	0.381	-0.792	0.717	0.646	0.324	-0.854	0.472	-0.346	0.083	0.266	-2.512	-1.649	0.408	1.104	1.093	0.659
ESpm14r	0.381	-0.791	0.717	0.646	0.323	-0.853	0.472	-0.346	0.082	0.266	-2.513	-1.648	0.409	1.104	1.093	0.658
ESpm15r	0.382	-0.791	0.716	0.646	0.323	-0.853	0.472	-0.345	0.082	0.267	-2.514	-1.648	0.409	1.104	1.093	0.658
BEHm1	-0.436	-0.753	0.550	1.548	0.072	-0.924	-0.292	-0.698	-0.251	-0.522	-1.556	0.652	-0.387	2.550	-0.078	0.524
BEHm2	-0.610	-1.193	0.873	0.760	0.310	-1.051	0.802	-0.444	0.979	-0.610	-2.521	0.760	-0.055	1.036	0.374	0.588
BEHm3	-0.024	-2.111	0.805	1.189	0.955	-0.776	0.262	-0.473	0.674	0.152	-2.111	-0.661	0.216	0.621	0.476	0.805
BEHm4	-1.248	-1.248	1.133	1.799	0.209	-1.248	0.633	-0.158	0.697	-0.331	-1.248	-1.248	0.190	0.366	0.839	0.865
BEHm5	-0.861	-0.988	1.101	-0.988	1.225	-0.861	0.508	-0.861	1.190	-0.861	-0.988	-0.988	0.363	1.037	1.530	0.442
BEHm6	-0.625	-0.803	1.214	-0.803	1.461	-0.625	-0.625	-0.625	1.654	-0.625	-0.803	-0.803	-0.625	1.746	0.070	0.821
BEHm7	-0.131	-0.575	3.424	-0.575	-0.131	-0.575	-0.131	-0.575	-0.131	-0.131	-0.575	-0.575	-0.131	-0.131	-0.131	1.074
BEHm8	-0.791	-0.791	3.148	-0.791	0.299	-0.791	0.299	-0.791	0.299	0.299	-0.791	-0.791	0.299	0.299	0.299	0.299
BELm1	0.167	-0.169	0.819	-3.223	0.848	0.089	0.143	-0.013	0.649	0.314	-0.410	-1.136	0.254	0.262	0.567	0.838
BELm2	0.387	-0.997	0.589	-1.948	0.617	-0.247	0.864	-0.286	1.129	0.387	-1.161	-1.948	0.312	0.806	0.806	0.690
BELm3	-0.936	-1.252	0.901	-1.252	1.478	-0.476	0.491	-0.434	1.331	0.131	-1.252	-1.252	0.133	0.256	1.233	0.901

BELm4	-1.075	-1.075	1.220	-1.075	0.780	-1.075	0.056	-0.047	1.543	-0.248	-1.075	-1.075	0.045	0.585	1.279	1.240
BEHv1	-0.101	-0.552	1.465	-1.000	0.917	-0.643	-0.204	-0.549	0.504	-0.073	-1.749	-1.448	-0.023	1.447	0.551	1.459
BEHv2	0.148	-0.444	0.897	-1.861	0.692	-0.448	0.929	-0.312	1.121	0.148	-1.525	-1.861	0.235	0.705	0.726	0.851
BEHv3	-0.122	-1.651	1.080	-0.836	1.255	-0.658	0.379	-0.562	0.992	0.067	-1.651	-1.231	0.119	0.928	0.810	1.080
BEHv4	-1.226	-1.226	1.441	-0.251	0.543	-1.226	0.783	-0.130	1.153	-0.272	-1.226	-1.226	0.041	0.502	0.987	1.336
BEHv5	-0.619	-1.101	0.849	-1.101	1.439	-0.619	-0.358	-0.619	1.412	-0.619	-1.101	-1.101	0.436	1.043	1.427	0.629
BEHv6	-0.363	-1.051	1.417	-1.051	1.729	-0.363	-0.363	-0.363	1.920	-0.363	-1.051	-1.051	-0.363	0.128	0.163	1.029
BEHv7	0.269	-0.984	2.767	-0.984	0.269	-0.984	0.269	-0.984	0.269	0.269	-0.984	-0.984	0.269	0.269	0.269	0.980
BEHv8	-1.067	-1.067	1.694	-1.067	0.722	-1.067	0.722	-1.067	0.722	0.722	-1.067	-1.067	0.722	0.722	0.722	0.722
BELv1	0.042	-0.541	1.070	-2.535	0.965	-0.151	0.158	-0.058	0.610	0.216	-0.820	-1.693	0.270	0.780	0.629	1.058
BELv2	0.190	-0.764	0.686	-2.007	0.495	-0.259	0.839	-0.083	1.046	0.190	-1.132	-2.007	0.310	1.229	0.692	0.577
BELv3	-0.442	-1.368	0.832	-1.368	1.441	-0.604	0.485	-0.418	1.236	0.342	-1.368	-1.368	0.412	0.255	1.099	0.832
BELv4	-1.064	-1.064	1.311	-1.064	0.465	-1.064	0.639	-0.185	1.362	-0.585	-1.064	-1.064	0.154	0.773	1.320	1.127
BEHe1	-0.089	-0.670	1.298	-1.381	0.852	-0.557	0.055	-0.345	0.414	-0.043	-1.434	-1.675	0.081	1.687	0.535	1.272
BEHe2	0.137	-0.441	0.833	-2.123	0.575	-0.242	0.884	0.010	1.041	0.137	-1.113	-2.123	0.337	0.717	0.679	0.690
BEHe3	-0.008	-1.649	0.898	-1.198	1.189	-0.498	0.459	-0.309	0.959	0.311	-1.649	-1.425	0.378	0.845	0.801	0.898
BEHe4	-1.216	-1.216	1.382	-0.676	0.482	-1.216	0.825	0.020	1.102	-0.219	-1.216	-1.216	0.318	0.533	1.137	1.176
BEHe5	-0.235	-1.500	0.950	-1.500	1.133	-0.235	0.148	-0.235	1.133	-0.235	-1.197	-1.500	0.525	0.885	1.327	0.536
BEHe6	0.082	-1.488	1.039	-1.488	1.323	0.082	0.082	0.082	1.430	0.082	-1.488	-1.488	0.082	0.724	0.117	0.829
BEHe7	0.615	-1.410	1.573	-1.410	0.615	-0.795	0.615	-0.702	0.615	0.615	-1.410	-1.410	0.615	0.615	0.615	0.647
BEHe8	-0.965	-1.195	0.874	-1.195	0.844	-1.195	0.844	-0.684	0.844	0.844	-1.195	-1.195	0.844	0.844	0.844	0.844
BELe1	0.005	-0.493	1.411	-1.912	1.066	-0.357	-0.077	-0.346	0.677	0.124	-1.319	-1.626	0.128	0.650	0.660	1.411
BELe2	0.073	-0.938	0.757	-1.491	0.603	-0.696	0.926	-0.609	1.208	0.073	-1.491	-1.491	0.110	1.493	0.733	0.741
BELe3	-0.867	-0.867	1.224	-0.867	1.799	-0.867	0.138	-0.867	1.394	-0.398	-0.867	-0.867	-0.336	-0.103	1.124	1.224
BEHp1	-0.177	-0.648	1.358	-0.707	0.841	-0.685	-0.294	-0.614	0.441	-0.134	-1.768	-1.242	-0.097	1.903	0.472	1.352
BEHp2	0.128	-0.553	0.894	-1.715	0.692	-0.523	0.941	-0.402	1.151	0.128	-1.675	-1.715	0.201	0.859	0.734	0.856
BEHp3	-0.185	-1.684	1.083	-0.643	1.273	-0.703	0.357	-0.617	1.009	0.028	-1.684	-1.154	0.076	0.937	0.824	1.083
BEHp4	-1.241	-1.241	1.419	-0.006	0.537	-1.241	0.752	-0.161	1.157	-0.300	-1.241	-1.241	-0.007	0.506	0.977	1.331
BEHp5	-0.540	-1.148	0.825	-1.148	1.433	-0.540	-0.451	-0.540	1.413	-0.540	-1.148	-1.148	0.455	1.056	1.376	0.644
BEHp6	-0.269	-1.121	1.393	-1.121	1.710	-0.269	-0.269	-0.269	1.891	-0.269	-1.121	-1.121	-0.269	-0.103	0.178	1.028
BEHp7	0.367	-1.062	2.511	-1.062	0.367	-1.062	0.367	-1.062	0.367	0.367	-1.062	-1.062	0.367	0.367	0.367	0.924
BEHp8	-1.082	-1.082	1.455	-1.082	0.765	-1.082	0.765	-1.082	0.765	0.765	-1.082	-1.082	0.765	0.765	0.765	0.765
BELp1	0.112	-0.418	1.090	-2.655	0.949	-0.097	0.249	0.025	0.606	0.256	-0.718	-1.706	0.328	0.253	0.653	1.072
BELp2	0.149	-0.724	0.720	-1.990	0.486	-0.278	0.850	-0.059	1.051	0.149	-1.181	-1.990	0.311	1.249	0.682	0.576
BELp3	-0.393	-1.349	0.850	-1.349	1.465	-0.684	0.488	-0.461	1.235	0.361	-1.349	-1.349	0.441	0.163	1.083	0.850
BELp4	-1.023	-1.023	1.381	-1.023	0.382	-1.023	0.683	-0.293	1.352	-0.731	-1.023	-1.023	0.113	0.718	1.384	1.147
LP1	-0.136	-1.080	1.064	0.660	0.660	-1.080	0.369	-0.474	0.072	-0.136	-2.310	-1.080	0.209	1.421	0.880	0.960
Eig1Z	-0.587	-1.112	1.698	-0.813	0.819	-0.933	0.731	-0.385	1.655	-0.437	-1.364	-1.194	0.208	0.052	0.898	0.764

Eig1m	-0.584	-1.109	1.697	-0.831	0.819	-0.929	0.732	-0.383	1.654	-0.434	-1.360	-1.197	0.210	0.053	0.898	0.764
Eig1v	-0.802	-1.358	1.529	-0.297	0.241	-0.958	1.847	-0.073	0.941	-0.466	-1.438	-1.169	0.527	0.549	0.732	0.196
Eig1e	-0.704	-1.241	1.634	-0.254	0.733	-1.057	0.649	-0.496	1.589	-0.550	-1.499	-1.103	0.111	0.693	0.816	0.677
Eig1p	-0.746	-1.290	1.535	-0.527	0.212	-0.866	2.051	0.050	0.871	-0.388	-1.339	-1.195	0.636	0.068	0.760	0.169
SEigZ	-0.390	-0.551	-0.390	3.126	-0.766	-0.390	-0.014	-0.014	-0.766	-0.390	-0.390	1.180	-0.014	0.926	-0.390	-0.766
SEigm	-0.390	-0.546	-0.390	3.138	-0.757	-0.390	-0.022	-0.022	-0.757	-0.390	-0.390	1.190	-0.022	0.898	-0.390	-0.757
SEigv	-0.131	0.567	-0.131	1.162	1.162	-0.131	-1.425	-1.425	1.162	-0.131	-0.131	1.162	-1.425	-1.315	-0.131	1.162
SEige	-0.263	-0.735	-0.263	2.303	-1.333	-0.263	0.803	0.803	-1.333	-0.263	-0.263	0.487	0.803	1.111	-0.263	-1.333
SEigp	-0.207	0.386	-0.207	1.742	0.979	-0.207	-1.393	-1.393	0.979	-0.207	-0.207	1.360	-1.393	-1.005	-0.207	0.979
AEigZ	-0.542	-1.035	1.662	-1.063	0.846	-0.875	0.697	-0.380	1.653	-0.397	-1.292	-1.263	0.193	0.112	0.891	0.794
AEigm	-0.538	-1.030	1.660	-1.086	0.846	-0.870	0.698	-0.376	1.651	-0.394	-1.286	-1.268	0.195	0.114	0.890	0.793
AEigv	-0.769	-1.366	1.497	-0.383	0.141	-0.921	1.910	0.044	0.821	-0.443	-1.387	-1.229	0.627	0.640	0.722	0.096
AEige	-0.691	-1.210	1.632	-0.325	0.770	-1.042	0.619	-0.518	1.620	-0.538	-1.481	-1.111	0.086	0.654	0.819	0.715
AEigp	-0.693	-1.275	1.491	-0.684	0.103	-0.809	2.108	0.190	0.734	-0.350	-1.261	-1.284	0.752	0.169	0.749	0.061
VEA1	-0.490	-1.092	1.615	0.016	0.894	-1.092	0.530	-0.452	0.608	-0.490	-1.877	-1.092	0.043	1.141	0.455	1.283
VEA2	0.232	1.086	-1.189	-0.354	-0.513	1.086	-0.742	0.271	-0.692	0.232	2.456	1.086	-0.335	-0.960	-0.781	-0.881
VRA1	-0.587	-1.051	1.863	-0.117	0.876	-1.051	0.391	-0.575	0.401	-0.587	-1.504	-1.051	-0.101	1.351	0.374	1.366
VRA2	-0.404	-0.973	1.403	-0.047	1.313	-0.973	0.270	-0.371	0.289	-0.404	-2.042	-0.973	-0.009	1.332	0.231	1.358
VED1	-0.482	-1.126	1.595	0.102	0.719	-1.126	0.625	-0.477	0.625	-0.482	-1.887	-1.126	0.097	1.149	0.623	1.172
VED2	0.252	1.067	-1.224	-0.285	-0.626	1.067	-0.687	0.262	-0.687	0.252	2.460	1.067	-0.295	-0.976	-0.687	-0.956
VRD1	-0.568	-1.063	1.847	-0.069	0.835	-1.063	0.401	-0.582	0.392	-0.568	-1.536	-1.063	-0.085	1.361	0.415	1.345
VRD2	-0.331	-0.989	1.357	0.073	1.220	-0.989	0.275	-0.376	0.256	-0.331	-2.155	-0.989	0.028	1.337	0.308	1.305
VEZ1	-0.467	-1.102	1.585	0.060	0.738	-1.107	0.628	-0.455	0.645	-0.460	-1.863	-1.245	0.120	1.125	0.640	1.158
VEZ2	0.270	1.096	-1.236	-0.316	-0.609	1.086	-0.682	0.280	-0.671	0.280	2.498	0.898	-0.274	-0.985	-0.671	-0.964
VRZ1	-0.572	-1.069	1.853	-0.051	0.835	-1.067	0.400	-0.586	0.390	-0.571	-1.543	-1.041	-0.088	1.343	0.414	1.353
VRZ2	-0.346	-1.005	1.367	0.115	1.222	-0.999	0.273	-0.393	0.247	-0.346	-2.185	-0.887	0.016	1.301	0.299	1.321
VEm1	-0.466	-1.101	1.585	0.059	0.738	-1.106	0.628	-0.454	0.646	-0.459	-1.860	-1.253	0.121	1.123	0.641	1.158
VEm2	0.271	1.098	-1.236	-0.325	-0.608	1.088	-0.681	0.282	-0.670	0.282	2.501	0.889	-0.273	-0.984	-0.670	-0.963
VRm1	-0.573	-1.069	1.853	-0.050	0.835	-1.067	0.400	-0.586	0.390	-0.572	-1.543	-1.040	-0.088	1.343	0.414	1.353
VRm2	-0.347	-1.007	1.367	0.121	1.222	-1.000	0.272	-0.393	0.246	-0.347	-2.186	-0.881	0.015	1.301	0.299	1.320
VEv1	-0.477	-1.134	1.526	0.106	0.727	-1.114	0.661	-0.465	0.634	-0.470	-1.923	-1.131	0.098	1.179	0.628	1.153
VEv2	0.265	1.072	-1.267	-0.284	-0.626	1.093	-0.667	0.275	-0.688	0.265	2.418	1.072	-0.294	-0.967	-0.688	-0.978
VRv1	-0.570	-1.064	1.855	-0.067	0.834	-1.065	0.392	-0.582	0.393	-0.570	-1.530	-1.060	-0.084	1.352	0.416	1.350
VRv2	-0.339	-1.001	1.376	0.077	1.220	-1.001	0.259	-0.384	0.259	-0.339	-2.131	-0.982	0.032	1.324	0.311	1.318
VEe1	-0.486	-1.124	1.578	0.100	0.726	-1.129	0.616	-0.473	0.633	-0.478	-1.886	-1.137	0.105	1.179	0.628	1.149
VEe2	0.252	1.072	-1.243	-0.288	-0.620	1.061	-0.693	0.262	-0.682	0.262	2.462	1.051	-0.288	-0.952	-0.682	-0.973
VRe1	-0.569	-1.064	1.852	-0.066	0.835	-1.062	0.401	-0.583	0.392	-0.568	-1.537	-1.060	-0.085	1.348	0.415	1.352
VRe2	-0.335	-0.989	1.365	0.077	1.222	-0.983	0.280	-0.381	0.253	-0.335	-2.160	-0.976	0.025	1.313	0.306	1.320

VEp1	-0.472	-1.131	1.521	0.101	0.730	-1.109	0.666	-0.462	0.636	-0.465	-1.925	-1.147	0.101	1.173	0.628	1.155
VEp2	0.267	1.075	-1.277	-0.293	-0.624	1.106	-0.666	0.277	-0.687	0.277	2.412	1.054	-0.293	-0.966	-0.687	-0.977
VRp1	-0.571	-1.065	1.857	-0.064	0.835	-1.066	0.391	-0.583	0.393	-0.571	-1.530	-1.056	-0.085	1.347	0.416	1.351
VRp2	-0.346	-1.004	1.380	0.084	1.224	-1.010	0.260	-0.385	0.260	-0.346	-2.130	-0.965	0.032	1.315	0.312	1.321

Table B.2: Eigenvector matrix from 2D descriptors

Descriptor	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
ZM1	0.067	0.009	-0.022	0.050	-0.004	-0.022	0.013	0.001	-0.008	0.012	-0.027	0.022	0.027	-0.023	0.021
ZM1V	0.042	0.023	0.122	0.081	0.015	-0.011	-0.003	0.116	0.009	-0.039	0.034	-0.107	0.015	-0.074	-0.025
ZM2	0.066	0.000	-0.018	0.071	-0.007	-0.016	0.027	0.001	-0.008	0.016	-0.035	0.027	0.027	-0.020	0.020
ZM2V	0.050	-0.009	0.074	0.053	0.009	0.039	-0.174	0.047	0.001	0.042	0.165	-0.050	0.054	-0.185	0.009
Qindex	0.054	0.035	-0.040	0.102	-0.048	-0.067	0.052	-0.015	-0.015	0.042	-0.064	0.048	0.053	-0.040	0.058
SNar	0.064	-0.034	-0.022	0.064	-0.022	0.032	0.016	0.032	-0.024	-0.011	0.006	0.002	0.025	0.006	-0.001
HNar	0.055	-0.057	-0.039	0.051	-0.066	0.058	0.041	0.076	-0.041	-0.018	0.065	0.022	0.049	0.010	-0.007
GNar	0.061	-0.039	-0.032	0.041	-0.059	0.051	0.044	0.068	-0.042	-0.006	0.042	0.000	0.038	0.011	-0.002
Xt	-0.025	0.067	0.022	-0.104	-0.067	0.174	0.018	-0.061	-0.125	0.026	0.004	0.078	0.078	-0.035	-0.009
Dz	0.068	0.009	0.009	0.022	0.040	-0.013	0.011	0.048	-0.026	-0.013	0.001	-0.004	0.005	-0.002	-0.017
Ram	0.039	0.099	-0.001	0.013	0.017	-0.129	-0.022	-0.071	0.005	0.072	-0.046	-0.017	0.001	-0.050	0.026
Pol	0.062	-0.043	0.016	0.073	0.037	0.023	-0.017	-0.003	0.003	0.039	-0.004	0.066	-0.001	-0.039	-0.019
LPRS	0.068	-0.014	-0.006	0.025	0.033	0.012	-0.017	0.003	0.002	-0.017	-0.010	0.007	0.010	-0.010	-0.003
VDA	0.067	-0.022	-0.001	0.019	0.052	0.022	-0.030	-0.005	0.010	-0.036	-0.013	0.001	0.003	-0.002	-0.009
MSD	-0.065	-0.037	0.029	0.006	0.050	0.028	-0.003	-0.014	0.054	-0.028	-0.043	0.001	-0.003	0.013	0.009
SMTI	0.064	-0.026	-0.007	0.077	0.026	0.025	-0.029	-0.017	-0.018	-0.033	-0.028	0.028	0.001	-0.006	0.001
SMTIV	0.058	-0.008	0.058	0.100	0.040	0.024	-0.030	0.030	-0.039	-0.039	0.007	-0.008	-0.022	-0.019	-0.024
GMTI	0.060	-0.036	-0.010	0.098	0.012	0.037	-0.032	-0.013	-0.033	-0.036	-0.026	0.030	-0.004	0.001	-0.003
GMTIV	0.053	-0.012	0.051	0.120	0.047	0.041	-0.042	0.032	-0.085	-0.042	0.000	0.031	-0.074	0.026	-0.051
Xu	0.069	-0.010	-0.006	-0.002	0.028	0.018	-0.011	0.018	-0.003	-0.014	0.001	-0.005	0.004	-0.004	-0.007
SPI	0.045	0.063	0.051	-0.080	0.114	-0.009	-0.039	-0.047	0.037	0.047	-0.115	-0.060	-0.040	-0.034	0.024
W	0.065	-0.025	0.000	0.056	0.050	0.025	-0.039	-0.025	-0.003	-0.042	-0.027	0.026	-0.001	-0.010	-0.003
WA	0.063	-0.026	0.011	-0.050	0.072	0.051	-0.028	0.024	0.018	-0.049	-0.007	-0.036	-0.020	0.019	-0.020
Har	0.068	-0.007	-0.016	0.050	0.003	-0.001	0.003	0.006	-0.010	0.002	-0.014	0.016	0.021	-0.016	0.009
Har2	0.068	-0.008	-0.013	0.053	0.012	-0.001	-0.003	-0.001	-0.007	-0.001	-0.018	0.022	0.019	-0.019	0.009
QW	0.064	-0.017	0.012	0.016	0.088	0.016	-0.041	-0.037	0.028	-0.059	-0.029	0.020	-0.001	-0.011	-0.005
TI1	-0.003	-0.021	-0.036	0.195	-0.135	0.024	0.042	0.015	-0.116	0.033	-0.040	0.042	-0.010	0.021	0.022
TI2	0.008	-0.041	0.076	-0.110	0.171	0.110	0.000	0.012	0.085	-0.069	-0.076	-0.020	-0.028	0.025	-0.015
HyDp	0.062	-0.034	0.005	0.054	0.067	0.039	-0.056	-0.035	-0.002	-0.069	-0.032	0.024	-0.016	0.003	-0.012
RHyDp	0.068	-0.007	-0.015	0.051	0.005	-0.002	0.002	0.004	-0.009	0.002	-0.016	0.018	0.021	-0.017	0.009
w	0.059	-0.033	-0.016	0.110	-0.010	0.034	-0.028	-0.005	-0.046	-0.014	-0.025	0.031	-0.001	-0.005	0.001
ww	0.054	-0.038	-0.017	0.125	-0.016	0.045	-0.045	-0.007	-0.066	-0.018	-0.024	0.041	-0.019	0.000	-0.006
Rww	0.036	0.047	0.036	-0.133	0.135	-0.044	-0.039	-0.039	0.074	-0.025	-0.013	0.011	-0.025	-0.030	-0.012
D/D	0.063	0.006	0.016	-0.039	0.101	-0.015	-0.040	-0.032	0.050	-0.032	-0.007	0.026	0.008	-0.035	-0.004
Wap	0.056	-0.033	-0.017	0.126	-0.021	0.037	-0.025	-0.004	-0.061	-0.012	-0.032	0.035	-0.011	0.001	0.001
WhetZ	0.062	-0.055	0.006	0.005	0.045	0.018	-0.053	-0.044	0.010	-0.041	0.029	0.076	-0.033	-0.019	-0.026

Whetm	0.062	-0.055	0.006	0.005	0.044	0.018	-0.052	-0.044	0.010	-0.041	0.029	0.076	-0.033	-0.019	-0.026
Whetv	0.063	-0.018	0.029	0.014	0.090	-0.001	-0.010	0.019	0.006	-0.076	0.091	0.018	0.071	-0.016	0.019
Whete	0.065	-0.036	-0.009	0.020	0.051	0.010	-0.013	-0.038	0.014	-0.059	-0.018	0.064	-0.016	0.013	-0.019
Whetp	0.060	-0.025	0.042	0.002	0.095	0.002	-0.039	0.027	0.000	-0.068	0.140	0.020	0.077	-0.040	0.018
J	0.045	0.076	-0.005	-0.119	0.032	-0.072	-0.038	-0.012	0.003	0.033	0.042	0.032	-0.017	-0.047	-0.015
JhetZ	0.013	0.116	-0.099	0.004	0.071	0.012	-0.010	0.028	-0.004	0.008	-0.026	-0.011	0.043	0.015	-0.020
Jhetm	0.011	0.115	-0.102	0.003	0.072	0.013	-0.012	0.028	-0.007	0.008	-0.023	-0.008	0.042	0.015	-0.021
Jhetv	0.051	0.058	-0.055	-0.018	-0.053	0.045	-0.090	-0.077	0.102	0.062	-0.109	0.053	-0.108	-0.010	-0.073
Jhete	0.043	0.101	0.005	-0.039	0.024	-0.007	-0.112	0.029	0.069	0.027	0.034	-0.045	0.039	-0.075	-0.018
Jhetp	0.045	0.076	-0.085	0.004	-0.028	0.041	-0.028	-0.065	0.084	0.045	-0.141	0.035	-0.082	0.021	-0.056
MAXDN	0.009	0.118	0.026	0.040	0.074	-0.087	0.115	0.049	-0.084	-0.047	-0.094	-0.032	-0.003	0.034	-0.010
MAXDP	0.018	0.067	0.137	0.031	0.028	-0.078	0.080	-0.066	-0.051	-0.123	0.040	0.029	-0.083	-0.128	0.020
DELS	0.021	0.090	0.078	0.035	0.059	-0.070	0.173	0.085	-0.042	-0.051	-0.077	-0.042	-0.070	0.086	-0.079
TIE	0.058	0.025	0.029	-0.002	0.037	-0.087	0.143	-0.049	0.047	0.008	-0.072	0.093	-0.050	0.021	-0.026
S0K	0.053	0.011	0.089	0.043	0.090	0.014	-0.056	-0.086	0.016	0.057	-0.034	-0.029	0.034	-0.016	-0.065
S1K	0.054	0.037	-0.031	-0.089	0.112	-0.027	-0.013	-0.004	-0.003	-0.019	0.002	0.028	-0.006	0.007	-0.021
S2K	0.019	-0.062	-0.020	-0.114	0.144	0.135	0.047	0.058	0.013	-0.090	-0.077	-0.013	-0.043	0.077	-0.014
S3K	0.011	-0.040	0.072	-0.123	0.144	0.017	0.073	0.059	0.148	0.106	0.037	0.198	0.067	-0.128	-0.005
PHI	0.003	-0.036	-0.031	-0.141	0.172	0.121	0.058	0.032	0.007	-0.058	-0.078	-0.028	-0.070	0.058	-0.013
BLI	-0.016	0.024	-0.108	-0.027	0.073	0.048	0.241	-0.169	-0.053	0.018	0.002	-0.037	-0.004	-0.037	0.073
PW2	0.056	0.064	-0.011	-0.076	-0.046	0.011	0.002	0.015	-0.083	0.032	0.024	-0.008	-0.030	-0.010	-0.008
PW3	0.061	-0.048	0.018	0.025	0.029	-0.003	0.072	0.103	0.028	0.079	0.004	0.063	-0.028	0.001	-0.042
PJI2	-0.012	0.089	0.036	-0.070	0.025	0.152	-0.073	-0.132	-0.086	-0.145	0.001	-0.131	0.128	-0.028	0.083
CSI	0.063	-0.047	-0.008	0.045	0.030	0.042	-0.033	0.008	-0.003	-0.031	0.002	0.016	0.006	-0.003	-0.014
ECC	0.064	-0.042	0.001	0.014	0.055	0.042	-0.038	0.005	0.013	-0.037	-0.002	0.010	-0.001	-0.003	-0.017
AECC	0.059	-0.053	0.009	-0.039	0.064	0.070	-0.022	0.046	0.023	-0.031	0.008	0.000	-0.009	0.004	-0.022
DECC	0.030	0.029	0.089	-0.056	0.119	0.160	-0.015	-0.045	-0.027	-0.023	-0.154	-0.060	-0.028	-0.003	0.035
MDDD	0.051	0.012	0.057	-0.050	0.145	0.034	-0.026	-0.036	0.046	-0.045	-0.129	-0.069	-0.072	0.035	-0.007
UNIP	0.065	-0.044	-0.015	0.025	0.026	0.032	-0.020	0.030	0.012	-0.037	0.027	-0.012	0.036	-0.008	0.002
CENT	0.058	0.023	0.034	0.042	0.105	-0.005	-0.046	-0.096	-0.004	-0.041	-0.116	0.069	-0.077	-0.001	-0.016
VAR	0.056	0.005	0.047	0.017	0.120	0.025	-0.088	-0.079	-0.001	-0.016	-0.082	0.039	-0.075	-0.019	-0.026
BAC	0.007	0.073	0.023	-0.154	0.110	-0.102	-0.068	-0.043	0.070	0.030	0.011	0.060	-0.044	-0.060	-0.020
Lop	0.025	0.026	0.077	-0.095	0.125	0.154	-0.014	-0.052	0.015	-0.039	-0.154	-0.115	0.008	0.001	0.049
ICR	0.032	0.020	0.078	-0.054	0.128	0.158	-0.042	-0.065	-0.025	-0.070	-0.150	-0.064	-0.039	0.016	0.025
MWC01	0.068	-0.012	-0.017	0.045	-0.002	0.007	0.005	0.015	-0.014	-0.003	-0.005	0.006	0.020	-0.007	0.004
MWC02	0.068	0.021	-0.019	-0.020	-0.016	-0.001	0.004	0.024	-0.034	0.008	0.015	-0.008	0.000	-0.007	-0.001
MWC03	0.068	0.017	-0.016	-0.017	-0.017	0.003	0.011	0.032	-0.034	0.013	0.016	-0.004	-0.002	-0.005	-0.005
MWC04	0.067	0.027	-0.019	-0.024	-0.021	-0.003	0.006	0.024	-0.040	0.010	0.016	-0.008	-0.004	-0.008	0.000

MWC05	0.068	0.023	-0.017	-0.023	-0.021	0.000	0.011	0.030	-0.039	0.015	0.017	-0.004	-0.006	-0.006	-0.004
MWC06	0.067	0.030	-0.018	-0.028	-0.024	-0.003	0.007	0.025	-0.043	0.012	0.017	-0.007	-0.007	-0.008	-0.001
MWC07	0.067	0.026	-0.017	-0.026	-0.023	0.000	0.011	0.029	-0.042	0.015	0.017	-0.004	-0.008	-0.006	-0.003
MWC08	0.067	0.031	-0.018	-0.030	-0.025	-0.002	0.008	0.025	-0.045	0.013	0.018	-0.006	-0.008	-0.008	-0.001
MWC09	0.067	0.029	-0.017	-0.028	-0.025	0.000	0.012	0.028	-0.044	0.016	0.017	-0.004	-0.009	-0.007	-0.003
MWC10	0.066	0.032	-0.018	-0.031	-0.026	-0.002	0.008	0.025	-0.047	0.014	0.018	-0.005	-0.009	-0.008	-0.001
TWC	0.068	0.020	-0.017	-0.014	-0.017	0.000	0.007	0.024	-0.036	0.010	0.013	-0.003	-0.002	-0.008	-0.001
SRW01	0.069	-0.006	-0.010	0.019	0.020	0.004	-0.009	0.009	-0.004	-0.006	-0.004	0.006	0.011	-0.012	-0.002
SRW02	0.068	-0.012	-0.017	0.045	-0.002	0.007	0.005	0.015	-0.014	-0.003	-0.005	0.006	0.020	-0.007	0.004
SRW04	0.067	0.014	-0.023	0.052	-0.005	-0.029	0.015	-0.003	-0.007	0.015	-0.032	0.026	0.029	-0.027	0.025
SRW06	0.064	0.022	-0.024	0.068	-0.006	-0.046	0.027	-0.013	0.000	0.032	-0.053	0.058	0.036	-0.045	0.040
SRW08	0.061	0.023	-0.024	0.085	-0.007	-0.049	0.039	-0.022	0.004	0.040	-0.075	0.079	0.041	-0.057	0.056
SRW10	0.059	0.024	-0.021	0.098	-0.008	-0.047	0.053	-0.030	0.005	0.042	-0.099	0.088	0.041	-0.063	0.071
MPC01	0.068	-0.012	-0.017	0.045	-0.002	0.007	0.005	0.015	-0.014	-0.003	-0.005	0.006	0.020	-0.007	0.004
MPC02	0.066	0.023	-0.024	0.053	-0.006	-0.041	0.018	-0.009	-0.005	0.021	-0.041	0.032	0.031	-0.033	0.032
piPC01	0.066	0.007	0.004	0.037	-0.040	0.056	-0.026	0.025	0.055	-0.017	-0.009	-0.014	0.013	0.015	-0.011
piPC02	0.066	0.032	-0.001	0.035	-0.044	0.023	-0.032	0.007	0.033	-0.002	-0.019	-0.038	-0.003	-0.020	-0.010
piPC03	0.061	-0.041	0.007	0.082	0.008	0.019	0.015	0.038	0.014	0.083	-0.019	0.002	0.040	-0.045	0.015
TPC	0.066	-0.025	-0.016	0.065	-0.004	0.022	-0.004	0.014	-0.022	-0.007	-0.010	0.005	0.014	-0.003	0.001
piID	0.061	-0.025	-0.007	0.095	-0.007	0.045	-0.056	0.002	-0.031	0.020	-0.038	-0.006	-0.003	-0.021	0.008
PCR	0.031	-0.005	0.033	0.138	-0.060	0.129	-0.171	-0.016	0.082	0.052	-0.087	-0.018	-0.009	0.003	0.016
CID	0.068	-0.014	-0.014	0.034	0.008	0.011	-0.004	0.015	-0.010	-0.007	-0.002	0.006	0.015	-0.007	-0.001
BID	0.069	0.000	-0.013	0.024	0.012	0.000	-0.008	0.002	-0.011	-0.005	-0.006	0.008	0.012	-0.012	0.000
X0	0.068	0.014	-0.002	-0.010	0.044	-0.017	-0.023	-0.006	0.008	0.000	-0.011	0.007	0.002	-0.023	-0.001
X1	0.067	-0.027	-0.009	0.032	0.017	0.024	-0.004	0.021	-0.006	-0.014	0.002	0.005	0.010	-0.001	-0.009
X2	0.063	0.052	-0.031	0.006	-0.006	-0.053	-0.011	-0.019	-0.014	0.009	-0.021	0.005	0.030	-0.037	0.036
X3	0.062	-0.040	0.004	0.080	0.002	0.006	0.049	0.030	0.002	0.055	-0.022	0.067	-0.006	-0.008	-0.025
X0A	-0.061	0.040	0.025	-0.025	0.056	-0.072	-0.051	-0.074	0.045	0.006	-0.038	-0.001	-0.037	-0.010	0.000
X1A	-0.061	-0.044	0.020	0.055	0.054	-0.031	-0.009	-0.023	0.080	-0.009	-0.030	0.016	0.005	0.007	-0.003
X2A	-0.017	0.061	0.021	-0.108	-0.071	0.194	0.020	-0.067	-0.123	0.007	0.017	0.047	0.104	-0.034	-0.001
X0v	0.061	0.025	-0.074	-0.032	0.049	-0.014	0.015	-0.055	0.027	-0.018	-0.012	0.002	0.050	-0.031	0.029
X1v	0.058	0.005	-0.063	0.013	0.011	-0.006	0.166	-0.039	0.018	-0.054	-0.071	-0.015	0.039	0.055	0.042
X2v	0.044	0.065	-0.094	0.012	0.007	-0.066	0.105	-0.016	0.012	-0.050	-0.130	0.033	0.045	0.092	0.026
X0Av	-0.021	0.058	-0.132	-0.060	0.076	0.002	0.042	-0.173	0.001	0.014	0.060	-0.028	0.021	-0.083	0.045
X1Av	-0.016	0.024	-0.108	-0.027	0.073	0.048	0.241	-0.170	-0.053	0.018	0.002	-0.036	-0.003	-0.036	0.073
X2Av	-0.005	0.061	-0.113	-0.057	0.041	0.117	0.137	-0.123	-0.110	0.066	0.073	-0.006	-0.113	-0.025	0.022
X0sol	0.058	0.052	-0.059	-0.007	0.075	-0.005	-0.033	0.007	-0.014	0.002	-0.014	0.021	0.018	-0.005	-0.018
X1sol	0.064	0.012	-0.049	0.051	0.041	0.030	0.053	0.018	-0.014	-0.014	-0.032	-0.007	0.013	0.024	0.000

X2sol	0.039	0.092	-0.091	0.014	0.033	-0.046	-0.005	0.024	-0.022	-0.030	-0.087	0.013	0.082	0.046	0.000
XMOD	0.056	0.038	-0.062	0.060	0.063	0.026	0.086	0.033	-0.026	-0.016	-0.042	-0.018	0.013	0.036	0.001
RDCHI	0.065	-0.042	-0.016	0.037	-0.002	0.042	0.013	0.042	-0.008	-0.023	0.013	-0.011	0.024	0.012	-0.007
RDSQ	0.066	-0.008	-0.015	0.070	0.003	-0.003	0.005	-0.002	-0.011	0.002	-0.027	0.025	0.022	-0.018	0.015
ISIZ	0.056	-0.066	0.015	-0.067	-0.005	-0.042	0.056	-0.071	0.035	-0.020	0.008	0.037	-0.012	-0.014	0.039
IAC	0.055	-0.028	0.084	-0.005	0.004	-0.040	0.142	-0.043	0.001	-0.025	0.026	-0.006	-0.052	-0.018	0.046
AAC	-0.014	0.030	0.122	0.097	0.010	0.075	0.198	-0.011	0.012	0.029	0.097	-0.031	-0.187	-0.026	0.040
IDE	0.060	-0.026	0.024	-0.066	0.046	0.095	0.002	0.052	-0.011	-0.009	0.008	-0.007	-0.012	-0.002	-0.015
IDM	0.069	0.009	-0.011	-0.024	0.002	0.011	-0.002	0.028	-0.024	0.002	0.015	-0.009	-0.003	-0.004	-0.008
IDDE	0.050	0.020	0.078	0.008	0.083	0.069	-0.018	-0.045	0.022	0.209	-0.129	-0.073	0.038	0.006	-0.102
IDDM	0.069	0.004	-0.012	-0.016	0.006	0.009	-0.002	0.029	-0.016	0.000	0.014	-0.009	0.002	-0.005	-0.007
IDET	0.063	-0.034	0.006	0.054	0.056	0.037	-0.044	-0.021	-0.002	-0.041	-0.021	0.023	-0.006	-0.010	-0.007
IDMT	0.063	-0.026	0.002	0.078	0.048	0.027	-0.044	-0.036	-0.014	-0.045	-0.035	0.043	-0.008	-0.014	0.001
IVDE	0.033	0.063	0.103	-0.037	0.074	0.089	0.039	-0.014	-0.002	0.191	-0.101	-0.081	0.070	-0.031	-0.032
IVDM	0.069	-0.015	-0.010	-0.003	0.006	0.026	0.007	0.040	-0.012	-0.005	0.015	-0.009	0.009	0.001	-0.009
HVcpx	0.056	-0.062	0.017	-0.041	0.049	0.092	0.004	0.074	0.012	-0.014	0.020	0.000	-0.009	0.010	-0.028
HDcpx	0.065	0.025	-0.007	-0.061	-0.019	0.042	0.005	0.034	-0.057	0.006	0.024	-0.010	-0.013	-0.003	-0.009
Uindex	0.066	0.005	0.006	-0.056	0.058	0.025	-0.018	-0.007	0.023	-0.010	0.019	0.022	0.046	-0.040	-0.004
Vindex	-0.028	0.079	0.010	-0.085	-0.073	0.146	0.011	-0.107	-0.111	0.030	0.012	0.097	0.125	-0.042	-0.024
Xindex	-0.011	0.082	0.019	-0.133	-0.056	0.146	0.017	-0.055	-0.111	0.043	0.025	0.098	0.087	-0.052	-0.017
Yindex	0.043	-0.013	0.060	-0.055	0.051	-0.081	0.110	0.110	0.059	0.218	0.034	0.224	-0.098	-0.100	-0.045
IC0	-0.014	0.030	0.122	0.097	0.010	0.075	0.198	-0.011	0.012	0.029	0.097	-0.031	-0.187	-0.026	0.040
TIC0	0.055	-0.028	0.084	-0.005	0.004	-0.040	0.142	-0.043	0.001	-0.025	0.026	-0.006	-0.052	-0.018	0.046
SIC0	-0.047	0.044	0.034	0.096	0.028	0.111	0.106	-0.020	0.048	0.056	0.103	0.030	-0.160	-0.001	0.027
CIC0	0.054	-0.065	0.002	-0.086	-0.027	-0.070	-0.018	-0.031	-0.016	-0.021	-0.030	-0.015	0.060	-0.021	0.017
BIC0	-0.014	0.077	0.014	-0.022	-0.006	0.213	0.120	-0.016	-0.104	0.091	0.128	0.047	-0.110	-0.069	0.015
IC1	0.011	-0.012	0.158	0.097	0.035	0.024	-0.032	-0.005	-0.077	0.176	-0.008	-0.032	-0.015	0.017	0.012
TIC1	0.054	-0.032	0.099	0.018	0.032	-0.026	-0.012	-0.050	-0.050	0.104	-0.008	0.006	-0.011	-0.020	0.033
SIC1	-0.033	0.010	0.111	0.128	0.043	0.071	-0.030	-0.004	-0.016	0.169	0.019	0.015	-0.037	0.046	0.012
CIC1	0.045	-0.050	-0.056	-0.118	-0.047	-0.063	0.069	-0.033	0.035	-0.123	0.006	-0.006	0.008	-0.041	0.023
BIC1	0.002	0.060	0.079	-0.020	-0.008	0.206	0.029	0.000	-0.176	0.176	0.065	0.044	-0.003	-0.030	-0.003
IC2	0.024	-0.036	0.160	0.017	0.044	-0.010	0.017	-0.046	-0.017	0.115	-0.043	-0.092	0.161	0.107	-0.074
TIC2	0.051	-0.038	0.098	-0.026	0.063	-0.028	0.016	-0.079	0.007	0.080	-0.044	-0.042	0.076	0.059	-0.027
SIC2	-0.004	-0.035	0.169	0.051	0.034	0.015	0.020	-0.043	0.011	0.122	-0.033	-0.072	0.172	0.128	-0.079
CIC2	0.026	-0.019	-0.139	-0.079	-0.077	-0.038	0.031	0.015	0.006	-0.145	0.051	0.079	-0.186	-0.154	0.118
BIC2	0.018	0.005	0.147	-0.047	-0.007	0.123	0.057	-0.035	-0.111	0.136	0.006	-0.037	0.170	0.063	-0.080
ATS1m	0.053	0.059	-0.072	0.032	0.050	0.024	0.094	0.035	-0.045	0.014	-0.005	-0.032	-0.001	0.007	0.012
ATS2m	0.042	0.085	-0.092	-0.010	0.049	0.016	0.011	0.021	-0.073	0.039	0.051	-0.002	-0.003	-0.027	-0.007

ATS3m	0.062	-0.041	0.031	0.028	0.047	-0.004	0.055	0.081	0.028	0.088	0.006	0.069	-0.026	-0.004	-0.058
ATS1v	0.067	0.001	-0.044	-0.002	-0.019	0.032	0.005	-0.014	-0.015	0.012	0.001	0.014	-0.008	0.008	-0.009
ATS2v	0.066	0.018	-0.056	0.004	-0.014	-0.015	-0.019	-0.019	-0.010	0.021	-0.003	0.022	-0.022	0.009	-0.014
ATS3v	0.062	-0.057	-0.001	0.029	0.031	0.015	-0.014	-0.011	0.063	0.047	0.048	0.049	0.082	-0.105	0.032
ATS1e	0.068	0.012	-0.017	0.005	0.014	0.004	0.019	0.057	-0.036	0.001	0.025	-0.021	0.015	-0.012	-0.003
ATS2e	0.064	0.049	-0.025	-0.019	-0.002	-0.020	0.006	0.031	-0.050	0.011	0.027	-0.020	0.015	-0.023	0.004
ATS3e	0.062	-0.041	0.031	0.028	0.047	-0.004	0.055	0.081	0.028	0.087	0.006	0.069	-0.025	-0.004	-0.057
ATS1p	0.065	0.013	-0.059	0.007	-0.011	0.036	0.031	-0.022	-0.018	0.013	-0.014	0.008	-0.012	0.014	-0.004
ATS2p	0.062	0.030	-0.077	0.009	-0.001	-0.010	-0.013	-0.021	-0.015	0.024	-0.007	0.024	-0.022	0.013	-0.017
ATS3p	0.061	-0.058	-0.003	0.028	0.030	0.017	-0.022	-0.017	0.065	0.043	0.052	0.048	0.089	-0.116	0.042
MATS1m	0.041	-0.085	-0.014	-0.071	-0.116	0.038	-0.016	-0.016	0.053	0.035	-0.081	-0.068	0.042	0.044	-0.041
MATS2m	0.028	-0.056	-0.138	0.012	0.043	0.020	0.001	-0.060	0.099	0.075	0.115	-0.169	0.000	-0.049	0.008
MATS3m	0.014	-0.086	-0.059	-0.075	-0.052	0.050	0.031	0.243	0.037	0.107	-0.143	-0.132	-0.114	0.024	0.096
MATS1v	0.024	-0.005	-0.150	-0.064	-0.017	0.118	-0.004	-0.040	-0.019	0.120	0.021	-0.042	-0.019	-0.008	-0.056
MATS2v	0.023	-0.024	-0.160	0.018	0.062	0.027	0.015	-0.052	0.075	0.072	0.098	-0.149	0.002	-0.035	0.004
MATS3v	0.003	-0.091	-0.055	-0.101	-0.038	0.059	-0.088	0.216	0.011	0.121	-0.030	-0.072	-0.108	-0.028	0.048
MATS1e	0.045	-0.077	-0.012	-0.055	-0.118	0.032	0.031	-0.019	0.062	0.023	-0.119	-0.085	0.046	0.064	-0.027
MATS2e	0.029	-0.054	-0.138	0.017	0.043	0.017	0.016	-0.061	0.102	0.071	0.103	-0.175	0.001	-0.042	0.012
MATS3e	0.000	-0.090	-0.053	-0.105	-0.033	0.059	-0.115	0.203	0.004	0.121	-0.001	-0.055	-0.104	-0.041	0.035
MATS1p	0.038	-0.089	-0.015	-0.081	-0.112	0.043	-0.053	-0.014	0.046	0.043	-0.050	-0.053	0.039	0.027	-0.051
MATS2p	0.031	-0.051	-0.137	0.022	0.041	0.015	0.034	-0.062	0.105	0.066	0.088	-0.181	0.003	-0.035	0.018
MATS3p	0.011	-0.089	-0.059	-0.084	-0.049	0.053	-0.003	0.239	0.030	0.112	-0.113	-0.117	-0.114	0.009	0.084
GATS1m	-0.017	0.116	-0.040	0.049	0.103	-0.031	0.079	0.084	-0.035	-0.059	-0.031	-0.002	0.061	0.048	0.039
GATS2m	-0.010	0.046	0.156	-0.071	-0.078	-0.004	0.018	0.037	-0.069	-0.039	-0.039	0.101	0.043	-0.041	0.061
GATS1v	-0.018	0.016	0.161	0.034	0.026	-0.094	0.037	0.112	0.016	-0.105	0.019	-0.014	0.060	-0.004	0.085
GATS2v	-0.010	0.045	0.156	-0.072	-0.078	-0.004	0.015	0.037	-0.070	-0.038	-0.037	0.102	0.043	-0.042	0.060
GATS1e	-0.030	0.099	-0.048	0.003	0.119	-0.011	-0.069	0.097	-0.066	-0.023	0.092	0.054	0.051	-0.015	-0.004
GATS2e	-0.009	0.048	0.156	-0.067	-0.079	-0.006	0.029	0.036	-0.067	-0.042	-0.049	0.096	0.044	-0.035	0.064
GATS1p	-0.019	0.116	-0.042	0.045	0.106	-0.029	0.064	0.086	-0.039	-0.055	-0.018	0.004	0.061	0.041	0.035
GATS2p	-0.012	0.043	0.155	-0.077	-0.077	-0.001	-0.003	0.038	-0.073	-0.033	-0.023	0.108	0.041	-0.049	0.055
EPS0	0.066	-0.024	-0.003	0.008	-0.002	0.082	0.013	0.028	-0.038	-0.009	0.005	0.031	0.024	-0.003	-0.013
EPS1	0.068	-0.006	-0.020	0.033	-0.015	0.027	0.008	0.014	-0.029	-0.006	0.001	0.000	0.035	-0.009	0.007
EEig01x	0.056	0.078	0.008	0.003	-0.055	0.004	-0.016	-0.011	0.082	-0.008	-0.029	0.043	0.012	0.046	0.019
EEig02x	0.062	-0.044	0.022	0.046	0.008	0.037	0.023	0.063	0.088	0.024	0.000	-0.029	0.007	-0.054	-0.051
EEig01d	0.031	0.112	0.023	0.029	-0.046	0.027	0.062	-0.006	0.167	-0.086	0.010	-0.004	0.025	0.023	0.010
EEig02d	0.049	-0.007	0.045	0.073	0.014	0.067	0.098	0.129	0.199	-0.024	0.052	-0.071	-0.105	0.037	-0.116
EEig03d	0.038	-0.025	-0.045	0.144	-0.020	0.053	0.081	-0.025	0.098	-0.081	0.045	-0.059	0.360	-0.127	-0.160
EEig04d	-0.008	0.028	0.028	0.163	0.011	0.094	-0.107	-0.018	-0.140	-0.198	-0.108	-0.078	-0.074	-0.211	0.432

EEig01r	0.062	0.051	0.008	-0.038	-0.038	-0.061	-0.004	-0.011	-0.015	0.008	-0.023	0.008	-0.014	-0.041	-0.004
EEig02r	0.061	-0.057	0.029	0.029	0.012	-0.006	0.053	0.052	0.073	0.011	-0.005	-0.030	0.042	-0.027	0.004
EEig03r	0.049	-0.068	-0.033	0.097	-0.055	0.048	0.041	-0.044	0.026	-0.031	0.023	-0.057	0.100	0.000	-0.040
ESpm02u	0.067	0.032	-0.019	-0.026	-0.022	-0.009	0.000	0.017	-0.041	0.011	0.018	-0.011	-0.006	-0.009	-0.001
ESpm03u	0.041	0.096	0.009	0.015	0.013	-0.124	-0.045	-0.071	-0.010	0.077	-0.014	-0.069	-0.015	-0.032	0.000
ESpm04u	0.064	0.045	-0.023	-0.026	-0.009	-0.060	-0.008	0.015	-0.024	0.015	0.005	-0.005	-0.035	-0.013	0.006
ESpm05u	0.044	0.090	0.015	0.025	0.016	-0.114	-0.057	-0.068	-0.017	0.087	-0.005	-0.080	-0.014	-0.023	-0.021
ESpm06u	0.062	0.049	-0.024	-0.026	-0.005	-0.082	-0.011	0.011	-0.017	0.014	0.004	-0.011	-0.048	-0.012	0.008
ESpm07u	0.045	0.088	0.017	0.029	0.017	-0.110	-0.062	-0.066	-0.019	0.091	-0.002	-0.084	-0.011	-0.023	-0.024
ESpm08u	0.061	0.051	-0.023	-0.025	-0.004	-0.092	-0.012	0.008	-0.014	0.013	0.005	-0.019	-0.056	-0.009	0.008
ESpm09u	0.045	0.087	0.017	0.030	0.017	-0.108	-0.064	-0.065	-0.020	0.093	-0.001	-0.086	-0.010	-0.024	-0.025
ESpm10u	0.060	0.051	-0.022	-0.024	-0.003	-0.097	-0.014	0.007	-0.013	0.012	0.006	-0.024	-0.061	-0.007	0.007
ESpm11u	0.045	0.086	0.017	0.031	0.018	-0.108	-0.065	-0.065	-0.020	0.094	0.000	-0.087	-0.009	-0.024	-0.025
ESpm12u	0.060	0.052	-0.022	-0.023	-0.003	-0.101	-0.015	0.007	-0.012	0.012	0.007	-0.028	-0.064	-0.006	0.006
ESpm13u	0.046	0.086	0.017	0.031	0.018	-0.107	-0.066	-0.065	-0.020	0.095	0.000	-0.087	-0.008	-0.025	-0.025
ESpm14u	0.060	0.052	-0.021	-0.022	-0.002	-0.103	-0.016	0.006	-0.012	0.012	0.008	-0.030	-0.066	-0.005	0.005
ESpm15u	0.046	0.086	0.017	0.031	0.018	-0.107	-0.066	-0.064	-0.020	0.095	0.000	-0.087	-0.008	-0.025	-0.025
ESpm01x	0.066	0.007	0.004	0.037	-0.040	0.056	-0.026	0.025	0.055	-0.017	-0.009	-0.014	0.013	0.015	-0.011
ESpm02x	0.065	0.034	0.000	0.002	-0.054	0.043	-0.027	0.019	0.053	-0.017	0.006	0.021	0.011	0.056	-0.002
ESpm03x	0.061	0.054	0.003	-0.011	-0.062	0.040	-0.037	0.006	0.054	-0.012	0.004	0.028	0.006	0.063	0.004
ESpm04x	0.059	0.062	0.006	-0.019	-0.065	0.038	-0.034	0.003	0.054	-0.012	0.004	0.032	0.002	0.067	0.005
ESpm05x	0.058	0.067	0.008	-0.024	-0.066	0.038	-0.032	0.001	0.053	-0.011	0.003	0.034	0.000	0.068	0.005
ESpm06x	0.057	0.070	0.009	-0.027	-0.067	0.038	-0.030	0.001	0.051	-0.011	0.003	0.035	-0.001	0.069	0.006
ESpm07x	0.057	0.071	0.010	-0.029	-0.067	0.039	-0.028	0.001	0.050	-0.010	0.003	0.036	-0.002	0.069	0.006
ESpm08x	0.056	0.072	0.010	-0.030	-0.067	0.039	-0.027	0.001	0.049	-0.010	0.003	0.037	-0.002	0.068	0.006
ESpm09x	0.056	0.072	0.011	-0.031	-0.067	0.040	-0.026	0.001	0.048	-0.010	0.003	0.037	-0.001	0.068	0.006
ESpm10x	0.056	0.073	0.011	-0.032	-0.067	0.040	-0.025	0.001	0.047	-0.010	0.003	0.037	-0.001	0.068	0.006
ESpm11x	0.056	0.073	0.011	-0.032	-0.067	0.041	-0.025	0.001	0.047	-0.010	0.004	0.037	-0.001	0.068	0.006
ESpm12x	0.056	0.073	0.011	-0.032	-0.067	0.041	-0.025	0.001	0.046	-0.010	0.004	0.037	-0.001	0.067	0.006
ESpm13x	0.056	0.073	0.011	-0.033	-0.067	0.041	-0.024	0.001	0.046	-0.010	0.004	0.038	-0.001	0.067	0.006
ESpm14x	0.056	0.073	0.011	-0.033	-0.068	0.042	-0.024	0.001	0.045	-0.010	0.004	0.038	-0.001	0.067	0.006
ESpm15x	0.056	0.073	0.011	-0.033	-0.068	0.042	-0.024	0.002	0.045	-0.009	0.004	0.038	0.000	0.067	0.006
ESpm01d	-0.004	0.116	0.058	0.073	0.015	0.064	0.049	0.077	0.116	-0.106	0.059	-0.100	-0.021	-0.051	-0.088
ESpm02d	0.055	0.066	0.009	0.010	-0.060	0.063	0.022	0.018	0.122	-0.053	0.053	0.022	-0.014	0.064	-0.009
ESpm03d	0.016	0.129	0.059	0.045	-0.015	0.027	0.000	-0.008	0.085	0.017	0.023	-0.023	-0.015	-0.049	0.050
ESpm04d	0.042	0.098	0.012	0.004	-0.058	0.056	0.026	-0.002	0.147	-0.067	0.068	0.007	-0.020	0.035	-0.010
ESpm05d	0.020	0.127	0.057	0.043	-0.021	0.031	-0.012	-0.011	0.081	0.034	0.019	-0.018	-0.026	-0.048	0.053
ESpm06d	0.039	0.105	0.015	0.002	-0.056	0.053	0.028	-0.006	0.145	-0.063	0.074	0.006	-0.023	0.030	-0.005

ESpm07d	0.021	0.126	0.056	0.042	-0.022	0.034	-0.015	-0.011	0.079	0.039	0.018	-0.019	-0.030	-0.047	0.054
ESpm08d	0.038	0.107	0.017	0.001	-0.055	0.053	0.029	-0.006	0.141	-0.060	0.075	0.006	-0.024	0.027	-0.002
ESpm09d	0.021	0.126	0.056	0.042	-0.023	0.036	-0.016	-0.011	0.078	0.041	0.018	-0.019	-0.031	-0.047	0.055
ESpm10d	0.037	0.108	0.018	0.001	-0.055	0.053	0.029	-0.006	0.139	-0.058	0.075	0.007	-0.024	0.025	0.000
ESpm11d	0.022	0.126	0.056	0.041	-0.023	0.037	-0.017	-0.011	0.077	0.042	0.018	-0.019	-0.031	-0.047	0.055
ESpm12d	0.037	0.109	0.019	0.001	-0.055	0.053	0.029	-0.006	0.137	-0.056	0.075	0.007	-0.024	0.024	0.001
ESpm13d	0.022	0.126	0.055	0.041	-0.023	0.037	-0.017	-0.011	0.077	0.043	0.017	-0.019	-0.031	-0.048	0.055
ESpm14d	0.037	0.109	0.020	0.001	-0.054	0.053	0.028	-0.006	0.136	-0.055	0.075	0.007	-0.024	0.023	0.001
ESpm15d	0.022	0.126	0.055	0.041	-0.024	0.038	-0.017	-0.011	0.077	0.043	0.017	-0.019	-0.031	-0.048	0.055
ESpm01r	0.066	-0.031	0.031	-0.004	-0.048	-0.005	0.005	0.013	0.015	-0.030	-0.003	-0.033	0.005	-0.027	-0.014
ESpm02r	0.068	0.019	-0.006	-0.027	-0.035	-0.012	-0.008	0.021	-0.023	-0.014	0.000	-0.017	0.010	-0.007	-0.015
ESpm03r	0.065	0.032	0.004	-0.044	-0.047	-0.030	-0.020	0.014	-0.025	-0.020	-0.014	-0.017	0.008	-0.002	-0.020
ESpm04r	0.064	0.039	0.006	-0.051	-0.048	-0.033	-0.020	0.012	-0.029	-0.017	-0.014	-0.016	0.000	-0.003	-0.020
ESpm05r	0.063	0.042	0.008	-0.055	-0.049	-0.036	-0.022	0.013	-0.030	-0.017	-0.018	-0.015	-0.002	-0.001	-0.021
ESpm06r	0.062	0.044	0.009	-0.057	-0.049	-0.037	-0.021	0.013	-0.032	-0.016	-0.020	-0.014	-0.003	0.000	-0.021
ESpm07r	0.062	0.046	0.010	-0.058	-0.050	-0.037	-0.021	0.014	-0.032	-0.016	-0.021	-0.013	-0.003	0.000	-0.021
ESpm08r	0.062	0.046	0.011	-0.059	-0.050	-0.037	-0.021	0.014	-0.033	-0.016	-0.021	-0.012	-0.003	0.000	-0.021
ESpm09r	0.062	0.047	0.011	-0.059	-0.050	-0.037	-0.021	0.014	-0.033	-0.016	-0.022	-0.012	-0.002	0.001	-0.021
ESpm10r	0.062	0.047	0.011	-0.059	-0.050	-0.037	-0.021	0.014	-0.033	-0.016	-0.022	-0.012	-0.002	0.001	-0.021
ESpm11r	0.062	0.047	0.012	-0.059	-0.050	-0.037	-0.021	0.015	-0.034	-0.016	-0.022	-0.012	-0.002	0.001	-0.021
ESpm12r	0.062	0.047	0.012	-0.060	-0.050	-0.037	-0.021	0.015	-0.034	-0.016	-0.022	-0.012	-0.002	0.001	-0.021
ESpm13r	0.061	0.047	0.012	-0.060	-0.051	-0.037	-0.021	0.015	-0.034	-0.016	-0.022	-0.012	-0.002	0.001	-0.021
ESpm14r	0.061	0.047	0.012	-0.060	-0.051	-0.037	-0.021	0.015	-0.034	-0.016	-0.022	-0.011	-0.002	0.001	-0.021
ESpm15r	0.061	0.047	0.012	-0.060	-0.051	-0.037	-0.021	0.015	-0.034	-0.016	-0.022	-0.011	-0.002	0.001	-0.021
BEHm1	0.041	0.074	-0.080	0.064	0.018	0.033	0.138	-0.015	-0.024	0.017	-0.089	-0.039	-0.016	0.039	0.028
BEHm2	0.055	0.030	-0.055	-0.039	0.076	0.097	0.074	-0.016	-0.034	0.037	0.077	-0.041	-0.024	-0.076	0.028
BEHm3	0.058	0.018	-0.067	-0.060	0.011	-0.033	-0.009	0.011	-0.138	-0.002	0.030	-0.101	-0.009	-0.057	-0.066
BEHm4	0.055	0.022	-0.055	-0.028	0.086	-0.051	-0.094	0.088	-0.004	0.027	-0.047	0.077	0.089	0.138	0.129
BEHm5	0.059	-0.046	0.005	-0.018	0.028	-0.042	0.084	-0.066	0.083	0.018	0.071	0.157	0.057	-0.025	-0.113
BEHm6	0.053	-0.060	-0.044	0.031	-0.025	0.012	0.120	-0.060	0.043	-0.101	-0.158	0.028	-0.038	0.068	-0.003
BEHm7	0.044	-0.031	0.023	0.109	0.022	0.054	-0.164	-0.075	-0.133	-0.046	0.016	0.073	-0.174	0.018	-0.086
BEHm8	0.052	-0.035	0.033	0.067	0.048	0.013	-0.080	-0.097	-0.123	-0.083	0.106	0.120	-0.105	0.245	0.006
BELm1	0.029	-0.094	0.099	-0.003	-0.085	0.003	0.052	-0.076	0.009	0.036	0.031	-0.027	-0.069	-0.062	0.044
BELm2	0.051	-0.053	0.078	-0.056	-0.047	-0.052	0.040	-0.050	0.017	-0.054	-0.019	-0.145	0.025	-0.071	0.016
BELm3	0.056	-0.070	0.008	-0.059	-0.009	-0.036	0.029	-0.048	-0.012	0.039	0.066	0.063	0.100	0.009	0.153
BELm4	0.059	-0.062	0.003	-0.029	0.024	-0.016	0.018	-0.031	0.040	0.092	-0.066	0.091	-0.099	0.027	0.197
BEHv1	0.064	-0.026	0.017	0.022	-0.082	0.010	0.028	-0.040	-0.009	0.012	-0.089	0.000	-0.022	0.032	0.005
BEHv2	0.055	-0.050	0.075	-0.047	-0.052	-0.003	0.018	-0.033	0.051	-0.053	0.003	-0.078	0.041	-0.013	0.033

BEHv3	0.064	-0.040	-0.002	-0.034	-0.028	-0.034	0.040	-0.043	-0.053	0.001	0.026	-0.071	0.001	-0.082	0.002
BEHv4	0.064	-0.036	-0.005	-0.017	0.057	-0.017	-0.037	0.026	0.026	0.065	-0.009	0.034	0.020	0.042	0.231
BEHv5	0.056	-0.057	-0.007	-0.036	-0.017	-0.048	0.099	-0.072	0.037	0.044	-0.038	0.150	-0.032	0.006	-0.161
BEHv6	0.051	-0.090	-0.033	-0.028	-0.030	0.025	0.003	-0.041	-0.023	-0.082	-0.063	0.027	-0.042	-0.015	-0.061
BEHv7	0.056	-0.032	0.035	0.057	-0.002	0.005	-0.106	-0.111	-0.088	-0.048	0.094	-0.045	-0.192	0.054	-0.123
BEHv8	0.059	-0.039	0.030	0.004	0.023	-0.036	0.001	-0.095	-0.040	0.035	0.146	-0.017	0.024	0.342	0.120
BELv1	0.046	-0.076	0.081	0.001	-0.084	-0.024	0.045	-0.036	-0.006	0.013	-0.025	-0.028	-0.030	-0.028	0.034
BELv2	0.051	-0.049	0.087	-0.036	-0.042	-0.043	0.071	-0.023	0.033	-0.069	-0.063	-0.111	0.027	-0.027	0.031
BELv3	0.056	-0.065	0.019	-0.068	-0.026	-0.046	0.026	-0.046	-0.030	0.027	0.089	-0.006	0.015	0.020	0.062
BELv4	0.061	-0.050	0.021	-0.017	0.053	-0.013	0.024	-0.022	0.068	0.081	-0.029	0.073	-0.026	-0.034	0.132
BEHe1	0.062	-0.033	0.041	0.027	-0.076	-0.017	0.060	-0.018	-0.003	-0.003	-0.089	-0.033	-0.006	0.016	0.026
BEHe2	0.049	-0.059	0.094	-0.041	-0.050	-0.026	0.025	-0.006	0.045	-0.064	-0.038	-0.069	0.038	-0.012	0.028
BEHe3	0.061	-0.045	0.023	-0.052	-0.037	-0.045	0.051	-0.029	-0.072	0.001	0.030	-0.091	-0.014	-0.054	0.001
BEHe4	0.063	-0.041	0.017	-0.023	0.057	-0.023	-0.014	0.025	0.022	0.087	0.012	0.048	-0.019	0.055	0.211
BEHe5	0.056	-0.058	0.035	-0.048	-0.018	-0.065	0.079	-0.047	-0.037	0.011	-0.027	0.059	-0.011	-0.099	-0.118
BEHe6	0.053	-0.071	0.021	-0.047	-0.048	-0.006	0.058	-0.003	-0.101	-0.081	-0.087	-0.097	-0.006	-0.072	-0.064
BEHe7	0.058	-0.034	0.055	-0.021	-0.032	-0.042	-0.013	-0.088	-0.080	-0.049	0.102	-0.141	-0.150	0.004	-0.113
BEHe8	0.058	-0.039	0.033	-0.033	0.006	-0.056	0.041	-0.050	-0.010	0.080	0.136	-0.083	0.009	0.310	0.193
BELe1	0.055	-0.064	0.050	0.001	-0.091	0.002	-0.001	-0.050	-0.013	0.024	-0.036	0.004	-0.032	-0.008	0.007
BELe2	0.060	-0.033	0.050	-0.032	-0.023	-0.022	0.084	-0.063	0.055	-0.057	-0.014	-0.136	0.024	-0.025	0.045
BELe3	0.054	-0.077	-0.036	-0.024	-0.013	-0.009	-0.022	-0.071	0.058	0.030	0.095	0.122	0.050	-0.043	0.085
BEHp1	0.064	-0.012	0.003	0.033	-0.073	0.008	0.065	-0.043	-0.011	0.007	-0.116	-0.011	-0.019	0.045	0.012
BEHp2	0.058	-0.043	0.066	-0.048	-0.046	0.001	0.031	-0.041	0.047	-0.050	0.007	-0.089	0.038	-0.018	0.035
BEHp3	0.065	-0.036	-0.012	-0.034	-0.022	-0.033	0.039	-0.044	-0.052	0.002	0.027	-0.063	0.007	-0.079	0.006
BEHp4	0.064	-0.031	-0.014	-0.018	0.060	-0.020	-0.044	0.030	0.027	0.058	-0.020	0.038	0.032	0.046	0.228
BEHp5	0.056	-0.058	-0.005	-0.039	-0.024	-0.049	0.101	-0.070	0.021	0.043	-0.057	0.138	-0.045	0.004	-0.165
BEHp6	0.050	-0.092	-0.025	-0.040	-0.032	0.023	-0.013	-0.034	-0.042	-0.078	-0.047	0.015	-0.038	-0.035	-0.071
BEHp7	0.058	-0.032	0.037	0.041	-0.009	-0.008	-0.087	-0.117	-0.074	-0.047	0.112	-0.074	-0.190	0.062	-0.130
BEHp8	0.059	-0.039	0.029	-0.005	0.019	-0.042	0.012	-0.093	-0.027	0.050	0.149	-0.036	0.042	0.348	0.133
BELp1	0.041	-0.084	0.088	-0.009	-0.084	-0.019	0.007	-0.031	-0.006	0.021	0.009	-0.017	-0.036	-0.046	0.026
BELp2	0.052	-0.048	0.087	-0.036	-0.040	-0.037	0.072	-0.018	0.034	-0.070	-0.065	-0.103	0.027	-0.021	0.032
BELp3	0.056	-0.065	0.017	-0.069	-0.027	-0.046	0.016	-0.048	-0.026	0.028	0.105	-0.011	-0.002	0.028	0.050
BELp4	0.061	-0.049	0.020	-0.012	0.059	-0.009	0.014	-0.033	0.079	0.082	-0.029	0.084	-0.010	-0.072	0.097
LP1	0.066	0.035	-0.020	-0.016	-0.025	-0.017	0.016	0.019	-0.038	0.021	0.007	0.005	-0.002	-0.013	0.004
Eig1Z	0.061	-0.056	0.011	-0.044	0.045	0.007	-0.038	-0.030	0.017	-0.036	0.040	0.053	-0.040	-0.005	-0.033
Eig1m	0.061	-0.057	0.011	-0.044	0.045	0.007	-0.038	-0.030	0.017	-0.036	0.040	0.052	-0.041	-0.005	-0.032
Eig1v	0.060	-0.015	0.037	-0.035	0.104	-0.013	-0.005	0.055	-0.007	-0.064	0.120	-0.007	0.076	-0.004	0.022
Eig1e	0.065	-0.034	-0.010	-0.031	0.057	0.001	-0.001	-0.023	0.019	-0.053	-0.005	0.043	-0.023	0.027	-0.028

Eig1p	0.056	-0.023	0.052	-0.046	0.106	-0.010	-0.035	0.064	-0.014	-0.054	0.168	-0.005	0.081	-0.028	0.023
SEigZ	-0.006	0.110	-0.095	0.019	0.099	-0.006	0.039	0.077	-0.066	-0.014	-0.032	0.006	0.032	0.065	-0.044
SEigm	-0.006	0.109	-0.097	0.019	0.099	-0.005	0.037	0.076	-0.065	-0.014	-0.031	0.007	0.032	0.065	-0.044
SEigv	-0.005	-0.027	-0.149	-0.016	-0.054	0.092	-0.112	-0.163	0.077	0.038	-0.018	0.066	0.045	-0.024	0.049
SEige	-0.006	0.112	-0.007	0.016	0.123	-0.050	0.061	0.164	-0.107	-0.027	0.009	-0.013	0.003	0.065	-0.078
SEigp	-0.005	-0.004	-0.164	-0.010	-0.032	0.086	-0.093	-0.144	0.060	0.034	-0.029	0.060	0.050	-0.014	0.040
AEigZ	0.060	-0.062	0.019	-0.041	0.035	0.006	-0.029	-0.036	0.024	-0.036	0.033	0.046	-0.041	-0.006	-0.025
AEigm	0.060	-0.063	0.020	-0.041	0.034	0.006	-0.029	-0.036	0.024	-0.036	0.033	0.046	-0.041	-0.006	-0.024
AEigv	0.059	-0.012	0.048	-0.033	0.105	-0.020	0.005	0.067	-0.013	-0.065	0.118	-0.012	0.070	-0.002	0.017
AEige	0.065	-0.037	-0.010	-0.031	0.053	0.002	-0.003	-0.028	0.022	-0.052	-0.006	0.044	-0.023	0.025	-0.026
AEigp	0.054	-0.022	0.067	-0.043	0.105	-0.019	-0.024	0.076	-0.020	-0.055	0.163	-0.011	0.073	-0.025	0.018
VEA1	0.069	-0.009	-0.017	0.010	-0.001	0.014	-0.001	0.029	-0.015	-0.002	0.009	-0.009	0.013	-0.003	-0.004
VEA2	-0.068	-0.017	0.005	0.046	-0.008	-0.013	0.004	-0.029	0.027	-0.001	-0.017	0.014	0.015	0.002	0.012
VRA1	0.068	-0.014	-0.017	0.047	-0.001	0.008	0.004	0.015	-0.013	-0.004	-0.006	0.006	0.020	-0.007	0.004
VRA2	0.067	-0.008	-0.027	0.017	-0.043	0.023	0.032	0.049	-0.039	0.005	0.024	-0.011	0.022	0.004	0.002
VED1	0.069	-0.003	-0.013	0.004	0.011	0.005	-0.004	0.021	-0.009	-0.002	0.007	-0.001	0.011	-0.009	-0.004
VED2	-0.068	-0.013	0.008	0.041	-0.001	-0.019	0.000	-0.033	0.030	-0.001	-0.017	0.016	0.013	-0.001	0.011
VRD1	0.068	-0.011	-0.017	0.044	0.000	0.006	0.003	0.013	-0.013	-0.003	-0.007	0.007	0.018	-0.008	0.003
VRD2	0.067	0.000	-0.026	0.010	-0.043	0.020	0.028	0.045	-0.044	0.006	0.022	-0.013	0.014	0.004	0.002
VEZ1	0.070	-0.005	-0.009	0.001	0.008	0.000	-0.007	0.024	-0.006	-0.008	0.002	0.002	0.015	-0.001	-0.006
VEZ2	-0.068	-0.015	0.012	0.039	-0.006	-0.027	-0.004	-0.029	0.035	-0.010	-0.026	0.017	0.021	0.011	0.009
VRZ1	0.068	-0.011	-0.018	0.044	0.001	0.007	0.002	0.013	-0.014	-0.001	-0.005	0.007	0.018	-0.009	0.003
VRZ2	0.067	0.001	-0.029	0.010	-0.041	0.025	0.027	0.043	-0.048	0.012	0.030	-0.012	0.012	-0.003	0.002
VEm1	0.070	-0.005	-0.009	0.001	0.008	0.000	-0.008	0.024	-0.006	-0.008	0.002	0.002	0.016	-0.001	-0.006
VEm2	-0.068	-0.015	0.013	0.039	-0.006	-0.027	-0.004	-0.029	0.035	-0.010	-0.026	0.017	0.021	0.011	0.010
VRm1	0.068	-0.011	-0.018	0.044	0.001	0.007	0.002	0.013	-0.014	-0.001	-0.005	0.007	0.018	-0.009	0.003
VRm2	0.067	0.001	-0.030	0.010	-0.040	0.025	0.027	0.043	-0.048	0.013	0.030	-0.012	0.012	-0.003	0.002
VEv1	0.069	-0.002	-0.012	0.000	0.011	0.005	0.001	0.023	-0.007	-0.003	0.008	-0.007	0.019	-0.011	0.002
VEv2	-0.068	-0.012	0.009	0.038	-0.002	-0.018	0.004	-0.033	0.029	-0.002	-0.016	0.013	0.018	-0.004	0.013
VRv1	0.068	-0.011	-0.017	0.045	0.000	0.006	0.002	0.013	-0.013	-0.002	-0.007	0.008	0.017	-0.008	0.002
VRv2	0.067	-0.001	-0.027	0.011	-0.042	0.019	0.026	0.044	-0.044	0.008	0.022	-0.011	0.010	0.004	-0.001
VEe1	0.069	-0.003	-0.013	0.003	0.011	0.004	-0.001	0.021	-0.008	-0.004	0.005	0.000	0.010	-0.004	-0.004
VEe2	-0.068	-0.013	0.008	0.041	-0.002	-0.021	0.003	-0.033	0.032	-0.003	-0.018	0.016	0.012	0.005	0.012
VRe1	0.068	-0.011	-0.017	0.044	0.000	0.006	0.002	0.013	-0.014	-0.002	-0.006	0.007	0.019	-0.009	0.003
VRe2	0.067	0.000	-0.026	0.010	-0.043	0.021	0.026	0.044	-0.045	0.008	0.024	-0.012	0.016	0.001	0.001
VEp1	0.069	-0.002	-0.012	-0.001	0.010	0.004	0.001	0.023	-0.007	-0.003	0.008	-0.008	0.020	-0.011	0.002
VEp2	-0.068	-0.012	0.009	0.037	-0.003	-0.019	0.004	-0.032	0.029	-0.003	-0.018	0.012	0.021	-0.002	0.015
VRp1	0.068	-0.011	-0.017	0.045	0.001	0.006	0.002	0.013	-0.014	-0.002	-0.007	0.008	0.016	-0.008	0.002

VRp2	0.067	-0.001	-0.027	0.011	-0.041	0.020	0.026	0.044	-0.044	0.010	0.024	-0.010	0.009	0.004	-0.001

Table B.3: 2D PCA factor matrix

Solvent	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
Acetone	-5.354	5.241	2.383	-1.530	-5.179	-0.663	-1.398	-1.083	0.365	-1.767	0.135	-1.801	-1.449	-1.370	-1.019
Aceto-nitrile	-14.975	4.300	4.375	1.291	-4.810	5.441	-1.343	0.084	4.023	-0.564	-0.118	1.642	0.297	0.985	0.140
Benzyl Alcohol	22.921	-3.603	2.826	7.480	2.902	2.516	-3.438	-1.138	-2.602	-1.730	0.210	1.468	-0.694	0.152	-0.209
Carbon Tetra-chloride	-1.022	15.459	-12.358	-1.313	2.572	-1.857	-3.055	1.635	-0.365	-0.784	-0.721	0.369	0.519	0.291	-0.146
Cyclo-hexane	9.083	-10.587	-8.743	-1.040	-6.235	-1.391	2.206	1.703	-0.372	-1.151	2.343	1.012	0.421	0.040	-0.141
Ethanol	-16.434	-2.713	3.268	-1.235	-1.287	1.218	0.725	-1.154	-3.389	-0.349	-1.684	0.484	2.310	-0.678	-0.475
Ethyl Acetate	8.869	2.445	6.579	-1.443	5.117	0.694	-0.464	1.689	1.638	-1.145	2.537	-1.634	1.450	-0.548	0.447
Ethylene Glycol	-8.069	-2.052	3.874	-2.484	0.818	1.006	0.898	5.698	-1.570	0.154	-1.264	0.734	-1.307	-0.538	0.796
Hexane	8.723	-11.232	-3.772	-7.768	4.433	2.341	0.560	-1.775	2.051	-1.413	-1.906	-0.707	-0.387	0.492	-0.158
Iso-propanol	-5.287	1.845	2.259	-2.928	-3.191	-2.414	-1.156	-2.463	-2.803	-0.199	0.542	-1.402	-0.354	1.331	1.447
Methanol	-28.703	-9.061	-0.939	7.175	3.268	-6.569	-0.820	-0.147	1.894	-0.214	-0.168	-0.275	-0.143	0.055	0.012
Methylene Dichloride	-18.282	2.691	-6.093	1.613	3.698	6.106	3.071	-2.132	-1.061	1.853	1.705	-0.114	-0.783	-0.464	0.105
Propylene Glycol	2.151	1.102	5.408	-2.650	1.458	-1.907	0.584	1.351	-0.954	2.405	0.813	-0.100	-0.263	1.454	-1.641
Sulfolane	18.294	7.869	1.121	5.876	-0.835	-1.598	6.776	-0.256	0.781	-0.672	-1.564	-0.661	0.083	0.286	0.149
t-Amyl Alcohol	11.388	3.182	2.942	-4.438	0.637	-4.712	-0.045	-2.711	1.411	1.861	0.227	2.734	-0.185	-1.053	0.421
Toluene	16.696	-4.885	-3.130	3.395	-3.366	1.789	-3.101	0.701	0.952	3.714	-1.086	-1.749	0.485	-0.436	0.270

Table B.4 below shows the QSAR models developed using 2D descriptors and containing all 14 factors. Because of the way PCA factors are calculated, F1 contains the most information from the descriptor matrix and F14 contains the least. In the Q^2 analysis where fewer factors are included, F14 is removed first and then F13 and then so on.

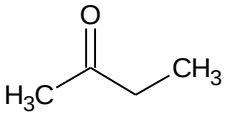
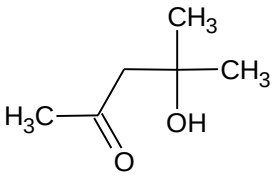
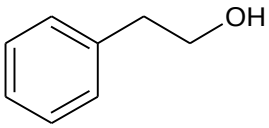
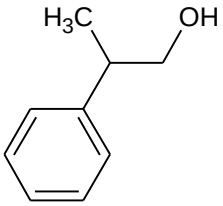
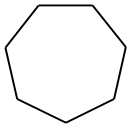
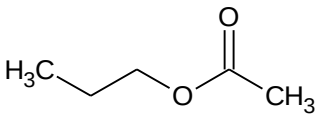
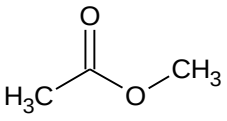
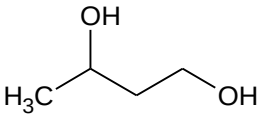
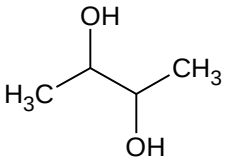
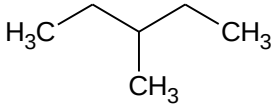
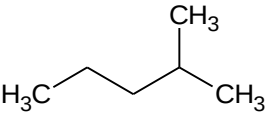
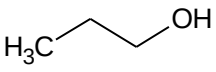
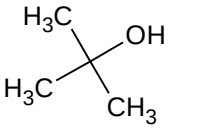
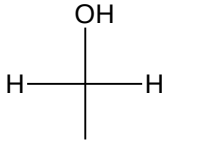
Table B.4: Equations developed with PCA method regression

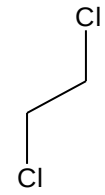
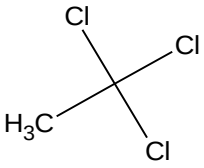
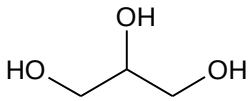
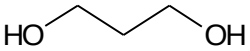
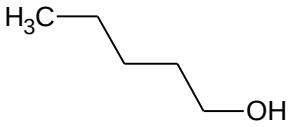
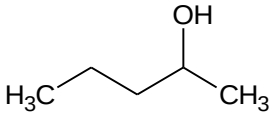
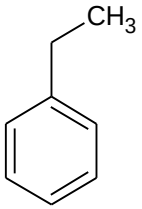
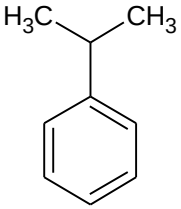
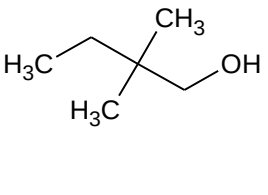
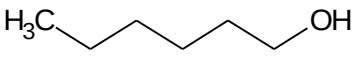
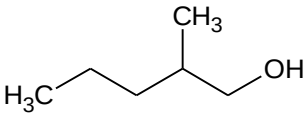
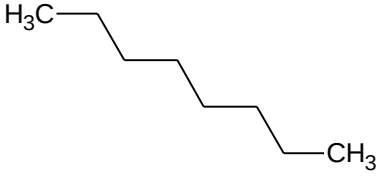
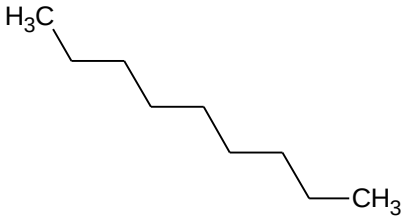
Descriptor Class	Equation to calculate normalized AR
2D	$0.033 \times (F1) + -0.0218 \times (F2) + -0.0916 \times (F3) + -0.1045 \times (F4) + -0.0119 \times (F5) + 0.045 \times (F6) + 0.0223 \times (F7) + -0.0268 \times (F8) + 0.1542 \times (F9) + -0.1187 \times (F10) + -0.0186 \times (F11) + -0.2495 \times (F12) + 0.1519 \times (F13) + -0.0151 \times (F14)$
3D GAFF	$0.034 \times (F1) + 0.0621 \times (F2) + -0.0295 \times (F3) + -0.0128 \times (F4) + 0.0604 \times (F5) + 0.0279 \times (F6) + 0.0291 \times (F7) + -0.0305 \times (F8) + 0.0294 \times (F9) + 0.0957 \times (F10) + 0.0129 \times (F11) + -0.0661 \times (F12) + 0.1925 \times (F13) + -0.016 \times (F14)$
2D & 3D GAFF	$0.026 \times (F1) + 0.0182 \times (F2) + 0.0657 \times (F3) + -0.0068 \times (F4) + 0.0284 \times (F5) + -0.0262 \times (F6) + 0.0599 \times (F7) + -0.0346 \times (F8) + 0.0207 \times (F9) + -0.0166 \times (F10) + 0.0475 \times (F11) + -0.0491 \times (F12) + 0.1127 \times (F13) + -0.0589 \times (F14)$
3D Ghemical	$0.038 \times (F1) + 0.0644 \times (F2) + -0.0131 \times (F3) + 0.041 \times (F4) + -0.0083 \times (F5) + 0.0072 \times (F6) + 0.0563 \times (F7) + -0.0617 \times (F8) + -0.0656 \times (F9) + 0.0505 \times (F10) + -0.0184 \times (F11) + 0.0465 \times (F12) + 0.1139 \times (F13) + -0.0956 \times (F14)$
2D & 3D Ghemical	$0.027 \times (F1) + 0.0131 \times (F2) + 0.064 \times (F3) + 0.0176 \times (F4) + -0.0163 \times (F5) + 0.0016 \times (F6) + -0.0655 \times (F7) + -0.0418 \times (F8) + 0.0197 \times (F9) + 0.0371 \times (F10) + 0.0379 \times (F11) + 0.0912 \times (F12) + 0.0317 \times (F13) + -0.0257 \times (F14)$
3D MMFF94s	$0.039 \times (F1) + 0.0599 \times (F2) + -0.0285 \times (F3) + 0.0416 \times (F4) + 0.0034 \times (F5) + 0.0219 \times (F6) + -0.0095 \times (F7) + -0.1013 \times (F8) + -0.0085 \times (F9) + 0.0589 \times (F10) + -0.0088 \times (F11) + 0 \times (F12) + -0.0339 \times (F13) + -0.1229 \times (F14)$
2D & 3D MMFF94s	$0.027 \times (F1) + 0.0087 \times (F2) + -0.0664 \times (F3) + 0.016 \times (F4) + -0.0141 \times (F5) + 0.0096 \times (F6) + 0.0782 \times (F7) + -0.0115 \times (F8) + -0.0274 \times (F9) + -0.0365 \times (F10) + -0.0503 \times (F11) + 0.0474 \times (F12) + -0.0421 \times (F13) + -0.0501 \times (F14)$

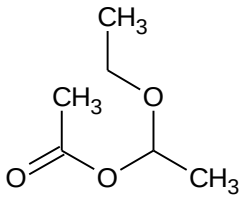
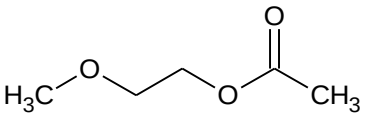
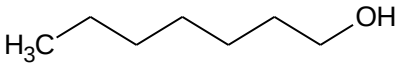
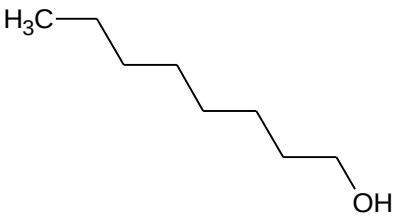
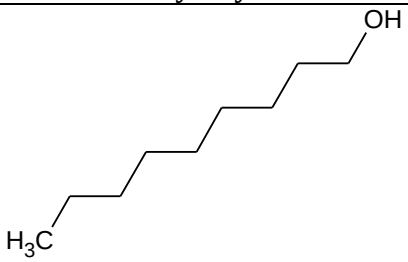
This procedure is repeated for each of the other sets of descriptors. The size of the matrices changes as more descriptors are added but the general procedure remains the same. Also, when the expansion solvents are added to the training set, the same procedure is repeated. Again, the additional solvents increase the size of the matrices, but the process remains the same.

Table B.5 shows the 35 expansion solvents and their two-dimensional structures. As in Table 4.4, the order within the table is the order in which they were added to the training set.

Table B.5: Expansion solvents and their two-dimensional structures

		
butanone	**diacetone alcohol**	**phenethyl alcohol**
		
2-phenylpropanol	**cycloheptane**	**cyclopentane**
		
propyl acetate	**methyl acetate**	**1,3-butanediol**
		
2,3-butanediol	**3-methylpentane**	**2-methylpentane**
		
propanol	***t*-butanol**	**methanediol**

		
1,2-Dichloroethane	**1,1,1-trichloroethane**	**glycerol**
		
1,3-propanediol	**1-pentanol**	**2-pentanol**
		
ethylbenzene	**cumene**	**2,2-dimethyl-1-butanol**
		
1-hexanol	**2-methyl-1-pentanol**	***n*-heptane**
		
***n*-octane**	***n*-nonane**	***n*-decane**

 **1-ethoxyethyl acetate**	 **2-Methoxyethyl acetate**	 **1-heptanol**
 **1-octanol**	 **1-nonanol**	